INTERNATIONAL DOCTORAL SCHOOL OF THE USC

Fabián
Suárez Lestón

PhD Thesis

Structure and activity of peptides at the water–lipid membrane interface. Molecular Dynamics insights

Santiago de Compostela, 2025

Doctoral Programme in Materials Science

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

DOCTORAL THESIS

STRUCTURE AND ACTIVITY OF PEPTIDES AT THE WATER–LIPID MEMBRANE INTERFACE. MOLECULAR DYNAMICS INSIGHTS

Author

Fabián Suárez Lestón

Supervisor/s: Rebeca García Fandiño, Ángel Piñeiro Guillén

Tutor: Ángel Piñeiro Guillén

PHD PROGRAMME IN MATERIALS SCIENCE

SANTIAGO DE COMPOSTELA – 2025

Contents

List of abbreviations . . . vi
List of figures . . . ix
List of tables . . . xv
List of publications . . . xix
Resumo . . . xx

1 Introduction **1**
1.1 Lipid membranes. The scaffold of life . . . 1
1.2 Antimicrobial peptides . . . 7
1.3 Experimental study of the interaction between antimicrobial peptides and membranes . . . 11
1.4 Simulation approaches to study peptide–membrane interactions . . . 13
1.5 Rational and data-driven design of antimicrobial peptides . . . 16
1.6 Integrating Molecular Dynamics data into predictive modeling . . . 18

2 Objectives and motivation **20**

3 Methods **23**
3.1 Molecular Dynamics simulations . . . 23
3.1.1 The method . . . 24
3.1.2 Force fields . . . 29
3.1.3 Molecular Dynamics beyond the *NVE* ensemble . . . 41
3.2 Simulation protocol . . . 46
3.2.1 Force field-specific parameters . . . 46
3.2.2 Atomistic simulations . . . 47
3.2.3 Coarse-grained simulations . . . 53
3.3 Processing and analysis of MD simulation data . . . 58
3.3.1 Transformation of coordinates . . . 58
3.3.2 General analyses . . . 62
3.3.3 Membrane-specific analyses . . . 64
3.3.4 Peptide-specific analyses . . . 67
3.3.5 Analyses associated with the peptide–lipid interaction . . . 73

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.4 Statistical methods . . . 77
3.4.1 General statistical concepts . . . 78
3.4.2 Statistical hypothesis tests . . . 82
3.4.3 Classification models . . . 91

4 Results and discussion **95**
4.1 Structural characterization of membrane-targeting antimicrobial peptides at the lipid membrane–water interface . . . 96
4.1.1 Peptide dynamics at the membrane–water interface . . . 98
4.1.2 Embedding of peptides into the bilayer . . . 108
4.1.3 Dynamical behavior of the peptide at the interface . . . 117
4.1.4 Performance and stability of the GROMOS-HIE and Reaction Field Schemes . . . 120
4.1.5 Long-timescale effects in the structure . . . 123
4.1.6 Conclusions . . . 125
4.2 Systematic simulation of peptide and membrane systems and database for the study of membrane-targeting antimicrobial peptides 127
4.2.1 Development of an automated framework for systematic peptide–membrane simulations . . . 128
4.2.2 Analysis of the interaction between peptides and lipid bilayers 130
4.2.3 The SuPepMem Database . . . 132
4.2.4 Conclusions . . . 134
4.3 Connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches . . . 138
4.3.1 Selection and statistical analysis of the MD-derived descriptors 139
4.3.2 Identification of sequence-induced bias . . . 154
4.3.3 Classification of the antimicrobial activity of the sequences . 160
4.3.4 Conclusions . . . 166

5 Conclusions **168**

Appendices

A Mathematical derivations **174**
A.1 Moments of inertia of an ellipsoid . . . 175
A.2 Translational and rotational diffusion of particles . . . 177
A.2.1 Translational diffusion . . . 177
A.2.2 Rotational diffusion . . . 180

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B Supplementary data for the structural characterization of membrane-targeting antimicrobial peptides at the lipid membrane–water interface **183**
B.1 Correlation between helicity metrics 184
B.2 Simulations starting from the shallow configuration 192
B.3 Simulations starting from the embedded configuration 217
B.4 Performance and stability of the GROMOS-HIE and Reaction Field Schemes 220
B.5 Long timescale simulations 227

C Relation of sequences simulated for the study of the antimicrobial activity of peptides **229**

D Supplementary data for the study of the connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches **232**

E Rights and permissions **264**
E.1 Rights for reproducing content from articles by the author 265
E.2 Rights for reproduced figures from third-party articles 266

Bibliography **267**

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Conflict of interests

The author declares having a contractual relationship with the company MD.USE Innovations S.L. The SuPepMem platform, a product of the research conducted for this project, was registered as intellectual property and transferred to MD.USE Innovations S.L.

The author declares financial support for his contract by the Xunta de Galicia and the Axencia Galega de Innovación (GAIN) under the Industrial PhD program (ref. 02_IN606D_2022_2667887).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Agradecementos

Wow… how long has it been since I got to make
music with everyone around a campfire?
I'm really happy we're all here.
Esker, Outer Wilds

As palabras coas que comeza este traballo son as últimas que escribo. E xa van unhas cantas. Mais non convén deixar pasa-la oportunidade de da-las grazas.

Son moitas persoas ás que debera facer referencia, e por moi diversos motivos. Porén, e apelando á indulxencia de quen non apareza, vou a adicarlle este espazo a todas aquelas que fixeron que isto que agora mesmo estás a ler poidas estalo lendo.

E as primeiras persoas ás que se lle pode atribuir algo son os meus directores, responsables non só de que eu comezase o doutoramento, senón tamén de que o rematase. Se hoxe hai un libro nas túas mans—ou un PDF na túa pantalla—é por eles.

Estes anos non foron para min un camiño de rosas. En absoluto. Pero pese a iso non podería ter escollido mellor compañía para percorrelo que o conxunto de entes que, nalgún momento ou outro, pasaron polo P1L11. Tanto os que alí "traballaban" coma os que cruzaban o corredor ou baixaban as escaleiras para ir tomar un café, un vermú, xogar ao tute ou, por que non, facer cencia.

Agradecer tamén a infinita paciencia das pobres estudantes que tiveron que lidiar coa miña presenza na clase e que—sorprendentemente—seguen a saudarme, máis aos profesores que o permitiron.

A quen me acolleu nun norte máis cálido e seco do esperado.

Aos *nakamas*, unidos por algo máis intenso que a interacción nuclear forte—o trauma da Física.

Á familia. Á oficial máis á efectiva. A vellas e novo.

A Pai e a Nai, a Pepe e a Manolita. Pobre homenaxe é esta, a de serdes as únicas persoas con nome propio, para quen me criou e me quixo, sen pedir máis que fose bo—e que me afeite.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of abbreviations

A . . . . . Alanine

AA . . . . . All atoms

AMBER . . . Assisted Model Builder with Energy Refinement

AMP . . . . . Antimicrobial peptide

APD3 . . . . . Antimicrobial Peptide Database (version 3)

APL . . . . . Area per lipid

AUC . . . . . Area under the curve

BF . . . . . Brown-Forsythe [test]

C . . . . . Cysteine

CD . . . . . Circular dichroism

CDF . . . . . Cumulative distribution function

CG . . . . . Coarse-grained

CHARMM . . Chemistry at Harvard macromolecular mechanics

COG . . . . . Center of geometry

COM . . . . . Center of mass

CPU . . . . . Central processing unit

CSV . . . . . Comma-separated values

D . . . . . Aspartic acid

dRMSD . . . Distance-based root mean squared deviation

DSSP . . . . . Define Secondary Structure of Proteins

E . . . . . Glutamic acid

EDM . . . . . Electric dipole moment

F . . . . . Phenylalanine

FDA . . . . . Food and Drug Administration

FF . . . . . Force field

FFT . . . . . Fast Fourier transform

FN . . . . . False negative

FP . . . . . False positive

G . . . . . Glycine

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GPU . . . . . Graphical Processing Unit

GROMACS . Groningen machine for chemical simulations

GROMOS . . Groningen molecular simulation

GUI . . . . . Graphical user interface

H . . . . . Histidine

HDM . . . . . Hydrophobic dipole moment

HIE . . . . . Hydrogen isotope exchange

I . . . . . Isoleucine

IUPAC . . . . International Union of Pure and Applied Chemistry

JSON . . . . . JavaScript Object Notation

K . . . . . Lysine

KDE . . . . . Kernel density estimator

L . . . . . Leucine

LINCS . . . . Linear constraint solver

LL . . . . . Lower leaflet

M . . . . . Methionine

MD . . . . . Molecular dynamics

ML . . . . . Machine learning

MWU . . . . . Mann-Whitney U [test]

M2 . . . . . Martini 2.2

M2P . . . . . Martini 2.2P

M3 . . . . . Martini 3

N . . . . . Asparagine

NC . . . . . Negative control

NMR . . . . . Nuclear magnetic resonance

OPC . . . . . Optimal point charge

P . . . . . Proline

PBC . . . . . Periodic boundary condition

PC . . . . . Phosphatidylcholine or
Principal Component

PCA . . . . . Principal component analysis

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PDF . . . . . Probability density function
PE . . . . . Phosphatidylethanolamine
PI . . . . . Phosphatidylinositol
PG . . . . . Phosphatidylglycerol
PS . . . . . Phosphatidylserine
PME . . . . . Particle-mesh Ewald
POPC 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine
POPE 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine
POPG 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol
PR . . . . . Precision–recall
Q . . . . . Glutamine
QM/MM . . . Quantum mechanics/molecular mechanics
QQ . . . . . Quantile–quantile
R . . . . . Arginine
REMD . . . . Replica-exchange molecular dynamics
RF . . . . . Reaction field
RMSD . . . . Root mean squared deviation
S . . . . . Serine
SPC . . . . . Simple point charge
SQL . . . . . Structured Query Language
SW . . . . . Shapiro-Wilk [test]
T . . . . . Threonine
TFE 2,2,2-trifluoroethanol
TIP3 . . . . . Three-point transferable intermolecular potential
TN . . . . . True negative
TP . . . . . True positive
UA . . . . . United atoms
UL . . . . . Upper leaflet
V . . . . . Valine
vdW . . . . . Van der Waals
W . . . . . Tryptophan
WSR . . . . . Wilcoxon signed-rank [test]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of figures

1.1 Self-assembly of lipids in water . . . 2
1.2 Structure of glycerophospholipids . . . 4
1.3 Lipid profile of different cytoplasmic membranes . . . 6
1.4 Structure of antimicrobial peptides . . . 8
1.5 Mechanisms of membrane disruption by antimicrobial peptides . . . 9

3.1 Workflow for an MD simulation . . . 25
3.2 Particle in an external potential . . . 26
3.3 The leapfrog integrator . . . 27
3.4 Effect of the periodic boundary conditions . . . 28
3.5 Bond and angle stretching . . . 30
3.6 Improper and proper dihedral angles . . . 32
3.7 Non-bonded interactions in molecular systems . . . 33
3.8 The use of a cutoff for the calculation of non-bonded interactions . . . 34
3.9 Charge distribution in the particle-mesh Ewald method . . . 36
3.10 Representation of a molecule using different levels of resolution . . . 37
3.11 3- and 4-point water models . . . 39
3.12 Fundamentals of the velocity-rescaling approach . . . 42
3.13 Fundamentals of the cell-rescaling approach . . . 44
3.14 Equilibration protocol for the atomistic simulations in the "shallow" configuration . . . 51
3.15 Equilibration protocol for the atomistic simulations in the "embedded" configuration . . . 54
3.16 Filtering similar sequences . . . 56
3.17 Equilibration protocol for the CG simulations . . . 57
3.18 The wrapping and unwrapping transformations . . . 59
3.19 Rotation of a system of particles . . . 62
3.20 Formation of H-bonds . . . 63
3.21 Identification of the lipids in each leaflet . . . 65
3.22 Schematic representation of the area per lipid . . . 66
3.23 The equivalent ellipsoid of the peptide . . . 72
3.24 Penetration of the peptide into the lipid bilayer . . . 74
3.25 The spin and tilt angles . . . 76
3.26 Schematic translational and rotational diffusion . . . 77
3.27 Averages in periodic domains . . . 79
3.28 Multiple linear regression . . . 81
3.29 Statistical hypothesis tests . . . 83
3.30 Logistic curve . . . 92

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.1 Time evolution of the peptide–membrane distance for the shallow starting configuration 99
4.2 Initial and final states of a system starting from the shallow configuration 100
4.3 Time evolution of peptide helicity calculated with the DSSP method for the shallow configuration 101
4.4 Time evolution of peptide helicity calculated with the αRMSD metric for the shallow configuration 102
4.5 Time evolution of peptide helicity calculated with the α-block metric for the shallow configuration 104
4.6 Mean number of peptide–lipid contacts in simulations starting from the shallow configuration 105
4.7 Mean number of hydrogen bonds formed between the peptide and the lipids and within the peptide itself in simulations starting from the shallow configuration 106
4.8 Relation between the ratios of the 1st to 2nd and 2nd to 3rd semiaxes of the equivalent ellipsoid representing each peptide in simulations starting from the shallow configuration 108
4.9 Time evolution of the peptide–membrane distance for the embedded starting configuration 110
4.10 Time evolution of peptide helicity calculated with the DSSP method for the embedded starting configuration 111
4.11 Time evolution of peptide helicity calculated with the αRMSD metric for the embedded starting configuration 112
4.12 Representative final structures of the peptides in four different simulations 113
4.13 Time evolution of peptide helicity calculated with the α-block metric for the embedded starting configuration 114
4.14 Relation between the ratios of the 1st to 2nd and 2nd to 3rd semiaxes of the equivalent ellipsoid representing each peptide in simulations starting from the embedded configuration 116
4.15 Mean number of water molecules that penetrate the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns of the simulation 117
4.16 Translational diffusion coefficients of the peptides in simulations starting from the shallow configuration, computed over the last 200 ns of the simulation 118
4.17 Translational diffusion coefficients of the peptides in simulations starting from the embedded configuration, computed over the last 200 ns of the simulation 119

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.18 Relation between the translational diffusion coefficients of the peptides and their average distance to the center of the membrane . 120
4.19 Mean distance between the peptide and the center of the membrane in simulations starting from the embedded configuration, averaged over the last 200 ns 122
4.20 Mean helicity of the peptides in simulations starting from the embedded configuration, averaged over the last 200 ns of simulation 123
4.21 Translational diffusion coefficient of the peptides in simulations starting from the embedded configuration, averaged over the last 200 ns of simulation 124
4.22 Time evolution of peptide helicity calculated with the DSSP method in long-timescale GROMOS-HIE simulations 125
4.23 Homepage of the SuPepMem Database GUI 133
4.24 Dedicated page to a simulation in the SuPepMem database. Peptide tab 135
4.25 Dedicated page to a simulation in the SuPepMem database. Membrane tab 136
4.26 Dedicated page to a simulation in the SuPepMem database. Analysis tab 137
4.27 Q–Q plot of the MD descriptors 144
4.28 Samples of MD descriptors 146
4.29 Fluctuation in peptide–membrane interactions and orientation . . . 147
4.30 Representative MD descriptors related to membrane perturbation . 149
4.31 Effect of peptide tilt on the apparent in-plane rotation of its axis . 150
4.32 Partial adsorption of a peptide into a POPC membrane 153
4.33 Q–Q plots of peptide physicochemical and structural properties . . 155
4.34 Distribution of peptide physicochemical and structural properties . 156
4.35 Penetration of the peptide into the bilayer as a function of the fraction of hydrophobic residues in the sequence 159
4.36 Contacts between the backbone of the peptide and water molecules as a function of the fraction of hydrophobic residues in the sequence 160
4.37 Evaluation of the performance of the classification models 164

B.1 Relation between the helicity of the peptide calculated with the αRMSD and α-block methods in simulations starting from the shallow configuration 185
B.2 Relation between the helicity of the peptide calculated with the DSSP and the αRMSD methods in simulations starting from the shallow configuration 186

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.3 Relation between the helicity of the peptide calculated with the DSSP and the α-block methods in simulations starting from the shallow configuration 187
B.4 Linear relations between the different helicity metrics in simulations starting from the shallow configuration 188
B.5 Time evolution of the number of contacts between the peptide and the lipids in simulations starting from the shallow configuration 192
B.6 Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–lipid contacts in simulations starting from the shallow configuration 193
B.7 Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–lipid contacts in simulations starting from the shallow configuration 194
B.8 Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–lipid contacts in simulations starting from the shallow configuration 195
B.9 Time evolution of the number of peptide–lipid H-bonds in simulations starting from the shallow configuration 199
B.10 Relation between the peptide–lipid H-bonds and contacts in simulations starting from the shallow configuration 200
B.11 Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration 202
B.12 Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration 203
B.13 Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration 204
B.14 Time evolution of the number of H-bonds within the peptide in simulations starting from the shallow configuration 208
B.15 Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration 209
B.16 Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration 210
B.17 Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration 211

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.18 Time evolution of distance between the C_α atoms of the first and last residues in the peptide in simulations starting from the shallow configuration 215
B.19 Relation between the length of the 1st axis of the equivalent ellipsoid and distance between the C_α atoms of the first and last residues in simulations starting from the shallow configuration . 216
B.20 Mean area per lipid in simulations starting from the embedded configuration, averaged over the last 200 ns 217
B.21 Mean membrane thickness in simulations starting from the embedded configuration, averaged over the last 200 ns 217
B.22 Standard deviation of the area per lipid in simulations starting from the embedded configuration, computed over the last 200 ns . . 218
B.23 Standard deviation of the membrane thickness in simulations starting from the embedded configuration, computed over the last 200 ns 218
B.24 Mean number of water molecules adsorbed at the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns of the simulation 219
B.25 Time evolution of the peptide–membrane distance for the embedded starting configuration 220
B.26 Mean area per lipid in simulations starting from the embedded configuration, averaged over the last 200 ns 221
B.27 Mean membrane thickness in simulations starting from the embedded configuration, averaged over the last 200 ns 221
B.28 Standard deviation of the area per lipid in simulations starting from the embedded configuration, computed over the last 200 ns . . 222
B.29 Standard deviation of the membrane thickness in simulations starting from the embedded configuration, computed over the last 200 ns 222
B.30 Mean number of water molecules that penetrate the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns of the simulation 223
B.31 Mean number of water molecules adsorbed into the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns of the simulation 223
B.32 Time evolution of peptide helicity calculated with the DSSP method for the embedded starting configuration 224
B.33 Time evolution of peptide helicity calculated with the αRMSD method for the embedded starting configuration 225

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.34 Time evolution of peptide helicity calculated with the α-block method for the embedded starting configuration 226
B.35 Time evolution of the peptide–membrane distance in long timescale GROMOS-HIE simulations 227
B.36 Time evolution of peptide helicity calculated with the αRMSD and α-block methods in long timescale GROMOS-HIE simulations . 228

D.1 Samples of MD descriptors 234
D.1 Samples of MD descriptors 235
D.1 Samples of MD descriptors 236
D.1 Samples of MD descriptors 237
D.2 Evaluation of the performance of the classification models trained with data projected onto subspaces of principal components 263

E.1 License of reproduction for the content of Section 4.2 265
E.2 License of reproduction for Figure 1.5 266

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of tables

3.1 Values of the parameters of the external harmonic potentials applied during the equilibration 52

4.1 Results of the least-squares multivariate linear regression between the fraction of hydrophobic residues in the sequence H, the nominal activity A, and the penetration of the peptide into the membrane 158

4.2 Results of the least-squares multivariate linear regression between the fraction of hydrophobic residues in the sequence H, its nominal activity A, and the mean number of contacts with water molecules 161

B.1 Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and α-block 189

B.2 Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and α-block 189

B.3 Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and αRMSD 190

B.4 Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and αRMSD 190

B.5 Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and α-block 191

B.6 Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and α-block 191

B.7 Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of contacts with lipids . 196

B.8 Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and the number of contacts with lipids 196

B.9 Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of contacts with lipids 197

B.10 Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of contacts with lipids 197

B.11 Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of contacts with lipids 198

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.12 Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of contacts with lipids 198
B.13 Values of Pearson's correlation coefficient r^2 between the peptide–lipid H-bonds and contacts 201
B.14 Values of Kendall's rank correlation coefficient τ between the peptide–lipid H-bonds and contacts 201
B.15 Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of H-bonds with lipids . 205
B.16 Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and the number of H-bonds with lipids 205
B.17 Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of H-bonds with lipids 206
B.18 Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of H-bonds with lipids 206
B.19 Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of H-bonds with lipids 207
B.20 Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of H-bonds with lipids 207
B.21 Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of H-bonds within the peptide 212
B.22 Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and the number of H-bonds within the peptide 212
B.23 Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of H-bonds within the peptide 213
B.24 Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of H-bonds within the peptide 213
B.25 Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of H-bonds within the peptide 214
B.26 Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of H-bonds within the peptide 214

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

C.1 List of peptides simulated in CG resolution 231

D.1 Structural descriptors of lipid bilayers in membrane-only systems . 233
D.2 Median of the samples of MD descriptors 238
D.3 Variance of the samples of MD descriptors 239
D.4 Standard deviation of the samples of MD descriptors 240
D.5 Value of the statistical W from the Shapiro-Wilk tests applied to the samples of MD descriptors. 241
D.6 p-value from the Shapiro-Wilk tests applied to the samples of MD descriptors 242
D.7 Differences in the values of MD descriptors based on antimicrobial activity for each combination of lipid composition and force field. Results of the Mann-Whitney U test 244
D.8 Differences in the variances of the samples of MD descriptors based on antimicrobial activity for each combination of lipid composition and force field. Results of the Brown-Forsythe test . . 245
D.9 Differences in the values of MD descriptors based on lipid composition for each combination of antimicrobial activity and force field. Results of the Wilcoxon signed-rank test 247
D.10 Differential behavior of peptides on the interaction with lipid membranes, based on their activity, for each force field. Results of the Mann-Whitney U test. 249
D.11 Difference in the variances of the samples of differences between lipid compositions based on activity for each force field. Results of the Brown-Forsythe test test. 250
D.12 Differences in the values of MD descriptors based on force field for each combination of antimicrobial activity and lipid composition. Results of the Wilcoxon signed-rank test 253
D.13 Results of the Shapiro-Wilk tests applied to the samples of peptide physicochemical properties 254
D.14 Differences in the value of physicochemical properties based on antimicrobial activity. Results of the Mann-Whitney U test 255
D.15 Differences in the variance of the samples of physicochemical properties based on antimicrobial activity. The results of the Brown-Forsythe test 256
D.16 Kendall rank correlation coefficient τ between MD descriptors and structural properties 256
D.17 Statistics of the samples of performance metrics 257
D.18 Results of the Shapiro-Wilk tests applied to the samples of performance metrics. 258

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

D.19 Differences in performance based on using all data or FF-specific data for the training of classification models. Results of the Mann-Whitney U test 259
D.20 Differences in variance of the performance metric samples obtained using all data or FF-specific data for the training of classification models. Results of the Brown-Forsythe test 259
D.21 Differences in performance based on using FF-specific data for the training of classification models. Results of the Mann-Whitney U test 260
D.22 Differences in variance of the performance metric samples obtained using FF-specific data for the training of classification models. Results of the Brown-Forsythe test 260
D.23 Differences in performance based on using data from both lipid compositions or one for each force field. Results of the Mann-Whitney U test 261
D.24 Differences in variance of the performance based on using data from both lipid compositions or one for each force field. Results of the Brown-Forsythe test 262

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

List of publications

- **Publication 1**: Reproduced in Chapter 4, Section 4.2.

Suárez-Lestón, F.,[a,b] Calvelo, M.,[a,c] Tolufashe, G. F.,[d] Muñoz, A.,[b] Veleiro, U.,[b] Porto, C.,[a] Bastos, M.,[d] Piñeiro, Á.,[b] Garcia-Fandino, R.[a,d] (2022). SuPepMem: A database of innate immune system peptides and their cell membrane interactions. *Comput. Struct. Biotechnol. J.*, **20**, 874–881. DOI: 10.1016/j.csbj.2022.01.025

Author affiliations:
[a] Department of Organic Chemistry, Center for Research in Biological Chemistry and Molecular Materials, Santiago de Compostela University, CIQUS, Spain
[b] Departamento de Física Aplicada, Facultade de Física, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain
[c] Departament de Química Inorgánica i Orgànica and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, Barcelona 08028, Spain
[d] CIQUP, Centro de Investigação em Química, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Porto, Portugal

Quality metrics (2022): Q1 (Biophysics, biochemistry); Impact Factor 6.0.

Contribution: The contributions of the author to this work involved the development of the code for the analysis of the MD simulations and database backend.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumo

A vida, tal e como a coñecemos, comeza cunha fronteira física que delimita o espazo celular e separa o interior do medio exterior, creando o equilibrio necesario para o funcionamento da célula. Esa fronteira é a membrana celular, unha estrutura complexa, flexible e dinámica que garante a súa integridade e regula o intercambio de materia e enerxía coa contorna.

O seu fundamento estrutural reside nos fosfolípidos, moléculas anfifílicas que conforman o esqueleto básico da membrana. En medios acuosos, estas moléculas autoensámblanse de maneira espontánea: as súas cadeas hidrofóbicas agréganse nun proceso entropicamente favorable, que pode ser estabilizado debido tanto ás interaccións ente os grupos polares das súas cabezas como as que se dan entre estes e a auga circundante.

Se ben a estrutura básica das membranas celulares é común, hai unha gran diversidade nas súas compoñentes, tanto nos fosfolípidos como noutras moléculas: outras clases de lípidos—e.g. esterois—, proteínas, carbohidratos e demais moléculas. Estas diferenzas atópanse tanto na comparación das membranas de distintas especies, coma naquelas dentro dun mesmo organismo, dun mesmo tecido ou incluso nunha mesma célula ao longo do seu ciclo. Así, a membrana citoplasmática do mamífero promedio é asimétrica: na capa exposta ao medio extracelular predominan as fosfatidilcolinas, que son lípidos zwitteriónicos, polo que a carga superficial é cero; polo contrario, a capa interna está principalmente composta por fostatidiletanolaminas, tamén zwitteriónicas, e as fosfatidilserinas, unha clase de lípidos aniónicos. Esta asimetría é vital para o normal funcionamento dos mecanismos celulares e a súa supervivencia, polo que é a propia célula a que o mantén activamente.

Porén, o mecanismo que mantén a devandita asimetría deixa de funcionar correctamente nalgúns escenarios, como é o caso do cancro. As células cancerosas adoitan amosar unha tendencia cara á simetrización de ambas capas da membrana, de modo que os lípidos aniónicos inicialmente restrinxidos á cara interna, comezan a expresarse tamén na superficie externa. Como consecuencia, a diferenza dunha membrana dunha célula sá, a superficie da célula cancerosa presenta unha carga negativa.

Esta característica—a exposición de lípidos con carga neta negativa—na superficie externa da membrana citoplasmática é común noutros casos, como é o das membranas bacterianas. *E. coli*, un dos patóxenos máis estudados, presenta unha membrana celular moi diferente da dun mamífero, pois está constituída principalmente por fosfatidiletanolaminas e fosfatidilglicerois, sendo estes últimos lípido tamén aniónicos.

En xeral, a composición dos fosfolípidos—e outros lípidos—nas membranas dun individuo san difiren notablemente das dos patóxenos que os infectan, ou de tecidos patolóxicos—e.g. cancro, inflamación. Este cambio na composición implica tamén

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

un cambio nas propiedades fisicoquímicas das membranas, como é o caso da carga anteriormente mencionada. Esta alteración pode ser, e é, aproveitada polo sistema inmune e por certas moléculas terapéuticas para recoñecer e atacar selectivamente patóxenos ou células patolóxicas, sen causar danos ás células sas.

Entre os mecanismos que aproveitan estas diferenzas na composición das membranas destaca o empregado polo sistema inmune de multitude de organismos, incluíndo os seres humanos. Nel participan os péptidos antimicrobianos (AMPs polas súas siglas en inglés): pequenas cadeas de aminoácidos capaces de recoñecer e destruír selectivamente determinadas células. As propiedades dos AMPs son moi diversas, se ben acostuman a ser curtos, a presentar unha carga neta positiva e un carácter anfifílico, con rexións moi hidrofóbicas. De particular interese resultan aqueles AMPs que, tras seren adsorbidos polas membranas celulares, cambian a súa conformación e adquiren unha estrutura helicoidal. A carga positiva dos AMPs favorece unha interacción electrostática coa membrana citoplasmática de bacterias, fungos, células cancerosas ou outras. Posteriormente, o seu pregamento e a adquisición dunha estrutura anfifílica amplifica a interacción cos lípidos, expoñendo a rexión hidrofóbica cara as colas dos mesmos, tamén hidrofóbicas, e a rexión polar cara o exterior da membrana, preto da auga e as cabezas polares.

Unha vez unha cantidade significativa de AMPs se acumulan na membrana, existen diversos mecanismos polos que se agregan e son quen de producir poros que permiten o fluxo de material do citoplasma ao medio extracelular, ou extraer lípidos da membrana. En ambos casos, a alteración da estrutura da bicapa devén na morte da célula atacada. A diferenza das terapias convencionais baseadas en fármacos, a acción dos AMPs non depende do recoñecemento dun receptor específico, senon das propiedades fisicoquímicas das membranas obxectivo. Este carácter non-específico fai que resulten moito menos vulnerables aos mecanismos clásicos de resistencia antimicrobiana. É por este motivo polo que nas últimas décadas se ven investindo unha gran cantidade de esforzo na investigación de estas moléculas, se ben a súa aplicación práctica se ve moi limitada por dificultades intrínsecas como son os altos custos da investigación, as dificultades de produción ou a eficacia reducida *in vivo*.

Investigar o efecto dos AMPs sobre os organismos implica caracterizar a súa interacción coa membrana celular principalmente en dous aspectos: a estrutura que adquiren unha vez se adsorben e os mecanismos específicos a través dos cales exercen a súa acción. O estudo experimental dos AMPs acostuma a abordarse desde un enfoque multidisciplinar, combinando diversas técnicas coa fin de proporcionar unha perspectiva máis completa do sistema e compensar as carencias e limitacións das distintas metodoloxías. Mesmo así, non é posible obter experimentalmente unha descrición a nivel atómico da interacción entre os AMPs e as membranas e a súa evolución no tempo.

Esta limitación pode superarse parcialmente mediante a simulación computacio-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

nal de sistemas con estas compoñentes, en particular a través da dinámica molecular (MD, polas súas siglas en inglés). A MD é un método de simulación que resolve as ecuacións clásicas do movemento para obter a evolución temporal das posicións e velocidades dun sistema de partículas que representa un sistema molecular. Sendo un método puramente clásico, non modela explicitamente ningún fenómeno de orixe cuántica, se ben por iso é quen de representar sistemas en escalas de tempo e dimensións moito maiores que outras aproximacións, o que a converte nunha ferramenta especialmente útil para o estudo de AMPs e bicapas lipídicas.

A interacción de AMPs coas membranas celulares leva décadas a estudarse a través da MD, se ben a complexidade dos sistemas simulados creceu conforme o fixo a capacidade de cálculo dos ordenadores modernos. Pese a iso, inda existen limitacións, tanto intrínsecas como prácticas, relacionadas co custo computacional e co alcance temporal das simulacións, que restrinxen os fenómenos de interacción péptido–membrana que poden ser capturados.

Na actualidade, a enorme cantidade de datos xerados na investigación vense explotando para crear modelos preditivos que axuden no desenvolvemento de novas terapias, como pode ser o caso de secuencias de AMPs con propiedades melloradas. Neste contexto, dispor de bases de datos con información a nivel atómico sobre as interaccións moleculares resulta esencial para a construción de modelos máis robustos e precisos.

Así, esta tese explora a utilidade das simulacións de MD para xerar conxuntos de datos sobre a interacción de AMPs dirixidos a membrana con bicapas lipídicas coa fin de producir modelos preditivos da súa estrutura en membrana máis da súa actividade antimicrobiana.

O capítulo da Introdución presenta en máis detalle a estrutura das membranas citoplasmáticas e a súa relevancia. Continúa cunha breve exposición sobre os AMPs, con especial énfase naqueles que atacan as membranas celulares, así como dalgunhas técnicas experimentais empregadas no seu estudo. Isto é seguido pola explicación sobre como a MD pode empregarse para complementar o estudo experimental de estes sistemas, cubrindo os aspectos que as distintas metodoloxías non son quen de abordar. Finaliza esta sección a aplicación de métodos preditivos no ámbito dos AMPs, así como a posibilidade de integrar os datos de MD no seu desenvolvemento.

Esta introdución serve como base para formular os Obxectivos deste traballo. Brevemente, estes abarcan o uso potencial dos datos de MD para o desenvolvemento de modelos preditivos de estrutura e actividade de péptidos en bicapas lipídicas. Con esa fin, o primeiro dos obxectivos específicos consiste en avaliar os resultados de simulacións de MD en resolución atomística—as partículas do sistema a resolver representan átomos dun sistema real—para determinar se poden usarse como base para o desenvolvemento de ditos modelos. O segundo obxectivo consiste na aplicación do anterior para a realización sistemática de simulacións de sistemas péptido–

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

membrana, e do establecemento dunha base de datos estruturada para almacenar descritores numéricos da interacción entre estas compoñentes. O terceiro e último obxectivo consiste na análise estatística dos resultados das simulacións anteriores de cara á identificación de patróns asociados coa actividade das secuencias, e do posterior desenvolvemento de modelos de clasificación dos resultados das simulacións de cara á identificación do potencial antimicrobiano dos péptidos empregados nelas.

O capítulo de Metodoloxía describe en detalle o conxunto de procedementos empregados para acadar os obxectivos formulados. Comeza coa exposición da técnica de MD e das súas características fundamentais, seguida dunha descripción dos algoritmos principais empregados nas simulacións. Segue a isto a lista pormenorizada das simulacións realizadas, detallando as constituíntes dos sistemas, o protocolo de simulación e os parámetros empregados en cada un dos pasos. Descríbense a continuación as análises realizadas sobre as traxectorias para a caracterización dos sistemas: estas inclúen desde o estudo dos cambios conformacionais dos péptidos en simulacións atomísticas empregando distintos criterios, a cuantificación da interacción péptido–membrana, a orientación relativa do péptido respecto da súa superficie, as súas propiedades dinámicas, ou os cambios estruturais que este induce nas bicapas. O capítulo conclúe coa descrición das técnicas estatísticas aplicadas aos resultados das análises anteriores, e que teñen como fin última identificar diferenzas significativas entre os distintos sistemas simulados.

O capítulo de Resultados estruturase en tres bloques, cada un deles asociado a un dos obxectivos específicos do traballo. No primeiro, leváronse a cabo unha serie de simulacións de MD en resolución atomística: cada un dos sistemas simulados está constituído por unha bicapa lipídica simétrica, unha única unidade dun AMP en conformación helicoidal, auga e ións. Cada un dos péptidos foi simulado partindo de cinco orientacións diferentes relativas á membrana, e empregando catro parametrizacións das interaccións (campos de forzas) distintas. Primeiramente aplicouse un protocolo de simulación que aproxima o péptido á membrana, de modo que a simulación de produción parta con este próximo a unha das superficies da bicapa.

Nestas simulacións observouse como o péptido permanece preto dos fosfatos dos lípidos, non chegando a penetrar na membrana. A evolución da helicidade foi estudada empregando diferentes métricas; se ben todas coinciden en amosar a perda da hélice inicial, os resultados para as distintas réplicas dun mesmo sistema diverxen. A comparación das distintas métricas indica unha alta correlación entre elas, pese a que os valores numéricos sexan diferentes. Estes resultados compáraronse cos obtidos nunha segunda serie de sistemas que parten do péptido inserido na bicapa. Pese a que o péptido emerxen a unha posición preto da dos sistemas anteriores, a estrutura resulta ser máis estable e consistente entre as distintas réplicas.

As discrepancias observadas nun dos campos de forzas empregados no estudio (AMBER) foron analizadas avaliada, a fin de investigar a súa orixe: un aumento

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

da hidratación das membranas neste sistema pode ser o motivo da perda da helicidade, da maior distancia á membrana e das maiores flutuacións atopadas nestas simulacións.

Máis alá das propiedades estruturais, investigouse o comportamento difusivo do péptido na membrana. Viuse como a masa do péptido máis o campo de forzas teñen un impacto no coeficiente de difusión asociado ao movemento do péptido, o mesmo que a distancia respecto do centro da membrana.

Verificouse ademais que a aceleración das simulacións mediante o aumento da masa dos átomos de hidróxeno en enlaces polares non supuxo unha alteración significativa nos resultados, o mesmo que substituír o uso de algúns algoritmos máis exactos por outros lixeiramente menos custosos.

En conxunto, estas análises poñen de manifesto as limitacións das simulacións atomísticas para acadar a converxencia estructural dos péptidos e reproducir fenómenos de repregamento en escalas de tempo da orde dos microsegundos.

Posto que a reproducibilidade e a eficiencia das simulacións atomísticas anteriores dificultan o seu uso como fonte de información sistemática da interacción péptido–membrana, o segundo obxectivo abordouse desde un nivel de resolución máis baixo: a resolución *coarse-grained*. Neste tipo de simulacións, cada partícula representa un conxunto de átomos, o que reduce o número de operacións a realizar para un mesmo sistema, aumenta o rendemento e simplifica a superficie de enerxía libre asociada ao sistema.

Coa fin de obter resultados reproducibles dun xeito sistemático, desenvolveuse un protocolo de ensamblado de sistemas, simulación e análise que permite xerar traxectorias asociadas a sistemas de unha unidade de péptido e unha bicapa lipídica en condicións uniformes, así como extraer delas unha serie de descritores numéricos asociados á súa interacción.

Os resultados destas simulacións serviron de base para o desenvolvemento de SuPepMem, unha base de datos que facilita o estudo do comportamento de AMPs en membranas modelo. Esta base de datos almacena metadatos asociados á simulación, como poden ser a composición do sistema, as condicións da simulación—e.g. temperatura, presión, concentración de ións—, os ficheiros necesarios para replicala, máis os resultados da análise realizada sobre cada unha das traxectorias. Estes preséntanse tanto en formato gráfico como numérico, o que permite tanto a inspección visual de aspectos relevantes da simulación como a obtención de parámetros numéricos que poden ser empregados no estudo sistemático dos sistemas. Desenvolveuse unha interface gráfica (`www.supepmem.com`) que permite o libre acceso aos recursos almacenados na base de datos. Implementásonse ademais unha serie de ferramentas de busca e comparación que fan da interface gráfica un recurso de utilidade propia no estudo dos sistemas.

Para abordar o terceiro obxectivo realizáronse unha serie de simulacións *coarse-*

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

grained que inclúen un total de 123 secuencias de péptidos—73 das cales corresponden a AMPs con actividade sobre membranas celulares—modeladas como hélices ideais. As simulacións incluiron dous modelos de membrana—mamífero san, e a bacteria *E. coli*—, e tres campos de forzas *coarse-grained* diferentes. O deseño destas simulacións, incorporadas á base de datos SuPepMem, tiña como obxectivo avaliar se a interacción entre unha única unidade de péptido e unha membrana reflicte indicios da actividade antimicrobiana da secuencia. Con esta fin, compáraronse os resultados das simulacións de AMPs coas de péptidos sen actividade coñecida baixo condicións uniformes—mesma membrana, mesmo campo de forzas—mediante técnicas estatísticas. Veuse que, se ben apareceron diferenzas significativas entre os resultados das simulacións, estas non son o suficientemente grandes como para que ningún dos descritores numéricos estudados poida servir como un indicador da potencial actividade da secuencia.

Atopáronse tamén diferenzas entre as simulacións dunha mesma secuencia cando a composición da bicapa varía, o que levou a estudar o comportamento diferencial dos péptidos ante as distintas membranas simuladas segundo sexa a actividade antimicrobiana asociada. Os resultados indicaron que, dos campos de forzas empregados nas simulacións, é o máis recente (Martini 3) o que reflicte maiores diferenzas entre os péptidos activos e non activos neste respecto, se ben estas seguiron sen ser o suficientemente destacables como para empregalas como criterio de clasificación.

Unha análise posterior das diferenzas entre campos de forzas confirmou a existencia de diferenzas sistemáticas nos resultados das simulacións. Mentres que na versión máis antiga (Martini 2.2) se veu unha adsorción case universal dos péptidos, Martini 3 non só corrixiu este efecto, senón que amosou unha maior selectividade dos péptidos: os péptidos activos tenderon a integrarse na membrana con maior frecuencia que aqueles sen actividade coñecida.

En conxunto, as análises estatísticas amosaron diferenzas significativas entre péptidos con e sen actividade, así como un efecto dependente da composición da bicapa e da parametrización dos campos de forzas. Porén, non se identificou ningunha magnitude derivada directamente da simulación que reflectise dun xeito inequívoco a potencial actividade antimicrobiana das secuencias.

Para dar cumprimento completo ao terceiro obxectivo formulado, investigouse o uso dos descritores numéricos derivados das simulacións para o desenvolvemento de modelos de clasificación da actividade antimicrobiana das secuencias. Escolleuse a regresión loxística como arquitectura do modelo ao combinar unha estrutura sinxela, un número reducido de parámetros axustables, e unha potencial interpretabilidade da transformación dos datos que aplica. Coa fin de determinar a combinación óptima de datos de entrada, creáronse conxuntos de datos de simulación que corresponden a cada unha das combinacións de composición da bicapa e campo de forzas, a todos os datos obtidos baixo un mesmo campo de forzas, e un último conxunto que combina

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

todos os datos xerados—as dúas membranas e os tres campos de forzas.

Definiuse como potencial aplicación o filtrado de secuencias, buscando que os modelos reducisen o número de simulacións de AMPs incorrectamente clasificadas. Baixo esta premisa, os modelos que amosaron uns mellores resultados foron aqueles adestrados con datos do campo de forzas máis recente (Martini 3), particularmente aquel que só empregou simulacións do modelo de membrana de bacteria. Verificouse ademais que estes resultados son independentes da dimensionalidade dos datos empregados, o que indica que non só as simulacións en resolución *coarse-grained* reflicten diferenzas que poden ser asociadas á actividade antimicrobiana dunha secuencia, senón que ademais a versión máis recente do campo de forzas é coa que estas diferenzas son máis salientables.

Se ben a aplicabilidade destes resultados se ve limitada por certas restricións intrínsecas ao enfoque escollido—e.g. unha soa unidade de péptido, unha estrutura helicoidal persistente—, os resultados poñen de manifesto o potencial uso dos datos de MD para o adestramento de modelos preditivos das propiedades de AMPs baseados na diferenciación das interaccións específicas con modelos de membrana celular.

O corpo deste traballo conclúe cun capítulo dedicado ás principais Conclusións derivadas das distintas partes nas que se estruturou o traballo realizado. En conxunto, abordouse o estudo de sistemas péptido–membrana desde un enfoque multiescala. Víronse as limitacións do uso de simulacións atomísticas únicas de cara á caracterización dos cambios conformacionais de péptidos, caracterizadas en xeral por unha baixa reproducibidade, así como algunhas diferenzas notables nos resultados dos distintos campos de forzas. Ao mesmo tempo, comprobouse como o método de incremento da masa dos hidróxenos produciu resultados comparables aos dunha simulación estándar, e como estas observacións se manteñen tras un tratamento simplificado das interaccións electrostáticas. En base a estes resultados, deseñouse un protocolo de ensamblaxe, simulación e análise de sistemas péptido–membrana en resolución *coarse-grained*, onde o maior rendemento da simulación e a simplificación da superficie de enerxía libre favorece a simulación sistemática e reproducible de multitude de sistemas. Os resultados destas simulacións almacenáronse nunha base de datos deseñada para esta fin, SuPepMem. Esta ferramenta de acceso libre contén información sobre os parámetros da simulación, a composición dos sistemas, mailos resultados dunha análise sistemática das traxectorias xeradas, así como os arquivos necesarios para a súa reprodución por parte de calquera usuario. Finalmente, avaliouse a utilidade desta plataforma a través do estudo estatístico das diferenzas entre sistemas onde as secuencias dos péptidos están asociadas ou non a AMPs. Se ben se observou a sensibilidade das simulacións a esta característica, foi necesario o desenvolvemento dun modelo de clasificación baseado nos datos de simulación para demostrar que os datos de MD poden servir para esta fin. En conxunto, os

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

resultados amosan que os datos de simulación poden empregarse tanto para xerar información cuantitativa sobre a interacción péptido–membrana non accesible experimentalmente, como para alimentar modelos preditivos destinados á identificación e caracterización de novas secuencias con actividade antimicrobiana.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 1.

Introduction

Come, sit with me, my fellow traveler.
Let's sit together and watch the stars die.
Chert, Outer Wilds

1.1 Lipid membranes. The scaffold of life

Few things are common to all known forms of life; one of them is that cells are their most basic functional unit, that cells are a medium—cytoplasm—enclosed by a boundary, the cell membrane, and that the main constituent of this structure are lipids.

Lipids According to IUPAC, lipid is "a loosely defined term for substances of biological origin that are soluble in nonpolar solvents"[1] which, as a definition, is not very practical. The identification and classification of lipids is complex, and different categories can be defined according to the necessities.

Focusing on those that are involved in the formation of the cell membrane, phospholipids are by far the most abundant—and the most relevant for this work.[2] The basic structure of phospholipids consists of an alcohol group connecting a hydrophilic head group, which contains a phosphate—hence their name—, with two hydrophobic tails (Figure 1.1, left).[3]

Self assembly of lipids in solution The amphiphilic nature of phospholipids leads to their aggregation in polar solvents, where both the hydrophobic interaction between the tails and the hydrophilic character of the head group play a key role.[4] Hydrophobic interactions are entropy-driven processes that occur naturally when nonpolar substances—or substances containing nonpolar regions—aggregate in polar media, as lipids do in water. Water molecules in a liquid state are forming a dynamic network of hydrogen bonds (H-bonds). When a nonpolar entity is incorporated,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

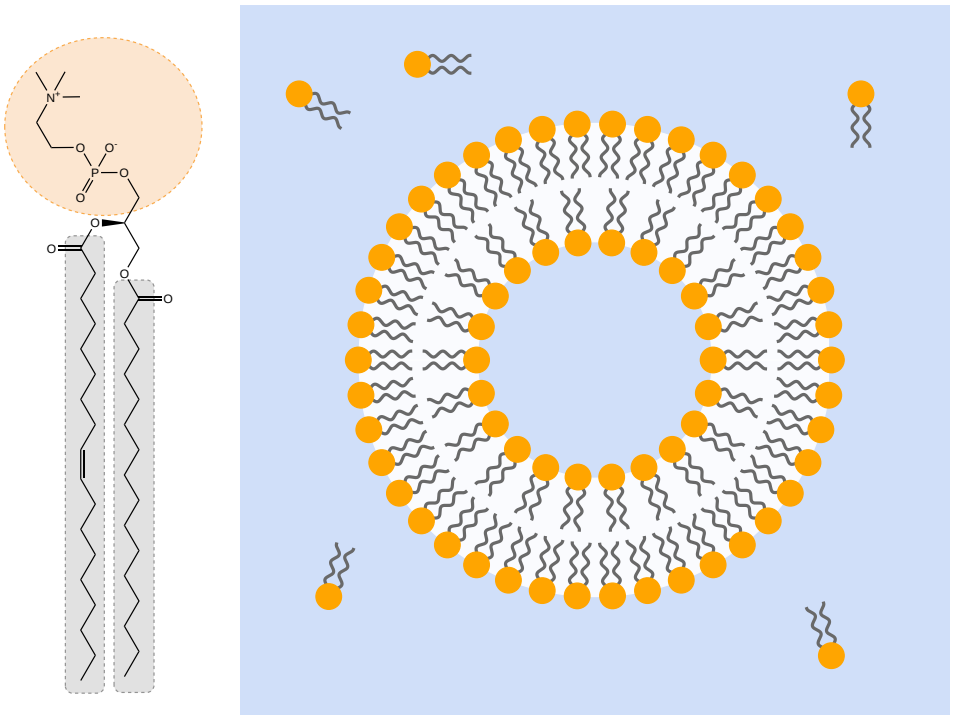


Figure 1.1: Self-assembly of lipids in water. (Left) Schematic representation of a phospholipid, where the polar head group (in orange), and the two lipid tails (in gray) can be elucidated. (Right) Hydrophobic interactions between the alkyl chains and the formation of H-bonds between the lipid head groups and water lead to the formation of stable vesicles, where lipids are distributed forming a bilayer. Unless stated otherwise, all figures in the present work have been created by the author. ChemDraw has been used to generate the molecular structure and Matplotlib for the vesicle scheme.[8]

because of the absence of polar groups capable of forming H-bonds with water, this lattice is disrupted. The water molecules in its vicinity will then rearrange so that a minimal number of H-bonds are lost, in exchange for the loss of certain freedom to move. When multiple nonpolar units are present, their contact minimizes the total hydrophobic surface exposed to water, leading to an overall increase in entropy compared to isolated molecules. Lipids will therefore aggregate so that their hydrophobic tails are in contact with each other and expose the polar head groups to water.[5] This can further increase the stability of the aggregates due to the formation of H-bonds, both between the head groups of adjacent molecules and with water.[6,7]

The shape of the structure as well, as their properties, depend on its constituents[9]; phospholipids will typically aggregate forming a bilayer, where the tails of the two layers (leaflets) are in contact, and the head groups are exposed to water. This bilayer becomes a closed surface, called a vesicle or liposome (Figure 1.1).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

When the size of these liposomes is large and its shape is regular, the local curvature of the surface can be low enough to be approximated by a flat surface.

The diversity of lipid molecules Even when referring only to phospholipids, whose structure is clearly defined and relatively simple, a plethora of molecules with different properties can arise. The most abundant lipids in cell membranes are glycerophospholipids, those in which the alcohol moiety is a glycerol,[10] but others like ceramides and their derivatives—e.g., sphingomyelins, where choline is part of the head group—also play key roles in cellular processes.[11]

The synthesis of glycerophospholipids—in eukaryotes and bacteria—requires the esterification of two fatty acids with glycerol in C_1 and C_2, which then become their hydrophobic tails.* These fatty acids, consisting of a carboxylic acid and an aliphatic chain, will condition some relevant structural properties of the membranes, such as its stability or fluidity: Longer chains favor the attractive interaction between lipids, reducing the fluidity of the membrane, while the kinks induced by the unsaturations reduce the packing of the lipids within the bilayer, which results in the opposite effect. Lipids in the membrane can undergo a phase transition between a gel phase, where the bilayer is more rigid, to a more fluid liquid phase; the chains affect the temperature at which this transition occurs: Longer tails and more saturation increase the transition temperature.[14,15] The combination of lipids of different tail lengths and saturation is a mechanism for the adaptation of cells to their environment and function.[16]

The phosphate group is in general doubly esterified, linked to the glycerol and to another group. An exception is phosphatidic acid, a minor constituent in cell membranes but a key precursor in the biosynthesis of other lipids[17]: In this simple phospholipid, only the glycerol is esterified with the phosphate group. The other relevant phospholipids can derive from the phosphatidic acid through esterification between the phosphate moiety and other groups. This constitutes the lipid head group (HG). The main types of glycerophospholipids found in cell membranes, classified according to the head group are (Figure 1.2): phosphatidylcholines (PC), where a choline endows the lipid with a zwitterionic character; phosphatidylethanolamines (PE), where the ethanolamine also imbues a zwitterionic character on the lipid; phosphatidylglycerols (PG), an anionic lipid due to the neutral charge of another glycerol moiety in its head group; phosphatidylserines (PS), an anionic lipid obtained as a result of the esterification of the phosphate with the side chain of serine, an amino acid with zwitterionic character at physiological pH; phosphatidylinositols (PI), a family of anionic lipids produced by the esterification with inositol, a sugar;

*In the stereospecific numbering (*sn*) method they corresponds to the positions *sn*1 and *sn*2. Throughout this work the *sn* numbering is used to identify the position of the substituents in glycerol.[12,13]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

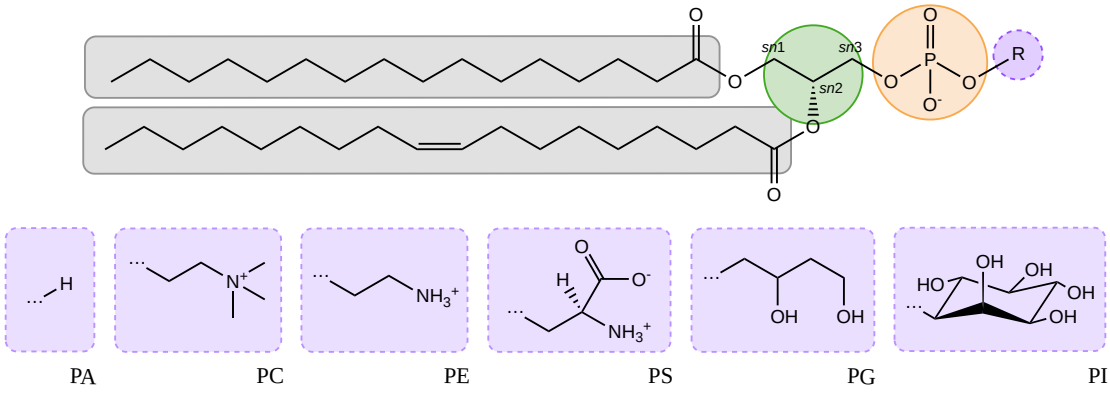


Figure 1.2: Structure of the main glycerophospholipids. All glycerophospholipids share the same basic structure: a glycerol group (in green) connected to two aliphatic chains (gray) and to a phosphate group (orange). The phosphate might also be connected with another functional group (purple). The position of the substituents in the glycerol is indicated by their *sn* position. ChemDraw has been used to generate the molecular structures.

and cardiolipins, the result of the esterification of two phosphatidic acid moieties with a glycerol.[18]

The structure of cytoplasmic membranes Glycerophospholipids provide the basic scaffold of the membrane that separates the cytoplasm from the external medium—the cytoplasmic membrane—, as well as for other structures such as organelles in eukaryotes, but are not their sole constituent.[2] Other lipids, such as sphingolipids and sterols, play a fundamental role in modulating the structural properties of cell membranes and their function[19]: Together with the aforementioned changes in fluidity, the local curvature can be modulated by the concentration of lipids of specific sizes in certain regions. Other components of the cell membrane, such as some proteins, require specific lipids in their environment to function.[20]

Membrane proteins are fundamental for the viability of the cell as a living entity, as their functions range from the transport of essential molecules to signaling.[21] The relative abundance of proteins in the membrane varies widely, from being a minor component of neuronal myelin to half of the mass in the eukaryotic cell membrane. In the mitochondrial membrane, proteins are the main constituent by mass.[22]

Cell membranes can be modeled as a 2-dimensional fluid of lipids, proteins, carbohydrates, and other molecules. This "fluid mosaic" model[23] accounts for the dynamical behavior of molecules in the plane of the membrane. This model set the foundations over which more complex concepts, such as the existence of domains within the membrane—regions with a different lipid composition—,[24] were incorporated.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A universal scaffold with diverse compositions In many organisms, the cytoplasmic membrane is further surrounded by a cell wall, whose composition varies substantially between plants, fungi, and bacteria.

A first-order distinction among bacteria is commonly made on the basis of the cell envelope architecture underlying the Gram stain.[25] In Gram-negative—diderm—bacteria, a thin peptidoglycan layer lies in the periplasm between the inner cytoplasmic membrane and an additional outer membrane. The outer membrane has an inner leaflet of phospholipids and an outer leaflet enriched in lipopolysaccharides. In contrast, Gram-positive—monoderm—bacteria lack an outer membrane and are surrounded by a much thicker peptidoglycan layer, typically containing teichoic and lipoteichoic acids.

Once the cell wall is removed, the underlying cytoplasmic membranes of both eukaryotes and bacteria share the same bilayer organization. However, their lipid composition differs significantly between species and cell types. In mammalian cells, membranes are asymmetric: the outer leaflet is enriched in PC, cholesterol, and sphingomyelin, while the inner leaflet mainly contains PE and PS.[26,27] This asymmetry leads to a neutral external surface and a negatively charged internal surface, favoring electrostatic interactions with cytoplasmic proteins.[28] Cholesterol further modulates membrane stability and fluidity.[29] Bacterial cytoplasmic membranes, by contrast, contain predominantly PE and PG, although their ratios vary widely across species[30,31] and even during the cell cycle.[32] A paradigmatic case is *Escherichia coli*, a Gram-negative bacterium, whose membrane is composed of PE ($\sim$ 75 %), PG ($\sim$ 20 %), and a smaller fraction of cardiolipin, a composition markedly different from that of mammalian membranes.[33]

Targeting alterations of the lipidic profile But while the "standard" lipid composition of cells can change to adapt to external conditions,[34–36] alterations of this profile can also be symptomatic of major issues. A scenario in which this naturally occurs is that of programmed cell death—apoptosis—in mammals: Leaflet asymmetry in a healthy cell is maintained thanks to specific enzymes.[37] In the apoptotic stage, the action of these enzymes begins to fail, so both leaflets become more symmetrical. This means that PS will manifest progressively on the external different of the cell membrane.[38] Exposure of PS to the intercellular medium is a signal that triggers the action of macrophages, which will engulf the apoptotic cell.[39]

A similar exposure of PS to the external surface of the cell membrane in mammals has also been observed in a very different scenario: in cancer cells. In this case, abnormal expression of PS acts as a protection mechanism, as it inhibits the immune response of the organism.[40] Cholesterol levels are another of the changes that cancer cells manifest differently from those healthy: Reduced cholesterol content can increase membrane fluidity, enhancing motility and facilitating metastasis.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

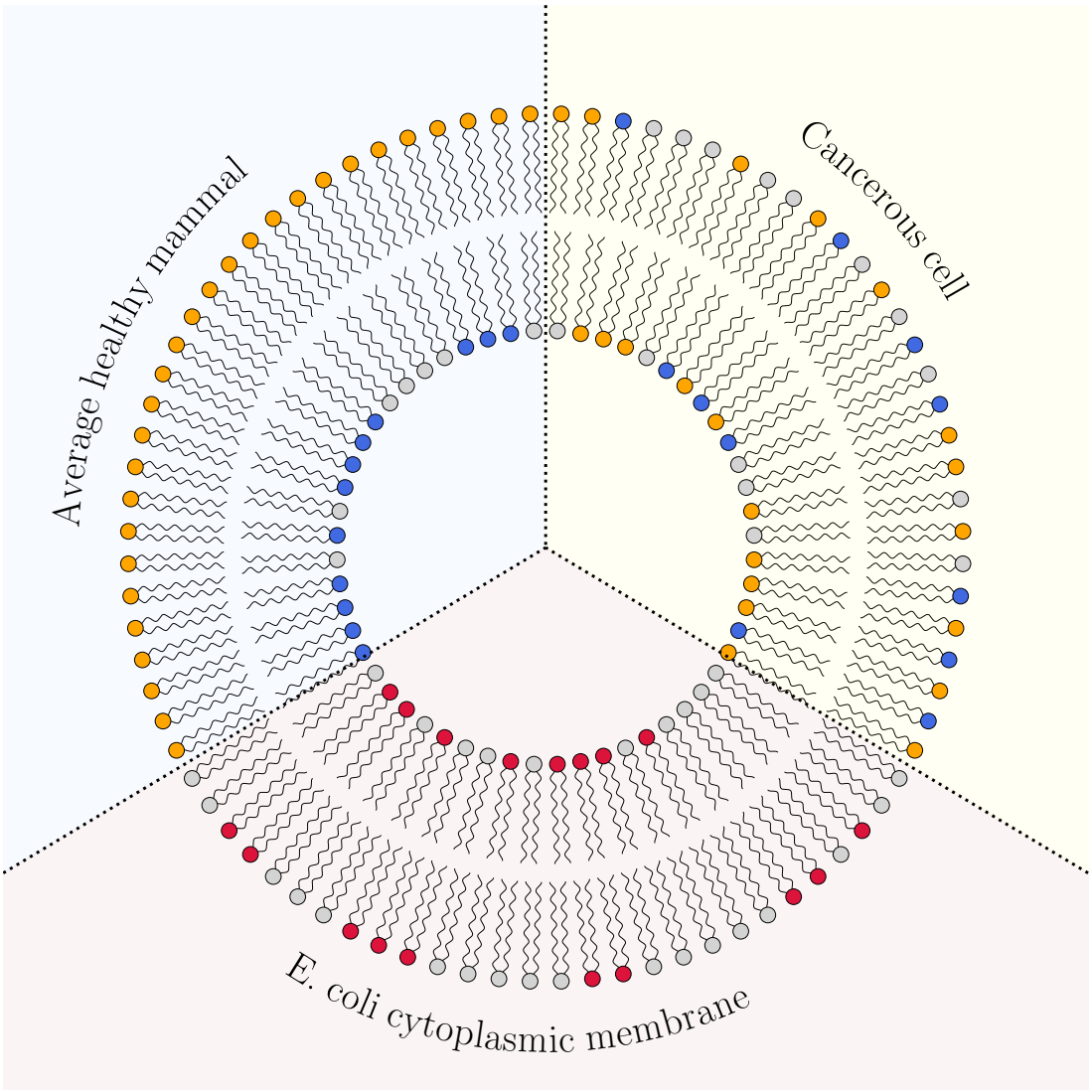


Figure 1.3: Lipid profile of different cytoplasmic membranes. While most organisms share the lipid bilayer structure as the basic scaffold of their cell membranes, the lipid composition can be very different. In an healthy mammal, the cell membrane is asymmetric, composed mostly of phosphatidylcholines (orange) in the outer leaflet and phosphatidylethanolamines (gray) and phosphatidylserines (blue) in the inner leaflet. This means that the inner surface is negatively charged, in comparison with a mostly neutral external surface. In cancer cells, this asymmetry is lost, and lipids become more evenly distributed, hence the external surface acquires a negative charge. The lipid composition of the cytoplasmic membrane of *E. coli* is very different from that of a mammal, containing phosphatidylethanolamines and phosphatidylglycerols (red), but as in the cancerous cell, it also displays an external negative charge. Scheme generated with Matplotlib.[8]

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Nevertheless, this effect is not universal: higher cholesterol levels in some types of cancer have been shown to be associated with a higher resistance to chemotherapy.[41]

Aging is another circumstance in which the lipids in the cytoplasmic membrane change: These alterations are complex and depend on several factors, such as the specific tissue studied or the gender.[42] It is, however, important to identify changes in the lipid profile of patients, as certain alterations, such as the variation in the concentration of certain lipids or their oxidation, can be associated with the development of several diseases associated with aging.[43,44]

While succinct, this relation of differences found between the lipid profile of a healthy cell and an altered one, or between the former and a pathogen (Figure 1.3), highlights why cell membrane composition is being investigated as a potential therapeutic target in the development of new therapies.[45]

1.2 Antimicrobial peptides

The label antimicrobial peptides (AMPs) refers to a family of short peptides found in all forms of life, from bacteria to more complex organisms. AMPs can target both pathogens, such as bacteria or fungi, and some pathological cells, e.g. cancerous cells. The way in which AMPs attack these cells depends on both the specific AMP and the target,[46] varying between a disruption of the cell membrane—the so-called membrane-targeting AMPs—and an inhibition of their metabolism. The ability of membrane-targeting AMPs to identify and attack cytoplasmic membranes is the main focus of this work.

The physicochemical properties are determined by the specific sequence of amino acids they are composed of—primary structure—, as well as the 3-dimensional structure they acquire—secondary structure. Most AMPs are cationic, which favors their interaction with negatively charged membranes of the organisms they target, and also amphipathic, enhancing their adsorption into the membranes.[46] Anionic AMPs, although less abundant, are also viable; their mechanism of action many times involves the formation of salt bridges with metal ions.[47]

The structure of AMPs A common distinction between AMPs is made by the secondary structure they acquire when interacting with their targets (Figure 1.4). The secondary structure, the amino acid sequence and the antimicrobial activity are interdependent on each other, as the arrangement of the residues in the space tends to the formation of an amphipathic structure that enhances the interaction between the AMP and the target cell.[48]

Among those for which the secondary structure could be resolved, the most abundant pattern is the α-helix. These peptides are usually rich in residues such as

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

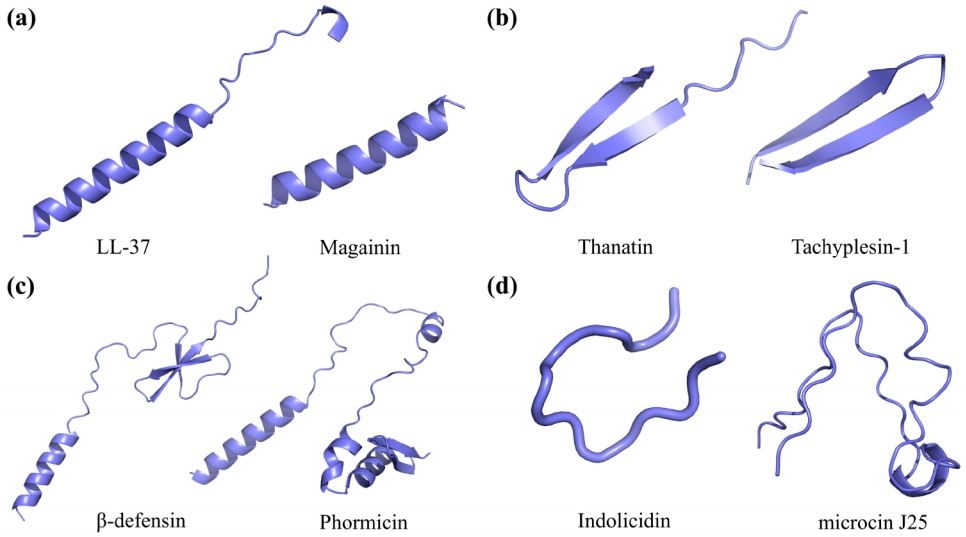


Figure 1.4: Structure of antimicrobial peptides. AMPs are typically present four main types of structure: (**a**) α-helical, (**b**) β-sheet, (**c**) mixed α-helix/β-sheet or (**c**) others. Image reproduced from the work by Ma *et al.*,[49] distributed under a Creative Commons CC BY license.

alanine (Ala, A), leucine (Leu, L), or lysine (Lys, K)—which are known to be helix-stabilizing amino acids. An important characteristic of these AMPs is that they undergo a conformational change upon adsorption into the lipid membrane of the target organism, as they are commonly found in a random coil state in solution.[50] As a result, the structure becomes amphiphilic, with hydrophobic and hydrophilic regions distributed on opposite sides along its axis. Examples of this type of peptides are magainins,[51] parasin[52] or LL-37.[53]

AMPs in which the secondary strucure is a β-sheet stabilize it, both in solution and when adsorbed, due to the presence of a variable (even) number of cysteine (Cys, C) residues forming disulfide bonds; human defensins are some of the AMPs which manifest this structure.[54] There exists also a group of AMPs whose secondary structure combines both α-helices and β-sheets, like the *Pisum sativum* defensin 1,[55] and another group in which no regular secondary structure is present, such is the case of indolicidin.[56] All of the examples mentioned are natural AMPs that can be found in very diverse organisms.

The mechanisms of membrane disruption Membrane-targeting AMPs are a subfamily of AMPs whose therapeutic action is done through the disruption of the structure of the cell membranes of their targets. In general terms, AMPs being amphipathic, with regions of high hydrophobicity, and cationic favor their interaction with the cell membrane of pathogens and pathological cells. There are two main

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

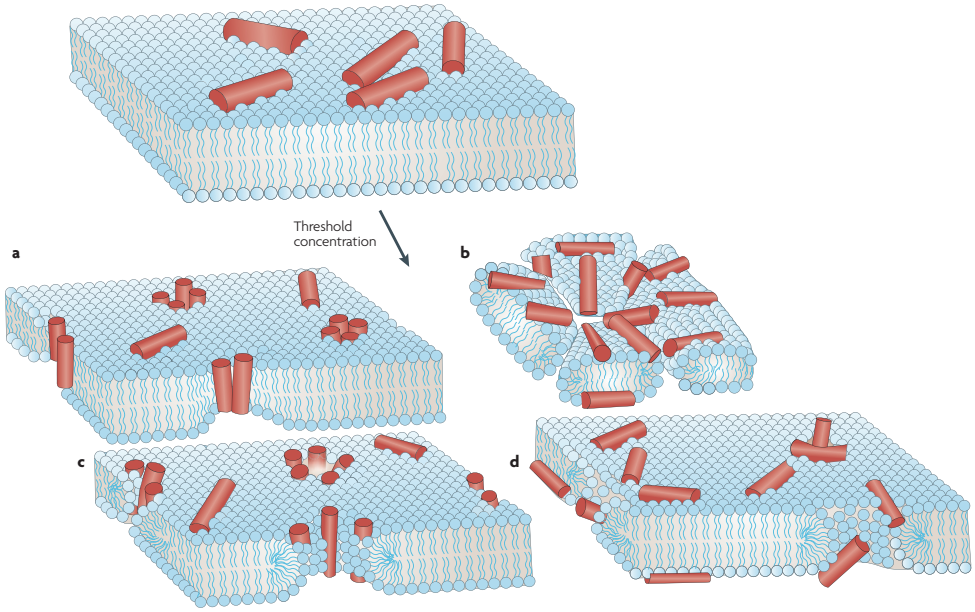


Figure 1.5: Mechanisms of membrane disruption by antimicrobial peptides. Membrane-targeting AMPs are adsorbed into the lipid membrane. Once a threshold concentration has been reach, peptides can induce the membrane disruption through different mechanisms: (**a**) barrel-stave pore model, (**b**) carpet model, (**c** and **d**) toroidal pore model. Image reproduced from the work by Melo, Ferre, and Castanho,[59] copyright by Springer Nature.

stages by which such an interaction takes place: the first is an electrostatic interaction between the AMP and the target. As already explained, the membranes of—not only—bacteria and cancerous tissues display an external net negative charge, which attracts a cationic peptide. Then, the amphiphilic character of the peptide favors adsorption due to the interaction between its hydrophobic domains with the nonpolar lipid tails, while the hydrophilic region may remain exposed to the surface.[57,58]

A single peptide adsorbed into the membrane would hardly affect the target, but after the accumulation of a significant amount of these molecules, the physical properties of the membrane will start to change. This alteration of the cell membrane could lead to the disruption of its structure, which will eventually lead to the death of the cell. Modification of the permeation properties of the cytoplasmic membrane can be achieved through different mechanisms (Figure 1.5), but only direct observation of the action of AMPs can be used to elucidate which one has occurred.

Some AMPs act by forming pores in the cytoplasmic membrane, which will lead to the leakage of material from the cytosol to the extracellular medium. Depending on the characteristics of the pores formed, two mechanisms are described[57]: the *barrel-stave* model requires a sufficient amount of peptides to acquire a trans-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

membrane orientation, aligned perpendicularly to the membrane surface, and their aggregation. Peptides are distributed within this cluster with their hydrophobic regions facing the lipid tails, whereas the hydrophilic parts create a channel that allows the flow of molecules across the membrane. The width of the formed pores is conditioned by the amount of peptides involved; the larger their number, the wider the whole in the membrane and the greater the leakage of molecules from the cytoplasm.

Similarly, the *toroidal-pore* model also allows the flow of cellular material by creating a channel across the membrane, but their topology and properties are different from the previous case. In these pores, the hydrophilic region of AMPs also interacts with polar lipid head groups. This leads to an alteration of the local curvature of the membrane, reorienting the lipids so that those of different leaflets become in contact with each other through their head groups. The inside of the pore then becomes lined with both AMPs and lipids.

Both mechanisms share the prior adsorption of peptides on the membrane surface, where the hydrophobic region of the peptide remains closer to the lipid tails. An alternative mechanism of action of AMPs consists in the aggregation of a significant amount of peptides while maintaining this orientation; in this *carpet* model the interaction between a large number of peptides and lipids alters the structural properties of the entire membrane, which can also induce the leakage of molecules across the membrane, without forming pores.[60] An extreme case is the *detergent* model, where after the accumulation of a very large amount of peptides they are able to induce the micellation of part of the lipid bilayer, effectively removing parts of the membrane. In the literature, they are sometimes not separated into different categories.[61]

It is worth highlighting how, unlike other molecules that bind to very specific receptors, AMPs interact with their targets and attack them due to the anomalous physicochemical properties of pathogens and pathological membranes—derived from a different composition in their membranes—when compared with those of the host. This makes AMPs not only less sensitive to antimicrobial resistance than conventional drugs, but also a broad-range tool to attack several diseases.[62]

AMP-based therapies Among the most interesting aspects of AMPs is that they are already part of the immune system of all organisms. Not only that, but after thousands of generations they are still effective. AMPs target pathogens in a much less specific way than other conventional drugs, so they are a promising candidate for the development of new drugs that overcome the issues associated with antimicrobial resistance that the latter suffer, e.g. reduction in membrane permeability or loss of efficacy due to modification of metabolic pathways or the receptors they bind.[63] It is estimated that by 2050, more than 8 million deaths will be associated with

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

antimicrobial resistance[64]; in this scenario, developing alternative ways to treat new but also current diseases becomes a top priority in public health.

However, the successful translation of AMP research into clinical applications has been limited. As of early 2025, only a few natural or derivative AMP-based antibiotics have received FDA approval,[65] including cyclic peptides such as daptomycin and polymyxins. Among linear AMPs composed exclusively of standard amino acids, gramicidin D—a mixture of closely related peptides discovered in the 1930s and approved in 1955—remains the only example in clinical use. Gramicidin D acts by forming channels in bacterial membranes, and is particularly effective against Gram-positive bacteria. Despite extensive research and numerous clinical trials, the development of new AMP-based drugs continues to face challenges related to stability, toxicity, and production cost.

There are several factors that hinder the development of AMP-based therapies.[61,66] The first is their instability, which makes their administration difficult: peptides are easily degraded or cleared by the organism. Then, despite their potent *in vitro* antimicrobial activity, once administered, their efficacy is reduced, again due to the change in the physiological conditions under which they are exposed in the organism. Some AMPs also have toxic effects, from inducing an unwanted immune response to damaging healthy cells. Finally, the high cost of the development of new therapies should be acknowledged, which, together with their limitations, reduces the viability of projects based on AMPs.

1.3 Experimental study of the interaction between antimicrobial peptides and membranes

The complexity of the study of AMPs and the development of new therapies based on these molecules requires the combination of different experimental techniques that can provide a more complex picture of their interaction with lipid membranes.

Most studies start not directly using real pathogens, but simplified models of cell membranes. This ranges from surfactant micelles, homogeneous or binary lipid compositions, etc.[58,67–72]; or even co-solvents, where 2,2,2-trifluoroethanol (TFE) stands out, as it is an helix-stabilizing medium.[58,68,71–74] These simplified models are easier to manipulate and study, but extending the results of their research to real pathogens is not straightforward. In addition, the vast diversity of models used makes the comparison between experimental results in the literature difficult. The structure of some membrane-targeting AMPs, especially those with a α-helix content, is very sensitive to the medium,[75,76] so one cannot just compare the values obtained from different sources without first ensuring that they were determined under the same conditions.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

To study the structure of peptides in interaction with membranes, the most widely used techniques are circular dichroism (CD) and nuclear magnetic resonance (NMR). CD spectroscopy can not only provide information about the structure of peptides,[77,78] but also indicate how they are distributed when bound to the membrane.[79] The interaction between circular polarized light and chiral centers in molecules is affected by the orientation of this polarization: In CD, the difference in the absorption of left- and right-circularly polarized light can be used to identify characteristic patterns related to the molecular structure. In the particular case of the helical content of a peptide, this can be estimated from the depth of two characteristic minima that can be observed at an incident wavelength around 208 and 222 nm, which originates in electronic transitions of the peptide bonds.[80,81] The main drawback of this method is that the recorded spectrum is the average over a sample of several molecules, or otherwise the recorded signal would be too dim, being impossible to obtain a detailed depiction of individual molecules. While it is possible to infer the secondary structure content of peptides using CD spectroscopy, this information comes from models and not from the CD spectrum recording atomic-level data of the samples.

An alternative to this method is the use of nuclear magnetic resonance spectroscopy instead.[82–84] In NMR, nuclei of non-zero magnetic spin are first aligned with a strong magnetic field and then perturbed by the action of a secondary oscillating field whose frequency matches that of the excitation of a reference isotope. The radiation emitted by these nuclei can be recovered and processed to obtain structural information about the sample. Different models can be applied to transform the NMR spectrum of a sample into a 3-dimensional average structure of the peptides in the sample.

Fluorescence spectroscopy is another technique used for the study of peptides, particularly to assess their binding to the membrane or their aggregation. Fluorescence is a phenomenon by which substances absorb electromagnetic radiation at a specific wavelength and emit part of the energy at a different wavelength. The fluorescent nature of the amino acids themselves, such as tryptophan (Trp, W), or of some modified groups,[85] can be used to identify when the peptide has adhered to the lipid membrane and can also identify how this interaction has occurred.[86] Alternatively, peptides can be covalently labeled with fluorescent tags to facilitate localization or kinetic measurements. However, such labeling must be interpreted with caution, as the attached fluorophore can alter the physicochemical properties of the peptide and, consequently, its interaction with the membrane.[87] The loading of vesicles with fluorophores is a widely used approach to quantify molecular leakage due to the action of AMPs targeting such structures.[88–90]

Atomic force microscopy (AFM) can be used to characterize the formation of pores in the lipid bilayer. In this approach, the topology of a structure can be

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

determined by a change in the depth a probe goes while scanning a surface: a pore in the membrane will allow the tip of this probe to penetrate deeper when compared with the thickness of the membrane.[91] Changes in the penetration of the tip are recorded by a change in the interference pattern of a laser that is reflected in the reverse of the probe. While useful in characterizing the defects induced by AMPs in the membrane—including determining the mechanism of membrane lysis—, it requires the bilayer to be distributed along a flat surface, which are not the conditions at which peptides interact with vesicles in solution.

The binding of a peptide to a membrane will generate an enthalpy change that can be recorded using calorimetry studies. The obtaining of thermodynamic quantities associated with the affinity of a peptide for a membrane can be exploited to find the better candidate to target a certain membrane and also to determine the selectivity of a certain peptide towards membranes in general, or to identify the formation of pores in vesicles, to mention some examples.[92,93]

This non-exhaustive relation of experimental techniques illustrates different ways in which the interaction of peptides and lipid membranes can be studied. Overall, none of these methods can provide time-resolved atomic-level insights about the interaction between peptides and lipid membranes. Although the characterization of the systems using different techniques is a better alternative, as it can be used to complement the weaknesses of others, the latter is still a major practical limitation, impossible to overcome with current techniques.

1.4 Simulation approaches to study peptide–membrane interactions

Unlike with experimental techniques, the simulation of systems involving peptides and lipid bilayers can provide atomistic, time-resolved trajectories of all particles in the system, which allows the study of peptide adsorption and folding at a level of detail beyond what is achieved experimentally. Due to the size of these systems—assuming at least one peptide, tens or hundreds of lipids and solvent—, and the timescales of adsorption and folding processes, most *ab initio* simulation methods are not of practical use.[94] For this reason, classical Molecular Dynamics (MD) has been widely used for the study of the interaction between peptides and membranes.

A more detailed description of MD is provided in the Methods chapter. Briefly, MD is a classical simulation technique—meaning that no explicit quantum treatment is included—in which the motion of all particles in a system is computed by integrating Newton's equations of motion numerically. When each particle represents an atom, MD provides an atomistic, time-resolved description of the structural and dynamical behavior of the system.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

MD relies on an empirical potential energy function (force field) that specifies both the analytical form of bonded and non-bonded interactions and the numerical parameters associated with each particle type. The choice and quality of the force field are critical for obtaining realistic behavior. Some force fields are tailored to specific classes of molecules—e.g. proteins, lipids—, whereas others are designed for broader transferability; in practice, there is a trade-off between system-specific accuracy and general applicability. The parameters are typically selected to reproduce experimental observable—e.g. densities, heats of vaporization, conformational preferences—, and/or quantum-mechanical data, within the constraints imposed by the chosen functional form. Because parametrization and validation are non-trivial, extensive benchmarking is required; if the selected force field does not capture the relevant interactions for the system of interest, the resulting trajectories will not be representative of the physical system.[95]

For the specific study of peptides through simulation, the simplest setup is that of a single molecule in a homogeneous solvent, either water or any other.[96–99] These studies are mostly focused in the conformational changes of the peptide in these environments, but this is not the only application possible.[100] Solvents in any simulation can be explicitly represented as a collection of molecules in interaction with the peptide, but also among them; is this last aspect in which most of the computing time is expended, although these interactions are not usually of interest for the final analysis. For this reason, an alternative is to represent the contribution of the solvent as an implicit contribution to the Hamiltonian of the system, being only the solute defined by a set of numerical coordinates.[101–103] This approach was especially popular in the early days of MD, when computers did not have the same power as today; nowadays it is easy to obtain a reasonably long simulation of a peptide in an explicit solvent using a moderate amount of resources.

The same strategy can be extended to systems that also include a lipid membrane: it is possible to model the membrane as a region with a certain electric permittivity, different from that of the solvent,[104–108] or to generate a structure where the individual lipids appear explicitly.[109–114] The limitations of the implicit treatment are the same as for the solvent: not having the explicit molecules means that the simulation is not able to incorporate the specific interaction between functional groups of the peptide and lipids—or solvent—; on top of that, while the lipids could form domains in an explicit simulation, this will not be observed in the case of an implicit membrane. Although useful, the result of the simulation could be very different from those when an explicit representation of the molecules is used instead.

The main limitation of fully atomistic MD simulations—where peptides, lipids, solvent, and ions are all explicitly represented—is system size. While this type of simulation should in principle yield better results, as it is closer to the physical real-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ity, having to calculate the interactions between too many molecules will hinder the performance of the calculation, to the point that some slow phenomena—e.g. folding or adsorption[115,116]—could not be observed in a reasonable time frame. The size and duration of the simulations depend on the available computational resources: earlier examples of MD simulations of these systems contained only a few tens of lipids and some thousands of water molecules.[117–121] The limited size of the membranes also conditioned their complexity. Similarly, the length of these simulations typically spanned up to a few nanoseconds, much below the timescales at which many relevant phenomena—again, folding or adsorption—occur.[122] Over the years, the increase in (super)computing power has allowed for the simulation of larger systems and for longer timescales: it is now possible to simulate several peptides in interaction with more complex membranes, constituted by hundreds of lipids—and other molecules—, achieving timescales in the order of the microseconds.[110,123–126] Nevertheless, both the size and time extent of MD simulations remain several orders of magnitude below the experimental ones.

These type of systems have been widely simulated, mainly to understand peptide adsorption, aggregation, conformational changes, and their effects on membrane structure.[112,121,123,127–129] There are, however, some difficulties and limitations that typical MD simulations cannot overcome: Observing a given event often requires that the initial configuration already lies near the relevant region of conformational space; otherwise, the transition may not occur within the simulated timescale. This is a consequence of the complex energy surface of these systems, where some regions of the phase space may not be directly accessible unless the simulation is guided in some way. For example, it is not uncommon to find in the literature works in which peptides are simulated at temperatures above the physiological.[95] At higher temperatures, the extra kinetic energy allows the peptide to overcome energy barriers, such as those in the membrane, enhancing the dynamics of the system and allowing for a wider sampling of the conformational space than in a "colder" simulation. Such simulations are not inherently incorrect,* but stable conformations identified under non-physiological temperatures may not coincide with those at physiological conditions.[130–134]

To overcome the limitations of MD in the sampling of the phase space, a collection of alternatives labeled enhanced sampling methods has been developed over the years. The foundation of these techniques is still that of MD, but they include some modifications in order to improve the sampling efficiency within relevant regions of the conformational space. Among these methodologies, we can cite replica-exchange MD, metadynamics, umbrella sampling, or accelerated MD.[135–137] Although these methodologies do indeed allow for the exploration of the energetic profile in a way

*It must be tested whether the formulas in the force field hold at temperatures far from the ones it was designed for.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

standard MD could not, this comes to a cost: the amount of resources needed is very large, trajectories are either not continuous or biased, which means that their use beyond sampling is heavily limited.[129,138,139]. At a different conceptual level, techniques such as hydrogen mass repartitioning[140] or hydrogen isotope exchange[141] methods can be used to accelerate the simulation and reach longer timescales by allowing larger integration steps through the slowing down of internal degrees of freedom.

A coarse-grained (CG) representation can also improve sampling indirectly by reducing the number of degrees of freedom in the system. This method is described in more detail in Section 3.1.2.3. In CG simulations, groups of neighboring atoms that suffer from fewer variations along a simulation are mapped into a single particle; the CG force field must then be parameterized in a way that the interaction between each of the particles is close to the mean interaction between their components. A molecule in CG will therefore require fewer particles to be described compared with its atomistic counterpart, which implies fewer calculations to be performed. This has a direct impact on the performance of the simulation; each integration step is computationally cheaper, allowing longer timescales to be reached with the same resources.[142] A consequence of the reduction of the number of degrees of freedom is the smoothing of the free energy surface, so that minima that could be found in an atomistic simulation may no longer be accessible.[143] It must also be taken into account how the resolution may limit the phenomena the simulation is capable of describing: for example, the way in which Martini* works is by performing roughly a four-to-one mapping between the heavy atoms in the system and the particles in the simulation. Light hydrogen atoms are mapped together with the heavier atoms they are bonded to, so in a Martini simulation no hydrogen atom appears and, consequently, explicit hydrogen bonding cannot be represented—although some of its effects can be incorporated through elastic network potentials. As mentioned earlier, in the discussion of the secondary structure of peptides, the formation of H-bonds is a key factor for the conformational changes and the stabilization of peptide structures. The loss of the ability to describe H-bonds means that it is not possible to accurately describe peptide folding. However, CG force fields are still useful and can give rise to relatively accurate results despite such limitations.[144–146]

1.5 Rational and data-driven design of antimicrobial peptides

The general physicochemical features of AMPs, such as cationic charge, amphiphilicity, and high hydrophobic content, are necessary but not sufficient to grant antimi-

*Described in more detail in Section 3.1.2.5.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

crobial activity. The design of new AMP sequences remains challenging, as there is no comprehensive framework to reliably predict antimicrobial potency from sequence alone. Although experimental and computational techniques can characterize peptide properties and activity, each candidate must still be synthesized and tested individually.

The first and primary source of AMP sequences are living organisms. As it was already mentioned, AMPs are part of the innate immune system of all organisms, and it was in them when they were first discovered.[147] This is still a path currently exploited in AMP research, where the proteome of a plethora of organisms is screened to find new, better candidates.[148–152]

Another widely used procedure consists in the mutation of an already known AMP sequence with the aim of improving any of its properties, from an enhanced therapeutic effect to a reduction in its toxicity. One must keep in mind the vast combinations of mutations that can be introduced in a sequence—19 possibilities per residue position, if using only the standard 20 proteinogenic amino acids—, but usually these mutations attend to particular changes in a few localized residues. For example, a common strategy consists in the substitution of residues by others with higher hydrophobicity to increase the interaction with the membrane or to increase the overall charge of the peptide, for the same purpose.[153] In other cases, researchers look for shorter sequences—easier to produce and cheaper—that preserve most of the therapeutic ability of their longer precursors.[154,155]

However, in recent years there has been a strong push toward the development of data-driven models that can filter sequences based on their potential antimicrobial activity. Databases compiling thousands of characterized peptides—along with their physicochemical properties, target organisms, activity levels, and structural information—[156–163] provide the foundation for training machine learning (ML) models of various architectures: random forest, support vector machines, deep neural networks, etc. These algorithms can estimate the likelihood that a given sequence exhibits antimicrobial activity, allowing rapid *in silico* screening before experimental validation.[164–167]

Apart from the potential activity of a sequence, prediction of the structure of peptides is a key aspect toward the characterization of their interaction with the lipid membrane. The development of tools such as AlphaFold2[168] has enabled the accurate prediction of protein structures from sequence alone. However, these algorithms were trained mainly on soluble, globular proteins and implicitly assume that each sequence adopts a single, well-defined fold under physiological conditions. This assumption breaks down for many membrane-targeted AMPs, whose structure is highly dependent on the physicochemical environment, especially the α-helices, which emerge only upon the interaction with lipid membranes—or membrane-mimetic media. Consequently, the structures predicted by AlphaFold2—

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

or similar models—cannot capture the environment-dependent folding transitions that are central to their mechanism of action.[169]

Despite their usefulness, data-driven approaches face important limitations. Most AMP databases contain only sequence-level information, and rarely include mechanistic data on peptide–membrane interactions. In addition, variability in experimental conditions across studies complicates reproducibility and reduces the reliability of trained models. Therefore, while ML methods can efficiently identify promising candidates, achieving a more physically grounded understanding of AMP activity will require complementary strategies capable of capturing the molecular processes underlying their function.

1.6 Integrating Molecular Dynamics data into predictive modeling

Most data-driven approaches for predicting molecular or biological properties rely on curated bioinformatic databases that compile sequence, structural, and activity information derived from experimental studies.[165,170–173] However, these repositories rarely capture microscopic processes that determine molecular activity or selectivity. MD simulations can address this limitation by providing atomistic, time-resolved data under well-defined and reproducible conditions, yielding quantitative descriptors—such as conformational flexibility, interaction energies, or diffusive behavior—of the peptide–membrane interaction.

Despite their potential, the direct use of MD-derived data to build predictive models is still limited. Most existing algorithms are trained exclusively on experimental or sequence-based data, whereas only a few exploit simulation outputs to infer molecular properties beyond the systems explicitly simulated.[174] In principle, MD data sets could serve as a consistent, high-resolution source of physically grounded information to train models capable of predicting stability, binding affinity, or membrane interaction propensity across related molecular systems.

Conversely, much greater effort has focused on applying machine learning to improve MD itself, either by developing ML-based force fields that reach higher-level accuracy,[175] enhancing sampling efficiency,[176] or automating the analysis of complex trajectories.[177] However, the inverse direction, where MD provides the data used to train models with predictive power outside the simulation, remains relatively unexplored.

In the specific case of peptide systems, several studies have begun to bridge this gap. Batra and coworkers[178] used CG simulations to extract information about peptide aggregation to find new sequences with improved self-assembly properties. Xu *et al.*[179] conducted similar work, where they combined experimental and CG

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

simulations to design new peptide hydrogels. There are examples of applications in which MD data are used to predict the structure of cyclic peptides without having to perform new simulations: using this method, Hui, Descoteaux, Miao and Lin[180] were able to generate accurate structures for penta-, hexa-, and heptapeptides. Van Teijlingen *et al.*[181] developed a protocol for the design of new pore-forming octapeptides that uses simulation data to improve the sequence-generating algorithm: after generating a set of candidates, their effect on a model lipid bilayer is tested in simulation. This data is then fed again into the model, which will generate new candidates. After 7 iterations, they synthesized the peptides and validated their ability to form pores.

Although these hybrid approaches remain system-specific and often lack extensive validation,[182] they illustrate the potential of integrating the mechanistic insight of MD with the generalization capability of data-driven methods. This synergy represents a promising direction toward developing more physically grounded predictive tools and accelerating the rational design of molecular systems.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 2.

Objectives and motivation

I admire your curiosity, friend, Let's find out together.
Solanum, Outer Wilds

Despite decades of research, the study of peptide–membrane interactions remains far from complete. In particular, there is no theoretical framework capable of reliably predicting either the antimicrobial activity of a peptide or its conformation upon binding to a lipid bilayer.

Experimental and computational methods provide valuable insights into these systems, but also present important limitations. On the experimental side, most techniques either require perturbing the systems or their conditions—e.g. crystallography or AFM—, rely on the averaged effect over a large number of molecules—as in NMR, or in CD spectroscopy—, or do not achieve an atomic level resolution for the entire system, as occurs with fluorescence approaches. On the computational side, systems that involve peptides, lipid bilayers and a surrounding solvent are still beyond the size and timescales accessible to *ab initio* simulations; classical Molecular Dynamics simulations have stood out as the better alternative, but their predictive capacity is still limited.

It is in this context that we aim to deepen our understanding of the interaction of peptides and lipid membranes through the study of Molecular Dynamics simulations, with a special emphasis on membrane-targeting antimicrobial peptides. Throughout this work, we employed different levels of resolution—atomistic and coarse-grained—, together with statistical analysis and data-driven approaches. The ultimate goal is to capture the structural changes that peptides undergo in the membrane environment, as well as to identify numerical descriptors that could connect the peptide–membrane interaction with their potential antimicrobial activity.

The work carried out for this purpose can be compartmentalized into three more specific objectives, each of them associated with one of the sections in the Results chapter.

Structural characterization of membrane-targeting antimicrobial peptides at the water–lipid membrane interface Our first objective is to explore and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

characterize the conformational behavior and dynamical properties of membrane-targeting peptides at the water–lipid interface. We will use for this purpose a set of atomistic MD simulations. These simulations will allow us to monitor the adsorption of the peptide into the surface of lipid bilayers, together with the evolution of its orientation and the structural rearrangement of the peptide when it is in contact with the lipids at an atomic detail.

We will pay special attention to how the force field, the initial configuration, and a specific enhanced sampling strategy—hydrogen isotope exchange—, influence the outcome of the simulation.

The aim of this study is to determine the feasibility of using atomistic MD simulations to conduct systematic studies of both the structural changes and the antimicrobial activity of peptides.

Systematic simulation of peptide and membrane systems and database for the study of membrane-targeting antimicrobial peptides The second specific objective of this project is to design and construct a standardized framework for the simulation of peptide–membrane systems. This is a key step toward the design of new peptides with enhanced therapeutic activity that use the interaction between peptides and membranes to rationally filter or generate candidates.

The construction of such resource implies both the development of specific code to prepare the systems for simulation, a simulation protocol that could be easily modified to adapt to future development needs, and a systematic and automated analysis of the simulation results that provides a complete picture of the interaction between the peptides and the lipid membranes.

The results of these efforts will be consolidated in a dedicated database, considered a resource for comparative analysis of peptides with and without antimicrobial activity, and under different conditions. This database will serve as a living repository, integrating raw simulation data with some processed descriptors. Its structure will enable future extensions, such as multi-peptide systems, alternative membrane models, or different resolutions. In doing so, it will provide the necessary foundation for hypothesis and data-driven approaches in peptide design.

Connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches The last specific objective consists in exploiting the aforementioned database to study whether the descriptors of the interaction between the peptide and model lipid membranes derived from MD simulations could be linked to antimicrobial activity, so they could be used for predictive modeling.

For this purpose, we will extract a comprehensive collection of numerical descriptors characterizing the interaction between peptides and different membrane

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

models. We will conduct a statistical analysis of these magnitudes, with special emphasis on the identification of systematic differences between the simulations of sequences with reported antimicrobial activity and those lacking it. With these results, we will train and measure the performance of simple classification models to elucidate the potential use of this resource in the development of predictive models of the antimicrobial activity aware of the specific peptide–membrane interactions.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 3.

Methods

The model will be useful as we continue monitoring Giant's Deep.
Also I very much wanted to make a model.
Spire, Outer Wilds

Having introduced the objectives of this work, the present chapter will provide a detailed description of the simulation methods used for the study of peptide–membrane systems, as well as the various analyses performed on the generated trajectories.

The chapter will first present an explanation of the fundamentals of MD, including the most relevant algorithms needed to perform the simulations that make up this work. This will be followed by a description of the systems studied and the protocols for their assembly, equilibration, and simulation. The relation of analyses performed on the simulation data will be outlined, followed by an explanation of the statistical methods applied to the results of such analyses to draw meaningful conclusions about the events in the simulations.

3.1 Molecular Dynamics simulations

The N-body problem has been widely studied since Isaac Newton first introduced it in the 17th century in the context of astronomy, but there is no general analytical solution for $N \geq 3$. The same applies to any system of N interacting particles in the molecular context.

The name of Molecular Dynamics (MD) refers to a series of computational simulation techniques for the study of the evolution of a system of interacting particles by means of numerical methods. In their 1957 pioneering work, Adler and Wainwright[183] simulated the interaction of just over a hundred particles with associated attractive square-well potentials to study the phase transition in a system of hard spheres. Now, due to the increase in computational power, it is possible to simulate hundreds of thousands of atoms at long timescales.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The term MD can encompass a variety of resolution and theory level approaches; however, it usually refers to *classical* MD. Within this framework, molecular systems are treated as purely classical systems, at most trying to incorporate more fundamental interactions without an explicit quantum treatment. The method is conceptually simple, as it consists in solving Newton's equations of motion

$$\dot{\mathbf{p}}_i = \mathbf{f}_i \,, \tag{3.1}$$

for a system of N particles. In Eq. 3.1, $\mathbf{p}_i$ refers to the momentum of the i-th particle in the system, whose position is $\mathbf{x}_i$; and $\mathbf{f}_i$ represents the total force acting on it as a result of the interaction with all other particles and also any external contribution. The notation of n dots over a magnitude indicates a derivative of order n with respect to time, so

$$\begin{cases} m_i \dot{\mathbf{x}}_i(t) = \mathbf{p}_i(t) & \text{(3.2a)} \\ m_i \ddot{\mathbf{x}}_i(t) = \dot{\mathbf{p}}_i(t) & \text{(3.2b)} \end{cases}$$

where m_i is the mass of the particle. We will see below how to numerically solve this equation to approximate the trajectory of all particles by evaluating their position and momenta in discrete time instants.

This classical approach enormously simplifies the calculations to be performed and allows one to reach larger timescales and simulate larger systems compared to methods that include a quantum treatment. Methodologies in which the whole system is evaluated at the quantum mechanical level are called *ab initio* Molecular Dynamics methods, such as Carr-Parrinello MD.[184] Having to include the dynamics of all the electrons of all atoms of the system restricts their applicability to systems of the order of a hundred atoms and trajectories up to a few picoseconds.[94]

Alternatively, one could study only a certain part of the system at a quantum level, while the rest is treated classically. This approach is called QM/MM.[185] Phenomena involving significant changes in the wavefunction, poorly described in classical MD, can then be studied for smaller specific regions of systems otherwise too large to be handled by *ab initio* methods, as in the case of enzymatic reactions.[186–188]

A simplified standard protocol for an MD simulation is given in Figure 3.1. The following sections will detail how each of the key steps is implemented, with emphasis in the specific methodologies and algorithms necessary to carry out the simulations designed for the present work.

3.1.1 The method

In the case of conservative forces, those that depend only on the position of the particle, there exists a potential U_i such as

$$\mathbf{f}_i = -\nabla U_i \,, \tag{3.3}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

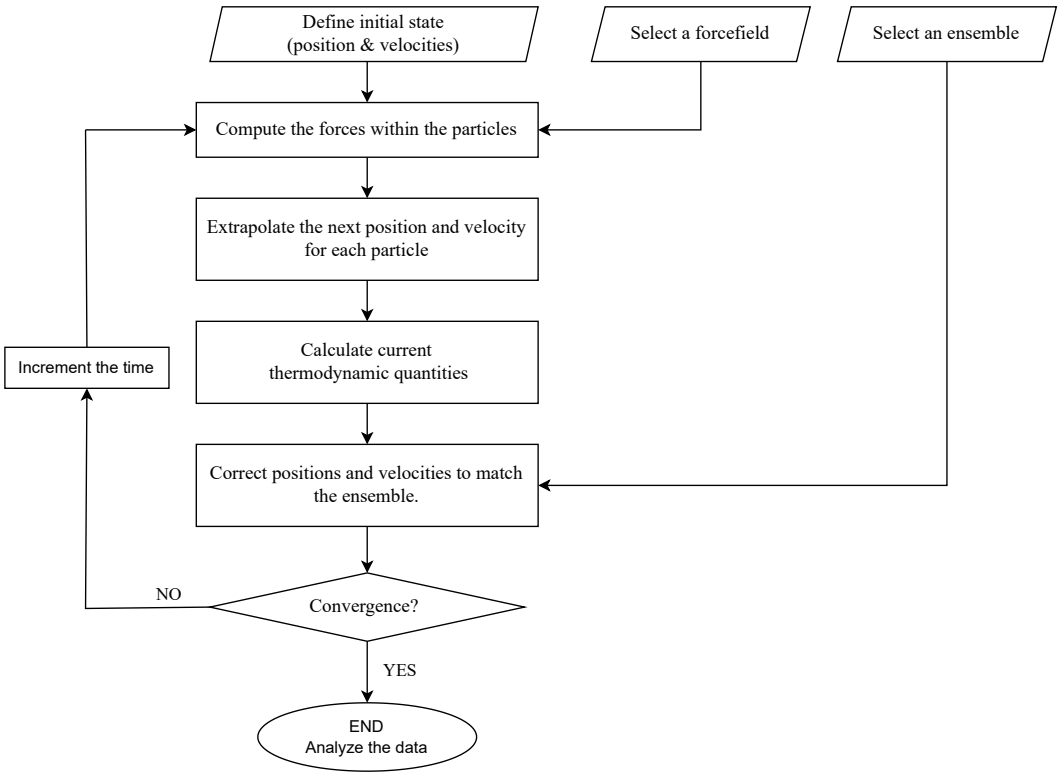


Figure 3.1: Workflow for an MD simulation. This simplified scheme summarizes the key steps in an MD simulation. Initially, the user must define the initial state of the system, together with the force field—with which interactions are computed—and the thermodynamic ensemble. Then, after evaluating the forces acting on the particles one can extrapolate the next position and velocity for each of the particles. Depending on the selected thermodynamic ensemble, these positions and velocities should be further modified to keep constant the temperature and/or pressure of the system. This process is repeated until some convergence criteria is achieved or until some predetermined maximum simulation time is reached, after which simulation data can be analyzed.

where ∇ is the operator nabla. Eq. 3.1 can then be rewritten as

$$\dot{\mathbf{p}}_i = -\nabla_i U \,. \tag{3.4}$$

Here U is the total potential, the sum of all the different interactions within the system, and ∇_i the derivative with respect to the coordinates of the i-th particle only (Figure 3.2).

Thus, for solving Eq. 3.4 one would need to know (**i**) the initial position and momenta of the particles in the system* and ((**ii**) the expression of its total potential.

*As Eq. 3.4 is a second-order differential equation in the position of the particles, an initial value for the position and also for its first derivative are needed.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

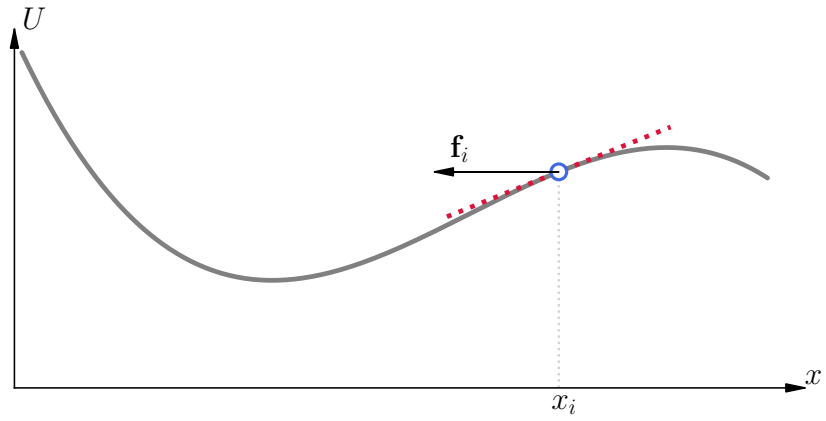


Figure 3.2: Particle in an external potential. The force (black arrow) acting on a particle is proportional to the slope (red dashed line) of the external potential (gray line) in the current position. All diagrams in this section, unless otherwise indicated, have been generated using Matplotlib.[8]

Defining the initial positions and momenta is easy, while doing the same for the potential is not trivial: It must faithfully reproduce the properties of the system while being simple enough for the calculation of the interactions to be feasible in a reasonable computing time. In classical MD, the potential is defined through a collection of parametrized functions whose form and values must be tailored to the system. The whole set of these functions and the values of the parameters used to describe a system are called **force field** (FF).

Integration of the equations of motion Upon the definition of a certain force field that allows to compute the potential of the whole system U, as well as an initial configuration of position and momenta, only a numerical method with which Eq. 3.4 can be solved would be needed to obtain the time evolution of the system. In most approaches, the state of the system is periodically evaluated at instants of time separated by a quantity Δt, and then extrapolated at $t + \Delta t$; this integration step is called the **time step**. Among the multiple algorithms available to solve this equation, simulations executed for this work used the **leapfrog integrator**, which is described by Eqs. 3.5.

$$\begin{cases} \mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \mathbf{p}_i(t + \Delta t/2)\dfrac{\Delta t}{m_i} & \text{(3.5a)} \\ \mathbf{p}_i(t + \Delta t/2) = \mathbf{p}_i(t - \Delta t/2) - \nabla_i U(t)\Delta t & \text{(3.5b)} \end{cases}$$

For a system where all forces are conservative, the potential depends only on the position of each particle, so will the force and hence its acceleration (Eq. 3.1).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

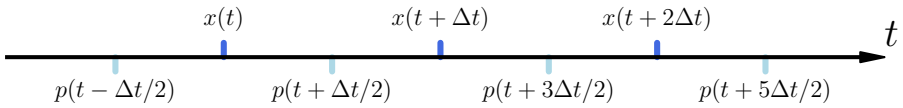


Figure 3.3: The leapfrog integrator. Both positions and momenta are evaluated after a time step Δt, but with a delay of $\Delta t/2$ between them.

The leapfrog algorithm integrates Eq. 3.4 based on a linear approximation of the derivatives in Eq. 3.2. It should be noted that position and momenta are evaluated with a delay of half a time step (Figure 3.3), and also depend on the value of the other in their previous iteration: The position explicitly depends on the previous momentum, while the momentum depends on the potential evaluated in the previous position of the particle (Eq. 3.3). Despite being so simple, it is numerically very stable for the systems under consideration and is one of the standards in MD.

Periodic boundary conditions The number of particles for which the system can be solved using a realistic amount of computational resources is limited. Simulating a finite-size system with a constant number of particles N implies, in principle, defining a volume where particles interact (simulation cell or region), but also a boundary to prevent particle leakage. An interaction between the particles and a wall is not expected in simulations that aim to model the behavior of a system in solution, far from the walls of the container.

Alternatively, one could avoid the interaction with the walls and also prevent the loss of mass by adding the same number of particles that exit the system, with the exact same properties, emerging from the opposite side of the simulation volume. This approach is equivalent to replicate the system in all directions of space, creating a periodic lattice where each element is exactly equal to the original simulation cell (Figure 3.4); when applied, the system is said to be under periodic boundary conditions (PBCs). The number of particles in the system is kept constant and no additional interactions to account for the walls of an artificial container are required; yet, the size of the simulation cell must be large enough to avoid any possible interaction between the different replicas of the same molecule in neighboring images, neither mediated by few intermediates, to avoid artifacts.

Choice of an appropriate time step Numerical methods introduce an error at each step of the calculation as a result of the approximations considered. This error cannot be avoided, but it can and must be minimized. As a second-order integrator, the truncation error for the leapfrog algorithm is $O(\Delta t^2)$. At the same time, the time

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

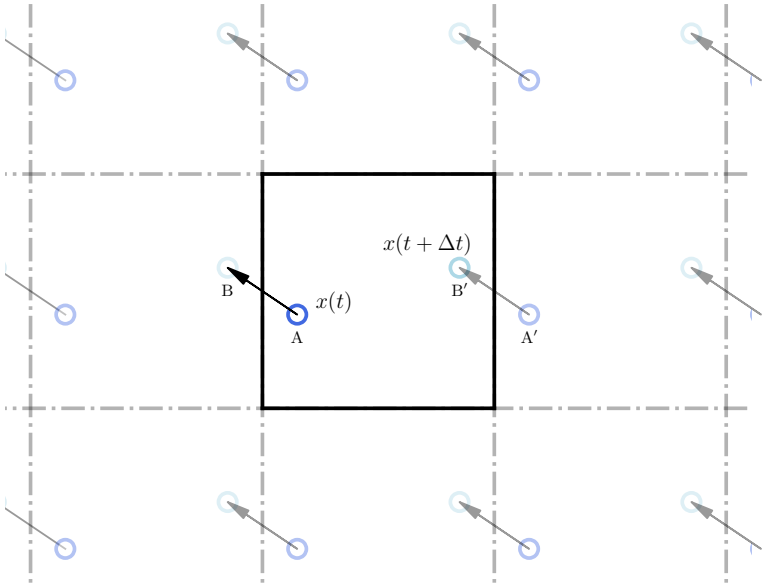


Figure 3.4: Effect of the periodic boundary conditions. When at a instant t the particle in the position A is displaced to position B at a time $t + \Delta t$, outside the original simulation cell (black solid line), a periodic image of the same particle, in the position A′ at time t emerges into the box in the position B′. The number of particles in the simulation cell (and in all periodic images) are then kept constant.

step must be low enough to accurately describe the movement of all particles in the system: the upper bound, according to the Nyquist-Shannon sampling theorem,[189] would be half the oscillation period of the fastest movement. In practice, however, it must be much smaller to compensate for the errors introduced by the integrator itself and ensure the stability of the simulation[190]; it is usually an order of magnitude smaller than this limiting frequency, or even less. In most biomolecular systems the fastest movement is the vibration of the C–H bond, whose wavenumber is about[191] 3000 cm^{-1}, so half its period would be 5 fs, but a time step of $\Delta t = 1$ fs is widely used.

Since the number of calculations required to reach a trajectory of a predetermined length is inversely proportional to the time step, and this parameter is restricted by the degrees of freedom with the highest frequency, different tricks and algorithms have been proposed to increase the period of the fastest movements. Some of them involve constraining some of the higher-frequency harmonic motions, effectively allowing the use of longer time steps. Namely, the SETTLE algorithm[192] was used to constrain water molecules, while the parallel implementation of the LINCS algorithm[193] was applied to constrain other bonds; details of when each of these algorithms was used are given in Section 3.2.1.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Energy minimization After introducing the molecules into the simulation box, by acting artificially on their coordinates, the system is likely to display a high potential energy, eventually leading to numerically unstable trajectories. In order to overcome this issue, a local minimization is typically performed. Different algorithms could be used for this aim. The **steepest descent** method consists of displacing the atoms in the direction of the maximum slope of the gradient:

$$\mathbf{x}_i \rightarrow \mathbf{x}_i - h\nabla_i U \tag{3.6}$$

where h is the step size. While for the practical implementation of this method more details must be considered, the fundamental workflow is the iterative application of the previous expression either for a fixed number of steps, until the maximum force is below a certain threshold, or until its change in consecutive steps is negligible. The step size h ensures the convergence of the method: If it was too large, the change in coordinates may skip the location of the minimum; if it was too small, the convergence may be too slow. In more refined approaches, the value of h would dynamically adapt its value to account for changes in the potential of the system.

3.1.2 Force fields

As previously defined, the term force field refers to the collection of equations and their parameters that aim to model the interaction between the particles within a system. The development of a new force field requires the selection of a form for the equations and the nature of the parameters employed. The goal can involve replicating a certain set of experimental properties, reproducing results from quantum chemical calculations, or both.

The development of new and better force fields is the focus of the research of many groups. Typically, force fields provide parameters for just a few molecules or groups of molecules. The parametrization of new molecules is far from trivial. All simulations performed for the present work made use of popular force fields for the simulation of biomolecular systems; Section 3.1.2.5 contains the detailed list.

3.1.2.1 The general structure of a force field

Regarding its structure, the different contributions to a force field can be separated in two main categories: those that model the interactions between bonded particles and those between particles that are not bonded. All the force fields employed in the present work follow this general structure:

$$U = \underbrace{U_{\text{bonds}} + U_{\text{angles}} + U_{\text{dihedrals}}}_{\text{Bonded interactions}} + \underbrace{U_{\text{vdW}} + U_{\text{Coulomb}}}_{\text{Non-bonded interactions}} \tag{3.7}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

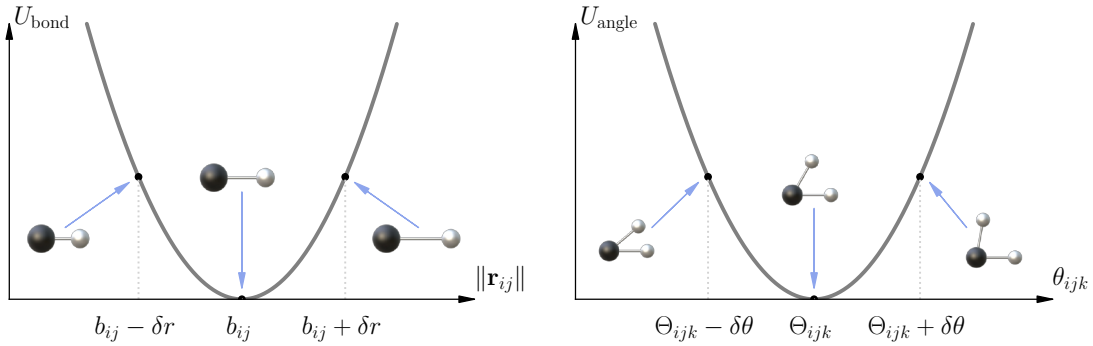


Figure 3.5: Bond and angle stretching. Harmonic potentials can be used to describe the oscillation of bond lengths (left) or angles (right) around an equilibrium position.

Bonds When the QM/MM method was presented, its application in the study of enzymatic reactions was mentioned because of their ability to incorporate changes in electron clouds that occur during the rupture or formation of a covalent bond. However, for many systems, it can be assumed that the bonds do not only do not break but also that they will remain close to their ground state. Thus, the most simple non-rigid approximation for the modeling of a bond between any two particles is a harmonic potential of the form

$$U_{\text{bond}}(\mathbf{r}_{ij}) = \frac{k^b_{ij}}{2} \left(\|\mathbf{r}_{ij}\| - b_{ij}\right)^2 , \tag{3.8}$$

where $\mathbf{r}_{ij}$ is the vector between the particle i and j, $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x_j}$, and $\|\mathbf{r}_{ij}\|$ is its norm; b_{ij} is the equilibrium distance between these same particles—the location of the minimum in the potential—and k^b_{ij} the force constant (Figure 3.5, left). While most of the force fields use the previous expression to model covalent bonds, others (e.g. GROMOS[194,195]) use

$$U_{\text{bond}}(\mathbf{r}_{ij}) = \frac{k^b_{ij}}{4} \left(\|\mathbf{r}_{ij}\|^2 - b^2_{ij}\right)^2 , \tag{3.9}$$

another harmonic potential but in the square of the distance between particles instead. In either of these two cases, the values of b_{ij} and k^b_{ij} can be tuned to modify the length of the bond in equilibrium and its rigidity, respectively; their value must be defined for any possible pair of bonded particles within the system.

Angles As with bonds, the angle between any three connected particles, i.e. those involved in two bonds with a common origin, is almost rigid in the ground state, fluctuating around its equilibrium value. Therefore, a harmonic potential seems also to be appropriate to describe such behavior

$$U_{\text{angle}}(\Theta_{ijk}) = \frac{k^a_{ijk}}{2} \left(\Theta_{ijk} - \theta_{ijk}\right)^2 , \tag{3.10}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where the angle defined by the particles i, j, k with reference in the second, Θ_{ijk}, can be computed as

$$\Theta_{ijk} = \arccos\left(\widehat{\mathbf{r}}_{ij} \cdot \widehat{\mathbf{r}}_{kj}\right) . \tag{3.11}$$

Here, $\widehat{\mathbf{r}}$ refers to the unit vector in the direction of $\mathbf{r}$, $\widehat{\mathbf{r}} = \mathbf{r}/\|\mathbf{r}\|$. The parameter k^a_{ijk} is again the force constant of the potential (Figure 3.5, right). Other force fields (again, GROMOS[194,195]) employ a function harmonic in the cosine of the angle instead

$$U_{\text{angle}}(\Theta_{ijk}) = \frac{k^a_{ijk}}{2}\left(\cos\Theta_{ijk} - \cos\theta_{ijk}\right)^2 . \tag{3.12}$$

Analogously to the potential for the bonds, θ_{ijk} defined the equilibrium angle between the triplet of particles and k^a_{ijk} its stiffness.

Dihedrals Dihedral angles are those formed between two planes, each of them defined by 3 particles, with two of them in common. If the four particles are consecutively connected, the formed angle is called a proper dihedral (Figure 3.6, lower left). In contrast, improper dihedrals are those in which all three of the particles are bound to the same fourth one. Improper dihedrals are defined to preserve the planarity of certain functional groups or to define and maintain their chirality (Figure 3.6, upper left). The most simple way to model improper dihedrals is a harmonic expression

$$U_{\text{improper}}(\phi_{ijkl}) = \frac{k^d_{ijkl}}{2}\left(\phi_{ijkl} - \varphi_{ijk}\right)^2 , \tag{3.13}$$

where the dihedral angle ϕ_{ijkl} between the planes defined by the triads of particles i, j, k and j, k, l is

$$\phi_{ijkl} = \arccos\left(\frac{(\widehat{\mathbf{r}}_{ij} \times \widehat{\mathbf{r}}_{kj})}{\|\widehat{\mathbf{r}}_{ij} \times \widehat{\mathbf{r}}_{kj}\|} \cdot \frac{(\widehat{\mathbf{r}}_{jk} \times \widehat{\mathbf{r}}_{kl})}{\|\widehat{\mathbf{r}}_{jk} \times \widehat{\mathbf{r}}_{kl}\|}\right) . \tag{3.14}$$

The term k^d_{ijkl} is the strength of the potential to act at the dihedral angle.

Proper dihedrals, on the other hand, must be constructed in a way that all minima and also the height of the rotational barriers are well reproduced. It must also be taken into account that the total shape of the potential term must be periodic, so $U_{\text{proper}}(\phi_{ijkl}) = U_{\text{proper}}(\phi_{ijkl} + 2\pi)$. A simple potential able to model n evenly separated minima in a proper dihedral, which could be, for example, the cis and trans conformations if $n = 2$, is

$$U_{\text{proper}}(\phi_{ijkl}) = k^d_{ijkl}\left(1 + \cos\left(n\phi_{ijkl} - \varphi_{ijk}\right)\right) . \tag{3.15}$$

A maximum in the potential appears at φ_{ijk}/n, whose height is $U_{\text{proper}}(\varphi_{ijk}/n) = 2k^d_{ijkl}$. A description of more complex potentials can be achieved by combining different functions (Figure 3.6, right).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

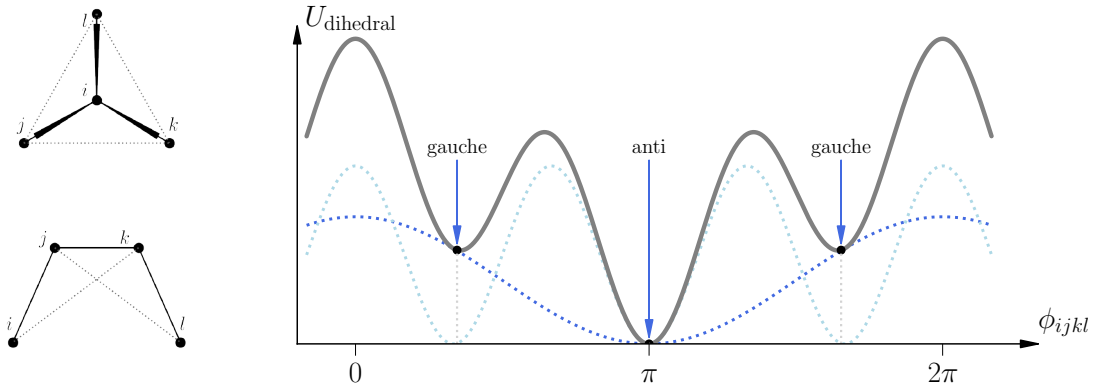


Figure 3.6: Improper and proper dihedral angles. (Upper left) Improper dihedrals are defined to keep the planarity or the chirality of groups, while (lower left) proper dihedrals are defined to represent the rotation around a bond; both correspond to the angle formed by the planes defined by the particles of indices i, j, k and j, k, l. (Right) A combination of periodic functions with the appropriate phases and frequencies can be used to describe the rotational barriers around a certain proper dihedral, so different rotamers can be reproduced.

Van der Waals interactions Particles still interact with each other when they are separated from each other by more than 3 covalent bonds and even when they belong to different molecules. Such interactions are usually described by two different contributions: the first refers to the electrostatic interaction between the partial charges between the particles as defined by the force field. All other interactions are commonly grouped into a single term, the van der Waals (vdW) interaction.

One of the most simple potentials to describe the van der Waals interaction is the Lennard-Jones potential: a combination of a very strong, short-range repulsive term modeling the Pauli repulsion between electron clouds, and a weaker long-range attractive contribution that aims to reproduce London dispersion forces.[196] The functional form of the Lennard-Jones potential is as follows:

$$U_{\text{vdW}}(\mathbf{r}_{ij}) = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{\|\mathbf{r}_{ij}\|}\right)^{12} - \left(\frac{\sigma_{ij}}{\|\mathbf{r}_{ij}\|}\right)^{6}\right], \tag{3.16}$$

The parameter ε_{ij} is the value of the potential at its minimum, while σ_{ij} is the distance at which the potential is zero (Figure 3.7). Instead of defining ε_{ij} and σ_{ij} for each possible pair of particles, several force fields establish a combination rule that is a direct function of the parameters ε_i and σ_i associated with independent particles, to account for the interaction between cross pairs. Despite its simplicity, the Lennard-Jones potential can be very accurate, making it one of the most used potentials for the description of these interactions.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

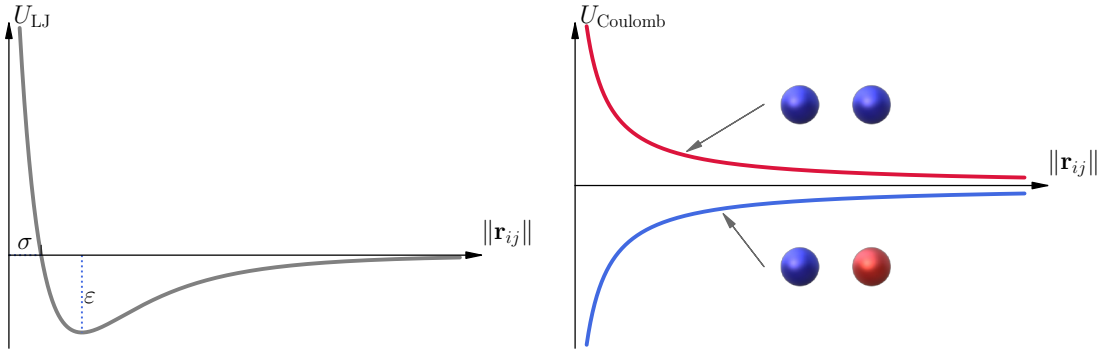


Figure 3.7: Non-bonded interactions in molecular systems. The Lennard-Jones potential (left) is widely used to represent the interatomic forces in MD simulations. While the Lennard-Jones potential combines a repulsive short-range interaction and an attractive long-range one, the Coulomb potential (right) is always attractive (blue) or repulsive (red), depending on the relative sign of the charges.

Electrostatics The last term corresponds to the direct electrostatic interaction between the partial charges of the particles, for which the classical Coulomb potential is used:

$$U_{\text{Coulomb}}\left(\mathbf{r}_{ij}\right) = \frac{q_i q_j}{4\pi\varepsilon\|\mathbf{r}_{ij}\|}\,, \tag{3.17}$$

The sign of the Coulomb potential depends on the charges of the particles q_i and q_j: positive for a repulsive potential when the sign of the charges is the same and negative for the attractive potential between opposite charges (Figure 3.7, right). The parameter $\varepsilon = \varepsilon_r\varepsilon_0$ is the permittivity of the medium, where ε_0 is the dielectric constant and ε_r the relative permittivity of the medium. The permittivity shields the interaction between charges; the effect is very relevant in an aqueous solution, as the electric permittivity of water is very high.

3.1.2.2 Calculation of non-bonded interactions

As previously described, bonded interactions involve groups of 2, 3 or 4 atoms, so they must be computed for each defined bond, angle or dihedral. Non-bonded interactions, however, exist between each and every pair of particles within the system. On top of that, there will also be an interaction between the particle and all the periodic images, including self-interaction terms. The calculation of the non-bonded interactions is a bottleneck in an MD simulation, as it scales with $\mathcal{O}(N^2)$, so different approaches can be followed aiming to reduce the computational cost.

Cutoff The most simple way in which one can deal with the practical calculation of the non-bonded interactions is to take advantage that they decrease with the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

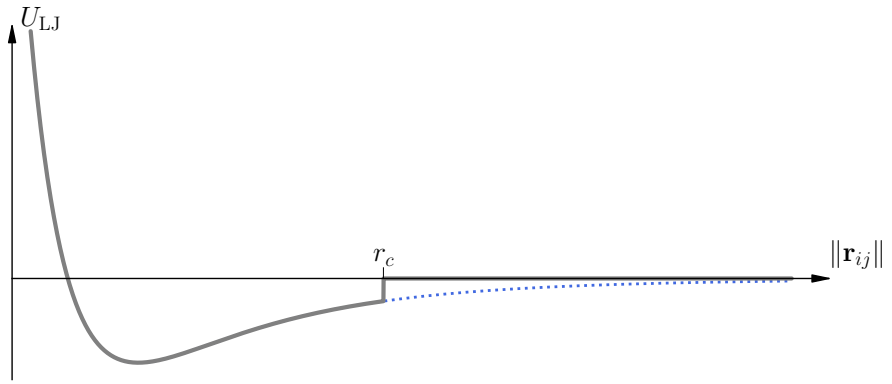


Figure 3.8: The use of a cutoff for the calculation of non-bonded interactions. Beyond a cutoff distance of r_c, the potential is assumed to be zero. To reduce the error, the cutoff must be several times the value of σ; in the figure, $r_c/\sigma \approx 1.55$ was selected so the effect is noticeable, but usually in MD $r_c/\sigma > 10$. Some corrections can be applied to shift the potential or the forces so that they become zero at r_c.

distance between particles (r^{-6} in the case of the Lennard-Jones potential, r^{-1} in the case of Coulomb) and set a cutoff beyond which the potential can be assumed to be negligible (Figure 3.8). A correction can then be applied so that the potential or the forces at the cutoff becomes zero, either by translating the potential (potential shift) or by transforming the potential or the forces (potential/force switch). The specific correction required to apply is associated with the force field selected and it is detailed in Section 3.1.2.5.

This approach, while acceptable for the van der Waals interaction, is poorer for the calculation of electrostatic interactions, where the decay is less steep, and better approaches can be used.

Reaction field Barker and Watts[197] proposed the reaction field scheme for the calculation of electrostatic interactions within a system: it is based on approximating the environment around a particle as a spherical cavity of radius r_c, within which the electrostatic interactions are treated explicitly; further than r_c, it is assumed that the medium is homogeneous with a relative dielectric constant $\varepsilon_{\mathrm{RF}}$. Then, Eq. 3.17 becomes

$$U_{\mathrm{RF}}(\mathbf{r}_{ij}) = \frac{1}{4\pi\varepsilon}\left[1 + \frac{\varepsilon_{\mathrm{RF}} - \varepsilon_r}{2\varepsilon_{\mathrm{RF}} + \varepsilon_r}\frac{\|\mathbf{r}_{ij}\|^3}{r_c^3}\right]\frac{q_i q_j}{\|\mathbf{r}_{ij}\|} - \frac{1}{4\pi\varepsilon}\frac{q_i q_j}{r_c}\frac{3\varepsilon_{RF}}{2\varepsilon_{\mathrm{RF}} + \varepsilon_r}\,. \tag{3.18}$$

At $\|\mathbf{r}_{ij}\| = r_c$ the slope of the potential, proportional to the force, becomes zero, so it is possible to just compute the interactions within the cutoff—first term in Eq. 3.18—and then add a contribution to the potential of the medium beyond

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

this shell—second term. In this scenario, a contribution accounting for the self-interaction should be added, which is equal in form to the second term but halved.

