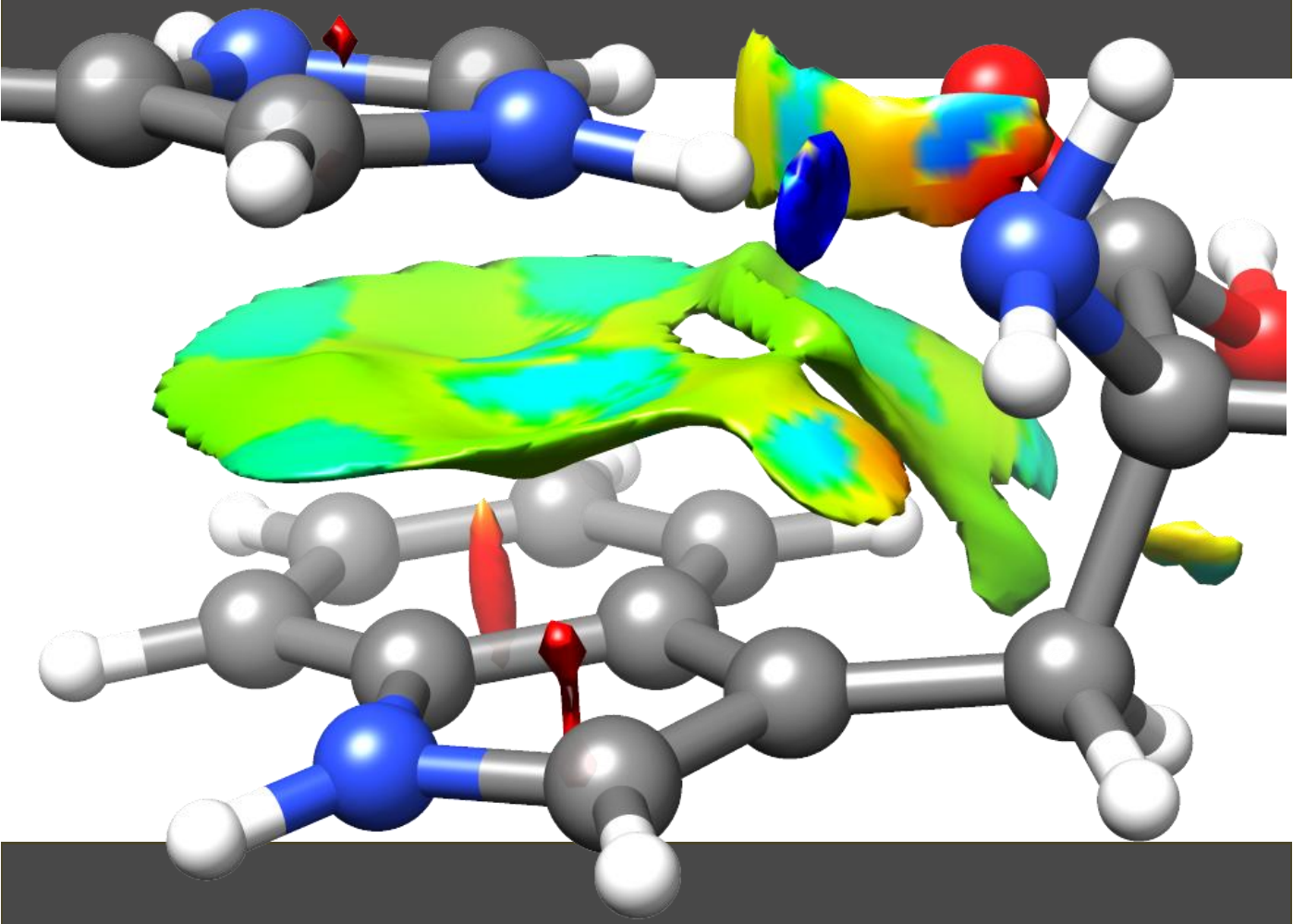
Computational Study of the Capability of the Side Chains of Amino Acids for Setting up Cation···π Interactions Relevant in Protein Stability and Structure



TESIS DE DOCTORADO

Ana Angustias Rodríguez Sanz

Departamento de Química Física
Facultade de Ciencias

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Lugo, xuño 2015

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

FACULTADE DE CIENCIAS

DEPARTAMENTO DE QUÍMICA-FÍSICA

Ph. D. Thesis

Computational study of the capability of the side chains of amino acids for setting up cation···π interactions relevant in protein stability and structure

Ana Angustias Rodríguez Sanz

Lugo, xuño 2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Autorización dos directores da tese

D. Enrique Manuel Cabaleiro Lago
Profesor/a do Departamento: Química-Física (Lugo)

D. Jesús Rodríguez Otero
Profesor/a do Departamento: Química-Física (CIQUS)

Como Directores da Tese de Doutoramento titulada: "*Computational study of the capability of the side chains of amino acids for setting up cation···π interactions relevant in protein stability and structure*".

Presentada por Dna. Ana Angustias Rodríguez Sanz

Alumna do Programa de Doutoramento en Ciencia e Tecnoloxía Química (D1121)

Autorizan a presentación da tese indicada, considerando que reúne os requisitos esixidos no artigo 34 do regulamento de Estudos de Doutoramento, e que como Director da mesma non incurre nas causas de abstención establecidas na lei 30/1992.

Lugo a 22 xuño de 2015

Asdo: Enrique M. Cabaleiro Lago

Asdo: Jesús Rodríguez Otero

Asdo: Ana Angustias Rodríguez Sanz

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Para mi abuela,

para mis padres,

para mis hermanos

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

En primer lugar me gustaría mostrar mi agradecimiento a mis directores de tesis, Jesús y Quique. Todavía recuerdo aquella conversación vía Skype que me dio tan buena impresión del proyecto y de vosotros como personas (aunque os empeñarais en quitarle importancia…). Gracias por hacer esto posible y sobre todo, gracias Quique por estar ahí siempre, con tu paciencia infinita a todos mis despistes y por todo lo que me enseñaste, por tu ayuda incondicional e inestimable.

Por supuesto doy las gracias también al Ministerio de Economía y Competitividad por la beca FPI, así como la beca de Estancias Breves, en Dublín, donde viví unos meses intensos y conocí gente a la que admiro. En primer lugar quiero dar las gracias a mi jefa Isabel, que sin conocerme me dio todo su apoyo y confianza en mi trabajo desde que entré en su grupo hasta que salí el último día. Mis compañeros de grupo también fueron un gran ejemplo a seguir. Además de ser unas personas maravillosas, su pasión por el trabajo y la ciencia me animaron a continuar. Gracias Elena, Viola, Patrick, Julian, Michela…y todas aquellas personas con las que me crucé en aquella estancia y que me apoyaron y ayudaron tanto, dentro y fuera del laboratorio.

Gracias también a mi compañera de despacho, Alba, a la que tanto le debo, y que sin ella todo habría sido mucho más complicado, ¡y por esos encargos de baberos! Este año, apareció una nueva compañera andaluza, Belén, que también hace mis horas frente al ordenador más amenas. Ortigueira, cómo triunfaste con tu mujer, ¡vaya pareja más chula que hacéis!

También quiero darle las gracias a Stu y Ronan, que aunque nos separen miles de kilómetros, os voy a tener ahí, no sólo para acribillaros a preguntas de inglés, si no que sé que tengo dos buenos amigos.

Por supuesto tengo que agradecer el apoyo constante de mi familia, mis padres, mis hermanos y mi abuela, que siempre estarán ahí para lo que necesite, igual que yo para ellos aunque estemos en la distancia. ¡Qué sería yo sin vosotros!

También tengo que agradecer el apoyo de mis amigos de Málaga, Bea, Paco, Rocío, Inma, Samantha… que aunque no estemos continuamente en contacto, cuando nos vemos es como si no hubieran pasado los días. ¡Por esas rutas en bici por la playa y esos viajes por el norte!

El día que llegué a Lugo sabía que empezaba una nueva etapa lejos de mi familia y amigos de siempre. Sin embargo, tuve la suerte de encontrar gente increíble desde el primer momento. Empezando por todas mis compañeras de piso que siempre han estado ahí para escuchar mis aventuras y desventuras con la tesis, aguantar mis cambios de ánimo y aún así darme todo su apoyo. Este último año he estado de maravilla, gracias Cris, Esther, Yoli e Iria. Y de entre todas

mis compañeras, sé que me llevo unas amigas de verdad para toda la vida, así que gracias Zoraida, Marina, Iria… me he sentido como en casa con vosotras; luego está Alfonso, que empezó dándome consejos sobre plantas carnívoras y prestándome su linterna cuando salía tarde de trabajar, y ahora vivo enganchada a sus historias con las que no puedo parar de reír; ¡y también están cómo no las Marquesitas!: Carmen, mi compi de defensa y rutas en bici por el río; María mi compi de aventuras por tierras hostiles; Patri, mi compi de viaje y mercadillos; y Elba, la consejera amorosa por excelencia. Es una suerte haberos conocido y que ahora forméis parte de mi vida, sólo espero que nunca salgáis de ella. Ya sabéis que allá donde vaya seréis bienvenidos.

También quería agradecer a las renombradas Morconas Morunas esas clases de baile y alguna cena que otra tan divertida que te hacen olvidar por unas horas todos los problemas. Y a Mar, por esos ratos de risas en la conserjería junto con el de las historias de antes. Y bueno, por qué no, también le doy las gracias a esos bichitos peludos de Nala, Bimba y Trufa, que siempre estaban ahí esperando detrás de la puerta cuando llegaba a casa para hacerme compañía. Incluso algunos aportando sus ideas en modelización…

En definitiva, agradecer a todas las personas que aún sin saberlo, han aportado su granito de arena con su apoyo para ayudarme a no abandonar y seguir adelante de una forma u otra.

Index

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Resumen

Origen y motivación del estudio

El propósito fundamental de esta tesis es seguir proporcionando información detallada sobre las características y naturaleza de interacciones intermoleculares de interés. Los estudios desarrollados proporcionan información energética, estructural y metodológica sobre determinadas interacciones no covalentes: en particular, sistemas en los que se ven involucradas especies aromáticas y catiónicas de modo que se establecen interacciones catión···π.

Las interacciones intermoleculares en general desempeñan funciones clave a nivel biológico, interviniendo en distintos tipos de procesos físicos y químicos, entre los que se pueden destacar el plegamiento de proteínas, los procesos de reconocimiento molecular o la actividad enzimática. En particular, las interacciones no covalentes en las que se ve involucrado un anillo aromático son bastante frecuentes en los sistemas biológicos, desempeñando un importante papel tanto en la estabilización del sistema como en la conformación de su estructura. En las proteínas es común encontrar aminoácidos aromáticos anclados en sitios activos a pesar de que tan sólo hay 4 aminoácidos esenciales con grupos aromáticos en su estructura dentro de un conjunto de 20. Esta mayor prevalencia parece indicar la importancia que los aminoácidos aromáticos tienen en las proteínas a la hora de interactuar con su entorno. Por otra parte, también es común encontrar aminoácidos protonados en sus cadenas laterales debido a que, a pH fisiológico, tienden a actuar como bases. Ejemplos de estos aminoácidos son la arginina, la cisteína, la prolina, la lisina, la tirosina o la histidina, cuyos pKa son 12.5, 8.3, 10.6, 10.5, 10.1 y 6.0, respectivamente. Estos valores de pKa pueden variar dependiendo del entorno y de otros grupos ionizables próximos, pero sirven como orientación para su clasificación. En esta tesis se ha centrado la atención principalmente en las interacciones entre aminoácidos aromáticos como son la fenilalanina, la tirosina, la histidina y el triptófano, y aminoácidos catiónicos con grupo amonio, la arginina, la prolina y la lisina. Según datos aportados por Gallivan y Dougherty a partir del estudio de 593 secuencias diferentes de proteínas en el 'Protein Data Bank', las interacciones catión···π son bastante frecuentes, encontrándose en una proporción media de una interacción por cada 77 residuos. Además, también se desprende de este estudio que la presencia de arginina en este tipo de interacciones es más probable que la de lisina, y que la probabilidad de participación de los aminoácidos aromáticos sigue el orden de prioridad Trp > Tyr > Phe. Tanto es así que se determinó que aproximadamente un cuarto de los triptófanos presentes estaban participando en interacciones catión···π. Es precisamente debido a esta funcionalidad que las interacciones intermoleculares han estado siempre en el punto de mira de numerosas investigaciones. Considerando entonces la importancia que tienen este tipo de interacciones a nivel biológico, si se consiguiera tener control sobre sus características y propiedades, se podrían condicionar

determinados procesos, por ejemplo mediante el uso de fármacos específicos o mediante el uso de procesos catalíticos específicos.

Para la realización de este estudio se han usado métodos computacionales, en concreto métodos derivados de la química cuántica. Muchas veces no es factible llevar a cabo este tipo de estudios tan específicos en los que se intentan caracterizar y cuantificar interacciones aisladas mediante la experimentación y se tiende, para estos casos, al análisis conjunto teórico-experimental o puramente teórico. Así, los métodos teóricos permiten una descripción más detallada de los fenómenos subyacentes, pudiéndose estudiar un sistema en concreto, permitiendo aislar el efecto de una interacción concreta respecto del valor global del sistema. Entre los métodos usados en el presente trabajo se encuentran métodos *ab initio* basados en la función de onda del sistema, tales como MP2 o CCSD(T), así como métodos basados en la teoría del funcional de la densidad (DFT). En este último caso se han empleado principalmente métodos corregidos para tratar la dispersión debido a que los métodos DFT no corregidos tienen ciertas dificultades para describirla. Fundamentalmente, se han usado nuevos funcionales como M06-2X, que proporciona descripciones bastante aceptables de las interacciones no covalentes. Además del cálculo de la propia energía de interacción, también se ha empleado la teoría de perturbación adaptada en simetría basada en DFT, SAPT(DFT), con lo que se obtienen las distintas contribuciones a la energía de interacción como una serie de términos con una naturaleza física bien determinada: las contribuciones de repulsión, dispersión, inducción y electrostática. Finalmente, también se han aplicado técnicas de análisis de la densidad electrónica para obtener más información acerca de las características de la interacción.

Cabe destacar que aunque este tipo de interacciones catión···π es bastante común en sistemas biológicos y se encuentra con frecuencia en sitios clave para determinados procesos, existe cierta discrepancia entre diferentes autores sobre la intensidad de las mismas. Mientras algunos autores afirman que las interacciones catión···π pueden llegar a adquirir una elevada intensidad, otros parecen indicar que son interacciones más débiles. Estas diferencias se deben en parte a que los estudios no han tenido en consideración el efecto del medio, y en particular el disolvente. Cuando éste ha sido tenido en cuenta los resultados parecen indicar que las interacciones catión···π pueden incluso llegar a ser más intensas que los puentes salinos, aportando una estabilización de hasta 5.5 kcal·mol^{-1}.

La estructura de las proteínas presenta distintas zonas de exposición al disolvente, desde las totalmente expuestas al mismo, hasta las enterradas por completo en la región hidrofóbica. A medio camino entre estos dos extremos, se encuentran cavidades de diferentes tamaños expuestas al medio. De esta forma, dependiendo del tamaño de la cavidad, la superficie de la proteína podrá estar en contacto con mayor o menor número de moléculas de agua, pudiendo modificar las

características de las interacciones no covalentes entre las especies que se encuentren próximas, y en particular las de la interacción catión···π. Además, también sería posible que la estructura de las especies interaccionantes se modificara dependiendo del grado de exposición al disolvente. A medida que se van introduciendo moléculas de disolvente, las distintas moléculas de disolvente compiten con la nube aromática por el catión, de modo que la interacción es cada vez más débil.

Estas evidencias constituyen el punto de partida de este proyecto de tesis, en el que se pretende, entre otras cosas, estudiar el efecto que el entorno podía provocar sobre las interacciones catión···π. Es por esto que uno de los aspectos que más ampliamente se ha estudiado en esta tesis ha sido el efecto de la introducción progresiva de moléculas de agua al sistema en el que aparece la interacción catión···π. El estudio de estos procesos se puede plantear de diversas maneras: una de ellas consiste en generar progresivamente nuevos sistemas con una molécula de disolvente más, a partir de las estructuras más estables obtenidas para un determinado nivel de microhidratación. La nueva molécula de agua se añade en diferentes posiciones, siguiendo la intuición química; otra alternativa consiste en hacer una exploración del espacio conformacional de cada uno de los sistemas hidratados para así obtener una muestra representativa de las estructuras más favorables en cada caso, que se emplearán como punto de partida de cálculos posteriores. El conjunto de estructuras obtenido se criba por medio de la energía para reducir la enorme cantidad de estructuras posibles y quedarnos con los sistemas más estables y por lo tanto los más probables. Todas las estructuras seleccionadas de esta forma son sometidas a cálculos mecano-cuánticos de optimización por los que se obtienen estructuras y energías con una mayor calidad.

En los siguientes apartados se hará un breve resumen de cada uno de los trabajos realizados en esta tesis, que incluirá también las principales conclusiones a las que se llegó en cada uno de los casos.

Efecto de la microhidratación sobre las interacciones metilamonio···fenol y amonio···fenol

Este estudio, correspondiente al capítulo 4, se basa en la microhidratación progresiva de complejos formados por un catión y una molécula de fenol, de modo que se van introduciendo hasta tres moléculas de agua para así poder valorar el efecto de la presencia de este número reducido de moléculas de disolvente sobre las características de la interacción catión···π. Existen algunos trabajos previos que tratan a cerca de de la interacción de cationes simples, principalmente alcalinos, con benceno. La interacción con benceno es relativamente simple ya que sólo ofrece una localización para la interacción favorable con cationes. El fenol, perteneciente a la cadena lateral de la tirosina, tiene por el contrario la peculiaridad de que además del sistema aromático presenta un grupo hidroxilo que puede actuar como segundo punto de anclaje para la interacción con el catión, aumentando así el número de posibles complejos

catión···π formados. Los cationes elegidos para este trabajo han sido los cationes amonio y metilamonio, en un intento de modelizar las características de la interacción con los residuos catiónicos presentes en la cadena lateral de la lisina. La utilización del catión metilamonio, con posibilidad de interaccionar tanto a través del grupo amonio como a través del grupo metilo, incrementa la complejidad del sistema estudiado. Por lo tanto, en este primer trabajo se trata de determinar las características de la interacción amonio···fenol como una primera aproximación al comportamiento de los contactos entre cadenas laterales de lisina y tirosina.

El procedimiento seguido en el estudio ha sido la microhidratación progresiva de los complejos. Como se ha indicado anteriormente, se obtienen las estructuras de los complejos más estables entre cationes amonio y fenol en ausencia de moléculas de agua; empleando estas estructuras, se introduce una primera molécula de agua en las posiciones más favorables y se obtienen las estructuras optimizadas más favorables para los complejos monohidratados. El proceso se continúa introduciendo de modo análogo las siguientes moléculas de agua.

La introducción de sucesivas moléculas de agua al sistema aumenta significativamente el número de posibles complejos, ya que se incrementa en gran medida la complejidad de la superficie de energía potencial. Sin embargo, los mínimos encontrados para ambos complejos, tanto con amonio como metilamonio, presentan características parecidas. Los complejos más estables localizados con ambas especies muestran un patrón cíclico de interacción muy similar, en el que suelen estar involucrados el sistema aromático, el grupo hidroxilo del fenol, el catión y al menos una molécula de agua. Este patrón cíclico se puede observar desde la introducción de la primera molécula de agua y permanece con la inclusión del resto de moléculas de agua. Por tanto, los resultados parecen indicar que al menos la primera molécula de agua interacciona de forma específica, formando un ciclo de enlaces de hidrógeno especialmente estable y que sobrevive tras la adición de más moléculas de agua.

A medida que se van introduciendo moléculas de agua, las diferencias de estabilidad entre complejos con diferentes estructuras van decreciendo. Así, la estabilización es mayor para la primera molécula de agua, mientras que la inclusión del resto de moléculas de agua provoca una estabilización cada vez menos intensa. Esto se debe a que la primera molécula de agua es capaz de interaccionar simultáneamente tanto con el catión como con la molécula de fenol. Después de este paso, la carga del catión se ha reducido en parte por lo que una segunda molécula de agua interaccionará con éste con menor intensidad. Sin embargo, el grupo hidroxilo también es capaz de interaccionar con nuevas moléculas de agua, involucrando al grupo OH en un nuevo enlace de hidrógeno con el oxígeno del agua de tipo φ-OH···O. Por lo tanto, ya a partir de la segunda molécula de agua, cada vez va siendo menos relevante el lugar que ocupa una nueva molécula de agua, igualándose cada vez más las estabilidades obtenidas por interacción directa con el catión o

mediante enlaces de hidrógeno con otras moléculas de agua o el fenol. La participación del grupo hidroxilo del fenol en la red de enlaces de hidrógeno formada es significativa.

Efecto de la microhidratación sobre la interacción guanidinio···π

Al igual que en el trabajo anterior, en este estudio, correspondiente al capítulo 5, se ha llevado a cabo un proceso de microhidratación gradual, pero en este caso en complejos formados por el catión guanidinio con diferentes sistemas aromáticos. Como se mencionó anteriormente, los aminoácidos catiónicos más básicos son la lisina, la prolina y la arginina. Una vez realizado el estudio anterior con los cationes amonio y metilamonio pertenecientes a la lisina, se ha considerado hacer un trabajo similar en el que el catión interaccionante fuera el guanidinio, localizado en la cadena lateral del aminoácido arginina. Sin embargo, a diferencia del trabajo anterior, se incluyeron en el estudio las moléculas de benceno, fenol e indol (pertenecientes a las cadenas laterales de fenilalanina, tirosina y triptófano respectivamente). En un estudio previo de nuestro grupo de investigación, se observa que la incorporación de moléculas de agua a un complejo formado por catión guanidinio y benceno provoca un descenso apreciable en la intensidad de la interacción catión···π, pero promueve además un cambio en la estructura de los complejos, de modo que se vuelven más favorables estructuras en las que el guanidinio se sitúa en paralelo al anillo aromático. Por otra parte, tanto fenol como indol ofrecen una segunda localización favorable para la interacción con el catión, que en el caso del indol se trata de un segundo anillo aromático de cinco miembros en el que uno de los átomos es nitrógeno.

Las estructuras más estables en ausencia de moléculas de agua presentan en todos los casos al catión guanidinio situado perpendicularmente a la especie aromática e interaccionando con la misma por medio de sus grupos NH_2, estableciendo diferentes enlaces de hidrógeno. Los complejos con benceno sólo presentan un mínimo, mientras que con fenol e indol surgen más posibilidades debido a la participación del grupo hidroxilo del fenol y en el anillo pirrólico del indol. En todo caso, los resultados muestran que la interacción se va haciendo más intensa a medida que pasamos de benceno, a fenol y a indol.

El análisis de la naturaleza de la interacción mediante cálculos SAPT(DFT) revela los motivos de estas diferentes estabilidades. Como era de esperar, la contribución más significativa a la estabilidad de los complejos proviene del término electrostático, pero se registran importantes contribuciones de la inducción y la dispersión a la estabilidad de los complejos. La contribución electrostática aumenta al pasar de benceno a fenol o indol debido a que se trata de especies polares. Asimismo, la inducción aumenta ya que se trata de especies más polarizables. Por último, en complejos con indol se registra un aumento significativo de la dispersión asociado a la mayor extensión de la nube aromática. Por lo tanto, el incremento de estabilidad observado se debe a incrementos en todas las contribuciones estabilizantes. La diferencia de estabilidades entre

los complejos con diferentes especies aromáticas no varía significativamente incluso con la inclusión de moléculas de agua. Sin embargo, estructuralmente adoptan patrones de interacción muy distintos, aumentando también drásticamente el número de estructuras posibles a medida que vamos introduciendo moléculas de agua. La estabilidad que se gana al introducir una nueva molécula de agua es aproximadamente la misma con las tres especies aromáticas, así que la diferencia de estabilidad de los complejos se debe principalmente a las diferencias de intensidad de la interacción guanidinio···π.

Interacción catión···π y efectos de la microhidratación en complejos formados por el catión pirrolidinio y especies aromáticas de las cadenas laterales de aminoácidos aromáticos

En el capítulo 6 se detalla este tercer y último estudio de microhidratación, en el que al igual que en los dos anteriores, se introducen gradualmente moléculas de agua, hasta un máximo de tres, a los sistemas formados por el catión pirrolidinio y diferentes especies aromáticas para evaluar de nuevo los efectos provocados sobre la interacción catión···π. El catión pirrolidinio se escogió siguiendo el mismo criterio que para los trabajos anteriores: es una especie fácilmente encontrada en los sistemas biológicos y en particular en el aminoácido prolina. Así, con este estudio se cerraría la sección dedicada a la microhidratación en la que se han tenido en cuenta algunos de los cationes más básicos pertenecientes a las cadenas laterales de los aminoácidos esenciales. De nuevo se usaron benceno, fenol e indol como especies aromáticas para caracterizar las interacciones catión···π.

Al igual que se observó para el caso del catión guanidinio, la interacción catión···π es más intensa con el indol, seguida de la interacción con el fenol. Así, la interacción benceno···catión es la menos favorecida de las tres ya que tan sólo puede establecer un punto de interacción a través de su nube electrónica. Por el contrario, los dos puntos de anclaje disponibles en el fenol e indol hacen posible la formación de una gran variedad de estructuras diferentes. Estos resultados se obtienen con independencia del método de cálculo usado. Dicho comportamiento también viene reflejado por los valores obtenidos mediante SAPT(DFT), que indican que aunque la interacción para las tres especies aromáticas es principalmente electrostática, dispersión e inducción también tienen una contribución no despreciable, principalmente para los complejos con indol. La interacción con fenol ya posee un carácter electrostático y de inducción más acusado que con el benceno debido a su grupo hidroxilo y su momento dipolar permanente. En el caso del indol, aunque también hay una importante contribución de la inducción, la contribución de la dispersión se hace visiblemente notable gracias a su sistema aromático extendido.

Las estructuras más estables observadas suelen presentar la interacción catión···π entre el grupo NH del pirrolidinio y las nubes electrónicas de los sistemas aromáticos. Sin embargo, hay

algunas excepciones en los complejos con fenol, en el que varias de las estructuras más estables presentan interacciones NH···O con el hidroxilo del sistema. Además, también se observan interacciones secundarias CH···π en bastantes casos en los complejos con fenol e indol, adoptando el catión en estos casos disposiciones paralelas respecto del anillo aromático. La introducción sucesiva de moléculas de agua en cada uno de los sistemas provoca una gran variedad de mínimos y en muchos casos con energías de coordinación muy similares, lo que indica que diferentes disposiciones dan lugar a una estabilidad similar. Además, aunque para el caso del benceno el patrón de interacción permanece más o menos simple, para los complejos con indol y fenol se hace más complejo. Sin embargo, sí que se puede observar que en la mayoría de los casos se conserva un patrón cíclico en el que están involucrados ambos puntos de anclaje para fenol e indol, con los que el catión normalmente se encuentra interaccionando, salvo alguna excepción. La inclusión de moléculas de agua también afecta a la estabilidad relativa entre los complejos de fenol e indol. Mientras que en ausencia de moléculas de agua los complejos con indol son claramente más estables que los de fenol, ya con dos moléculas de agua, las diferencias de estabilidad son prácticamente inexistentes.

Finalmente, es interesante mencionar también que entre los diferentes métodos de cálculo usados, los que proporcionan una mejor relación precisión/coste computacional son el M06-2X/6-31+G* y el SCSN-MP2/AVDZ, siendo el primero especialmente indicado para los complejos no hidratados.

Interacción entre el cation guanidinio y aminoácidos aromáticos

Una vez concluido el estudio de microhidratación se propuso aumentar el tamaño del sistema con la inclusión de aminoácidos completos (Trp, Phe y Tyr), de modo que se pudieran evaluar las características de la interacción catión···π en un sistema más complejo. Se han tenido en cuenta tanto las estructuras zwitteriónicas como las que tienen los grupos amino y ácido en forma neutra. Para este trabajo se volvió a escoger el guanidinio como catión debido a su relevancia y a su peculiar estructura plana, que abre la posibilidad para la formación de diferentes tipos de interacciones.

Los resultados indican que las interacciones entre el guanidinio y los aminoácidos Tyr y Phe, tanto en su forma zwitteriónica como neutra, son muy similares en estabilidad, mientras que la interacción con el Trp es algo más intensa. En todos los casos, los complejos más estables presentan interacciones cation···π, si bien se han localizado complejos con la forma zwitteriónica del amino ácido que son prácticamente tan estables como los formados con la forma neutra. El análisis de los distintos componentes de la energía de interacción muestra que los complejos con el amino ácido plegado de modo que permita el contacto catión···π son los más favorecidos debido a que el contacto catión···π aporta mayor estabilización en forma de dispersión e

inducción. Este aporte extra de estabilidad debido al contacto catión···π es capaz de superar el coste energético necesario para plegar el amino ácido, que es menor en otras estructuras más extendidas. Por lo tanto, para describir correctamente este tipo de interacciones también es necesaria una evaluación apropiada de las distintas contribuciones a la estabilidad, ya que la estabilidad final resulta del balance de contribuciones que a menudo operan en sentidos opuestos.

En este trabajo se ha probado una batería de métodos de cálculo de diversa naturaleza, proporcionando todos ellos resultados bastante apropiados. En todo caso, los mejores resultados se obtienen con el método MP2.X, que es el único capaz de reproducir los valores de estabilidad relativa entre formas zwitteriónicas y neutras de los complejos. Respecto a otros métodos menos costosos, tanto M06-2X, B3LYP-D como SCSN-MP2/aug-cc-pVDZ son las mejores elecciones si tenemos en cuenta la relación precisión/coste computacional.

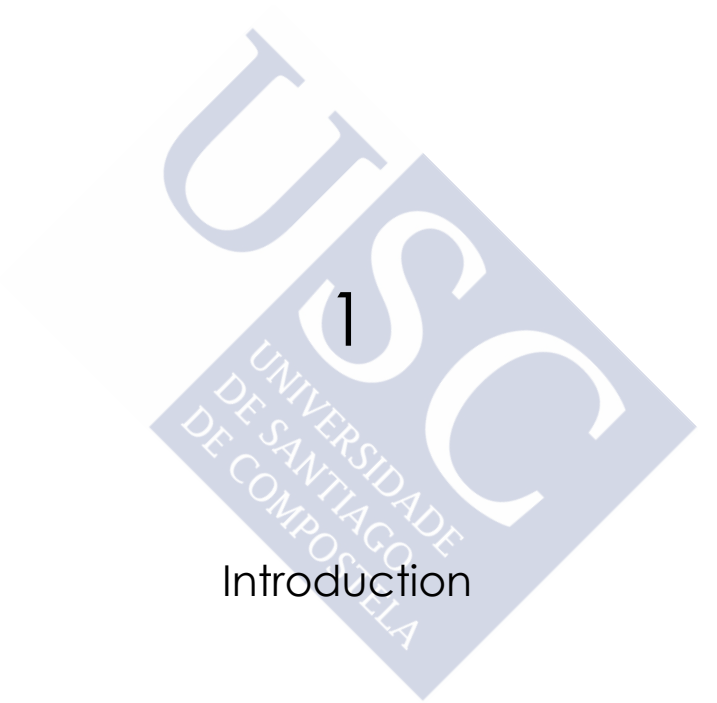
Interacción entre el cation imidazolio y aminoácidos aromáticos

En este trabajo se ha estudiado la interacción entre el catión imidazolio y la serie de aminoácidos aromáticos, incluyendo la histidina. El catión imidazolio es un anillo pentagonal que contiene dos grupos N-H. Se trata por tanto de un catión plano que puede interaccionar favorablemente mediante stacking con los diferentes anillos aromáticos de los aminoácidos. El procedimiento seguido es similar al planteado para el estudio de los complejos con catión guanidinio, empleando los métodos que ya han mostrado ser capaces de proporcionar una descripción adecuada de este tipo de sistemas.

Los resultados indican que los complejos formados por el catión imidazolio con Phe, Tyr y Trp presentan características muy similares entre sí, y a la vez coincidentes con las obtenidas para los complejos con catión guanidinio. Las estructuras más estables presentan contactos catión···π, aunque complejos con los aminoácidos en forma zwitteriónica presentan estabilidades muy similares. Nuevamente, la estabilidad de los complejos aumenta al aumentar el tamaño del anillo aromático implicado en la interacción debido a que proporciona mayores contribuciones de inducción y dispersión, que compensan el coste necesario para plegar el aminoácido. Los complejos con His presentan un comportamiento totalmente diferente. No se encuentran estructuras con contactos cation···π entre los mínimos más estables, que corresponden a estructuras con el aminoácido en forma zwitteriónica y con el catión interaccionando con el aminoácido sólo mediante enlaces de hidrógeno. Los resultados indican que, tanto en fase gas como en presencia de disolvente, es posible la formación de estructuras en las que se establecen contactos catión···π con Phe, Tyr y Trp, mientras que ese tipo de estructuras no es posible para los complejos con His.

En conjunto, podemos decir que los trabajos presentados en esta tesis ayudan a profundizar en el conocimiento de las características de los contactos catión···π que aparecen de forma habitual en sistemas biológicos. De modo general, los resultados indican que la presencia de un pequeño número de moléculas de disolvente que puede interaccionar con el contacto catión···π puede tener un impacto considerable en sus características. Así, la interacción se debilita, pero más relevantes son los cambios en la estructura de los complejos que se deben a la necesidad de acomodar las moléculas de disolvente próximas. Por otra parte, los contactos catión···π considerados están dominados por una combinación de efectos electrostáticos y de inducción, si bien el papel de la dispersión también puede ser relevante, especialmente en los cationes más estructurados. La preferencia de los aminoácidos por estructuras con contactos catión···π se debe a mayores contribuciones de la inducción y especialmente dispersión a la estabilidad, capaces de vencer la resistencia del aminoácido a adoptar conformaciones más plegadas.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA



1

Introduction

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

1.1. Amino acids

The current thesis is mainly focused on interactions involving amino acids and, especially, on the cation···π interaction observed among aromatic amino acids and cations. The main reason which gave rise to this study is the vital importance of amino acids, the building blocks of proteins, as well as their properties and characteristics in different environments.

Within every living organism, proteins carry out essential functions, so much that without them life would be non-viable. There are different kinds of proteins as regulatory proteins, receptor proteins, inmuno proteins, structural proteins, transfer proteins… Nevertheless, most of them participate in the so-called enzymatic catalysis or chemical conversions in and around the cell. These molecular machines, involved consequently in the cellular metabolism, are the responsible of obtaining chemical energy and also of fabricating useful compounds for the organism itself.[1] They are so important that the instructions for their synthesis are encoded within the genetic code of every living organism, transmitted from generation to generation. In this way, there is a collinear and unique correspondence between the sequence of amino acids of proteins and the sequence of nucleotides of the nucleic acid: three adjacent nucleotides code for one amino acid (Figure 1.1).

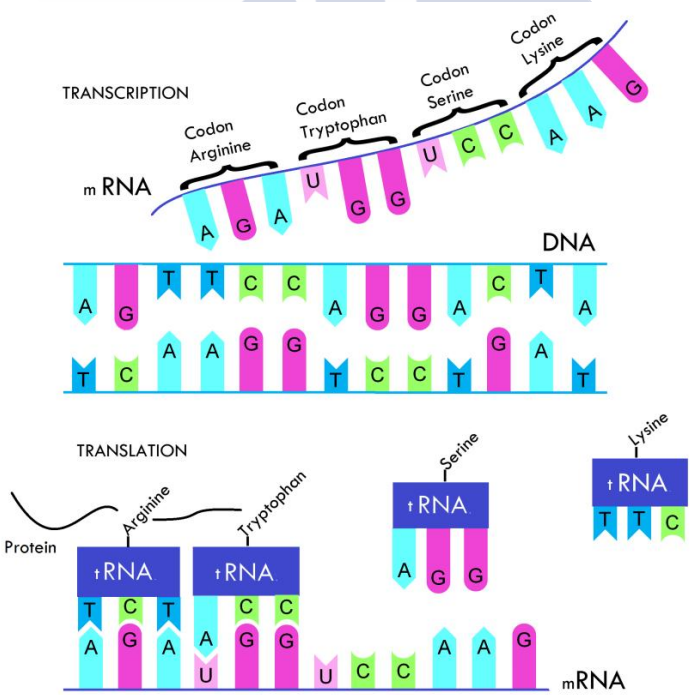


Figure 1.1. Scheme of gene-code and the relationship between amino acids and DNA.

Amino acids, both isolated and within a polypeptide chain, can be observed as *L* or *D* forms in their C_α carbon; that is, the carbon linked to the amino and acid groups of the molecule, as well as to its characteristic side chain. However, it is interesting to note that only those with *L* form are gene-coded and, thus, directly expressed.[2] Also, spontaneous racemization (*L* ↔ *D* transition, Figure 1.2) is not observed in proteins. If amino acids in *D* form are necessary, specific enzymes are required for their synthesis. This fact has not got a unique explanation, although presumably it occurred by chance. Nonetheless, once the *L* configuration was established as the main one, the *D* configuration became obsolete. The genetic code is at the center of the organizational scheme of life, so that it is universal for every living material. If a change in that code occurs (for example the *D* configuration), this scheme is greatly disrupted, resulting on the extinction of the organism in which the change had occurred. For that reason, the conservation of the code and thus the types of amino acids is essential.

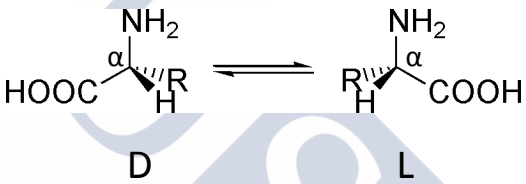


Figure 1.2. Racemization of amino acids (not observed spontaneously in proteins).

Such preference for one kind of structure, characteristic of living material, is also one of the searched aspects in extraterrestrial material when seeking for living organisms, for instance in meteorites from space. Until now, there are not conclusive data since the large majority of amino acids found in meteorites were observed in their racemic form, or if otherwise were found with enantiomeric excess for one of its forms, it was not possible to accurately determine if its origin is due to terrestrial contamination, being a controversial aspect among different authors.[3-9]

To understand the reason of the high specificity observed in living organisms as commented above, it is first necessary to answer what life is. According to Norman R. Pace,[10] "life is a self-replicating, evolving system expected to be based on organic chemistry", composed of macromolecules. That is, a living organism is such, when it has the capacity not only of auto-replication, but also of evolution. This replication process, according to the same author, requires large molecules with a complex molecular structure, where the information of how the system is built is stored. Such information is given by the high degree of specificity in the interactions of molecules that carry out the replication.

This high degree of specificity can be obtained with a certain number of molecules, easily differentiable, that show a series of characteristics and that will be the building blocks used by every known living organism. In the case of proteins, such a group is formed by 20 standard amino acids, their main building blocks. Such amino acids are likely the simplest carbon structures that can deliver the functional groups used by living systems, with properties such as intrinsic chirality, reactivity with water, and also the ability to attach to each other by means of the peptide unit, resulting in a repeating structure. Besides, there exists another advantage, since amino acids can be formed from simple organic molecules, being formed even in extraterrestrial bodies such as meteorites.

Moreover, these 20 amino acids differ only with respect to their side chains and thus, it is in these side chains where all the information is stored. Each side chain is so specific that it cannot be easily replaced by another one, since they have different characteristics or interact in a different manner with other surrounding molecules. Actually, they also can establish interactions within the same molecule with other side chains located in different areas of the molecule. This fact confers to each specific side chain certain characteristics that, when the molecule is going to be built, will be taken into account to give the desirable functionality and structure to the final protein.[2] As a result, the interactions involved in molecules as essential and specific as proteins play a key role, so a change or a loss of functionality can be produced if those interactions are not proper. On the other hand, the knowledge of such interactions as well as their properties also can be used to modulate the activity of proteins and employ them, for instance, on the development of drugs.

This work is thus focused towards one kind of those interactions present in proteins, the one involving aromatic groups (for example between side chains of aromatic amino acids) and also cations from positively charged side chains. This cation$\cdots\pi$ interaction will be discussed deeper in following sections.

1.2. Non covalent interactions

Although the term non covalent interaction is relatively modern (the existence of non covalent interactions is known since van der Waals proposed his equation of state for real gases at the end of the XIX century),[11] the idea of matter constituted by atoms intertangled among them is not that new. In the fifth century BC, the atomistic philosophy arose with Leucippus and Democritus' theories. This philosophy was an important influence on Greek and Roman thoughts. In the first century BC, the Roman Epicurean Lucretius wrote the poem *De rerum natura*, where references to the early concept of molecular interactions and viscosity can be seen.[12]

Again, things that seem to be hard and stiff must be composed of deeply indented and hooked atoms and held firm by their intertangling branches. In the front of this class stand diamonds(...) Liquids on the other hand must owe their fluid consistency to component atoms that are smooth and round.

We see that wine flows through a strainer as fast as it is poured in; but sluggish oil loiters. This is no doubt either because oil consists of larger atoms or because they are hooked and intertangled and therefore cannot separate so rapidly, so as to trickle though the holes one by one.

Since then, both molecular theory and also the concept of interaction have been modified up to get the current concepts. Thus, in general terms, it can be said that matter is constituted, as we see, by means of two kinds of interactions: those which are involved in the formation of molecules, where the final product (or molecule) has its own characteristics and different from the constituent atoms (the so-called covalent interactions or bonding interactions); and the interactions, therefore at a larger distance between these molecules, which do not entail changes on the intrinsic characteristics of the molecules (the so-called non covalent interactions or intermolecular interactions). Only taking into account these two kinds of interactions acting together, it is possible to roughly understand the complexity of biological systems.[13]

Processes in which the former interactions are modified are the so-called chemical reactions. These processes give rise to different bond reorganizations and thus, new molecules with different properties from the initial ones are formed. Therefore, the study of chemical reactions is a basic subject in this discipline. However, non covalent interactions, though with a smaller strength, also play a key role in chemistry. Although individually these interactions are weak, jointly they can be of great importance, being the responsible of the final structure of proteins, molecular recognition, autoorganization processes of macromolecules and cells and, ultimately, of every process in which movement or communication between molecules or cells is required. It is for that reason that non covalent interactions are essential for every structural or functional aspect in living things.