Particle-mesh Ewald Due to the longer range of electrostatic interactions, an alternative is to consider all electrostatic interactions, including also those with periodic images:

$$U_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon} \sum_{\mathbf{n}^*} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{\|\mathbf{r}_{ij} + \mathbf{n}\|} \,. \tag{3.19}$$

Here $\mathbf{n}$ is the lattice vector for the periodic images: each of its components is $n_i \mathbf{L}_i$, being $n_i \in \mathbb{N}$ the index of the periodic image in the direction defined by the i-th triclinic vector of the simulation cell $\mathbf{L}_i$. The asterisk indicates that all self-interaction terms that appear when $\mathbf{n} = 0$ should be excluded from the sums. As the error function and its complementary satisfy

$$\text{erf}(x) + \text{erfc}(x) = 1 \,, \tag{3.20}$$

Eq. 3.19 can be rewritten into short-range and long-range contributions,

$$U_{\text{Coulomb}} = \frac{1}{4\pi\varepsilon} \left[\underbrace{\sum_{\mathbf{n}^*} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j \text{erfc}\left(\beta \|\mathbf{r}_{ij} + \mathbf{n}\|\right)}{\|\mathbf{r}_{ij} + \mathbf{n}\|}}_{\text{Short range}} + \underbrace{\sum_{\mathbf{n}} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j \text{erf}\left(\beta \|\mathbf{r}_{ij} + \mathbf{n}\|\right)}{\|\mathbf{r}_{ij} + \mathbf{n}\|}}_{\text{Long range}} \right] , \tag{3.21}$$

as erfc(x) $\to 0$ as $x \to \infty$. The parameter β will determine how fast the short-range term converges. Ewald[198] proved that the long-range term can be transformed into a sum in the Fourier space, so the previous expression becomes

$$\begin{aligned} U_{\text{Coulomb}} =& \frac{1}{4\pi\varepsilon} \sum_{\mathbf{n}^*} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j \text{erfc}\left(\beta \|\mathbf{r}_{ij} + \mathbf{n}\|\right)}{\|\mathbf{r}_{ij} + \mathbf{n}\|} \\ &+ \frac{1}{4\pi\varepsilon} \frac{1}{\pi V} \sum_{\mathbf{m}*} \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j e^{-(\pi \mathbf{m}/\beta)^2} e^{i 2\pi \mathbf{m} \cdot \mathbf{r}_{ij}}}{\|\mathbf{m}\|^2} \\ &+ \frac{1}{4\pi\varepsilon} \frac{\beta}{\sqrt{\pi}} \sum_{i=1}^{N} q_i^2 \end{aligned} \tag{3.22}$$

The derivation of this formula can be found elsewhere.[199] By choosing an appropriate β, two short cutoffs can be selected: one for the direct space, which is the term in Eq. 3.22 that loops through $\mathbf{n}$, and the other for the reciprocal space, which is the sum involving $\mathbf{m}$ (Figure 3.9). The vector $\mathbf{m}$ is a vector in the reciprocal lattice

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

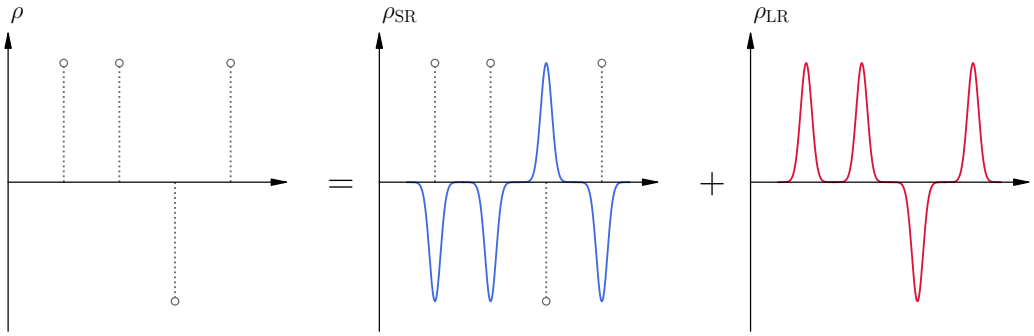


Figure 3.9: Charge distribution in the particle-mesh Ewald method. (Left) The charge density of point particles is that of a Dirac delta located in the position of the particle. The decomposition of the short and long-range contributions are equivalent to have (center) in the direct space the same distribution of Dirac-like charge density and an complementary Gaussian distribution, in the same position and with the same charge but opposite sign; and (right) Gaussian distribution of charges, with the same sign and magnitude as the original, in the reciprocal space . This treatment allows for a rapid convergence of the series, while keeping the periodicity of the system.

whose components are $m_i\mathbf{L}_i^*$, where $m_i \in \mathbb{N}$. $\mathbf{L}_i^*$ are the reciprocal primitive vectors, columns of $\left(\mathbf{L}^\mathsf{T}\right)^{-1}$ multiplied by 2π; here, $\mathbf{A}^\mathsf{T}$ stands for the transpose matrix, where $(\mathbf{A}^\mathsf{T})_{ij} = \mathbf{A}_{ji}$, while $\mathbf{A}^{-1}$ is the inverse of $\mathbf{A}$, so $\mathbf{A}\mathbf{A}^{-1} = \mathbb{1}$ the identity matrix. Eq. 3.22 is called the Ewald summation.

An improvement in terms of efficiency is achieved through the particle-mesh Ewald method (PME),[200] in which charges in the reciprocal space are evaluated in a grid, which is then transformed with a 3D fast Fourier transform (FFT) algorithm. The implementation of this algorithm used in our simulations is known as smooth PME, where the interpolation in the grid is done through cardinal B-splines.[201]

3.1.2.3 Levels of resolution

Up to this point, most of the previous text referred to "particles" and not "atoms", despite being focused on the simulation of molecular systems. The choice of this terminology was deliberate. Force fields and simulations in which the particles in the system represent the individual atoms of the different molecules present are called **atomistic** or, more specifically, **all atoms** (AA).

Yet, at no point did the aforementioned methods need this to be the case: for instance, hydrogen atoms in aliphatic chains are tightly bonded to carbon atoms and also do not play an important role in the description of the system beyond the addition of some mass and their contribution to the electronic shells. In fact, both the mass and the partial charges associated with such atoms are much smaller than

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

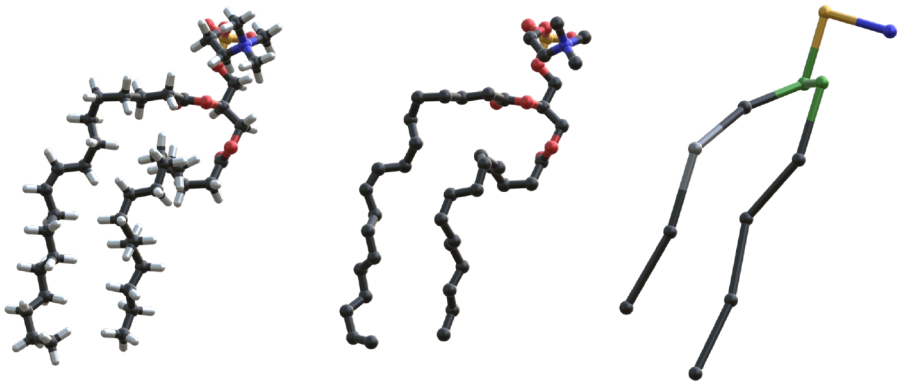


Figure 3.10: Representation of a molecule using different levels of resolution. The same lipid, 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine, is represented as it would appear in an all-atoms simulation (left), in an united-atoms simulation (center), where all nonpolar hydrogen atoms are merge with the heavier atoms they are bonded to; and in a coarse-grained (right) simulation. Here the particles represent groups of several atoms. The number of particles used to model the molecule is reduced from 134, to 54 and 12, respectively. Structures were rendered using Blender.

those in C atoms. So, one would be tempted to merge them with the C atoms to which they are bonded and associate the CH, CH_2 and CH_3 groups to a single particle each, reducing the number of degrees of freedom in the system for the sake of the performance of the calculation, without significantly altering the accuracy of the representation of the real system. This level of resolution is known as **united-atoms** (UA)*.

Following the same strategy, but going a step beyond, it is also possible to map to a single particle the properties of a collection of connected heavy atoms and all of their bonded hydrogen. The reduction of the number of degrees of freedom and hence of the number of calculations to perform is much higher than with UA. On top of that, since in such a representation all particles would be much heavier than any of the previous atoms, the upper bound of the time step, as discussed in Section 3.1.1, becomes much larger. Therefore, much larger systems can be studied and much longer timescales can be reached at the same computational cost. The payoff is a high loss in the resolution of the phenomena these simulations can describe: for example, the formation and breaking of hydrogen bonds (H-bonds), which are of great interest and importance in peptides and proteins, may no longer be studied

*Sometimes the term *atomistic* is used to refer to both AA and UA representations, as most of the particles in the latter still refer to recognizable atoms.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

under this approach. This way of modeling a system is known as a **coarse-grained** (CG) representation, and the name to refer to the particles in this case is the term *beads*. In Figure 3.10, the same molecule is represented at the three mentioned levels of resolution.

3.1.2.4 Water models

Water is the most abundant molecule in living organisms and will be so in our systems. The parametrization of water molecules is for this reason especially important and it is usually done separately from the rest of the molecules. These specific parameters for water are known as water models. Most force fields are parametrized based on some specific water model. The use of a different water model from the one used for the development of the force field should be tested to ensure that the simulation results are appropriate.

The simplicity of the structure of the water molecule, with only two H atoms bonded to one O, contrasts with the complexity of their properties. It is very difficult for a classical water model to reproduce all properties of water at the same time, so different water models focus on reproducing some aspects: density, transition temperatures, viscosity, compressibility, specific heat, surface tension, etc.[202] Some of them include extra virtual particles that only have a certain charge, which changes the charge distribution of the molecule and allows a better reproduction of the properties of the bulk. The water models used in the present work are 3- and 4-point models, a classification based on the number of particles used to define each molecule (Figure 3.11).*

3.1.2.5 Relation of force fields used

The simulations that constitute the present work have been carried out using different force fields. Here is an overview of them:

The AMBER force fields AMBER stands for "Assisted Model Builder with Energy Refinement", and refers to both a package for MD simulations and force fields for biomolecular simulations.

The AMBER force fields are grouped by types of molecules: there are force fields for proteins, for DNA, for RNA, for lipids, etc., and there are many versions available; yet, compatibility between all of them is not guaranteed. We used the latest version of the force fields for proteins and lipids, ff19SB[203] and Lipid21,[204] respectively, in combination with the 4-point OPC water model.[205] Although the original development of the AMBER force fields was done with the TIP3 water

*Here the Martini water models, which will be detailed below, were not included.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

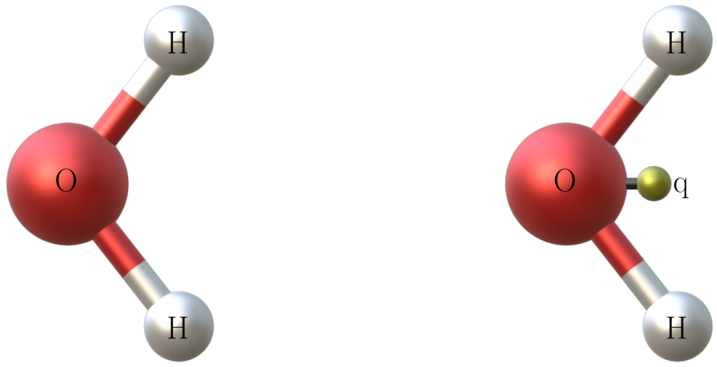


Figure 3.11: 3- and 4-point water models. The structure of the 3-point water molecule, in the left, is that of the real water molecule, one particle representing the oxygen atom, which is bonded to two different hydrogen atoms. However, the angle formed by the H–O–H bonds as well as the bond lengths can vary between models to better reproduce some properties of interest. The same applies for the 4-point water models, where an extra particle (labeled as "q" in the representation) is added, typically in the bisection of the H–O–H angle. The partial charge of the oxygen atom is displaced to this particle, aiming to better reproduce the distribution of charge in the molecule. Structures were rendered using Blender.

model,[206] in the same publication where ff19SB is reported, the use of OPC was tested and recommended. Since then, this combination has already been widely used in the literature.[207–213]

The CHARMM force fields CHARMM, an acronym for "Chemistry at Harvard Macromolecular Mechanics", is, as AMBER, both the name of a program for MD simulations and a set of force fields.

The latest version of the CHARMM force fields is CHARMM36m,[214] a slight modification of CHARMM36,[215] which includes parameters for amino acids, lipids, ions and some small organic molecules. This force field was used in combination with the modified version of the TIP3P water model.[216] Since the standard TIP3P will not be used at any point in the present work, there is no margin for ambiguity, and we will refer to this model as TIP3P without further specification.

The GROMOS force fields GROMOS (Groningen Molecular Simulation), once again refers to a simulation package (almost discontinued) and to the force fields for simulation of biomolecules. Unlike AMBER and CHARMM, GROMOS are UA

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

force fields, meaning that only polar hydrogen atoms are explicitly defined.

GROMOS force fields were originally parameterized using a twin-range cutoff, in which forces between particles are updated every single or every N time steps depending on their distance. This algorithm has been deprecated in earlier versions of GROMACS,[217] so simulations using this force field must use another method to treat non-bonded interactions, which gives different results from the intended.[218] Nonetheless, the parameters of its 54a7 release[195] for amino acids and ions, together with the CKP parameters[219,220] for lipids and the recommended SPC water model were used to perform part of the simulations.[221]

A proposed method to speed up MD simulations consists of mimicking the isotopic substitution process[141,222]: Replacement of ^{1}H by ^{2}H, specifically in polar hydrogens, is used experimentally for many purposes, such as the structural study of molecules through neutron scattering,[223] NMR,[224] or mass spectroscopy,[225] among other examples. Although isotopic substitution might affect some molecular properties,[226,227] the induced slowdown of the vibrational frequencies of H-containing bonds is very interesting from a computational point of view, as one could benefit from the use of a longer time step. To do so, all H atoms should be substituted for the heavier isotope, so its application to an UA force field, where only the polar H atoms are explicitly represented, seems more appropriate. The substitution of ^{1}H by ^{4}H, an isotope with a mass of 4.02643(11) u and a half-life of $1.39 \cdot 10^{-22}$ s,[228] is a generalization of the experimental deuteration, and has already been tested in cyclodextrins[222], transmembrane cyclic peptide nanotubes[141], and lipid bilayer systems.[229] No significant differences have been found with respect to the original systems, while a 3-fold increase in performance of the simulation is obtained, as it scales linearly with the change in the time step.

The simulations in which polar hydrogen had been replaced by ^{4}H are going to be referred to as GROMOS-HIE, as proposed by Conde-Torres *et al.*,[141] leaving GROMOS to label the unmodified cases.

The Martini force fields Martini are some of the most widely used CG force fields for the study of biomolecular systems. The Martini 2.2 force field[230] applies a four-to-one mapping between the real system and its representation, meaning that around 4 heavy atoms, together with all their bonded H, are modeled by a single bead (particle). This is less consistent in the case of the most recent version of Martini, Martini 3[231], where beads of different sizes were added to better represent smaller groups of 2 or 3 heavy atoms. Another exception involves water molecules: Martini usually represents 4 water molecules by a simple particle. An improved force field with a polarizable water model[232] based on Martini 2 that uses 3 particles instead is also available.

All three aforementioned versions of the Martini force field were used in this

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

work: MARTINI 2.2, MARTINI 2.2P (the modification of MARTINI 2.2 compatible with their polarizable water model), and MARTINI 3.

3.1.3 Molecular Dynamics beyond the *NVE* ensemble

As Newton's equations of motion (Eq. 3.4) conserve the total energy of the system E, and, by default, they are solved within a fixed simulation cell of volume V (with PBCs) for a fixed number of particles N, the aforementioned setup allow us to compute the evolution of the system in the NVE (microcanonical) ensemble, i.e. a closed and isolated system.

However, most processes in nature occur at a constant temperature T and/or pressure P, rather than at a fixed energy or volume. Therefore, most simulations of biomolecular systems in equilibrium are performed in the NVT or NPT ensembles. To sample them instead of NVE, different algorithms were proposed to conserve these quantities: algorithms that act on T are called **thermostatic algorithms** or **thermostats**, while those that act on pressure are called **barostatic algorithms** or **barostats**.

For sampling correctly the NVT or NPT ensembles is not enough to keep them at a fixed temperature and/or pressure, but also allowing fluctuations of these quantities to be compatible with the theoretical ones derived from statistical mechanics. Next, the thermostats and barostats that were used at any point of the simulation process will be introduced.

3.1.3.1 Thermostats

From a microscopic point of view, the temperature of a system of particles is a function of their kinetic energy. Being $\mathbf{p}_i$ are the momenta of a system of N particles of mass m_i, their kinetic energy is

$$K = \frac{1}{2}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\,. \tag{3.23}$$

According to the equipartition theorem, each degree of freedom contributes with $kT/2$ to the kinetic energy, with k the Boltzmann constant, so $K = N_f kT/2$, where N_f is the total number of degrees of freedom.[233] For a 3-dimensional system with l constraints, $N_f = 3N - l$. Therefore, the temperature of the particles is:

$$T = \frac{1}{kN_f}\sum_{i=1}^{N}\frac{\mathbf{p}_i^2}{m_i}\,, \tag{3.24}$$

So, modifying the velocities of the particles allows to control the total temperature. This is mostly done in two ways: by scaling all velocities to match the target temperature, or by making some particles interact with a thermal bath so that only

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

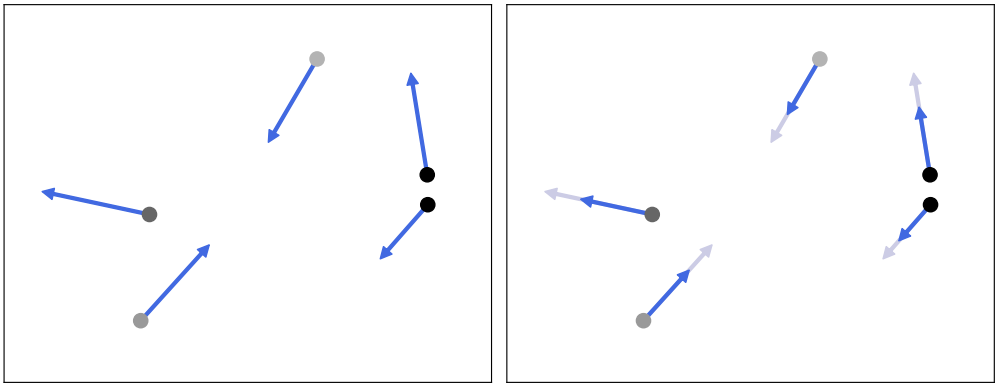


Figure 3.12: Fundamentals of the velocity-rescaling approach. In the left, a set of particles and their velocities (blue lines). The current temperature of the system T, derived from the kinetic energy of the particles, is higher than a target T_{ref}. The ratio of these temperatures has been found to be $\alpha = \sqrt{T_{ref}/T} = \sqrt{2/3}$, so multiplying each individual velocity by this factor leads to the system in the right, where the temperature is T_{ref}.

their velocities are modified; the former being the principle behind all thermostats used in the present work.

The stochastic velocity-rescaling thermostat An improvement over the original velocity-rescaling[234] and Berendsen thermostats,[235] it is also known as the Bussi-Donadio-Parrinello thermostat.[236] All three aforementioned thermostatic algorithms follow the same philosophy: given the temperature of the system T (computed according to Eq. 3.24) and a target reference temperature T_{ref}, it is possible to define a parameter $\alpha = \alpha\,(T, T_{\text{ref}})$, which scales the velocities of all particles in the system so that T tends to the target temperature T_{ref} (Figure 3.12).

While for the original velocity rescaling thermostat α was just the ratio $\sqrt{T_{\text{ref}}/T}$, so the temperature after the action of the thermostat became exactly T_{ref}. Berendsen's α includes a relaxation time t_r to smooth the action of the thermostat. However, neither of them reproduces the correct thermal fluctuations of the NVT ensemble.

Bussi's idea is to, given that the kinetic energy of a system is proportional to its temperature ($K = N_f T/2$), draw a value from its distribution for the canonical ensemble[237]

$$P(K) = \frac{K^{N_f/2-1} e^{-K/(kT)}}{\Gamma(N_f/2)(kT)^{N_f/2}}\,. \tag{3.25}$$

As doing this at each step would induce large fluctuations in velocities, they instead proposed to evolve the kinetic energy while following the distribution given by Eq.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.25. The differential equation proposed for the propagation of the kinetic energy is

$$dK = (K_{\text{ref}} - K)\frac{dt}{t_r} + 2\sqrt{\frac{K K_{\text{ref}}}{N_f t_r}}dW\,, \tag{3.26}$$

where t_r is an equivalent to Berendsen's relaxation time and dW is a Wiener process*. By solving the previous equation, they arrive at the formula of the rescaling factor

$$\alpha^2 = e^{-\Delta t/t_r} + \frac{K_{\text{ref}}}{N_f K}\left(1 - e^{-\Delta t/t_r}\right)\left(R_1^2 + \sum_{i=2}^{N_f} R_i^2\right) + 2e^{-\Delta t/(2t_r)}\sqrt{\frac{K_{\text{ref}}}{N_f K}\left(1 - e^{-\Delta t/t_r}\right)}R_1\,. \tag{3.27}$$

where R_i are random numbers drawn from a normal distribution of unit variance and zero mean. Scaling all velocities by α, the temperature of the system in the simulation will reproduce the distribution of the canonical ensemble.

The stochastic velocity-rescaling algorithm was the only thermostat used in all simulations performed for the present work.

3.1.3.2 Barostats

Using the Virial theorem, one can arrive at the following expression for the stress tensor[238]

$$\mathbf{P} = \frac{2}{V}\left(\frac{1}{2}\sum_{i=1}^{N}\frac{\mathbf{p}_i \otimes \mathbf{p}_i}{m_i} - \frac{1}{2}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\mathbf{r}_{ij} \otimes \mathbf{f}_{ij}\right)\,, \tag{3.28}$$

being $\mathbf{f}_{ij}$ the interaction force between two particles and $\mathbf{r}_{ij}$ their relative position. The symbol $\otimes$ refers to the Kronecker product, which for vectors is equivalent to $\mathbf{a} \otimes \mathbf{b} = \mathbf{a}\mathbf{b}^{\mathsf{T}}$. Notice how the trace of the first term corresponds to the kinetic energy of the system (Eq. 3.23). As the forces—derivatives of the potential U_{ij}—depend on the relative positions of the particles involved, acting on the latter will affect the stress tensor. In the same way, scaling the whole volume of the simulation cell will also lead to changes in the stress tensor.

In a system in equilibrium with the environment, the elements on the diagonal of the stress tensor $\mathbf{P}$ correspond to the pressure acting on each of the surfaces; furthermore, the pressure p in an isotropic system should be the same in all 3 components, so $p = \text{Tr}(\mathbf{P})/3$.

Next, the barostatic algorithms used for the present work will be introduced. A correct scaling of the pressure has to be done according to the symmetry of the system through a **semi-isotropic scaling**: In planar interfacial systems, the deformations in the plane of the interface (the XY plane by construction) must be

*A Wiener process can be interpreted as the continuum limit of a random walk. The increments in W at a time τ follow a zero-mean and τ-variance Gaussian distribution.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

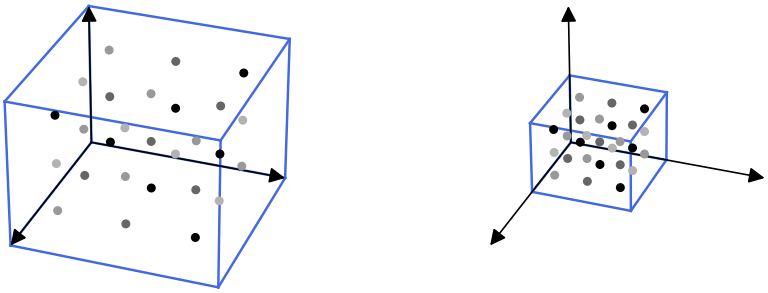


Figure 3.13: Fundamentals of the cell-rescaling approach. In the left, a set of particles is occupying the volume defined by the simulation cell (blue lines). The internal pressure of the system $\mathbf{P}$, computed according to Eq. 3.28, is lower than the target. Scaling the position of the particles and the dimensions of the cell leads to the desired $\mathbf{P}_{\text{ref}}$.

controlled independently from its normal direction. The practical implementation is to compute the average pressure at the interface $p_{XY} = (\mathbf{P}_{XX} + \mathbf{P}_{YY})/2$, and in the normal direction $p_Z = \mathbf{P}_{ZZ}$ and obtain a different scaling factor for each of them. The reference stress tensor would be

$$\mathbf{P}_{\text{ref}} = \begin{pmatrix} P_{\text{ref},XY} & 0 & 0 \\ 0 & P_{\text{ref},XY} & 0 \\ 0 & 0 & P_{\text{ref},Z} \end{pmatrix} \tag{3.29}$$

The scaling matrix $\mathbf{M}$ is that that transforms the coordinates and volume of the system so that the final pressure matches the reference values (Figure 3.13):

$$\begin{cases} \mathbf{x} \to \mathbf{M}\mathbf{x} & \text{(3.30a)} \\ V \to \det(\mathbf{M})V & \text{(3.30b)} \end{cases}$$

The Berendsen barostat In addition to the Berendsen thermostat mentioned above, this barostat acts on the components of the pressure tensor by scaling all coordinates in the simulation cell and its volume accordingly.[235] The scaling matrix is

$$\mathbf{M} = \mathbb{1} - \frac{\Delta t}{3t_r}\boldsymbol{\beta} \odot (\mathbf{P} - \mathbf{P}_{\text{ref}}) \ . \tag{3.31}$$

with $\mathbb{1}$ the identity matrix and t_r the relaxation time of the barostat. The symbol $\odot$ stands for the Hadamard product of two matrices, so $(\mathbf{A} \odot \mathbf{B})_{ij} = a_{ij}b_{ij}$. The matrix $\boldsymbol{\beta}$ is the isothermal compressibility matrix that modulates how the volume of the system changes in response to changes in pressure. As the reference stress

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

tensor, in a semi-isotropic scale $\boldsymbol{\beta}$ is a diagonal matrix, and the elements associated with the dimensions X and Y are the same

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{XY} & 0 & 0 \\ 0 & \beta_{XY} & 0 \\ 0 & 0 & \beta_Z \end{pmatrix}, \tag{3.32}$$

so will be $\mathbf{M}$.

The Berendsen barostat does not correctly sample the pressure fluctuations as described by the NPT ensemble but it is very stable in matching the reference pressure. For these reasons, its use is limited to some initial equilibration steps in the simulation protocol.

The stochastic cell-rescaling barostat Following the same philosophy as with the stochastic velocity-rescaling thermostat, Bernetti and Bussi[239] proposed a barostat that correctly samples the NPT ensemble by adding a stochastic term to reproduce the fluctuations in pressure. Although Berendsen's barostat could be extended for semi-isotropic systems easily, the derivation of the formulae for isotropic (all dimensions are scaled evenly) or semi-isotropic systems is different for this barostat. In the isotropic case, the proposal for the evolution of the strain of volume of the simulation cell with respect to an arbitrary reference volume V_0, $\varepsilon = \ln(V/V_0)$, is

$$d\varepsilon = -\frac{\beta}{t_r}\left(P_{\text{ref}} - P\right) dt + \sqrt{\frac{2kT\beta}{Vt_r}} dW\,, \tag{3.33}$$

Its semi-isotropic equivalent consists of two equations, one for the strain in the area, $\varepsilon_A = \ln(A/A_0)$, and another for the strain in the height, $\varepsilon_L = \ln(L/L_0)$. If the plane of the interface is the plane XY:

$$d\varepsilon_A = -\frac{\beta_{XY}}{3t_r}\left(P_{\text{ref},XY} - \frac{\gamma}{L} - \frac{P_{XX} + P_{YY}}{2}\right) dt + \sqrt{\frac{4kT\beta_z}{3Vt_r}} dW \tag{3.34a}$$

$$d\varepsilon_L = -\frac{\beta_Z}{3t_r}\left(P_{\text{ref},Z} - P_{ZZ}\right) dt + \sqrt{\frac{2kT\beta_z}{3Vt_r}} dW \tag{3.34b}$$

where γ is an external surface tension, always zero in our case. By integrating the differential equations associated with the implementation of the barostat, one obtains the change in the volume of the simulation cell. Then, positions must be scaled accordingly.

The Parrinello-Rahman barostat The way in which Parrinello and Rahman[240] proposed controlling the volume of the simulation cell, and therefore scaling the pressure, is to evolve the vectors of the simulation cell.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Being $\mathbf{L}$ the cell matrix, where each of the columns is the triclinic vectors that define the simulation cell, $\mathbf{L}_i$, the equations of motion associated with the cell vectors are[241]

$$\ddot{\mathbf{L}} = V\mathbf{W}^{-1}\left(\mathbf{L}^{\mathsf{T}}\right)^{-1}\left(\mathbf{P}-\mathbf{P}_{\text{ref}}\right) . \tag{3.35}$$

Here $\mathbf{W}^{-1}$ is a matrix whose components are

$$(\mathbf{W}^{-1})_{ij} = \frac{4\pi\beta_{ij}}{3t_r \max(\|L_x\|,\|L_y\|,\|L_z\|)} . \tag{3.36}$$

This change in the simulation box affects the Hamiltonian of the system and, hence, the equations of motion of the particles which have to be solved:

$$\dot{\mathbf{p}}_i = -\nabla_i U - \mathbf{M}\dot{\mathbf{x}}_i , \tag{3.37}$$

where

$$\mathbf{M} = \mathbf{L}^{-1}\left[\mathbf{L}(\dot{\mathbf{L}^{\mathsf{T}}}) + \dot{\mathbf{L}}\mathbf{L}^{\mathsf{T}}\right]\left(\mathbf{L}^{\mathsf{T}}\right)^{-1} \tag{3.38}$$

3.2 Simulation protocol

The simulations performed for the present work can be grouped into two different sets: **(i)** the atomistic simulations to study the evolution of the structure of peptides at the interface between water and a lipid bilayer; and **(ii)** the CG simulations to study the experimentally-proven antimicrobial activity of a sequence has an impact on the peptide–membrane interaction.

Both of them share the same elements: a lipid bilayer, a single unit of a peptide, water as a solvent, and ions. However, the simulation parameters, the protocols used to assembly the systems, to equilibrate them, and to perform the simulations are specific to each setup.

The process of system assembly and preparation process was performed using a combination of custom Bash and Python scripts, employing for the latter the libraries MDAnalysis,[242,243] Numpy,[244] Pandas,[245] and Scipy,[246] as well as other libraries included in the standard package.[247]

3.2.1 Force field-specific parameters

Some simulation parameters are common to all simulations in the same force field. The specifics on the approaches used for the calculation of the non-bonded interactions, as well as the bonds restrained are presented next. AMBER, CHARMM and Martini parameters are those of the inputs generated by CHARMM-GUI.[248–251]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

AMBER ff19SB+Lipids21 (AA) The van der Waals interactions were computed using a 0.9 nm cutoff and long-range correction for energy and pressure. Electrostatics were computed using the particle-mesh Ewald method, with a 0.9 nm cutoff for real-space calculations, and a grid spacing in the reciprocal space of 0.12 nm and a 4-th order interpolation. Bonds involving hydrogen atoms were constrained using the LINCS algorithm.

CHARMM36m (AA) Van der Waals interactions were computed using a cutoff scheme with a switching function for the forces that act between 1.0 nm and 1.2 nm. Electrostatics were computed with the PME method, with a 1.2 nm cutoff for the real-space calculations, while other parameters are the same as in AMBER. The LINCS algorithm is also employed to constrain covalent bonds with hydrogen atoms.

GROMOS54a7 (UA) As mentioned previously, the algorithms used originally to parameterize this force field are not physically accurate.[218] Instead, the same parameters as in AMBER were used, but the potentials were shifted so that they become zero at the cutoff distance of 1 nm, the same cutoff as used for the real-space calculations in PME.[141,229] The LINCS algorithm was used to constrain all bonds.

MARTINI (CG) The calculation of the van der Waals interactions for all three CG force fields—Martini 2.2, Martini 2.2P and Martini 3—was done within a cutoff of 1.1 nm, shifting the potential so it became zero at this distance. Electrostatics, on the other hand, were computed using the reaction field method, with a cutoff of 1.1 nm and a relative permittivity of $\varepsilon = 15$ for Martini 2.2 and Martini 3 or $\varepsilon = 2.5$ for Martini 2.2P, and $\varepsilon_{\mathrm{RF}} = \infty$ in all cases.

3.2.2 Atomistic simulations

Lipid bilayer assembly Systems of 152 lipids per leaflet were assembled using the *BilayerBuilder* tool in CHARMM-GUI,[248,252,253] with a 3:1 ratio of POPE (1-palmitoyl-2-oleoyl-sn-*glycero*-3-phosphoethanolamine) and POPG (1-palmitoyl-2-oleoyl-sn-*glycero*-3-phosphoglycerol), symmetric for each leaflet. This is roughly the ratio of the most common lipids in the inner lipid bilayer of the Gramm-negative bacteria *Escherichia coli*.[33] The membrane was oriented with its normals parallel to the Z axis. 70 water molecules per lipid were added, namely the OPC[205] water model for AMBER simulations or the TIP3P[216] water model for CHARMM. Sodium counterions were added to neutralize the charge inside the simulation box.

The UA membrane for GROMOS simulations was built from the CHARMM one by removing all nonpolar H atoms. The atoms were sorted and renamed according to the GROMOS-CPK topologies for lipids[219,220] and for the SPC water model.[221]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

All three structures were minimized and equilibrated following the protocol given by CHARMM-GUI. Each equilibration step started from the last structure and velocities of the previous, and unless stated otherwise, the simulation parameters and restraints were preserved between steps:

1.- A 5000-step steepest descent minimization, or until a convergence in maximum force of 1000 kJ mol^{-1} nm^{-1}. The change in the Z coordinate of the phosphorus atoms in the phospholipids was restrained by a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}, the same value—in kJ mol^{-1} rad^{-2}—used to restrain the dihedrals around the saturation in the lipid tails.

2.- A 125 ps NVT equilibration with a 1 fs time step. The temperature was kept at 300 K using the stochastic velocity-rescale thermostat,[236] coupling the membrane and the solvent separately, both with a time constant of 1 ns. The initial velocities of the particles were generated by randomly drawing them from the Maxwell-Boltzmann distribution at the same temperature.

3.- A 125 ps NVT equilibration. The force constants in the restraints were reduced to 400 kJ mol^{-1} nm^{-2}, or 400 kJ mol^{-1} rad^{-2} for dihedrals.

4.- A 125 ps NPT equilibration. The pressure was controlled by the semi-isotropic implementation of the cell-rescaling barostat,[239] set at 1 bar, with a value of the isothermal compressibility of $4.5{\cdot}10^{-5}$ bar^{-1} for both the XY plane and the Z direction. The time constant of the barostat was 5 ps. The force constant in restraining the position of the lipids was preserved, but reduced for the dihedrals to 200 kJ mol^{-1} rad^{-2}

5.- A 500 ps NPT equilibration. The time step was increased to 2 fs and all force constants were reduced to 200 kJ mol^{-1} nm^{-2}.

6.- A 500 ps NPT equilibration. The force constants of the potentials acting on the position of the phosphorus atoms were further reduced to 40 kJ mol^{-1} nm^{-2}, while those of the saturations were reduced 100 kJ mol^{-1} rad^{-2}.

7.- A 500 ps NPT equilibration with no restraints acting on the lipids.

The final structure of the equilibration process was used for the assembly of all subsequent systems. The same structure was used for both GROMOS and GROMOS-HIE.

Assembly of peptide–membrane systems Initial structures for the peptides were generated with a Python script based on the PeptideBuilder[254] library. Peptides were constructed by concatenating amino acids following the sequences, so the backbone dihedral angles were $\phi = -60°$ and $\psi = -40{\cdot}0°$* obtaining right-handed

*The usual naming is used: ϕ refers to the dihedral angle around the N-Cα bond, while ψ is the angle around the Cα-C bond.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

α-helices. The protonation state of the different groups coincides with the most abundant at physiological pH: both termini are charged, and so are the side chains of Arg, Lys, Asp, and Glu.[255]

AMBER and CHARMM topologies for the peptides were generated together with those for the lipids and ions in CHARMM-GUI, while those for GROMOS were built with the GROMACS pdb2gmx tool.[256]

The structure of the peptide was then inserted into that of the lipid bilayer, at a minimum distance from the lipids of 10 Å and with an specific orientation of the peptide around its axis (spin, see Section 3.3.4). The water molecules within 6 Å of the peptide were removed to avoid overlap and sodium or chlorine ions were added to balance the charge within the box.

Unfolding simulations To study the structure of peptides at the water–lipid bilayer interface, a first set of systems was assembled by placing one copy of one of the following peptides: parasin I (KGRGKQGGKVRAKAKTRSS),[257] magainin-2 (GIGKFLHSAKKFGKAFVGEIMNS),[258] or octa-2 (RRWWRWWR)[259,260]*; in the same position with respect to the membrane but with 5 different evenly spaced orientations around their axes, spanning the whole 2π rad range. These peptides have shown significant antimicrobial activity, so they were expected to interact with the lipid bilayers. One copy of each of these systems was generated for each of the atomistic force fields—AMBER, CHARMM, GROMOS and GROMOS-HIE—, adding to a total of 60 systems. The adsorption of a peptide into a membrane is known to be slow,[115] so to facilitate it without biasing the production run, an external harmonic potential acting on the peptide was added during the equilibration process to bring it closer to the lipid surface. The whole protocol used to prepare these simulations, which we will refer to as **shallow** configuration, is as follows (Figure 3.14):

1.- A 5000-step steepest descent minimization, or until a convergence in maximum force of 1000 $\mathrm{kJ\,mol^{-1}\,nm^{-1}}$. The change in the Z coordinate of the phosphorus atoms in the phospholipids was restrained by a harmonic potential with a force constant of 1000 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$, the same value—in $\mathrm{kJ\,mol^{-1}\,rad^{-2}}$—used to restrain the dihedrals around the saturation in the lipid tails. The positions of the heavy atoms in the backbone and side chains were also restrained with force constants of 4000 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ and 2000 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$, respectively.

2.- A 200 ps NVT equilibration with a 1 fs time step. The temperature was kept at 300 K using the stochastic velocity-rescale thermostat,[236] coupling the

*Henceforth, parasin I and magainin-2 will be simply referred to as parasin and magainin, respectively.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

membrane, the peptide, and the solvent separately, all with a time constant of 1 ns. The initial velocities of the particles were generated by randomly drawing them from the Maxwell-Boltzmann distribution at the same temperature.

3.- A 1 ns *NPT* equilibration. The time step was increased to 2 fs. The pressure was controlled by the semi-isotropic implementation of the cell-rescaling barostat,[239] set at 1 bar, with a value of the isothermal compressibility of $4.5 \cdot 10^{-5}$ bar^{-1} for both the XY plane and the Z direction. The time constant of the barostat was 5 ps. The position restraints acting on the peptide and lipids were removed, while the force constants acting on the lipid dihedrals were reduced to 400 kJ mol^{-1} rad^{-2}. Additional harmonic restraints were imposed on the distance between all 5-neighboring C_α atoms and between the H_N and O atoms involved in stabilizing the helix, with a force constant of 400 kJ mol^{-1} nm^{-2}. Using PLUMED,[261,262] an additional external harmonic potential with a constant of 3000 kJ mol^{-1} nm^{-2} pulled the peptide toward the membrane, at an equilibrium position of 2 nm over the mean position of the phosphorus atoms of the lipids, close to those in the upper leaflet.

4.- A 5 ns *NPT* equilibration. The external harmonic potential was removed, as well as the restraints in the dihedrals. The force constants acting on the C_α–C_α and H_N–O distances were reduced to 50 kJ mol^{-1} nm^{-2}.

5.- A 5 ns *NPT* equilibration. Force constants were reduced to 25 kJ mol^{-1} nm^{-2}.

The protocol for GROMOS-HIE simulations suffered from a few differences: In Steps 4 and 5, the time step increased to 4 fs and 5 fs, respectively.

The production run for each of the systems continued from the last state of the equilibration, but removing all restraints. Each simulation was extended for 900 ns in total simulation time, with a 2 fs time step for AMBER, CHARMM, and GROMOS, and 6 fs for GROMOS-HIE. To compensate for the differences in the time step and to allow a direct comparison among the force fields, simulation states were stored every 6 ps in all cases.

A second batch of simulations comprised the exact same systems: all three peptides, starting from five different orientations around its axes, and five force fields. The equilibration protocol followed now included a external harmonic potential that induced the insertion of the peptide parallel to the membrane below the surface; such a protocol, which lead to **embedded** configurations, is as follows (Figure 3.15):

1.- A 5000-step steepest descent minimization, or until a convergence in maximum force of 1000 kJ mol^{-1} nm^{-1}. The change in the Z coordinate of the phosphorus atoms in the phospholipids was restrained by a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}, the same value—in kJ mol^{-1} rad^{-2}—used to restrain the dihedrals around the saturation in the lipid tails. The position

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

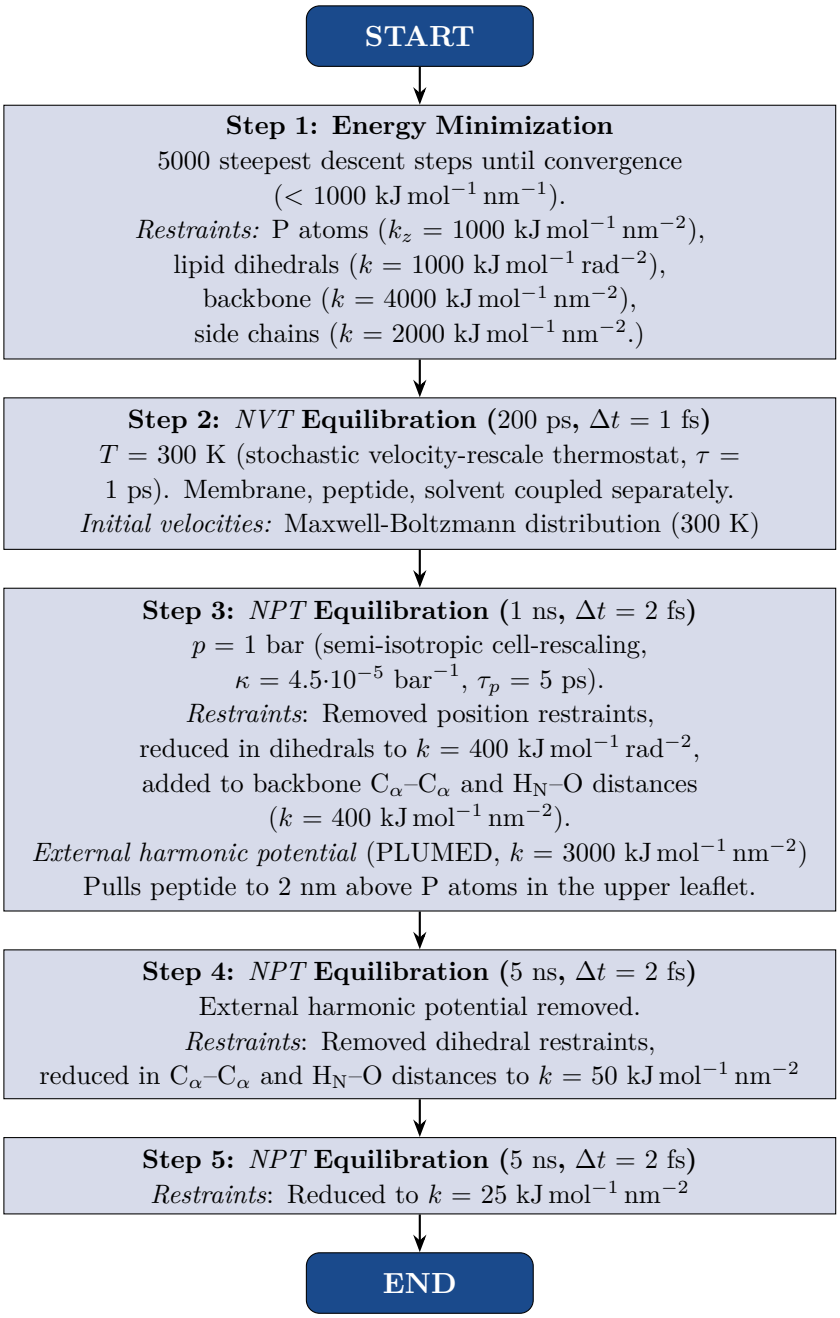


Figure 3.14: Equilibration protocol for the atomistic simulations in the "shallow" configuration. The value of time step in the equilibration of GROMOS-HIE simulations was 4 fs and 5 fs in steps 4 and 5, respectively.

of the heavy atoms in the backbone and side chains were also restrained with force constants of 4000 kJ mol^{-1} nm^{-2} and 2000 kJ mol^{-1} nm^{-2}, respectively.

2.- A 200 ps *NVT* equilibration with a 1 fs time step. The temperature was kept at 300 K using the stochastic velocity-rescale thermostat,[236] coupling the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Time (ps)	**Pulling**		**Alignment**	
	k (kJ mol^{-1} nm^{-2})	h_0 (nm)[†]	k (kJ mol^{-1} nm^{-2})	d_0 (nm)[‡]
0	3000	2.0	3000	0
40	4000	1.5	–	–
80	5000	1.0	0	–
120	3000	–	–	–

[†] Distance below the mean position of the P atoms in the upper leaflet.

[‡] Distance between the mean position of the 4 C_α in each termini and the centroid of the peptide.

Table 3.1: Values of the parameters of the external harmonic potentials applied during the equilibration. The first column is the instant at which the values change; followed by the force constant k and the equilibrium position of each of the potentials: the first used to embed the peptide into the membrane, and the second to keep it aligned with the surface. The symbol "–" indicates no change in the parameter.

membrane, the peptide, and the solvent, all with a time constant of 1 ps. The initial velocities were generated by randomly drawing them from the Maxwell-Boltzmann distribution at the same temperature.

3.- A 50 ns *NPT* equilibration. The time step was increased to 2 fs. The pressure was controlled by the semi-isotropic implementation of the cell-rescaling barostat,[239] set at 1 bar, with a value for the isothermal compressibility of $4.5 \cdot 10^{-5}$ bar^{-1} for both the XY plane and the Z direction. The time constant of the barostat was 5 ps. The position restraints acting on the peptide and lipids were removed, while the force constants acting on lipid dihedrals were reduced to 400 kJ mol^{-1} rad^{-2}. Additional harmonic restraints were imposed on the distance between all 5-neighboring C_α atoms between the H_N and O atoms involved in stabilizing the helix, with a force constant of 4000 kJ mol^{-1} nm^{-2}. An additional external potential pulled the peptide towards the membrane, embedding it below the surface defined by the phosphorus atoms. The force constant as well as the equilibrium position changed over time. Auxiliary potentials were added to keep the axis of the peptide parallel to the surface while the insertion occurs; they acted by maintaining the Z component of the mean position of C_α atoms in the first and last turn equal to that of the center of geometry of the backbone. The parameters of the external potentials are summarized in Table 3.1.

4.- A 5 ns *NPT* equilibration. All external harmonic potentials were removed, as well as the restraints in the dihedrals. The force constants acting on the C_α–C_α and H_N–O distances were reduced to 1000 kJ mol^{-1} nm^{-2}.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.- A 5 ns *NPT* equilibration. The force constants of the restraints were reduced to 500 kJ mol^{-1} nm^{-2}.
6.- A 5 ns *NPT* equilibration. Force constants were reduced to 50 kJ mol^{-1} nm^{-2}.
7.- A 5 ns *NPT* equilibration. Force constants were reduced to 10 kJ mol^{-1} nm^{-2}.

The time step in the GROMOS-HIE simulations was progressively increased: in Step 4 it was 4 fs, while in Steps 5 to 7 it was increased to 5 fs. Simulations with the GROMOS-HIE force field were done in duplicate, once using the PME method for the calculation of the electrostatics and the other using the reaction field method. The value of the cutoff distance of 1 nm used is the same as for the van der Waals interactions. For the medium beyond the cavity, the permittivity of the SPC water model, $\varepsilon_{\mathrm{RF}} = 68$,[263] was used.

Production runs had the same characteristics as in the previous batch of simulations: 900 ns in total simulation time, with a 2 fs time step for AMBER, CHARMM, and GROMOS, and 6 fs for GROMOS-HIE.

A third batch of simulations was performed using only the GROMOS-HIE approach. Two systems per peptide were assembled, both with the same initial orientation but equilibrated with each of the previous protocols. Another membrane-targeting AMP, chensinin-1b (SKVWRHWRRFWHRAHRLH), was also included.[264] Each of these simulations were extended for 15 µs.

All 128 atomistic simulations were performed using a version of GROMACS 2024.2[256] patched with PLUMED[261,262] on the Finisterrae III supercomputer. The first batch of production simulations was carried out in an Intel Xeon Ice Lake 8352Y processor of 32 cores, and 4 Gb of memory per system were used; while for the second and third batches, as well as for the equilibration of all systems, one NVIDIA A100 GPU and the same processors and memory were used.

3.2.3 Coarse-grained simulations

Lipid bilayer preparation Systems of 250 lipids per leaflet and approximately 100 water molecules per lipid* were assembled using the *BilayerBuilder* tool in CHARMM-GUI[248,251]. Symmetric bilayers with two different compositions were generated: a pure POPC (1-palmitoyl-2-oleoyl-sn-*glycero*-3-phosphocholine) lipid bilayer, a model for the membrane of a healthy mammal,[2,265] and another with the same 3:1 POPE:POPG mixture as used in previous atomistic simulations. Sodium counterions were added to the latter to neutralize the total charge within the system. The structures were generated in triplicate, each according to the representation of three different force fields: Martini 2.2, Martini 2.2P, and Martini 3.

*Each Martini water "molecule" models 4 real water molecules; thus, there were 25 Martini water particles per lipid.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

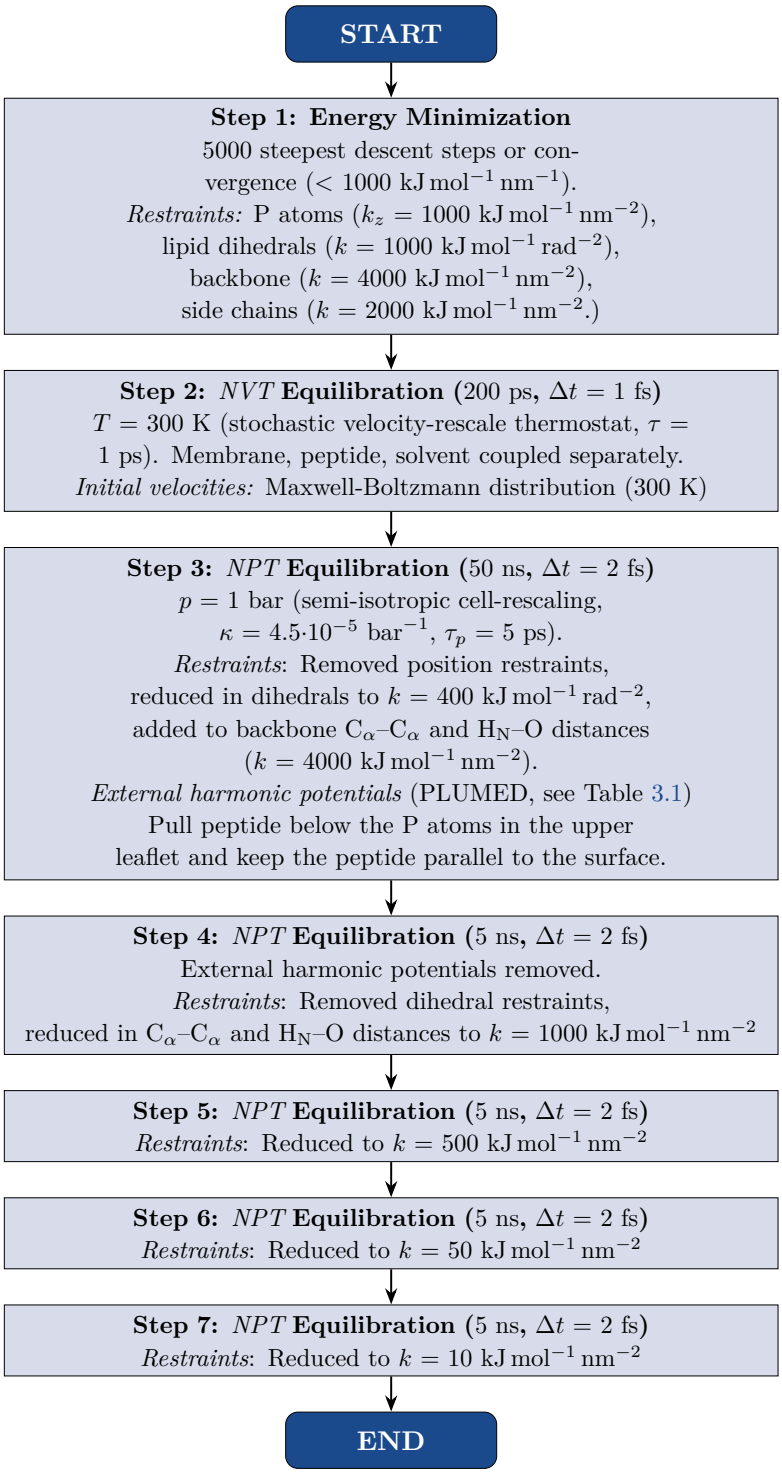


Figure 3.15: Equilibration protocol for the atomistic simulations in the "embedded" configuration. The value of time step in the equilibration of GROMOS-HIE simulations is 4 fs in step 4 and 5 fs in steps 5 to 7.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Each structure was minimized and equilibrated following the protocol defined by the CHARMM-GUI team. As in the atomistic case, each step started from the final state of the previous one, and any unspecified parameters must be assumed to be preserved from the prior step.

1.- A 10000-step steepest descent minimization, or until a convergence in the maximum force of 1000 $\mathrm{kJ\,mol^{-1}\,nm^{-1}}$.

2.- A 1 ns *NPT* equilibration, with a 2 fs time step. The temperature was kept at 300 K using the stochastic velocity-rescaling thermostat,[236] coupling the membrane and the solvent separately, both with a time constant of 1 ps. The pressure was controlled by the semi-isotropic implementation of the Berendsen barostat,[235] set at 1 bar, with a value for the isothermal compressibility of $4.5 \cdot 10^{-5}$ $\mathrm{bar^{-1}}$ for both the XY plane and the Z direction. The time constant of the barostat was 5 ps. An external harmonic potential restrained the Z coordinate of the beads describing the phosphate group of the lipids, with a force constant of 200 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$. The initial velocities were generated by randomly drawing them from the Maxwell-Boltzmann distribution at the same reference temperature.

3.- A 1 ns *NPT* equilibration. The time step was increased to 5 fs. The force constant restraining the phosphate groups was reduced to 100 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$.

4.- A 1 ns *NPT* equilibration. The time step was increased to 10 fs. The force constant restraining the phosphate groups was reduced to 50 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$.

5.- A 750 ps *NPT* equilibration. The time step was increased to 15 fs. The force constant restraining the phosphate groups was reduced to 20 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$.

6.- A 1 ns *NPT* equilibration. The time step was increased to 20 fs. The force constant restraining the phosphate groups was reduced to 10 $\mathrm{kJ\,mol^{-1}\,nm^{-2}}$.

Assembly of peptide–membrane systems Initial structures of the peptides were created with the same script as for the atomistic simulations. The mapping to the Martini representation, both for structure and topology, was done using tools developed by the Martini team: the script *martinize*[230] for Martini 2.2 and 2.2P, and the martinize2 library[266] for Martini 3. The topologies of the peptides included parameters to preserve the secondary structure of the peptide as a full α-helix. The protonation state was set to match the most probable one at neutral pH.

Sequences of membrane-targeting antimicrobial peptides with some content of α-helix structure were obtained from the APD3 database,[156] while nonactive sequences were extracted from the data set curated by Tucs *et al.*[267]. All sequences were first filtered to range from 10 to 30 residues.

To ensure the diversity of the simulated sequences, each pair of sequences within

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

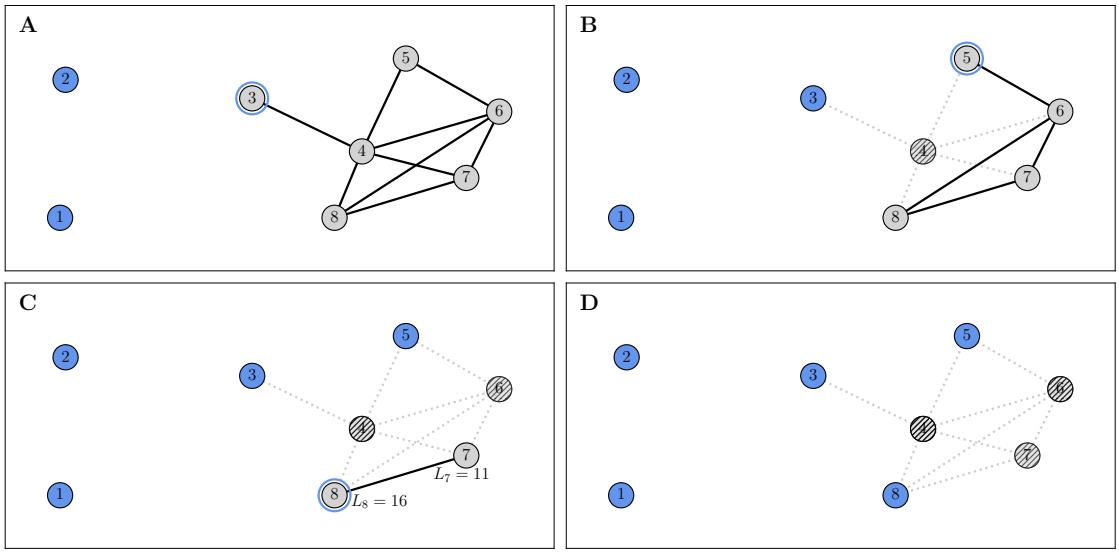


Figure 3.16: Filtering similar sequences. **A.-**A graph is built where each node is a sequence within the data set. An axes connects two nodes when the similarity metric is above 0.8. The isolated sequences 1 and 2 are automatically preserved; then, for each connected subgraph (sequences 3 to 8), the least connected sequence (3) is identified. **B.-** All sequences connected to 3 are removed, so it becomes isolated and a new subgraph is generated. The sequence with the least connections, 5, is once again identified. **C.-**The process is repeated, so 5 becomes an isolated sequence. All sequences in the resulting subgraph have the same number of connections, so the longest sequence is selected: sequence 8 has 16 residues, while 7 is constituted by only 11 residues. **D.-**The remaining sequence connected to 8 is removed. All selected sequences (1, 2, 3, 5 and 8) are isolated and the similarity metric between them is always below the threshold, indicative of a sufficient difference.

the data set was compared using BLASTp.[268] A measure of similarity between the peptide sequences was obtained by multiplying the percentage of identity and the alignment length—both are indicated in the result of the analysis conducted by BLASTp—relative to the size of the shorter sequence. This "normalized similarity factor" ranges between 0 and 1. For each data set, a graph was built in which each sequence is a node, connected to all others by an axis whenever this metric is above a threshold of 0.8. The isolated nodes were assumed to be sufficiently different from the other sequences, while for the subsets of connected nodes, an iterative process was done to further filter them: The node with the least axes was the most different within the subgraph, so it was be preserved; or the longest sequence if more than one node had the same number of axes. Then, all nodes connected to them were removed and the process was repeated until all nodes in the graph were isolated (Figure 3.16).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The final selection contained 123 sequences—73 of them corresponding to membrane-targeting antimicrobial peptides. A copy of each of them was inserted into a system with each of the membranes, in the bulk and at a minimum distance in the Z direction of 10 Å with respect to the upper leaflet. The water molecules within 6 Å of the peptide were removed to avoid overlap. Sodium or chlorine counterions were added to balance the charge in the simulation box.

A total of 738 systems were assembled as a combination of all sequences (123), lipid compositions (2), and force fields (3). Each of them followed a simple minimization and equilibration process (Figure 3.17):

1.- A 20000-step steepest descent minimization, or until a convergence in the maximum force of 1000 kJ mol^{-1} nm^{-1}.

2.- A 750 ns NPT equilibration, with a 15 fs time step. The temperature was kept at 300 K using the stochastic velocity-rescaling thermostat,[236] coupling the membrane, the peptide and the solvent separately, all with a time constant of 1 ps. The pressure was controlled by the semi-isotropic implementation of the Berendsen barostat,[235] set at 1 bar, with a value for the isothermal compressibility of $3{\cdot}10^{-4}$ bar^{-1} for both the XY plane and the Z direction. The time constant of the barostat was 5 ps. The initial velocities were generated by randomly drawing them from the Maxwell-Boltzmann distribution at the reference temperature.

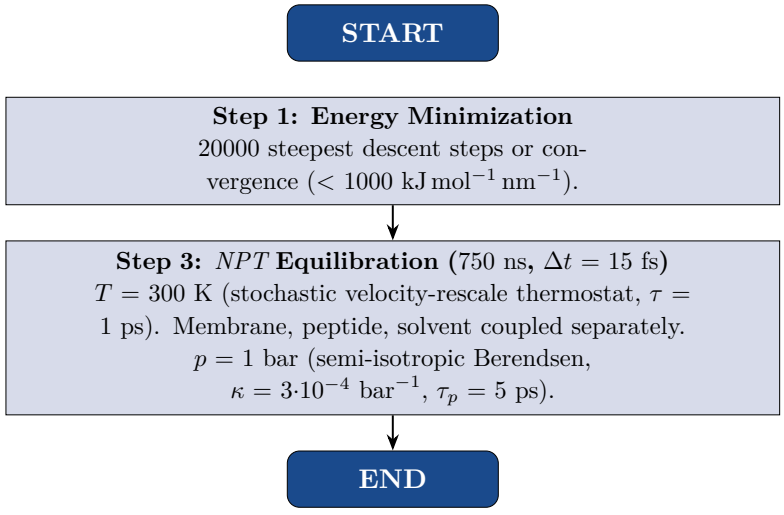


Figure 3.17: Equilibration protocol for the CG simulations. Membrane structures were previously equilibrated, reducing the need for an extensive equilibration at this point.

The production run of each of the systems was extended for 5 µs, using a 25 fs time step. The pressure was controlled with the semi-isotropic implementation of the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Parrinello-Rahman barostat,[240] to ensure a correct sampling of the *NPT* ensemble. The time constant of the barostat was 12 ps, while all other parameters were the same as in the equilibration run.

All CG simulations were performed using GROMACS 2021.5[256] on the Finisterrae III supercomputer. Each of them was executed using an Intel Xeon Ice Lake 8352Y processor of 32 cores and 4 Gb of memory.

3.3 Processing and analysis of MD simulation data

The previous simulations generated a time-series of the positions and velocities of the particles in each system. These trajectories must be analyzed to extract information about key aspects of the behavior of the molecules. This section details the different analyses applied on the trajectories, together with some pre-processing methods that had to be used in conjunction with them.

3.3.1 Transformation of coordinates

As a result of the use of PBCs when performing a simulation, the stored coordinates of the particles are located within the boundaries of the original simulation cell. For the same reason, any translation of the coordinates of a particle any integer number of times the vectors of the simulation cell $\mathbf{L}_i$ is equivalent to not having translated the particle

$$\mathbf{x}' = \mathbf{x} + (n_x\mathbf{L}_x + n_y\mathbf{L}_y + n_z\mathbf{L}_z) \equiv \mathbf{x}, \qquad n_x, n_y, n_z \in \mathbb{R}. \tag{3.39}$$

Nonetheless, it is sometimes convenient or even essential to apply some transformations to the coordinates for a later analysis of the data.

Molecule wrapping Wrapping refers to the transformation of the coordinates of the particles in the system so that afterwards they are contained within the limits of the simulation box. This transformation is what PBCs do on the coordinates of the particles "leaving" the simulation cell.

If x and x' are respectively the wrapped and the original—unwrapped—coordinates along a certain dimension, where the size of the simulation cell is L, the transformation from x' into x requires translating the particle a certain (integer) number of times the size of the simulation cell, so in the end $x \in [0, L)$:

$$x = x' - \left\lfloor \frac{x'}{L} \right\rfloor L = x' \mod L\,, \tag{3.40}$$

where the floor function $\lfloor\cdot\rfloor$ returns the closest integer number that is less than or equal to the argument and mod the modulo operator. Thus, x is the remainder of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

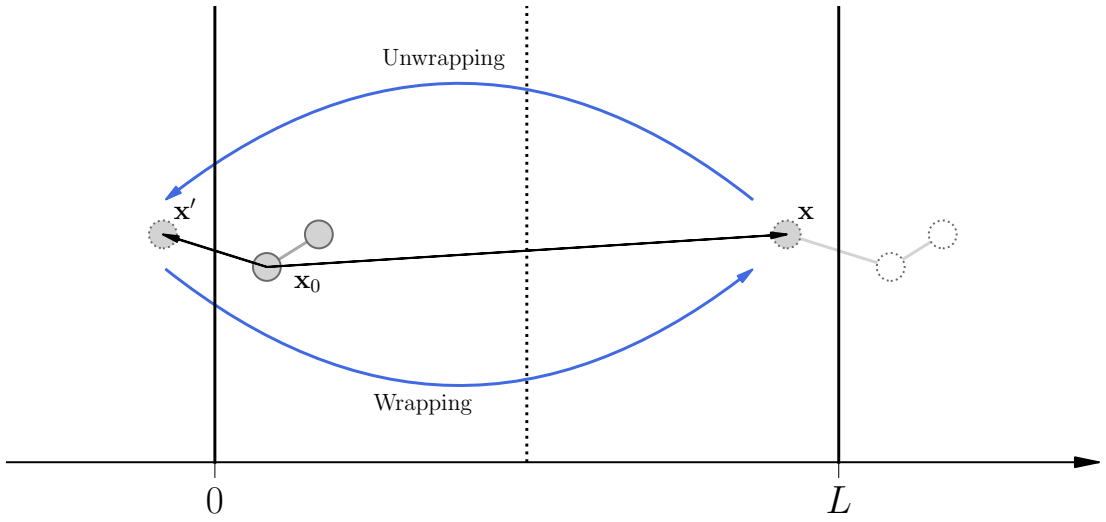


Figure 3.18: The wrapping and unwrapping transformations. The wrapping transformation puts all the molecules into the bounds of the simulation cell ($\mathbf{x}' \rightarrow \mathbf{x}$); to recover the original structure, the unwrapping transformation needs to use another position $\mathbf{x}_0$ as reference, ensuring that the distance of the unwrapped particle with respect to the reference is smaller than half the simulation cell.

the division of the original coordinate by the corresponding dimension of the cell (Figure 3.18).

Molecule unwrapping The inverse transformation is not unique, as the memory of the previous position has been lost. Still, it can be done for a set of particles with respect to a reference; for example, to reconstruct a molecule split by PBCs. Doing so implies finding the periodic image in which the distance to any other particle is minimal (Figure 3.18).

Let x_0 be the reference position, which could be that of a certain particle. The factor

$$k = \left\lfloor \frac{x - x_0}{L} + \frac{1}{2} \right\rfloor \tag{3.41}$$

is the number of times that L should be subtracted from x so that the distance from the reference, $\|(x - kL) - x_0\|$, is minimal. This implies that the unwrapped position x' satisfies

$$\|x' - x_0\| < \frac{L}{2}\,; \tag{3.42}$$

if it were larger than $L/2$, by adding or subtracting L the previous condition would

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

be satisfied. Thus, the unwrapped coordinate is given by

$$x' = x - \left\lfloor \frac{x - x_0}{L} + \frac{1}{2} \right\rfloor L \,. \tag{3.43}$$

Notice how the position of the reference particle remains unchanged, as $x'_0 = x_0$.

Our implementation is based on first defining the whole molecule to reconstruct and, if it is too large, some auxiliary fragments—such as amino acids in a peptide. Each of the fragments is reconstructed separately according to Eq. 3.43, and then the same is done to the whole molecule, using the centroids of the fragments instead. In the end, the whole molecule is unwrapped.

Molecule centering For both analysis and visualization purposes, it is often convenient to translate the system so that an atom or molecule occupies a certain position, typically in the middle of the box, along one or more dimensions.

If $\mathbf{x}_r$ is the vector whose components are the current positions of the references to translate, and $\mathbf{x}_t$ the target positions, the transformation to apply is

$$\mathbf{x} \to (\mathbf{x_t} - \mathbf{x_r}) + \mathbf{x} \tag{3.44}$$

Regardless of the reference molecule, its target position and the coordinates along which the transformation is applied, the whole system must be translated. The transformation can be made so that translations across different dimensions are uncoupled; for example, the system could be translated so that the lipid bilayer would be located in the middle of the Z dimension of the simulation box, without affecting the original positions in the XY plane.

Time-dependent unwrapping As mentioned above, there is an ambiguity with respect to the choice of reference to use for the unwrapping transformation given by Eq. 3.43. There are properties, however, which are sensitive to the temporal evolution of the position of the particles, like the diffusion, for which obtaining a time-unwrapped trajectory is key.

In an smooth trajectory, the displacement of a particle between two consecutive instants, t_{i-1} and t_i, should be small. If x'_j is the unwrapped coordinate of the particle in the instant t_j, then

$$\left\| x'_i - x'_{i-1} \right\| < \frac{L}{2} \,, \tag{3.45}$$

This condition reminds of Eq. 3.42, where now instead of a distance between two nearby particles in the same time step, the position of the same particle in consecutive time steps is being compared. Therefore, Eq. 3.43 should be also valid to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

reconstruct the smooth trajectory of the particle, now being

$$x'_i = x_i - \left\lfloor \frac{x_i - x'_{i-1}}{L} + \frac{1}{2} \right\rfloor L \,, \tag{3.46}$$

where the reference used to unwrap the position is that in the previous time step t_{i-1}, x'_{i-1}. Usually, the coordinates of the particles in an MD simulation are not saved at every single step, as the data generated would be immense, but at a certain frequency. If this frequency is too low—the separation between states is too large—, the condition given by Eq. 3.45 could not be fulfilled, so trying to unwrap the trajectory will lead to artifacts.

The previous expression is not universally valid, as the volume of the simulation cell in the *NPT* ensemble changes over time. This means that L is also a function of time and, in general, $L_i \neq L_{i-1}$. To account for the changes in the simulation cell, Kulke and Vermaas[269] proposed the alternative scheme

$$x'_i = x'_{i-1} + (x_i - x_{i-1}) - \left\lfloor \frac{x_i - x_{i-1}}{L_i} + \frac{1}{2} \right\rfloor L_i - \left\lfloor \frac{x_{i-1} - x'_{i-1}}{L_{i-1}} + \frac{1}{2} \right\rfloor (L_i - L_{i-1}) \tag{3.47}$$

A matrix formulation, completely equivalent to Eq. 3.47, was also provided, and is the one implemented in our code:

$$\mathbf{x}'_i = \boldsymbol{\alpha}\mathbf{x}'_{i-1} + \mathbf{x}_i - \boldsymbol{\alpha}\mathbf{x}_{i-1} - \mathbf{L}_i \left\lfloor \mathbf{L}_i^{-1}\mathbf{x}_i - \mathbf{L}_{i-1}^{-1}\mathbf{x}_{i-1} + \frac{1}{2} \right\rfloor \,, \tag{3.48}$$

where $\mathbf{x}_j$ and $\mathbf{x}'_j$ are the wrapped and unwrapped coordinates in time of a particle, respectively; $\mathbf{L}_j$ the matrix whose columns are the vectors of the box at a time t_j, and $\boldsymbol{\alpha} = \mathbf{L}_i \mathbf{L}_{i-1}^{-1}$. This formulation allows to unwrap the coordinates of all particles of a molecule or system in just one step.

Rotation While the system is invariant under translations due to the PBCs, the same does not apply to a rotation of the coordinates. Nevertheless, sometimes it is necessary to align a molecule with an axis, for which a rotation is unavoidable.

Let $\mathbf{x}$ be the position of a particle in $\mathbb{R}^3$ and let $\widehat{\mathbf{v}}$ be an associated unit vector; if $\widehat{\mathbf{w}}$ is a direction in space with which $\widehat{\mathbf{v}}$ wants to be aligned to obtain the coordinates $\mathbf{x}_{\text{rot}}$, a rotation can be performed in a plane whose normal—the axis of rotation—is defined by the cross product

$$\widehat{\mathbf{n}} = \widehat{\mathbf{v}} \times \widehat{\mathbf{w}} \,. \tag{3.49}$$

On the other hand, the angle between $\widehat{\mathbf{v}}$ and $\widehat{\mathbf{w}}$ in the plane defined by $\widehat{\mathbf{n}}$ can be derived from their dot product, as

$$\theta = \arccos\left(\widehat{\mathbf{v}} \cdot \widehat{\mathbf{w}}\right) \,. \tag{3.50}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

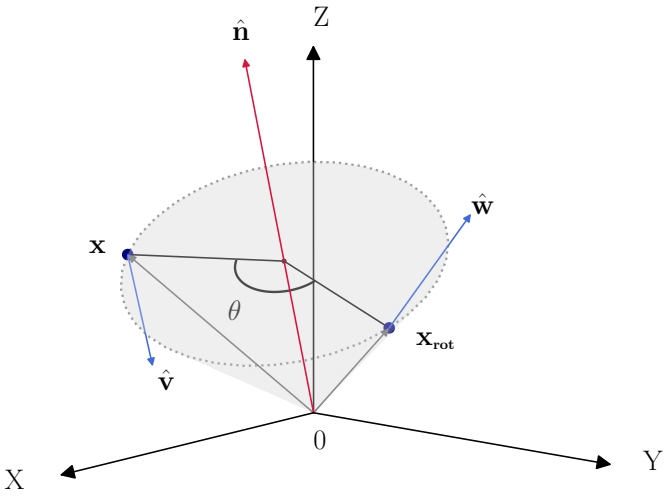


Figure 3.19: Rotation of a system of particles. To orient the particle $\mathbf{x}$ such as its associated vector $\widehat{\mathbf{v}}$ becomes parallel to $\widehat{\mathbf{w}}$ requires to apply a rotation of an angle θ (Eq. 3.50) around the axis defined by $\widehat{\mathbf{n}}$ (Eq. 3.49).

Thus, $\mathbf{x}_{\text{rot}}$ can be obtained through a rotation of $\mathbf{x}$ an angle θ around $\widehat{\mathbf{n}}$ (Figure 3.19), which can be easily done by applying Rodrigues' rotation formula[270]

$$\mathbf{x}_{\text{rot}} = \mathbf{x}\cos\theta + (\widehat{\mathbf{n}} \times \mathbf{x})\sin\theta + \widehat{\mathbf{n}}\,(\widehat{\mathbf{n}} \cdot \mathbf{x})\,(1 - \cos\theta)\ . \tag{3.51}$$

It must be taken into account that the rotation axes in the previous formula pass through the origin, so it is many times necessary to combine the rotation with previous and later translations.

This transformation was not implemented from scratch, unlike the previous ones, but instead we used the *from_rotvec* function of the Scipy Python library.[246]

3.3.2 General analyses

The previous transformations were used to facilitate the analysis of the MD simulation results. Next, the list of analyses performed on the trajectories will be described. Unless stated otherwise, all of them were implemented from scratch in custom Python scripts, using the libraries Numpy,[244] Scipy,[246] NetworkX,[271] Pandas,[245] and MDAnalysis.[242,243]

Hydrogen bonds Hydrogen bonds (H-bonds) are a type of interaction of major importance in biological systems. It is an inter- or intramolecular interaction of mainly electrostatic origin. It occurs when a hydrogen atom, covalently bonded to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

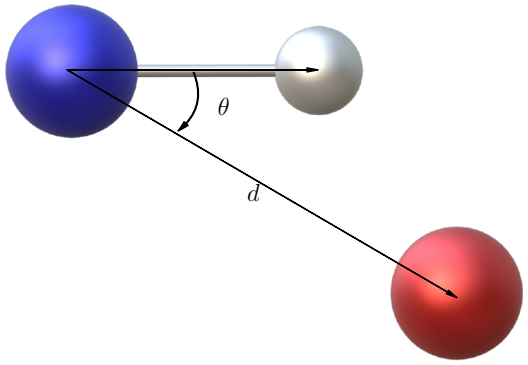


Figure 3.20: Formation of H-bonds. The donor (blue), hydrogen (white) and acceptor (red) atoms are forming an H-bond in the simulation when the distance between the donor and the acceptor d is less than 3.5 Å and the angle between hydrogen and acceptor relative to the donor θ is below 30°. The figure illustrates this limit case.

an electronegative atom (donor, typically O, N, or F) is also attracted to the electron pair in another electronegative atom (acceptor).

Neither of the force fields used in this work include additional terms to the model of H-bonds; instead they appear as a result of the combination of non-bonded interactions.

The decision on the existence of an H-bond involving a triad of atoms was made based on geometric criteria: If $\mathbf{x}_{\mathrm{H}}$, $\mathbf{x}_{\mathrm{d}}$ and $\mathbf{x}_{\mathrm{a}}$ are the positions of the hydrogen atom, and the donor and acceptor candidates,* an H-bond involving these 3 atoms occurs whenever the acceptor–donor distance is below 3.5 Å and the angle defined with the center in the donor is below 30° (Figure 3.20):

$$\text{H-Bond}(\mathrm{H},\mathrm{d},\mathrm{a}) \;\Leftrightarrow\; \begin{cases} \|\mathbf{r}_{\mathrm{a,d}}\| < 3.5 \text{ Å} \\ \arccos(\widehat{\mathbf{r}}_{\mathrm{H,d}} \cdot \widehat{\mathbf{r}}_{\mathrm{a,d}}) < 30^\circ \end{cases} , \tag{3.52}$$

where $\mathbf{r}_{i,j} = \mathbf{x}_i - \mathbf{x}_j$. The values used were the same as those in the analysis tool implemented in GROMACS.[256] The code developed to evaluate the H-bonds checks for all possible pairs of donor–acceptor candidates based on both the species and their distance, for which the angle is then computed.

Contacts Two particles, a and b, are in contact if their separation is below a certain threshold d_{cont}:

$$\text{Contact}(\mathrm{a},\mathrm{b}) \;\Leftrightarrow\; \|\mathbf{r}_{\mathrm{a,b}}\| < d_{\mathrm{cont}} \,. \tag{3.53}$$

*The both donor and acceptor can only be N or O (other electronegative species do not appear in our systems), the former being bonded to an H atom.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The criterion used in this work was to set $d_{\text{cont}} = 6$ Å, which is slightly higher than the approximate half of the cutoff distance used for long-range interactions.* By extension, two fragments or molecules are in contact when Eq. 3.53 is satisfied for any two particles in these groups. The class *NeighborSearch* of MDAnalysis was used to evaluate the number of contacts.

3.3.3 Membrane-specific analyses

While the previous analyses were applied to many different types of molecules, the following are specific to lipid bilayer systems.

Leaflet identification Identifying which lipids belong to each of the leaflets in the bilayer is not trivial in most cases. Fortunately, the bilayers in the simulations performed maintained a flat, symmetric, and mostly unperturbed structure. The clear distribution of lipids in the two leaflets allow the use of a simple approach to identify them: For a membrane composed of J different species, having N_j lipids of the species j, let z_i^j be the Z coordinate of each of the lipids. The projection of the center of geometry (centroid) of the membrane onto the Z axis is the mean position of all z_i^j

$$Z_{\text{mem}} = \frac{1}{N_{\text{lip}}} \sum_{j=1}^{J} \sum_{i=1}^{N_j} z_i^j \,, \tag{3.54}$$

being N_{lip} the total number of lipids in the bilayer. If the membrane is symmetric—or almost—and flat—or almost—, which means that the number of lipids of each species in each leaflet is the same, then Z_{mem} is the midpoint of the lipid bilayer, close to the end of the lipid tails. The position of the phosphorus atom—or the phosphate-associated bead in CG—was selected for reference of the lipids.

Thus, the lipids belonging to either leaflet (upper, UL; or lower, LL) can be identified by determining whether z_i^j lies above or below the midpoint of the membrane Z_{mem} (Figure 3.21):

$$\text{UL} = \left\{(i,j) \;\middle|\; j \in [1,J],\ i \in [1,N_j],\ z_i^j > Z_{\text{mem}}\right\} \tag{3.55a}$$

$$\text{LL} = \left\{(i,j) \;\middle|\; j \in [1,J],\ i \in [1,N_j],\ z_i^j < Z_{\text{mem}}\right\} \tag{3.55b}$$

Area per lipid The area per lipid is the inverse of the lipid concentration per unit of area in the membrane: the higher the lipid density, the lower the separation between individual molecules, so will be the fraction of surface area associated with each of them.

*This cutoff ranges between 0.9 and 1.2 nm, depending on the force field.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

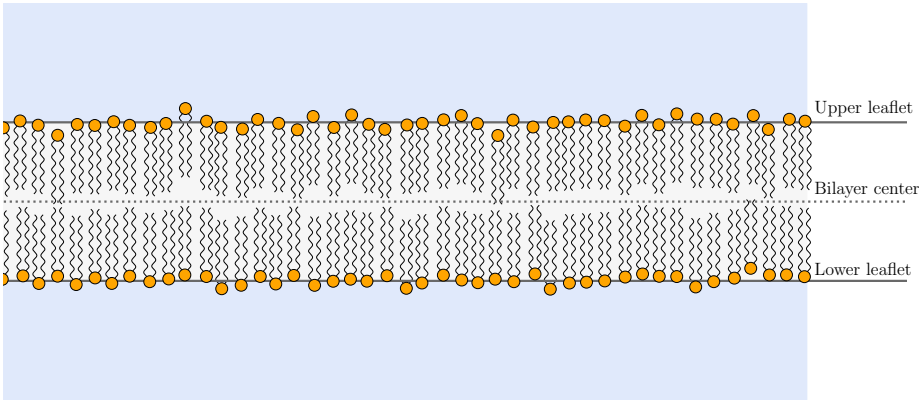


Figure 3.21: Identification of the lipids in each leaflet. Using the phosphate groups as reference for each lipid, the position of the center of the bilayer (dashed black line) is computed. Then, the lipids belonging to the upper leaflet are those whose phosphates are above the center of the bilayer, while the lower leaflet corresponds to those that are below. The thickness of the bilayer can be computed as the difference between the mean position of the phosphates in each of the leaflets (solid black lines).

The region of the membrane surface closer to each lipid than to any other is given by the Voronoi tessellation of the surface (Figure 3.22). If $\mathbf{s}$ is the position of a lipid, and S is the set of all lipids, then its Voronoi region is—in two dimensions—

$$R_{\mathbf{s}} = \left\{ \mathbf{x} \in \mathbb{R}^2 \;\middle|\; \|\mathbf{x} - \mathbf{s}\| \leq \|\mathbf{x} - \mathbf{s}'\| \quad \forall \mathbf{s}, \mathbf{s}' \in S, \mathbf{s}' \neq \mathbf{s} \right\} . \tag{3.56}$$

This method can be applied to both a flat membrane with PBCs or to any convex geometry, such as a vesicle, after a parametrization of the 3-dimensional structure on a 2D surface.[272,273]

For the particular case of a flat and symmetric lipid bilayer, if A_i^j is the area of the Voronoi region associated with the i-th lipid of the j-th species,

$$\sum_{j=1}^{J} \sum_{i=1}^{N_j} A_i^j = 2A_T \,, \tag{3.57}$$

being A_T the area of the plane of the simulation cell along which the membrane extends: the sum of the areas associated with each of the lipids in each leaflet is equal to the total area of the surface (Figure 3.22), so adding the areas of the lipids in both leaflets will give rise to twice its value. Therefore, the mean area per lipid across all species is equal to

$$\langle A_L \rangle = \frac{1}{N_{\text{lip}}} \sum_{j=1}^{J} \sum_{i=1}^{N_j} A_i^j = 2\frac{A_T}{N_{\text{lip}}} \,. \tag{3.58}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

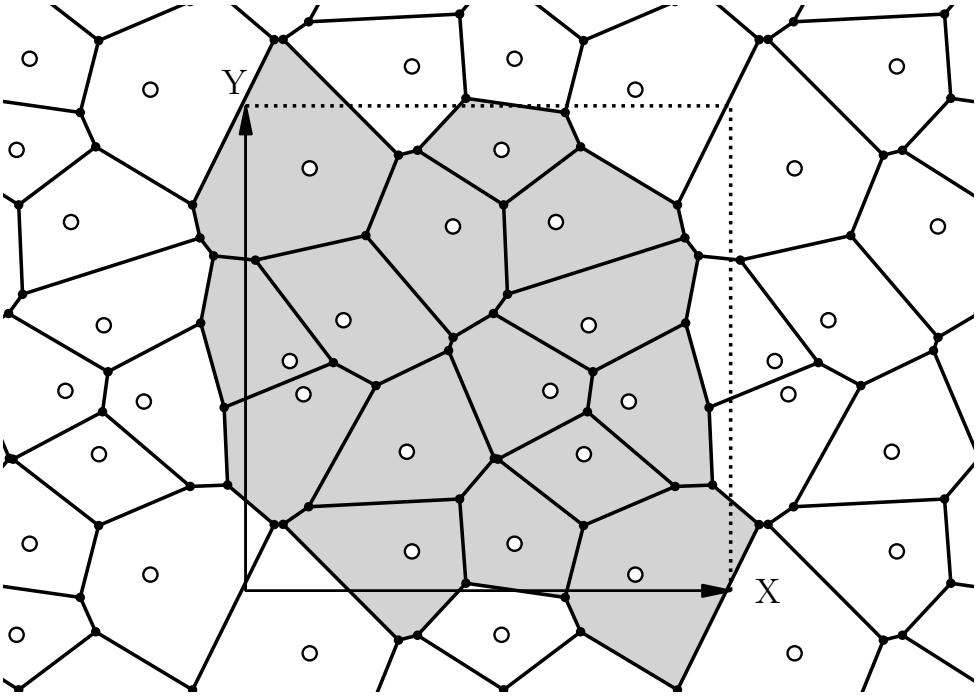


Figure 3.22: Schematic representation of the area per lipid. The Voronoi tessellation of the membrane (solid black lines) can be used to define the area of the simulation cell (dashed lines) closer to one lipid (white circles) than to any other. The total area of each of the lipids contained within the original simulation cell (gray shaded areas) is equal to the area of the cell in the plane where the tessellation is computed. In consequence, the area per lipid averaged over all lipids when both leaflets are considered is equal to twice the area of the XY plane divided by the total number of lipids.

The mean area per lipid for a flat bilayer depends only on the size of the simulation cell, and it is agnostic to the lipid composition or any flip-flop processes that alter the ratio of lipids between leaflets. It is possible to calculate the mean area per lipid for each individual leaflet as

$$\langle A_L \rangle^l = \frac{A_T}{N^l_{\text{lip}}}\,, \tag{3.59}$$

where the number of lipids in the leaflet l, N^l_{lip}, is the cardinality—the number of elements in a set—of the leaflet, as given in Eq. 3.55.

In the NPT ensemble, the dimensions of the simulation cell change over time, so $\langle A_L \rangle = \langle A_L \rangle(t)$. In the future, the result of Eq. 3.58 will be simply referred to as the area per lipid A_L, leaving the term "mean" to be used only for the time averages of this quantity.

Bilayer thickness Determine the extension of a membrane in the transverse direction can be somewhat arbitrary: one may choose to compute it as the maximum

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

distance between any two atoms along the normal—which will be very sensitive to the movement of individual lipids—, the distance between the points at which the density of water drops below a certain value—a magnitude which is noisy when computed for a single state—, etc.

As upon visual inspection, all membranes appeared to maintain a flat geometry during the simulation time, the instantaneous bilayer thickness was calculated as the difference in the mean Z coordinate of the phosphorus atoms in both leaflets:

$$\Delta Z_{\text{mem}} = \left(\frac{1}{N_{\text{lip}}^{\text{UL}}} \sum_{i,j \in \text{UL}} z_i^j \right) - \left(\frac{1}{N_{\text{lip}}^{\text{LL}}} \sum_{i,j \in \text{LL}} z_i^j \right) , \tag{3.60}$$

where $N_{\text{lip}}^{\text{UL}}$ and $N_{\text{lip}}^{\text{LL}}$ are, respectively, the number of lipids in the upper and lower leaflets (Figure 3.21). In CG simulations, as with leaflet identification, beads representing the phosphate groups were used instead.

3.3.4 Peptide-specific analyses

One of the most relevant aspects of peptides is their structure. The determination of the structure of a peptide arises from the identification of certain patterns found in the distribution of their amino acids in space: helices of different radii and pitches, parallel or antiparallel sheets, coils, etc. There is no unique way in which the structure can be determined, rather different conventions on experimental geometries.

For the present work, the α-helix pattern is key. It is characterized by a distribution of amino acids in a right-handed helix conformation, stabilized by the formation of H-bonds within the backbone, between amino acids 4 residues apart ($i \to i+4$). An idealized α-helix has a radius of 2.3 Å and a pitch—displacement in the axes after a full rotation—of 5.4 Å.[274]

DSSP Stands for "Define Secondary Structure of Proteins", and it is one of the most widely used algorithms for the assignment of the secondary structure of a protein. Originally developed for its application in protein crystallographic data,[275] each residue of the sequence is labeled after one of the following structures: α-helix, 3_{10}-helix, π-helix, β-ladder, β-sheet, bend or coil.

The α-helix motif is associated with the existence of a specific type of H-bond within the backbone. DSSP applies a semi-empirical energetic criterion to determine the presence of a H-bond between two residues:

$$\text{H-bond}_{\text{DSSP}}(i,j) \;\Leftrightarrow\; 27.888 \left(\frac{1}{\|\mathbf{r}_{\text{NO}}\|} + \frac{1}{\|\mathbf{r}_{\text{CH}_\text{N}}\|} - \frac{1}{\|\mathbf{r}_{\text{OH}_\text{N}}\|} - \frac{1}{\|\mathbf{r}_{\text{CN}}\|} \right) < -0.5 \,, \tag{3.61}$$

where the distances involving the C and O atoms of the backbone of residue i, and the N and H_N atoms of the backbone of residue j are given in Å. Both terms in Eq.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.61 have units of kcal mol^{-1}. An H-bond between residues i and $i+4$, is said to be a 4-turn at position i.

The minimal α-helix spans the residues from i to $i+3$ (both included) when consecutive 4-turns are identified at positions $i-1$ and i; that is, there are H-bonds that involve the C and O of the residues $i-1$ and i and N and H_N of $i+3$ and $i+4$, respectively. This minimal 4-residue-long α-helix grows if more consecutive 4-turns occur at positions $i+1$, $i+2$, etc.; or by the concatenation of contiguous helices.

This algorithm assigns only one type of structure to each of the residues, prioritizing the α-helix above any other structure. The fraction of α-helix in the structure given by DSSP is the discrete quantity

$$\alpha_{\text{DSSP}} = \frac{1}{L}\sum_{i=1}^{L} n_i^{\alpha}, \qquad \alpha_{\text{DSSP}} \in \left[\frac{4}{L}, \frac{L-2}{L}\right]^{*} \tag{3.62}$$

where n_i^{α} is 1 if residue i is assigned to a α-helical pattern and 0 otherwise.

αRMSD Pietrucci and Laio[276] proposed a metric to determine how close a fragment of a peptide structure is to a β-sheet by comparing its structure with respect to an ideal reference. They also indicated how to extend the same concept for α-helical structures: Let $\mathbf{X}^i$ be the coordinates of the backbone and the C_β atoms of 6 consecutive residues within a peptide chain, starting at the residue i. Then, the quantity

$$n\left(\mathbf{X}^i, \mathbf{\Omega}\right) = \frac{1 - \text{RMSD}\left(\mathbf{X}^i, \mathbf{\Omega}\right)^8}{1 - \text{RMSD}\left(\mathbf{X}^i, \mathbf{\Omega}\right)^{12}}, \tag{3.63}$$

determines how close the coordinates of $\mathbf{X}^i$ are to the ideal structure $\mathbf{\Omega}$. To measure their deviation, the root mean square deviation (RMSD) between the coordinates is computed:

$$\text{RMSD}\left(\mathbf{X}^i, \mathbf{\Omega}\right) = \sqrt{\frac{1}{6N}\sum_{j=1}^{6N}\left\|\mathbf{x}_j^i - \mathbf{x}_j^{\Omega}\right\|^2} \tag{3.64}$$

where $\mathbf{x}_k^i$ and $\mathbf{x}_k^{\Omega}$ are the positions of the k-th atoms of $\mathbf{X}^i$ and $\mathbf{\Omega}$, respectively, and 6 comes from the number of residues being compared. The closest $\mathbf{X}^i$ is to $\mathbf{\Omega}$, the lower the RMSD and so the closest n is to one. Doing this for the whole peptide

*One could expect this range to be $[0, 1]$, but as the smallest helix possible consists in 4 residues the lower bound is $4/L$. As mentioned above, a helix starting in position i requires a 4-turn in $i-1$, so a helix starting in the first residue is impossible; on the other hand, a 4-turn in position $L-4$ is the closest to the C-terminus possible, so the helix can extend up to $L-1$. Thus, the ends cannot be involved; hence the upper bound of $(L-2)/L$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

gives an estimate on how similar its structure is to a full α-helix

$$\alpha_{\text{RMSD}} = \frac{1}{L-5} \sum_{i=1}^{L-5} n\left(\mathbf{X}^i, \mathbf{\Omega}\right), \tag{3.65}$$

Unlike α_{DSSP}, α_{RMSD} is a continuous value that ranges from 0 to 1 as the geometries become closer to each other.

The coordinates of the molecules are referred to an arbitrary system of reference, so the calculation of their RMSD usually implies the alignment of the structures with the reference, so the value obtained is minimal. Doing this is not only expensive in terms of the resources used, but also adds an iterative process that may not converge to the minimum easily.

To avoid this, the matrix of interatomic distances $\mathbf{R}^i$ within the fragment was calculated. Its components are

$$\left(\mathbf{R}^i\right)_{nm} = r^i_{nm} = \|\mathbf{x}^i_n - \mathbf{x}^i_m\| \tag{3.66}$$

the distance between the atoms of the indices n and m. Doing the same for the reference coordinates, $\mathbf{\Omega}$, lead to $\mathbf{R}^\Omega$. Unlike in the original proposal, we decided to compare only the backbone atoms, to avoid dealing with an exception for the C_β atom when the fragment contains a Gly,* and we also tried to apply this metric to only the C_α atoms. Thus, a distance-based RMSD (dRMSD) was calculated as[277]

$$\text{dRMSD}\left(\mathbf{X}^i, \mathbf{\Omega}\right) = \sqrt{\frac{2}{6N(6N-1)} \sum_{n=1}^{6N} \sum_{m=n+1}^{6N} \left(r^i_{nm} - r^\Omega_{nm}\right)^2}, \tag{3.67}$$

where $\left(\mathbf{R}^\Omega\right)_{nm} = r^\Omega_{nm}$. The sums in Eq. 3.67 exploit both the symmetry of the interatomic distance matrix, $r^i_{nm} = r^i_{nm}$, and the fact that the elements of the diagonal are zero.

Our implementation of this metric differs from the original proposal (Eq. 3.65) on the use of dRMSD (Eq. 3.67) to compare the structures instead. The reference structure for the ideal helix used derived from the 6OBI record[278,279] from the Protein Data Bank.[280] This structure was selected because it is a long, continuous, and stable α-helix. After discarding the first and last four residues of the structure, $\mathbf{R}^\Omega$ was calculated as the mean of all interatomic distance matrices for the remaining fragments. When comparing only the C_α atoms, $\mathbf{R}^\Omega$ was constructed from the parameters of an ideal helix: a radius of 2.3 Å an angle between the residues of 100° and a distance between consecutive residues along the axis of 1.5 Å.[274]

*None of the peptides in the atomistic simulations contained Pro.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

α-block Fiorin, Klein and Hénin[281] proposed an alternative metric for the determination of the content of α-helix in a structure based on the scoring of the angles between C_α atoms and the formation of H-bonds. Its expression is

$$\alpha_{\text{block}} = \frac{1-\beta}{L-2}\sum_{i=1}^{L-2}\text{angf}(i) + \frac{\beta}{L-4}\sum_{i=1}^{L-4}\text{hbf}(i)\,, \tag{3.68}$$

where β is a parameter used to balance the contributions of the scoring of angles and distances. The default value of $\beta = 0.5$ was used, which means that both terms will contribute equally.

The angular function evaluates how much the angle formed by consecutive C_α atoms, whose indices are $i, i+1, i+2$, deviates from the reference in the ideal helix:

$$\text{angf}(i) = \frac{1-\left(\arccos\left(\widehat{\mathbf{r}}^{\alpha}_{i,i+1}\cdot\widehat{\mathbf{r}}^{\alpha}_{i+2,i+1}\right)-\theta_\alpha\right)^2/\delta\theta_\alpha^2}{1-\left(\arccos\left(\widehat{\mathbf{r}}^{\alpha}_{i,i+1}\cdot\widehat{\mathbf{r}}^{\alpha}_{i+2,i+1}\right)-\theta_\alpha\right)^4/\delta\theta_\alpha^4}\,, \tag{3.69}$$

being $\widehat{\mathbf{r}}^{\alpha}_{i,j}$ the unit vector pointing from the j-th Cα to the i-th. θ_α is the reference angle and $\delta\theta_\alpha$ a tolerance for this angle. Both parameters were assigned the values of 88° and 15°, as in the original implementation.

The H-bond scoring function, on the other hand, compares the distance from the backbone i-th O atom to the $(i+4)$-th N atom against the reference value d_0:

$$\text{hbf}(i) = \frac{1-\left(\|\mathbf{x}^{\text{O}}_i - \mathbf{x}^{\text{N}}_{i+4}\|/d_0\right)^6}{1-\left(\|\mathbf{x}^{\text{O}}_i - \mathbf{x}^{\text{N}}_{i+4}\|/d_0\right)^{12}}\,. \tag{3.70}$$

The reference distance was once again set to the recommended value of 3.3 Å. The sums in Eq. 3.68 iterate over all the possible triads of contiguous C_α atoms and all pairs of O and N atoms 4 residues apart. The values of both angf(i) and hbf(i) range between 0, if angles or distances are far from the reference values, and 1, if they are exactly these values; therefore, the value of α_{block} is also restricted to $(0, 1]$.

Helical axes When the peptide forms a perfect helix, its axis can be used as a reference for the orientation of the whole structure. Its coordinates, $\widehat{\mathbf{h}}$, can be derived from the distribution of the C_α atoms in space by fitting their coordinates by the least-squares method to a three-dimensional straight line. This method is discussed in Section 3.4.1.

The resulting vector of the fitting, $\widehat{\mathbf{h}}_0$, can be oriented in the N-to-C direction by applying

$$\widehat{\mathbf{h}} = \frac{\widehat{\mathbf{r}}_{\text{CN}}\cdot\widehat{\mathbf{h}}_0}{\left|\widehat{\mathbf{r}}_{\text{CN}}\cdot\widehat{\mathbf{h}}_0\right|}\widehat{\mathbf{h}}_0\,, \tag{3.71}$$

where $\widehat{\mathbf{r}}_{\text{CN}}$ is the unit vector that points from C_α of the first residue (N-terminus) to the one of the last (C-terminus).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Distribution of mass. Principal axes of the peptide Conformational changes in the structure of the peptide will impact the way its mass is distributed in space. In MD simulation, atoms are point particles with associated properties, such as mass, charge, etc. Then it is possible to compute the tensor of inertia $\mathbf{I}$ of the peptide as

$$\mathbf{I} = \sum_{i=1}^{N} m_i \left(\|\mathbf{x}_i\|^2 \mathbb{1} - \mathbf{x}_i^{\mathsf{T}} \mathbf{x}_i \right) , \tag{3.72}$$

where the sum iterates over the N atoms in the peptide. Moments of inertia of the peptide are the solutions I_i to the eigenvalue problem.

$$\det \left(I_i \mathbb{1} - \mathbf{I} \right) = 0 \, , \tag{3.73}$$

and their corresponding principal axes of inertia $\mathbf{s}_i$ satisfy

$$\left(I_i \mathbb{1} - \mathbf{I} \right) \mathbf{s}_i = 0 \, . \tag{3.74}$$

Once the tensor of inertia was calculated, its diagonalization was performed using the *diagonalize* function of Sympy. The overall shape of the peptide was approximated by that of its equivalent ellipsoid: a solid with mass $M = \sum_{i=1}^{N} m_i$—the same as the peptide—whose semiaxes coincide with its principal axes of inertia and span a length of

$$c_i = \sqrt{\frac{5}{2M} \sum_{k=1}^{3} (-1)^{\delta_{ik}} I_k} \, . \tag{3.75}$$

The complete derivation of this expression is provided in Appendix A.1. Observing changes in the shape of the peptide is much more intuitive and easy to interpret than doing it for moments of inertia (Figure 3.23).

Distribution of charge. Electric dipole moment All atoms in the system have an associated partial charge whose exact value depends on the force field used for the simulation. Therefore, a conformational change of the peptide will affect not only the distribution of mass but also the distribution of charge, even for a peptide with net zero charge.

Being q_i the partial charge of the i-th atom whose coordinates are $\mathbf{x}_i$, its electric dipole moment $\mathbf{p}$ evaluated in position $\mathbf{r}$ is

$$\mathbf{p}(\mathbf{r}) = \sum_{i=1}^{N} q_i \left(\mathbf{x}_i - \mathbf{r} \right) \, . \tag{3.76}$$

The electric dipole moment is constant—its dependence in $\mathbf{r}$ vanishes—when the net charge of the peptide is zero. In any other case, the reference conditions the value

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

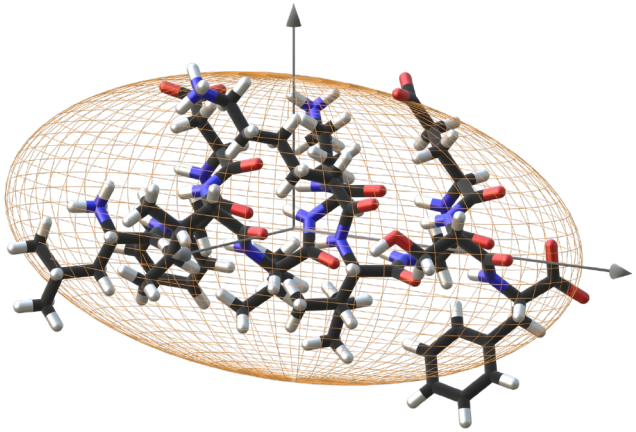


Figure 3.23: The equivalent ellipsoid of the peptide. This solid (orange mesh) has the same mass and inertia moments as the peptide. The length of the semiaxes of this solid are used as a first-order estimation of the dimensions of the peptide. The figure was rendered using Blender.

of $\mathbf{p}$. For consistency among the different properties, the center of geometry of the peptide was used as reference.

When the axis of the peptide is known, it is especially illustrative to decompose the electric dipole moment into a component parallel to the axes $\widehat{\mathbf{h}}$ and another component perpendicular, $\mathbf{p}_{\parallel}$ and $\mathbf{p}_{\perp}$ respectively:

$$\begin{cases} \mathbf{p}_{\parallel} = \left(\mathbf{p} \cdot \widehat{\mathbf{h}}\right) \widehat{\mathbf{h}} \\ \mathbf{p}_{\perp} = \mathbf{p} - \left(\mathbf{p} \cdot \widehat{\mathbf{h}}\right) \widehat{\mathbf{h}} \end{cases} \tag{3.77}$$

Distribution of hydrophobic residues. Hydrophobic dipole moment Analogously to the electric dipole moment, a hydrophobic dipole moment can be computed to determine how the hydrophobic residues are distributed in space. Hydrophobicity is the tendency for a group to prefer a nonpolar environment over a polar one, but there is no unique way to obtain a quantitative measure of this property. Several scales are available for the hydrophobicity of amino acids, calculated on the basis of different criteria.[282]

Being h_i the hydrophobicity of an amino acid, the hydrophobic dipole moment $\boldsymbol{\mu}$ evaluated in $\mathbf{r}$ is

$$\boldsymbol{\mu}(\mathbf{r}) = \sum_{i=1}^{N} h_i \left(\mathbf{x}_i - \mathbf{r}\right) \ . \tag{3.78}$$

The decomposition of the hydrophobic dipole moment into the parallel and transversal component to the axes of the peptide $\widehat{\mathbf{h}}$ is the same as for the electric

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

dipole moment

$$\begin{cases} \boldsymbol{\mu}_{\parallel} = \left(\boldsymbol{\mu} \cdot \widehat{\mathbf{h}}\right) \widehat{\mathbf{h}} \\ \boldsymbol{\mu}_{\perp} = \boldsymbol{\mu} - \left(\boldsymbol{\mu} \cdot \widehat{\mathbf{h}}\right) \widehat{\mathbf{h}} \end{cases} \tag{3.79}$$

As for the values of the hydrophobicity of the residues, the hydrophobicity scale by Fauchere and Pliska[283] was used. Because this property was only calculated for the peptides in the CG simulations, the hydrophobicity was assigned to the position of the bead of the backbone of each residue.

3.3.5 Analyses associated with the peptide–lipid interaction

Minimal distance The most straightforward way to see if a peptide may interact with a lipid membrane is to check how close they are to each other. The minimal distance between the particles of the peptide and the membrane, whose coordinates are, respectively, $\mathbf{x}$ and $\mathbf{X}$, is.

$$d_{\min}\left(\mathbf{x}, \mathbf{X}\right) = \min_{i,j \in N_p, N_m} \left(\|\mathbf{x}_i - \mathbf{X}_j\|\right) , \tag{3.80}$$

where N_p is the total number of particles in the peptide and N_m, in the membrane.

Peptide penetration into the membrane An attractive interaction between the peptide and the membrane can lead to internalization of the first into the second. The depth of the penetration of the peptide will depend on both the peptide, the lipid composition, and the environment.

As by construction the bilayer extended in the XY plane, the simplest way to calculate the penetration of the peptide into the membrane is to measure the difference in the Z coordinate of the center of geometry of the peptide and the membrane.

Let z_p, Z_{UL} and Z_{LL} be the Z coordinate of the center of geometry of the peptide, of the upper and of the lower leaflet, respectively (Eqs. 3.54 together with 3.55). The penetration depth Δz depends on which leaflet the peptide is closer to:

$$\Delta z = \begin{cases} Z_{\mathrm{UL}} - z_p & \text{if } \ |z_p - Z_{\mathrm{UL}}| < |z_p - Z_{\mathrm{LL}}| \\ z_p - Z_{\mathrm{LL}} & \text{if } \ |z_p - Z_{\mathrm{UL}}| > |z_p - Z_{\mathrm{LL}}| \end{cases} \tag{3.81}$$

The value of Δz will be positive when $z_p \in (Z_{\mathrm{LL}}, Z_{\mathrm{UL}})$, which corresponds to the center of geometry of the peptide having crossed the plane of reference of either leaflet (Figure 3.24).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

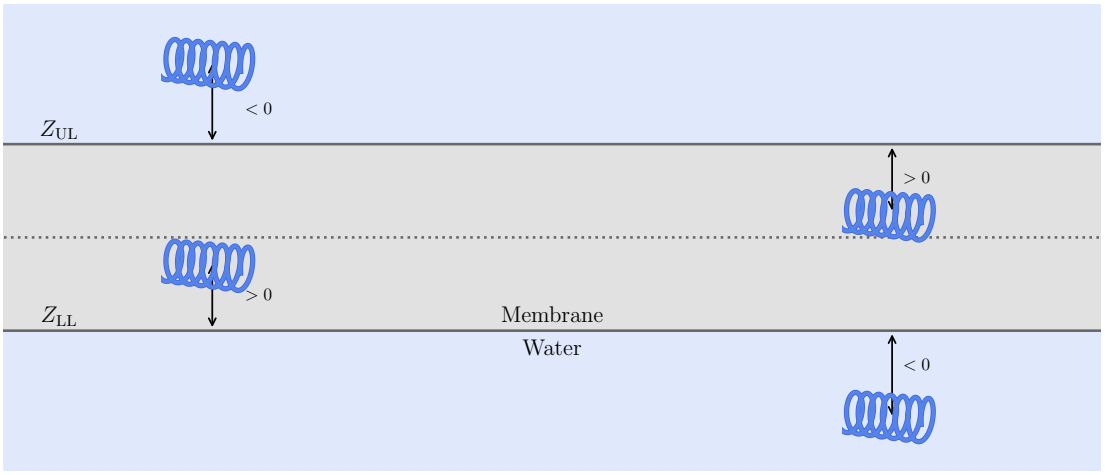


Figure 3.24: Penetration of the peptide into the lipid bilayer. The sign of the penetration depth is positive when the peptide is inserted into the membrane. To achieve this, the way of calculating it must depend in the closest leaflet to the peptide.

Tilt The tilt is defined as the angle formed by the principal axis of the peptide $\widehat{\mathbf{h}}$ and the normal to the membrane. Lipid bilayers have two normals, one per leaflet. Since the bilayers in the simulations performed remained almost perfectly flat across the XY plane throughout the entire trajectory, the membrane normals $\widehat{\mathbf{N}}$ can be approximated by

$$\widehat{\mathbf{N}} = \pm\widehat{\mathbf{k}} \tag{3.82}$$

where $\widehat{\mathbf{k}}$ is the unit vector in the Z direction. The value of the sign will depend on the leaflet the peptide is closer to at each instant: positive for the upper leaflet and negative for the lower leaflet.

The angle between the axis of the peptide and the normal is then

$$\theta = \arccos\left(\widehat{\mathbf{N}} \cdot \widehat{\mathbf{h}}\right) \tag{3.83}$$

The argument of arccos is just the Z-component of $\widehat{\mathbf{h}}$ with positive or negative sign, depending on the normal used. The tilt can span values in the range $[0, \pi]$; 0 if the peptide axis is aligned with the normal with orientation N-to-C and π if the orientation is the opposite. A value of $\theta = \pi/2$ means that the axis is parallel to either surface of the membrane (Figure 3.25).

Spin We refer to the angle of rotation of the peptide around its principal axis as the spin angle. Based on the definition by Esteban-Martín and Salgado,[121] the vector

$$\mathbf{t} = \left(\left(\widehat{\mathbf{h}} \times \widehat{\mathbf{N}}\right) \times \widehat{\mathbf{h}}\right) \tag{3.84}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

is perpendicular to the peptide axis $\widehat{\mathbf{h}}$ and points to the normal of the surface of the membrane, so it changes with displacements of the peptides and any rotation of the axes. Then another vector, perpendicular to $\widehat{\mathbf{h}}$ and associated with the peptide, is needed; that way, only the rotation of the peptide around its axes will change the value of the spin. Choosing the transversal hydrophobic moment $\boldsymbol{\mu}_\perp$ to play this role, the angle with respect to the reference is (Figure 3.25)

$$\psi_0 = \arccos\left(\widehat{\boldsymbol{\mu}}_\perp \cdot \widehat{\mathbf{t}}\right). \tag{3.85}$$

Unlike the tilt, which only has a sense for the range of values $[0, \pi]$ due to the rotational symmetry of $\widehat{\mathbf{h}}$ around the normal, the spin must be evaluated in the whole range $[0, 2\pi)$. Thus, Eq. 3.85 must be corrected for

$$\psi = \left(\frac{\left(\widehat{\boldsymbol{\mu}}_\perp \times \widehat{\mathbf{t}}\right) \cdot \widehat{\mathbf{h}}}{\left|\left(\widehat{\boldsymbol{\mu}}_\perp \times \widehat{\mathbf{t}}\right) \cdot \widehat{\mathbf{h}}\right|} \arccos\left(\widehat{\boldsymbol{\mu}}_\perp \cdot \widehat{\mathbf{t}}\right) \right) \mod 2\pi\,. \tag{3.86}$$

The value of the fraction is 1 when the angle between $\widehat{\mathbf{t}}$ and $\widehat{\boldsymbol{\mu}}_\perp$ is lower than π—cross product parallel to the axis—, or -1 if greater—antiparallel. This extends the possible angle values to $(-\pi, \pi]$. The function $\cdot \bmod 2\pi$ wraps this result in the range $[0, 2\pi)$.

The value of ψ will be π when the transversal hydrophobic moment points to the surface of the membrane.

Translational and rotational diffusion of the peptide The interaction of the peptide with the membrane will condition how the former moves across its surface. Once adsorbed into the membrane, the peptide is assumed to display a nearly-Brownian 2-dimensional motion. Consequently, the probability to observe a displacement of the peptide of magnitude d in a time τ, could be described by the Rayleigh distribution[284]

$$P_{\mathrm{tr}}(d, \tau) = \frac{d}{2D\tau} e^{-d^2/(4D\tau)}\,, \tag{3.87}$$

where τ is the time window such as $d = \|\mathbf{x}(t+\tau) - \mathbf{x}(t)\|$, and D is the translational (or lateral) diffusion coefficient. On the other hand, if the peptide is helical, the rotation of its axis $\widehat{\mathbf{h}}$ between two states separated a short time τ should follow[285]

$$P_{\mathrm{rot}}(\phi, \tau) = \frac{1}{\sqrt{4\pi D_r \tau}} e^{-\phi^2/(4D_r\tau)}\,, \tag{3.88}$$

being

$$\phi = \frac{\widehat{\mathbf{h}}(t) \times \widehat{\mathbf{h}}(t+\tau)}{\left|\widehat{\mathbf{h}}(t) \times \widehat{\mathbf{h}}(t+\tau)\right|} \arccos\left(\widehat{\mathbf{h}}(t) \cdot \widehat{\mathbf{h}}(t+\tau)\right) \tag{3.89}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

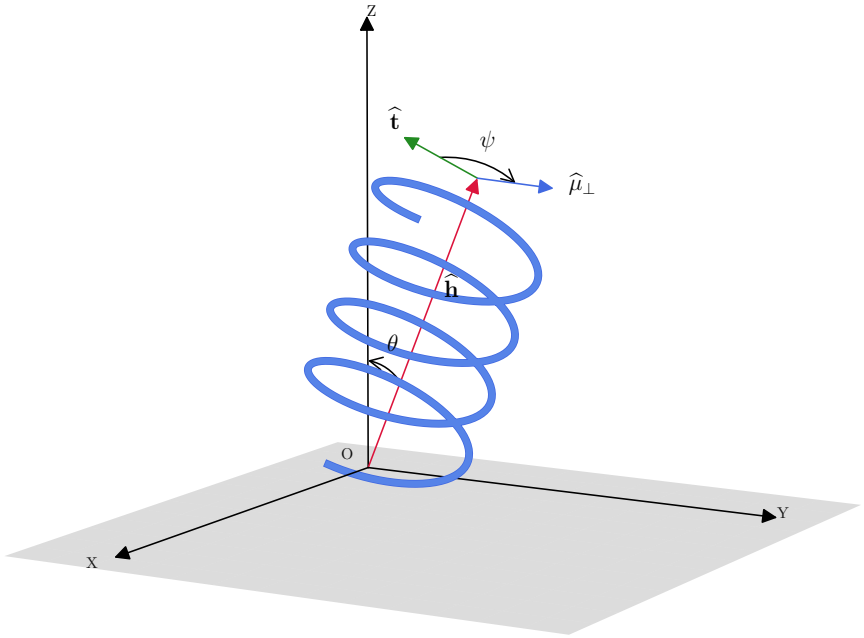


Figure 3.25: The spin and tilt angles. The tilt θ is the angle formed between the axis of the peptide $\widehat{\mathbf{h}}$ and the normal to the membrane (here, the positive Z direction), while the spin ψ is the angle of rotation around the axes of the peptide of its transversal hydrophobic moment $\widehat{\mathbf{h}}_\perp$ and an auxiliary vector $\widehat{\mathbf{t}}$.

the angle between these two states of the axis in the range $(-\pi, \pi]$, and D_r the rotational diffusion coefficient. These angles were be computed for the projections of the axis onto the surface of the membrane.

The full derivation of the previous equations can be found in Appendix A.2. The histograms of the displacement of the center of mass of the peptide and of the rotation of the axis of the peptide for a certain τ were fit by the least-squares method to Eqs. 3.87 and 3.87 to estimate the values of D and D_r for each simulation (Figure 3.26).

Density of molecules The density in a region can be estimated as the ratio between the total mass of particles in a region and the volume of such a region. The density inside the simulation cell can be computed by first splitting its volume along one or two dimensions and computing the density of different types of molecules in their interior.

The density profile is calculated by first splitting the simulation cell volume into thin slices along the Z direction, then computing the time-averaged mass density of different molecules over a fragment of the trajectory, and representing it against the location of the slice in the box. This illustrates how the mass is distributed, on

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

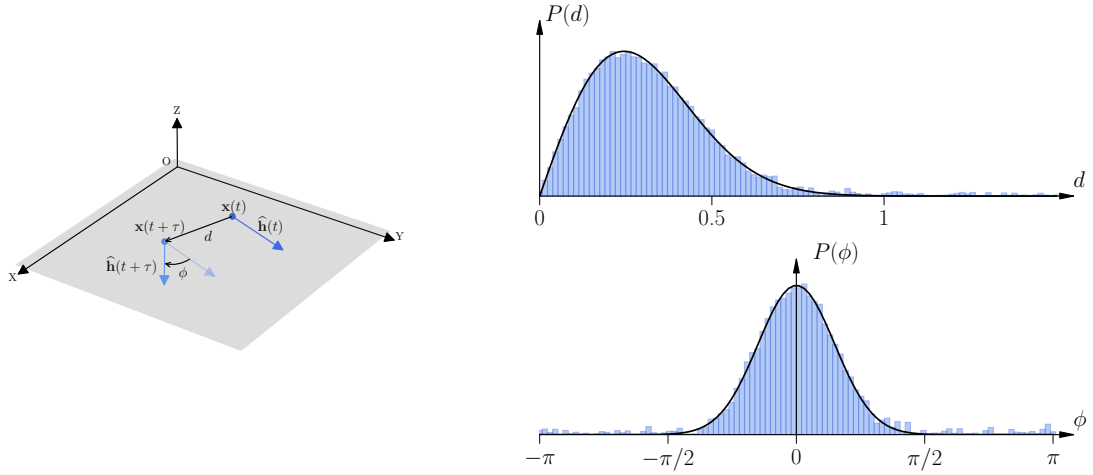


Figure 3.26: Schematic translational and rotational diffusion. (Left) In a time window τ, the displacement of the the centroid of the peptide in the plane of the membrane (gray plane) is d, while its helical axis rotates an angle ϕ. (Right) The probability distribution of the displacement (upper) and rotation of the axis (lower) can be used to estimate the value of the translational and rotational diffusion coefficients.

average, in the Z dimension.

The lipid density map, on the other hand, is used to observe the distribution of lipids around the peptide when it is adsorbed into the membrane. For this reason, each state of the trajectory is rotated so that the peptide axis is oriented in the same direction—Y axis, by choice. The positions of the phosphate groups were used as a reference for the position of the peptide; the value for each lipid and frame is stored, and three density maps are generated: one for the upper leaflet, the other for the lower leaflet, both representing the density of lipids in the XY plane; and a third with the lateral density in the YZ plane.

3.4 Statistical methods

Most of the previous analyses generated a time series of the evolution of the different magnitudes studied throughout the simulation. Further analyses are necessary to extract statistically significant conclusions about the behavior of molecules in the simulated systems.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.4.1 General statistical concepts

Ensemble and temporal average It has been several times that the formula for the average value of a magnitude over a set of elements—called the ensemble* average—has been used

$$\langle x \rangle = \frac{1}{N} \sum_{n=1}^{N} x_n \,. \tag{3.90}$$

If a magnitude x is a function of time, $x = x(t)$, its temporal average over a period T is defined as

$$\bar{x} = \frac{1}{T} \int_{t_0}^{t_0+T} x(t')dt' \,. \tag{3.91}$$

However, the sampling of magnitudes in MD simulation are not continuous, but are evaluated at discrete time intervals. Therefore, the practical implementation of the temporal average is done as the ensemble average, so

$$\bar{x} = \frac{1}{T} \int_{t_0}^{t_0+T} x(t')dt' \approx \frac{1}{N} \sum_{n=1}^{N} x(t_0 + n\Delta t) \,, \tag{3.92}$$

where $T = N\Delta t$. Here Δt is not necessarily the time step of the simulation: as the amount of data generated in a single simulation is huge, the states of the system are stored every several time steps. For the analysis, even more intermediate states can be excluded without significantly altering the results. Thus, Δt refers to the separation between consecutive states considered for the analysis.

Variance and standard deviation The variance is the expected value of the squared deviation of the values of a sample from the mean. The variance of a sample is estimated as

$$s^2(x) = \frac{1}{N-1} \sum_{n=1}^{N} \left(x_n - \langle x \rangle\right)^2 \,. \tag{3.93}$$

The variance s^2 is here the squared of the standard deviation s, given by

$$s(x) = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N} \left(x_n - \langle x \rangle\right)^2} \,. \tag{3.94}$$

Directional statistics The previous formulas are widely known and used, but they cannot be applied in all cases. Let us imagine an angular magnitude ϕ, periodic

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

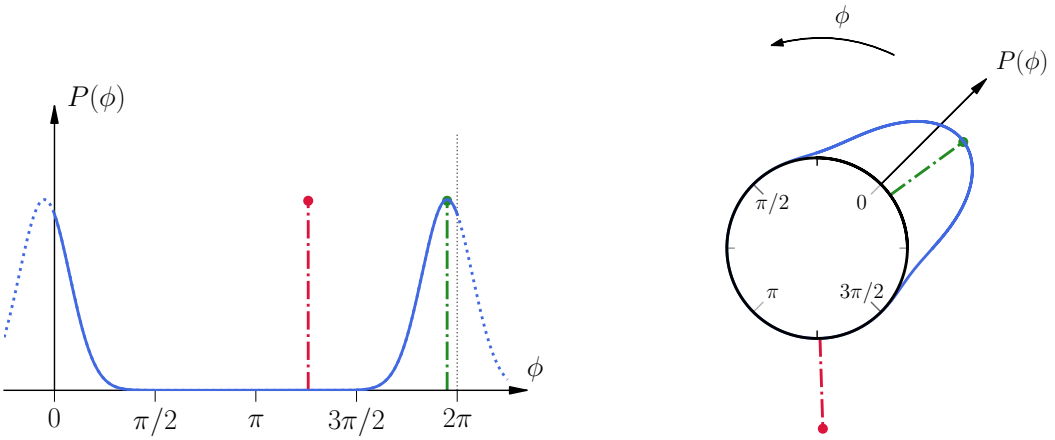


Figure 3.27: Averages in periodic domains. (Left) Computing the expected value of normal distribution (blue line) with mean $\phi \approx 0$ (green vertical line) will lead to an erroneous result (red vertical line) if Eq. 3.90 is used. (Right) The same curves as in the left panel, but in a polar representation.

in $[0, 2\pi)$ and symmetrically distributed around $\phi \approx 0$. Applying Eq. 3.90 to compute the average will result in $\bar{\phi} \approx \pi$, a clearly wrong answer (Figure 3.27).

Directional (or circular) statistics provides the tools to deal with such magnitudes.[286] Calculation of the mean and variance of the samples is done by a transformation into the space of complex numbers. The mean is computed as

$$\bar{\phi} = \arg\left(\frac{1}{N}\sum_{n=1}^{N} e^{i\phi_n}\right), \tag{3.95}$$

where arg() is the function that returns the argument of a complex number. The result is restricted to the interval $[0, 2\pi)$. The variance of the sample, on the other hand, can be estimated by

$$s^2(\phi) = 1 - \left\| \frac{1}{N}\sum_{n=1}^{N} e^{i\phi_n} \right\|, \tag{3.96}$$

being $\|\cdot\|$ the modulus of the complex number, so if $r \in \mathbb{C}$, $r = \|r\|e^{i\arg(r)}$. In circular statistics, the variance is no longer the square of the standard deviation, as the latter is estimated as

$$s(\phi) = \sqrt{-2\ln\left\| \frac{1}{N}\sum_{n=1}^{N} e^{i\phi_n} \right\|} \tag{3.97}$$

[*] Ensemble can refer to both the concept from statistical mechanics and to a collection of elements.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Linear regression A linear regression is a model to estimate how one or more variables (independent) induce a change in another variable (dependent), these changes being proportional to the change experimented by the independent variables. The most simple case connects an independent variable x with a dependent variable y:

$$f(x) = \beta_0 + \beta x \,. \tag{3.98}$$

Here, β_0 is the intercept and β the slope. A generalization of the previous model to N independent variables x_i, $i = 1, \ldots, N$, and has the form

$$f(\mathbf{x}) = \beta_0 + \sum_{i=1}^{N} \beta_i x_i + \sum_{i=1}^{N-1} \sum_{j=i}^{N} \beta_{ij} x_i x_j \,, \tag{3.99}$$

where β_i is the slope associated with x_i, and β_{ij} is the interaction term between x_i and x_j. The interaction terms appear when the effect of an independent variable in y is conditioned by the value of another (Figure 3.28).

For a collection of M data points, $(x_{m1}, \ldots, x_{mN}, y_m)$, $m = 1, \ldots, M$, where $\mathbf{x}_m = (x_{m1}, \ldots, x_{mN})$ is the vector with the values of the independent variables and y_m the value of the dependent variable, the values of the K parameters in the model β_k (where the intercept, slopes, and interaction terms can be included) can be estimated if $K < M$. If so, the least-squares fit of the data points to the model guarantees that the optimal values of the coefficients are found when the sum of the squares of the residuals S is minimal. The residuals are the difference between the dependent variable and the value given by the model,

$$r_m = y_m - f(\mathbf{x}_m) \qquad\qquad m = 1, \ldots, M \tag{3.100}$$

As this metric S,

$$S = \sum_{m=1}^{M} r_m^2 = \sum_{m=1}^{M} (y_m - f(\mathbf{x}_m))^2 \tag{3.101}$$

depends on the parameters of the model, $S = S(\boldsymbol{\beta})$, finding its minimum leads to the solution of a system of K equations:

$$\sum_{m=1}^{M} r_m \frac{\partial r_m}{\partial \beta_k} = 0 \qquad\qquad k = 1, \ldots, K \tag{3.102}$$

Along this work different practical implementations of the least-squares method were used: For all the cases in Section 4.1, and for the parametrization of geometries from a series of coordinates in Sections 4.2 and 4.3, the *least_squares* function of Scipy[246] was used with its default values default parameters[287]; for the study of the relationship between variables in Section 4.3 the OLS class of the statsmodels[288] library was used instead, as it allows the addition of interaction terms and also provides an estimate of the error in the coefficients and applies some statistical tests on them.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

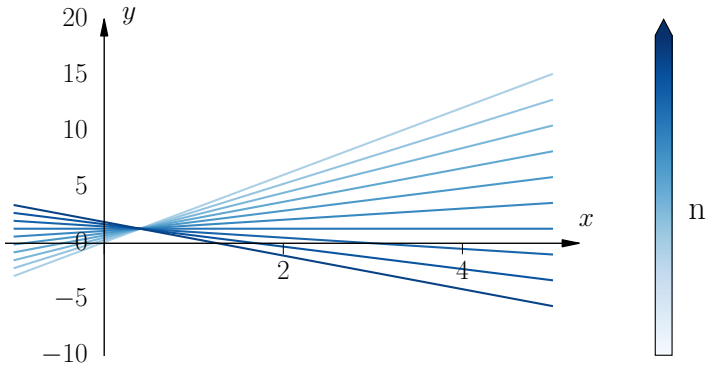


Figure 3.28: Multiple linear regression. In this example, x and n are the independent variables. As the value of n change, so will the intercept and the slope line.

Bootstrap The calculation of both fitting parameters and their uncertainties in regression models depends on whether the residuals are normally distributed.*

Namely, for the estimation of the fitting parameters of simple regression models in Section 4.1 the bootstrap method to the residuals was applied. Bootstrap[289] is a method to study the distribution of a statistic by resampling a data set. If the original data set contains M measurements, B new samples are extracted with replacement i.e. the same data point may appear more than once. Then a certain magnitude of interest is calculated for each of the B samples, so the properties of this magnitude can be inferred from this newly generated distribution.

In the case of residual bootstrap[290] for a simple linear regression (Eq. 3.98), the algorithm is as follows: First, an estimation of the parameters $\tilde{\beta}_0$ and $\tilde{\beta}_1$ and the residuals (Eq. 3.100) for the original data set were obtained by least-squares minimization. The residuals were then resampled, so for the m-th element a new dependent variable was generated as

$$y'_m = \tilde{\beta}_0 + \tilde{\beta}_1 x_m + r_m \,, \tag{3.103}$$

where x_m is the m-th measurement of the independent variable. The b-th (re)sample of M elements (x_m, y'_m) was then again fitted to obtain the bootstrap estimates of the parameters $\tilde{\beta}_0^b$ and $\tilde{\beta}_1^b$. Finally, the value of the parameters and their error were estimated from the sample of bootstrap estimates; for a number of resamples large enough under the central limit theorem, $\tilde{\beta}_i^b$ will be close to a normal distribution from which the mean and standard deviation can be derived.

*In many cases the distribution of the error may still be non-normal and the results will still hold.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Coefficient of determination The coefficient of determination r^2 indicates how much of the variation of the dependent variable is explained by a model,

$$r^2 = 1 - \frac{\sum_{m=1}^{M} (y_m - f(\mathbf{x}_m))}{\sum_{m=1}^{M} (y_m - \langle y \rangle)} \tag{3.104}$$

A value of the coefficient of determination of 1 indicates that model fits perfectly the data; in general, the higher the value of r^2, the better the model is able to explain the variability on the dependent variables.

For a simple linear regression between one independent variable and another dependent variable, the value of r^2 becomes the Pearson correlation coefficient, an indicator of the correlation between the variables, how well the data *could be* explained though a linear relation. For any other model, this does not hold.

Parentheses notation To express the value of an estimated parameter together with its uncertainty—standard deviation—the more compact parentheses notation was employed in this work: If the value of a parameter was $x = 1.343$ and its standard deviation was $u(x) = 0.034$, then it could be expressed as $x = 1.343(34)$. The number in parentheses corresponds to the uncertainty and refers to the last digits of the parameter value. Both the parameter and its uncertainty are always rounded to the same number of decimal places, so the uncertainty is rounded to two significant figures.

3.4.2 Statistical hypothesis tests

A statistical hypothesis test is a method used to determine whether the available data verify a certain hypothesis. The general framework is as follows. First, an statement about the data must be made, the **null hypothesis** (H_0). This null hypothesis involves a property of the population behind the sample and has an associated **alternative hypothesis** (H_1) that negates the former. Then, a statistic whose probability distribution under H_0 is known must be selected. The value of the statistic for the sample is then calculated, and the probability of having obtained this value can be inferred from its probability distribution: If the probability is sufficiently low, this value is unlikely to appear if H_0 is true, so H_0 can be rejected in favor of H_1 (Figure 3.29). This probability is the ***p*-value** of the test, and the rejection threshold must be made by a prior selection of a **significance threshold** $\boldsymbol{\alpha}$. Not rejecting H_0 does not mean accepting it as true, it means that not enough evidence has been found to discard it.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

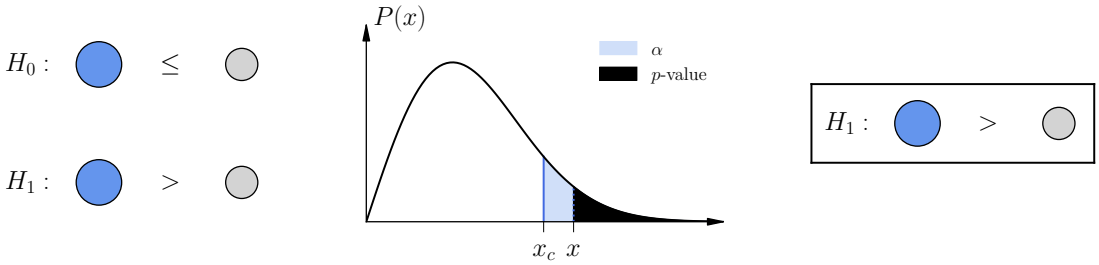


Figure 3.29: Statistical hypothesis tests. (Left) A null hypothesis about the data H_0 is formulated, together with an alternative hypothesis H_1. A statistical for which its probability density function under the null hypothesis $P(x)$ is known is selected and its value is calculated. (Center) The probability of obtaining such value of the statistical—p-value—is computed, and it is compared against a significance threshold α selected beforehand. (Right) If $p < \alpha$ H_0 can be rejected in favor of H_1.

The selection of the significance threshold is determined by the rate of false positives—H_0 is rejected despite being true, also known as type I errors—allowed: the more conservative, the lower the value of α, but also less statistically significant results will be obtained. We have set the value of $\alpha = 0.05$ for all tests performed, unless otherwise stated.

The way the p-value is calculated depends both on the distribution of the test statistic and on the nature of H_0. For a value of the statistic x distributed according to the function f in the domain (A, B):

$$p = \begin{cases} \int_x^B f(t)dt & \text{in a one-sided right-tail distribution} \\ \int_A^x f(t)dt & \text{in a one-sided left-tail distribution} \\ 2\min\left(\int_x^B f(t)dt, \int_A^x f(t)dt\right) & \text{in a two-sided distribution} \end{cases} \tag{3.105}$$

The p-value was evaluated for a two-sided distribution in most of the tests performed along the present work. In the case of a significant result for a test where H_0 involves an equality between two values, sometimes both one-sided alternatives could be evaluated to determine the sign of the difference, if this sign was relevant. If this was the case, only one of the tests will be statistically significant; the sign of the difference and the p-value of this test would be reported. Otherwise, the results of the two-sided test would be given.

Statistical hypothesis tests are commonly grouped into two categories, parametric and nonparametric methods. Parametric methods rely not only on the sample but also on the properties of the distribution they derive from (parent distribution), which leads to more robust results. When this distribution is not known,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

nonparametric methods are used instead, as they use only the available data to get conclusions about the populations. The results are less robust, but they can be universally applied.

All the functions used to implement the statistical hypothesis tests are, unless otherwise stated, those provided by the library SciPy.[246]

The Q–Q plot A quantile–quantile (Q–Q) plot is a graphical method to compare two probability distributions. Quantiles are values of a variable such that its cumulative distribution function (CDF), $F(x)$, at this point has a certain value. Therefore, a theoretical quantile function is the inverse of the CDF, $Q = F^{-1}(x)$.

By comparing the quantiles from two theoretical functions, from one sample and one function, or from two samples, it is possible determine how close their distributions are. It was used along this work to compare the quantiles derived from a sample with the theoretical ones for a normal distribution. The samples were first standardized by subtracting the mean and dividing by the standard deviation each of the points, so when compared to a normal distribution of 0 mean and 1 variance, the quantiles should lie in a diagonal of slope 1 if the sample was drawn from such a distribution. The further the points are from the diagonal, the more different the distributions. While the Q–Q plot provides a useful visual reference, it must be complemented with a more robust test of the normality.

The Shapiro-Wilk test of normality Among the methods available to test whether the data come from a normal parent distribution, the Shapiro-Wilk test[291] was chosen because it showed a higher power, i.e. a higher probability of rejecting H_0 when H_1 is true, among other commonly used tests.[292]

The hypotheses formulated in the Shapiro-Wilk are

$$\begin{cases} H_0 : \text{The sample is drawn from a normal distribution.} \\ H_1 : \text{The sample is not drawn from the normal distribution.} \end{cases}$$

One of the drawbacks of this test is its high sensitivity to outliers, especially as the size of the sample studied grows.

The Brown-Fosythe test While the original proposal of this test made by Levene[293] is parametric, as it assumes that the sample derives from a normal distribution, Brown and Forsythe[294] proved that it can be extended to handle non-normal distributions. The Levene test can be used to identify a difference between the variances of the unknown distribution of two or more samples.

$$\begin{cases} H_0 : \text{The variances of all distributions are equal.} \\ H_1 : \text{Any of the variances of the distributions is different.} \end{cases}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

For K different samples, with m_k measurements in the k-th sample, x_{km} is the m-th measurement in the sample, and z_{km} is the distance between each of the measurements and the mean value of the sample, $z_{km} = |x_{km} - \langle x \rangle_k|$. Then the statistic in the original proposal by Levene is

$$W = \frac{M-K}{K-1} \frac{\sum_{k=1}^{K} m_k \left(\langle z \rangle_k - \langle z \rangle\right)^2}{\sum_{k=1}^{K} m_k \left(z_{km} - \langle z \rangle_k\right)^2}, \tag{3.106}$$

where $\langle z \rangle_k$ is the average of z_{km} in the k-th sample, while $\langle z \rangle$ is the average over all samples. $M = \sum_{k=1}^{K} m_k$ is the total number of measurements. The modification proposed by Brown and Forsythe, which is the one used in this work, consists in the substitution of the average by the median in all terms of Eq. 3.106.

The test statistic W follows, regardless of the version of the test used, the Fisher-Snedecor distribution

$$f(x; \nu_1, \nu_2) = \frac{1}{\mathrm{B}\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} x^{\nu_1/2-1} \left(1 + \frac{\nu_1}{\nu_2} x\right)^{-(\nu_1+\nu_2)/2}, \tag{3.107}$$

where $\mathrm{B}(x_1, x_2) = \int_0^1 t^{x_1-1}(1-t)^{x_2-1} dt$ is the Euler beta function, and ν_1 and ν_2 the degrees of freedom; in this case, $\nu_1 = K - 1$ and $\nu_2 = M - K$. The rejection of H_0 would imply that the variance of at least one of the samples being compared is different from the others.

3.4.2.1 Parametric methods

The following parametric methods were used at some point in the present work. All of them are derived for normally distributed data, so a previous evaluation of the normality of the samples was made.

The Student t-test Designed to compare the mean of a normally distributed sample with a reference value, the Student t-test[295] can be also applied to compare the means of two samples. Although the exact formulation of the t-test depends on what exactly one wants to compare, its use in this work was limited to evaluate the results of a linear regression (Eq. 3.99): Having estimated the values β_k of the K coefficients of a regression model, each with an associated standard error s_{β_k}, a t-test can be formulated to evaluate whether each term contributes to a variation in the dependent variable

$$\begin{cases} H_0 : \text{The term multiplied by } \beta_k \text{ does not correlate with the dependent variable.} \\ H_1 : \text{The term multiplied by } \beta_k \text{ correlates with the dependent variable.} \end{cases}$$

In other words, the t-test can be used to find those β_k that are sufficiently different from 0.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The value of the statistic t associated with a coefficient is

$$t_{\beta_k} = \frac{\beta_k - \beta_k^0}{s_{\beta_k}}, \tag{3.108}$$

where β_k^0 is a reference value, dependent on H_0; in this particular case, $\beta_k^0 = 0$. The probability density function associated with the test statistic t is Student's t distribution

$$f(x;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2}, \tag{3.109}$$

being $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ the gamma function, and ν the number of degrees of freedom, $\nu = M - K$.

Therefore, finding enough evidence to reject H_0 would imply that the value of β_k is sufficiently different from 0 to consider its associated variable has an impact on the dependent variable of the model.

The Fisher F-test The F-tests are a family of tests to compare the variance of samples. Specifically, for the result of a regression model, it allows us to find if any of the independent variables are significant. In general, having two models with K_1 and K_2 parameters, such $K_1 < K_2$, but also model 1 is contained within model 2, a F-test will evaluate[296]

$$\begin{cases} H_0 : \text{All coefficients of model 2 not in model 1 are zero} \\ H_1 : \text{At least one coefficient in model 2 and not in model 1 is non-zero.} \end{cases}$$

In the most simple case, model 1 would be a constant equal to $\langle y \rangle$, so the result of comparing model 1 against any other linear model would determine if any coefficient is sufficiently different from zero. If model i is defined by f_i, the sum of squares of the residuals of the model, S_i, is computed according to Eq. 3.101. The F statistic is

$$F = \frac{S_1 - S_2}{S_2}\frac{M - K_2}{K_2 - K_1}. \tag{3.110}$$

where M is the size of the sample. The statistic follows the Fisher-Snedecor distribution (Eq. 3.107) with $\nu_1 = K_2 - K_1$ and $\nu_2 = M - K_2$. Rejecting H_0 means that f_2 (the linear model) fits the data significantly better than f_1 (a constant term), so any of the independent variables impact the dependent variable.

3.4.2.2 Nonparametric methods

Whenever there is not enough evidence that the data are compatible with a normal distribution, nonparametric methods must be used. The methods used in the subsequent analysis were:

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The Mann-Whitney U test Having two samples for two populations, the Mann-Whitney U test[297] can be applied to find if they both can be described by the same parent distribution.

If X_1 and X_2 are the two populations under study, the hypothesis evaluated by the Mann-Whitney U test are:

$$\begin{cases} H_0 : \text{The distribution of } X_1 \text{ and } X_2 \text{ are equal.} \\ H_1 : \text{The distribution of } X_1 \text{ and } X_2 \text{ are not equal.} \end{cases}$$

A clarification of what "equal" means in this context must be made: What the test evaluates is whether the probability of sampling values of X_1 that are larger than those of those samples of X_2 is the same as the probability of them being lower. The previous hypotheses can be therefore written in an equivalent formulation:

$$\begin{cases} H_0 : \ P(X_1 < X_2) = P(X_1 > X_2) \\ H_1 : \ P(X_1 < X_2) \neq P(X_1 > X_2) \end{cases}$$

With enough statistical evidence for the rejection of the null hypothesis, it can be assumed that one of the distributions tends to be higher or lower than the other. For the determination of the sign of the difference, it is possible to make use of alternative hypotheses:

$$\begin{cases} H_0 : \ P(X_1 < X_2) \geq P(X_1 > X_2) \quad | \quad P(X_1 < X_2) \leq P(X_1 > X_2) \\ H_1 : \ P(X_1 < X_2) < P(X_1 > X_2) \quad | \quad P(X_1 < X_2) > P(X_1 > X_2) \end{cases}$$

Finding that $P(X_1 < X_2)$ is higher than $P(X_1 > X_2)$ would mean that the cumulative distribution function (CDF) associated with X_1, $F(X_1)$ is stochastically greater than that of X_2, $F(X_2)$. If both samples have the same variance, that would mean that the distribution of X_1 shifts towards lower values.

The Mann-Whintey U test, and many other tests, use the ranks of the samples to evaluate the hypotheses. Ranking the points in a sample of M elements is to assign them the position (or rank) they have in their ordered sequence, 1 being the rank of the lowest element and M that of the largest. Therefore, the rank of the k-th element in the set x_k is

$$R(x_k) = \sum_{m=1}^{M} H(x_k - x_m)\,, \tag{3.111}$$

where $H(x)$ is the Heaviside step function, equal to 1 if the argument is non-negative and 0 elsewhere.

The test statistic U is computed as

$$U = \min(U_1, U_2), \qquad U_j = S_j - \frac{M_j(M_j+1)}{2} \tag{3.112}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The notation S_j represents the sum of the ranks of the elements of the j-th sample, whose size is M_j, after ranking both samples combined.

Computing the p-value as in the original proposal is too expensive for large samples. However, as the size of the samples increases, it is possible to take advantage of the fact that the distribution of U tends to a normal distribution.

The Wilcoxon signed-rank test When the two samples have not only equal sizes but there is a relation between each element in X_1 and only one other element in X_2, the Wilcoxon signed-rank test[298] is a better approach than Mann-Whitney U to evaluate a difference between the distributions. In that case, elements of X_1 and X_2 are said to be paired.

The hypotheses of the Wilcoxon signed-rank test are based in the distribution of the differences between all pairs of measurements for both samples

$$\begin{cases} H_0 : P(D > 0) = P(D < 0) \\ H_1 : P(D > 0) \neq P(D < 0) \end{cases}$$

If D is non-symmetric around zero, the CDF of one population must be stochastically greater than the other. The sign of this skewness can be determined by the evaluation the one-sided hypotheses:

$$\begin{cases} H_0 : \ P(D < 0) \geq P(D > 0) & | & P(D < 0) \leq P(D > 0) \\ H_1 : \ P(D < 0) < P(D > 0) & | & P(D < 0) > P(D > 0) \end{cases}$$

The statistic associated with the Wilcoxon signed-rank test is

$$T = \sum_{m=1}^{M} \operatorname{sign}(x_m - y_m) R(|x_m - y_m|) \,, \tag{3.113}$$

As with the Mann-Whitney U test, the calculation of the p-value can be made by approximating the distribution of T to a normal distribution when the size of the sample is enough.

Kernel density estimate When the parent distribution behind a sample is not known, the kernel density estimation (KDE) can be used to approximate the PDF behind the sample by doing a convolution of the measurements with another function, called the kernel. If a sample consists of the measurements $X = (x_1, \ldots, x_N)$, the density function of the sample can be built as

$$\phi_X(x) = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n) \,, \tag{3.114}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

being $\delta(x)$ the Dirac delta function. If $K_h(x)$ is a kernel that depends on an auxiliary smoothing parameter h, its convolution with $\phi_X(x)$ is

$$\begin{aligned}(\phi_X * K_h)(x) &:= \int_{\mathbb{R}} \phi_X(\tau) K_h(x-\tau) d\tau \\ &= \int_{\mathbb{R}} \frac{1}{N} \sum_{n=1}^{N} \delta(\tau - x_i) K_h(x-\tau) d\tau \\ &= \frac{1}{N} \sum_{n=1}^{N} K_h(x - x_i)\end{aligned} \tag{3.115}$$

This result implies that the KDE of a sample of discrete measurements is the sum of kernel functions centered on each of the measurements. The KDE of the sample is this convolution,

$$K(x) = \frac{1}{N} \sum_{n=1}^{N} K_h(x - x_i)\,. \tag{3.116}$$

Many functions can be used as the kernel; for this work, a standard Gaussian function was selected to be used as such. The selection of the value of h—bandwidth—, has a significant impact on the resulting KDE: larger values may oversimplify the shape of the distribution, while the distribution will tend to a series of spikes located in the measurements if it is too low. The optimal value of h is that which minimizes the difference between the KDE and the real—yet unknown—PDF. Alternatively, different methods can be used to estimate their value. We opted for Scott's rule,[299] where the bandwidth is

$$h = \sigma N^{-1/5}\,, \tag{3.117}$$

being σ the standard deviation of the sample.

The KDE of two or more samples was used to estimate their overlap. Assuming the KDEs to be faithful estimators of the parent distributions, so $f(x) \approx K(x)$, the overlap between the PDFs of two samples can be approximated by

$$S(f_1, f_2) \approx \int_{-\infty}^{\infty} \min(K_1(x), K_2(x)) dx\,. \tag{3.118}$$

The same formula can be generalized for the overlap of more than two samples. The overlap is an estimation of how similar the samples are: Since the area under any PDF is equal to 1, $S = 1$ if $f_1(x) = f_2(x)\ \forall x \in \mathbb{R}$. On the other hand, $S = 0$ when either PDF is equal to zero in the values of x such that the other PDF is non-zero. Although the value of S estimates the similarity of the samples, further tests must be applied to find statistically significant differences between them.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The Kendall τ test Given two magnitudes, X and Y, there exists a monotonic correlation between them when the growth in one is associated with either a growth or a decrease in the other. The Kendall rank correlation coefficient[300], usually denoted as τ, is a measurement of this type of correlation.

Let (x_i, y_i) and (x_j, y_j), $i < j$, be a pair of simultaneous observations of magnitudes X and Y. This pair of observations are said to be concordant if

$$\operatorname{sign}(x_j - x_i) = \operatorname{sign}(y_j - y_i)\,; \tag{3.119}$$

i.e. the change in both magnitudes has the same sign. Two pairs are discordant when, conversely, their sign is opposite.

For a set of M joint observations of X and Y, a total of $\binom{M}{2} = \frac{M(M-1)}{2}$ non-redundant pairings can be made. Kendall's τ is the difference between the number of concordant and discordant pairs relative to the number of pairs, so it can be computed as:

$$\tau = \frac{2}{M(M-1)} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \operatorname{sign}(x_j - x_i) \cdot \operatorname{sign}(y_j - y_i)\,. \tag{3.120}$$

If all pairs are concordant, then $\tau = 1$, which means that the ranking of the observations of X is equal to that of the observations of Y; when the ranking is inverse, then $\tau = -1$, which is obtained when all pairs are discordant. The value $\tau = 0$ indicates that the same number of concordant and discordant pairs have been found, so no correlation between X and Y can be expected.

To evaluate this last statement, i.e. finding a sufficient evidence about the correlation of these two magnitudes, the Kendall τ test can be used.

$$\begin{cases} H_0 : \text{The Kendall rank correlation coefficient } \tau \text{ between } X \text{ and } Y \text{ is zero.} \\ H_1 : \text{The Kendall rank correlation coefficient } \tau \text{ between } X \text{ and } Y \text{ is non-zero.} \end{cases}$$

For small samples, the distribution of τ under the null hypothesis can be calculated exactly, but for large samples it tends to a normal with 0-mean and variance $2(2M+5)/(9M(M-1))$.[300]

The rejection of the null hypothesis cannot be interpreted as a sign of a strong correlation, but as strong evidence *of the existence* of a correlation: For a sample of 50 measurements, the critical value of τ for a significance level of $\alpha = 0.05$ is $\tau_c \approx 0.194$[*]; while lowering to $\alpha = 0.01$, the critical value is $\tau_c \approx 0.322$[†]. As the samples become larger, the value of τ_c for a given α will be reduced. Thus, an additional criterion for strong correlation is $|\tau| > \tau_r$, τ_r being a reference value that we set in $\tau_r = 0.6$[‡].

[*] Equivalent to ratio between concordant and discordant pairs of ~ 1.48.

[†] Approximately 1.95 concordant pairs per discordant pair.

[‡] The value that arises from a 4:1 ratio of concordant and discordant pairs; or vice versa for $\tau_r = -0.6$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3.4.3 Classification models

The Yeo-Johnson transformation The applicability of different algorithms and techniques can be limited when the data available are non-normal, from a worse performance to unreliable results. To handle these scenarios, different transformations can be applied, such that the data become more normal-like. After testing their use for the training of the classification model detailed below, applying the Yeo-Johnson transform on the data led to a better performance.

Yeo and Johnson[301] proposed a monotonic transformation of the data that can be applied to samples of magnitudes in $\mathbb{R}$, so the result is closer to a normal distribution. This transformation depends on a parameter λ (power transform) and has the functional form

$$f(x) = \begin{cases} \dfrac{(x+1)^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0,\ y \geq 0 \\ \ln(x+1) & \text{if } \lambda = 0,\ y \geq 0 \\ \dfrac{1-(1-x)^{2-\lambda}}{2-\lambda} & \text{if } \lambda \neq 2,\ y < 0 \\ -\ln(1-x) & \text{if } \lambda = 2,\ y < 0 \end{cases} \tag{3.121}$$

which is continuous in both λ and x. Although λ can be arbitrarily selected, the implementation of the Yeo-Johnson transform available in Scikit-learn[302] optimizes its value, so the outcome is closest to a normal distribution.

Logistic regression The least-squares method can be in principle used to fit data points to any model; for example, to a logistic function

$$f(x) = \frac{1}{1 + e^{-\beta_0 - \beta x}}\,. \tag{3.122}$$

The parameters β_0 and β control the shape of the curve (Figure 3.30). The model can be extended to a multidimensional case of $K-1$ independent variables.

$$f(\mathbf{x}) = \frac{1}{1 + e^{-\beta_0 - \boldsymbol{\beta}\cdot\mathbf{x}}}\,, \tag{3.123}$$

where $\mathbf{x}$ is the vector representation of the independent variables and $\boldsymbol{\beta}$ the vector with their associated coefficients. Eq. 3.123 is completely equivalent to Eq. 3.122 after a linear transformation $x : \mathbb{R}^{K-1} \mapsto \mathbb{R}$:

$$x(\mathbf{x}) = \frac{1}{\beta}\boldsymbol{\beta}\cdot\mathbf{x} \tag{3.124}$$

The logistic regression is used as a simple binary classification model: given two classes, A and B, the value $f(\mathbf{x})$ would be the probability that the point $\mathbf{x}$ will be

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

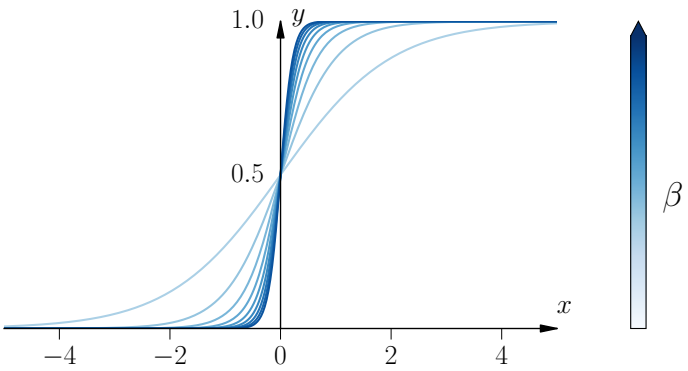


Figure 3.30: Logistic curve. In this 1-dimensional example, x and is an independent variables. As the value of β change, so will the shape of the logistic curve. The parameter β_0 shifts the value at which the curve intersects the y axis.

associated with class A, while $1 - f(\mathbf{x})$ is the probability of belonging to B. The parameters $\boldsymbol{\beta}$ are fitted so the given probability of the input belonging to one of the classes is higher for the class they actually belong to. We used the implementation of the Scikit-learn[302] class LogisticRegression. The parameters are found through the optimization of the cost function

$$S = -\frac{1}{M}\sum_{m=1}^{M}\left(y_m \ln(f(x_m)) + (1-y_m)\ln(1-f(x_m))\right) + \frac{1}{2M}\|\boldsymbol{\beta}'\|^2 . \qquad (3.125)$$

for a set of M data points. The optimal parameters $\boldsymbol{\beta}'$ (here the prime indicates that the intercept is also included) are obtained through the L-BFGS-B algorithm.[303]

Evaluation of the performance of the classification model The process of fitting the parameters of the classification models is known as model training. To evaluate the model, a classification of data that was never used in the training stage is performed, so it is possible to estimate its predictive power.

For a binary classification, there are 4 possible outcomes: if the model is looking for the identification of elements that belong to the class A (positives), an element of A identified as part of this class is a true positive (TP), while an element of B correctly assigned to its class is a true negative (TN). Both are the result of a satisfactory classification; when this does not happen, false positives (FP) appear when an element of B is classified as part of A, or false negatives (FN), in the case of an element of A assigned to B. The ideal model is that with no FP or FN, and to determine how the model deviates from this case, different metrics were developed.[304]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The **accuracy** is the ratio of correct classifications among all

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \,. \tag{3.126}$$

Accuracy weights both categories equally, as only correct predictions are considered: The higher the accuracy, the more elements were correctly classified. However, this metric may be misleading in imbalanced data sets, where one category is more abundant than the other, disguising a poor predictive power thanks to an uneven number of elements in the classes.

The **precision** determines the fraction of positive predictions that are correct, so

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \,. \tag{3.127}$$

Unlike accuracy, precision is focused on positive results: High precision means that most of the elements classified as positive are, in fact, positive. Evaluating a model based on its precision implies penalizing the number of false positives—elements of B classified as belonging to A—, but does not consider the number of false negatives—misclassified elements of A.

The **recall**, on the other hand, determines the fraction of positives correctly predicted:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \,. \tag{3.128}$$

The difference with respect to precision is that the recall penalizes the number of false negatives: a high recall implies that most positives are correctly identified. However, there may be false positives that are not taken into account.

Using the precision or the recall to evaluate a model will depend on which is less relevant for the purpose of the model. A common way in which both of these metrics are considered altogether is the calculation of the F_1 score, the harmonic mean of precision and recall:

$$F_1 = \frac{2}{1/\text{Precision} + 1/\text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{3.129}$$

The value of the F_1 score will always be between that of precision and the recall of the model.

All of the previous metrics depend on the threshold used to determine when an element belongs to a class. Although in many cases this is set to 0.5, assigning the element to the most probable class, it can be shifted to try to reduce the number of false positives or negatives, depending on what is best for the applicability of the model. This will, of course, change the values of all previous metrics. By exploring the whole range of thresholds, a precision–recall curve can be built, which can then be used to evaluate the model beyond a single threshold or to select an appropriate

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

value to increase the performance of the model in a desired aspect. The ideal model is that for which both precision and recall are 1; therefore, the **area under the precision–recall curve (AUC-PR)** would be also 1. For real models, this AUC-PR will be lower, but the closer its value is to 1, the better the precision under changes in the recall.

For practical purposes, we used the average precision score given by the namesake function of Scikit-learn[302] as an estimator of the AUC-PR. Its calculation is done by weighted mean of the precision for different thresholds, using the change in recall from the previous threshold as weight:

$$\text{AUC-PR} = \sum_n (\text{Recall}_n - \text{Recall}_{n-1}) \cdot \text{Precision}_n \tag{3.130}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 4.

Results and discussion

AHHH! That doesn't make any sense! I don't understand!
I am very excited for you, though! P-please stop yelling!
Riebeck, Outer Wilds

The study of peptide–membrane interactions requires the combination of different levels of molecular modeling and analysis to capture both structural detail and emergent behavior. The following chapter summarizes the outcomes of these computational studies, focusing on the results obtained from MD simulations performed at atomistic and CG resolutions. The discussion is organized into three main sections, each corresponding to a specific objective described in Chapter 2.

The first section examines atomistic simulations aimed at characterizing the structural and dynamical behavior of peptides at the water–lipid interface. This includes a description of adsorption events, conformational rearrangements, and the influence of simulation parameters, including the use of the hydrogen isotope exchange (HIE) strategy.

The second section details the systematic simulation and analysis of peptide–membrane systems, and introduces the SuPepMem database, a structured and extensible repository that integrates simulation data for comparative studies of membrane-targeting AMPs.

The final section explores the relationship between the numerical descriptors derived from these simulations and the reported antimicrobial activity of the corresponding sequences. Statistical and data-driven analyses were applied to assess whether these descriptors could be used as predictive indicators of antimicrobial behavior.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.1 Structural characterization of membrane-targeting antimicrobial peptides at the lipid membrane–water interface

Predicting the conformation of peptides whose structure is strongly dependent on the environment remains a major challenge in structural biology. While recent deep-learning methods such as AlphaFold2[168] have achieved remarkable accuracy for soluble proteins, their predictions correspond to static structures and do not reflect the influence of complex, heterogeneous environments such as lipid membranes. For membrane-targeting AMPs, the adopted conformation and orientation are not intrinsic properties of the sequence but a response to the conditions at the membrane interface. Understanding these processes at atomic resolution is essential not only to elucidate their mechanism of action but also to evaluate whether MD simulations can generate the kind of information required to extend or complement current structure-prediction frameworks.

A full description of peptide folding from an unstructured state lies beyond the timescales accessible through conventional atomistic MD simulations, which typically cover up to the microsecond timescale. Instead, this study focuses on evaluating the stability and structural evolution of initially helical peptides upon interaction with a lipid bilayer, analyzing how the secondary structure responds to the interfacial environment. This approach aims to provide insight into the balance between intrinsic sequence preferences and membrane-induced effects that modulate the peptide conformation.

Three representative peptides were selected, based primarily on their experimentally reported structural behavior in the membrane environment, particularly their tendency—or lack thereof—to adopt a α-helical conformation upon membrane interaction. Parasin I, a cationic peptide derived from catfish skin mucus,[257] has been reported to display membrane-targeting activity. Upon binding to sodium dodecylsulfate micelles, its helicity has been estimated to be $\sim$ 47 % by both CD and NMR spectroscopy.[52] Magainin-2, one of the most widely characterized antimicrobial peptides from amphibian skin, forms stable amphipathic α-helices upon membrane binding, and serves as a model for helix-mediated membrane permeabilization. In a 3:1 POPE:POPG model membrane, the NMR spectrum of magainin indicated a helicity above 90 %.[258] Octa-2, a short synthetic peptide composed of alternating arginine and tryptophan residues, displays strong membrane affinity, but its structure has not been characterized yet.[259,260] Another peptide, chensinin-1b, shows potent bactericidal activity and exhibits a partially helical conformation in membrane-mimetic environments—close to 30 %, determined by CD and NMR in a 50 % TFE solution[264,305]—and was therefore included in extended simulations as a

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

control.

The simulated systems consisted of a single unit of each of the previous peptides and a symmetric bilayer containing 152 lipids per leaflet with a 3:1 POPE:POPG composition, representative of the inner membrane of *E. coli*.[33] The bilayer was fully hydrated—70 water molecules per lipid—and the charge of the systems was neutralized with sodium counterions.

Two distinct initial configurations were prepared in order to explore the different contributions of adsorption and insertion to peptide–membrane interaction. The first, referred to as the **shallow configuration**, places the peptide in the aqueous phase just above the lipid headgroups, allowing the study of the early stages of membrane recognition and adsorption. The second, the **embedded configuration**, starts with the peptide inserted into the bilayer, providing a means to assess how the structure and orientation of the peptide evolve once direct contact with the hydrophobic region is established. Comparison of these two starting points enables identification of whether the final conformation depends mainly on the intrinsic stability of the peptide structure or on the conditions during insertion.

Each peptide was modeled as a right-handed α-helix with charged termini, and the protonation state of their side chains coincides with the most probable state under physiological conditions. For both initial configurations, five evenly spaced peptide orientations were simulated, spanning the full 2π rad rotation around the peptide axis; these orientations define the five independent replicas used in the analysis. Each setup was reproduced under four atomistic force fields—AMBER, CHARMM, GROMOS, and GROMOS-HIE—leading to a total of 60 systems per configuration.

The systems were equilibrated through the protocols described in Section 3.2.2, gradually releasing position and dihedral constraints. The HIE scheme enabled the use of an extended time step of 6 fs while maintaining numerical stability. Production trajectories were extended for 900 ns in all cases, with frames stored every 6 ps.

The robustness and computational efficiency of the HIE approach was tested by comparing with standard GROMOS simulations in terms of stability and consistency of structural and dynamical magnitudes. A complementary set of simulations employing the reaction field method for the treatment of electrostatics was also conducted to determine whether simplified long-range interaction treatments could reproduce the results obtained with the more computationally demanding PME method.

Extended trajectories of up to 15 µs were run for both shallow and embedded configurations to probe long-timescale effects and evaluate whether conventional MD is able to capture reversible folding–unfolding transitions, or transitions between adsorption and insertion states within feasible computational limits.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The following section presents the results obtained from these simulations: first, the adsorption of the peptide and its interfacial behavior; second, the structural rearrangements associated with the insertion into the bilayer; and finally, the dynamical and long-timescale conformational effects derived from extended trajectories.

4.1.1 Peptide dynamics at the membrane–water interface

The *shallow* configuration represents conditions in which the peptides interact only weakly with the lipid surface, remaining largely solvated in water while partially in contact with the polar headgroups. In this regime, the peptide experiences an interfacial environment that closely resembles that of a fully solvated peptide in bulk solution. As all sequences were initially modeled as right-handed α-helices, this setup allows assessing the intrinsic stability of their secondary structure under an initial minor membrane perturbation.

In all simulations, the peptides remained close to the water–lipid interface (Figure 4.1), influenced by the electrostatic attraction between the cationic residues and the anionic POPG headgroups. Nonetheless, the degree of adsorption differs across force fields. In both GROMOS and GROMOS-HIE, the peptides remained at about 2 nm from the center of the bilayer, maintaining a stable position. In CHARMM, octa-2 penetrated slightly more into the bilayer in four of the five replicas, while magainin showed significant variability among replicas. AMBER simulations show the weaker binding: the peptides tended to move away from the surface and, in one case, detached completely before readsorbing on the opposite leaflet. In all cases, the peptides remained adsorbed at the water–lipid interface, mainly interacting with the lipid headgroups. The observed binding is consistent with an electrostatic attraction between cationic peptides and anionic lipids. The preference of the peptides for this region was expected, as the selected sequences correspond to membrane-targeting AMPs. However, the peptides did not penetrate deeply into the membrane: all remained more or less close to the surface, extending along the interface rather than penetrating and acquiring a transmembrane conformation (Figure 4.2).

Challenges in the quantification of helicity A major issue when analyzing these systems lies in quantifying the degree of α-helical structure in the short and flexible peptides. Most secondary structure metrics were originally developed for folded proteins, where α-helices are generally long, stable, and embedded in a tertiary scaffold. When applied to short amphipathic peptides, these discrete classifications can misrepresent structural fluctuations.

The most widely used criterion, the DSSP algorithm, identifies α-helical residues based on backbone H-bond patterns ($i \rightarrow i + 4$). This approach assigns a discrete label to each residue, producing a stepwise helicity profile that is particularly prob-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

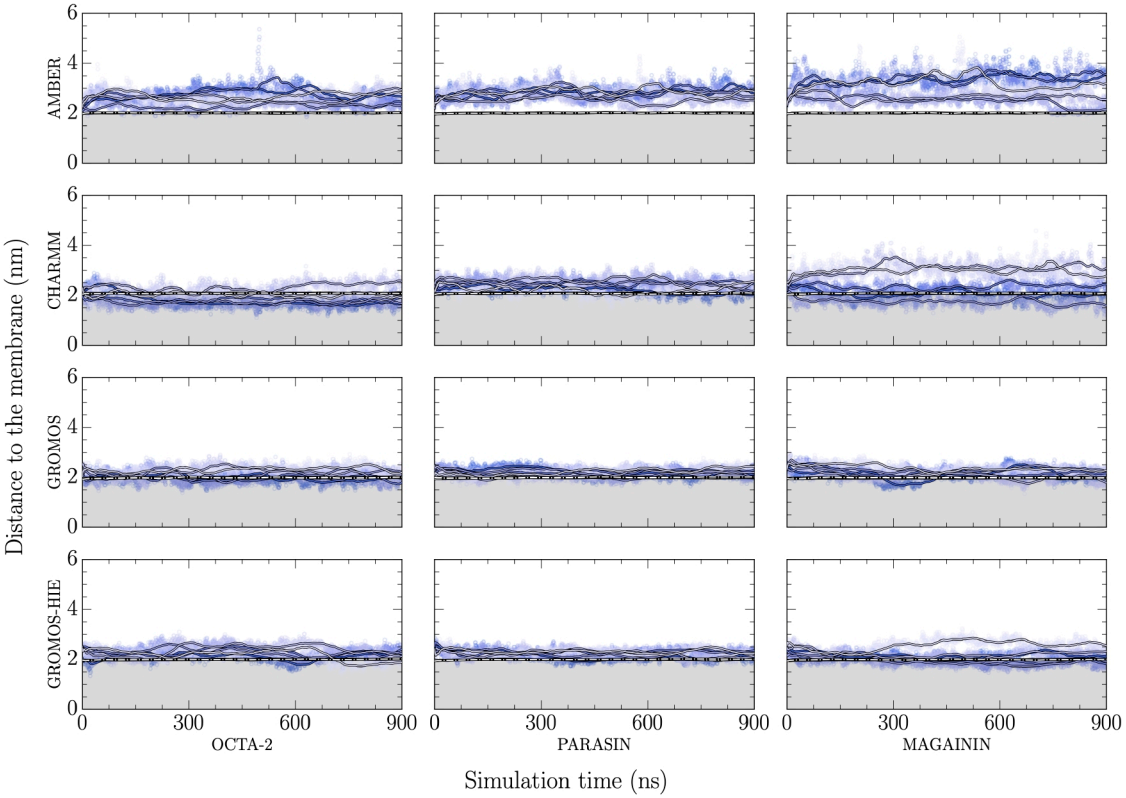


Figure 4.1: Time evolution of the peptide–membrane distance for the shallow starting configuration. Each panel shows the evolution of the distance along the Z coordinate—perpendicular to the bilayer plane—between the centroid of the peptide and the center of the membrane, as a function of simulation time. The plots are arranged by force field (rows) and peptide (columns). In each panel, the thin blue lines represent the rolling mean (window of 100 steps) of the instantaneous peptide–membrane distance, sampled every 6 ns, for five independent simulation replicas. The gray shaded area indicates the hydrophobic core of the membrane, while the dashed white line at approximately 2 nm marks the average position of the phosphorus atoms of the lipid headgroups, corresponding to the membrane surface. Smaller distances correspond to deeper peptide insertion within the bilayer, while larger values indicate adsorption near or above the interface. All plots share identical axes and scale, allowing direct comparison between peptides and force fields. Unless otherwise stated, the same conventions apply to subsequent figures showing time-dependent properties. This graph and all subsequent graphs were created using Matplotlib[8] and Seaborn.[306]

lematic for short sequences. For example, in an eight-residue peptide such as octa-2, only a handful of values are possible for the fractional helicity,* and a minor geometric fluctuation—such as breaking a single H-bond—can abruptly change the assigned

*These are $0, 0.5, 0.625$, and 0.75.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

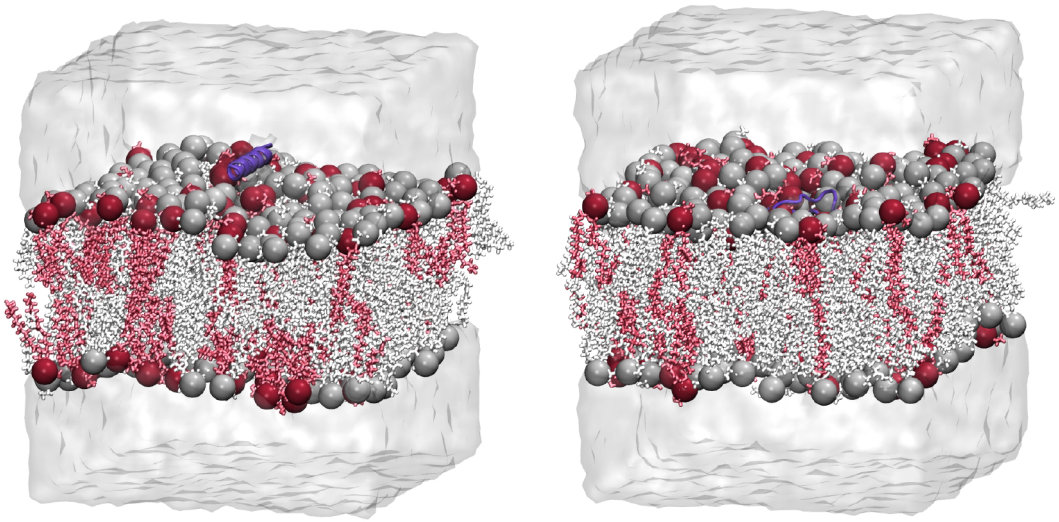


Figure 4.2: Initial (left) and final (right) states of a system starting from the shallow configuration. POPE lipids are represented in gray, and POPG lipids in red. Phosphorus atoms are highlighted as spheres. The backbone of the peptide is represented as a purple ribbon. The set of water molecules is represented as a gray volume. These and all subsequent system snapshots were rendered using VMD.[307]

structure of several residues. Consequently, DSSP tends to exaggerate conformational variability in small peptides (Figure 4.3).

The general trend observed is that the different replicas—which differ in the initial orientation of the peptide around its axes prior to equilibration—led to different structures. For the shortest peptide, no consistency is observed between the simulations performed under the same force field, except for the GROMOS simulations. In AMBER, the peptides unfolded in the first 100 ns of the simulation, with two replicas recovering some helical content shortly afterwards. One of them unfolded once again after 700 ns of simulation time, the same instant at which the peptide partially folded in other replica. Rather than being due to a large change in the 3-dimensional structure, the gain in helicity derived from the formation of the appropriate H-bonds within the backbone.

Two of the octa-2 CHARMM replicas preserved the full helix throughout the simulation, and in other two the peptide alternated between complete and partial unfolding, while the last of the replicas explored a wider range of intermediate states. Standard GROMOS simulations are by far the most consistent and stable: in four out of the five replicas, octa-2 unfolded from the very beginning and never acquired a more defined structure, while the remaining one did preserve the helix. As in AMBER, GROMOS-HIE simulations suffer from large variations in the structure; at the end of the 900 ns of simulation, only one replica exhibits a mostly folded

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

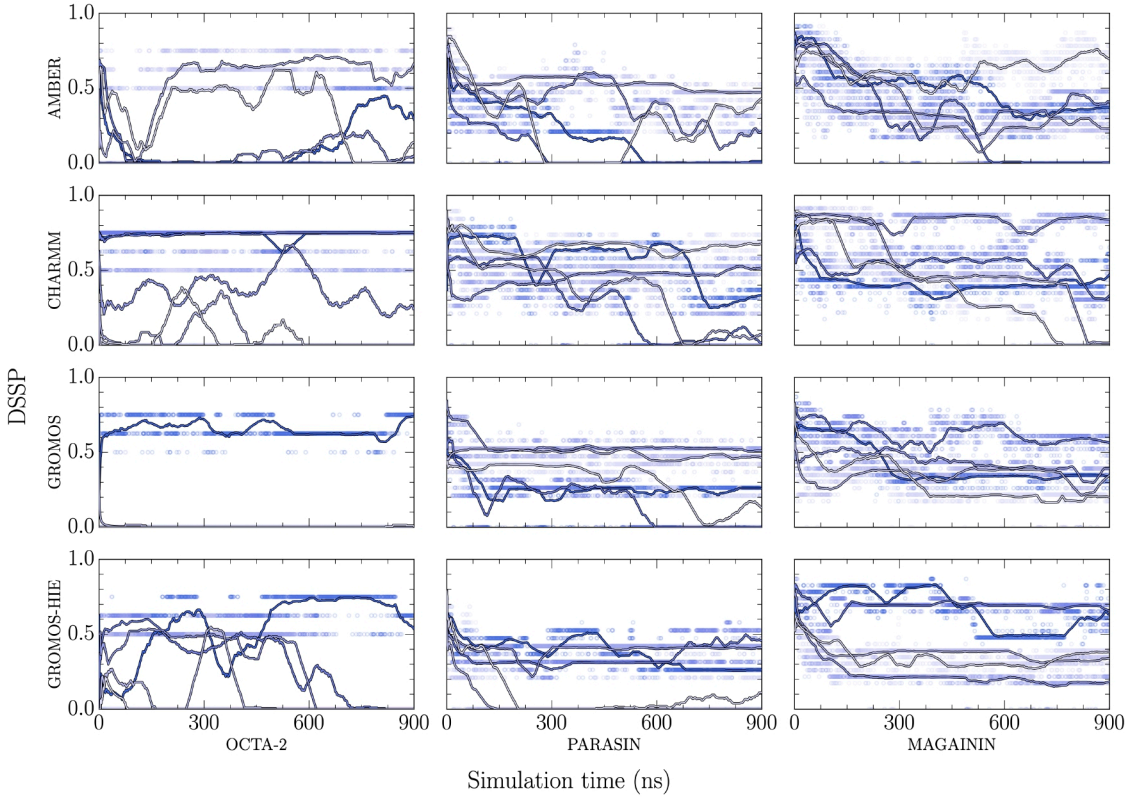


Figure 4.3: Time evolution of peptide helicity calculated with the DSSP method for the shallow configuration. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using DSSP. The plots are arranged by force field and peptide. The thin blue lines represent the helicity evolution of the five independent simulation replicas, calculated as a rolling average of the instantaneous DSSP values (displayed as faint points), to emphasize the overall trends.

state. Interestingly, in this case the peptide unfolded completely at the beginning of the simulation and progressively recovered the helix.

In longer peptides, more helicity values are accessible, but large differences across replicas can still be observed in all force fields. This makes it very difficult to draw conclusions about the effect of the force field on the structure. In the case of parasin, three of the AMBER simulations stabilized into defined values of the helicity (two of them at zero and the other around 0.5), two in CHARMM (around 0.5 and 0.7), three in GROMOS (0, 0.25 and two around 0.5), and four in GROMOS-HIE (two at 0, others at 0.25 and 0.4). The similarities between the results of GROMOS and GROMOS-HIE for this peptide must be noted.

The structure of magainin seems even more variable that that of parasin. In AMBER and CHARMM not only large fluctuations in the helicity were obtained, but also large differences between the different replicas; in GROMOS and GROMOS-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

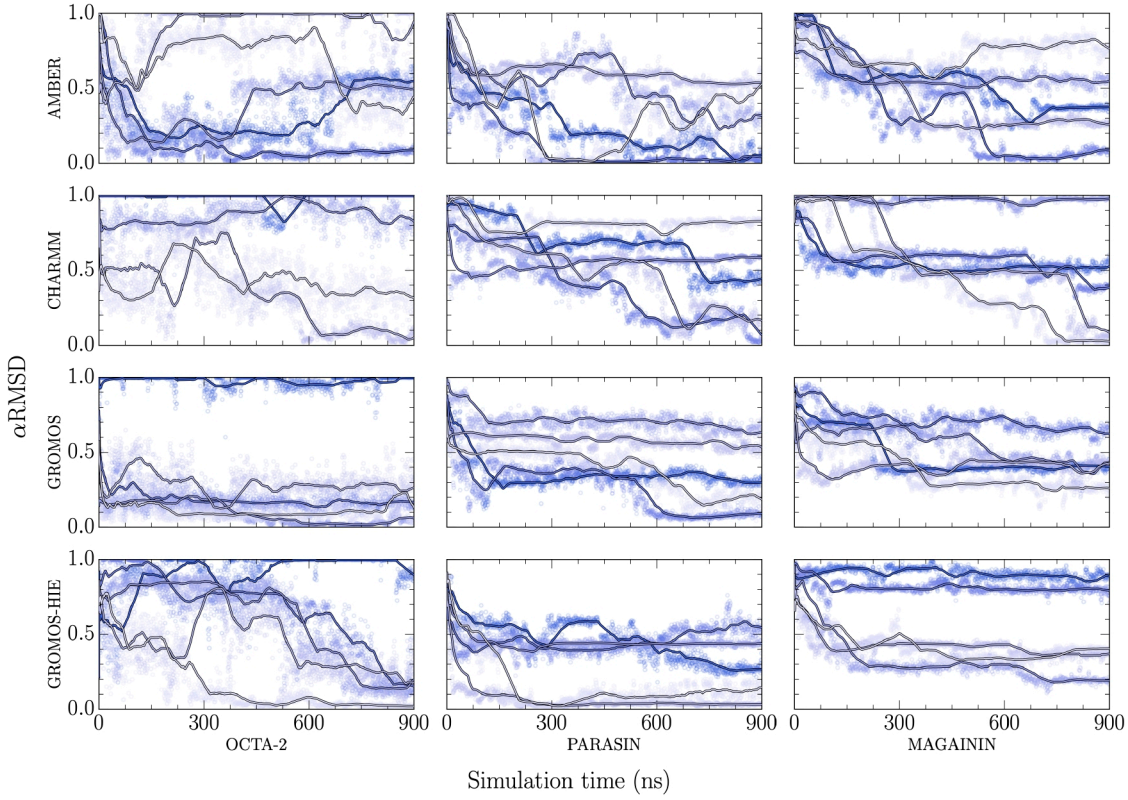


Figure 4.4: Time evolution of peptide helicity calculated with the αRMSD metric for the shallow configuration. Each panel shows the variation of αRMSD values (Y-axis) over simulation time (X-axis), providing an alternative measure of α-helical content. The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the five independent replicas, smoothed as rolling averages of the instantaneous αRMSD values (shown as faint points).

HIE fluctuations were much smaller and the replicas stabilized around different values of the helicity, but the ranges spanned are still very wide.

Given the limitations of discrete metrics such as DSSP, a complementary approach was employed to obtain a smoother, geometry-based measure of helicity: the αRMSD.[276] This quantity evaluates the root mean square deviation of the peptide backbone from an ideal α-helical reference structure, providing a continuous estimate of how closely the conformation resembles a canonical helix. Unlike DSSP, it does not rely on H-bond assignments, and thus it is less sensitive to transient bond rearrangements or small distortions at the peptide termini.

The evolution of αRMSD throughout the trajectories (Figure 4.4) reveals that peptides retained some degree of helical organization even after apparent unfolding. Profiles obtained with this metric are considerably smoother, and replicas that appeared completely disordered according to DSSP maintain intermediate αRMSD

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

values, reflecting local structural organization despite its flexibility. In this sense, αRMSD captures gradual conformational transitions that DSSP tends to identify as abrupt losses of secondary structure.

Nevertheless, αRMSD has interpretative limitations of its own. While values close to unity correspond unambiguously to a well-defined α-helix, deviations from this reference do not map uniquely to specific geometries—different bent, kinked, or partially folded conformations may yield similar αRMSD values. Hence, although the metric provides a more realistic and continuous view of helicity evolution, it cannot alone distinguish among the different partially ordered states that short membrane-active peptides can adopt at the interface.

An alternative between these two methods can be the collective variable proposed by Fioring, Klein, and Hénin,[281] here referred to as α-block. This descriptor combines the detection of backbone H-bonds with the analysis of angular deviations among consecutive C_α atoms, yielding a continuous measurement of helical order. Unlike DSSP, which assigns binary states to each residue, α-block quantifies the extent to which the local backbone geometry resembles an α-helix, thus reducing artificial discontinuities while maintaining sensitivity to structural rearrangements.

In the present simulations, α-block exhibits smoother temporal behavior and narrower fluctuations than either DSSP or αRMSD (Figure 4.5). The estimated helicity typically ranges between 0.1 and 0.9, suggesting that none of the peptides remain fully folded or completely unstructured during the trajectories. Although its absolute value lacks a direct structural interpretation, the α-block variable provides a numerically stable way to monitor partial helical content in short and flexible peptides.

The smoother and more consistent trends obtained with the continuous metrics suggest a possible physical interpretation of the shallow configuration. Under these conditions, peptides remained near the boundary between the aqueous phase and the lipid headgroups, where they are only weakly influenced by the membrane environment. In this interfacial regime—neither fully solvated nor truly embedded—, the peptides may experience competing interactions that prevent their stabilization into a single conformation. The resulting energy landscape could therefore be shallow and highly complex, allowing the coexistence of multiple partially folded states. In this context, the observed reduction in the α-helical content is probably not an artifact of the simulation setup but rather a reflection of the flexibility of amphipathic peptides when interfacial interactions are insufficient to promote a stable folded state.

A detailed comparison between the helicity metrics employed in this study—DSSP, αRMSD, and α-block—is provided in Appendix B.1. Briefly, the three methods yield highly consistent trends across peptides and force fields. The two continuous metrics, αRMSD and α-block, show a strong linear correlation ($r^2 \approx 0.97$,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

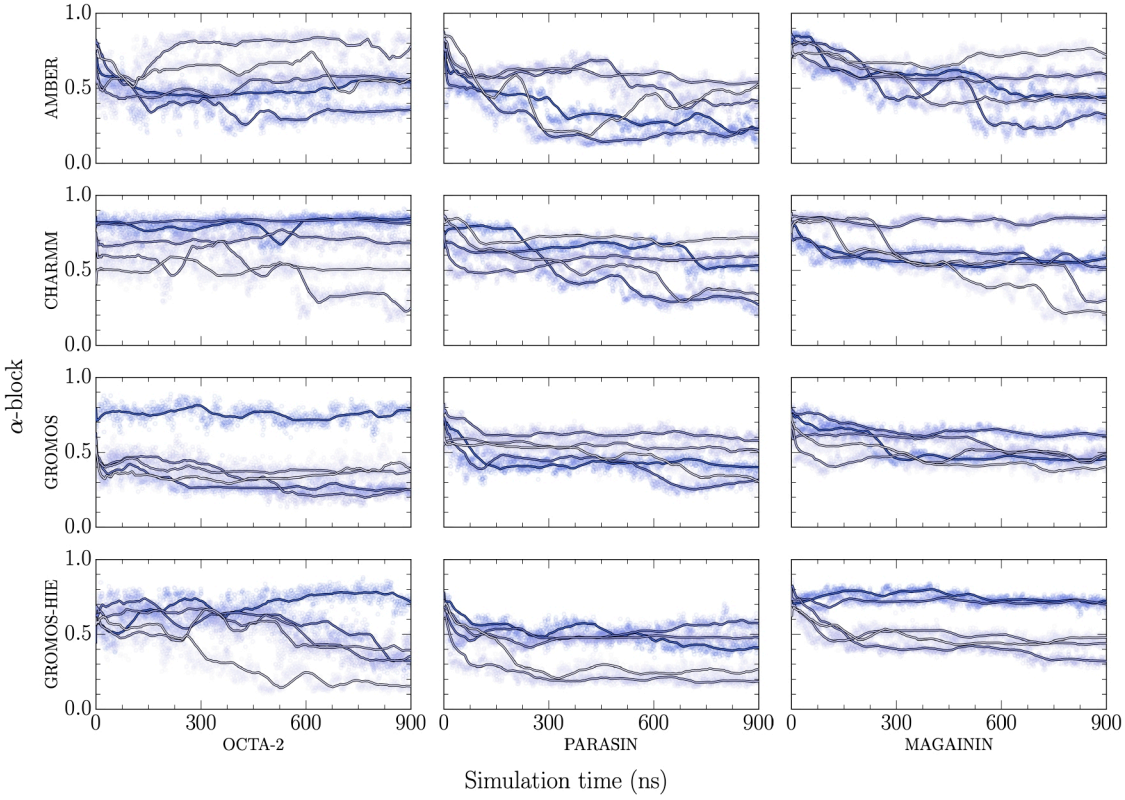


Figure 4.5: Time evolution of peptide helicity calculated with the α-block metric for the shallow configuration. Each panel shows the variation of α-block values (Y-axis) over simulation time (X-axis), providing an additional assessment of α-helical stability. The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the five independent simulation replicas, displayed as rolling averages of the instantaneous α-block values (shown as faint points).

$\tau \approx 0.84$), confirming that both capture the similar underlying structural information despite their different definitions. Thus, α-block tends to underestimate the helicity of well-folded peptides and, at the same time, overestimate that of mostly extended conformations. DSSP also correlates well with αRMSD ($r^2 \approx 0.95$, $\tau \approx 0.81$), although its discrete nature limits a strict 1-to-1 correspondence. In general, the analysis supports the consistency in the trends, but also highlights the differences in the numerical values.

Peptide–membrane interaction The conformational changes observed at the interface could be associated with a change in the way the peptide interacts with the membrane. A first way to measure such an interaction is the number of contacts established between the peptide and the lipids, represented in Figure B.5. In these plots, a contact is counted when an atom of a lipid is within 6 Å of any of the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

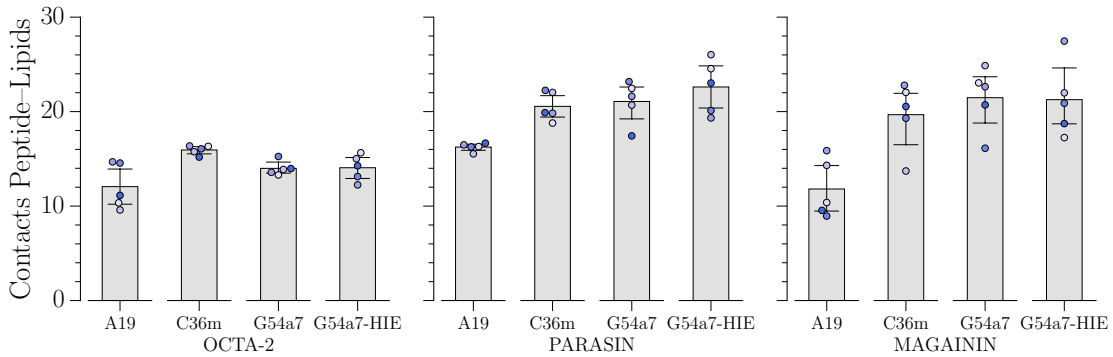


Figure 4.6: Mean number of peptide–lipid contacts in simulations starting from the shallow configuration. Each plot corresponds to a different peptide (octa-2, parasin, and magainin), with bars grouped by force field (A19 for AMBER, C36m for CHARMM, G54a7 and G54a7-HIE for GROMOS). Circles represent the values obtained from individual replicas, while bar heights indicate their mean and error bars represent one standard deviation.

peptide atoms.

As expected, the shorter the peptide, the fewer contacts formed. However, in AMBER and CHARMM there are simulations in which the number of contacts formed is comparable for the shortest and longest sequences, mostly due to the larger separation of the latter from the membrane previously discussed (Figure 4.1). This increased separation can also be associated with a larger fluctuation of the number of contacts along the simulation, since a wider distribution is observed in AMBER simulations.

Despite these fluctuations, the magnitude was much more stable along the simulation than any helicity metric, and it seems to stabilize in the last 200 ns of simulation. The average number of peptide–lipid contacts over this time interval for each independent replica is represented in Figure 4.6. For the shortest peptide, CHARMM yields highly consistent results among replicas, and the largest number of contacts overall. The other force fields show greater variability, but no significant differences between them are noted. For parasin and magainin, all force fields produce similar trends, with AMBER showing consistently fewer contacts than the rest. This weaker interaction in AMBER simulations might be related to the higher variability previously noted in the peptide conformation.

Nevertheless, no evidence suggests that variations in the number of contacts directly affected the peptide helicity (Figures B.6 to B.8), as confirmed by the low Pearson and Kendall correlation coefficients (Tables B.7 to B.12).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

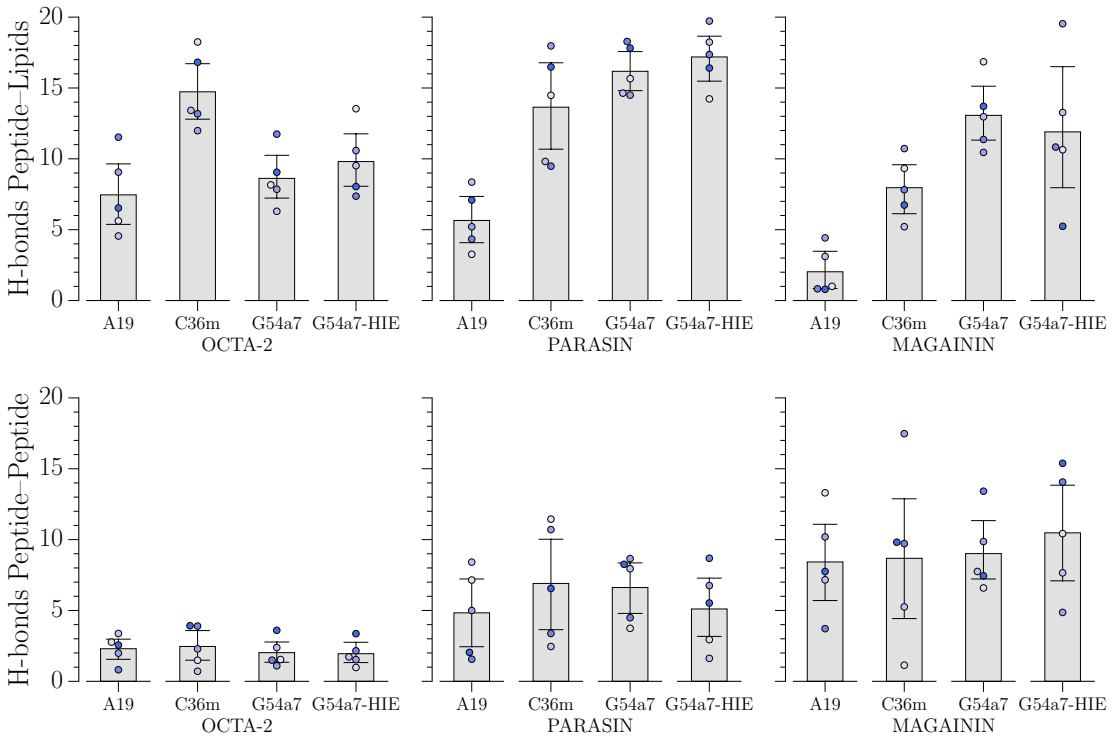


Figure 4.7: Mean number of hydrogen bonds formed between the peptide and the lipids (top panels) and within the peptide itself (bottom panels) in simulations starting from the shallow configuration. Each plot corresponds to a different peptide (octa-2, parasin, and magainin), with bars grouped by force field (A19 for AMBER, C36m for CHARMM, G54a7 and G54a7-HIE for GROMOS). Circles represent the values obtained from individual replicas, while bar heights indicate their mean and error bars represent one standard deviation.

Hydrogen bonding and secondary structure Unlike simple contacts, the formation of H-bonds provides more direct evidence of peptide–lipid interactions. The temporal evolution of the number of such bonds (Figure B.9) follows the same general trends as those observed for contacts: a higher mean value in CHARMM for the shortest peptide, and lower values in AMBER simulations of the longer ones. Overall, the number of H-bonds formed is smaller than the number of contacts and shows greater variability between replicas (Figure 4.7, upper). Although GROMOS and GROMOS-HIE appear to produce slightly more peptide–lipid H-bonds than CHARMM for the longer peptides, this relationship is not robust.

No strong correlation is observed between the number of peptide–lipid H-bonds and either the total number of contacts (Figure B.10) or the helicity metrics (Figures B.11 to B.13). Being close to the membrane does not necessarily imply H-bond formation, nor does it require a structural rearrangement of the peptide. However,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

for magainin in GROMOS and GROMOS-HIE, a weak negative trend is detected: fewer helical peptides tend to form more H-bonds with lipids, likely because of the higher flexibility of the unfolded state, which allows additional configurational rearrangements.

In contrast, H-bonds formed within the peptide display a clear relationship with secondary structure stability. Their number is consistent across replicas and force fields for each peptide (Figures B.14 and 4.7, bottom), and although individual trajectories may have explored limited regions of the conformational space, moderate positive correlations emerge when considering the entire data set. This confirms that intramolecular hydrogen bonding is the primary contributor to the stabilization of the α-helical conformation (Figures B.15 to B.17).

Peptides tend to form compact structures Once the helix was lost, the peptide backbone gained flexibility, allowing the molecule to adopt either more extended or more compact conformations. As described in Section 3.3.4, the overall shape of each peptide was approximated by an ellipsoid having the same moments of inertia as the atomic configuration. The relative lengths of its semiaxes provide a simple yet informative descriptor of peptide geometry.

An ideal α-helical peptide is expected to adopt a roughly cylindrical shape, long and slender, with two shorter axes of similar magnitude and a much longer principal axis. Consequently, the ratio between the first and second semiaxes should be large—and proportional to the length of the peptide—, while the ratio between the second and third should approach unity. This condition is not strictly satisfied for octa-2, whose three axes are similar in length due to its short sequence. In contrast, for parasin and magainin, the initial configurations lie near the reference line corresponding to cylindrical symmetry, consistent with their helical starting conformations (Figure 4.8).

During the simulations, most peptides evolved toward more spherical geometries, as reflected in a decrease in the first-to-second semiaxis ratio and an increase in the second-to-third ratio. Some exceptions are observed: for example, one CHARMM replica of magainin lost its helical structure completely and became largely extended, with only a slight bend around residues Ala9 and Lys10; this is reflected as an increase in the first-to-second semiaxis ratio. In other replicas, the helix was preserved, resulting in little change in overall shape.

The compaction of the 3-dimensional structure is further evidenced by a reduction in the end-to-end distance relative to the initial helical state (Figure B.18). This effect is less pronounced for octa-2, given its short length, and for simulations in which the overall shape remained stable, such as GROMOS runs of magainin and GROMOS-HIE runs of parasin. As expected, the length of the principal axis of inertia aligns closely with that of the end-to-end vector in both extended and he-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

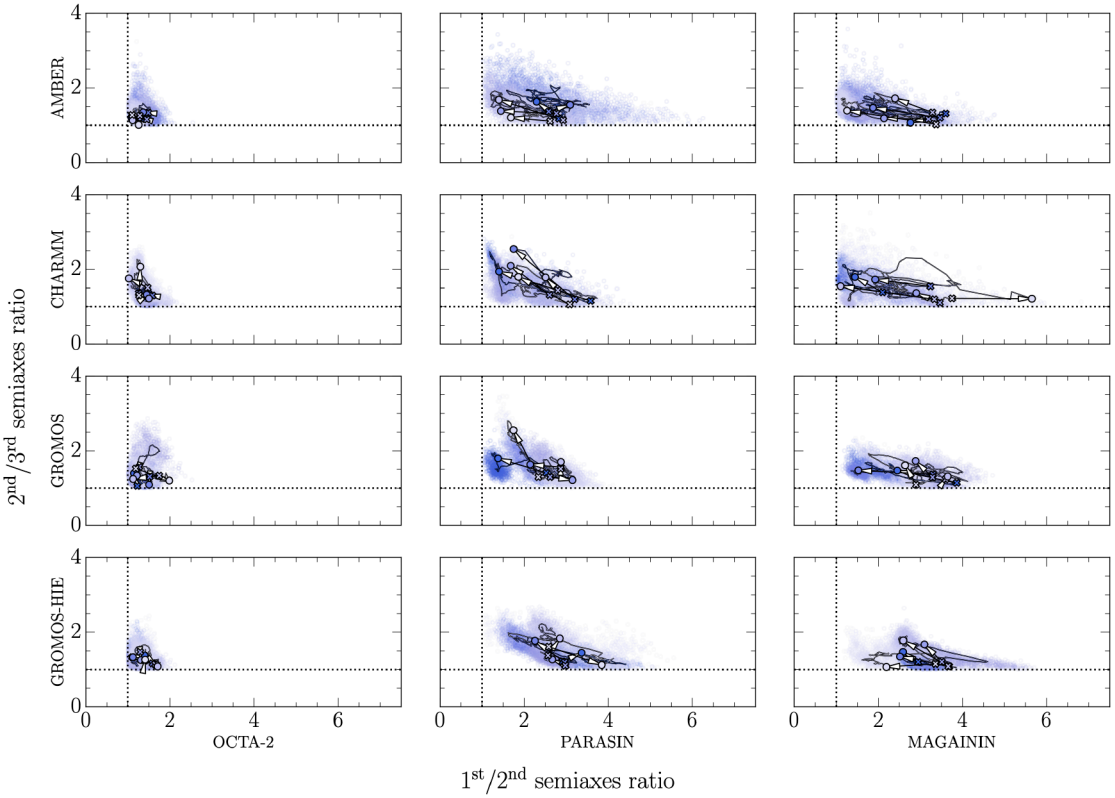


Figure 4.8: Relation between the ratios of the 1st to 2nd and 2nd to 3rd semiaxes of the equivalent ellipsoid representing each peptide in simulations starting from the shallow configuration. The initial and final states of each simulation are marked with a cross and a circle, respectively, connected by a line showing the rolling mean over a 100-frame window. Dashed reference lines indicate characteristic geometries: the horizontal line corresponds to cylindrical symmetry (2nd and 3rd semiaxes equal and shorter than the 1st), while the vertical represents disk-like shapes (1st and 2nd semiaxes equal and larger than the 3rd). States below each of these lines are geometrically forbidden.

lical conformations, yielding a strong correlation between these descriptors (Figure B.19).

4.1.2 Embedding of peptides into the bilayer

The results obtained from the shallow configuration indicate that the peptides tend to lose their initial α-helical structure and remain adsorbed near the membrane surface, without achieving stable insertion into the bilayer. This behavior, although variable across replicas and force fields, is consistent with the notion that adsorption and folding are sequential rather than simultaneous processes. In other words, peptides may first interact with the membrane in a disordered or partially un-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

folded state before acquiring a more ordered conformation once deeper insertion occurs. Such a mechanism has been reported in previous studies. For example, Nymeyer, Woolf, and García[138] investigated the adsorption of the transmembrane model peptide WALP-16* into a DPPC bilayer using replica-exchange MD. Their results suggested a two-step process: initial adsorption of a disordered peptide at the membrane surface, followed by folding into an α-helix once inserted within the hydrophobic core. Although that system and the simulation conditions differ from those employed here, the conceptual picture is compatible with many proposed models of membrane–peptide interaction.[308]

The *embedded* configuration was designed to approximate the second stage of this process; that is, the behavior of peptides that have already established a significant contact with the lipid tails. To reach this regime within feasible computational times, the equilibration protocol included bias potentials to ensure that the peptide was initially aligned parallel to the bilayer plane and positioned below the headgroup region. All external potentials were progressively removed before production simulations, allowing the system to evolve freely from that initial partially inserted state.

During these unbiased trajectories, peptides generally remained associated with the membrane, but exhibited partial emergence toward the interface over time (Figure 4.9). Its extent varied across force fields: In AMBER, peptides tended to drift further away from the bilayer center, occasionally extending their termini into the aqueous phase; visual inspection indicated that these apparent detachments are associated with local unfolding events, while the central residues remained near the phosphate groups.

Structural analysis Unlike in the simulations starting from the shallow configuration, the secondary structure of the embedded peptides remained considerably more stable and consistent between replicas.

The helicity estimated by DSSP shows much smaller fluctuations throughout the trajectories: except for a few cases, the structures reached stable states after approximately 500 ns, fluctuating slightly around neighboring discrete helicity values (Figure 4.10). AMBER simulations display the largest variability, with octa-2 unfolding completely in three out of five replicas, while in the remaining two it alternated between the four possible helicity states. In contrast, magainin retained a high degree of structural order, remaining fully helical in two replicas, partially folded ($\sim$ 65 % helical) in another two, and almost completely unfolded in one.

Results in the other force fields were more consistent. Octa-2 remains the least stable of the three peptides: In three CHARMM simulations it preserved its helical

*A 16-residue model peptide. Its sequence is GWWLALALALALAWWA.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

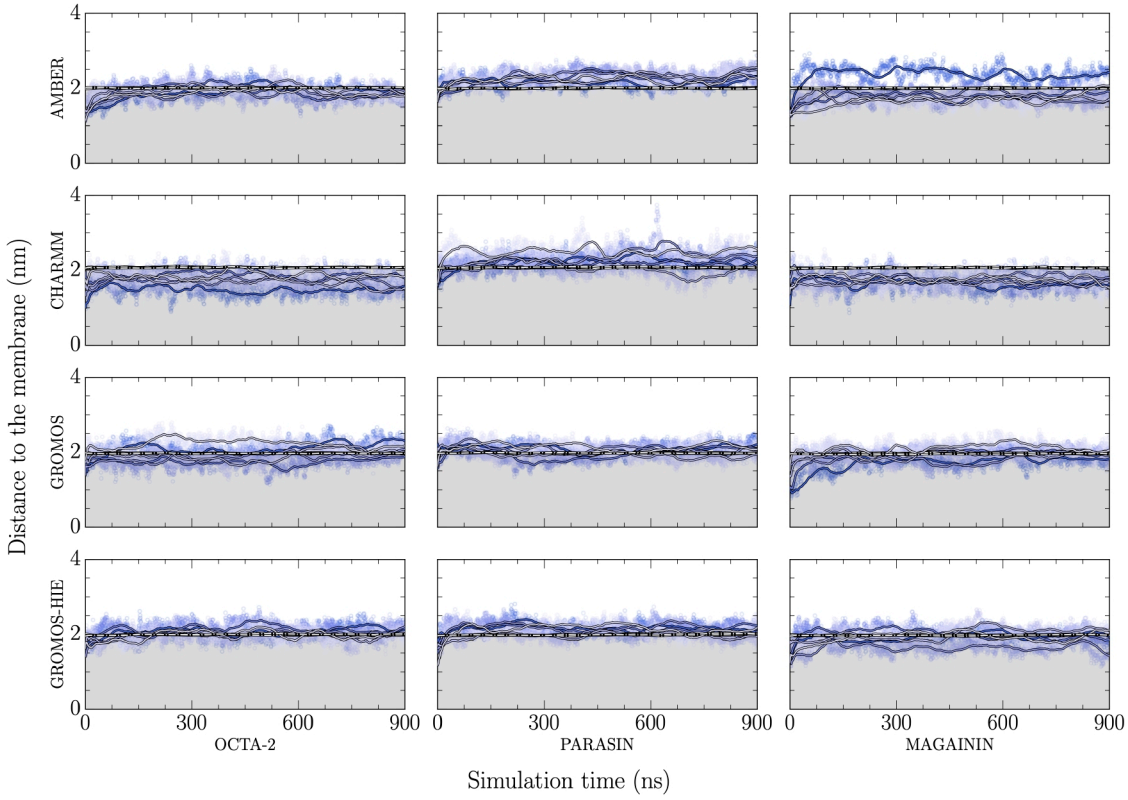


Figure 4.9: Time evolution of the peptide–membrane distance for the embedded starting configuration. Each panel shows the evolution of the distance along the Z coordinate—perpendicular to the bilayer plane—between the centroid of the peptide and the center of the membrane, as a function of simulation time. The plots are arranged by force field (rows) and peptide (columns). In each panel, the thin blue lines represent the rolling mean (window of 100 steps) of the instantaneous peptide–membrane distance, sampled every 6 ns, for five independent simulation replicas. The gray shaded area indicates the hydrophobic core of the membrane, while the dashed white line at approximately 2 nm marks the average position of the phosphorus atoms of the lipid headgroups, corresponding to the membrane surface. Smaller distances correspond to deeper peptide embedding within the bilayer, while larger values indicate migration toward or above the interface.

structure, whereas in all GROMOS and GROMOS-HIE trajectories it became completely unfolded. Interestingly, the only cases where another peptide fully lost its secondary structure occurred in CHARMM, where two parasin simulations reached an estimated helicity of zero. For this peptide, the overall behavior is consistent across AMBER, GROMOS, and GROMOS-HIE, although the helicity varies considerably among replicas, ranging between 30 % and 75 %. Magainin again shows the highest reproducibility in all force fields except AMBER, where greater dispersion is observed; in the remaining force fields, its helicity lies within the $60 - 80$ %

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

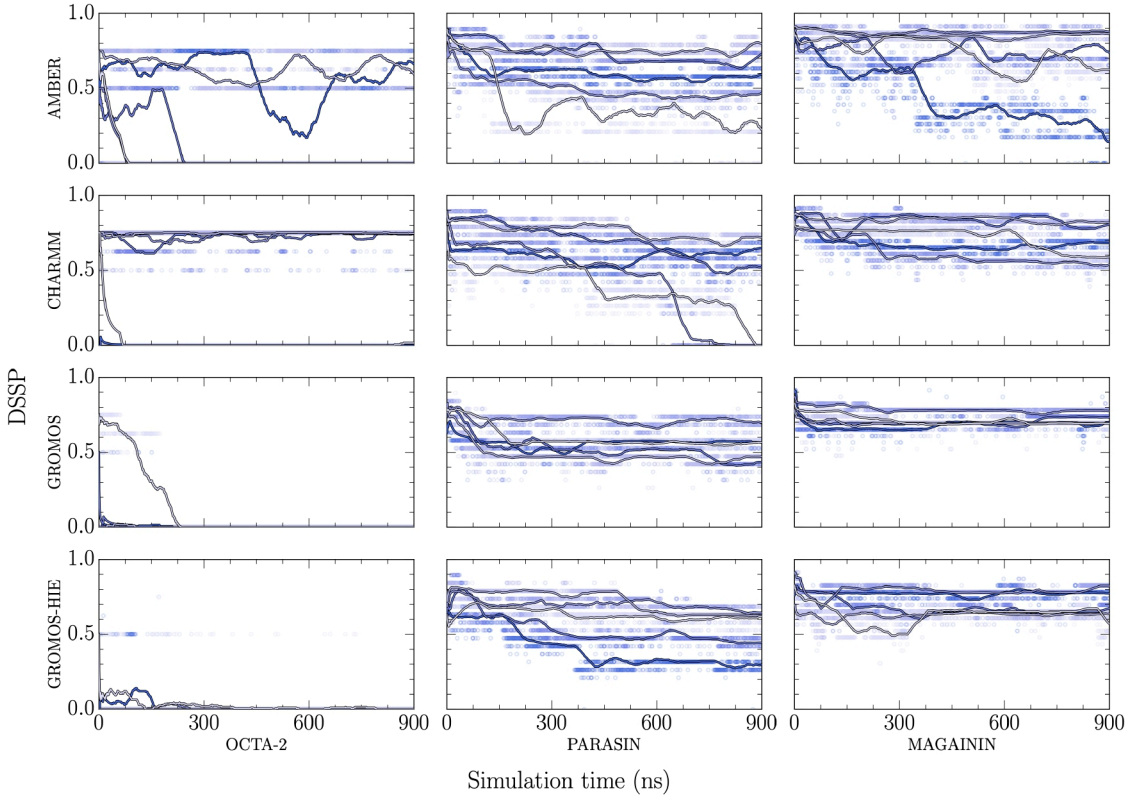


Figure 4.10: Time evolution of peptide helicity calculated with the DSSP method for the embedded starting configuration. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using DSSP. The plots are arranged by force field and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, calculated as rolling averages of the instantaneous DSSP values (shown as faint points), to highlight the overall trends. All panels share identical axes and scales, enabling direct comparison across peptides and force fields.

range.

When helicity is evaluated using the continuous αRMSD metric (Figure 4.11), its absolute values differ from those of DSSP, although the qualitative trends remain comparable. Both descriptors identify octa-2 as largely unfolded, yet two simulations—one in CHARMM and one in GROMOS-HIE—stand out. In these, the DSSP metric reports a helicity close to zero for most of the trajectory, whereas that of αRMSD stabilized after the first 200 ns around 0.3 and 0.5, respectively. This indicates that, although the H-bonds characteristic of an α-helix are not formed, the spatial arrangement of the backbone atoms was close to that of a helical conformation (Figure 4.12, upper left). In all other simulations, both helicity metrics follow similar trends, although αRMSD values tend to be systematically higher.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

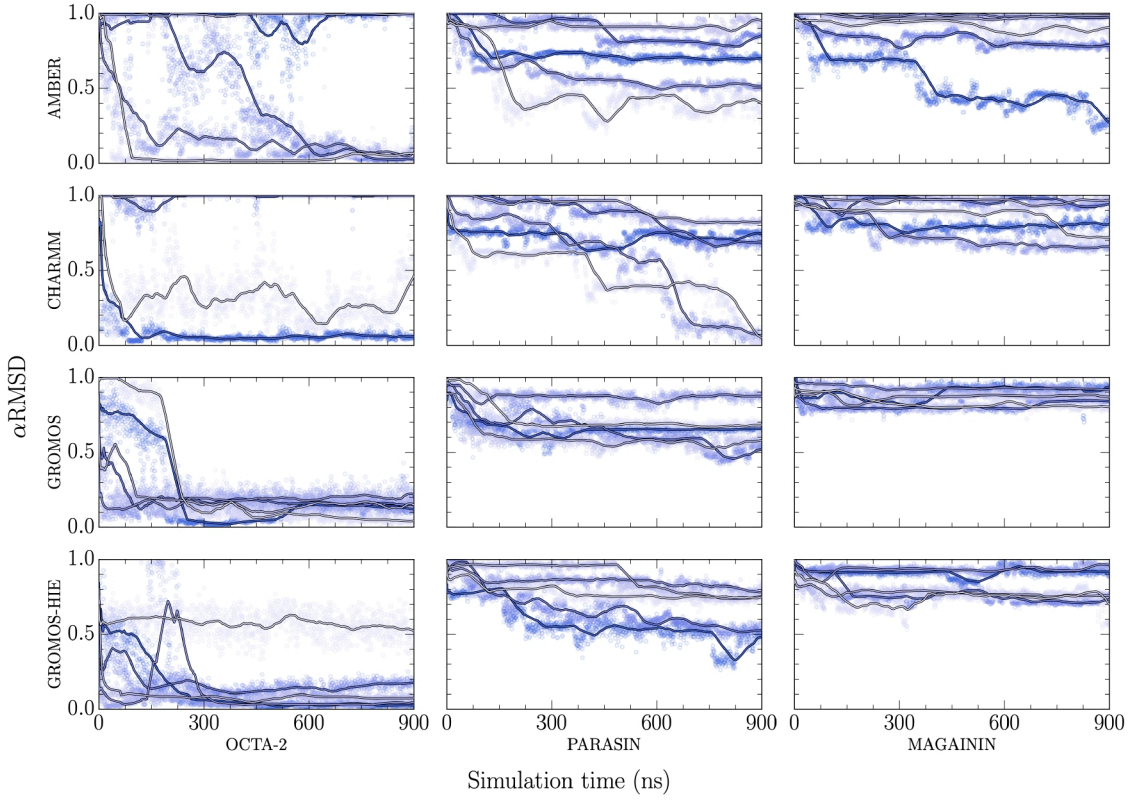


Figure 4.11: Time evolution of peptide helicity calculated with the αRMSD metric for the embedded starting configuration. Each panel shows the variation of αRMSD values (Y-axis) over simulation time (X-axis), providing a continuous measure of α-helical content. The plots are arranged by force field and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, computed as rolling averages of the instantaneous αRMSD values (displayed as faint points), to emphasize overall trends. All panels share identical scales to allow direct comparison across peptides and force fields.

These observations suggest that αRMSD is much more lenient in identifying helical structures than DSSP. Its continuous definition captures finer geometric details, but the physical interpretation of its absolute values is less straightforward, whereas DSSP is a well-established metric with clearly defined structural meaning for each value.

The helicity estimated using the α-block metric is even more consistent across replicas and force fields (Figure 4.13), showing smaller fluctuations within individual simulations than αRMSD. However, this increased consistency is not necessarily advantageous. For example, in GROMOS-HIE simulations of octa-2, DSSP identified all replicas as completely unfolded, while αRMSD detected that one peptide retained a geometry close to a helix. α-block similarly reports a non-zero helicity

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

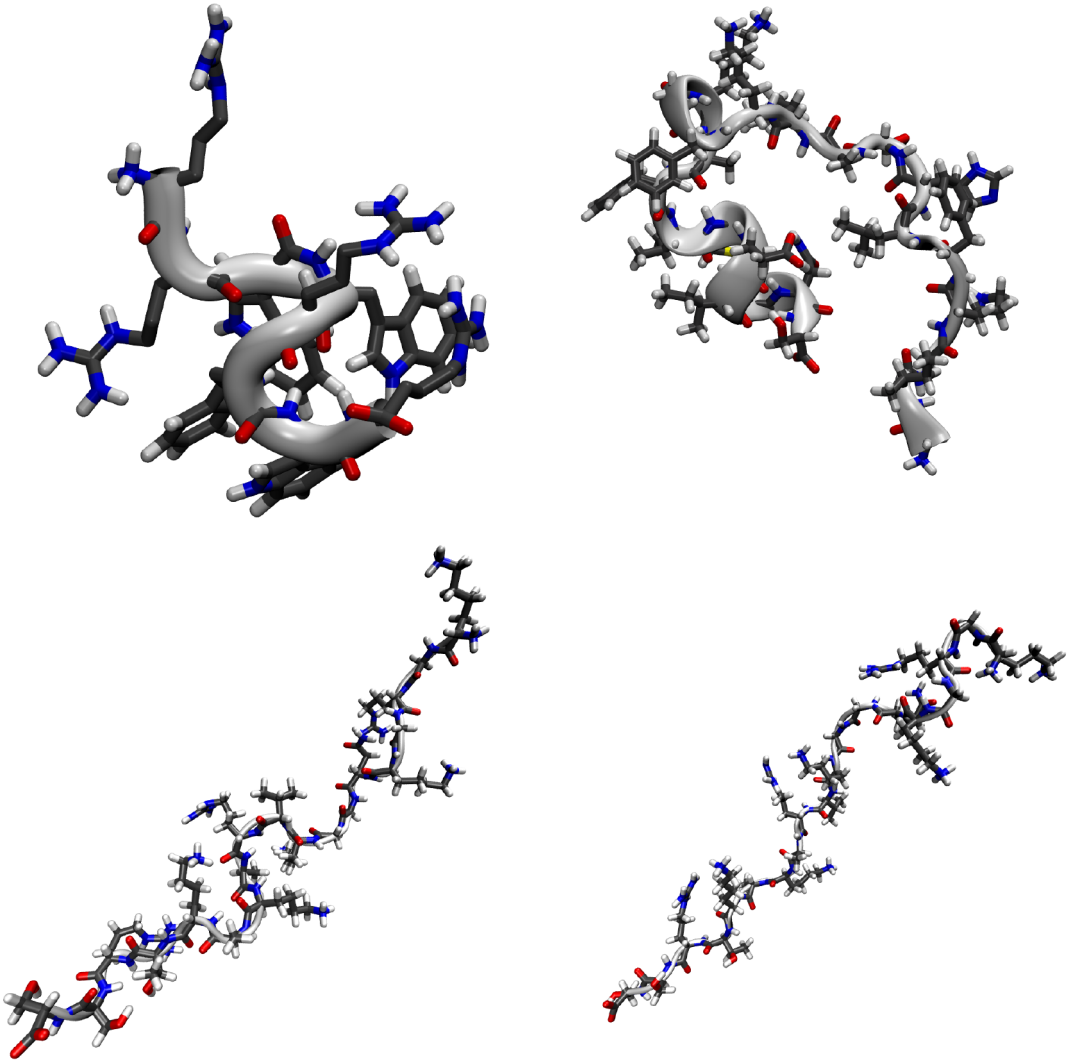


Figure 4.12: Representative final structures of the peptides in four different simulations. The peptides are not shown at the same scale. Backbone structures are highlighted as white ribbons. (Upper left) Octa-2 in the 5th GROMOS-HIE simulation: according to DSSP, the peptide lacks helical content, whereas αRMSD yields a helicity of 0.5. The backbone conformation resembles a helix, but no H-bonds are formed, explaining the discrepancy between the metrics. (Upper right) MAGAININ in the 1st AMBER simulation, where all metrics indicate partial helicity. (Lower panels) PARASIN in the 3rd (left) and 5th (right) CHARMM simulations, showing nearly zero helicity in both DSSP and αRMSD, consistent with predominantly extended conformations. All structures were rendered using VMD.[307]

for this peptide, but for clearly unfolded structures it tends to overestimate helicity relative to αRMSD.

These observations confirm that the use of α-block as a helicity metric is limited.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

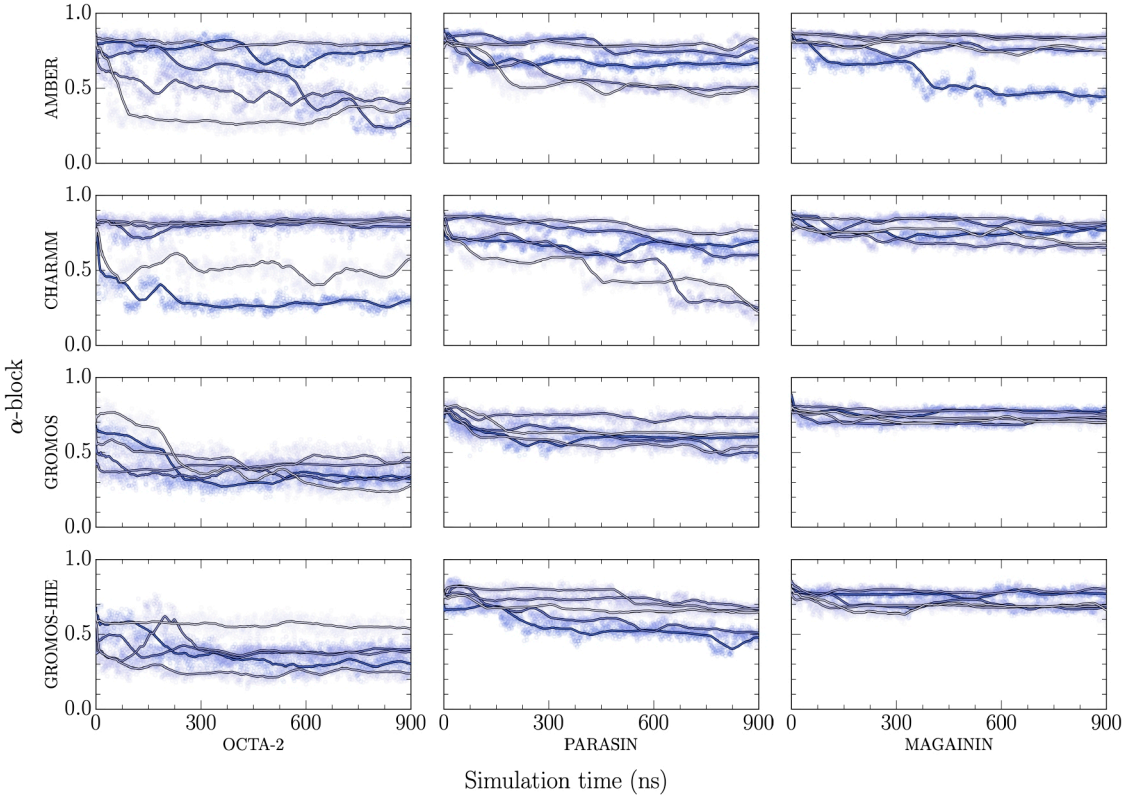


Figure 4.13: Time evolution of peptide helicity calculated with the α-block metric for the embedded starting configuration. Each panel shows the variation of α-block values (Y-axis) over simulation time (X-axis), providing an alternative continuous measure of α-helical content. Thin blue lines represent the helicity evolution of five independent simulation replicas, computed as rolling averages of the instantaneous α-block values (shown as faint points), to emphasize overall trends. All panels share identical axes and scales to facilitate direct comparison across peptides and force fields.

As shown in the first set of simulations (Eq. B.2), the effective range of α-block is narrower than that of αRMSD: it yields lower helicity values for well-folded states and higher values for unfolded structures.

Focusing on individual simulations of the longer peptides, two of them deviate from the consensus of the others: two CHARMM simulations of parasin and an AMBER simulation of magainin (Figure 4.12, bottom row and upper right, respectively). In the latter, a long region of extended residues is found between the N-terminus and Lys11. A small helix spans up to Ala15, but it is more compact than an α-helix—it is a 3_{10}-helix*. A longer α-helix appears between Phe16 and the C-terminus. DSSP only identifies this last section as helical, while the value of α-block (~ 0.5) seems to indicate that half of the structure is helical, despite most

*The 3_{10} helix is thinner as a result of a pattern of H-bonds ($i \to i+3$).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

of the residues clearly being part of coils. The value given by αRMSD ($\sim$ 0.3) is much closer to what one could assume to be the helicity of the structure shown. The other two CHARMM simulations of parasin converged to extended conformations, and again the lower αRMSD values—close to zero—reflect this structural state more accurately than α-block, which tends to overestimate residual helicity in such cases.

Therefore, despite its limitations, αRMSD appears to be the most suitable metric to quantify the helical content of the structures obtained in our simulations. Its continuous nature allows it to capture subtle conformational changes, while its numerical values remain largely consistent with qualitative assessments based on visual inspection.

Change in the shape of the peptide Unlike in the first batch of simulations, where a trend toward the acquisition of a more compact 3-dimensional arrangement was observed, the conservation of higher helicity when starting from the embedded configuration implies that fewer variations with respect to the initial structure should appear (Figure 4.14).

The shorter octa-2 still maintained roughly the same global shape, independently of the degree of helicity measured. Parasin simulations in CHARMM where the peptide acquired an extended conformation (Figure 4.12, bottom row) follow the opposite behavior as the rest of the replicas, leaning towards higher first-to-second semiaxis ratios. Others formed more rounded shapes as a result of the partial unfolding of the termini and their later retraction towards the helix. This reduced the lengths of the first semiaxis and increased those of the second.

Lastly, as most magainin simulations preserved most of the helix, the shape of the peptide did not change, except for two cases in AMBER. One is that of the simulation in which this peptide unfolded the most: The first half of the peptide acquired a coiled conformation, which then moved close to the helix (Figure 4.12, upper right). In the other case, only one terminus unfolded partially, but the coiled region then extended perpendicular to the peptide, reducing the length of the first semiaxis and also increasing that of the second.

Differences in the interfaces Overall, regardless of the simulation protocol, peptides in AMBER exhibited the least structural stability and were more prone to unfolding. This behavior may arise from differences in the parametrization of the peptide–lipid interaction, variations in the interfacial properties of the membrane, or a combination of both.

The area per lipid and thickness of the lipid bilayers are comparable in the AMBER, GROMOS, and GROMOS-HIE simulations, while the CHARMM membranes show slightly larger lateral spacing between lipids (Figures B.20 and B.21). The fluctuations of these properties, quantified as the standard deviation of their tempo-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

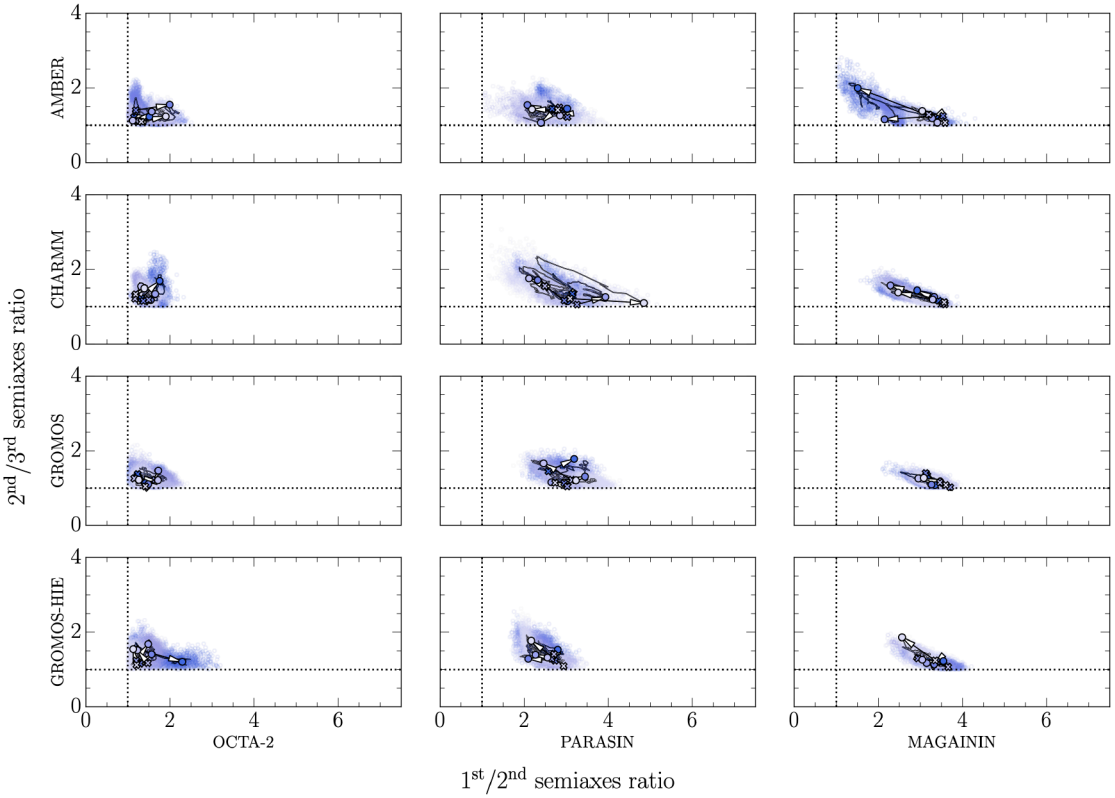


Figure 4.14: Relation between the ratios of the 1st to 2nd and 2nd to 3rd semiaxes of the equivalent ellipsoid representing each peptide in simulations starting from the embedded configuration. The initial and final states of each simulation are marked with a cross and a circle, respectively, connected by a line showing the rolling mean over a 100-frame window. Dashed reference lines indicate characteristic geometries: the horizontal line corresponds to cylindrical symmetry (2nd and 3rd semiaxes equal and shorter than the 1st), while the vertical line represents disk-like shapes (1st and 2nd semiaxes equal and larger than the 3rd). States below these lines are geometrically forbidden

ral distributions (Figures B.22 and B.23), are also similar for AMBER, GROMOS, and GROMOS-HIE, but consistently higher in CHARMM. Thus, differences in the membrane surface structure are unlikely to explain the distinct peptide behavior observed in AMBER.

An alternative explanation involves variations in membrane permeability to water, influenced by the combination of force fields and water models—AMBER simulations being the only ones performed with a 4-point water model. For each trajectory frame, the density profiles of water, total lipids, and phosphorus atoms were calculated along the Z axis of the simulation box and then averaged over the final 200 ns, corresponding to the structural equilibrium of the peptides. The bulk water density corresponds to that of the water model in the simulation at 300 K, and de-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

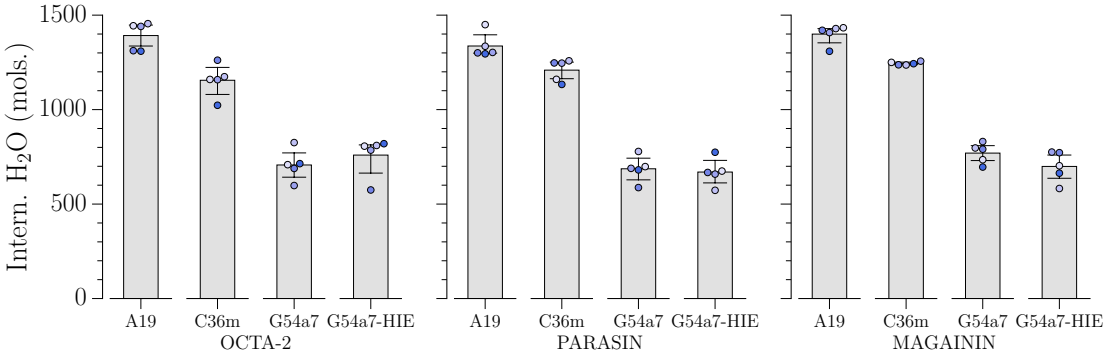


Figure 4.15: Mean number of water molecules that penetrate the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation. Water adsorption was estimated from the integral of the water density profile within the maxima of the lipid phosphorus atom distribution.

creases progressively across the bilayer as the lipid density increases. By integrating the water density within the region of non-zero lipid density, the mean number of water molecules adsorbed at the membrane interface was obtained (Figure B.24). This value depends primarily on the force field rather than on the peptide sequence, being higher in AMBER and CHARMM simulations and lower in GROMOS and GROMOS-HIE.

When the number of water molecules that penetrate deeper into the bilayer is considered, the trend changes markedly. By integrating the water density within the maxima of the phosphorus distribution—approximately corresponding to the peptide insertion depth—it is observed that more water penetrates the AMBER membranes than the CHARMM ones, which in turn contain more internal water than those simulated with GROMOS or GROMOS-HIE (Figure 4.15). The peptides in AMBER also remained farther from the bilayer center, making them more exposed to water molecules. This increased hydration may have interfered with both peptide stability and peptide–lipid interactions.

4.1.3 Dynamical behavior of the peptide at the interface

The interaction between the peptide and the lipids can affect not only its structure but also its dynamical behavior. For a peptide adsorbed at the surface of a lipid bilayer—a 2-dimensional interface—, its displacement should follow a characteristic Brownian diffusion regime. To isolate the translational motion of the peptide as a whole and minimize the contribution from the internal conformational changes, the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

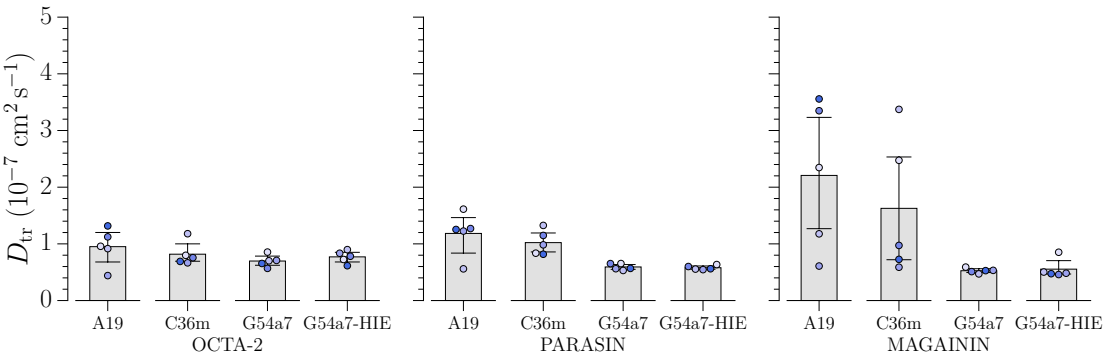


Figure 4.16: Translational diffusion coefficients of the peptides in simulations starting from the shallow configuration, computed over the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values obtained from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation.

distribution of displacements within the last 200 ns of simulation was fitted to Eq. 3.87, using displacements projected onto the membrane plane and within $\tau = 1$ ns intervals.

Figure 4.16 shows the translational diffusion coefficients estimated by least-squares fitting for the simulations starting from the shallow configuration. The results indicate a generally consistent dynamical behavior across replicas and force fields. Exceptions are observed in AMBER and CHARMM simulations of parasin and, especially, magainin, where the estimated diffusion coefficients span nearly an order of magnitude. For the shorter peptide, AMBER and CHARMM exhibit a much narrower range of values. GROMOS and GROMOS-HIE display highly consistent behavior, both between each other and among replicas: the longer peptides, comparable in size, show similar diffusion rates, which are lower than those of octa-2. Consequently, this shorter peptide was able to displace farther within the same time interval.

When simulations begin with the peptide positioned below the plane of the lipid phosphate groups, its diffusion becomes considerably more consistent and markedly slower, as shown in Figure 4.17. This effect is particularly evident in AMBER and CHARMM simulations. In AMBER, the largest contrast is observed: peptides now display the slowest absolute diffusion rates. In contrast, CHARMM continues to yield the highest diffusion coefficients. GROMOS and GROMOS-HIE results are consistent with those of the shallow configuration, showing comparable D_{tr} values, which remain higher for the shorter peptide. Only a few simulations deviate from the consensus within each force field, corresponding to cases where pronounced peptide unfolding was previously observed. However, no clear correlation can be

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

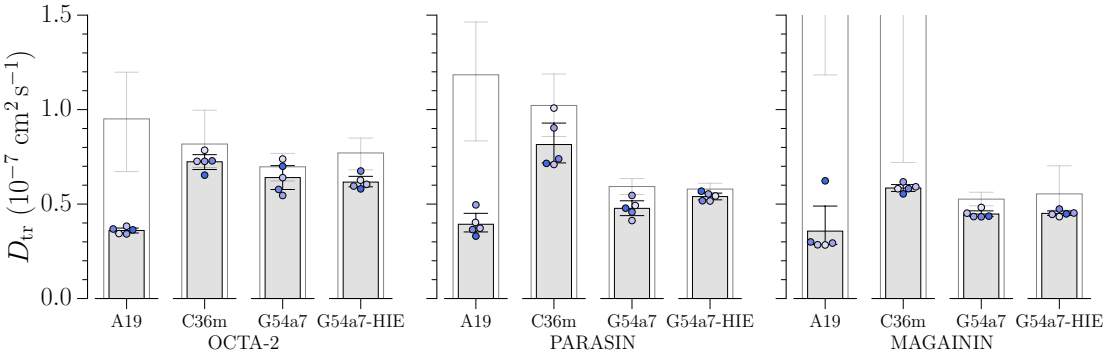


Figure 4.17: Translational diffusion coefficients of the peptides in simulations starting from the embedded configuration, computed over the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values obtained from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation. The gray bars represent the mean and standard deviation of peptide diffusion coefficients from the shallow configuration (Figure 4.16) as a reference.

established between diffusion behavior and peptide helicity, as this relationship is not consistently manifested across systems.

Regardless of the starting configuration, the mass of the peptide appears to influence the diffusion, which could be expected as the Brownian motion comes from the transference of momentum from the surrounding molecules. Although neither the increase in mass due to isotopic substitution in GROMOS-HIE—which ranges from 7.2 % for magainin to 11.7 % for octa-2—, nor the mass difference between magainin and parasin—approximately a 25 %—is sufficient to have a noticeable impact, the 43 % increase in mass from octa-2 to parasin slows its movement.

To analyze the differences in diffusion between the two starting configurations, the dependence of the translational diffusion coefficient on the position of the peptide within the membrane was examined (Figure 4.18). The position of the peptide corresponds to the average distance from the center of the membrane over the same time interval used to compute the diffusion coefficients. The results clearly show that peptide detachment from the membrane reduces the constraints imposed by the lipid interactions, thus enhancing mobility. In contrast, the motion of deeply embedded peptides is hindered by the surroundings. Sampling within the membrane interior is limited—no peptide reached the midpoint of the bilayer—, but the diffusion coefficient is expected to approach an asymptotic limit as the peptide moves away from the membrane. At sufficiently large distances, where lipid interactions become negligible, D_{tr} should converge to the diffusion coefficient of the peptide in

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

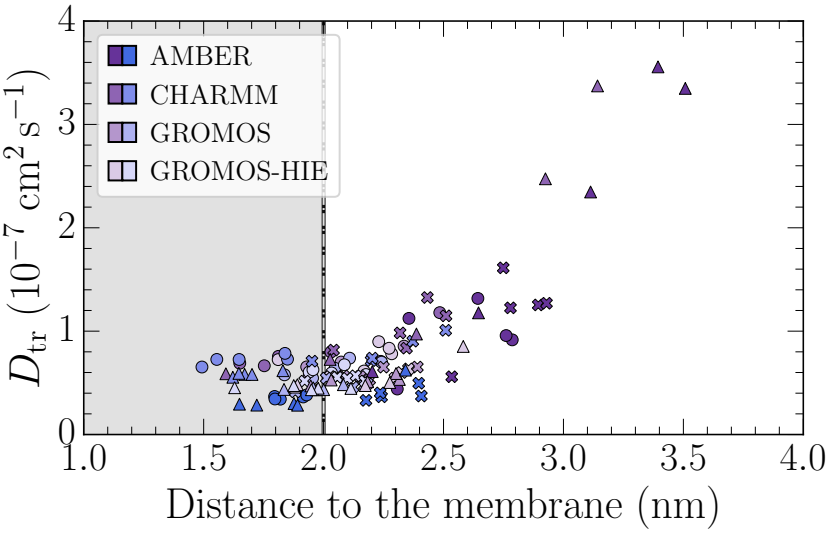


Figure 4.18: Relation between the translational diffusion coefficients of the peptides and their average distance to the center of the membrane. Both magnitudes were computed over the last 200 ns of simulation. The marker indicates the peptide, circles for octa-2, crosses for parasin and triangles for magainin; while its color represents the initial configuration of the simulation, purple for the shallow configuration and blue for the embedded. The gray shaded area represents the internal region of the membrane, delimited by a dashed white line at approximately 2 nm, which is the average position of the phosphorus atoms of the lipids.

bulk water.*

4.1.4 Performance and stability of the GROMOS-HIE and Reaction Field Schemes

A central limitation of atomistic peptide–membrane simulations lies in the high computational cost required to reach the timescales where relevant conformational changes occur. In particular, the temporal resolution needed to integrate the motion of hydrogen atoms—typically restricting the time step to 2 fs—makes long trajectories extremely demanding, especially for large solvated bilayer systems. This limitation motivated the exploration of alternative approaches capable of extending the simulation time step without compromising physical accuracy. Overcoming these limitations is essential to enable the massive production of MD simulations that could provide statistically meaningful results for data-driven approaches, or serve as physically grounded training sets for machine learning models aimed at predicting

*As the diffusion was computed in a 2-dimensional case, this limiting value would not match a that of a 3-dimensional isotropic diffusion.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

peptide–membrane interactions.

In the previous sections, the hydrogen isotope exchange (HIE) scheme was validated as a robust method to improve computational efficiency while producing results comparable to those of standard GROMOS simulations. By increasing the mass of polar hydrogen atoms, the integration time step could be extended to 6 fs, causing a direct 3-fold reduction in the number of required integration steps without introducing numerical instability or noticeable artifacts in the properties analyzed. Comparison with standard GROMOS simulations confirmed that such isotopic substitution does not significantly alter peptide conformations, membrane structure, or peptide–lipid interactions within the simulated timescales. The largest relative mass increase—approximately 12 % in octa-2, lower in longer peptides—had no detectable impact on helicity, insertion depth, or diffusion behavior, although possible effects on phenomena like H-bond kinetics were not explicitly assessed. These results establish the HIE approach as a physically consistent means to accelerate atomistic simulations.

Having confirmed that the HIE scheme preserves structural reliability, the next logical step was to evaluate whether further performance gains could be achieved by simplifying the treatment of long-range electrostatics, which remains the principal computational bottleneck in MD. The particle-mesh Ewald (PME) method provides an accurate description of electrostatic interactions, but comes with a considerable computational workload, particularly when only CPUs are available. Although GPU acceleration mitigates this cost, PME calculations still dominate the runtime for solvated membrane systems. As a potential alternative, we examined the reaction field (RF) method, in which electrostatic interactions beyond a cutoff are replaced by a dielectric continuum. This distance was selected to be equal to that used for van der Waals interactions. This approach can reduce the computational cost to up to 30 %,[309] at the expense of a more approximate treatment of long-range Coulomb interactions.

To assess the impact on the simulation results, the same systems previously analyzed under PME were simulated using the RF method under the GROMOS-HIE force field. The simulations started with the embedded configuration, as previously it yielded the most reproducible results. The obtained trajectories show the equilibrium position of the peptides relative to the center of the bilayer to be indistinguishable from that obtained in the PME simulations (Figures B.25). The peptides remained embedded near the interface (Figure 4.19), and the structural descriptors of the bilayer—area per lipid, thickness, and hydration—show no significant variations (Figures B.26 to B.31). The slightly deeper insertion observed for magainin in a few replicas is likely coincidental rather than an effect of the electrostatic approximation.

The estimations of the helicity are also consistent across electrostatic schemes

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

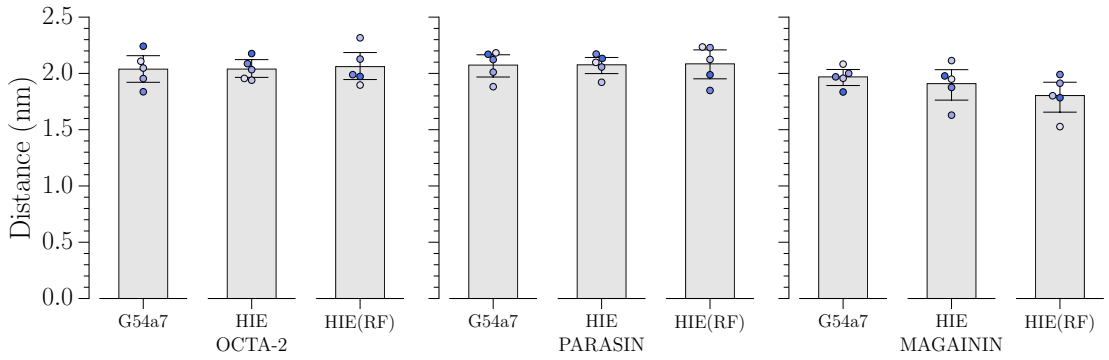


Figure 4.19: Mean distance between the peptide and the center of the membrane in simulations starting from the embedded configuration, averaged over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

(Figures B.32, to B.34). During the final 200 ns of simulation, the helicity values obtained under RF remained stable and are comparable to those from PME-based simulations. Although the RF trajectories show slightly higher variability among replicas, the averaged results across force fields and peptides are compatible (Figure 4.20). For octa-2, some replicas retained non-zero helicity according to DSSP—supported by αRMSD and α-block metrics—indicating the stabilization of the structure through backbone H-bonds. Parasin shows the opposite tendency, with greater unfolding in the outlier replica, while magainin exhibits the most reproducible behavior across all conditions.

The translational diffusion coefficients of the peptides (Figure 4.21) further confirm the consistency between the methods: the diffusion coefficients scale inversely with the length of the peptide and are independent of the the treatment of electrostatics. Together, these analyses indicate that the reaction field method reproduces key structural and dynamical features of the system with a negligible deviation from the PME reference.

Overall, the combination of the HIE scheme and the reaction field method for the calculation of electrostatics offers a practical route to substantially accelerate atomistic simulations of peptide–membrane systems without compromising the results. The gain in performance—arising from both the extended time step and the simplified electrostatics—enables the generation of longer trajectories or larger systems within the same computational cost. This makes the approach particularly attractive for large-scale simulation needs and for the construction of simulation databases.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

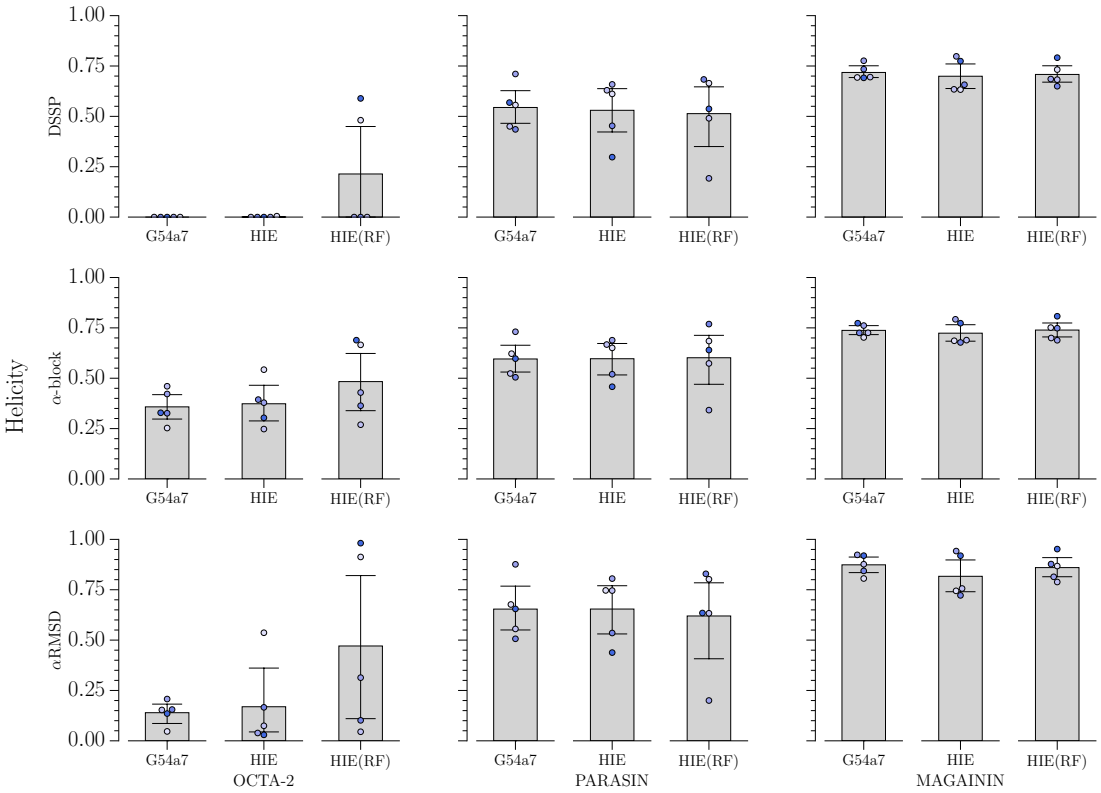


Figure 4.20: Mean helicity of the peptides in simulations starting from the embedded configuration, averaged over the last 200 ns **of simulation.** Each row corresponds to a different helicity metric—DSSP, α-block, and αRMSD—, and each column to a peptide. Bars are grouped by force field—G54a7, G54a7-HIE, and G54a7-HIE(RF). Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation. The time evolution of helicity for each case is presented in Figures B.32 to B.34.

4.1.5 Long-timescale effects in the structure

To assess whether longer simulations could reveal structural transitions beyond those accessible in the 900 ns-trajectories, additional MD runs were performed using the GROMOS-HIE force field—and PME to compute the electrostatics, as all simulations were conducted using GPUs—, which provided the best compromise between stability and computational efficiency. These extended simulations, covering up to 15 µs per system, were carried out for all previously studied peptides, starting from both the shallow and embedded configurations, and including an additional peptide, chensinin-1b.

Throughout these trajectories, all peptides remained associated with the membrane interface. No detachment into the aqueous phase or complete insertion into

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

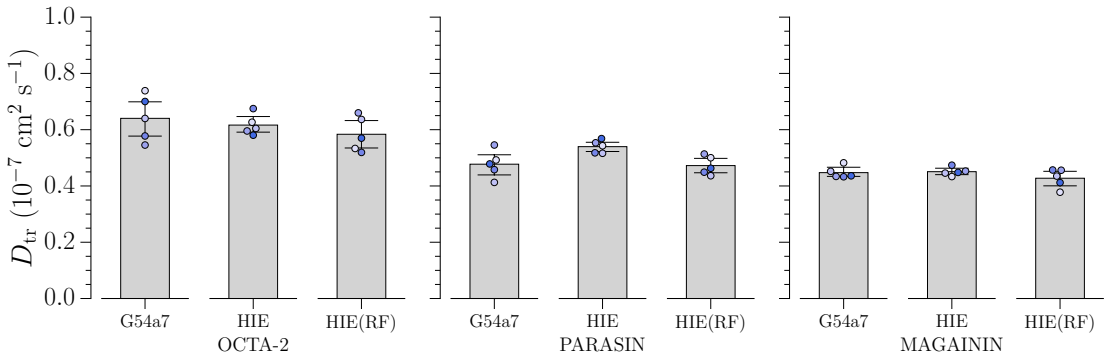


Figure 4.21: Translational diffusion coefficient of the peptides in simulations starting from the embedded configuration, averaged over the last 200 ns **of simulation.** The coefficient represents the lateral motion of the peptide within the plane of the membrane. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

the bilayer core was observed (Figure B.35); instead, the position of the peptides fluctuated within the interfacial region. This behavior is physically reasonable: In the absence of phenomena such as peptide aggregation or transmembrane potential, there is no net driving force for a single amphipathic peptide to penetrate deeply into the bilayer or be released into the aqueous phase. The simulations therefore probe the equilibrium within the adsorbed, interfacial regime rather than the full insertion or the dissociation from the membrane.

The evolution of the secondary structure, monitored through the DSSP helicity metric (Figure 4.22), shows no clear evidence of reversible folding–unfolding events over the 15 µs window. Both octa-2 and parasin remained largely unstructured after initial unfolding, while shallow magainin exhibited transient helical segments but no sustained recovery of the initial α-helical conformation. Only the embedded chensinin and, to a lesser extent, magainin trajectories explored a minor partial refolding event, whose amplitude is insufficient to be considered a significant structural transition. These observations suggest that within the timescales accessible to unbiased atomistic simulations, the conformational dynamics of adsorbed peptides are dominated by local fluctuations around metastable interfacial states.

If the exploration of structural changes is performed using the αRMSD metric instead (Figure B.36), more subtle variations can be observed. However, the only significant difference is a relevant partial refolding in the parasin simulation starting from the embedded configuration. Although this episode does not invalidate the previous conclusions, it does highlight the importance of using more accurate continuous metrics to capture detailed structural changes.

Although extending the trajectory length by more than an order of magnitude

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

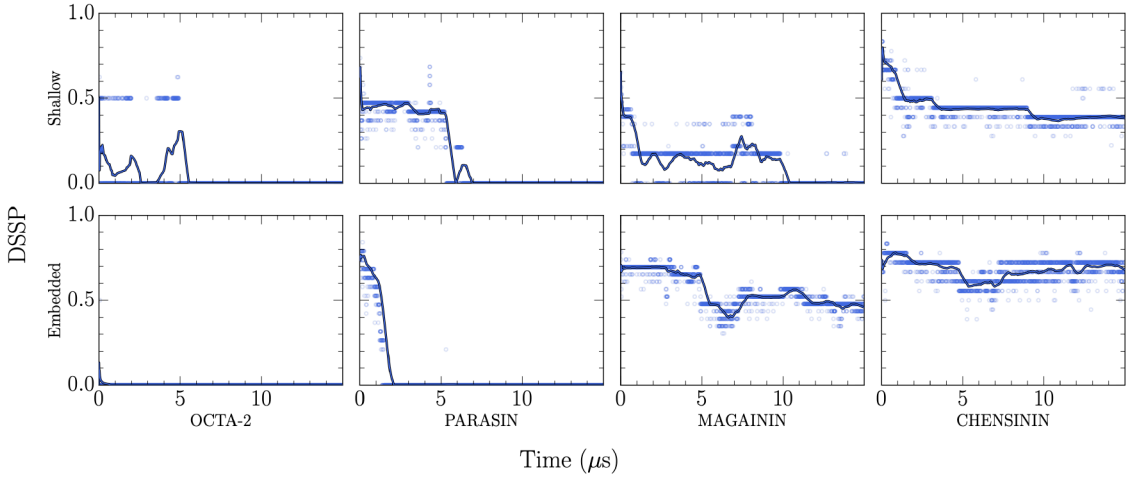


Figure 4.22: Time evolution of peptide helicity calculated with the DSSP method in long-timescale GROMOS-HIE simulations. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using DSSP. The plots are arranged by initial configuration and peptide. The thin blue lines represent the helicity evolution of the five independent simulation replicas, calculated as a rolling average of the instantaneous DSSP values (displayed as faint points), to emphasize the overall trends.

improved the sampling, the accessible conformational space remains limited. The systems appear kinetically trapped in shallow free energy minima separated by barriers that are unlikely to be crossed spontaneously within tens of microseconds. Importantly, this kinetic trapping does not indicate a methodological artifact but rather reflects the physical characteristics of the system: peptide–membrane binding and secondary structure reorganization involve slow processes that exceed the typical timescales of single-peptide MD.

4.1.6 Conclusions

The conformational behavior of short amphipathic peptides at the interfaces between the water and lipid membranes was examined using atomistic MD simulations, revealing the delicate relationship between the peptide sequence, the membrane environment, and methodological approaches. The results emphasize both the interpretative value and the inherent limitations of conventional MD to describe the structural dynamics of membrane-active peptides.

In the shallow regime, the peptides rapidly lost their initial α-helical structure and remained adsorbed near the membrane surface without achieving a stable insertion into the bilayer. Although the specific extent of unfolding varied between replicas and force fields, the overall trend indicates that adsorption and folding are

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

coupled processes. This interpretation is consistent with experimental and computational models of membrane–peptide interaction, which propose an initial disordered adsorption at the lipid headgroups followed by helix formation upon deeper penetration into the hydrophobic region.

In contrast, the embedded configuration is closer to the second stage of this process, where the peptides are already in extensive contact with the lipid tails. When initialized below the headgroup region, the peptides remained associated with the bilayer but exhibited partial emergence toward the interface. The degree of this change varied among force fields, with AMBER tending to yield more hydrated, less embedded states. These differences highlight how subtle variations in the balance between peptide–lipid and peptide–water interactions can shift the conformational equilibrium, even when overall qualitative behavior remains consistent.

The comparative analysis of helicity metrics emphasized the difficulty of quantifying helicity in short, flexible peptides. Discrete algorithms such as DSSP, designed for folded proteins, exaggerate conformational fluctuations, whereas continuous measures like αRMSD and α-block capture structural changes with greater numerical stability. α-block proven to have a narrower effective range than αRMSD, yielding lower helicity values for well-folded states and higher values for unfolded structures. This observation, together with a strong correlation between both metrics at intermediate helicity values, suggests that αRMSD can serve as a robust, lower-noise alternative for tracking partial helicity, though none of these descriptors can replace direct structural interpretation.

The GROMOS-HIE scheme provided results equivalent to those of the standard GROMOS force field, while allowing a 3-fold increase in the integration time step, significantly improving computational efficiency without noticeable artifacts. Substituting the PME treatment of electrostatics with a reaction field approach also preserved the essential structural and dynamical features, although with slightly larger dispersion between replicas. These findings indicate that methodical simplifications can be implemented safely, to the extent that their consequences are carefully validated.

Long simulations on the order of tens of microseconds confirmed that the peptides remain stably adsorbed at or within the interface, and no significant regression to higher folding states was detected. The absence of reversible folding reflects the existence of kinetic barriers and the energetic degeneracy of the interfacial environment. Even at extended timescales, conventional atomistic MD remains limited in its ability to achieve an exhaustive conformational sampling.

Overall, the results outline the practical limits of atomistic MD for peptide–membrane systems. Although the method is capable of describing local interactions and provides mechanistic insight into the early stages of adsorption, its temporal and spatial limitations hinder the observation of slow structural reorganizations. These

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

findings point toward the need for multiscale, enhanced sampling, or CG strategies in which atomistic simulations could be used to explore narrower regions of interest of the conformational space.

4.2 Systematic simulation of peptide and membrane systems and database for the study of membrane-targeting antimicrobial peptides

While atomistic MD simulations can provide atomically detailed insights into peptide–lipid interactions, their high computational cost and limited accessible timescales hinder the systematic exploration of the sequence space and membrane diversity. As discussed in the previous section, conventional MD can capture structural changes but cannot yet produce the large data sets required for comparative or predictive studies on peptide–membrane recognition. The accessible timescales remain orders of magnitude shorter than those associated with folding, insertion, and cooperative processes, making it impractical to generalize results across many peptide and membrane types.

To overcome these limitations, a CG modeling strategy was adopted. This approach reduces the level of atomic detail while preserving the physicochemical interactions that govern adsorption, insertion, and membrane perturbation. By enabling simulations on the microsecond timescale for hundreds of systems, the CG approach makes it possible to generate data sets that complement atomistic models.