The existence of this kind of intermolecular interactions becomes evident with the condensed phases of matter, which indicate the presence of attractive forces between molecules. In turn, the tough compression of liquids also makes think about the existence of repulsive interactions between molecules. Thus, the set of all intermolecular interactions, both attractive and repulsive, allows for instance the formation of molecular aggregates, as well as macromolecular stabilization.

In fact, there is one specific chemical discipline dedicated only to these kinds of interactions, Supramolecular Chemistry that, according to J. M. Lehn, is defined as "Chemistry of non covalent bonding".[14, 15] It is focused on the development of big complex chemical systems using molecules that interact among them by means of intermolecular interactions. Therefore, the development of Supramolecular Chemistry depends on the knowledge and proper description of this kind of interactions, being able to be applied in a variety of fields as Catalysis, Medicine or Material Science due to its high versatility.

However, non covalent interactions also play a role in the macroscopic world. An example of that can be seen in the ability of geckos to rapidly climb smooth vertical surfaces up. The study of the adhesive force values support the hypothesis that individual keratinous hairs set on their feet, operate by means of non covalent interactions. The sum of all those interactions (the gecko has a few hundred thousand of these keratinous hairs in its feet) can support their weight.[16]

Concerning to the nature of intermolecular forces it can be said that their origin is mostly electromagnetic. This origin is due to the fact that, although the forces which arise between particles are of different nature (gravitational, electromagnetic, strong nuclear forces and weak nuclear forces), most of them are extremely short ranged (strong and weak nuclear forces, 10^{-4} nm) or conversely, extremely long-ranged (at 0.4 nm gravitational forces amount to only $7 \cdot 10^{-52}$ J) so they can be neglected at typical molecular dimensions. Consequently, the origin of intermolecular forces must be electromagnetic.

Different classifications of non covalent interactions can be done, taking into account the nature, force, direction, etc. although commonly they are classified depending on the distance between interacting objects where the interaction becomes significant: "long range" and "short range" forces.[13, 17, 18] This classification has a foundation in theory, since long and short range forces are described by means of different functions: r^{-n} for long range, and $e^{-\alpha r}$ for short range forces. Both of them are present at short distances, but only the long range forces are present when molecules are far apart.

Within the group of **long range forces** are electrostatic, induction, dispersion, resonance and magnetic forces. The effect of *electrostatic interactions* is the result from slow-moving or stationary charge distribution of molecules, being a classical effect. Depending on the charge distribution and orientation of interacting molecules, this effect can be attractive or repulsive, but it is always pairwise additive. Those charge distributions may be monopoles (punctual charges) or multipoles (distribution of static charges). Thus, a representation of the electrostatic interaction can be obtained from the sum of the contributions related to each interacting molecule. Nevertheless such a representation applies only for large separations between the interacting systems, since at short distances overlapping interferes on it. The aforementioned sum is

theoretically infinite, involving every possible interaction between different contributions from interacting molecules. Therefore, in this expression terms like charge-charge, charge-dipole, dipole-dipole, charge-quadrupole, dipole-quadrupole, etc., should appear. However, in practice, an infinite sum is unviable, so the interaction will be defined only with the most important contributing terms. Such leading terms depend on the nature of the interacting system. The dependence on the distance for each term of interacting multipoles is given by $R^{-(l_1 + l_2 + 1)}$, where R is the distance and l_1 and l_2 indicate the order of the multipoles corresponding to the interacting distributions *1* and *2* ($l_{\text{monopole}} = 0$, $l_{\text{dipole}} = 1$, $l_{\text{quadrupole}} = 2$, etc.). Therefore, the prevailing term in the expression will be always that with the lowest global order. According also to intuition, the leading term, for instance, for purely ionic compounds will be the charge-charge term (R^{-1}) or the dipole-dipole term (R^{-3}) for polar fluids. Other examples are systems constituted by center-symmetric molecules, as H_2, N_2 or benzene where the prevailing term is the quadrupole-quadrupole (R^{-5}) interaction, or the case of tetrahedral molecules such as methane where the leading term is the octopole-octopole (R^{-7}) interaction.

Inductive interactions are also included in the long range forces, typically showing an R^{-6} dependency. As in electrostatic forces, induction effects are due to the interaction between charge distributions, but in this case these distributions are deformed by the external field generated by surrounding molecules. The inductive effect appears whenever a permanent electric moment exists, since it can modify the charge distribution of other molecules close by. However, if an electrostatic effect is also present, inductive forces are partially screened by electrostatics so they are relatively less relevant. The so-called Debye interaction is the most common case of inductive effect,[19] where a dipolar molecule, interacting with an nonpolar molecule, generates a distortion on the electronic charge distribution of the latter, inducing a dipolar moment. Thus, an attractive force arises from the interaction of both dipoles: the permanent of the dipolar molecule and the induced one on the nonpolar molecule. The fact that the induction energy is triggered by an external field makes it always negative for molecules in their ground electronic states. Besides, induction interactions are non-additive because the presence of other molecules affects the interaction between two of them.

When electric moments exist, either permanent or induced, it can be easily understood how they may interact between them. However, the interaction between two nonpolar molecules is not that intuitive. Indeed, there was not a clear explanation of the existence of condensed states (liquid or solids) of nonpolar atoms or molecules as temperature is lowered until 1930, when F. London described the dispersive or *London dispersion forces*.[20] This effect was first described by means of second order perturbation theory, indicating a relationship between the optic dispersion in gases and these forces, hence its name. Thus, it has not a classical origin, but a quantum

mechanics origin, arising from quantum-induced instantaneous polarized multipoles in molecules.

In symmetrical molecules like hydrogen there does not seem to be any distortion in the charge distribution to produce dipole moments. However, that is only true on average. The movement of the electrons in a molecule makes the charge distribution fluctuate continuously, and at any one instant, the electronic distribution might be not symmetrical, arising then an electric moment. This instant dipole can induce another dipole in a surrounding molecule, and thus, they can interact with each other. This synchronized movement of electrons, the so-called electron correlation, can occur even over a large number of molecules as long as the molecules are close together, spreading thus the dispersion effect. Indeed, London forces are the only long-range interactions present in all molecular interactions, being the most relevant contribution except in small polar molecules. The correlation effect favors lower-energy configurations, decreasing then the energy of the system. Also, the resulting effect is attractive since correlation becomes stronger when molecules are close each other. The charge fluctuation in a molecule that causes correlation is related to the polarizability. In this way, multipoles, either instantaneous or induced, are directly related to molecular polarizabilities and consequently dispersion interactions are too. The size and molecular shape also affect the dispersion interaction, since polarizability depends on them, dispersion forces usually increasing with the size of the molecules.

The last effects included as part of the long-range forces are resonance and magnetic effects. On the one hand, resonance interaction occurs when at least one of the molecules is in a degenerate state, or if the system is formed by identical molecules, when one of them is in an excited state. Therefore, it cannot be observed in closed-shell molecules in their ground states. On the other hand, magnetic effects occur when there are unpaired spins. Anyway, both of them are negligible for the systems in this study.

Short-range effects, contrary to long-range effects, need a more complex description, being the most important contribution the so-called *exchange-repulsion effect*.[21] If the distance between nuclei is short enough, the overlap of the charge clouds becomes appreciable and large distortions are required by the Pauli Exclusion Principle. Thus, it is not possible to treat the system as composed of completely separated molecules in their respective unperturbed states, as for long-range forces. To satisfy this principle, a correction arising from the antisymmetry of the wavefunction is included in the wavefunction, the exchange term. Although the origin of some distortion is due to the Coulombic repulsion, this seems to be a secondary effect. For the systems treated in this work, with closed electronic shells, the electronic clouds of the interacting atoms tend to avoid each other. The decrease on charge density between the interacting atoms and therefore the reduction of the screening of the nuclear charges by the electrons, results in a

repulsion effect between the nuclei and in the last instance, between molecules. However, although molecules at short distances mostly suffer this repulsive effect, this repulsion is also combined with an attractive force coming from the fact that the electrons, instead of moving around only one atom, can increase its free movement over both atoms (or molecules).

At short distances, charge transfer and charge penetration have to be considered too. The former effect was firstly introduced by Mulliken, and consists on an attractive force between an acceptor (molecule deficient on electrons) and the electronic cloud of a donor (molecule rich on electrons). This effect may take place in the region between both molecules and greatly increases the interaction energy, though it can also be considered as a part of induction forces acting at short-range.[13, 17] The latter effect arises from the overlap of electron clouds, and corresponds to the difference between the electrostatic and coulombic energies, which start to be different at short distances, since the multipole expansion does not provide a satisfactory description at these distances, resulting in an error. Thus, charge penetration can be seen as the attraction between the semi-shielded nuclei of one molecule and the electronic cloud associated to another molecule, making this combination unviable for description by using the usual multipolar expansion. Also, the magnitude of penetration, although attractive, decreases exponentially with increasing distance. Thus, different contributions can be defined at short distances depending on the method applied, though the overall effect is repulsive.

Taking into account the cation$\cdots\pi$ interactions studied in this work, it can be said that the main contributions will be electrostatics, induction and, in larger systems, also dispersion forces. On the one hand, the interaction between the charge of the cation present in the system and the multipole distribution of the other species will be mainly electrostatic. On the other hand, the polarizable aromatic cloud of the aromatic species will be affected by the polarizing ion, leading to a high induction effect.

1.3. Interactions with aromatic systems

When two species of the same or different nature are in contact, interactions between them appear and new bonds can be formed or not. These latter interactions, depending on the involved species, could belong to one or more kinds and could be of different strength. Thus, interaction between ions is purely electrostatic, while for instance in the interactions between polar species there are both electrostatic and induction contributions. In Table 1.1 different kinds of interactions as well as their approximate energy value in a simple contact between two species are shown.

Table 1.1. Nature and strength of different non covalent interactions.

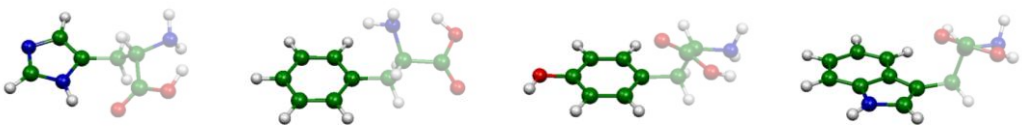
Interaction	Binding Energy (kcal mol^{-1})	Nature of the Interaction
Ion-Ion	25 – 85	Electrostatic
Ion-Dipole	10 – 50	Electrostatic and induction
Dipole-Dipole	1 – 10	Electrostatic and induction
Hydrogen Bond	1 – 30	Electrostatic and induction (dipole-dipole)
XH···π	1-5	Electrostatic and induction (weak HB)
Cation···π	1 – 20	Electrostatic and induction
π···π Stacking	1 – 10	Weak electrostatic and dispersion
Anion···π	5 – 10	Electrostatic and induction
Lone pair···π	1 – 5	Electrostatic
Halogen Bond	1 – 45	Weak electrostatic
van der Waals Forces	< 1	Dispersion

However, it is important mentioning that frequently not only isolated interactions are established, but several of them jointly contribute to the stability of the system, greatly affecting their properties. A remarkable case is that one related with the DNA chain structure, stabilized by many stacking interactions together with hydrogen bonds between nitrogen bases.

Nevertheless, the present work is focused on interactions involving aromatic systems, especially cation···π interactions. The presence of the delocalized electronic cloud of the aromatic units, together with their planar structure, confers to these systems some characteristic properties favoring them to be present in lots of structures and key processes of biochemistry, as protein-ligand interactions, neurotransmitters or ionic channels.[22, 23] For that reason, the design of new drugs based on this kind of interactions has also gained importance.[22-24]

As already mentioned in previous sections, amino acids are fundamental structures whose importance relies on the fact that they are the constituent units of proteins. Some of them present aromatic units in their side chains, so it is also possible to establish interactions through them. The way in which these aromatic units interacts and their strength will be determined by the interacting species as well as by the environment in which the interaction is established.

Therefore, each aromatic amino acid has a characteristic aromatic unit which could be used in different aspects, as shown in Figure 1.3. These amino acids can interact with other species by means of the aromatic rings in their side chain, so it is crucial to understand how aromatic species interact.



His (imidazole) **Phe** (benzene) **Tyr** (phenol) **Trp** (indole)

Figure 1.3. Aromatic amino acids and their side chains.

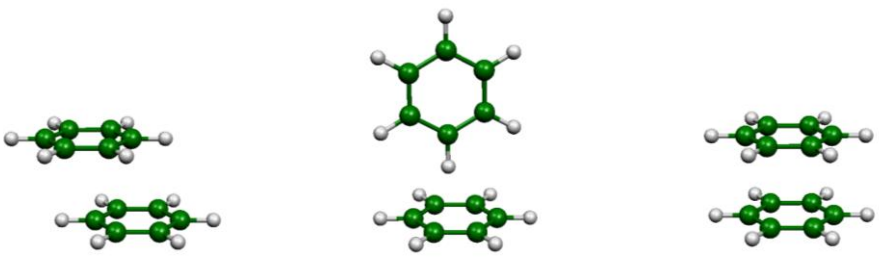
Commonly, an aromatic ring can establish interactions with different species, which in most cases can be classified as one of the following three types: π···π interactions, XH···π interactions and ion···π interactions. These three types of non covalent interactions will be briefly described below.[25]

Stacking or π···π interactions

In aromatic systems, the electron density is located into π orbitals associated with the carbon framework.[26] On the other hand, the σ electronic cloud, corresponding to the bonds with the peripheral hydrogens of the aromatic system, is positively polarized. Thus, two different regions coexist in the aromatic systems: an electron poor region (σ cloud) and a rich electron region, associated with the π system. This model was firstly proposed by Hunter and Sanders (1990) based on the combination of the van der Waals and electrostatic forces with the aim of explaining the different configurations observed and to predict their interaction energies.[27]

Each one of the previous regions is able to interact with different kind of species. For instance, if another aromatic system is in the neighborhood, different situations may occur: a perpendicular interaction between σ and π electronic clouds (T-shaped or Edge to face); or parallel interactions. In this latter situation it is possible to observe two different orientations: one with parallel displaced systems and another one with aromatic rings totally overlapped (face-to-

face or sandwich), although only few examples exist for the sandwich arrangement.[28, 29] The most common situations are the T-shaped or the parallel displaced ones, where molecules, although lie above each other, are offset enabling the complementary electron poor and rich regions to match up. Figure 1.4 shows the different types of interactions between two aromatic systems, benzene in this example, and how parallel arrangement competes with other dispositions.[30, 31]



Offset face to face or **Parallel displaced** | **Edge to face** or **T-shaped** | **Face to face** or **Sandwich**

Figure 1.4. Types of π···π interactions between benzene molecules.

Different factors contribute to the strength of these interactions, such as the presence of substituents in the rings like electron-donors or electron acceptors, the ring being or not a heterocycle, or the number of condensed aromatic rings involved in the structure. Thus, the strength of the interaction tends to increase when electron-attractor groups are present on the ring, since the π···π repulsion decreases because this substituent attracts the electronic cloud, diminishing its density.[32, 33] Also it increases with the higher number of rings involved[34] or when the rings are heterocycles.[35] In the latter case, as in the example of the electron-attractor substituents, the heteroatom, more electronegative than the carbon atom, causes a decrease of the electronic π density. This is a common effect when a nitrogen atom is substituting a carbon atom of the six-membered ring, as in the case of pyridine.[35] Anyway, a recent review of this interpretation has been done by Wheeler pointing towards direct through space effects to explain substituent effects.[36, 37]

Although stacking interactions were considered to be energetically much less significant than H-bonding (0-10 kcal mol^{-1}), accurate calculations have revealed that when they are associated,

like in DNA, surprisingly large stabilization energies are observed comparable with those of strong H-bonding. It is widely known that stacking interactions together with hydrogen bond play a key role in DNA stabilization as well as in its characteristic helical structure.[22, 23] This fact makes stacking interactions of great importance in biology and supramolecular chemistry, being not surprising the large number of drugs which has been developed based on them.[38-40] For instance, in cancer therapy a new target has emerged, the so-called c-Met or hepatocyte growth factor receptor (HGFR). This receptor tyrosine kinase has an abnormal activation in many cancer cells,[41, 42] and selective c-Met inhibitors have been developed in which π···π stacking interactions have an important role in the binding to the target.

Furthermore, with the discovery of fullerenes and related carbon-rich materials, a novel type of stacking involving curved carbon networks has been introduced. This kind of materials with accessible concave surfaces appear to be good candidates for the formation of π···π stacked supramolecular assemblies with fullerenes. The buckycatcher synthesized by Sygula el al,[43] which acts as a receptor of fullerene C_{60} molecule, is an outstanding example of this kind of interactions and its applicability. Also, since carbon nanotubes (CNTs) have been widely utilized as novel drug carriers, the π···π stacking interaction has been used by some authors as a way to bind the drug to the CNTs structure.[44]

XH···π interactions

In this interaction a π electron cloud of an aromatic ring and an X-H group are involved, usually C-H, N-H or O-H groups. CH···π interactions were described in 1952 by Tamres, when he observed that benzene and analogous compounds dissolve exothermically in chloroform.[45] In 1957, Reeves and Schneider proposed this interaction as a type of H-bond, taking into account their NMR results. Since then, CH···π interactions have been described in a vast number of small molecular systems, being Nishio's book an excellent compilation of these observations.[46, 47]

Interactions between π systems and NH or OH bonds have been also described with similar characteristics, although quantitatively the strength of the different interactions follows the order OH···π > NH···π > CH···π.[24] Spectroscopic measurements of benzene···water and benzene···ammonia complexes reported by Suzuki et al.[48] and Rodham et al.[49] showed that benzene acts as proton acceptor and that water and ammonia are positioned above the benzene plane. The NH···π interaction between a NH bond and a π-system was first reported in 1959 by Oki and Imamura using the measurements of the IR spectra of N-benzylaniline and its derivatives,[50] although it was also observed in other systems,[51-56] being a remarkable example the presence of this interaction in the hemoglobin-drug complex, reported by Perutz et al.[57] The preferred location of the amino groups with respect to the aromatic ring was reported by Burley and Petsko, indicating that these groups tend to locate above and below the aromatic system and

close to the center.[58] On the other hand, the OH···π (or OH/π) interaction plays an important role since it allows the interaction between non polar groups as aromatic systems and the polar water molecules of the solvent.[59] This kind of interaction is the responsible of the fact that benzene, a typical nonpolar organic molecule, is soluble in water, although also the delocalization of the π cloud greatly contributes to the affinity towards the polar solvent.[60]

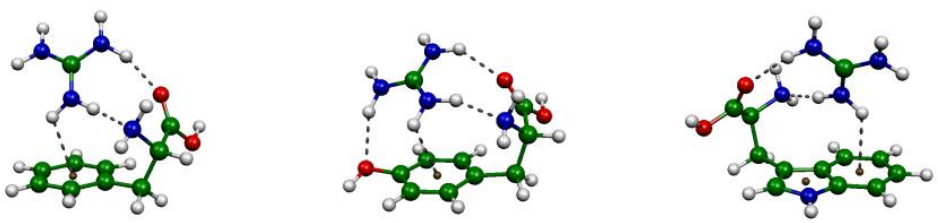
In XH···π contacts, the main contributions to the strength are the charge transfer and the electrostatic effects, which mainly control the directionality of the interactions, giving rise to a linear interaction between the CH bond and the *p* orbitals of the ring. However, dispersion effects also contribute in a non despicable way, suggesting that XH···π contacts have different nature than prototypical hydrogen bonds.[61-64] Recent experimental and theoretical studies in the gas phase[65, 66] show that the main contribution to this interaction is in many cases due to dispersion effects instead of electrostatics ones, as in the typical hydrogen bonds. The electrostatic contribution only becomes dominant when the hydrogen involved in the X-H···π interaction is very acid, for example, in cases as chloroform or acetylene. This fact makes the X-H···π interaction stable in both polar and non-polar solvents. Besides, the different dependence against the directionality of the interaction, corroborates their different nature.[67] While typical hydrogen bonds have a high directionality, the X-H···π interactions are not that strict with respect to the orientation of the species involved in it. In fact, it is not rare to find non linear arrangements on X-H···π contacts.[24, 68]

The overall protein stability is in many cases not more than a few kcal mol^{-1}, so that the overall stabilization energy of about 0.5-1.0 kcal mol^{-1} per interaction can be an important contributor.[67] In fact, the formation of many complexes of proteins with special ligands or cofactors in which X-H···π interactions have a special participation have been described by different authors.[46, 69-73] An example of applicability of this interaction was the design of serine protease inhibitors.[74, 75]

It is worth mentioning that substituent effects on XH···π interactions can impact deeply on their properties, but their magnitude as well as their physical origin depend strongly on the nature of the X group. Particularly, the higher polarity of the XH bond, the higher the effects caused by the substituents will be.[76] XH···π interactions will be of great importance in this work, since the aromatic units of the aromatic side chains can interact by means of this kind of interaction with solvent molecules or the XH residues present in the molecules (intra or intermolecular interactions depending on the case).

Cation···π interactions

Aromatic systems such as benzene or phenol are commonly found in different biological or non-biological contexts. In the biological context this kind of systems can be observed for instance as a part of aromatic amino acids as tyrosine, tryptophan or phenylalanine.[26] The existence of charged species surrounding them, even forming part of the same protein (like cationic side chains of arginine or lysine) allows for the possibility of interactions between the charged groups and the electronic cloud of a π system. A particular and usual case is that one in which a cation interacts with the delocalized electron density that lies above and below the plane of an aromatic ring. In fact, an important role of this kind of interactions is that one related with the building of the correct tertiary structures in proteins. In Figure 1.5 examples of different aromatic amino acids interacting with a guanidinium cation, present in the side chain of arginine, are shown. Cation···π interactions can be appreciated acting together with hydrogen bonds.



Gnd-Phe **Gnd-Tyr** **Gnd-Trp**

Figure 1.5. Examples of cation···π interaction.

The cation···π interaction is usually explained in terms of electrostatic effects.[77-80] Accordingly, this kind of interaction arises from the electrostatic interaction between the negative first nonzero multipole moment of the arene (quadrupole moment in the case benzene), and the positively charged cation. Other molecules like ethylene or acetylene and its derivates also have a quadrupole moment, which can interact in the same way by means of cation···π interactions. Although this is the larger contribution, other effects are also important, as the polarization of the π-electron system by the cation.[81] However, the dispersion contribution is typically small.[82]

The cation···π interaction usually has a large magnitude, more appreciable in the gas phase, as one might expect. For instance, the first evidence of this cation···π interaction, published in 1981

by Kebarle,[83] reported that K^+ ions are better stabilized in the gas phase by benzene (ΔH^o = -19 kcal mol^{-1}) than by water (ΔH^o = -18 kcal mol^{-1}). However, contrarily to other strong non covalent forces like ion pairs or hydrogen bonds, cation···π interactions retain the strength and the specificity in a better way in water (and even across a range of solvents), and they can be even strong in this environment.[84] For instance, since benzene is a hydrophobic nonpolar molecule, the first step of desolvation is not as big a problem. Therefore, the global stability will be dominated by the desolvation energy of the cation and also by the tendency of the cation to bind the benzene molecule. This fact is the responsible of the differences in the cation···π interaction tendencies when they are measured in gas or aqueous phases. Although Li^+ cation has the strongest cation···π interaction in the gas phase among all the alkaline cations, when the measurement is done in water, it is the K^+ cation the one presenting the largest tendency to interact with benzene. K^+ cation has much lower desolvation energy than Li^+ while still remains as a good π binder. Consequences of these properties are shown in reports which indicate that most cation···π interactions in proteins are located at the surface of proteins, where salt effects would be the weakest, pointing again to the importance of the cation···π interactions on protein structure.[85]

Due to their remarkable properties, the interest on cation···π interactions rapidly increased, giving rise to a wide amount of experimental and computational works focused on their study. Thus, only a few years after Kebarle's report, Meot-Ner demonstrated that the more complex organic cations like ammonium or tetramethylammonium (TMA) are also good π binders in the gas phase.[86] Besides, *ab initio* studies of this kind of interaction have been done to obtain the theoretical minimum energy structures.[87]

Cation···π interactions are also present and take part on numerous structures and biological processes. For instance, cation···π interactions are common in proteins or their complexes with different ligands or even with DNA.[78, 80, 87, 88] Also, cation···π contacts usually participate in the bond between neurotransmitters and receptors. A remarkable example of this is the nicotinic receptor. This protein usually binds the acetylcholine (Ach) neurotransmitter, triggering a process in which the voluntary movements of muscles are involved. This binding is established between the quaternary ammonium of the Ach and one specific tryptophan of the receptor by means of a cation···π interaction. However, the nicotine molecule, protonated in the biologic system, shows a similar structure as the neurotransmitter, with a quaternary ammonium being able to bind the receptor also by means of a cation···π interaction, resulting in a physiological response.[89] Another example, postulated by Dougherty in the early 1990's,[90] was the possibility of the participation of cation···π interactions in highly selective K^+ channels. As commented above, among the alkalines, K^+ shows the larger cation···π interaction in water, allowing the channels to be more selective for K^+ than for Na^+. Once proved the efficiency of the interaction between an

aromatic ring and a cation due to its abundance in many kinds of systems, this interaction has also been used to increase the π-face selectivity in asymmetric catalysis[87, 91] or as binding places for alkaline metals. In the latter case, molecules similar to crown ethers but with aromatic systems located in the structure have been used.[87, 92]

Regarding to substituent effects in cation···π interactions, it is worth mentioning that not only electrostatic forces are involved, but also polarization effects on the aromatic systems play an important role.[93] Thus, both effects are important for an appropriate description of the cation···π interaction, either for heterocyclic arenes or not. The size of the aromatic systems greatly affects the polarizability, leading to the increase of the induction effects. In fact, the induction effects would be enough to stabilize cation···π complexes even in cases in which the electrostatic term is repulsive.[94] Substituents also influence the polarizability of the arenes and in general, it can be assumed that while electron-donor (π-electron rich) groups strengthen the cation···π interactions, the electron-acceptor (π-electron poor) groups weaken it, but not only from the local C–X dipoles, but also due to the resonance effects over the aromatic ring. However, these substituent effects can also be seen from another point of view, considering that these effects arise from interactions of the cation with the local dipole (or higher-order multipoles) associated with the substituent.[95]

Although it is not of relevance in this work, it is worth mentioning that interactions between anions and aromatic rings are also possible. Due to the great capability of substituent effects, when an aromatic system is completely substituted with electron-acceptors (for instance halogen, nitro or cyano groups), the π cloud becomes electron-deficient so that attractive interactions with anions is possible.[96, 97]

1.4. References

(1) A. V. Finkelstein, O. B. Ptitsyn, *Protein Physics: A Course of Lectures*, Academic Press, London, **2002**.

(2) G. E. Schulz, R. H. Schirmer, *Principles of Protein Structure*, Springer, Berlin, **1979**.

(3) M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House, J. P. Dworkin. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13995-13998.

(4) D. P. Glavin, J. L. Bada, K. L. F. Brinton, G. D. McDonald. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 8835-8838.

(5) J. R. Cronin, S. Pizzarello. *Science* **1997**, *275*, 951-955.

(6) J. E. Elsila, D. P. Glavin, J. P. Dworkin, Z. Martins, J. L. Bada. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E3288.

(7) J. L. Bada. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, E85.

(8) S. Pizzarello, D. L. Schrader, A. A. Monroe, D. S. Lauretta. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11949-11954.
(9) K. A. Kvenvolden, J. G. Lawless, C. Ponnamperuma. *Proc. Natl. Acad. Sci. U. S. A.* **1971**, *68*, 486-490.
(10) N. R. Pace. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 805-808.
(11) P. Schuster. *Ber. Bunsenges. Phys. Chem.* **1983**, *87*, 291-292.
(12) Lucretius, *On the Nature of the Universe*, Penguin Classics, London, **1994**.
(13) P. Hobza, R. Zahradník, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and in the biodisciplines*, Elsevier, Amsterdam, **1988**.
(14) J. M. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, Wiley, Weinheim, **1995**.
(15) J. W. Steed, J. L. Atwood, *Supramolecular Chemistry*, Wiley, Chichester, **2009**.
(16) K. Autumn, Y. A. Liang, S. T. Hsieh, W. Zesch, W. P. Chan, T. W. Kenny, R. Fearing, R. J. Full. *Nature* **2000**, *405*, 681-685.
(17) A. J. Stone, *The theory of intermolecular forces*, Oxford University Press, Oxford, **2013**.
(18) I. Kaplan, *Intermolecular Interactions: Physical Picture , Computational Methods*, John Wiley & Sons, Chichester, **2006**.
(19) P. Debye. *Journal of the Society of Chemical Industry* **1929**, *48*, 1036-1037.
(20) F. London. *Zeitschrift für Physik* **1930**, *63*, 245-279.
(21) J. O. Hirschfelder, *Advances in Chemical Physics, Intermolecular Forces*, Wiley, Hoboken, **2009**.
(22) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.
(23) L. M. Salonen, M. Ellermann, F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-4842.
(24) S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe. *J. Am. Chem. Soc.* **2000**, *122*, 11450-11458.
(25) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.
(26) P. J. Cragg, *Supramolecular Chemistry: From Biological Inspiration to Biomedical Applications*, Springer Netherlands, Dordrecht, **2010**.
(27) C. A. Hunter, J. K. M. Sanders. *J. Am. Chem. Soc.* **1990**, *112*, 5525-5534.
(28) A. N. Sokolov, T. Friščić, S. Blais, J. A. Ripmeester, L. R. MacGillivray. *Crystal Growth & Design* **2006**, *6*, 2427-2428.
(29) A. N. Sokolov, T. Friščić, L. R. MacGillivray. *J. Am. Chem. Soc.* **2006**, *128*, 2806-2807.
(30) M. O. Sinnokrot, E. F. Valeev, C. D. Sherrill. *J. Am. Chem. Soc.* **2002**, *124*, 10887-10893.
(31) S. Tsuzuki, K. Honda, T. Uchimaru, M. Mikami, K. Tanabe. *J. Am. Chem. Soc.* **2001**, *124*, 104-112.
(32) C. A. Hunter. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1584-1586.
(33) F. Cozzi, M. Cinquini, R. Annuziata, J. S. Siegel. *J. Am. Chem. Soc.* **1993**, *115*, 5330-5331.
(34) R. Goddard, M. W. Haenel, W. C. Herndon, C. Krueger, M. Zander. *J. Am. Chem. Soc.* **1995**, *117*, 30-41.

(35) K. Kano, H. Minamizono, T. Kitae, S. Negi. *J. Phys. Chem. A J Phys Chem A* **1997**, *101*, 6118-6124.
(36) S. E. Wheeler, J. W. G. Bloom. *J. Phys. Chem. A J Phys Chem A* **2014**, *118*, 6133-6147.
(37) S. E. Wheeler. *Acc. Chem. Res.* **2012**, *46*, 1029-1038.
(38) P. Jurečka, P. Hobza. *J. Am. Chem. Soc.* **2003**, *125*, 15608-15613.
(39) P. Jurečka, P. Hobza. *Chem. Phys. Lett.* **2002**, *365*, 89-94.
(40) P. Hobza, J. Šponer. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.
(41) Y. Ma, G. Sun, D. Chen, X. Peng, Y.-L. Chen, Y. Su, Y. Ji, J. Liang, X. Wang, L. Chen, J. Ding, B. Xiong, J. Ai, M. Geng, J. Shen. *J. Med. Chem.* **2015**, *58*, 2513-2529.
(42) P. M. Comoglio, S. Giordano, L. Trusolino. *Nat Rev Drug Discov* **2008**, *7*, 504-516.
(43) A. Sygula, F. R. Fronczek, R. Sygula, P. W. Rabideau, M. M. Olmstead. *J. Am. Chem. Soc.* **2007**, *129*, 3842-3843.
(44) M. Marcaccio, F. Paolucci, *Making and Exploiting Fullerenes, Graphene, and Carbon Nanotubes*, Springer, Berlin, **2014**.
(45) M. Tamres. *J. Am. Chem. Soc.* **1952**, *74*, 3375-3378.
(46) M. Nishio, M. Hirota, Y. Umezawa, *The CH/π Interaction: Evidence, Nature, and Consequences*, Wiley, New York, **1998**.
(47) M. Nishio, *in* http://www.tim.hi-ho.ne.jp/dionisio/.
(48) S. Suzuki, P. G. Green, R. E. Bumgarner, S. Dasgupta, W. A. Goddard, G. A. Blake. *Science* **1992**, *257*, 942-945.
(49) D. A. Rodham, S. Suzuki, R. D. Suenram, F. J. Lovas, S. Dasgupta, W. A. Goddard, G. A. Blake. *Nature* **1993**, *362*, 735-737.
(50) M. Oki, H. Iwamura. *Bull. Chem. Soc. Jpn.* **1959**, *32*, 955-959.
(51) I. Suzuki, M. Tsuboi, T. Shimanouchi, S. Mizushima. *J. Chem. Phys.* **1959**, *31*, 1437-1438.
(52) I. Suzuki, M. Tsuboi, T. Shimanouchi. *J. Chem. Phys.* **1960**, *32*, 1263-1264.
(53) M. Oki, K. Mutai. *Bull. Chem. Soc. Jpn.* **1960**, *33*, 784-789.
(54) M. Oki, K. Mutai. *Bull. Chem. Soc. Jpn.* **1965**, *38*, 387-392.
(55) M. Oki, K. Mutai. *Bull. Chem. Soc. Jpn.* **1965**, *38*, 393-399.
(56) M. Oki, K. Mutai. *Bull. Chem. Soc. Jpn.* **1966**, *39*, 809-813.
(57) M. F. Perutz, G. Fermi, D. J. Abraham, C. Poyart, E. Bursaux. *J. Am. Chem. Soc.* **1986**, *108*, 1064-1078.
(58) S. K. Burley, G. A. Petsko. *FEBS Lett.* **1986**, *203*, 139-143.
(59) B. D. Ostojic, G. V. Janjic, S. D. Zaric. *Chem. Commun.* **2008**, 6546-6548.
(60) H. Takahashi, D. Suzuoka, A. Morita. *J. Chem. Theory Comput.* **2015**, *11*, 1181-1194.
(61) H. Suezawa, S. Ishihara, O. Takahashi, K. Saito, Y. Kohno, M. Nishio. *New J. Chem.* **2003**, *27*, 1609-1613.
(62) Y. Umezawa, M. Nishio. *Biopolymers* **2005**, *79*, 248-258.
(63) M. Harigai, M. Kataoka, Y. Imamoto. *J. Am. Chem. Soc.* **2006**, *128*, 10646-10647.
(64) A. Gil, V. Branchadell, J. Bertran, A. Oliva. *J. Phys. Chem. B* **2007**, *111*, 9372-9379.

(65) A. Fujii, K. Shibasaki, T. Kazama, R. Itaya, N. Mikami, S. Tsuzuki. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2836-2843.
(66) S. Tsuzuki, A. Fujii. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2584-2594.
(67) M. Brandl, M. S. Weiss, A. Jabs, J. Sühnel, R. Hilgenfeld. *J. Mol. Biol.* **2001**, *307*, 357-377.
(68) M. Nishio. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873-900.
(69) M. Nakanishi-Matsui, Y.-W. Zheng, D. J. Sulciner, E. J. Weiss, M. J. Ludeman, S. R. Coughlin. *Nature* **2000**, *404*, 609-613.
(70) P. Chakrabarti, U. Samanta. *J. Mol. Biol.* **1995**, *251*, 9-14.
(71) Y. Umezawa, M. Nishio. *Biorg. Med. Chem.* **1998**, *6*, 493-504.
(72) Y. Umezawa, M. Nishio. *Biorg. Med. Chem.* **1998**, *6*, 2507-2515.
(73) M. Muraki, H. Morii, K. Harata. *Protein Eng.* **2000**, *13*, 385-389.
(74) I. Maeda, Y. Shimohigashi, K. Ikesue, T. Nose, Y. Ide, K. Kawano, M. Ohno. *J. Biochem.* **1996**, *119*, 870-877.
(75) Y. Shimohigashi, T. Nose, Y. Yamauchi, I. Maeda. *Peptide Science* **1999**, *51*, 9-17.
(76) J. W. G. Bloom, R. K. Raju, S. E. Wheeler. *J. Chem. Theory Comput.* **2012**, *8*, 3167-3174.
(77) E. V. Anslyn, D. A. Dougherty, *Modern Physical Organic Chemistry*, University Science Books, Sausalito, **2006**.
(78) J. C. Ma, D. A. Dougherty. *Chem. Rev.* **1997**, *97*, 1303-1324.
(79) J. P. Gallivan, D. A. Dougherty. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459-9464.
(80) N. Zacharias, D. A. Dougherty. *Trends Pharmacol. Sci.* **2002**, *23*, 281-287.
(81) E. Cubero, F. J. Luque, M. Orozco. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 5976-5980.
(82) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.
(83) J. Sunner, K. Nishizawa, P. Kebarle. *J. Phys. Chem.* **1981**, *85*, 1814-1820.
(84) D. A. Dougherty. *Science* **1996**, *271*, 163-168.
(85) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.
(86) M. Meot-Ner, C. A. Deakyne. *J. Am. Chem. Soc.* **1985**, *107*, 474-479.
(87) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-2138.
(88) D. A. Dougherty. *J. Nutr.* **2007**, *137*, 1504S-1508S.
(89) W. Zhong, J. P. Gallivan, Y. Zhang, L. Li, H. A. Lester, D. A. Dougherty. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12088-12093.
(90) R. Kumpf, D. Dougherty. *Science* **1993**, *261*, 1708-1710.
(91) S. Yamada, C. Morita. *J. Am. Chem. Soc.* **2002**, *124*, 8184-8185.
(92) G. W. Gokel, L. J. Barbour, R. Ferdani, J. Hu. *Acc. Chem. Res.* **2002**, *35*, 878-886.
(93) A. Bauza, P. M. Deya, A. Frontera, D. Quinonero. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1322-1326.
(94) J. A. Carrazana-García, J. Rodríguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2011**, *115*, 2774-2782.
(95) S. E. Wheeler. *Acc. Chem. Res.* **2013**, *46*, 1029-1038.
(96) B. L. Schottel, H. T. Chifotides, K. R. Dunbar. *Chem. Soc. Rev.* **2008**, *37*, 68-83.
(97) A. Frontera, D. Quiñonero, P. M. Deyà. *WIREs Comput. Mol. Sci.* **2011**, *1*, 440-459.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

2

Objectives

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

The interaction between cations and aromatic species in the side chains of some amino acids is one of the key factors that control the stability of proteins. Different studies suggest that this type of interaction is present in a variety of systems indicating its relevance in many processes of chemical and biological recognition, such as nerve transmission or transport through the membrane. Considering the relevance of these interactions, the main objective of this thesis is to gain a deeper comprehension of the characteristics of systems that establish interactions between aromatic and cationic species by applying computational chemistry methods.

It is already know that the interaction between cations and aromatic systems is strong in the gas phase as a consequence of large contributions from electrostatics. The electrostatic contribution mainly comes from charge-dipole and charge-quadrupole contributions, and it usually dominates the interaction in model systems employing small cations and aromatic species. However, the aromatic electron cloud is easily polarizable, so large contributions from induction have been also observed, especially when large aromatic systems are involved. Therefore there is a general agreement as to consider the cation$\cdots\pi$ interaction as controlled by a combination of electrostatic and induction.

However, the data found in the literature show a significant variability with regard to the intensity of such interactions and thus the importance they actually have for protein stability. From some of these results, the interaction between the cation and one of the aromatic amino acids is intense and can make a decisive contribution to the stability of the system. On the other hand, other studies indicate that the role played by such interactions is marginal, contributing only a few tenths of kcal mol^{-1} to the stability of the system, although such estimates are often purely qualitative. Overall, everything seems to indicate that the environment of the interacting species may affect the characteristics of the interaction, either causing changes in intensity or changes in the geometrical disposition of the interacting fragments.

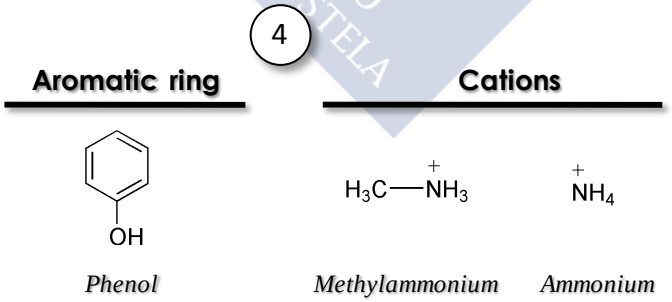
This thesis proposes conducting a thorough study of this kind of interactions in systems of interest, trying to clarify essentially which are the main features of such interactions. Therefore, it also aims to quantify the role played by the environment on these interactions, and whether it can act as a modulator of them, either weakening or causing significant structural changes. Since the study of cation$\cdots\pi$ interactions is a wide field for study, this thesis will be focused on interactions involving aromatic and charged species that can be on the side chain of amino acids.

Since the problems considered in the thesis are quite complex for their treatment using quantum mechanics methods, a gradual approach to the problem has been chosen. Thus, a series of initial studies has been carried out using simplified models for the amino acids where only the aromatic ring is considered in the interaction. Using these simplified models, studies have been conducted in order to ascertain the role played by the solvent molecules closer to the cation$\cdots\pi$

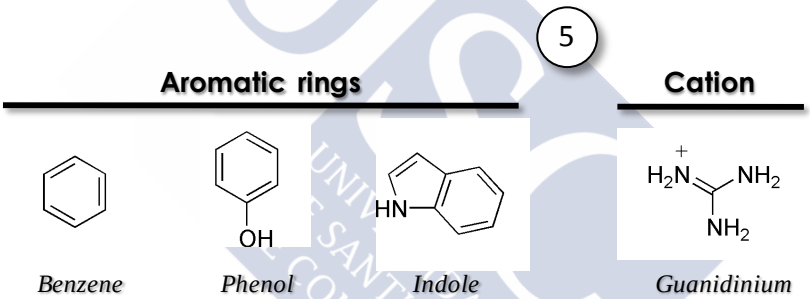
contact upon its characteristics. These studies using simplified models for the aromatic amino acid are the subject of chapters 4 to 6. Chapters 7 and 8 of this thesis show the results for a more complex model using the whole amino acid structure in the calculations. The goals of the different chapters are summarized below.