The objective is twofold: (**i**) to develop an automated computational framework capable of assembling, simulating, and analyzing peptide–membrane systems in a standardized and reproducible manner; and (**ii**) to apply this framework to perform systematic CG simulations of a diverse set of membrane-active peptides interacting with model lipid bilayers, and to organize the resulting data into an open-access database.

This effort culminated in the creation of SuPepMem, an open-access database that compiles MD simulations, quantitative descriptors, and analysis tools for peptide–membrane systems. This platform combines reproducible simulation protocols with an automated analysis of the trajectories, allowing the study of how sequence and membrane composition determine the behavior of the peptides at the molecular level.

Part of the results of this section—the first stage of the development of the database—were published in the Computational and Structural Biotechnology Journal.[310] The contributions of the author to this work involved the development of the code for the analysis of the MD simulations and database backend. In addition, the SuPepMem platform was registered as intellectual property and transferred

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

to MD.USE Innovations S.L., the industrial partner participating in this doctoral project, for its long-term maintenance.

4.2.1 Development of an automated framework for systematic peptide–membrane simulations

Coarse-grained modeling framework and force-field selection The systematic simulation of peptide–membrane interactions required a protocol capable of exploring hundreds of systems under consistent physical conditions. To achieve this, a CG representation based on the Martini family of force fields was used.

The Martini approach reduces atomic detail by mapping an average of four heavy atoms to a single interaction bead. This mapping decreases the number of degrees of freedom and smooths the free energy surface, allowing the use of time steps in the order of $20 - 30$ fs, and the generation of trajectories in the order of microseconds with moderate resources. Despite this simplification, Martini retains the essential balance of the forces that dominate the interaction between peptide and lipid, their adsorption, and their interfacial organization.

In the first stage, which corresponds to the simulations the original publication[310] refers to, the Martini 2.2 and Martini 2.2P force fields were used. In a later phase, which corresponds to the simulations performed for this work, the Martini 3 force field was also included. Martini 2.2 has been widely validated for lipid membrane systems, providing a robust description of peptide–lipid interactions. However, its single-bead water model limits the representation of dielectric screening, particularly in charged environments.[230] The polarizable extension, Martini 2.2P, overcomes this limitation by representing water molecules with a three-bead model, which introduces an effective dipole moment. This refinement improves the treatment of electrostatics in charged regions and interfaces.[232] Martini 3 introduces a new set of bead types and reparametrized interactions, including the description of hydrophobic and polar interactions, which can influence peptide–membrane affinity.[145,231] The adoption of the Martini CG representation provided a practical compromise between computational efficiency and physical realism, making it possible to simulate hundreds of peptide–membrane combinations under standardized and reproducible conditions.

Selection of sequences The investigation of the peptide–membrane interactions required the definition of a representative and chemically diverse set of peptide sequences that encompasses both membrane-active and nonactive structures.

In the first phase of the database development, the sequences were extracted from the Database of Antimicrobial Peptides (DRAMP 2.0),[161] which contains thousands of natural and synthetic AMPs characterized experimentally. The peptides

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

were chosen according to three criteria: significant membrane activity, sequence below 80 residues, and predicted α-helical conformation upon interaction with lipid membranes.

In the second phase, the database was expanded to include additional membrane-active sequences from the Antimicrobial Peptide Database (APD3),[156] which contains thousands of natural and synthetic AMPs, and from a data set of nonactive synthetic controls by Tucs *et al.*[267]. For simplicity, the sequences obtained from APD3 are going to be referred to as active sequences—or similar terms—, and those extracted from the data set by Tucs *et al.* as nonactive sequences. Both data sets are very diverse in length and in physicochemical properties, but there are also sequences that are very similar within each of them, differing only in a few residues.

Sequences from both data sets were filtered according to two main criteria: Only peptides containing 10 to 30 residues were selected; this range includes a significant number of sequences with different properties, while ensuring that all systems could be simulated in boxes of similar size (250 lipids per leaflet). Then, sequence diversity was achieved by applying a custom similarity-filtering approach based on pairwise BLASTp alignments. The detailed protocol is reported in Section 3.2.3; briefly, each of the data sets was mapped to a network where the sequences were represented as nodes, with edges connecting highly similar sequences. The networks were iteratively pruned to maximize sequence diversity by removing similar sequences.

After this filtering, the active data set was reduced to 73 sequences, while the number of remaining nonactive sequences was much larger. Thus, 50 nonactive sequences were selected to be simulated.

Each of these sequences was converted into an ideal 3-dimensional α-helical model using a Python-based workflow and mapped to the Martini CG representations. The helical structure was restricted throughout the MD trajectories; results in the literature highlighted the simplification of the free energy surfaces associated with peptide adsorption into a lipid bilayer in Martini 2.2 simulations under such a structural conformation, which improves the reproducibility and convergence of the results.[311]

Membrane composition and simulation setup The initial batch of simulations comprised six different membrane models, with the aim of capturing the diversity of biological lipid environments relevant to membrane–targeting peptides. These models were selected to represent a range of electrostatic and compositional conditions, spanning from neutral mammalian membranes to charged bacterial and cancer-associated membranes.

In the second phase, a diversity of sequences and force fields was prioritized over the lipid compositions. Thus, the simulations included only two of the pervious membrane models: POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine),

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

a zwitterionic lipid representative of healthy mammalian membranes,[2,265] used as a neutral baseline model; and POPE:POPG 3:1, the same model of *E. coli* lipid membrane used in our previous atomistic simulations.

The simulation of the systems was performed assuming physiological pH, with fully charged peptide termini and standard ionization states for the side chains: Arg and Lys residues were protonated, and Asp and Glu residues were deprotonated; POPG—together with DOPS in the simulations of the first stage—lipids were assigned a negative charge, while POPC and POPE—and DOPC, DPSM, and DOPE—remained neutral. Charge neutrality within the simulation box was maintained by adding a neutralizing number of sodium or chlorine counterions, without additional salt.

All membranes were constructed using the CHARMM-GUI Martini Maker,[248,251] obtaining bilayers of 250 lipids per leaflet. The second set of simulations included 25 water particles per lipid. The resulting systems contained one peptide positioned parallel to the membrane surface at ~ 10 Å from the headgroup region, surrounded by standard or polarizable water models, depending on the force field.

To manage the large number of systems—hundreds of peptide–membrane combinations—, an automated workflow was developed to assemble, equilibrate and simulate each system under the same conditions. Once a membrane model was generated, it could be reused across simulations of different peptide sequences, significantly simplifying the preparation of the systems.

4.2.2 Analysis of the interaction between peptides and lipid bilayers

Foreseeing a scenario of having to manage a significant number of simulations, it became necessary to define a standardized set of analyses that describe the most relevant aspects of the simulation results. The detailed explanation of each of the analyses and transformations applied is provided in Section 3.3.

The trajectory needed to be pre-processed before several quantitative descriptors were calculated. Membranes extend along the XY plane and are used as a reference; therefore, their movement in the Z direction must be corrected by an appropriate translation at each step. As a result of the action of the peptide, the membrane can expand or contract in its plane and can also widen and thin down in its normal direction. The area per lipid is a direct measurement of the degree of compactness of the membrane since it is the inverse of the lipid surface concentration, whereas the variation in thickness was estimated by the evolution of the difference in the mean position of the phosphate-associated particles for both leaflets.

Concerning the peptide, the restraints to the secondary structure imposed in CG simulations make it behave almost like a rigid body, so its movement in the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

3-dimensional space can be characterized by six coordinates: three of them are associated with its position and the other three with its orientation. The symmetries in the system enforce a separate study of the movement of the peptide parallel and perpendicular to the plane of the membrane. The latter was recorded together with that of other groups—such as the phosphates—in an analysis of their position along the Z direction. It is in this analysis where the approximation of the peptide toward the membrane can be observed. The movement in the plane of the membrane is not particularly relevant unless the peptide interacts with it. For that purpose, a measurement of the displacement of the peptide was used to characterize its diffusive behavior.

The rotation of the axis of the peptide—by construction, the axis of the helix—in the plane of the membrane should also manifest diffusive behavior. The joint distribution of the displacement of the peptide and the rotation of its axis, projected onto the membrane plane, is a dynamic fingerprint of the peptide, influenced by the interaction with lipids. Thus, the only two angles needed to characterize the relative orientation between the peptide and the membrane are its tilt—the angle it forms with the normal direction to the membrane—and the spin—the rotation of the peptide around the helical axis. The latter was determined as the angle formed between the transversal component of the hydrophobic dipole moment of the peptide and a vector defined perpendicular to its axis and pointing outwards from the membrane surface.

While the restraints on the backbone structure preserve the helix throughout the simulation, there is a residual flexibility allowed. The side chains are also able to move freely. This can affect the spatial distribution of the charges and hydrophobicities. Thus, the time evolution of the electric and hydrophobic dipole moments was also computed. As these are 3-dimensional quantities, the resultants were projected onto the axis of the peptide to extract parallel and perpendicular contributions. The moduli of these contributions indicate how these magnitudes are distributed along the peptide and also if they suffer from any variation during the simulation.

Regarding the interaction between peptides and membranes, the number of contacts is a widely extended metric throughout the literature used to quantify it. Contacts are the number of molecules of a certain species, namely water and lipids, around the reference—the peptide. The number of contacts with lipid heads and lipid tails was computed separately to account for the difference in peptide penetration. It was also considered relevant to distinguish the interaction between lipids and the different types of amino acids and also the individual residues, aiming to provide a more complete picture of the system.

Finally, different density analyses were also implemented. The first is the density profile along the Z axis, which reflects the average distribution of the mass of different molecules in the last microsecond of simulation. The other is a density map for the

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

lipids surrounding the peptide; while this analysis can be applied in all cases, it makes sense only when the peptide is adsorbed into the membrane. To obtain it, the system was previously rotated so that the axis of the peptide was aligned with the same direction throughout the trajectory. With this analysis, the regions where each specific lipid species tends to concentrate can be observed.

4.2.3 The SuPepMem Database

To enable large-scale comparative analyses of the outcomes of simulations, a dedicated database named SuPepMem was implemented. For its initial development, the database was populated with simulations performed by previous members of the research group. Such simulations are those the original publication refers to, and spanned a more limited space of sequences and used only two force fields, but covering six different membrane compositions.

All necessary simulation inputs, including topologies, coordinate files, and detailed metadata—force field, peptide sequence, membrane composition, ion content, trajectory length, etc.—, are stored within a relational SQL database. Each trajectory is linked to the set of analysis outputs automatically computed by the pipeline: positional and orientational metrics, contacts, structural properties of the membrane—area per lipid and thickness—, density distributions, dipole moments, and mobility fingerprints. These descriptors are preserved in both formatted files—JSON and CSV—and graphs, enabling access and reuse. The input files necessary to fully reproduce any simulation are made available to users, even if the entire trajectory is not publicly downloadable. A graphical user interface (GUI) was developed so that any user can access the simulation data through the site `https://supepmem.com` (Figure 4.23).[310]

Search tools There are two SQL-based search methods that allow the user to access the data. A simple search can be performed if the query is related to the constituents of the system—peptides, lipids, ions, water models—, peptide sequence or activity, or the membrane model.

The advanced search allows for a more detailed query: besides the previous features, peptide properties such as the charge or length, simulation parameters like the temperature or pressure, or the results of some analyses can also be used to further filter the simulations of interest.

Regardless of the search method used, the simulations matching the query will be displayed. The user can either access the details of each of the simulations or select a sample of simulations of interest and compare the results of their analysis. For this comparison, the platform generates a table with numerical descriptors from some of the analyses that can be exported in CSV format, together with histograms

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

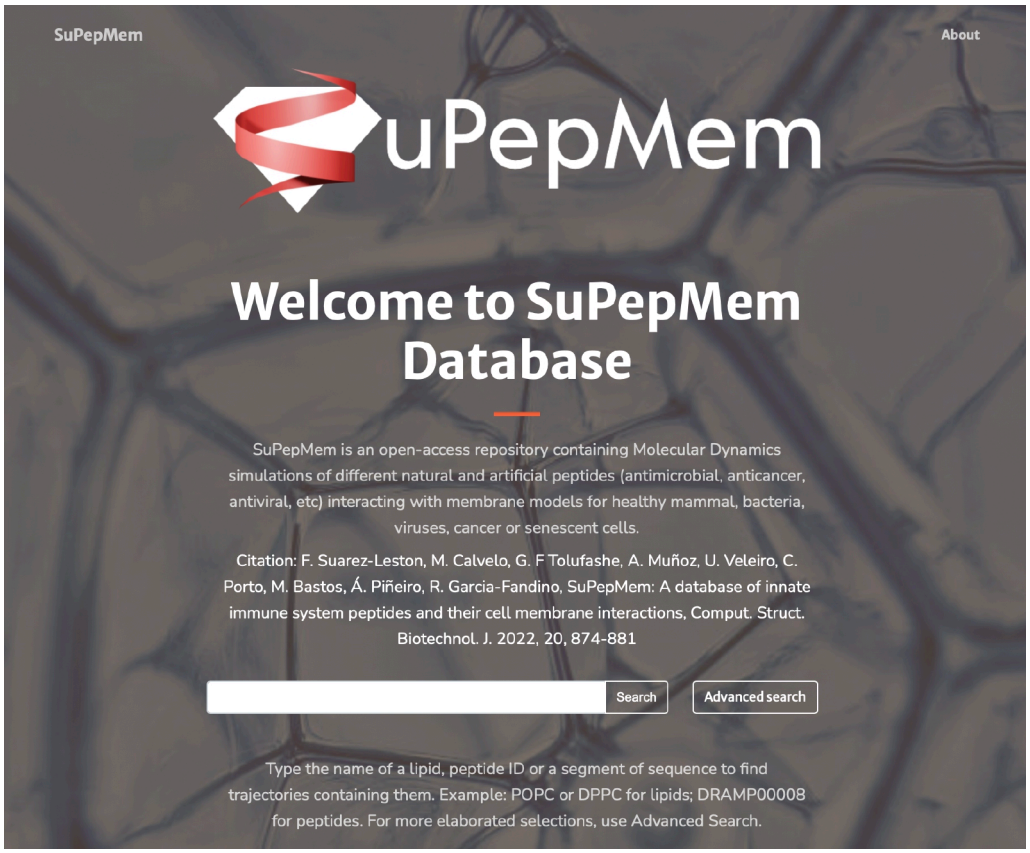


Figure 4.23: Homepage of the SuPepMem Database GUI. Available in `https://supepmem.com`, the GUI allows any user to access the simulation data. Several tools were implemented to facilitate the search, comparison and analysis of the simulations stored in the database.

where the distributions of these parameters can be easily studied.

Detailed simulation display and results access Each simulation stored in SuPepMem has a dedicated page that contains the simulation metadata, input files, structural visualization, and quantitative analysis results. At the top of the page, a header shows key parameters of the system: an ID for the peptide, lipid composition, force field version, trajectory length, simulation parameters, and software used. A link is provided to download the input files—topology, coordinates, parameters, and configuration files—needed to reproduce the simulation.

Below the header, the interface is organized into three main tabs: *Peptide*, *Membrane*, and *Analysis*. The *Peptide* tab (Figure 4.24) offers multiple structural representations: an interactive 3-dimensional view of the CG model, a 2-dimensional schematic of its atomistic counterpart, and a helical-wheel diagram illustrating the distribution of residues. The amino acid sequence of the peptide and key physico-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

chemical properties—such as net charge, hydrophobic moment, and length—are also shown.

The *Membrane* tab (Figure 4.25) presents a 3-dimensional rendering of the bilayer model and detailed composition data for each leaflet, including the number and proportions of each lipid species.

The *Analysis* tab (Figure 4.26) displays the structural and quantitative results of each simulation. The final structure of the system, including peptide, lipids, solvent, and ions is shown. Two residue-colored contact maps show, for each residue, the number of contacts established with the lipid bilayer. This is followed by a series of plots and tables summarizing all derived descriptors—such as peptide insertion depth, tilt and spin angles, peptide displacement fingerprint, contacts by molecule type, density profiles, dipole moments, etc.—, with mean and standard deviation values computed over the final microsecond of the trajectory.

This interface design supports both the browsing of individual systems and the export of comparative analysis between multiple simulations. All numeric descriptors are downloadable in standard formats, enabling external processing.

4.2.4 Conclusions

The implementation of a standardized workflow for the preparation, simulation, and automated analysis of peptide–lipid bilayer systems has enabled the creation of SuPepMem, a database containing MD simulation results for the study of membrane-targeting AMPs.

SuPepMem stores data from CG simulations in which a single peptide unit interacts with different membrane models. The design of both the code for the preparation of the simulations, and the database structure anticipates future extensions: incorporation of simulations executed at other resolutions, in different lipid environments, and involving multiple peptide units. In this sense, SuPepMem is conceived as a living resource, subject to continuous growth and refinement as new simulation data are produced or contributed by external users.

Current simulations have been performed using several Martini force fields. Across all available systems, consistent protocols for the construction of peptides, the assembly of the system, equilibration, and simulation, together with the automated determination of numerical descriptors, ensure uniformity, which is essential for reliable large-scale analysis of the results. The database stores both simulation metadata and input files required for reproduction, along with precomputed descriptor sets in standard formats, facilitating reuse, traceability, and integration into computational workflows. Users can browse, filter, compare, and download descriptor tables and system inputs through the GUI.

Some limitations remain. First, the current simulation data set does not explic-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

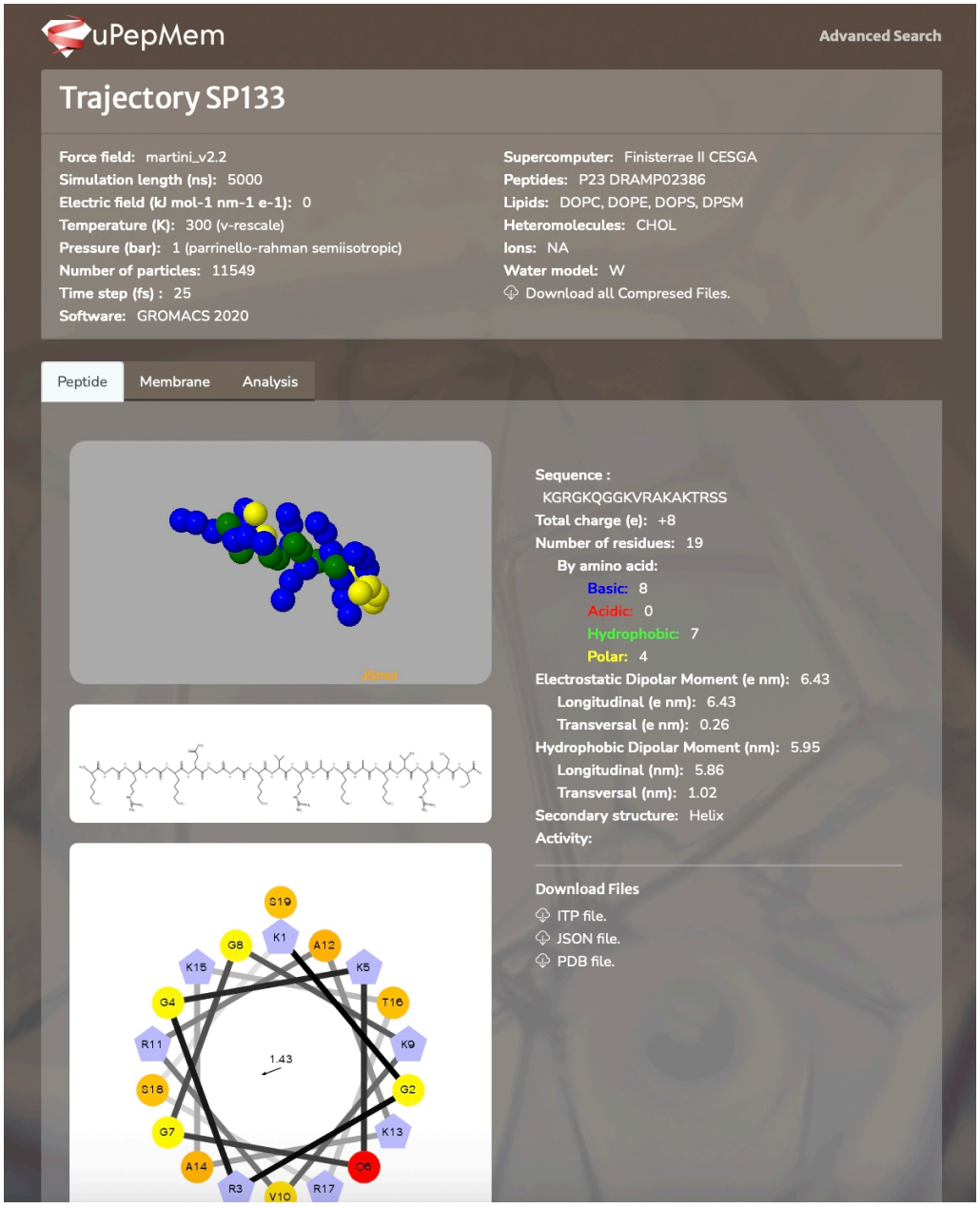


Figure 4.24: Dedicated page to a simulation in the SuPepMem database. Peptide tab. The header contains simulation metadata. Below, different resolution structures and a helical-wheel diagram, together with relevant physicochemical properties provide information about the peptide used in the simulation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

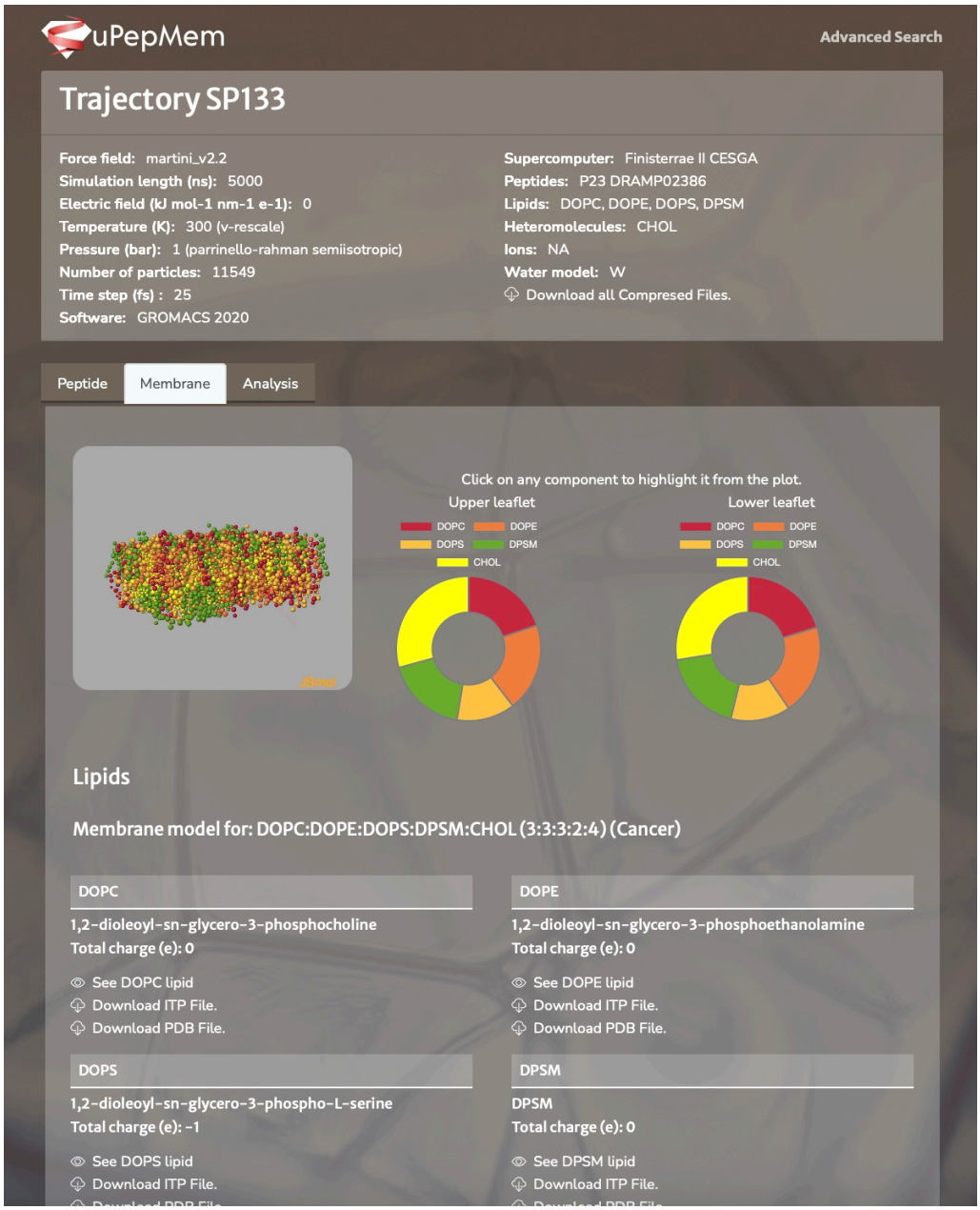


Figure 4.25: Dedicated page to a simulation in the SuPepMem database. Membrane tab. The header contains simulation metadata. Below, the structure and composition of the lipids of the membrane used in the simulation are detailed.

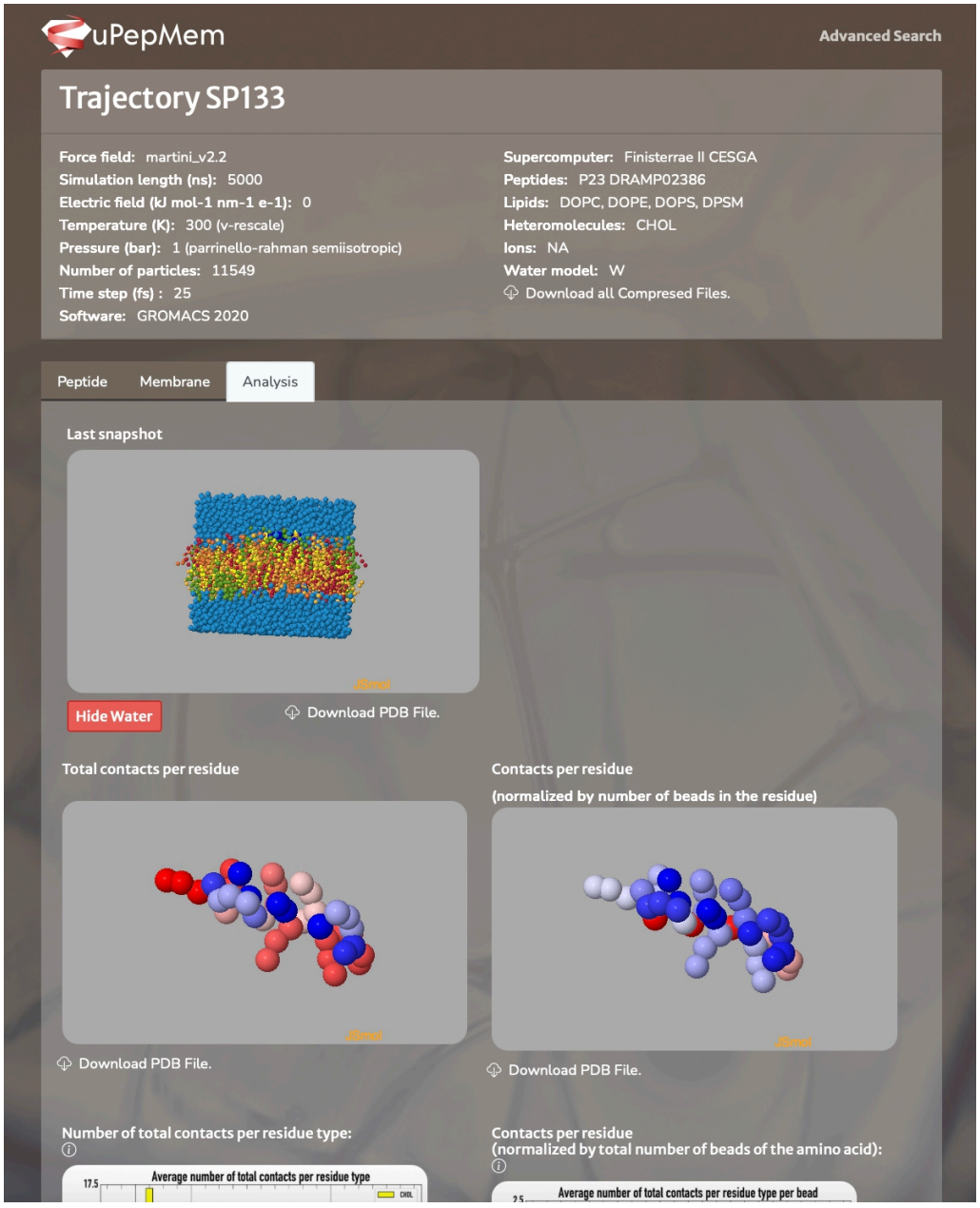


Figure 4.26: Dedicated page to a simulation in the SuPepMem database. Analysis tab. The header contains simulation metadata. Below, the structure of the final frame of the simulation and the graphs of the analyses performed are complemented by a set of numerical descriptors (not shown).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

itly model peptide folding–unfolding transitions, since all peptides were restrained to an α-helical state. Second, only single-peptide units were simulated; cooperative effects arising in multi-peptide systems could not be captured. Third, environmental variables such as ionic strength, salt concentration, and pH are not widely sampled. Fourth, the CG resolution inherently omits certain atomistic details—including precise hydrogen-bonding networks and explicit water structuring around residues. These limitations should be considered when interpreting simulation outcomes and deriving general insights from the data.

SuPepMem offers a basis for exploring correlations between peptide physicochemical attributes—e.g. net charge, hydrophobic moment, amphipathicity, residue composition—and membrane interaction—adsorption tendency, insertion depth, contact frequency, mobility, alteration of the bilayer structure, etc. The generated data set of numerical descriptors supports a statistical analysis and allows its use for data-based approaches aimed at predicting antimicrobial activity or membrane affinity prior to experimental testing.

4.3 Connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches

The SuPepMem repository, introduced in the previous section, contains a systematically generated set of CG MD simulations covering multiple peptides, bilayer compositions, and force fields. From these trajectories, a selection of quantitative descriptors was obtained, including the peptide insertion depth and its orientation within the membrane, the number of contacts with lipid headgroups and tails and water, the translational and rotational diffusion coefficients, and membrane-perturbation metrics such as the change in the area per lipid and bilayer thickness.

Although membrane-lytic activity is typically the result of the cooperative action of multiple peptide units, the present analysis examines whether the descriptors derived from the simulations of a single peptide already capture signals associated with the antimicrobial potential of the sequences.

This section is structured around two goals: (**i**) to identify descriptors that show statistically significant differences between the simulation of sequences labeled as active and nonactive, and (**ii**) to exploit these differences to enable a binary classification of the activity of peptides—active v. nonactive—using simple models.

The following discussion details the descriptors used, the pre-processing criteria, the statistical analyses performed, and the results of the predictive models trained.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.3.1 Selection and statistical analysis of the MD-derived descriptors

From the CG simulations available in SuPepMem, a subset was selected for further analysis. This corresponds to the simulations performed for the second stage of the development of the database, the details of which are given in Section 3.2.3. Briefly, they comprise a total of 123 peptide sequences—73 active, 50 nonactive—modeled as α-helices, two different lipid compositions—POPC and POPE:POPG 3:1—, and three force fields—Martini 2.2 (M2), Martini 2.2P (M2P) and Martini 3 (M3). All of these simulations were performed under a consistent set of simulation parameters and system assembly and analysis protocols; this allows for a direct comparison between their results.

From the list of analyses performed, a series of numerical descriptors were extracted. This includes both the mean value and the standard deviation of the sample in the last microsecond of simulation, with a separation between frames of 2.5 ns, of the following analyses: area per lipid and the thickness of the lipid bilayer, peptide penetration depth, tilt and spin around the helical axes, total, longitudinal, and transversal components of the electric and hydrophobic dipole moments (EDM and HDM, respectively), and the number of contacts with the lipid heads, with the lipid tails, and with water. The values of the translational and rotational diffusion coefficients of the peptide, together with their standard errors, were also included as descriptors. The details of the calculation of each of these magnitudes are given in Section 3.3.

The focus of the comparative study is on how the peptide interacts and affects the lipid bilayer, so some of the previous descriptors were transformed as follows: The area per lipid and the bilayer thickness values were expressed as differences relative to peptide-free membrane simulations (Table D.1).* The number of contacts—with either lipid heads, lipid tails, or water—were normalized by peptide length. For both EDM and EDM, the change in each component relative to the length of the peptide was used; the value at the start of the simulation was subtracted from the instantaneous values obtained, and divided by the length of the sequence. These ensures that descriptors reflects the changes induced or experienced by the peptide rather than the dominant reference structure.

A total of 32 quantitative descriptors were extracted from each simulation, which were then organized into 12 data sets, defined by the combination of (**i**) nominal antimicrobial activity—active or nonactive—, (**ii**) bilayer composition—POPC or POPE:POPG—, and (**iii**) force field—M2, M2P, or M3. The samples of the MD

*Throughout this section they are going to be referred to indistinctly by the name of the property or by highlighting that it refers to a change with respect to the reference value, e.g. the *area per lipid* or the *change in the area per lipid*.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

descriptors obtained for each of these data sets are represented in Figure D.1.

Normality As already mentioned in Section 3.4, knowledge of the parent distribution behind the data determines the statistical hypothesis tests that can be applied on the samples. Consequently, the first step consisted in evaluated the extend to which the samples of each MD descriptors—for each of the 12 data sets of simulations—are compatible with a normally distributed population.

Figure 4.27 shows the Q–Q plots obtained by comparing the empirical quantiles of each sample with the theoretical ones of a normal distribution. The closer the data points lie to the diagonal, the more consistent the sample is with normality. Although many descriptors seem to be accurately described by such a distribution, e.g. the mean change in the thickness of the bilayer, others are far different. To complement the visual inspection, the Shapiro–Wilk (SW) test was applied to each sample. The summarized results are reported in Tables D.5 and D.6. As the null hypothesis of the SW test is that the population behind the data is normally distributed, a significant result of the test is indicative of this condition not being fulfilled. Only three descriptors did not yield significant results in any of the tested groups: the mean value and standard deviation of the thickness of the bilayer and the standard deviation of the area per lipid. For all remaining descriptors, at least one of the samples has been found to have a population that is likely non-normal.

The mean change in area per lipid is not compatible with a normal distribution only in two cases, both corresponding to a POPC membrane: M2P simulations of active sequences and M3 simulations of active and nonactive sequences ($0.01 < p < 0.05$ for all of them). Although in the Q–Q plot the data appear to be slightly deviated from the diagonal, they could still be considered to follow a normal distribution. In all other cases, the interpretation of the Q–Q plot mostly agrees with the results of the test: significant results of the test are associated with cases where the distribution of the data is clearly non-normal.

As most of the descriptors are not compatible with a normal parent population, a conservative approach was adopted. Subsequent statistical analyses were performed without assuming normality for any of the samples. Therefore, nonparametric tests were used to study the properties and differences between samples.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

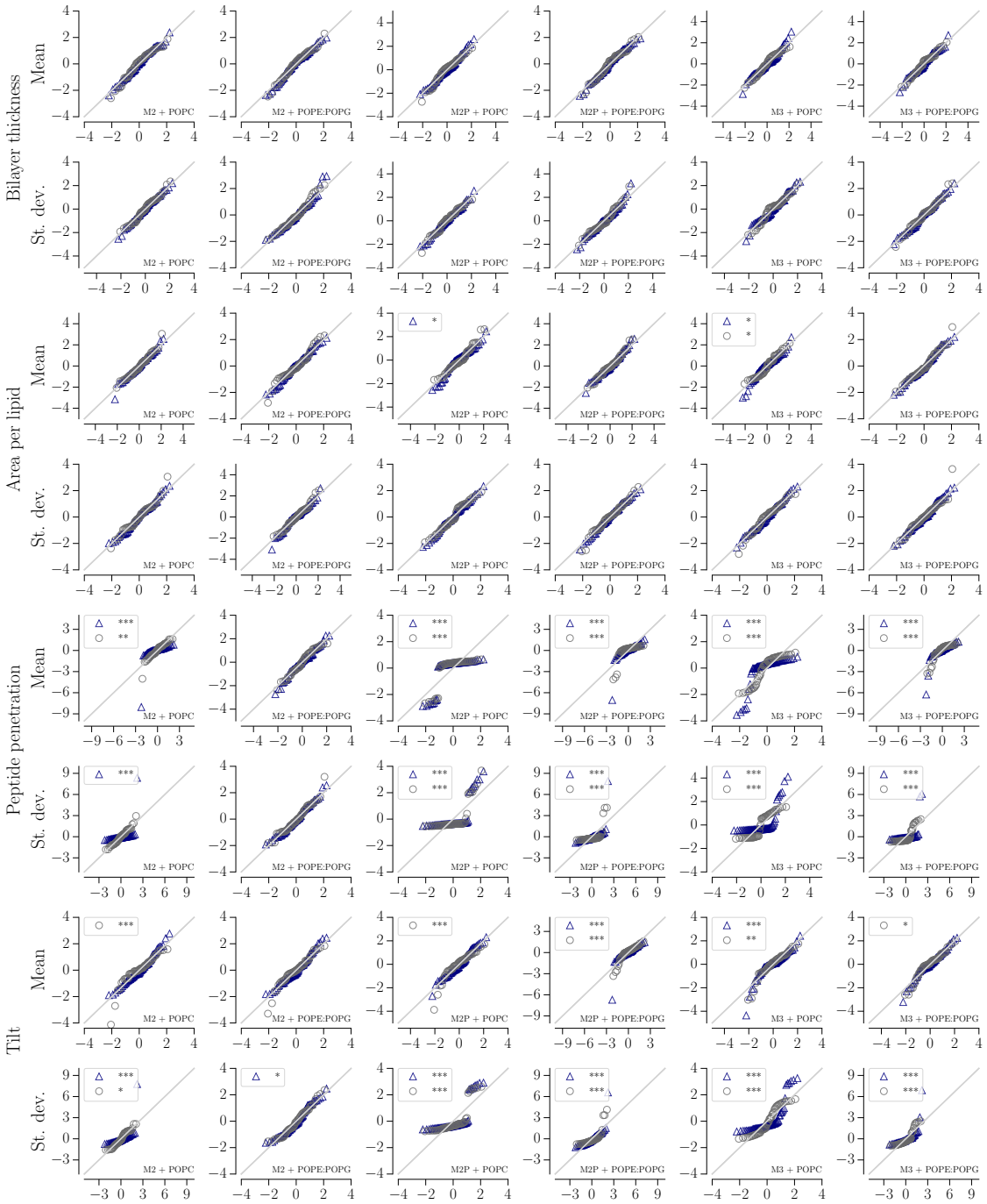


Figure 4.27: Q–Q plot of the MD descriptors. Continues in the next page.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

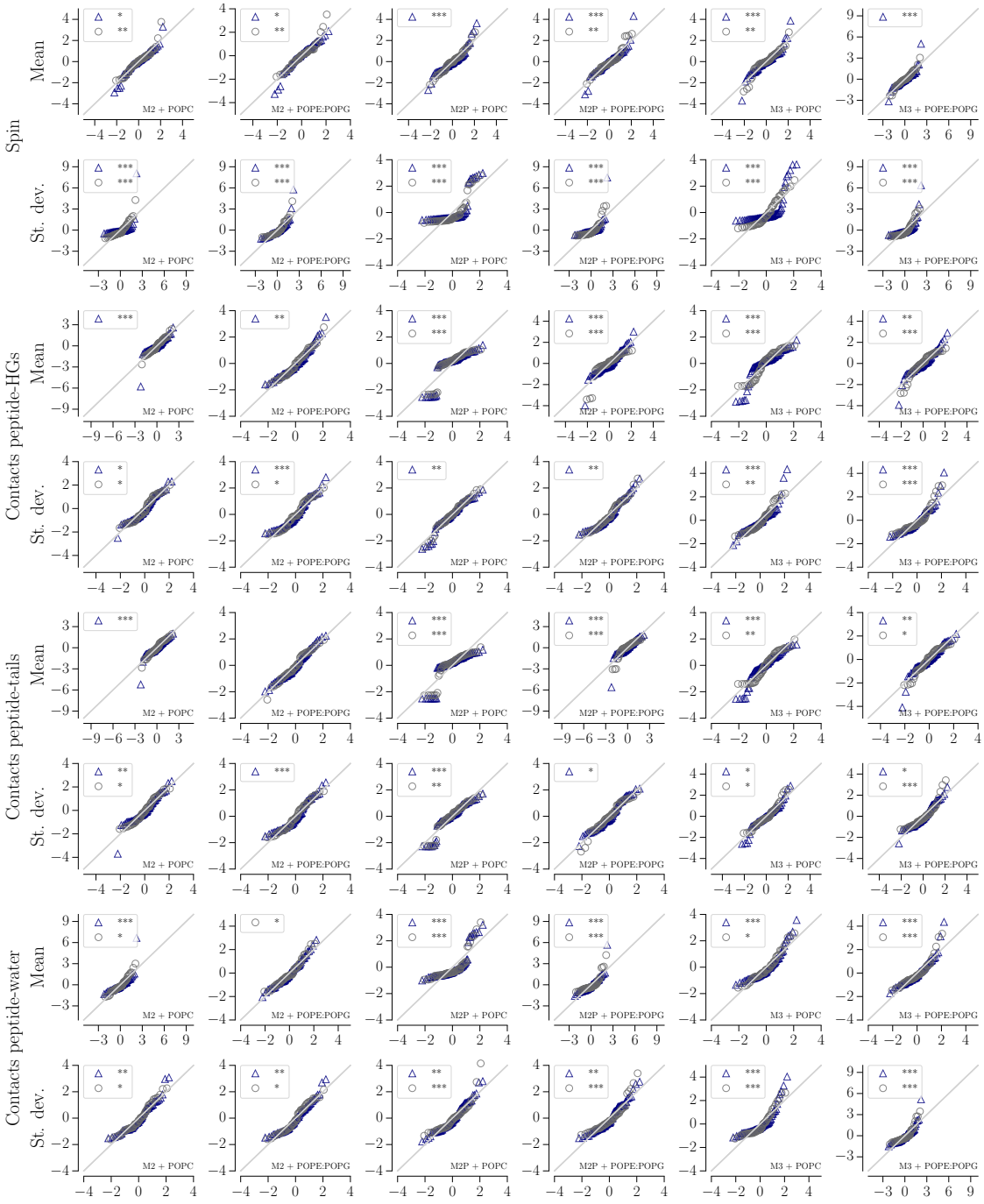


Figure 4.27: Q–Q plot of the MD descriptors. Continued from previous pages, continues in the next page.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

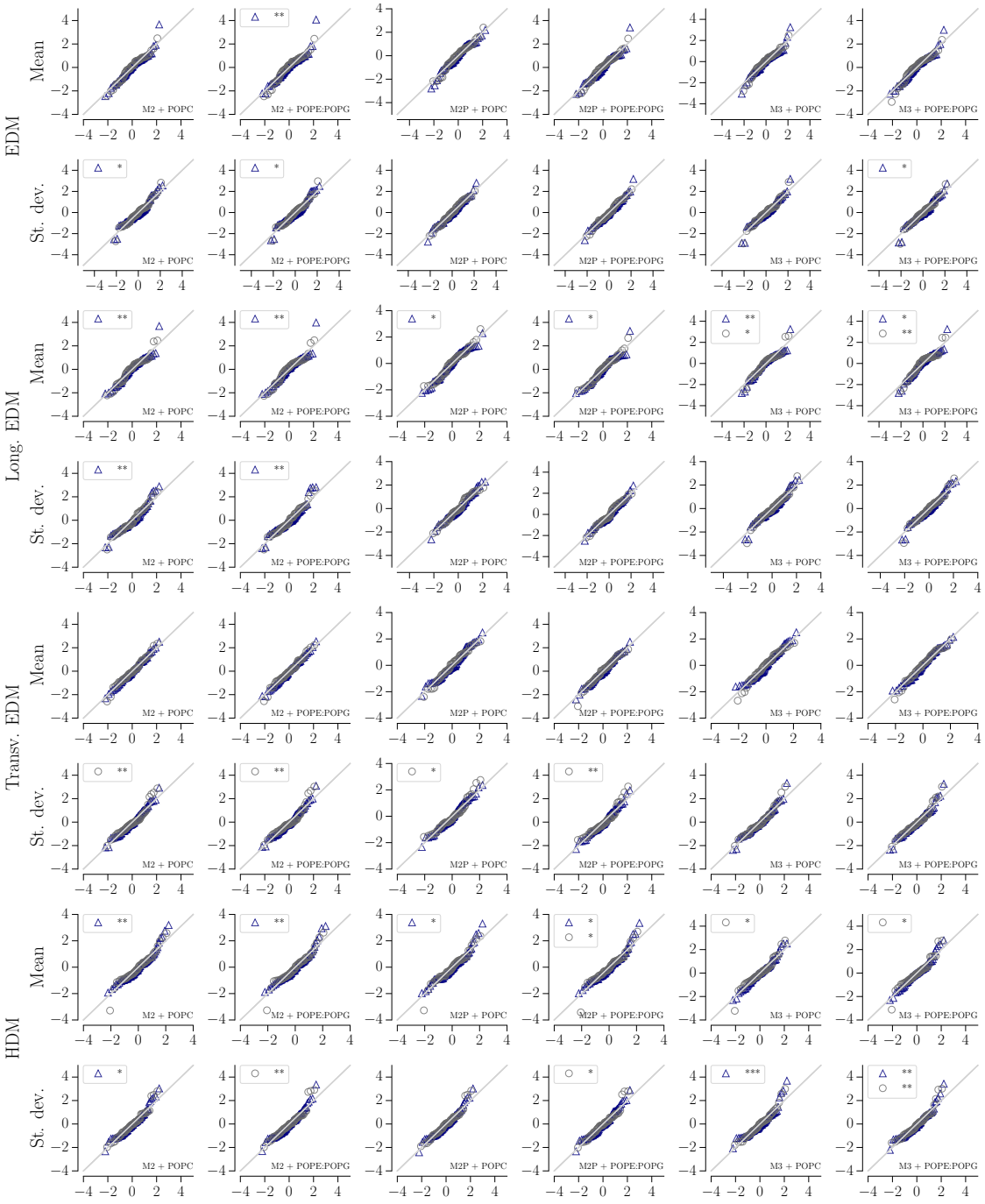


Figure 4.27: Q–Q plot of the MD descriptors. Continued from previous pages, continues in the next page.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

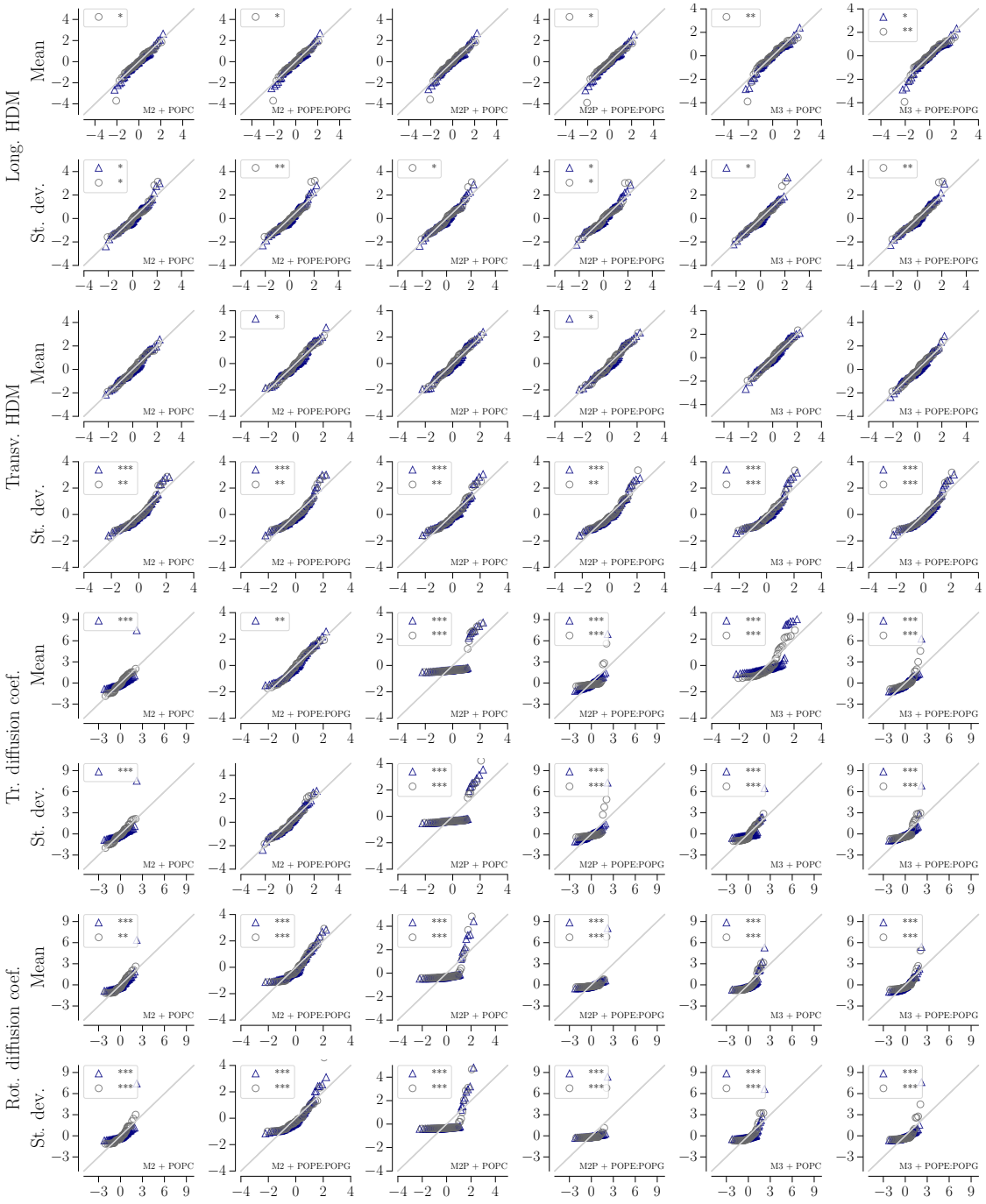


Figure 4.27: Q–Q plot of the MD descriptors. Continued from previous pages. Panels are grouped by MD descriptors (rows) and lipid composition and force field (columns). Theoretical (X-axis) and empirical (Y-axis) quantiles are represented for the samples of active (blue triangles) and nonactive (gray circles) sequences. The gray diagonal line corresponds to a perfect agreement between the data-derived and theoretical quantiles. The legend indicates a significant result in the Shapiro-Wilk test: *** if $p < 0.001$, ** if $p < 0.01$, and * if $p < 0.05$.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Effect of the nominal antimicrobial activity The selected MD descriptors were designed to quantify the interaction between the peptide and the lipid bilayer. Therefore, it will be assumed that any systematic difference in these descriptors between active and nonactive sequences constitutes strong evidence of the sensitivity of the simulation to antimicrobial activity.

A total of 73 active and 50 nonactive peptides were simulated under six different conditions: in the presence of membranes of two distinct lipid compositions and with three different force fields. To investigate the differences in descriptors between sets of simulations of active and nonactive sequences, performed under the same force field and interacting with the same lipid bilayer, the Mann-Whitney U test (MWU) was applied to each case.

As detailed in Section 3.4, a two-sided MWU test was first performed to assess whether the two samples differ significantly. When a significant result is obtained, one-sided MWU tests were applied to determine the direction of the difference. In addition, Brown-Forsythe (BF) tests were applied to evaluate the equality of variances. When the BF test indicated comparable variances and the data distributions displayed similar shapes, differences detected by MWU were interpreted as shifts in central tendency rather than changes in distribution shape.

The results of the MWU and BF tests are summarized in Tables D.7 and D.8, respectively. There is a substantial diversity of results among descriptors (Figure 4.28): while some showed clear differences based on the activity associated with the sample under all conditions, others did so only for some cases. There are also MD descriptors for which no significant differences could be identified in any case.

Among the most consistent results, both the mean change in area per lipid and also the penetration depth of the peptide into the membrane were found to be higher for active peptides. This means that active peptides not only went deeper into the membrane but also occupied a larger volume, displacing the lipids laterally more than in the case of nonactive sequences. However, the magnitude of this effect is very small: the change in area per lipid with respect to the membrane alone is on the order of tenths of $Å^2$—the area per lipid of a phospholipid in the bilayer is on the order of tens of $Å^2$—, but the difference due to the activity can be even lower. The penetration into the membrane is harder to compare, as peptides sometimes are not embedded in the lipid bilayer.

Consistent with deeper insertion, contacts with lipid tails increased and contacts with water decreased for active peptides, which is reflected in the results of the MWU tests. The number of contacts with the lipid heads did not change significantly between active and nonactive sequences, suggesting that they were all close to the surface of the membrane. M3 simulations constitute an exception, as this quantity was observed to be higher for active peptides (Figure 4.28, right); KDEs for nonactive peptide show broader distributions—particularly for POPC simulations—

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

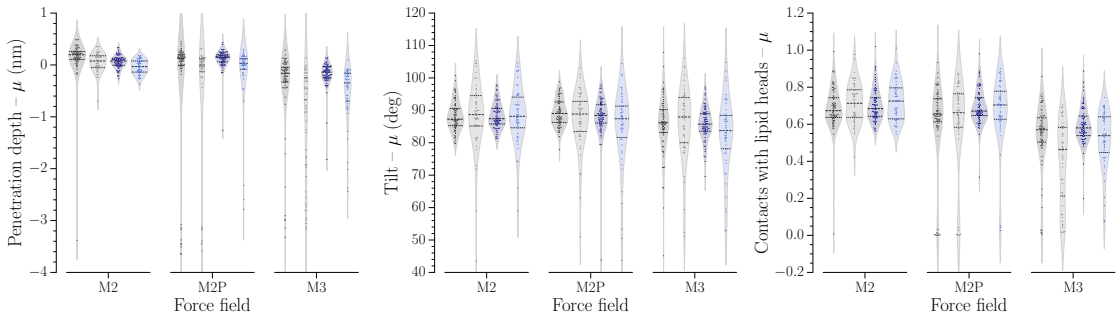


Figure 4.28: Samples of MD descriptors. (Left) Mean values of the penetration depth of peptides into the lipid bilayers, (center) mean values of the peptide tilt angle, and (right) mean number of contacts between peptides and the lipid heads. Within each panel, samples are grouped by force field. Gray symbols are associated with simulations of the POPC membrane, and blue symbols with those of the POPE:POPG 3:1. The brightness signifies the nominal antimicrobial activity of the sequences with darker shades for active sequences and lighter shades for nonactive. Points represent the values of the descriptor obtained for individual simulations, while "violins" represent the kernel density estimate (KDE) of the samples, where wider regions indicate a higher probability density. The median and quartiles are marked with dashed lines in the distributions. The same conventions apply to similar plots. The Mann-Whitney U tests indicate that active peptides penetrate deeper into the membrane in all scenarios. In contrast, no significant differences have been found between the samples of the tilt. The difference in the number of contacts with lipid heads is only statistically relevant for M3 simulations.

, which also coincides with a broader distribution of the standard deviation of the tilt toward higher values (Figure 4.29, left and center panels). Both of these observations are indicative of simulations where the peptide underwent several processes of adsorption into the membrane and release to the bulk (Figure 4.29, right panel), so the average number of contacts became smaller and its fluctuation, much larger: When the peptide is in the bulk, it has more freedom to rotate, and while the average tilt is also around 90°,* its fluctuation is much higher than when it is bound to the membrane, where the out-of-plane movement is far more restricted.

The diffusion coefficients are consistently lower for active peptides. This parameter measures how easy it is for a body to move freely: the higher the diffusion coefficient, the further distance it reaches, on average, in a certain time interval. The diffusion coefficient of a tightly bound peptide—whose movement is more hindered—would be lower than that of a free-to-move one. Therefore, this result suggests that the interaction with the surrounding lipids limits the lateral displacement and rotation of the axes of active peptides in the plane of the membrane.

*For a peptide in an homogeneous and isotropic medium, the expected mean value of the angle with respect to any plane would be 90°, as its PDF is an uniform distribution.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

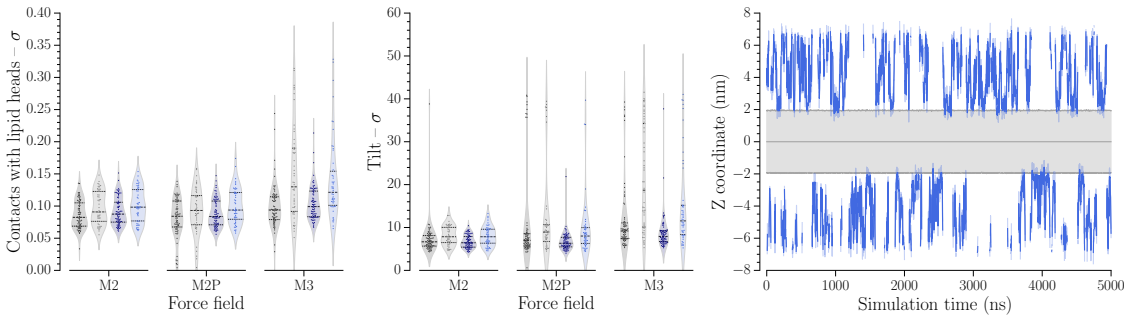


Figure 4.29: Fluctuation in peptide–membrane interactions and orientation. (Left) Distribution of the standard deviation of the number of contacts between each peptide and the lipid tails, and (center) distribution of the fluctuation of in the peptide tilt angle. The same color scheme as in Figure 4.28 is used: gray for POPC and blue for POPE:POPG bilayers, with darker shades for active peptides and lighter ones for nonactive sequences. Each "violin" represents the kernel density estimate of the data, where wider regions indicate a higher probability density. Broader KDEs for the nonactive sequences reflect larger structural fluctuations, consistent with transient peptide–membrane interactions. (Right) Evolution over the simulation of the Z coordinate of the centroid of the NC02424 peptide (dark blue line) in the presence of a POPC membrane (gray-shaded area), relative to the center of the latter. The positions of the N- and C- termini are represented in a lighter blue. The jumps due to PBCs are shown as discontinuities in the curves.

No significant differences were found in the angular descriptors associated with peptide orientation—tilt and spin—as a function of the activity, except for the M2P force field in the POPC bilayers. Here, the mean value of the spin angle is higher for nonactive sequences (Table D.2): the median value of the distribution for active sequences is 175.34°, while for nonactive sequences it is 156.53°. In general, the median of the distribution of the mean spin angle for active peptides in all samples is closer 180°, indicative of an alignment of the transversal HDM towards the nonpolar center of the bilayer.

While the results of the tests showed significant differences between the samples for many descriptors in all three force fields, it is in M3 simulations, followed by M2P, where the most number of them was recorded. However, in order for an MD descriptor to serve as a reliable predictor of antimicrobial activity, the differences between classes should not only be statistically significant, but also large enough to be discriminative. To evaluate how close the samples are to each other, their overlap was computed according to Eq. 3.118. This magnitude ranges from 0 to 1; the more similar the samples are to each other, the higher the overlap. Therefore, a low—ideally zero—overlap between the samples of any descriptor would mean that they are easy to discriminate, and the descriptor is a potential identifier of the activity of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

a sequence, at least based on the data available. The estimated overlap values were already included in Table D.7. Except for some cases, it is consistently above 0.5. This means that most of the area under each of the KDEs coincides with the other, therefore making the identification of the activity based on them not possible. Some exceptions are the samples of rotational diffusion coefficients—and their standard error—, whose overlap is smaller. These distributions are characterized by abundant outliers, mostly in M3 simulations, whose origin will be discussed later.

Therefore, the sole comparison of MD descriptors, while giving rise to several statistically significant and noticeable differences, is insufficient to make an educated guess about the antimicrobial activity of a sequence.

Effect of the lipid composition As discussed in the Introduction, membrane-targeting AMPs are able to selectively disrupt pathogenic and pathological membranes while leaving healthy eukaryotic membranes largely unaffected. The two lipid compositions considered in this study represent two very simple models of different types of cells: PC is the dominant phospholipid in the outer leaflet of mammalian cells,[2,265] while PE and PG are the most abundant in the lipid bilayer of the *E. Coli* bacteria.[33] While the lipid composition across membranes is very diverse, pathogens display more negatively charged lipids on their external surface. Consequently, the preferential interaction of peptides with one or the other bilayer should be reflected in MD descriptors.

Two simulations of the same peptide and under the same force field with different lipid composition can be assumed to be paired. In this context, the Wilcoxon signed-rank (WSR) test is appropriate to identify whether the differences in the MD descriptors within all pairs are skewed toward positive or negative values, which would mean that the value of that descriptor is higher or lower when the simulation involves one of the lipid compositions. As in previous analyses, two-sided tests were first performed to detect overall differences, followed by one-sided tests to identify the direction of the shift, if found. The differences will always be computed by subtracting the value of the descriptor in the POPE:POPG system to that in the POPC one,

$$\Delta x = x|_{\text{POPC}} - x|_{\text{POPE:POPG}}. \tag{4.1}$$

Among the results of the WSR tests, reported in Table D.9, those concerning changes in the area per lipid and thickness experienced by the lipid bilayer are the most striking: In both M2 and M3 simulations, the perturbations induced by the peptide are larger* in POPC than in POPE:POPG, regardless of the nominal

*Here the sign of the difference in the table may be confusing. A reduction in thickness when the peptide interacts with the membrane will have a negative sign. If the reduction of the thickness in POPC is larger—in absolute value—than in POPE:POPG, the value obtained after applying Eq. 4.1 will be negative.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

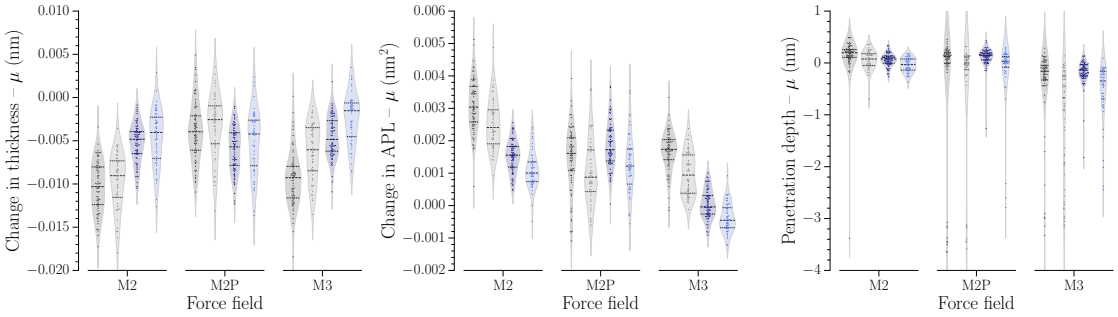


Figure 4.30: Representative MD descriptors related to membrane perturbation. From left to right, violin plots show the mean variation of the mmebrane thickness, of the area per lipid, and the mean peptide penetration depth. The same color scheme as in Figure 4.28 is used: gray for POPC and blue for POPE:POPG simulations; darker shades for active peptides and lighter for nonactives. Wider regions in each violin indicate a higher probability density (KDE). According to the Wilcoxon signed-rank test, in M2P simulations peptides exert a larger structural perturbation on the POPE:POPG bilayer than on POPC, whereas the opposite trend is observed for M2 and M3. In M2 simulations, peptides penetrate deeper into the POPC membrane than into POPE:POPG. These effects are independent of peptide activity.

antimicrobial activity. The overlap of the samples is also much lower in M2 and M3 simulations than in M2P (Figure 4.30, left and center).

The depth at which the peptide penetrates is conditioned by the force field used in the simulation (Figure 4.30, right): In M2, the peptides penetrate deeper in the POPC bilayer than in the POPE:POPG one, while in M2P and M3 it is the other way around. A tendency of the M2 force field to overbind peptides to membranes has already been reported in the literature.[145,312] Both active and nonactive peptides display mostly a net positive charge, which favors their interaction with negatively charged beads in lipids, i.e. the phosphate groups—present in all three types of phospholipids—that mark the reference position for the calculation of penetration. The zwitterionic character of POPC and POPE molecules comes from a positive head group bead. The differences between force fields must therefore arise for the hydrophobic interaction between the molecules.[313–315]

The standard deviation of the area per lipid, the penetration of the peptide, and the thickness of the bilayer are higher for POPC simulations in all cases—except for the latter in M3—, indicating that the peptide has more freedom to move in this membrane compared to POPE:POPG. The larger diffusion coefficients, both translational and rotational, in POPC membranes also point in that direction. Rotational diffusion coefficients must be interpreted with care, and it is important to note how they were computed: both the movement of the centroid of the peptide and the rotation of its axes are projected into a 2-dimensional plane—the XY plane—,

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

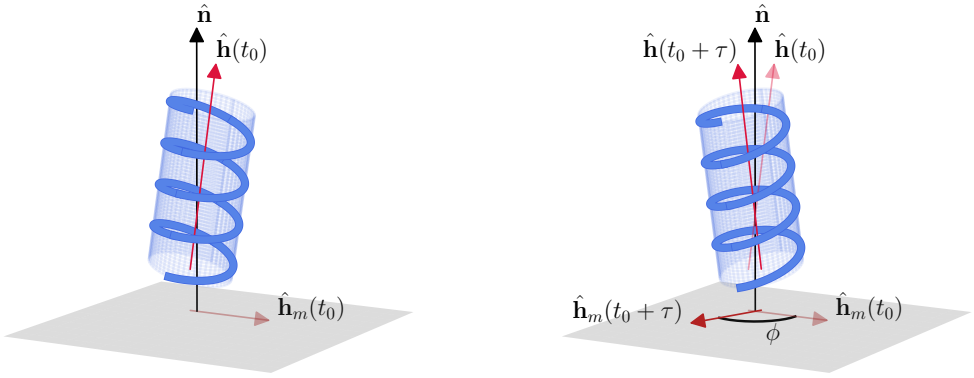


Figure 4.31: Effect of peptide tilt on the apparent in-plane rotation of its axis. (Left) At a initial time $t = t_0$ the peptide is not adsorbed into the membrane, and its principal axis $\widehat{\mathbf{h}}$ is almost parallel to the normal of the bilayer $\widehat{\mathbf{n}}$, resulting in a large tilt angle. (Right) After a time interval τ, the peptide undergoes only a small change in its overall orientation, but the apparent rotation of its projection onto the membrane plane, ϕ, becomes much larger. The coupling between the tilt and the in-plane rotation distorts the distribution of ϕ angles, making A.48 inapplicable under these conditions.

where the membrane extends. When a peptide is not bound to the membrane, the method for the calculation of the translational diffusion coefficient could still be used, but without any physical justification, as its movement in the bulk should be modeled as a 3-dimensional random walk. The interaction with the lipids—a more viscous medium—drags the movement of the peptide, so the diffusion coefficient in water is higher than that of the peptide when adsorbed into the lipid bilayer. The rotational diffusion coefficient D_r, on the other hand, can only be calculated for a 2-dimensional movement after ensuring that the angular changes are small. Otherwise, the out-of-plane rotation of the peptide can induce artifacts in the distribution. A clear example is a case where the tilt of the peptide is either $\theta \approx 0$ or $\theta \approx \pi$, i.e. the peptide almost perpendicular to the plane where the rotational diffusion is computed: a small variation in its orientation can lead in a huge variation of the rotation of the projection of the axes into the plane, as illustrated in Figure 4.31. As a result, the PDF no longer obeys the model given by Eq. 3.88, so the estimated value of D_r obtained from the least-squares fit of the data is meaningless.

The number of contacts with the lipid tails and water was previously connected with the depth at which the peptide penetrates the membrane. Therefore, as peptides in M2 entered deeper into the POPC bilayer, contacts with the POPC tails

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

increased and those with water decreased. As there are simulations with almost no contacts with POPC head groups, the samples of contacts with head groups are always larger for the POPE:POPG bilayer.

A very relevant aspect of the previous analysis is the sign of the change in the descriptors between the simulations with either lipid composition (Eq. 4.1) not being affected by the activity of the peptide. At most, the results may be not conclusive for one of these groups, but for no descriptor in any force field, the difference between the simulations with POPC and with POPE:POPG is positive for active peptides and negative for nonactive peptides, or vice versa. Yet, it is still possible for the magnitude of the difference to be connected to the activity, e.g. a larger—or smaller—change in the descriptor between POPC and POPE:POPG simulations for active peptides than the same change for nonactive ones.

To test this possibility, the differences between POPC and POPE:POPG descriptors were compared between active and nonactive peptides using the Mann-Whitney U test. Tables D.10 and D.11 summarize the results of the MWU and BF tests, respectively. The number of differences found varies greatly between the force fields: Only a few descriptors differed between active and nonactive peptides in M2 and M2P, whereas most descriptors showed substantial changes in M3. In M2 simulations, differences are limited to the change in the modulus of the EDM and the standard deviation of the transversal component of the HDM. For M2P, significant differences appear only in the spin and in the rotational diffusion coefficient. In contrast, M3 simulations revealed activity-dependent variations between POPC and POPE:POPG bilayers in several descriptors: the area per lipid and the membrane thickness—both showing larger changes for active peptides—, the penetration depth—smaller for nonactives due to weaker adsorption into POPC—, and the number of contacts with both lipids and water.

In summary, the peptides in M3 exhibited a notable differential behavior between the two lipid compositions, primarily due to the number of simulations of nonactive sequences in which the peptides were not adsorbed into the POPC bilayer. All simulations were assembled following the same protocol and in the same conditions—the only difference comes from the dielectric constant, imposed by the force field—, and the simulation timescales were long enough for the adsorption process to occur, suggesting that this difference originates from the reparametrization of the force field rather from any bias in the setup or the protocol.

Effect of the force field The third variable examined in this study was how the force field affected the results of the simulations. While some force-field-related trends were already noted in the previous sections, the present analysis focused on identifying and quantifying systematic differences attributable to the choice of parametrization, as we did previously for the activity and for the lipid composition.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Unlike the previous factors, which involved comparisons between only two classes—active v. nonactive, POPC v. POPE:POPG—, analyzing three force fields requires performing three pairwise tests, one for each combination.

Simulations were paired according to sequence and lipid composition, and the WSR test was used to detect statistically significant differences between each pair of force fields. The results of the tests are summarized in Table D.12.

The interaction with the peptide reduced the thickness of the POPC lipid bilayer, the magnitude of this change being much more important in the M2 simulations. The change in this magnitude for M2P simulations is the lower. The more recent M3 force field gave rise to the largest difference in behavior: while active sequences induced an effect comparable to that observed in M2, the impact of nonactive peptides was much smaller—intermediate between M2 and M2P. The other lipid bilayer, composed of a mixture of POPE and POPG, suffered from a larger thinning in M2P simulations and the lowest in M2. In this case, no differences were observed when comparing the samples of active and nonactive sequences.

The complementary property to thickness is the area per lipid, which typically increased—although not always—as an effect of the interaction with the peptide. M3 simulations are those where minimal variation with respect to the unperturbed membrane was recorded, together with those of M2P and the POPC membrane. Therefore, while the POPC membrane was more expanded in M2 simulations, the POPE:POPG membrane was significantly more stretched in M2P simulations. The standard deviations of both the area per lipid and of the thickness are always higher in M3, followed by M2P and last by M2.

As previously discussed, peptides penetrate deeper into POPC membranes under the M2 force field as a result of an overestimation of nonpolar interactions. In the POPE:POPG membranes, in contrast, the higher depth corresponds to that in the M2P simulations. M3 simulations did always lead to the more shallow interaction with the membrane, even when not taking into account the unadsorbed peptides: while most of the M2 and M2P simulations yielded a positive peptide penetration depth—the peptide below the mean position of the phosphate-associated beads—, in M3 they were mostly negative. This position allows for more flexibility in the movement of the peptide, which is manifested in a higher fluctuation of the depth.

When the peptide is adsorbed and parallel to the surface of the bilayer, or when it is in solution, the expected value of the tilt is 90°—see footnote in Page 146. However, in some simulations—especially in M2 and M3—there is a tendency to yield lower mean values of this angular descriptor. This behavior is a consequence of systems in which the peptide was only partially adsorbed onto the membrane (Figure 4.32): the positively charged N-terminus of the peptide remained in contact with the bilayer surface, while the rest of the peptide extended to the aqueous phase. As a result, the tilt angle oscillated between 0 and 90°, lowering its average value.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

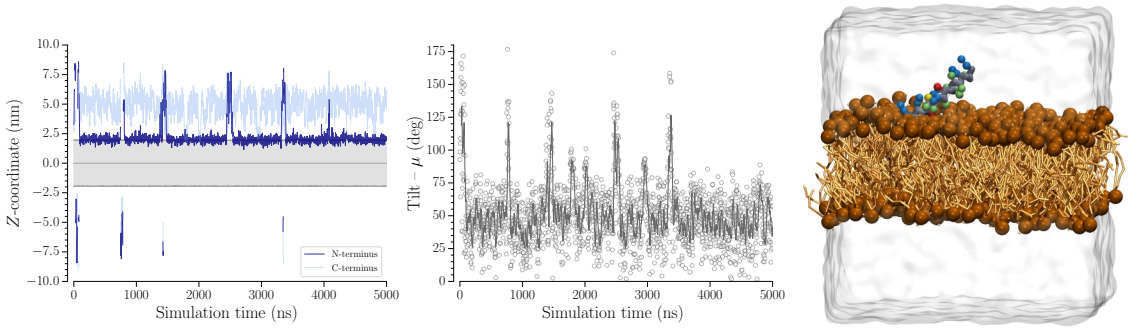


Figure 4.32: Partial adsorption of a peptide into a POPC membrane. Simulation of the peptide AP03406 and a POPC bilayer using the MARTINI 3 force field. (Left) Time evolution of the Z coordinate of the peptide termini. The gray shaded region marks the bilayer interior, defined by the mean position of the phosphate groups in both leaflets. The N-terminus (dark blue) remains adsorbed within the membrane for most of the simulation, whereas the C-terminus (light blue) stays predominantly in the aqueous phase. (Center) Time evolution of the peptide tilt angle. Because the peptide is not parallel to the membrane, the instantaneous tilt values (gray circles) cluster below 90°. The transient release of the N-terminus observed in the left panel corresponds to the spikes in tilt highlighted by the rolling mean of the magnitude, (solid line) computed in a 100-frame window. (Right) Final snapshot of the simulation, showing the peptide partially adsorbed at the membrane interface.

This partial adsorption also contributed to the increased fluctuations observed in both the penetration depth and the number of contacts with the lipid headgroups, the lipid tails, and water; all these quantities were significantly higher in the M2 simulations.

In addition, systems in which the peptide remained unbound to the membrane further decrease the mean number of lipid contacts while increasing contacts with water. Contacts with lipid headgroups were more frequent in M2 simulations, whereas contact with lipid tails follows the same trend as the penetration depth: higher in POPC bilayers for M2, and in POPE:POPG bilayers for M2P. None of these relations are affected by the nominal antimicrobial activity of the sequence.

The reduced penetration of peptides observed in the M3 simulation, together with the larger number of cases in which the peptide remains fully solvated, led to higher translational and rotational diffusion coefficients under this force field. Between M2 and M2P, diffusion coefficients were generally higher in the former, although the translational diffusion across the POPC membrane is comparable for both.

While the use of different force fields had an impact on the value of the MD descriptors, no property exhibited large or consistent changes when comparing the samples of active and nonactive peptides. The most notable exception concerns

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the bilayer thickness, where the effect induced by peptides is similar in M2 and M3, but nonactive peptides caused a stronger thinning in M2. The evolution of the Martini force field from M2 to M3 involved substantial modifications in the treatment of non-bonded interactions. These changes translate into a qualitative shift in peptide–membrane behavior: from the near-universal membrane binding observed in M2, to a much more diverse adsorption landscape in M3, where peptide binding depends sensibly on both sequence and membrane composition.

4.3.2 Identification of sequence-induced bias

The statistical analyses presented in the previous section revealed significant differences in several MD descriptors between simulations of active and nonactive peptides. However, these comparisons did not account for the differences in the structural or compositional properties of the peptide sequences themselves. Because the sequence selection criteria were limited to peptide length and sufficient sequence diversity within each data set, it is possible that some of the differences in MD descriptors previously noted arose from a bias attributable to an uneven sample of some of the peptide properties, rather than from their antimicrobial activity.

To evaluate this possibility, the distributions of a set of sequence-derived properties were analyzed and compared between the two activity-based groups. The properties considered include: the length of the sequence, the net charge of the peptide, the values of the electric and hydrophobic dipole moment—total, and also their longitudinal and transversal components—, and the number of amino acids, both of each individual amino acid and the amount of polar, hydrophobic, acidic, and basic amino acids in the sequence. All quantities were normalized by the sequence length, except for the sequence length itself.

The Q–Q plots of each of these properties, represented in Figure 4.33, revealed that many of them deviate strongly from normality, particularly the relative frequencies of individual amino acids. SW tests (Table D.13) confirmed that most distributions are not compatible with a normal parent population.

In consequence, differences between active and nonactive peptides were assessed using the MWU and BF tests (Tables D.14 and D.15). The results (Figure 4.34) show that active peptides in the selection possess a higher charge than the nonactive ones, both in sign and in magnitude—most peptides are positively charged. However, this does not translate into a well-ordered pattern of charged residues, since no significant differences were observed between the components of the electric dipole moment.

In contrast, the sequence composition differs more markedly: nonactive sequences are enriched in polar residues—Cys, Asn, and Tyr are also more abundant individually than in active sequences—, whereas active peptides contain a significantly higher fraction of hydrophobic residues—the occurrence of Ala, Leu, and

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

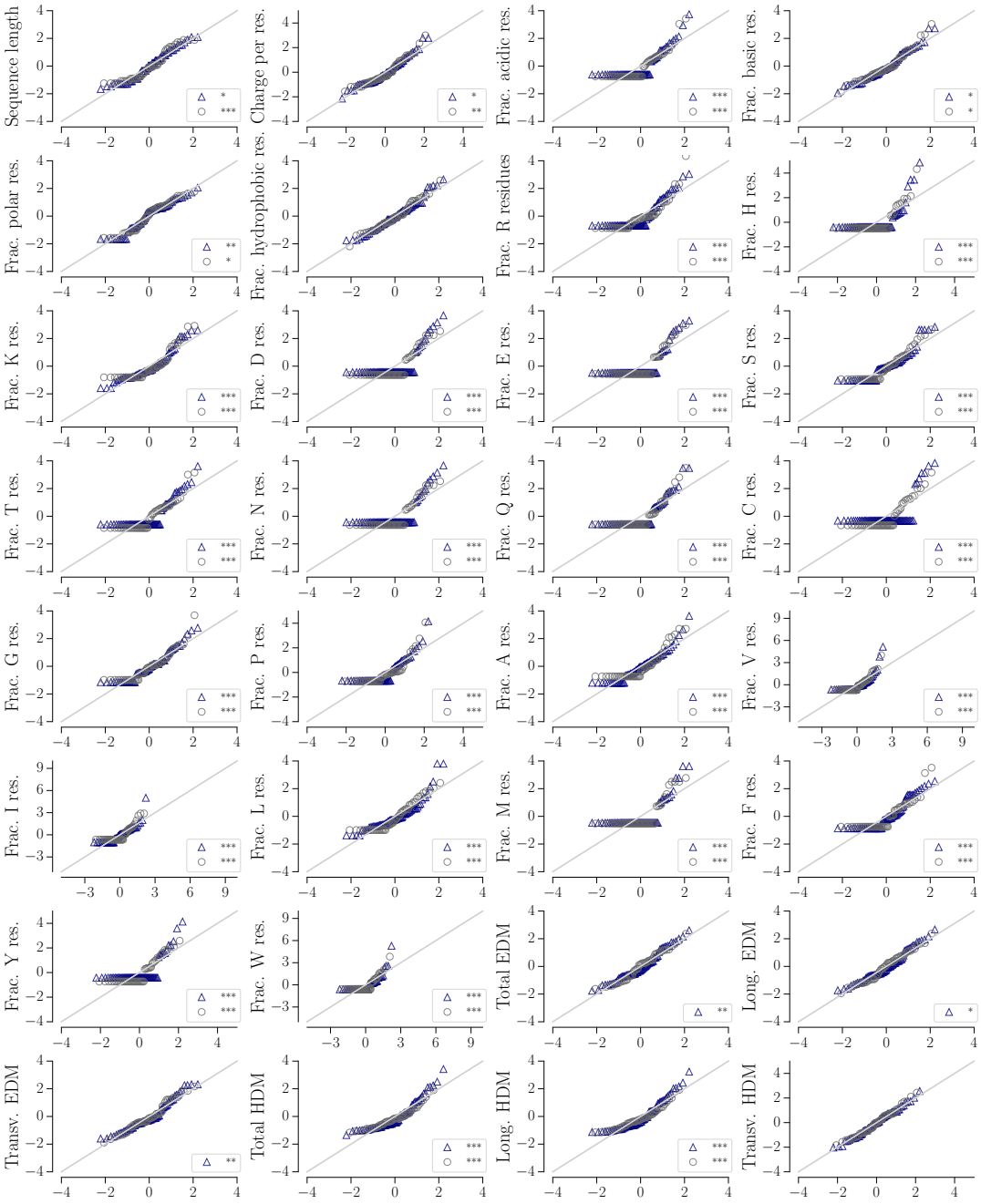


Figure 4.33: Q–Q plots of peptide physicochemical and structural properties. Each panel compares the empirical (Y-axis) and theoretical (X-axis) quantiles of the corresponding property. Blue triangles represent active sequences, whereas gray circles correspond to nonactive ones. The gray diagonal line indicates the case of perfect normality; deviations from this line reveal non-normal distributions for several descriptors. Statistical significance according to the Shapiro-Wilk test is indicated as follows: *** if $p < 0.001$, ** if $p < 0.01$, and * if $p < 0.05$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

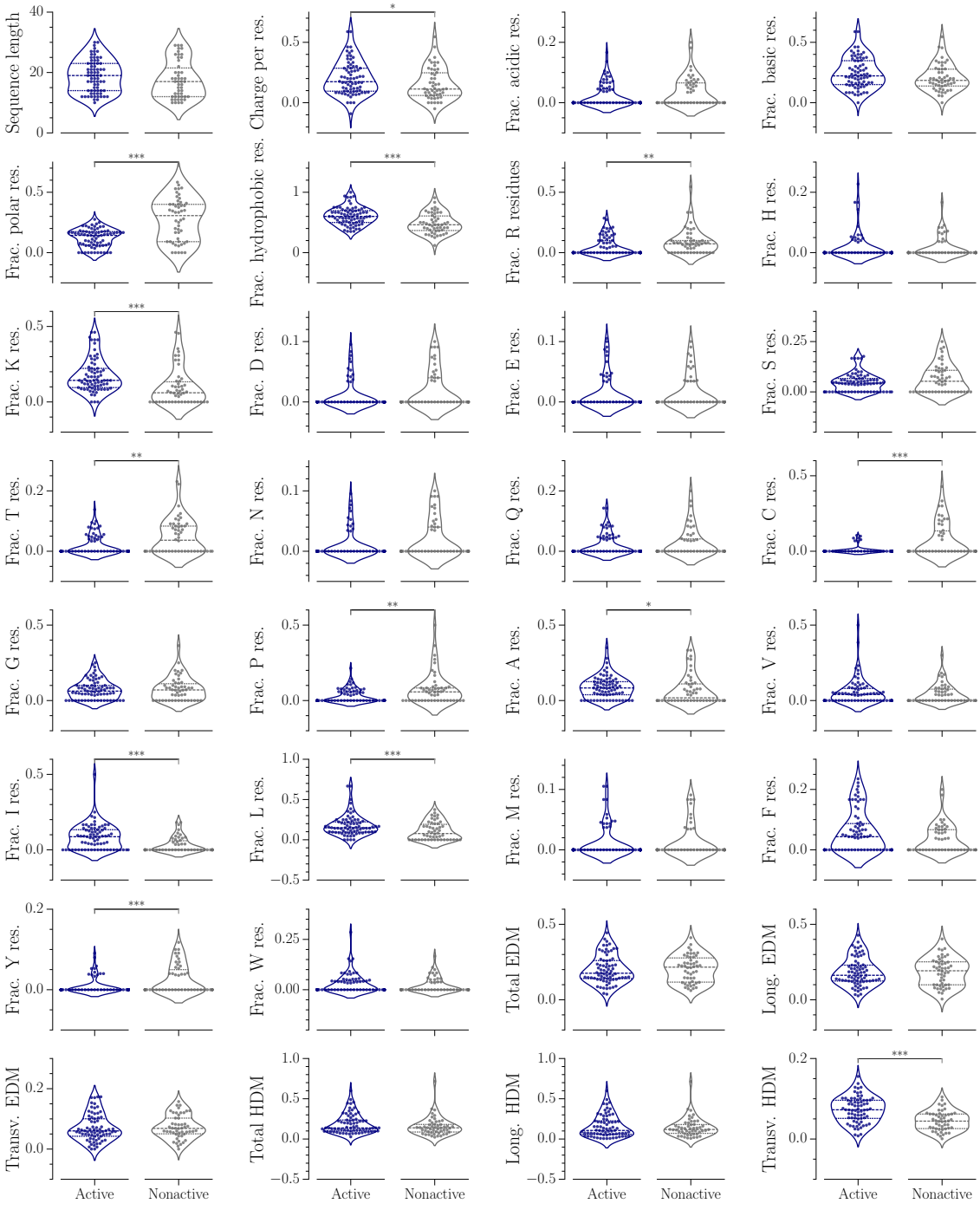


Figure 4.34: Distribution of peptide physicochemical and structural properties. Each violin plot represents the distribution (KDE) of a given property across all peptides, grouped by nominal antimicrobial activity: actives in blue, nonactives in gray. Individual data points correspond to the property values for each peptide, while dashed lines indicate the median and quartiles. Statistical significance according to the Shapiro-Wilk test is indicated as follows: *** if $p < 0.001$, ** if $p < 0.01$, and * if $p < 0.05$.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Ile residues is also higher. The higher number of hydrophobic residues is accompanied by a higher transversal hydrophobic dipole moment, suggesting that they are arranged along the axis of the peptide, forming an amphipathic structure. For all other components of the HDM, no sufficient evidence of a difference has been found. With regard to other amino acids, Lys is more common in AMPs, while Arg and Pro are more common in nonactive sequences. The case of Pro is especially significant because it disrupts the helical structure; in these CG simulations, it has been treated as any other amino acid.

All of the properties mentioned could have potentially biased the simulation results, thereby affecting the comparison between active and nonactive sequences. To assess which descriptors could be directly influenced, the Kendall rank correlation coefficient τ was calculated between each MD descriptor and each structural property previously identified as different between active and nonactive peptides. A high value of $|\tau|$ indicates a strong correlation, suggesting that the structural feature could have influenced, and thus biased, the corresponding MD descriptor. As described in Section 3.4.2, correlations were considered strong when $|\tau| > 0.6$. The analysis was performed separately for each combination of lipid composition and force field, without making a distinction based on activity.

Only two structural properties manifested strong correlations with two MD descriptors: The penetration depth of the peptide into the membrane increased with a higher fraction of hydrophobic residues, and the variation in the longitudinal component of the electric dipole moment grew with the net charge—both properties are higher in active sequences. As shown in Table D.16, $|\tau|$ did not exceed the threshold in all cases, but did so in most cases, and the correlations were consistently stronger for active sequences. Since no significant differences in the change in the longitudinal EDM were previously detected between active and nonactive peptides—regardless of lipid composition or force field (Table D.7)—, there is no evidence that this descriptor was directly influenced by the difference in net charge of the peptides.

If the correlation threshold is lowered to $|\tau| > 0.5$, three additional MD descriptors become correlated with these same two structural properties: the number of contacts with water molecules and with lipid tails—both with the hydrophobic fraction—, and the standard deviation of the transverse EDM—increasing with charge. However, the latter two correlations barely exceeded the threshold in a single case, suggesting that such connections were likely spurious. Consequently, only the mean number of contacts with water molecules was added to the list of potentially biased descriptors.

To determine whether the change in the descriptor was caused by uneven sampling or if other variables were involved, linear models were fitted for each potentially biased descriptor according to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sample	r^2	F-test	α		β_H		β_A		$\beta_{H,A}$	
		p-value	Value	p-value	Value	p-value	Value	p-value	Value	p-value
M2 + POPC	0.663	$1.55 \cdot 10^{-27}$	−0.241(40)	$2.00 \cdot 10^{-8}$	0.654(78)	$1.70 \cdot 10^{-13}$	−0.078(62)	$2.11 \cdot 10^{-1}$	0.17(11)	$1.30 \cdot 10^{-1}$
M2 + POPE:POPG	0.621	$5.86 \cdot 10^{-25}$	−0.306(37)	$2.54 \cdot 10^{-13}$	0.555(73)	$6.73 \cdot 10^{-12}$	−0.038(56)	$4.98 \cdot 10^{-1}$	0.11(10)	$2.65 \cdot 10^{-1}$
M2P + POPC	0.564	$1.60 \cdot 10^{-18}$	−0.256(53)	$4.75 \cdot 10^{-6}$	0.55(10)	$5.11 \cdot 10^{-7}$	−0.101(79)	$2.05 \cdot 10^{-1}$	0.22(14)	$1.07 \cdot 10^{-1}$
M2P + POPE:POPG	0.597	$2.03 \cdot 10^{-22}$	−0.293(43)	$6.26 \cdot 10^{-10}$	0.648(83)	$3.30 \cdot 10^{-12}$	0.045(64)	$4.87 \cdot 10^{-1}$	−0.02(11)	$8.56 \cdot 10^{-1}$
M3 + POPC	0.578	$1.01 \cdot 10^{-15}$	−0.687(88)	$1.69 \cdot 10^{-11}$	0.80(15)	$1.41 \cdot 10^{-6}$	−0.03(11)	$8.15 \cdot 10^{-1}$	0.12(19)	$5.22 \cdot 10^{-1}$
M3 + POPE:POPG	0.473	$4.86 \cdot 10^{-14}$	−0.567(67)	$1.61 \cdot 10^{-13}$	0.61(12)	$1.92 \cdot 10^{-6}$	0.028(88)	$7.52 \cdot 10^{-1}$	0.02(15)	$8.77 \cdot 10^{-1}$

Table 4.1: Results of the least-squares multivariate linear regression between the fraction of hydrophobic residues in the sequence H, the nominal activity A, and the penetration of the peptide into the membrane. The coefficient of determination r^2 and the p-value of the corresponding F-test assess the goodness of fit of each model. The regression coefficients (α, β_H, $\beta_{H,A}$) are reported together with their standard errors (in parentheses) in units of nm, and the p-values of the associated t-tests, indicating their statistical significance within the model.

$$y(X, A) = \alpha + \beta_X X + \beta_A A + \beta_{XA} X \cdot A, \tag{4.2}$$

where y is an MD descriptor, X a structural property, and A is a categorical variable that takes two values, 1 or 0, depending on whether the sequence has or does not have a nominal antimicrobial activity, so

$$\begin{cases} y(X)\big|_{A=0} = \alpha + \beta_X X \\ y(X)\big|_{A=1} = (\alpha + \beta_A) + (\beta_X + \beta_{XA})X \end{cases} \tag{4.3}$$

Depending on the values of the parameters β_A and β_{XA}, if they are significantly different from zero, the two lines can have different intercepts and slopes, so activity somehow affects the value of the descriptor. If this is not the case—the parameters β_A and β_{XA} are not significantly different from zero—, then the activity has no direct impact on the descriptor, so the differences observed in the latter can only be associated with the sampling of the space of sequences.

Figure 4.35 illustrates the relationship between penetration depth and fraction of hydrophobic residues. A linear relation between these two magnitudes is clear once the points associated with peptides not adsorbed into the membrane—further than 0.6 nm from the membrane—are omitted. The result of fitting the remaining data points to the model in Eq. 4.2, together with some statistical tests, are given in Table 4.1.

The values of r^2 are indicative of a moderate linear trend, while the significant results of the F-test suggest that at least one of the independent variables—fraction of hydrophobic residues or activity or both—has an impact on the penetration value. The significant result of the t-test in the coefficient β_H suggests that the fraction of hydrophobic residues does indeed condition how deep the peptide is adsorbed. In contrast, neither β_A nor β_{HA} is sufficiently different from zero ($p > 0.05$), indicating

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

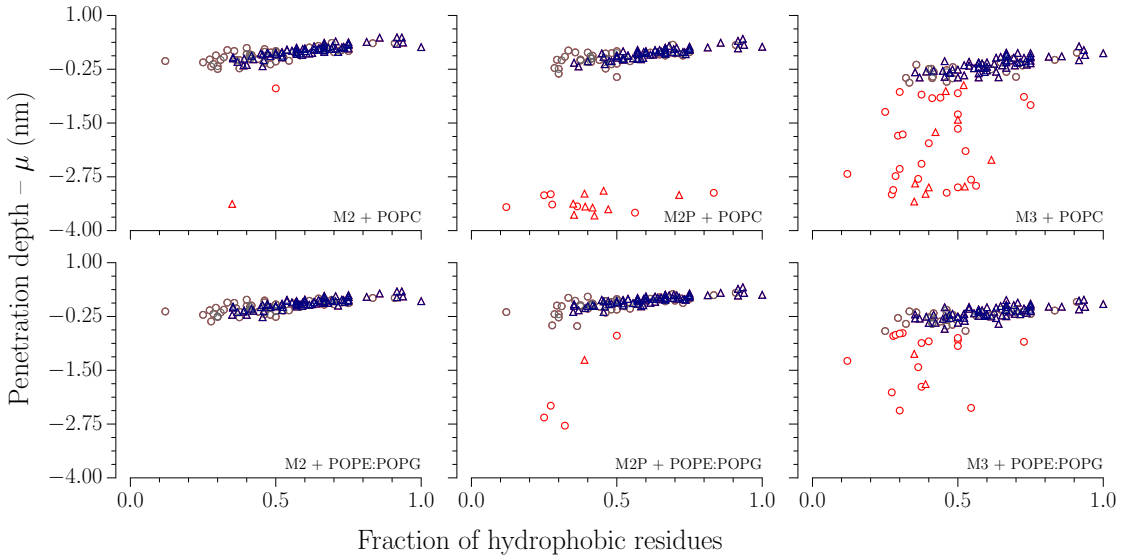


Figure 4.35: Penetration of the peptide into the bilayer as a function of the fraction of hydrophobic residues in the sequence. Each panel corresponds to a combination of lipid composition (rows) and force field (columns). The peptide penetration depth (Y-axis) is plotted against the fraction of hydrophobic residues of the peptide (X-axis) for each of the simulations. Blue triangles correspond to the simulation of active peptides, and gray circles to nonactives. Points in red were not used for the fitting, as they differ from the linear regime.

a lack of evidence of the activity modifying either the intercept or the slope of the model. In consequence, the observed difference in the penetration depth between active and nonactive peptides can only be attributed to the unequal sampling of the number of hydrophobic residues in the sequences.

The same analysis was applied to the relationship between the number of water contacts and the fraction of hydrophobic residues (Figure 4.36). The data exhibit a generally linear trend, except for a few outliers corresponding to peptides not adsorbed into the membrane. The linear fitting was restricted to those below a mean number of contacts with water molecules—per residue—below 3. The results are given in Table 4.2.

The low p-value of the t-test associated with the coefficient β_H confirms that the fraction of hydrophobic residues conditions the number of contacts with water; this time, the higher the value of this property, the less the peptide will be exposed to water molecules. But unlike in the previous case, some of the coefficients associated with the activity are now significant. Except for the M2P and POPE:POPG simulations, the term $\beta_{H,A}$ is non-zero ($p < 0.05$). The coefficient β_A is also non-zero ($p < 0.05$) in all but the M2 and M2P simulations of POPE:POPG. In consequence, the differences in the number of contacts with water molecules cannot be attributed

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

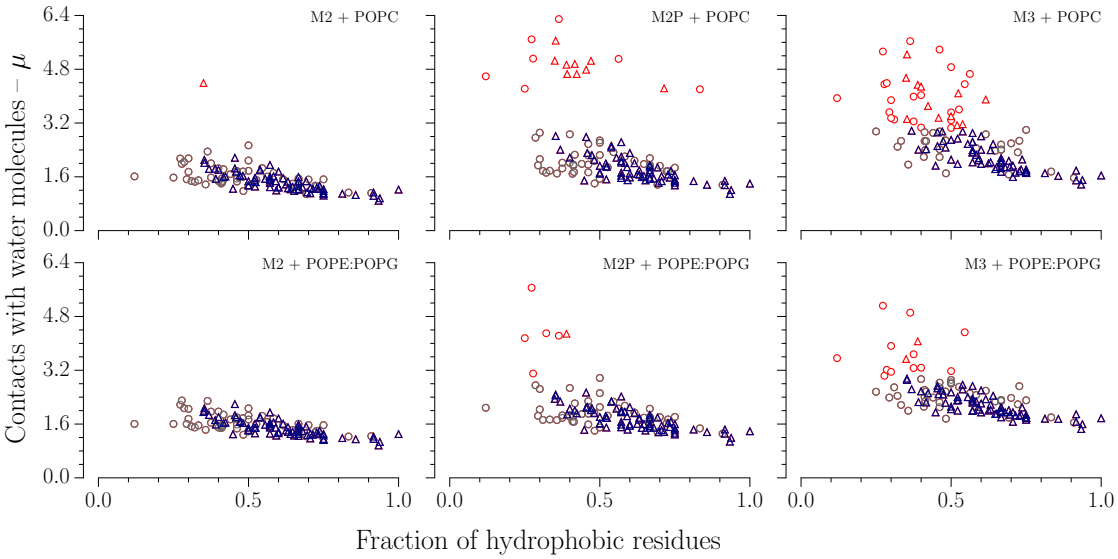


Figure 4.36: Contacts between the backbone of the peptide and water molecules as a function of the fraction of hydrophobic residues in the sequence. Each panel corresponds to a combination of lipid composition (rows) and force field (columns). The mean number of contacts with water molecules (Y-axis) is plotted against the fraction of hydrophobic residues of the peptide (X-axis) for each of the simulations. Blue triangles correspond to the simulation of active peptides, and gray circles to nonactives. Points in red were not used for the fitting, as they differ from the linear regime.

only to the sampling of the sequences, as its relation with the fraction of hydrophobic residues is different for active and nonactive sequences.

4.3.3 Classification of the antimicrobial activity of the sequences

The preceding analyses identified several MD descriptors that turned out to show statistically significant differences between the samples of simulations of active and nonactive peptides. Nevertheless, the size of these effects was small and the overlap of the distributions remained substantial, limiting the inference of the activity based on a single descriptor. To overcome such a limitation, we aimed to find a linear combination of the descriptors to obtain a new property so that the difference between the samples becomes large enough to be used to identify the antimicrobial activity of the sequence.

Model design Given the limited sequence space simulated—123 peptides—, such a metric should have the fewer parameters possible. The initial set of 32 MD de-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Sample	r^2	F-test	α		β_H		β_A		$\beta_{H,A}$	
		p-value	Value	p-value	Value	p-value	Value	p-value	Value	p-value
M2 + POPC	0.516	$2.26 \cdot 10^{-18}$	2.024(89)	$6.24 \cdot 10^{-45}$	−0.89(17)	$1.19 \cdot 10^{-6}$	0.30(14)	$2.91 \cdot 10^{-2}$	−0.58(24)	$1.76 \cdot 10^{-2}$
M2 + POPE:POPG	0.511	$1.97 \cdot 10^{-18}$	2.051(83)	$1.16 \cdot 10^{-48}$	−0.84(16)	$9.18 \cdot 10^{-7}$	0.24(13)	$6.44 \cdot 10^{-2}$	−0.48(22)	$3.31 \cdot 10^{-2}$
M2P + POPC	0.398	$2.27 \cdot 10^{-11}$	2.42(16)	$3.17 \cdot 10^{-27}$	−0.82(31)	$1.04 \cdot 10^{-2}$	0.62(24)	$1.29 \cdot 10^{-2}$	−1.14(42)	$7.97 \cdot 10^{-3}$
M2P + POPE:POPG	0.385	$4.92 \cdot 10^{-12}$	2.60(16)	$2.62 \cdot 10^{-32}$	−1.27(30)	$4.76 \cdot 10^{-5}$	0.17(23)	$4.56 \cdot 10^{-1}$	−0.43(40)	$2.92 \cdot 10^{-1}$
M3 + POPC	0.537	$4.78 \cdot 10^{-14}$	3.00(23)	$2.87 \cdot 10^{-22}$	−1.16(39)	$4.24 \cdot 10^{-3}$	0.78(29)	$8.34 \cdot 10^{-3}$	−1.37(48)	$5.61 \cdot 10^{-3}$
M3 + POPE:POPG	0.538	$7.17 \cdot 10^{-17}$	2.80(14)	$1.65 \cdot 10^{-36}$	−0.88(26)	$9.26 \cdot 10^{-4}$	0.49(19)	$1.02 \cdot 10^{-2}$	−0.98(32)	$3.00 \cdot 10^{-3}$

Table 4.2: Results of the least-squares multivariate linear regression between the fraction of hydrophobic residues in the sequence H, its nominal activity A, and the mean number of contacts with water molecules. The coefficient of determination r^2 and the p-value of the corresponding F-test assess the goodness of fit of each model. The regression coefficients (α, β_H, $\beta_{H,A}$) are reported together with their standard errors (in parentheses), both adimensional, and the p-values of the associated t-tests, indicating their statistical significance within the model.

scriptors was reduced based on the results of the previous discussion: the mean penetration of the peptide, although different between active and nonactive sequences, appeared to be biased; no significant difference between the samples of tilt have been observed, the same as for many components of EDM and HDM. In addition, changes in EDM and HDM with respect to the initial structure, because of the stiffness of the peptides, are not considered to be especially relevant for a comparison. Therefore, removing both the mean and the standard deviation of these magnitudes, the number of remaining descriptors is 16.

The most simple model that uses all these descriptors would contain a total of 17 adjustable parameters—one per descriptor plus an offset. At the same time, a reasonable objective would be to produce a binary classification of the MD data into two categories: potentially active and potentially nonactive. With all of this in mind, a logistic regression model (Eq. 3.123) was selected as a first approach for the prediction of the antimicrobial activity of a sequence based on MD data. The basics of the implementation were already mentioned in Section 3.4.1, while additional methodological details relevant to model training are described below.

Model optimization typically involves partitioning the available data into a training set—used to adjust the parameters of the model—and a test set—used to evaluate the performance after the training—, roughly in a 4:1 ratio. When dealing with reduced data sets, a common approach is the use of a k-fold cross-validation: the data set is split into k different groups (folds), so $k-1$ of them form the training set and the remaining one becomes the test set. Each combination of these groups is used to train a different model, so the test set used to validate each of them is different; the average of the performance of these models is then computed. This aims to reduce the sensitivity of the results to the initial partition of the data, as it may condition the overall results.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

As the number of active and nonactive peptides is different, a stratified cross-validation protocol was applied to preserve the class proportion of the original data within each fold. Despite this adjustment, the model performance can still depend on the initial partition of the data. To further reduce this dependence, 5-fold cross validation process was repeated a total of 200 times,* each time with a different seed, to increase the diversity of the input data. The ensemble of performance metrics resulting from these iterations was then used to evaluate the predictive power.

To explore the influence of the input data, 10 different data sets were constructed by systematically combining the simulation data:

- **All**, containing all the simulation data, i.e. each peptide has associated the values of the descriptors of all its simulations—with both lipid compositions and the three force fields.
- **M2**, **M2P** and **M3**, comprising all data from simulations performed under the same force field.
- Six additional data sets correspond to each unique combination of lipid composition and force field.

The number of adjustable parameters in the models grows accordingly with the number of descriptors used: from 17 in the most simple case of only data from one force field and the same lipid composition, to 33 when all data from each force field is used, and 97† for the model that incorporates all data.

Evaluation and comparative analysis of the model performance To investigate the predictive power of the trained models, a number of different metrics were selected. For each trained model, all metrics were computed and then averaged over the 5 folds. The ensemble of 200 independent repetitions provided a distribution of these averaged metrics, which was used to derive statistical estimates of the model performance. Because of the imbalance between the number of active and nonactive peptides simulated, some metrics—particularly accuracy and precision—are not fully representative when considered in isolation. Nevertheless, their comparative trends across different models remain informative.

A practical objective for these models could be the identification of potentially active peptides, maximizing the number of true positives while minimizing the false

*The way in which n elements can be arranged into m groups, ignoring their order, is $\frac{n!}{(n/m)!^m m!}$. For $n = 50$ (peptides) and $m = 5$ (folds), there are of the order of 10^{29} combinations.

†Close to the number of simulations used, there is a significant chance for this model to overfit the data.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

negatives. Among the metrics described in Section 3.4, the ***recall*** of the model will increase as the better the model is at fulfilling this task. Since the recall is computed from the assignment of the activity in the subset of the active sequences (Eq. 3.128), it is agnostic to the imbalance in the data set. A secondary objective concerns the reliability of positive classifications, quantified by the precision, which measures the proportion of correctly identified active peptides among all peptides predicted as active. Precision is, however, sensitive to class imbalance. An alternative in such scenarios is the construction of precision–recall curves, where these two metrics are obtained for different values of the classification threshold; the area under the precision–recall curve (AUC-PR) is then used as a secondary performance indicator. A larger value corresponds to a more consistent and reliable classification across decision thresholds.

The samples of the different metrics obtained are represented in Figure 4.37. Table D.17 contains some relevant statistics associated with each of them. In general, the closer each of the metrics is to 1, the better the model is in the measured aspect; the value of 0.5 is what could be expected from a (uniform) random classification in a balanced data set. The medians of all distributions are sensibly larger than this value, all above 0.7. At the same time, the variances of the samples are not that large, which shows a notable consistency in the metrics among the different repetitions despite the small data set.

Normality of the metric distributions was evaluated using the Shapiro-Wilk test (Table D.18). Several metrics deviated from normality; therefore, Mann-Whitney U tests were used to compare the performance between pairs of models.

The model trained with all available data did not display a substantial performance improvement with respect to those that used all data from a single force field. The comparison between these models, given in Table D.19, confirms such observations: models trained with all data only perform better than those trained with only M2 simulations. Its recall is similar to that of the models of M2P and M3 data. The rest of the metrics are higher in the models trained with M3 simulations, except for the area under the precision–recall curve, which is notably higher in M2P.

The pairwise comparison between the models trained with simulations of each force field (Table D.21) points in the same direction: the metrics in M2 are the worst, and M3-based models outperformed M2P in all metrics, except the area under the PR curve. These results indicate that the use of M2 data drags the performance of the models, and that the addition of parameters in the more complex model is not justified a by better performance. The use of either M2P or M3 data is enough to get better results.

The impact of the lipid composition was evaluated by comparing models trained with data from both membrane types with those trained with a single bilayer. The outcomes of these comparisons, given in Table D.23, highlight that, while most

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

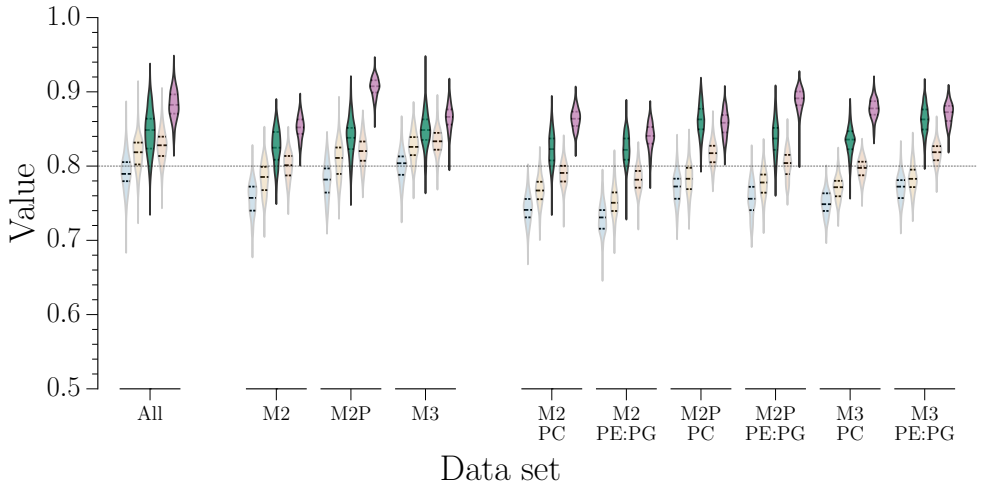


Figure 4.37: Evaluation of the performance of the classification models. Performance metrics are grouped by data set used to train and test: "All", trained with all simulation data; M2, M2P and M3, trained with all the data from each of these force fields; and the models that used data from each combination of lipid composition and force field (PC for the POPC membrane and PE:PG for the POPE:POPG bilayer). For each data set, the five distributions are ordered; from left to right, accuracy, precision, recall (highlighted, in green), $F1$-score, and area under the precision–recall curve (highlighted, in purple). The Y-axis starts at 0.5 for ease of comparison among the different metrics; a horizontal line at 0.8 is included for visual reference.

metrics are better when both compositions are used in the model, there are some notable exceptions: the recall is not sufficiently different from that of the models fed with only M2 data and a single lipid composition, and the same happens for the M2P model trained with only POPE:POPG simulations with respect to the M2P model. On the other hand, the recall in models trained with M2P+POPC and M3+POPE:POPG tends to be higher than when both compositions are used. The area under the precision-recall curves is also higher for the M3 models that use a single lipid composition.

Overall, the better performance corresponds to models trained with M3 simulation data—the most recent force field—; among them, the model which only used simulation data from the POPE:POPG bilayer yielded the best results. This model outperformed all others, except for the area under the precision-recall curve—higher in models trained with M2P data. Assuming its application as an AMP candidate filter, the highest recall, together with the higher computational efficiency of M3 simulations with respect to M2P, makes this model the most promising candidate for future development. Although performance differences could change with a larger sampling of the space of sequences, the observed trends point to M3 as the path to

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

improve the classification and discard the use of M2 data.

The previous study was carried out on models that had a different number of adjustable parameters, as one was required per MD descriptor used. To compare the performance of models trained with exactly the same data sets but with the same number of parameters, a principal component analysis (PCA) of the space of descriptors was performed. The original data was projected onto subsets of either 16—the same number of descriptors as in the most simple data sets—or 7—almost half the number of descriptors—first principal components (PCs), which bear most of the variance in the data. Logistic regression models trained on PC-transformed data thus contained the same number of parameters, the number of selected PCs plus one.

Performance metrics samples are represented in Figure D.2. The simple models performed exactly the same when the 16 PCs were selected as they did when trained with the original data; this was expected since both use effectively the same data, differing only in a linear transformation. Models trained with both lipid compositions suffered from slight changes, while for the most complex model the drop in AUC-PR curve is noticeable. Reducing the number of PCs to 7 altered the samples of all metrics, being the most relevant aspect the increase in recall of the M2 models and the drop in both accuracy and precision of the M2P-based ones. Still, the model trained with M3 simulations of both lipid compositions holds the best recall, even increasing it with respect to those trained with the whole data set and no projection onto the PCs.

Limitations Several aspects must be considered when interpreting these results. The first concerns the limited size of the data set, which included only 123 peptide sequences. Although this number is relatively large for an MD study, it remains insufficient to construct a statistically robust and generalizable predictive model. The ratio between the number of simulations and the number of adjustable parameters is low, limiting its predictive power and increasing the potential for overfitting. Nevertheless, the statistical analyses indicate that the MD descriptors differ significantly between active and nonactive sequences, suggesting that the simulations capture physically meaningful differences in peptide–membrane interactions that classification models can exploit.

A second limitation arises from the simplified representation of the simulated systems. Each configuration involved a single peptide interacting with a symmetric lipid bilayer, surrounded by equivalent aqueous environments on both sides. The peptides were modeled as rigid, ideal α-helices, without conformational flexibility or rearrangements of their secondary structure. These simulations facilitate a systematic comparison across multiple conditions, but they depart from realistic *in vivo* and *in vitro* scenarios, where membrane asymmetry, curvature, and peptide confor-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

mational changes and aggregation play decisive roles. The assumption of rigidity is particularly critical, as the actual interfacial structures of most AMPs are unknown, and those containing helix-disrupting residues such as Pro or Gly may adopt markedly different conformations in real membranes. This limitation could alter the balance between hydrophobic and electrostatic interactions, ultimately affecting the computed MD descriptors.

Despite these simplifications, the present framework provides a proof of concept that the CG MD descriptors retain sufficient physical meaning to distinguish peptides from both classes. The results support the feasibility of extracting predictive information from large-scale CG simulations, even under idealized conditions. The approach offers a controlled basis from which more realistic models can be developed.

Future work should extend the space of sequences studied and these analyses by incorporating experimentally determined peptide conformations, atomistic representations, and simulations that account for the flexibility and cooperative adsorption of the peptides. These improvements will clarify the relevance of the descriptors and to which extent CG simulations can reproduce biologically meaningful patterns of antimicrobial activity.

4.3.4 Conclusions

The statistical analysis of CG simulations of solvated systems, including a unit of a peptide and a lipid bilayer, showed that, when grouped according to the expected antimicrobial activity of the peptide sequence, several descriptors exhibit statistically significant differences. A detailed examination across all simulated conditions—covering three Martini force fields and two membrane compositions—showed that these differences are systematic, yet individually insufficient to discriminate active from nonactive sequences with confidence.

Comparative analyses confirmed that the interaction between the peptides and the two lipid bilayers, POPC and POPE:POPG 3:1, differs consistently, reflecting the expected membrane selectivity of antimicrobial peptides. The extent of this differential interaction depends strongly on the force field, being most pronounced in Martini 3 simulations. Despite this improvement, the magnitude of the observed changes remains too small to allow a reliable classification based solely on individual descriptors.

The comparison between force field versions provided further insight into how successive Martini parametrizations affect peptide–membrane interactions. The newer force fields, particularly Martini 3, corrected the overestimation of hydrophobic interactions characteristic of Martini 2, thus reducing artificial peptide adsorption and aggregation at the membrane surface.

To integrate the information contained in all descriptors, logistic regression mod-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

els were constructed using data sets of increasing complexity—from single force field and membrane combinations to the complete data set using all simulation conditions. Among the models tested, those trained on MARTINI 3 data, particularly for simulations with the anionic POPE:POPG bilayer, achieved the best performance in filtering active sequences. This result remained consistent when the data was projected onto a reduced space defined by principal components, indicating that the discriminative capacity of the models it is not merely a consequence of the dimensionality of the descriptor space.

Although the resulting models are not directly applicable for predictive purposes, the overall results support the use of using MD-derived descriptors as informative variables for activity classification. The present work thus constitutes a proof of concept showing that dynamic, simulation-based observables encode relevant information about antimicrobial behavior. Future efforts should focus on expanding the sequence space, incorporating flexible and experimentally-derived peptide structures, and integrating multiscale or experimental data to enhance both predictive power and biological relevance.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 5.

Conclusions

It's different than I'd envisioned.
Whatever happens next, I do not think it is to be feared.
The prisoner, Outer Wilds

This work confronts a complex question in molecular biophysics: how short, amphipathic peptides recognize and interact with lipid membranes, and how these events give rise to the antimicrobial activity that characterizes a large class of biologically relevant sequences. The approach followed combined atomistic and CG MD simulations with systematic data analysis and statistical methods, aiming to connect the microscopic detail of molecular structures with the properties that define their biological function.

The simulated systems are a simplification of the biological reality: a single peptide in interaction with a symmetric mono- or bi-component bilayer of phospholipids surrounded by water and ions. Thus, the results revealed a level of complexity in the interactions and structural transformations, which suggests that minimal models still capture relevant phenomena associated with the peptide–membrane interactions, the conformational flexibility of peptides, and the membrane response.

This work was structured around three differentiated but related aspects: The first part comprised the atomistic characterization of peptide adsorption and the conformational dynamics they undergo. The second stage was focused on the establishment of a standardized protocol for the simulation of peptide–membrane systems and their analysis, which led to the development of a database that contains data from simulations performed under systematic conditions. The final part used these data to investigate whether simulations of a single peptide can anticipate their antimicrobial activity.

Structural characterization of membrane-targeting antimicrobial peptides at the lipid membrane–water interface The first part of this study addressed the behavior of membrane-targeting peptides at the water–lipid interface. Three different membrane-targeting peptide sequences and a model for a bacterial cell were selected for this purpose. Five different independent replicas were performed for

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

each of them, starting with a different orientation of the peptide around its axis. The effect of the initial placement of the peptide was tested by using two distinct initial configurations: a shallow interfacial location and a deeper embedding into the bilayer. To achieve both of these states, an external harmonic potential was imposed during the equilibration stage, so the production simulations started from unbiased states and were extended for 900 ns each.

The shallow configuration led to peptides close to the plane defined by the phosphate groups of the lipids, but no penetration events were observed. This occurred in line with proposed mechanisms by which the adsorption of a highly unstructured membrane-targeting AMP is followed by the folding into an amphipathic structure.

The structure of peptides that started from this configuration was largely variable across replicas. Such variability is indicative of a complex energy surface, and also of the folding in the membrane being a process that occurs beyond the microsecond time scale.

The diversity of structures sampled allowed for a rigorous comparison of different helicity metrics. The DSSP algorithm yields a finite, discrete number of helicity values; this prevents it from capturing subtle conformational changes while overestimating the effect of minor structural variations. Near-helical states lacking specific H-bonds within the backbone were not recognized as such by this method. Continuous alternatives like αRMSD or α-block were able to depict such effects with greater consistency. These metrics were shown to be highly correlated, but not equivalent. A major discrepancy in the assignment of the helicity arose when peptides were well-folded into a helix or when completely extended; α-block compressed the range of helicities to a more restrained domain than the theoretical one, overestimating the helicity of the extended peptides and underestimating that of the full helices. αRMSD did not only show to be numerically stable, but yielded values within all its domain and their correspondence with the expected value upon visual inspection was far more satisfactory. Furthermore, it was observed that computing αRMSD using as reference the whole backbone—as originally proposed—or using only the C_α atoms yields nearly identical results, enabling a more efficient quantification of the structural changes.

The loss of the helicity was followed by a general compactification of the structure. The analysis of the differences by force field showed that the greater divergence came from AMBER simulations: peptides remained farther from the bilayer center, the helicities were less consistent and formed fewer contacts and H-bonds with lipids. This did not affect the number of H-bonds within the peptide, which was roughly the same across force fields for a given sequence.

To overcome both the limited adsorption and the slow dynamics of the folding process, a second batch of simulations with the peptide was performed, starting from an embedded configuration. During these simulations, peptides emerged near

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

the phosphates, but the conformational dynamics was notably different from those of the previous simulations: While still diverse, the variability was significantly reduced and more consistent results were obtained. The helical content was higher than in the previous simulations, indicating that the penetration depth conditions the sampled states—at least in the simulated timescales.

Regardless of the initial configuration of the simulation, AMBER remained the force field with which the higher dispersion of the helicity values and the highest discrepancies from all other force fields were observed. The structure of the membranes was then analyzed, where a higher adsorption of water molecules into AMBER bilayers was recorded. Thus, the systematic variability in AMBER can be associated with a difference in the local environment of the interface, weakening the peptide–lipid interactions in favor of those with water.

The dynamics of the peptides at the interface was studied by their diffusive behavior. The estimated values of the diffusion coefficients were not shown to be correlated with the structure or direct interaction with lipid molecules. Instead, a clear dependence on the insertion depth was observed: the deeper the peptide, the more hindered its lateral motion.

A methodological approach, the hydrogen isotope exchange method, was tested simultaneously with other force fields. The results indicate that GROMOS-HIE performed as well as the standard GROMOS force field: structural outcomes—despite their variability—and dynamical properties, were consistently compatible with the standard formulation of the force field. The 3-fold increase in performance did not significantly affect any of the magnitudes studied, so the use of an alternative treatment of electrostatics was also tested to look for further performance increases. The results were also satisfactory, as the values obtained were compatible with those obtained from both GROMOS-HIE and standard GROMOS simulations; only a slightly higher variation across replicas was observed for some properties.

Finally, long simulations were performed under the GROMOS-HIE force field to evaluate the limitations of the timescales previously used. The overall tendency was a progressive loss of the helicity, with only minor events of partial refolding being observed only for some simulations. These results reinforce a view by which standard MD by itself is not adequate for the exhaustive exploration of peptide–membrane systems. As an alternative, enhanced sampling methods could be used to explore the whole conformational space, in combination with shorter unbiased simulations of relevant states.

Systematic simulation of peptide and membrane systems and database for the study of membrane-targeting antimicrobial peptides The limited reproducibility and slow convergence observed in atomistic simulations of peptide–membrane systems highlighted their complexity of the free energy landscape. To

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

overcome this, a computational workflow was implemented for the preparation, simulation, and analysis of systems at a CG resolution. The reduction in the internal degrees of freedom of the peptide structure leads to a simplified free energy surface that improves the reproducibility and systematic study of these systems. The workflow automates all processes involved in the simulation, from the generation of peptide structures given their sequence, the assembly of the system structure, and the preparation of the simulation inputs to the systematic analysis of the final trajectories. This standardization allows the creation of data sets of MD-associated numerical descriptors obtained under controlled and consistent conditions, facilitating the later use of the data for systematic studies of the simulation results.

Using this framework, a total of 738 simulations were performed. This number is a combination of 123 peptides—73 of which were reported to have antimicrobial activity—, two lipid bilayer compositions—POPC and POPE:POPG 3:1—, and three MARTINI force fields—versions 2.2, 2.2P, and 3. All systems consisted of a single peptide, modeled as a full α-helix, in the presence of a symmetric lipid bilayer of 250 lipids per leaflet, surrounded by water and neutralizing counterions. The CG resolution allowed to reach the microsecond timescales, enabling for sufficient sampling of the interaction and adsorption processes.

From each of the trajectories, a set of quantitative descriptors was derived, describing the structural and dynamical aspects of peptide–membrane interactions. These included changes induced by the peptide in the membrane, relative position and orientation of the peptide with respect to the surface of the bilayer and its dynamics, and specific contacts with lipids and water molecules. Together, these metrics provide a basis for comparing how the peptide sequence, the lipid composition, or the force field used influence the simulation results.

All simulation results were consolidated in the SuPepMem database (www.supepmem.com). This repository contains MD results related to membrane-targeting AMPs and their interactions with model membranes. The database contains simulation metadata and input files together with the numerical descriptors extracted from the automated analysis. The web interface allows users to browse, filter, and compare systems, as well as download descriptor data and input files. The architecture of the database was designed to allow for future extensions, including a wider variety of systems, multi-peptide simulations, and multiscale resolution.

Connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches For the final part of the present work, the data generated in the CG simulations previously detailed was used to investigate whether the descriptors reflect effects associated with an earlier antimicrobial activity of the peptide. Each of the 738 simulations—consisting of a single peptide interacting with a lipid membrane—was processed to extract a set of

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

numerical descriptors that would then be analyzed to explore the differences across three dimensions: (i) nominal antimicrobial activity of the sequence involved, (ii) lipid membrane composition, and (iii) force field used.

The outcome of the statistical analyses performed revealed significant differences between active and nonactive sequences for all combinations of lipid composition and force field. Such differences were however small in magnitude, as indicated by a substantial overlap of most samples.

The analyses also revealed a markedly different behavior when the simulation involved a POPC—neutral—or a POPE:POPG 3:1 bilayer—anionic. The differential behavior of peptides toward the membrane based on the antimicrobial activity was the most prominent in Martini 3 simulations, while for both Martini 2.2 and Martini 2.2P it was barely manifested. This suggests that the updated parametrization given by Martini 3 is able to capture the sensitivity of the simulated membranes to the physicochemical properties of peptides.

The effect of the force field was further studied by a pairwise comparison of the MD descriptors. The overestimation of the peptide–membrane interaction of the earlier Martini 2.2 force field was corrected in Martini 3. Under this force field, the interaction between membrane-targeting peptides and the lipid membranes differed from that of peptides with no reported antimicrobial activity.

Despite the systematic occurrence of differences between the samples of descriptors on the basis of the antimicrobial activity of the sequences, none of them was large enough to confer them with sufficient discriminative power for the identification of membrane-targeting peptides. To integrate the collective information across all descriptors, a set of logistic regression models was trained using data sets of increasing complexity. The simplest models were constructed from the data associated with simulations of a single lipid composition and force field; intermediate models combined both membrane compositions within one force field; and the most complex included all simulation data. The training was done after a prior simplification of the descriptor space, removing both the least relevant descriptors and also those for which the differences previously found were related to a methodological bias.

For each data set, performance was evaluated using standard classification metrics. Models trained exclusively on data from Martini 3 simulations involving the POPE:POPG 3:1 bilayer achieved the highest performance in filtering active sequences. To verify that the results did not depend on the dimensionality of the data set, the simulation data were projected onto reduced subspaces of their principal components and used to train another set of models. Although the results differed, the overall trends remained consistent, confirming that the discriminative power originates from the descriptors.

Despite the limited size of the data set and the simplified simulation conditions, the results indicate that descriptors obtained from MD simulations contain informa-

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

tion that can be exploited to separate the two classes of sequences. The performance of the classification models, while modest, establishes a proof of concept for the use of simulation results as quantitative predictors of biological function. Importantly, the force field—and other conditions—used in the simulations must provide a physically consistent basis for the identification of such patterns.

Several limitations hinder the applicability of these models. An unbalanced and small space of peptide sequences, a simplification of the peptide structure that forbids the depiction of conformational changes that can contribute to its activity, or a limited set of simulation conditions are aspects that constrained the current study. Future work would require the expansion of the sampling of sequences and their diversity, the integration of different types of simulation data—e.g. atomistic or enhanced-sampling—, and the combination with experimental measurements of structure and activity. These aspects would help to assess the potential of MD data as a source of molecular information for the development of data-based predictive models.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix A.

Mathematical derivations

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A.1 Moments of inertia of an ellipsoid

Volume of an ellipsoid The parametric equation of an ellipsoid of semiaxes a, b, and c, aligned with the X, Y, and Z axes, respectively, is

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1. \tag{A.1}$$

By doing the change of variables $x_i'=c_i x_i$, where $x_i=x,y,z$ and $c_i=a,b,c$, the previous equation is transformed into that of a 3-dimensional sphere

$$x'^2+y'^2+z'^2=1, \tag{A.2}$$

from which the polar coordinates of a point in the surface of the ellipsoid can be deduced

$$\begin{cases} x=ar\sin\theta\cos\phi \\ y=br\sin\theta\sin\phi \\ z=cr\cos\theta \end{cases} \tag{A.3}$$

and, by analogy, the volume element of the ellipsoid is

$$d\tau=abcr^2\sin\theta dr d\theta d\phi. \tag{A.4}$$

Therefore, the volume of the ellipsoid is given by the integral

$$\begin{aligned} V=\iiint d\tau &= abc\iiint r^2\sin\theta dr dr\theta d\phi \\ &= abc\int_0^1 r^2dr\int_0^\theta \sin\theta d\theta\int_0^{2\pi} d\phi \end{aligned} \tag{A.5}$$

so

$$\boxed{V=\frac{4\pi}{3}abc} \tag{A.6}$$

Moments of inertia of an ellipsoid The components of the tensor of inertia of a solid object of density $\rho(\mathbf{x})$ are given by

$$I_{ij}=\iiint \rho(\mathbf{x})\left(\|\mathbf{x}\|^2\delta_{ij}-x_ix_j\right)d\tau. \tag{A.7}$$

This implies that the off-diagonal elements $I_{i,j}$, $j\neq i$, are integrals of $\rho(\mathbf{x})x_ix_j$. For a homogeneous ellipsoid, this means that every off-diagonal element involves an integral of $\sin\psi\cos\psi$ between the limits $[0,\pi)$ if $\psi=\theta$, or $[0,2\pi)$ if $\psi=\phi$. As $\sin\psi\cos\psi$ is an odd function and periodic in $[0,\pi)$, all these integrals and, consequently, the off-diagonal elements are zero, which means that the axes of the ellipsoid are also its principal axes of inertia.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

For the diagonal elements, Eq. A.7 becomes

$$I_{ii} = \rho \iiint \left(x_j^2 + x_k^2\right) d\tau, \tag{A.8}$$

where $i \neq j \neq k$. Knowing the result of the integral of any squared component of a sphere is, due to rotational symmetry,

$$\begin{aligned} \iiint x'^2 d\tau' &= \int_0^1 r^4 dr \int_0^\pi \sin^3\theta d\theta \int_0^{2\pi} \cos^2\phi d\phi \\ &= \frac{1}{5} \cdot \frac{2}{3} \cdot \pi \\ &= \frac{4\pi}{15}\,, \end{aligned} \tag{A.9}$$

the integral over the coordinates of an ellipsoid is this value multiplied by the length of the semiaxis along which it is oriented. Thus,

$$I_{ii} = \frac{4\pi}{15} \rho abc \left(c_j^2 + c_k^2\right) \tag{A.10}$$

where c_j and c_k are the lengths of the semiaxes of the ellipsoid that are not oriented along i.

As for an homogeneous ellipsoid $\rho = M/V$, using the result in Eq. A.6, the moment of inertia along the i-th axes of an ellipsoid is

$$\boxed{I_{ii} = \frac{M}{5}\left(c_j^2 + c_k^2\right)} \tag{A.11}$$

Semiaxes of an ellipsoid from the moments of inertia Considering the relation given by Eq. A.11, the values of the semiaxes of an ellipsoid can be derived from its moments of inertia as

$$\boxed{c_i = \sqrt{\frac{5}{2M}\left(I_{jj} + I_{kk} - I_{ii}\right)}} \tag{A.12}$$

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A.2 Translational and rotational diffusion of particles

The probability density function associated with the position of a particle in a N-dimensional space $\mathbf{r}$ at an instant t, $P(\mathbf{r},t)$, governed by a diffusive process, is the solution to Fick's diffusion equation[316]

$$\frac{\partial P(\mathbf{r},t)}{\partial t} = D\nabla^2 P(\mathbf{r},t). \tag{A.13}$$

being D the diffusion coefficient.

A.2.1 Translational diffusion

Solution of the diffusion equation in 1-dimension In the 1-dimensional case, Eq. A.13 is reduced to

$$\frac{\partial P(x,t)}{\partial t} = D\frac{\partial^2 P(x,t)}{\partial x^2}, \tag{A.14}$$

where $x \in \mathbb{R}$ is the position along that dimension. The probability density function can be expressed in terms of the inverse Fourier transform of the PDF in the spatial frequencies, $\widehat{P}(k,t)$,

$$P(x,t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ikx}\widehat{P}(k,t)dk; \tag{A.15}$$

then

$$\frac{\partial P(x,t)}{\partial t} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ikx}\frac{\partial \widehat{P}(k,t)}{\partial t}dk, \tag{A.16}$$

$$\frac{\partial^2 P(x,t)}{\partial x^2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(\frac{\partial^2}{\partial x^2}e^{-ikx}\right)\widehat{P}(k,t)dk. \tag{A.17}$$

Substituting Eqs. A.16 and A.17 into Eq. A.14 leads to

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ikx}\left(\frac{\partial \widehat{P}(k,t)}{\partial t} + Dk^2\widehat{P}(k,t)\right)dk = 0, \tag{A.18}$$

which implies that

$$\frac{\partial \widehat{P}(k,t)}{\partial t} + Dk^2\widehat{P}(k,t) = 0. \tag{A.19}$$

Solving this simple first-order linear differential equation leads to

$$\widehat{P}(k,t) = C(k)e^{-Dk^2 t}, \tag{A.20}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where $C(k)$ is a function that has to be determined by the boundary conditions that can only depend on k. If at time $t = 0$ the particle is located at $x = 0$, then $P(x, 0) = \delta(x)$, and as the Fourier transform of the Dirac delta is equal to $1/\sqrt{2\pi}$, then

$$\widehat{P}(k, 0) = C(k) = \frac{1}{\sqrt{2\pi}}, \tag{A.21}$$

a constant factor, so

$$\widehat{P}(k, t) = \frac{1}{\sqrt{2\pi}} e^{-Dk^2 t}. \tag{A.22}$$

Now it only remains to compute the inverse Fourier transform as given by Eq. A.15,

$$\begin{aligned} P(x, t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ikx} \frac{1}{\sqrt{2\pi}} e^{-Dk^2 t} dk \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{2Dt}} e^{-x^2/(4Dt)} \end{aligned} \tag{A.23}$$

Thus, the 1-dimensional probability density function associated with a diffusive process is a Gaussian distribution of mean $\mu = 0$ and variance $\sigma^2 = 2Dt$:

$$\boxed{P(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-x^2/(4Dt)}} \tag{A.24}$$

Probability density function associated to a N-dimensional displacement If the components of the Brownian motion of a particle in a N-dimensional space, whose coordinates are $\{x_i\}_{i=1,\dots,N}$, are distributed according to a 0-mean σ^2-variance Gaussian distribution, and are independent from each other, then, as previously demonstrated, the probability density function associated with the position along the i-th dimension at time t is

$$P(x_i, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x_i^2/(2\sigma^2)}, \tag{A.25}$$

where $\sigma = 2Dt$. Let x be the norm of the N-dimensional position vector, $x = \left(\sum_{i=1}^{N} x_i^2\right)^{1/2}$. The cumulative distribution function associated with the distance x is

$$F_N(x, \sigma) = \int_{V_N} \prod_{i=1}^{N} P(x_i, \sigma) d\tau_N, \tag{A.26}$$

being V_N the volume enclosed by the (N-1)-sphere of radius x. The previous expression can be rewritten in spherical coordinates, so it becomes

$$F_N(x, \sigma) = \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{S_{N-1}} dA_N \int_0^x \rho^{N-1} e^{-\rho^2/(2\sigma^2)} d\rho. \tag{A.27}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where the term $\int_{S_{N-1}} dA_N$ is the surface of the (N-1)-sphere, which can be substituted for its analytical expression[317]

$$\int_{S_{N-1}} dA_N = \frac{2\pi^{N/2}}{\Gamma\left(N/2\right)}, \tag{A.28}$$

being Γ the gamma function, so Eq. A.27 becomes

$$F_N(x,\sigma) = \frac{2}{(2\sigma^2)^{N/2}\,\Gamma\left(N/2\right)} \int_0^x \rho^{N-1} e^{-\rho^2/(2\sigma^2)} d\rho. \tag{A.29}$$

Applying the fundamental theorem of calculus to find the probability density function associated with a displacement x:

$$\begin{aligned} P_N(x,\sigma) = \frac{d}{dx}\left[F_N(x,\sigma)\right] &= \frac{2}{(2\sigma^2)^{N/2}\,\Gamma\left(N/2\right)} \frac{d}{dx}\left[\int_0^x \rho^{N-1} e^{-\rho^2/(2\sigma^2)} d\rho\right] \\ &= \frac{2}{(2\sigma^2)^{N/2}\,\Gamma\left(N/2\right)} r^{N-1} e^{-x^2/(2\sigma^2)}. \end{aligned} \tag{A.30}$$

For $\sigma^2 = 2Dt$, being t the time window associated with the displacement and D the diffusion coefficient of the particle, Eq. A.30 becomes the probability density function associated with a displacement x to occur at a time t in a N-dimensional space

$$\boxed{P_N(r,t) = \frac{r^{N-1}}{2^{N-1}(Dt)^{N/2}\Gamma(N/2)} e^{-r^2/(4Dt)}}, \tag{A.31}$$

which, evaluated for up to three dimensions becomes:

$$\begin{cases} P_{1D}(x,t) = \dfrac{1}{\sqrt{\pi Dt}} e^{-x^2/(4Dt)} \\ P_{2D}(x,t) = \dfrac{x}{2Dt} e^{-x^2/(4Dt)} \\ P_{3D}(x,t) = \dfrac{x^2}{2\sqrt{\pi (Dt)^3}} e^{-x^2/(4Dt)} \end{cases} \tag{A.32}$$

Verification For assessing that Eq. A.31 is in fact a probability density function, first it must be ensured how the units of $P(x,t)$ are those of the inverse of x; in this case, inverse of length:

$$[P(x,t)] = \frac{1}{(L^2T \cdot T^{-1})^{N/2}} L^{N-1} = L^{-1}. \tag{A.33}$$

Also $P(x,t)$ needs to be normalized, i.e. the integral over the domain should be equal to 1:

$$\int_0^\infty P(x,t)dx = \frac{1}{2^{N-1}(Dt)^{N/2}\Gamma\left(N/2\right)} \int_0^\infty x^{N-1} e^{-x^2/(4Dt)} dx; \tag{A.34}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

by the definition of the Gamma function,

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \tag{A.35}$$

which, for $t = x^2/(4Dt)$, evaluated in $z = N/2$, becomes

$$\Gamma\left(N/2\right) = \frac{1}{2^{N-1}(Dt)^{N/2}} \int_0^\infty x^{N-1} e^{-x^2/(4Dt)} dx, \tag{A.36}$$

so after clearing the integral,

$$\int_0^\infty x^{N-1} e^{-x^2/(4Dt)} dx = 2^{N-1}(Dt)^{N/2}\Gamma\left(N/2\right), \tag{A.37}$$

and substituting its value into Eq. A.34, it is finally proven that

$$\int_0^\infty P(x,t) = \frac{1}{2^{N-1}\Gamma\left(N/2\right)} \left(\frac{1}{Dt}\right)^{N/2} 2^{N-1}(Dt)^{N/2}\Gamma\left(N/2\right) = 1 \tag{A.38}$$

A.2.2 Rotational diffusion

Solution of the rotational diffusion equation in 2 dimensions Given that the orientation of a vector in a N-dimensional space can be seen as a point in a unit (N-1)-sphere, Eq. A.13 can be transformed into a rotational diffusion equation, which for the 2-dimensional case becomes

$$\frac{\partial P(\phi,t)}{\partial t} = D_r \nabla^2 P(\phi,t). \tag{A.39}$$

Notice how D_r is now a rotational diffusion coefficient, whose value—and units—is different from that of D in the original equation.

The solution to Eq. A.39 can be expressed in terms of spherical harmonics $Y_{lm}(\theta,\phi)$:

$$P(\phi,t) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} A_{lm}(t) Y_{lm}(\theta,\phi)\Big|_{\theta=\pi/2}, \tag{A.40}$$

whose non-zero values are obtained when $Y_{lm}(\theta,\phi) \not\propto \cos(\theta)$, that is, $m = \pm l$, so A.40 is simplified into

$$P(\phi,t) = \sum_{m=-\infty}^{\infty} A_m(t) e^{im\phi}. \tag{A.41}$$

Substituting this result into Eq. A.39 leads to the expression

$$\sum_{m=-\infty}^{\infty} \frac{\partial A_m(t)}{\partial t} e^{im\phi} = D_r \sum_{m=-\infty}^{\infty} (-m^2) A_m(t) e^{im\phi}, \tag{A.42}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

which, due to the orthonormality of the spherical harmonics, implies that

$$\frac{\partial A_m(t)}{\partial t} = -D_r m^2 A_m(t), \ \forall m \in (-\infty, \infty), \tag{A.43}$$

so

$$A_m(t) = a_m e^{-D_r m^2 t}. \tag{A.44}$$

Therefore, the probability distribution function becomes

$$P(\phi, t) = \sum_{m=-\infty}^{\infty} a_m e^{-D_r m^2 t} e^{im\phi}, \tag{A.45}$$

which looks like the Fourier expansion of $P(\phi, t)$, where the coefficients of the Fourier series are

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\phi, t) e^{-im\phi} d\phi = a_m e^{-D_r m^2 t}. \tag{A.46}$$

If at time $t = 0$, the orientation of the vector corresponds to $\phi = 0$, then $P(\phi, 0) = \delta(\phi)$, so

$$\begin{aligned} a_m &= \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\phi, 0) e^{-im\phi} d\phi \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \delta(\phi) e^{-im\phi} d\phi \\ &= \frac{1}{2\pi}. \end{aligned} \tag{A.47}$$

Thus, the probability density function associated with the diffusive orientation of a vector in 2 dimensions is

$$\boxed{P(\phi, t) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im\phi} e^{-m^2 D_r t}} \tag{A.48}$$

Short time approximation As proven by Jain and Sebastian,[285] Eq. A.48 can be rewritten in terms of the Jacobi theta function

$$\vartheta_3(z, q) = \sum_{m=-\infty}^{\infty} q^{m^2} e^{i2mz}, \tag{A.49}$$

so

$$P(\phi, t) = \frac{1}{2\pi} \vartheta_3 \left(\frac{\phi}{2}, e^{-D_r t} \right). \tag{A.50}$$

Using the identity

$$\vartheta_3(z, q) = \frac{\sqrt{i\pi} e^{z^2 / \ln q}}{\sqrt{-i \ln q}} \vartheta_3 \left(\frac{i\pi z}{\ln q}, e^{\pi^2 / \ln q} \right), \tag{A.51}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

then the previous expression is transformed into

$$\begin{aligned} P(\phi,t) &= \frac{1}{2\pi}\vartheta_3\left(\frac{\phi}{2}, e^{-D_r t}\right) \\ &= \frac{1}{2\pi}\frac{\sqrt{i\pi}}{\sqrt{iD_r T}} e^{-\phi^2/(4D_r t)}\vartheta_3\left(\frac{-i\pi\phi}{2D_r t}, e^{-\pi^2/(D_r T)}\right) \end{aligned} \tag{A.52}$$

In the short-time regime, where $D_r t \ll 1$, and, consequently, $|\phi| \ll \pi$,

$$\begin{aligned} \vartheta_3\left(\frac{-i\pi\phi}{2D_r t}, e^{-\pi^2/(D_r T)}\right) &= \sum_{m=-\infty}^{\infty} e^{-m^2\pi^2/(D_r t)} e^{m\pi\phi/(D_r t)} \\ &= \sum_{m=-\infty}^{\infty} e^{m\pi(\phi - m\pi)/(D_r t)} \\ &= 1 + \sum_{\substack{m=-\infty \\ m\neq 0}}^{\infty} e^{m\pi(\phi - m\pi)/(D_r t)} \approx 1. \end{aligned} \tag{A.53}$$

So, expression A.52 becomes, at short time windows,

$$\boxed{P(\phi,t) \approx \frac{1}{\sqrt{4\pi D_r t}} e^{-\phi^2/(4D_r t)}}, \tag{A.54}$$

which is a Gaussian distribution of mean $\mu = 0$ and variance $\sigma^2 = 2D_r t$.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix B.

Supplementary data for the structural characterization of membrane-targeting antimicrobial peptides at the lipid membrane–water interface

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.1 Correlation between helicity metrics

The variation of the helicity metrics in the shallow configuration can be exploited to investigate the correlation between the different helicity metrics.

Correlation between αRMSD and α-block The representation of the instantaneous values of the helicity given by both continuous metrics shows a clear linear relation between them (Figure B.1). Such an observation is further confirmed by the high values of both Pearson's and Kendall's rank correlation coefficients (Tables B.1 and B.2); their values being close to 1 indicate a strong, monotonic, and positive linear correlation between the metrics. Thus, an increase in the helicity given by α-block should correspond to an increment in that of αRMSD.

The distribution of the values of helicity is affected by the sequence: For octa-2, states of helicity close to 0.5 according to α-block may correspond to almost any value in αRMSD. On top of that, the relation between the magnitudes differs close to the limits of the metrics.

The helicity values of all simulations, except those of octa-2, are fitted to a linear model

$$y = a + bx, \tag{B.1}$$

where the helicity given by αRMSD has been selected as the independent variable. The range of data points used is restricted to be within 0.25 and 0.75 for α-block, as in this interval the trend is clearly linear. As the fits of the data for each combination of peptide and force field yield similar results (not shown), a single linear model is constructed from all the data (Figure B.4, upper left). The relation between the metrics is estimated to be

$$\alpha_{\text{RMSD}} = -0.349(37) + 1.593(72)\alpha_{\text{block}} \qquad r^2 = 0.967,\ \tau = 0.840 \tag{B.2}$$

where the parentheses indicate the error in each of the coefficients, estimated through the bootstrap of the residuals. This model indicates that the coincidence between the values of the helicity occurs in around 0.5; above this value, α-block yields lower estimates of the helicity than αRMSD, and vice versa.

Despite the linear model is able to capture the relation between the metrics for most of their domain, the behavior near the limits is clearly conditioned by the values of α-block. This metric seems to fail to assign the higher and lower values of helicity to peptide structures.

Correlation between αRMSD for the whole backbone and only C_α atoms The original proposal of αRMSD includes the calculation of the RMSD between all elements within the backbone. This is more expensive in terms of computation time

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

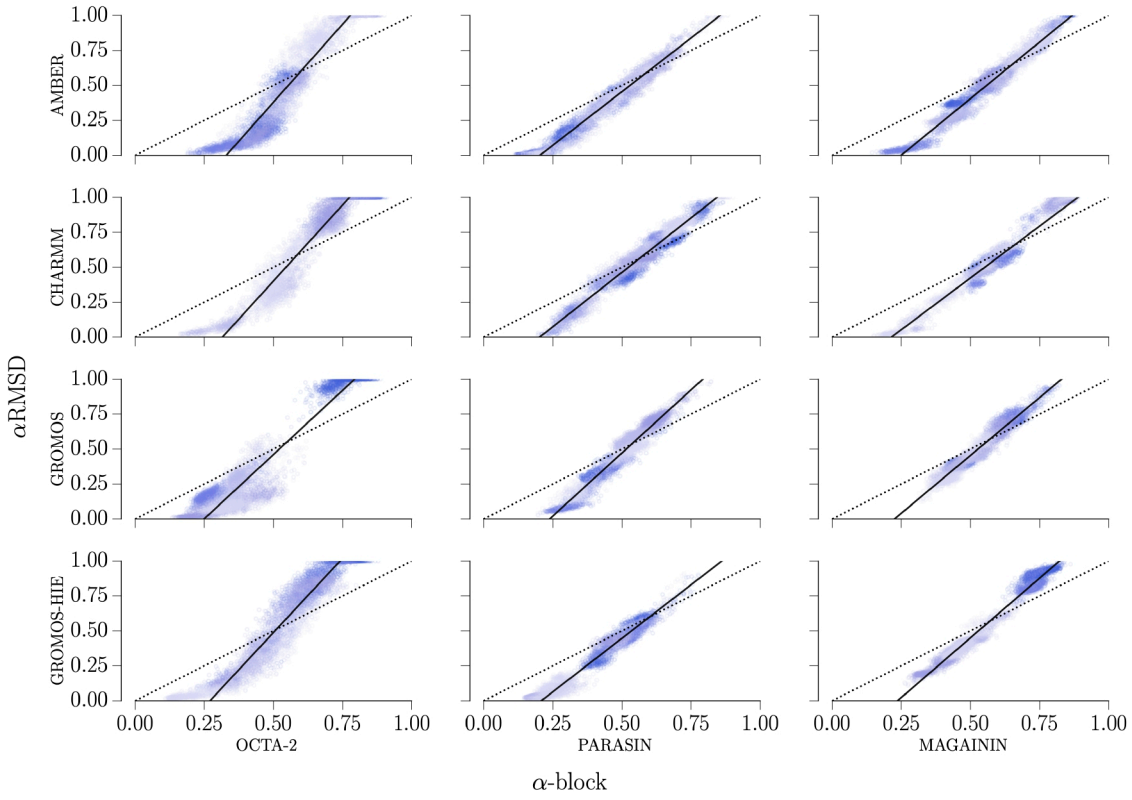


Figure B.1: Relation between the helicity of the peptide calculated with the αRMSD and α-block methods in simulations starting from the shallow configuration. The plots are arranged by force field and peptide. The saturation reflects the density of points. The solid black lines correspond to the result of the linear regressions when the helicity given by α-block is restricted to the interval $(0.25, 0.75)$. A 95 % confidence interval for each of the regressions, estimated through the bootstrap of the residues, is represented, but it is not clearly noticeable. The dashed black line corresponds to the 1-to-1 equivalence between the metrics, and is included as a reference.

than doing it for a smaller selection, like using only the C_α atoms as a reference. The equivalence between both approaches was evaluated by analyzing their correlation, which we found to be practically identical (Figure B.4, upper right). The estimated functional form of their relation is

$$\alpha_{\mathrm{RSMD}} = 0.0121(53) + 0.995(98)\alpha^{\mathrm{C}_\alpha}_{\mathrm{RMSD}} \qquad r^2 = 0.997,\ \tau = 0.947 \qquad \text{(B.3)}$$

This confirms that the relevant conformational changes in the peptide are reflected in the rearrangement of the C_α atoms. In consequence, it is not necessary to compute the differences within all atoms in the backbone, as it is sufficient to do it for the C_α atoms only. Nonetheless, all further αRMSD values will correspond to the estimate given for all the backbone.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

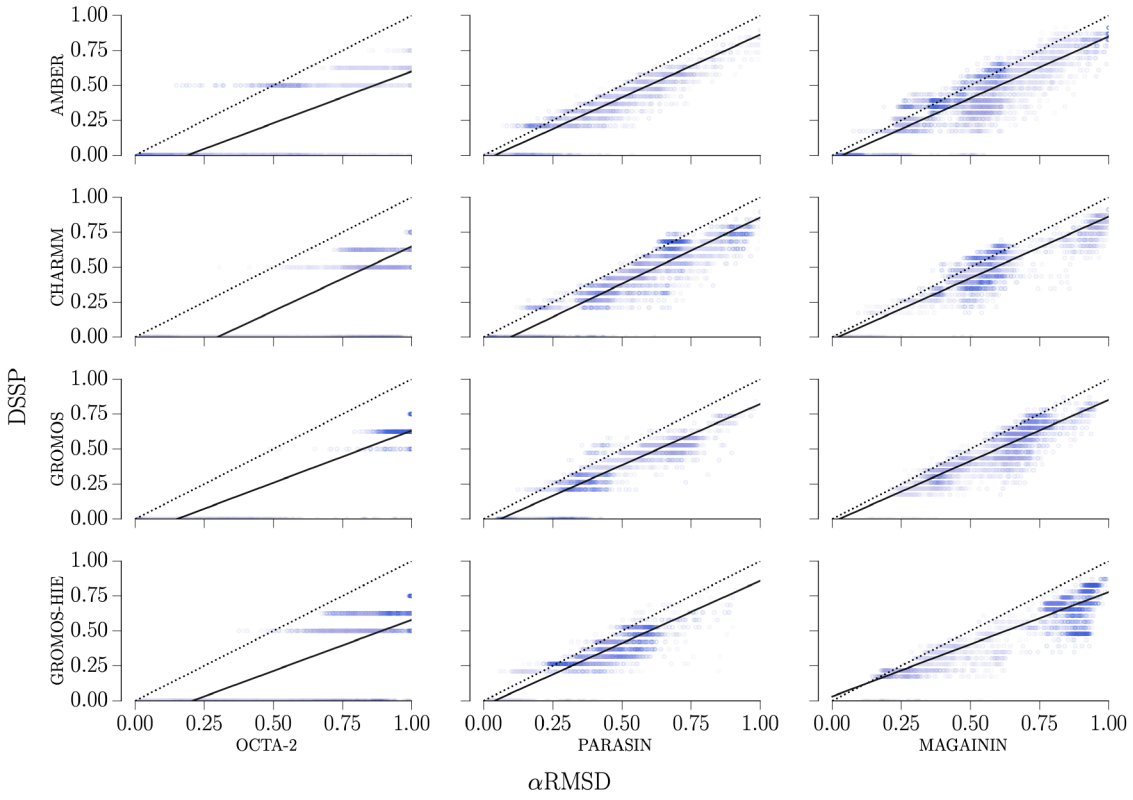


Figure B.2: Relation between the helicity of the peptide calculated with the DSSP and the αRMSD methods in simulations starting from the shallow configuration. The plots are arranged by force field and peptide.

Correlation between DSSP and αRMSD The study of the correlation between DSSP—which can only yield a finite number of discrete helicity values—and any other continuous metric is hindered by the mathematical impossibility of defining a bijection between them.

Nonetheless, the relation between the values of helicity given by DSSP and αRMSD, while weaker than in the previous case, does is also evident (Figure B.2). One of the main limitations of DSSP is clearly manifested in the results for octa-2: a value of the helicity of 0 for DSSP corresponds to very different structural arrangements when αRMSD is used instead. Not only that, the overlap between the ranges of helicity spanned by αRMSD for adjacent values in DSSP is very large. Pearson's correlation coefficients still suggest a linear relation, while the values of Kendall's τ are indicative of a much weaker monotonic correlation (Tables B.3 and B.4). It is still possible to model the data through a linear model, which yields the relation (Figure B.4, lower left)

$$\alpha_{\text{DSSP}} = -0.331(21) + 0.873(41)\alpha_{\text{RMSD}} \qquad r^2 = 0.945,\ \tau = 0.811 \qquad \text{(B.4)}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

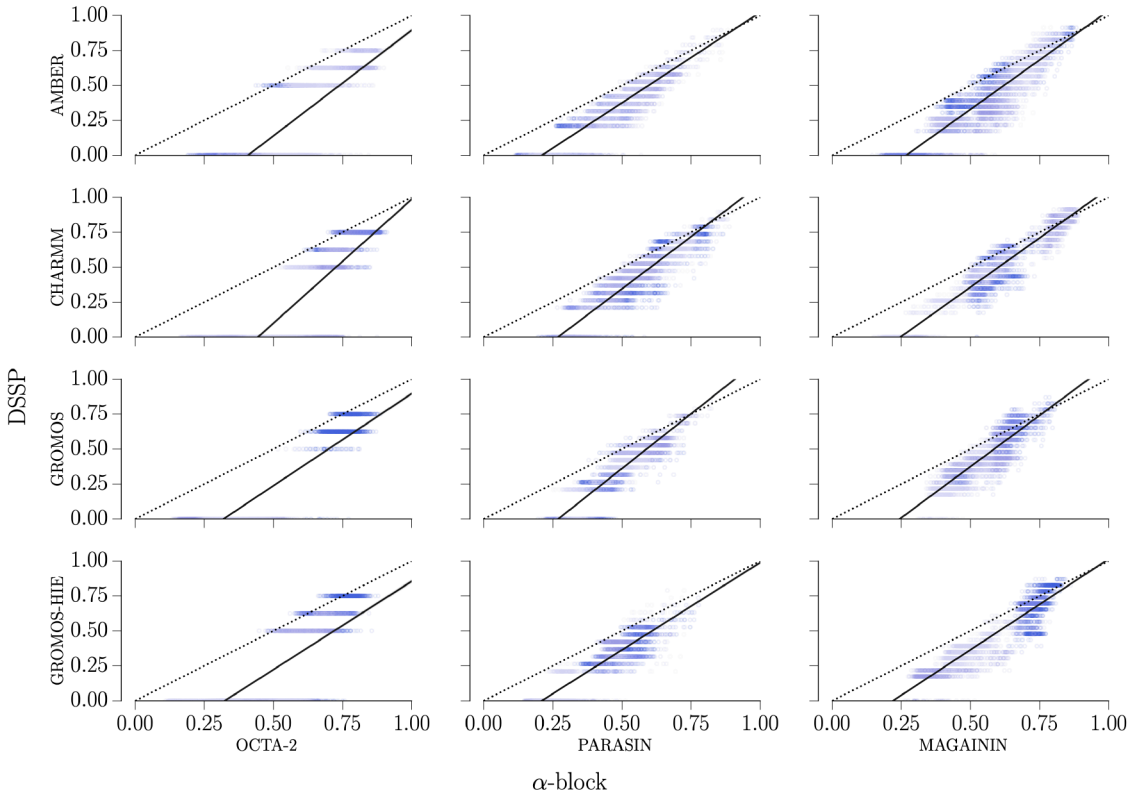


Figure B.3: Relation between the helicity of the peptide calculated with the DSSP and the α-block methods in simulations starting from the shallow configuration. The plots are arranged by force field and peptide.

Correlation between DSSP and α-block In the comparison between the helicities given for the same structures by DSSP and α-block, the same issues as in the comparison with αRMSD are observed. Wide ranges of helicity given by α-block are compatible with the same estimation obtained by DSSP, and the overlap between these intervals for adjacent values of DSSP is very large.

The values of both the Pearson correlation coefficient and Kendall τ (Tables B.5 and B.6) are not that different from the previous case. The linear relation obtained after fitting the simulation data of parasin and magainin is (Figure B.4, lower right)

$$\alpha_{\mathrm{DSSP}} = -0.313(36) + 1.350(65)\alpha_{\mathrm{block}} \qquad r^2 = 0.936,\ \tau = 0.789 \qquad \text{(B.5)}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

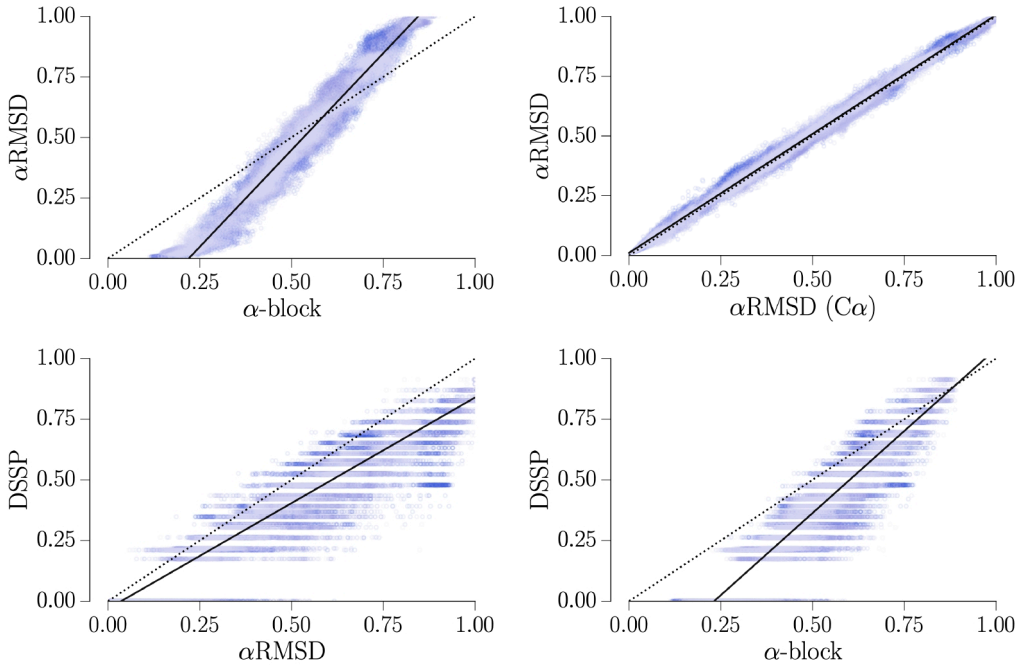


Figure B.4: Linear relations between the different helicity metrics in simulations starting from the shallow configuration. The panels correspond to the relation between αRMSD and α-block (upper left), between the αRMSD of the whole backbone and only the C_α atoms (upper right), DSSP and αRMSD (lower left), and DSSP and α-block (lower right). The plots contain no distinction between peptides or force fields. Octa-2 simulation data is not included. The solid black lines correspond to the results of the linear regressions. A 95 % confidence interval for each of the regressions, estimated through the bootstrap of the residues, is represented, but it is not clearly noticeable. The dashed black line represents the 1-to-1 correspondence between the metrics, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.665	0.729	0.861	0.861	0.716	0.899
	Parasin	0.954	0.924	0.920	0.804	0.965	0.966
	Magainin	0.946	0.972	0.602	0.963	0.833	0.962
C36m	Octa-2	0.506	0.028	0.503	0.887	0.618	0.926
	Parasin	0.948	0.974	0.706	0.954	0.766	0.955
	Magainin	0.729	0.836	0.471	0.982	0.985	0.976
G54a7	Octa-2	0.477	0.616	0.796	0.502	0.554	0.898
	Parasin	0.845	0.929	0.779	0.664	0.920	0.944
	Magainin	0.939	0.765	0.883	0.777	0.933	0.943
G54a7-HIE	Octa-2	0.764	0.888	0.933	0.874	0.857	0.912
	Parasin	0.901	0.858	0.692	0.803	0.899	0.956
	Magainin	0.574	0.363	0.941	0.807	0.863	0.975

Table B.1: Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and α-block. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.588	0.668	0.752	0.721	0.666	0.841
	Parasin	0.827	0.783	0.825	0.677	0.883	0.888
	Magainin	0.794	0.899	0.444	0.803	0.746	0.871
C36m	Octa-2	0.580	0.370	0.479	0.844	0.542	0.830
	Parasin	0.812	0.830	0.645	0.869	0.573	0.854
	Magainin	0.524	0.679	0.386	0.823	0.906	0.843
G54a7	Octa-2	0.595	0.470	0.726	0.452	0.481	0.622
	Parasin	0.603	0.818	0.588	0.609	0.788	0.854
	Magainin	0.526	0.620	0.744	0.698	0.759	0.832
G54a7-HIE	Octa-2	0.738	0.763	0.792	0.760	0.793	0.852
	Parasin	0.789	0.775	0.411	0.632	0.528	0.817
	Magainin	0.578	0.478	0.741	0.725	0.724	0.858

Table B.2: Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and α-block. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.438	0.599	0.100	0.688	0.582	0.695
	Parasin	0.850	0.883	0.839	0.647	0.919	0.928
	Magainin	0.813	0.935	0.303	0.868	0.519	0.821
C36m	Octa-2	0.603	0.051	0.149	0.307	0.356	0.720
	Parasin	0.718	0.891	0.520	0.884	0.522	0.887
	Magainin	0.382	0.526	0.492	0.952	0.972	0.931
G54a7	Octa-2	0.267	0.000	0.000	0.024	0.057	0.896
	Parasin	0.412	0.747	0.539	0.388	0.843	0.874
	Magainin	0.917	0.370	0.475	0.474	0.798	0.828
G54a7-HIE	Octa-2	0.550	0.591	0.778	0.422	0.114	0.681
	Parasin	0.810	0.504	0.278	0.661	0.785	0.893
	Magainin	0.371	0.272	0.649	0.739	0.436	0.847

Table B.3: Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and αRMSD. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.498	0.204	0.271	0.608	0.611	0.677
	Parasin	0.768	0.601	0.795	0.680	0.824	0.845
	Magainin	0.691	0.801	0.380	0.656	0.548	0.745
C36m	Octa-2	0.381	0.173	0.318	0.380	0.407	0.754
	Parasin	0.676	0.788	0.576	0.790	0.471	0.806
	Magainin	0.369	0.579	0.499	0.740	0.879	0.775
G54a7	Octa-2	0.526	—	—	0.070	0.063	0.572
	Parasin	0.350	0.674	0.417	0.534	0.783	0.821
	Magainin	0.647	0.501	0.555	0.582	0.654	0.789
G54a7-HIE	Octa-2	0.690	0.644	0.668	0.475	0.135	0.705
	Parasin	0.758	0.508	0.264	0.339	0.630	0.756
	Magainin	0.508	0.461	0.565	0.687	0.544	0.771

Table B.4: Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and αRMSD. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas. Missing values could not be calculated.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.347	0.284	0.131	0.676	0.447	0.637
	Parasin	0.790	0.804	0.792	0.609	0.870	0.887
	Magainin	0.800	0.903	0.353	0.845	0.531	0.810
C36m	Octa-2	0.395	0.012	0.221	0.223	0.357	0.724
	Parasin	0.730	0.882	0.435	0.861	0.499	0.876
	Magainin	0.364	0.573	0.354	0.951	0.963	0.929
G54a7	Octa-2	0.222	0.000	0.000	0.045	0.032	0.828
	Parasin	0.369	0.705	0.499	0.319	0.760	0.822
	Magainin	0.879	0.413	0.422	0.391	0.755	0.799
G54a7-HIE	Octa-2	0.586	0.528	0.763	0.365	0.044	0.581
	Parasin	0.740	0.426	0.276	0.482	0.693	0.845
	Magainin	0.583	0.348	0.607	0.628	0.530	0.869

Table B.5: Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and α-block. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.482	0.204	0.320	0.572	0.601	0.659
	Parasin	0.719	0.594	0.766	0.643	0.799	0.818
	Magainin	0.678	0.796	0.397	0.625	0.560	0.737
C36m	Octa-2	0.369	0.111	0.405	0.377	0.413	0.752
	Parasin	0.686	0.772	0.501	0.791	0.466	0.802
	Magainin	0.425	0.573	0.369	0.735	0.869	0.779
G54a7	Octa-2	0.403	—	—	0.073	0.063	0.566
	Parasin	0.329	0.692	0.407	0.494	0.721	0.778
	Magainin	0.554	0.532	0.517	0.526	0.636	0.749
G54a7-HIE	Octa-2	0.647	0.643	0.690	0.441	0.131	0.688
	Parasin	0.710	0.505	0.319	0.320	0.530	0.701
	Magainin	0.616	0.454	0.572	0.622	0.556	0.789

Table B.6: Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and α-block. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas. Missing values could not be calculated.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.2 Simulations starting from the shallow configuration

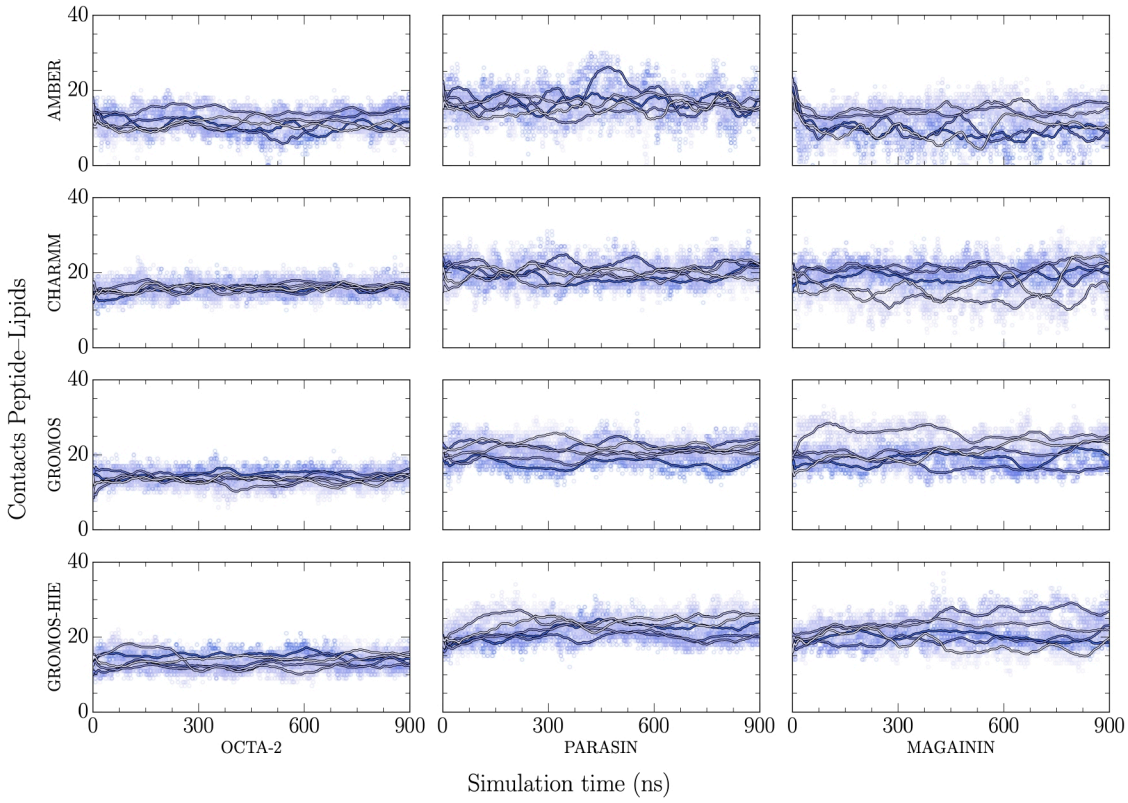


Figure B.5: Time evolution of the number of contacts between the peptide and the lipids in simulations starting from the shallow configuration. Each panel shows the variation of the number of contacts (Y-axis) over the simulation time (X-axis). The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the number of contacts in the five independent simulation replicas, calculated as a rolling average of the instantaneous values (displayed as faint points).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

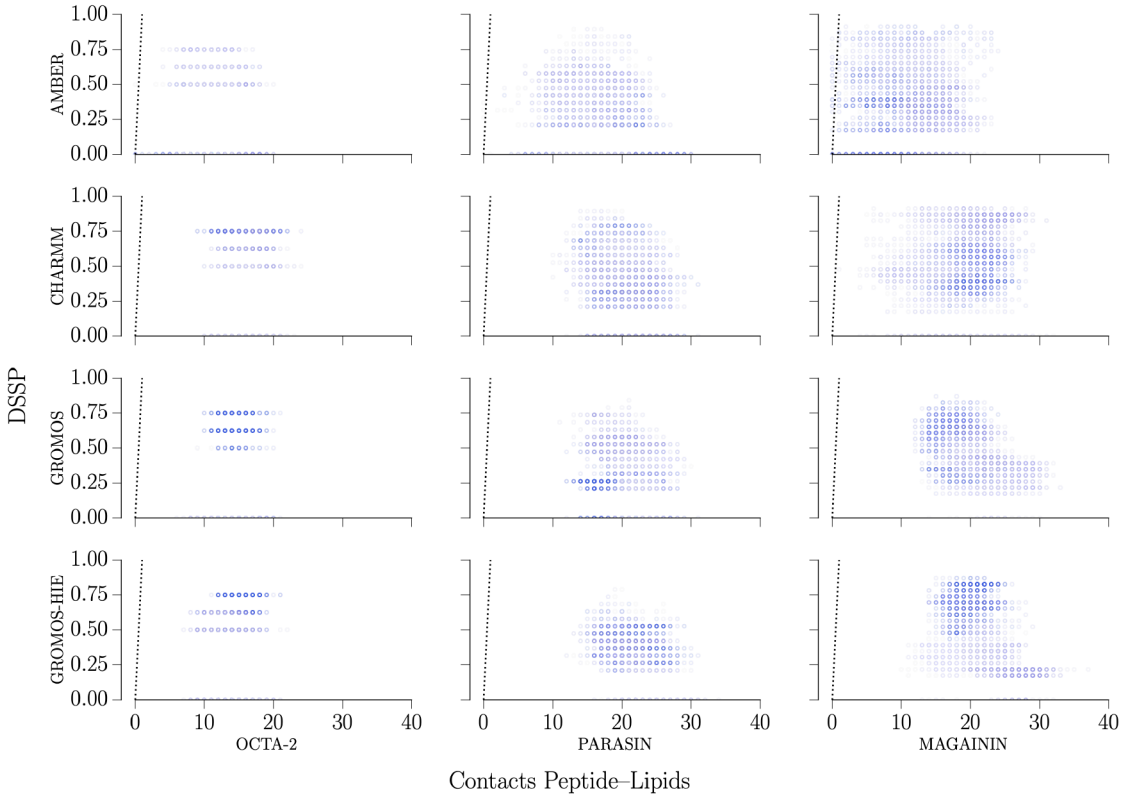


Figure B.6: Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–lipid contacts in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid contacts (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

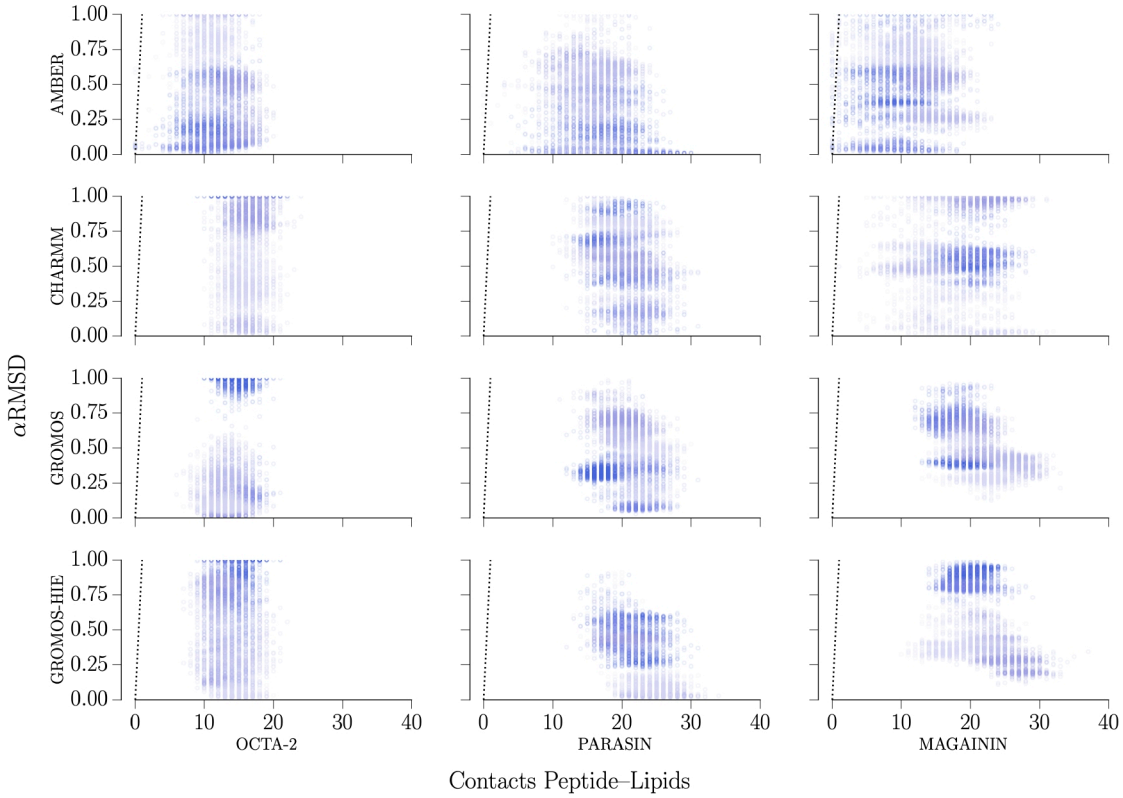


Figure B.7: Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–lipid contacts in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid contacts (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

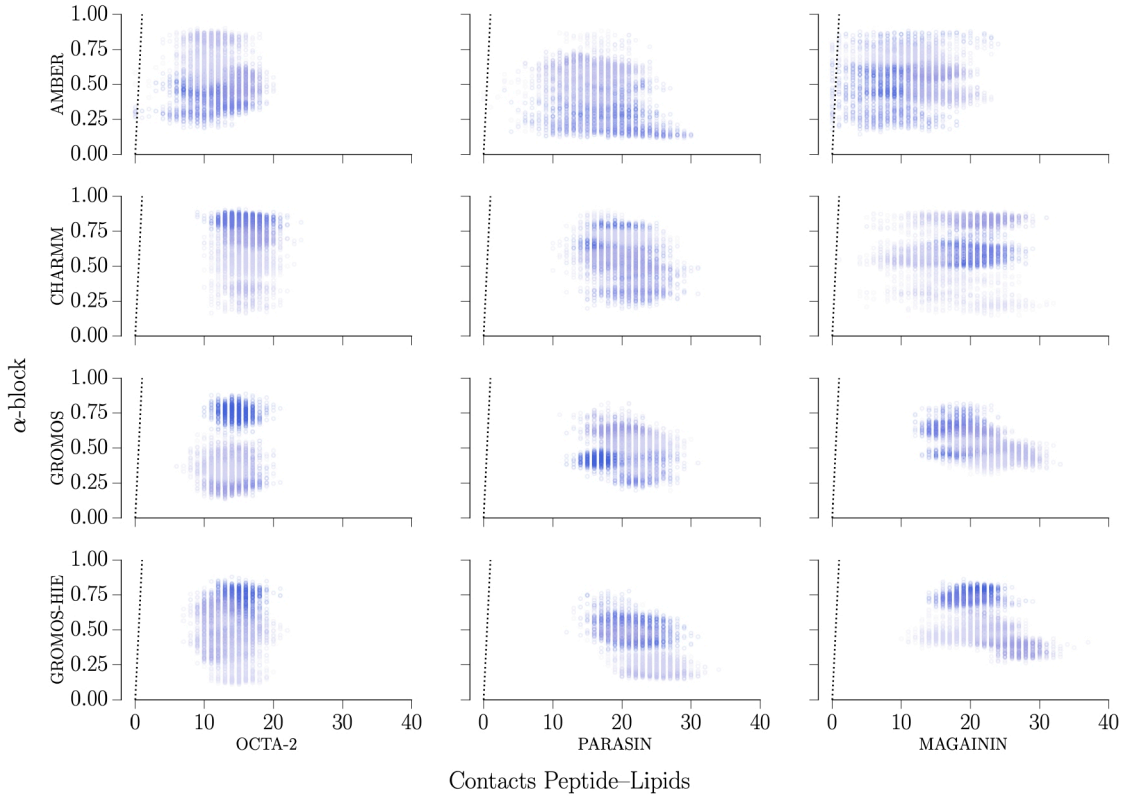


Figure B.8: Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–lipid contacts in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid contacts (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.005	0.016	0.004	0.014	0.058	0.013
	Parasin	0.000	0.013	0.083	0.038	0.006	0.017
	Magainin	0.029	0.028	0.010	0.025	0.007	0.006
C36m	Octa-2	0.003	0.002	0.005	0.015	0.000	0.003
	Parasin	0.051	0.040	0.093	0.222	0.068	0.107
	Magainin	0.001	0.000	0.000	0.015	0.215	0.024
G54a7	Octa-2	0.000	0.000	0.000	0.000	0.000	0.038
	Parasin	0.004	0.001	0.042	0.014	0.030	0.013
	Magainin	0.157	0.035	0.044	0.051	0.312	0.259
G54a7-HIE	Octa-2	0.001	0.018	0.003	0.016	0.014	0.000
	Parasin	0.055	0.009	0.020	0.074	0.128	0.319
	Magainin	0.099	0.016	0.171	0.019	0.061	0.164

Table B.7: Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.059	0.110	-0.061	0.095	0.178	-0.104
	Parasin	-0.027	-0.101	-0.210	0.197	-0.065	-0.070
	Magainin	0.068	0.125	0.092	-0.102	0.081	-0.103
C36m	Octa-2	-0.061	0.027	0.074	-0.106	-0.018	-0.068
	Parasin	-0.097	-0.151	-0.206	-0.339	-0.172	-0.228
	Magainin	0.053	0.005	0.005	0.128	-0.341	0.122
G54a7	Octa-2	-0.039	—	—	0.013	-0.001	0.169
	Parasin	0.050	0.026	-0.089	0.094	0.175	-0.080
	Magainin	-0.264	0.172	-0.136	-0.203	-0.424	-0.367
G54a7-HIE	Octa-2	-0.006	-0.085	-0.029	-0.107	-0.075	0.023
	Parasin	-0.168	-0.067	-0.091	-0.170	-0.241	-0.431
	Magainin	0.255	-0.162	-0.388	0.064	0.214	-0.314

Table B.8: Values of Kendall's rank correlation coefficient τ fbetween the values of helicity given by DSSP and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.035	0.014	0.034	0.029	0.075	0.005
	Parasin	0.001	0.022	0.084	0.034	0.008	0.024
	Magainin	0.031	0.026	0.009	0.028	0.039	0.000
C36m	Octa-2	0.010	0.001	0.002	0.039	0.000	0.001
	Parasin	0.011	0.033	0.144	0.231	0.147	0.092
	Magainin	0.023	0.003	0.005	0.027	0.230	0.028
G54a7	Octa-2	0.016	0.001	0.003	0.031	0.000	0.037
	Parasin	0.025	0.000	0.111	0.006	0.038	0.020
	Magainin	0.182	0.076	0.098	0.229	0.355	0.304
G54a7-HIE	Octa-2	0.004	0.021	0.003	0.065	0.158	0.004
	Parasin	0.083	0.002	0.067	0.061	0.228	0.359
	Magainin	0.059	0.036	0.283	0.035	0.101	0.177

Table B.9: Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.111	0.149	-0.132	0.140	0.191	-0.056
	Parasin	-0.015	-0.257	-0.201	0.180	-0.071	-0.093
	Magainin	0.056	0.096	-0.021	-0.137	0.136	-0.048
C36m	Octa-2	-0.106	-0.013	0.031	-0.141	0.018	-0.056
	Parasin	-0.042	-0.089	-0.210	-0.306	-0.245	-0.195
	Magainin	-0.065	0.023	0.019	0.105	-0.335	0.147
G54a7	Octa-2	-0.115	-0.068	0.056	0.138	-0.035	0.111
	Parasin	0.130	0.007	-0.207	0.056	0.153	-0.101
	Magainin	-0.224	0.212	-0.194	-0.321	-0.385	-0.380
G54a7-HIE	Octa-2	-0.013	-0.126	-0.029	-0.166	-0.307	-0.023
	Parasin	-0.176	-0.025	-0.104	-0.043	-0.258	-0.407
	Magainin	0.215	-0.137	-0.378	0.062	0.285	-0.286

Table B.10: Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.017	0.032	0.020	0.033	0.060	0.001
	Parasin	0.001	0.048	0.059	0.046	0.006	0.028
	Magainin	0.022	0.025	0.004	0.034	0.022	0.000
C36m	Octa-2	0.023	0.007	0.000	0.027	0.001	0.004
	Parasin	0.011	0.033	0.119	0.194	0.121	0.100
	Magainin	0.007	0.000	0.008	0.022	0.242	0.024
G54a7	Octa-2	0.011	0.000	0.010	0.028	0.015	0.020
	Parasin	0.012	0.000	0.085	0.005	0.026	0.011
	Magainin	0.199	0.109	0.098	0.222	0.341	0.286
G54a7-HIE	Octa-2	0.005	0.013	0.003	0.042	0.174	0.005
	Parasin	0.118	0.002	0.087	0.033	0.271	0.375
	Magainin	0.135	0.024	0.343	0.009	0.090	0.182

Table B.11: Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.084	0.126	-0.113	0.124	0.167	-0.041
	Parasin	-0.013	-0.263	-0.167	0.176	-0.063	-0.087
	Magainin	0.038	0.097	0.002	-0.144	0.118	-0.049
C36m	Octa-2	-0.108	-0.050	-0.013	-0.130	-0.003	-0.067
	Parasin	-0.050	-0.098	-0.219	-0.295	-0.206	-0.204
	Magainin	0.032	0.007	0.063	0.110	-0.339	0.155
G54a7	Octa-2	-0.095	-0.016	0.075	0.121	-0.096	0.040
	Parasin	0.104	0.020	-0.165	0.049	0.125	-0.069
	Magainin	-0.304	0.234	-0.172	-0.323	-0.388	-0.377
G54a7-HIE	Octa-2	-0.048	-0.073	-0.029	-0.107	-0.275	-0.026
	Parasin	-0.215	-0.029	-0.134	-0.043	-0.336	-0.426
	Magainin	0.283	-0.099	-0.440	0.056	0.243	-0.290

Table B.12: Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of contacts with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

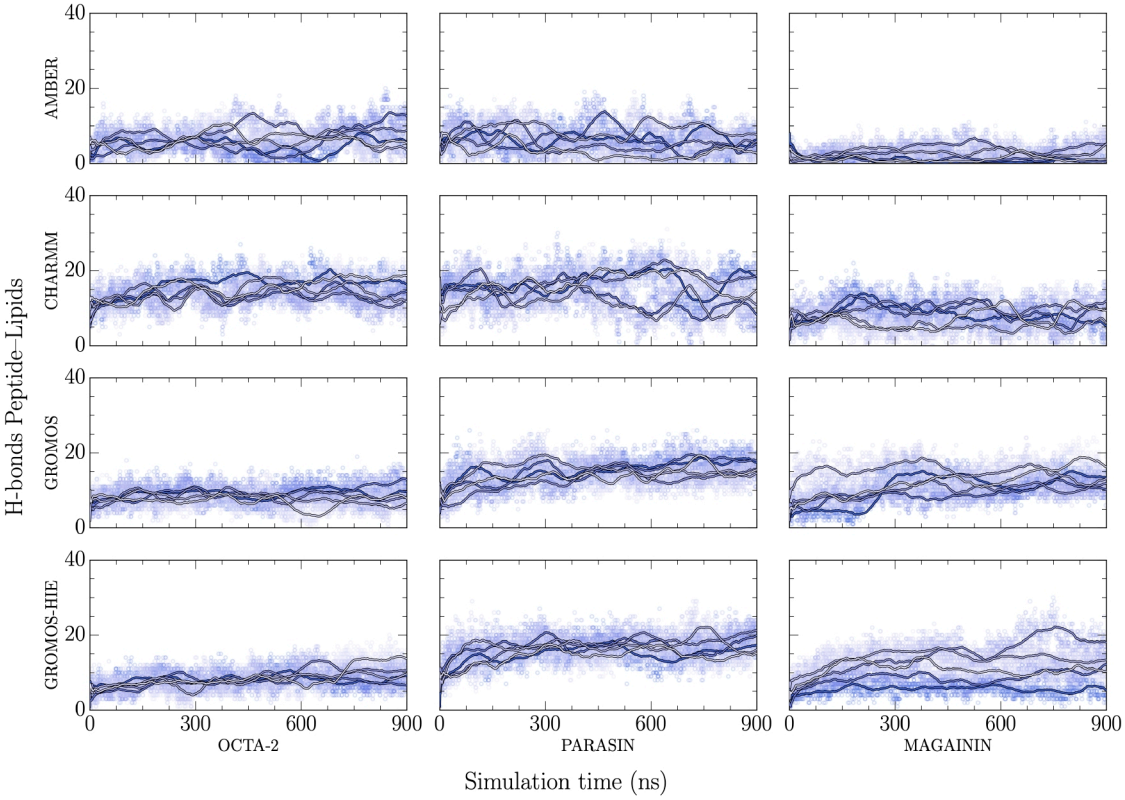


Figure B.9: Time evolution of the number of peptide–lipid H-bonds in simulations starting from the shallow configuration. Each panel shows the variation of the number of H-bonds (Y-axis) over the simulation time (X-axis). The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the number of H-bonds in the five independent simulation replicas, calculated as a rolling average of the instantaneous values (displayed as faint points).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

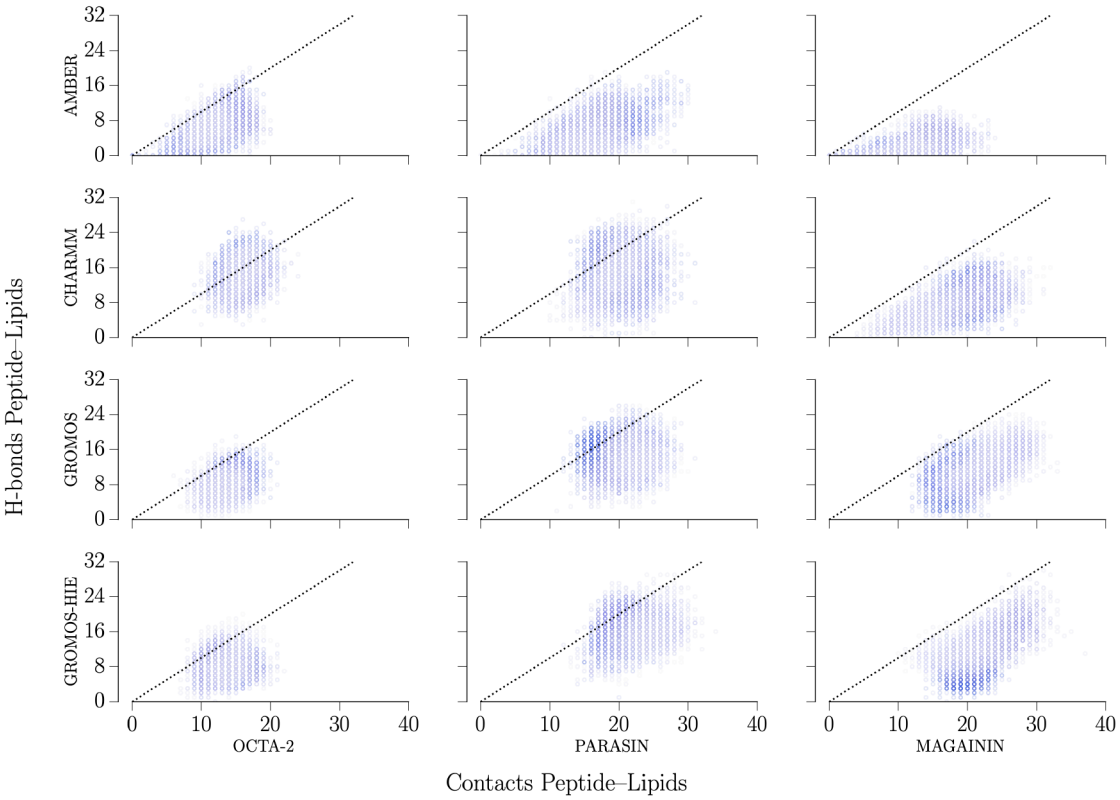


Figure B.10: Relation between the peptide–lipid H-bonds and contacts in simulations starting from the shallow configuration. Each panel shows the relation between H-bonds (Y-axis) and the number of contacts (X-axis) formed between peptide and lipids. The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.214	0.474	0.002	0.108	0.183	0.303
	Parasin	0.259	0.430	0.235	0.092	0.043	0.254
	Magainin	0.149	0.188	0.105	0.067	0.190	0.326
C36m	Octa-2	0.104	0.042	0.055	0.142	0.081	0.058
	Parasin	0.047	0.026	0.010	0.089	0.199	0.021
	Magainin	0.016	0.202	0.074	0.321	0.421	0.358
G54a7	Octa-2	0.013	0.032	0.106	0.000	0.046	0.067
	Parasin	0.030	0.003	0.126	0.030	0.023	0.011
	Magainin	0.288	0.009	0.064	0.145	0.336	0.336
G54a7-HIE	Octa-2	0.001	0.014	0.055	0.001	0.160	0.017
	Parasin	0.157	0.058	0.050	0.040	0.170	0.022
	Magainin	0.033	0.096	0.383	0.148	0.101	0.341

Table B.13: Values of Pearson's correlation coefficient r^2 between the peptide–lipid H-bonds and contacts. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.353	0.558	-0.029	0.247	0.302	0.420
	Parasin	0.369	0.483	0.366	0.235	0.144	0.359
	Magainin	0.280	0.341	0.272	0.199	0.366	0.491
C36m	Octa-2	0.225	0.159	0.163	0.275	0.204	0.172
	Parasin	0.160	0.092	0.056	0.213	0.334	0.100
	Magainin	0.090	0.316	0.200	0.419	0.490	0.445
G54a7	Octa-2	0.070	0.129	0.246	-0.024	0.153	0.184
	Parasin	0.111	0.035	0.249	0.123	0.117	0.067
	Magainin	0.416	-0.065	0.194	0.287	0.440	0.428
G54a7-HIE	Octa-2	0.026	0.080	0.174	0.006	0.296	0.083
	Parasin	0.274	0.174	0.117	0.102	0.269	0.073
	Magainin	0.158	0.194	0.478	0.266	0.226	0.407

Table B.14: Values of Kendall's rank correlation coefficient τ between the peptide–lipid H-bonds and contacts. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

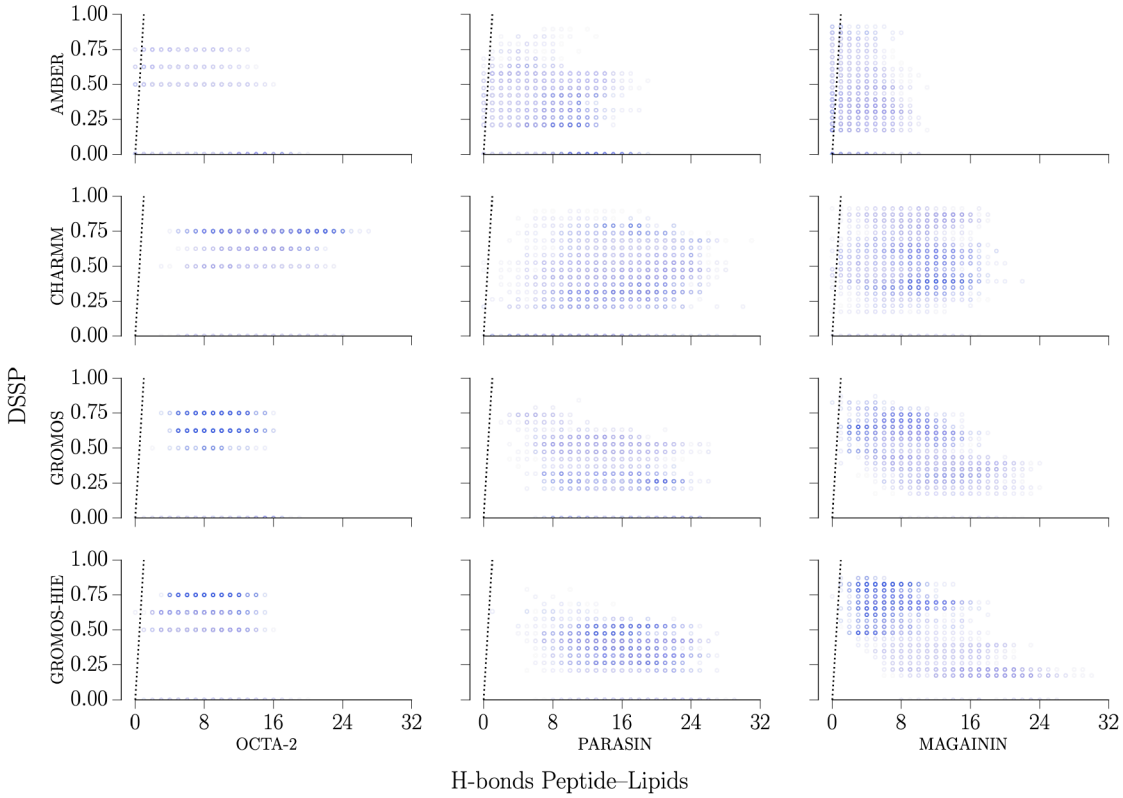


Figure B.11: Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid H-bonds (X-axis). The plots are arranged by force field and peptide. TThe saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

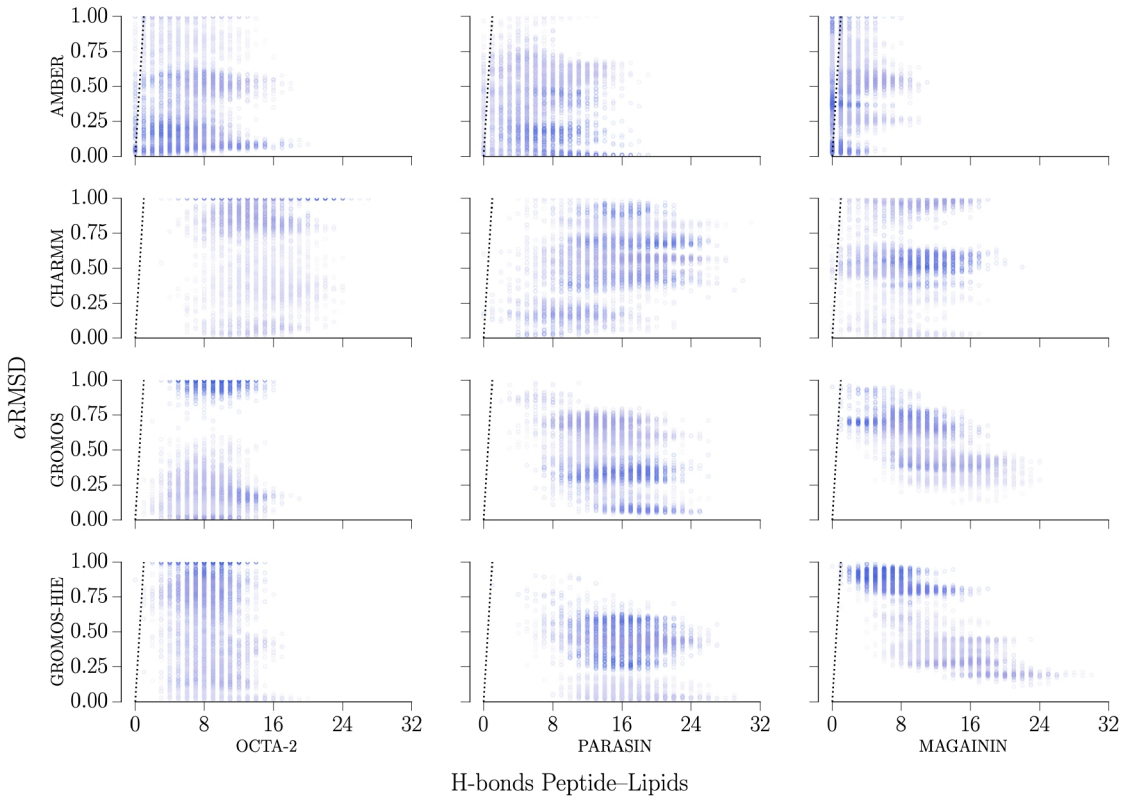


Figure B.12: Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid H-bonds (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

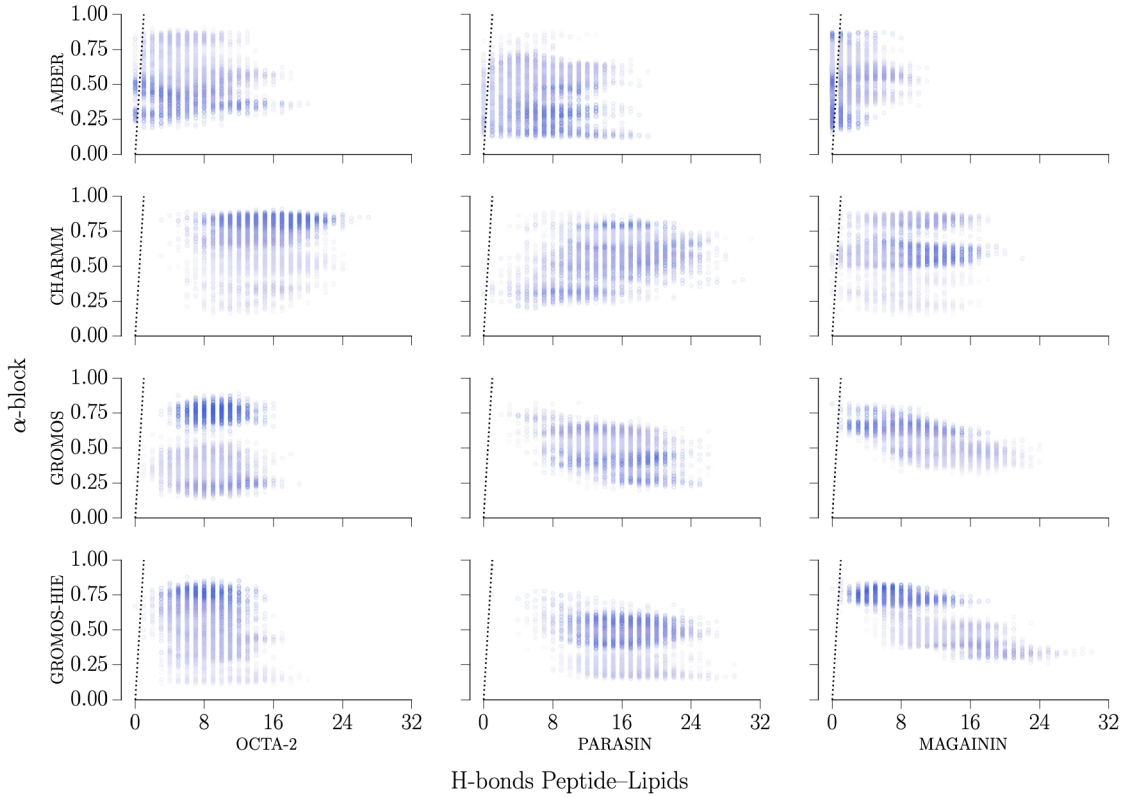


Figure B.13: Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–lipid H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–lipid H-bonds (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.019	0.017	0.000	0.004	0.060	0.002
	Parasin	0.002	0.004	0.001	0.000	0.120	0.032
	Magainin	0.026	0.014	0.060	0.004	0.003	0.014
C36m	Octa-2	0.011	0.003	0.043	0.021	0.005	0.003
	Parasin	0.009	0.275	0.014	0.000	0.048	0.129
	Magainin	0.018	0.000	0.000	0.032	0.102	0.017
G54a7	Octa-2	0.006	0.000	0.000	0.001	0.001	0.020
	Parasin	0.021	0.181	0.134	0.018	0.177	0.097
	Magainin	0.640	0.065	0.064	0.009	0.345	0.321
G54a7-HIE	Octa-2	0.007	0.018	0.311	0.075	0.010	0.032
	Parasin	0.027	0.032	0.000	0.053	0.171	0.003
	Magainin	0.002	0.038	0.298	0.023	0.006	0.433

Table B.15: Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.119	-0.122	0.011	0.072	0.198	-0.025
	Parasin	-0.026	0.050	-0.040	0.036	0.215	0.152
	Magainin	0.112	0.090	-0.148	-0.014	0.042	-0.105
C36m	Octa-2	0.097	0.021	0.180	-0.126	0.059	0.050
	Parasin	0.037	0.434	0.116	0.012	-0.121	0.217
	Magainin	-0.042	0.011	-0.013	0.130	-0.251	0.087
G54a7	Octa-2	0.029	—	—	-0.011	0.024	0.120
	Parasin	-0.104	-0.323	-0.135	0.117	-0.269	-0.201
	Magainin	-0.517	-0.224	-0.190	0.056	-0.458	-0.422
G54a7-HIE	Octa-2	0.039	-0.138	-0.489	-0.213	-0.094	-0.128
	Parasin	-0.086	-0.145	0.019	-0.077	-0.327	-0.007
	Magainin	0.037	-0.221	-0.484	0.034	-0.021	-0.494

Table B.16: Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.052	0.027	0.026	0.001	0.054	0.000
	Parasin	0.000	0.006	0.000	0.004	0.103	0.027
	Magainin	0.023	0.014	0.128	0.004	0.013	0.000
C36m	Octa-2	0.004	0.000	0.006	0.014	0.004	0.000
	Parasin	0.000	0.316	0.004	0.000	0.106	0.112
	Magainin	0.016	0.002	0.000	0.043	0.116	0.019
G54a7	Octa-2	0.000	0.037	0.001	0.003	0.008	0.018
	Parasin	0.033	0.259	0.165	0.005	0.208	0.120
	Magainin	0.691	0.253	0.112	0.029	0.452	0.425
G54a7-HIE	Octa-2	0.026	0.011	0.284	0.092	0.159	0.061
	Parasin	0.053	0.024	0.090	0.097	0.234	0.009
	Magainin	0.000	0.130	0.501	0.041	0.007	0.570

Table B.17: Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.150	0.069	0.091	0.090	0.157	0.010
	Parasin	0.018	-0.128	-0.022	0.053	0.164	0.120
	Magainin	0.084	0.060	-0.192	-0.026	0.063	-0.034
C36m	Octa-2	0.082	0.036	0.042	-0.048	0.035	-0.009
	Parasin	0.015	0.432	0.080	0.011	-0.193	0.177
	Magainin	-0.077	0.003	-0.013	0.117	-0.253	0.108
G54a7	Octa-2	-0.016	-0.062	0.029	0.022	0.033	0.083
	Parasin	0.022	-0.370	-0.161	0.057	-0.280	-0.216
	Magainin	-0.424	-0.312	-0.254	-0.095	-0.483	-0.465
G54a7-HIE	Octa-2	0.045	-0.130	-0.346	-0.207	-0.381	-0.150
	Parasin	-0.124	-0.082	-0.091	-0.099	-0.283	-0.057
	Magainin	0.015	-0.214	-0.523	0.066	0.046	-0.560

Table B.18: Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.020	0.006	0.020	0.011	0.032	0.000
	Parasin	0.000	0.000	0.002	0.002	0.097	0.022
	Magainin	0.017	0.013	0.140	0.005	0.010	0.000
C36m	Octa-2	0.021	0.002	0.023	0.001	0.026	0.000
	Parasin	0.001	0.281	0.037	0.001	0.104	0.134
	Magainin	0.008	0.003	0.001	0.040	0.123	0.017
G54a7	Octa-2	0.002	0.131	0.010	0.002	0.000	0.006
	Parasin	0.035	0.283	0.181	0.003	0.183	0.149
	Magainin	0.676	0.217	0.083	0.021	0.453	0.410
G54a7-HIE	Octa-2	0.007	0.019	0.286	0.098	0.204	0.083
	Parasin	0.067	0.026	0.076	0.094	0.247	0.012
	Magainin	0.004	0.024	0.550	0.003	0.016	0.552

Table B.19: Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.098	0.027	0.088	0.103	0.135	0.014
	Parasin	0.021	-0.146	0.010	0.035	0.171	0.119
	Magainin	0.080	0.066	-0.215	-0.025	0.064	-0.035
C36m	Octa-2	0.102	0.026	0.110	-0.039	0.128	0.015
	Parasin	0.019	0.408	0.101	0.045	-0.168	0.216
	Magainin	-0.035	-0.026	-0.022	0.127	-0.258	0.112
G54a7	Octa-2	0.029	-0.209	0.068	0.027	-0.026	-0.003
	Parasin	-0.001	-0.373	-0.199	-0.023	-0.276	-0.245
	Magainin	-0.441	-0.277	-0.219	-0.083	-0.486	-0.454
G54a7-HIE	Octa-2	0.020	-0.155	-0.371	-0.217	-0.317	-0.171
	Parasin	-0.131	-0.094	-0.038	-0.145	-0.303	-0.064
	Magainin	0.048	-0.072	-0.556	0.096	-0.016	-0.538

Table B.20: Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of H-bonds with lipids. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

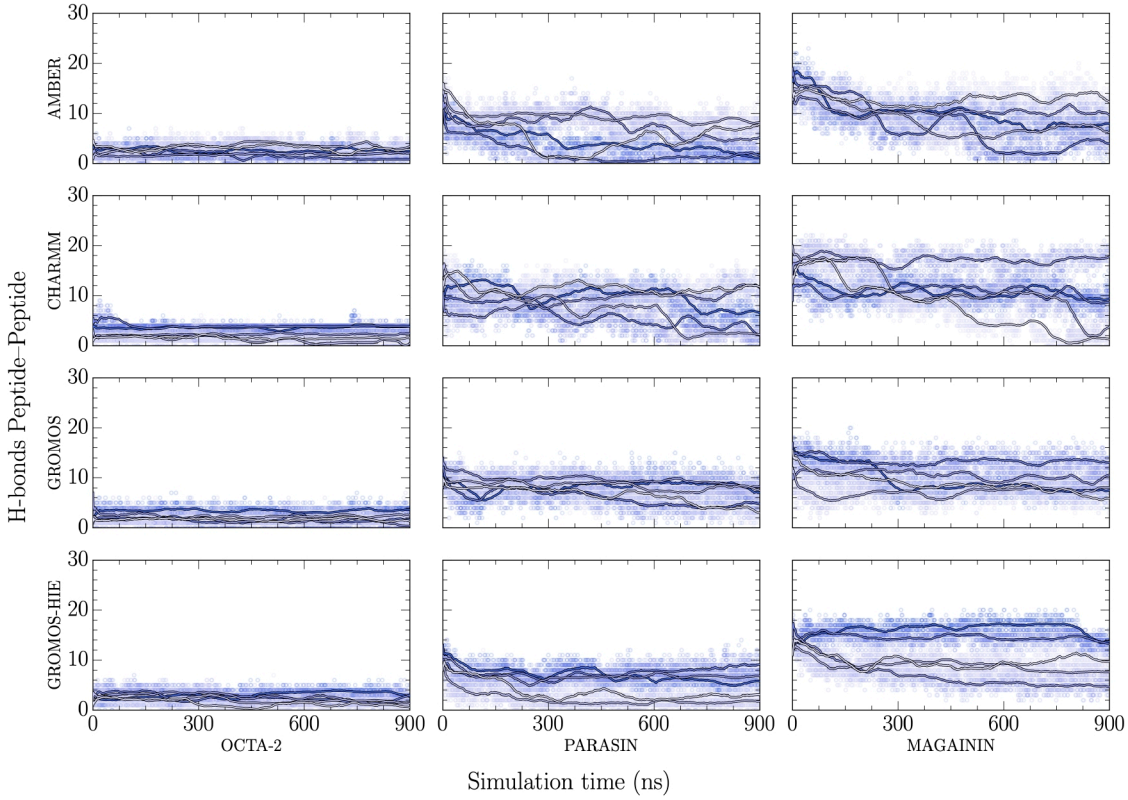


Figure B.14: Time evolution of the number of H-bonds within the peptide in simulations starting from the shallow configuration. Each panel shows the variation of the H-bonds (Y-axis) over the simulation time (X-axis). The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the number of H-bonds in the five independent simulation replicas, calculated as a rolling average of the instantaneous values (displayed as faint points).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

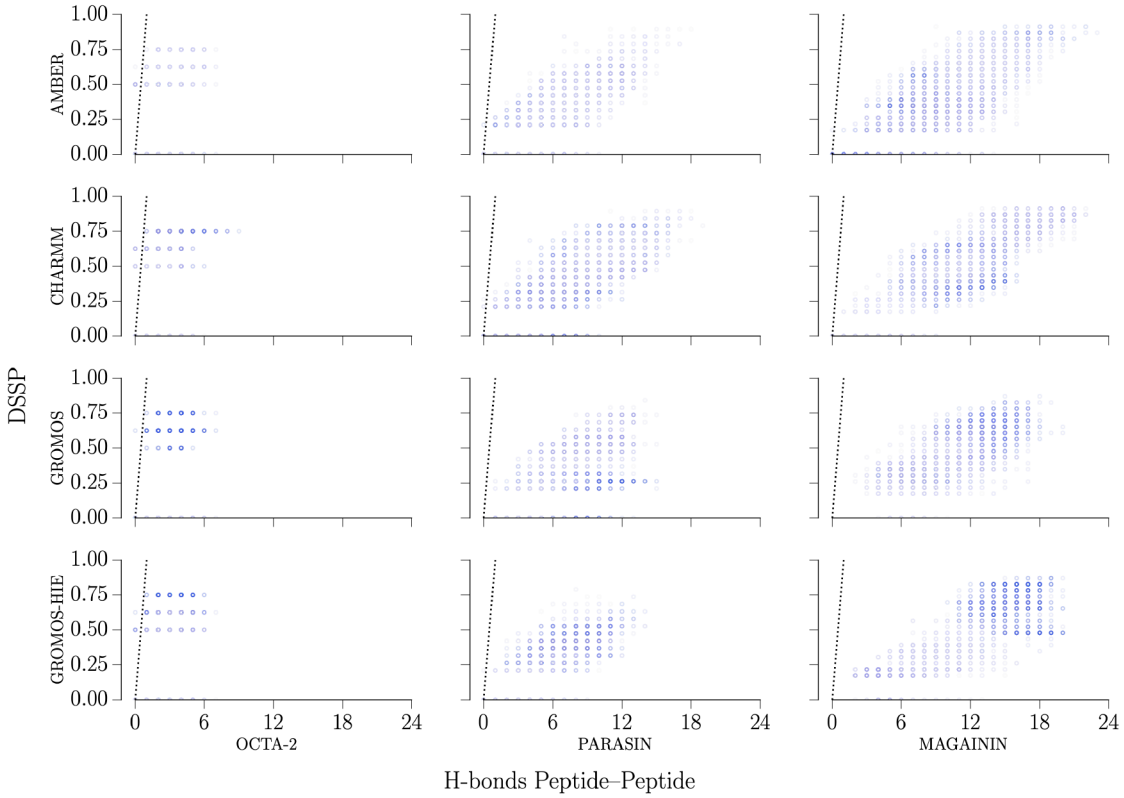


Figure B.15: Relation between the helicity of the peptide calculated with the DSSP method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–peptide H-bonds (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

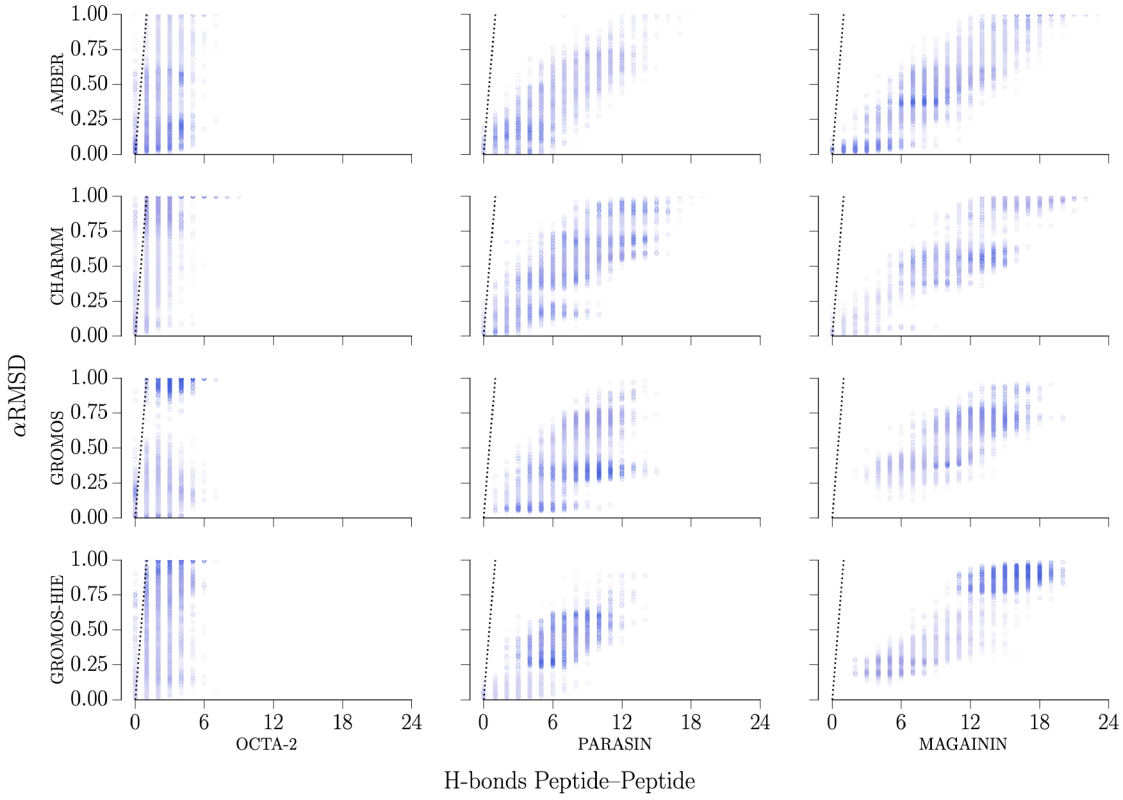


Figure B.16: Relation between the helicity of the peptide calculated with the αRMSD method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–peptide H-bonds (X-axis). The plots are arranged by force field and peptide. TThe saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

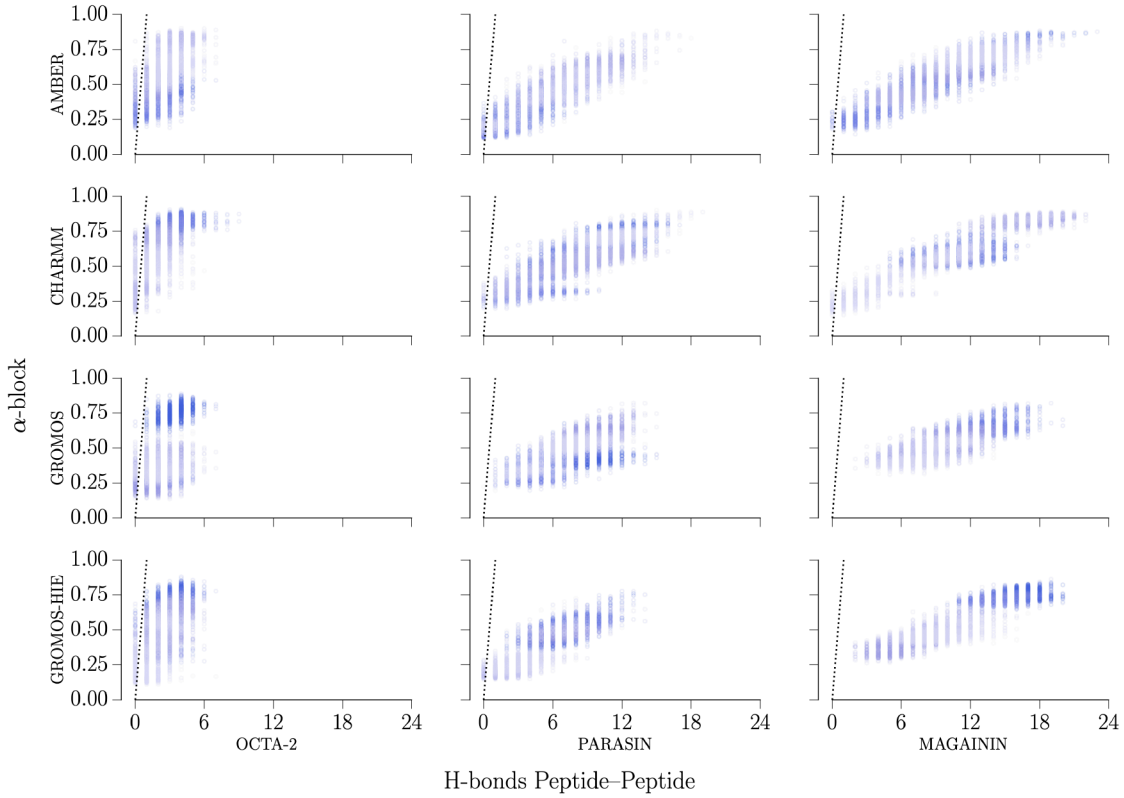


Figure B.17: Relation between the helicity of the peptide calculated with the α-block method and the number of peptide–peptide H-bonds in simulations starting from the shallow configuration. Each panel shows the relation between α-helical content (Y-axis) and the number of peptide–peptide H-bonds (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.001	0.032	0.010	0.109	0.098	0.219
	Parasin	0.557	0.558	0.532	0.206	0.694	0.766
	Magainin	0.541	0.812	0.076	0.642	0.126	0.636
C36m	Octa-2	0.117	0.004	0.002	0.085	0.002	0.506
	Parasin	0.539	0.579	0.349	0.763	0.210	0.685
	Magainin	0.056	0.231	0.107	0.846	0.907	0.812
G54a7	Octa-2	0.029	0.000	0.000	0.000	0.001	0.289
	Parasin	0.041	0.326	0.138	0.135	0.470	0.397
	Magainin	0.669	0.075	0.109	0.114	0.270	0.530
G54a7-HIE	Octa-2	0.187	0.085	0.208	0.132	0.020	0.275
	Parasin	0.443	0.207	0.133	0.173	0.425	0.681
	Magainin	0.003	0.010	0.378	0.326	0.182	0.697

Table B.21: Values of Pearson's correlation coefficient r^2 between the values of helicity given by DSSP and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.040	0.154	-0.097	0.212	0.260	0.403
	Parasin	0.604	0.574	0.605	0.357	0.668	0.742
	Magainin	0.522	0.742	0.168	0.545	0.274	0.619
C36m	Octa-2	0.339	0.078	-0.039	0.253	0.047	0.673
	Parasin	0.556	0.621	0.466	0.730	0.310	0.657
	Magainin	0.135	0.352	0.194	0.679	0.847	0.682
G54a7	Octa-2	0.172	—	—	-0.019	0.034	0.467
	Parasin	0.134	0.472	0.239	0.330	0.578	0.479
	Magainin	0.551	0.221	0.268	0.327	0.400	0.567
G54a7-HIE	Octa-2	0.410	0.279	0.401	0.316	0.108	0.459
	Parasin	0.538	0.378	0.245	0.261	0.480	0.655
	Magainin	-0.036	0.187	0.480	0.491	0.351	0.688

Table B.22: Values of Kendall's rank correlation coefficient τ between the values of helicity given by DSSP and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.005	0.124	0.061	0.203	0.148	0.316
	Parasin	0.668	0.661	0.583	0.254	0.752	0.817
	Magainin	0.671	0.874	0.144	0.734	0.206	0.770
C36m	Octa-2	0.107	0.019	0.007	0.329	0.011	0.538
	Parasin	0.535	0.626	0.384	0.793	0.267	0.685
	Magainin	0.205	0.224	0.101	0.852	0.927	0.854
G54a7	Octa-2	0.112	0.015	0.148	0.018	0.014	0.318
	Parasin	0.031	0.375	0.272	0.134	0.553	0.429
	Magainin	0.710	0.092	0.204	0.304	0.377	0.622
G54a7-HIE	Octa-2	0.220	0.119	0.289	0.096	0.279	0.362
	Parasin	0.489	0.254	0.211	0.330	0.520	0.747
	Magainin	0.051	0.102	0.625	0.373	0.225	0.837

Table B.23: Values of Pearson's correlation coefficient r^2 between the values of helicity given by αRMSD and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.054	0.297	0.196	0.343	0.271	0.434
	Parasin	0.606	0.562	0.591	0.337	0.683	0.744
	Magainin	0.573	0.766	0.219	0.620	0.330	0.698
C36m	Octa-2	0.285	0.043	0.064	0.478	0.041	0.639
	Parasin	0.547	0.600	0.483	0.707	0.338	0.626
	Magainin	0.286	0.379	0.230	0.700	0.838	0.707
G54a7	Octa-2	0.285	0.082	0.241	-0.093	0.074	0.387
	Parasin	0.149	0.456	0.332	0.278	0.568	0.475
	Magainin	0.464	0.167	0.312	0.404	0.460	0.592
G54a7-HIE	Octa-2	0.443	0.335	0.416	0.200	0.361	0.488
	Parasin	0.532	0.349	0.182	0.300	0.389	0.659
	Magainin	0.129	0.238	0.486	0.469	0.300	0.737

Table B.24: Values of Kendall's rank correlation coefficient τ between the values of helicity given by αRMSD and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.014	0.229	0.087	0.249	0.183	0.351
	Parasin	0.691	0.681	0.639	0.277	0.767	0.830
	Magainin	0.672	0.886	0.170	0.743	0.281	0.784
C36m	Octa-2	0.210	0.002	0.041	0.362	0.010	0.601
	Parasin	0.577	0.649	0.441	0.797	0.383	0.717
	Magainin	0.175	0.343	0.243	0.881	0.927	0.875
G54a7	Octa-2	0.205	0.038	0.188	0.013	0.015	0.349
	Parasin	0.047	0.379	0.301	0.196	0.610	0.443
	Magainin	0.738	0.139	0.217	0.302	0.397	0.635
G54a7-HIE	Octa-2	0.316	0.140	0.300	0.093	0.264	0.375
	Parasin	0.532	0.289	0.319	0.388	0.618	0.785
	Magainin	0.105	0.096	0.656	0.430	0.288	0.843

Table B.25: Values of Pearson's correlation coefficient r^2 between the values of helicity given by α-block and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

Force field	Peptide	Replica					
		#1	#2	#3	#4	#5	All
A19	Octa-2	0.083	0.372	0.207	0.344	0.307	0.442
	Parasin	0.624	0.570	0.629	0.362	0.689	0.754
	Magainin	0.577	0.776	0.254	0.612	0.375	0.707
C36m	Octa-2	0.320	0.053	0.143	0.499	0.081	0.644
	Parasin	0.576	0.615	0.477	0.718	0.439	0.652
	Magainin	0.228	0.453	0.358	0.708	0.848	0.728
G54a7	Octa-2	0.363	0.137	0.294	-0.074	0.067	0.413
	Parasin	0.146	0.466	0.362	0.338	0.598	0.484
	Magainin	0.539	0.230	0.328	0.399	0.449	0.599
G54a7-HIE	Octa-2	0.451	0.355	0.419	0.186	0.328	0.486
	Parasin	0.552	0.385	0.316	0.339	0.542	0.698
	Magainin	0.191	0.230	0.531	0.482	0.375	0.742

Table B.26: Values of Kendall's rank correlation coefficient τ between the values of helicity given by α-block and the number of H-bonds within the peptide. Each column corresponds to the data extracted from one of the replicas. The values under the column 'All' are associated with the combined data across all replicas.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

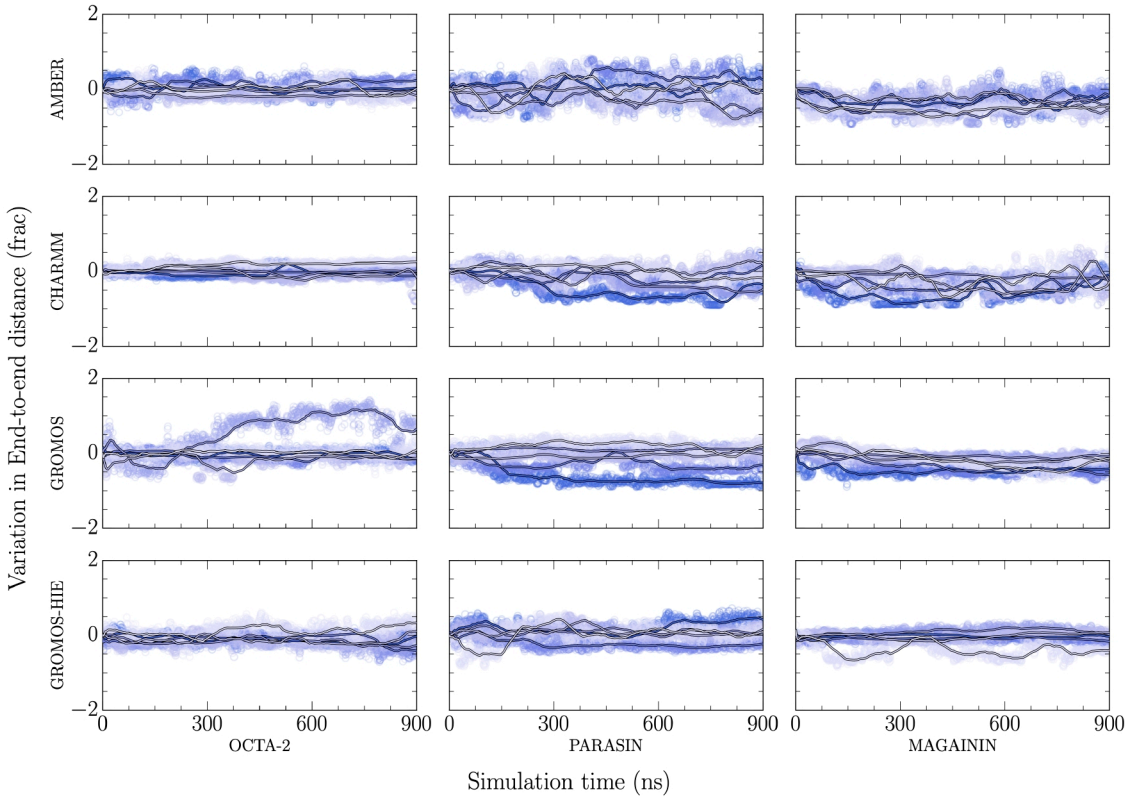


Figure B.18: Time evolution of distance between the C_α atoms of the first and last residues in the peptide in simulations starting from the shallow configuration. Each panel shows the variation of the distance with respect to the initial structure and relative to that value (Y-axis) over the simulation time (X-axis). The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the variation in end-to-end distance in the five independent simulation replicas, calculated as a rolling average of the instantaneous values (displayed as faint points).

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

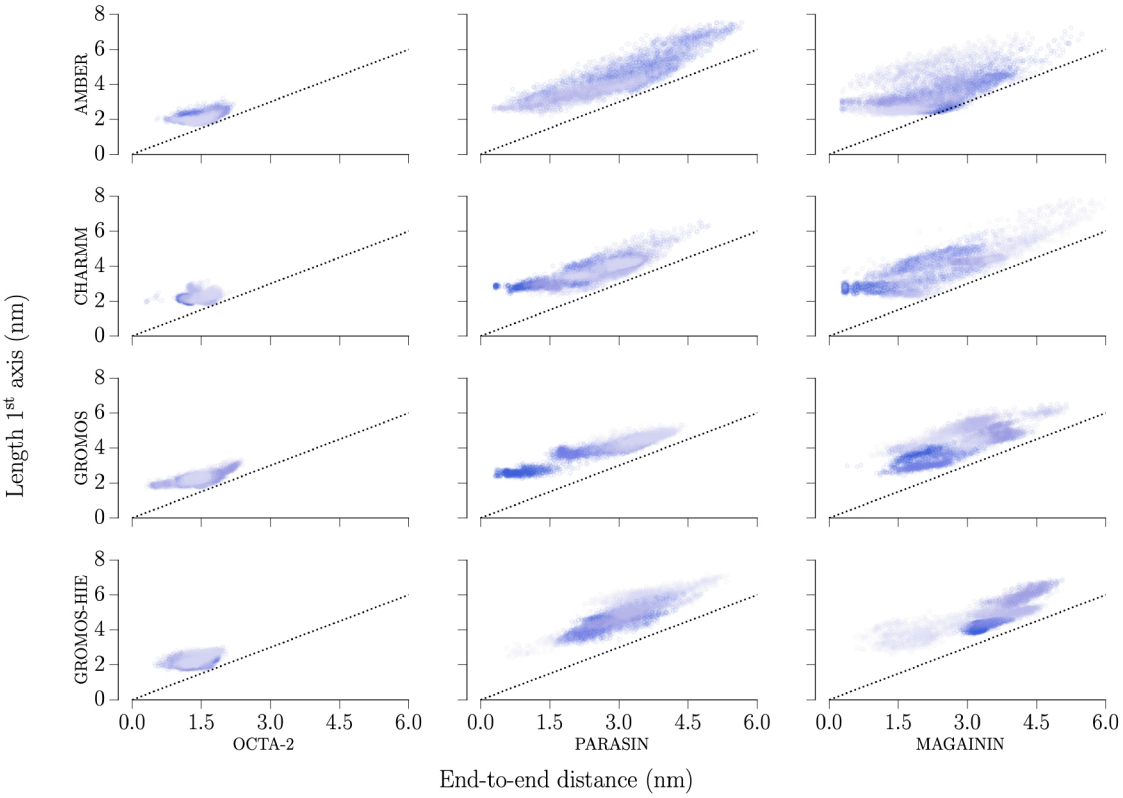


Figure B.19: Relation between the length of the 1st axis of the equivalent ellipsoid and distance between the C_α atoms of the first and last residues in simulations starting from the shallow configuration. Each panel shows the relation between the length of the 1st axis of the ellipsoid (Y-axis) and the distance between C_α atoms (X-axis). The plots are arranged by force field and peptide. The saturation reflects the density of points. The dashed black line indicates a 1-to-1 correspondence between the magnitudes, and is included as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.3 Simulations starting from the embedded configuration

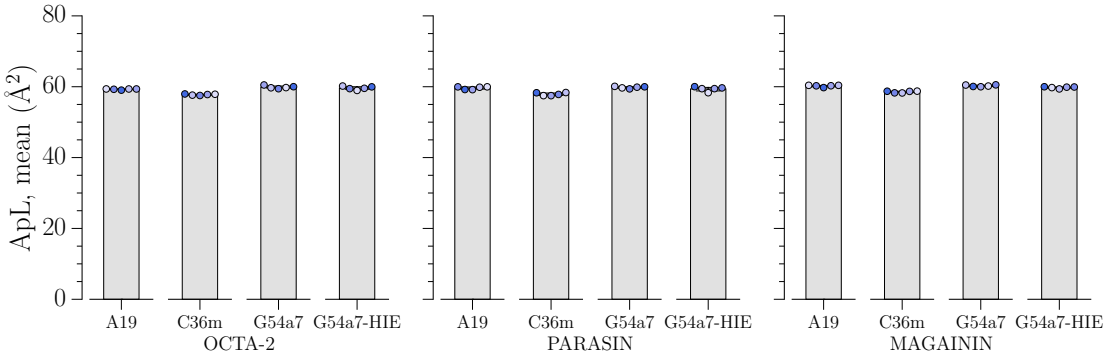


Figure B.20: Mean area per lipid in simulations starting from the embedded configuration, averaged over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field (A19 for AMBER, C36m for CHARMM, G54a7 for GROMOS, and G54a7-HIE for GROMOS-HIE). Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

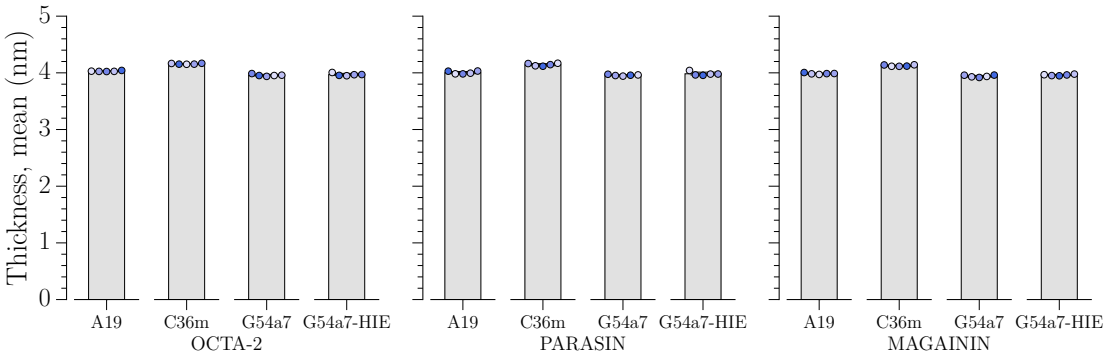


Figure B.21: Mean membrane thickness in simulations starting from the embedded configuration, averaged over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

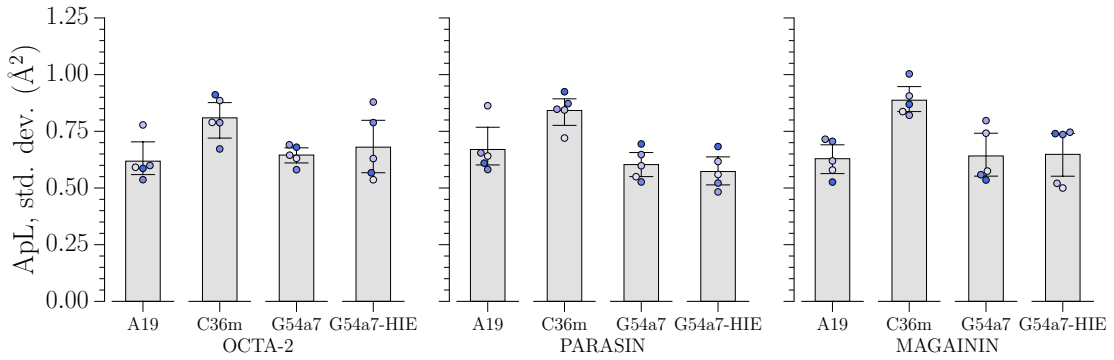


Figure B.22: Standard deviation of the area per lipid in simulations starting from the embedded configuration, computed over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

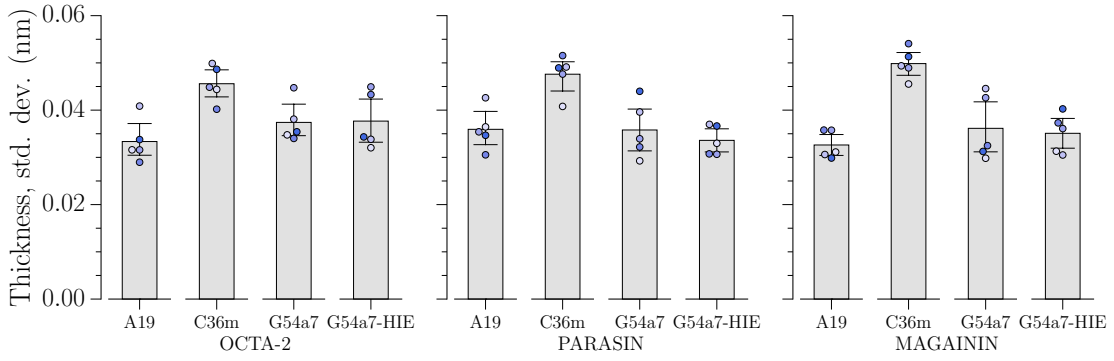


Figure B.23: Standard deviation of the membrane thickness in simulations starting from the embedded configuration, computed over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

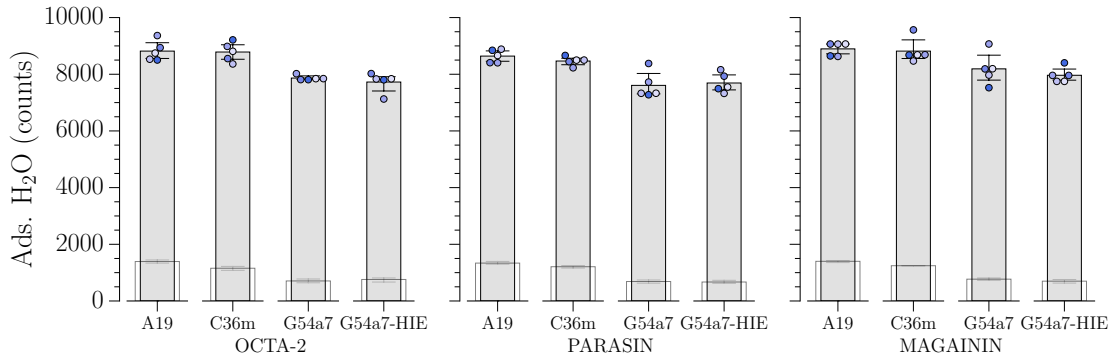


Figure B.24: Mean number of water molecules adsorbed at the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation. Water adsorption was estimated from the integral of the water density profile within the region of non-zero lipid density. The gray bars represent the mean and standard deviation of water molecules that penetrate the membrane (Figure B.24) as a reference.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.4 Performance and stability of the GROMOS-HIE and Reaction Field Schemes

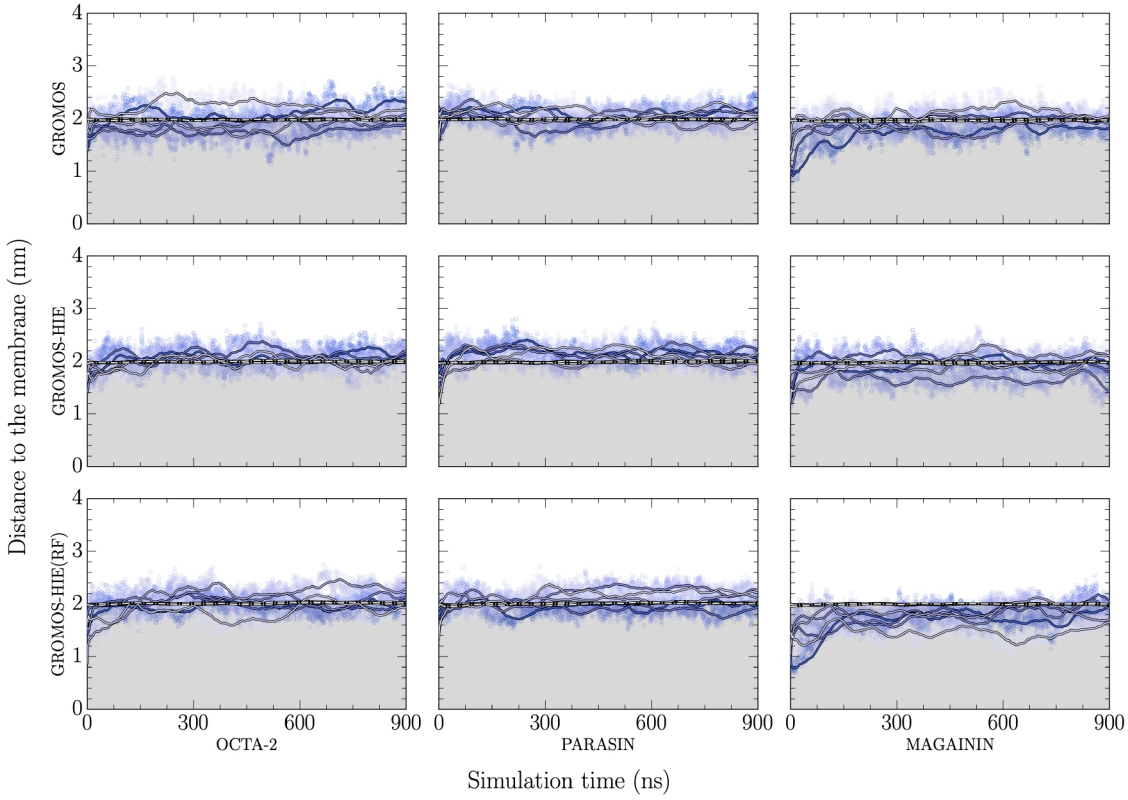


Figure B.25: Time evolution of the peptide–membrane distance for the embedded starting configuration. Each panel shows the evolution of the distance along the Z coordinate—perpendicular to the bilayer plane—between the centroid of the peptide and the center of the membrane, as a function of simulation time. The plots are arranged by force field and peptide. The thin blue lines represent the evolution of the peptide–membrane distance, calculated as a rolling average of the instantaneous values (displayed as faint points), for five independent simulation replicas. The gray shaded area indicates the hydrophobic core of the membrane, while the dashed white line at approximately 2 nm marks the average position of the phosphorus atoms of the lipid headgroups, corresponding to the membrane surface.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

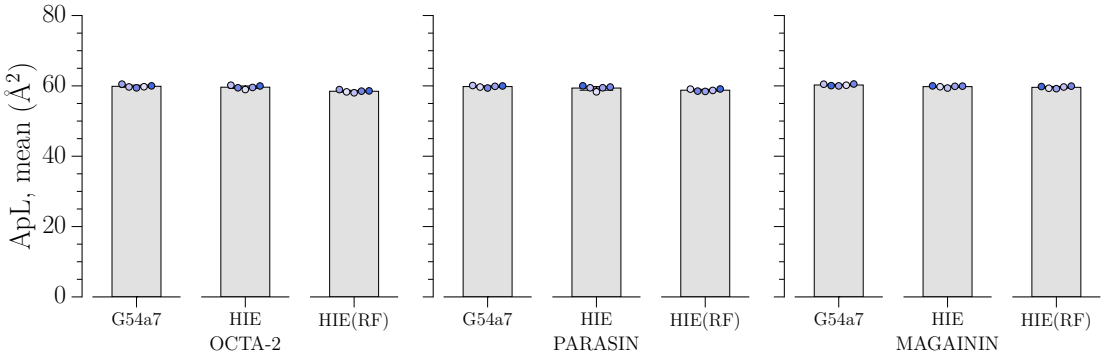


Figure B.26: Mean area per lipid in simulations starting from the embedded configuration, averaged over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

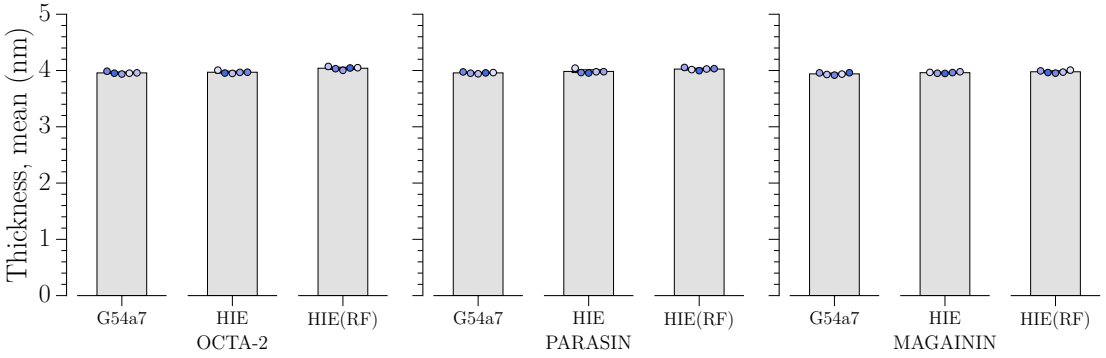


Figure B.27: Mean membrane thickness in simulations starting from the embedded configuration, averaged over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

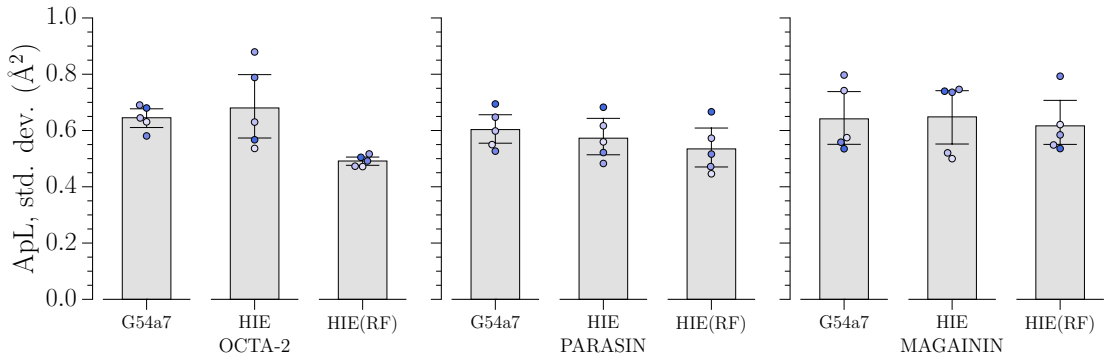


Figure B.28: Standard deviation of the area per lipid in simulations starting from the embedded configuration, computed over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

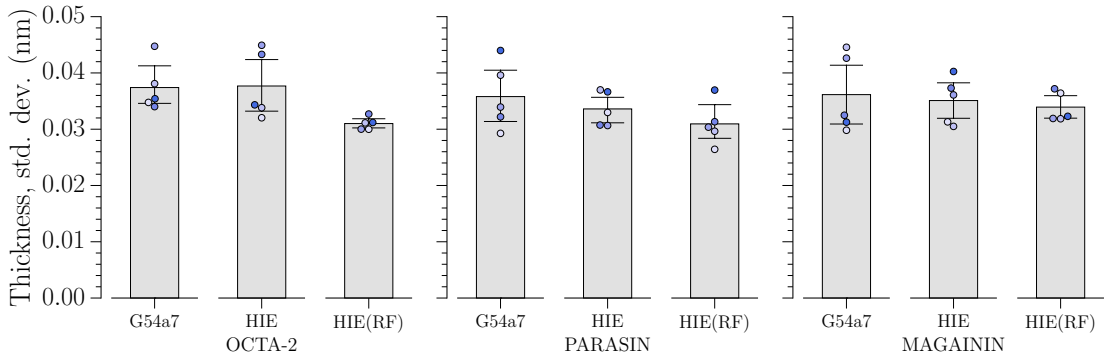


Figure B.29: Standard deviation of the membrane thickness in simulations starting from the embedded configuration, computed over the last 200 ns**.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent individual replica values, bar heights indicate their mean, and error bars correspond to one standard deviation.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

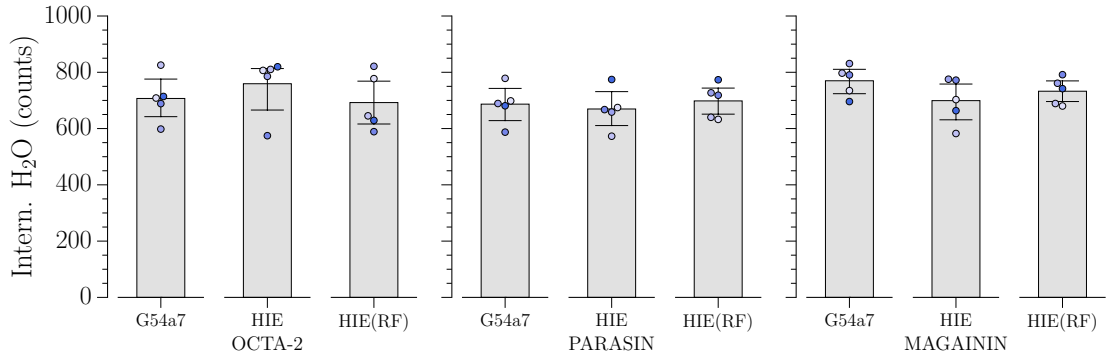


Figure B.30: Mean number of water molecules that penetrate the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation. Water adsorption was estimated from the integral of the water density profile within the maxima of the lipid phosphorus atom distribution.

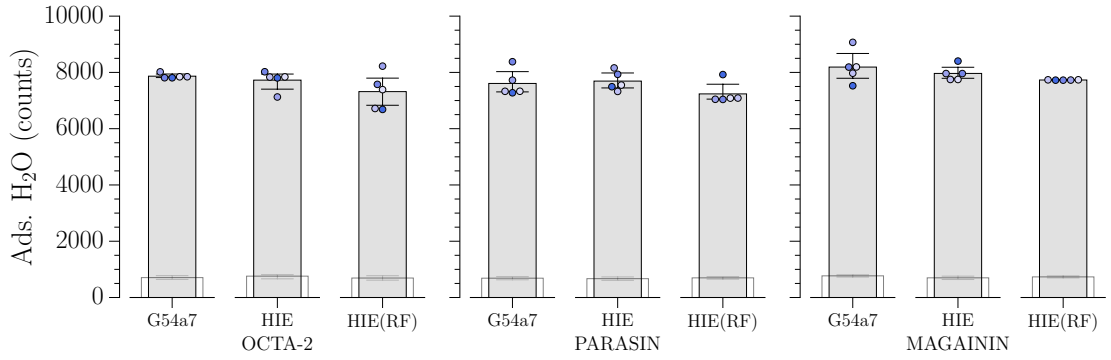


Figure B.31: Mean number of water molecules adsorbed into the membrane in simulations starting with the peptide embedded below the surface, computed in the last 200 ns **of the simulation.** Each plot corresponds to a different peptide, with bars grouped by force field. Circles represent values from individual replicas, bar heights indicate their mean, and error bars correspond to one standard deviation. Water adsorption was estimated from the integral of the water density profile within region of non-zero lipid density. The gray bars represent the mean and standard deviation of the number of water molecules contained in the membrane (Figure B.30) as reference

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

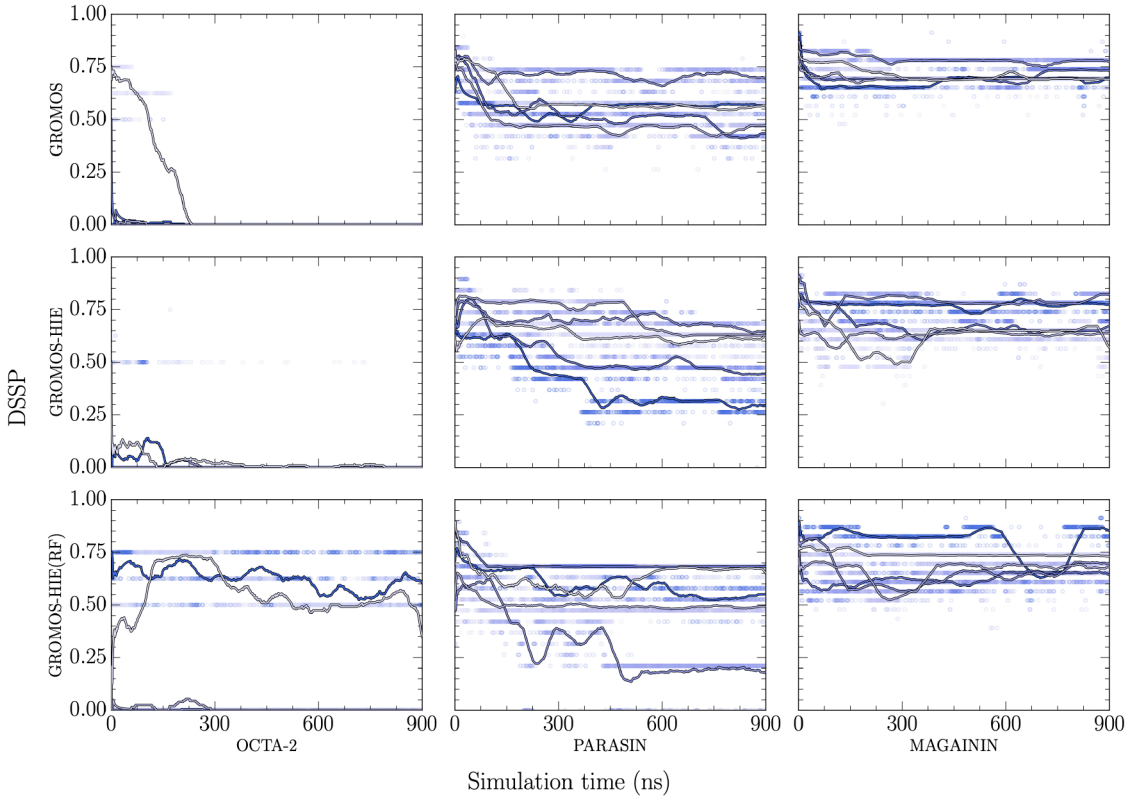


Figure B.32: Time evolution of peptide helicity calculated with the DSSP method for the embedded starting configuration. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using DSSP. The plots are arranged by force field and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, calculated as rolling averages of the instantaneous DSSP values (shown as faint points), to highlight the overall trends. All panels share identical axes and scales, enabling direct comparison across peptides and force fields.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

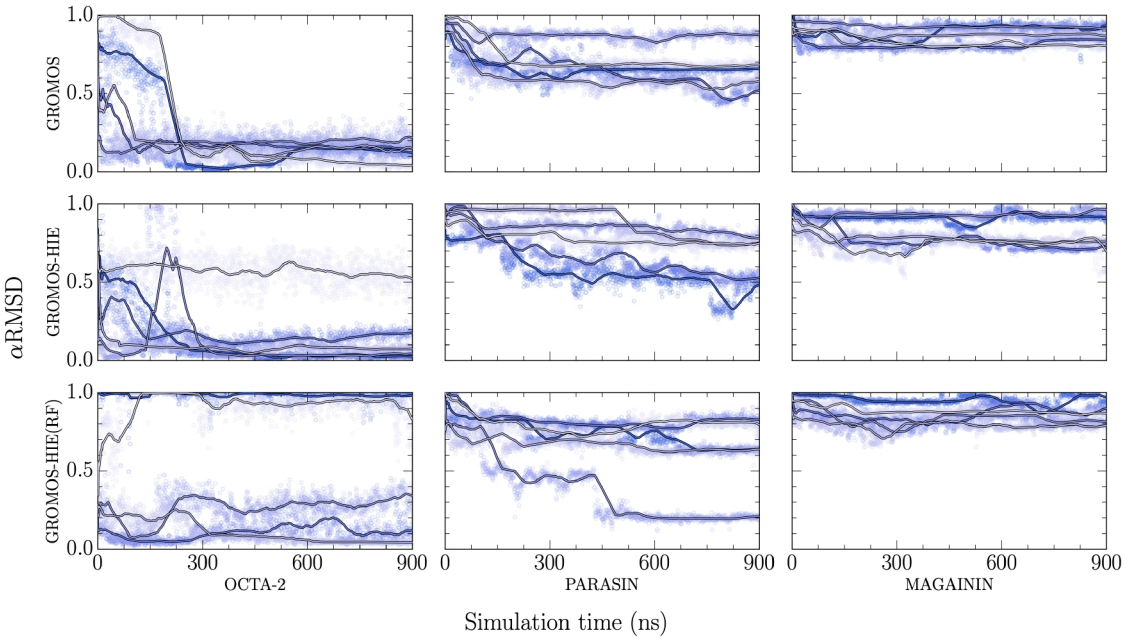


Figure B.33: Time evolution of peptide helicity calculated with the αRMSD method for the embedded starting configuration. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using αRMSD. The plots are arranged by force field and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, calculated as rolling averages of the instantaneous αRMSD values (shown as faint points), to highlight the overall trends. All panels share identical axes and scales, enabling direct comparison across peptides and force fields.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

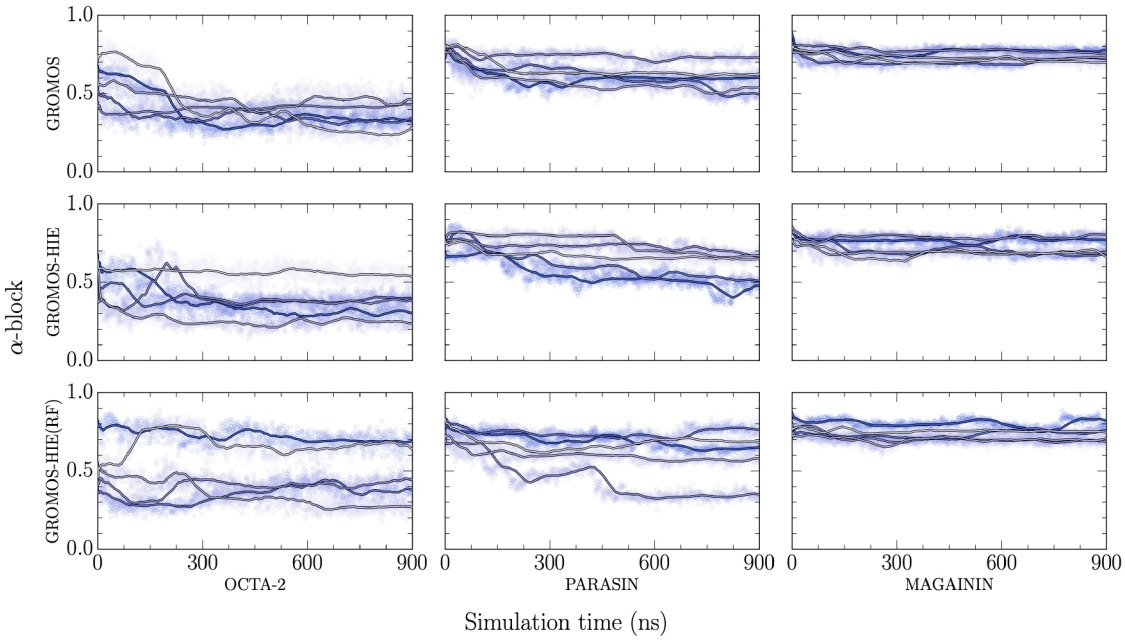


Figure B.34: Time evolution of peptide helicity calculated with the α-block method for the embedded starting configuration. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using α-block. The plots are arranged by force field and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, calculated as rolling averages of the instantaneous α-block values (shown as faint points), to highlight the overall trends. All panels share identical axes and scales, enabling direct comparison across peptides and force fields.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

B.5 Long timescale simulations

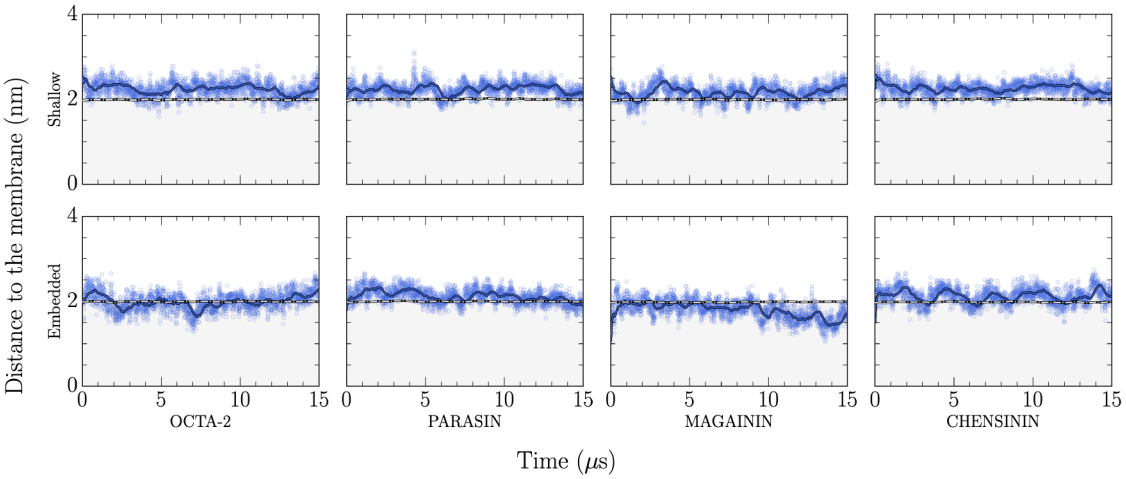


Figure B.35: Time evolution of the peptide–membrane distance in long timescale GROMOS-HIE simulations. Each panel shows the evolution of the distance along the Z coordinate—perpendicular to the bilayer plane—between the centroid of the peptide and the center of the membrane, as a function of simulation time. The plots are arranged by initial configuration and peptide. The thin blue lines represent the evolution of the peptide–membrane distance, calculated as a rolling average of the instantaneous values (displayed as faint points). The gray shaded area indicates the hydrophobic core of the membrane, while the dashed white line at approximately 2 nm marks the average position of the phosphorus atoms of the lipid headgroups, corresponding to the membrane surface.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

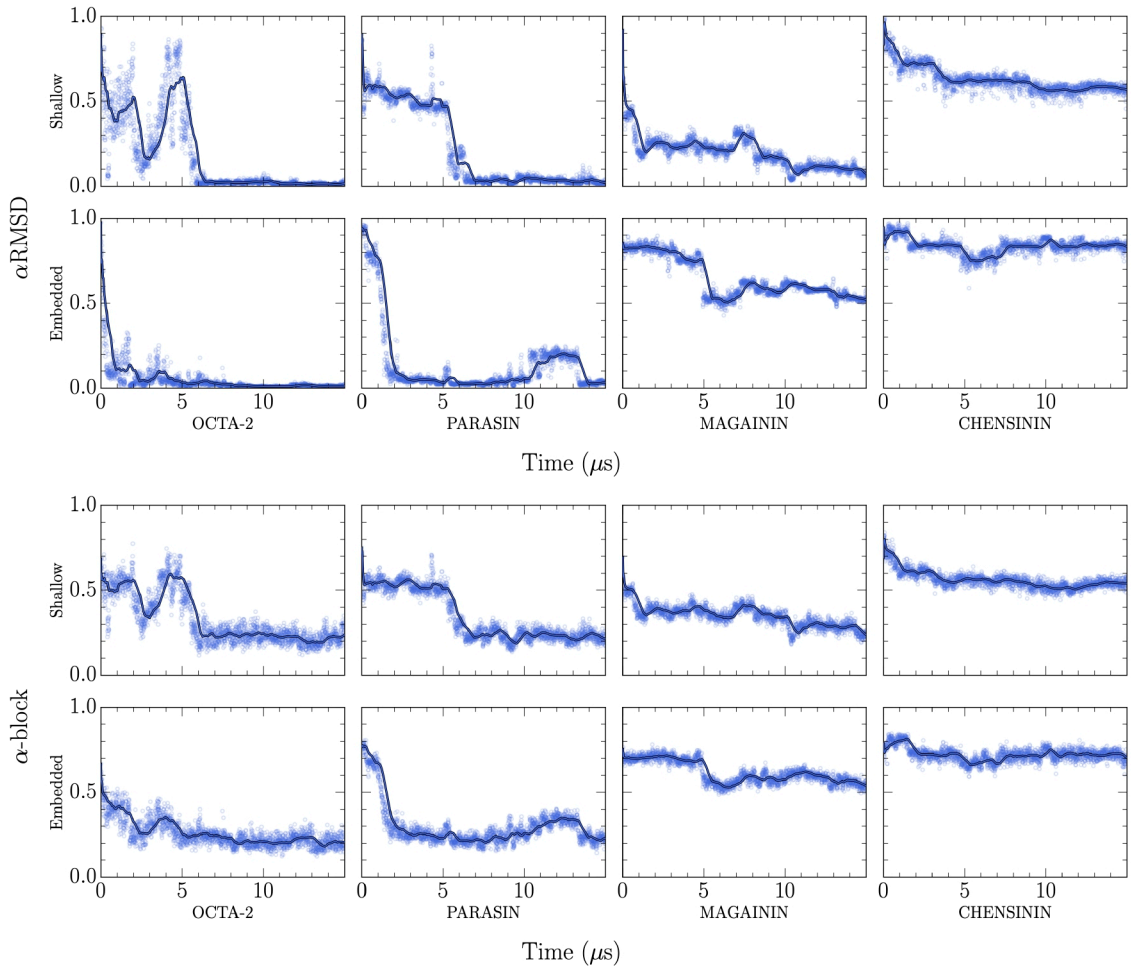


Figure B.36: Time evolution of peptide helicity calculated with the αRMSD (upper) and α-block (lower) methods in long timescale GROMOS-HIE simulations. Each panel shows the variation of the α-helical content (Y-axis) over simulation time (X-axis), computed using αRMSD or α-block. The plots are arranged by starting configuration and peptide. Thin blue lines represent the helicity evolution of five independent simulation replicas, calculated as rolling averages of the instantaneous helicity values (shown as faint points), to highlight the overall trends. All panels share identical axes and scales, enabling direct comparison across peptides and initial configuration.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix C.