Chapter 4 is devoted to the simplest of the systems considered, corresponding to complexes formed by phenol with ammonium and methylammonium cations. Therefore, the systems mimic the interaction between a phenol unit in the side chain of tyrosine and the cationic end of lysine.

Most studies in literature on cation···π interactions deal with benzene complexes employing simple cations as alkali metals. The hydration pattern of an alkali cation···benzene complex is pretty simple, the water molecules coordinating to the cation while a coordination location is occupied by the benzene unit. However, the situation can be already rather different when a slightly different aromatic species such as phenol is employed. The presence of the hydroxyl group makes another position available for hydration thus leading to different hydration patterns. A microhydration study has been carried out, so a small number of water molecules has been added stepwise to the cation···π in order to ascertain their impact on the structure and energetics of the cation···π moiety. The results will help to determine the main characteristics of the competition between the aromatic cloud and the hydroxyl group in order to interact with both the cation and the water molecules. Also, the role played by the water molecules closest to the cation···π contact, and especially their impact in structure and stability, will be revealed.

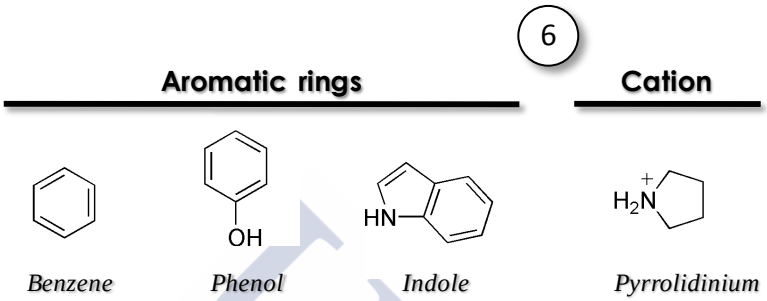


Chapter 5 deals with complexes formed by guanidinium cation and benzene, phenol and indol, in an attempt to recover information on the interaction between the cationic end of arginine and the aromatic amino acids phenylalanine, tyrosine and tryptophan. The procedure to follow is similar to the one employed in chapter 4. Thus, a small number of water molecules is included in an stepwise manner in the complexes formed by guanidinium cation and the three aromatic species. The results obtained will allow a direct comparison of the interaction of the three different aromatic molecules with guanidinium, giving hints about its strength. Also, guanidinium can interact with the aromatic rings adopting both a perpendicular and stacked orientation. While the perpendicular complex is the most stable in the gas phase, it has been indicated that in benzene complexes the presence of water molecules can change the order of stability favoring the parallel arrangement. The results will give more information about this competition between different orientations for the three aromatic units considered, as well as the possible role of water molecules in changing the stability order.

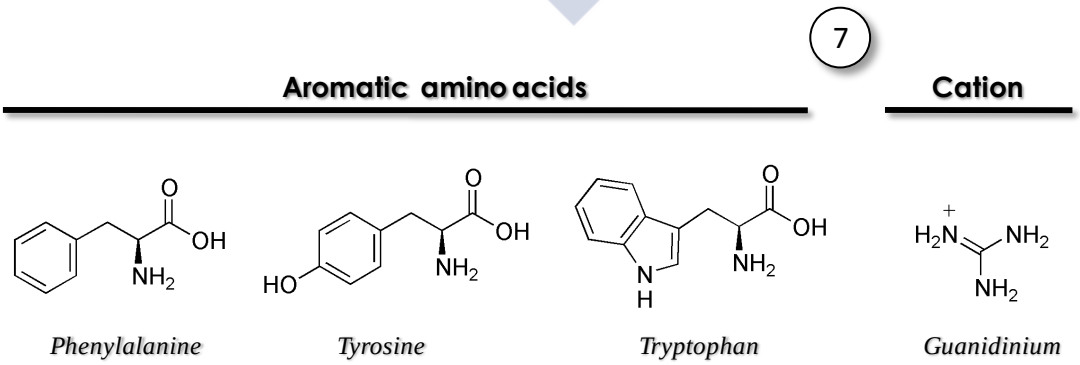


Chapter 6 further complicates the cationic unit employed, which in this case corresponds to pyrrolidinium cation, the protonated end of proline. Pyrrolidinium cation shows an ammonium group within a saturated five-membered ring, exhibiting larger flexibility than any of the cations considered in previous chapters. Besides, due to the ring structure of the cation it could be expected that parallel stacked structures will be favored. Reaching this point a question arises as to whether the methods employed in previous chapters are capable of properly describe the systems studied. In principle, the cation$\cdots\pi$ interaction, controlled by electrostatics and induction, could be described with common DFT methods. However, the possibility of forming stacked structures with larger contributions from dispersion suggests that more appropriate methods should be employed. Therefore, complexation energies will be obtained at the CCSD(T)/CBS

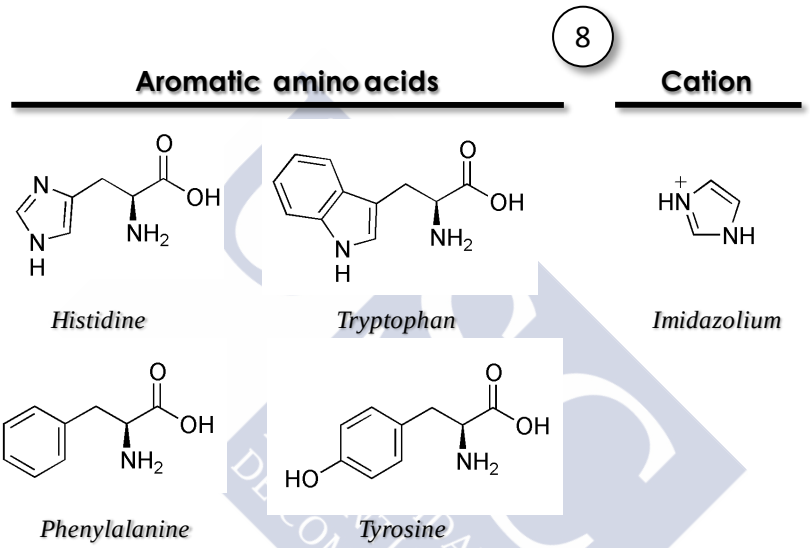
level, being used as benchmark in order to check the performance of cheaper methods. Therefore, in this chapter the goals are twofold: determining the characteristics of the interaction and the role of water molecules, as in previous chapters, and checking the performance of different methods in this kind of complexes.



In **chapter 7,** whole aromatic amino acids are considered, focusing on their interaction with guanidinium cation. Considering the whole amino acid introduces the problem of the large number of conformers which can form complexes with guanidinium cation and similar stability. Therefore, an automatic procedure for searching the conformational space of the amino acids and their complexes with guanidinium is employed based on an empirical force field. The results are further refined by using different quantum chemistry methods. As in the previous chapter, benchmark-quality results have been obtained for the stabilities in order to check the performance of cheaper methods. The results obtained will give information on the interaction and, specially, of the role of the amino acid group trying to interact with the guanidinium cation, thus competing with the cation···π contact.



Finally, in **chapter 8** a similar study employing the whole amino acids is performed, but in this case considering imidazolium as cation, trying to model the protonated side chain of histidine. Complexes of imidazolium cation with phenylalanine, tyrosine and tryptophan as in the previous chapter, as well as those formed with histidine are considered. An exhaustive exploration is performed for these complexes, employing the methods that have already proved to properly reproduce the interaction in chapter 7. Also, the effect of the solvent as modeled by a polarizable continuum model upon the characteristics of the complexes is evaluated.



USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3

Methodology

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

3.1. Methods based on the wavefunction

Macroscopic systems can be easily described by Newtonian mechanics, where the energy can apparently vary continuously. However, when a system has to be described at a microscopic level, a different mechanics is required, in which quantization is taken account: this is quantum mechanics.[1]

The fundamental postulate of quantum mechanics states that a so-called wavefunction exists, Ψ, for any state which describes the system, and that any observable property *a* of the system can be obtained by applying an appropriate operator upon that wavefunction. Thus, the goal of quantum mechanics is finding the system wavefunction. This postulate can be expressed as the following equation:

$$\hat{A}\Psi = a\Psi \qquad \text{(eq. 3.1)}$$

This equation is called the Schrödinger equation when the operator is the Hamiltonian, and the property obtained is the energy of the system:

$$\hat{H}\Psi = E\Psi \qquad \text{(eq. 3.2)}$$

Usually, the Hamiltonian operator without relativistic effects is taken into account, constructed from five contributions to the total energy of a system, being those contributions the kinetic energies of the electrons and nuclei; the attraction of the electrons to the nuclei; and the interelectronic and internuclear repulsions.

In the process of obtaining the best wavefunction which describes our system, its quality can be judged by evaluation of the energy eigenvalues associated to it. As stated by the variational principle, the wavefunction with the lowest energy will be the most accurate and presumably the best one. This trial wavefunction is usually represented by a combination of appropriate functions (LCAO approach), the basis set, that are commonly atomic orbitals (AOs) in chemistry. A large number of methods based on the wavefunction have been developed. In the next sections some of those will be briefly exposed.

3.1.1. *Ab initio* methods

The term *ab initio* comes from the Latin expression “from the beginning”. This kind of methods corresponds to those computational methods which draw upon theoretical principles, without relying in any experimental data to obtain the results. Quantum mechanics needs approximate methods to solve multiple electron systems, since the equations can be solved exactly only for

one electron systems. Thus, *ab initio* methods, as other many ones, were developed for this purpose using, in most cases, mathematical approximations.[2] In this section, the most popular *ab initio* methods and those ones used for obtaining the results in this present work are shown.

3.1.1.1. Hartree-Fock approximation

The Hartree-Fock approximation (HF) is a central method to quantum chemistry, and though it does not include electronic correlation, it has an important role as starting point for more accurate approximations.[3] This approximation takes into account the Born-Oppenheimer approach and also the non-relativistic Hamiltonian. Thus, the method can be described as the process of obtaining the best Slater determinant[3] set up by the molecular orbitals constituting the *N*-electron wavefunction of the system, taking into account the proper antisymmetry which satisfies the Pauli exclusion principle. All of these molecular orbitals are formed by linear combination of atomic orbitals (one-electron functions), the coefficients obtained as a result of the Hartree-Fock method.

According to Hartree-Fock theory, it is possible to approximate the wavefunction of the system through just a single Slater determinant formed by an specific set of *N* spinorbitals $\{\chi_i\}$ that results in the best approximation to the ground state of the *N*-electron system described by this Hamiltonian, those which minimize the electronic energy as the variational principle states. Thus, the simplest antisymmetric wavefunction, which can be used to describe the ground state of *N*-electron systems, is a single Slater determinant,

$$\left|\Psi_{HF}\right\rangle = \left|\chi_1\chi_2\cdots\chi_a\chi_b\cdots\chi_N\right\rangle. \qquad \text{(eq. 3.3)}$$

Particularly, the closed-shell Hartree-Fock electronic energy can be obtained from eq. 3.2 by using the electronic Hamiltonian, resulting in a three-term expression. According to the variational principle, the best wavefunction is the one which gives the lowest possible energy

$$E_{HF} = \left\langle \Psi_{HF} \middle| \hat{H} \middle| \Psi_{HF} \right\rangle, \qquad \text{(eq. 3.4)}$$

where $\hat{H}$ is the full electronic Hamiltonian. By minimizing E_{HF} with respect to the choice of spinorbitals, one can derive an equation, called the Hartree-Fock equation, which determines the optimal spinorbitals, and which is an eigenvalue equation of the form:

$$\hat{f}\chi_a = \varepsilon\chi_a, \qquad \text{(eq. 3.5)}$$

where $\hat{f}$ is the Fock operator, an effective one-electron operator, which includes the average potential experienced by the a^{th} electron due to the presence of the other electrons. By means of the Hartree-Fock approximation, the complicated many-electron problem is replaced by a one-electron problem in which electron repulsion is treated in an averaged way. The Hartree-Fock eq. 3.5 is nonlinear because the potential depends on the spinorbitals of the other electrons; thus, it has to be solved iteratively by means the so-called self-consistent-field (SCF) method. This process starts from a trying function (by making an initial guess for the spinorbitals) to obtain the average field. After solving eq. 3.5, a new set of spinorbitals is obtained. Then, the process is repeated until self-consistency is reached, giving a set of $\{\chi\}$ orthonormal Hartree-Fock spinorbitals with orbital energies $\{\varepsilon\}$ which are filled up with the electrons of the system. The Hartree-Fock ground state wavefunction, that is the best variational monodeterminantal approximation to the ground state of the system, is built by means of the single Slater determinant in which the *N*-spinorbitals with the lowest energies have been included.

Although in principle there is an infinite number of solutions, Hartree-Fock equation is solved by introducing a finite set of *K* spatial basis functions that, with their corresponding spin functions α and β, are expanded and substituted into the eigenvalue eq. 3.5 to obtain the *2K* spinorbitals $\{\chi_a\}$. The HF electronic energy for a closed shell system results in the following expression

$$E_{HF} = 2\sum_{a}^{N/2}\left\langle a\left|\hat{h}\right|a\right\rangle + \sum_{ab}^{N/2} 2\langle ab|ab\rangle - \langle ab|ba\rangle . \qquad \text{(eq. 3.6)}$$

The three terms define, respectively, the sum of the one-electron energies of the molecular orbitals; the Coulomb electron-electron interaction energy, *J*, and finally, the so-called exchange energy, *K*, arising from the antisymmetric restriction. Thus, the same Hartree-Fock energy of a closed shell system given in terms of coulomb and exchange integrals would be expressed as

$$E_{HF} = 2\sum_{a}^{N/2}\hat{h}_{aa} + \sum_{ab}^{N/2} 2\hat{J}_{ab} - \hat{K}_{ab} . \qquad \text{(eq. 3.7)}$$

In this expression, two kinds of operators can be distinguished, corresponding to a one-electron operator in the case of $\hat{h}$ and two-electron operators in the case of $\hat{J}$ and $\hat{K}$. These two latter operators include the average electron repulsion on one electron described by the spinorbital χ_a.

This method, although describes most of the system electronic energy, it does not take into account the correlation energy due to the electrons with different spin, since one of the

requirements is to consider the rest of the surrounding electrons as producers of an average potential. This fact, together with the use of a single Slater determinant to describe the wavefunction of the system, makes the HF method not suitable for obtaining accurate results.

The correlation energy not included within the HF method corresponds to the following expression:

$$E_{corr} = E_o - E_{HF}\,, \qquad \text{(eq. 3.8)}$$

being E_o the exact energy, i.e. the lowest eigenvalue of the Hamiltonian equation, the exact nonrelativistic ground state energy of the system within the Born-Oppenheimer approximation; E_{HF} is the Hartree-Fock limit energy.

Thus, to include this correlation energy it is common to expand the wavefunction from a single Slater determinant to a linear combination of a large number of them:

$$\left|\Psi\right\rangle = \sum_k c_k \left|\psi_k\right\rangle. \qquad \text{(eq. 3.9)}$$

This sum includes the excited configurations of the system, just changing the disposal of the different occupied and virtual orbitals arranged in the starting Slater determinant. Thus, the rest of Slater determinants involved on the combination represent the different configurations of the electrons, including the doubly, triply and higher excited determinants. This procedure is called Configuration Interaction (CI). Unfortunately, it is not possible to implement this procedure for a complete solution to the many-electron problem, since an infinite basis set cannot be handled. However, the procedure called *full CI* is possible, leading to solutions that are exact within the one-electron subspace spanned by the set of spinorbitals. However, even for relatively small systems and minimal basis sets, the number of determinants that must be included in a full CI calculation is extremely large. Thus, in practice, one must truncate the full CI expansion and use only a small section, but to do this truncation entails the problem of size inconsistency because not all the possible excitations are included, making this procedure unsuitable. To circumvent this problem other size-consistent methods have been developed in order to approximate its solutions to the FCI solution.

3.1.1.2. Møller-Plesset Perturbation Theory

As commented above, one of the limitations of the HF method is that it does not include electron correlation. Thus, only the average effect of electron repulsion is taken into account, but not the explicit electron-electron interaction. This is an important fact, since including correlation

generally improves the accuracy of computed energies and molecular geometries.[2] In order to include this effect, a number of types of methods were developed beginning with a HF calculation and then correcting for correlation. The Møller-Plesset Perturbation Theory is one of these methods, among others like multi-configurational self-consistent field (MCSCF), configuration interaction (CI) or coupled cluster theory (CC).

Møller-Plesset perturbation theory, typically referred to by the acronym MP*n*, was developed in 1934 by Christian Møller and Milton S. Plesset, and it improves the Hartree-Fock method by adding electron correlation effects by means of Rayleigh-Schrödinger perturbation theory (RS).[1] The subscript *n* denotes the order at which perturbation theory is truncated, e.g. MP2, MP3, etc. This MP perturbation theory is a particular case of many-body perturbation theory, (MBPT)[4] in which the operator used is the Hamiltonian to which a small perturbation *V* is added:

$$\hat{H} = \hat{H}_o + \lambda \hat{V} . \qquad \text{(eq. 3.10)}$$

In this expression, λ is an arbitrary parameter that controls the size of the perturbation, and the perturbation $\hat{V}$ is associated with the correlation potential. This method is not variational but is size consistent at each level. The unperturbed Hamiltonian $\hat{H}_o$ is taken to be the sum of the one-electron Fock operators for the MP approach as in the following expression:[1]

$$\hat{H}_o = \sum_i \hat{f}_i . \qquad \text{(eq. 3.11)}$$

Thus, it can be easily seen that the perturbation can be expressed as the difference between the perturbed and the unperturbed Hamiltonians;

$$\lambda \hat{V} = \hat{H} - \sum_i \hat{f}_i = \hat{H} - \hat{H}_o . \qquad \text{(eq. 3.12)}$$

In this MP method, the new energy and wavefunction must change continuously when the perturbation is continuously increased, and thus, they can be written as a Taylor expansion in powers of the perturbation parameter λ;

$$\psi_i = \psi_i^{(o)} + \lambda \psi_i^{(1)} + \lambda^2 \psi_i^{(2)} + \lambda^3 \psi_i^{(3)} + \ldots \qquad \text{(eq. 3.13)}$$

$$E_i = E_i^{(o)} + \lambda E_i^{(1)} + \lambda^2 E_i^{(2)} + \lambda^3 E_i^{(3)} + \ldots \qquad \text{(eq. 3.14)}$$

Thus, $\Psi^{(1)}$, $\Psi^{(2)}$,… and $E^{(1)}$, $E^{(2)}$, …are the first-order, second-order, etc., corrections, and truncating the expansion in those ones, the zero-, first-, second-, nth-order perturbation equations

are obtained. The most used Møller-Plesset calculations are MP2, MP3 and MP4, which includes doubly (MP2 and MP3) or until quadruply-excited determinants (MP4). The zero-order equation is just the Schrödinger equation for the unperturbed problem.[1] If every excited-determinant were included in the expansion, the FCI solution would be ideally achieved.[5]

In addition, $\psi^{(0)}$ is taken to be the HF wavefunction, a Slater determinant formed from the occupied orbitals,

$$\hat{H}_o \psi^{(0)} = \sum_i^{occ} \varepsilon_i \psi^{(0)} \quad , \qquad \text{(eq. 3.15)}$$

where the orbital energies are the usual eigenvalues of the specific one-electron Fock operators. The energy correction obtained through first-order in Moller-Plesset perturbation theory (MP1) corresponds to the Hartree-Fock energy. In this way, as MP1 does not provide more information about the total energy than HF method, it is necessary to consider the second-order correction to obtain an estimate of correlation energy, and that is:[5]

$$E(MP2) = \sum_{i<j}^{occ} \sum_{a<b}^{vir} \frac{\left(\left\langle \phi_i \phi_j \middle| \phi_a \phi_b \right\rangle - \left\langle \phi_i \phi_j \middle| \phi_b \phi_a \right\rangle \right)^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \quad . \qquad \text{(eq. 3.16)}$$

It could be said that with each increase on the calculation order, i.e. from MP2 to MP3 and so on, the solution would be more similar to the FCI solution. However, there is no reason to expect that an MP*n* calculation will give a value for the correlation energy that is particularly good, since the perturbation is the full electron-electron repulsion energy, which is a rather large contributor to the total energy, and the perturbation theory works best when the perturbation is small. Also, the non-variational MP*n* methodology makes uncertain, in principle, the estimated value of MP2 for the correlation energy. Furthermore, in practice, only low orders of perturbation theory can be routinely carried out.

Using MP*n* theory, the more typical observation for a given property is oscillatory when the order of perturbation is increased.[1] Ideally, the MP*n* results would show convergent behavior as a function of *n*, but it does not usually occur, and for that reason it can be difficult to extrapolate. MP2 tends to underestimate the energy, especially in complexes with large dispersion contributions, and MP3, although gives an overestimated value of the energy, in the direction of HF results, tends to offer rather little improvement over MP2. MP4 increases the difference again, but in favorable cases by only a small margin. A typical behavior of the property according to the method is represented in Figure 3.1.[5]

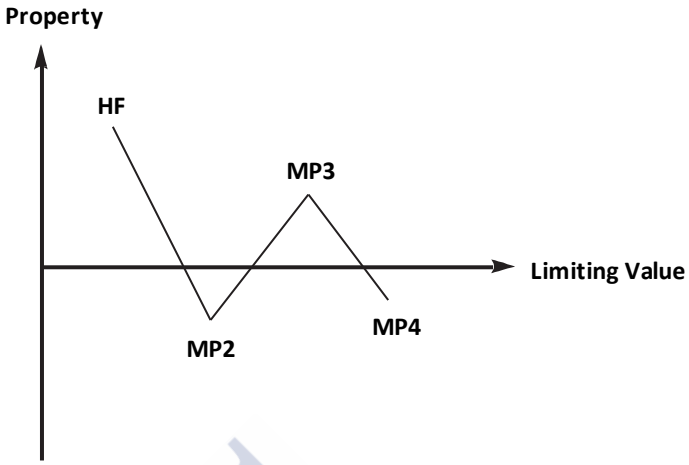


Figure 3.1. Typical behavior of MP_n energies with increasing order.

In practical calculations, MP2, MP3 and MP4 typically account for 80-90%, 90-95% and 95-98% of the correlation energy, respectively, employing a good basis set, but it is MP2 the most economical wavefunction method for including electron correlation, with a computational cost similar (though larger) to that required for calculating the HF energy.[5] Also, the MP2 method has a good behavior when intermolecular interactions are involved in a system, since it includes correlation and more specifically dispersion. Thus, MP2 method is a good choice from which carrying out an improvement. Based on this idea, an empirical modification with better results, the so-called Spin Component Scaled MP2 method (SCS-MP2), was proposed by Grimme.[6, 7] This approach consists on performing an empirical scaling of the contributions to correlation energy. Thus, electron pairs with equal spin contribute to the energy by a different factor than the electron pairs with different spin. This approach together with the one proposed by Hill and Platts (SCSN-MP2),[8] can be summarized in the following general expression:

$$E_{X-MP2} = E_{HF} + p_t E_{(\alpha\alpha+\beta\beta)} + p_s E_{(\alpha\beta)} \,. \qquad \text{(eq. 3.17)}$$

Both approaches are based on the same idea, although they use different coefficients to obtain the energy. In the first case, for the SCS-MP2 approach, the coefficients are p_t=*0.33* and p_s=*1.20*; meanwhile for the second approach, SCSN-MP2, the coefficients are *1.76* and *0.00*, respectively. The choice of the coefficients is made with the aim to improve the MP2 method,[7, 9-13] and also to take into account the basis incompleteness of the basis employed in each case. Other similar approaches have been developed,[14-19] since in most systems it has been observed to be a good way to obtain good results with a moderate computational cost.

3.1.1.3. Coupled Cluster method

The coupled cluster approach, although belongs to another category of post Hartee Fock methods than CI or MBPT, it is closely related to them. In the coupled cluster method, like in CI or MBPT, excitations are included in the expression of the wavefunction. However, in this case, is not by means of a linear combination of configurations but through an exponential Ansatz, including the so-called cluster operator, $\hat{T}$. This approach ensures the correct separability of the different fragments and also the correct size consistency of the method.[2, 20]

In this way, the basis of the coupled cluster method is that the wavefunction can be described by a product between the HF wavefunction and an exponential operator to include the different excited determinants,[1]

$$\psi = e^{\hat{T}}\psi_{HF} \text{; with } \hat{\Omega} = e^{\hat{T}}, \qquad \text{(eq. 3.18)}$$

where $\hat{\Omega}$ is the wave operator, which transforms the reference wavefunction into the exact one. The expression of this operator differs from the methods indicated above, being an exponential of the cluster operator for the coupled cluster method. This cluster operator $\hat{T}$ is defined as:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_n, \qquad \text{(eq. 3.19)}$$

where n is the number of electrons and the $\hat{T}_i$ operators are the excitation operators for the respective i excitation state. Thus, $\hat{T}_i$ generates all the possible determinants having i excitations from the reference. As the cluster operator is included in the exponential coefficient in eq. 3.18, a Taylor expansion can be done, being the final expression for the coefficient the following one;

$$e^{\hat{T}} = \hat{\Omega} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + ... \qquad \text{(eq. 3.20)}$$

For example, for two excitations, the effect of the operator $\hat{T}_2$ is expressed as:

$$\hat{T}_2\psi_{HF} = \sum_{i<j}^{occ}\sum_{a<b}^{vir} t_{ij}^{ab}\psi_{ij}^{ab}, \qquad \text{(eq. 3.21)}$$

where t_{ij}^{ab} are the amplitudes that must be determined by the CC calculations.

If all cluster operators up to $\hat{T}_n$ were included in $\hat{T}$, the coupled cluster wavefunction would be equivalent to full CI because all possible excited determinants would be generated. But, as in CI, a truncation of the expansion is necessary. The advantage of coupled cluster over full CI appears when this truncation of the expansion is done. Truncation in CI at the *i* level implies that only states up to *i* excitations are taken into account, making the method non-size consistent. However, when the truncation in CC is done in a specific level of excitation, higher levels are taken into account due to the own definition of the cluster operator. For instance, if only the double-excitation operator is consider (CCD), $\hat{T}=\hat{T}_2$ the Taylor expansion would be:

$$\Psi_{\mathbf{CCD}} = e^{\hat{T}}\psi_{HF} = \left(1+\hat{T}_2+\frac{\hat{T}_2^{\,2}}{2!}+\frac{\hat{T}_2^{\,3}}{3!}+\ldots\right)\psi_{HF}. \qquad \text{(eq. 3.22)}$$

Thus, as $\hat{T}_2$ is the double-excited operator, its products give the multiple of two excitations. The third term into the parenthesis corresponds to the quadruple excitations, the fourth corresponds to the hextuple excitations, etc, ensuring size consistency.[5] If also the single excitations $\hat{T}_1$ are included, the method is called CCSD. The expression for the CCSD wavefunction is then described by the following expression:

$$\Psi_{\mathbf{CCSD}} = e^{(\hat{T}_1+\hat{T}_2)}\psi_{HF} = \left(\begin{array}{l} 1+(\hat{T}_1+\hat{T}_2)+\frac{1}{2!}(\hat{T}_1+\hat{T}_2)^2 \\ +\frac{1}{3!}(\hat{T}_1+\hat{T}_2)^3+\ldots \end{array}\right)\psi_{HF}. \qquad \text{(eq.3.23)}$$

Thus, the exponential operator may be arranged by groups with the same excited states,

$$e^{\hat{T}} = 1+\hat{T}_1+\left(\hat{T}_2+\frac{1}{2}\hat{T}_1^{\,2}\right)+\left(\hat{T}_3+\hat{T}_1\hat{T}_2+\frac{1}{6}\hat{T}_1^{\,3}\right)+ \left(\hat{T}_4+\hat{T}_1\hat{T}_3+\frac{1}{2}\hat{T}_2^{\,2}+\frac{1}{2}\hat{T}_2\hat{T}_1^{\,2}+\frac{1}{24}\hat{T}_1^{\,4}\right)+\ldots \qquad \text{(eq. 3.24)}$$

In this way, it can be appreciated that the first term corresponds to the reference HF wavefunction and the second to all singly excited determinants. In the first parentheses, with the doubly excited determinants, two kinds of terms appear: *connected* ($\hat{T}_2$) and *disconnected* ($\hat{T}_1^{\,2}$), and in the rest of parenthesis it can be observed the same, with "true" ($\hat{T}_i$) or "product" excitations which result in an *i* excitation. Physically, a connected type such as $\hat{T}_4$, corresponds to

four electrons interacting simultaneous, while a disconnected term such as $\hat{T}_2^2$, corresponds to two non-interacting pairs of interacting electrons.

This CCSD method is the most common choice since it includes up to double excitation operators, and has a computational cost not too high.[1, 2, 20] However, there are other orders of CC expansion as CCSDT[21] and so on, in which higher excitation operators are included. The drawback is the huge computational cost they imply. Another option to estimate the effects of the connected triples is to use CCSD(T) calculations,[22] in which the triple excitations are not included by adding the triple-excited operator but by means of a perturbation approach. Indeed, CCSD(T) calculations constitute a very robust method and one of the most used as standard.[23]

3.2. Improving the results

Any calculation performed using a wavefunction-based method is obviously affected by errors. Apparently, these errors come from two different sources:

1) from the one-electron basis functions, the building blocks of the *N*-electron wavefunction.
2) from the method itself, and its difficulties for describing the correlation properly.

In the first case, the error is related to the fact that an incomplete basis set is employed. The larger the basis set employed, the smaller would be the associated error. In the second case, methods approaching to FCI solution as MP*n* or CC need to be truncated to make the process viable. This truncation causes loss of information, carrying then an associated error compared to FCI results. By including higher orders of MP*n* or CC, the error would decrease, but the computational costs would increase considerably.

Therefore, taking into account these sources of errors, the best way to reduce them is the choice of a good correlation method combined with a large basis set, always according to our computing resources. In practice, when accurate results are desired, it is usual to employ MP2 or CCSD(T), depending on the size of the system and the available resources.

3.2.1. Extrapolation to basis limit

As commented in the previous section, the error associated to the one-electron basis set could be reduced by using larger basis sets. However, the continuous shift towards larger basis sets does not produce a convergence as quick as it would be desirable. Conversely, it converges very slowly as the size of the basis set is increased, being impractical or computationally expensive the solution of using very large basis set.

Besides this procedure, explicitly correlated methods can be employed, which tend more quickly to the limit than the standard ones.[24, 25] An intermediate solution consists on extrapolating the results obtained with moderate sized basis sets, obtaining then an estimation of the limiting value.[26-32]

However, this last solution requires a smooth behavior of the energy (smoothly varying as the basis set increases) to construct a fitting function, and then approaching the limiting value. For example, the correlation consistent cc-pVXZ family of basis sets proposed by Dunning is a good choice for this purpose.[33] Once the basis set is chosen, it has to be taken into account that the behavior of the energy can be split into two different behaviors, that one due to the HF energy and that one due to the correlation contributions to the energy.

Firstly, regarding to the HF energies, it can be assumed an approximately exponential behavior in molecules as in the case of atoms.[24, 34, 35] Thus, the following expression for the behavior of the HF energy is frequently employed:

$$E_X^{HF} \approx E_{CBS}^{HF} + Ae^{-BX}, \qquad \text{(eq. 3.25)}$$

where X is the cardinal number of the basis set. Thus, if three calculations are performed with correlation consistent basis set of increasing X, the limiting value could be estimated as:

$$E_{CBS}^{HF} \approx \frac{E_X^{HF} - bE_{X-1}^{HF}}{1-b}; b = e^{-B} = \frac{E_X^{HF} - E_{X-1}^{HF}}{E_{X-1}^{HF} - E_{X-2}^{HF}}. \qquad \text{(eq. 3.26)}$$

Although this method gives a good approximation, it is still computational expensive since it requires three calculations of increasing size. Other two-point extrapolation schemes have been developed, as the one proposed by Karton and Martin;[36, 37]

$$E_X^{HF} = E_{CBS}^{HF} + A(X+1)\exp(-9\sqrt{X}). \qquad \text{(eq. 3.27)}$$

These errors associated to HF energies are not the main problem, since HF energies converge quite quickly with the basis set, especially considering energy differences, as interaction energies. Also, since errors due to correlation are usually larger, basis sets as cc-pVTZ seem to be enough for a proper description of the HF contribution.

However, regarding to the correlation energies, it has to be said that the dependency on the basis set size is stronger and also that the convergence with the basis set is even slower. For that reason, larger basis set would be needed. The correlation energy usually behaves as:

$$E_{CBS} \approx E_X + AX^{-3}. \qquad \text{(eq. 3.28)}$$

The correlation energy can be obtained then by performing two calculations (since it contains only two unknowns):

$$E_{CBS} = E_X + AX^{-3}, \qquad \text{(eq. 3.29)}$$

$$E_{CBS} = E_Y + AY^{-3}. \qquad \text{(eq. 3.30)}$$

So,

$$E_{exact} = \frac{X^3 E_X - Y^3 E_Y}{X^3 - Y^3}. \qquad \text{(eq. 3.31)}$$

The advantage of this scheme relies on the fact that employing lower order basis sets as cc-pVTZ and cc-pVQZ, better results are obtained than using directly a huge basis set as cc-pV6Z, reducing considerably the computational costs.

3.2.2. Benchmark values

Usually, MP2 estimations are used instead of better ones as CCSD(T) to obtain values at the CBS limit for the correlation energy. For these procedures, CCSD(T) method requires computational resources that in cases of systems of moderate size are unviable. However, although MP2 is less demanding, also entails an error due to the *N*-electron model employed.

Therefore, in order to estimate the CCSD(T)/CBS values but with a reasonable computational cost, some approaches have been devised, many of them by Hobza.[38, 39] These methods proposed by Hobza and collaborators, are based on the fact that the differences between the correlation energy contributions to the interaction energy of a dimer using MP2 and CCSD(T) are pretty independent of the basis set size.[18, 38, 39] Thus, it is possible to perform an approximation of the CCSD(T)/CBS correlation energy by using a correction with a small basis set (term in parentheses), since the difference between CCSD(T) and MP2 is assumed constant:

$$\Delta E^{corr,CBS}_{CCSD(T)} = \Delta E^{corr,CBS}_{MP2} + \left(\Delta E^{corr,smallbasis}_{CCSD(T)} - \Delta E^{corr,smallbasis}_{MP2} \right), \qquad \text{(eq. 3.32)}$$

and where the MP2 contribution to the correlation energy is estimated to basis limit with the extrapolation procedures described above. Although this procedure is usually employed to obtain

benchmark values for the interaction energies of complexes, it has to be said that still entails errors associated to the small basis calculation.[40-42]

The following step within this approach is to avoid the calculation of the CCSD(T)/smallbasis value, still demanding for moderate size systems. Procedures for estimating the CCSD(T) corrections have been proposed by Hobza and Rezac (MP2.5[43] and the following improvements MP2.X[44]). The MP2.X approach consists of substituting the CCSD(T) calculations by cheaper MP3 ones, and also uses and empirical scaling coefficient. This coefficient is obtained by fitting to the CCSD(T)/CBS estimates of a set of complexes of different nature. The limit is estimated as:

$$\Delta E_{CCSD(T)}^{corr,CBS} = \Delta E_{MP2}^{corr,CBS} + C\left(\Delta E_{MP3}^{corr,smallbasis} - \Delta E_{MP2}^{corr,smallbasis} \right), \quad \text{(eq. 3.33)}$$

given a proper *C* coefficient in every case, it has been seen that the accuracy of the MP2.X results is almost independent of the small basis set employed. Besides, even with a basis set as small as 6-31G* the results are similar to those obtained from CCSD(T) calculations.[45, 46] Thus, the computational costs are considerably reduced by using both a smaller basis set and a lower-level method.

3.3. Density Functional Theory methods

The density functional methods arose from the idea of an interpretable or more intuitive method than those derived from the wavefunction.[1] Thus, instead of studying the wavefunction of the system, these methods work with a useful physical observable in determining the energy of the system. This physical observable is the electron density ρ because it lets the construction of the Hamiltonian operator depending only on the positions and atomic numbers of the nuclei and the total number of electrons; integrated over all space, ρ also gives the number of electrons *N* of the system. The electron density is a good choice since as the nuclei are effectively point charges, the local maxima in the electron density correspond to these positions. Besides, the advantage of the electron density over the wavefunction for the calculations relies on the fact that, since the electron density is the square of the wavefunction, the electron density has the same number of variables when the number of electrons increase, so it is independent of the system size, making the electron density a useful tool for that purpose. The problem is then to find the exact functional which gives the relation between the density and the ground state energy of the system. DFT methods have been developed with this aim.[47-52]

There were some previous approximations to the DFT methods in which only the electron density was used as a variable to try to evaluate the molecular energy, knowing that it is

separable into kinetic and potential components. Hohenberg and Kohn in 1964 were the ones who proved two critical theorems for establishing DFT as a legitimate quantum chemical methodology, and basically as a proof that the electron density ρ determined completely the ground state electronic energy.

The first Hohenberg–Kohn theorem states that knowing the electron density of a stationary non-degenerate ground state of any system, any observable existing in this state can be calculated from it.[5, 53] This is because observables like the energy are functionals of the electron density of the ground state of the system.[1] Thus, the electron density determines the Hamiltonian (except for an additive constant) so there is a direct relationship between the density and the wavefunction through the external potential:

$$E[\rho] = T[\rho] + V_{Ne}[\rho] + V_{ee}[\rho]. \qquad \text{(eq. 3.34)}$$

$T[\rho]$ and $V_{ee}[\rho]$ do not depend on the external potential and can be included within the Hohenberg-Kohn functional $F_{HK}[\rho]$, with which

$$E_v[\rho] = \int \rho(\mathbf{r})\, v(\mathbf{r})\, d\mathbf{r} + F_{HK}[\rho], \qquad \text{(eq. 3.35)}$$

where $E_v[\rho]$ indicates that, for a specific external potential $v(\mathbf{r})$, the energy is a functional of the density.

On the other hand, the second Hohenberg–Kohn theorem is a variational theorem, which states that determining the density that minimizes the energy of the ground state it is possible to theoretically obtain, in an exact way, the electron density of a non-degenerate ground state.

3.3.1. The Kohn and Sham method

With those two theorems in hand, the energy may be minimized to determine the density of the ground state.[20] The main problem is the unknown expression of the relation between F_{HK} and the density, in particular the *T[ρ]* form. Trying to solve this problem, Kohn and Sham[54] proposed an ingenious method, similar in structure to the Hartree-Fock method, to calculate the energy from ρ, using a non-interacting electron system as a reference which, after being applied an external potential $v_s(\boldsymbol{r})$, provides a wavefunction, ψ_s, which has the same density as the real system.

As commented above, Hartree-Fock and density functional theories have a very similar mathematical development, and thus, both need the Hamiltonian of the system to get the energy.[5] However, in DFT methods an approximate Hamiltonian is employed instead of the exact one, since an ideal system is considered, without electron-electron interactions, providing a potentially

exact result with the same density than the real interacting electron system. The Hamiltonian of this system only contains the single-electron terms, meanwhile the exact wavefunction is the Slater determinant, constructed from the so-called Kohn-Sham orbitals:

$$\hat{H}_s = \sum_{i=1}^{N} \hat{h}(i) = \sum_{i=1}^{N} -\frac{1}{2}\nabla^2(i) + \sum_{i=1}^{N} \hat{v}_s(i)\,, \qquad \text{(eq. 3.36)}$$

$$\psi_s = \frac{1}{\sqrt{N!}} \left| \chi_1(1)\chi_2(2)\chi_3(3)...\chi_N(N) \right|. \qquad \text{(eq. 3.37)}$$

These unknown Kohn-Sham orbitals, which minimize the energy, may be determined by numerical methods, or expanded in a set of basis functions, analogously to the HF method:

$$\left[-\frac{1}{2}\nabla^2 + v_s(\mathbf{r}) \right] \chi_i = \varepsilon_i \chi_i\,; \qquad \left\langle \chi_i \middle| \chi_j \right\rangle = \delta_{ij}\,. \qquad \text{(eq. 3.38)}$$

Also, the exact electron density for this system is expressed as a linear combination of these orbitals, and it is used to calculate the energy,

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_{OCC}} \left| \chi_i(\mathbf{r}) \right|^2. \qquad \text{(eq. 3.39)}$$

Following this method, the energy functional of the real system may be divided into three parts (implicitly including correlation energy in all the terms): kinetic energy, $T[\rho]$, and potential energy split into the attraction between the nuclei and electrons, $E_{ne}[\rho]$, and electron–electron repulsion, $E_{ee}[\rho]$ (the nuclear–nuclear repulsion is a constant within the Born-Oppenheimer approximation). Also, the $E_{ee}[\rho]$ term may be divided into Coulomb ($J[\rho]$) and exchange ($K[\rho]$) parts.

On the other hand, this Kohn-Sham procedure proposes the division of the kinetic energy functional into two terms, one which can be calculated exactly, and a small correction term due to the difference between the exact kinetic energy and that calculated by assuming non-interacting orbitals, it is taken account into the exchange-correlation term. Thus, the general DFT energy expression would be:

$$E_{\mathrm{DFT}}[\rho] = T_{\mathrm{s}}[\rho] + E_{\mathrm{ne}}[\rho] + J[\rho] + E_{\mathrm{xc}}[\rho]\,, \qquad \text{(eq. 3.40)}$$

where $T_s[\rho]$ corresponds to the kinetic energy functional described by the Slater molecular orbitals, and the rest expressions of the components of the eq. 3.40 are below:

$$T_s[\rho] = \sum_{i=1}^{N_{occ}} \left\langle \chi_i \left| -\frac{1}{2}\nabla^2 \right| \chi_i \right\rangle, \qquad \text{(eq. 3.41)}$$

$$E_{ne}[\rho] = -\sum_{a}^{N_{nuclei}} \int \frac{Z_a(\mathbf{R}_a)\rho(\mathbf{r})}{|\mathbf{R}_a - \mathbf{r}|} d\mathbf{r}, \qquad \text{(eq. 3.42)}$$

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \qquad \text{(eq. 3.43)}$$

Due to the fact that the electron interaction exists, the kinetic energy provided by $T_s[\rho]$ is not the total kinetic energy (although HF theory yields the 99% approximately) and it is necessary to take it into account the difference, although small, between the exact kinetic energy of the real system and that calculated by assuming non-interacting electrons, the correlation kinetic energy. This difference is included in eq. 3.44 , in which it is also included the difference between E_{ee} and J potential energy terms:

$$E_{xc}[\rho] = \left(T[\rho] - T_s[\rho]\right) + \left(E_{ee}[\rho] - J[\rho]\right). \qquad \text{(eq. 3.44)}$$

The first parentheses corresponds to the kinetic correlation energy and in the second one coexist both exchange and potential correlation energy, although the exchange energy is by far the largest contributor to E_{xc}. All the contributions to the energy without a simple expression as a function of the electron density are included within the $E_{xc}[\rho]$. However, this exact $E_{xc}[\rho]$ functional is not known, and thus, as commented above, the aim of DFT methods is to design functionals connecting the electron density with the energy.[47-52] Compared with HF theory, DFT provides much better results. Also, if the exact $E_{xc}[\rho]$ were known DFT would provide the exact total energy including the correlation. However, there is an important problem with density functional methods that is their inability to systematically improve the results and the known failure to describe certain important features, such as van der Waals interactions. Anyway, different functionals used for DFT methods have been developed assuming different approximations being then suitable for different kind of systems with acceptable accuracy. The method to construct the different functionals vary from quantum mechanics to methods in which parametric functions are used to best reproduce experimental results.[2] Which functional is the best one will have to be settled by comparing the performance with experiments or high-level wave mechanics calculations.

3.3.2. Functional types

In early attempts on the DFT development all the energy components were expressed as a functional of the electron density; however, it was seen that this led to no good performance. Thus, instead of this, the modern methods use the electron density, represented by means of an auxiliary set of orbitals to calculate the electron kinetic energy (Kohn and Sham, 1965),[54] and use approximations to calculate the exchange–correlation energy, which is the only unknown functional. There are many stages of development for the functionals. The so-called Local Density Approximation, LDA, is the simplest approximation, and it is based only on the electron density. In this approximation, the electron density can be treated as a uniform electron gas or, equivalently, as a slowly varying function, with the advantage that the calculations of the correlation energy of a uniform gas have been already determined by Monte Carlo methods for a number of different densities. For that reason, LDA approximation works very well in systems in which the density is maintained approximately constant. In 1980, Vosko, Wilk and Nusair (VWN)[55] constructed a suitable analytic interpolation formula in order to use these results in DFT calculations.

In the LDA approximation $\varepsilon_{xc}[\rho]$ is a functional that depends exclusively on the density. Contributions to correlation and to exchange are usually treated separately as:

$$\varepsilon_{xc}^{LDA}[\rho] = \varepsilon_{x}^{LDA}[\rho] + \varepsilon_{c}^{LDA}[\rho], \qquad \text{(eq. 3.45)}$$

being the exchange energy per particle

$$\varepsilon_{x}^{LDA}[\rho] = -\frac{9}{4}\alpha\left(\frac{3}{8\pi}\right)^{\frac{1}{3}}\rho^{\frac{1}{3}}. \qquad \text{(eq. 3.46)}$$

Thus, the correlation energy per particle can take different values. The simplest method is known as X_α, takes $\varepsilon_c[\rho]=0$ and $\alpha=2/3$, and thereby this method includes electron exchange but not correlation.

When open-shell systems are used, the method is called the local spin density approximation (LSDA). Both LDA and LSDA in general underestimate the exchange energy by ~10%, creating larger errors than the whole correlation energy.[5] Also, electron correlation is overestimated, often by a factor close to 2, implying at the same time the overestimation of the bond strengths. However, LSDA methods often provide results with similar accuracy to that obtained by wave mechanics HF methods, and work very well in systems in which the density is maintained

approximately constant.[5, 52] The main problem is that they still fail to describe van der Waals complexes.

Different improvements have been developed regarding to the accuracy of the method or the necessity to consider, in many cases, a non-uniform electron gas.[1, 2] Besides, the LDA methods are also not suitable for systems with weak bonds or for making reliable thermochemical predictions, overestimating in general the bond energy by approximately 30%.[20] The simplest improvement is obtained by means of the so-called gradient-corrected methods, in which the electron density and its gradient are used to construct the set of exchange-correlation functionals. These methods are also known as Gradient Corrected or Generalized Gradient Approximation (GGA) methods, being also sometimes referred to as non-local methods. To develop these methods, Becke proposed to include an additional term starting from the LDA functional, and this is the usual way to proceed. There are many other non-local corrections for both the exchange part and the correlation part, different from Becke's correction, although the most often used for exchange is Becke's. Among the corrections developed for the exchange energy there are for example PW86,[56] developed by Perdew and Wang, or B88 proposed by Becke,[57] which include a β parameter determined by fitting to known atomic data. Among the corrections aimed to improve the correlation energy, one popular functional is LYP,[58, 59] proposed by Lee, Yang and Parr, where parameters are determined by fitting to data for helium. The widely used B-LYP functional comes from the combination of these exchange and correlation functionals.[57-59]

Another way to correct the functionals is by using hybrid methods. These methods combine functionals from other methods with part of HF calculations, usually the exchange integrals. This approximation can be developed when a uniform gas of non-interacting electrons is considered since there is no correlation energy but only exchange energy; indeed the exact exchange energy is given by HF method (the wavefunction for these systems corresponds to the single Slater determinant). Thus, the exchange and correlation energies could be expressed as the sum of the LSDA and a gradient correction term of exchange and correlation respectively, and including also the exact exchange for the exchange energy, as in the following expression:

$$E_{xc}^{hybrid\,method} = (1-a)E_x^{LSDA} + a\,E_x^{exact} + b\,\Delta E_x^{GGA} + E_c^{LSDA} + c\,\Delta E_c^{GGA} \quad \text{(eq. 3.47)}$$

The *a*, *b* and *c* parameters depend on the chosen forms for E_x^{GGA} and E_c^{GGA}, and are determined experimentally. Becke 3 parameter functional (B3) method is an example of such hybrid methods, in which the popular GGA exchange functional B88 is used as a correction to the

LSDA exchange energy. Also, this B3 is part of the B3LYP functional, in which the GGA correlation term corresponds to the Lee-Yang-Parr correlation functional:

$$E_{xc}^{B3LYP} = (1-a)E_x^{LSDA} + a\,E_x^{exact} + b\,\Delta E_x^{B88} + (1-c)E_c^{LSDA} + c\,E_c^{LYP}. \qquad \text{(eq. 3.48)}$$

This B3LYP functional is one of the most famous functionals of DFT and therefore one of the most employed.[58-60] Subsequent versions denoted B97 and B98 employed ten fitting parameters,[27] but the improvements were rather marginal relative to the three parameters version.

Obtaining the hybrid meta-GGA functionals corresponds to the following step in the development of the functionals, combining meta-GGA functionals, in which the kinetic energy is included in the functional, and HF exchange. These methods are quite efficient and they have been very successful for the analysis of many problems in chemistry. There are some functionals in this group as the so-called Minnessota functionals, developed by Zhao and Truhlar, that are even competitive with the famous B3LYP and others approaches for a wide range of chemical systems.[61-67]

One of the most commonly used functionals is the M06 family, which it is constituted by four functionals. Every functional differs from each other according to the optimized parameters obtained to be used with a different percentage of HF exchange, being similar the form of the DFT part though.[61, 62] M06-L is a non-hybrid functional, and thus, does not include HF exchange. M06 and M06-2X include 27% and 54% HF exchange respectively, being good functionals for the description of the overall chemistry, except for multireference systems such as many systems containing transition metals. The last functional constituting this set is M06-HF, with 100% of HF exchange.

Specifically, M06-2X and, from the also known M05 family, M05-2X, have been shown to describe non covalent interactions better than density functionals which are currently in common use.[68] Description of dispersion interactions, critical in non covalent complexes, are inherently long-range electron correlation effects, which are not captured by the popular local or semilocal density functionals, and thus, limit the application of DFT to non covalently bound complexes.[69-75] The highly parametrized M05-2X and M06-2X functionals are an alternative, like other several approaches which handle dispersion effects,[61, 62, 76, 77] to improve this description. According to Zhao and Truhlar, the M05 and M06 series of functionals implicitly account for "medium-range" electron correlation (complexes separated by about 5 Å or less) because of the way they are parametrized, and this is sufficient to describe the dispersion interactions within many complexes, especially with M06-2X.[62, 68, 78]

In the fifth step of the Jacob's ladder (in which improved density functionals for the exchange-correlation energy are constructed) are the double hybrid functionals, firstly proposed by Donald Truhlar in 2004 and the more well-known by Stefan Grimme in 2006.[78, 79] These ones use both HF and second-order perturbation theory (double hybrid) to perform corrections to the exchange and correlation parts respectively. Unoccupied KS orbitals are included in the calculation in a similar way as in MP2. The use of the perturbational correction in this approximation is made with the aim of reducing the deficiencies of the functionals. As a consequence, both diffuse orbitals and kinetic barriers can be improved. Besides, due to the presence of this perturbational term, dispersion forces (van der Waals) can be more accurately computed. However, even with these improvements, these functionals result in too weak dispersion forces.

3.3.3. Dispersion-corrected DFT methods

Although the DFT methods include electron correlation effects in an approximate manner, almost all gradient corrected density functionals are unable to describe dispersive interactions.[70, 71, 80, 81] These DFT methods tend to underestimate the energies for many types of functionals. However, since dispersive interactions are involved in interactions as important as hydrogen bonds or π stacking, it was necessary to develop new ways to improve the results.

The main problem of DFT methods when dispersion interaction wants to be analyzed lies in its intrinsic local or semi-local density functional structure. Dispersion interactions are within the long-range, nonlocal effects and, although local density functionals can properly describe short and even medium-range effects, they are not capable to provide accurate results for the long-range ones. To do so, a fully nonlocal functional must be applied to the system.

A different and simple approach for introducing dispersion in DFT is to supply the already available functionals with an empirical dispersion term in the so-called DFT-D methods:[80-88]

$$E_{DFT-D} = E_{KS-DFT} + E_{disp} \quad \text{(eq. 3.49)}$$

This empirical dispersion term, properly parameterized, highly improves the capability of the functional to describe non covalent interactions[89] without appreciably increasing the computational cost, recovering part of the dispersion contribution not described by the initial non-corrected functional.

The DFT+Dispersion method was initially proposed by Wu and Yang[90] and assiduously developed by Grimme,[80, 82, 83] using a dispersion energy series to construct the dispersion energy term. Particularly for DFT-D2 method, this empirical dispersion correction is given by

$$E_{disp} = -s_6 \sum_{i=1}^{N_{at}-1} \sum_{j=i+1}^{N_{at}} \frac{C_6^{ij}}{R_{ij}^6} f_{dmp}(R_{ij}), \qquad \text{(eq. 3.50)}$$

where N_{at} is the number of atoms in the system, C_6^{ij} denotes the dispersion coefficient for atom pair ij, s_6 is a global scaling factor that only depends on the functional used, and R_{ij} is an interatomic distance. In order to avoid near-singularities for small R, a damping function f_{dmp} must be used:

$$f_{dmp}(R_{ij}) = \frac{1}{1 + e^{-d(R_{ij}/R_r - 1)}}, \qquad \text{(eq. 3.51)}$$

where R_r is the sum of atomic vdW radii, and d provides larger corrections at intermediate distances.

The dispersion correction to the DFT method greatly improves the results obtained for non covalent interactions, particularly those with significant dispersion contributions as in stacked systems. However, the model can be improved as shown by Grimme in the so-called DFT-D3 method. The most relevant improvement of the DFT-D3 method over the D2 version is the definition of environment coefficients, capable of adapting to different atomic environments within molecular systems.[83,84]

Other approaches to correct the dispersion problem are the density functional approaches based on the vdW-DF-04 of Langreth and Lundqvist[91,92] and approaches which go beyond pairwise additivity, such as many-body dispersion and the random-phase approximation (RPA).[93]

3.4. Interaction energy

All this work is focused on the interaction between different molecules and ions, so knowing the meaning of and how the interaction energy is obtained is essential. The concept of interaction energy arises when the Born-Oppenheimer approximation is applied to the energy of a system formed by more than one molecule.[94]

The energy of the system is calculated by treating the system as two sets of separated nuclei, in which each molecule contributes to the final energy taking also into account the interaction among them. Thus, the energy of a system formed by two molecules could be expressed as a combination of the energies of the two molecules plus the interaction:

$$E_{AB} = E_A + E_B + \Delta E_{AB}^{int} \ . \qquad \text{(eq. 3.52)}$$

Commonly, the supermolecule approach is employed to calculate the interaction energy of a system, by using the same nuclear positions in computing the contributions from the individual molecules and the whole system.[94-97] Thus, this interaction energy can be expressed as:

$$\Delta E_{AB}^{int}(R) = E_{AB}(R) - E_A(R) - E_B(R) \ . \qquad \text{(eq. 3.53)}$$

However, in this work, complexation is a key objective, so in order to describe it, it is necessary to introduce an additional quantity, the deformation energy:[98, 99]

$$E_{def}(R) = E_A(R) + E_B(R) - E_A(R_0) - E_B(R_0) \ . \qquad \text{(eq. 3.54)}$$

This contribution, also-called relaxation energy[95-97] or preparation energy,[100] takes into account the difference in energy observed depending on the geometries of the molecules before and after the formation of the final system. The reason for using this notation comes from the fact that isolated molecules undergo a change or deformation in their geometries due to the interaction when they constitute the complex.[95-97]

Other notations commented above refer to the same concept and therefore have equivalent meaning. Relaxation energy comes from dissociation, the geometries being relaxed when they arrive to the isolated state.[95-97] On the other hand, preparation energy comes from the point of view of complex formation, so it is necessary to change the geometry of the isolated fragments to those geometries which are proper to form the final complex. Thus, the energy used in this preparation process is the so-called preparation energy.[100]

The sum of the energy wasted on getting the optimal geometry to form the complex plus the energy associated to the interaction of the different fragments is then the complexation energy. It is worth nothing that if the opposite process is studied (the dissociation of the complex to the isolated fragments) this energy change is usually called binding energy and corresponds to the negative of the complexation energy. Also, for each of the energies needed for obtaining the previous magnitudes, other effects such as zero point energy corrections or thermal effects are added as usual. Briefly, the complexation energy is obtained as:

$$\Delta E_{AB}^{complex}(R) = \Delta E_{AB}^{int}(R) + E_{def}(R) \ . \qquad \text{(eq. 3.55)}$$

And thus, using equations 3.53 and 3.54, it can be concluded that

$$\Delta E_{AB}^{complex}(R) = E_{AB}(R) - E_A(R_0) - E_B(R_0)\,. \qquad \text{(eq. 3.56)}$$

Although this last description is simple, it can be more appropriate a decomposition of the complexation energy into interaction and deformation energies. This way gives a better interpretation of the system, especially when large deformation effects are present as a consequence of important geometry changes. In such situations, the large deformation effects could hide a very large interaction.[96, 97, 101, 102] Also, as it will be commented in the next section, it is advisable to make this separation due to some problems in the practical calculation of complexation energies.

3.4.1. Basis Set Superposition Error

Basis Set Superposition Error or BSSE is a common problem in quantum chemistry when interaction energies have to be calculated.[94, 95, 103, 104] As commented above, the interaction energy has to be calculated by using eq. 3.53, subtracting the energy of the isolated fragments from the energy of the final complex. The wavefunction of the different fragments is constructed with a finite basis set to perform the quantum chemistry calculation in a practical way. However, this fact entails an error.

If the complexation of the dimer AB is considered, basis sets centered onto the atoms of A or B are employed respectively for the isolated monomers A or B. However, when they approach each other to form the dimer, the superposition between basis sets is inevitable. At close distance, monomer A (or B) can also employ the basis set centered onto the atoms belonging to monomer B (or A). This increase in the number of basis functions in that situation makes the energy of the dimer decrease. Thus, the energy of the dimer seems lower than actually is. The problem relies on the fact that the superposition of basis sets for the monomers A and B is different depending on the distance between them, and as Duijneveldt says,[105] the extra basis functions would not be a error in itself if at large distances, the monomers could get the benefit of increasing its basis sets by employing the basis set of the other. But this does not occur, and there is an inconsistent treatment of the monomers, the origin of the BSSE. Thus, BSSE depends on the geometry of the systems, and its value is different for each arrangement of nuclei. Besides BSSE, there are other errors associated to the basis set and the method used.

To circumvent this error, Boys and Bernardi developed the counterpoise method, explained below.[106, 107] As in eq. 3.53, the interaction energy of a dimer AB, the uncorrected one, would be

obtained by using the same geometry for all the fragments, the geometry of the dimer (the superscripts), and also the basis set corresponding to each fragment (in parentheses):

$$\Delta E_{AB}^{int}(AB) = E_{AB}^{dimer}(AB) - E_{A}^{dimer}(A) - E_{B}^{dimer}(B). \qquad \text{(eq. 3.57)}$$

The counterpoise correction consists on obtaining the three energies using the same basis set in all calculations since the problem relied on the use of different sets of functions. Thus, the corrected interaction energy can be expressed as:

$$\Delta E_{AB}^{int}(AB) = E_{AB}^{dimer}(AB) - E_{A}^{dimer}(AB) - E_{B}^{dimer}(AB) \qquad \text{(eq. 3.58)}$$

Therefore, the BSSE is the difference between the corrected and the uncorrected interactions energies, resulting in a positive value that destabilizes the complex. It is expressed by the following equation:

$$BSSE = E_{A}^{dimer}(A) + E_{B}^{dimer}(B) - E_{A}^{dimer}(AB) - E_{B}^{dimer}(AB). \qquad \text{(eq. 3.59)}$$

Although the counterpoise method is used commonly to solve the problem of BSSE, a lot of literature argues about the possibility that it overcorrects the error due to BSSE, since the whole basis set of B (occupied and virtual) is allowed to be used by A. Indeed, recent studies have suggested its inconvenience.[108, 109] However, as commented above, it constitutes a common method within the theoretical chemistry community to reduce the error and also seems to allow the system to converge in a better way to the basis limit. For these reasons it has been employed throughout this work. The main drawback of using counterpoise method is probably the significant increase of the computation costs due to its dependency on the geometry. It means that for each geometry or nuclei configuration of the system, a counterpoise calculation has to be done.

In any case, as in equation 3.55, the complexation energy is the sum of the interaction plus the deformation energies. Using the corrected expression of interaction energy, the counterpoise-corrected complexation energy is calculated by:

$$\begin{aligned}\Delta E_{AB}^{comp}(AB) = {} & E_{AB}^{dimer}(AB) - E_{A}^{dimer}(AB) - E_{B}^{dimer}(AB) + \\ & + E_{A}^{dimer}(A) + E_{B}^{dimer}(B) - E_{A}^{isolated}(A) - E_{B}^{isolated}(B)\end{aligned}. \qquad \text{(eq. 3.60)}$$

It is worth noting that this error, as commented above, is due to the finite basis sets employed in the calculations, so that is not exclusive of intermolecular interactions but of any electronic structure calculation.[34, 110] Knowing this, BSSE is also present in the calculations for an isolated

molecule, taking as reference the isolated constituting atoms. However, it has been observed that intramolecular BSSE for small molecules gives negligible errors. Different behavior has the BSSE for large molecules in which there exists interaction among different areas or sections within the same molecule. In these situations, the best choice to reduce BSSE is to employ the largest basis set as possible.[111, 112]

3.4.2. Many-body effects

When more than a pair of molecules is involved in a system, the calculation of the interaction energy becomes a harder process due to interactions among all the fragments.[113] The total energy of the system could be expressed as the sum over all fragments, but also including a correction from all pairs, trios, etc., of interacting fragments;

$$E_{ijk\ldots} = \sum_i E_i + \sum_i \sum_{j>i} E_{ij} + \sum_i \sum_{j>i} \sum_{k>i,j} E_{ijk} + \ldots . \qquad \text{(eq. 3.61)}$$

On the other hand, the interaction energy can be expressed similarly to eq. 3.57, in which the interaction energy is the energy of the whole complex minus the energy corresponding to the constituting molecules. In this case, instead a dimer, it is considered the trimer ABC as example. Applying then the counterpoise method, in which also every fragment has the same geometry and basis set than the whole complex, the interaction energy can be expressed as:

$$\Delta E^{int}_{ABC} = E^{trimer}_{ABC}(ABC) - E^{trimer}_{A}(ABC) - E^{trimer}_{B}(ABC) - E^{trimer}_{C}(ABC) \qquad \text{(eq. 3.62)}$$

However, another more useful expression of the interaction energy of the cluster, in this case the trimer, can be used. In a first approach, since the monomer terms do not affect to the interaction energy (only to deformation), they are not considered. Instead of that, this interaction energy can be expressed as a pairwise sum of interaction between all pairs of molecules, plus a correction involving trios in this case:

$$\Delta E^{int}_{ABC} = \Delta E^{int}_{AB} + \Delta E^{int}_{AC} + \Delta E^{int}_{BC} + E_{3-body}, \qquad \text{(eq.3.63)}$$

where the $E_{3\text{-}body}$ corresponds to the correction due to the trio ABC, and the rest of the terms are the interaction energies of each pair of molecules, and can be expressed as:

$$\Delta E^{int}_{AB} = E^{trimer}_{AB}(ABC) - E^{trimer}_{A}(ABC) - E^{trimer}_{B}(ABC), \qquad \text{(eq. 3.64)}$$

$$\Delta E_{AC}^{int} = E_{AC}^{trimer}(ABC) - E_{A}^{trimer}(ABC) - E_{C}^{trimer}(ABC), \qquad \text{(eq. 3.65)}$$

$$\Delta E_{BC}^{int} = E_{BC}^{trimer}(ABC) - E_{B}^{trimer}(ABC) - E_{C}^{trimer}(ABC). \qquad \text{(eq. 3.66)}$$

Similarly, the interaction energy for larger systems can be obtained, just increasing the number of correction terms ($E_{4\text{-}body}$, $E_{5\text{-}body}$, etc) within the eq. 3.63 until the order of the system.[99, 104, 114, 115]

In the case of a trimer, the three-body contribution to the interaction energy of the trimer would be:

$$E_{3-body} = \Delta E_{ABC}^{int} - \Delta E_{AB}^{int} - \Delta E_{AC}^{int} - \Delta E_{BC}^{int}. \qquad \text{(eq. 3.67)}$$

This quantity is usually small in trimers, but there are systems in which that term becomes significant, as in those systems with important polarizing effects or hydrogen bonds. However, it becomes clear that when the number of constituting molecules of the system increases, also these many-body effects become more relevant.

Finally, it is important to indicate that also BSSE seems to be related to the number of fragments forming the cluster, since an increase of the BSSE has been observed when more molecules were included, even when the value of BSSE for a given dimer was small. For that reason, it is mandatory to perform a proper treatment of BSSE in large clusters.

3.5. Symmetry-Adapted Perturbation Theory (SAPT)

Although the magnitude of the interaction energy for a given geometry is easily obtained applying the supermolecule method with the wavefunction-based or DFT methods, detailed information about the nature, origins or characteristics of the interaction itself is not provided. Thus, methods which give this kind of information would be desirable.[94, 116] Then, focusing on the intermolecular forces, it would be interesting to employ methods which provide us with values for the different contributions, i.e. electrostatics, repulsion, polarization and dispersion effects.

Interaction energies, compared with those energies due to the chemical bond or with the intrinsic electronic energies of atoms or molecules, are more than an order of magnitude smaller in the best case. Thus, as it is a small contribution to the energy of the whole system, it can be treated as a perturbation, considering the most important contributions to the energy of the system are those from isolated monomers. Particularly, Symmetry-Adapted Perturbation Theory

(SAPT) is a well-known method among those based on perturbation theory which, applying a perturbational scheme (Rayleigh-Schrödinger) to a complex, provides expressions for the different contributions: electrostatics, induction and dispersion effects. The simplest partitioning of the Hamiltonian for a pair of molecules A and B corresponds to the following expression:

$$\hat{H} = \hat{H}_A + \hat{H}_B + \hat{V} = \hat{H}_0 + \hat{V}\,, \qquad \text{(eq. 3.68)}$$

where the operators $\hat{H}_{\mathrm{A}}$ and $\hat{H}_{\mathrm{B}}$ are the Hamiltonians of the isolated reference systems, those from the non-interacting molecules, $\hat{H}_0$ corresponding to the unperturbed Hamiltonian of the system and the operator $\hat{V}$ is the perturbation, in our case, the interaction energy.

The different contributions to the interaction energy are obtained as different first order, second order, … corrections of the energy as the result of applying the Rayleigh-Schrödinger perturbation theory under the previous assumptions. Thus, the first order correction energy corresponds to the electrostatic energy, since it shows the coulombic interaction between the electron density of both monomers,

$$E_{el}^{(1)} = \left\langle \psi_0^A \psi_0^B \middle| \hat{V} \middle| \psi_0^A \psi_0^B \right\rangle. \qquad \text{(eq. 3.69)}$$

On the other hand, induction and dispersion contributions appear for the correction to second order, depending on single and double excitations. For the interacting monomers A and B, the induction contribution is computed as the single excitations within one monomer induced by the close presence of the other monomer, obtaining two equivalent expressions, one for each monomer. For example, the induction of single excitations within monomer A is expressed by the following expression:

$$E_{ind}^{A} = -\sum_{m\neq 0} \frac{\left\langle \psi_0^A \psi_0^B \middle| \hat{V} \middle| \psi_m^A \psi_0^B \right\rangle^2}{E_m^A - E_0^A}. \qquad \text{(eq. 3.70)}$$

Finally, dispersion contribution is considered as the term within second order correction which takes into account double excitations:

$$E_{disp} = -\sum_{m\neq 0; n\neq 0} \frac{\left\langle \psi_0^A \psi_0^B \middle| \hat{V} \middle| \psi_m^A \psi_n^B \right\rangle^2}{E_m^A + E_n^B - E_0^A - E_0^B}. \qquad \text{(eq. 3.71)}$$

Thus, by means of these contributions together with the multipole expansion, intermolecular interactions can be described with a better qualitative and quantitative information, being possible describing them as functions of molecular properties as multipoles or polarizabilities.[94, 116] However, this Rayleigh-Schrödinger perturbation theory is only successful for molecules a long distance apart, being completely inappropriate at short range.[94] There are some reasons which lead to method failure and although one of them is due to the breaking down of the multipole expansion, the most fundamental failure comes from the fact that the repulsion between molecules at short range is not taken into account. This is an important consequence of ignoring the exchange contribution, and thus, the overlapping of wavefunctions when molecules are close enough is missed.

The problem of this loss lies in the wrong symmetry of the reference wavefunction when the Rayleigh-Schrödinger perturbation theory is used to describe intermolecular interactions.[94, 116] This reference wavefunction $\psi_0 = \psi_0^A \psi_0^B$ is not antisymmetric upon exchange of electrons between A and B, so it cannot satisfy the Pauli principle. Thus, the description of repulsion forces when molecules are close is defective.

Symmetrized Rayleigh-Schrödinger theory, or nowadays commonly called Symmetry Adapted Perturbation Theory (SAPT), is the most used method to solve the antisymmetry problem. Thus, the exchange-repulsion energy is achieved by forcing the antisymmetry in the energy expressions, modifying the electron density in the proper way which causes a repulsive force on the nuclei (exchange-repulsion energy).[117-119]

As in Rayleigh-Schrödinger perturbation theory at large distances, the SAPT procedure consists on a series of contributions which can be associated to physical effects when the order is low. However, the forced antisymmetrization produces new exchange-repulsion terms which now accompany to the old long-range terms. Thus, when the wavefunction of closed-shell molecules overlaps significantly, a strong repulsion appears overcoming the problem. In this way, the interaction energy by this procedure can be expressed as:

$$E_{\text{int}} = E_{\text{el}}^{(1)} + E_{\text{exch}}^{(1)} + E_{\text{ind}}^{(2)} + E_{\text{exch–ind}}^{(2)} + E_{\text{disp}}^{(2)} + E_{\text{exch–disp}}^{(2)} + \ldots . \qquad \text{(eq. 3.72)}$$

The monomers' wavefunctions are not computed exactly, but separating HF and correlation contributions, including the intramolecular electron correlation. This intramolecular contribution is an important term that cannot be avoided since molecular properties required as SAPT inputs are significantly affected by electron correlation.[18, 117-119] To solve this, the Hamiltonian can be expressed as a double perturbation in both the interaction and the intramonomer correlation,

$$\hat{H} = \hat{F}_A + \hat{W}_A + \hat{F}_B + \hat{W}_B + \hat{V} \,. \qquad \text{(eq. 3.73)}$$

The Hamiltonian is written as a sum of monomer Fock operators, $\hat{F}$, the potential of each monomer, $\hat{W}$, and the interaction potential, $\hat{V}$. Therefore, now it is possible to write the interaction energy as:

$$E_{\text{int}} = \sum_{i=1;\, j=0} \left(E_{pol}^{ij} + E_{exch}^{ij} \right), \qquad \text{(eq. 3.74)}$$

where *ij* is the order of $\hat{V}$ and $\hat{W}$. The accuracy of the interaction's description will increase with higher orders in both expansions. Commonly, the expansion is truncated at second-order in $\hat{V}$, since the third- and higher order terms are usually negligible, but different truncations can be done giving rise to different models. Indeed, induction effects are often associated to higher terms, so that in polar systems the truncations would be done at higher orders. Another way to take into account such effects is by means of a correction term:

$$\delta_{HF} = E_{\text{int}}^{HF} - E_{el}^{(10)} - E_{exch}^{(10)} - E_{ind}^{(20)} - E_{exch-ind}^{(20)} \,, \qquad \text{(eq. 3.75)}$$

where E_{int}^{HF} is the Hartree-Fock interaction energy calculated using the supermolecular method.

Although at HF level dispersion is not included, by applying perturbation theory based on HF wavefunctions this term can be obtained by SAPT. Thus, the HF+D method can be obtained by correcting the HF calculation with the dispersion contribution, so the interaction energy using HF wavefunction up to second order can be expressed as the following sum of SAPT contributions.

$$E_{\text{int}}^{HF+D} = E_{el}^{(10)} + E_{exch}^{(10)} + E_{ind}^{(20)} + E_{exch-ind}^{(20)} + \delta_{HF} + E_{disp}^{(20)} + E_{exch-disp}^{(20)} \,. \qquad \text{(eq. 3.76)}$$

For a more accurate description of the interactions, intramonomer correlation effects could be included. For example, the so-called SAPT2 level, including similar corrections to MP2, corresponds to:

$$E_{\text{int}}^{SAPT} = E_{\text{int}}^{HF+D} + E_{el}^{(12)} + E_{exch}^{(11)} + E_{exch}^{(12)} + E_{ind}^{(22)} + E_{exch-ind}^{(22)} \,. \qquad \text{(eq. 3.77)}$$

The next step would be including contributions to third order both in the intermolecular perturbation and the intramonomer correlation perturbation. However, this procedure requires relatively significant computer resources. Other approaches have been developed, as SAPT(DFT)

that will be commented in the following section. Also, it is worth mentioning that Hohenstein and Sherrill have developed a density fitting based approach in order to reduce the cost of SAPT calculations.[18, 120, 121] This approach has been included into the PSI4 program.[122]

3.5.1. SAPT(DFT)

As commented above, to include intramonomer correlation effects within SAPT, together with the triple perturbation theory used in the SAPT method, would require high computational costs would be required, and when larger molecular systems are employed it becomes prohibitive. A solution to this difficulty could be a SAPT approach utilizing a density functional theory (DFT) description of monomers, since if a correlated description of monomers is used, the costly correlation terms could be avoided.[123] Such a method, now called SAPT(DFT), has been first proposed by Williams and Chabalowski (2001) and later developed by others.[123-125] This procedure consists on changing the HF orbitals and energies by their Kohn-Sham counterparts. Since the intramonomer correlation effects are already taken care of in DFT calculations, only the perturbation on the intermolecular interaction will suffice on the SAPT calculation.[123]

This improvement allows SAPT to be performed on much larger systems than previously did, although the initial results with this procedure were disappointing.[94] The incorrect asymptotic behavior of the exchange-correlation functionals seems to be the main reason of these inaccurate results, with the consequence of one too diffuse electron density that leads to errors in both the first-order terms. In addition, the second-order terms were calculated using uncoupled perturbation theory and were also poor.

To circumvent this problem, two very similar proposals were soon developed independently by Hesselmann and Jansen[126, 127] and Szalewicz,[128, 129] focused on the correction for the asymptotic behavior. In Jansen's proposal the functional employed is combined with the LB94 functional, which has a correct asymptotic behavior at long-range, but fails at short-range,[127] arising a new problem consisting on inconsistency at intermediate distances between the behavior of the two functionals.[130, 131] In order to solve this, a gradient-regulated connection method developed by Grüning et al. was employed.[130] This is called Adiabatic Correction AC and requires the sum $IP + \varepsilon_{HOMO}$ as the input parameter, which vanishes for exact KS DFT. The value of ε_{HOMO} is obtained from a calculation with the uncorrected *xc* functional, whereas the ionization potentials can either be taken from experiment or calculated from the difference of KS DFT calculations of the neutral and the ionized systems.

After solving these accuracy problems, the SAPT(DFT) interaction energy can be expressed as:

$$E_{\text{int}}^{SAPT-DFT} = E_{el}^{(10)} + E_{exch}^{(10)} + E_{ind,resp}^{(20)} + E_{exch-ind,resp}^{(20)} + E_{disp}^{(20)} + E_{exch-disp}^{(20)} + \delta_{HF} \quad \text{(eq. 3.78)}$$

where a δ_{HF} correction term is also included which takes into account higher-order contributions. With this implementation of DFT within SAPT theory, the computational costs decrease considerably for the description of intramonomer correlation, allowing the use of this method to larger systems than those allowed for SAPT based in MBPT or CC. To reduce even more the computational costs, the density fitting approach can be employed. Furthermore, the performance of these DFT-based methods seems to be better than the more conventional SAPT using triple perturbation theory. Also, simplified extrapolation schemes can be used to reach the complete basis set limit in the framework of SAPT(DFT) calculations.[132]

3.6. Electron density analysis

Molecular properties often can be expressed in terms of a property density.[133] In these cases, it is possible to obtain the contribution of an atom to that molecular property by integrating this density over the bounded volume of that atom in the molecule. Thus, a set of physical properties characterize each atom in a molecule or crystal, corresponding then to molecular properties, and being added up to those properties of the whole molecular system. In the next sections methods based on this principle will be exposed.

3.6.1. QTAIM

One of the aims that led the development of the Atoms in Molecules Theory (AIM) was that most methods are only focused on obtaining approximate solutions to Schrödinger's equation. As Bader indicated, these approximate solutions, although provide a very useful tool to predict and understand the properties of some systems of interest (energy, geometry and other many properties), are not couched in the conceptual language of chemistry, as for instance the Lewis model of the electron pair or the molecular structure hypothesis.[134] Both of them treat the systems from chemical observations, and quantum mechanics lacked the incorporation of these observations into it. On the one hand, since Lewis model of the electron pair, it is known that the localized or delocalized nature of the electrons is related with aspects as resonance, reactivity, geometry or bonding. On the other hand, the molecular structure hypothesis relates the molecular structure of a system with its properties.

The molecular structure hypothesis comes from Dalton's atomic hypothesis formulated in 1807, in which the atoms retain their identity even in chemical combination with other atoms. This hypothesis evolved incorporating the geometrical aspects, providing the concept of molecular structure, according to which the structure of a molecule is imparted by the bonds, which form a network linking the different atoms in a molecule. This hypothesis came from the observation that atoms or groups of atoms have characteristic sets of properties which in general encompass a narrow range, so that those atoms or groups of atoms can be identified or a given behavior can be predicted. It was found that, if the distribution of charge of a given atom or group of atoms remained constant, their properties remained also constant, even their contribution to the total energy of the system. Also, if the distribution of charge changes the properties change by the same amount. According to Bader, this is because there exists a direct relationship between the properties of an atom and its spatial form even when the atom belongs to different systems, although obviously more evident when the atom or the group of atoms are transferrable. Thus, the properties of atoms in molecules can be experimentally measured, and with more precision those which are transferrable. With that premise arose the idea that the development of an AIM method was possible, with the purpose to relate molecular properties to those of its constituent atoms.

The development of this method was made possible by the knowledge that the electron density distribution, and its topology in particular, plays an essential role in experimental chemistry, providing a good explanation and understanding of the observations like chemical structure, functional groups, transferability, chemical reactivity or chemical bonding.[134] Also, the real three dimensional space was at last introduced into theoretical chemistry by this theory, since electron density provides a description of charge throughout real space, and thus relates the concepts of state functions in Hilbert space and the physical model of matter in space, according to Bader's words.

Bader and coworkers started to develop the foundations of AIM theory in the seventies, on the basis of their own works on molecular electron density distributions made a decade before. This theory is also-called quantum theory of atoms in molecules (QTAIM) in more recent literature due to its rigorous basis in quantum mechanics.

3.6.1.1. The topology of the electron density

The nucleus of each atom plays a role as a charge attractor, being the principal force involved in the final topology of the electron density, and therefore, providing it of a maximum of electron density at the position of each nucleus. In this way, the final density distribution in a molecule is described basically by the balance of the forces that the neighboring nuclei exert on the electrons. To know how the topology of the electron density, $\rho(\mathbf{r})$, is, it should be necessary to indicate how

it changes, and for that purpose the gradient of the electron density, noted as $\nabla\rho(\mathbf{r})$, is employed. As a definition of a gradient of any scalar at a point in space, $\nabla\rho(\mathbf{r})$ denotes a vector pointing in the direction in which $\rho(\mathbf{r})$ undergoes the greatest rate of increase, and having a magnitude equal to the rate of increase in that direction. In this way, the gradient is going to be used to define the "critical points" (CP) whose study will determine the topology of the electron density of each system. Thus, a "critical point" in the electron density is defined as the point in space where the gradient of the electron density (the first derivatives respect to the coordinates) vanish.

The results about the topology of the electron density are provided by the study of the critical points associated to the electron density of the system. At the same time, critical points are classified according to two properties called rank (ω) and signature (σ), and denoting each critical point as the couple of their values by (ω, σ). The rank is defined as the number of non-zero curvatures of ρ at the critical point. The signature is defined as the algebraic sum of the signs of the curvatures, contributing each curvature with ±1 depending on whether it is a positive or negative curvature. The most common value of ω is 3, since values of $\omega < 3$ are mathematically unstable and tend to vanish or bifurcate, and are not usually found in equilibrium charge distributions. In this way, there are four types of stable critical points with $\omega = 3$, corresponding with different elements of chemical structure:

a) *Nuclear critical point (NCP):* **(3, –3)** Three negative curvatures: ρ is a local maximum.
b) *Bond critical point (BCP):* **(3, –1)** Two negative curvatures: ρ is a maximum in the plane defined by the two associated axes, and a minimum along the axis perpendicular to this plane.
c) *Ring critical point (RCP):* **(3, +1)** Two positive curvatures: ρ is a minimum in the plane defined by the two associated axes, and a maximum along the axis perpendicular to this plane.
d) *Cage critical point (CCP):* **(3, +3)** Three positive curvatures: ρ is a local minimum.

In a molecule or crystal there exists a relationship between the number and the type of critical points:

$$n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP} = \begin{Bmatrix} 1 \ Poincaré - Hopf \ \rightarrow \ isolated\ molecules \\ 0 \ Morse\ equation \ \rightarrow Infinite\ crystals \end{Bmatrix},$$

where *n* denotes the number of critical points of each type, and the set $\{ n_{NCP} - n_{BCP} + n_{RCP} - n_{CCP}\}$ for a given system is called as the "characteristic set".

A useful concept is the *bond path (BP)*, which is described as the single line of maximum electron density linking the nuclei of two chemically bonded atoms. This line can be curved to a

greater or lesser degree depending on the ring strain of the molecule. Also, if some bond paths arrange in a plane forming a ring, a ring critical point will appear among them in the same plane. Raising the dimensionality, if some of those rings arrange forming a polyhedron, a cage critical point will appear among them into the volume enclosed by the rings. The bond path gives hints on the type of chemical bonding, i.e. weak, strong, closed-shell, and open shell interactions.

Other concept commonly used is the bond critical point (BCP), which describes the point into the bond path in which the electron density of the two atoms linked has the same value. In other words, the bond critical point is the point where $\nabla\rho(r) = 0$, along the bond path, and also the point with the lowest value of electron density. Thus, there exists a zero-flux surface defined by all of $\nabla\rho(r)$ trajectories verifying that $\nabla\rho(r) = 0$ and which intersects the bond path in the BCP. That occurs between each pair of linked atoms.

With only those two concepts, bond path and bond critical point, the *molecular graph* can be constructed with the aim to visualize the molecular structure. This molecular graph constitutes the representation of the bond paths linking the nuclei of bonded atoms in equilibrium geometry with the associated critical points.

3.6.2. NCI index

The non covalent interaction index (NCI) was developed with the aim of having a method to view and analyze non covalent interactions.[135, 136] This index provides the most relevant features of the interaction in a quick and simple way, without unnecessary extra data.

The background of the method relies on the fact that the so-called reduced density gradient, *s*, takes different values depending on the kind of interaction. This variable comes from the electron density, ρ, and its first derivative, and has the following expression:

$$s = \frac{1}{2\left(3\pi^2\right)^{1/3}} \frac{|\nabla\rho|}{\rho^{4/3}}. \qquad \text{(eq. 3.79)}$$

In regions far from the molecules with very small values of density, the reduced density gradient will have very large positive values. Conversely, within regions in which there are interactions (covalent or non covalent), the reduced density gradient will assume very small values, nearly zero.

To identify non covalent interactions among all small reduced gradients, *s* is plotted versus ρ, revealing that these interactions exhibit one or more spikes in the low-density, low-gradient region, a signature of non covalent interactions. Apparently, this feature comes from the

formation of an intermolecular complex, since the reduced gradient changes from very large values in the monomers to close to zero when the complex is formed.