Relation of sequences simulated for the study of the antimicrobial activity of peptides

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ID	Sequence	Properties					
		Length	Total charge	#Acidic res.	#Basic res.	#Polar res.	#Hydrophobic res.
AP00013	LFDIIKKIAESF	12	0	2	2	1	7
AP00015	LFDIVKKVVGALGSL	15	1	1	2	1	11
AP00101	VQWFSKFLGRIL	12	2	0	2	2	8
AP00144	IGKFLHSAKKFGKAFVGEIMNS	22	3	1	5	3	13
AP00146	IGAVLKVLTTGLPALISWIKRKRQQ	25	5	0	5	5	15
AP00339	FGWLIKGAIHAGKAIHGLIHRRRH	24	5	0	9	0	15
AP00367	GLRSLGRKILRAWKKYGPIIVPIIRIG	27	7	0	7	2	18
AP00414	IGSALKKALPVAKKIGKIALPIAKAALP	28	6	0	6	1	21
AP00476	LNALKKVFQGIHEAIKLINNHVQ	23	2	1	5	5	12
AP00492	QRVEELSKFSKKGAAARRRK	20	6	2	8	3	7
AP00499	GALAVVVWLWLWLW	14	0	0	0	0	14
AP00542	LGTILGLLKSL	11	1	0	1	2	8
AP00555	GRGKQGGKVRAKAKTRSS	18	7	0	7	4	7
AP00890	DWKKVDWKKVSKKTCKVMLKACKFLG	26	7	2	9	4	11
AP00952	IGKFLKKAKKFGKAFVKILKK	21	9	0	9	0	12
AP00955	IIRPIIQIIKQKIR	14	4	0	4	2	8
AP01010	MWSGMWRRKLKKLRNALKKKLKGE	24	9	1	10	2	11
AP01011	LFGKLIKKFGRKAISYAVKKARGKH	25	9	0	10	2	13
AP01012	WKSMAKKLKEYMEKLKQRA	19	5	2	7	3	7
AP01015	LKDKVKSMGEKLKQYIQTWKAKF	23	5	2	7	5	9
AP01016	FFGKMKEYFKKFGASFKRRFANLKKRL	27	9	1	10	3	13
AP01223	FKDLLKGAAKALVKTVLF	18	3	1	4	1	12
AP01445	MGSALRIAAKVLPAALCQIFKKC	23	4	0	4	4	15
AP01447	LPGLIAGIAKML	12	1	0	1	0	11
AP01448	LPIALKALGSIFPKIL	16	2	0	2	1	13
AP01545	FGHLFKLATKIIPSLFQ	17	2	0	3	3	11
AP02104	QFITDLIKKAVDFFKGLFGNK	21	2	2	4	3	12
AP02118	LPAALAGIGGILGKLF	16	1	0	1	0	15
AP02139	LLSLLSLLGKLL	12	1	0	1	2	9
AP02143	SLQMGATAIKQVKKLFKKKGG	21	6	0	6	4	11
AP02151	LYKKLLKKLLKSAKKLG	17	7	0	7	2	8
AP02162	IKIPWGKVKDFLVGGMKAV	19	3	1	4	0	14
AP02218	FKAIWSGIKSLF	12	2	0	2	2	8
AP02228	KGWFKAMKSIAKFIAKEKLKEHL	23	5	2	8	1	12
AP02379	LLRPLLQLLKQKLR	14	4	0	4	2	8
AP02401	APIAKYLATALAKWALKQGFAKLKS	25	5	0	5	4	16
AP02637	LCASLRARHTIPQCKKFGRR	20	6	0	7	5	8
AP02645	EDLPNFGHIQVKVFNHGEHIHH	22	-2	3	6	3	10
AP02662	PHKKKLAVYPVFLFYLFLSWFSLIV	25	3	0	4	4	17
AP02666	VAGEKLWLLPHLLKMLLTPTP	21	1	1	3	2	15
AP02872	IFSKLAGKKIKNLLISGLKG	20	5	0	5	3	12
AP02883	MASLLAKVLPHVVKLIK	17	3	0	4	1	12
AP02970	LQHIIGALSHFF	12	0	0	2	2	8
AP03013	NPLKLFLPSTWVHFFKFLR	19	3	0	4	3	12
AP03059	IVPSIFLLKKAFCIALKKC	19	4	0	4	3	12
AP03098	IVELTLPRVSVRL	13	1	1	2	2	8
AP03112	LSLLLSLGLKLL	12	1	0	1	2	9
AP03200	FGHLLRGIVSVGKHIHGLITG	21	2	0	5	2	14
AP03245	LPALAGIAGLLGKIF	15	1	0	1	0	14
AP03258	LPLIAGLFGKIF	12	1	0	1	0	11
AP03307	LWSKIKNVAAAAGKAALGAL	20	3	0	3	2	15
AP03324	VITDTLKGVAKTVAAELLRKAHCKLTNSC	29	3	2	6	8	13
AP03406	WLSKTYKKLENSAKKRIAEGIAIALRGGPR	30	6	2	8	5	15
AP03427	LWKTLLKNVGKAAGKAVLNAVTDMVNQ	27	3	1	4	6	16
AP03428	VDLKKIANILNSIF	14	1	1	2	3	8
AP03449	KFLARTGKFL	10	3	0	3	1	6
AP03488	LWENFKNAGKKFILNILDKIRCRVAGGCRT	30	5	2	7	6	15
AP03489	KRWRRLIFNYF	11	4	0	4	2	5
AP03517	VVNVLVKVLPPPVV	14	1	0	1	1	12
AP03654	WLRRIKAWLRRIKA	14	6	0	6	0	8
AP03679	LPLLAGLAANFLPQIICKIARKC	23	3	0	3	4	16
AP03721	LWKSILKNAGKAALNEINQIV	21	2	1	3	5	12

Table C.1: List of peptides simulated in CG resolution. Continues on the next page.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

ID	Sequence	Properties					
		Length	Total charge	#Acidic res.	#Basic res.	#Polar res.	#Hydrophobic res.
AP03756	RFKRFRKKFKKLFKKLS	17	10	0	10	1	6
AP03758	GLRSLGRKILRAWKKYGPQATPATRQ	26	7	0	7	6	13
AP03759	RLKRLRKKLKKLLKKLS	17	10	0	10	1	6
AP03763	NKMAYNVGKAISRIMRRVR	19	6	0	6	4	9
AP03825	KWLKKLLKWLKKL	13	6	0	6	0	7
AP03826	KWVKKVVKWVKKV	13	6	0	6	0	7
AP03941	LSALWGVAKSLF	12	1	0	1	2	9
AP03942	PFLLSLIPSAISALKKL	17	2	0	2	3	12
AP04269	ILGKLWKGVKSIF	13	3	0	3	1	9
AP04597	LWKLIPKAIKKVKSLIKK	18	7	0	7	1	10
AP04628	RRLSKVLGKLRDKIATWRQRLLTK	24	8	1	9	4	10
NC00130	LNVPPSWFLSQR	12	1	0	1	4	7
NC00188	CKRKGSSCRRTSYDCCTGSCRNGKC	25	6	1	7	14	3
NC00196	NGCCRAGDCCSRFEIKENDF	20	-1	4	3	7	6
NC00210	GIRGNCCMFHTCPIDYSRFYCP	22	1	1	3	9	9
NC00252	FAKLLAKLAKKSL	13	4	0	4	1	8
NC00258	KKEKDIMKKTI	11	3	2	5	1	3
NC00268	KIKWFKTMKSLAKFLAKEQMKKHLGE	26	6	2	9	3	12
NC00356	QFRPSYQIPP	10	1	0	1	4	5
NC00415	PGLKGKRGDSGSPATWTTRG	20	3	1	4	5	10
NC00565	CWLCNGRCGR	10	2	0	2	4	4
NC00592	FAKKLAKLAKKALAL	15	5	0	5	0	10
NC00701	LPHPVLHMGPLR	12	1	0	3	0	9
NC00961	SVEGVISTIKDFAVKVCCSVSLKFCCPTA	29	1	2	3	10	14
NC01227	TPPAGPDVGPR	11	0	1	1	1	8
NC01255	FRVDEEFQSPFASQSRGYFLFRPRN	25	1	3	4	7	11
NC01402	AVKQPAVVYP	10	1	0	1	2	7
NC01521	LVVGDQLCYRVLIKCLMNK	19	2	1	3	5	10
NC01525	TLTKAFEQTLLPNSATVL	18	0	1	1	7	9
NC01594	YGRRARRAARR	11	6	0	6	1	4
NC01730	GRRINRLILPRN	12	4	0	4	2	6
NC01800	KITLKLAIKAWKLALKAA	18	5	0	5	1	12
NC01995	FAKKLAKLAKKLAKLAL	17	6	0	6	0	11
NC02039	DSGCFGRRLDRIGSLSGLGCNVLRRY	26	3	2	5	7	12
NC02046	HPTWPQKSVWHGSDPNGRRLTESY	24	1	2	5	8	9
NC02053	STNITVTLKKFPL	13	2	0	2	5	6
NC02212	KLALKLALKAWKAALKLA	18	5	0	5	0	13
NC02252	GVDITVIRPNH	11	0	1	2	2	6
NC02346	QSCCATPSCARLYEKVY	17	1	1	2	9	5
NC02424	CGRCNGRCLL	10	2	0	2	4	4
NC02428	CCRLSCGLGCHPCC	14	1	0	2	7	5
NC02437	GFPGTPGLPGF	11	0	0	0	1	10
NC02460	KYKKALKKLAKLL	13	6	0	6	1	6
NC02461	GCTPEYCSMWCKVKVSQNYCVKNCKCPGR	29	4	1	5	14	9
NC02533	SATGAPWKMWVR	12	2	0	2	2	8
NC02545	GRILSFIKAGLAEHL	15	1	1	3	1	10
NC02629	KTVLLRKLLKLLVRKI	16	6	0	6	1	9
NC02924	VKLPPPVKLPPPVKLPPP	18	3	0	3	0	15
NC02940	ANTPCGPYTHDCPVKR	16	1	1	3	6	6
NC03040	SPWSQCTASCGGGVQTR	17	1	0	1	9	7
NC03115	GCCSFAACRKYRPEMCG	17	2	1	3	6	7
NC03407	TRDCKTKGSVCFASSECCIQDCWFVCLY	28	0	3	3	13	9
NC03613	ECTRSGGACNSHDQCCANFCRKATSTCM	28	1	2	4	15	7
NC03623	QTCCGYRMCIPC	12	1	0	1	7	4
NC03629	MRKWFHNVLSSGQLLADKWPAWDYNRK	27	3	2	6	6	13
NC03705	HIQLSPFSQSWR	12	1	0	2	5	5
NC03728	RGGRLSYSRRRFSTSTGR	18	6	0	6	7	5
NC03788	GWTLNSAGYLLGKFLPLILRKIVTAL	26	3	0	3	5	18
NC03835	YGGFMKPYTQQSHKPLITLLKHVTLKNEQ	29	3	1	6	10	12
NC03982	NTRGSWSNKRLSPR	14	4	0	4	6	4
NC04227	SPWSQCSVRCGRGQRSRQVR	20	5	0	5	9	6

Table C.1: List of peptides simulated in CG resolution. Continued from the previous page. The ID for active sequences has the form APXXXXX, as given in the APD3 database[156]. The data set for nonactive sequences of Tucs *et al.*[267] was not labeled, so IDs of the form NCXXXXX were generated for the ease of the handling of simulations and analyses. Length is defined as the number of residues; charge is indicated in multiples of e, the elementary charge of an electron.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix D.

Supplementary data for the study of the connection between the peptide–membrane interaction descriptors and antimicrobial activity through data-driven approaches

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude	Martini 2.2		Martini 2.2P		Martini 3	
	POPC	POPE:POPG 3:1	POPC	POPE:POPG 3:1	POPC	POPE:POPG 3:1
Bilayer thickness (nm)	3.930(28)	4.059(28)	3.941(26)	3.994(24)	3.891(29)	4.028(27)
Area per lipid (nm^2)	0.6407(59)	0.6048(56)	0.6478(59)	0.6345(57)	0.6477(60)	0.6147(62)

Table D.1: Structural descriptors of lipid bilayers in membrane-only systems. The membranes simulated are the same as those used in the simulations with peptides. The values in parentheses correspond to the standard deviation of the samples. Mean value and standard deviation were computed for the last 100 ns of the simulation.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

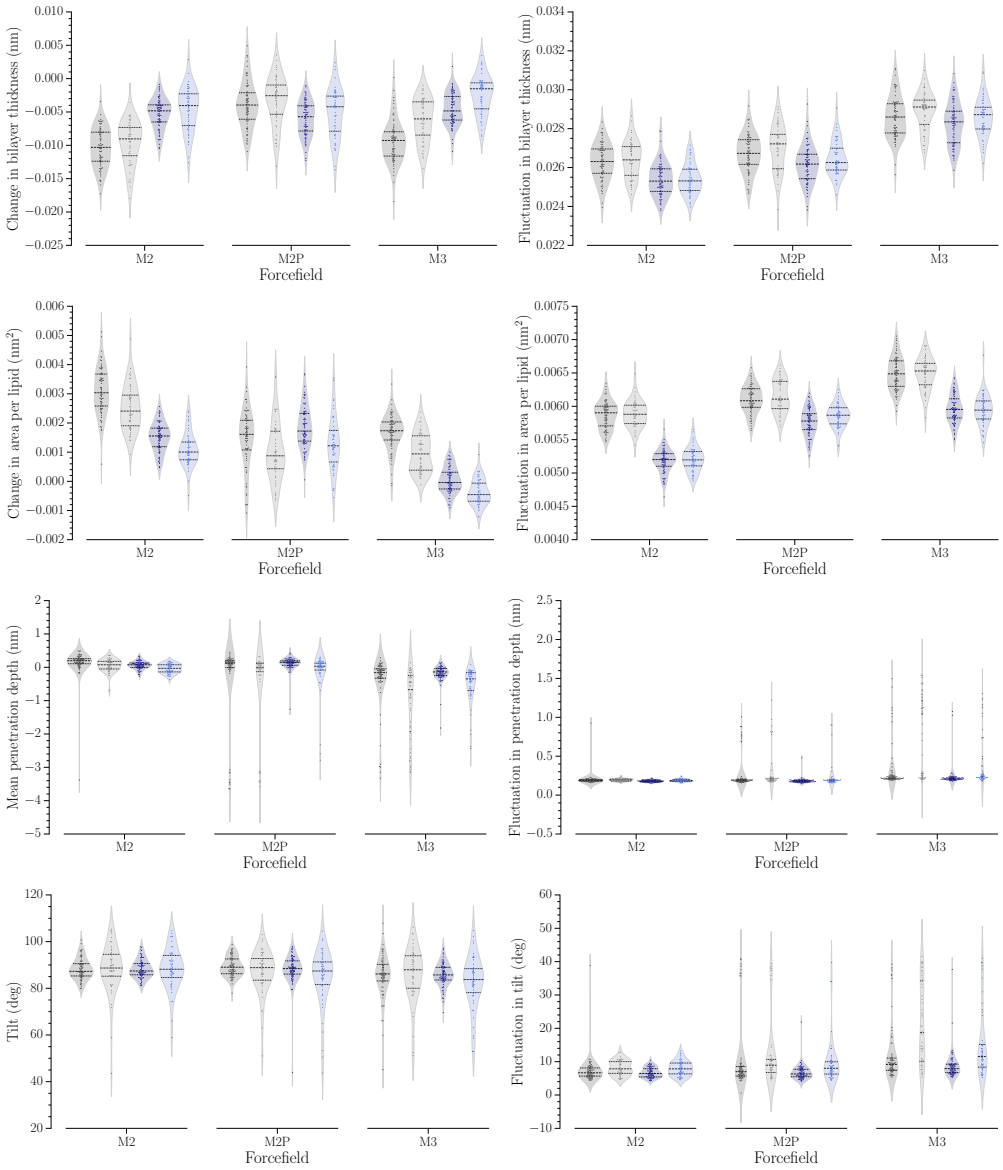


Figure D.1: Samples of MD descriptors. Continues in the next page.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

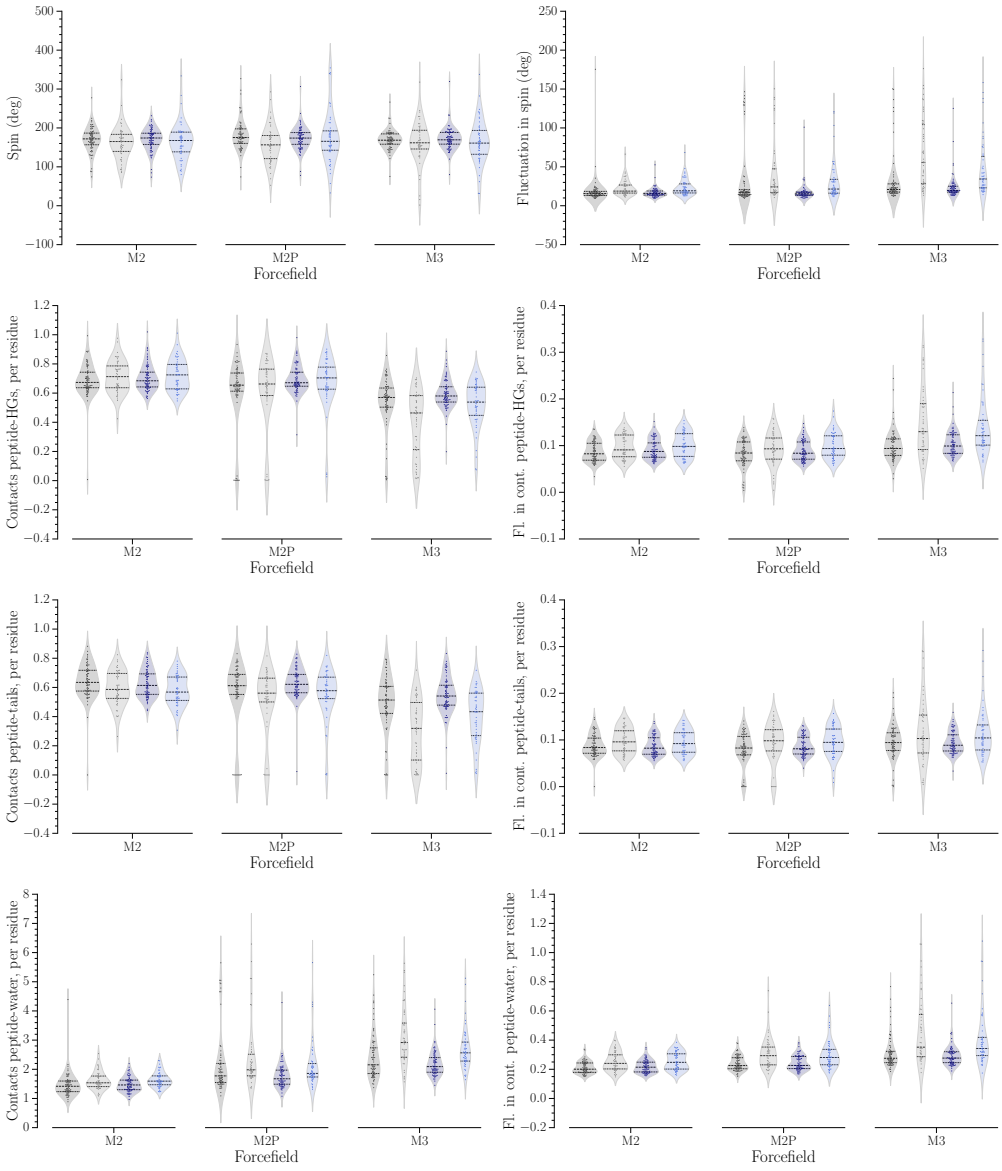


Figure D.1: Samples of MD descriptors. Continued from previous pages, Continues in the next page.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

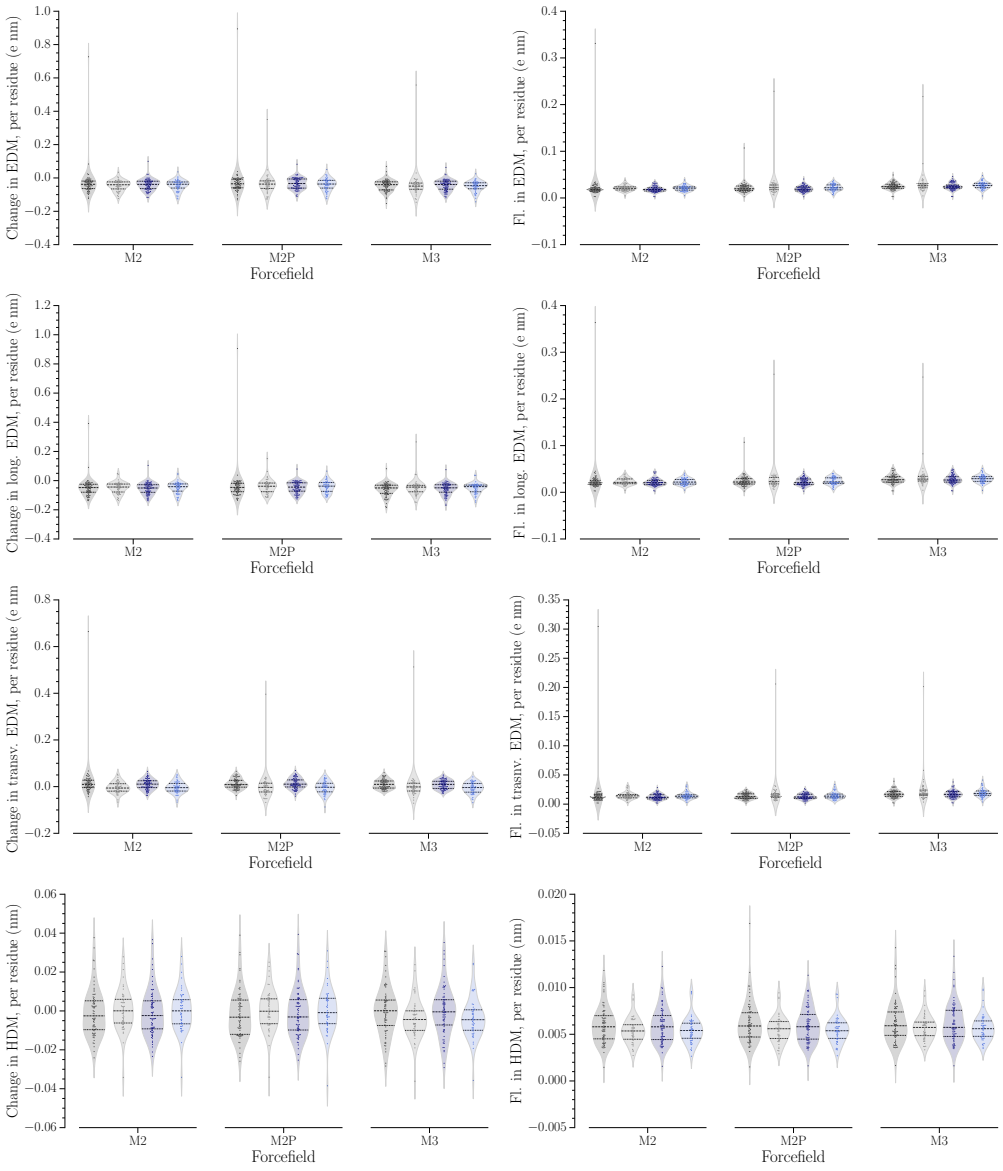


Figure D.1: Samples of MD descriptors. Continued from previous pages, Continues in the next page.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

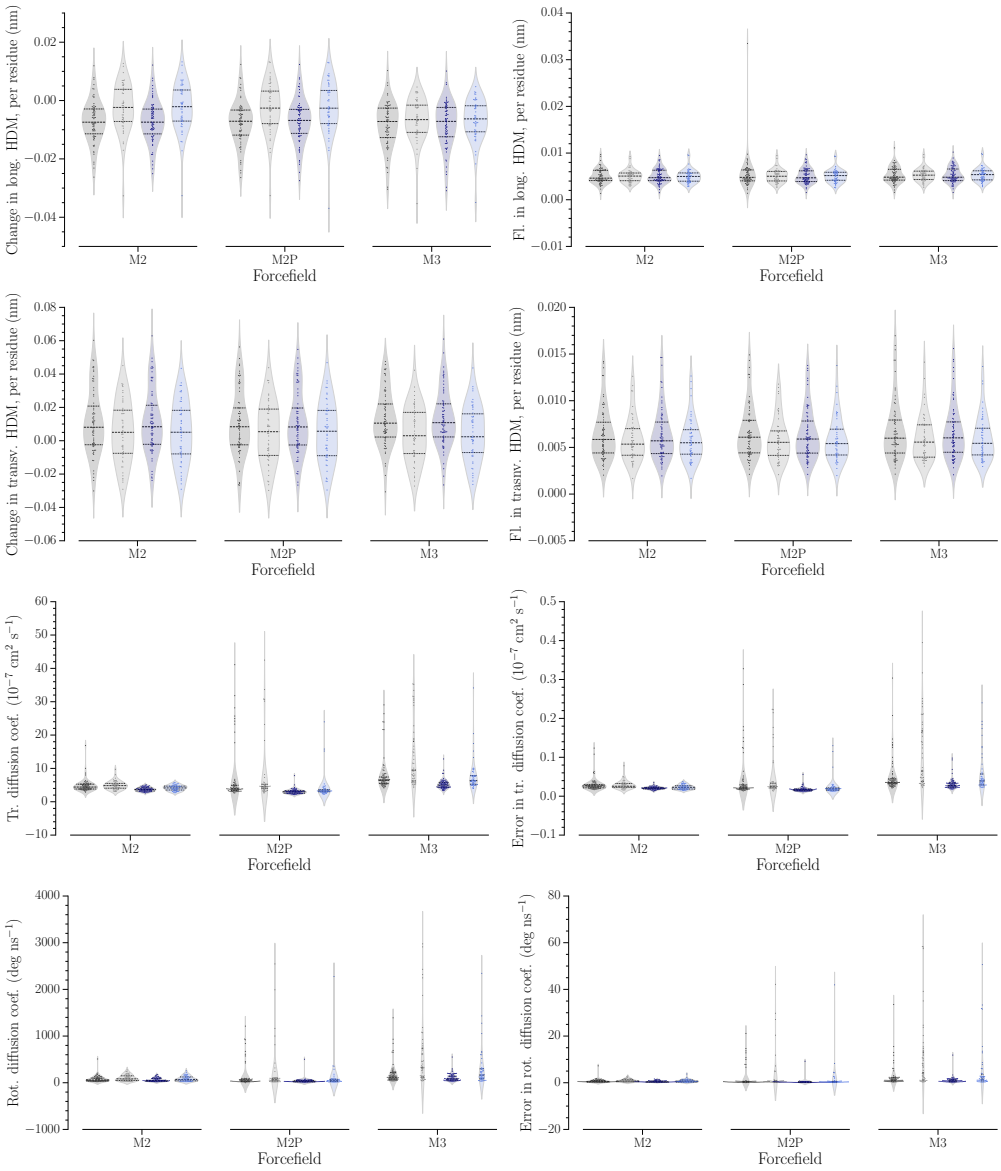


Figure D.1: Samples of MD descriptors. Continued from previous pages. Within each panel, samples are grouped by force field. Gray symbols are associated with simulations of the POPC membrane, and blue symbols with those of the POPE:POPG 3:1. The brightness signifies the nominal antimicrobial activity of the sequences with darker shades for active sequences and lighter shades for nonactives. Points represent the values of the descriptor obtained for individual simulations, while "violins" represent the kernel density estimate (KDE) of the samples, where wider regions indicate a higher probability density. The median and quartiles are marked with dashed lines in the distributions. The left column corresponds to the mean value of the magnitude (or related), the right column, to their fluctuations, estimated by the standard deviation of the sample—or the error, in the case of the diffusion coefficients—; both computed for the last 1 µs of simulation.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC
Bilayer thickness	μ	$1.03 \cdot 10^{-2}$	$9.04 \cdot 10^{-3}$	$4.83 \cdot 10^{-3}$	$4.04 \cdot 10^{-3}$	$3.96 \cdot 10^{-3}$	$2.55 \cdot 10^{-3}$	$5.71 \cdot 10^{-3}$	$4.22 \cdot 10^{-3}$	$9.27 \cdot 10^{-3}$	$6.04 \cdot 10^{-3}$	$4.84 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$
	σ	$2.63 \cdot 10^{-2}$	$2.64 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	$2.67 \cdot 10^{-2}$	$2.72 \cdot 10^{-2}$	$2.62 \cdot 10^{-2}$	$2.63 \cdot 10^{-2}$	$2.86 \cdot 10^{-2}$	$2.91 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$
Area per lipid	μ	$3.04 \cdot 10^{-3}$	$2.41 \cdot 10^{-3}$	$1.55 \cdot 10^{-3}$	$1.00 \cdot 10^{-3}$	$1.61 \cdot 10^{-3}$	$8.76 \cdot 10^{-4}$	$1.72 \cdot 10^{-3}$	$1.22 \cdot 10^{-3}$	$1.73 \cdot 10^{-3}$	$9.40 \cdot 10^{-4}$	$4.06 \cdot 10^{-5}$	$4.55 \cdot 10^{-4}$
	σ	$5.90 \cdot 10^{-3}$	$5.88 \cdot 10^{-3}$	$5.20 \cdot 10^{-3}$	$5.20 \cdot 10^{-3}$	$6.08 \cdot 10^{-3}$	$6.11 \cdot 10^{-3}$	$5.78 \cdot 10^{-3}$	$5.87 \cdot 10^{-3}$	$6.49 \cdot 10^{-3}$	$6.53 \cdot 10^{-3}$	$5.95 \cdot 10^{-3}$	$5.94 \cdot 10^{-3}$
Peptide penetration	μ	$1.97 \cdot 10^{-1}$	$7.62 \cdot 10^{-2}$	$7.62 \cdot 10^{-2}$	$3.09 \cdot 10^{-2}$	$1.23 \cdot 10^{-1}$	$1.40 \cdot 10^{-3}$	$1.48 \cdot 10^{-1}$	$3.14 \cdot 10^{-2}$	$1.60 \cdot 10^{-1}$	$6.71 \cdot 10^{-1}$	$1.40 \cdot 10^{-1}$	$3.47 \cdot 10^{-1}$
	σ	$1.89 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$	$1.79 \cdot 10^{-1}$	$1.86 \cdot 10^{-1}$	$1.90 \cdot 10^{-1}$	$2.08 \cdot 10^{-1}$	$1.79 \cdot 10^{-1}$	$1.89 \cdot 10^{-1}$	$2.19 \cdot 10^{-1}$	$9.12 \cdot 10^{-1}$	$2.11 \cdot 10^{-1}$	$2.41 \cdot 10^{-1}$
Tilt	μ	$8.73 \cdot 10^{+1}$	$8.87 \cdot 10^{+1}$	$8.74 \cdot 10^{+1}$	$8.82 \cdot 10^{+1}$	$8.90 \cdot 10^{+1}$	$8.89 \cdot 10^{+1}$	$8.84 \cdot 10^{+1}$	$8.74 \cdot 10^{+1}$	$8.61 \cdot 10^{+1}$	$8.79 \cdot 10^{+1}$	$8.57 \cdot 10^{+1}$	$8.37 \cdot 10^{+1}$
	σ	$6.66 \cdot 10^{+0}$	$7.84 \cdot 10^{+0}$	$6.44 \cdot 10^{+0}$	$7.84 \cdot 10^{+0}$	$7.00 \cdot 10^{+0}$	$8.93 \cdot 10^{+0}$	$6.32 \cdot 10^{+0}$	$8.01 \cdot 10^{+0}$	$9.18 \cdot 10^{+0}$	$1.87 \cdot 10^{+1}$	$7.91 \cdot 10^{+0}$	$1.15 \cdot 10^{+1}$
Spin	μ	$1.72 \cdot 10^{+2}$	$1.65 \cdot 10^{+2}$	$1.74 \cdot 10^{+2}$	$1.68 \cdot 10^{+2}$	$1.75 \cdot 10^{+2}$	$1.57 \cdot 10^{+2}$	$1.74 \cdot 10^{+2}$	$1.66 \cdot 10^{+2}$	$1.69 \cdot 10^{+2}$	$1.62 \cdot 10^{+2}$	$1.69 \cdot 10^{+2}$	$1.61 \cdot 10^{+2}$
	σ	$1.57 \cdot 10^{+1}$	$1.88 \cdot 10^{+1}$	$1.58 \cdot 10^{+1}$	$1.92 \cdot 10^{+1}$	$1.67 \cdot 10^{+1}$	$2.42 \cdot 10^{+1}$	$1.52 \cdot 10^{+1}$	$2.15 \cdot 10^{+1}$	$2.09 \cdot 10^{+1}$	$5.56 \cdot 10^{+1}$	$1.97 \cdot 10^{+1}$	$3.44 \cdot 10^{+1}$
Contacts peptide-HGs	μ	$6.73 \cdot 10^{-1}$	$7.13 \cdot 10^{-1}$	$6.84 \cdot 10^{-1}$	$7.25 \cdot 10^{-1}$	$6.54 \cdot 10^{-1}$	$6.62 \cdot 10^{-1}$	$6.71 \cdot 10^{-1}$	$7.04 \cdot 10^{-1}$	$5.70 \cdot 10^{-1}$	$4.64 \cdot 10^{-1}$	$5.80 \cdot 10^{-1}$	$5.38 \cdot 10^{-1}$
	σ	$8.24 \cdot 10^{-2}$	$9.09 \cdot 10^{-2}$	$8.75 \cdot 10^{-2}$	$9.83 \cdot 10^{-2}$	$8.41 \cdot 10^{-2}$	$9.31 \cdot 10^{-2}$	$8.37 \cdot 10^{-2}$	$9.38 \cdot 10^{-2}$	$9.38 \cdot 10^{-2}$	$1.30 \cdot 10^{-1}$	$9.91 \cdot 10^{-2}$	$1.21 \cdot 10^{-1}$
Contacts peptide-tails	μ	$6.35 \cdot 10^{-1}$	$5.86 \cdot 10^{-1}$	$6.14 \cdot 10^{-1}$	$5.68 \cdot 10^{-1}$	$6.12 \cdot 10^{-1}$	$5.61 \cdot 10^{-1}$	$6.21 \cdot 10^{-1}$	$5.78 \cdot 10^{-1}$	$5.14 \cdot 10^{-1}$	$3.19 \cdot 10^{-1}$	$5.41 \cdot 10^{-1}$	$4.34 \cdot 10^{-1}$
	σ	$8.37 \cdot 10^{-2}$	$9.59 \cdot 10^{-2}$	$8.23 \cdot 10^{-2}$	$9.22 \cdot 10^{-2}$	$8.27 \cdot 10^{-2}$	$9.83 \cdot 10^{-2}$	$8.10 \cdot 10^{-2}$	$9.48 \cdot 10^{-2}$	$9.44 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$	$8.86 \cdot 10^{-2}$	$1.04 \cdot 10^{-1}$
Contacts peptide-water	μ	$1.42 \cdot 10^{+0}$	$1.54 \cdot 10^{+0}$	$1.46 \cdot 10^{+0}$	$1.59 \cdot 10^{+0}$	$1.77 \cdot 10^{+0}$	$1.98 \cdot 10^{+0}$	$1.67 \cdot 10^{+0}$	$1.86 \cdot 10^{+0}$	$2.15 \cdot 10^{+0}$	$2.92 \cdot 10^{+0}$	$2.09 \cdot 10^{+0}$	$2.56 \cdot 10^{+0}$
	σ	$2.01 \cdot 10^{-1}$	$2.40 \cdot 10^{-1}$	$2.14 \cdot 10^{-1}$	$2.48 \cdot 10^{-1}$	$2.26 \cdot 10^{-1}$	$2.93 \cdot 10^{-1}$	$2.26 \cdot 10^{-1}$	$2.81 \cdot 10^{-1}$	$2.75 \cdot 10^{-1}$	$3.52 \cdot 10^{-1}$	$2.77 \cdot 10^{-1}$	$3.43 \cdot 10^{-1}$
EDM	μ	$3.80 \cdot 10^{-2}$	$4.07 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$	$3.78 \cdot 10^{-2}$	$3.50 \cdot 10^{-2}$	$3.69 \cdot 10^{-2}$	$3.35 \cdot 10^{-2}$	$3.65 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$	$4.89 \cdot 10^{-2}$	$3.91 \cdot 10^{-2}$	$4.60 \cdot 10^{-2}$
	σ	$1.81 \cdot 10^{-2}$	$2.06 \cdot 10^{-2}$	$1.81 \cdot 10^{-2}$	$2.10 \cdot 10^{-2}$	$1.97 \cdot 10^{-2}$	$2.23 \cdot 10^{-2}$	$1.87 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	$2.43 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$2.39 \cdot 10^{-2}$	$2.68 \cdot 10^{-2}$
Long. EDM	μ	$4.76 \cdot 10^{-2}$	$4.40 \cdot 10^{-2}$	$5.06 \cdot 10^{-2}$	$4.14 \cdot 10^{-2}$	$4.65 \cdot 10^{-2}$	$3.85 \cdot 10^{-2}$	$4.45 \cdot 10^{-2}$	$3.79 \cdot 10^{-2}$	$5.02 \cdot 10^{-2}$	$4.45 \cdot 10^{-2}$	$4.94 \cdot 10^{-2}$	$4.11 \cdot 10^{-2}$
	σ	$2.02 \cdot 10^{-2}$	$2.12 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$2.18 \cdot 10^{-2}$	$2.20 \cdot 10^{-2}$	$2.31 \cdot 10^{-2}$	$2.08 \cdot 10^{-2}$	$2.23 \cdot 10^{-2}$	$2.69 \cdot 10^{-2}$	$2.83 \cdot 10^{-2}$	$2.63 \cdot 10^{-2}$	$2.92 \cdot 10^{-2}$
Transv. EDM	μ	$9.60 \cdot 10^{-3}$	$6.07 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$4.32 \cdot 10^{-3}$	$8.73 \cdot 10^{-3}$	$3.29 \cdot 10^{-3}$	$1.06 \cdot 10^{-2}$	$2.77 \cdot 10^{-3}$	$9.70 \cdot 10^{-3}$	$1.32 \cdot 10^{-3}$	$9.61 \cdot 10^{-3}$	$3.91 \cdot 10^{-3}$
	σ	$1.18 \cdot 10^{-2}$	$1.42 \cdot 10^{-2}$	$1.17 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$1.29 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$1.22 \cdot 10^{-2}$	$1.37 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.72 \cdot 10^{-2}$	$1.67 \cdot 10^{-2}$	$1.81 \cdot 10^{-2}$
HDM	μ	$2.52 \cdot 10^{-3}$	$5.79 \cdot 10^{-5}$	$2.38 \cdot 10^{-3}$	$4.40 \cdot 10^{-5}$	$3.21 \cdot 10^{-3}$	$1.51 \cdot 10^{-4}$	$3.09 \cdot 10^{-3}$	$8.75 \cdot 10^{-4}$	$8.47 \cdot 10^{-5}$	$4.41 \cdot 10^{-3}$	$4.61 \cdot 10^{-4}$	$4.50 \cdot 10^{-3}$
	σ	$5.80 \cdot 10^{-3}$	$5.36 \cdot 10^{-3}$	$5.80 \cdot 10^{-3}$	$5.40 \cdot 10^{-3}$	$5.89 \cdot 10^{-3}$	$5.60 \cdot 10^{-3}$	$5.81 \cdot 10^{-3}$	$5.38 \cdot 10^{-3}$	$5.92 \cdot 10^{-3}$	$5.73 \cdot 10^{-3}$	$5.74 \cdot 10^{-3}$	$5.60 \cdot 10^{-3}$
Long. HDM	μ	$7.42 \cdot 10^{-3}$	$2.36 \cdot 10^{-3}$	$7.40 \cdot 10^{-3}$	$2.10 \cdot 10^{-3}$	$7.10 \cdot 10^{-3}$	$2.59 \cdot 10^{-3}$	$6.80 \cdot 10^{-3}$	$2.62 \cdot 10^{-3}$	$7.18 \cdot 10^{-3}$	$6.53 \cdot 10^{-3}$	$7.14 \cdot 10^{-3}$	$6.28 \cdot 10^{-3}$
	σ	$4.62 \cdot 10^{-3}$	$5.07 \cdot 10^{-3}$	$4.73 \cdot 10^{-3}$	$4.99 \cdot 10^{-3}$	$4.79 \cdot 10^{-3}$	$5.04 \cdot 10^{-3}$	$4.70 \cdot 10^{-3}$	$5.18 \cdot 10^{-3}$	$4.83 \cdot 10^{-3}$	$5.27 \cdot 10^{-3}$	$4.83 \cdot 10^{-3}$	$5.38 \cdot 10^{-3}$
Transv. HDM	μ	$8.09 \cdot 10^{-3}$	$5.05 \cdot 10^{-3}$	$8.38 \cdot 10^{-3}$	$5.10 \cdot 10^{-3}$	$8.38 \cdot 10^{-3}$	$5.42 \cdot 10^{-3}$	$8.27 \cdot 10^{-3}$	$5.69 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$3.00 \cdot 10^{-3}$	$1.08 \cdot 10^{-2}$	$2.38 \cdot 10^{-3}$
	σ	$5.85 \cdot 10^{-3}$	$5.35 \cdot 10^{-3}$	$5.70 \cdot 10^{-3}$	$5.50 \cdot 10^{-3}$	$6.09 \cdot 10^{-3}$	$5.54 \cdot 10^{-3}$	$5.89 \cdot 10^{-3}$	$5.41 \cdot 10^{-3}$	$5.99 \cdot 10^{-3}$	$5.57 \cdot 10^{-3}$	$6.02 \cdot 10^{-3}$	$5.44 \cdot 10^{-3}$
Tr. diffusion coef.	μ	$4.41 \cdot 10^{+0}$	$4.84 \cdot 10^{+0}$	$3.67 \cdot 10^{+0}$	$4.14 \cdot 10^{+0}$	$3.87 \cdot 10^{+0}$	$4.52 \cdot 10^{+0}$	$3.01 \cdot 10^{+0}$	$3.24 \cdot 10^{+0}$	$6.45 \cdot 10^{+0}$	$9.46 \cdot 10^{+0}$	$5.02 \cdot 10^{+0}$	$6.33 \cdot 10^{+0}$
	σ	$2.62 \cdot 10^{-2}$	$2.56 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$2.28 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$2.53 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$1.88 \cdot 10^{-2}$	$3.54 \cdot 10^{-2}$	$6.47 \cdot 10^{-2}$	$2.77 \cdot 10^{-2}$	$3.92 \cdot 10^{-2}$
Rot. diffusion coef.	μ	$5.73 \cdot 10^{+1}$	$8.31 \cdot 10^{+1}$	$4.67 \cdot 10^{+1}$	$7.10 \cdot 10^{+1}$	$3.95 \cdot 10^{+1}$	$6.43 \cdot 10^{+1}$	$2.86 \cdot 10^{+1}$	$4.27 \cdot 10^{+1}$	$1.20 \cdot 10^{+2}$	$3.20 \cdot 10^{+2}$	$8.32 \cdot 10^{+1}$	$1.64 \cdot 10^{+2}$
	σ	$5.07 \cdot 10^{-1}$	$7.16 \cdot 10^{-1}$	$4.19 \cdot 10^{-1}$	$6.49 \cdot 10^{-1}$	$3.75 \cdot 10^{-1}$	$5.65 \cdot 10^{-1}$	$2.36 \cdot 10^{-1}$	$3.53 \cdot 10^{-1}$	$1.11 \cdot 10^{+0}$	$4.28 \cdot 10^{+0}$	$8.20 \cdot 10^{-1}$	$1.56 \cdot 10^{+0}$

Table D.2: Median of the samples of MD descriptors. The samples are grouped by force field—M2, M2P or M3—, lipid composition—POPC or POPE:POPG—, and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Unless indicated otherwise, bilayer thickness and peptide penetration are always given in nm, area per lipid in nm^2, tilt and spin in degrees, EDM and its components in e nm, HDM and its components in nm, the translational diffusion coefficient in 10^{-7} $cm^2\,s^{-1}$ and the rotational diffusion coefficient in $deg^2\,ns^{-1}$.

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC
Bilayer thickness	μ	$8.56 \cdot 10^{-6}$	$1.05 \cdot 10^{-5}$	$4.98 \cdot 10^{-6}$	$1.07 \cdot 10^{-5}$	$1.16 \cdot 10^{-5}$	$1.34 \cdot 10^{-5}$	$6.77 \cdot 10^{-6}$	$1.39 \cdot 10^{-5}$	$1.01 \cdot 10^{-5}$	$1.06 \cdot 10^{-5}$	$5.66 \cdot 10^{-6}$	$8.40 \cdot 10^{-6}$
	σ	$8.68 \cdot 10^{-7}$	$9.11 \cdot 10^{-7}$	$7.25 \cdot 10^{-7}$	$6.05 \cdot 10^{-7}$	$9.13 \cdot 10^{-7}$	$1.37 \cdot 10^{-6}$	$9.51 \cdot 10^{-7}$	$8.01 \cdot 10^{-7}$	$1.18 \cdot 10^{-6}$	$9.47 \cdot 10^{-7}$	$1.17 \cdot 10^{-6}$	$9.68 \cdot 10^{-7}$
Area per lipid	μ	$6.38 \cdot 10^{-7}$	$6.11 \cdot 10^{-7}$	$2.40 \cdot 10^{-7}$	$3.32 \cdot 10^{-7}$	$1.02 \cdot 10^{-6}$	$9.83 \cdot 10^{-7}$	$5.10 \cdot 10^{-7}$	$8.26 \cdot 10^{-7}$	$3.77 \cdot 10^{-7}$	$4.28 \cdot 10^{-7}$	$1.93 \cdot 10^{-7}$	$1.95 \cdot 10^{-7}$
	σ	$4.15 \cdot 10^{-8}$	$3.95 \cdot 10^{-8}$	$3.10 \cdot 10^{-8}$	$3.23 \cdot 10^{-8}$	$4.21 \cdot 10^{-8}$	$5.94 \cdot 10^{-8}$	$4.31 \cdot 10^{-8}$	$3.12 \cdot 10^{-8}$	$6.06 \cdot 10^{-8}$	$5.78 \cdot 10^{-8}$	$4.52 \cdot 10^{-8}$	$5.26 \cdot 10^{-8}$
Peptide penetration	μ	$1.94 \cdot 10^{-1}$	$3.58 \cdot 10^{-2}$	$1.46 \cdot 10^{-2}$	$1.93 \cdot 10^{-2}$	$1.37 \cdot 10^{+0}$	$1.40 \cdot 10^{+0}$	$3.88 \cdot 10^{-2}$	$4.21 \cdot 10^{-1}$	$6.80 \cdot 10^{-1}$	$1.16 \cdot 10^{+0}$	$6.93 \cdot 10^{-2}$	$3.40 \cdot 10^{-1}$
	σ	$7.65 \cdot 10^{-3}$	$2.76 \cdot 10^{-4}$	$2.03 \cdot 10^{-4}$	$2.57 \cdot 10^{-4}$	$4.27 \cdot 10^{-2}$	$6.32 \cdot 10^{-2}$	$1.42 \cdot 10^{-3}$	$2.69 \cdot 10^{-2}$	$8.24 \cdot 10^{-2}$	$2.62 \cdot 10^{-1}$	$1.93 \cdot 10^{-2}$	$1.26 \cdot 10^{-1}$
Tilt	μ	$2.04 \cdot 10^{+1}$	$1.20 \cdot 10^{+2}$	$1.52 \cdot 10^{+1}$	$8.07 \cdot 10^{+1}$	$1.76 \cdot 10^{+1}$	$8.90 \cdot 10^{+1}$	$4.42 \cdot 10^{+1}$	$1.54 \cdot 10^{+2}$	$8.67 \cdot 10^{+1}$	$1.35 \cdot 10^{+2}$	$2.59 \cdot 10^{+1}$	$1.33 \cdot 10^{+2}$
	σ	$1.66 \cdot 10^{+1}$	$4.98 \cdot 10^{+0}$	$2.27 \cdot 10^{+0}$	$5.09 \cdot 10^{+0}$	$1.10 \cdot 10^{+2}$	$1.10 \cdot 10^{+2}$	$5.32 \cdot 10^{+0}$	$5.29 \cdot 10^{+1}$	$7.15 \cdot 10^{+1}$	$1.49 \cdot 10^{+2}$	$1.78 \cdot 10^{+1}$	$1.08 \cdot 10^{+2}$
Spin	μ	$1.08 \cdot 10^{+3}$	$1.88 \cdot 10^{+3}$	$8.77 \cdot 10^{+2}$	$2.33 \cdot 10^{+3}$	$1.61 \cdot 10^{+3}$	$2.36 \cdot 10^{+3}$	$9.61 \cdot 10^{+2}$	$4.70 \cdot 10^{+3}$	$6.46 \cdot 10^{+2}$	$3.22 \cdot 10^{+3}$	$8.68 \cdot 10^{+2}$	$3.30 \cdot 10^{+3}$
	σ	$3.82 \cdot 10^{+2}$	$1.06 \cdot 10^{+2}$	$3.87 \cdot 10^{+1}$	$1.22 \cdot 10^{+2}$	$1.50 \cdot 10^{+3}$	$1.53 \cdot 10^{+3}$	$1.26 \cdot 10^{+2}$	$6.77 \cdot 10^{+2}$	$1.06 \cdot 10^{+3}$	$2.02 \cdot 10^{+3}$	$2.54 \cdot 10^{+2}$	$1.33 \cdot 10^{+3}$
Contacts peptide-HGs	μ	$1.40 \cdot 10^{-2}$	$1.22 \cdot 10^{-2}$	$8.07 \cdot 10^{-3}$	$1.12 \cdot 10^{-2}$	$5.70 \cdot 10^{-2}$	$6.71 \cdot 10^{-2}$	$9.21 \cdot 10^{-3}$	$3.63 \cdot 10^{-2}$	$3.36 \cdot 10^{-2}$	$5.08 \cdot 10^{-2}$	$1.03 \cdot 10^{-2}$	$2.52 \cdot 10^{-2}$
	σ	$4.55 \cdot 10^{-4}$	$6.57 \cdot 10^{-4}$	$4.58 \cdot 10^{-4}$	$6.81 \cdot 10^{-4}$	$9.09 \cdot 10^{-4}$	$1.23 \cdot 10^{-3}$	$4.67 \cdot 10^{-4}$	$7.53 \cdot 10^{-4}$	$1.12 \cdot 10^{-3}$	$5.27 \cdot 10^{-3}$	$7.25 \cdot 10^{-4}$	$4.00 \cdot 10^{-3}$
Contacts peptide-tails	μ	$1.51 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$	$8.95 \cdot 10^{-3}$	$1.12 \cdot 10^{-2}$	$5.12 \cdot 10^{-2}$	$5.11 \cdot 10^{-2}$	$1.18 \cdot 10^{-2}$	$3.26 \cdot 10^{-2}$	$3.78 \cdot 10^{-2}$	$4.65 \cdot 10^{-2}$	$1.74 \cdot 10^{-2}$	$3.31 \cdot 10^{-2}$
	σ	$5.77 \cdot 10^{-4}$	$7.05 \cdot 10^{-4}$	$4.29 \cdot 10^{-4}$	$6.20 \cdot 10^{-4}$	$1.26 \cdot 10^{-3}$	$1.79 \cdot 10^{-3}$	$4.43 \cdot 10^{-4}$	$1.07 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	$4.95 \cdot 10^{-3}$	$5.62 \cdot 10^{-4}$	$2.70 \cdot 10^{-3}$
Contacts peptide-water	μ	$1.95 \cdot 10^{-1}$	$9.55 \cdot 10^{-2}$	$6.43 \cdot 10^{-2}$	$7.03 \cdot 10^{-2}$	$1.19 \cdot 10^{+0}$	$1.32 \cdot 10^{+0}$	$1.98 \cdot 10^{-1}$	$6.99 \cdot 10^{-1}$	$6.29 \cdot 10^{-1}$	$9.94 \cdot 10^{-1}$	$1.86 \cdot 10^{-1}$	$5.30 \cdot 10^{-1}$
	σ	$1.68 \cdot 10^{-3}$	$4.28 \cdot 10^{-3}$	$1.94 \cdot 10^{-3}$	$3.68 \cdot 10^{-3}$	$2.56 \cdot 10^{-3}$	$1.10 \cdot 10^{-2}$	$2.56 \cdot 10^{-3}$	$1.00 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$5.19 \cdot 10^{-2}$	$4.92 \cdot 10^{-3}$	$3.82 \cdot 10^{-2}$
EDM	μ	$9.21 \cdot 10^{-3}$	$1.05 \cdot 10^{-3}$	$1.20 \cdot 10^{-3}$	$1.11 \cdot 10^{-3}$	$1.30 \cdot 10^{-2}$	$4.54 \cdot 10^{-3}$	$1.18 \cdot 10^{-3}$	$1.37 \cdot 10^{-3}$	$1.26 \cdot 10^{-3}$	$8.49 \cdot 10^{-3}$	$1.13 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$
	σ	$1.37 \cdot 10^{-3}$	$3.74 \cdot 10^{-5}$	$3.85 \cdot 10^{-5}$	$3.63 \cdot 10^{-5}$	$1.44 \cdot 10^{-4}$	$9.04 \cdot 10^{-4}$	$4.36 \cdot 10^{-5}$	$5.10 \cdot 10^{-5}$	$5.86 \cdot 10^{-5}$	$8.20 \cdot 10^{-4}$	$5.98 \cdot 10^{-5}$	$6.00 \cdot 10^{-5}$
Long. EDM	μ	$4.28 \cdot 10^{-3}$	$1.55 \cdot 10^{-3}$	$1.61 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$	$1.40 \cdot 10^{-2}$	$2.41 \cdot 10^{-3}$	$1.52 \cdot 10^{-3}$	$1.64 \cdot 10^{-3}$	$1.95 \cdot 10^{-3}$	$3.48 \cdot 10^{-3}$	$1.66 \cdot 10^{-3}$	$1.36 \cdot 10^{-3}$
	σ	$1.67 \cdot 10^{-3}$	$5.59 \cdot 10^{-5}$	$6.27 \cdot 10^{-5}$	$5.22 \cdot 10^{-5}$	$1.57 \cdot 10^{-4}$	$1.12 \cdot 10^{-3}$	$6.33 \cdot 10^{-5}$	$6.77 \cdot 10^{-5}$	$9.30 \cdot 10^{-5}$	$1.07 \cdot 10^{-3}$	$8.87 \cdot 10^{-5}$	$7.14 \cdot 10^{-5}$
Transv. EDM	μ	$6.40 \cdot 10^{-3}$	$5.98 \cdot 10^{-4}$	$4.87 \cdot 10^{-4}$	$6.13 \cdot 10^{-4}$	$4.72 \cdot 10^{-4}$	$3.90 \cdot 10^{-3}$	$5.08 \cdot 10^{-4}$	$7.79 \cdot 10^{-4}$	$3.95 \cdot 10^{-4}$	$5.88 \cdot 10^{-3}$	$3.75 \cdot 10^{-4}$	$5.99 \cdot 10^{-4}$
	σ	$1.19 \cdot 10^{-3}$	$3.58 \cdot 10^{-5}$	$2.62 \cdot 10^{-5}$	$3.69 \cdot 10^{-5}$	$2.62 \cdot 10^{-5}$	$7.65 \cdot 10^{-4}$	$2.44 \cdot 10^{-5}$	$3.91 \cdot 10^{-5}$	$4.36 \cdot 10^{-5}$	$7.50 \cdot 10^{-4}$	$4.14 \cdot 10^{-5}$	$5.45 \cdot 10^{-5}$
HDM	μ	$1.47 \cdot 10^{-4}$	$1.13 \cdot 10^{-4}$	$1.46 \cdot 10^{-4}$	$1.13 \cdot 10^{-4}$	$1.54 \cdot 10^{-4}$	$1.13 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	$1.52 \cdot 10^{-4}$	$1.02 \cdot 10^{-4}$	$1.62 \cdot 10^{-4}$	$1.07 \cdot 10^{-4}$
	σ	$3.80 \cdot 10^{-6}$	$1.89 \cdot 10^{-6}$	$3.60 \cdot 10^{-6}$	$2.06 \cdot 10^{-6}$	$5.05 \cdot 10^{-6}$	$1.94 \cdot 10^{-6}$	$3.50 \cdot 10^{-6}$	$1.93 \cdot 10^{-6}$	$4.85 \cdot 10^{-6}$	$1.77 \cdot 10^{-6}$	$4.39 \cdot 10^{-6}$	$1.93 \cdot 10^{-6}$
Long. HDM	μ	$5.24 \cdot 10^{-5}$	$6.83 \cdot 10^{-5}$	$5.15 \cdot 10^{-5}$	$6.80 \cdot 10^{-5}$	$5.81 \cdot 10^{-5}$	$7.29 \cdot 10^{-5}$	$5.76 \cdot 10^{-5}$	$8.05 \cdot 10^{-5}$	$6.11 \cdot 10^{-5}$	$5.51 \cdot 10^{-5}$	$6.29 \cdot 10^{-5}$	$5.27 \cdot 10^{-5}$
	σ	$2.44 \cdot 10^{-6}$	$1.83 \cdot 10^{-6}$	$2.46 \cdot 10^{-6}$	$2.05 \cdot 10^{-6}$	$1.36 \cdot 10^{-5}$	$2.01 \cdot 10^{-6}$	$2.55 \cdot 10^{-6}$	$2.00 \cdot 10^{-6}$	$2.81 \cdot 10^{-6}$	$1.95 \cdot 10^{-6}$	$2.77 \cdot 10^{-6}$	$2.08 \cdot 10^{-6}$
Transv. HDM	μ	$3.73 \cdot 10^{-4}$	$3.33 \cdot 10^{-4}$	$3.64 \cdot 10^{-4}$	$3.31 \cdot 10^{-4}$	$3.69 \cdot 10^{-4}$	$3.29 \cdot 10^{-4}$	$3.60 \cdot 10^{-4}$	$3.36 \cdot 10^{-4}$	$2.71 \cdot 10^{-4}$	$2.68 \cdot 10^{-4}$	$2.87 \cdot 10^{-4}$	$2.71 \cdot 10^{-4}$
	σ	$7.68 \cdot 10^{-6}$	$5.77 \cdot 10^{-6}$	$7.73 \cdot 10^{-6}$	$5.63 \cdot 10^{-6}$	$8.12 \cdot 10^{-6}$	$5.62 \cdot 10^{-6}$	$7.32 \cdot 10^{-6}$	$5.73 \cdot 10^{-6}$	$1.05 \cdot 10^{-5}$	$6.04 \cdot 10^{-6}$	$8.89 \cdot 10^{-6}$	$5.91 \cdot 10^{-6}$
Tr. diffusion coef.	μ	$3.51 \cdot 10^{+0}$	$1.71 \cdot 10^{+0}$	$3.79 \cdot 10^{-1}$	$6.86 \cdot 10^{-1}$	$6.10 \cdot 10^{+1}$	$8.82 \cdot 10^{+1}$	$6.34 \cdot 10^{-1}$	$1.44 \cdot 10^{+1}$	$2.76 \cdot 10^{+1}$	$9.35 \cdot 10^{+1}$	$2.16 \cdot 10^{+0}$	$2.49 \cdot 10^{+1}$
	σ	$2.94 \cdot 10^{-4}$	$1.15 \cdot 10^{-4}$	$1.44 \cdot 10^{-5}$	$3.59 \cdot 10^{-5}$	$3.38 \cdot 10^{-3}$	$3.29 \cdot 10^{-3}$	$4.03 \cdot 10^{-5}$	$4.96 \cdot 10^{-4}$	$2.06 \cdot 10^{-3}$	$7.93 \cdot 10^{-3}$	$1.95 \cdot 10^{-4}$	$2.59 \cdot 10^{-3}$
Rot. diffusion coef.	μ	$4.58 \cdot 10^{+3}$	$4.35 \cdot 10^{+3}$	$1.57 \cdot 10^{+3}$	$3.32 \cdot 10^{+3}$	$6.07 \cdot 10^{+4}$	$2.37 \cdot 10^{+5}$	$3.34 \cdot 10^{+3}$	$1.03 \cdot 10^{+5}$	$5.15 \cdot 10^{+4}$	$5.78 \cdot 10^{+5}$	$6.75 \cdot 10^{+3}$	$1.78 \cdot 10^{+5}$
	σ	$7.71 \cdot 10^{-1}$	$4.08 \cdot 10^{-1}$	$1.24 \cdot 10^{-1}$	$4.13 \cdot 10^{-1}$	$1.64 \cdot 10^{+1}$	$7.13 \cdot 10^{+1}$	$1.11 \cdot 10^{+0}$	$3.58 \cdot 10^{+1}$	$2.20 \cdot 10^{+1}$	$2.22 \cdot 10^{+2}$	$2.01 \cdot 10^{+0}$	$1.04 \cdot 10^{+2}$

Table D.3: Variance of the samples of MD descriptors. The samples are grouped by force field—M2, M2P or M3—, lipid composition—POPC or POPE:POPG—, and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Same units as in Table D.2, but squared.

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC
Bilayer thickness	μ	$2.93 \cdot 10^{-3}$	$3.24 \cdot 10^{-3}$	$2.23 \cdot 10^{-3}$	$3.28 \cdot 10^{-3}$	$3.40 \cdot 10^{-3}$	$3.66 \cdot 10^{-3}$	$2.60 \cdot 10^{-3}$	$3.73 \cdot 10^{-3}$	$3.17 \cdot 10^{-3}$	$3.25 \cdot 10^{-3}$	$2.38 \cdot 10^{-3}$	$2.90 \cdot 10^{-3}$
	σ	$9.32 \cdot 10^{-4}$	$9.54 \cdot 10^{-4}$	$8.52 \cdot 10^{-4}$	$7.78 \cdot 10^{-4}$	$9.55 \cdot 10^{-4}$	$1.17 \cdot 10^{-3}$	$9.75 \cdot 10^{-4}$	$8.95 \cdot 10^{-4}$	$1.09 \cdot 10^{-3}$	$9.73 \cdot 10^{-4}$	$1.08 \cdot 10^{-3}$	$9.84 \cdot 10^{-4}$
Area per lipid	μ	$7.99 \cdot 10^{-4}$	$7.82 \cdot 10^{-4}$	$4.90 \cdot 10^{-4}$	$5.76 \cdot 10^{-4}$	$1.01 \cdot 10^{-3}$	$9.91 \cdot 10^{-4}$	$7.14 \cdot 10^{-4}$	$9.09 \cdot 10^{-4}$	$6.14 \cdot 10^{-4}$	$6.54 \cdot 10^{-4}$	$4.40 \cdot 10^{-4}$	$4.42 \cdot 10^{-4}$
	σ	$2.04 \cdot 10^{-4}$	$1.99 \cdot 10^{-4}$	$1.76 \cdot 10^{-4}$	$1.80 \cdot 10^{-4}$	$2.05 \cdot 10^{-4}$	$2.44 \cdot 10^{-4}$	$2.08 \cdot 10^{-4}$	$1.77 \cdot 10^{-4}$	$2.46 \cdot 10^{-4}$	$2.40 \cdot 10^{-4}$	$2.13 \cdot 10^{-4}$	$2.29 \cdot 10^{-4}$
Peptide penetration	μ	$4.40 \cdot 10^{-1}$	$1.89 \cdot 10^{-1}$	$1.21 \cdot 10^{-1}$	$1.39 \cdot 10^{-1}$	$1.17 \cdot 10^{+0}$	$1.18 \cdot 10^{+0}$	$1.97 \cdot 10^{-1}$	$6.49 \cdot 10^{-1}$	$8.25 \cdot 10^{-1}$	$1.08 \cdot 10^{+0}$	$2.63 \cdot 10^{-1}$	$5.83 \cdot 10^{-1}$
	σ	$8.75 \cdot 10^{-2}$	$1.66 \cdot 10^{-2}$	$1.42 \cdot 10^{-2}$	$1.60 \cdot 10^{-2}$	$2.07 \cdot 10^{-1}$	$2.51 \cdot 10^{-1}$	$3.76 \cdot 10^{-2}$	$1.64 \cdot 10^{-1}$	$2.87 \cdot 10^{-1}$	$5.11 \cdot 10^{-1}$	$1.39 \cdot 10^{-1}$	$3.55 \cdot 10^{-1}$
Tilt	μ	$4.51 \cdot 10^{+0}$	$1.09 \cdot 10^{+1}$	$3.89 \cdot 10^{+0}$	$8.99 \cdot 10^{+0}$	$4.20 \cdot 10^{+0}$	$9.44 \cdot 10^{+0}$	$6.65 \cdot 10^{+0}$	$1.24 \cdot 10^{+1}$	$9.31 \cdot 10^{+0}$	$1.16 \cdot 10^{+1}$	$5.09 \cdot 10^{+0}$	$1.16 \cdot 10^{+1}$
	σ	$4.08 \cdot 10^{+0}$	$2.23 \cdot 10^{+0}$	$1.51 \cdot 10^{+0}$	$2.26 \cdot 10^{+0}$	$1.05 \cdot 10^{+1}$	$1.05 \cdot 10^{+1}$	$2.31 \cdot 10^{+0}$	$7.27 \cdot 10^{+0}$	$8.46 \cdot 10^{+0}$	$1.22 \cdot 10^{+1}$	$4.22 \cdot 10^{+0}$	$1.04 \cdot 10^{+1}$
Spin	μ	$3.28 \cdot 10^{+1}$	$4.34 \cdot 10^{+1}$	$2.96 \cdot 10^{+1}$	$4.83 \cdot 10^{+1}$	$4.01 \cdot 10^{+1}$	$4.86 \cdot 10^{+1}$	$3.10 \cdot 10^{+1}$	$6.86 \cdot 10^{+1}$	$2.54 \cdot 10^{+1}$	$5.67 \cdot 10^{+1}$	$2.95 \cdot 10^{+1}$	$5.74 \cdot 10^{+1}$
	σ	$1.95 \cdot 10^{+1}$	$1.03 \cdot 10^{+1}$	$6.22 \cdot 10^{+0}$	$1.11 \cdot 10^{+1}$	$3.87 \cdot 10^{+1}$	$3.91 \cdot 10^{+1}$	$1.12 \cdot 10^{+1}$	$2.60 \cdot 10^{+1}$	$3.26 \cdot 10^{+1}$	$4.49 \cdot 10^{+1}$	$1.59 \cdot 10^{+1}$	$3.65 \cdot 10^{+1}$
Contacts peptide-HGs	μ	$1.18 \cdot 10^{-1}$	$1.11 \cdot 10^{-1}$	$8.98 \cdot 10^{-2}$	$1.06 \cdot 10^{-1}$	$2.39 \cdot 10^{-1}$	$2.59 \cdot 10^{-1}$	$9.60 \cdot 10^{-2}$	$1.91 \cdot 10^{-1}$	$1.83 \cdot 10^{-1}$	$2.25 \cdot 10^{-1}$	$1.02 \cdot 10^{-1}$	$1.59 \cdot 10^{-1}$
	σ	$2.13 \cdot 10^{-2}$	$2.56 \cdot 10^{-2}$	$2.14 \cdot 10^{-2}$	$2.61 \cdot 10^{-2}$	$3.01 \cdot 10^{-2}$	$3.51 \cdot 10^{-2}$	$2.16 \cdot 10^{-2}$	$2.74 \cdot 10^{-2}$	$3.34 \cdot 10^{-2}$	$7.26 \cdot 10^{-2}$	$2.69 \cdot 10^{-2}$	$6.33 \cdot 10^{-2}$
Contacts peptide-tails	μ	$1.23 \cdot 10^{-1}$	$1.17 \cdot 10^{-1}$	$9.46 \cdot 10^{-2}$	$1.06 \cdot 10^{-1}$	$2.26 \cdot 10^{-1}$	$2.26 \cdot 10^{-1}$	$1.09 \cdot 10^{-1}$	$1.80 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$	$2.16 \cdot 10^{-1}$	$1.32 \cdot 10^{-1}$	$1.82 \cdot 10^{-1}$
	σ	$2.40 \cdot 10^{-2}$	$2.65 \cdot 10^{-2}$	$2.07 \cdot 10^{-2}$	$2.49 \cdot 10^{-2}$	$3.55 \cdot 10^{-2}$	$4.23 \cdot 10^{-2}$	$2.10 \cdot 10^{-2}$	$3.27 \cdot 10^{-2}$	$3.67 \cdot 10^{-2}$	$7.03 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$	$5.19 \cdot 10^{-2}$
Contacts peptide-water	μ	$4.42 \cdot 10^{-1}$	$3.09 \cdot 10^{-1}$	$2.54 \cdot 10^{-1}$	$2.65 \cdot 10^{-1}$	$1.09 \cdot 10^{+0}$	$1.15 \cdot 10^{+0}$	$4.45 \cdot 10^{-1}$	$8.36 \cdot 10^{-1}$	$7.93 \cdot 10^{-1}$	$9.97 \cdot 10^{-1}$	$4.32 \cdot 10^{-1}$	$7.28 \cdot 10^{-1}$
	σ	$4.09 \cdot 10^{-2}$	$6.54 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$	$6.06 \cdot 10^{-2}$	$5.06 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$	$5.06 \cdot 10^{-2}$	$1.00 \cdot 10^{-1}$	$1.13 \cdot 10^{-1}$	$2.28 \cdot 10^{-1}$	$7.01 \cdot 10^{-2}$	$1.95 \cdot 10^{-1}$
EDM	μ	$9.60 \cdot 10^{-2}$	$3.24 \cdot 10^{-2}$	$3.46 \cdot 10^{-2}$	$3.34 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$	$6.74 \cdot 10^{-2}$	$3.44 \cdot 10^{-2}$	$3.70 \cdot 10^{-2}$	$3.56 \cdot 10^{-2}$	$9.22 \cdot 10^{-2}$	$3.37 \cdot 10^{-2}$	$3.20 \cdot 10^{-2}$
	σ	$3.71 \cdot 10^{-2}$	$6.11 \cdot 10^{-3}$	$6.21 \cdot 10^{-3}$	$6.02 \cdot 10^{-3}$	$1.20 \cdot 10^{-2}$	$3.01 \cdot 10^{-2}$	$6.60 \cdot 10^{-3}$	$7.14 \cdot 10^{-3}$	$7.65 \cdot 10^{-3}$	$2.86 \cdot 10^{-2}$	$7.73 \cdot 10^{-3}$	$7.74 \cdot 10^{-3}$
Long. EDM	μ	$6.55 \cdot 10^{-2}$	$3.93 \cdot 10^{-2}$	$4.01 \cdot 10^{-2}$	$3.98 \cdot 10^{-2}$	$1.18 \cdot 10^{-1}$	$4.91 \cdot 10^{-2}$	$3.90 \cdot 10^{-2}$	$4.05 \cdot 10^{-2}$	$4.41 \cdot 10^{-2}$	$5.90 \cdot 10^{-2}$	$4.07 \cdot 10^{-2}$	$3.69 \cdot 10^{-2}$
	σ	$4.08 \cdot 10^{-2}$	$7.47 \cdot 10^{-3}$	$7.92 \cdot 10^{-3}$	$7.23 \cdot 10^{-3}$	$1.25 \cdot 10^{-2}$	$3.35 \cdot 10^{-2}$	$7.96 \cdot 10^{-3}$	$8.23 \cdot 10^{-3}$	$9.64 \cdot 10^{-3}$	$3.27 \cdot 10^{-2}$	$9.42 \cdot 10^{-3}$	$8.45 \cdot 10^{-3}$
Transv. EDM	μ	$8.00 \cdot 10^{-2}$	$2.44 \cdot 10^{-2}$	$2.21 \cdot 10^{-2}$	$2.48 \cdot 10^{-2}$	$2.17 \cdot 10^{-2}$	$6.24 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$	$2.79 \cdot 10^{-2}$	$1.99 \cdot 10^{-2}$	$7.67 \cdot 10^{-2}$	$1.94 \cdot 10^{-2}$	$2.45 \cdot 10^{-2}$
	σ	$3.45 \cdot 10^{-2}$	$5.98 \cdot 10^{-3}$	$5.12 \cdot 10^{-3}$	$6.07 \cdot 10^{-3}$	$5.12 \cdot 10^{-3}$	$2.77 \cdot 10^{-2}$	$4.94 \cdot 10^{-3}$	$6.25 \cdot 10^{-3}$	$6.60 \cdot 10^{-3}$	$2.74 \cdot 10^{-2}$	$6.44 \cdot 10^{-3}$	$7.38 \cdot 10^{-3}$
HDM	μ	$1.21 \cdot 10^{-2}$	$1.07 \cdot 10^{-2}$	$1.21 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$1.24 \cdot 10^{-2}$	$1.06 \cdot 10^{-2}$	$1.22 \cdot 10^{-2}$	$1.14 \cdot 10^{-2}$	$1.23 \cdot 10^{-2}$	$1.01 \cdot 10^{-2}$	$1.27 \cdot 10^{-2}$	$1.03 \cdot 10^{-2}$
	σ	$1.95 \cdot 10^{-3}$	$1.37 \cdot 10^{-3}$	$1.90 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$2.25 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$	$1.87 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$	$2.20 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$	$2.10 \cdot 10^{-3}$	$1.39 \cdot 10^{-3}$
Long. HDM	μ	$7.24 \cdot 10^{-3}$	$8.26 \cdot 10^{-3}$	$7.18 \cdot 10^{-3}$	$8.25 \cdot 10^{-3}$	$7.62 \cdot 10^{-3}$	$8.54 \cdot 10^{-3}$	$7.59 \cdot 10^{-3}$	$8.97 \cdot 10^{-3}$	$7.82 \cdot 10^{-3}$	$7.42 \cdot 10^{-3}$	$7.93 \cdot 10^{-3}$	$7.26 \cdot 10^{-3}$
	σ	$1.56 \cdot 10^{-3}$	$1.35 \cdot 10^{-3}$	$1.57 \cdot 10^{-3}$	$1.43 \cdot 10^{-3}$	$3.69 \cdot 10^{-3}$	$1.42 \cdot 10^{-3}$	$1.60 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$	$1.68 \cdot 10^{-3}$	$1.40 \cdot 10^{-3}$	$1.67 \cdot 10^{-3}$	$1.44 \cdot 10^{-3}$
Transv. HDM	μ	$1.93 \cdot 10^{-2}$	$1.82 \cdot 10^{-2}$	$1.91 \cdot 10^{-2}$	$1.82 \cdot 10^{-2}$	$1.92 \cdot 10^{-2}$	$1.81 \cdot 10^{-2}$	$1.90 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$	$1.64 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$1.65 \cdot 10^{-2}$
	σ	$2.77 \cdot 10^{-3}$	$2.40 \cdot 10^{-3}$	$2.78 \cdot 10^{-3}$	$2.37 \cdot 10^{-3}$	$2.85 \cdot 10^{-3}$	$2.37 \cdot 10^{-3}$	$2.70 \cdot 10^{-3}$	$2.39 \cdot 10^{-3}$	$3.25 \cdot 10^{-3}$	$2.46 \cdot 10^{-3}$	$2.98 \cdot 10^{-3}$	$2.43 \cdot 10^{-3}$
Tr. diffusion coef.	μ	$1.87 \cdot 10^{+0}$	$1.31 \cdot 10^{+0}$	$6.16 \cdot 10^{-1}$	$8.28 \cdot 10^{-1}$	$7.81 \cdot 10^{+0}$	$9.39 \cdot 10^{+0}$	$7.96 \cdot 10^{-1}$	$3.79 \cdot 10^{+0}$	$5.25 \cdot 10^{+0}$	$9.67 \cdot 10^{+0}$	$1.47 \cdot 10^{+0}$	$4.99 \cdot 10^{+0}$
	σ	$1.72 \cdot 10^{-2}$	$1.07 \cdot 10^{-2}$	$3.80 \cdot 10^{-3}$	$5.99 \cdot 10^{-3}$	$5.81 \cdot 10^{-2}$	$5.74 \cdot 10^{-2}$	$6.35 \cdot 10^{-3}$	$2.23 \cdot 10^{-2}$	$4.54 \cdot 10^{-2}$	$8.91 \cdot 10^{-2}$	$1.40 \cdot 10^{-2}$	$5.09 \cdot 10^{-2}$
Rot. diffusion coef.	μ	$6.77 \cdot 10^{+1}$	$6.60 \cdot 10^{+1}$	$3.96 \cdot 10^{+1}$	$5.76 \cdot 10^{+1}$	$2.46 \cdot 10^{+2}$	$4.86 \cdot 10^{+2}$	$5.78 \cdot 10^{+1}$	$3.22 \cdot 10^{+2}$	$2.27 \cdot 10^{+2}$	$7.60 \cdot 10^{+2}$	$8.22 \cdot 10^{+1}$	$4.22 \cdot 10^{+2}$
	σ	$8.78 \cdot 10^{-1}$	$6.39 \cdot 10^{-1}$	$3.52 \cdot 10^{-1}$	$6.43 \cdot 10^{-1}$	$4.04 \cdot 10^{+0}$	$8.44 \cdot 10^{+0}$	$1.06 \cdot 10^{+0}$	$5.99 \cdot 10^{+0}$	$4.69 \cdot 10^{+0}$	$1.49 \cdot 10^{+1}$	$1.42 \cdot 10^{+0}$	$1.02 \cdot 10^{+1}$

Table D.4: Standard deviation of the samples of MD descriptors. The samples are grouped by force field—M2, M2P or M3—, lipid composition—POPC or POPE:POPG—, and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Same units as in Table D.2.

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC
Bilayer thickness	μ	$9.85 \cdot 10^{-1}$	$9.60 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.75 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$
	σ	$9.92 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$	$9.69 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$9.90 \cdot 10^{-1}$	$9.73 \cdot 10^{-1}$	$9.88 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.57 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.89 \cdot 10^{-1}$
Area per lipid	μ	$9.89 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.85 \cdot 10^{-1}$	$9.62 \cdot 10^{-1}$	$9.55 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$	$9.79 \cdot 10^{-1}$	$9.59 \cdot 10^{-1}$	$9.48 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$
	σ	$9.82 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.79 \cdot 10^{-1}$	$9.90 \cdot 10^{-1}$	$9.61 \cdot 10^{-1}$	$9.89 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$	$9.88 \cdot 10^{-1}$	$9.72 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$
Peptide penetration	μ	$3.50 \cdot 10^{-1}$	$9.22 \cdot 10^{-1}$	$9.83 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$	$5.30 \cdot 10^{-1}$	$6.07 \cdot 10^{-1}$	$5.04 \cdot 10^{-1}$	$5.83 \cdot 10^{-1}$	$8.44 \cdot 10^{-1}$	$6.48 \cdot 10^{-1}$	$7.63 \cdot 10^{-1}$
	σ	$2.19 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$	$4.69 \cdot 10^{-1}$	$5.34 \cdot 10^{-1}$	$3.95 \cdot 10^{-1}$	$3.97 \cdot 10^{-1}$	$4.78 \cdot 10^{-1}$	$8.17 \cdot 10^{-1}$	$2.53 \cdot 10^{-1}$	$6.69 \cdot 10^{-1}$
Tilt	μ	$9.77 \cdot 10^{-1}$	$8.78 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$9.55 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$8.99 \cdot 10^{-1}$	$6.76 \cdot 10^{-1}$	$8.67 \cdot 10^{-1}$	$8.99 \cdot 10^{-1}$	$9.24 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.43 \cdot 10^{-1}$
	σ	$4.29 \cdot 10^{-1}$	$9.53 \cdot 10^{-1}$	$9.58 \cdot 10^{-1}$	$9.63 \cdot 10^{-1}$	$5.41 \cdot 10^{-1}$	$6.21 \cdot 10^{-1}$	$6.83 \cdot 10^{-1}$	$6.05 \cdot 10^{-1}$	$6.43 \cdot 10^{-1}$	$8.89 \cdot 10^{-1}$	$5.40 \cdot 10^{-1}$	$7.76 \cdot 10^{-1}$
Spin	μ	$9.56 \cdot 10^{-1}$	$9.34 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$	$9.36 \cdot 10^{-1}$	$9.14 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.21 \cdot 10^{-1}$	$9.23 \cdot 10^{-1}$	$9.26 \cdot 10^{-1}$	$9.29 \cdot 10^{-1}$	$8.68 \cdot 10^{-1}$	$9.72 \cdot 10^{-1}$
	σ	$2.91 \cdot 10^{-1}$	$8.19 \cdot 10^{-1}$	$7.45 \cdot 10^{-1}$	$8.11 \cdot 10^{-1}$	$5.24 \cdot 10^{-1}$	$6.90 \cdot 10^{-1}$	$4.37 \cdot 10^{-1}$	$6.98 \cdot 10^{-1}$	$5.48 \cdot 10^{-1}$	$9.02 \cdot 10^{-1}$	$5.30 \cdot 10^{-1}$	$7.87 \cdot 10^{-1}$
Contacts peptide-HGs	μ	$7.89 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.38 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$7.05 \cdot 10^{-1}$	$7.48 \cdot 10^{-1}$	$9.27 \cdot 10^{-1}$	$7.70 \cdot 10^{-1}$	$8.25 \cdot 10^{-1}$	$9.02 \cdot 10^{-1}$	$9.39 \cdot 10^{-1}$	$9.09 \cdot 10^{-1}$
	σ	$9.61 \cdot 10^{-1}$	$9.38 \cdot 10^{-1}$	$9.34 \cdot 10^{-1}$	$9.47 \cdot 10^{-1}$	$9.50 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.41 \cdot 10^{-1}$	$9.58 \cdot 10^{-1}$	$8.69 \cdot 10^{-1}$	$9.13 \cdot 10^{-1}$	$8.96 \cdot 10^{-1}$	$8.38 \cdot 10^{-1}$
Contacts peptide-tails	μ	$8.70 \cdot 10^{-1}$	$9.79 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$	$7.06 \cdot 10^{-1}$	$7.91 \cdot 10^{-1}$	$8.24 \cdot 10^{-1}$	$8.33 \cdot 10^{-1}$	$9.08 \cdot 10^{-1}$	$9.21 \cdot 10^{-1}$	$9.35 \cdot 10^{-1}$	$9.46 \cdot 10^{-1}$
	σ	$9.42 \cdot 10^{-1}$	$9.53 \cdot 10^{-1}$	$9.30 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$8.97 \cdot 10^{-1}$	$9.13 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$	$9.68 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$9.43 \cdot 10^{-1}$	$9.59 \cdot 10^{-1}$	$8.69 \cdot 10^{-1}$
Contacts peptide-water	μ	$6.71 \cdot 10^{-1}$	$9.45 \cdot 10^{-1}$	$9.72 \cdot 10^{-1}$	$9.49 \cdot 10^{-1}$	$6.91 \cdot 10^{-1}$	$7.07 \cdot 10^{-1}$	$8.06 \cdot 10^{-1}$	$7.11 \cdot 10^{-1}$	$8.73 \cdot 10^{-1}$	$9.39 \cdot 10^{-1}$	$8.92 \cdot 10^{-1}$	$8.76 \cdot 10^{-1}$
	σ	$9.41 \cdot 10^{-1}$	$9.50 \cdot 10^{-1}$	$9.47 \cdot 10^{-1}$	$9.45 \cdot 10^{-1}$	$9.51 \cdot 10^{-1}$	$8.56 \cdot 10^{-1}$	$9.40 \cdot 10^{-1}$	$8.68 \cdot 10^{-1}$	$7.70 \cdot 10^{-1}$	$8.59 \cdot 10^{-1}$	$8.27 \cdot 10^{-1}$	$7.85 \cdot 10^{-1}$
EDM	μ	$3.80 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$	$9.51 \cdot 10^{-1}$	$9.61 \cdot 10^{-1}$	$3.31 \cdot 10^{-1}$	$6.59 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$4.33 \cdot 10^{-1}$	$9.73 \cdot 10^{-1}$	$9.60 \cdot 10^{-1}$
	σ	$2.09 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.60 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$5.83 \cdot 10^{-1}$	$3.17 \cdot 10^{-1}$	$9.88 \cdot 10^{-1}$	$9.88 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$3.88 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$
Long. EDM	μ	$6.44 \cdot 10^{-1}$	$9.55 \cdot 10^{-1}$	$9.46 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$3.54 \cdot 10^{-1}$	$9.20 \cdot 10^{-1}$	$9.58 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$	$9.41 \cdot 10^{-1}$	$7.30 \cdot 10^{-1}$	$9.55 \cdot 10^{-1}$	$9.31 \cdot 10^{-1}$
	σ	$2.30 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.42 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$	$6.78 \cdot 10^{-1}$	$3.29 \cdot 10^{-1}$	$9.85 \cdot 10^{-1}$	$9.75 \cdot 10^{-1}$	$9.69 \cdot 10^{-1}$	$3.74 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$
Transv. EDM	μ	$3.20 \cdot 10^{-1}$	$9.87 \cdot 10^{-1}$	$9.92 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$	$4.97 \cdot 10^{-1}$	$9.88 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.67 \cdot 10^{-1}$	$3.77 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$	$9.79 \cdot 10^{-1}$
	σ	$1.92 \cdot 10^{-1}$	$9.31 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$9.20 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$2.87 \cdot 10^{-1}$	$9.83 \cdot 10^{-1}$	$9.18 \cdot 10^{-1}$	$9.74 \cdot 10^{-1}$	$3.91 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$	$9.55 \cdot 10^{-1}$
HDM	μ	$9.53 \cdot 10^{-1}$	$9.58 \cdot 10^{-1}$	$9.52 \cdot 10^{-1}$	$9.60 \cdot 10^{-1}$	$9.56 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$	$9.58 \cdot 10^{-1}$	$9.53 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$	$9.51 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.45 \cdot 10^{-1}$
	σ	$9.64 \cdot 10^{-1}$	$9.62 \cdot 10^{-1}$	$9.68 \cdot 10^{-1}$	$9.25 \cdot 10^{-1}$	$8.88 \cdot 10^{-1}$	$9.59 \cdot 10^{-1}$	$9.73 \cdot 10^{-1}$	$9.43 \cdot 10^{-1}$	$9.19 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$9.31 \cdot 10^{-1}$
Long. HDM	μ	$9.89 \cdot 10^{-1}$	$9.49 \cdot 10^{-1}$	$9.89 \cdot 10^{-1}$	$9.52 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.59 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$9.37 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$9.18 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.19 \cdot 10^{-1}$
	σ	$9.64 \cdot 10^{-1}$	$9.40 \cdot 10^{-1}$	$9.69 \cdot 10^{-1}$	$9.21 \cdot 10^{-1}$	$4.69 \cdot 10^{-1}$	$9.52 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.40 \cdot 10^{-1}$	$9.65 \cdot 10^{-1}$	$9.57 \cdot 10^{-1}$	$9.72 \cdot 10^{-1}$	$9.36 \cdot 10^{-1}$
Transv. HDM	μ	$9.70 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$	$9.62 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$	$9.83 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$9.84 \cdot 10^{-1}$
	σ	$9.15 \cdot 10^{-1}$	$9.25 \cdot 10^{-1}$	$9.04 \cdot 10^{-1}$	$9.23 \cdot 10^{-1}$	$9.07 \cdot 10^{-1}$	$9.31 \cdot 10^{-1}$	$9.12 \cdot 10^{-1}$	$9.17 \cdot 10^{-1}$	$8.70 \cdot 10^{-1}$	$8.93 \cdot 10^{-1}$	$9.02 \cdot 10^{-1}$	$8.96 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$6.16 \cdot 10^{-1}$	$9.22 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$	$5.28 \cdot 10^{-1}$	$7.43 \cdot 10^{-1}$	$4.07 \cdot 10^{-1}$	$5.79 \cdot 10^{-1}$	$8.06 \cdot 10^{-1}$	$8.04 \cdot 10^{-1}$	$6.00 \cdot 10^{-1}$
	σ	$5.55 \cdot 10^{-1}$	$8.17 \cdot 10^{-1}$	$9.69 \cdot 10^{-1}$	$9.48 \cdot 10^{-1}$	$4.95 \cdot 10^{-1}$	$5.72 \cdot 10^{-1}$	$6.73 \cdot 10^{-1}$	$4.26 \cdot 10^{-1}$	$5.56 \cdot 10^{-1}$	$8.53 \cdot 10^{-1}$	$6.71 \cdot 10^{-1}$	$6.83 \cdot 10^{-1}$
Rot. diffusion coef.	μ	$6.63 \cdot 10^{-1}$	$9.12 \cdot 10^{-1}$	$8.75 \cdot 10^{-1}$	$9.05 \cdot 10^{-1}$	$4.82 \cdot 10^{-1}$	$4.35 \cdot 10^{-1}$	$3.30 \cdot 10^{-1}$	$2.59 \cdot 10^{-1}$	$6.02 \cdot 10^{-1}$	$7.14 \cdot 10^{-1}$	$7.80 \cdot 10^{-1}$	$6.15 \cdot 10^{-1}$
	σ	$4.82 \cdot 10^{-1}$	$8.97 \cdot 10^{-1}$	$8.70 \cdot 10^{-1}$	$8.01 \cdot 10^{-1}$	$4.40 \cdot 10^{-1}$	$4.25 \cdot 10^{-1}$	$1.89 \cdot 10^{-1}$	$2.37 \cdot 10^{-1}$	$4.46 \cdot 10^{-1}$	$6.84 \cdot 10^{-1}$	$4.44 \cdot 10^{-1}$	$5.35 \cdot 10^{-1}$

Table D.5: Value of the statistical W from the Shapiro-Wilk tests applied to the samples of MD descriptors. The samples are grouped by force field—M2, M2P or M3—, lipid composition—POPC or POPE:POPG—, and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC	AMP	NC
Bilayer thickness	μ	$5.67 \cdot 10^{-1}$	$9.22 \cdot 10^{-2}$	$1.97 \cdot 10^{-1}$	$5.36 \cdot 10^{-1}$	$5.85 \cdot 10^{-1}$	$1.39 \cdot 10^{-1}$	$3.63 \cdot 10^{-1}$	$3.52 \cdot 10^{-1}$	$1.38 \cdot 10^{-1}$	$6.23 \cdot 10^{-2}$	$3.34 \cdot 10^{-1}$	$3.43 \cdot 10^{-1}$
	σ	$9.19 \cdot 10^{-1}$	$3.96 \cdot 10^{-1}$	$7.10 \cdot 10^{-2}$	$2.48 \cdot 10^{-1}$	$8.25 \cdot 10^{-1}$	$3.07 \cdot 10^{-1}$	$6.95 \cdot 10^{-1}$	$1.51 \cdot 10^{-1}$	$2.05 \cdot 10^{-1}$	$6.75 \cdot 10^{-2}$	$5.81 \cdot 10^{-1}$	$9.13 \cdot 10^{-1}$
Area per lipid	μ	$7.64 \cdot 10^{-1}$	$4.41 \cdot 10^{-1}$	$5.48 \cdot 10^{-1}$	$1.12 \cdot 10^{-1}$	$\mathbf{1.06 \cdot 10^{-2}}$	$6.30 \cdot 10^{-2}$	$4.86 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$	$\mathbf{1.89 \cdot 10^{-2}}$	$\mathbf{2.84 \cdot 10^{-2}}$	$3.82 \cdot 10^{-1}$	$3.27 \cdot 10^{-1}$
	σ	$3.86 \cdot 10^{-1}$	$5.87 \cdot 10^{-1}$	$2.18 \cdot 10^{-1}$	$5.00 \cdot 10^{-1}$	$8.42 \cdot 10^{-1}$	$9.80 \cdot 10^{-2}$	$7.74 \cdot 10^{-1}$	$7.13 \cdot 10^{-1}$	$7.04 \cdot 10^{-1}$	$2.80 \cdot 10^{-1}$	$5.85 \cdot 10^{-1}$	$6.19 \cdot 10^{-2}$
Peptide penetration	μ	$\mathbf{2.41 \cdot 10^{-16}}$	$\mathbf{2.80 \cdot 10^{-3}}$	$4.38 \cdot 10^{-1}$	$1.52 \cdot 10^{-1}$	$\mathbf{1.63 \cdot 10^{-14}}$	$\mathbf{2.05 \cdot 10^{-11}}$	$\mathbf{1.14 \cdot 10^{-12}}$	$\mathbf{9.21 \cdot 10^{-12}}$	$\mathbf{4.37 \cdot 10^{-13}}$	$\mathbf{1.09 \cdot 10^{-5}}$	$\mathbf{6.11 \cdot 10^{-12}}$	$\mathbf{1.36 \cdot 10^{-7}}$
	σ	$\mathbf{8.21 \cdot 10^{-18}}$	$3.91 \cdot 10^{-1}$	$2.94 \cdot 10^{-1}$	$1.52 \cdot 10^{-1}$	$\mathbf{8.26 \cdot 10^{-15}}$	$\mathbf{2.28 \cdot 10^{-11}}$	$\mathbf{8.58 \cdot 10^{-16}}$	$\mathbf{4.92 \cdot 10^{-13}}$	$\mathbf{1.10 \cdot 10^{-14}}$	$\mathbf{2.23 \cdot 10^{-6}}$	$\mathbf{1.91 \cdot 10^{-17}}$	$\mathbf{2.40 \cdot 10^{-9}}$
Tilt	μ	$2.12 \cdot 10^{-1}$	$\mathbf{1.00 \cdot 10^{-4}}$	$9.32 \cdot 10^{-2}$	$5.26 \cdot 10^{-2}$	$2.88 \cdot 10^{-1}$	$\mathbf{4.31 \cdot 10^{-4}}$	$\mathbf{2.12 \cdot 10^{-11}}$	$\mathbf{4.61 \cdot 10^{-5}}$	$\mathbf{2.37 \cdot 10^{-5}}$	$\mathbf{3.36 \cdot 10^{-3}}$	$2.04 \cdot 10^{-1}$	$\mathbf{1.84 \cdot 10^{-2}}$
	σ	$\mathbf{2.35 \cdot 10^{-15}}$	$\mathbf{4.69 \cdot 10^{-2}}$	$\mathbf{1.55 \cdot 10^{-2}}$	$1.20 \cdot 10^{-1}$	$\mathbf{9.53 \cdot 10^{-14}}$	$\mathbf{4.03 \cdot 10^{-10}}$	$\mathbf{2.89 \cdot 10^{-11}}$	$\mathbf{2.29 \cdot 10^{-10}}$	$\mathbf{4.83 \cdot 10^{-12}}$	$\mathbf{2.19 \cdot 10^{-4}}$	$\mathbf{8.89 \cdot 10^{-14}}$	$\mathbf{2.54 \cdot 10^{-7}}$
Spin	μ	$\mathbf{1.17 \cdot 10^{-2}}$	$\mathbf{8.07 \cdot 10^{-3}}$	$\mathbf{1.20 \cdot 10^{-2}}$	$\mathbf{9.18 \cdot 10^{-3}}$	$\mathbf{1.07 \cdot 10^{-4}}$	$6.20 \cdot 10^{-1}$	$\mathbf{2.09 \cdot 10^{-4}}$	$\mathbf{2.97 \cdot 10^{-3}}$	$\mathbf{3.47 \cdot 10^{-4}}$	$\mathbf{5.14 \cdot 10^{-3}}$	$\mathbf{1.69 \cdot 10^{-6}}$	$2.75 \cdot 10^{-1}$
	σ	$\mathbf{4.93 \cdot 10^{-17}}$	$\mathbf{2.41 \cdot 10^{-6}}$	$\mathbf{6.22 \cdot 10^{-10}}$	$\mathbf{1.61 \cdot 10^{-6}}$	$\mathbf{5.10 \cdot 10^{-14}}$	$\mathbf{5.40 \cdot 10^{-9}}$	$\mathbf{2.99 \cdot 10^{-15}}$	$\mathbf{7.73 \cdot 10^{-9}}$	$\mathbf{1.22 \cdot 10^{-13}}$	$\mathbf{5.47 \cdot 10^{-4}}$	$\mathbf{6.41 \cdot 10^{-14}}$	$\mathbf{4.46 \cdot 10^{-7}}$
Contacts peptide-HGs	μ	$\mathbf{7.60 \cdot 10^{-9}}$	$6.55 \cdot 10^{-1}$	$\mathbf{1.32 \cdot 10^{-3}}$	$1.80 \cdot 10^{-1}$	$\mathbf{8.11 \cdot 10^{-11}}$	$\mathbf{6.75 \cdot 10^{-8}}$	$\mathbf{4.16 \cdot 10^{-4}}$	$\mathbf{1.93 \cdot 10^{-7}}$	$\mathbf{7.58 \cdot 10^{-8}}$	$\mathbf{5.54 \cdot 10^{-4}}$	$\mathbf{1.62 \cdot 10^{-3}}$	$\mathbf{9.40 \cdot 10^{-4}}$
	σ	$\mathbf{2.50 \cdot 10^{-2}}$	$\mathbf{1.14 \cdot 10^{-2}}$	$\mathbf{8.99 \cdot 10^{-4}}$	$\mathbf{2.51 \cdot 10^{-2}}$	$\mathbf{5.88 \cdot 10^{-3}}$	$4.66 \cdot 10^{-1}$	$\mathbf{2.04 \cdot 10^{-3}}$	$7.22 \cdot 10^{-2}$	$\mathbf{1.83 \cdot 10^{-6}}$	$\mathbf{1.33 \cdot 10^{-3}}$	$\mathbf{1.89 \cdot 10^{-5}}$	$\mathbf{7.48 \cdot 10^{-6}}$
Contacts peptide-tails	μ	$\mathbf{2.04 \cdot 10^{-6}}$	$4.96 \cdot 10^{-1}$	$9.52 \cdot 10^{-2}$	$3.87 \cdot 10^{-1}$	$\mathbf{8.59 \cdot 10^{-11}}$	$\mathbf{5.47 \cdot 10^{-7}}$	$\mathbf{6.72 \cdot 10^{-8}}$	$\mathbf{5.38 \cdot 10^{-6}}$	$\mathbf{5.92 \cdot 10^{-5}}$	$\mathbf{2.67 \cdot 10^{-3}}$	$\mathbf{1.01 \cdot 10^{-3}}$	$\mathbf{2.38 \cdot 10^{-2}}$
	σ	$\mathbf{2.18 \cdot 10^{-3}}$	$\mathbf{4.64 \cdot 10^{-2}}$	$\mathbf{5.75 \cdot 10^{-4}}$	$5.11 \cdot 10^{-2}$	$\mathbf{1.97 \cdot 10^{-5}}$	$\mathbf{1.37 \cdot 10^{-3}}$	$\mathbf{1.20 \cdot 10^{-2}}$	$1.97 \cdot 10^{-1}$	$\mathbf{1.02 \cdot 10^{-2}}$	$\mathbf{1.76 \cdot 10^{-2}}$	$\mathbf{1.71 \cdot 10^{-2}}$	$\mathbf{5.20 \cdot 10^{-5}}$
Contacts peptide-water	μ	$\mathbf{1.69 \cdot 10^{-11}}$	$\mathbf{2.05 \cdot 10^{-2}}$	$1.10 \cdot 10^{-1}$	$\mathbf{3.13 \cdot 10^{-2}}$	$\mathbf{4.06 \cdot 10^{-11}}$	$\mathbf{1.11 \cdot 10^{-8}}$	$\mathbf{2.08 \cdot 10^{-8}}$	$\mathbf{1.30 \cdot 10^{-8}}$	$\mathbf{2.51 \cdot 10^{-6}}$	$\mathbf{1.20 \cdot 10^{-2}}$	$\mathbf{1.36 \cdot 10^{-5}}$	$\mathbf{8.71 \cdot 10^{-5}}$
	σ	$\mathbf{1.88 \cdot 10^{-3}}$	$\mathbf{3.48 \cdot 10^{-2}}$	$\mathbf{3.87 \cdot 10^{-3}}$	$\mathbf{2.13 \cdot 10^{-2}}$	$\mathbf{6.16 \cdot 10^{-3}}$	$\mathbf{2.25 \cdot 10^{-5}}$	$\mathbf{1.75 \cdot 10^{-3}}$	$\mathbf{4.89 \cdot 10^{-5}}$	$\mathbf{2.43 \cdot 10^{-9}}$	$\mathbf{2.66 \cdot 10^{-5}}$	$\mathbf{8.34 \cdot 10^{-8}}$	$\mathbf{4.00 \cdot 10^{-7}}$
EDM	μ	$\mathbf{5.64 \cdot 10^{-16}}$	$3.24 \cdot 10^{-1}$	$\mathbf{6.23 \cdot 10^{-3}}$	$9.80 \cdot 10^{-2}$	$\mathbf{1.42 \cdot 10^{-16}}$	$\mathbf{1.65 \cdot 10^{-9}}$	$5.33 \cdot 10^{-2}$	$6.59 \cdot 10^{-1}$	$5.24 \cdot 10^{-2}$	$\mathbf{1.26 \cdot 10^{-12}}$	$1.18 \cdot 10^{-1}$	$8.97 \cdot 10^{-2}$
	σ	$\mathbf{6.52 \cdot 10^{-18}}$	$5.86 \cdot 10^{-1}$	$\mathbf{2.18 \cdot 10^{-2}}$	$5.34 \cdot 10^{-1}$	$\mathbf{4.34 \cdot 10^{-13}}$	$\mathbf{6.90 \cdot 10^{-14}}$	$7.32 \cdot 10^{-1}$	$8.75 \cdot 10^{-1}$	$5.50 \cdot 10^{-2}$	$\mathbf{3.84 \cdot 10^{-13}}$	$\mathbf{3.89 \cdot 10^{-2}}$	$6.30 \cdot 10^{-1}$
Long. EDM	μ	$\mathbf{5.03 \cdot 10^{-12}}$	$5.64 \cdot 10^{-2}$	$\mathbf{3.42 \cdot 10^{-3}}$	$5.04 \cdot 10^{-2}$	$\mathbf{2.70 \cdot 10^{-16}}$	$\mathbf{2.41 \cdot 10^{-3}}$	$\mathbf{1.56 \cdot 10^{-2}}$	$3.93 \cdot 10^{-1}$	$\mathbf{1.86 \cdot 10^{-3}}$	$\mathbf{2.95 \cdot 10^{-8}}$	$\mathbf{1.03 \cdot 10^{-2}}$	$\mathbf{6.02 \cdot 10^{-3}}$
	σ	$\mathbf{1.07 \cdot 10^{-17}}$	$4.54 \cdot 10^{-1}$	$\mathbf{2.13 \cdot 10^{-3}}$	$7.22 \cdot 10^{-1}$	$\mathbf{2.25 \cdot 10^{-11}}$	$\mathbf{9.11 \cdot 10^{-14}}$	$5.32 \cdot 10^{-1}$	$3.70 \cdot 10^{-1}$	$6.62 \cdot 10^{-2}$	$\mathbf{2.76 \cdot 10^{-13}}$	$1.97 \cdot 10^{-1}$	$7.17 \cdot 10^{-1}$
Transv. EDM	μ	$\mathbf{1.08 \cdot 10^{-16}}$	$8.64 \cdot 10^{-1}$	$9.45 \cdot 10^{-1}$	$8.33 \cdot 10^{-1}$	$1.35 \cdot 10^{-1}$	$\mathbf{7.56 \cdot 10^{-12}}$	$7.46 \cdot 10^{-1}$	$6.07 \cdot 10^{-1}$	$5.43 \cdot 10^{-2}$	$\mathbf{2.91 \cdot 10^{-13}}$	$1.85 \cdot 10^{-1}$	$5.09 \cdot 10^{-1}$
	σ	$\mathbf{4.32 \cdot 10^{-18}}$	$\mathbf{5.99 \cdot 10^{-3}}$	$2.96 \cdot 10^{-1}$	$\mathbf{2.29 \cdot 10^{-3}}$	$2.81 \cdot 10^{-1}$	$\mathbf{3.41 \cdot 10^{-14}}$	$4.49 \cdot 10^{-1}$	$\mathbf{2.05 \cdot 10^{-3}}$	$1.41 \cdot 10^{-1}$	$\mathbf{4.21 \cdot 10^{-13}}$	$2.07 \cdot 10^{-1}$	$5.49 \cdot 10^{-2}$
HDM	μ	$\mathbf{8.21 \cdot 10^{-3}}$	$7.36 \cdot 10^{-2}$	$\mathbf{7.28 \cdot 10^{-3}}$	$8.92 \cdot 10^{-2}$	$\mathbf{1.19 \cdot 10^{-2}}$	$1.65 \cdot 10^{-1}$	$\mathbf{1.55 \cdot 10^{-2}}$	$\mathbf{4.45 \cdot 10^{-2}}$	$4.79 \cdot 10^{-1}$	$\mathbf{3.65 \cdot 10^{-2}}$	$2.18 \cdot 10^{-1}$	$\mathbf{2.12 \cdot 10^{-2}}$
	σ	$\mathbf{3.41 \cdot 10^{-2}}$	$1.03 \cdot 10^{-1}$	$5.78 \cdot 10^{-2}$	$\mathbf{3.61 \cdot 10^{-3}}$	$\mathbf{8.99 \cdot 10^{-6}}$	$7.79 \cdot 10^{-2}$	$1.22 \cdot 10^{-1}$	$\mathbf{1.72 \cdot 10^{-2}}$	$\mathbf{1.71 \cdot 10^{-4}}$	$1.49 \cdot 10^{-1}$	$\mathbf{9.21 \cdot 10^{-3}}$	$\mathbf{5.92 \cdot 10^{-3}}$
Long. HDM	μ	$7.92 \cdot 10^{-1}$	$\mathbf{3.02 \cdot 10^{-2}}$	$7.72 \cdot 10^{-1}$	$\mathbf{3.97 \cdot 10^{-2}}$	$3.44 \cdot 10^{-1}$	$8.37 \cdot 10^{-2}$	$6.20 \cdot 10^{-1}$	$\mathbf{1.06 \cdot 10^{-2}}$	$8.80 \cdot 10^{-2}$	$\mathbf{2.05 \cdot 10^{-3}}$	$\mathbf{4.30 \cdot 10^{-2}}$	$\mathbf{2.23 \cdot 10^{-3}}$
	σ	$\mathbf{3.44 \cdot 10^{-2}}$	$\mathbf{1.33 \cdot 10^{-2}}$	$6.56 \cdot 10^{-2}$	$\mathbf{2.55 \cdot 10^{-3}}$	$\mathbf{8.19 \cdot 10^{-15}}$	$\mathbf{4.06 \cdot 10^{-2}}$	$\mathbf{4.22 \cdot 10^{-2}}$	$\mathbf{1.35 \cdot 10^{-2}}$	$\mathbf{4.09 \cdot 10^{-2}}$	$6.84 \cdot 10^{-2}$	$1.06 \cdot 10^{-1}$	$\mathbf{9.26 \cdot 10^{-3}}$
Transv. HDM	μ	$8.42 \cdot 10^{-2}$	$5.98 \cdot 10^{-1}$	$\mathbf{2.89 \cdot 10^{-2}}$	$5.38 \cdot 10^{-1}$	$\mathbf{4.85 \cdot 10^{-2}}$	$6.69 \cdot 10^{-1}$	$\mathbf{4.40 \cdot 10^{-2}}$	$6.22 \cdot 10^{-1}$	$2.37 \cdot 10^{-1}$	$7.27 \cdot 10^{-1}$	$2.27 \cdot 10^{-1}$	$7.32 \cdot 10^{-1}$
	σ	$\mathbf{1.16 \cdot 10^{-4}}$	$\mathbf{3.60 \cdot 10^{-3}}$	$\mathbf{3.89 \cdot 10^{-5}}$	$\mathbf{3.05 \cdot 10^{-3}}$	$\mathbf{5.29 \cdot 10^{-5}}$	$\mathbf{6.22 \cdot 10^{-3}}$	$\mathbf{8.41 \cdot 10^{-5}}$	$\mathbf{1.92 \cdot 10^{-3}}$	$\mathbf{2.07 \cdot 10^{-6}}$	$\mathbf{2.74 \cdot 10^{-4}}$	$\mathbf{3.13 \cdot 10^{-5}}$	$\mathbf{3.66 \cdot 10^{-4}}$
Tr. diffusion coef.	μ	$\mathbf{1.61 \cdot 10^{-12}}$	$\mathbf{2.84 \cdot 10^{-3}}$	$8.50 \cdot 10^{-2}$	$2.59 \cdot 10^{-1}$	$\mathbf{2.33 \cdot 10^{-14}}$	$\mathbf{1.89 \cdot 10^{-11}}$	$\mathbf{5.52 \cdot 10^{-10}}$	$\mathbf{6.41 \cdot 10^{-13}}$	$\mathbf{3.79 \cdot 10^{-13}}$	$\mathbf{1.20 \cdot 10^{-6}}$	$\mathbf{1.88 \cdot 10^{-8}}$	$\mathbf{1.97 \cdot 10^{-10}}$
	σ	$\mathbf{1.57 \cdot 10^{-13}}$	$\mathbf{2.18 \cdot 10^{-6}}$	$6.83 \cdot 10^{-2}$	$\mathbf{2.84 \cdot 10^{-2}}$	$\mathbf{1.90 \cdot 10^{-14}}$	$\mathbf{7.71 \cdot 10^{-11}}$	$\mathbf{1.81 \cdot 10^{-11}}$	$\mathbf{1.04 \cdot 10^{-12}}$	$\mathbf{1.60 \cdot 10^{-13}}$	$\mathbf{1.88 \cdot 10^{-5}}$	$\mathbf{1.65 \cdot 10^{-11}}$	$\mathbf{4.22 \cdot 10^{-9}}$
Rot. diffusion coef.	μ	$\mathbf{1.16 \cdot 10^{-11}}$	$\mathbf{1.23 \cdot 10^{-3}}$	$\mathbf{2.98 \cdot 10^{-6}}$	$\mathbf{7.14 \cdot 10^{-4}}$	$\mathbf{1.26 \cdot 10^{-14}}$	$\mathbf{1.35 \cdot 10^{-12}}$	$\mathbf{1.39 \cdot 10^{-16}}$	$\mathbf{1.82 \cdot 10^{-14}}$	$\mathbf{9.30 \cdot 10^{-13}}$	$\mathbf{1.50 \cdot 10^{-8}}$	$\mathbf{4.30 \cdot 10^{-9}}$	$\mathbf{3.32 \cdot 10^{-10}}$
	σ	$\mathbf{1.25 \cdot 10^{-14}}$	$\mathbf{3.84 \cdot 10^{-4}}$	$\mathbf{1.97 \cdot 10^{-6}}$	$\mathbf{9.25 \cdot 10^{-7}}$	$\mathbf{3.36 \cdot 10^{-15}}$	$\mathbf{1.03 \cdot 10^{-12}}$	$\mathbf{4.06 \cdot 10^{-18}}$	$\mathbf{1.13 \cdot 10^{-14}}$	$\mathbf{4.06 \cdot 10^{-15}}$	$\mathbf{4.31 \cdot 10^{-9}}$	$\mathbf{3.70 \cdot 10^{-15}}$	$\mathbf{2.34 \cdot 10^{-11}}$

Table D.6: p-value from the Shapiro-Wilk tests applied to the samples of MD descriptors. The samples are grouped by force field—M2, M2P or M3—, lipid composition—POPC or POPE:POPG—, and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a non-normal parent distribution.

Magnitude		Martini 2.2 POPC	Martini 2.2 POPE:POPG 3:1	Martini 2.2P POPC	Martini 2.2P POPE:POPG 3:1	Martini 3 POPC	Martini 3 POPE:POPG 3:1
Bilayer thickness	μ	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.865$ $p = 1.47 \cdot 10^{-1}$	$\sim^*$ $U = 1.51 \cdot 10^{+3}$ $S = 0.767$ $p = 1.04 \cdot 10^{-1}$	$\sim$ $U = 1.53 \cdot 10^{+3}$ $S = 0.828$ $p = 1.24 \cdot 10^{-1}$	$\sim^*$ $U = 1.49 \cdot 10^{+3}$ $S = 0.776$ $p = 8.41 \cdot 10^{-2}$	$<$ $U = 8.52 \cdot 10^{+2}$ $S = 0.601$ $p = \mathbf{2.76 \cdot 10^{-7}}$	$<$ $U = 9.83 \cdot 10^{+2}$ $S = 0.688$ $p = \mathbf{7.35 \cdot 10^{-6}}$
	σ	$\sim$ $U = 1.70 \cdot 10^{+3}$ $S = 0.933$ $p = 5.11 \cdot 10^{-1}$	$\sim$ $U = 1.79 \cdot 10^{+3}$ $S = 0.910$ $p = 8.71 \cdot 10^{-1}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.838$ $p = 1.49 \cdot 10^{-1}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.877$ $p = 1.44 \cdot 10^{-1}$	$\sim$ $U = 1.46 \cdot 10^{+3}$ $S = 0.797$ $p = 5.71 \cdot 10^{-2}$	$\sim$ $U = 1.51 \cdot 10^{+3}$ $S = 0.870$ $p = 1.05 \cdot 10^{-1}$
Area per lipid	μ	$>$ $U = 2.58 \cdot 10^{+3}$ $S = 0.726$ $p = \mathbf{5.23 \cdot 10^{-5}}$	$>$ $U = 2.65 \cdot 10^{+3}$ $S = 0.655$ $p = \mathbf{1.04 \cdot 10^{-5}}$	$>$ $U = 2.39 \cdot 10^{+3}$ $S = 0.712$ $p = \mathbf{1.83 \cdot 10^{-3}}$	$>$ $U = 2.58 \cdot 10^{+3}$ $S = 0.711$ $p = \mathbf{5.69 \cdot 10^{-5}}$	$>$ $U = 2.86 \cdot 10^{+3}$ $S = 0.583$ $p = \mathbf{4.72 \cdot 10^{-8}}$	$>$ $U = 2.74 \cdot 10^{+3}$ $S = 0.644$ $p = \mathbf{1.21 \cdot 10^{-6}}$
	σ	$\sim$ $U = 1.76 \cdot 10^{+3}$ $S = 0.928$ $p = 7.40 \cdot 10^{-1}$	$\sim$ $U = 1.73 \cdot 10^{+3}$ $S = 0.913$ $p = 6.38 \cdot 10^{-1}$	$\sim$ $U = 1.60 \cdot 10^{+3}$ $S = 0.847$ $p = 2.46 \cdot 10^{-1}$	$<$ $U = 1.34 \cdot 10^{+3}$ $S = 0.812$ $p = \mathbf{6.49 \cdot 10^{-3}}$	$\sim$ $U = 1.74 \cdot 10^{+3}$ $S = 0.895$ $p = 6.82 \cdot 10^{-1}$	$\sim$ $U = 1.90 \cdot 10^{+3}$ $S = 0.911$ $p = 6.90 \cdot 10^{-1}$
Peptide penetration	μ	$>$ $U = 2.51 \cdot 10^{+3}$ $S = 0.764$ $p = \mathbf{2.20 \cdot 10^{-4}}$	$>$ $U = 2.52 \cdot 10^{+3}$ $S = 0.711$ $p = \mathbf{1.58 \cdot 10^{-4}}$	$>$ $U = 2.46 \cdot 10^{+3}$ $S = 0.911$ $p = \mathbf{5.95 \cdot 10^{-4}}$	$>^*$ $U = 2.62 \cdot 10^{+3}$ $S = 0.544$ $p = \mathbf{2.15 \cdot 10^{-5}}$	$>^*$ $U = 2.73 \cdot 10^{+3}$ $S = 0.638$ $p = \mathbf{1.45 \cdot 10^{-6}}$	$>^*$ $U = 2.75 \cdot 10^{+3}$ $S = 0.569$ $p = \mathbf{8.95 \cdot 10^{-7}}$
	σ	$\sim$ $U = 1.58 \cdot 10^{+3}$ $S = 0.613$ $p = 2.12 \cdot 10^{-1}$	$<$ $U = 1.29 \cdot 10^{+3}$ $S = 0.822$ $p = \mathbf{3.01 \cdot 10^{-3}}$	$<$ $U = 1.22 \cdot 10^{+3}$ $S = 0.840$ $p = \mathbf{8.95 \cdot 10^{-4}}$	$<^*$ $U = 1.31 \cdot 10^{+3}$ $S = 0.413$ $p = \mathbf{4.10 \cdot 10^{-3}}$	$<^*$ $U = 8.51 \cdot 10^{+2}$ $S = 0.486$ $p = \mathbf{2.68 \cdot 10^{-7}}$	$<^*$ $U = 8.16 \cdot 10^{+2}$ $S = 0.461$ $p = \mathbf{1.04 \cdot 10^{-7}}$
Tilt	μ	$\sim^*$ $U = 1.66 \cdot 10^{+3}$ $S = 0.690$ $p = 4.06 \cdot 10^{-1}$	$\sim^*$ $U = 1.75 \cdot 10^{+3}$ $S = 0.672$ $p = 7.13 \cdot 10^{-1}$	$\sim^*$ $U = 1.95 \cdot 10^{+3}$ $S = 0.707$ $p = 5.11 \cdot 10^{-1}$	$\sim^*$ $U = 2.08 \cdot 10^{+3}$ $S = 0.681$ $p = 1.95 \cdot 10^{-1}$	$\sim$ $U = 1.68 \cdot 10^{+3}$ $S = 0.840$ $p = 4.73 \cdot 10^{-1}$	$\sim^*$ $U = 2.13 \cdot 10^{+3}$ $S = 0.672$ $p = 1.16 \cdot 10^{-1}$
	σ	$<$ $U = 1.27 \cdot 10^{+3}$ $S = 0.803$ $p = \mathbf{2.02 \cdot 10^{-3}}$	$<^*$ $U = 1.22 \cdot 10^{+3}$ $S = 0.749$ $p = \mathbf{9.60 \cdot 10^{-4}}$	$<$ $U = 1.28 \cdot 10^{+3}$ $S = 0.877$ $p = \mathbf{2.74 \cdot 10^{-3}}$	$<^*$ $U = 1.13 \cdot 10^{+3}$ $S = 0.558$ $p = \mathbf{1.89 \cdot 10^{-4}}$	$<^*$ $U = 9.23 \cdot 10^{+2}$ $S = 0.586$ $p = \mathbf{1.73 \cdot 10^{-6}}$	$<^*$ $U = 9.02 \cdot 10^{+2}$ $S = 0.524$ $p = \mathbf{1.02 \cdot 10^{-6}}$
Spin	μ	$\sim$ $U = 2.10 \cdot 10^{+3}$ $S = 0.827$ $p = 1.52 \cdot 10^{-1}$	$\sim^*$ $U = 2.04 \cdot 10^{+3}$ $S = 0.776$ $p = 2.72 \cdot 10^{-1}$	$>$ $U = 2.37 \cdot 10^{+3}$ $S = 0.763$ $p = \mathbf{2.45 \cdot 10^{-3}}$	$\sim^*$ $U = 1.98 \cdot 10^{+3}$ $S = 0.643$ $p = 4.17 \cdot 10^{-1}$	$\sim^*$ $U = 1.97 \cdot 10^{+3}$ $S = 0.657$ $p = 4.48 \cdot 10^{-1}$	$\sim^*$ $U = 2.10 \cdot 10^{+3}$ $S = 0.638$ $p = 1.50 \cdot 10^{-1}$
	σ	$<$ $U = 1.02 \cdot 10^{+3}$ $S = 0.771$ $p = \mathbf{1.57 \cdot 10^{-5}}$	$<^*$ $U = 1.01 \cdot 10^{+3}$ $S = 0.672$ $p = \mathbf{1.43 \cdot 10^{-5}}$	$<$ $U = 1.14 \cdot 10^{+3}$ $S = 0.837$ $p = \mathbf{2.25 \cdot 10^{-4}}$	$<^*$ $U = 8.34 \cdot 10^{+2}$ $S = 0.529$ $p = \mathbf{1.70 \cdot 10^{-7}}$	$<^*$ $U = 7.45 \cdot 10^{+2}$ $S = 0.564$ $p = \mathbf{1.36 \cdot 10^{-8}}$	$<^*$ $U = 7.63 \cdot 10^{+2}$ $S = 0.519$ $p = \mathbf{2.30 \cdot 10^{-8}}$
Contacts peptide-HGs	μ	$\sim$ $U = 1.61 \cdot 10^{+3}$ $S = 0.847$ $p = 2.61 \cdot 10^{-1}$	$\sim$ $U = 1.67 \cdot 10^{+3}$ $S = 0.844$ $p = 4.17 \cdot 10^{-1}$	$\sim$ $U = 1.74 \cdot 10^{+3}$ $S = 0.903$ $p = 6.75 \cdot 10^{-1}$	$\sim^*$ $U = 1.80 \cdot 10^{+3}$ $S = 0.757$ $p = 8.96 \cdot 10^{-1}$	$>^*$ $U = 2.47 \cdot 10^{+3}$ $S = 0.703$ $p = \mathbf{4.78 \cdot 10^{-4}}$	$>^*$ $U = 2.30 \cdot 10^{+3}$ $S = 0.741$ $p = \mathbf{6.97 \cdot 10^{-3}}$
	σ	$<$ $U = 1.39 \cdot 10^{+3}$ $S = 0.824$ $p = \mathbf{1.26 \cdot 10^{-2}}$	$\sim$ $U = 1.45 \cdot 10^{+3}$ $S = 0.824$ $p = 5.45 \cdot 10^{-2}$	$\sim$ $U = 1.49 \cdot 10^{+3}$ $S = 0.859$ $p = 8.78 \cdot 10^{-2}$	$<$ $U = 1.38 \cdot 10^{+3}$ $S = 0.829$ $p = \mathbf{1.03 \cdot 10^{-2}}$	$<^*$ $U = 1.06 \cdot 10^{+3}$ $S = 0.576$ $p = \mathbf{4.13 \cdot 10^{-5}}$	$<^*$ $U = 1.11 \cdot 10^{+3}$ $S = 0.648$ $p = \mathbf{1.19 \cdot 10^{-4}}$
Contacts peptide-tails	μ	$>$ $U = 2.25 \cdot 10^{+3}$ $S = 0.827$ $p = \mathbf{1.39 \cdot 10^{-2}}$	$>$ $U = 2.28 \cdot 10^{+3}$ $S = 0.825$ $p = \mathbf{9.12 \cdot 10^{-3}}$	$>$ $U = 2.28 \cdot 10^{+3}$ $S = 0.856$ $p = \mathbf{9.69 \cdot 10^{-3}}$	$>^*$ $U = 2.36 \cdot 10^{+3}$ $S = 0.756$ $p = \mathbf{2.91 \cdot 10^{-3}}$	$>^*$ $U = 2.71 \cdot 10^{+3}$ $S = 0.639$ $p = \mathbf{2.56 \cdot 10^{-6}}$	$>^*$ $U = 2.65 \cdot 10^{+3}$ $S = 0.650$ $p = \mathbf{1.04 \cdot 10^{-5}}$
	σ	$\sim$ $U = 1.47 \cdot 10^{+3}$ $S = 0.829$ $p = 7.11 \cdot 10^{-2}$	$\sim$ $U = 1.46 \cdot 10^{+3}$ $S = 0.833$ $p = 6.34 \cdot 10^{-2}$	$<$ $U = 1.39 \cdot 10^{+3}$ $S = 0.813$ $p = \mathbf{1.24 \cdot 10^{-2}}$	$<^*$ $U = 1.44 \cdot 10^{+3}$ $S = 0.762$ $p = \mathbf{2.44 \cdot 10^{-2}}$	$\sim^*$ $U = 1.58 \cdot 10^{+3}$ $S = 0.686$ $p = 2.16 \cdot 10^{-1}$	$<^*$ $U = 1.39 \cdot 10^{+3}$ $S = 0.712$ $p = \mathbf{1.23 \cdot 10^{-2}}$
Contacts peptide-water	μ	$<$ $U = 1.22 \cdot 10^{+3}$ $S = 0.766$ $p = \mathbf{9.94 \cdot 10^{-4}}$	$<$ $U = 1.21 \cdot 10^{+3}$ $S = 0.769$ $p = \mathbf{8.35 \cdot 10^{-4}}$	$<$ $U = 1.30 \cdot 10^{+3}$ $S = 0.848$ $p = \mathbf{3.46 \cdot 10^{-3}}$	$<$ $U = 1.20 \cdot 10^{+3}$ $S = 0.765$ $p = \mathbf{5.95 \cdot 10^{-4}}$	$<$ $U = 1.02 \cdot 10^{+3}$ $S = 0.692$ $p = \mathbf{1.57 \cdot 10^{-5}}$	$<^*$ $U = 9.14 \cdot 10^{+2}$ $S = 0.658$ $p = \mathbf{1.38 \cdot 10^{-6}}$
	σ	$<^*$ $U = 1.17 \cdot 10^{+3}$ $S = 0.715$ $p = \mathbf{3.76 \cdot 10^{-4}}$	$<^*$ $U = 1.19 \cdot 10^{+3}$ $S = 0.729$ $p = \mathbf{5.33 \cdot 10^{-4}}$	$<^*$ $U = 9.95 \cdot 10^{+2}$ $S = 0.666$ $p = \mathbf{9.72 \cdot 10^{-6}}$	$<^*$ $U = 1.16 \cdot 10^{+3}$ $S = 0.720$ $p = \mathbf{3.05 \cdot 10^{-4}}$	$<^*$ $U = 9.68 \cdot 10^{+2}$ $S = 0.617$ $p = \mathbf{5.16 \cdot 10^{-6}}$	$<^*$ $U = 9.54 \cdot 10^{+2}$ $S = 0.571$ $p = \mathbf{3.69 \cdot 10^{-6}}$
EDM	μ	$\sim$ $U = 1.99 \cdot 10^{+3}$ $S = 0.798$ $p = 3.88 \cdot 10^{-1}$	$\sim$ $U = 1.92 \cdot 10^{+3}$ $S = 0.928$ $p = 6.45 \cdot 10^{-1}$	$\sim$ $U = 1.97 \cdot 10^{+3}$ $S = 0.886$ $p = 4.69 \cdot 10^{-1}$	$\sim$ $U = 1.99 \cdot 10^{+3}$ $S = 0.879$ $p = 4.09 \cdot 10^{-1}$	$\sim$ $U = 1.98 \cdot 10^{+3}$ $S = 0.795$ $p = 4.35 \cdot 10^{-1}$	$\sim$ $U = 2.01 \cdot 10^{+3}$ $S = 0.895$ $p = 3.32 \cdot 10^{-1}$
	σ	$<$ $U = 1.42 \cdot 10^{+3}$ $S = 0.563$ $p = \mathbf{1.98 \cdot 10^{-2}}$	$\sim$ $U = 1.46 \cdot 10^{+3}$ $S = 0.823$ $p = 6.27 \cdot 10^{-2}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.681$ $p = 1.43 \cdot 10^{-1}$	$<$ $U = 1.44 \cdot 10^{+3}$ $S = 0.852$ $p = \mathbf{2.25 \cdot 10^{-2}}$	$\sim$ $U = 1.50 \cdot 10^{+3}$ $S = 0.668$ $p = 9.99 \cdot 10^{-2}$	$\sim$ $U = 1.48 \cdot 10^{+3}$ $S = 0.842$ $p = 7.27 \cdot 10^{-2}$

Table D.7: Differences in the values of MD descriptors based on antimicrobial activity for each combination of lipid composition and force field. Results of the Mann-Whitney U test. Continues on the next page.