Once the non covalent interactions have been identified, it is necessary to know the type of interaction in every case. Although it can be said that low density regions are related to the weakest interactions (such as van der Waals) and the high-density regions are related to the strongest interactions, it is necessary to resort to density derivates, to the Laplacian of the density, $\nabla^2\rho$. Strong interactions have been commonly distinguished with this tool.[137] Often, the Laplacian is decomposed in three principal axes of maximal variation, the three eigenvalues λ_i of the electron density Hessian matrix, and used to understand bonding in more detail.[134, 138] Thus, the Laplacian of the electron density can be expressed as:

$$\nabla^2\rho = \lambda_1 + \lambda_2 + \lambda_3 \;\; ; \;\; (\lambda_1 \leq \lambda_2 \leq \lambda_3), \qquad \text{(eq. 3.80)}$$

being λ_1 ,λ_2 and λ_3 the spatial components of the Laplacian. At the nuclei all values are negative, while away from them $\lambda_3 > 0$. In molecules, λ_3 indicates the variation of the density along the internuclear direction while λ_1 and λ_2 do the same along the plane normal to the λ_3 vector. The Laplacian of the density is used to distinguish between concentrated density regions, characteristic of covalent interactions, in which its sign is negative, and depleted density regions, characteristic of nonbonded or weak interactions, in which $\nabla^2\rho > 0$.[134] Also, the value of λ_2 can be used to distinguish between bonded and nonbonded contacts. Thus, if in that interatomic region $\lambda_2 > 0$, it can be said that there are nonbonded contacts, such as steric repulsion (with a density depletion), while if in that region the value is negative ($\lambda_2 < 0$), it indicates bonding interactions, such as hydrogen bonds (with an accumulation of density perpendicular to the bond). Also, van der Waals interactions can be detected, presenting a negligible density overlap that gives $\lambda_2 \leq 0$. Therefore, the analysis of λ_2 gives an idea about the type of the nonbonded interactions and the density indicates their strength.[136]

It is then possible to locate non covalent interactions by generating gradient isosurfaces enclosing the corresponding region of the real space, which are the basis of NCI. As commented before, both λ_2 and ρ give an idea about the kind and the strength of the interactions, so that the value (sign of $\lambda_2)\cdot\rho$ has been used as a key value to distinguish between them. Thus, isosurfaces are colored according to this value. Bonding interactions (such as dipole-dipole interactions or hydrogen bonding) present large and negative values of (sign of $\lambda_2)\cdot\rho$; meanwhile nonbonding interactions present large and positive values. Also, very weak interactions (van der Waals) show values near zero.

3.7. Solvation effects

From a microscopic point of view, solvation involves the formation of a set of interactions between a solute and a solvent as well as changes on the interactions of the solvent molecules in the vicinity of the solute.[139] Thus, to better understand the solvation process, the determination of the structure adopted by solvent molecules around the solute is necessary, since the nature and strength of the system interactions are intimately related to the macroscopic properties of the solvated system. Due to this important role of the solvent within the whole system, a lot of works have been focused directly on reveal the solvent shell structure around the solute,[140-157] and mostly on the structure of the hydration shell around ions[158-167] and on the solvation of hydrophobic solutes.[168-172]

When the target of the study is a weak interaction, as in our case the cation···π interaction, the influence of the solvent becomes even more important because the presence of solvent molecules can significantly change the strength of the interaction. Real systems in living organisms are complex in general, although in most of the cases the molecules are surrounded mostly by solvent molecules, being by far the most common solvent water. Thus, if the behavior of those interactions in a more realistic environment needs to be assessed, the solvent must be included in the system.

Knowing that, the solvent can be taken into account by means of two different methods, depending on the situation and the type of study carried out. On the one hand, the microhydration method; that is, the controlled addition of a given number of water molecules to a neutral or charged molecule or atom.[173] On the other hand, methods based in a continuum model of the solvent, surrounding uniformly our system. Combinations are also possible, for example by explicitly considering the first solvation shell and treating the rest by a continuum model.

3.7.1. Microhydration

In this work, one of the goals is to know how a specific weak interaction, the cation···π interaction, modifies its properties, as geometry or strength, especially in the presence of solvent. Cation···π interactions, as commented in the introduction, are typical interactions on amino acid systems, and hence in proteins, as some of them have an aromatic system in its side chain and also others can have protonated groups depending on environmental pH. Amino acids are relatively small molecules; however, those can join together to build proteins, which have a considerable size. Besides, proteins could be seen like a kind of sponges, with cavities, hollows and tunnels through the molecule. Thus, cation···π interactions could be sited in those cavities, in the surface of the molecule or just buried in the bulk of the protein. In terms of solvation, that

means that the degree of exposure to the solvent of a given cation···π interaction could vary in a wide range of values, being fully exposed to the solvent when the interaction is in the surface, partially exposed when sited in a cavity, or even it is possible that the solvent does not affect to the interaction if it is in a hydrophobic cavity with no solvent access.[174] Furthermore, it is known that the strength of the cation···π interaction decreases as more solvent molecules are included, as expected taking into account the decrease of the effective charge carried by the cation.[175-183] The role of the solvent upon cation···π interactions in amino acids could be assessed taking only into account the few types of aromatic systems and cations that could be involved, present on the amino acid side chains. Thus, the study can be carried out by introducing one by one water molecules in the system and observing how the properties are changing.

When the number of interacting water molecules increases, the amount of structures that can be conceived and that should be treated blows up, being an obvious problem for theoretical calculations aiming at determining the characteristics of microhydrated complexes. In these cases, the chemical knowledge may be a first step to face the problem. Indeed, one can settle the solvent molecules in reasonable starting positions to begin the optimization process when clear acidic or basic groups can be identified. However, it becomes impossible to track if no clear hydrogen bond donor or acceptor can be identified, or if one aims to go beyond first-shell effects. To avoid this limitation, several works carry out a preliminary molecular dynamics (MD) or Monte Carlo (MC) step to generate starting geometries that are subsequently optimized with quantum mechanical tools.[95, 107, 174, 180] Whilst MD and MC tend to provide a long list of structures, this second computational strategy is able to yield possible complexes without resorting to a priori considerations. However, the initial MD/MC step is often followed by two stages: first, a quick screening performed at semiempirical or low-quality *ab initio* levels and after that, optimizations of the most stable minima are carried out. Therefore, although the use of previous molecular dynamics steps is probably less error-prone than the use of chemical intuition when a large number of solvent molecules are involved in the studied system, there is no proof that the most relevant complexes formed by solute and solvent molecules can systematically be obtained by this technique.[173] It could be possible to not obtain one or some of the key minima with the selected force field approximation.

Another procedure, based on determining the structures of microhydrated anionic and cationic amino acids, has been proposed if the number of water molecules is small.[184-188] This approach consists in that the most important complexes with *n+1* water molecules can be generated from the similar most stable aggregates with *n* water molecules. This approach indeed provided complexation enthalpies and entropies in excellent agreement with experiment,[185, 187, 188] but is certainly not adapted to model bulk solvation shells it fails when the lowest-energy complex containing *n* water molecules is not related to the most stable *n+1* structure. In any case, for the

systems considered in this thesis, an approach based on placing water molecules in the most favorable positions for hydrogen bonding seems to be appropriate in order to include most stable structures of the complexes.

Another problem appears when the effects upon a system immersed in the bulk of the solvent are studied. Then, the microhydration method becomes unfeasible, as it should be necessary to include large numbers of water molecules in the system, and subsequently carry out the *ab initio* calculations. This procedure would result in high computational costs so in that case it would be better to use a continuum model.

3.7.2. The Polarizable Continuum Model (PCM)

Basically, continuum models consider the solvent as a continuum dielectric polarizable medium with a dielectric constant ε, instead of taking into account every molecule of the solvent as an individual molecule, and with the solute placed in a suitably shaped hole in the medium. Thus, the microscopic structure of the solvent cannot be known, but its average statistic response by means of the electric polarization. Also, the charge distribution of the solute is the source of the polarization of the dielectric, arising an electrostatic interaction between them.[189] With this approximation, the problem becomes much more affordable because the huge number of degrees of freedom associated to solvent is drastically reduced.

Particularly, the Polarizable Continuum Model (PCM) developed by Tomasi and coworkers is one of the most frequently used continuum solvation methods and has seen numerous variations over the years.[190] Essentially, the PCM model calculates the molecular free energy in solution as the sum over three contributions taken into account for the solvation process,

$$G_{sol} = G_{el} + G_{dr} + G_{cav}, \qquad \text{(eq. 3.81)}$$

where these components represent the electrostatic (G_{el}), the dispersion-repulsion (G_{dr}), and the cavitation energy (G_{cav}). All three contributions represent physical influences of solvent molecules upon the solute as part of the solvation process.[116] Thus, in this process the electrostatic term measures the work spent in building up the charge distribution of the solute in solution, representing the mutual interaction between them; the dispersion-repulsion term introduces the non-electrostatic interaction between solute and solvent; and the cavitation term measures the energy spent on the formation of the cavity in the dielectric medium, large enough to accommodate the solute within the solvent. This latter term is unfavorable to solvation, since this is accomplished by breaking down the cohesive forces between solvent molecules. However, the dispersion-repulsion term is favorable to solvation, since the solute is created in regions

where the dispersion forces are stronger than the repulsive ones. These two terms, cavitation and dispersion-repulsion, are often referred to as steric contributions.

All three terms are calculated using a cavity defined through interlocking van der Waals spheres, taking the atomic radii as a suitable factor (a typical value is 1.2) times a van der Waals radius, and centered at atomic positions. The vdW radius of each atom is a function of atom type, connectivity, overall charge of the molecule, and the number of attached hydrogen atoms. Such a surface may have small "pockets" where no solvent molecules can enter and a more appropriate descriptor may be defined as the surface traced out by a spherical particle of a given radius (a typical radius of 1.4 Å to model a water molecule) rolling on the van der Waals surface. This is denoted as the Solvent Accessible Surface (SAS).

Thus, to calculate the cavitation energy G_{cav}, the surface defined by the van der Waals spheres should be used, meanwhile to calculate the dispersion-repulsion contribution G_{dr} to the solution free energy, the solvent accessible surface is employed. On the other hand, the electrostatic contribution to the free energy in solution G_{el}, uses an approximate version of the solvent excluding surface constructed through scaling all radii by a constant factor (e.g. 1.2 for water) and then adding some more spheres not centered on atoms in order to arrive at a somewhat smoother surface. Also, to locate and calculate surface charges, it is necessary to make an approach through systematic division of the spherical surface in small regions of known area, and to carry out the calculation of one point charge per surface element afterwards.

These PCM methods (or D-PCM), are often called self-consistent reaction field methods (SCRF), since the solvation process described by forming the cavity, including the solvent and the later interaction between charge distributions (from the solute and the polarized of the solvent) is a self-consistent process. Thus, the solution passes by means of an iterative procedure in which favorable interactions occur, being possible to model the most important electrostatic effects due to the solvent.

Analytically, this electrostatic problem can be solved using the Poisson equation, since it is the usual way to find the electric potential for a given charge distribution described by the density function:

$$\nabla^2 V(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \qquad \text{(eq. 3.82)}$$

where ρ is the solute charge density, and V the electrostatic potential. To solve this equation, it has to be assumed that all the solute charge is inside the cavity, so that the second term vanishes outside the cavity. That means that the cavity inside the dielectric in the continuum model belongs "electrostatically" to the solute, and not to the solvent.[189, 191] Thus, the electrostatic

problem of eq. 3.82 can be solved in terms of the electrostatic potential *V*, which is the sum of the solute potential (inside the cavity), plus the contribution due to the response of the medium to the presence of the solute inside the cavity (potential due to the rest of the superficial cavity charges),

$$V(\mathbf{r}) = V_s(\mathbf{r}) + V_\sigma(\mathbf{r}) \quad \text{(eq. 3.83)}$$

The apparent charges potential, called the solvent reaction potential, V_σ, interacts with the real charges of the molecule, giving the total potential of the solvation process.

In order to solve the equation, some boundary conditions must be applied; for instance, the potential has to be continuous and the electric field must be discontinuous through the surface of the cavity. Although the problem has been simplified by using the charge distribution enclosed by a surface, the integration of the potential through that surface is most times complex, so it is obtained by approximate procedures. One of the most used methods in the numerical solution of the equation is the Boundary Element Method (BEM),[116] based on the discretization of the cavity in a finite number of elements with a set of apparent surface charges (ASC),

$$V_\sigma(\mathbf{r}) = \int_\Gamma \frac{\sigma(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} d^2\mathbf{s}. \quad \text{(eq. 3.84)}$$

Thus, eq. 3.84 can be discretized in that finite number of elements producing the same reaction potential V_σ, which is included into the Hamiltonian to analyze it computationally by means of quantum mechanics. Also, the value obtained for the electrostatic potential will be used to calculate the free energy of the solvation process by means of the following equation:

$$G^{el} = -\frac{1}{2}\int \rho(\mathbf{r}) V_\sigma(\mathbf{r})\, d\mathbf{r} \quad \text{(eq. 3.85)}$$

Besides the basic PCM method described above (D-PCM), other different continuum methods have been developed in order to solve the same solvation problem, being one of the most employed the Conductor-Like Screening Model (COSMO).[192] In this model, the solvent is treated like a conductor, so the total potential vanishes on the cavity surface, simplifying the electrostatic equations. In order to recover the effects of the finite value of the dielectric constant of the medium, the apparent charge distribution for ideal conductor is scaled by a function depending on the dielectric constant of the solvent. In COSMO, the solvent is treated as a dielectric, with charges defining the reaction field obtained applying conductor-like boundary conditions and properly scaled to represent the value of the dielectric constant of the solvent.

3.8. References

(1) C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models, 2nd Edition*, John Wiley & Sons, Chichester, **2004**.

(2) D. C. Young, *Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems*, John Wiley & Sons, Inc., New York, **2001**.

(3) S. Attila, S. O. Neil, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications, New York, **2012**.

(4) R. J. Bartlett. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359-401.

(5) F. Jensen, *Introduction to computational chemistry*, John Wiley and Sons, Chichester, **2001**.

(6) S. Grimme. *J. Chem. Phys.* **2003**, *118*, 9095-9102.

(7) S. Grimme, L. Goerigk, R. F. Fink. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 886-906.

(8) J. G. Hill, J. A. Platts. *J. Chem. Theory Comput.* **2006**, *3*, 80-85.

(9) M. Gerenkamp, S. Grimme. *Chem. Phys. Lett.* **2004**, *392*, 229-235.

(10) T. P. M. Goumans, A. W. Ehlers, K. Lammertsma, E.-U. Wuerthwein, S. Grimme. *Chem. - Eur. J.* **2004**, *10*, 6468-6475.

(11) S. Grimme. *J. Phys. Chem. A* **2005**, *109*, 3067-3077.

(12) J. Antony, S. Grimme. *J. Phys. Chem. A* **2007**, *111*, 4862-4868.

(13) K. E. Riley, J. A. Platts, J. Řezáč, P. Hobza, J. G. Hill. *J. Phys. Chem. A J Phys Chem A* **2012**, *116*, 4159-4169.

(14) S. Grimme. *J. Comput. Chem.* **2003**, *24*, 1529-1537.

(15) I. Hyla-Kryspin, S. Grimme. *Organometallics* **2004**, *23*, 5581-5592.

(16) M. Pitonak, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9611-9614.

(17) T. Takatani, E. G. Hohenstein, C. D. Sherrill. *J. Chem. Phys.* **2008**, *128*, 124111/1-124111/7.

(18) E. G. Hohenstein, C. D. Sherrill. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304-326.

(19) A. Hellweg, S. A. Grun, C. Hattig. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4119-4127.

(20) J. A. Bort, J. B. Rusca, *Theoretical and Computational Chemistry: Foundations, Methods and Techniques*, Publicacions de la Universitat Jaume I, Castelló de la Plana, **2007**.

(21) T. Van Voorhis, M. Head-Gordon. *J. Chem. Phys.* **2000**, *113*, 8873-8879.

(22) K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon. *Chem. Phys. Lett.* **1989**, *157*, 479-483.

(23) T. Helgaker, W. Klopper, A. Halkier, K. Bak, P. Jørgensen, J. Olsen, *Highly Accurate Ab Initio Computation of Thermochemical Data*, in Quantum-Mechanical Prediction of Thermochemical Data, *Vol. 22* (Ed. J. Cioslowski), Springer Netherlands, New York, **2001**, pp.1-30.

(24) C. Hättig, W. Klopper, A. Köhn, D. P. Tew. *Chem. Rev.* **2011**, *112*, 4-74.

(25) L. Kong, F. A. Bischoff, E. F. Valeev. *Chem. Rev.* **2011**, *112*, 75-107.

(26) T. Helgaker, W. Klopper, H. Koch, J. Noga. *J. Chem. Phys.* **1997**, *106*, 9639-9646.

(27) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson. *Chem. Phys. Lett.* **1998**, *286*, 243-252.
(28) D. G. Truhlar. *Chem. Phys. Lett.* **1998**, *294*, 45-48.
(29) A. J. C. Varandas. *J. Chem. Phys.* **2007**, *126*, 244105/1-244105/15.
(30) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, J. Olsen. *Chem. Phys. Lett.* **1999**, *302*, 437-446.
(31) D. G. Liakos, F. Neese. *J. Phys. Chem. A J Phys Chem A* **2012**, *116*, 4801-4816.
(32) F. Neese, E. F. Valeev. *J. Chem. Theory Comput.* **2010**, *7*, 33-43.
(33) R. A. Kendall, T. H. Dunning, R. J. Harrison. *J. Chem. Phys.* **1992**, *96*, 6796-6806.
(34) T. Helgaker, P. Jørgensen, J. Olsen, *Molecular electronic-structure theory* John Wiley & Sons, Chichester, **2000**.
(35) W. Klopper, W. Kutzelnigg. *J. Mol. Struct.: THEOCHEM* **1986**, *135*, 339-356.
(36) A. Karton, J. L. Martin. *Theor. Chem. Acc.* **2006**, *115*, 330-333.
(37) F. Jensen. *Theor. Chem. Acc.* **2005**, *113*, 267-273.
(38) K. E. Riley, P. Hobza. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3-17.
(39) P. Jurecka, J. Sponer, J. Cerny, P. Hobza. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985-1993.
(40) J. Řezáč, K. E. Riley, P. Hobza. *J. Chem. Theory Comput.* **2014**, *10*, 1359-1360.
(41) J. Řezáč, K. E. Riley, P. Hobza. *J. Chem. Theory Comput.* **2011**, *7*, 2427-2438.
(42) P. Hobza, J. Šponer. *J. Am. Chem. Soc.* **2002**, *124*, 11802-11808.
(43) M. Pitonak, P. Neograd, J. Cerny, S. Grimme, P. Hobza. *ChemPhysChem* **2009**, *10*, 282-289.
(44) K. E. Riley, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2011**, *13*, 21121-21125.
(45) R. Sedlak, K. E. Riley, J. Řezáč, M. Pitoňák, P. Hobza. *ChemPhysChem* **2013**, *14*, 698-707.
(46) K. E. Riley, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2012**, *14*, 13187-13193.
(47) R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, USA, Oxford, **1989**.
(48) L. J. Bartolotti, K. Flurchick, *An Introduction to Density Functional Theory*, in Rev. Comput. Chem., *Vol.*, John Wiley & Sons, Inc., Hoboken, **2007**, pp.187-216.
(49) A. St-Amant, *Density Functional Methods in Biomolecular Modeling*, in Rev. Comput. Chem., *Vol.*, John Wiley & Sons, Inc., Hoboken, **2007**, pp.217-259.
(50) T. Ziegler. *Chem. Rev.* **1991**, *91*, 651-667.
(51) E. J. Baerends, O. V. Gritsenko. *J. Phys. Chem. A J Phys Chem A* **1997**, *101*, 5383-5403.
(52) W. Koch, M. C. Holthausen, *A Chemist's guide to density functional theory*, Wiley-VCH, Weinhim, **2000**.
(53) P. Hohenberg, W. Kohn. *Phys.* **1964**, *136*, B864-B871.
(54) W. Kohn, L. J. Sham. *Phys.* **1965**, *140*, A1133-A1138.
(55) S. H. Vosko, L. Wilk, M. Nusair. *Can. J. Phys.* **1980**, *58*, 1200-1211.
(56) J. P. Perdew. *Phys. Rev. B* **1986**, *33*, 8822-8824.
(57) A. D. Becke. *Phys. Rev. A* **1988**, *38*, 3098-3100.

(58) C. Lee, W. Yang, R. G. Parr. *Phys. Rev. B* **1988**, *37*, 785-789.
(59) B. Miehlich, A. Savin, H. Stoll, H. Preuss. *Chem. Phys. Lett.* **1989**, *157*, 200-206.
(60) A. D. Becke. *J. Chem. Phys.* **1993**, *98*, 1372-1377.
(61) Y. Zhao, D. G. Truhlar. *J. Chem. Phys.* **2006**, *125*, 194101/1-194101/18.
(62) Y. Zhao, D. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
(63) Y. Zhao, D. G. Truhlar. *J. Chem. Theory Comput.* **2008**, *4*, 1849-1868.
(64) R. Peverati, D. G. Truhlar. *The Journal of Physical Chemistry Letters* **2011**, *3*, 117-124.
(65) R. Peverati, D. G. Truhlar. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2810-2817.
(66) R. Peverati, D. G. Truhlar. *Phys. Chem. Chem. Phys.* **2012**, *14*, 16187-16191.
(67) R. Peverati, D. G. Truhlar. *Phys. Chem. Chem. Phys.* **2012**, *14*, 13171-13174.
(68) E. G. Hohenstein, S. T. Chill, C. D. Sherrill. *J. Chem. Theory Comput.* **2008**, *4*, 1996-2000.
(69) J. Cerny, P. Hobza. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1624-1626.
(70) S. Tsuzuki, H. P. Lüthi. *J. Chem. Phys.* **2001**, *114*, 3949-3957.
(71) M. J. Allen, D. J. Tozer. *J. Chem. Phys.* **2002**, *117*, 11113-11120.
(72) P. Hobza, J. šponer, T. Reschel. *J. Comput. Chem.* **1995**, *16*, 1315-1325.
(73) S. Kristyán, P. Pulay. *Chem. Phys. Lett.* **1994**, *229*, 175-180.
(74) N. Kurita, H. Sekino. *Int. J. Quantum Chem* **2003**, *91*, 355-362.
(75) E. R. Johnson, R. A. Wolkow, G. A. DiLabio. *Chem. Phys. Lett.* **2004**, *394*, 334-338.
(76) Y. Zhao, N. E. Schultz, D. G. Truhlar. *J. Chem. Phys.* **2005**, *123*, 161103/1-161103/4.
(77) Y. Zhao, N. E. Schultz, D. G. Truhlar. *J. Chem. Theory Comput.* **2006**, *2*, 364-382.
(78) Y. Zhao, B. J. Lynch, D. G. Truhlar. *J. Phys. Chem. A J Phys Chem A* **2004**, *108*, 4786-4791.
(79) S. Grimme. *J. Chem. Phys.* **2006**, *124*, 034108/1-034108/16.
(80) S. Grimme. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
(81) U. Zimmerli, M. Parrinello, P. Koumoutsakos. *J. Chem. Phys.* **2004**, *120*, 2693-2699.
(82) S. Grimme. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
(83) S. Grimme, J. Antony, S. Ehrlich, H. Krieg. *J. Chem. Phys.* **2010**, *132*, 154104/1-154104/19.
(84) S. Grimme, S. Ehrlich, L. Goerigk. *J. Comput. Chem.* **2011**, *32*, 1456-1465.
(85) S. Grimme. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211-228.
(86) A. Krishtal, C. Van Alsenoy, P. Geerlings. *J. Chem. Phys.* **2014**, *140*, 184105/1-184105/14.
(87) S. N. Steinmann, C. Corminboeuf. *J. Chem. Theory Comput.* **2010**, *6*, 1990-2001.
(88) S. N. Steinmann, C. Corminboeuf. *J. Chem. Phys.* **2011**, *134*, 044117/1-044117/5.
(89) L. A. Burns, Á. V. Mayagoitia, B. G. Sumpter, C. D. Sherrill. *J. Chem. Phys.* **2011**, *134*, 084107/1-084107/25.
(90) Q. Wu, W. Yang. *J. Chem. Phys.* **2002**, *116*, 515-524.
(91) O. A. Vydrov, T. Van Voorhis. *J. Chem. Phys.* **2010**, *133*, 244103/1-244103/9.
(92) M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist. *Phys. Rev. Lett.* **2004**, *92*, 246401.
(93) F. D. John, G. Tim. *J. Phys.: Condens. Matter* **2012**, *24*, 073201.

(94) A. J. Stone, *The theory of intermolecular forces*, Oxford University Press, Oxford, **2013**.
(95) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.
(96) K. Szalewicz, B. Jeziorski. *J. Chem. Phys.* **1998**, *109*, 1198-1200.
(97) S. S. Xantheas. *J. Chem. Phys.* **1996**, *104*, 8821-8824.
(98) E. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *Theor. Chem. Acc.* **2011**, *128*, 531-539.
(99) E. M. Cabaleiro-Lago, M. A. Ríos. *J. Chem. Phys.* **2000**, *112*, 2155-2163.
(100) M. v. Hopffgarten, G. Frenking. *WIREs Comput. Mol. Sci.* **2012**, *2*, 43-62.
(101) A. Campo-Cacharrón, E. M. Cabaleiro-Lago, J. Rodríguez-Otero. *ChemPhysChem* **2012**, *13*, 570-577.
(102) E. M. Cabaleiro-Lago, Á. Peña-Gallego, J. Rodríguez-Otero. *J. Chem. Phys.* **2008**, *128*, 194311/1-194311/8.
(103) G. Chalasinski, M. M. Szczesniak. *Chem. Rev.* **1994**, *94*, 1723-1765.
(104) C. D. Sherrill, *Computations of Non covalent π Interactions,* in Rev. Comput. Chem., *Vol.*, John Wiley & Sons, Inc., Hoboken, **2009**, pp.1-38.
(105) F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, J. H. van Lenthe. *Chem. Rev.* **1994**, *94*, 1873-1885.
(106) H. B. Jansen, P. Ros. *Chem. Phys. Lett.* **1969**, *3*, 140-143.
(107) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.
(108) J. Alvarez-Idaboy, A. Galano. *Theor. Chem. Acc.* **2010**, *126*, 75-85.
(109) Ł. M. Mentel, E. J. Baerends. *J. Chem. Theory Comput.* **2013**, *10*, 252-267.
(110) T. Helgaker, W. Klopper, D. P. Tew. *Mol. Phys.* **2008**, *106*, 2107-2143.
(111) R. M. Balabin. *J. Chem. Phys.* **2010**, *132*, 211103/1-211103/4.
(112) D. Asturiol, M. Duran, P. Salvador. *J. Chem. Theory Comput.* **2009**, *5*, 2574-2581.
(113) M. J. Elrod, R. J. Saykally. *Chem. Rev.* **1994**, *94*, 1975-1997.
(114) K. Walczak, J. Friedrich, M. Dolg. *J. Chem. Phys.* **2011**, *135*, 134118/1-134118/11.
(115) J. F. Ouyang, M. W. Cvitkovic, R. P. A. Bettens. *J. Chem. Theory Comput.* **2014**.
(116) I. Kaplan, *Intermolecular Interactions : Physical Picture , Computational Methods*, John Wiley & Sons, Chichester, **2006**.
(117) B. Jeziorski, R. Moszynski, K. Szalewicz. *Chem. Rev.* **1994**, *94*, 1887-1930.
(118) R. Moszynski. *Mol. Phys.* **1996**, *88*, 741-758.
(119) R. Moszynski, B. Jeziorski, A. Ratkiewicz, S. a. Rybak. *J. Chem. Phys.* **1993**, *99*, 8856-8869.
(120) E. G. Hohenstein, C. D. Sherrill. *J. Chem. Phys.* **2010**, *133*, 014101/1-014101/12.
(121) E. G. Hohenstein, C. D. Sherrill. *J. Chem. Phys.* **2010**, *132*, 184111/1-184111/10.
(122) J. M. Turney, A. C. Simmonett, R. M. Parrish, E. G. Hohenstein, F. A. Evangelista, J. T. Fermann, B. J. Mintz, L. A. Burns, J. J. Wilke, M. L. Abrams, N. J. Russ, M. L. Leininger, C. L. Janssen, E. T. Seidl, W. D. Allen, H. F. Schaefer, R. A. King, E. F. Valeev, C. D. Sherrill, T. D. Crawford. *WIREs Comput. Mol. Sci.* **2012**, *2*, 556-565.
(123) H. L. Williams, C. F. Chabalowski. *J. Phys. Chem. A J Phys Chem A* **2001**, *105*, 646-659.

(124) A. J. Misquitta, K. Szalewicz. *Chem. Phys. Lett.* **2002**, *357*, 301-306.
(125) A. J. Misquitta, B. Jeziorski, K. Szalewicz. *Phys. Rev. Lett.* **2003**, *91*, 033201.
(126) A. Heßelmann, G. Jansen. *Chem. Phys. Lett.* **2002**, *357*, 464-470.
(127) G. Jansen. *WIREs Comput. Mol. Sci.* **2014**, *4*, 127-144.
(128) A. J. Misquitta, R. Podeszwa, B. Jeziorski, K. Szalewicz. *J. Chem. Phys.* **2005**, *123*, 214103/1-214103/14.
(129) K. Szalewicz. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254-272.
(130) M. Grüning, O. V. Gritsenko, S. J. A. van Gisbergen, E. J. Baerends. *J. Chem. Phys.* **2001**, *114*, 652-660.
(131) D. J. Tozer, N. C. Handy. *Mol. Phys.* **2003**, *101*, 2669-2675.
(132) J. Řezáč, P. Hobza. *J. Chem. Theory Comput.* **2011**, *7*, 685-689.
(133) Cherif F. Matta (Editor), Russell J. Boyd (Editor), *The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design*, WILEY-VCH, Weinheim, **2007**.
(134) R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford **1990**.
(135) J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.
(136) E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang. *J. Am. Chem. Soc.* **2010**, *132*, 6498-6506.
(137) R. F. W. Bader, H. Essén. *J. Chem. Phys.* **1984**, *80*, 1943-1960.
(138) R. F. W. Bader. *J. Phys. Chem. A J Phys Chem A* **1998**, *102*, 7314-7323.
(139) M. Orozco, F. J. Luque. *Chem. Rev.* **2000**, *100*, 4187-4226.
(140) Q. Liu, J. W. Brady. *J. Phys. Chem. B* **1997**, *101*, 1317-1321.
(141) M. A. Vincent, I. J. Palmer, I. H. Hillier. *J. Mol. Struct.: THEOCHEM* **1997**, *394*, 1-9.
(142) P. Manivet, M. Masella. *Chem. Phys. Lett.* **1998**, *288*, 642-646.
(143) S. Koneshan, J. C. Rasaiah, R. M. Lynden-Bell, S. H. Lee. *J. Phys. Chem. B* **1998**, *102*, 4193-4204.
(144) H.-P. Cheng. *J. Phys. Chem. A J Phys Chem A* **1998**, *102*, 6201-6204.
(145) M. Diraison, P. Millie, S. Pommeret, T. Gustavsson, J. C. Mialocq. *Chem. Phys. Lett.* **1998**, *282*, 152-158.
(146) C. Alemán, S. E. Galembeck. *Chem. Phys.* **1998**, *232*, 151-159.
(147) R. Biswas, S. Bhattacharyya, B. Bagchi. *J. Phys. Chem. B* **1998**, *102*, 3252-3256.
(148) R. Gratias, H. Kessler. *J. Phys. Chem. B* **1998**, *102*, 2027-2031.
(149) N. U. Zhanpeisov, J. Leszczynski. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 8317-8327.
(150) N. U. Zhanpeisov, J. Leszczynski. *J. Phys. Chem. A J Phys Chem A* **1998**, *102*, 6167-6172.
(151) M. L. S. Mendoza, M. A. Aguilar, F. J. O. del Valle. *J. Mol. Struct.: THEOCHEM* **1998**, *426*, 181-190.
(152) K. A. Swiss, R. A. Firestone. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 5369-5372.
(153) R. Takasu, K. Hashimoto, R. Okuda, K. Fuke. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 349-354.

(154) M. Masamura. *J. Mol. Struct.: THEOCHEM* **1999**, *466*, 85-93.
(155) V. Tran, B. J. Schwartz. *J. Phys. Chem. B* **1999**, *103*, 5570-5580.
(156) K. Coutinho, N. Saavedra, S. Canuto. *J. Mol. Struct.: THEOCHEM* **1999**, *466*, 69-75.
(157) T. van Mourik, S. L. Price, D. C. Clary. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 1611-1618.
(158) T.-M. Chang, L. X. Dang. *J. Phys. Chem. B* **1997**, *101*, 10518-10526.
(159) J. M. Martínez, R. R. Pappalardo, E. S. Marcos. *J. Phys. Chem. A J Phys Chem A* **1997**, *101*, 4444-4448.
(160) G. D. Scholes, R. D. Harcourt, I. R. Gould, D. Phillips. *J. Phys. Chem. A J Phys Chem A* **1997**, *101*, 678-684.
(161) A. Tongraar, K. R. Liedl, B. M. Rode. *Chem. Phys. Lett.* **1998**, *286*, 56-64.
(162) H. Sato, F. Hirata, A. B. Myers. *J. Phys. Chem. A J Phys Chem A* **1998**, *102*, 2065-2071.
(163) K. Hashimoto, T. Kamimoto. *J. Am. Chem. Soc.* **1998**, *120*, 3560-3570.
(164) J. M. Martínez, R. R. Pappalardo, E. S. Marcos, K. Refson, S. Díaz-Moreno, A. Muñoz-Páez. *J. Phys. Chem. B* **1998**, *102*, 3272-3282.
(165) T.-M. Chang, L. X. Dang. *J. Phys. Chem. B* **1999**, *103*, 4714-4720.
(166) A. Tongraar, B. M. Rode. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 8524-8527.
(167) H. D. Pranowo, B. M. Rode. *J. Phys. Chem. A J Phys Chem A* **1999**, *103*, 4298-4302.
(168) S. W. Rick, B. J. Berne. *J. Phys. Chem. B* **1997**, *101*, 10488-10493.
(169) P. E. Smith. *J. Phys. Chem. B* **1999**, *103*, 525-534.
(170) B. Madan, K. Sharp. *Biophys. Chem.* **1999**, *78*, 33-41.
(171) Y.-K. Cheng, P. J. Rossky. *Biopolymers* **1999**, *50*, 742-750.
(172) Y.-P. Pang, J. L. Miller, P. A. Kollman. *J. Am. Chem. Soc.* **1999**, *121*, 1717-1725.
(173) D. Jacquemin, C. Michaux, E. A. Perpète, G. Frison. *J. Phys. Chem. B* **2011**, *115*, 3604-3613.
(174) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.
(175) P. E. Mason, C. E. Dempsey, G. W. Neilson, S. R. Kline, J. W. Brady. *J. Am. Chem. Soc.* **2009**, *131*, 16689-16696.
(176) C. D. Tatko, M. L. Waters. *Protein Sci.* **2003**, *12*, 2443-2452.
(177) A. Rodríguez-Sanz, J. Carrazana-García, E. Cabaleiro-Lago, J. Rodríguez-Otero. *J. Mol. Model.* **2013**, *19*, 1985-1994.
(178) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.
(179) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.
(180) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.
(181) A. Campo-Cacharrón, E. Cabaleiro-Lago, J. Rodríguez-Otero. *Theor. Chem. Acc.* **2012**, *131*, 1-13.
(182) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-2138.
(183) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.

(184) C. Michaux, J. Wouters, D. Jacquemin, E. A. Perpète. *Chem. Phys. Lett.* **2007**, *445*, 57-61.
(185) C. Michaux, J. Wouters, E. A. Perpète, D. Jacquemin. *J. Phys. Chem. B* **2008**, *112*, 2430-2438.
(186) C. Michaux, J. Wouters, E. A. Perpète, D. Jacquemin. *J. Am. Soc. Mass. Spectrom.* **2009**, *20*, 632-638.
(187) C. Michaux, J. Wouters, E. A. Perpète, D. Jacquemin. *J. Phys. Chem. B* **2008**, *112*, 9896-9902.
(188) C. Michaux, J. Wouters, E. A. Perpète, D. Jacquemin. *J. Phys. Chem. B* **2008**, *112*, 7702-7705.
(189) J. Tomasi, M. Persico. *Chem. Rev.* **1994**, *94*, 2027-2094.
(190) J. Tomasi, B. Mennucci, R. Cammi. *Chem. Rev.* **2005**, *105*, 2999-3094.
(191) C. J. F. Böttcher, O. C. van Belle, P. Bordewijk, A. Rip, *Theory of electric polarization*, Elsevier Scientific Pub. Co., Amsterdam, **1978**.
(192) A. Klamt, G. Schuurmann. *J. Chem. Soc. Perkin Trans. 2* **1993**, 799-805.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4

Microhydration study: complexes between NH_4^+ and $CH_3NH_3^+$ with phenol

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

4.1. Introduction

Non covalent interactions are an essential tool for the molecular architecture in natural systems, as in certain key processes as molecular recognition or protein folding.[1-3] These interactions allow the construction of nanoscale molecular assemblies of a complexity greater than the initial molecules and held together and organized by non covalent interactions.[4-6] One widespread type of non covalent interaction is the cation···π interaction between a cation and an aromatic unit.[1, 3, 7-11] It is very common to find protonated amino acids which may establish cation···π interactions with the side chain of other aromatic amino acids present in the environment, for example, as part of the same protein. Thus, it is common for the side chains of arginine and lysine to interact with phenylalanine, tyrosine and tryptophan.[1, 7, 9, 11]

Though cation···π interactions are strong interactions in the gas phase, their strength can be dramatically affected by environmental effects, most importantly the presence of solvent around a given cation···π contact.[1, 12-23] Several studies have shown that cation···π contacts in proteins can be formed with a great variety of degrees of exposure to solvent, ranging from fully exposed contacts to others buried in hydrophobic regions of the protein. Therefore, there can be cation···π contacts which are only partially exposed to the solvent, forming a microhydrated structure that can present different properties compared to those of the fully exposed cation···π interaction.

Theoretical methods are especially well suited for studying this kind of effect, since the progressive hydration of a given cation···π interaction can be modeled, providing information at a microscopic level which is usually not affordable from experiment.[24, 25] The cation···π interaction weakens as water molecules are included,[26-30] and the solvent can promote structural changes in the mutual orientation of the cation and the aromatic unit, as recently shown by our group in guanidinium···benzene complexes.[30] An important interaction among amino acid side chains is established by the cationic lysine side chains, which bear an ammonium cation, and the aromatic units in the side chains of aromatic amino acids.[9, 11] The interaction between methylammmonium and benzene has been studied as a simple model for this kind of interaction, showing that the inclusion of water molecules weakens the methylammonium···benzene interaction. Also, the inclusion of the third water molecule breaks the direct contact between the cation and benzene.[29] However, it can be expected that interactions with other aromatic units as the phenol unit in tyrosine will present a much more complex behavior than benzene. Contrary to benzene, phenol possesses two different regions where a cation can establish a stabilizing interaction. Also, the hydroxyl group can act both as donor or acceptor in hydrogen bonds with the water molecules included in the cluster. Therefore, even when phenol is very similar to benzene, the presence of the hydroxyl group introduces a greater complexity on the potential

energy surface of the cluster, allowing for a greater variety of stable structures, as already shown in phenol···water clusters as compared to benzene ones.[31-38]

With the aim of modeling the interaction between lysine and tyrosine side chains, complexes formed by ammonium and methylammonium cations with one phenol unit have been computationally studied employing *ab initio* methods. The effect of a small number of water molecules in the complex has been considered in order to determine how their presence can alter the characteristics of the interaction.

4.2. Computational details

The first step for the study of the clusters formed by ammonium and methylammonium with phenol in the presence of water molecules is the location of the most stable minimum energy structures on the potential energy surface of each of the complexes studied in this work. Starting structures have been prepared taking into account the chemical characteristics of the species involved in the formation of the clusters. Phenol presents two regions for a favorable interaction with cations: the aromatic ring and the hydroxyl oxygen. Therefore, complexes formed by phenol and one ammonium or methylamonium cation have been prepared exploring these regions. The geometries of these clusters have been optimized at the MP2/6-31+G* level of calculation, the resulting structures being used for subsequent optimizations at the MP2/6-31+G(2d,p) level. Once the stationary points on the potential energy surface of each of the complexes has been located, a frequency analysis at the MP2/6-31+G(2d,p) level has been carried out for ensuring that no imaginary frequencies are obtained and therefore the structure corresponds to a minimum on the potential energy surface. As observed in previous work, this level of calculation is expected to provide a reasonable description of the interaction in this kind of systems.[39, 40] Also, the values obtained at the MP2/6-31+G(2p,d) level for water···water, ammonium···water and methylammonium···water dimers amount to -4.6, -19.8 and -17.6 kcal mol^{-1}, respectively, which compares pretty well with those obtained at the CCSD(T)/aug-cc-pVTZ level (-4.4, -20.6 and 18.3 kcal mol^{-1}).

Water molecules have been introduced in the clusters following chemical intuition and a systematic procedure, taking into account that water molecules will favorably interact with any of: a) the N-H free units of the ammonium cations, b) the phenol hydroxyl group, c) another water molecule present in the system. This procedure gives rise to a great variety of minima, which were selected according to complexation energy. When several minima show similar hydrogen bonds but slightly differ in the position of the methyl group in methylammonium, for example, only the most stable structure is considered.

For all the minima located, the complexation energy has been obtained by employing the supermolecule method,[41] using the counterpoise method to avoid basis set superposition error (BSSE).[41, 42] Therefore, the complexation energy is obtained as:

$$\Delta E_{complex} = E^{complex}(ijk...) - \sum_i E_i^{isolated}(i) - \sum_i \left(E_i^{complex}(ijk...) - E_i^{complex}(i) \right) \quad \text{(eq. 4.1)}$$

where terms in parentheses indicate the basis set, superscripts the geometry employed in the calculations and *i*, *j*, *k* the fragments constituting the complex. This quantity describes the formation of the complex from isolated fragments, and therefore is related to the total stability of the complex, which will increase as more water molecules are included, since more favorable interactions will be established. Since part of the interest of the present work relays on the effect of water molecules on the cation···π interaction, another quantity has been obtained, denoted by ΔE_{hyd}, corresponding to the process:

$$\phi\text{-OH} + M^+(H_2O)_n \rightarrow \phi\text{-OH}\cdots M^+\cdots(H_2O)_n .$$

This quantity is related to the formation of the complex from phenol and an already hydrated cation, describing the strength of the interaction between the phenol molecule and the rest of the system. It is expected to provide more information about the effects of water molecules on the cation···π interaction. All calculations have been performed by using Gaussian 09.[43]

4.3. Results

A discussion taking into account microhydration level has been chosen, so complexes of both ammonium and methylammonium with a given number of water molecules will be compared. Figure 4.1 shows the minimum energy structures found for the complexes formed by phenol and ammonium or methylammonium as obtained with the MP2/6-31+G(2d,p) level of calculation. As expected, in accordance with previous results for phenol complexes with alkali cations,[44] two structures have been found corresponding to the location of the ammonium cation over the phenyl ring and contacting with the hydroxyl oxygen. In minimum PA0-1 (the nomenclature reflects the phenol unit P, ammonium A or methylammonium M cation, the number of water molecules present in the system 0, and a numerical identifier for each structure) the ammonium cation interacts with the hydroxyl oxygen, forming a hydrogen bond at a distance of about 1.65 Å. This is a quite short hydrogen bond distance reflecting the larger intensity of the interaction due to the cationic nature of ammonium. On the other hand, in minimum PA0-2 the

cation is placed over the phenyl ring, forming a N-H···π hydrogen bond at a distance of 2.20 Å from the ring center. Also, the MP2/6-31+G(2d,p) level of calculation provides a third structure which is similar to PA0-2 but with the ammonium cation rotated by about 90º to be aligned with the O-C bond of phenol (N-H···π at 2.14 Å).

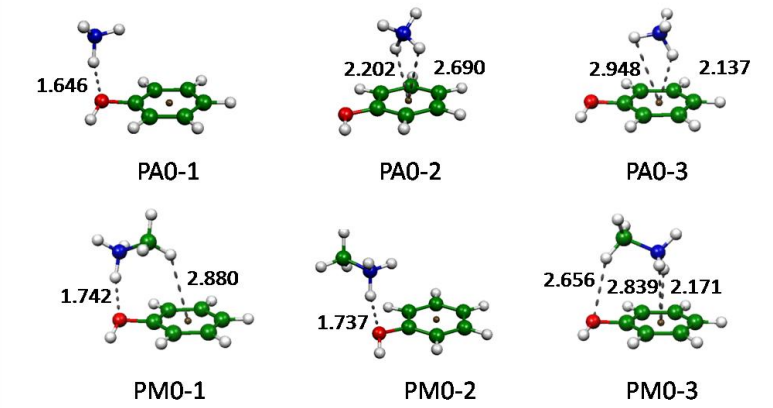


Figure 4.1. Structures of the minima found for the complexes formed by ammonium and methylammonium cation with phenol as obtained at the MP2/6-31+G(2d,p) level of calculation. Distances of selected contacts shown in Å.

As regards complexation energies, these are shown in Table 4.1. It can be observed that the most stable minimum corresponds to the interaction of the cation with the hydroxyl group, showing a complexation energy amounting to -19.9 kcal mol^{-1} at the MP2/6-31+G(2d,p) level of calculation, whereas minima PA0-2 and PA0-3 are less stable by about 1.5 kcal mol^{-1}. Inclusion of zero point energy corrections (ZPE) does not change this picture. Similar structural patterns have been obtained for methylammonium complexes, though in this case the methyl group establishes a secondary interaction with the aromatic ring or hydroxyl oxygen. Therefore, minimum PM0-1 presents the ammonium group of methylammonium interacting with the hydroxyl oxygen at a distance of 1.74 Å, whereas the methyl group participates in a C-H···π hydrogen bond at a distance of 2.88 Å from the ring center. In minimum PM0-2 the methyl group of ammonium is oriented away from the aromatic ring, with a N-H···O hydrogen bond at 1.74 Å. Finally, in minimum PM0-3 methylamonium cation is rotated 180º relative to PM0-1, so the ammonium group forms a N-H···π hydrogen bond at 2.17 Å from the ring center, whereas the methyl group contacts the hydroxyl oxygen forming a C-H···O contact at 2.66 Å.

Table 4.1. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the clusters formed by phenol and the cations studied in this work as obtained at the MP2/6-31+G(2d,p) level.

	Ammonium				Methylammonium		
	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}		$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}
PA0-1	-19.88	-18.87	-18.98	**PM0-1**	-18.18	-14.00	-13.84
PA0-2	-18.31	-17.34	-17.22	**PM0-2**	-17.84	-13.79	-13.51
PA0-3	-18.28	-17.27	-17.17	**PM0-3**	-17.73	-13.62	-13.62

The complexation energies shown in Table 4.1 indicate that the interaction is weaker in methylammonium complexes than in analogous ammonium ones, reaching -18.2 kcal mol^{-1} in the most stable minimum, corresponding again to contact with the hydroxyl oxygen. Minima PM0-2 and PM0-3 are within 0.5 kcal mol^{-1} above the most stable structure, so energy differences between minima are smaller than in ammonium complexes. The structures and energies are similar to those obtained by other authors for benzene complexes.[29, 45-49] BSSE amounts to around 1-2 kcal mol^{-1} in these complexes, though its effect in relative stabilities is smaller since all minima present similar values. When one water molecule is incorporated to the complexes, a variety of minima have been located, the most stable among which are presented in Figure 4.2. It is important to notice that, especially in the case of methylammonium complexes, there are several minima with similar geometrical arrangements and almost equal interaction energies. When this happens only the most stable structure of each kind is presented. For example, in structure PM1-1 the methyl group points in the opposite direction of O-H bond. There is another minimum with the methyl group pointing in the same direction, with a very similar energy, which for the sake of simplicity is not considered in the discussion of results. As observed in Figure 4.2, both ammonium and methylammonium present similar minima for the complexes containing one water molecule. Two of the minima with each cation present a cyclic pattern, where the water molecule interacts with one of the free N-H groups of the cation, but simultaneously acts as hydrogen donor in a hydrogen bond to the hydroxyl oxygen (PA1-1 and PM1-1) or to the phenyl ring (PA1-2 and PM1-2). This kind of structures with participation of the hydroxyl group, not possible in benzene clusters, is a key feature of the complexes formed with phenol. The rest of the structures shown in Figure 4.2 are similar, but in this case the water molecule does not interact directly with the phenol molecule. Considering complexation energies shown in Table 4.2, the most stable structure for ammonium complexes is PA1-1, with a complexation energy amounting to -37.8 kcal mol^{-1}.

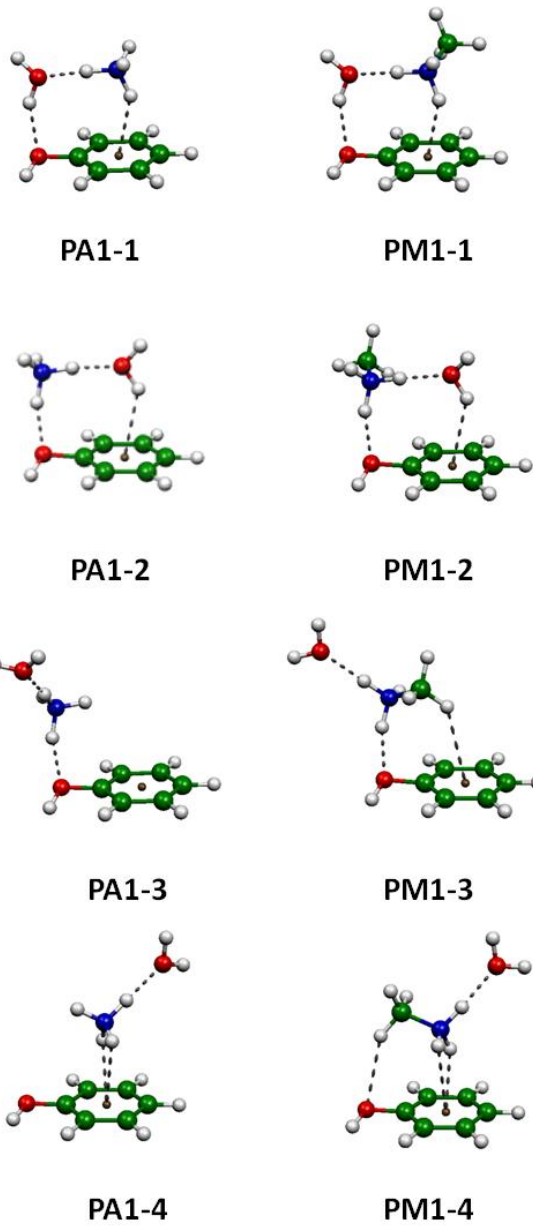


Figure 4.2. Selected most stable minima for the complexes formed by ammonium and methylammonium with phenol in the presence of one water molecule as obtained at the MP2/6-31+G(2d,p) level of calculation.

However, the second most stable structure PA1-2 is only 0.5 kcal mol^{-1} less stable. After these two structures, there is an energy gap of about 2 kcal mol^{-1} to the next stable minimum. It is worth noting that inclusion of the first water molecule makes the most stable minimum to be that with the cation over the phenyl ring, contrary to the behavior observed in the cluster without water molecules. This is a consequence of the secondary interaction established by the water molecule and the phenol moiety. The water molecule can establish a hydrogen bond with the hydroxyl oxygen, which is stronger than the hydrogen bond formed with the aromatic ring (typical values for water···water hydrogen bonds are around 4-5 kcal mol^{-1} whereas for O-H···π amount to 2-3 kcal mol^{-1}).[50] Therefore, even though the interaction between the cation and the phenol molecule is weaker over the ring, this loss is compensated with the additional strength of the O-H···O hydrogen bond formed. This effect can be clearly seen in structures PA1-3 and PA1-4; since there are no hydrogen bonds between water and phenol, the most stable structure is PA1-3, with the cation over the hydroxyl group. Also, the complexation energy difference between PA1-1 and PA1-4 allows an estimation of the contribution of the hydrogen bond to phenol of about 3.5 kcal mol^{-1} (1.9 kcal mol^{-1} for the water contact with the aromatic ring). Methylamonium complexes behave in a similar manner, showing complexation energies around 2.5-3 kcal mol^{-1} less negative than the corresponding ammonium minima. As before, no changes are observed after introduction of ZPE or enthalpy corrections.

Table 4.2. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the clusters containing one water molecule (see Figure 4.2) as obtained at the MP2/6-31+G(2d,p) level.

	Ammonium					**Methylammonium**			
	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}		$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}
PA1-1	-37.77	-34.37	-35.17	-18.56	**PM1-1**	-35.26	-29.00	-29.35	-18.03
PA1-2	-37.30	-34.30	-34.91	-18.03	**PM1-2**	-34.17	-28.20	-28.38	-16.95
PA1-3	-35.44	-32.81	-33.02	-16.07	**PM1-3**	-32.70	-27.12	-26.88	-15.36
PA1-4	-34.25	-31.70	-31.76	-14.94	**PM1-4**	-32.38	-26.94	-26.59	-15.04

The energy differences between analogous structures of ammonium and methylamonium is larger than in complexes without water, suggesting that ammonium cation is able to polarize more efficiently the water molecule, giving an extra stabilization to the complexes compared to methylammonium. Another way of quantifying the effect of the new water molecule is focusing on the energy change observed when a water molecule is added to the complexes without water.

Therefore, the formation of PA1-1 from PA0-3 implies an energy gain of -19.8 kcal mol^{-1}, whereas in forming PA1-2 from PA0-1 only changes by -17.4 kcal mol^{-1}. These 2.5 kcal mol^{-1} reflect the commented above about the different strength of O-H···π and O-H···O contacts. On the other hand, forming PA1-4 from PA0-3 changes the energy by -15.5 kcal mol^{-1}, so the formation of the water···phenol hydrogen bond gives an extra stabilization of 2-3.5 kcal mol^{-1}. The same is observed for methylammonium complexes, with changes of -17 kcal mol^{-1} and -15 kcal mol^{-1} in the presence and absence of the water···phenol hydrogen bond, respectively. The energy gain is smaller than in ammonium complexes, because in the formation of the water···phenol hydrogen bond, a contact between the methyl group and phenol has to be broken.

Considering the values obtained for ΔE_{hyd} corresponding to the interaction between phenol and the rest of the complex considered as a single unit, it can be appreciated that the values registered are similar to those obtained in the absence of water (of course are less negative than complexation energies obtained for complexes with one water molecule since the cation···water interaction is not included). Therefore, contrary to the usual trends observed in other systems where the presence of one water molecule decreases the strength of the interaction because it competes with the aromatic molecule for interacting with the cation,[26-30] in minima PA1-1 and PM1-1 ΔE_{hyd} amounts to -18.6 and -18.0 kcal mol^{-1}, respectively. Though the cation···π interaction is weakened as shown by the values obtained for PA1-3 and PM1-3 which exhibit decreases in the interaction strength of more than 3 kcal mol^{-1}, the formation of the O-H···O hydrogen bond in PA1-1 and PM1-1 compensates for this effect. In benzene complexes, the decrease in strength amounts to around 1.7 kcal mol^{-1} when the first water molecule is included.[29]

The inclusion of the second water molecule increases the complexity of the potential energy surface with lots of minima obtained. The six more stable ones for ammonium and methylammonium complexes are shown in Figure 4.3, with their complexation energies listed in Table 4.3. It can be appreciated that most of the minima shown in Figure 4.3 present a cyclic pattern of hydrogen bonds similar to those observed for the complexes containing one water molecule. Therefore, two different patterns arise: one presenting a π···H-N-H···O-H···O-ϕ hydrogen bond network, and the other with a series of ϕ-O···H-N-H···O-H···π contacts. Most of the structures in Figure 4.3 present these patterns (exceptions are PA2-5 and PM2-4), with the second water molecule occupying one of the free N-H units of ammonium cations, or hydrogen bonded as acceptor to the hydroxyl group of phenol. Therefore, as the second water molecule is included, the hydroxyl group starts participating in the hydrogen bond network of several of the most stable structures. This is an indication that the energy differences between coordinating the water molecule to the cation or to the hydroxyl group have diminished, being competitive with each other.

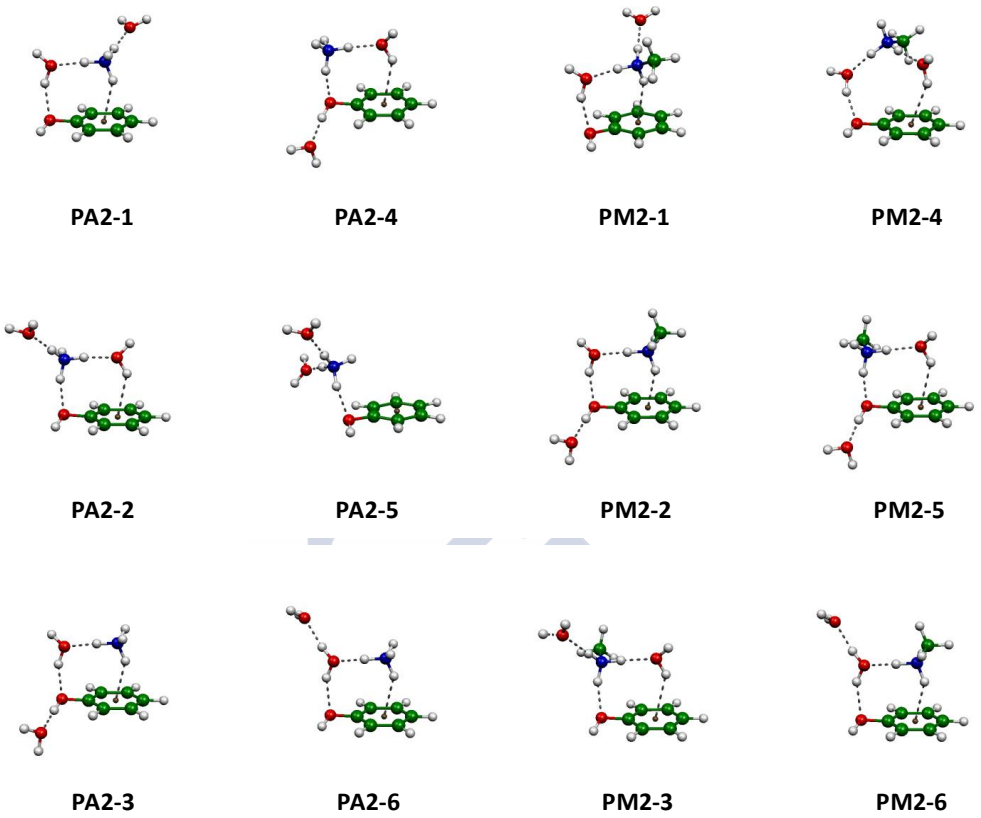


Figure 4.3. Selected most stable minima for the complexes formed by ammonium and methylammonium with phenol in the presence of two water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

Table 4.3 lists the values obtained for the complexation energies of the minima shown in Figure 4.3. Considering ammonium complexes, it can be observed that the two most stable structures differ in stability by only 0.5 kcal mol^{-1}, so including more water molecules decreases the difference in stability between structures presenting O-H···O and O-H···π contacts. The next structure in order of stability PA2-3 already presents a ϕ-OH···O hydrogen bond, and is 1.6 kcal mol^{-1} less stable than the most stable minimum found, and isoenergetic with the analogous structure bearing a O-H···π contact PA2-4. Minimum PA2-5 is unique since no contact between water molecules and the phenol moiety is established, with a drop in stability of 2.5 kcal mol^{-1} with respect to the most stable minimum. Finally, in PA2-6 the second water

molecule is hydrogen bonded to the previous one. The complexation energy for this structure is 2.9 kcal mol^{-1} less negative than that of the most stable minimum. Therefore, when a second water molecule is included in the cluster, there are three main possibilities: bonding to a N-H group, bonding to the hydroxyl oxygen and bonding to the previous water molecule. As deduced from the comparison of PA2-1, PA2-3 and PA2-6 the first option is the most favorable, the second being 1.6 kcal mol^{-1} less stable, whereas the third one decreases complexation energy by 2.9 kcal mol^{-1} with respect to bonding to the N-H group, or 1.3 kcal mol^{-1} with respect to hydroxyl oxygen.

Table 4.3. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the clusters containing two water molecules (see Figure 4.3) as obtained at the MP2/6-31+G(2d,p) level.

	Ammonium					**Methylammonium**			
	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}		$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}
PA2-1	-51.18	-46.38	-47.08	-15.83	**PM2-1**	-48.14	-40.50	-40.72	-16.00
PA2-2	-50.57	-46.01	-46.61	-15.16	**PM2-2**	-47.20	-38.81	-39.58	-14.55
PA2-3	-49.59	-44.06	-45.31	-13.81	**PM2-3**	-46.97	-39.70	-39.67	-14.80
PA2-4	-49.43	-44.45	-45.44	-13.56	**PM2-4**	-46.48	-38.56	-39.12	-14.56
PA2-5	-48.76	-44.59	-44.80	-13.31	**PM2-5**	-46.29	-38.37	-38.88	-13.61
PA2-6	-48.31	-43.13	-44.28	-13.25	**PM2-6**	-45.72	-37.58	-38.28	-13.86

In the case of methylammonium clusters, the behavior is pretty similar, though some differences arise. The most stable structure PM2-1 is analogous to the most stable minimum of ammonium complexes. However, the second most stable minimum, with a complexation energy only 0.8 kcal mol^{-1} less negative already presents a ϕ-OH···O contact, with methylamonium over the phenyl ring. It is worth noting that whereas in PM2-1 all N-H groups of the cation are occupied, in PM2-2 there is a free one, with the small impact in energies indicating that complexation with the cation or phenol already gives similar stabilization to the complex. In fact there is an inversion in the order of stability between structures PM2-2 and PM2-3 with respect to that observed in ammonium complexes, though the difference between these two structures amounts to only 0.3 kcal mol^{-1}. Structure PM2-4 is also very close in energy, being 1.6 kcal mol^{-1} less stable than the global minimum. In this structure both water molecules are bound to methylammonium N-H groups, establishing both O-H···O and O-H···π contacts with phenol. Therefore in this structure the interaction takes place between a hydrated methylammonium

cation and a phenol molecule, with no direct interaction between aromatic molecule and cation, in a similar way to that observed in benzene clusters with three water molecules.[29] Finally, in PM2-6 the new water molecule is hydrogen bonded to the previous one, with a stability loss of 2.4 kcal mol^{-1} with respect to the most stable minimum. So, even when the behavior is similar with both cations, in methylammonium complexes there are smaller differences for water to be coordinated to any of the favorable locations within the cluster, especially between N-H and phenol O-H group. In complexes with two water molecules, inclusion of ZPE does alter the order of stability already discussed, though when this happens it is a consequence of the structures being almost isoenergetic.

ΔE_{hyd} values shown in Table 4.3 are significantly less negative than those obtained in complexes with one water molecule. Since the second water molecule does not directly interact with phenol, there is no compensating effect for the decrease in the cation$\cdots\pi$ interaction as a consequence of the cation charge being shared among all neutral species in the complex. In fact, the charges obtained from a natural bond orbital (NBO) analysis indicate that the charge of the ammonium cation amounts to 0.93 in the ammonium$\cdots$phenol complex, decreasing to 0.91, 0.89 and 0.89 when including up to three water molecules (values for methylammonium are 0.94 a.u. in the complex with phenol and 0.92, 0.89 and 0.89 a.u. as water is included). Therefore, ΔE_{hyd} drops by about 2 kcal mol^{-1} with respect to the values observed in complexes with one water molecule.

The inclusion of a third molecule extremingly complicates the search for minima, since an overwhelming amount of minima can be located, the most stable of which are shown in Figure 4.4. The corresponding complexation energies are shown in Table 4.4. In the case of ammonium complexes, most structures show the cyclic hydrogen bond pattern observed in the smaller clusters, but other possibilities arise presenting a variety of hydrogen bond networks, with water molecules interacting among themselves (PA3-2 and PA3-8, for example), or participation of the hydroxyl group. The energy differences among structures are very small, with up to 7 minima within a complexation energy interval of only 1 kcal mol^{-1}. In any case the most stable minimum exhibits the same structure observed in smaller clusters with the third water molecule occupying the last free N-H group of ammonium cation. A similar pattern is observed in PA3-2, only 0.2 kcal mol^{-1} less stable. However, PA3-3, with a complexation energy only 0.4 kcal mol^{-1} less negative than the most stable structure, already shows a ϕ-OH$\cdots$O hydrogen bond, revealing a further decrease on the difference between coordinating to the cation or to the hydroxyl group. Therefore, already with three water molecules, there are plenty of structures with similar stability because the stability differences of the favorable regions for water coordination have diminished to be almost negligible. Due to the simultaneous participation of phenol hydroxyl group and aromatic ring in the hydrogen bond network, the structural patterns are similar to those found in ammonium$\cdots$water clusters with n+2 water molecules.[51]

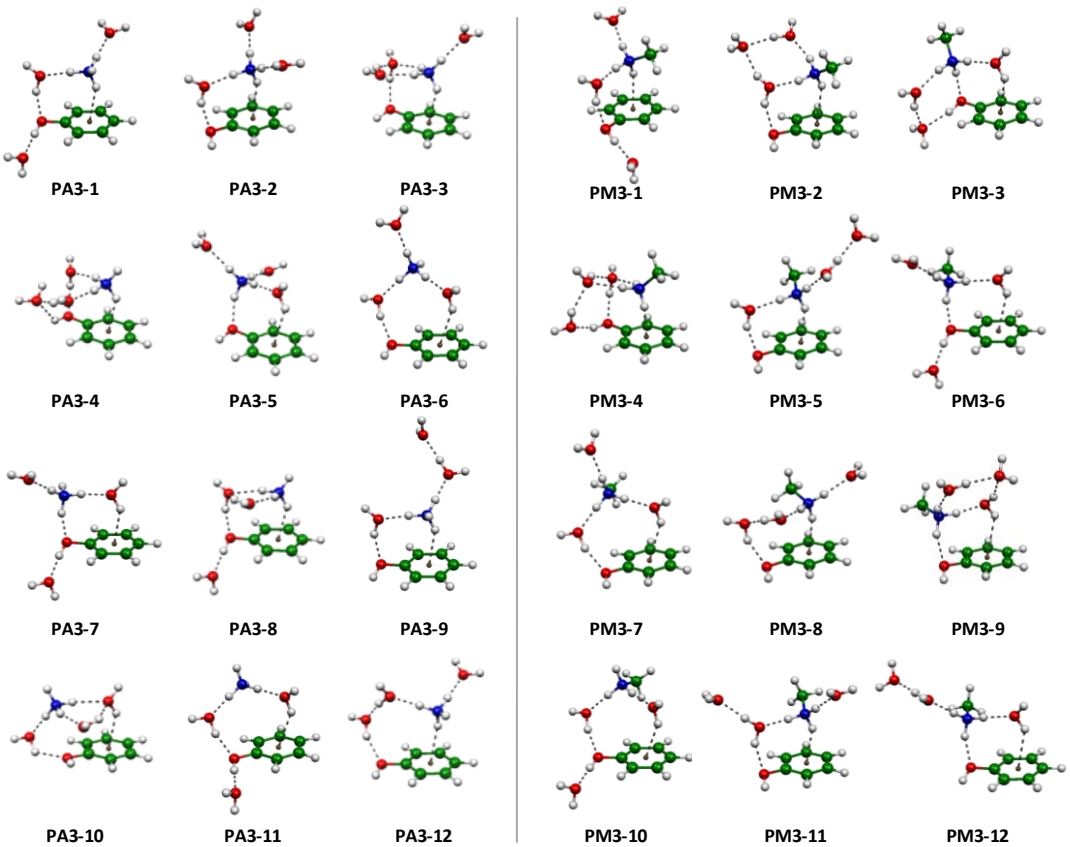


Figure 4.4. Selected most stable minima for the complexes formed by ammonium and methylammonium with phenol in the presence of three water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

The results obtained for methylammonium complexes are similar, though only four structures are within a 1 kcal mol^{-1} range from the most stable minimum. The main difference with ammonium complexes arises because in methylammonium cluster already with two water molecules and phenol there are no N-H free groups for the third water to be coordinated, so it must be incorporated to the hydrogen bond network. Therefore, the most stable structure already shows a ϕ-OH···O hydrogen bond, as also does PM3-3, the third most stable minimum. The rest of stable structures present O-H···O hydrogen bonds among water molecules or with water acting as hydrogen bond donor to the hydroxyl group. It becomes clear that including more water molecules already implies the formation of hydrogen bonds between water molecules on a second solvation shell or necessarily introduces ϕ-OH···O contacts. Almost any position

occupied by the water molecule will lead to minima with complexation energies of similar magnitude. In fact, analyzing the energy changes when including water molecules to the phenol···cation complexes, it becomes clear that the stabilization drops significantly when water molecules are included in the complex. Thus, when the first water molecule is included, the interaction changes by -17.9 kcal mol^{-1} in ammonium complexes and by -17.6 kcal mol^{-1} in methylammonium ones, as a consequence of both a new N-H···O contact but also due to the presence of a new O-H···O hydrogen bond to the hydroxyl group. The inclusion of the second water molecule stabilizes the complex in a significantly smaller amount, reaching -13.4 and -12.4 kcal mol^{-1} for ammonium and methylamonium complexes, respectively. So, the stabilization drops by 4.5 kcal mol^{-1} in ammonium and by 5.2 kcal mol^{-1} in methylammonium. This happens because the inclusion of the second water molecule only introduces a new N-H···O contact.

Table 4.4. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the clusters containing three water molecules (see Figure 4.4) as obtained at the MP2/6-31+G(2d,p) level.

	Ammonium					**Methylammonium**			
	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}		$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hyd}
PA3-1	-63.06	-56.79	-57.42	-13.31	**PM3-1**	-59.18	-49.47	-50.03	-13.54
PA3-2	-62.91	-55.65	-56.94	-13.59	**PM3-2**	-58.82	-49.07	-49.75	-14.43
PA3-3	-62.73	-56.56	-57.88	-13.60	**PM3-3**	-58.27	-48.17	-49.05	-13.14
PA3-4	-62.62	-53.84	-56.25	-13.67	**PM3-4**	-58.14	-46.35	-48.29	-13.30
PA3-5	-62.55	-56.43	-57.00	-13.04	**PM3-5**	-58.02	-48.36	-48.97	-13.10
PA3-6	-62.30	-55.83	-56.75	-12.98	**PM3-6**	-57.88	-48.68	-48.93	-12.19
PA3-7	-62.10	-55.53	-56.49	-12.10	**PM3-7**	-57.86	-48.72	-48.99	-12.99
PA3-8	-61.90	-54.10	-55.82	-12.14	**PM3-8**	-57.86	-47.90	-48.70	-13.07
PA3-9	-61.71	-54.69	-55.96	-12.46	**PM3-9**	-57.70	-48.52	-48.81	-13.28
PA3-10	-61.19	-53.59	-55.26	-11.91	**PM3-10**	-57.28	-47.41	-48.26	-11.82
PA3-11	-61.07	-53.92	-55.43	-11.31	**PM3-11**	-57.14	-47.69	-48.16	-12.22
PA3-12	-61.05	-54.25	-55.37	-11.83	**PM3-12**	-56.93	-47.52	-47.97	-11.98

Finally, when the third water molecule is included, the energy change amounts to -11.9 kcal mol^{-1} for ammonium and -11.5 kcal mol^{-1} for methylammonium, with an extra drop of 1.5 and 0.9 kcal mol^{-1}, respectively. These values indicate that the first water molecule is tightly bound within the complex whereas as more water molecules are included they are more loosely held in the cluster.

The same trends are observed in ΔE_{hyd} values, which become less negative as the more water molecules are included. The values for ΔE_{hyd} of the most stable complexes of ammonium amount to -19.9, -18.6, -15.8 and -13.3 kcal mol^{-1} for complexes from 0 to 3 water molecules. It becomes clear that the first water molecule hardly affects the interaction with phenol due to the compensation of the loss in cation···phenol interaction by the O-H···O hydrogen bond. However, as more water molecules are included the interaction strength changes by larger quantities, exceeding 2 kcal mol $^{-1}$. In the case of methylammonium complexes (-18.2, -18.0, -16.0 and -13.5 kcal mol^{-1}) the effect is similar, though in this case the first water molecule is able of totally recovering the loss of strength in the interaction. The inclusion of more water molecules produces a decrease in the interaction strength of more than 2 kcal mol^{-1}. Therefore, as observed in other systems,[26-30] the inclusion of water weakens the cation···π interaction. Nevertheless, the participation of the hydroxyl group interacting as hydrogen acceptor to one water molecule makes the weakening only evident in complexes with at least two water units.

4.4. Conclusions

Clusters formed by one phenol molecule and an ammonium or methylamonium cation in the presence of up to three water molecules have been computationally studied at the MP2/6-31+G(2d,p) level of calculation. Both ammonium and methylammonium form complexes interacting with the aromatic ring and the hydroxyl group of phenol with similar stabilities. However, in methylammonium complexes, secondary interactions are established between the methyl group and phenol.

The presence of water molecules greatly increases the complexity of the potential energy surfaces of the clusters though the minima located show similar characteristics for both cations. In any case, as one water molecule is incorporated to the system, the most stable minima present cyclic patterns with the water molecule bound to the cation and simultaneously establishing a hydrogen bond with the phenol molecule via the hydroxyl group or the aromatic ring. The inclusion of more water molecules does not break this pattern which is observed in the most stable structures of all clusters studied.

As more water molecules are included, the energy differences of the favorable interaction sites allowed to the new water molecule (contact with N-H of the cation, O-H of phenol or another water molecule) become smaller, so in clusters with two and three water molecules, several of the most stable minima present a ϕ-OH···O hydrogen bond between the phenol hydroxyl group and a water molecule.

The energy change upon formation of a complex with *n* water molecules from the *n-1* one decreases as more water molecules are included. The stabilization is especially significant for the first water molecule since it interacts simultaneously with the cation and the phenol molecule. The incorporation of the second and third molecules is accompanied by significantly smaller changes. Therefore, though the presence of water weakens the phenol···cation interaction, at least two water molecules are needed to produce a noticeable effect. The first water molecule partially recovers the loss in phenol···cation interaction by means of the hydrogen bond to phenol oxygen.

The results obtained in the present study can help understanding the interaction between ammonium cations and the side chain of tyrosine, especially in environments where the amino acid is only partially exposed to the solvent.

4.5. References

(1) L. M. Salonen, M. Ellermann, F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-4842.

(2) P. Hobza, R. Zaradnik, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and the biodisciplines*, Elsevier, Amsterdam, **1988**.

(3) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.

(4) J.-M. Lehn, *Supramolecular Chemistry: concepts and perspectives*, VCH, Weinheim, **1995**.

(5) F. Voegtle, Editor, *Supramolecular Chemistry: An Introduction*, Maruzen Co., Ltd., New York, **1995**.

(6) F. Voegtle, Editor, *Comprehensive Supramolecular Chemistry, Volume 2: Molecular Recognition: Receptors for Molecular Guests*, Pergamon, Oxford, **1996**.

(7) J. C. Ma, D. A. Dougherty. *Chem. Rev.* **1997**, *97*, 1303-1324.

(8) N. S. Scrutton, A. R. Raine. *Biochem. J.* **1996**, *319*, 1-8.

(9) J. P. Gallivan, D. A. Dougherty. *Proc. Nart. Acad. Sci. USA* **1999**, *96*, 9459-9464.

(10) D. A. Dougherty. *J. Nutr.* **2007**, *137*, 1504S-1508S.

(11) M. L. Waters. *Biopolymers (Peptide Science)* **2004**, *76*, 435-445.

(12) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.

(13) M. A. Anderson, B. Ogbay, R. Arimoto, W. Sha, O. G. Kisselev, D. P. Cistola, G. R. Marshall. *J. Am. Chem. Soc.* **2006**, *128*, 7531-7541.

(14) B. W. Berry, M. M. Elvekrog, C. Tommos. *J. Am. Chem. Soc.* **2007**, *129*, 5308-5309.
(15) R. M. Hughes, M. L. Benshoff, M. L. Waters. *Chem. Eur. J.* **2007**, *13*, 5753-5764.
(16) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2005**, *127*, 6518-6519.
(17) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 13586-13591.
(18) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 12735-12742.
(19) H. Khandelia, Y. N. Kaznessis. *J. Phys. Chem. B* **2007**, *111*, 242-250.
(20) P. E. Mason, C. E. Dempsey, G. W. Neilson, S. R. Kline, J. W. Brady. *J. Am. Chem. Soc.* **2009**, *131*, 16689-16696.
(21) A. J. Riemen, M. L. Waters. *Biochemistry* **2009**, *48*, 1525-1531.
(22) Z. Shi, C. A. Olson, N. R. Kallenbach. *J. Am. Chem. Soc.* **2002**, *124*, 3284-3291.
(23) C. D. Tatko, M. L. Waters. *Protein Sci.* **2003**, *12*, 2443-2452.
(24) K. E. Riley, P. Hobza. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3-17.
(25) C. D. Sherrill, *Computations of Non covalent π Interactions,* in Rev. Comput. Chem., *Vol.*, John Wiley & Sons, Inc., **2009**, pp.1-38.
(26) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.
(27) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.
(28) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.
(29) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.
(30) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.
(31) D. Feller, M. W. Feyereisen. *J. Comput. Chem.* **1993**, *14*, 1027-1035.
(32) H. Watanabe, S. Iwata. *J. Chem. Phys.* **1996**, *105*, 420-431.
(33) M. Gerhards, K. Kleinermanns. *J. Chem. Phys.* **1995**, *103*, 7392-7400.
(34) R. Wu, B. Brutschy. *Chem. Phys. Lett.* **2004**, *390*, 272-278.
(35) D. M. Benoit, D. C. Clary. *J. Phys. Chem. A* **2000**, *104*, 5590-5599.
(36) T. Ebata, A. Fujii, N. Mikami. *Int. J. Mass Spectrom. Ion Processes* **1996**, *159*, 111-124.
(37) C. Janzen, D. Spangenberg, W. Roth, K. Kleinermanns. *J. Chem. Phys.* **1999**, *110*, 9898-9907.
(38) W. Roth, M. Schmitt, C. Jacoby, D. Spangenberg, C. Janzen, K. Kleinermanns. *Chem. Phys.* **1998**, *239*, 1-9.
(39) M. S. Marshall, R. P. Steele, K. S. Thanthiriwatte, C. D. Sherrill. *J. Phys. Chem. A* **2009**, *113*, 13628-13632.
(40) H. M. Lee, P. Tarakeshwar, J. Park, M. R. Kołaski, Y. J. Yoon, H.-B. Yi, W. Y. Kim, K. S. Kim. *J. Phys. Chem. A* **2004**, *108*, 2949-2958.
(41) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.
(42) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.

(43) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. M. J. A, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. Gaussian 09, Revision B.01, Gaussian, Inc., *Wallingford, CT* **2009**.

(44) T. D. Vaden, J. M. Lisy. *J. Chem. Phys.* **2004**, *120*, 721-730.

(45) J. Y. Lee, S. J. Lee, H. S. Choi, S. J. Cho, K. S. Kim, T.-K. Ha. *Chem. Phys. Lett.* **1995**, *232*, 67-71.

(46) D. Majumdar, J. Leszczynski. *Comput. Lett.* **2007**, *3*, 257-265.

(47) M. Meot-Ner, C. A. Deakyne. *J. Am. Chem. Soc.* **1985**, *107*, 474-479.

(48) A. Pullman, G. Berthier, R. Savinelli. *J. Mol. Struct.: THEOCHEM* **2001**, *537*, 163-172.

(49) L. Pejov. *Chem. Phys.* **2002**, *285*, 177-193.

(50) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.

(51) Y.-S. Wang, H.-C. Chang, J.-C. Jiang, S. H. Lin, Y. T. Lee, H.-C. Chang. *J. Am. Chem. Soc.* **1998**, *120*, 8777-8788.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

5

Microhydration study: complexes between guanidinium cation and aromatic species

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

5.1. Introduction

Non covalent interactions with the participation of aromatic units are crucial in biochemistry.[1-4] The combination of a polarizable electron cloud and planar structure results in aromatic species to interact differently to aliphatic groups, so their presence can provide specificity. Among the different kinds of interactions involving aromatic molecules the cation···π interaction is the strongest, at least in the gas phase. Cation···π interactions are nowadays recognized as one of the key factors in determining the characteristics of proteins, together with hydrogen bonds, stacking interactions and salt bridges.[5-7]

The importance of cation···π interactions in controlling protein structure can be easily understood considering that several amino acids contain aromatic groups on their side chains.[3, 4, 6-8] On the other hand, protonation of the amino acids can be promoted by the pH of the environment in which proteins are immersed. Thus, amino acids with basic groups like histidine, lysine or arginine, are susceptible of protonation on their N-H groups. When it occurs, these protonated amino acids could establish cation···π interactions with the side chains or aromatic amino acids like phenylalanine, tyrosine or tryptophan.[3, 6-8]

The nature of cation···π interactions can be explained as a combination of large electrostatic contributions arising from the interaction of the cation with the quadrupole of the aromatic species, together with rather large induction effects as a consequence of the polarization of the aromatic cloud by the cation.[5, 7, 9, 10] However, the presence of solvent molecules can significantly change the strength of the interaction, which decreases as more solvent molecules are included.[7] It has been reported, however, that cation···π interactions fully exposed to solvent can be stronger than salt bridges, so their contribution can be crucial to protein stability.[11, 12] In any case, different studies give different results for the cation···π role in solution, the differences being considered a consequence of the different degree of exposure of the cation···π group to the solvent.[13-21] The structure of proteins can present hollows or cavities of different sizes, so there could be different levels of exposition to the solvent by the constituent amino acids. Thus, a given cation···π interaction, could be affected to a different extent depending on the amount of solvent molecules surrounding it. Theoretical methods are especially well suited for studying this kind of effect, since the progressive hydration of a given cation···π interaction can be modeled, providing information at a microscopic level which is usually non-affordable from experiment.

Several studies have dealt with the effects of microhydration on cation···π interactions, showing that the progressive solvation of the cation reduces the strength of the interaction.[7, 22-27] In a previous work, the effect of microhydration on the interaction of guanidinium cation with benzene was analyzed, showing that the presence of a small number of water molecules can alter

the stabilities and geometrical arrangements of the complexes.[22] On the other hand, hydration of ammonium···phenol complexes has shown that their behavior is more complex than benzene ones as a consequence of the participation of the hydroxyl group in the hydrogen bond network.[23] The participation of the hydroxyl group has also been confirmed in hydrated complexes formed by phenol and alkali cations.[27, 28] In a similar way, the pyrrol ring in indole constitutes a new favorable site for the interaction with cations, absent in benzene, which can also reflect on a more complex behavior.

In the present work, the effect of microhydration on cation···π interactions has been studied by introducing up to three water molecules in complexes formed by guanidinium cation and benzene, phenol or indole, as a way of modeling cation···π interactions among amino acids. The results obtained could help understanding the effect of changing the aromatic unit and microhydration level on the characteristics of cation···π interactions relevant in proteins.

5.2. Computational details

Complexes formed by a guanidinium cation, benzene, phenol or indole, and up to three water molecules have been optimized using the M06-2X/6-31+G* level of calculation.[29] After a stationary point has been located a frequency analysis has been carried out at the same level of calculation in order to ensure that the structure corresponds to a minimum in the potential energy surface of the cluster.

Different starting structures have been considered taking into account proposals found in literature and also structures based on chemical knowledge (basically systematically expanding the hydrogen bond network).[22-24, 26, 30] The process has been as follows: first, complexes formed by guanidinium and each of the aromatic molecules have been optimized starting from different structures where the orientation of guanidinium with respect to the aromatic units has been changed (parallel, parallel-displaced, T-shaped, …).[22, 30] Once the complexes between guanidinium and the aromatic molecules are located, a water molecule has been incorporated to the complex in each of the favorable positions available, such as the NH_2 groups of guanidinium or the hydroxyl group of phenol. After the minima with one water molecule have been located, the procedure is repeated with two and three water molecules. Basically the water molecules are incorporated interacting directly with guanidinium, forming hydrogen bonds with a previous water molecule or forming a hydrogen bond with the O-H group of phenol or with the N-H group of indole. Also, O-H···π hydrogen bonds to the aromatic rings have been considered.