Magnitude		Martini 2.2 POPC	Martini 2.2 POPE:POPG 3:1	Martini 2.2P POPC	Martini 2.2P POPE:POPG 3:1	Martini 3 POPC	Martini 3 POPE:POPG 3:1
Long. EDM	μ	$\sim$ $U = 1.70 \cdot 10^{+3}$ $S = 0.887$ $p = 5.08 \cdot 10^{-1}$	$\sim$ $U = 1.62 \cdot 10^{+3}$ $S = 0.903$ $p = 2.85 \cdot 10^{-1}$	$\sim$ $U = 1.68 \cdot 10^{+3}$ $S = 0.847$ $p = 4.69 \cdot 10^{-1}$	$\sim$ $U = 1.72 \cdot 10^{+3}$ $S = 0.935$ $p = 5.73 \cdot 10^{-1}$	$\sim$ $U = 1.69 \cdot 10^{+3}$ $S = 0.915$ $p = 4.89 \cdot 10^{-1}$	$\sim$ $U = 1.74 \cdot 10^{+3}$ $S = 0.917$ $p = 6.56 \cdot 10^{-1}$
	σ	$\sim$ $U = 1.61 \cdot 10^{+3}$ $S = 0.599$ $p = 2.61 \cdot 10^{-1}$	$\sim$ $U = 1.62 \cdot 10^{+3}$ $S = 0.884$ $p = 2.83 \cdot 10^{-1}$	$\sim$ $U = 1.66 \cdot 10^{+3}$ $S = 0.693$ $p = 4.00 \cdot 10^{-1}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.900$ $p = 1.49 \cdot 10^{-1}$	$\sim$ $U = 1.68 \cdot 10^{+3}$ $S = 0.722$ $p = 4.57 \cdot 10^{-1}$	$\sim$ $U = 1.65 \cdot 10^{+3}$ $S = 0.898$ $p = 3.74 \cdot 10^{-1}$
Transv. EDM	μ	$>$ $U = 2.40 \cdot 10^{+3}$ $S = 0.744$ $p = \mathbf{1.52 \cdot 10^{-3}}$	$>$ $U = 2.38 \cdot 10^{+3}$ $S = 0.790$ $p = \mathbf{2.12 \cdot 10^{-3}}$	$>$ $U = 2.50 \cdot 10^{+3}$ $S = 0.682$ $p = \mathbf{2.78 \cdot 10^{-4}}$	$>$ $U = 2.49 \cdot 10^{+3}$ $S = 0.757$ $p = \mathbf{3.11 \cdot 10^{-4}}$	$>$ $U = 2.42 \cdot 10^{+3}$ $S = 0.646$ $p = \mathbf{1.10 \cdot 10^{-3}}$	$>$ $U = 2.44 \cdot 10^{+3}$ $S = 0.789$ $p = \mathbf{8.49 \cdot 10^{-4}}$
	σ	$\sim$ $U = 1.48 \cdot 10^{+3}$ $S = 0.571$ $p = 7.27 \cdot 10^{-2}$	$\sim$ $U = 1.48 \cdot 10^{+3}$ $S = 0.869$ $p = 7.52 \cdot 10^{-2}$	$\sim$ $U = 1.59 \cdot 10^{+3}$ $S = 0.569$ $p = 2.31 \cdot 10^{-1}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.885$ $p = 1.46 \cdot 10^{-1}$	$\sim$ $U = 1.60 \cdot 10^{+3}$ $S = 0.630$ $p = 2.44 \cdot 10^{-1}$	$\sim$ $U = 1.54 \cdot 10^{+3}$ $S = 0.907$ $p = 1.50 \cdot 10^{-1}$
HDM	μ	$\sim$ $U = 1.58 \cdot 10^{+3}$ $S = 0.861$ $p = 2.17 \cdot 10^{-1}$	$\sim$ $U = 1.62 \cdot 10^{+3}$ $S = 0.875$ $p = 2.92 \cdot 10^{-1}$	$\sim$ $U = 1.53 \cdot 10^{+3}$ $S = 0.841$ $p = 1.35 \cdot 10^{-1}$	$\sim$ $U = 1.59 \cdot 10^{+3}$ $S = 0.877$ $p = 2.19 \cdot 10^{-1}$	$>$ $U = 2.23 \cdot 10^{+3}$ $S = 0.789$ $p = \mathbf{1.93 \cdot 10^{-2}}$	$>$ $U = 2.23 \cdot 10^{+3}$ $S = 0.794$ $p = \mathbf{1.96 \cdot 10^{-2}}$
	σ	$\sim^*$ $U = 2.12 \cdot 10^{+3}$ $S = 0.829$ $p = 1.26 \cdot 10^{-1}$	$\sim^*$ $U = 2.10 \cdot 10^{+3}$ $S = 0.823$ $p = 1.55 \cdot 10^{-1}$	$\sim^*$ $U = 2.13 \cdot 10^{+3}$ $S = 0.816$ $p = 1.22 \cdot 10^{-1}$	$\sim^*$ $U = 2.09 \cdot 10^{+3}$ $S = 0.821$ $p = 1.73 \cdot 10^{-1}$	$\sim^*$ $U = 2.02 \cdot 10^{+3}$ $S = 0.801$ $p = 3.29 \cdot 10^{-1}$	$\sim^*$ $U = 2.07 \cdot 10^{+3}$ $S = 0.779$ $p = 2.03 \cdot 10^{-1}$
Long. HDM	μ	$<$ $U = 1.15 \cdot 10^{+3}$ $S = 0.729$ $p = \mathbf{2.52 \cdot 10^{-4}}$	$<$ $U = 1.14 \cdot 10^{+3}$ $S = 0.731$ $p = \mathbf{2.25 \cdot 10^{-4}}$	$<$ $U = 1.13 \cdot 10^{+3}$ $S = 0.712$ $p = \mathbf{1.89 \cdot 10^{-4}}$	$<$ $U = 1.21 \cdot 10^{+3}$ $S = 0.739$ $p = \mathbf{8.20 \cdot 10^{-4}}$	$\sim$ $U = 1.58 \cdot 10^{+3}$ $S = 0.898$ $p = 2.14 \cdot 10^{-1}$	$\sim$ $U = 1.59 \cdot 10^{+3}$ $S = 0.899$ $p = 2.33 \cdot 10^{-1}$
	σ	$\sim$ $U = 1.84 \cdot 10^{+3}$ $S = 0.897$ $p = 9.28 \cdot 10^{-1}$	$\sim$ $U = 1.86 \cdot 10^{+3}$ $S = 0.894$ $p = 8.47 \cdot 10^{-1}$	$\sim$ $U = 1.89 \cdot 10^{+3}$ $S = 0.784$ $p = 7.40 \cdot 10^{-1}$	$\sim$ $U = 1.84 \cdot 10^{+3}$ $S = 0.887$ $p = 9.61 \cdot 10^{-1}$	$\sim$ $U = 1.78 \cdot 10^{+3}$ $S = 0.858$ $p = 8.35 \cdot 10^{-1}$	$\sim$ $U = 1.78 \cdot 10^{+3}$ $S = 0.868$ $p = 8.35 \cdot 10^{-1}$
Transv. HDM	μ	$\sim$ $U = 2.11 \cdot 10^{+3}$ $S = 0.865$ $p = 1.43 \cdot 10^{-1}$	$\sim$ $U = 2.13 \cdot 10^{+3}$ $S = 0.852$ $p = 1.13 \cdot 10^{-1}$	$\sim$ $U = 2.08 \cdot 10^{+3}$ $S = 0.863$ $p = 1.99 \cdot 10^{-1}$	$\sim$ $U = 2.10 \cdot 10^{+3}$ $S = 0.858$ $p = 1.58 \cdot 10^{-1}$	$>$ $U = 2.33 \cdot 10^{+3}$ $S = 0.808$ $p = \mathbf{4.83 \cdot 10^{-3}}$	$>$ $U = 2.31 \cdot 10^{+3}$ $S = 0.814$ $p = \mathbf{6.21 \cdot 10^{-3}}$
	σ	$\sim$ $U = 2.05 \cdot 10^{+3}$ $S = 0.911$ $p = 2.56 \cdot 10^{-1}$	$\sim$ $U = 1.98 \cdot 10^{+3}$ $S = 0.935$ $p = 4.38 \cdot 10^{-1}$	$\sim$ $U = 2.08 \cdot 10^{+3}$ $S = 0.909$ $p = 1.85 \cdot 10^{-1}$	$\sim$ $U = 2.04 \cdot 10^{+3}$ $S = 0.926$ $p = 2.78 \cdot 10^{-1}$	$\sim$ $U = 2.03 \cdot 10^{+3}$ $S = 0.912$ $p = 2.97 \cdot 10^{-1}$	$\sim$ $U = 2.03 \cdot 10^{+3}$ $S = 0.917$ $p = 2.85 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$\sim$ $U = 1.58 \cdot 10^{+3}$ $S = 0.895$ $p = 1.99 \cdot 10^{-1}$	$<^*$ $U = 1.27 \cdot 10^{+3}$ $S = 0.751$ $p = \mathbf{2.12 \cdot 10^{-3}}$	$<$ $U = 1.41 \cdot 10^{+3}$ $S = 0.853$ $p = \mathbf{1.68 \cdot 10^{-2}}$	$<$ $U = 1.24 \cdot 10^{+3}$ $S = 0.505$ $p = \mathbf{1.40 \cdot 10^{-3}}$	$<^*$ $U = 1.05 \cdot 10^{+3}$ $S = 0.592$ $p = \mathbf{3.19 \cdot 10^{-5}}$	$<^*$ $U = 8.28 \cdot 10^{+2}$ $S = 0.531$ $p = \mathbf{1.44 \cdot 10^{-7}}$
	σ	$\sim$ $U = 1.72 \cdot 10^{+3}$ $S = 0.897$ $p = 5.83 \cdot 10^{-1}$	$\sim^*$ $U = 1.46 \cdot 10^{+3}$ $S = 0.749$ $p = 6.42 \cdot 10^{-2}$	$<$ $U = 1.39 \cdot 10^{+3}$ $S = 0.916$ $p = \mathbf{1.30 \cdot 10^{-2}}$	$<$ $U = 1.36 \cdot 10^{+3}$ $S = 0.565$ $p = \mathbf{7.92 \cdot 10^{-3}}$	$<^*$ $U = 1.02 \cdot 10^{+3}$ $S = 0.547$ $p = \mathbf{1.64 \cdot 10^{-5}}$	$<^*$ $U = 8.63 \cdot 10^{+2}$ $S = 0.477$ $p = \mathbf{3.69 \cdot 10^{-7}}$
Rot. diffusion coef.	μ	$<$ $U = 1.41 \cdot 10^{+3}$ $S = 0.826$ $p = \mathbf{1.68 \cdot 10^{-2}}$	$<^*$ $U = 1.40 \cdot 10^{+3}$ $S = 0.793$ $p = \mathbf{1.35 \cdot 10^{-2}}$	$<$ $U = 1.42 \cdot 10^{+3}$ $S = 0.676$ $p = \mathbf{1.84 \cdot 10^{-2}}$	$<$ $U = 1.27 \cdot 10^{+3}$ $S = 0.342$ $p = \mathbf{2.26 \cdot 10^{-3}}$	$<^*$ $U = 1.02 \cdot 10^{+3}$ $S = 0.461$ $p = \mathbf{1.92 \cdot 10^{-5}}$	$<^*$ $U = 1.11 \cdot 10^{+3}$ $S = 0.439$ $p = \mathbf{1.22 \cdot 10^{-4}}$
	σ	$<$ $U = 1.40 \cdot 10^{+3}$ $S = 0.860$ $p = \mathbf{1.46 \cdot 10^{-2}}$	$<^*$ $U = 1.36 \cdot 10^{+3}$ $S = 0.745$ $p = \mathbf{8.15 \cdot 10^{-3}}$	$<$ $U = 1.39 \cdot 10^{+3}$ $S = 0.641$ $p = \mathbf{1.25 \cdot 10^{-2}}$	$<$ $U = 1.20 \cdot 10^{+3}$ $S = 0.301$ $p = \mathbf{7.12 \cdot 10^{-4}}$	$<^*$ $U = 1.01 \cdot 10^{+3}$ $S = 0.446$ $p = \mathbf{1.34 \cdot 10^{-5}}$	$<^*$ $U = 1.08 \cdot 10^{+3}$ $S = 0.288$ $p = \mathbf{6.32 \cdot 10^{-5}}$

Table D.7: Differences in the values of MD descriptors based on antimicrobial activity for each combination of lipid composition and force field. Results of the Mann-Whitney U test. Continued from the previous page. The samples are grouped by force field and lipid composition. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Each test compares the distribution of MD descriptors associated with active peptides against that of nonactives. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for active sequences is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.8).

Magnitude		Martini 2.2				Martini 2.2P				Martini 3			
		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1		POPC		POPE:POPG 3:1	
		Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value	Stat.	p-value
Bilayer thickness	μ	$1.86 \cdot 10^{-1}$	$6.67 \cdot 10^{-1}$	$7.57 \cdot 10^{+0}$	$\mathbf{6.85 \cdot 10^{-3}}$	$1.27 \cdot 10^{-1}$	$7.22 \cdot 10^{-1}$	$5.87 \cdot 10^{+0}$	$\mathbf{1.69 \cdot 10^{-2}}$	$1.11 \cdot 10^{+0}$	$2.94 \cdot 10^{-1}$	$1.43 \cdot 10^{+0}$	$2.34 \cdot 10^{-1}$
	σ	$9.68 \cdot 10^{-2}$	$7.56 \cdot 10^{-1}$	$5.63 \cdot 10^{-1}$	$4.55 \cdot 10^{-1}$	$2.16 \cdot 10^{+0}$	$1.44 \cdot 10^{-1}$	$2.28 \cdot 10^{-1}$	$6.34 \cdot 10^{-1}$	$1.57 \cdot 10^{+0}$	$2.13 \cdot 10^{-1}$	$8.10 \cdot 10^{-1}$	$3.70 \cdot 10^{-1}$
Area per lipid	μ	$2.97 \cdot 10^{-3}$	$9.57 \cdot 10^{-1}$	$4.33 \cdot 10^{-1}$	$5.12 \cdot 10^{-1}$	$1.03 \cdot 10^{-3}$	$9.74 \cdot 10^{-1}$	$1.57 \cdot 10^{+0}$	$2.13 \cdot 10^{-1}$	$3.34 \cdot 10^{+0}$	$7.00 \cdot 10^{-2}$	$3.26 \cdot 10^{-2}$	$8.57 \cdot 10^{-1}$
	σ	$1.91 \cdot 10^{-1}$	$6.63 \cdot 10^{-1}$	$3.56 \cdot 10^{-1}$	$5.52 \cdot 10^{-1}$	$2.98 \cdot 10^{+0}$	$8.70 \cdot 10^{-2}$	$2.62 \cdot 10^{+0}$	$1.08 \cdot 10^{-1}$	$3.78 \cdot 10^{-1}$	$5.40 \cdot 10^{-1}$	$2.50 \cdot 10^{-2}$	$8.75 \cdot 10^{-1}$
Peptide penetration	μ	$3.38 \cdot 10^{-2}$	$8.54 \cdot 10^{-1}$	$3.45 \cdot 10^{+0}$	$6.58 \cdot 10^{-2}$	$5.76 \cdot 10^{-2}$	$8.11 \cdot 10^{-1}$	$5.81 \cdot 10^{+0}$	$\mathbf{1.74 \cdot 10^{-2}}$	$1.06 \cdot 10^{+1}$	$\mathbf{1.44 \cdot 10^{-3}}$	$1.31 \cdot 10^{+1}$	$\mathbf{4.26 \cdot 10^{-4}}$
	σ	$5.52 \cdot 10^{-1}$	$4.59 \cdot 10^{-1}$	$4.80 \cdot 10^{-1}$	$4.90 \cdot 10^{-1}$	$5.51 \cdot 10^{-1}$	$4.59 \cdot 10^{-1}$	$4.80 \cdot 10^{+0}$	$\mathbf{3.03 \cdot 10^{-2}}$	$5.94 \cdot 10^{+1}$	$\mathbf{4.01 \cdot 10^{-12}}$	$1.65 \cdot 10^{+1}$	$\mathbf{8.75 \cdot 10^{-5}}$
Tilt	μ	$1.33 \cdot 10^{+1}$	$\mathbf{3.99 \cdot 10^{-4}}$	$1.88 \cdot 10^{+1}$	$\mathbf{3.04 \cdot 10^{-5}}$	$1.37 \cdot 10^{+1}$	$\mathbf{3.30 \cdot 10^{-4}}$	$1.08 \cdot 10^{+1}$	$\mathbf{1.31 \cdot 10^{-3}}$	$3.06 \cdot 10^{+0}$	$8.26 \cdot 10^{-2}$	$1.89 \cdot 10^{+1}$	$\mathbf{2.90 \cdot 10^{-5}}$
	σ	$4.08 \cdot 10^{-2}$	$8.40 \cdot 10^{-1}$	$1.10 \cdot 10^{+1}$	$\mathbf{1.20 \cdot 10^{-3}}$	$1.67 \cdot 10^{-1}$	$6.84 \cdot 10^{-1}$	$8.43 \cdot 10^{+0}$	$\mathbf{4.40 \cdot 10^{-3}}$	$2.25 \cdot 10^{+1}$	$\mathbf{5.76 \cdot 10^{-6}}$	$1.83 \cdot 10^{+1}$	$\mathbf{3.78 \cdot 10^{-5}}$
Spin	μ	$2.47 \cdot 10^{+0}$	$1.19 \cdot 10^{-1}$	$6.92 \cdot 10^{+0}$	$\mathbf{9.64 \cdot 10^{-3}}$	$2.58 \cdot 10^{+0}$	$1.11 \cdot 10^{-1}$	$1.60 \cdot 10^{+1}$	$\mathbf{1.08 \cdot 10^{-4}}$	$1.51 \cdot 10^{+1}$	$\mathbf{1.70 \cdot 10^{-4}}$	$1.48 \cdot 10^{+1}$	$\mathbf{1.89 \cdot 10^{-4}}$
	σ	$1.20 \cdot 10^{-1}$	$7.29 \cdot 10^{-1}$	$8.35 \cdot 10^{+0}$	$\mathbf{4.57 \cdot 10^{-3}}$	$7.70 \cdot 10^{-1}$	$3.82 \cdot 10^{-1}$	$1.33 \cdot 10^{+1}$	$\mathbf{4.00 \cdot 10^{-4}}$	$1.58 \cdot 10^{+1}$	$\mathbf{1.18 \cdot 10^{-4}}$	$1.82 \cdot 10^{+1}$	$\mathbf{3.98 \cdot 10^{-5}}$
Contacts peptide-HGs	μ	$1.06 \cdot 10^{+0}$	$3.06 \cdot 10^{-1}$	$3.23 \cdot 10^{+0}$	$7.48 \cdot 10^{-2}$	$5.91 \cdot 10^{-1}$	$4.44 \cdot 10^{-1}$	$6.70 \cdot 10^{+0}$	$\mathbf{1.08 \cdot 10^{-2}}$	$7.19 \cdot 10^{+0}$	$\mathbf{8.37 \cdot 10^{-3}}$	$8.95 \cdot 10^{+0}$	$\mathbf{3.36 \cdot 10^{-3}}$
	σ	$3.42 \cdot 10^{+0}$	$6.70 \cdot 10^{-2}$	$2.55 \cdot 10^{+0}$	$1.13 \cdot 10^{-1}$	$1.26 \cdot 10^{+0}$	$2.63 \cdot 10^{-1}$	$3.30 \cdot 10^{+0}$	$7.18 \cdot 10^{-2}$	$2.87 \cdot 10^{+1}$	$\mathbf{4.08 \cdot 10^{-7}}$	$1.28 \cdot 10^{+1}$	$\mathbf{4.98 \cdot 10^{-4}}$
Contacts peptide-tails	μ	$2.77 \cdot 10^{-1}$	$6.00 \cdot 10^{-1}$	$7.94 \cdot 10^{-1}$	$3.75 \cdot 10^{-1}$	$1.73 \cdot 10^{-1}$	$6.78 \cdot 10^{-1}$	$5.09 \cdot 10^{+0}$	$\mathbf{2.58 \cdot 10^{-2}}$	$4.29 \cdot 10^{+0}$	$\mathbf{4.05 \cdot 10^{-2}}$	$8.11 \cdot 10^{+0}$	$\mathbf{5.17 \cdot 10^{-3}}$
	σ	$2.21 \cdot 10^{+0}$	$1.40 \cdot 10^{-1}$	$3.24 \cdot 10^{+0}$	$7.45 \cdot 10^{-2}$	$1.62 \cdot 10^{+0}$	$2.05 \cdot 10^{-1}$	$8.34 \cdot 10^{+0}$	$\mathbf{4.60 \cdot 10^{-3}}$	$1.44 \cdot 10^{+1}$	$\mathbf{2.36 \cdot 10^{-4}}$	$1.23 \cdot 10^{+1}$	$\mathbf{6.23 \cdot 10^{-4}}$
Contacts peptide-water	μ	$3.39 \cdot 10^{-1}$	$5.61 \cdot 10^{-1}$	$2.12 \cdot 10^{-2}$	$8.85 \cdot 10^{-1}$	$9.62 \cdot 10^{-3}$	$9.22 \cdot 10^{-1}$	$2.92 \cdot 10^{+0}$	$9.01 \cdot 10^{-2}$	$2.53 \cdot 10^{+0}$	$1.15 \cdot 10^{-1}$	$5.81 \cdot 10^{+0}$	$\mathbf{1.74 \cdot 10^{-2}}$
	σ	$1.20 \cdot 10^{+1}$	$\mathbf{7.54 \cdot 10^{-4}}$	$1.05 \cdot 10^{+1}$	$\mathbf{1.52 \cdot 10^{-3}}$	$1.24 \cdot 10^{+1}$	$\mathbf{6.17 \cdot 10^{-4}}$	$8.87 \cdot 10^{+0}$	$\mathbf{3.50 \cdot 10^{-3}}$	$1.54 \cdot 10^{+1}$	$\mathbf{1.48 \cdot 10^{-4}}$	$1.37 \cdot 10^{+1}$	$\mathbf{3.18 \cdot 10^{-4}}$
EDM	μ	$7.55 \cdot 10^{-1}$	$3.87 \cdot 10^{-1}$	$6.17 \cdot 10^{-2}$	$8.04 \cdot 10^{-1}$	$5.05 \cdot 10^{-2}$	$8.23 \cdot 10^{-1}$	$2.92 \cdot 10^{-3}$	$9.57 \cdot 10^{-1}$	$7.29 \cdot 10^{-1}$	$3.95 \cdot 10^{-1}$	$4.55 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$
	σ	$6.29 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$	$8.26 \cdot 10^{-3}$	$9.28 \cdot 10^{-1}$	$9.51 \cdot 10^{-1}$	$3.32 \cdot 10^{-1}$	$4.54 \cdot 10^{-1}$	$5.02 \cdot 10^{-1}$	$1.87 \cdot 10^{+0}$	$1.74 \cdot 10^{-1}$	$1.20 \cdot 10^{-2}$	$9.13 \cdot 10^{-1}$
Long. EDM	μ	$5.48 \cdot 10^{-1}$	$4.60 \cdot 10^{-1}$	$1.65 \cdot 10^{-2}$	$8.98 \cdot 10^{-1}$	$3.06 \cdot 10^{-1}$	$5.81 \cdot 10^{-1}$	$2.82 \cdot 10^{-4}$	$9.87 \cdot 10^{-1}$	$2.99 \cdot 10^{-2}$	$8.63 \cdot 10^{-1}$	$4.72 \cdot 10^{-1}$	$4.94 \cdot 10^{-1}$
	σ	$6.18 \cdot 10^{-1}$	$4.33 \cdot 10^{-1}$	$1.93 \cdot 10^{-2}$	$8.90 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$	$3.23 \cdot 10^{-1}$	$1.48 \cdot 10^{-1}$	$7.01 \cdot 10^{-1}$	$1.24 \cdot 10^{+0}$	$2.68 \cdot 10^{-1}$	$4.36 \cdot 10^{-1}$	$5.10 \cdot 10^{-1}$
Transv. EDM	μ	$6.21 \cdot 10^{-1}$	$4.32 \cdot 10^{-1}$	$1.72 \cdot 10^{-1}$	$6.79 \cdot 10^{-1}$	$2.78 \cdot 10^{+0}$	$9.81 \cdot 10^{-2}$	$2.11 \cdot 10^{+0}$	$1.49 \cdot 10^{-1}$	$1.75 \cdot 10^{+0}$	$1.89 \cdot 10^{-1}$	$2.12 \cdot 10^{+0}$	$1.48 \cdot 10^{-1}$
	σ	$5.49 \cdot 10^{-1}$	$4.60 \cdot 10^{-1}$	$9.84 \cdot 10^{-2}$	$7.54 \cdot 10^{-1}$	$1.54 \cdot 10^{+0}$	$2.16 \cdot 10^{-1}$	$8.69 \cdot 10^{-1}$	$3.53 \cdot 10^{-1}$	$2.55 \cdot 10^{+0}$	$1.13 \cdot 10^{-1}$	$4.00 \cdot 10^{-1}$	$5.28 \cdot 10^{-1}$
HDM	μ	$1.08 \cdot 10^{+0}$	$3.00 \cdot 10^{-1}$	$8.11 \cdot 10^{-1}$	$3.70 \cdot 10^{-1}$	$1.11 \cdot 10^{+0}$	$2.95 \cdot 10^{-1}$	$4.19 \cdot 10^{-1}$	$5.19 \cdot 10^{-1}$	$2.39 \cdot 10^{+0}$	$1.25 \cdot 10^{-1}$	$2.11 \cdot 10^{+0}$	$1.49 \cdot 10^{-1}$
	σ	$5.25 \cdot 10^{+0}$	$\mathbf{2.37 \cdot 10^{-2}}$	$4.87 \cdot 10^{+0}$	$\mathbf{2.91 \cdot 10^{-2}}$	$4.24 \cdot 10^{+0}$	$\mathbf{4.16 \cdot 10^{-2}}$	$5.24 \cdot 10^{+0}$	$\mathbf{2.39 \cdot 10^{-2}}$	$6.64 \cdot 10^{+0}$	$\mathbf{1.12 \cdot 10^{-2}}$	$6.65 \cdot 10^{+0}$	$\mathbf{1.11 \cdot 10^{-2}}$
Long. HDM	μ	$7.54 \cdot 10^{-1}$	$3.87 \cdot 10^{-1}$	$8.06 \cdot 10^{-1}$	$3.71 \cdot 10^{-1}$	$6.29 \cdot 10^{-1}$	$4.29 \cdot 10^{-1}$	$1.16 \cdot 10^{+0}$	$2.83 \cdot 10^{-1}$	$3.54 \cdot 10^{-1}$	$5.53 \cdot 10^{-1}$	$5.70 \cdot 10^{-1}$	$4.52 \cdot 10^{-1}$
	σ	$1.04 \cdot 10^{+0}$	$3.10 \cdot 10^{-1}$	$7.81 \cdot 10^{-1}$	$3.79 \cdot 10^{-1}$	$1.20 \cdot 10^{+0}$	$2.76 \cdot 10^{-1}$	$9.44 \cdot 10^{-1}$	$3.33 \cdot 10^{-1}$	$1.42 \cdot 10^{+0}$	$2.36 \cdot 10^{-1}$	$1.31 \cdot 10^{+0}$	$2.54 \cdot 10^{-1}$
Transv. HDM	μ	$1.47 \cdot 10^{-4}$	$9.90 \cdot 10^{-1}$	$1.89 \cdot 10^{-2}$	$8.91 \cdot 10^{-1}$	$3.90 \cdot 10^{-4}$	$9.84 \cdot 10^{-1}$	$5.64 \cdot 10^{-2}$	$8.13 \cdot 10^{-1}$	$2.19 \cdot 10^{-1}$	$6.41 \cdot 10^{-1}$	$1.26 \cdot 10^{-1}$	$7.24 \cdot 10^{-1}$
	σ	$8.47 \cdot 10^{-1}$	$3.59 \cdot 10^{-1}$	$6.46 \cdot 10^{-1}$	$4.23 \cdot 10^{-1}$	$7.12 \cdot 10^{-1}$	$4.01 \cdot 10^{-1}$	$6.44 \cdot 10^{-1}$	$4.24 \cdot 10^{-1}$	$1.23 \cdot 10^{+0}$	$2.69 \cdot 10^{-1}$	$1.03 \cdot 10^{+0}$	$3.12 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$3.78 \cdot 10^{-2}$	$8.46 \cdot 10^{-1}$	$6.66 \cdot 10^{+0}$	$\mathbf{1.10 \cdot 10^{-2}}$	$3.28 \cdot 10^{-1}$	$5.68 \cdot 10^{-1}$	$3.85 \cdot 10^{+0}$	$5.22 \cdot 10^{-2}$	$1.53 \cdot 10^{+1}$	$\mathbf{1.52 \cdot 10^{-4}}$	$7.69 \cdot 10^{+0}$	$\mathbf{6.43 \cdot 10^{-3}}$
	σ	$1.35 \cdot 10^{-1}$	$7.14 \cdot 10^{-1}$	$1.36 \cdot 10^{+1}$	$\mathbf{3.41 \cdot 10^{-4}}$	$1.46 \cdot 10^{-1}$	$7.03 \cdot 10^{-1}$	$3.56 \cdot 10^{+0}$	$6.17 \cdot 10^{-2}$	$2.50 \cdot 10^{+1}$	$\mathbf{2.00 \cdot 10^{-6}}$	$1.42 \cdot 10^{+1}$	$\mathbf{2.52 \cdot 10^{-4}}$
Rot. diffusion coef.	μ	$1.86 \cdot 10^{+0}$	$1.75 \cdot 10^{-1}$	$6.00 \cdot 10^{+0}$	$\mathbf{1.57 \cdot 10^{-2}}$	$1.47 \cdot 10^{+0}$	$2.28 \cdot 10^{-1}$	$2.79 \cdot 10^{+0}$	$9.74 \cdot 10^{-2}$	$1.79 \cdot 10^{+1}$	$\mathbf{4.60 \cdot 10^{-5}}$	$1.28 \cdot 10^{+1}$	$\mathbf{5.11 \cdot 10^{-4}}$
	σ	$3.96 \cdot 10^{-1}$	$5.30 \cdot 10^{-1}$	$6.21 \cdot 10^{+0}$	$\mathbf{1.41 \cdot 10^{-2}}$	$1.85 \cdot 10^{+0}$	$1.76 \cdot 10^{-1}$	$2.71 \cdot 10^{+0}$	$1.03 \cdot 10^{-1}$	$1.80 \cdot 10^{+1}$	$\mathbf{4.27 \cdot 10^{-5}}$	$1.21 \cdot 10^{+1}$	$\mathbf{7.06 \cdot 10^{-4}}$

Table D.8: Differences in the variances of the samples of MD descriptors based on antimicrobial activity for each combination of lipid composition and force field. Results of the Brown-Forsythe test. The samples are grouped by force field and lipid composition . Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Each test compares the variance of the samples of MD descriptors associated with active and nonactive peptides. The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in the variances.

Magnitude		Martini 2.2 AMP	Martini 2.2 NC	Martini 2.2P AMP	Martini 2.2P NC	Martini 3 AMP	Martini 3 NC
Bilayer thickness	μ	$<$ $T = 6.00 \cdot 10^{+0}$ $S = 0.388$ $p = \mathbf{7.26 \cdot 10^{-14}}$	$<$ $T = 2.00 \cdot 10^{+0}$ $S = 0.476$ $p = \mathbf{2.66 \cdot 10^{-15}}$	$>$ $T = 2.23 \cdot 10^{+3}$ $S = 0.765$ $p = \mathbf{7.43 \cdot 10^{-7}}$	$>$ $T = 1.08 \cdot 10^{+3}$ $S = 0.800$ $p = \mathbf{2.39 \cdot 10^{-6}}$	$<$ $T = 7.10 \cdot 10^{+1}$ $S = 0.364$ $p = \mathbf{1.00 \cdot 10^{-12}}$	$<$ $T = 1.80 \cdot 10^{+1}$ $S = 0.606$ $p = \mathbf{2.25 \cdot 10^{-13}}$
	σ	$>$ $T = 2.20 \cdot 10^{+3}$ $S = 0.620$ $p = \mathbf{1.35 \cdot 10^{-6}}$	$>$ $T = 1.11 \cdot 10^{+3}$ $S = 0.589$ $p = \mathbf{4.39 \cdot 10^{-7}}$	$>$ $T = 1.87 \cdot 10^{+3}$ $S = 0.762$ $p = \mathbf{2.11 \cdot 10^{-3}}$	$>$ $T = 8.76 \cdot 10^{+2}$ $S = 0.725$ $p = \mathbf{1.03 \cdot 10^{-2}}$	$\sim$ $T = 1.07 \cdot 10^{+3}$ $S = 0.871$ $p = 1.28 \cdot 10^{-1}$	$\sim$ $T = 4.90 \cdot 10^{+2}$ $S = 0.817$ $p = 1.57 \cdot 10^{-1}$
Area per lipid	μ	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.244$ $p = \mathbf{5.90 \cdot 10^{-14}}$	$>$ $T = 1.27 \cdot 10^{+3}$ $S = 0.332$ $p = \mathbf{1.78 \cdot 10^{-15}}$	$<$ $T = 7.02 \cdot 10^{+2}$ $S = 0.836$ $p = \mathbf{1.82 \cdot 10^{-4}}$	$<$ $T = 3.19 \cdot 10^{+2}$ $S = 0.843$ $p = \mathbf{8.58 \cdot 10^{-4}}$	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.160$ $p = \mathbf{5.66 \cdot 10^{-14}}$	$>$ $T = 1.27 \cdot 10^{+3}$ $S = 0.277$ $p = \mathbf{8.88 \cdot 10^{-16}}$
	σ	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.081$ $p = \mathbf{5.66 \cdot 10^{-14}}$	$>$ $T = 1.27 \cdot 10^{+3}$ $S = 0.105$ $p = \mathbf{1.78 \cdot 10^{-15}}$	$>$ $T = 2.56 \cdot 10^{+3}$ $S = 0.460$ $p = \mathbf{1.65 \cdot 10^{-11}}$	$>$ $T = 1.16 \cdot 10^{+3}$ $S = 0.528$ $p = \mathbf{1.43 \cdot 10^{-8}}$	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.305$ $p = \mathbf{7.26 \cdot 10^{-14}}$	$>$ $T = 1.26 \cdot 10^{+3}$ $S = 0.259$ $p = \mathbf{7.82 \cdot 10^{-14}}$
Peptide penetration	μ	$>$ $T = 2.63 \cdot 10^{+3}$ $S = 0.636$ $p = \mathbf{1.13 \cdot 10^{-12}}$	$>$ $T = 1.23 \cdot 10^{+3}$ $S = 0.737$ $p = \mathbf{2.81 \cdot 10^{-11}}$	$<$ $T = 4.83 \cdot 10^{+2}$ $S = 0.406$ $p = \mathbf{9.25 \cdot 10^{-7}}$	$<$ $T = 2.81 \cdot 10^{+2}$ $S = 0.721$ $p = \mathbf{1.99 \cdot 10^{-4}}$	$\sim$ $T = 1.08 \cdot 10^{+3}$ $S = 0.594$ $p = 1.33 \cdot 10^{-1}$	$<$ $T = 1.60 \cdot 10^{+2}$ $S = 0.671$ $p = \mathbf{3.86 \cdot 10^{-7}}$
	σ	$>$ $T = 2.55 \cdot 10^{+3}$ $S = 0.541$ $p = \mathbf{2.22 \cdot 10^{-11}}$	$>$ $T = 1.01 \cdot 10^{+3}$ $S = 0.858$ $p = \mathbf{1.11 \cdot 10^{-4}}$	$>$ $T = 2.45 \cdot 10^{+3}$ $S = 0.361$ $p = \mathbf{6.76 \cdot 10^{-10}}$	$>$ $T = 1.12 \cdot 10^{+3}$ $S = 0.770$ $p = \mathbf{3.86 \cdot 10^{-7}}$	$>$ $T = 2.12 \cdot 10^{+3}$ $S = 0.625$ $p = \mathbf{1.29 \cdot 10^{-5}}$	$>$ $T = 1.16 \cdot 10^{+3}$ $S = 0.665$ $p = \mathbf{1.55 \cdot 10^{-8}}$
Tilt	μ	$\sim$ $T = 1.21 \cdot 10^{+3}$ $S = 0.933$ $p = 4.43 \cdot 10^{-1}$	$\sim$ $T = 5.60 \cdot 10^{+2}$ $S = 0.945$ $p = 4.61 \cdot 10^{-1}$	$\sim$ $T = 1.12 \cdot 10^{+3}$ $S = 0.922$ $p = 2.09 \cdot 10^{-1}$	$\sim$ $T = 5.88 \cdot 10^{+2}$ $S = 0.899$ $p = 6.39 \cdot 10^{-1}$	$\sim$ $T = 1.27 \cdot 10^{+3}$ $S = 0.811$ $p = 6.50 \cdot 10^{-1}$	$>$ $T = 9.34 \cdot 10^{+2}$ $S = 0.862$ $p = \mathbf{1.83 \cdot 10^{-3}}$
	σ	$>$ $T = 1.85 \cdot 10^{+3}$ $S = 0.808$ $p = \mathbf{3.22 \cdot 10^{-3}}$	$>$ $T = 8.63 \cdot 10^{+2}$ $S = 0.945$ $p = \mathbf{1.45 \cdot 10^{-2}}$	$>$ $T = 2.29 \cdot 10^{+3}$ $S = 0.497$ $p = \mathbf{1.24 \cdot 10^{-7}}$	$>$ $T = 1.01 \cdot 10^{+3}$ $S = 0.844$ $p = \mathbf{8.14 \cdot 10^{-5}}$	$>$ $T = 2.28 \cdot 10^{+3}$ $S = 0.702$ $p = \mathbf{1.48 \cdot 10^{-7}}$	$>$ $T = 1.13 \cdot 10^{+3}$ $S = 0.749$ $p = \mathbf{1.42 \cdot 10^{-7}}$
Spin	μ	$\sim$ $T = 1.09 \cdot 10^{+3}$ $S = 0.966$ $p = 1.54 \cdot 10^{-1}$	$\sim$ $T = 4.65 \cdot 10^{+2}$ $S = 0.953$ $p = 9.71 \cdot 10^{-2}$	$\sim$ $T = 1.20 \cdot 10^{+3}$ $S = 0.895$ $p = 3.93 \cdot 10^{-1}$	$<$ $T = 3.92 \cdot 10^{+2}$ $S = 0.882$ $p = \mathbf{8.55 \cdot 10^{-3}}$	$<$ $T = 8.14 \cdot 10^{+2}$ $S = 0.923$ $p = \mathbf{1.59 \cdot 10^{-3}}$	$\sim$ $T = 5.99 \cdot 10^{+2}$ $S = 0.917$ $p = 7.16 \cdot 10^{-1}$
	σ	$<$ $T = 9.45 \cdot 10^{+2}$ $S = 0.686$ $p = \mathbf{1.29 \cdot 10^{-2}}$	$\sim$ $T = 4.81 \cdot 10^{+2}$ $S = 0.965$ $p = 1.33 \cdot 10^{-1}$	$>$ $T = 1.89 \cdot 10^{+3}$ $S = 0.513$ $p = \mathbf{1.59 \cdot 10^{-3}}$	$>$ $T = 8.65 \cdot 10^{+2}$ $S = 0.842$ $p = \mathbf{1.38 \cdot 10^{-2}}$	$\sim$ $T = 1.28 \cdot 10^{+3}$ $S = 0.719$ $p = 7.19 \cdot 10^{-1}$	$>$ $T = 9.54 \cdot 10^{+2}$ $S = 0.795$ $p = \mathbf{9.21 \cdot 10^{-4}}$
Contacts peptide-HGs	μ	$<$ $T = 3.56 \cdot 10^{+2}$ $S = 0.936$ $p = \mathbf{3.75 \cdot 10^{-8}}$	$<$ $T = 2.72 \cdot 10^{+2}$ $S = 0.967$ $p = \mathbf{2.13 \cdot 10^{-4}}$	$<$ $T = 2.31 \cdot 10^{+2}$ $S = 0.745$ $p = \mathbf{6.10 \cdot 10^{-10}}$	$<$ $T = 1.74 \cdot 10^{+2}$ $S = 0.875$ $p = \mathbf{9.30 \cdot 10^{-7}}$	$<$ $T = 7.87 \cdot 10^{+2}$ $S = 0.818$ $p = \mathbf{9.75 \cdot 10^{-4}}$	$<$ $T = 1.56 \cdot 10^{+2}$ $S = 0.755$ $p = \mathbf{2.98 \cdot 10^{-7}}$
	σ	$<$ $T = 2.14 \cdot 10^{+2}$ $S = 0.903$ $p = \mathbf{2.08 \cdot 10^{-10}}$	$<$ $T = 2.67 \cdot 10^{+2}$ $S = 0.924$ $p = \mathbf{1.11 \cdot 10^{-4}}$	$\sim$ $T = 1.19 \cdot 10^{+3}$ $S = 0.864$ $p = 3.87 \cdot 10^{-1}$	$\sim$ $T = 5.11 \cdot 10^{+2}$ $S = 0.877$ $p = 2.26 \cdot 10^{-1}$	$<$ $T = 7.29 \cdot 10^{+2}$ $S = 0.886$ $p = \mathbf{3.17 \cdot 10^{-4}}$	$\sim$ $T = 6.28 \cdot 10^{+2}$ $S = 0.852$ $p = 9.31 \cdot 10^{-1}$
Contacts peptide-tails	μ	$>$ $T = 2.50 \cdot 10^{+3}$ $S = 0.908$ $p = \mathbf{1.22 \cdot 10^{-10}}$	$>$ $T = 1.21 \cdot 10^{+3}$ $S = 0.927$ $p = \mathbf{1.74 \cdot 10^{-10}}$	$<$ $T = 6.42 \cdot 10^{+2}$ $S = 0.787$ $p = \mathbf{4.91 \cdot 10^{-5}}$	$<$ $T = 2.53 \cdot 10^{+2}$ $S = 0.909$ $p = \mathbf{1.03 \cdot 10^{-4}}$	$<$ $T = 3.21 \cdot 10^{+2}$ $S = 0.833$ $p = \mathbf{7.58 \cdot 10^{-9}}$	$<$ $T = 1.20 \cdot 10^{+1}$ $S = 0.794$ $p = \mathbf{6.22 \cdot 10^{-14}}$
	σ	$>$ $T = 1.97 \cdot 10^{+3}$ $S = 0.927$ $p = \mathbf{3.04 \cdot 10^{-4}}$	$>$ $T = 1.03 \cdot 10^{+3}$ $S = 0.951$ $p = \mathbf{4.96 \cdot 10^{-5}}$	$\sim$ $T = 1.10 \cdot 10^{+3}$ $S = 0.822$ $p = 1.60 \cdot 10^{-1}$	$\sim$ $T = 4.52 \cdot 10^{+2}$ $S = 0.890$ $p = 7.40 \cdot 10^{-2}$	$>$ $T = 1.75 \cdot 10^{+3}$ $S = 0.834$ $p = \mathbf{1.35 \cdot 10^{-2}}$	$\sim$ $T = 6.16 \cdot 10^{+2}$ $S = 0.812$ $p = 8.41 \cdot 10^{-1}$
Contacts peptide-water	μ	$<$ $T = 2.12 \cdot 10^{+2}$ $S = 0.869$ $p = \mathbf{1.94 \cdot 10^{-10}}$	$<$ $T = 1.64 \cdot 10^{+2}$ $S = 0.911$ $p = \mathbf{4.07 \cdot 10^{-6}}$	$>$ $T = 2.65 \cdot 10^{+3}$ $S = 0.728$ $p = \mathbf{3.86 \cdot 10^{-13}}$	$>$ $T = 1.18 \cdot 10^{+3}$ $S = 0.863$ $p = \mathbf{4.11 \cdot 10^{-9}}$	$>$ $T = 2.06 \cdot 10^{+3}$ $S = 0.745$ $p = \mathbf{4.69 \cdot 10^{-5}}$	$>$ $T = 1.20 \cdot 10^{+3}$ $S = 0.780$ $p = \mathbf{7.25 \cdot 10^{-10}}$
	σ	$<$ $T = 5.46 \cdot 10^{+2}$ $S = 0.934$ $p = \mathbf{4.87 \cdot 10^{-6}}$	$\sim$ $T = 4.61 \cdot 10^{+2}$ $S = 0.924$ $p = 8.94 \cdot 10^{-2}$	$\sim$ $T = 1.24 \cdot 10^{+3}$ $S = 0.972$ $p = 5.58 \cdot 10^{-1}$	$\sim$ $T = 5.14 \cdot 10^{+2}$ $S = 0.912$ $p = 2.37 \cdot 10^{-1}$	$\sim$ $T = 1.08 \cdot 10^{+3}$ $S = 0.840$ $p = 1.37 \cdot 10^{-1}$	$\sim$ $T = 5.37 \cdot 10^{+2}$ $S = 0.869$ $p = 3.37 \cdot 10^{-1}$
EDM	μ	$\sim$ $T = 1.33 \cdot 10^{+3}$ $S = 0.793$ $p = 9.28 \cdot 10^{-1}$	$<$ $T = 3.20 \cdot 10^{+2}$ $S = 0.966$ $p = \mathbf{8.89 \cdot 10^{-4}}$	$<$ $T = 4.89 \cdot 10^{+2}$ $S = 0.745$ $p = \mathbf{1.09 \cdot 10^{-6}}$	$<$ $T = 3.48 \cdot 10^{+2}$ $S = 0.896$ $p = \mathbf{2.30 \cdot 10^{-3}}$	$\sim$ $T = 1.08 \cdot 10^{+3}$ $S = 0.959$ $p = 1.40 \cdot 10^{-1}$	$\sim$ $T = 4.47 \cdot 10^{+2}$ $S = 0.765$ $p = 6.63 \cdot 10^{-2}$
	σ	$\sim$ $T = 1.22 \cdot 10^{+3}$ $S = 0.571$ $p = 4.80 \cdot 10^{-1}$	$\sim$ $T = 5.43 \cdot 10^{+2}$ $S = 0.974$ $p = 3.67 \cdot 10^{-1}$	$>$ $T = 1.76 \cdot 10^{+3}$ $S = 0.904$ $p = \mathbf{1.15 \cdot 10^{-2}}$	$\sim$ $T = 5.05 \cdot 10^{+2}$ $S = 0.670$ $p = 2.04 \cdot 10^{-1}$	$\sim$ $T = 1.19 \cdot 10^{+3}$ $S = 0.962$ $p = 3.69 \cdot 10^{-1}$	$\sim$ $T = 6.09 \cdot 10^{+2}$ $S = 0.690$ $p = 7.89 \cdot 10^{-1}$

Table D.9: Differences in the values of MD descriptors based on lipid composition for each combination of antimicrobial activity and force field. Results of the Wilcoxon signed-rank test. Continues on the next page.

Magnitude		Martini 2.2 AMP	Martini 2.2 NC	Martini 2.2P AMP	Martini 2.2P NC	Martini 3 AMP	Martini 3 NC
Long. EDM	μ	$\sim$ $T = 1.14 \cdot 10^{+3}$ $S = 0.917$ $p = 2.59 \cdot 10^{-1}$	$<$ $T = 3.58 \cdot 10^{+2}$ $S = 0.980$ $p = \mathbf{3.16 \cdot 10^{-3}}$	$<$ $T = 6.18 \cdot 10^{+2}$ $S = 0.782$ $p = \mathbf{2.83 \cdot 10^{-5}}$	$<$ $T = 2.94 \cdot 10^{+2}$ $S = 0.955$ $p = \mathbf{3.36 \cdot 10^{-4}}$	$<$ $T = 4.72 \cdot 10^{+2}$ $S = 0.962$ $p = \mathbf{6.84 \cdot 10^{-7}}$	$<$ $T = 3.58 \cdot 10^{+2}$ $S = 0.881$ $p = \mathbf{3.16 \cdot 10^{-3}}$
	σ	$\sim$ $T = 1.18 \cdot 10^{+3}$ $S = 0.590$ $p = 3.63 \cdot 10^{-1}$	$\sim$ $T = 5.02 \cdot 10^{+2}$ $S = 0.964$ $p = 1.94 \cdot 10^{-1}$	$>$ $T = 1.76 \cdot 10^{+3}$ $S = 0.932$ $p = \mathbf{1.17 \cdot 10^{-2}}$	$\sim$ $T = 4.80 \cdot 10^{+2}$ $S = 0.677$ $p = 1.30 \cdot 10^{-1}$	$\sim$ $T = 1.06 \cdot 10^{+3}$ $S = 0.976$ $p = 1.17 \cdot 10^{-1}$	$\sim$ $T = 5.96 \cdot 10^{+2}$ $S = 0.672$ $p = 6.95 \cdot 10^{-1}$
Transv. EDM	μ	$\sim$ $T = 1.02 \cdot 10^{+3}$ $S = 0.741$ $p = 6.67 \cdot 10^{-2}$	$\sim$ $T = 5.73 \cdot 10^{+2}$ $S = 0.986$ $p = 5.40 \cdot 10^{-1}$	$<$ $T = 9.90 \cdot 10^{+2}$ $S = 0.951$ $p = \mathbf{2.37 \cdot 10^{-2}}$	$\sim$ $T = 5.59 \cdot 10^{+2}$ $S = 0.866$ $p = 4.55 \cdot 10^{-1}$	$>$ $T = 1.93 \cdot 10^{+3}$ $S = 0.945$ $p = \mathbf{7.64 \cdot 10^{-4}}$	$\sim$ $T = 6.26 \cdot 10^{+2}$ $S = 0.774$ $p = 9.16 \cdot 10^{-1}$
	σ	$\sim$ $T = 1.26 \cdot 10^{+3}$ $S = 0.529$ $p = 6.03 \cdot 10^{-1}$	$\sim$ $T = 6.11 \cdot 10^{+2}$ $S = 0.970$ $p = 8.04 \cdot 10^{-1}$	$\sim$ $T = 1.11 \cdot 10^{+3}$ $S = 0.956$ $p = 1.90 \cdot 10^{-1}$	$\sim$ $T = 4.72 \cdot 10^{+2}$ $S = 0.626$ $p = 1.12 \cdot 10^{-1}$	$\sim$ $T = 1.19 \cdot 10^{+3}$ $S = 0.966$ $p = 3.81 \cdot 10^{-1}$	$\sim$ $T = 5.88 \cdot 10^{+2}$ $S = 0.672$ $p = 6.39 \cdot 10^{-1}$
HDM	μ	$\sim$ $T = 1.28 \cdot 10^{+3}$ $S = 0.991$ $p = 7.15 \cdot 10^{-1}$	$\sim$ $T = 5.89 \cdot 10^{+2}$ $S = 0.984$ $p = 6.46 \cdot 10^{-1}$	$\sim$ $T = 1.26 \cdot 10^{+3}$ $S = 0.983$ $p = 6.34 \cdot 10^{-1}$	$\sim$ $T = 5.32 \cdot 10^{+2}$ $S = 0.965$ $p = 3.14 \cdot 10^{-1}$	$\sim$ $T = 1.15 \cdot 10^{+3}$ $S = 0.983$ $p = 2.61 \cdot 10^{-1}$	$\sim$ $T = 5.94 \cdot 10^{+2}$ $S = 0.985$ $p = 6.81 \cdot 10^{-1}$
	σ	$\sim$ $T = 1.27 \cdot 10^{+3}$ $S = 0.967$ $p = 6.66 \cdot 10^{-1}$	$\sim$ $T = 5.15 \cdot 10^{+2}$ $S = 0.965$ $p = 2.41 \cdot 10^{-1}$	$\sim$ $T = 1.01 \cdot 10^{+3}$ $S = 0.959$ $p = 6.51 \cdot 10^{-2}$	$\sim$ $T = 5.09 \cdot 10^{+2}$ $S = 0.970$ $p = 2.19 \cdot 10^{-1}$	$\sim$ $T = 1.24 \cdot 10^{+3}$ $S = 0.969$ $p = 5.55 \cdot 10^{-1}$	$\sim$ $T = 6.16 \cdot 10^{+2}$ $S = 0.963$ $p = 8.41 \cdot 10^{-1}$
Long. HDM	μ	$\sim$ $T = 1.23 \cdot 10^{+3}$ $S = 0.992$ $p = 5.15 \cdot 10^{-1}$	$\sim$ $T = 5.87 \cdot 10^{+2}$ $S = 0.992$ $p = 6.32 \cdot 10^{-1}$	$\sim$ $T = 1.14 \cdot 10^{+3}$ $S = 0.970$ $p = 2.38 \cdot 10^{-1}$	$\sim$ $T = 5.31 \cdot 10^{+2}$ $S = 0.967$ $p = 3.09 \cdot 10^{-1}$	$\sim$ $T = 1.20 \cdot 10^{+3}$ $S = 0.989$ $p = 4.14 \cdot 10^{-1}$	$\sim$ $T = 6.29 \cdot 10^{+2}$ $S = 0.983$ $p = 9.39 \cdot 10^{-1}$
	σ	$<$ $T = 9.87 \cdot 10^{+2}$ $S = 0.973$ $p = \mathbf{2.28 \cdot 10^{-2}}$	$\sim$ $T = 5.49 \cdot 10^{+2}$ $S = 0.971$ $p = 3.99 \cdot 10^{-1}$	$\sim$ $T = 1.11 \cdot 10^{+3}$ $S = 0.850$ $p = 1.94 \cdot 10^{-1}$	$\sim$ $T = 5.35 \cdot 10^{+2}$ $S = 0.972$ $p = 3.28 \cdot 10^{-1}$	$\sim$ $T = 1.21 \cdot 10^{+3}$ $S = 0.979$ $p = 4.30 \cdot 10^{-1}$	$\sim$ $T = 5.58 \cdot 10^{+2}$ $S = 0.979$ $p = 4.49 \cdot 10^{-1}$
Transv. HDM	μ	$\sim$ $T = 1.20 \cdot 10^{+3}$ $S = 0.986$ $p = 4.14 \cdot 10^{-1}$	$\sim$ $T = 4.48 \cdot 10^{+2}$ $S = 0.993$ $p = 6.78 \cdot 10^{-2}$	$\sim$ $T = 1.26 \cdot 10^{+3}$ $S = 0.986$ $p = 6.27 \cdot 10^{-1}$	$\sim$ $T = 5.39 \cdot 10^{+2}$ $S = 0.987$ $p = 3.47 \cdot 10^{-1}$	$\sim$ $T = 1.07 \cdot 10^{+3}$ $S = 0.976$ $p = 1.27 \cdot 10^{-1}$	$\sim$ $T = 6.08 \cdot 10^{+2}$ $S = 0.989$ $p = 7.81 \cdot 10^{-1}$
	σ	$\sim$ $T = 1.00 \cdot 10^{+3}$ $S = 0.976$ $p = 5.61 \cdot 10^{-2}$	$\sim$ $T = 5.06 \cdot 10^{+2}$ $S = 0.980$ $p = 2.08 \cdot 10^{-1}$	$\sim$ $T = 1.09 \cdot 10^{+3}$ $S = 0.968$ $p = 1.54 \cdot 10^{-1}$	$\sim$ $T = 5.27 \cdot 10^{+2}$ $S = 0.961$ $p = 2.91 \cdot 10^{-1}$	$\sim$ $T = 1.13 \cdot 10^{+3}$ $S = 0.971$ $p = 2.30 \cdot 10^{-1}$	$\sim$ $T = 5.88 \cdot 10^{+2}$ $S = 0.983$ $p = 6.39 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$>$ $T = 2.63 \cdot 10^{+3}$ $S = 0.625$ $p = \mathbf{1.00 \cdot 10^{-12}}$	$>$ $T = 1.15 \cdot 10^{+3}$ $S = 0.744$ $p = \mathbf{4.20 \cdot 10^{-8}}$	$>$ $T = 2.63 \cdot 10^{+3}$ $S = 0.303$ $p = \mathbf{1.17 \cdot 10^{-12}}$	$>$ $T = 1.22 \cdot 10^{+3}$ $S = 0.569$ $p = \mathbf{5.73 \cdot 10^{-11}}$	$>$ $T = 2.65 \cdot 10^{+3}$ $S = 0.550$ $p = \mathbf{4.72 \cdot 10^{-13}}$	$>$ $T = 1.22 \cdot 10^{+3}$ $S = 0.609$ $p = \mathbf{9.05 \cdot 10^{-11}}$
	σ	$>$ $T = 2.50 \cdot 10^{+3}$ $S = 0.531$ $p = \mathbf{1.51 \cdot 10^{-10}}$	$>$ $T = 1.05 \cdot 10^{+3}$ $S = 0.762$ $p = \mathbf{1.91 \cdot 10^{-5}}$	$>$ $T = 2.53 \cdot 10^{+3}$ $S = 0.318$ $p = \mathbf{4.79 \cdot 10^{-11}}$	$>$ $T = 1.20 \cdot 10^{+3}$ $S = 0.552$ $p = \mathbf{6.57 \cdot 10^{-10}}$	$>$ $T = 2.49 \cdot 10^{+3}$ $S = 0.562$ $p = \mathbf{1.68 \cdot 10^{-10}}$	$>$ $T = 1.09 \cdot 10^{+3}$ $S = 0.649$ $p = \mathbf{1.59 \cdot 10^{-6}}$
Rot. diffusion coef.	μ	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.854$ $p = \mathbf{5.66 \cdot 10^{-14}}$	$>$ $T = 1.20 \cdot 10^{+3}$ $S = 0.915$ $p = \mathbf{7.25 \cdot 10^{-10}}$	$>$ $T = 2.64 \cdot 10^{+3}$ $S = 0.424$ $p = \mathbf{7.60 \cdot 10^{-13}}$	$>$ $T = 1.19 \cdot 10^{+3}$ $S = 0.776$ $p = \mathbf{1.54 \cdot 10^{-9}}$	$>$ $T = 2.70 \cdot 10^{+3}$ $S = 0.692$ $p = \mathbf{5.66 \cdot 10^{-14}}$	$>$ $T = 1.27 \cdot 10^{+3}$ $S = 0.705$ $p = \mathbf{8.88 \cdot 10^{-16}}$
	σ	$>$ $T = 2.37 \cdot 10^{+3}$ $S = 0.760$ $p = \mathbf{1.11 \cdot 10^{-8}}$	$>$ $T = 1.06 \cdot 10^{+3}$ $S = 0.905$ $p = \mathbf{1.15 \cdot 10^{-5}}$	$>$ $T = 2.58 \cdot 10^{+3}$ $S = 0.415$ $p = \mathbf{6.68 \cdot 10^{-12}}$	$>$ $T = 1.16 \cdot 10^{+3}$ $S = 0.802$ $p = \mathbf{1.81 \cdot 10^{-8}}$	$>$ $T = 2.44 \cdot 10^{+3}$ $S = 0.541$ $p = \mathbf{1.05 \cdot 10^{-9}}$	$>$ $T = 1.24 \cdot 10^{+3}$ $S = 0.739$ $p = \mathbf{2.46 \cdot 10^{-12}}$

Table D.9: Differences in the values of MD descriptors based on lipid composition for each combination of antimicrobial activity and force field. Results of the Wilcoxon signed-rank test. Continued from the previous page. The samples are grouped by force field—M2, M2P or M3—and antimicrobial activity—AMP for active sequences, NC for nonactive sequences. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Each test compares the distribution of the differences between MD descriptors associated with POPC and POPE:POPG simulations. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution of differences $x_{\text{POPC}} - x_{\text{POPE:POPG}}$ is skewed toward positive values, i.e. larger values for POPC simulations, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$.

Magnitude		Martini 2.2	Martini 2.2P	Martini 3
Bilayer thickness	μ	$\sim$ $U = 1.78 \cdot 10^{+3}$ $S = 0.942$ $p = 8.07 \cdot 10^{-1}$	$\sim$ $U = 1.78 \cdot 10^{+3}$ $S = 0.890$ $p = 8.15 \cdot 10^{-1}$	$<$ $U = 1.32 \cdot 10^{+3}$ $S = 0.788$ $p = \mathbf{4.62 \cdot 10^{-3}}$
	σ	$\sim$ $U = 1.76 \cdot 10^{+3}$ $S = 0.911$ $p = 7.48 \cdot 10^{-1}$	$\sim$ $U = 1.81 \cdot 10^{+3}$ $S = 0.913$ $p = 9.36 \cdot 10^{-1}$	$\sim$ $U = 1.83 \cdot 10^{+3}$ $S = 0.925$ $p = 9.73 \cdot 10^{-1}$
Area per lipid	μ	$\sim$ $U = 2.13 \cdot 10^{+3}$ $S = 0.868$ $p = 1.19 \cdot 10^{-1}$	$\sim$ $U = 1.85 \cdot 10^{+3}$ $S = 0.898$ $p = 9.12 \cdot 10^{-1}$	$>$ $U = 2.34 \cdot 10^{+3}$ $S = 0.800$ $p = \mathbf{4.22 \cdot 10^{-3}}$
	σ	$\sim$ $U = 1.88 \cdot 10^{+3}$ $S = 0.930$ $p = 7.75 \cdot 10^{-1}$	$\sim$ $U = 1.95 \cdot 10^{+3}$ $S = 0.928$ $p = 5.18 \cdot 10^{-1}$	$\sim$ $U = 1.70 \cdot 10^{+3}$ $S = 0.930$ $p = 5.35 \cdot 10^{-1}$
Peptide penetration	μ	$\sim$ $U = 2.09 \cdot 10^{+3}$ $S = 0.506$ $p = 1.73 \cdot 10^{-1}$	$\sim$ $U = 1.75 \cdot 10^{+3}$ $S = 0.932$ $p = 6.97 \cdot 10^{-1}$	$>^*$ $U = 2.50 \cdot 10^{+3}$ $S = 0.726$ $p = \mathbf{2.78 \cdot 10^{-4}}$
	σ	$>$ $U = 2.24 \cdot 10^{+3}$ $S = 0.478$ $p = \mathbf{1.56 \cdot 10^{-2}}$	$\sim$ $U = 1.53 \cdot 10^{+3}$ $S = 0.841$ $p = 1.35 \cdot 10^{-1}$	$<^*$ $U = 1.16 \cdot 10^{+3}$ $S = 0.588$ $p = \mathbf{3.17 \cdot 10^{-4}}$
Tilt	μ	$\sim$ $U = 1.78 \cdot 10^{+3}$ $S = 0.563$ $p = 8.03 \cdot 10^{-1}$	$\sim$ $U = 2.11 \cdot 10^{+3}$ $S = 0.699$ $p = 1.39 \cdot 10^{-1}$	$<$ $U = 1.37 \cdot 10^{+3}$ $S = 0.816$ $p = \mathbf{1.00 \cdot 10^{-2}}$
	σ	$\sim$ $U = 1.79 \cdot 10^{+3}$ $S = 0.540$ $p = 8.63 \cdot 10^{-1}$	$\sim$ $U = 1.81 \cdot 10^{+3}$ $S = 0.951$ $p = 9.32 \cdot 10^{-1}$	$<^*$ $U = 1.41 \cdot 10^{+3}$ $S = 0.781$ $p = \mathbf{1.68 \cdot 10^{-2}}$
Spin	μ	$\sim$ $U = 1.97 \cdot 10^{+3}$ $S = 0.845$ $p = 4.57 \cdot 10^{-1}$	$>$ $U = 2.32 \cdot 10^{+3}$ $S = 0.816$ $p = \mathbf{5.53 \cdot 10^{-3}}$	$\sim^*$ $U = 1.46 \cdot 10^{+3}$ $S = 0.522$ $p = 5.71 \cdot 10^{-2}$
	σ	$\sim$ $U = 1.82 \cdot 10^{+3}$ $S = 0.459$ $p = 1.00 \cdot 10^{+0}$	$\sim$ $U = 1.79 \cdot 10^{+3}$ $S = 0.927$ $p = 8.47 \cdot 10^{-1}$	$<^*$ $U = 1.40 \cdot 10^{+3}$ $S = 0.711$ $p = \mathbf{1.48 \cdot 10^{-2}}$
Contacts peptide-HGs	μ	$\sim$ $U = 1.62 \cdot 10^{+3}$ $S = 0.689$ $p = 2.85 \cdot 10^{-1}$	$\sim$ $U = 1.84 \cdot 10^{+3}$ $S = 0.938$ $p = 9.59 \cdot 10^{-1}$	$>$ $U = 2.29 \cdot 10^{+3}$ $S = 0.793$ $p = \mathbf{8.62 \cdot 10^{-3}}$
	σ	$\sim$ $U = 1.60 \cdot 10^{+3}$ $S = 0.856$ $p = 2.48 \cdot 10^{-1}$	$\sim$ $U = 1.91 \cdot 10^{+3}$ $S = 0.744$ $p = 6.79 \cdot 10^{-1}$	$\sim^*$ $U = 1.67 \cdot 10^{+3}$ $S = 0.619$ $p = 4.23 \cdot 10^{-1}$
Contacts peptide-tails	μ	$\sim$ $U = 1.79 \cdot 10^{+3}$ $S = 0.811$ $p = 8.57 \cdot 10^{-1}$	$\sim$ $U = 1.97 \cdot 10^{+3}$ $S = 0.928$ $p = 4.57 \cdot 10^{-1}$	$>$ $U = 2.51 \cdot 10^{+3}$ $S = 0.758$ $p = \mathbf{2.16 \cdot 10^{-4}}$
	σ	$\sim$ $U = 1.64 \cdot 10^{+3}$ $S = 0.915$ $p = 3.34 \cdot 10^{-1}$	$\sim$ $U = 1.51 \cdot 10^{+3}$ $S = 0.759$ $p = 1.02 \cdot 10^{-1}$	$\sim^*$ $U = 1.94 \cdot 10^{+3}$ $S = 0.625$ $p = 5.59 \cdot 10^{-1}$
Contacts peptide-water	μ	$\sim$ $U = 1.71 \cdot 10^{+3}$ $S = 0.748$ $p = 5.42 \cdot 10^{-1}$	$\sim$ $U = 1.86 \cdot 10^{+3}$ $S = 0.937$ $p = 8.55 \cdot 10^{-1}$	$<$ $U = 1.35 \cdot 10^{+3}$ $S = 0.813$ $p = \mathbf{6.87 \cdot 10^{-3}}$
	σ	$\sim^*$ $U = 1.60 \cdot 10^{+3}$ $S = 0.809$ $p = 2.56 \cdot 10^{-1}$	$\sim^*$ $U = 1.52 \cdot 10^{+3}$ $S = 0.508$ $p = 1.22 \cdot 10^{-1}$	$\sim^*$ $U = 1.88 \cdot 10^{+3}$ $S = 0.647$ $p = 7.71 \cdot 10^{-1}$
EDM	μ	$>$ $U = 2.28 \cdot 10^{+3}$ $S = 0.226$ $p = \mathbf{9.90 \cdot 10^{-3}}$	$\sim$ $U = 1.76 \cdot 10^{+3}$ $S = 0.772$ $p = 7.36 \cdot 10^{-1}$	$\sim$ $U = 1.94 \cdot 10^{+3}$ $S = 0.326$ $p = 5.52 \cdot 10^{-1}$
	σ	$\sim$ $U = 1.77 \cdot 10^{+3}$ $S = 0.164$ $p = 7.95 \cdot 10^{-1}$	$\sim$ $U = 1.90 \cdot 10^{+3}$ $S = 0.547$ $p = 6.90 \cdot 10^{-1}$	$\sim$ $U = 1.94 \cdot 10^{+3}$ $S = 0.244$ $p = 5.59 \cdot 10^{-1}$
Long. EDM	μ	$\sim$ $U = 2.10 \cdot 10^{+3}$ $S = 0.306$ $p = 1.59 \cdot 10^{-1}$	$\sim$ $U = 1.76 \cdot 10^{+3}$ $S = 0.561$ $p = 7.59 \cdot 10^{-1}$	$\sim$ $U = 1.67 \cdot 10^{+3}$ $S = 0.460$ $p = 4.20 \cdot 10^{-1}$
	σ	$\sim$ $U = 1.76 \cdot 10^{+3}$ $S = 0.173$ $p = 7.48 \cdot 10^{-1}$	$\sim$ $U = 1.89 \cdot 10^{+3}$ $S = 0.526$ $p = 7.28 \cdot 10^{-1}$	$\sim$ $U = 1.95 \cdot 10^{+3}$ $S = 0.278$ $p = 5.11 \cdot 10^{-1}$
Transv. EDM	μ	$\sim$ $U = 2.19 \cdot 10^{+3}$ $S = 0.177$ $p = 6.12 \cdot 10^{-2}$	$\sim$ $U = 1.73 \cdot 10^{+3}$ $S = 0.445$ $p = 6.38 \cdot 10^{-1}$	$>$ $U = 2.25 \cdot 10^{+3}$ $S = 0.335$ $p = \mathbf{1.42 \cdot 10^{-2}}$
	σ	$\sim$ $U = 1.95 \cdot 10^{+3}$ $S = 0.173$ $p = 5.32 \cdot 10^{-1}$	$\sim$ $U = 1.72 \cdot 10^{+3}$ $S = 0.259$ $p = 6.05 \cdot 10^{-1}$	$\sim$ $U = 2.00 \cdot 10^{+3}$ $S = 0.257$ $p = 3.63 \cdot 10^{-1}$

Table D.10: Differential behavior of peptides on the interaction with lipid membranes, based on their activity, for each force field. Results of the Mann-Whitney U test. Continues on the next page.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		Martini 2.2		Martini 2.2P		Martini 3	
HDM	μ	$\sim$	$U = 1.68 \cdot 10^{+3}$	$\sim$	$U = 1.64 \cdot 10^{+3}$	$\sim$	$U = 2.04 \cdot 10^{+3}$
		$S = 0.926$	$p = 4.69 \cdot 10^{-1}$	$S = 0.865$	$p = 3.39 \cdot 10^{-1}$	$S = 0.824$	$p = 2.78 \cdot 10^{-1}$
	σ	$\sim$	$U = 2.05 \cdot 10^{+3}$	$\sim$	$U = 1.93 \cdot 10^{+3}$	$\sim$	$U = 1.72 \cdot 10^{+3}$
		$S = 0.902$	$p = 2.44 \cdot 10^{-1}$	$S = 0.515$	$p = 5.76 \cdot 10^{-1}$	$S = 0.924$	$p = 5.87 \cdot 10^{-1}$
Long. HDM	μ	$\sim$	$U = 1.97 \cdot 10^{+3}$	$\sim$	$U = 1.54 \cdot 10^{+3}$	$\sim$	$U = 1.71 \cdot 10^{+3}$
		$S = 0.853$	$p = 4.54 \cdot 10^{-1}$	$S = 0.838$	$p = 1.42 \cdot 10^{-1}$	$S = 0.905$	$p = 5.59 \cdot 10^{-1}$
	σ	$\sim$	$U = 1.70 \cdot 10^{+3}$	$\sim$	$U = 1.90 \cdot 10^{+3}$	$\sim$	$U = 1.82 \cdot 10^{+3}$
		$S = 0.913$	$p = 5.32 \cdot 10^{-1}$	$S = 0.337$	$p = 7.20 \cdot 10^{-1}$	$S = 0.941$	$p = 9.98 \cdot 10^{-1}$
Transv. HDM	μ	$\sim$	$U = 1.45 \cdot 10^{+3}$	$\sim$	$U = 1.74 \cdot 10^{+3}$	$\sim$	$U = 2.02 \cdot 10^{+3}$
		$S = 0.839$	$p = 5.38 \cdot 10^{-2}$	$S = 0.921$	$p = 6.45 \cdot 10^{-1}$	$S = 0.751$	$p = 3.17 \cdot 10^{-1}$
	σ	$>$	$U = 2.25 \cdot 10^{+3}$	$\sim$	$U = 1.86 \cdot 10^{+3}$	$\sim$	$U = 2.04 \cdot 10^{+3}$
		$S = 0.832$	$p = \mathbf{1.46 \cdot 10^{-2}}$	$S = 0.679$	$p = 8.67 \cdot 10^{-1}$	$S = 0.919$	$p = 2.78 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$\sim$	$U = 2.09 \cdot 10^{+3}$	$\sim$	$U = 1.69 \cdot 10^{+3}$	$<^*$	$U = 1.42 \cdot 10^{+3}$
		$S = 0.880$	$p = 1.68 \cdot 10^{-1}$	$S = 0.883$	$p = 4.79 \cdot 10^{-1}$	$S = 0.651$	$p = \mathbf{1.86 \cdot 10^{-2}}$
	σ	$\sim$	$U = 2.06 \cdot 10^{+3}$	$\sim$	$U = 1.68 \cdot 10^{+3}$	$\sim^*$	$U = 1.50 \cdot 10^{+3}$
		$S = 0.895$	$p = 2.37 \cdot 10^{-1}$	$S = 0.944$	$p = 4.57 \cdot 10^{-1}$	$S = 0.591$	$p = 8.98 \cdot 10^{-2}$
Rot. diffusion coef.	μ	$\sim$	$U = 1.49 \cdot 10^{+3}$	$<$	$U = 1.41 \cdot 10^{+3}$	$<^*$	$U = 8.99 \cdot 10^{+2}$
		$S = 0.799$	$p = 8.13 \cdot 10^{-2}$	$S = 0.785$	$p = \mathbf{1.58 \cdot 10^{-2}}$	$S = 0.521$	$p = \mathbf{9.42 \cdot 10^{-7}}$
	σ	$\sim$	$U = 1.70 \cdot 10^{+3}$	$\sim$	$U = 1.50 \cdot 10^{+3}$	$<^*$	$U = 1.06 \cdot 10^{+3}$
		$S = 0.760$	$p = 5.35 \cdot 10^{-1}$	$S = 0.761$	$p = 9.47 \cdot 10^{-2}$	$S = 0.600$	$p = \mathbf{3.96 \cdot 10^{-5}}$

Table D.10: Differential behavior of peptides on the interaction with lipid membranes, based on their activity, for each force field. Results of the Mann-Whitney U test. Continued from the previous page. The samples are grouped by force field. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Each test compares the distribution of the samples of the differences in MD descriptors between the simulations with POPC and with POPE:POPG associated with active peptides against that of nonactives. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for active sequences is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.11).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		Martini 2.2		Martini 2.2P		Martini 3	
		Stat.	p-value	Stat.	p-value	Stat.	p-value
Bilayer thickness	μ	$5.89 \cdot 10^{-2}$	$8.09 \cdot 10^{-1}$	$3.41 \cdot 10^{-1}$	$5.60 \cdot 10^{-1}$	$4.44 \cdot 10^{-1}$	$5.06 \cdot 10^{-1}$
	σ	$1.01 \cdot 10^{+0}$	$3.17 \cdot 10^{-1}$	$4.03 \cdot 10^{-3}$	$9.50 \cdot 10^{-1}$	$1.11 \cdot 10^{+0}$	$2.94 \cdot 10^{-1}$
Area per lipid	μ	$1.35 \cdot 10^{+0}$	$2.47 \cdot 10^{-1}$	$1.34 \cdot 10^{-1}$	$7.15 \cdot 10^{-1}$	$5.04 \cdot 10^{-1}$	$4.79 \cdot 10^{-1}$
	σ	$1.11 \cdot 10^{-2}$	$9.16 \cdot 10^{-1}$	$1.68 \cdot 10^{-1}$	$6.83 \cdot 10^{-1}$	$3.39 \cdot 10^{-4}$	$9.85 \cdot 10^{-1}$
Peptide penetration	μ	$4.69 \cdot 10^{-1}$	$4.95 \cdot 10^{-1}$	$2.47 \cdot 10^{-3}$	$9.60 \cdot 10^{-1}$	$4.92 \cdot 10^{+0}$	$\mathbf{2.85 \cdot 10^{-2}}$
	σ	$5.04 \cdot 10^{-1}$	$4.79 \cdot 10^{-1}$	$1.53 \cdot 10^{-1}$	$6.97 \cdot 10^{-1}$	$1.63 \cdot 10^{+1}$	$\mathbf{9.35 \cdot 10^{-5}}$
Tilt	μ	$2.45 \cdot 10^{+0}$	$1.20 \cdot 10^{-1}$	$2.93 \cdot 10^{+0}$	$8.96 \cdot 10^{-2}$	$9.19 \cdot 10^{-1}$	$3.40 \cdot 10^{-1}$
	σ	$3.94 \cdot 10^{-1}$	$5.31 \cdot 10^{-1}$	$1.06 \cdot 10^{-2}$	$9.18 \cdot 10^{-1}$	$4.16 \cdot 10^{+0}$	$\mathbf{4.37 \cdot 10^{-2}}$
Spin	μ	$2.55 \cdot 10^{-1}$	$6.14 \cdot 10^{-1}$	$5.01 \cdot 10^{-1}$	$4.81 \cdot 10^{-1}$	$8.63 \cdot 10^{+0}$	$\mathbf{3.96 \cdot 10^{-3}}$
	σ	$3.87 \cdot 10^{-1}$	$5.35 \cdot 10^{-1}$	$1.62 \cdot 10^{-3}$	$9.68 \cdot 10^{-1}$	$5.65 \cdot 10^{+0}$	$\mathbf{1.91 \cdot 10^{-2}}$
Contacts peptide-HGs	μ	$1.41 \cdot 10^{-1}$	$7.08 \cdot 10^{-1}$	$8.20 \cdot 10^{-4}$	$9.77 \cdot 10^{-1}$	$2.45 \cdot 10^{+0}$	$1.20 \cdot 10^{-1}$
	σ	$1.11 \cdot 10^{+0}$	$2.94 \cdot 10^{-1}$	$1.29 \cdot 10^{+0}$	$2.58 \cdot 10^{-1}$	$1.35 \cdot 10^{+1}$	$\mathbf{3.58 \cdot 10^{-4}}$
Contacts peptide-tails	μ	$6.61 \cdot 10^{-2}$	$7.97 \cdot 10^{-1}$	$1.48 \cdot 10^{-2}$	$9.03 \cdot 10^{-1}$	$1.41 \cdot 10^{+0}$	$2.38 \cdot 10^{-1}$
	σ	$2.62 \cdot 10^{-2}$	$8.72 \cdot 10^{-1}$	$9.54 \cdot 10^{-1}$	$3.31 \cdot 10^{-1}$	$1.48 \cdot 10^{+1}$	$\mathbf{1.95 \cdot 10^{-4}}$
Contacts peptide-water	μ	$4.21 \cdot 10^{-2}$	$8.38 \cdot 10^{-1}$	$8.30 \cdot 10^{-3}$	$9.28 \cdot 10^{-1}$	$9.12 \cdot 10^{-1}$	$3.41 \cdot 10^{-1}$
	σ	$4.87 \cdot 10^{+0}$	$\mathbf{2.92 \cdot 10^{-2}}$	$7.15 \cdot 10^{+0}$	$\mathbf{8.54 \cdot 10^{-3}}$	$1.25 \cdot 10^{+1}$	$\mathbf{5.87 \cdot 10^{-4}}$
EDM	μ	$6.12 \cdot 10^{-1}$	$4.36 \cdot 10^{-1}$	$4.94 \cdot 10^{-2}$	$8.24 \cdot 10^{-1}$	$1.45 \cdot 10^{+0}$	$2.31 \cdot 10^{-1}$
	σ	$7.24 \cdot 10^{-1}$	$3.97 \cdot 10^{-1}$	$6.92 \cdot 10^{-1}$	$4.07 \cdot 10^{-1}$	$2.06 \cdot 10^{+0}$	$1.53 \cdot 10^{-1}$
Long. EDM	μ	$5.64 \cdot 10^{-1}$	$4.54 \cdot 10^{-1}$	$2.24 \cdot 10^{-1}$	$6.37 \cdot 10^{-1}$	$1.45 \cdot 10^{+0}$	$2.30 \cdot 10^{-1}$
	σ	$7.64 \cdot 10^{-1}$	$3.84 \cdot 10^{-1}$	$6.87 \cdot 10^{-1}$	$4.09 \cdot 10^{-1}$	$1.98 \cdot 10^{+0}$	$1.62 \cdot 10^{-1}$
Transv. EDM	μ	$7.17 \cdot 10^{-1}$	$3.99 \cdot 10^{-1}$	$1.04 \cdot 10^{+0}$	$3.10 \cdot 10^{-1}$	$1.34 \cdot 10^{+0}$	$2.49 \cdot 10^{-1}$
	σ	$6.36 \cdot 10^{-1}$	$4.27 \cdot 10^{-1}$	$1.49 \cdot 10^{+0}$	$2.25 \cdot 10^{-1}$	$2.26 \cdot 10^{+0}$	$1.36 \cdot 10^{-1}$
HDM	μ	$2.14 \cdot 10^{-1}$	$6.44 \cdot 10^{-1}$	$3.39 \cdot 10^{-3}$	$9.54 \cdot 10^{-1}$	$1.23 \cdot 10^{+0}$	$2.69 \cdot 10^{-1}$
	σ	$2.36 \cdot 10^{-2}$	$8.78 \cdot 10^{-1}$	$9.78 \cdot 10^{-1}$	$3.25 \cdot 10^{-1}$	$6.48 \cdot 10^{-1}$	$4.22 \cdot 10^{-1}$
Long. HDM	μ	$2.57 \cdot 10^{-1}$	$6.13 \cdot 10^{-1}$	$4.74 \cdot 10^{-1}$	$4.92 \cdot 10^{-1}$	$7.06 \cdot 10^{-2}$	$7.91 \cdot 10^{-1}$
	σ	$1.66 \cdot 10^{-1}$	$6.84 \cdot 10^{-1}$	$7.82 \cdot 10^{-1}$	$3.78 \cdot 10^{-1}$	$1.06 \cdot 10^{+0}$	$3.04 \cdot 10^{-1}$
Transv. HDM	μ	$7.43 \cdot 10^{-1}$	$3.90 \cdot 10^{-1}$	$1.59 \cdot 10^{-2}$	$9.00 \cdot 10^{-1}$	$2.22 \cdot 10^{+0}$	$1.39 \cdot 10^{-1}$
	σ	$5.60 \cdot 10^{-1}$	$4.56 \cdot 10^{-1}$	$1.43 \cdot 10^{+0}$	$2.34 \cdot 10^{-1}$	$8.26 \cdot 10^{-2}$	$7.74 \cdot 10^{-1}$
Tr. diffusion coef.	μ	$1.44 \cdot 10^{-1}$	$7.05 \cdot 10^{-1}$	$1.25 \cdot 10^{-1}$	$7.24 \cdot 10^{-1}$	$9.47 \cdot 10^{+0}$	$\mathbf{2.58 \cdot 10^{-3}}$
	σ	$3.59 \cdot 10^{-1}$	$5.50 \cdot 10^{-1}$	$2.29 \cdot 10^{-2}$	$8.80 \cdot 10^{-1}$	$1.47 \cdot 10^{+1}$	$\mathbf{2.06 \cdot 10^{-4}}$
Rot. diffusion coef.	μ	$1.13 \cdot 10^{-3}$	$9.73 \cdot 10^{-1}$	$3.78 \cdot 10^{-1}$	$5.40 \cdot 10^{-1}$	$1.06 \cdot 10^{+1}$	$\mathbf{1.43 \cdot 10^{-3}}$
	σ	$2.18 \cdot 10^{-2}$	$8.83 \cdot 10^{-1}$	$4.60 \cdot 10^{-1}$	$4.99 \cdot 10^{-1}$	$9.75 \cdot 10^{+0}$	$\mathbf{2.24 \cdot 10^{-3}}$

Table D.11: Difference in the variances of the samples of differences between lipid compositions based on activity for each force field. Results of the Brown-Forsythe test test. The samples, computed as the differences in descriptors in simulations of POPC and POPE:POPG, are grouped by force field. Lines with gray background and the symbol μ refer to the samples of the mean value of the property; the σ symbol corresponds to those of their standard deviation. Each test compares the variance of the samples of differences in MD descriptors associated with active and nonactive peptides. The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in the variances.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		POPC				POPE:POPG 3:1			
		Active		Nonactive		Active		Nonactive	
Bilayer thickness	M2,M2P	$<$ $S = 0.355$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.364$	$1.00 \cdot 10^{+0}$ $\mathbf{1.78 \cdot 10^{-15}}$	$>$ $S = 0.858$	$1.89 \cdot 10^{+3}$ $\mathbf{1.40 \cdot 10^{-3}}$	$>$ $S = 0.916$	$8.83 \cdot 10^{+2}$ $\mathbf{8.55 \cdot 10^{-3}}$
	M2,M3	$\sim$ $S = 0.875$	$1.01 \cdot 10^{+3}$ $6.35 \cdot 10^{-2}$	$<$ $S = 0.632$	$1.33 \cdot 10^{+2}$ $\mathbf{6.07 \cdot 10^{-8}}$	$<$ $S = 0.847$	$8.84 \cdot 10^{+2}$ $\mathbf{5.16 \cdot 10^{-3}}$	$<$ $S = 0.758$	$2.33 \cdot 10^{+2}$ $\mathbf{2.32 \cdot 10^{-5}}$
	M2P,M3	$>$ $S = 0.406$	$2.58 \cdot 10^{+3}$ $\mathbf{8.38 \cdot 10^{-12}}$	$>$ $S = 0.681$	$1.12 \cdot 10^{+3}$ $\mathbf{2.13 \cdot 10^{-7}}$	$<$ $S = 0.796$	$4.80 \cdot 10^{+2}$ $\mathbf{8.52 \cdot 10^{-7}}$	$<$ $S = 0.726$	$1.30 \cdot 10^{+2}$ $\mathbf{4.87 \cdot 10^{-8}}$
Area per lipid	M2,M2P	$<$ $S = 0.837$	$9.30 \cdot 10^{+2}$ $\mathbf{1.04 \cdot 10^{-2}}$	$<$ $S = 0.749$	$3.64 \cdot 10^{+2}$ $\mathbf{3.80 \cdot 10^{-3}}$	$<$ $S = 0.671$	$5.30 \cdot 10^{+2}$ $\mathbf{3.23 \cdot 10^{-6}}$	$<$ $S = 0.551$	$9.80 \cdot 10^{+1}$ $\mathbf{3.77 \cdot 10^{-9}}$
	M2,M3	$<$ $S = 0.283$	$1.60 \cdot 10^{+1}$ $\mathbf{1.10 \cdot 10^{-13}}$	$<$ $S = 0.259$	$9.00 \cdot 10^{+0}$ $\mathbf{2.93 \cdot 10^{-14}}$	$<$ $S = 0.170$	$3.00 \cdot 10^{+0}$ $\mathbf{6.41 \cdot 10^{-14}}$	$<$ $S = 0.118$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$ $S = 0.405$	$9.10 \cdot 10^{+1}$ $\mathbf{2.19 \cdot 10^{-12}}$	$<$ $S = 0.424$	$6.40 \cdot 10^{+1}$ $\mathbf{1.41 \cdot 10^{-10}}$	$<$ $S = 0.356$	$3.40 \cdot 10^{+1}$ $\mathbf{2.28 \cdot 10^{-13}}$	$<$ $S = 0.311$	$1.30 \cdot 10^{+1}$ $\mathbf{7.82 \cdot 10^{-14}}$
	M2,M2P	$>$ $S = 0.380$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.401$	$1.27 \cdot 10^{+3}$ $\mathbf{1.78 \cdot 10^{-15}}$	$<$ $S = 0.774$	$5.03 \cdot 10^{+2}$ $\mathbf{1.59 \cdot 10^{-6}}$	$<$ $S = 0.763$	$3.78 \cdot 10^{+2}$ $\mathbf{5.76 \cdot 10^{-3}}$
	M2,M3	$>$ $S = 0.315$	$2.70 \cdot 10^{+3}$ $\mathbf{6.68 \cdot 10^{-14}}$	$>$ $S = 0.351$	$1.27 \cdot 10^{+3}$ $\mathbf{6.22 \cdot 10^{-15}}$	$>$ $S = 0.170$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.157$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$\sim$ $S = 0.792$	$1.02 \cdot 10^{+3}$ $7.35 \cdot 10^{-2}$	$\sim$ $S = 0.805$	$6.19 \cdot 10^{+2}$ $8.63 \cdot 10^{-1}$	$>$ $S = 0.133$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.237$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$
	M2,M2P	$<$ $S = 0.599$	$3.68 \cdot 10^{+2}$ $\mathbf{3.31 \cdot 10^{-8}}$	$<$ $S = 0.590$	$1.46 \cdot 10^{+2}$ $\mathbf{1.52 \cdot 10^{-7}}$	$<$ $S = 0.154$	$1.00 \cdot 10^{+1}$ $\mathbf{8.57 \cdot 10^{-14}}$	$<$ $S = 0.109$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2,M3	$<$ $S = 0.191$	$6.00 \cdot 10^{+0}$ $\mathbf{7.26 \cdot 10^{-14}}$	$<$ $S = 0.202$	$8.00 \cdot 10^{+0}$ $\mathbf{2.22 \cdot 10^{-14}}$	$<$ $S = 0.054$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.110$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$ $S = 0.446$	$1.37 \cdot 10^{+2}$ $\mathbf{1.27 \cdot 10^{-11}}$	$<$ $S = 0.543$	$1.15 \cdot 10^{+2}$ $\mathbf{1.55 \cdot 10^{-8}}$	$<$ $S = 0.697$	$4.38 \cdot 10^{+2}$ $\mathbf{2.63 \cdot 10^{-7}}$	$<$ $S = 0.807$	$4.13 \cdot 10^{+2}$ $\mathbf{1.49 \cdot 10^{-2}}$
Peptide penetration	M2,M2P	$>$ $S = 0.586$	$2.61 \cdot 10^{+3}$ $\mathbf{2.11 \cdot 10^{-12}}$	$>$ $S = 0.453$	$1.15 \cdot 10^{+3}$ $\mathbf{3.36 \cdot 10^{-8}}$	$<$ $S = 0.752$	$1.18 \cdot 10^{+2}$ $\mathbf{6.19 \cdot 10^{-12}}$	$<$ $S = 0.573$	$3.20 \cdot 10^{+2}$ $\mathbf{8.89 \cdot 10^{-4}}$
	M2,M3	$>$ $S = 0.527$	$2.70 \cdot 10^{+3}$ $\mathbf{6.15 \cdot 10^{-14}}$	$>$ $S = 0.283$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$	$>$ $S = 0.464$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.418$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$>$ $S = 0.727$	$2.22 \cdot 10^{+3}$ $\mathbf{9.25 \cdot 10^{-7}}$	$>$ $S = 0.640$	$1.08 \cdot 10^{+3}$ $\mathbf{2.68 \cdot 10^{-6}}$	$>$ $S = 0.362$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.592$	$1.16 \cdot 10^{+3}$ $\mathbf{1.96 \cdot 10^{-8}}$
	M2,M2P	$\sim$ $S = 0.585$	$1.31 \cdot 10^{+3}$ $8.20 \cdot 10^{-1}$	$<$ $S = 0.243$	$2.81 \cdot 10^{+2}$ $\mathbf{1.99 \cdot 10^{-4}}$	$\sim$ $S = 0.842$	$1.29 \cdot 10^{+3}$ $7.56 \cdot 10^{-1}$	$\sim$ $S = 0.346$	$6.33 \cdot 10^{+2}$ $9.69 \cdot 10^{-1}$
	M2,M3	$<$ $S = 0.448$	$1.00 \cdot 10^{+0}$ $\mathbf{5.90 \cdot 10^{-14}}$	$<$ $S = 0.079$	$4.00 \cdot 10^{+0}$ $\mathbf{6.22 \cdot 10^{-15}}$	$<$ $S = 0.355$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.155$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$ $S = 0.750$	$3.36 \cdot 10^{+2}$ $\mathbf{1.22 \cdot 10^{-8}}$	$<$ $S = 0.493$	$6.30 \cdot 10^{+1}$ $\mathbf{1.26 \cdot 10^{-10}}$	$<$ $S = 0.457$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.567$	$1.04 \cdot 10^{+2}$ $\mathbf{6.30 \cdot 10^{-9}}$
Tilt	M2,M2P	$<$ $S = 0.901$	$8.62 \cdot 10^{+2}$ $\mathbf{3.62 \cdot 10^{-3}}$	$\sim$ $S = 0.920$	$4.95 \cdot 10^{+2}$ $1.72 \cdot 10^{-1}$	$<$ $S = 0.895$	$9.33 \cdot 10^{+2}$ $\mathbf{1.09 \cdot 10^{-2}}$	$\sim$ $S = 0.893$	$4.48 \cdot 10^{+2}$ $6.78 \cdot 10^{-2}$
	M2,M3	$>$ $S = 0.765$	$2.04 \cdot 10^{+3}$ $\mathbf{7.85 \cdot 10^{-5}}$	$>$ $S = 0.901$	$8.74 \cdot 10^{+2}$ $\mathbf{1.09 \cdot 10^{-2}}$	$>$ $S = 0.800$	$2.36 \cdot 10^{+3}$ $\mathbf{1.67 \cdot 10^{-8}}$	$>$ $S = 0.788$	$1.09 \cdot 10^{+3}$ $\mathbf{2.13 \cdot 10^{-6}}$
	M2P,M3	$>$ $S = 0.691$	$2.25 \cdot 10^{+3}$ $\mathbf{4.26 \cdot 10^{-7}}$	$\sim$ $S = 0.880$	$4.92 \cdot 10^{+2}$ $1.63 \cdot 10^{-1}$	$>$ $S = 0.780$	$2.40 \cdot 10^{+3}$ $\mathbf{4.24 \cdot 10^{-9}}$	$>$ $S = 0.864$	$8.98 \cdot 10^{+2}$ $\mathbf{5.60 \cdot 10^{-3}}$
	M2,M2P	$\sim$ $S = 0.626$	$1.20 \cdot 10^{+3}$ $3.93 \cdot 10^{-1}$	$<$ $S = 0.553$	$3.86 \cdot 10^{+2}$ $\mathbf{7.24 \cdot 10^{-3}}$	$>$ $S = 0.922$	$1.82 \cdot 10^{+3}$ $\mathbf{5.33 \cdot 10^{-3}}$	$\sim$ $S = 0.706$	$5.90 \cdot 10^{+2}$ $6.53 \cdot 10^{-1}$
	M2,M3	$<$ $S = 0.616$	$4.70 \cdot 10^{+1}$ $\mathbf{3.86 \cdot 10^{-13}}$	$<$ $S = 0.291$	$1.00 \cdot 10^{+0}$ $\mathbf{1.78 \cdot 10^{-15}}$	$<$ $S = 0.709$	$6.00 \cdot 10^{+0}$ $\mathbf{7.26 \cdot 10^{-14}}$	$<$ $S = 0.469$	$6.00 \cdot 10^{+0}$ $\mathbf{1.24 \cdot 10^{-14}}$
	M2P,M3	$<$ $S = 0.802$	$4.91 \cdot 10^{+2}$ $\mathbf{1.15 \cdot 10^{-6}}$	$<$ $S = 0.622$	$1.36 \cdot 10^{+2}$ $\mathbf{7.54 \cdot 10^{-8}}$	$<$ $S = 0.705$	$1.00 \cdot 10^{+0}$ $\mathbf{5.90 \cdot 10^{-14}}$	$<$ $S = 0.710$	$9.80 \cdot 10^{+1}$ $\mathbf{3.77 \cdot 10^{-9}}$
Spin	M2,M2P	$\sim$ $S = 0.911$	$1.26 \cdot 10^{+3}$ $6.07 \cdot 10^{-1}$	$\sim$ $S = 0.896$	$5.01 \cdot 10^{+2}$ $1.91 \cdot 10^{-1}$	$\sim$ $S = 0.959$	$1.17 \cdot 10^{+3}$ $3.26 \cdot 10^{-1}$	$\sim$ $S = 0.862$	$5.70 \cdot 10^{+2}$ $5.21 \cdot 10^{-1}$
	M2,M3	$\sim$ $S = 0.838$	$1.06 \cdot 10^{+3}$ $1.10 \cdot 10^{-1}$	$\sim$ $S = 0.857$	$6.35 \cdot 10^{+2}$ $9.85 \cdot 10^{-1}$	$\sim$ $S = 0.899$	$1.12 \cdot 10^{+3}$ $2.07 \cdot 10^{-1}$	$\sim$ $S = 0.896$	$5.19 \cdot 10^{+2}$ $2.57 \cdot 10^{-1}$
	M2P,M3	$>$ $S = 0.796$	$1.96 \cdot 10^{+3}$ $\mathbf{4.11 \cdot 10^{-4}}$	$\sim$ $S = 0.888$	$5.98 \cdot 10^{+2}$ $7.09 \cdot 10^{-1}$	$\sim$ $S = 0.931$	$1.01 \cdot 10^{+3}$ $6.35 \cdot 10^{-2}$	$\sim$ $S = 0.930$	$4.45 \cdot 10^{+2}$ $6.35 \cdot 10^{-2}$
	M2,M2P	$\sim$ $S = 0.653$	$1.08 \cdot 10^{+3}$ $1.37 \cdot 10^{-1}$	$<$ $S = 0.558$	$3.56 \cdot 10^{+2}$ $\mathbf{2.97 \cdot 10^{-3}}$	$\sim$ $S = 0.860$	$1.11 \cdot 10^{+3}$ $1.88 \cdot 10^{-1}$	$\sim$ $S = 0.718$	$4.51 \cdot 10^{+2}$ $7.24 \cdot 10^{-2}$
	M2,M3	$<$ $S = 0.682$	$1.54 \cdot 10^{+2}$ $\mathbf{2.39 \cdot 10^{-11}}$	$<$ $S = 0.365$	$1.10 \cdot 10^{+1}$ $\mathbf{4.88 \cdot 10^{-14}}$	$<$ $S = 0.639$	$1.08 \cdot 10^{+2}$ $\mathbf{4.22 \cdot 10^{-12}}$	$<$ $S = 0.543$	$2.80 \cdot 10^{+1}$ $\mathbf{1.32 \cdot 10^{-12}}$
	M2P,M3	$<$ $S = 0.864$	$5.22 \cdot 10^{+2}$ $\mathbf{2.62 \cdot 10^{-6}}$	$<$ $S = 0.711$	$1.79 \cdot 10^{+2}$ $\mathbf{1.26 \cdot 10^{-6}}$	$<$ $S = 0.688$	$8.60 \cdot 10^{+1}$ $\mathbf{1.81 \cdot 10^{-12}}$	$<$ $S = 0.774$	$1.96 \cdot 10^{+2}$ $\mathbf{3.36 \cdot 10^{-6}}$
Contacts peptide-HGs	M2,M2P	$>$ $S = 0.781$	$1.90 \cdot 10^{+3}$ $\mathbf{5.35 \cdot 10^{-4}}$	$>$ $S = 0.761$	$9.10 \cdot 10^{+2}$ $\mathbf{3.92 \cdot 10^{-3}}$	$>$ $S = 0.954$	$1.79 \cdot 10^{+3}$ $\mathbf{7.38 \cdot 10^{-3}}$	$>$ $S = 0.865$	$8.98 \cdot 10^{+2}$ $\mathbf{5.60 \cdot 10^{-3}}$
	M2,M3	$>$ $S = 0.568$	$2.70 \cdot 10^{+3}$ $\mathbf{5.90 \cdot 10^{-14}}$	$>$ $S = 0.391$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$	$>$ $S = 0.541$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.503$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$>$ $S = 0.699$	$2.29 \cdot 10^{+3}$ $\mathbf{1.10 \cdot 10^{-7}}$	$>$ $S = 0.590$	$1.17 \cdot 10^{+3}$ $\mathbf{5.32 \cdot 10^{-9}}$	$>$ $S = 0.561$	$2.70 \cdot 10^{+3}$ $\mathbf{5.66 \cdot 10^{-14}}$	$>$ $S = 0.590$	$1.16 \cdot 10^{+3}$ $\mathbf{1.81 \cdot 10^{-8}}$
	M2,M2P	$\sim$ $S = 0.863$	$1.26 \cdot 10^{+3}$ $6.34 \cdot 10^{-1}$	$\sim$ $S = 0.851$	$6.09 \cdot 10^{+2}$ $7.89 \cdot 10^{-1}$	$>$ $S = 0.944$	$2.26 \cdot 10^{+3}$ $\mathbf{2.79 \cdot 10^{-7}}$	$\sim$ $S = 0.943$	$4.91 \cdot 10^{+2}$ $1.60 \cdot 10^{-1}$
	M2,M3	$<$ $S = 0.825$	$3.83 \cdot 10^{+2}$ $\mathbf{5.22 \cdot 10^{-8}}$	$<$ $S = 0.536$	$1.35 \cdot 10^{+2}$ $\mathbf{7.02 \cdot 10^{-8}}$	$<$ $S = 0.811$	$2.00 \cdot 10^{+1}$ $\mathbf{1.29 \cdot 10^{-13}}$	$<$ $S = 0.677$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$ $S = 0.839$	$3.36 \cdot 10^{+2}$ $\mathbf{1.22 \cdot 10^{-8}}$	$<$ $S = 0.622$	$1.05 \cdot 10^{+2}$ $\mathbf{6.85 \cdot 10^{-9}}$	$<$ $S = 0.770$	$1.20 \cdot 10^{+1}$ $\mathbf{9.30 \cdot 10^{-14}}$	$<$ $S = 0.676$	$5.00 \cdot 10^{+0}$ $\mathbf{8.88 \cdot 10^{-15}}$

Table D.12: Differences in the values of MD descriptors based on force field for each combination of antimicrobial activity and lipid composition. Results of the Wilcoxon signed-rank test. Continues in the next page.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		POPC				POPE:POPG 3:1			
		Active		Nonactive		Active		Nonactive	
Contacts peptide-tails	M2,M2P	$>$ $S = 0.846$	$2.40 \cdot 10^{+3}$ $\mathbf{4.31 \cdot 10^{-9}}$	$>$ $S = 0.823$	$1.07 \cdot 10^{+3}$ $\mathbf{5.82 \cdot 10^{-6}}$	$<$ $S = 0.920$	$8.22 \cdot 10^{+2}$ $\mathbf{1.82 \cdot 10^{-3}}$	$\sim$ $S = 0.854$	$5.82 \cdot 10^{+2}$ $5.98 \cdot 10^{-1}$
	M2,M3	$>$ $S = 0.615$	$2.63 \cdot 10^{+3}$ $\mathbf{8.65 \cdot 10^{-14}}$	$>$ $S = 0.435$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$	$>$ $S = 0.723$	$2.69 \cdot 10^{+3}$ $\mathbf{8.22 \cdot 10^{-14}}$	$>$ $S = 0.587$	$1.27 \cdot 10^{+3}$ $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$>$ $S = 0.715$	$2.29 \cdot 10^{+3}$ $\mathbf{1.28 \cdot 10^{-7}}$	$>$ $S = 0.591$	$1.20 \cdot 10^{+3}$ $\mathbf{7.25 \cdot 10^{-10}}$	$>$ $S = 0.671$	$2.69 \cdot 10^{+3}$ $\mathbf{9.30 \cdot 10^{-14}}$	$>$ $S = 0.657$	$1.19 \cdot 10^{+3}$ $\mathbf{1.17 \cdot 10^{-9}}$
Contacts peptide-water	M2,M2P	$\sim$ $S = 0.855$	$1.16 \cdot 10^{+3}$ $3.03 \cdot 10^{-1}$	$\sim$ $S = 0.818$	$5.02 \cdot 10^{+2}$ $1.94 \cdot 10^{-1}$	$\sim$ $S = 0.958$	$1.09 \cdot 10^{+3}$ $1.46 \cdot 10^{-1}$	$<$ $S = 0.870$	$4.15 \cdot 10^{+2}$ $\mathbf{1.56 \cdot 10^{-2}}$
	M2,M3	$<$ $S = 0.791$	$5.88 \cdot 10^{+2}$ $\mathbf{2.31 \cdot 10^{-5}}$	$<$ $S = 0.596$	$4.04 \cdot 10^{+2}$ $\mathbf{1.18 \cdot 10^{-2}}$	$<$ $S = 0.851$	$2.00 \cdot 10^{+2}$ $\mathbf{1.27 \cdot 10^{-10}}$	$<$ $S = 0.746$	$2.21 \cdot 10^{+2}$ $\mathbf{1.28 \cdot 10^{-5}}$
	M2P,M3	$<$ $S = 0.861$	$3.71 \cdot 10^{+2}$ $\mathbf{3.63 \cdot 10^{-8}}$	$<$ $S = 0.745$	$3.58 \cdot 10^{+2}$ $\mathbf{3.16 \cdot 10^{-3}}$	$<$ $S = 0.872$	$3.39 \cdot 10^{+2}$ $\mathbf{1.34 \cdot 10^{-8}}$	$<$ $S = 0.837$	$2.11 \cdot 10^{+2}$ $\mathbf{7.60 \cdot 10^{-6}}$
EDM	M2,M2P	$<$ $S = 0.586$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.536$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$	$<$ $S = 0.693$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.608$	$3.00 \cdot 10^{+0}$ $\mathbf{4.44 \cdot 10^{-15}}$
	M2,M3	$<$ $S = 0.383$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.260$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$	$<$ $S = 0.309$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.283$	0.00 $\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$ $S = 0.718$	$5.52 \cdot 10^{+2}$ $\mathbf{5.67 \cdot 10^{-6}}$	$<$ $S = 0.629$	$1.77 \cdot 10^{+2}$ $\mathbf{1.12 \cdot 10^{-6}}$	$<$ $S = 0.594$	$7.00 \cdot 10^{+0}$ $\mathbf{7.57 \cdot 10^{-14}}$	$<$ $S = 0.596$	$1.22 \cdot 10^{+2}$ $\mathbf{2.67 \cdot 10^{-8}}$
	M2,M2P	$<$ $S = 0.784$	$1.03 \cdot 10^{+2}$ $\mathbf{3.49 \cdot 10^{-12}}$	$<$ $S = 0.778$	$1.06 \cdot 10^{+2}$ $\mathbf{7.45 \cdot 10^{-9}}$	$<$ $S = 0.814$	$2.80 \cdot 10^{+1}$ $\mathbf{1.79 \cdot 10^{-13}}$	$<$ $S = 0.818$	$6.80 \cdot 10^{+1}$ $\mathbf{2.16 \cdot 10^{-10}}$
	M2,M3	$<$ $S = 0.504$	$1.10 \cdot 10^{+1}$ $\mathbf{8.93 \cdot 10^{-14}}$	$<$ $S = 0.484$	$5.70 \cdot 10^{+1}$ $\mathbf{6.44 \cdot 10^{-11}}$	$<$ $S = 0.552$	0.00 $\mathbf{5.66 \cdot 10^{-14}}$	$<$ $S = 0.538$	$1.90 \cdot 10^{+1}$ $\mathbf{2.73 \cdot 10^{-13}}$
	M2P,M3	$<$ $S = 0.698$	$1.31 \cdot 10^{+2}$ $\mathbf{1.01 \cdot 10^{-11}}$	$<$ $S = 0.632$	$1.95 \cdot 10^{+2}$ $\mathbf{3.18 \cdot 10^{-6}}$	$<$ $S = 0.697$	$1.80 \cdot 10^{+1}$ $\mathbf{1.19 \cdot 10^{-13}}$	$<$ $S = 0.705$	$6.20 \cdot 10^{+1}$ $\mathbf{1.13 \cdot 10^{-10}}$
Long. EDM	M2,M2P	$<$ $S = 0.920$	$7.36 \cdot 10^{+2}$ $\mathbf{3.65 \cdot 10^{-4}}$	$\sim$ $S = 0.820$	$4.79 \cdot 10^{+2}$ $1.28 \cdot 10^{-1}$	$<$ $S = 0.878$	$5.09 \cdot 10^{+2}$ $\mathbf{1.86 \cdot 10^{-6}}$	$<$ $S = 0.924$	$4.21 \cdot 10^{+2}$ $\mathbf{1.81 \cdot 10^{-2}}$
	M2,M3	$>$ $S = 0.797$	$2.09 \cdot 10^{+3}$ $\mathbf{2.23 \cdot 10^{-5}}$	$>$ $S = 0.792$	$9.62 \cdot 10^{+2}$ $\mathbf{6.90 \cdot 10^{-4}}$	$>$ $S = 0.956$	$1.98 \cdot 10^{+3}$ $\mathbf{2.69 \cdot 10^{-4}}$	$>$ $S = 0.914$	$8.92 \cdot 10^{+2}$ $\mathbf{6.65 \cdot 10^{-3}}$
	M2P,M3	$>$ $S = 0.732$	$2.26 \cdot 10^{+3}$ $\mathbf{2.95 \cdot 10^{-7}}$	$>$ $S = 0.895$	$9.86 \cdot 10^{+2}$ $\mathbf{2.76 \cdot 10^{-4}}$	$>$ $S = 0.848$	$2.35 \cdot 10^{+3}$ $\mathbf{1.78 \cdot 10^{-8}}$	$>$ $S = 0.855$	$1.03 \cdot 10^{+3}$ $\mathbf{4.96 \cdot 10^{-5}}$
	M2,M2P	$<$ $S = 0.651$	$7.28 \cdot 10^{+2}$ $\mathbf{3.11 \cdot 10^{-4}}$	$<$ $S = 0.592$	$3.89 \cdot 10^{+2}$ $\mathbf{7.87 \cdot 10^{-3}}$	$<$ $S = 0.885$	$9.46 \cdot 10^{+2}$ $\mathbf{1.31 \cdot 10^{-2}}$	$<$ $S = 0.851$	$3.50 \cdot 10^{+2}$ $\mathbf{2.46 \cdot 10^{-3}}$
	M2,M3	$<$ $S = 0.592$	$7.30 \cdot 10^{+1}$ $\mathbf{1.08 \cdot 10^{-12}}$	$<$ $S = 0.564$	$3.00 \cdot 10^{+0}$ $\mathbf{4.44 \cdot 10^{-15}}$	$<$ $S = 0.641$	$1.00 \cdot 10^{+0}$ $\mathbf{5.90 \cdot 10^{-14}}$	$<$ $S = 0.632$	$1.00 \cdot 10^{+0}$ $\mathbf{1.78 \cdot 10^{-15}}$
	M2P,M3	$<$ $S = 0.745$	$2.17 \cdot 10^{+2}$ $\mathbf{2.31 \cdot 10^{-10}}$	$<$ $S = 0.869$	$1.76 \cdot 10^{+2}$ $\mathbf{1.05 \cdot 10^{-6}}$	$<$ $S = 0.724$	$1.08 \cdot 10^{+2}$ $\mathbf{4.22 \cdot 10^{-12}}$	$<$ $S = 0.748$	$1.05 \cdot 10^{+2}$ $\mathbf{6.85 \cdot 10^{-9}}$
Transv. EDM	M2,M2P	$<$ $S = 0.840$	$6.19 \cdot 10^{+2}$ $\mathbf{2.89 \cdot 10^{-5}}$	$<$ $S = 0.902$	$4.33 \cdot 10^{+2}$ $\mathbf{2.42 \cdot 10^{-2}}$	$<$ $S = 0.907$	$4.26 \cdot 10^{+2}$ $\mathbf{1.86 \cdot 10^{-7}}$	$<$ $S = 0.895$	$3.69 \cdot 10^{+2}$ $\mathbf{4.42 \cdot 10^{-3}}$
	M2,M3	$>$ $S = 0.905$	$2.00 \cdot 10^{+3}$ $\mathbf{1.71 \cdot 10^{-4}}$	$>$ $S = 0.900$	$9.79 \cdot 10^{+2}$ $\mathbf{3.63 \cdot 10^{-4}}$	$\sim$ $S = 0.949$	$1.06 \cdot 10^{+3}$ $1.14 \cdot 10^{-1}$	$>$ $S = 0.906$	$8.80 \cdot 10^{+2}$ $\mathbf{9.28 \cdot 10^{-3}}$
	M2P,M3	$>$ $S = 0.805$	$2.29 \cdot 10^{+3}$ $\mathbf{1.07 \cdot 10^{-7}}$	$>$ $S = 0.875$	$1.01 \cdot 10^{+3}$ $\mathbf{1.20 \cdot 10^{-4}}$	$>$ $S = 0.880$	$2.22 \cdot 10^{+3}$ $\mathbf{1.00 \cdot 10^{-6}}$	$>$ $S = 0.835$	$1.01 \cdot 10^{+3}$ $\mathbf{8.14 \cdot 10^{-5}}$
	M2,M2P	$<$ $S = 0.662$	$7.70 \cdot 10^{+2}$ $\mathbf{7.08 \cdot 10^{-4}}$	$<$ $S = 0.629$	$4.22 \cdot 10^{+2}$ $\mathbf{1.86 \cdot 10^{-2}}$	$<$ $S = 0.898$	$9.82 \cdot 10^{+2}$ $\mathbf{2.14 \cdot 10^{-2}}$	$<$ $S = 0.879$	$3.86 \cdot 10^{+2}$ $\mathbf{7.24 \cdot 10^{-3}}$
	M2,M3	$<$ $S = 0.655$	$7.30 \cdot 10^{+1}$ $\mathbf{1.08 \cdot 10^{-12}}$	$<$ $S = 0.610$	$1.40 \cdot 10^{+1}$ $\mathbf{9.77 \cdot 10^{-14}}$	$<$ $S = 0.695$	$1.00 \cdot 10^{+0}$ $\mathbf{5.90 \cdot 10^{-14}}$	$<$ $S = 0.707$	$1.50 \cdot 10^{+1}$ $\mathbf{1.22 \cdot 10^{-13}}$
	M2P,M3	$<$ $S = 0.776$	$2.30 \cdot 10^{+2}$ $\mathbf{3.64 \cdot 10^{-10}}$	$<$ $S = 0.885$	$1.94 \cdot 10^{+2}$ $\mathbf{3.01 \cdot 10^{-6}}$	$<$ $S = 0.761$	$1.63 \cdot 10^{+2}$ $\mathbf{3.32 \cdot 10^{-11}}$	$<$ $S = 0.780$	$1.53 \cdot 10^{+2}$ $\mathbf{2.44 \cdot 10^{-7}}$
	M2,M2P	$<$ $S = 0.732$	$9.01 \cdot 10^{+2}$ $\mathbf{6.73 \cdot 10^{-3}}$	$\sim$ $S = 0.796$	$6.01 \cdot 10^{+2}$ $7.30 \cdot 10^{-1}$	$<$ $S = 0.933$	$5.77 \cdot 10^{+2}$ $\mathbf{1.06 \cdot 10^{-5}}$	$\sim$ $S = 0.914$	$6.04 \cdot 10^{+2}$ $7.52 \cdot 10^{-1}$
	M2,M3	$\sim$ $S = 0.690$	$1.12 \cdot 10^{+3}$ $1.99 \cdot 10^{-1}$	$\sim$ $S = 0.764$	$6.28 \cdot 10^{+2}$ $9.31 \cdot 10^{-1}$	$\sim$ $S = 0.936$	$1.23 \cdot 10^{+3}$ $5.08 \cdot 10^{-1}$	$\sim$ $S = 0.923$	$5.43 \cdot 10^{+2}$ $3.67 \cdot 10^{-1}$
	M2P,M3	$\sim$ $S = 0.924$	$1.03 \cdot 10^{+3}$ $8.28 \cdot 10^{-2}$	$\sim$ $S = 0.942$	$6.08 \cdot 10^{+2}$ $7.81 \cdot 10^{-1}$	$>$ $S = 0.892$	$2.12 \cdot 10^{+3}$ $\mathbf{1.06 \cdot 10^{-5}}$	$\sim$ $S = 0.931$	$4.35 \cdot 10^{+2}$ $5.06 \cdot 10^{-2}$
	M2,M2P	$<$ $S = 0.529$	$9.30 \cdot 10^{+2}$ $\mathbf{1.04 \cdot 10^{-2}}$	$\sim$ $S = 0.605$	$4.68 \cdot 10^{+2}$ $1.03 \cdot 10^{-1}$	$\sim$ $S = 0.953$	$1.05 \cdot 10^{+3}$ $9.41 \cdot 10^{-2}$	$\sim$ $S = 0.959$	$5.41 \cdot 10^{+2}$ $3.57 \cdot 10^{-1}$
	M2,M3	$<$ $S = 0.598$	$7.40 \cdot 10^{+1}$ $\mathbf{1.13 \cdot 10^{-12}}$	$<$ $S = 0.562$	$1.60 \cdot 10^{+1}$ $\mathbf{1.50 \cdot 10^{-13}}$	$<$ $S = 0.705$	$3.00 \cdot 10^{+0}$ $\mathbf{6.41 \cdot 10^{-14}}$	$<$ $S = 0.705$	$2.00 \cdot 10^{+0}$ $\mathbf{2.66 \cdot 10^{-15}}$
	M2P,M3	$<$ $S = 0.765$	$1.76 \cdot 10^{+2}$ $\mathbf{5.34 \cdot 10^{-11}}$	$<$ $S = 0.879$	$1.65 \cdot 10^{+2}$ $\mathbf{5.32 \cdot 10^{-7}}$	$<$ $S = 0.736$	$1.48 \cdot 10^{+2}$ $\mathbf{1.91 \cdot 10^{-11}}$	$<$ $S = 0.728$	$9.60 \cdot 10^{+1}$ $\mathbf{3.17 \cdot 10^{-9}}$