Complexation energies for the minima thus located have been obtained at the M06-2X/6-31+G* and the MP2/aug-cc-pVDZ levels of calculation. Taking into account that MP2 is known to overbind clusters where aromatic units are present,[31-33] especially when stacked, empirical corrected MP2 variants have been considered. These so-called Spin Component Scaled MP2 methods are based on an empirical scaling of the contributions of parallel and antiparallel electron pairs to the correlation energy.[34] Therefore, the scaled MP2 energy is obtained as:

$$\mathrm{SCS\text{-}X} = \mathrm{E_{RHF}} + \mathrm{pt\,E_{2,(\alpha\alpha+\beta\beta)}} + \mathrm{ps\,E_{2,(\alpha\beta)}}, \qquad \text{(eq. 5.1)}$$

where pt = 0.33 ; ps = 1.20 for SCS-MP2[34] and pt = 1.76 ; ps = 0 for SCSN-MP2[35]. Of course for the original MP2 both parameters are 1.

In order to avoid the basis set superposition error (BSSE),[36] the counterpoise method has been used to obtain the complexation energy of all the minima located.[36, 37] Therefore, interaction energies are obtained employing the full basis set of the complex as:

$$\Delta E_{\mathrm{int}} = E^{complex}(ijk...) - \sum_i E_i^{complex}(ijk...). \qquad \text{(eq. 5.2)}$$

Terms in parentheses indicate the basis set while superscripts refer to the geometry employed in the calculations. As the geometry of the molecules changes when the cluster is formed, an additional contribution describing this effect must be included, obtained as the energy difference between the molecules in the cluster geometry and in isolation.

$$E_{def} = \sum_i \left(E_i^{complex}(i) - E_i^{isolated}(i)\right). \qquad \text{(eq. 5.3)}$$

Finally, the complexation energy is the combination of these two quantities:

$$\Delta E_{complex} = \Delta E_{\mathrm{int}} + E_{def}. \qquad \text{(eq. 5.4)}$$

In order to obtain more information about the nature of the interaction with the different aromatic systems, Symmetry Adapted Perturbation Theory[38, 39] calculations based on DFT (SAPT(DFT)) have been carried out for the complexes without water.[40, 41] Since many body effects are not easily treated with this method, the analysis has been restricted to non-hydrated systems. These calculations have been performed at the LPBE0AC/aug-cc-pVDZ level employing Molpro.[42] This involves a correction by adding a shift to the asymptotic part of the potential. This shift is obtained as the sum of the ionization potential and the energy of the

highest occupied molecular orbital, as obtained at the PBE0/aug-cc-pVDZ level of calculation. The DFT-SAPT calculations were performed with the aug-cc-pVDZ basis set, employing the cc-pVTZ/JKFIT for Hartree–Fock and aug-cc-pVDZ/MP2FIT for the second-order dispersion terms. The rest of the calculations have been performed by using Gaussian09.[43]

5.3. Results

After optimization a variety of minima have been obtained, their number quickly growing as more water molecules are included. Therefore, results will be presented for a selection of the most stable minima found, numbered and sorted by their stability at the MP2/aug-cc-pVDZ level.

Figure 5.1 shows the minima found for the complexes formed by the guanidinium cation and the aromatic systems employed in this work: benzene, indole and phenol, whereas Table 5.1 list the values obtained for complexation energies.

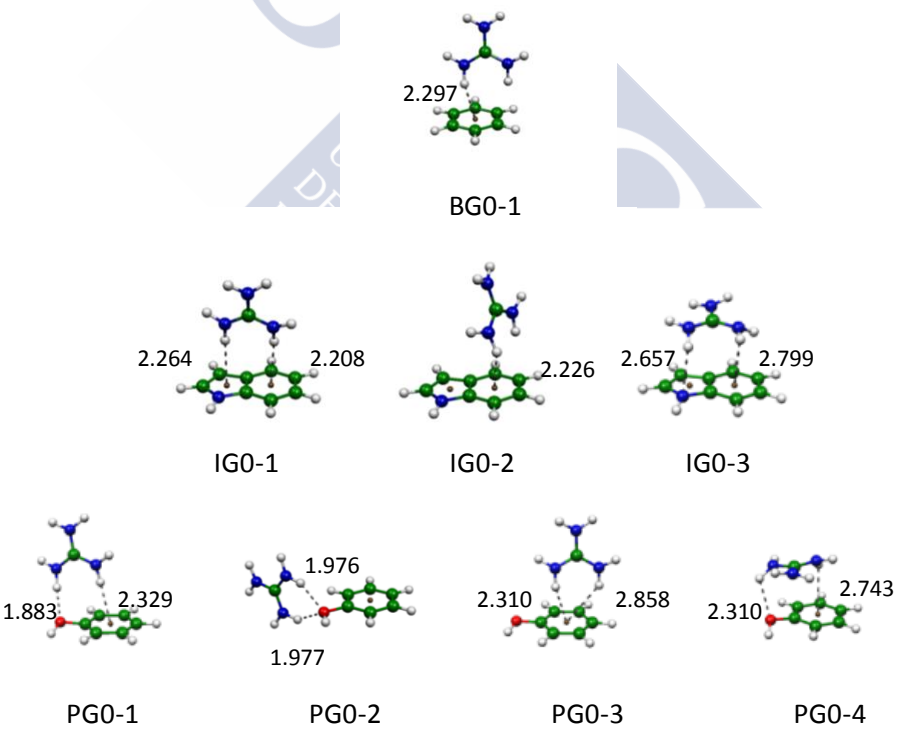


Figure 5.1. Structures of the minima obtained for complexes containing guanidinium cation and the aromatic molecules. Selected distances are shown in Å.

One minimum has been found for guanidinium-benzene complex, corresponding to a T-shaped structure where guanidinium cation interacts with benzene by means of a double contact with the NH_2 groups, in agreement with previously published results.[22, 30] In this minimum BG0-1 (the nomenclature reflects the benzene B, indole I, or phenol P unit, guanidinium G cation, the number of water molecules present in the system (0 in this case), and a numerical identifier for each structure) guanidinium interacts with the C-C midbonds, establishing a N-H···π contact at about 2.30 Å to the center of the ring. The complexation energy amounts to around -14 kcal mol^{-1}.

In the case of indole complexes three different structures have been found. IG0-1 corresponds to a minimum with guanidinium in a T-shaped structure interacting simultaneously with both rings of indole; IG0-2 is similar, but the cation is rotated and interacts only with the phenyl ring of indole; IG0-3 corresponds to a parallel-stacked structure with guanidinium interacting with both rings of indole. In the latter case, the distances from guanidinium to the ring center are longer, since there are no N-H···π hydrogen bonds (see Figure 5.1). As regards complexation energies IG0-1, as expected, is the most stable minimum, reaching around -21 kcal mol^{-1} with the different methods. That is, the presence of a larger aromatic molecule increases the strength of the interaction by about -7 kcal mol^{-1} with respect to the structure with benzene. IG0-2 is about 2-3 kcal mol^{-1} less stable, whereas IG0-3 is the least stable of the three minima, with a complexation energy of -16 kcal mol^{-1}.

Table 5.1. Complexation energies (kcal mol^{-1}) for the complexes formed by guanidinium cation and the aromatic molecules considered in this work as obtained at the M06-2X/6-31+G* and different versions of MP2/aug-cc-pVDZ.

	$\Delta E_{M06\text{-}2X}$	ΔE_{MP2}	$\Delta E_{SCS\text{-}MP2}$	$\Delta E_{SCS(N)\text{-}MP2}$
BG0-1	-13.42	-14.19	-12.39	-13.83
IG0-1	-20.51	-21.26	-18.70	-20.70
IG0-2	-17.93	-18.81	-16.58	-18.24
IG0-3	-17.06	-16.40	-13.72	-15.06
PG0-1	-17.20	-17.42	-15.56	-17.07
PG0-2	-17.22	-16.13	-14.85	-15.96
PG0-3	-14.04	-14.85	-12.97	-14.55
PG0-4	-13.68	-12.34	-10.09	-11.23

Phenol complexes with guanidinium show the four minima presented in Figure 5.1. It can be observed that PG0-1, PG0-3 and PG0-4 approximately correspond to the minima found for indole complexes, with the hydroxyl group of phenol playing the role of the second aromatic ring in indole. Besides, there is a minimum, PG0-2, showing guanidinium interacting only with the hydroxyl oxygen by means of two equivalent N-H···O hydrogen bonds. The complexation energies in Table 5.1 show that the most stable minimum is PG0-1, reaching -17 kcal mol^{-1}, but closely followed by PG0-2. In fact, in the case of the M06-2X/6-31+G* results, PG0-2 is as stable as PG0-1, but it can be observed that, compared to MP2 values, M06-2X seems to overestimate the interaction in structures showing N-H···O hydrogen bonds. As in indole complexes, PG0-3 and the parallel stacked PG0-4 are less stable by a couple of kcal mol^{-1}. Overall it can be observed that the strength of the interaction increases from benzene to phenol to indole, as a consequence of the double simultaneous interaction with the ring and the larger size of indole.

As regards the method employed it can be confirmed that the results obtained with any of the methods are pretty similar with the exception of SCS-MP2/aug-cc-pVDZ, which tends to underestimate the strength of the interaction compared with the others. Also, there are almost no differences between the regular MP2 and SCSN-MP2, and both agree quite well with the M06-2X/6-31+G* results.

In order to understand why the interaction increases from benzene to indole, SAPT(DFT) calculations have been performed for the minima shown in Figure 5.1. The results are displayed in Figure 5.2. First, in benzene complex it becomes clear that the main contribution to the interaction energy is electrostatic (-10.0 kcal mol^{-1}). This reflects the N-H···π hydrogen bonds formed by the charged guanidinium cation and the benzene ring. Also, as a consequence of a large electrostatic interaction, a significant induction contribution is also observed (-8.4 kcal mol^{-1}), as a result of the deformation of the polarizable π cloud by the charged species nearby. However, the contribution of dispersion is significant, reaching –6.3 kcal mol^{-1}, despite the T-shaped orientation of guanidinium relative to benzene. Finally, a large repulsion contribution of a magnitude similar to the electrostatic one is found. These results agree with those previously published,[22, 30] and confirm the important role of dispersion in this kind of complexes. As comparison, in benzene-sodium complexes, the dispersion contribution barely reaches -1.5 kcal mol^{-1}.[10, 44]

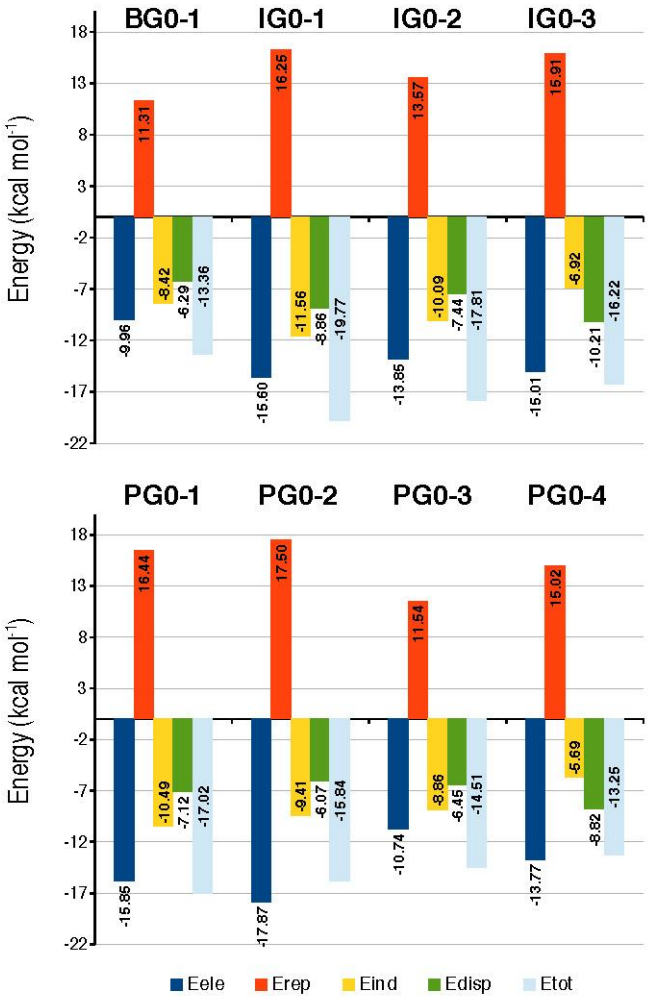


Figure 5.2. SAPT(DFT) energy decomposition for the complexes shown in Figure 5.1.

Indole complexes behave quite differently. As observed in Figure 5.2, the contribution from electrostatics is larger than that observed in benzene for any of the minima shown in Figure 5.1. That is, the presence of a permanent dipole in indole produces a stronger electrostatic contribution that reaches -15.6 kcal mol^{-1} in the most stable complex. Quite surprisingly, the electrostatic contribution is almost as large in the parallel-stacked minimum IG0-3, of about -15.0 kcal mol^{-1}, whereas decreases when guanidinium cation interacts only with the phenyl ring in indole (-13.9 kcal mol^{-1}), though in all cases is still larger than in benzene complex. As expected,

taking into account the larger size of the polarizable aromatic cloud, induction contributions are large, amounting to –11.6 kcal mol^{-1} in the most stable complex, and decreasing to -10.0 kcal mol^{-1} in IG0-2. However, in the case of the parallel stacked minimum IG0-3 there is an important decrease of polarization, which does not reach -7 kcal mol^{-1} and therefore is similar to that in benzene. Dispersion is also larger than in benzene complex, amounting to –8.9 and -7.4 kcal mol^{-1} in minima IG0-1 and IG0-2, respectively. However, in minimum IG0-3, the parallel orientation of guanidinium relative to indole produces a larger contribution of dispersion, which is the largest among the minima considered, reaching -10.2 kcal mol^{-1}, and partially compensating the smaller contribution of polarization. Repulsion is huge in all structures though it favors IG0-2.

Phenol complexes display similar characteristics. The electrostatic contribution is large, reaching -15.9 kcal mol^{-1} in the most stable complex, but amounting to -17.9 kcal mol^{-1} in PG0-2, the largest electrostatic contribution among minima in Figure 5.1. This is a consequence of the double N-H···O hydrogen bond formed in this minimum. The smaller electrostatic contribution, as in indole, occurs in PG0-3 where guanidinium interacts only with the phenyl ring. In accordance with the large electrostatic contribution, induction terms are also important, amounting to around -10 kcal mol^{-1} in the most stable structures. As in indole complexes, the parallel orientation of guanidinium produces a decrease in polarization contribution, which only amounts to -5.7 kcal mol^{-1}. Dispersion is also large, being midway between the values observed for benzene and indole, and reaching -7.1 kcal mol^{-1} in the most stable complex. As in indole, the largest contribution is observed in the parallel PG0-4, amounting to -8.8 kcal mol^{-1} and compensating the loss in induction.

Overall, it becomes clear that in the interaction of guanidinium with aromatic molecules, the electrostatic contribution is usually the leading term, but with important contributions from induction and dispersion. Moving from benzene to phenol to indole, all contributions to the interaction energy are increased, especially electrostatic and dispersion ones. This latter contribution is more significant in parallel minima, so it is crucial in order to explain the formation of stacked structures.

Figure 5.3 shows the selected most stable minima found for the complexes formed by guanidinium and each of the aromatic species studied in the presence of one water molecule. Guanidinium can easily accommodate up to three water molecules by means of double NH···O hydrogen bonds, and at most only one of these positions is blocked in the minima in Figure 5.1. Thus, it can be expected the water molecule to be bonded to guanidinium cation whereas the cation itself interacts with the aromatic unit as in Figure 5.1. However, as already commented before, in benzene complexes a new possibility arises, with water intercalated between the cation

and the aromatic ring. The most stable structure (Table 5.2) simply corresponds to BG0-1 with one water molecule coordinated to guanidinium, reaching -29 kcal mol^{-1}. The other structure BG1-2 with water hydrogen bonded to benzene is less stable by about 4 kcal mol^{-1}.

Complexes with indole are more difficult to analyze since more structures are possible, so only a selection of the most stable ones are shown in Figure 5.3. The most stable structure with MP2/aug-cc-pVDZ corresponds, as in benzene, to IG0-1 plus one water molecule coordinated to guanidinium (IG1-1), whereas the cation···π contact remains almost unperturbed. In the case of IG1-2 and IG1-3, both come from IG0-3, with the cation oriented in parallel to the aromatic ring and one water molecule bonded to it. However, since the aromatic cloud of indole is more extended than that of benzene, the water molecule also interacts simultaneously with the other ring of indole (with the pyrrol ring in IG1-2 and the phenyl one in IG1-3).

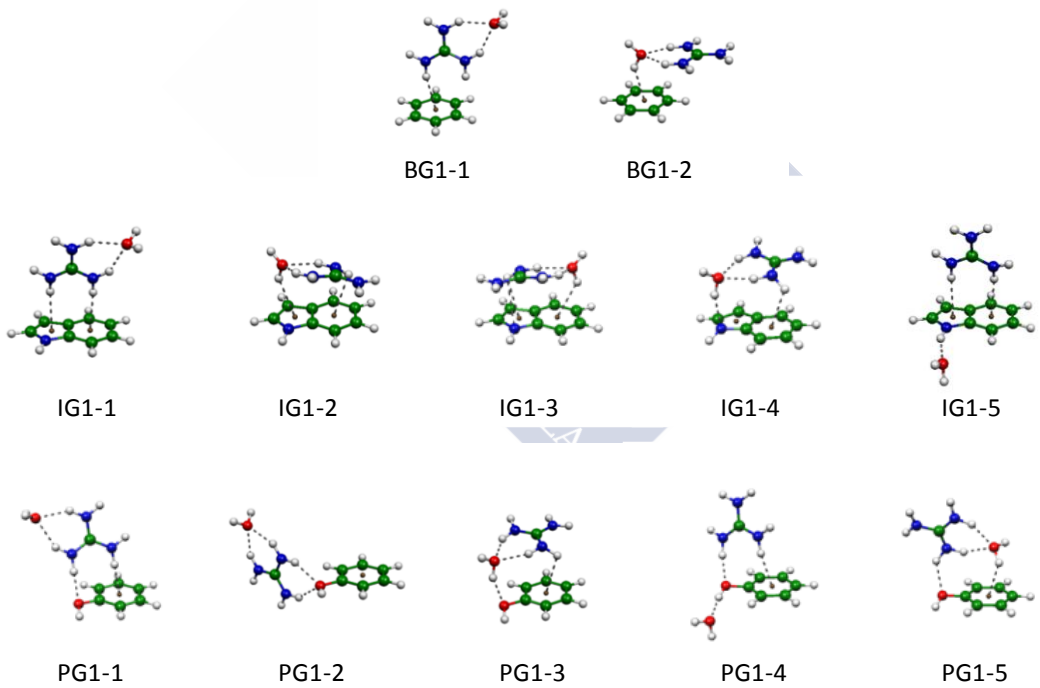


Figure 5.3. Structures of selected minima obtained for complexes containing one water molecule.

Table 5.2. Complexation energies (kcal mol^{-1}) for selected complexes containing one water molecule as obtained at the M06-2X/6-31+G* and different versions of MP2/aug-cc-pVDZ.

	$\Delta E_{M06\text{-}2X}$	ΔE_{MP2}	$\Delta E_{SCS\text{-}MP2}$	$\Delta E_{SCS(N)\text{-}MP2}$
BG1-1	-30.79	-29.05	-26.42	-29.22
BG1-2	-28.38	-25.23	-22.39	-25.08
IG1-1	-37.01	-35.29	-31.93	-35.21
IG1-2	-35.26	-32.54	-28.59	-31.90
IG1-3	-34.25	-31.40	-27.44	-30.78
IG1-4	-34.20	-31.10	-27.43	-30.28
IG1-5	-31.12	-30.82	-27.35	-30.52
PG1-1	-34.04	-31.76	-29.00	-31.89
PG1-2	-34.41	-30.80	-28.64	-31.08
PG1-3	-33.90	-30.11	-26.64	-29.71
PG1-4	-30.89	-29.14	-26.14	-29.25
PG1-5	-31.91	-28.20	-25.32	-28.16

With any of the methods employed, coordination of guanidinium to the phenyl ring is favored. The same happens in phenol complexes with one water molecule. The most stable minima PG1-1 and PG1-2 correspond to PG0-1 and PG0-2 with the water molecule coordinated to one free position in guanidinium. About 2 kcal mol^{-1} less stable there is a structure like PG1-5 with guanidinium interacting with phenol by only one NH_2 group. Also, it is worth noting that when water coordinates to the phenol hydroxyl group by means of a hydrogen bond (PG1-4), the stability is similar, so this kind of structures start being competitive with others where the water molecule is directly coordinated to the cation. Comparing PG1-1 and PG1-4 it can be observed that the energy difference between coordinating to the cation or the hydroxyl group only introduces an energy difference of 3 kcal mol^{-1} (as already observed in other cation-phenol complexes[23, 27]). In any case, in phenol complexes, the three most stable structures are virtually isoenergetic even though the interaction pattern is totally different, with guanidinium interacting with the aromatic ring, with the hydroxyl group or in a parallel orientation with water hydrogen bonded to phenol.

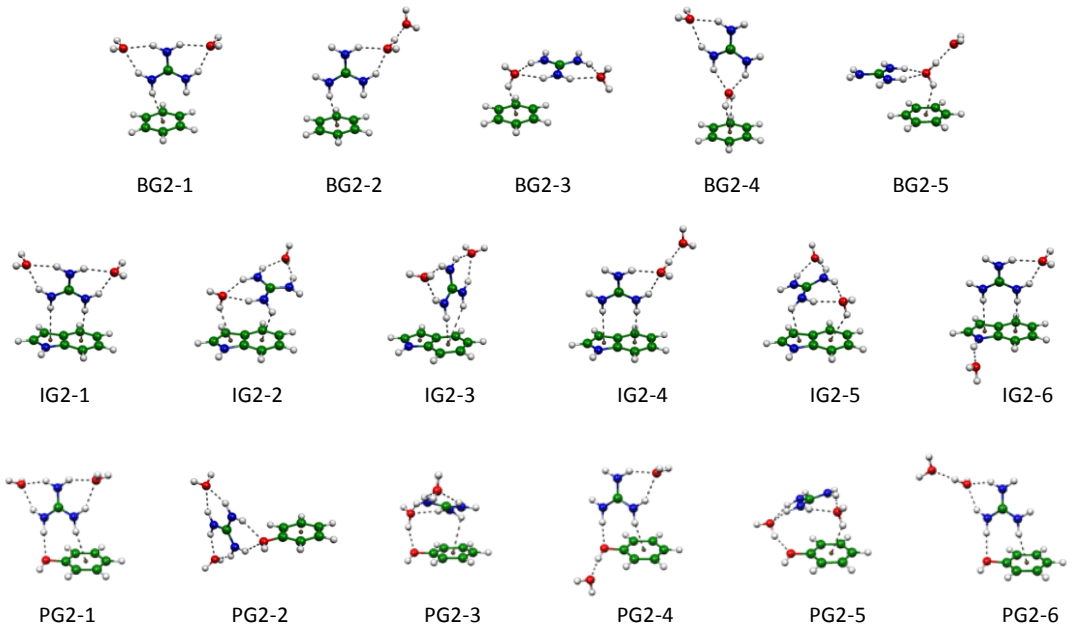


Figure 5.4. Structures of selected minima obtained for complexes containing two water molecules.

Including the second water molecule increases the number of minima with similar complexation energies. A selection of these is shown in Figure 5.4, with complexation energies listed in Table 5.3. In benzene complex, the most stable minimum found (-43 kcal mol^{-1}) corresponds to the T-shaped BG0-1 minimum with the two water molecules occupying the free coordination sites of guanidinium. It would be expected that most stable structures with two water molecules will present similar characteristics, with water molecules interacting directly with the cation. However, a minimum like BG2-2 is less stable than BG2-1 by 2 kcal mol^{-1} but with similar stability as other structures as BG2-3 and BG2-4. The peculiarity is that in BG2-2, one of the water molecules forms a hydrogen bond to the other water unit without interacting directly with the cation. That is, already with two water molecules, hydrogen bonding between water molecules starts being competitive with the cation···water interaction. A similar behavior is observed in complexes with indole. In this case, the most stable structure corresponds to IG2-1 with the two water molecules coordinated to guanidinium and complexation energy of -48 kcal mol^{-1}. Other possibilities with water interacting with the ring (IG2-2) or with water···water hydrogen bond (IG2-4) are around 2 kcal mol^{-1} less stable. In phenol complexes three structures present the same stability (around -45 kcal mol^{-1}), corresponding to contacts of guanidinium with the phenyl ring and hydroxyl group, double contact with hydroxyl, and finally

a parallel structure with a OH···O hydrogen bond to the hydroxyl group. These minima correspond to the most stable monohydrated clusters, with the second water molecule in the free NH_2 groups of guanidinium cation.

Table 5.3. Complexation energies (kcal mol^{-1}) for selected complexes containing two water molecules as obtained at the M06-2X/6-31+G* and different versions of MP2/aug-cc-pVDZ.

	$\Delta E_{M06\text{-}2X}$	ΔE_{MP2}	$\Delta E_{SCS\text{-}MP2}$	$\Delta E_{SCS(N)\text{-}MP2}$
BG2-1	-46.54	-42.52	-39.12	-43.06
BG2-2	-43.52	-39.77	-36.39	-40.45
BG2-3	-44.88	-39.34	-35.70	-39.61
BG2-4	-44.97	-39.17	-36.10	-39.96
BG2-5	-40.20	-35.45	-31.84	-35.70
IG2-1	-52.05	-48.06	-43.96	-48.34
IG2-2	-51.44	-46.43	-41.76	-46.27
IG2-3	-49.74	-46.14	-42.43	-46.55
IG2-4	-49.38	-45.70	-41.61	-46.12
IG2-5	-50.22	-45.35	-40.75	-45.08
IG2-6	-47.16	-44.40	-40.15	-44.56
PG2-1	-49.44	-44.84	-41.36	-45.36
PG2-2	-50.06	-44.08	-41.08	-44.72
PG2-3	-49.86	-44.05	-39.93	-44.21
PG2-4	-47.20	-43.05	-39.24	-43.63
PG2-5	-49.84	-42.75	-38.25	-42.82
PG2-6	-46.57	-42.32	-38.79	-42.94

Guanidinium cation can establish three simultaneous contacts by means of its NH_2 groups. In the most stable minima discussed so far one of these positions is occupied by the aromatic molecules, whereas the other two are available for accommodating up to two water molecules. The inclusion of the third water molecule changes this behavior because now at least one of the water molecules or the aromatic unit cannot interact directly with the NH_2 groups of guanidinium. Figure 5.5 shows the most stable minima found with three water molecules, and

Table 5.4 lists their corresponding complexation energies. The most stable benzene cluster (-53 kcal mol^{-1}) corresponds to a T-shaped structure with one of the water molecules hydrogen bonded to a previous one, being part of a second solvation shell. In any case, this structure is almost isoenergetic with BG3-2, with a trihydrated guanidinium cation interacting in parallel orientation with benzene, or BG3-3, with a trihydrated guanidinium cation interacting with benzene by means of a O-H···π hydrogen bond. Other structures similar to these, but with one water molecule in a second solvation shell are less stable by 2-3 kcal mol^{-1}.

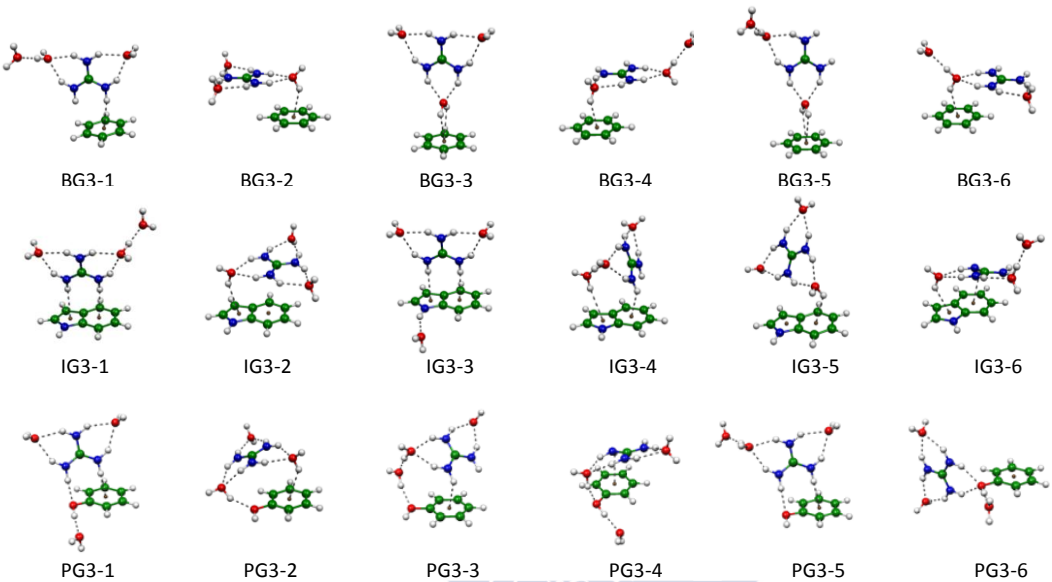


Figure 5.5. Structures of selected minima obtained for complexes containing three water molecules.

Indole complexes behave similarly, the most stable structure (-58 kcal mol^{-1}) being T-shaped but with similar stability to a parallel one as IG3-2, and even with a T-shaped structure with one water molecule participating in an N-H···O hydrogen bond as IG3-3. The energy difference between the O-H···O hydrogen bond and the N-H···O one amounts to 1-2 kcal mol^{-1}. Finally, phenol complexes are slightly different. The most stable minimum (-56 kcal mol^{-1}) is a T-shaped structure with guanidinium coordinated to two water molecules whereas the third one interacts with the hydroxyl oxygen in an O-H···O hydrogen bond. In any case, several minima in different orientations of guanidinium (parallel, T-shaped) and different hydrogen bond patterns show similar stabilities.

Table 5.4. Complexation energies (kcal mol^{-1}) for selected complexes containing three water molecules as obtained at the M06-2X/6-31+G* and different versions of MP2/aug-cc-pVDZ.

	$\Delta E_{M06\text{-}2X}$	ΔE_{MP2}	$\Delta E_{SCS\text{-}MP2}$	$\Delta E_{SCS(N)\text{-}MP2}$
BG3-1	-58.53	-52.55	-48.42	-53.56
BG3-2	-59.91	-52.06	-47.57	-52.60
BG3-3	-60.04	-51.98	-48.17	-53.11
BG3-4	-57.26	-49.78	-45.31	-50.49
BG3-5	-57.25	-49.56	-45.76	-50.86
BG3-6	-56.00	-48.81	-44.38	-49.41
IG3-1	-63.72	-57.87	-53.07	-58.63
IG3-2	-64.38	-57.14	-51.59	-57.09
IG3-3	-61.74	-56.77	-51.83	-57.31
IG3-4	-63.07	-56.21	-50.73	-56.58
IG3-5	-63.32	-56.05	-51.39	-56.72
IG3-6	-62.88	-56.01	-50.56	-56.30
PG3-1	-62.03	-55.59	-51.02	-56.53
PG3-2	-64.70	-55.50	-50.24	-55.93
PG3-3	-62.14	-55.00	-50.03	-55.81
PG3-4	-62.62	-54.87	-49.55	-55.41
PG3-5	-61.32	-54.83	-50.61	-55.83
PG3-6	-62.46	-54.67	-50.70	-55.89

Overall, it becomes clear that in all cases the preferred structure of the minima corresponds to a T-shaped one with guanidinium interacting with the phenyl ring of benzene, both rings of indole or the phenyl ring and oxygen (or only oxygen) in phenol. As water molecules are introduced this behavior does not change, since the guanidinium cation still offers two vacant positions for interaction with water. However, already for two water molecules, other hydrogen bonds start being competitive with the cation, so hydrogen bonds involving the hydroxyl group of phenol or the N-H group in indole start being present among the most stable structures. With

three water molecules two situations arise. First, the three water molecules coordinate directly to guanidinium, so the hydrated cation interacts parallelly with the aromatic molecule. In the case of phenol and indole complexes, this possibility is also favored by the formation of O-H···X hydrogen bonds with one of the water molecules (X=O, π) at the same time the cation interacts with the phenyl ring. The second option is a T-shaped minimum where guanidinium interacts with the aromatic ring and two water molecules. This structure appears among the most stable minima in all clusters. With this basic structural motif, different minima are possible depending on whether the third water molecule binds to the hydroxyl group of phenol, the NH group of indole or to another water molecule. Benzene clearly favors coordination in the second shell, whereas for indole and phenol, the participation of N-H and O-H groups is relevant. Also, it can be observed that indole shows a larger amount of parallel clusters among the most stable structures than any of the other aromatic molecules, as a consequence of its larger size and more dispersive character of the interaction.

Figure 5.6 shows in its top section how complexation energies change as more water molecules are included in the cluster. It can be observed that the complexation energies steadily become more negative as more water molecules are included as a consequence of the larger number of favorable contacts established in the cluster. The energy differences between complexes containing different aromatic molecules are roughly constant, though slightly decreasing, when including water. More interesting is the evolution of the energy change when one water molecule is included in the complex (Figure 5.6, bottom).

First, the interaction of guanidinium cation with benzene is of similar magnitude as the interaction with water, whereas for phenol and indole, the interaction with the aromatic molecule is clearly stronger than with water. Therefore, structures without direct contact between the cation and the aromatic unit should be more penalized. On the other hand, every new water molecule incorporated to the complex introduces a smaller stabilization, since the cation charge is shared by more and more units. The first and second molecules can easily be accommodated by the guanidinium cation, so the stability gain rounds 13-14 kcal mol^{-1} in all complexes. However, there is a sharp decrease when the third water molecule is included. As commented above, the third molecule has to be located in a second solvation shell or the guanidinium cation cannot interact directly with the aromatic ring. Therefore, the stability introduced by the third water molecule decreases to around 10 kcal mol^{-1} in all cases.

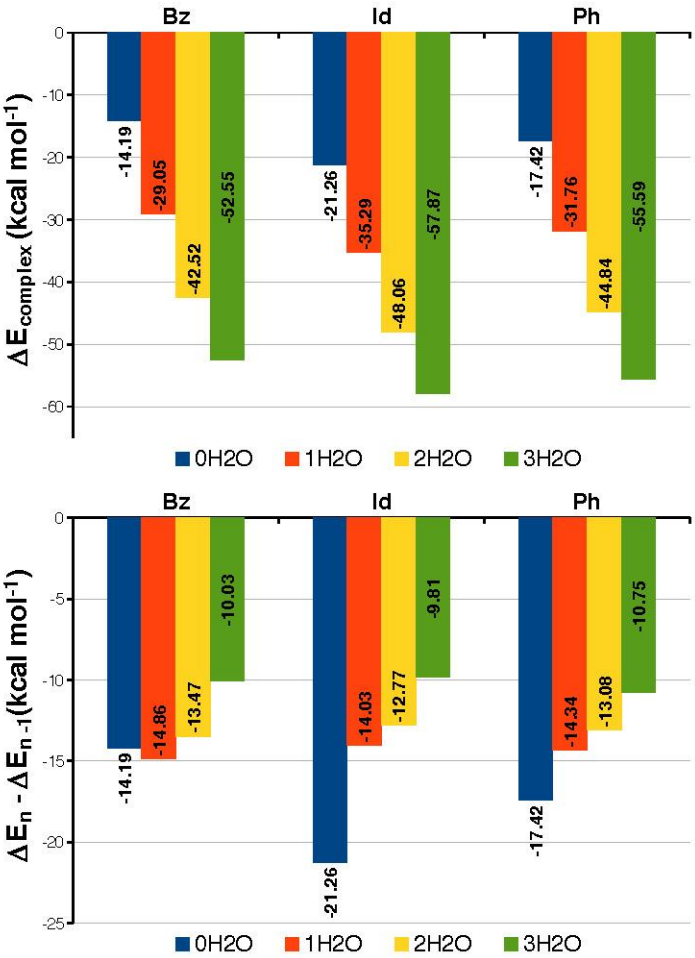


Figure 5.6. Variation of energy as water molecules are included as obtained at the MP2/aug-cc-pVDZ level of calculation. Top: complexation energies. Bottom: difference between the complex with *n* water molecules and another with *n-1* water molecules.

5.4. Conclusions

A computational study has been carried out in order to characterize the interaction between guanidinium cation and benzene, phenol and indole, as a model to know the characteristics of the cation···π interaction involving aromatic amino acid side chains. The results have been obtained at the M06-2X/6-31+G*, MP2/aug-cc-pVDZ and two empirical scaled variants of the latter, all producing similar values for the complexation energies and predicting similar behaviors for the complexes studied (except SCS-MP2 which gives underestimated complexation energies).

The most stable minima found for all non-hydrated complexes correspond to T-shaped minima with the guanidinium cation interacting with the aromatic rings by means of N-H···π contacts. Only one minimum has been located for benzene whereas other structures are possible for phenol and indole due to the participation of the hydroxyl group in phenol or the pyrrol ring in indole. In any case, with all the methods considered, the strength of the interaction grows when passing from benzene (-14 kcal mol^{-1}) to phenol (-17 kcal mol^{-1}) to indole (-21 kcal mol^{-1}).

SAPT(DFT) calculations indicate that in all these minima, as expected, the leading attractive contribution comes from electrostatics, but with significant contributions from both induction and dispersion. In fact, all three attractive interactions grow from benzene complexes to indole ones, being responsible of their larger stability. The presence of permanent dipole moments in phenol and indole produces an increase in electrostatic interaction as compared with benzene. Induction also increases, though to a lesser extent, in phenol and indole complexes because they are more polarizable than benzene. Dispersion only increases significantly in indole complexes, and especially in minima with guanidinium parallel to the aromatic ring.

The inclusion of a small number of water molecules produces a great variety of minima in all complexes, especially in phenol and indole ones. Most stable structures usually correspond to T-shaped minima with water occupying the free NH_2 groups of guanidinium cation. However, as more water molecules are included, the differences among hydrating the cation, forming a hydrogen bond with another water molecule, or forming a hydrogen bond with phenol or indole decrease, leading to a variety of minima with similar stability but displaying totally different structural patterns. The stability introduced by a new water molecule in the complex is roughly the same in the three aromatic molecules considered in this work, so the stability differences among complexes with different aromatic molecules are mainly a consequence of the different strength of the cation···π interaction.

In summary, guanidinium cation interacts more strongly with indole than with phenol or benzene. Even though the presence of a small number of water molecules does not affect significantly to the relative stability of complexes with different aromatic units (the energy differences are slightly reduced), it has a deep impact in their characteristics, allowing more possibilities to form equally stable structures with very different structural patterns. These results could help understand the behavior of cation···π contacts involving amino acid side chains, especially in environments partially exposed to solvent.

5.5. References

(1) P. Hobza, R. Zaradnik, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and the biodisciplines*, Elsevier, Amsterdam, **1988**.
(2) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.
(3) L. M. Salonen, M. Ellermann, F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-4842.
(4) M. L. Waters. *Biopolymers (Peptide Science)* **2004**, *76*, 435-445.
(5) J. C. Ma, D. A. Dougherty. *Chem. Rev.* **1997**, *97*, 1303-1324.
(6) D. A. Dougherty. *Acc. Chem. Res.* **2013**, *46*, 885-893.
(7) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-2138.
(8) D. A. Dougherty. *J. Nutr.* **2007**, *137*, 1504S-1508S.
(9) J. A. Carrazana-García, J. s. Rodríguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2011**, *115*, 2774-2782.
(10) I. Soteras, M. Orozco, F. J. Luque. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2616-2624.
(11) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.
(12) M. A. Anderson, B. Ogbay, R. Arimoto, W. Sha, O. G. Kisselev, D. P. Cistola, G. R. Marshall. *J. Am. Chem. Soc.* **2006**, *128*, 7531-7541.
(13) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2005**, *127*, 6518-6519.
(14) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 12735-12742.
(15) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 13586-13591.
(16) R. M. Hughes, M. L. Benshoff, M. L. Waters. *Chem. Eur. J.* **2007**, *13*, 5753-5764.
(17) B. W. Berry, M. M. Elvekrog, C. Tommos. *J. Am. Chem. Soc.* **2007**, *129*, 5308-5309.
(18) H. Khandelia, Y. N. Kaznessis. *J. Phys. Chem. B* **2006**, *111*, 242-250.
(19) Z. Shi, C. A. Olson, N. R. Kallenbach. *J. Am. Chem. Soc.* **2002**, *124*, 3284-3291.
(20) P. E. Mason, C. E. Dempsey, G. W. Neilson, S. R. Kline, J. W. Brady. *J. Am. Chem. Soc.* **2009**, *131*, 16689-16696.
(21) C. D. Tatko, M. L. Waters. *Protein Sci.* **2003**, *12*, 2443-2452.

(22) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.

(23) A. Rodríguez-Sanz, J. Carrazana-García, E. Cabaleiro-Lago, J. Rodríguez-Otero. *J. Mol. Model.* **2013**, *19*, 1985-1994.

(24) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.

(25) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.

(26) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.

(27) A. Campo-Cacharrón, E. Cabaleiro-Lago, J. Rodríguez-Otero. *Theor. Chem. Acc.* **2012**, *131*, 1-13.

(28) T. D. Vaden, J. M. Lisy. *J. Chem. Phys.* **2004**, *120*, 721-730.

(29) Y. Zhao, D. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215-241.

(30) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.

(31) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.

(32) E. G. Hohenstein, C. D. Sherrill. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304-326.

(33) C. D. Sherrill, *Computations of Non covalent π Interactions,* in Rev. Comput. Chem., *Vol.*, John Wiley & Sons, Inc., **2009**, pp.1-38.

(34) S. Grimme. *J. Chem. Phys.* **2003**, *118*, 9095-9102.

(35) J. G. Hill, J. A. Platts. *J. Chem. Theory Comput.* **2007**, *3*, 80-85.

(36) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.

(37) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.

(38) B. Jeziorski, R. Moszynski, K. Szalewicz. *Chem. Rev.* **1994**, *94*, 1887-1930.

(39) C. D. Sherrill. *Acc. Chem. Res.* **2013**, *46*, 1020-1028.