Table D.12: Differences in the values of MD descriptors based on force field for each combination of antimicrobial activity and lipid composition. Results of the Wilcoxon signed-rank test. Continued from the previous page, continues in the next page.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude		POPC				POPE:POPG 3:1			
		Active		Nonactive		Active		Nonactive	
HDM	M2,M2P	$>$	$2.05 \cdot 10^{+3}$	$\sim$	$6.05 \cdot 10^{+2}$	$>$	$1.75 \cdot 10^{+3}$	$\sim$	$5.71 \cdot 10^{+2}$
		$S = 0.973$	$\mathbf{5.88 \cdot 10^{-5}}$	$S = 0.976$	$7.59 \cdot 10^{-1}$	$S = 0.982$	$\mathbf{1.44 \cdot 10^{-2}}$	$S = 0.970$	$5.27 \cdot 10^{-1}$
	M2,M3	$\sim$	$1.21 \cdot 10^{+3}$	$>$	$1.08 \cdot 10^{+3}$	$\sim$	$1.20 \cdot 10^{+3}$	$>$	$1.06 \cdot 10^{+3}$
		$S = 0.906$	$4.37 \cdot 10^{-1}$	$S = 0.811$	$\mathbf{2.54 \cdot 10^{-6}}$	$S = 0.909$	$4.05 \cdot 10^{-1}$	$S = 0.821$	$\mathbf{1.04 \cdot 10^{-5}}$
	M2P,M3	$\sim$	$1.03 \cdot 10^{+3}$	$>$	$1.05 \cdot 10^{+3}$	$\sim$	$1.12 \cdot 10^{+3}$	$>$	$1.03 \cdot 10^{+3}$
		$S = 0.896$	$7.90 \cdot 10^{-2}$	$S = 0.807$	$\mathbf{1.91 \cdot 10^{-5}}$	$S = 0.905$	$2.01 \cdot 10^{-1}$	$S = 0.816$	$\mathbf{4.96 \cdot 10^{-5}}$
	M2,M2P	$\sim$	$1.17 \cdot 10^{+3}$	$<$	$4.00 \cdot 10^{+2}$	$\sim$	$1.28 \cdot 10^{+3}$	$\sim$	$6.07 \cdot 10^{+2}$
		$S = 0.962$	$3.10 \cdot 10^{-1}$	$S = 0.954$	$\mathbf{1.06 \cdot 10^{-2}}$	$S = 0.973$	$6.90 \cdot 10^{-1}$	$S = 0.967$	$7.74 \cdot 10^{-1}$
	M2,M3	$<$	$7.95 \cdot 10^{+2}$	$<$	$1.57 \cdot 10^{+2}$	$<$	$5.12 \cdot 10^{+2}$	$<$	$2.62 \cdot 10^{+2}$
		$S = 0.936$	$\mathbf{1.13 \cdot 10^{-3}}$	$S = 0.905$	$\mathbf{3.18 \cdot 10^{-7}}$	$S = 0.941$	$\mathbf{2.02 \cdot 10^{-6}}$	$S = 0.935$	$\mathbf{8.89 \cdot 10^{-5}}$
	M2P,M3	$\sim$	$1.03 \cdot 10^{+3}$	$<$	$3.58 \cdot 10^{+2}$	$<$	$7.24 \cdot 10^{+2}$	$<$	$3.12 \cdot 10^{+2}$
		$S = 0.953$	$7.62 \cdot 10^{-2}$	$S = 0.946$	$\mathbf{3.16 \cdot 10^{-3}}$	$S = 0.939$	$\mathbf{2.86 \cdot 10^{-4}}$	$S = 0.935$	$\mathbf{6.65 \cdot 10^{-4}}$
Long. HDM	M2,M2P	$\sim$	$1.14 \cdot 10^{+3}$	$\sim$	$6.09 \cdot 10^{+2}$	$\sim$	$1.13 \cdot 10^{+3}$	$\sim$	$6.13 \cdot 10^{+2}$
		$S = 0.973$	$2.43 \cdot 10^{-1}$	$S = 0.974$	$7.89 \cdot 10^{-1}$	$S = 0.973$	$2.23 \cdot 10^{-1}$	$S = 0.951$	$8.18 \cdot 10^{-1}$
	M2,M3	$>$	$1.82 \cdot 10^{+3}$	$>$	$1.06 \cdot 10^{+3}$	$>$	$1.80 \cdot 10^{+3}$	$>$	$1.07 \cdot 10^{+3}$
		$S = 0.932$	$\mathbf{5.16 \cdot 10^{-3}}$	$S = 0.792$	$\mathbf{8.02 \cdot 10^{-6}}$	$S = 0.928$	$\mathbf{6.63 \cdot 10^{-3}}$	$S = 0.789$	$\mathbf{5.23 \cdot 10^{-6}}$
	M2P,M3	$\sim$	$1.02 \cdot 10^{+3}$	$>$	$1.05 \cdot 10^{+3}$	$>$	$1.80 \cdot 10^{+3}$	$>$	$1.04 \cdot 10^{+3}$
		$S = 0.937$	$7.09 \cdot 10^{-2}$	$S = 0.787$	$\mathbf{1.56 \cdot 10^{-5}}$	$S = 0.937$	$\mathbf{6.23 \cdot 10^{-3}}$	$S = 0.775$	$\mathbf{2.21 \cdot 10^{-5}}$
	M2,M2P	$<$	$8.25 \cdot 10^{+2}$	$\sim$	$4.70 \cdot 10^{+2}$	$\sim$	$1.35 \cdot 10^{+3}$	$\sim$	$6.05 \cdot 10^{+2}$
		$S = 0.832$	$\mathbf{1.93 \cdot 10^{-3}}$	$S = 0.949$	$1.07 \cdot 10^{-1}$	$S = 0.972$	$9.85 \cdot 10^{-1}$	$S = 0.953$	$7.59 \cdot 10^{-1}$
	M2,M3	$<$	$6.98 \cdot 10^{+2}$	$<$	$1.84 \cdot 10^{+2}$	$<$	$6.62 \cdot 10^{+2}$	$<$	$1.76 \cdot 10^{+2}$
		$S = 0.918$	$\mathbf{1.67 \cdot 10^{-4}}$	$S = 0.906$	$\mathbf{1.69 \cdot 10^{-6}}$	$S = 0.938$	$\mathbf{7.68 \cdot 10^{-5}}$	$S = 0.914$	$\mathbf{1.05 \cdot 10^{-6}}$
	M2P,M3	$\sim$	$1.05 \cdot 10^{+3}$	$<$	$3.33 \cdot 10^{+2}$	$<$	$7.67 \cdot 10^{+2}$	$<$	$2.79 \cdot 10^{+2}$
		$S = 0.863$	$1.01 \cdot 10^{-1}$	$S = 0.934$	$\mathbf{1.40 \cdot 10^{-3}}$	$S = 0.927$	$\mathbf{6.69 \cdot 10^{-4}}$	$S = 0.928$	$\mathbf{1.84 \cdot 10^{-4}}$
Transv. HDM	M2,M2P	$>$	$1.89 \cdot 10^{+3}$	$\sim$	$6.34 \cdot 10^{+2}$	$>$	$1.81 \cdot 10^{+3}$	$\sim$	$5.20 \cdot 10^{+2}$
		$S = 0.980$	$\mathbf{1.51 \cdot 10^{-3}}$	$S = 0.981$	$9.77 \cdot 10^{-1}$	$S = 0.971$	$\mathbf{5.68 \cdot 10^{-3}}$	$S = 0.985$	$2.61 \cdot 10^{-1}$
	M2,M3	$<$	$8.78 \cdot 10^{+2}$	$\sim$	$5.06 \cdot 10^{+2}$	$<$	$8.70 \cdot 10^{+2}$	$\sim$	$5.39 \cdot 10^{+2}$
		$S = 0.884$	$\mathbf{4.69 \cdot 10^{-3}}$	$S = 0.932$	$2.08 \cdot 10^{-1}$	$S = 0.896$	$\mathbf{4.13 \cdot 10^{-3}}$	$S = 0.935$	$3.47 \cdot 10^{-1}$
	M2P,M3	$<$	$7.91 \cdot 10^{+2}$	$\sim$	$5.76 \cdot 10^{+2}$	$<$	$8.12 \cdot 10^{+2}$	$\sim$	$5.44 \cdot 10^{+2}$
		$S = 0.879$	$\mathbf{1.05 \cdot 10^{-3}}$	$S = 0.943$	$5.59 \cdot 10^{-1}$	$S = 0.880$	$\mathbf{1.54 \cdot 10^{-3}}$	$S = 0.938$	$3.72 \cdot 10^{-1}$
	M2,M2P	$\sim$	$1.16 \cdot 10^{+3}$	$\sim$	$5.19 \cdot 10^{+2}$	$\sim$	$1.09 \cdot 10^{+3}$	$\sim$	$6.27 \cdot 10^{+2}$
		$S = 0.981$	$2.98 \cdot 10^{-1}$	$S = 0.972$	$2.57 \cdot 10^{-1}$	$S = 0.976$	$1.59 \cdot 10^{-1}$	$S = 0.963$	$9.24 \cdot 10^{-1}$
	M2,M3	$<$	$8.32 \cdot 10^{+2}$	$<$	$3.49 \cdot 10^{+2}$	$<$	$5.45 \cdot 10^{+2}$	$<$	$4.06 \cdot 10^{+2}$
		$S = 0.956$	$\mathbf{2.18 \cdot 10^{-3}}$	$S = 0.945$	$\mathbf{2.38 \cdot 10^{-3}}$	$S = 0.959$	$\mathbf{4.75 \cdot 10^{-6}}$	$S = 0.958$	$\mathbf{1.24 \cdot 10^{-2}}$
	M2P,M3	$<$	$9.20 \cdot 10^{+2}$	$\sim$	$4.83 \cdot 10^{+2}$	$<$	$7.69 \cdot 10^{+2}$	$<$	$3.88 \cdot 10^{+2}$
		$S = 0.955$	$\mathbf{8.97 \cdot 10^{-3}}$	$S = 0.945$	$1.38 \cdot 10^{-1}$	$S = 0.955$	$\mathbf{6.95 \cdot 10^{-4}}$	$S = 0.964$	$\mathbf{7.66 \cdot 10^{-3}}$
Tr. diffusion coef.	M2,M2P	$\sim$	$1.03 \cdot 10^{+3}$	$\sim$	$5.72 \cdot 10^{+2}$	$>$	$2.49 \cdot 10^{+3}$	$>$	$1.07 \cdot 10^{+3}$
		$S = 0.503$	$7.81 \cdot 10^{-2}$	$S = 0.417$	$5.34 \cdot 10^{-1}$	$S = 0.584$	$\mathbf{1.81 \cdot 10^{-10}}$	$S = 0.585$	$\mathbf{4.44 \cdot 10^{-6}}$
	M2,M3	$<$	$1.34 \cdot 10^{+2}$	$<$	$2.00 \cdot 10^{+0}$	$<$	0.00	$<$	0.00
		$S = 0.508$	$\mathbf{1.13 \cdot 10^{-11}}$	$S = 0.282$	$\mathbf{2.66 \cdot 10^{-15}}$	$S = 0.474$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.352$	$\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$	$5.35 \cdot 10^{+2}$	$<$	$2.14 \cdot 10^{+2}$	$<$	0.00	$<$	$1.36 \cdot 10^{+2}$
		$S = 0.677$	$\mathbf{3.68 \cdot 10^{-6}}$	$S = 0.661$	$\mathbf{8.90 \cdot 10^{-6}}$	$S = 0.256$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.486$	$\mathbf{7.54 \cdot 10^{-8}}$
	M2,M2P	$\sim$	$1.11 \cdot 10^{+3}$	$\sim$	$6.30 \cdot 10^{+2}$	$>$	$2.27 \cdot 10^{+3}$	$>$	$9.87 \cdot 10^{+2}$
		$S = 0.546$	$1.84 \cdot 10^{-1}$	$S = 0.471$	$9.47 \cdot 10^{-1}$	$S = 0.579$	$\mathbf{2.21 \cdot 10^{-7}}$	$S = 0.651$	$\mathbf{2.65 \cdot 10^{-4}}$
	M2,M3	$<$	$2.60 \cdot 10^{+2}$	$<$	$1.40 \cdot 10^{+1}$	$<$	$4.70 \cdot 10^{+1}$	$<$	$3.00 \cdot 10^{+0}$
		$S = 0.570$	$\mathbf{1.02 \cdot 10^{-9}}$	$S = 0.247$	$\mathbf{9.77 \cdot 10^{-14}}$	$S = 0.476$	$\mathbf{3.86 \cdot 10^{-13}}$	$S = 0.313$	$\mathbf{4.44 \cdot 10^{-15}}$
	M2P,M3	$<$	$6.21 \cdot 10^{+2}$	$<$	$1.93 \cdot 10^{+2}$	$<$	$2.80 \cdot 10^{+1}$	$<$	$8.20 \cdot 10^{+1}$
		$S = 0.774$	$\mathbf{3.03 \cdot 10^{-5}}$	$S = 0.613$	$\mathbf{2.84 \cdot 10^{-6}}$	$S = 0.377$	$\mathbf{1.79 \cdot 10^{-13}}$	$S = 0.481$	$\mathbf{8.78 \cdot 10^{-10}}$
Rot. diffusion coef.	M2,M2P	$>$	$2.08 \cdot 10^{+3}$	$>$	$9.46 \cdot 10^{+2}$	$>$	$2.63 \cdot 10^{+3}$	$>$	$1.04 \cdot 10^{+3}$
		$S = 0.591$	$\mathbf{2.96 \cdot 10^{-5}}$	$S = 0.429$	$\mathbf{1.22 \cdot 10^{-3}}$	$S = 0.756$	$\mathbf{1.08 \cdot 10^{-12}}$	$S = 0.541$	$\mathbf{2.44 \cdot 10^{-5}}$
	M2,M3	$<$	0.00	$<$	0.00	$<$	0.00	$<$	0.00
		$S = 0.547$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.238$	$\mathbf{8.88 \cdot 10^{-16}}$	$S = 0.708$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.379$	$\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$	$4.50 \cdot 10^{+2}$	$<$	$1.22 \cdot 10^{+2}$	$<$	0.00	$<$	$1.22 \cdot 10^{+2}$
		$S = 0.730$	$\mathbf{3.70 \cdot 10^{-7}}$	$S = 0.646$	$\mathbf{2.67 \cdot 10^{-8}}$	$S = 0.496$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.695$	$\mathbf{2.67 \cdot 10^{-8}}$
	M2,M2P	$>$	$2.08 \cdot 10^{+3}$	$>$	$9.02 \cdot 10^{+2}$	$>$	$2.63 \cdot 10^{+3}$	$>$	$9.91 \cdot 10^{+2}$
		$S = 0.461$	$\mathbf{3.10 \cdot 10^{-5}}$	$S = 0.280$	$\mathbf{4.98 \cdot 10^{-3}}$	$S = 0.715$	$\mathbf{1.08 \cdot 10^{-12}}$	$S = 0.346$	$\mathbf{2.25 \cdot 10^{-4}}$
	M2,M3	$<$	$4.00 \cdot 10^{+0}$	$<$	0.00	$<$	$1.20 \cdot 10^{+1}$	$<$	0.00
		$S = 0.399$	$\mathbf{6.68 \cdot 10^{-14}}$	$S = 0.136$	$\mathbf{8.88 \cdot 10^{-16}}$	$S = 0.566$	$\mathbf{9.30 \cdot 10^{-14}}$	$S = 0.209$	$\mathbf{8.88 \cdot 10^{-16}}$
	M2P,M3	$<$	$4.60 \cdot 10^{+2}$	$<$	$1.37 \cdot 10^{+2}$	$<$	0.00	$<$	$1.32 \cdot 10^{+2}$
		$S = 0.803$	$\mathbf{4.90 \cdot 10^{-7}}$	$S = 0.623$	$\mathbf{8.10 \cdot 10^{-8}}$	$S = 0.584$	$\mathbf{5.66 \cdot 10^{-14}}$	$S = 0.669$	$\mathbf{5.64 \cdot 10^{-8}}$

Table D.12: Differences in the values of MD descriptors based on force field for each combination of antimicrobial activity and lipid composition. Results of the Wilcoxon signed-rank test. Continued from previous pages. Lines with gray background refer to the samples of the mean value of the property; lines with no background correspond to those of their standard deviation. Each test compares the distribution of the differences between MD descriptors associated with two force fields. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution of differences $x_{\mathrm{FF1}} - x_{\mathrm{FF2}}$ is skewed toward positive values, i.e. larger values for simulations of the first force field, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Data set	Sample	
	Active	Nonactive
Sequence length	$9.56 \cdot 10^{-1}$ $p = \mathbf{1.18 \cdot 10^{-2}}$	$9.00 \cdot 10^{-1}$ $p = \mathbf{4.67 \cdot 10^{-4}}$
Charge per res.	$9.58 \cdot 10^{-1}$ $p = \mathbf{1.55 \cdot 10^{-2}}$	$9.27 \cdot 10^{-1}$ $p = \mathbf{4.20 \cdot 10^{-3}}$
Frac. acidic res.	$6.73 \cdot 10^{-1}$ $p = \mathbf{1.80 \cdot 10^{-11}}$	$7.58 \cdot 10^{-1}$ $p = \mathbf{1.08 \cdot 10^{-7}}$
Frac. basic res.	$9.65 \cdot 10^{-1}$ $p = \mathbf{3.86 \cdot 10^{-2}}$	$9.44 \cdot 10^{-1}$ $p = \mathbf{1.94 \cdot 10^{-2}}$
Frac. polar res.	$9.41 \cdot 10^{-1}$ $p = \mathbf{1.87 \cdot 10^{-3}}$	$9.39 \cdot 10^{-1}$ $p = \mathbf{1.27 \cdot 10^{-2}}$
Frac. hydroph. res.	$9.73 \cdot 10^{-1}$ $p = 1.13 \cdot 10^{-1}$	$9.72 \cdot 10^{-1}$ $p = 2.70 \cdot 10^{-1}$
Frac. R residues	$7.20 \cdot 10^{-1}$ $p = \mathbf{1.65 \cdot 10^{-10}}$	$7.73 \cdot 10^{-1}$ $p = \mathbf{2.26 \cdot 10^{-7}}$
Frac. H res.	$4.63 \cdot 10^{-1}$ $p = \mathbf{6.76 \cdot 10^{-15}}$	$5.60 \cdot 10^{-1}$ $p = \mathbf{5.23 \cdot 10^{-11}}$
Frac. K res.	$9.09 \cdot 10^{-1}$ $p = \mathbf{6.55 \cdot 10^{-5}}$	$7.84 \cdot 10^{-1}$ $p = \mathbf{3.80 \cdot 10^{-7}}$
Frac. D res.	$4.95 \cdot 10^{-1}$ $p = \mathbf{1.94 \cdot 10^{-14}}$	$6.64 \cdot 10^{-1}$ $p = \mathbf{1.95 \cdot 10^{-9}}$
Frac. E res.	$5.45 \cdot 10^{-1}$ $p = \mathbf{1.08 \cdot 10^{-13}}$	$6.24 \cdot 10^{-1}$ $p = \mathbf{4.51 \cdot 10^{-10}}$
Frac. S res.	$8.52 \cdot 10^{-1}$ $p = \mathbf{5.03 \cdot 10^{-7}}$	$8.54 \cdot 10^{-1}$ $p = \mathbf{2.04 \cdot 10^{-5}}$
Frac. T res.	$6.50 \cdot 10^{-1}$ $p = \mathbf{6.54 \cdot 10^{-12}}$	$8.01 \cdot 10^{-1}$ $p = \mathbf{9.25 \cdot 10^{-7}}$
Frac. N res.	$4.95 \cdot 10^{-1}$ $p = \mathbf{1.94 \cdot 10^{-14}}$	$6.64 \cdot 10^{-1}$ $p = \mathbf{1.95 \cdot 10^{-9}}$
Frac. Q res.	$6.39 \cdot 10^{-1}$ $p = \mathbf{4.17 \cdot 10^{-12}}$	$6.69 \cdot 10^{-1}$ $p = \mathbf{2.39 \cdot 10^{-9}}$
Frac. C res.	$3.48 \cdot 10^{-1}$ $p = \mathbf{2.30 \cdot 10^{-16}}$	$7.02 \cdot 10^{-1}$ $p = \mathbf{9.01 \cdot 10^{-9}}$

Data set	Sample	
	Active	Nonactive
Frac. G res.	$9.11 \cdot 10^{-1}$ $p = \mathbf{7.58 \cdot 10^{-5}}$	$8.61 \cdot 10^{-1}$ $p = \mathbf{3.08 \cdot 10^{-5}}$
Frac. P res.	$7.13 \cdot 10^{-1}$ $p = \mathbf{1.18 \cdot 10^{-10}}$	$7.26 \cdot 10^{-1}$ $p = \mathbf{2.49 \cdot 10^{-8}}$
Frac. A res.	$9.15 \cdot 10^{-1}$ $p = \mathbf{1.15 \cdot 10^{-4}}$	$7.48 \cdot 10^{-1}$ $p = \mathbf{6.84 \cdot 10^{-8}}$
Frac. V res.	$6.63 \cdot 10^{-1}$ $p = \mathbf{1.14 \cdot 10^{-11}}$	$7.41 \cdot 10^{-1}$ $p = \mathbf{4.85 \cdot 10^{-8}}$
Frac. I res.	$8.30 \cdot 10^{-1}$ $p = \mathbf{1.06 \cdot 10^{-7}}$	$6.97 \cdot 10^{-1}$ $p = \mathbf{7.24 \cdot 10^{-9}}$
Frac. L res.	$8.47 \cdot 10^{-1}$ $p = \mathbf{3.41 \cdot 10^{-7}}$	$8.70 \cdot 10^{-1}$ $p = \mathbf{5.76 \cdot 10^{-5}}$
Frac. M res.	$5.26 \cdot 10^{-1}$ $p = \mathbf{5.59 \cdot 10^{-14}}$	$5.69 \cdot 10^{-1}$ $p = \mathbf{7.02 \cdot 10^{-11}}$
Frac. F res.	$8.15 \cdot 10^{-1}$ $p = \mathbf{3.89 \cdot 10^{-8}}$	$7.50 \cdot 10^{-1}$ $p = \mathbf{7.28 \cdot 10^{-8}}$
Frac. Y res.	$4.73 \cdot 10^{-1}$ $p = \mathbf{9.51 \cdot 10^{-15}}$	$7.39 \cdot 10^{-1}$ $p = \mathbf{4.49 \cdot 10^{-8}}$
Frac. W res.	$6.46 \cdot 10^{-1}$ $p = \mathbf{5.56 \cdot 10^{-12}}$	$6.55 \cdot 10^{-1}$ $p = \mathbf{1.41 \cdot 10^{-9}}$
Total EDM	$9.53 \cdot 10^{-1}$ $p = \mathbf{8.43 \cdot 10^{-3}}$	$9.63 \cdot 10^{-1}$ $p = 1.15 \cdot 10^{-1}$
Long. EDM	$9.57 \cdot 10^{-1}$ $p = \mathbf{1.37 \cdot 10^{-2}}$	$9.73 \cdot 10^{-1}$ $p = 3.11 \cdot 10^{-1}$
Transv. EDM	$9.36 \cdot 10^{-1}$ $p = \mathbf{1.15 \cdot 10^{-3}}$	$9.62 \cdot 10^{-1}$ $p = 1.07 \cdot 10^{-1}$
Total HDM	$8.73 \cdot 10^{-1}$ $p = \mathbf{2.54 \cdot 10^{-6}}$	$7.81 \cdot 10^{-1}$ $p = \mathbf{3.28 \cdot 10^{-7}}$
Long. HDM	$8.87 \cdot 10^{-1}$ $p = \mathbf{8.50 \cdot 10^{-6}}$	$7.98 \cdot 10^{-1}$ $p = \mathbf{7.83 \cdot 10^{-7}}$
Transv. HDM	$9.88 \cdot 10^{-1}$ $p = 7.18 \cdot 10^{-1}$	$9.77 \cdot 10^{-1}$ $p = 4.28 \cdot 10^{-1}$

Table D.13: Results of the Shapiro-Wilk tests applied to the samples of peptide physicochemical properties. The samples are grouped by antimicrobial activity. The two values reported are (upper) the value of the statistic W and (lower) the p-value of the test performed on each sample. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a non-normal parent distribution.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude	Results			
Sequence length	$\sim$	$U = 2.12 \cdot 10^{+3}$	$S = 0.853$	$p = 1.24 \cdot 10^{-1}$
Charge per res.	$>$	$U = 2.26 \cdot 10^{+3}$	$S = 0.844$	$p = \mathbf{1.20 \cdot 10^{-2}}$
Frac. acidic res.	$\sim$	$U = 1.56 \cdot 10^{+3}$	$S = 0.828$	$p = 1.26 \cdot 10^{-1}$
Frac. basic res.	$\sim$	$U = 2.16 \cdot 10^{+3}$	$S = 0.870$	$p = 8.54 \cdot 10^{-2}$
Frac. polar res.	$<^*$	$U = 9.27 \cdot 10^{+2}$	$S = 0.457$	$p = \mathbf{1.85 \cdot 10^{-6}}$
Frac. hydrophobic res.	$>$	$U = 2.61 \cdot 10^{+3}$	$S = 0.680$	$p = \mathbf{2.43 \cdot 10^{-5}}$
Frac. R residues	$<$	$U = 1.34 \cdot 10^{+3}$	$S = 0.765$	$p = \mathbf{4.21 \cdot 10^{-3}}$
Frac. H res.	$\sim$	$U = 1.76 \cdot 10^{+3}$	$S = 0.900$	$p = 6.49 \cdot 10^{-1}$
Frac. K res.	$>$	$U = 2.69 \cdot 10^{+3}$	$S = 0.656$	$p = \mathbf{3.90 \cdot 10^{-6}}$
Frac. D res.	$\sim$	$U = 1.55 \cdot 10^{+3}$	$S = 0.762$	$p = 5.93 \cdot 10^{-2}$
Frac. E res.	$\sim$	$U = 1.73 \cdot 10^{+3}$	$S = 0.917$	$p = 5.10 \cdot 10^{-1}$
Frac. S res.	$\sim^*$	$U = 1.64 \cdot 10^{+3}$	$S = 0.759$	$p = 3.28 \cdot 10^{-1}$
Frac. T res.	$<^*$	$U = 1.30 \cdot 10^{+3}$	$S = 0.677$	$p = \mathbf{1.02 \cdot 10^{-3}}$
Frac. N res.	$\sim$	$U = 1.55 \cdot 10^{+3}$	$S = 0.762$	$p = 5.93 \cdot 10^{-2}$
Frac. Q res.	$\sim$	$U = 1.70 \cdot 10^{+3}$	$S = 0.834$	$p = 4.60 \cdot 10^{-1}$
Frac. C res.	$<^*$	$U = 1.29 \cdot 10^{+3}$	$S = 0.349$	$p = \mathbf{4.06 \cdot 10^{-5}}$

Magnitude	Results			
Frac. G res.	$\sim$	$U = 1.85 \cdot 10^{+3}$	$S = 0.926$	$p = 9.11 \cdot 10^{-1}$
Frac. P res.	$<^*$	$U = 1.32 \cdot 10^{+3}$	$S = 0.690$	$p = \mathbf{2.49 \cdot 10^{-3}}$
Frac. A res.	$>$	$U = 2.24 \cdot 10^{+3}$	$S = 0.756$	$p = \mathbf{1.51 \cdot 10^{-2}}$
Frac. V res.	$\sim$	$U = 1.94 \cdot 10^{+3}$	$S = 0.910$	$p = 5.28 \cdot 10^{-1}$
Frac. I res.	$>^*$	$U = 2.56 \cdot 10^{+3}$	$S = 0.668$	$p = \mathbf{3.67 \cdot 10^{-5}}$
Frac. L res.	$>$	$U = 2.42 \cdot 10^{+3}$	$S = 0.738$	$p = \mathbf{9.82 \cdot 10^{-4}}$
Frac. M res.	$\sim$	$U = 1.76 \cdot 10^{+3}$	$S = 0.893$	$p = 6.70 \cdot 10^{-1}$
Frac. F res.	$\sim^*$	$U = 2.13 \cdot 10^{+3}$	$S = 0.811$	$p = 9.21 \cdot 10^{-2}$
Frac. Y res.	$<^*$	$U = 1.35 \cdot 10^{+3}$	$S = 0.622$	$p = \mathbf{8.14 \cdot 10^{-4}}$
Frac. W res.	$\sim$	$U = 1.96 \cdot 10^{+3}$	$S = 0.888$	$p = 4.12 \cdot 10^{-1}$
Total EDM	$\sim$	$U = 1.78 \cdot 10^{+3}$	$S = 0.875$	$p = 7.99 \cdot 10^{-1}$
Long. EDM	$\sim$	$U = 1.79 \cdot 10^{+3}$	$S = 0.869$	$p = 8.59 \cdot 10^{-1}$
Transv. EDM	$\sim$	$U = 1.67 \cdot 10^{+3}$	$S = 0.897$	$p = 4.26 \cdot 10^{-1}$
Total HDM	$\sim$	$U = 2.11 \cdot 10^{+3}$	$S = 0.871$	$p = 1.47 \cdot 10^{-1}$
Long. HDM	$\sim$	$U = 1.82 \cdot 10^{+3}$	$S = 0.848$	$p = 9.98 \cdot 10^{-1}$
Transv. HDM	$>^*$	$U = 2.72 \cdot 10^{+3}$	$S = 0.646$	$p = \mathbf{1.86 \cdot 10^{-6}}$

Table D.14: Differences in the value of physicochemical properties based on antimicrobial activity. Results of the Mann-Whitney U test. Each test compares the distribution of physicochemical properties associated with active peptides against that of nonactives. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for active sequences is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.15).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Magnitude	Results		Magnitude	Results	
	Stat.	p-value		Stat.	p-value
Sequence length	$7.21 \cdot 10^{-1}$	$3.98 \cdot 10^{-1}$	Frac. G res.	$1.04 \cdot 10^{+0}$	$3.09 \cdot 10^{-1}$
Charge per res.	$2.50 \cdot 10^{-1}$	$6.18 \cdot 10^{-1}$	Frac. P res.	$1.07 \cdot 10^{+1}$	$\mathbf{1.37 \cdot 10^{-3}}$
Frac. acidic res.	$2.52 \cdot 10^{+0}$	$1.15 \cdot 10^{-1}$	Frac. A res.	$1.44 \cdot 10^{+0}$	$2.32 \cdot 10^{-1}$
Frac. basic res.	$1.47 \cdot 10^{+0}$	$2.28 \cdot 10^{-1}$	Frac. V res.	$3.70 \cdot 10^{-1}$	$5.44 \cdot 10^{-1}$
Frac. polar res.	$5.49 \cdot 10^{+1}$	$\mathbf{1.87 \cdot 10^{-11}}$	Frac. I res.	$8.95 \cdot 10^{+0}$	$\mathbf{3.36 \cdot 10^{-3}}$
Frac. hydroph. res.	$8.02 \cdot 10^{-1}$	$3.72 \cdot 10^{-1}$	Frac. L res.	$1.63 \cdot 10^{-1}$	$6.87 \cdot 10^{-1}$
Frac. R residues	$1.25 \cdot 10^{+0}$	$2.66 \cdot 10^{-1}$	Frac. M res.	$2.35 \cdot 10^{-1}$	$6.29 \cdot 10^{-1}$
Frac. H res.	$1.38 \cdot 10^{-3}$	$9.70 \cdot 10^{-1}$	Frac. F res.	$4.60 \cdot 10^{+0}$	$\mathbf{3.40 \cdot 10^{-2}}$
Frac. K res.	$1.26 \cdot 10^{-1}$	$7.23 \cdot 10^{-1}$	Frac. Y res.	$1.22 \cdot 10^{+1}$	$\mathbf{6.58 \cdot 10^{-4}}$
Frac. D res.	$3.90 \cdot 10^{+0}$	$5.05 \cdot 10^{-2}$	Frac. W res.	$7.31 \cdot 10^{-1}$	$3.94 \cdot 10^{-1}$
Frac. E res.	$2.52 \cdot 10^{-1}$	$6.17 \cdot 10^{-1}$	Total EDM	$4.03 \cdot 10^{-8}$	$10.00 \cdot 10^{-1}$
Frac. S res.	$1.22 \cdot 10^{+1}$	$\mathbf{6.58 \cdot 10^{-4}}$	Long. EDM	$2.18 \cdot 10^{-1}$	$6.42 \cdot 10^{-1}$
Frac. T res.	$2.04 \cdot 10^{+1}$	$\mathbf{1.50 \cdot 10^{-5}}$	Transv. EDM	$3.01 \cdot 10^{-1}$	$5.84 \cdot 10^{-1}$
Frac. N res.	$3.90 \cdot 10^{+0}$	$5.05 \cdot 10^{-2}$	Total HDM	$1.02 \cdot 10^{+0}$	$3.14 \cdot 10^{-1}$
Frac. Q res.	$1.41 \cdot 10^{+0}$	$2.38 \cdot 10^{-1}$	Long. HDM	$2.50 \cdot 10^{+0}$	$1.16 \cdot 10^{-1}$
Frac. C res.	$2.30 \cdot 10^{+1}$	$\mathbf{4.69 \cdot 10^{-6}}$	Transv. HDM	$4.04 \cdot 10^{+0}$	$\mathbf{4.65 \cdot 10^{-2}}$

Table D.15: Differences in the variance of the samples of physicochemical properties based on antimicrobial activity. The results of the Brown-Forsythe test. The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in variances.

Data set	Depth		Cont. water		Long. EDM		Cont. tails	
	Hydroph.	resid.	Hydroph.	resid.	Charge		Hydroph.	resid.
M2 + POPC	0.640	0.717 0.442	−0.537	−0.642 −0.341	−0.613	−0.659 −0.538	0.449	0.460 0.375
M2 + POPE:POPG	0.611	0.660 0.462	−0.539	−0.615 −0.372	−0.650	−0.699 −0.556	0.442	0.431 0.390
M2P + POPC	0.555	0.661 0.373	−0.464	−0.606 −0.230	−0.543	−0.624 −0.414	0.445	0.462 0.381
M2P + POPE:POPG	0.633	0.677 0.505	−0.503	−0.588 −0.337	−0.574	−0.632 −0.477	0.456	0.414 0.452
M3 + POPC	0.569	0.605 0.380	−0.562	−0.680 −0.350	−0.573	−0.643 −0.472	0.506	0.442 0.436
M3 + POPE:POPG	0.535	0.520 0.406	−0.567	−0.676 −0.358	−0.599	−0.649 −0.534	0.475	0.350 0.487

Table D.16: Kendall rank correlation coefficient τ between MD descriptors and structural properties. Only pairs of MD descriptor and physicochemical property where $|\tau| \geq 0.5$ are reported. The values of τ for (left) the whole sample, (upper right) the sample of active peptides and (lower right) nonactives are given.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Data set	Metrics				
	Accuracy	Precision	Recall	$F1$ score	AUC-PR
All data	$7.90 \cdot 10^{-1}$	$8.19 \cdot 10^{-1}$	$8.49 \cdot 10^{-1}$	$8.28 \cdot 10^{-1}$	$8.82 \cdot 10^{-1}$
	$2.41 \cdot 10^{-2}$	$2.25 \cdot 10^{-2}$	$2.87 \cdot 10^{-2}$	$2.09 \cdot 10^{-2}$	$1.97 \cdot 10^{-2}$
M2	$7.57 \cdot 10^{-1}$	$7.85 \cdot 10^{-1}$	$8.25 \cdot 10^{-1}$	$8.01 \cdot 10^{-1}$	$8.52 \cdot 10^{-1}$
	$2.17 \cdot 10^{-2}$	$2.15 \cdot 10^{-2}$	$2.30 \cdot 10^{-2}$	$1.81 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$
M2P	$7.82 \cdot 10^{-1}$	$8.11 \cdot 10^{-1}$	$8.38 \cdot 10^{-1}$	$8.20 \cdot 10^{-1}$	$9.08 \cdot 10^{-1}$
	$2.35 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$	$2.48 \cdot 10^{-2}$	$1.97 \cdot 10^{-2}$	$1.25 \cdot 10^{-2}$
M3	$8.04 \cdot 10^{-1}$	$8.26 \cdot 10^{-1}$	$8.49 \cdot 10^{-1}$	$8.33 \cdot 10^{-1}$	$8.66 \cdot 10^{-1}$
	$2.04 \cdot 10^{-2}$	$1.95 \cdot 10^{-2}$	$2.41 \cdot 10^{-2}$	$1.78 \cdot 10^{-2}$	$1.68 \cdot 10^{-2}$
M2 + POPC	$7.41 \cdot 10^{-1}$	$7.67 \cdot 10^{-1}$	$8.23 \cdot 10^{-1}$	$7.91 \cdot 10^{-1}$	$8.64 \cdot 10^{-1}$
	$1.92 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$2.22 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	$1.38 \cdot 10^{-2}$
M2 + POPE:POPG	$7.31 \cdot 10^{-1}$	$7.51 \cdot 10^{-1}$	$8.22 \cdot 10^{-1}$	$7.82 \cdot 10^{-1}$	$8.41 \cdot 10^{-1}$
	$2.10 \cdot 10^{-2}$	$2.13 \cdot 10^{-2}$	$2.01 \cdot 10^{-2}$	$1.64 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$
M2P + POPC	$7.72 \cdot 10^{-1}$	$7.83 \cdot 10^{-1}$	$8.63 \cdot 10^{-1}$	$8.17 \cdot 10^{-1}$	$8.58 \cdot 10^{-1}$
	$2.05 \cdot 10^{-2}$	$2.02 \cdot 10^{-2}$	$2.05 \cdot 10^{-2}$	$1.61 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$
M2P + POPE:POPG	$7.56 \cdot 10^{-1}$	$7.78 \cdot 10^{-1}$	$8.37 \cdot 10^{-1}$	$8.04 \cdot 10^{-1}$	$8.91 \cdot 10^{-1}$
	$2.16 \cdot 10^{-2}$	$1.96 \cdot 10^{-2}$	$2.48 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$1.60 \cdot 10^{-2}$
M3 + POPC	$7.49 \cdot 10^{-1}$	$7.72 \cdot 10^{-1}$	$8.36 \cdot 10^{-1}$	$7.98 \cdot 10^{-1}$	$8.78 \cdot 10^{-1}$
	$1.61 \cdot 10^{-2}$	$1.55 \cdot 10^{-2}$	$1.79 \cdot 10^{-2}$	$1.35 \cdot 10^{-2}$	$1.33 \cdot 10^{-2}$
M3 + POPE:POPG	$7.72 \cdot 10^{-1}$	$7.83 \cdot 10^{-1}$	$8.63 \cdot 10^{-1}$	$8.19 \cdot 10^{-1}$	$8.73 \cdot 10^{-1}$
	$1.83 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	$1.80 \cdot 10^{-2}$	$1.45 \cdot 10^{-2}$	$1.42 \cdot 10^{-2}$

Table D.17: Statistics of the samples of performance metrics. Samples correspond to the mean of the performance metrics across the 5-folds for the 200 independent iterations of train–test process.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Data set	Metrics				
	Accuracy	Precision	Recall	$F1$ score	AUC-PR
All data	$9.87 \cdot 10^{-1}$ $p = 6.00 \cdot 10^{-2}$	$9.86 \cdot 10^{-1}$ $p = \mathbf{4.94 \cdot 10^{-2}}$	$9.88 \cdot 10^{-1}$ $p = 1.04 \cdot 10^{-1}$	$9.97 \cdot 10^{-1}$ $p = 9.67 \cdot 10^{-1}$	$9.96 \cdot 10^{-1}$ $p = 8.91 \cdot 10^{-1}$
M2	$9.79 \cdot 10^{-1}$ $p = \mathbf{4.77 \cdot 10^{-3}}$	$9.89 \cdot 10^{-1}$ $p = 1.38 \cdot 10^{-1}$	$9.73 \cdot 10^{-1}$ $p = \mathbf{7.19 \cdot 10^{-4}}$	$9.84 \cdot 10^{-1}$ $p = \mathbf{2.02 \cdot 10^{-2}}$	$9.92 \cdot 10^{-1}$ $p = 3.49 \cdot 10^{-1}$
M2P	$9.81 \cdot 10^{-1}$ $p = \mathbf{7.84 \cdot 10^{-3}}$	$9.90 \cdot 10^{-1}$ $p = 1.81 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$ $p = \mathbf{1.30 \cdot 10^{-2}}$	$9.90 \cdot 10^{-1}$ $p = 1.89 \cdot 10^{-1}$	$9.80 \cdot 10^{-1}$ $p = \mathbf{5.26 \cdot 10^{-3}}$
M3	$9.87 \cdot 10^{-1}$ $p = 5.54 \cdot 10^{-2}$	$9.94 \cdot 10^{-1}$ $p = 5.68 \cdot 10^{-1}$	$9.71 \cdot 10^{-1}$ $p = \mathbf{3.68 \cdot 10^{-4}}$	$9.89 \cdot 10^{-1}$ $p = 1.33 \cdot 10^{-1}$	$9.89 \cdot 10^{-1}$ $p = 1.18 \cdot 10^{-1}$
M2 + POPC	$9.90 \cdot 10^{-1}$ $p = 1.61 \cdot 10^{-1}$	$9.94 \cdot 10^{-1}$ $p = 5.43 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$ $p = \mathbf{1.02 \cdot 10^{-2}}$	$9.93 \cdot 10^{-1}$ $p = 4.24 \cdot 10^{-1}$	$9.95 \cdot 10^{-1}$ $p = 6.89 \cdot 10^{-1}$
M2 + POPE:POPG	$9.83 \cdot 10^{-1}$ $p = \mathbf{1.75 \cdot 10^{-2}}$	$9.92 \cdot 10^{-1}$ $p = 3.69 \cdot 10^{-1}$	$9.69 \cdot 10^{-1}$ $p = \mathbf{2.08 \cdot 10^{-4}}$	$9.93 \cdot 10^{-1}$ $p = 4.92 \cdot 10^{-1}$	$9.81 \cdot 10^{-1}$ $p = \mathbf{7.67 \cdot 10^{-3}}$
M2P + POPC	$9.87 \cdot 10^{-1}$ $p = 5.97 \cdot 10^{-2}$	$9.95 \cdot 10^{-1}$ $p = 7.99 \cdot 10^{-1}$	$9.76 \cdot 10^{-1}$ $p = \mathbf{1.79 \cdot 10^{-3}}$	$9.96 \cdot 10^{-1}$ $p = 9.25 \cdot 10^{-1}$	$9.90 \cdot 10^{-1}$ $p = 1.60 \cdot 10^{-1}$
M2P + POPE:POPG	$9.77 \cdot 10^{-1}$ $p = \mathbf{2.47 \cdot 10^{-3}}$	$9.90 \cdot 10^{-1}$ $p = 1.63 \cdot 10^{-1}$	$9.82 \cdot 10^{-1}$ $p = \mathbf{1.35 \cdot 10^{-2}}$	$9.81 \cdot 10^{-1}$ $p = \mathbf{9.32 \cdot 10^{-3}}$	$9.07 \cdot 10^{-1}$ $p = \mathbf{7.58 \cdot 10^{-10}}$
M3 + POPC	$9.81 \cdot 10^{-1}$ $p = \mathbf{7.34 \cdot 10^{-3}}$	$9.95 \cdot 10^{-1}$ $p = 6.79 \cdot 10^{-1}$	$9.66 \cdot 10^{-1}$ $p = \mathbf{1.07 \cdot 10^{-4}}$	$9.92 \cdot 10^{-1}$ $p = 3.83 \cdot 10^{-1}$	$9.90 \cdot 10^{-1}$ $p = 1.64 \cdot 10^{-1}$
M3 + POPE:POPG	$9.87 \cdot 10^{-1}$ $p = 5.80 \cdot 10^{-2}$	$9.94 \cdot 10^{-1}$ $p = 5.56 \cdot 10^{-1}$	$9.70 \cdot 10^{-1}$ $p = \mathbf{2.60 \cdot 10^{-4}}$	$9.95 \cdot 10^{-1}$ $p = 7.87 \cdot 10^{-1}$	$9.86 \cdot 10^{-1}$ $p = \mathbf{4.29 \cdot 10^{-2}}$

Table D.18: Results of the Shapiro-Wilk tests applied to the samples of performance metrics. The two values reported are (upper) the value of the statistic W and (lower) the p-value of the test performed on each sample. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a non-normal parent distribution.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Data sets	Metrics									
	Accuracy		Precision		Recall		$F1$ score		AUC-PR	
All v. M2	$>$	$U = 3.46 \cdot 10^{+4}$	$>$	$U = 3.45 \cdot 10^{+4}$	$>^*$	$U = 2.80 \cdot 10^{+4}$	$>$	$U = 3.37 \cdot 10^{+4}$	$>^*$	$U = 3.57 \cdot 10^{+4}$
	$S = 0.452$	$p =$ **$5.84 \cdot 10^{-37}$**	$S = 0.474$	$p =$ **$1.83 \cdot 10^{-36}$**	$S = 0.716$	$p =$ **$2.07 \cdot 10^{-12}$**	$S = 0.504$	$p =$ **$1.05 \cdot 10^{-32}$**	$S = 0.400$	$p =$ **$4.26 \cdot 10^{-42}$**
All v. M2P	$>$	$U = 2.46 \cdot 10^{+4}$	$>$	$U = 2.43 \cdot 10^{+4}$	$\sim^*$	$U = 2.22 \cdot 10^{+4}$	$>$	$U = 2.41 \cdot 10^{+4}$	$<^*$	$U = 6.01 \cdot 10^{+3}$
	$S = 0.840$	$p =$ **$3.06 \cdot 10^{-5}$**	$S = 0.862$	$p =$ **$1.01 \cdot 10^{-4}$**	$S = 0.880$	$p = 5.25 \cdot 10^{-2}$	$S = 0.860$	$p =$ **$1.91 \cdot 10^{-4}$**	$S = 0.468$	$p =$ **$5.33 \cdot 10^{-34}$**
All v. M3	$<$	$U = 1.59 \cdot 10^{+4}$	$<$	$U = 1.52 \cdot 10^{+4}$	$\sim^*$	$U = 1.91 \cdot 10^{+4}$	$<^*$	$U = 1.65 \cdot 10^{+4}$	$>^*$	$U = 2.98 \cdot 10^{+4}$
	$S = 0.839$	$p =$ **$1.84 \cdot 10^{-4}$**	$S = 0.831$	$p =$ **$1.55 \cdot 10^{-5}$**	$S = 0.850$	$p = 4.58 \cdot 10^{-1}$	$S = 0.853$	$p =$ **$1.37 \cdot 10^{-3}$**	$S = 0.653$	$p =$ **$1.60 \cdot 10^{-17}$**

Table D.19: Differences in performance based on using all data or FF-specific data for the training of classification models. Results of the Mann-Whitney U test. The test compares the performance of the model trained with all data against those that only used data from a single force field. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for the model trained with all data is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.20

Data set	Metrics				
	Accuracy	Precision	Recall	$F1$ score	AUC-PR
All v. M2	$5.40 \cdot 10^{-1}$	$4.81 \cdot 10^{-3}$	$8.24 \cdot 10^{+0}$	$1.79 \cdot 10^{+0}$	$1.46 \cdot 10^{+1}$
	$p = 4.63 \cdot 10^{-1}$	$p = 9.45 \cdot 10^{-1}$	$p =$ **$4.32 \cdot 10^{-3}$**	$p = 1.82 \cdot 10^{-1}$	$p =$ **$1.51 \cdot 10^{-4}$**
All v. M2P	$5.55 \cdot 10^{-2}$	$1.33 \cdot 10^{+0}$	$5.16 \cdot 10^{+0}$	$2.13 \cdot 10^{-1}$	$3.89 \cdot 10^{+1}$
	$p = 8.14 \cdot 10^{-1}$	$p = 2.50 \cdot 10^{-1}$	$p =$ **$2.36 \cdot 10^{-2}$**	$p = 6.44 \cdot 10^{-1}$	$p =$ **$1.13 \cdot 10^{-9}$**
All v. M3	$3.54 \cdot 10^{+0}$	$2.33 \cdot 10^{+0}$	$1.05 \cdot 10^{+1}$	$4.65 \cdot 10^{+0}$	$6.67 \cdot 10^{+0}$
	$p = 6.05 \cdot 10^{-2}$	$p = 1.28 \cdot 10^{-1}$	$p =$ **$1.33 \cdot 10^{-3}$**	$p =$ **$3.17 \cdot 10^{-2}$**	$p =$ **$1.01 \cdot 10^{-2}$**

Table D.20: Differences in variance of the performance metric samples obtained using all data or FF-specific data for the training of classification models. Results of the Brown-Forsythe test. The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in variances.

Data set	Metrics									
	Accuracy		Precision		Recall		$F1$ score		AUC-PR	
M2 v. M2P	$<$	$U = 8.80 \cdot 10^{+3}$	$<$	$U = 9.07 \cdot 10^{+3}$	$<$	$U = 1.35 \cdot 10^{+4}$	$<$	$U = 9.35 \cdot 10^{+3}$	$<^*$	$U = 1.20 \cdot 10^{+2}$
	$S = 0.595$	$p = \mathbf{1.65 \cdot 10^{-22}}$	$S = 0.608$	$p = \mathbf{1.59 \cdot 10^{-21}}$	$S = 0.782$	$p = \mathbf{1.03 \cdot 10^{-8}}$	$S = 0.626$	$p = \mathbf{1.59 \cdot 10^{-20}}$	$S = 0.064$	$p = \mathbf{1.46 \cdot 10^{-66}}$
M2 v. M3	$<$	$U = 2.86 \cdot 10^{+3}$	$<$	$U = 2.74 \cdot 10^{+3}$	$<$	$U = 1.01 \cdot 10^{+4}$	$<$	$U = 3.63 \cdot 10^{+3}$	$<$	$U = 1.07 \cdot 10^{+4}$
	$S = 0.326$	$p = \mathbf{5.05 \cdot 10^{-50}}$	$S = 0.339$	$p = \mathbf{1.08 \cdot 10^{-50}}$	$S = 0.658$	$p = \mathbf{6.19 \cdot 10^{-18}}$	$S = 0.365$	$p = \mathbf{8.15 \cdot 10^{-46}}$	$S = 0.667$	$p = \mathbf{5.95 \cdot 10^{-16}}$
M2P v. M3	$<^*$	$U = 1.14 \cdot 10^{+4}$	$<^*$	$U = 1.13 \cdot 10^{+4}$	$<$	$U = 1.65 \cdot 10^{+4}$	$<$	$U = 1.23 \cdot 10^{+4}$	$>^*$	$U = 3.90 \cdot 10^{+4}$
	$S = 0.692$	$p = \mathbf{4.34 \cdot 10^{-14}}$	$S = 0.705$	$p = \mathbf{2.69 \cdot 10^{-14}}$	$S = 0.875$	$p = \mathbf{1.25 \cdot 10^{-3}}$	$S = 0.724$	$p = \mathbf{1.81 \cdot 10^{-11}}$	$S = 0.192$	$p = \mathbf{4.98 \cdot 10^{-61}}$

Table D.21: Differences in performance based on using FF-specific data for the training of classification models. Results of the Mann-Whitney U test. The test compares the performance of the models trained with data from two different force fields. The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for the model trained with data of the first force field is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.22

Data set	Metrics				
	Accuracy	Precision	Recall	$F1$ score	AUC-PR
M2 v. M2P	$1.09 \cdot 10^{+0}$	$1.64 \cdot 10^{+0}$	$2.35 \cdot 10^{-1}$	$8.17 \cdot 10^{-1}$	$7.07 \cdot 10^{+0}$
	$p = 2.98 \cdot 10^{-1}$	$p = 2.01 \cdot 10^{-1}$	$p = 6.28 \cdot 10^{-1}$	$p = 3.66 \cdot 10^{-1}$	$p = \mathbf{8.13 \cdot 10^{-3}}$
M2 v. M3	$1.68 \cdot 10^{+0}$	$2.39 \cdot 10^{+0}$	$4.33 \cdot 10^{-1}$	$8.92 \cdot 10^{-1}$	$1.23 \cdot 10^{+0}$
	$p = 1.95 \cdot 10^{-1}$	$p = 1.23 \cdot 10^{-1}$	$p = 5.11 \cdot 10^{-1}$	$p = 3.45 \cdot 10^{-1}$	$p = 2.68 \cdot 10^{-1}$
M2P v. M3	$5.05 \cdot 10^{+0}$	$7.71 \cdot 10^{+0}$	$1.12 \cdot 10^{+0}$	$3.16 \cdot 10^{+0}$	$1.26 \cdot 10^{+1}$
	$p = \mathbf{2.52 \cdot 10^{-2}}$	$p = \mathbf{5.74 \cdot 10^{-3}}$	$p = 2.90 \cdot 10^{-1}$	$p = 7.62 \cdot 10^{-2}$	$p = \mathbf{4.24 \cdot 10^{-4}}$

Table D.22: Differences in variance of the performance metric samples obtained using FF-specific data for the training of classification models. Results of the Brown-Forsythe test. The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in variances.

Force field	Metrics									
	Accuracy		Precision		Recall		$F1$ score		AUC-PR	
Martini 2.2	$>$	$U = 2.77 \cdot 10^{+4}$	$>^*$	$U = 2.88 \cdot 10^{+4}$	$\sim$	$U = 2.13 \cdot 10^{+4}$	$>$	$U = 2.66 \cdot 10^{+4}$	$<$	$U = 1.15 \cdot 10^{+4}$
	$S = 0.700$	$p = \mathbf{1.75 \cdot 10^{-11}}$	$S = 0.658$	$p = \mathbf{9.98 \cdot 10^{-15}}$	$S = 0.908$	$p = 2.53 \cdot 10^{-1}$	$S = 0.748$	$p = \mathbf{4.47 \cdot 10^{-9}}$	$S = 0.709$	$p = \mathbf{1.34 \cdot 10^{-13}}$
	$>$	$U = 3.24 \cdot 10^{+4}$	$>$	$U = 3.39 \cdot 10^{+4}$	$\sim^*$	$U = 2.12 \cdot 10^{+4}$	$>$	$U = 3.06 \cdot 10^{+4}$	$>$	$U = 2.79 \cdot 10^{+4}$
	$S = 0.540$	$p = \mathbf{4.72 \cdot 10^{-27}}$	$S = 0.484$	$p = \mathbf{1.51 \cdot 10^{-33}}$	$S = 0.892$	$p = 2.98 \cdot 10^{-1}$	$S = 0.615$	$p = \mathbf{3.58 \cdot 10^{-20}}$	$S = 0.717$	$p = \mathbf{3.41 \cdot 10^{-12}}$
Martini 2.2P	$>^*$	$U = 2.56 \cdot 10^{+4}$	$>^*$	$U = 3.15 \cdot 10^{+4}$	$<^*$	$U = 9.85 \cdot 10^{+3}$	$\sim^*$	$U = 2.20 \cdot 10^{+4}$	$>^*$	$U = 3.96 \cdot 10^{+4}$
	$S = 0.791$	$p = \mathbf{7.01 \cdot 10^{-7}}$	$S = 0.563$	$p = \mathbf{1.03 \cdot 10^{-23}}$	$S = 0.634$	$p = \mathbf{8.16 \cdot 10^{-19}}$	$S = 0.870$	$p = 8.72 \cdot 10^{-2}$	$S = 0.119$	$p = \mathbf{4.62 \cdot 10^{-65}}$
	$>$	$U = 3.14 \cdot 10^{+4}$	$>^*$	$U = 3.37 \cdot 10^{+4}$	$\sim$	$U = 2.08 \cdot 10^{+4}$	$>$	$U = 2.97 \cdot 10^{+4}$	$>^*$	$U = 3.30 \cdot 10^{+4}$
	$S = 0.591$	$p = \mathbf{2.89 \cdot 10^{-23}}$	$S = 0.484$	$p = \mathbf{1.40 \cdot 10^{-32}}$	$S = 0.959$	$p = 5.04 \cdot 10^{-1}$	$S = 0.659$	$p = \mathbf{3.43 \cdot 10^{-17}}$	$S = 0.532$	$p = \mathbf{7.75 \cdot 10^{-30}}$
Martini 3	$>^*$	$U = 3.88 \cdot 10^{+4}$	$>^*$	$U = 3.95 \cdot 10^{+4}$	$>^*$	$U = 2.76 \cdot 10^{+4}$	$>^*$	$U = 3.78 \cdot 10^{+4}$	$<^*$	$U = 1.14 \cdot 10^{+4}$
	$S = 0.188$	$p = \mathbf{4.91 \cdot 10^{-60}}$	$S = 0.142$	$p = \mathbf{6.30 \cdot 10^{-64}}$	$S = 0.694$	$p = \mathbf{1.96 \cdot 10^{-11}}$	$S = 0.252$	$p = \mathbf{9.12 \cdot 10^{-54}}$	$S = 0.690$	$p = \mathbf{5.28 \cdot 10^{-14}}$
	$>$	$U = 3.40 \cdot 10^{+4}$	$>$	$U = 3.77 \cdot 10^{+4}$	$<^*$	$U = 1.23 \cdot 10^{+4}$	$>^*$	$U = 3.03 \cdot 10^{+4}$	$<$	$U = 1.60 \cdot 10^{+4}$
	$S = 0.463$	$p = \mathbf{4.50 \cdot 10^{-34}}$	$S = 0.304$	$p = \mathbf{4.70 \cdot 10^{-53}}$	$S = 0.697$	$p = \mathbf{1.43 \cdot 10^{-11}}$	$S = 0.625$	$p = \mathbf{2.57 \cdot 10^{-19}}$	$S = 0.845$	$p = \mathbf{2.90 \cdot 10^{-4}}$

Table D.23: Differences in performance based on using data from both lipid compositions or one for each force field. Results of the Mann-Whitney U test. The test compares the performance of the models trained with data from the two membranes or only POPC (upper row, gray background) or POPE:POPG (lower row, no background). The four values reported are: (Upper left) The most statistically significant relation between the distributions; "$>$" if the distribution for the model trained with data of both lipid compositions is stochastically greater, "$<$" if the opposite occurs, and "$\sim$" if no significant difference has been found. (Upper right) The value of the statistical U and (lower right) p-value of the test from which the previous relation has been derived. (Lower left) The overlap between the samples, estimated through their KDEs. The threshold for statistical significance used was $\alpha = 0.05$. An asterisk next to the symbol in the upper left indicates a significant difference in the variances, determined by the Brown-Forsythe test (Table D.24

Data set	Metrics				
	Accuracy	Precision	Recall	$F1$ score	AUC-PR
Martini 2.2	$3.85 \cdot 10^{+0}$	$8.13 \cdot 10^{+0}$	$1.44 \cdot 10^{-1}$	$2.98 \cdot 10^{+0}$	$8.28 \cdot 10^{-1}$
	$p = 5.06 \cdot 10^{-2}$	$p = \mathbf{4.58 \cdot 10^{-3}}$	$p = 7.05 \cdot 10^{-1}$	$p = 8.48 \cdot 10^{-2}$	$p = 3.63 \cdot 10^{-1}$
	$5.41 \cdot 10^{-1}$	$2.64 \cdot 10^{-1}$	$4.26 \cdot 10^{+0}$	$1.97 \cdot 10^{+0}$	$2.35 \cdot 10^{+0}$
	$p = 4.62 \cdot 10^{-1}$	$p = 6.07 \cdot 10^{-1}$	$p = \mathbf{3.96 \cdot 10^{-2}}$	$p = 1.61 \cdot 10^{-1}$	$p = 1.26 \cdot 10^{-1}$
Martini 2.2P	$4.13 \cdot 10^{+0}$	$5.13 \cdot 10^{+0}$	$5.35 \cdot 10^{+0}$	$6.50 \cdot 10^{+0}$	$2.02 \cdot 10^{+1}$
	$p = \mathbf{4.29 \cdot 10^{-2}}$	$p = \mathbf{2.41 \cdot 10^{-2}}$	$p = \mathbf{2.12 \cdot 10^{-2}}$	$p = \mathbf{1.12 \cdot 10^{-2}}$	$p = \mathbf{8.98 \cdot 10^{-6}}$
	$1.69 \cdot 10^{+0}$	$7.82 \cdot 10^{+0}$	$7.30 \cdot 10^{-2}$	$8.05 \cdot 10^{-1}$	$4.01 \cdot 10^{+0}$
	$p = 1.94 \cdot 10^{-1}$	$p = \mathbf{5.43 \cdot 10^{-3}}$	$p = 7.87 \cdot 10^{-1}$	$p = 3.70 \cdot 10^{-1}$	$p = \mathbf{4.60 \cdot 10^{-2}}$
Martini 3	$6.62 \cdot 10^{+0}$	$9.90 \cdot 10^{+0}$	$1.03 \cdot 10^{+1}$	$7.31 \cdot 10^{+0}$	$6.65 \cdot 10^{+0}$
	$p = \mathbf{1.04 \cdot 10^{-2}}$	$p = \mathbf{1.78 \cdot 10^{-3}}$	$p = \mathbf{1.45 \cdot 10^{-3}}$	$p = \mathbf{7.15 \cdot 10^{-3}}$	$p = \mathbf{1.03 \cdot 10^{-2}}$
	$1.39 \cdot 10^{+0}$	$1.25 \cdot 10^{+0}$	$6.79 \cdot 10^{+0}$	$4.42 \cdot 10^{+0}$	$2.53 \cdot 10^{+0}$
	$p = 2.38 \cdot 10^{-1}$	$p = 2.65 \cdot 10^{-1}$	$p = \mathbf{9.54 \cdot 10^{-3}}$	$p = \mathbf{3.61 \cdot 10^{-2}}$	$p = 1.12 \cdot 10^{-1}$

Table D.24: Differences in variance of the performance based on using data from both lipid compositions or one for each force field. Results of the Brown-Forsythe test. The test compares the variance of the samples of performance metrics of the models trained with data from the two membranes or only POPC (upper row, gray background) or POPE:POPG (lower row, no background). The value of the statistical and the p-value of each test are given. The p-values below the significance level of $\alpha = 0.05$ are highlighted; this is considered strong evidence of a difference in variances.

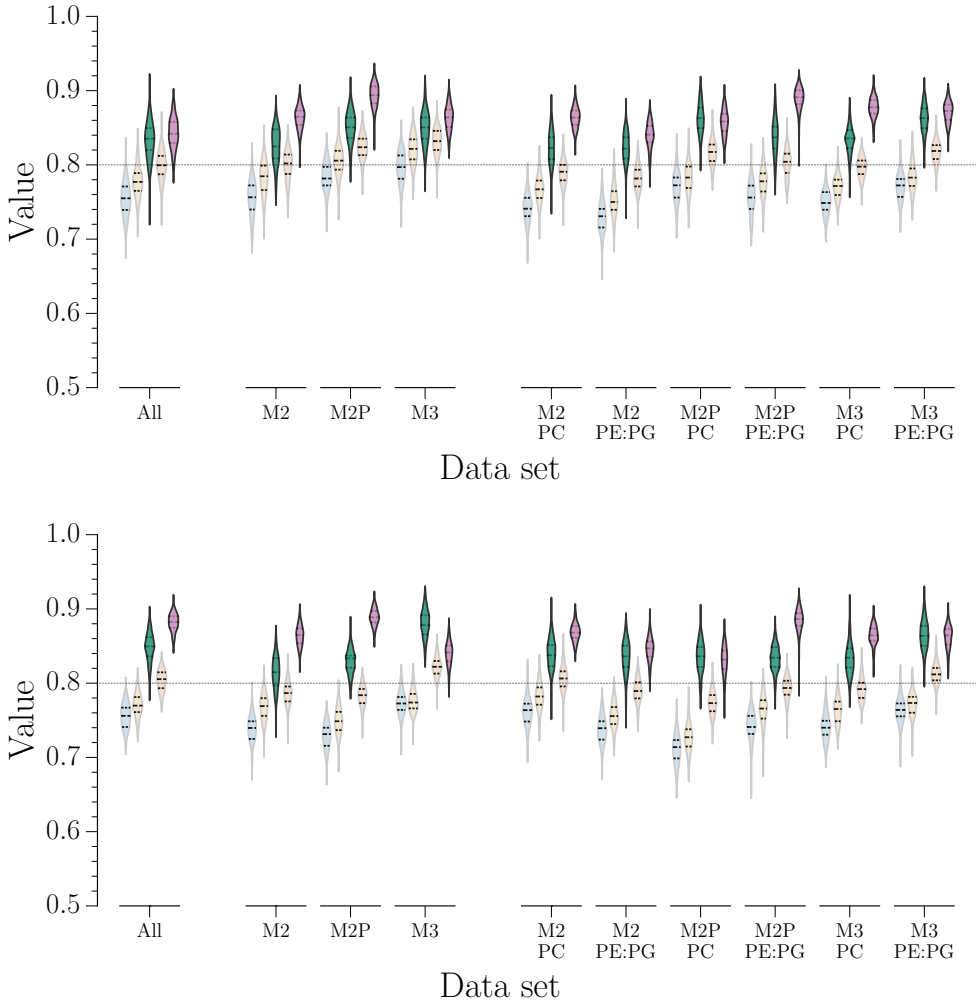


Figure D.2: Evaluation of the performance of the classification models trained with data projected onto subspaces of principal components. A different number of principal components were used, 16 (upper) or 7 (lower). Performance metrics are grouped by data set used to train and test: "All", trained with all simulation data; M2, M2P and M3, trained with all the data from each of these force fields; and the models that used data from each combination of lipid composition and force field (PC for the POPC membrane and PE:PG for the POPE:POPG bilayer). For each data set, the five distributions are ordered; from left to right, accuracy, precision, recall (highlighted, in green), $F1$-score, and area under the precision–recall curve (highlighted, in purple). The Y-axis starts at 0.5 for ease of comparison among the different metrics; a horizontal line at 0.8 is included for visual reference..

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Appendix E.

Rights and permissions

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

E.1 Rights for reproducing content from articles by the author



Figure E.1: License of reproduction for the content of Section 4.2.[310]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

E.2 Rights for reproduced figures from third-party articles

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Oct 07, 2025

This Agreement between Fabián Suárez Lestón ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	6123620120930
License date	Oct 07, 2025
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Microbiology
Licensed Content Title	Antimicrobial peptides: linking partition, activity and high membrane-bound concentrations
Licensed Content Author	Manuel N. Melo et al
Licensed Content Date	Dec 31, 1969
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title of new work	Structure and activity of peptides at the water–lipid membrane interface. Molecular Dynamics insights
Institution name	Universidade de Santiago de Compostela
Expected presentation date	Dec 2025
Portions	Figure 1
The Requesting Person / Organization to Appear on the License	Fabián Suárez Lestón
Requestor Location	Rúa de Jenaro de la Fuente, s/n Santiago de Compostela, 15705 Spain
Payment Type	Invoice
Email Address	fabian.suarez.leston@usc.es
Billing Address	Mr. Fabián Suárez Lestón Rúa de Jenaro de la Fuente, s/n Santiago de Compostela, Spain 15705
Total	0.00 USD
Terms and Conditions	

Figure E.2: License of reproduction for Figure 1.5.[59]

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Bibliography

1. International Union of Pure and Applied Chemistry (IUPAC). (2025). *Lipids*. doi:10.1351/goldbook.L03571.

2. van Meer, G., Voelker, D. R., Feigenson, G. W. (2008). Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.*, **9**(2), 112–124.

3. Eibl, H. (1984). Phospholipids as Functional Constituents of Biomembranes. *Angew. Chemie Int. Ed. English*, **23**(4), 257–271.

4. Franks, F. (1975). The Hydrophobic Interaction. Pp. 1–94. In: Franks, F. (ed.), *Water A Compr. Treatise*. Boston, MA: Springer US.

5. Tanford, C. (1973). *The Hydrophobic Effect: Formation of micelles and biological membranes*. Nashville, TN: John Wiley & Sons.

6. Boggs, J. M. (1987). Lipid intermolecular hydrogen bonding: influence on structural organization and membrane function. *Biochim. Biophys. Acta - Rev. Biomembr.*, **906**(3), 353–404.

7. Raghavan, K., Reddy, M. R., Berkowitz, M. L. (1992). A molecular dynamics study of the structure and dynamics of water between dilauroylphosphatidylethanolamine bilayers. *Langmuir*, **8**(1), 233–240.

8. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.*, **9**(3), 90–95.

9. Israelachvili, J. N., Mitchell, D., Ninham, B. W. (1977). Theory of self-assembly of lipid bilayers and vesicles. *Biochim. Biophys. Acta - Biomembr.*, **470**(2), 185–201.

10. Hishikawa, D., Hashidate, T., Shimizu, T., Shindou, H. (2014). Diversity and function of membrane glycerophospholipids generated by the remodeling pathway in mammalian cells. *J. Lipid Res.*, **55**(5), 799–807.

11. Goñi, F. M., Alonso, A. (2006). Biophysics of sphingolipids I. Membrane properties of sphingosine, ceramides and other simple sphingolipids. *Biochim. Biophys. Acta - Biomembr.*, **1758**(12), 1902–1921.

12. Fahy, E., Subramaniam, S., Brown, H. A., Glass, C. K., Merrill, A. H., *et al.* (2005). A comprehensive classification system for lipids. *J. Lipid Res.*, **46**(5), 839–861.

13. Fahy, E., Subramaniam, S., Murphy, R. C., Nishijima, M., Raetz, C. R., *et al.* (2009). Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.*, **50**, S9–S14.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

14. Marsh, D. (2010). Structural and thermodynamic determinants of chain-melting transition temperatures for phospholipid and glycolipids membranes. *Biochim. Biophys. Acta - Biomembr.*, **1798**(1), 40–51.

15. Heimburg, T. (2019). Phase Transitions in Biological Membranes. Pp. 39–61. In: Demetzos, C., Pippa, N. (eds.), *Thermodynamics and Biophysics of Biomedical Nanosystems: Applications and Practical Considerations*. Singapore: Springer Singapore.

16. Panconi, L., Lorenz, C. D., May, R. C., Owen, D. M., Makarova, M. (2023). Phospholipid tail asymmetry allows cellular adaptation to anoxic environments. *J. Biol. Chem.*, **299**(9), 105134.

17. Cockcroft, S. (2021). Mammalian lipids: structure, synthesis and function. *Essays Biochem.*, **65**(5), 813–845.

18. Stillwell, W. (2016). *An Introduction to Biological Membranes: Composition, Structure and Function*. 2nd edn. London: Elsevier.

19. Harayama, T., Riezman, H. (2018). Understanding the diversity of membrane lipid composition. *Nat. Rev. Mol. Cell Biol.*, **19**(5), 281–296.

20. Opekarová, M., Tanner, W. (2003). Specific lipid requirements of membrane proteins—a putative bottleneck in heterologous expression. *Biochim. Biophys. Acta - Biomembr.*, **1610**(1), 11–22.

21. Hedin, L. E., Illergård, K., Elofsson, A. (2011). An Introduction to Membrane Proteins. *J. Proteome Res.*, **10**(8), 3324–3331.

22. Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., *et al.* (2017). *Molecular Biology of the Cell*. W.W. Norton & Company.

23. Singer, S. J., Nicolson, G. L. (1972). The Fluid Mosaic Model of the Structure of Cell Membranes. *Science*, **175**(4023), 720–731.

24. Karnovsky, M. J., Kleinfeld, A. M., Hoover, R. L., Klausner, R. D. (1982). The concept of lipid domains in membranes. *J. Cell Biol.*, **94**(1), 1–6.

25. Kleanthous, C., Armitage, J. P. (2015). The bacterial cell envelope. *Philos. Trans. R. Soc. B Biol. Sci.*, **370**(1679), 20150019.

26. Vance, J. E. (2015). Phospholipid Synthesis and Transport in Mammalian Cells. *Traffic*, **16**(1), 1–18.

27. Lorent, J. H., Levental, K. R., Ganesan, L., Rivera-Longsworth, G., Sezgin, E., *et al.* (2020). Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.*, **16**(6), 644–652.

28. Clarke, R. J. (2023). Electrostatic switch mechanisms of membrane protein trafficking and regulation. *Biophys. Rev.*, **15**(6), 1967–1985.

29. Maxfield, F. R., van Meer, G. (2010). Cholesterol, the central lipid of mammalian cells. *Curr. Opin. Cell Biol.*, **22**(4), 422–429.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

30. Sohlenkamp, C., Geiger, O. (2016). Bacterial membrane lipids: diversity in structures and pathways. *FEMS Microbiol. Rev.*, **40**(1), 133–159.

31. Rajagopal, M., Walker, S. (2017). Envelope Structures of Gram-Positive Bacteria. Pp. 1–44. In: Bagnoli, F., Rappuoli, R. (eds.), *Protein and Sugar Export and Assembly in Gram-positive Bacteria*. Cham: Springer International Publishing.

32. Furse, S., Wienk, H., Boelens, R., de Kroon, A. I., Killian, J. A. (2015). *E. coli* MG1655 modulates its phospholipid composition through the cell cycle. *FEBS Lett.*, **589**(19B), 2726–2730.

33. Raetz, C. R., Dowhan, W. (1990). Biosynthesis and function of phospholipids in Escherichia coli. *J. Biol. Chem.*, **265**(3), 1235–1238.

34. Reynolds, A. M., Lee, R. E., Costanzo, J. P. (2014). Membrane adaptation in phospholipids and cholesterol in the widely distributed, freeze-tolerant wood frog, Rana sylvatica. *J. Comp. Physiol. B*, **184**(3), 371–383.

35. Seydlova, G., Beranova, J., Bibova, I., Dienstbier, A., Drzmisek, J., Masin, J., Fiser, R., Konopasek, I., Vecerek, B. (2017). The extent of the temperature-induced membrane remodeling in two closely related Bordetella species reflects their adaptation to diverse environmental niches. *J. Biol. Chem.*, **292**(19), 8048–8058.

36. Hassan, N., Anesio, A. M., Rafiq, M., Holtvoeth, J., Bull, I., Haleem, A., Shah, A. A., Hasan, F. (2020). Temperature Driven Membrane Lipid Adaptation in Glacial Psychrophilic Bacteria. *Front. Microbiol.*, **11**, 1–10.

37. van Meer, G. (2011). Dynamic Transbilayer Lipid Asymmetry. *Cold Spring Harb. Perspect. Biol.*, **3**(5), a004671.

38. Bratton, D. L., Fadok, V. A., Richter, D. A., Kailey, J. M., Guthrie, L. A., Henson, P. M. (1997). Appearance of Phosphatidylserine on Apoptotic Cells Requires Calcium-mediated Nonspecific Flip-Flop and Is Enhanced by Loss of the Aminophospholipid Translocase. *J. Biol. Chem.*, **272**(42), 26159–26165.

39. Kobayashi, T., Menon, A. K. (2018). Transbilayer lipid asymmetry. *Curr. Biol.*, **28**(8), R386–R391.

40. Kur, I.-M., Weigert, A. (2024). Phosphatidylserine externalization as immune checkpoint in cancer. *Pflügers Arch. - Eur. J. Physiol.*, **476**(12), 1789–1802.

41. Szlasa, W., Zendran, I., Zalesińska, A., Tarek, M., Kulbacka, J. (2020). Lipid composition of the cancer cell membrane. *J. Bioenerg. Biomembr.*, **52**(5), 321–342.

42. Almeida, I., Magalhães, S., Nunes, A. (2021). Lipids: biomarkers of healthy aging. *Biogerontology*, **22**(3), 273–295.

43. Chew, H., Solomon, V. A., Fonteh, A. N. (2020). Involvement of Lipids in Alzheimer's Disease Pathology and Potential Therapies. *Front. Physiol.*, **11**, 1–28.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

44. Ali, J., Aziz, M. A., Rashid, M. M. O., Basher, M. A., Islam, M. S. (2022). Propagation of age-related diseases due to the changes of lipid peroxide and antioxidant levels in elderly people: A narrative review. *Heal. Sci. Reports*, **5**(3).

45. Conde-Torres, D., Blanco-González, A., Seco-González, A., Suárez-Lestón, F., Cabezón, A., Antelo-Riveiro, P., *et al.* (2024). Unraveling lipid and inflammation interplay in cancer, aging and infection for novel theranostic approaches. *Front. Immunol.*, **15**, 1–18.

46. Huan, Y., Kong, Q., Mou, H., Yi, H. (2020). Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields. *Front. Microbiol.*, **11**, 582779.

47. Harris, F., Dennison, S., Phoenix, D. (2009). Anionic Antimicrobial Peptides from Eukaryotic Organisms. *Curr. Protein Pept. Sci.*, **10**(6), 585–606.

48. Li, S., Wang, Y., Xue, Z., Jia, Y., *et al.* (2021). The structure-mechanism relationship and mode of actions of antimicrobial peptides: A review. *Trends Food Sci. Technol.*, **109**, 103–115.

49. Ma, X., Wang, Q., Ren, K., Xu, T., Zhang, Z., Xu, M., Rao, Z., Zhang, X. (2024). A Review of Antimicrobial Peptides: Structure, Mechanism of Action, and Molecular Optimization Strategies. *Fermentation*, **10**(11), 540.

50. Takahashi, D., Shukla, S. K., Prakash, O., Zhang, G. (2010). Structural determinants of host defense peptides for antimicrobial activity and target cell selectivity. *Biochimie*, **92**(9), 1236–1241.

51. Matsuzaki, K. (1998). Magainins as paradigm for the mode of action of pore forming polypeptides. *Biochim. Biophys. Acta - Rev. Biomembr.*, **1376**(3), 391–400.

52. Koo, Y. S., Kim, J. M., Park, I. Y., Yu, B. J., Jang, S. A., Kim, K.-S., Park, C. B., Cho, J. H., Kim, S. C. (2008). Structure–activity relations of parasin I, a histone H2A-derived antimicrobial peptide. *Peptides*, **29**(7), 1102–1108.

53. Wang, G. (2008). Structures of Human Host Defense Cathelicidin LL-37 and Its Smallest Antimicrobial Peptide KR-12 in Lipid Micelles. *J. Biol. Chem.*, **283**(47), 32637–32643.

54. Panteleev, P. V., Bolosov, I. A., Balandin, S. V., Ovchinnikova, T. V. (2015). Structure and Biological Functions of β-Hairpin Antimicrobial Peptides. *Acta Naturae*, **7**(1), 37–47.

55. Almeida, M. S., Cabral, K. M., Kurtenbach, E., Almeida, F. C., Valente, A. P. (2002). Solution structure of Pisum sativum defensin 1 by high resolution NMR: plant defensins, identical backbone with different mechanisms of action. *J. Mol. Biol.*, **315**(4), 749–757.

56. Falla, T. J., Karunaratne, D. N., Hancock, R. E. (1996). Mode of Action of the Antimicrobial Peptide Indolicidin. *J. Biol. Chem.*, **271**(32), 19298–19303.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

57. Brogden, K. A. (2005). Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.*, **3**(3), 238–250.

58. Travkova, O. G., Moehwald, H., Brezesinski, G. (2017). The interaction of antimicrobial peptides with membranes. *Adv. Colloid Interface Sci.*, **247**, 521–532.

59. Melo, M. N., Ferre, R., Castanho, M. A. (2009). Antimicrobial peptides: Linking partition, activity and high membrane-bound concentrations. *Nat. Rev. Microbiol.*, **7**(3), 245–250.

60. Bucataru, C., Ciobanasu, C. (2024). Antimicrobial peptides: Opportunities and challenges in overcoming resistance. *Microbiol. Res.*, **286**(April), 127822.

61. Gani, Z., Kumar, A., Raje, M., Raje, C. I. (2025). Antimicrobial peptides: An alternative strategy to combat antimicrobial resistance. *Drug Discov. Today*, **30**(2), 104305.

62. Ageitos, J. M., Sánchez-Pérez, A., Calo-Mata, P., Villa, T. G. (2017). Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria. *Biochem. Pharmacol.*, **133**, 117–138.

63. Ahmed, S. K., Hussein, S., Qurbani, K., Ibrahim, R. H., Fareeq, A., Mahmood, K. A., Mohamed, M. G. (2024). Antimicrobial resistance: Impacts, challenges, and future prospects. *J. Med. Surgery, Public Heal.*, **2**, 100081.

64. Naghavi, M., Vollset, S. E., Ikuta, K. S., Swetschinski, L. R., Gray, A. P., *et al.* (2024). Global burden of bacterial antimicrobial resistance 1990–2021: a systematic analysis with forecasts to 2050. *Lancet*, **404**(10459), 1199–1226.

65. Al Musaimi, O. (2025). FDA-Approved Antibacterials and Echinocandins. *Antibiotics*, **14**(2), 166.

66. Xuan, J., Feng, W., Wang, J., Wang, R., Zhang, B., *et al.* (2023). Antimicrobial peptides for combating drug-resistant bacterial infections. *Drug Resist. Updat.*, **68**, 100954.

67. Beschiaschvili, G., Seelig, J. (1990). Melittin Binding to Mixed Phosphatidylglycerol/Phosphatidylcholine Membranes. *Biochemistry*, **29**(1), 52–58.

68. Wieprecht, T., Dathe, M., Beyermann, M., Krause, E., Maloy, W. L., MacDonald, D. L., Bienert, M. (1997). Peptide hydrophobicity controls the activity and selectivity of magainin 2 amide in interaction with membranes. *Biochemistry*, **36**(20), 6124–6132.

69. Oren, Z., Lerman, J. C., Gudmundsson, G. H., Agerberth, B., Shai, Y. (1999). Structure and organization of the human antimicrobial peptide LL-37 in phospholipid membranes: relevance to the molecular basis for its non-cell-selective activity. *Biochem. J.*, **341**(3), 501.

70. Wimley, W. C. (2010). Describing the Mechanism of Antimicrobial Peptide Action with the Interfacial Activity Model. *ACS Chem. Biol.*, **5**(10), 905–917.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

71. Sun, J., Xia, Y., Li, D., Du, Q., Liang, D. (2014). Relationship between peptide structure and antimicrobial activity as studied by de novo designed peptides. *Biochim. Biophys. Acta - Biomembr.*, **1838**(12), 2985–2993.

72. Wang, J., Mura, M., Zhou, Y., Pinna, M., Zvelindovsky, A. V., Dennison, S. R., Phoenix, D. A. (2014). The cooperative behaviour of antimicrobial peptides in model membranes. *Biochim. Biophys. Acta - Biomembr.*, **1838**(11), 2870–2881.

73. Souza, P. F., Marques, L. S., Oliveira, J. T., Lima, P. G., Dias, L. P., Neto, N. A., Lopes, F. E., Sousa, J. S., Silva, A. F., Caneiro, R. F., Lopes, J. L., Ramos, M. V., Freitas, C. D. (2020). Synthetic antimicrobial peptides: From choice of the best sequences to action mechanisms. *Biochimie*, **175**, 132–145.

74. Behera, L. M., Ghosh, M., Gupta, P. K., Rana, S. (2024). A rationally engineered small antimicrobial peptide with potent antibacterial activity. *J. Cell. Biochem.*, **125**(2), 1–26.

75. Barrera, F. N., Fendos, J., Engelman, D. M. (2012). Membrane physical properties influence transmembrane helix formation. *Proc. Natl. Acad. Sci.*, **109**(36), 14422–14427.

76. Sun, S., Zhao, G., Huang, Y., Cai, M., Shan, Y., Wang, H., Chen, Y. (2016). Specificity and mechanism of action of alpha-helical membrane-active peptides interacting with model and biological membranes by single-molecule force spectroscopy. *Sci. Rep.*, **6**(1), 29145.

77. Kelly, S. M., Jess, T. J., Price, N. C. (2005). How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics*, **1751**(2), 119–139.

78. Greenfield, N. J. (2006). Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.*, **1**(6), 2876–2890.

79. Bürck, J., Wadhwani, P., Fanghänel, S., Ulrich, A. S. (2016). Oriented Circular Dichroism: A Method to Characterize Membrane-Active Peptides in Oriented Lipid Bilayers. *Acc. Chem. Res.*, **49**(2), 184–192.

80. Holzwarth, G., Doty, P. (1965). The Ultraviolet Circular Dichroism of Polypeptides. *J. Am. Chem. Soc.*, **87**(2), 218–228.

81. Rodger, A., Marshall, D. (2021). Beginners guide to circular dichroism. *Biochem. (Lond).*, **43**(2), 58–64.

82. Kellenbach, E., Rundlöf, T. (2019). Determination of the identity, content and purity of therapeutic peptides by NMR spectroscopy. Pp. 381–420. In: *Peptide Therapeutics*. RSC drug discovery. Cambridge: Royal Society of Chemistry.

83. Booth, V. (2021). Deuterium Solid State NMR Studies of Intact Bacteria Treated With Antimicrobial Peptides. *Front. Med. Technol.*, **2**, 1–9.

84. Sani, M.-A., Rajput, S., Keizer, D. W., Separovic, F. (2024). NMR techniques for investigating antimicrobial peptides in model membranes and bacterial cells. *Methods*, **224**, 10–20.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

85. Boaro, A., Ageitos, L., Torres, M., Bartoloni, F. H., de la Fuente-Nunez, C. (2020). Light-Emitting Probes for Labeling Peptides. *Cell Reports Phys. Sci.*, **1**(12), 100257.

86. Arias, M., Vogel, H. J. (2017). Fluorescence and Absorbance Spectroscopy Methods to Study Membrane Perturbations by Antimicrobial Host Defense Peptides. Pp. 141–157. In: Hansen, P. R. (ed.), *Antimicrobial Peptides. Methods in Molecular Biology.* Methods in Molecular Biology, vol. 1548. New York, NY: Springer New York.

87. Hedegaard, S. F., Derbas, M. S., Lind, T. K., Kasimova, M. R., Christensen, M. V., Michaelsen, M. H., Campbell, R. A., Jorgensen, L., Franzyk, H., Cárdenas, M., Nielsen, H. M. (2018). Fluorophore labeling of a cell-penetrating peptide significantly alters the mode and degree of biomembrane interaction. *Sci. Rep.*, **8**(1), 6327.

88. Gregory, S. M., Cavenaugh, A., Journigan, V., Pokorny, A., Almeida, P. F. (2008). A Quantitative Model for the All-or-None Permeabilization of Phospholipid Vesicles by the Antimicrobial Peptide Cecropin A. *Biophys. J.*, **94**(5), 1667–1680.

89. Orioni, B., Bocchinfuso, G., Kim, J. Y., Palleschi, A., Grande, G., Bobone, S., Park, Y., Kim, J. I., Hahm, K.-s., Stella, L. (2009). Membrane perturbation by the antimicrobial peptide PMAP-23: A fluorescence and molecular dynamics study. *Biochim. Biophys. Acta - Biomembr.*, **1788**(7), 1523–1533.

90. Moniruzzaman, M., Alam, J. M., Dohra, H., Yamazaki, M. (2015). Antimicrobial Peptide Lactoferricin B-Induced Rapid Leakage of Internal Contents from Single Giant Unilamellar Vesicles. *Biochemistry*, **54**(38), 5802–5814.

91. Hammond, K., Ryadnov, M. G., Hoogenboom, B. W. (2021). Atomic force microscopy to elucidate how peptides disrupt membranes. *Biochim. Biophys. Acta - Biomembr.*, **1863**(1), 183447.

92. Henriksen, J. R., Andresen, T. L. (2011). Thermodynamic profiling of peptide membrane interactions by isothermal titration calorimetry: A search for pores and micelles. *Biophys. J.*, **101**(1), 100–109.

93. Beck, K., Nandy, J., Hoernke, M. (2025). Strong Membrane Permeabilization Activity Can Reduce Selectivity of Cyclic Antimicrobial Peptides. *J. Phys. Chem. B*, **129**(9), 2446–2460.

94. Röthlisberger, U., Sprik, M., Hutter, J. (2002). Time and Length Scales in ab initio Molecular Dynamics. Pp. 413–442. In: Nielaba, P., Mareschal, M., Ciccotti, G. (eds.), *Bridging Time Scales: Molecular Simulations for the Next Decade.* Berlin, Heidelberg: Springer Berlin Heidelberg.

95. Reid, L. M., Verma, C. S., Essex, J. W. (2019). The role of molecular simulations in understanding the mechanisms of cell-penetrating peptides. *Drug Discov. Today*, **24**(9), 1821–1835.

96. Daura, X., Mark, A. E., van Gunsteren, W. F. (1999). Peptide folding simulations: no solvent required? *Comput. Phys. Commun.*, **123**(1-3), 97–102.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

97. Bürgi, R., Daura, X., Mark, A. E., Van Gunsteren, W. F., Bellanda, M., Mammi, S., Peggion, E. (2000). Reversible Peptide Folding in DMSO by Molecular Dynamics Simulation. Pp. 342–343. In: *Molecular Dynamics on Parallel Computers*. Wold Scirntific.

98. Duarte, A. M. S., van Mierlo, C. P. M., Hemminga, M. A. (2008). Molecular Dynamics Study of the Solvation of an α-Helical Transmembrane Peptide by DMSO. *J. Phys. Chem. B*, **112**(29), 8664–8671.

99. Yang, C., Li, J., Li, Y., Zhu, X. (2009). The effect of solvents on the conformations of Amyloid β-peptide (1–42) studied by molecular dynamics simulation. *J. Mol. Struct. THEOCHEM*, **895**(1–3), 1–8.

100. Park, S.-M., Nguyen, P. H., Stock, G. (2009). Molecular dynamics simulation of cooling: Heat transfer from a photoexcited peptide to the solvent. *J. Chem. Phys.*, **131**(18).

101. Schaefer, M., Bartels, C., Karplus, M. (1998). Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *J. Mol. Biol.*, **284**(3), 835–848.

102. Efremov, R. G., Nolde, D. E., Vergoten, G., Arseniev, A. S. (1999). Peptides in membranes: assessment of environmental effects via simulations using an implicit solvation model. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)*, **101**(1-3), 170–174.

103. Ferrara, P., Apostolakis, J., Caflisch, A. (2002). Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins Struct. Funct. Bioinforma.*, **46**(1), 24–33.

104. Bechor, D., Ben-Tal, N. (2001). Implicit Solvent Model Studies of the Interactions of the Influenza Hemagglutinin Fusion Peptide with Lipid Bilayers. *Biophys. J.*, **80**(2), 643–655.

105. Lazaridis, T., Leveritt, J. M., PeBenito, L. (2014). Implicit membrane treatment of buried charged groups: Application to peptide translocation across lipid bilayers. *Biochim. Biophys. Acta - Biomembr.*, **1838**(9), 2149–2159.

106. Lipkin, R. B., Lazaridis, T. (2015). Implicit Membrane Investigation of the Stability of Antimicrobial Peptide β-Barrels and Arcs. *J. Membr. Biol.*, **248**(3), 469–486.

107. Nepal, B., Leveritt, J., Lazaridis, T. (2018). Membrane Curvature Sensing by Amphipathic Helices: Insights from Implicit Membrane Modeling. *Biophys. J.*, **114**(9), 2128–2141.

108. Conde-Torres, D., Mussa-Juane, M., Faílde, D., Gómez, A., García-Fandiño, R., Piñeiro, Á. (2024). Classical Simulations on Quantum Computers: Interface-Driven Peptide Folding on Simulated Membrane Surfaces. *Comput. Biol. Med.*, **182**, 109157.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

109. Yesylevskyy, S., Marrink, S. J., Mark, A. E. (2009). Alternative mechanisms for the interaction of the cell-penetrating peptides penetratin and the TAT peptide with lipid bilayers. *Biophys. J.*, **97**(1), 40–49.

110. Talandashti, R., Mehrnejad, F., Rostamipour, K., Doustdar, F., Lavasanifar, A. (2021). Molecular Insights into Pore Formation Mechanism, Membrane Perturbation, and Water Permeation by the Antimicrobial Peptide Pleurocidin: A Combined All-Atom and Coarse-Grained Molecular Dynamics Simulation Study. *J. Phys. Chem. B*, **125**(26), 7163–7176.

111. Chen, C. H., Starr, C. G., Guha, S., Wimley, W. C., Ulmschneider, M. B., Ulmschneider, J. P. (2021). Tuning of a Membrane-Perforating Antimicrobial Peptide to Selectively Target Membranes of Different Lipid Composition. *J. Membr. Biol.*, **254**(1), 75–96.

112. Nguyen, M. T. H., Biriukov, D., Tempra, C., Baxova, K., Martinez-Seara, H., *et al.* (2022). Ionic Strength and Solution Composition Dictate the Adsorption of Cell-Penetrating Peptides onto Phosphatidylcholine Membranes. *Langmuir*, **38**(37), 11284–11295.

113. Hasannejad-Asl, B., Heydari, S., Azod, F., Pooresmaeil, F., Esmaeili, A., Bolhassani, A. (2024). Peptide-Membrane Docking and Molecular Dynamic Simulation of In Silico Detected Antimicrobial Peptides from Portulaca oleracea's Transcriptome. *Probiotics Antimicrob. Proteins.*

114. Catalina-Hernandez, E., Aguilella-Arzo, M., Peralvarez-Marin, A., Lopez-Martin, M. (2025). Computational Insights into Membrane Disruption by Cell-Penetrating Peptides. *J. Chem. Inf. Model.*, **65**(3), 1549–1559.

115. Ennaceur, S. M., Hicks, M. R., Pridmore, C. J., Dafforn, T. R., Rodger, A., Sanderson, J. M. (2009). Peptide Adsorption to Lipid Bilayers: Slow Processes Revealed by Linear Dichroism Spectroscopy. *Biophys. J.*, **96**(4), 1399–1407.

116. Ulmschneider, M. B., Ulmschneider, J. P. (2008). Folding Peptides into Lipid Bilayer Membranes. *J. Chem. Theory Comput.*, **4**(11), 1807–1809.

117. Shen, L., Bassolino, D., Stouch, T. (1997). Transmembrane helix structure, dynamics, and interactions: multi-nanosecond molecular dynamics simulations. *Biophys. J.*, **73**(1), 3–20.

118. Bernèche, S., Nina, M., Roux, B. (1998). Molecular Dynamics Simulation of Melittin in a Dimyristoylphosphatidylcholine Bilayer Membrane. *Biophys. J.*, **75**(4), 1603–1618.

119. La Rocca, P., Biggin, P. C., Tieleman, D. P., Sansom, M. S. (1999). Simulation studies of the interaction of antimicrobial peptides and lipid bilayers. *Biochim. Biophys. Acta - Biomembr.*, **1462**(1-2), 185–200.

120. Kandasamy, S. K., Larson, R. G. (2006). Molecular dynamics simulations of model trans-membrane peptides in lipid bilayers: A systematic investigation of hydrophobic mismatch. *Biophys. J.*, **90**(7), 2326–2343.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

121. Esteban-Martín, S., Salgado, J. (2007). The dynamic orientation of membrane-bound peptides: Bridging simulations and experiments. *Biophys. J.*, **93**(12), 4278–4288.

122. Muñoz, V., Cerminara, M. (2016). When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem. J.*, **473**(17), 2545–2559.

123. Wang, Y., Zhao, T., Wei, D., Strandberg, E., Ulrich, A. S., Ulmschneider, J. P. (2014). How reliable are molecular dynamics simulations of membrane active antimicrobial peptides? *Biochim. Biophys. Acta - Biomembr.*, **1838**(9), 2280–2288.

124. Jalali, S., Yang, Y., Mahmoudinobar, F., Singh, S. M., Nilsson, B. L., Dias, C. (2022). Using all-atom simulations in explicit solvent to study aggregation of amphipathic peptides into amyloid-like fibrils. *J. Mol. Liq.*, **347**, 118283.

125. Rizo, J., Sari, L., Qi, Y., Im, W., Lin, M. M. (2022). All-atom molecular dynamics simulations of Synaptotagmin-SNARE-complexin complexes bridging a vesicle and a flat lipid bilayer. *Elife*, **11**, 1–28.

126. Yang, Y., Dias, C. L. (2023). Peptide–Membrane Binding: Effects of the Amino Acid Sequence. *J. Phys. Chem. B*, **127**(4), 912–920.

127. Wang, Y., Schlamadinger, D. E., Kim, J. E., McCammon, J. A. (2012). Comparative molecular dynamics simulations of the antimicrobial peptide CM15 in model lipid bilayers. *Biochim. Biophys. Acta - Biomembr.*, **1818**(5), 1402–1409.

128. Lyu, Y., Xiang, N., Zhu, X., Narsimhan, G. (2017). Potential of mean force for insertion of antimicrobial peptide melittin into a pore in mixed DOPC/DOPG lipid bilayer by molecular dynamics simulation. *J. Chem. Phys.*, **146**(15), 155101.

129. Gumbart, J. C., Ulmschneider, M. B., Hazel, A., White, S. H., Ulmschneider, J. P. (2018). Computed Free Energies of Peptide Insertion into Bilayers are Independent of Computational Method. *J. Membr. Biol.*, **251**(3), 345–356.

130. Thompson, P. A., Eaton, W. A., Hofrichter, J. (1997). Laser Temperature Jump Study of the Helix Coil Kinetics of an Alanine Peptide Interpreted with a 'Kinetic Zipper' Model. *Biochemistry*, **36**(30), 9200–9210.

131. Wieprecht, T., Apostolov, O., Beyermann, M., Seelig, J. (1999). Thermodynamics of the α-helix-coil transition of amphipathic peptides in a membrane environment: implications for the peptide-membrane binding equilibrium. *J. Mol. Biol.*, **294**(3), 785–794.

132. Irbäck, A., Mohanty, S. (2005). Folding Thermodynamics of Peptides. *Biophys. J.*, **88**(3), 1560–1569.

133. Wang, T., Wade, R. C. (2007). On the Use of Elevated Temperature in Simulations To Study Protein Unfolding Mechanisms. *J. Chem. Theory Comput.*, **3**(4), 1476–1483.

134. Hungerland, J., Frederiksen, A., Gerhards, L., Solov'yov, I. A. (2022). Studying folding $\leftrightarrow$ unfolding dynamics of solvated alanine polypeptides using molecular dynamics. *Eur. Phys. J. D*, **76**(8), 154.

135. Kästner, J. (2011). Umbrella sampling. *WIREs Comput. Mol. Sci.*, **1**(6), 932–942.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

136. Bernardi, R. C., Melo, M. C., Schulten, K. (2015). Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta - Gen. Subj.*, **1850**(5), 872–877.

137. Hénin, J., Lelièvre, T., Shirts, M. R., Valsson, O., Delemotte, L. (2022). Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living J. Comput. Mol. Sci.*, **4**(1), 1–60.

138. Nymeyer, H., Woolf, T. B., Garcia, A. E. (2005). Folding is not required for bilayer insertion: Replica exchange simulations of an α-helical peptide with an explicit lipid bilayer. *Proteins Struct. Funct. Genet.*, **59**(4), 783–790.

139. Wang, K., Shao, X., Cai, W. (2022). Binding Models of Aβ42 Peptide with Membranes Explored by Molecular Simulations. *J. Chem. Inf. Model.*, **62**(24), 6482–6493.

140. Hopkins, C. W., Le Grand, S., Walker, R. C., Roitberg, A. E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.*, **11**(4), 1864–1874.

141. Conde, D., Garrido, P. F., Calvelo, M., Piñeiro, Á., Garcia-Fandino, R. (2022). Molecular Dynamics Simulations of Transmembrane Cyclic Peptide Nanotubes Using Classical Force Fields, Hydrogen Mass Repartitioning, and Hydrogen Isotope Exchange Methods: A Critical Comparison. *Int. J. Mol. Sci.*, **23**(6), 3158.

142. Guenza, M. G. (2025). Everything You Want to Know About Coarse-Graining and Never Dared to Ask: Macromolecules as a Key Example. *WIREs Comput. Mol. Sci.*, **15**(2).

143. Ingólfsson, H. I., Lopez, C. A., Uusitalo, J. J., de Jong, D. H., Gopal, S. M., Periole, X., Marrink, S. J. (2014). The power of coarse graining in biomolecular simulations. *WIREs Comput. Mol. Sci.*, **4**(3), 225–248.

144. Su, J., Marrink, S. J., Melo, M. N. (2020). Localization Preference of Antimicrobial Peptides on Liquid-Disordered Membrane Domains. *Front. Cell Dev. Biol.*, **8**, 1–11.

145. van Hilten, N., Stroh, K. S., Risselada, H. J. (2022). Efficient Quantification of Lipid Packing Defect Sensing by Amphipathic Peptides: Comparing Martini 2 and 3 with CHARMM36. *J. Chem. Theory Comput.*, **18**(7), 4503–4514.

146. Valério, M., Mendonça, D. A., Morais, J., Buga, C. C., Cruz, C. H., *et al.* (2022). Parainfluenza Fusion Peptide Promotes Membrane Fusion by Assembling into Oligomeric Porelike Structures. *ACS Chem. Biol.*, **17**(7), 1831–1843.

147. Nakatsuji, T., Gallo, R. L. (2012). Antimicrobial Peptides: Old Molecules with New Ideas. *J. Invest. Dermatol.*, **132**(3), 887–895.

148. Mihaylova-Garnizova, R., Davidova, S., Hodzhev, Y., Satchanska, G. (2024). Antimicrobial Peptides Derived from Bacteria: Classification, Sources, and Mechanism of Action against Multidrug-Resistant Bacteria. *Int. J. Mol. Sci.*, **25**(19), 10788.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

149. Dinata, R., Arati, C., Manikandan, B., Abinash, G., Pori, B., *et al.* (2024). Insect Antimicrobial Peptides in Host Defense and Their Therapeutic Application. Pp. 103–135. In: Baindara, P., Mandal, S. M. (eds.), *Evolution of Antimicrobial Peptides: From Self-Defense to Therapeutic Applications*. Cham: Springer Nature Switzerland.

150. Li, J., Hu, S., Jian, W., Xie, C., Yang, X. (2021). Plant antimicrobial peptides: structures, functions, and applications. *Bot. Stud.*, **62**(1), 5.

151. Valero, Y., Saraiva-Fraga, M., Costas, B., Guardiola, F. A. (2020). Antimicrobial peptides from fish: beyond the fight against pathogens. *Rev. Aquac.*, **12**(1), 224–253.

152. Fazly Bazzaz, B. S., Seyedi, S., Hoseini Goki, N., Khameneh, B. (2021). Human Antimicrobial Peptides: Spectrum, Mode of Action and Resistance Mechanisms. *Int. J. Pept. Res. Ther.*, **27**(1), 801–816.

153. Gagat, P., Ostrówka, M., Duda-Madej, A., Mackiewicz, P. (2024). Enhancing Antimicrobial Peptide Activity through Modifications of Charge, Hydrophobicity, and Structure. *Int. J. Mol. Sci.*, **25**(19), 10821.

154. de la Fuente-Núñez, C., Korolik, V., Bains, M., Nguyen, U., Breidenstein, E. B. M., *et al.* (2012). Inhibition of Bacterial Biofilm Formation and Swarming Motility by a Small Synthetic Cationic Peptide. *Antimicrob. Agents Chemother.*, **56**(5), 2696–2704.

155. Barreto-Santamaría, A., Rivera, Z. J., García, J. E., Curtidor, H., Patarroyo, M. E., *et al.* (2020). Shorter Antibacterial Peptide Having High Selectivity for E. coli Membranes and Low Potential for Inducing Resistance. *Microorganisms*, **8**(6), 867.

156. Wang, G., Li, X., Wang, Z. (2016). APD3: the antimicrobial peptide database as a tool for research and education. *Nucleic Acids Res.*, **44**(D1), D1087–D1093.

157. Brahmachary, M. (2004). ANTIMIC: a database of antimicrobial sequences. *Nucleic Acids Res.*, **32**(90001), 586–589.

158. Fjell, C. D., Hancock, R. E., Cherkasov, A. (2007). AMPer: A database and an automated discovery tool for antimicrobial peptides. *Bioinformatics*, **23**(9), 1148–1155.

159. Piotto, S. P., Sessa, L., Concilio, S., Iannelli, P. (2012). YADAMP: yet another database of antimicrobial peptides. *Int. J. Antimicrob. Agents*, **39**(4), 346–351.

160. Waghu, F. H., Barai, R. S., Gurung, P., Idicula-Thomas, S. (2016). CAMP R3 : a database on sequences, structures and signatures of antimicrobial peptides: Table 1. *Nucleic Acids Res.*, **44**(D1), D1094–D1097.

161. Fan, L., Sun, J., Zhou, M., Zhou, J., Lao, X., Zheng, H., Xu, H. (2016). DRAMP: a comprehensive data repository of antimicrobial peptides. *Sci. Rep.*, **6**(1), 24482.

162. Ye, G., Wu, H., Huang, J., Wang, W., Ge, K., *et al.* (2020). LAMP2: a major update of the database linking antimicrobial peptides. *Database*, **2020**, 1–6.

163. Gawde, U., Chakraborty, S., Waghu, F. H., Barai, R. S., Khanderkar, A., *et al.* (2023). CAMPR4: a database of natural and synthetic antimicrobial peptides. *Nucleic Acids Res.*, **51**(D1), D377–D383.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

164. Capecchi, A., Cai, X., Personne, H., Köhler, T., van Delden, C., Reymond, J.-L. (2021). Machine learning designs non-hemolytic antimicrobial peptides. *Chem. Sci.*, **12**(26), 9221–9232.

165. Wan, F., Torres, M. D. T., Peng, J., de la Fuente-Nunez, C. (2024). Deep-learning-enabled antibiotic discovery through molecular de-extinction. *Nat. Biomed. Eng.*, **8**(7), 854–871.

166. Bhangu, S. K., Welch, N., Lewis, M., Li, F., Gardner, B., *et al.* (2025). Machine Learning-Assisted Prediction and Generation of Antimicrobial Peptides. *Small Sci.*, **5**(6).

167. Santos-Júnior, C. D., Torres, M. D., Duan, Y., Rodríguez del Río, Á., Schmidt, T. S., *et al.* (2024). Discovery of antimicrobial peptides in the global microbiome with machine learning. *Cell*, **187**(14), 3761–3778.e16.

168. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., *et al.* (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, **596**(7873), 583–589.

169. McDonald, E. F., Jones, T., Plate, L., Meiler, J., Gulsevin, A. (2023). Benchmarking AlphaFold2 on peptide structure prediction. *Structure*, **31**(1), 111–119.e2.

170. Yue, T., He, J., Li, Y. (2025). Machine-Learning-Assisted Molecular Design of Innovative Polymers. *Accounts Mater. Res.*, **6**(8), 1033–1045.

171. Strothmann, R., Amanpur, M., Neveselý, T., Hecht, S., Reuter, K., Margraf, J. T. (2025). Machine learning driven design of spiropyran photoswitches. *Digit. Discov.*

172. Li, Z., Li, M., Luo, Y., Cao, H., Liu, H., Fang, Y. (2025). Machine learning for accelerated prediction of lattice thermal conductivity at arbitrary temperature. *Digit. Discov.*, **4**(1), 204–210.

173. Kevlishvili, I., Duan, C., Kulik, H. J. (2023). Classification of Hemilabile Ligands Using Machine Learning. *J. Phys. Chem. Lett.*, **14**(49), 11100–11109.

174. Shahbazi, F., Esfahani, M. N., Keshmiri, A., Jabbari, M. (2024). Assessment of machine learning models trained by molecular dynamics simulations results for inferring ethanol adsorption on an aluminium surface. *Sci. Rep.*, **14**(1), 20437.

175. Unke, O. T., Chmiela, S., Sauceda, H. E., Gastegger, M., Poltavsky, I., *et al.* (2021). Machine Learning Force Fields. *Chem. Rev.*, **121**(16), 10142–10186.

176. Mehdi, S., Smith, Z., Herron, L., Zou, Z., Tiwary, P. (2024). Enhanced Sampling with Machine Learning. *Annu. Rev. Phys. Chem.*, **75**(1), 347–370.

177. Brownless, A.-L. R., Rheaume, E., Kuo, K. M., Kamerlin, S. C. L., Gumbart, J. C. (2025). Using Machine Learning to Analyze Molecular Dynamics Simulations of Biomolecules. *J. Phys. Chem. B*, **129**(22), 5375–5385.

178. Batra, R., Loeffler, T. D., Chan, H., Srinivasan, S., Cui, H., *et al.* (2022). Machine learning overcomes human bias in the discovery of self-assembling peptides. *Nat. Chem.*, **14**(12), 1427–1435.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

179. Xu, T., Wang, J., Zhao, S., Chen, D., Zhang, H., *et al.* (2023). Accelerating the prediction and discovery of peptide hydrogels with human-in-the-loop. *Nat. Commun.*, **14**(1), 3880.

180. Hui, T., Descoteaux, M. L., Miao, J., Lin, Y.-S. (2023). Training Neural Network Models Using Molecular Dynamics Simulation Results to Efficiently Predict Cyclic Hexapeptide Structural Ensembles. *J. Chem. Theory Comput.*, **19**(14), 4757–4769.

181. van Teijlingen, A., Edwards, D. C., Hu, L., Lilienkampf, A., Cockroft, S. L., Tuttle, T. (2024). An active machine learning discovery platform for membrane-disrupting and pore-forming peptides. *Phys. Chem. Chem. Phys.*, **26**(25), 17745–17752.

182. Rosenberger, D., Smith, J. S., Garcia, A. E. (2021). Modeling of Peptides with Classical and Novel Machine Learning Force Fields: A Comparison. *J. Phys. Chem. B*, **125**(14), 3598–3612.

183. Alder, B. J., Wainwright, T. E. (1957). Phase Transition for a Hard Sphere System. *J. Chem. Phys.*, **27**(5), 1208–1209.

184. Car, R., Parrinello, M. (1985). Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.*, **55**(22), 2471–2474.

185. Warshel, A., Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.*, **103**(2), 227–249.

186. Vennelakanti, V., Nazemi, A., Mehmood, R., Steeves, A. H., Kulik, H. J. (2022). Harder, better, faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins. *Curr. Opin. Struct. Biol.*, **72**, 9–17.

187. Ahmadi, S., Barrios Herrera, L., Chehelamirani, M., Hostaš, J., Jalife, S., Salahub, D. R. (2018). Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review. *Int. J. Quantum Chem.*, **118**(9), 1–34.

188. Tichauer, R. H., Favre, G., Cabantous, S., Landa, G., Hemeryck, A., Brut, M. (2018). Water Distribution within Wild-Type NRas Protein and Q61 Mutants during Unrestrained QM/MM Dynamics. *Biophys. J.*, **115**(8), 1417–1430.

189. Shannon, C. E. (1949). Communication in the Presence of Noise. *Proc. IRE*, **37**(1), 10–21.

190. Feenstra, K. A., Hess, B., Berendsen, H. J. C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *J. Comput. Chem.*, **20**(8), 786–798.

191. Kamogawa, K., Kitagawa, T. (1986). Raman difference spectroscopy of the C-H stretching vibrations: Frequency shifts and excess quantities for acetone/water and acetonitrile/water solutions. *J. Phys. Chem.*, **90**(6), 1077–1081.

192. Miyamoto, S., Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.*, **13**(8), 952–962.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

193. Hess, B. (2008). P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.*, **4**(1), 116–122.

194. Oostenbrink, C., Villa, A., Mark, A. E., Van Gunsteren, W. F. (2004). A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, **25**(13), 1656–1676.

195. Schmid, N., Eichenberger, A. P., Choutko, A., Riniker, S., Winger, M., Mark, A. E., Van Gunsteren, W. F. (2011). Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.*, **40**(7), 843–856.

196. Fischer, J., Wendland, M. (2023). On the history of key empirical intermolecular potentials. *Fluid Phase Equilib.*, **573**, 113876.

197. Barker, J., Watts, R. (1973). Monte Carlo studies of the dielectric properties of water-like models. *Mol. Phys.*, **26**(3), 789–792.

198. Ewald, P. P. (1921). Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, **369**(3), 253–287.

199. Frenkel, D., Smit, B. (2002). Long-Range Interactions. Pp. 291–320. In: Frenkel, D., Smit, B. (eds.), *Understanding Molecular Simulation*, 2nd edn. San Diego: Academic Press.

200. Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An N · log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**(12), 10089–10092.

201. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.*, **103**(19), 8577–8593.

202. Soczó, B., Pethes, I. (2025). Comparison of water models for structure prediction. *J. Mol. Liq.*, **433**(1), 127975.

203. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., *et al.* (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *J. Chem. Theory Comput.*, **16**(1), 528–552.

204. Dickson, C. J., Walker, R. C., Gould, I. R. (2022). Lipid21: Complex Lipid Membrane Simulations with AMBER. *J. Chem. Theory Comput.*, **18**(3), 1726–1736.

205. Izadi, S., Anandakrishnan, R., Onufriev, A. V. (2014). Building Water Models: A Different Approach. *J. Phys. Chem. Lett.*, **5**(21), 3863–3871.

206. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**(2), 926–935.

207. Karska, N., Zhukov, I., Lipińska, A. D., Rodziewicz-Motowidło, S., Krupa, P. (2023). Why does the herpes simplex 1 virus-encoded UL49.5 protein fail to inhibit the TAP-dependent antigen presentation? *Biochim. Biophys. Acta - Biomembr.*, **1865**(8), 184200.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

208. Knotts, G. M., Lile, S. K., Campbell, E. M., Agee, T. A., Liyanage, S. D., *et al.* (2024). An all-atom model of the human cardiac sodium channel in a lipid bilayer. *Sci. Rep.*, **14**(1), 26857.

209. Chen, S.-Y., Fiedler, M. K., Gronauer, T. F., Omelko, O., von Wrisberg, M.-K., *et al.* (2024). Unraveling the mechanism of small molecule induced activation of Staphylococcus aureus signal peptidase IB. *Commun. Biol.*, **7**(1), 895.

210. Lile, S. K., Campbell, E. M., Gyawu, R. F., Smith, J. A., Gwaltney, S. R., Johnson, C. N. (2025). Potential C-terminal domain interactions of the cardiac voltage gated sodium channel. *Sci. Rep.*, **15**(1), 14100.

211. Engberg, O., Saha Roy, D., Krupa, P., Banerjee, S., Chaudhary, A., *et al.* (2025). Molecules in the Serotonin-Melatonin Synthesis Pathway Have Distinct Interactions with Lipid Membranes. *J. Phys. Chem. B*, **129**(10), 2687–2700.

212. Gonzalez-Sanabria, N., Contreras, G. F., Rojas, M., Duarte, Y., Gonzalez-Nilo, F. D., *et al.* (2025). *The BK channel-NS1619 agonist complex reveals molecular insights on allosteric activation gating.* bioRxiv:10.1101/2025.03.27.645783.

213. Tang, Q., Sinclair, M., Bisignano, P., Zhang, Y., Tajkhorshid, E., Mchaourab, H. S. (2025). *Lipid-mediated mechanism of drug extrusion by a heterodimeric ABC exporter.* bioRxiv:10.1101/2025.02.26.640354.

214. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., *et al.* (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods*, **14**(1), 71–73.

215. Huang, J., MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.*, **34**(25), 2135–2145.

216. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., *et al.* (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B*, **102**(18), 3586–3616.

217. Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, **29**(7), 845–854.

218. Hess, B., van der Spoel, D., Abraham, M. J., Lindahl, E. (2019). On the importance of accurate algorithms for reliable molecular dynamics simulations. *ChemRxiv.*

219. Piggot, T. J., Piñeiro, Á., Khalid, S. (2012). Molecular dynamics simulations of phosphatidylcholine membranes: A comparative force field study. *J. Chem. Theory Comput.*, **8**(11), 4593–4609.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

220. Piggot, T. J., Piñeiro, Á., Khalid, S. (2017). Correction to Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *J. Chem. Theory Comput.*, **13**(4), 1862–1865.

221. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. (1981). Interaction Models for Water in Relation to Protein Hydration. Pp. 331–342. In: Pullman, B. (ed.), *Intermolecular Forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry Held in Jerusalem, Israel, April 13–16, 1981*. Dordrecht: Springer Netherlands.

222. Garrido, P. F., Calvelo, M., Garcia-Fandiño, R., Piñeiro, Á. (2020). Rings, hexagons, petals, and dipolar moment sink-sources: The fanciful behavior of water around cyclodextrin complexes. *Biomolecules*, **10**(3), 431.

223. Haertlein, M., Moulin, M., Devos, J. M., Laux, V., Dunne, O., Trevor Forsyth, V. (2016). Biomolecular Deuteration for Neutron Structural Biology and Dynamics. Pp. 113–157. In: *Methods Enzymol.*, 1st edn., vol. 566. Elsevier Inc.

224. Reif, B. (2022). Deuteration for High-Resolution Detection of Protons in Protein Magic Angle Spinning (MAS) Solid-State NMR. *Chem. Rev.*, **122**(10), 10019–10035.

225. De Meutter, J., Goormaghtigh, E. (2021). Evaluation of protein secondary structure from FTIR spectra improved after partial deuteration. *Eur. Biophys. J.*, **50**(3-4), 613–628.

226. Piszczek, G., Lee, J. C., Tjandra, N., Lee, C.-R., Seok, Y.-J., Levine, R. L., Peterkofsky, A. (2011). Deuteration of Escherichia coli Enzyme INtr alters its stability. *Arch. Biochem. Biophys.*, **507**(2), 332–342.

227. Liu, C., Yao, Z.-S., Tao, J. (2025). Impact of H/D isotopic effects on the physical properties of materials. *Inorg. Chem. Front.*, **12**(3), 876–896.

228. Audi, G., Bersillon, O., Blachot, J., Wapstra, A. H. (2003). The NUBASE evaluation of nuclear and decay properties. *Nucl. Phys. A*, **729**(1), 3–128.

229. Blanco-González, A., Wurl, A., Mendes Ferreira, T., Piñeiro, Á., Garcia-Fandino, R. (2024). Simulating Bacterial Membrane Models at the Atomistic Level: A Force Field Comparison. *J. Chem. Theory Comput.*, **20**(18).

230. de Jong, D. H., Singh, G., Bennett, W. F. D., Arnarez, C., Wassenaar, T. A., *et al.* (2013). Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.*, **9**(1), 687–697.

231. Souza, P. C. T., Alessandri, R., Barnoud, J., Thallmair, S., Faustino, I., *et al.* (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods*, **18**(4), 382–388.

232. Yesylevskyy, S. O., Schäfer, L. V., Sengupta, D., Marrink, S. J. (2010). Polarizable water model for the coarse-grained MARTINI force field. *PLoS Comput. Biol.*, **6**(6), 1–17.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

233. Nolting, W. (2018). *Theoretical Physics 8: Statistical Physics*. Cham: Springer International Publishing.

234. Woodcock, L. (1971). Isothermal molecular dynamics calculations for liquid salts. *Chem. Phys. Lett.*, **10**(3), 257–261.

235. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**(8), 3684–3690.

236. Bussi, G., Donadio, D., Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.*, **126**(1), 14101.

237. Dimelow, R. J., Masters, A. J. (2007). Deriving the distribution of N -particle kinetic energies in a classical system. *Mol. Simul.*, **33**(14), 1165–1166.

238. Tuckerman, M. (2010). *Statistical mechanics: Theory and molecular simulation*. Oxford Graduate Texts. London, England: Oxford University Press.

239. Bernetti, M., Bussi, G. (2020). Pressure control using stochastic cell rescaling. *J. Chem. Phys.*, **153**.

240. Parrinello, M., Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, **52**(12), 7182–7190.

241. Nosé, S., Klein, M. (1983). Constant pressure molecular dynamics for molecular systems. *Mol. Phys.*, **50**(5), 1055–1076.

242. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., Beckstein, O. (2011). MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, **32**(10), 2319–2327.

243. Gowers, R., Linke, M., Barnoud, J., Reddy, T., Melo, M., *et al.* (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. 15th Python Sci. Conf.*, 98–105.

244. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., *et al.* (2020). Array programming with NumPy. *Nature*, **585**(7825), 357–362.

245. McKinney, W. 2010. Data Structures for Statistical Computing in Python. Pp. 56–61. In: *Proceedings of the 9th Python in Science Conference*, vol. 1.

246. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., *et al.* (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*, **17**(3), 261–272.

247. Van Rossum, G. (2020). *The Python Library Reference, release 3.8.2*. Python Software Foundation.

248. Jo, S., Kim, T., Iyer, V. G., Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.*, **29**(11), 1859–1865.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

249. Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., *et al.* (2016). CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.*, **12**(1), 405–413.

250. Lee, J., Hitzenberger, M., Rieger, M., Kern, N. R., Zacharias, M., Im, W. (2020). CHARMM-GUI supports the Amber force fields. *J. Chem. Phys.*, **153**(3).

251. Qi, Y., Ingólfsson, H. I., Cheng, X., Lee, J., Marrink, S. J., Im, W. (2015). CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.*, **11**(9), 4486–4494.

252. Jo, S., Lim, J. B., Klauda, J. B., Im, W. (2009). CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophys. J.*, **97**(1), 50–58.

253. Wu, E. L., Cheng, X., Jo, S., Rui, H., Song, K. C., *et al.* (2014). CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.*, **35**(27), 1997–2004.

254. Tien, M. Z., Sydykova, D. K., Meyer, A. G., Wilke, C. O. (2013). Peptidebuilder: A simple python library to generate model peptides. *PeerJ*, **2013**(1), e80.

255. Bernini, A., Temussi, P. (2011). Nuclear Magnetic Resonance of Amino Acids, Peptides, and Proteins. Pp. 97–153. In: *Amino Acids, Peptides and Proteins in Organic Chemistry*. John Wiley & Sons, Ltd.

256. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., *et al.* (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, **1-2**, 19–25.

257. Park, I. Y., Park, C. B., Kim, M. S., Kim, S. C. (1998). Parasin I, an antimicrobial peptide derived from histone H2A in the catfish, Parasilurus asotus. *FEBS Lett.*, **437**(3), 258–262.

258. Zasloff, M. (1987). Magainins, a class of antimicrobial peptides from Xenopus skin: isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci.*, **84**(15), 5449–5453.

259. Strøm, M. B., Rekdal, Ø., Svendsen, J. S. (2002). Antimicrobial activity of short arginine- and tryptophan-rich peptides. *J. Pept. Sci.*, **8**(8), 431–437.

260. Mechesso, A. F., Su, Y., Xie, J., Wang, G. (2023). Enhanced Antimicrobial Screening Sensitivity Enabled the Identification of an Ultrashort Peptide KR-8 for Engineering of LL-37mini to Combat Drug-Resistant Pathogens. *ACS Infect. Dis.*, **9**(11), 2215–2225.

261. The PLUMED Consortium. (2019). Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods*, **16**(8), 670–673.

262. Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G. (2014). PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.*, **185**(2), 604–613.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

263. Alper, H. E., Levy, R. M. (1989). Computer simulations of the dielectric properties of water: Studies of the simple point charge and transferrable intermolecular potential models. *J. Chem. Phys.*, **91**(2), 1242–1251.

264. Sun, Y., Dong, W., Sun, L., Ma, L., Shang, D. (2015). Insights into the membrane interaction mechanism and antibacterial properties of chensinin-1b. *Biomaterials*, **37**, 299–311.

265. Virtanen, J. A., Cheng, K. H., Somerharju, P. (1998). Phospholipid composition of the mammalian red cell membrane can be rationalized by a superlattice model. *Proc. Natl. Acad. Sci.*, **95**(9), 4964–4969.

266. Kroon, P. C., Grunewald, F., Barnoud, J., van Tilburg, M., Souza, P. C. T., *et al.* (2023). Martinize2 and Vermouth: Unified Framework for Topology Generation. *Elife*, **12**, 1–39.

267. Tucs, A., Tran, D. P., Yumoto, A., Ito, Y., Uzawa, T., Tsuda, K. (2020). Generating Ampicillin-Level Antimicrobial Peptides with Activity-Aware Generative Adversarial Networks. *ACS Omega*, **5**(36), 22847–22851.

268. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., *et al.* (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, **10**(1), 421.

269. Kulke, M., Vermaas, J. V. (2022). Reversible Unwrapping Algorithm for Constant-Pressure Molecular Dynamics Simulations. *J. Chem. Theory Comput.*, **18**(10), 6161–6171.

270. Rodrigues, O. (1840). Des lois géométriques qui régissent les déplacements d'un système solide dans l'espace, et de la variation des coordonnées provenant de ces déplacements considérés indépendamment des causes qui peuvent les produire. *J. Math. Pures Appl.*, **1**, 380–440.

271. Hagberg, A. A., Schult, D. A., Swart, P. J. 2008. Exploring Network Structure, Dynamics, and Function using NetworkX. Pp. 11–15. In: *Proceedings of the 7th Python in Science Conference.*

272. Shinoda, W., Okazaki, S. (1998). A Voronoi analysis of lipid area fluctuation in a bilayer. *J. Chem. Phys.*, **109**(4), 1517–1521.

273. Bhatia, H., Ingólfsson, H. I., Carpenter, T. S., Lightstone, F. C., Bremer, P.-T. (2019). MemSurfer: A Tool for Robust Computation and Characterization of Curved Membranes. *J. Chem. Theory Comput.*, **15**(11), 6411–6421.

274. Guo, Z., Kraka, E., Cremer, D. (2013). Description of local and global shape properties of protein helices. *J. Mol. Model.*, **19**(7), 2901–2911.

275. Kabsch, W., Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**(12), 2577–2637.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

276. Pietrucci, F., Laio, A. (2009). A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.*, **5**(9), 2197–2201.

277. Nishikaw, K., Ooi, T. (1974). Comparison of homologous tertiary structures of proteins. *J. Theor. Biol.*, **43**(2), 351–374.

278. Barnes, C. A., Shen, Y., Ying, J., Takagi, Y., Torchia, D. A., *et al.* (2019). *Remarkable rigidity of the single alpha-helical domain of myosin-VI revealed by NMR spectroscopy.* doi:10.2210/pdb6OBI/pdb.

279. Barnes, C. A., Shen, Y., Ying, J., Takagi, Y., Torchia, D. A., *et al.* (2019). Remarkable Rigidity of the Single α-Helical Domain of Myosin-VI As Revealed by NMR Spectroscopy. *J. Am. Chem. Soc.*, **141**(22), 9004–9017.

280. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., *et al.* (2000). The Protein Data Bank. *Nucleic Acids Research*, **28**(1), 235–242.

281. Fiorin, G., Klein, M. L., Hénin, J. (2013). Using collective variables to drive molecular dynamics simulations. *Mol. Phys.*, **111**(22-23), 3345–3362.

282. Simm, S., Einloft, J., Mirus, O., Schleiff, E. (2016). 50 years of amino acid hydrophobicity scales: revisiting the capacity for peptide classification. *Biol. Res.*, **49**(1), 31.

283. Fauchere, J. L., Pliska, V. (1983). Hydrophobic parameters pi of amino acid side chains from the partitioning of N-acetyl-amino acid amides. *Eur. J. Med. Chem.*, **18**(4), 369–375.

284. Niemelä, P. S., Miettinen, M. S., Monticelli, L., Hammaren, H., Bjelkmar, P., *et al.* (2010). Membrane proteins diffuse as dynamic complexes with lipids. *J. Am. Chem. Soc.*, **132**(22), 7574–7575.

285. Jain, R., Sebastian, K. L. (2017). Diffusing diffusivity: Rotational diffusion in two and three dimensions. *J. Chem. Phys.*, **146**(21), 1–10.

286. Mardia, K. V., Jupp, P. E. (1999). *Directional Statistics.* Wiley Series in Probability and Statistics. Chichester, England: John Wiley & Sons.

287. Branch, M. A., Coleman, T. F., Li, Y. (1999). A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems. *SIAM J. Sci. Comput.*, **21**(1), 1–23.

288. Seabold, S., Perktold, J. 2010. Statsmodels: Econometric and Statistical Modeling with Python. Pp. 92–96. In: van der Walt, S., Millman, J. (eds.), *Proceedings of the 9th Python in Science Conference.*

289. Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Stat.*, **7**(1), 1–26.

290. Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *Ann. Stat.*, **14**(4), 1261–1295.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

291. Shapiro, S. S., Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**(3-4), 591–611.

292. Mohd Razali, N., Bee Wah, Y. (2011). Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Anal.*, **2**(1), 13–14.

293. Levene, H. (1960). Robust Tests for Equality of Variances. Pp. 278–292. In: Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., Mann, H. B. (eds.), *Contibutions to Probab. Stat. Essays Honor Harold Hotell.* Stanford University Press.

294. Brown, M. B., Forsythe, A. B. (1974). Robust Tests for the Equality of Variances. *J. Am. Stat. Assoc.*, **69**(346), 364–367.

295. Student. (1908). The Probable Error of a Mean. *Biometrika*, **6**(1), 1.

296. Bonamente, M. (2022). *Statistics and analysis of scientific data.* Graduate texts in physics. Singapore: Springer Nature Singapore.

297. Mann, H. B., Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Ann. Math. Stat.*, **18**(1), 50–60.

298. Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bull.*, **1**(6), 80–83.

299. Scott, D. W. (2015). *Multivariate density estimation.* 2nd edn. Wiley Series in Probability and Statistics. Chichester, England: Wiley-Blackwell.

300. Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**(1-2), 81–93.

301. Yeo, I.-K. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87**(4), 954–959.

302. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., *et al.* (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, **12**(85), 2825–2830.

303. Zhu, C., Byrd, R. H., Lu, P., Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, **23**(4), 550–560.

304. Canbek, G., Sagiroglu, S., Temizel, T. T., Baykal, N. (2017). Binary classification performance measures/metrics: A comprehensive visualized roadmap to gain new insights. Pp. 821–826. In: *2017 Int. Conf. Comput. Sci. Eng.* IEEE.

305. Dong, W., Dong, Z., Mao, X., Sun, Y., Li, F., Shang, D. (2016). Structure-activity analysis and biological studies of chensinin-1b analogues. *Acta Biomater.*, **37**, 59–68.

306. Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, **6**(60), 3021.

307. Humphrey, W., Dalke, A., Schulten, K. (1996). VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, **14**, 33–38.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

308. Seelig, J. (2004). Thermodynamics of lipid–peptide interactions. *Biochim. Biophys. Acta - Biomembr.*, **1666**(1-2), 40–50.

309. Ge, Y., Hahn, D. F., Mobley, D. L. (2021). A Benchmark of Electrostatic Method Performance in Relative Binding Free Energy Calculations. *J. Chem. Inf. Model.*, **61**(3), 1048–1052.

310. Suarez-Leston, F., Calvelo, M., Tolufashe, G. F., Muñoz, A., Veleiro, U., *et al.* (2022). SuPepMem: A database of innate immune system peptides and their cell membrane interactions. *Comput. Struct. Biotechnol. J.*, **20**, 874–881.

311. Conde-Torres, D., Calvelo, M., Rovira, C., Piñeiro, Á., Garcia-Fandino, R. (2024). Unlocking the specificity of antimicrobial peptide interactions for membrane-targeted therapies. *Comput. Struct. Biotechnol. J.*, **25**, 61–74.

312. Singh, G., Tieleman, D. P. (2011). Using the wimley-white hydrophobicity scale as a direct quantitative test of force fields: The MARTINI coarse-grained model. *J. Chem. Theory Comput.*, **7**(7), 2316–2324.

313. Javanainen, M., Martinez-Seara, H., Vattulainen, I. (2017). Excessive aggregation of membrane proteins in the Martini model. *PLoS One*, **12**(11), e0187936.

314. Majumder, A., Straub, J. E. (2021). Addressing the Excessive Aggregation of Membrane Proteins in the MARTINI Model. *J. Chem. Theory Comput.*, **17**(4), 2513–2521.

315. van Teijlingen, A., Smith, M. C., Tuttle, T. (2023). Short Peptide Self-Assembly in the Martini Coarse-Grain Force Field Family. *Acc. Chem. Res.*, **56**(6), 644–654.

316. Fick, A. (1855). Ueber Diffusion. *Ann. Phys.*, **170**(1), 59–86.

317. Stone, M., Goldbart, P. (2009). *Mathematics for physics.* Cambridge, England: Cambridge University Press.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

The composition of the cell membrane and its physicochemical properties vary across species and conditions. Antimicrobial peptides exploit these differences to selectively attack and destroy pathogens and pathological cells without affecting the host. Despite having been studied experimentally for decades, there are aspects of their interaction with membranes that remain beyond the reach of current methods.

This work investigates antimicrobial peptides from a computational perspective, particularly through Molecular Dynamics simulations and data-driven methods, in order to determine whether the results of systematic simulation of peptide–membrane systems can provide a basis for developing predictive models of the structures that peptides adopt upon membrane interaction and of their potential antimicrobial activity.