(40) A. J. Misquitta, K. Szalewicz. *J. Chem. Phys.* **2005**, *122*, 214109/1-214109/19.

(41) A. Heßelmann, G. Jansen, M. Schütz. *J. Chem. Phys.* **2005**, *122*, 014103/1-014103/17.

(42) H.-J. W. and, P. J. K. and, F. R. M. and, M. S. u. t. and, P. C. and, G. K. and, T. K. and, R. L. and, A. M. and, G. R. and, T. B. A. and, R. D. A. and, A. B. and, A. B. and, D. L. C. and, M. J. O. D. and, A. J. D. and, F. E. and, E. G. and, C. H. and, A. H. and, G. H. and, T. H. and, G. J. and, C. K. o. and, Y. L. and, A. W. L. and, R. A. M. and, A. J. M. and, S. J. M. and, W. M. and, M. E. M. and, A. N. and, P. P. and, K. P. u. and, R. P. and, M. R. and, T. S. and, H. S. and, A. J. S. and, R. T. and, T. T. and, M. W. and, A. Wolf, *Molpro version 2010.1, a package of ab initio programs, see http://www.molpro.net (2010).*

(43) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg,

M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. M. J. A, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. Gaussian 09, Revision B.01, Gaussian, Inc., *Wallingford, CT* **2009**.

(44) D. Kim, S. Hu, P. Tarakeshwar, K. S. Kim, J. M. Lisy. *J. Phys. Chem. A* **2003**, *107*, 1228-1238.

6

Microhydration study: complexes between pyrrolidinium cation and aromatic species

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

6.1. Introduction

Non covalent interactions are crucial in a great variety of phenomena in physics, chemistry and biochemistry.[1] The interaction among molecules not bound by chemical bonds makes possible the formation of supramolecular units which are the basis of many aspects of the promising fields of supramolecular chemistry and nanotechnology.[2, 3] In biological systems, non covalent interactions are responsible of phenomena as molecular recognition and play a crucial role in enzymatic catalysis and protein folding, among other processes as ion transport or molecular recognition.[1, 4-6]

The ubiquity of these non covalent interactions points out the need for a deep knowledge and understanding of the characteristics of the interaction at a microscopic level.[7, 8] The application of computational chemistry techniques is especially suited for this purpose, since they allow the microscopic description of the system with a detail that is often not affordable from experiment. Most especially, methods based in quantum chemistry or its approximations allow the isolation of the specific effects produced by a given interaction, and a detailed dissection of its nature.[9-11]

Among the different kinds of non covalent contacts present in biological systems, a special place is occupied by interactions involving aromatic units.[4-6, 12] When aromatic molecules participate in a non covalent interaction, it usually corresponds to one of the following three types: π···π interaction, XH···π hydrogen bond or ion···π interaction.[4, 13] In the case of proteins, the presence of aromatic species is almost guaranteed, since there are several amino acids which bear an aromatic moiety in their side chains. Therefore, the side chains of phenylalanine, tyrosine and tryptophan are usually found to be interacting with participation or their aromatic groups (benzene, phenol and indole, respectively). Quite commonly, these side chain groups interact with cationic species present in the environment which allow the formation of a stabilizing cation···π interaction. In fact, the presence of other cationic groups in the side chains of amino acids as arginine of lysine makes a quite frequent event the presence of cation···π interactions within a protein.[6, 14]

After the seminal work of Dougherty et al.,[14-18] the cation···π interaction is nowadays recognized as one of the main structural stabilizing motifs in the structure of proteins, with an important participation in a variety of recognition processes.[19] Though the cation···π interaction is strong in the gas phase as corresponds to the interaction of a bare cation with a polarizable and electron rich aromatic cloud,[13, 19] the presence of solvent molecules can significantly alter its characteristics.[19] Several studies have shown that coordination of electron-rich solvent molecules to the cation decreases the intensity of the interaction with the cation, as expected taking into account the decrease of the effective charge carried by the cation.[20-26] Also, solvent molecules

can promote significant structural changes as recently shown for the guanidinium···benzene interaction.[23] In any case, there is some controversy about the real role played by cation···π interaction in the stability of proteins, since different works have pointed out a negligible effect whereas others indicate that the contribution to the stability is very important.[27-33] These differences have been attributed to environmental effects, mostly due to the presence of the solvent.[19, 31] However, it is known that the degree of exposure to the solvent of a given cation···π interaction can vary in a wide range of values, from contacts fully exposed to the solvent where the cation···π contact behaves as immersed in bulk solvent, to situations with the cation···π buried in a hydrophobic cavity with no access for solvent molecules.[22] In this respect, the application of computational chemistry methods in systems which are gradually microhydrated can reveal hints about the effects caused by the presence of the most nearby solvent molecules and how their presence affects the characteristics of the cation···π interaction.

Most studies to date have been carried out employing benzene as model, and mostly with alkali cations.[19, 24] To our knowledge, few studies have dealt with the interaction with more complex cations as methylammonium, tetramethylammonium or guanidinium.[20-23, 26, 34, 35] The presence of a more complex cation introduces some significant differences, such as a larger structural complexity resulting in a larger variety of minima, but also on the nature of the interaction, which becomes more dispersive in nature, as shown in the case of guanidinium interaction with benzene.[23]

In the present work, a study is performed about the interaction of a structured cation with the aromatic residues of the aromatic amino acids: benzene, phenol and indole. The cation chosen for this purpose is the pyrrolidinium cation. Pyrrolidinium is an ammonium cation within a five-membered saturated ring. The pyrrolidine ring is present in a variety of natural products, most significantly as part of the structure of the amino acid proline. Also, it is part of the structure of nicotine, and in fact it is believed that nicotine addiction begins with the activation of the receptor by means of a cation···π contact with the pyrrolidinium ring of nicotine.[16, 18] Pyrrolidinium is also frequently used as constituent in a variety of ionic liquids.[36] Taking into account the ring nature of this cation it can be expected that the tendency to form stacked structures with aromatic molecules, and therefore establish contacts with a more dispersive nature, will be larger than in simpler cations. The effect of introducing several water molecules on the interaction has been analyzed in order to obtain some clues about the role water molecules play in modulating the cation···π interactions. The results obtained will help understanding the main features of these relevant interactions and the role of the closer environment upon their characteristics.

6.2. Computational details

Complexes formed by pyrrolidinium cation and one aromatic unit among benzene, phenol and indole, have been computationally studied by means of ab initio and density functional theory methods. The geometry of each complex has been optimized at the M06-2X/6-31+G* level of calculation.[37] Once a stationary point has been reached, a frequency analysis has been carried out to ensure that the structure corresponds to a minimum in the potential energy surface of the complex.

The structures found for the complexes formed by the cation and each aromatic species has been chosen as starting point for introducing water molecules in a stepwise procedure. Water molecules have been located in the most favorable regions of the cation···aromatic complex; that is, directly bonded to the N-H groups or expanding the hydrogen bond network by interacting with previous water molecules. It is worth noting that in the case of phenol and indole, the O-H and N-H groups can also participate in the hydrogen bond network, so structures with water hydrogen-bonded to phenol or indole have also been considered.

Once a minimum has been found, the complexation energy has been obtained by applying the supermolecule method, using the counterpoise procedure to avoid any basis set superposition error.[38, 39] Therefore, interaction energies are obtained as:

$$\Delta E_{\text{int}} = \Delta E^{complex}(ijk...) - \sum_i \Delta E^{complex}(ijk...) \qquad \text{(eq. 6.1)}$$

where terms in parentheses indicate the basis set and superscripts the geometry employed in the calculations. As the geometry of the molecules changes when the cluster is formed, an additional contribution describing this effect must be included, obtained as the energy difference between the molecules in the cluster geometry and in isolation.

$$E_{def} = \sum_i \left(E_i^{complex}(i) - E_i^{isolated}(i) \right) \qquad \text{(eq. 6.2)}$$

$$\Delta E_{complex} = \Delta E_{int} + \Delta E_{def} \qquad \text{(eq. 6.3)}$$

Since no reference values for the complexation energies are available for these systems, high-level calculations have been carried out for non-hydrated complexes. Thus, energies have been estimated at the CCSD(T) complete basis set limit (CBS) employing an extrapolation of the MP2 correlation energy obtained with the aug-cc-pVDZ and aug-cc-pVTZ basis sets. Following a

widely used extrapolation procedure,[10, 40] the contribution of the correlation energy to the complexation energy is obtained as:

$$\Delta E_{corr,MP2}^{CBS} = \frac{X^3}{X^3-(X-1)^3}\Delta E_{corr,MP2}^{AVXZ} - \frac{(X-1)^3}{X^3-(X-1)^3}\Delta E_{corr,MP2}^{AV(X-1)Z}; X = 3 \quad \text{(eq. 6.4)}$$

and assuming HF values already converged to the basis limit, the MP2 complexation energy to basis limit is:

$$\Delta E_{MP2}^{CBS} = \Delta E_{HF}^{AVTZ} + \Delta E_{corr,MP2}^{CBS}. \quad \text{(eq. 6.5)}$$

Finally, the deficiencies on the MP2 method are corrected by using CCSD(T)/AVDZ calculations:

$$\Delta E_{CCSD(T)}^{CBS} = \Delta E_{MP2}^{CBS} + \left(\Delta E_{CCSD(T)}^{AVDZ} - \Delta E_{MP2}^{AVDZ} \right). \quad \text{(eq. 6.6)}$$

The results obtained with this extrapolation procedure have been employed to assess the performance of cheaper methods that will be employed in the hydrated systems.

Besides the M06-2X/6-31+G* already employed in geometry optimizations, several empirically scaled variants of MP2 have been also used for obtaining complexation energies; namely SCS-MP2 and SCSN-MP2.[41] In these methods the contributions to correlation energies from same-spin and different-spin electron pairs are empirically scaled in order to improve the performance of the native MP2. In SCS-MP2 the scaling factors are 1.20 and 0.33 for opposite-spin and same-spin components,[41] whereas in SCSN-MP2 the factors are 0.00 and 1.76.[42]

Finally, a new variant proposed by Hobza in order to obtain high quality results avoiding the CCSD(T) costly calculations has also been considered.[43] In this MP2.X method the extrapolation to basis limit is performed as commented above, but the CCSD(T) correction is substituted by a MP3 correction adequately scaled in order to reproduce the CCSD(T) values. In the present work MP3/6-31G* level has been employed for this purpose with an scaling factor of 0.86; that is:

$$E_{MP2.X}^{CBS} = E_{MP2}^{CBS} + 0.86\left(E_{MP3}^{6-31G^*} - E_{MP2}^{6-31G^*} \right). \quad \text{(eq. 6.7)}$$

Also, in order to obtain more information about the characteristics of the interaction, the interaction energies for non-hydrated complexes have been decomposed by applying the SAPT method with intramonomer correlation effects described at the DFT level (SAPT(DFT)).[44-46] So, using the geometries optimized at the M06-2X/6-31+G* level, density fitted DFT-SAPT calculations were carried out, providing information on the individual physical components of the interaction energy. For these calculations the LPBE0AC exchange-correlation was used, involving a shift parameter obtained as the sum of the ionization potential and the energy of the highest occupied molecular orbital. Orbital energies and ionization potentials have been obtained by using the PBE0 functional with the aug-cc-pVDZ basis set. The DFT-SAPT calculations were performed with the aug-cc-pVDZ basis set, employing the cc-pVTZ/JKFIT for Hartree–Fock,and aug-cc-pVDZ/MP2FIT for the second-order dispersion terms. All SAPT and CCSD(T) calculations have been performed with Molpro.[47]

M06-2X and MP3 calculations have been done with Gaussian09,[48] whereas MP2 calculations have been performed with Turbomole 6.3.[49] In order to save computational time in MP2 calculations, the resolution of the identity approach has been employed both for the HF and correlation energies. That is; RI-JK-MP2 calculations have been performed using the aug-cc-pVXZ auxiliary basis set for correlation and the def2-TZVPP auxiliary basis set for both coulomb and exchange in the calculation of HF energies.[50, 51]

6.3. Results

6.3.1. Pyrrolidinium···π complexes

In this section the results obtained for complexes formed by pyrrolidinium and benzene, phenol or indole in the absence of water molecules will be presented. Figure 6.1 shows the minimum energy structures found for the complexes formed by pyrrolidinium and the aromatic species considered in this study, whereas Table 6.1 lists the values obtained for the complexation energies of these structures as obtained with different computational methods.

Two very similar structures have been found for pyrrolidinium···benzene complex PB (the nomenclature will indicate P for pyrrolidinium, B, P or I for benzene, phenol or indole, and a numerical identifier indicating the structure). As expected, in both minima the interaction takes place by means of hydrogen bond contacts of the N-H groups of the cation with the aromatic cloud of benzene. However, whereas in PB-1 there is only one NH···π contact at about 2.0 Å, in PB-2 the pyrrolidinium cation adopts a more vertical orientation with respect to benzene, with two N-H···π contacts, though one of them is pretty long (2.8 Å).

In the case of phenol complexes the behavior is much more complicated since phenol presents two favorable positions for interacting with the cation, the hydroxyl oxygen and the aromatic ring. Besides, pyrrolidinium can interact with one of these positions by means of the ammonium group while simultaneously establishes favorable contacts trough the positively charged C-H groups of the carbon skeleton. Therefore, up to six different minima have been located for pyrrolidinium···phenol (PP) complexes, as shown in Figure 6.1.

These six minima are grouped by pairs of similar structures which differ on whether N-H groups interact with the oxygen or the phenyl ring. Thus, in minima PP-5 and PP-6 only one contact is registered by means of a N-H···X interaction, which is oriented towards the hydroxyl oxygen in PP-5 and towards the phenyl ring in PP-6. The second N-H unit is oriented in order to interact favorably with the free OH or phenyl ring, but the distance is quite large, about 3.3-3.6 Å. In the rest of minima two simultaneous contacts with phenol are observed, one with the N-H group and the other by means of one of the C-H groups of the cation. In minima PP-1 and PP-2 the C-H group is next to nitrogen, whereas in PP-3 and PP-4 is two bonds apart. On the other hand, in PP-1 and PP-3 there are N-H···O contacts whereas in PP-2 and PP-4 there are N-H···π ones. The distances of the N-H···O hydrogen bonds are shorter than 1.9 Å, reaching values over 2 Å for N-H···π ones. The distances are usually larger for contacts involving C-H units, as expected.

The behavior is similar in indole complexes. Indole also presents two favorable positions for interacting with cations corresponding to the two rings. Therefore, minima have been found showing N-H···π contacts to the phenyl or pyrrol rings of indole. Structures showing only one N-H···X contact as in phenol have been tried, but they collapse to PI-5, with pyrrolidinium located roughly over the bridge between the two rings and establishing two N-H···π contacts with both rings. The distance to the phenyl ring if much shorter (2.1 Å) than to the pyrrol ring (2.7 Å).

Table 6.1 summarizes the complexation energies for the complexes shown in Figure 6.1 as obtained with a variety of methods. Skipping by the moment method performance and considering only CCSD(T)/CBS values, it can be observed that both minima located for PB present complexation energies of -16 kcal mol^{-1} since the energy surface is very shallow regarding rotations of the N-H groups over the phenyl ring. The interaction is slightly stronger in phenol complexes, reaching -17.5 kcal mol^{-1} in the most stable structure corresponding to PP-3. In this structure there is a N-H···O hydrogen bond whereas a C-H group two bonds apart participates in a contact to the π cloud. Rotating the pyrrolidinium cation as to establish a N-H···π plus a C-H···O contact is only penalized by 0.6 kcal mol^{-1} (PP-4), showing that both the hydroxyl groups and the aromatic ring are equally capable of interacting favorably with cations, as observed in previous work.[25, 26]

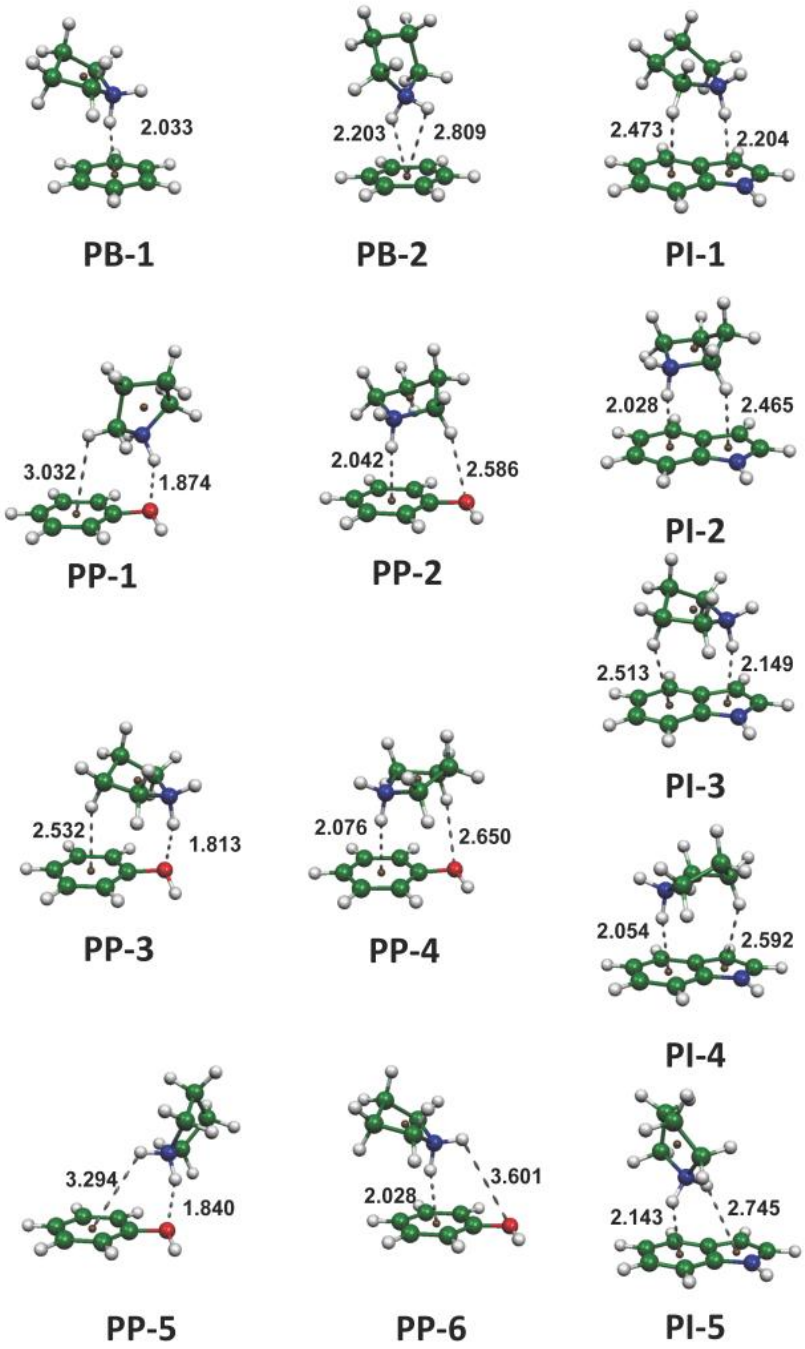


Figure 6.1. Minimum energy structures obtained for the complexes formed by pyrrolidinium cation with benzene, phenol and indole, as obtained at the M06-2X/6-31+G* level of calculation. Distances in Å.

This is confirmed by the energies obtained for PP-5 and PP-6, minima with only one N-H···X contact and no participation of the carbon skeleton in the interaction. The complexation energies are identical and similar to those obtained in benzene complexes. In any case, all minima with phenol are comprised within an energy range of only 1.3 kcal mol^{-1}.

Indole complexes behave in a similar way but in this case there is a significant increment on the complexation energies which reach more than -20 kcal mol^{-1} in all cases. This is a consequence of the larger polarizability and dispersion interaction in indole. In indole complexes the interaction with the phenyl ring is clearly favored. Thus, both minima PI-2 and PI-4, with the N-H group pointing to the phenyl ring whereas a C-H group interacts with the pyrrol ring are almost isoenergetic, reaching around -21.8 kcal mol^{-1}. Minima PI-5, where the interaction takes place only by means of the two N-H groups, is only 0.5 kcal mol^{-1} less stable. However, the interaction of the N-H units with the pyrrol ring as in P I-1 and PI-3 is penalized with 1.1-1.3 kcal mol^{-1}. Again, as in phenol clusters all minima are comprised in a narrow energy range of around 1 kcal mol^{-1}, showing the versatility of pyrrolidinium for interacting with species which offer two sites for cation interaction.

As regards the different methods tested, several trends are worth commenting. Values obtained at the CCSD(T)/CBS level as described in computational details have been considered as reference in order to check the performance of the other methods. Also included are values obtained by using the cc-pVDZ and the hcc-pVDZ (diffuse functions on heavy atoms) basis sets on the CCSD(T) calculations in order to estimate the effect of a better correlation treatment (eq. 6.6). It can be observed that removing the diffuse functions on hydrogen atoms hardly affects the quality of the results, so it could be an interesting option in order to save computation time in larger systems. On the other hand employing the small cc-pVDZ basis set gives worse results, though still close to the reference ones. The mean absolute error (MAE) increases from 0.04 kcal mol^{-1} to 0.14 kcal mol^{-1} when no diffuse functions are used. On the other hand, a less costly method as MP2.X provides results of similar quality with a MAE of 0.10 kcal mol^{-1}. In any case, MP2.X is still quite costly so other options have been analyzed as shown in Table 6.1.

First, it is worth noting the impressive performance of M06-2X/6-31+G*, providing very good results with a MAE of 0.16 kcal mol^{-1} at a much reduced computational cost. Among the different MP2 variants considered, it can be appreciated that MP2 performance deteriorates as the basis set increases, giving worse results when extrapolated to basis limit (MAE=1.71 kcal mol^{-1}) than with the AVDZ basis set (MAE=0.57 kcal mol^{-1}). SCS-MP2 improves when extrapolated but still showing significant deviations. In the case of the SCSN-MP2 method, it shows an interesting MAE of 0.28 kcal mol^{-1} with the AVDZ basis set which slightly deteriorates to 0.44 when extrapolated to basis limit.

Table 6.1. Complexation energies (kcal mol^{-1}) for pyrrolidinium complexes with benzene, phenol and indole as obtained with different methods.

	M06-2X	MP2[a]						MP2.X[b]	CCSD(T)		
	6-31+G*	AVDZ	SCS/AVDZ	SCSN/AVDZ	CBS	SCS/CBS	SCSN/CBS	CBS	CBSC[c]	CBSB[d]	CBS
PB-1	-15.83	-16.53	-14.33	-15.84	-17.67	-15.34	-16.50	-16.29	-15.86	-15.96	-15.99
PB-2	-15.79	-16.47	-14.46	-15.97	-17.53	-15.40	-16.54	-16.34	-15.85	-15.96	-15.99
PP-1	-16.92	-16.37	-14.61	-15.80	-17.31	-15.43	-16.29	-16.38	-16.26	-16.37	-16.41
PP-2	-17.00	-17.44	-15.00	-16.73	-18.71	-16.12	-17.49	-17.05	-16.86	-16.97	-17.00
PP-3	-17.79	-17.71	-15.38	-16.89	-18.86	-16.40	-17.49	-17.47	-17.37	-17.43	-17.49
PP-4	-16.91	-17.31	-14.72	-16.48	-18.63	-15.89	-17.27	-16.88	-16.71	-16.82	-16.86
PP-5	-16.30	-16.01	-14.54	-15.61	-16.82	-15.25	-16.01	-16.10	-16.01	-16.14	-16.18
PP-6	-16.23	-16.90	-14.62	-16.27	-18.07	-15.66	-16.99	-16.50	-16.26	-16.38	-16.40
PI-1	-20.56	-21.46	-18.55	-20.40	-22.65	-19.60	-21.06	-20.62	-20.43	-20.51	-20.56
PI-2	-22.09	-22.82	-19.70	-21.70	-24.19	-20.90	-22.49	-21.91	-21.74	-21.83	-21.87
PI-3	-20.61	-21.51	-18.41	-20.29	-22.81	-19.56	-21.00	-20.63	-20.45	-20.50	-20.56
PI-4	-21.70	-22.63	-19.47	-21.52	-24.01	-20.69	-22.30	-21.75	-21.56	-21.64	-21.69
PI-5	-21.06	-22.12	-19.59	-21.42	-23.24	-20.56	-22.02	-21.48	-21.16	-21.28	-21.30
MAE[e]	0.16	0.57	1.92	0.28	1.71	0.89	0.44	0.10	0.14	0.04	

[a] Different versions of MP2, AVDZ and CBS stand for aug-cc-pVDZ basis set and extrapolated to complete basis set limit. SCS and SCSN are empirically scaled MP2 methods.

[b] With the 6-31G* basis set.

[c] With the xx-pVDZ basis set in eq. 6.6.

[d] With the haug-cc-pVDZ basis set (diffuse function only in heavy atoms) in eq. 6.6.

[e] Mean absolute error.

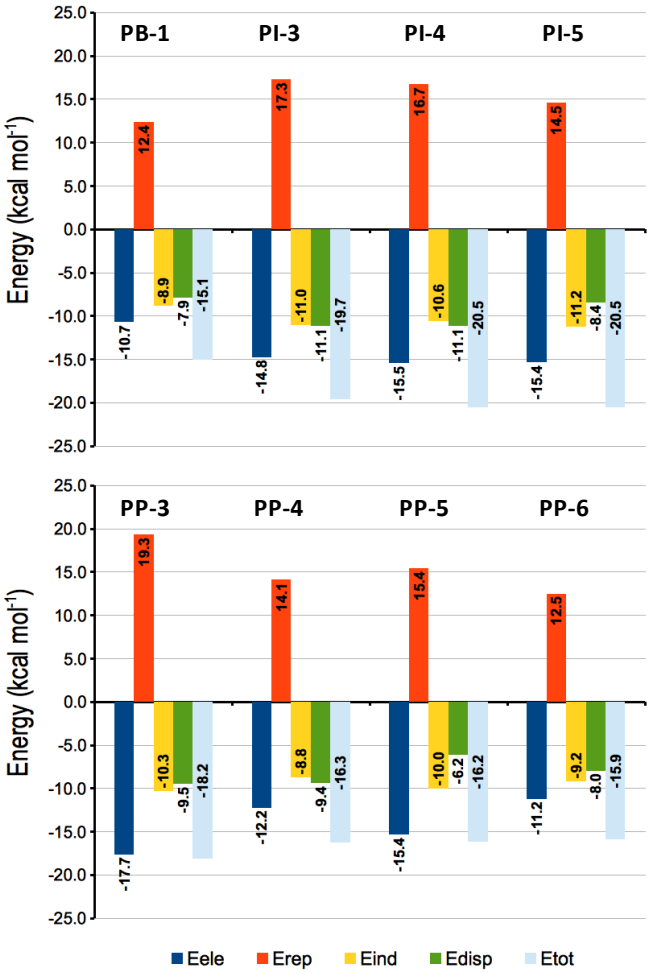


Figure 6.2. SAPT(DFT) energy decomposition for non-hydrated complexes formed by pyrrolidinium cation with benzene, phenol and indole.

Therefore, considering the different methods shown in Table 6.1, the choices would be the M06-2X/6-31+G* and the SCSN-MP2/AVDZ levels of calculation which give the best balance between performance and computational cost. Other options should be discarded except in the case of benchmark calculations which could be done with CCSD(T)/hAVDZ of even MP2.X.

In order to gain more insight in the origin of the interaction, SAPT(DFT) calculations have been carried out for the complexes shown in Figure 6.1, as explained in Computational Details. Figure 6.2 summarizes the results obtained for selected complexes (all results can be consulted in Appendix 3). Benzene complexes are mainly bound by the electrostatic contribution as expected for these cation···π systems. However, there are significant contributions from induction and also dispersion. The electrostatic and repulsion terms almost cancel each other, the total interaction energy almost matching the combination of induction and dispersion. These results are similar to those found for guanidinium····benzene complexes though dispersion is slightly larger.[23, 24]

In the case of phenol complexes, the interaction is stronger. In all complexes, the electrostatic contribution is larger than in benzene as a consequence of the presence of the hydroxyl group. Note however that in minimum PP-6, the energy partitioning almost matches perfectly that observed in benzene, since in this minimum the pyrrolidinium cation interacts only with the aromatic ring by means of a N-H···π contact. When the interaction is established by means of a N-H···O contact as in PP-5, the electrostatic contribution is larger, as also is the induction component (these two contributions usually go paired). However, since the pyrrolidinium cation is not oriented towards the ring the dispersion contribution is the smallest observed in phenol complexes. In minima PP-3 and PP-4 the behavior is slightly different as a consequence of the double C-H···X and N-H···X contacts. PP-3 is the most stable minima as a consequence of the largest electrostatic contribution (though the balance with repulsion is positive), together with a large induction contribution. In this structure, since both rings of phenol and pyrrolidinium are stacked, the dispersion contribution is larger, reaching -10 kcal mol^{-1}. When the N-H group contacts with the ring, as in PP-4, the interaction weakens as a consequence of smaller electrostatic and induction contributions, whereas dispersion remains almost constant since both molecules are stacked. In the case of indole complexes a similar pattern arises though, as expected, dispersion contributions are the largest among the complexes studied in this work, reaching -11 kcal mol^{-1}. Comparing the results for indole and phenol complexes it becomes clear that the stronger interaction observed with indole is a consequence of the larger dispersion term since from the electrostatic or inductive point of view the contributions in both complexes are similar. The only minimum with a slightly different behavior is PI-5, since it does not present stacked molecules. However, the loss in dispersion is partially compensated by a larger amount of induction and smaller repulsion leading to a similar global stabilization than PI-4.

Therefore, even though the stability of different complexes can be similar, the interaction obeys to a delicate balance of the different contributions. Though the electrostatic contribution is the most important one, the play among induction, dispersion and also repulsion is determinant for the global stability of the complex.

6.3.2. Pyrrolidinium···π···water complexes

In this section, the characteristics of complexes containing water molecules will be presented and discussed. It has been observed that the first water molecule in this kind of cluster can establish specific interactions with the aromatic unit, thus leading to a specially enhanced strength of the interaction, so a more detailed analysis has been performed for monohydrated complexes.

Figure 6.3 shows the structures of complexes found including one water molecule, whereas complexation energies are shown in Table 6.2. In cases where similar minima with slightly different energies have been found, only one representative structure is shown (these similar structures usually correspond to different orientations of the water molecule relative to pyrrolidinium cation). Starting with the minima obtained for pyrrolidinium···benzene complexes the water molecule has been introduced in possible favorable positions leading to the two minima shown in Figure 6.3. The first minimum PB-1W-1 corresponds to the expected structure with the water molecule occupying the free N-H group of the cation, with a complexation energy of -29.7 kcal mol^{-1} at the MP2.X level of calculation (the highest quality method employed in these systems). The other minimum PB-1W-2, with water interacting with the ring of course presents smaller stability amounting to about -25.3 kcal mol^{-1}.

Among the different minima found for phenol complexes, four different patterns can be identified. The most stable structures correspond to minima where a hydrogen bond network is developed leading to a cyclic arrangement with participation of the aromatic ring and the hydroxyl oxygen acting as hydrogen acceptor. These structures PP-1W-1 and PP-1W-2 present complexation energies amounting to -31.0 and -33.6 kcal mol^{-1} respectively, favoring the contact of the water molecule with the hydroxyl oxygen over the O-H···π contact. This has to be a consequence of the greater strength of O-H···O interactions with respect to O-H···π ones. Another pattern corresponds to a stacked complex between pyrrolidinium and phenol, where water molecule occupies the free NH group as in benzene complex (PP-1W-3 and PP-1W-4). Also, minima with water intercalated between pyrrolidinium and phenol are found (PP-1W-5 and PP-1W-6). Finally, structures were located where a stacked pyrrolidinium···phenol complex interacts with water by means of the hydroxyl group of phenol acting as donor (PP-1W-7 and PP-1W-8). All these structures present complexation energies rounding -30 kcal mol^{-1}.

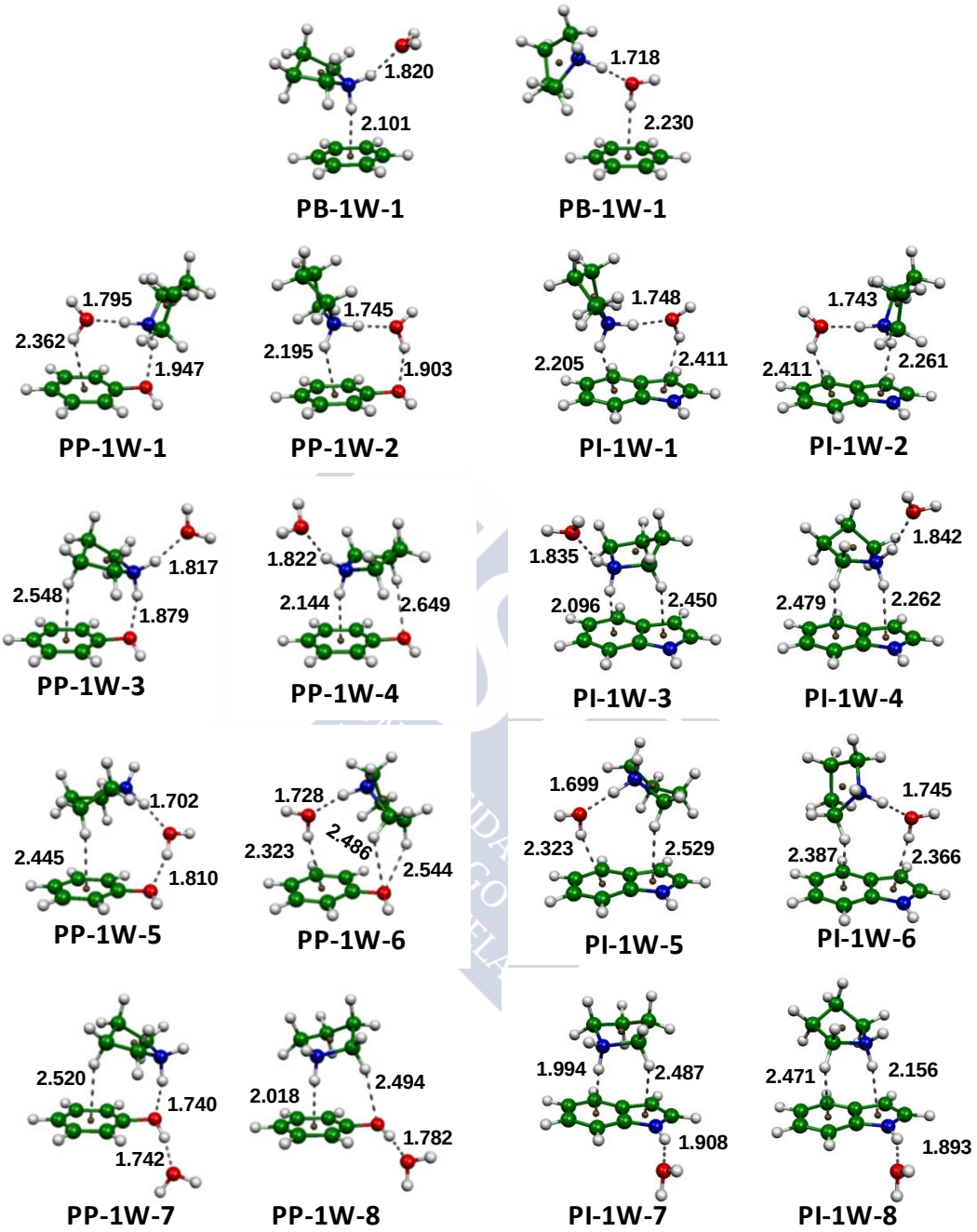


Figure 6.3. Selected minima for complexes formed by pyrrolidinium with benzene, phenol, indole and one water molecule. Distances in Å.

A pretty similar behavior is observed in indole complexes. Again, the most stable structures present a cyclic pattern reaching -36.4 and -34.3 kcal mol^{-1} (PI-1W-1 and PI-1W-2), favoring water contact with the pyrrol ring. Stacked structures with water on the second N-H group reach -33.6 and -34.9 kcal mol^{-1} (PI-1W-3 and PI-1W-4). The rest of structures with water between both rings or bonded to the N-H group of indole are less stable, with complexation energies around -31 kcal mol^{-1}. Therefore, among the different structural arrangements found, it becomes clear that for aromatic units with two favorable interaction sites, the most favored minima correspond to structures with a cyclic hydrogen bond pattern: π(phenyl)···H-N-H···O-H···O in phenol and π(phenyl)···H-N-H···O-H···π(pyrrol) in indole. These structures are not formed from the most stable non-hydrated structures, which correspond to stacked minima leading to the second most stable structures in monohydrated complexes. Therefore, the presence of the water molecule affects the structural arrangement of the cation···π contact in order to facilitate the interaction of water with the aromatic unit.

As regards method performance, assuming that MP2.X still reproduces adequately the CCSD(T)/CBS results, similar trends are observed as in the case of non-hydrated complexes. However, whereas the performance of SCSN-MP2/AVDZ is still good with a MAE of 0.37 kcal mol^{-1} for the values listed in Table 6.2, M06-2X deteriorates reaching a value of 1.58 kcal mol^{-1}. Taking into account the excellent performance in non-hydrated complexes, this behavior has to be related with problems reproducing the interaction between pyrrolidinium and water. In any case, M06-2X always overestimates complexation energies by 1-2 kcal mol^{-1} (3-7 % of the total complexation energy), the relative energies not being specially affected, so for clusters with more water molecules only M06-2X/6-31+G* and SCSN-MP2 values will be presented.

Table 6.3 lists the pair energy decomposition of the complexes with one water molecule, allowing a better description of the interactions established within the cluster. Complexation energies for each pair of units have been obtained employing the geometry and the whole basis set of the complex, the three-body contribution being obtained as the difference between the interaction energy of the trimer and the sum of all the pair interaction energies. As regards benzene complexes it can be observed that the most stable minimum PB-1W-1 shows two strong interactions of pyrrolidinium cation with benzene and water of similar strength amounting roughly to -16 kcal mol^{-1}. The interaction between water and benzene, as expected, is almost zero (0.4 kcal mol^{-1}) as a result of the large distance between both units. In minimum PB-1W-2 the pyrrolidinium···water interaction remains with the same characteristics but now the cation interacts more weakly with benzene reaching only -6 kcal mol^{-1}. This lost is partially compensated by the interaction of water with benzene, contributing -2 kcal mol^{-1} by means of an O-H···π contact.

Table 6.2. Complexation energies (kcal mol^{-1}) for pyrrolidinium complexes with benzene, phenol and indole including one water molecule as obtained with different methods.

	M06-2X	**MP2**[a]				**MP2.X**[b]
	6-31+G*	**AVDZ**	**SCSN/AVDZ**	**CBS**	**SCS/CBS**	**CBS**
PB-1W-1	-30.62	-29.61	-29.24	-31.50	-30.40	-29.69
PB-1W-2	-26.69	-24.76	-24.59	-26.76	-25.83	-25.28
PP-1W-1	-33.18	-30.43	-30.25	-32.55	-31.55	-30.97
PP-1W-2	-35.43	-33.32	-33.14	-35.59	-34.57	-33.56
PP-1W-3	-32.67	-30.73	-30.29	-32.62	-31.46	-30.85
PP-1W-4	-31.90	-30.43	-29.92	-32.47	-31.24	-30.33
PP-1W-5	-32.39	-30.30	-30.04	-32.51	-31.47	-30.63
PP-1W-6	-29.78	-27.48	-27.24	-29.64	-28.61	-27.92
PP-1W-7	-31.98	-30.01	-29.54	-32.12	-30.80	-29.95
PP-1W-8	-30.28	-28.68	-28.36	-30.81	-29.68	-28.50
PI-1W-1	-37.67	-36.52	-36.09	-38.85	-37.56	-36.42
PI-1W-2	-35.84	-34.45	-33.82	-36.73	-35.21	-34.33
PI-1W-3	-36.47	-35.46	-34.67	-37.55	-35.99	-34.91
PI-1W-4	-35.08	-34.09	-33.32	-36.01	-34.52	-33.61
PI-1W-5	-32.49	-31.02	-30.43	-33.25	-31.82	-31.21
PI-1W-6	-32.72	-31.03	-30.45	-33.15	-31.77	-31.01
PI-1W-7	-33.02	-32.59	-31.70	-34.52	-32.80	-31.58
PI-1W-8	-31.51	-31.36	-30.55	-33.14	-31.57	-30.48
MAE[c]	1.58	0.43	0.37	1.96	0.69	

[a]Different versions of MP2. AVDZ and CBS stand for aug-cc-pVDZ stand for aug-cc-pVDZ basis set and extrapolated to complete basis set limit. SCS and SCSN are empirically scaled MP2 methods.
[b]With the 6-31G* basis set.
[c]Mean absolute error.

Table 6.3. Pair energy contributions (kcal mol^{-1}) to monohydrated complexes as obtained with SCN-MP2/AVDZ.

	Pyrr···π	**Pyrr···H_2O**	**H_2O···π**	**3-Body**
PB-1W-1	-15.80	-15.90	0.37	1.60
PB-1W-2	-6.25	-15.66	-1.88	-1.81
PP-1W-1	-16.36	-15.55	-1.68	1.23
PP-1W-2	-16.69	-15.97	-2.70	0.79
PP-1W-3	-18.70	-15.96	0.45	1.69
PP-1W-4	-17.11	-15.88	0.39	1.62
PP-1W-5	-10.51	-16.14	-3.46	-1.43
PP-1W-6	-9.76	-15.76	-2.08	-1.07
PP-1W-7	-18.82	-5.20	-6.25	-2.05
PP-1W-8	-17.21	-3.78	-6.61	-1.93
PI-1W-1	-20.86	-15.71	-2.03	1.40
PI-1W-2	-18.65	-15.39	-2.18	1.06
PI-1W-3	-21.92	-15.85	0.52	1.94
PI-1W-4	-20.79	-15.78	0.43	1.98
PI-1W-5	-12.15	-15.70	-2.35	-1.51
PI-1W-6	-12.43	-15.62	-2.43	-1.29
PI-1W-7	-22.13	-3.50	-5.49	-1.41
PI-1W-8	-21.01	-3.37	-5.49	-1.57

Phenol clusters are more complex in account of the larger number of different minima. In the most stable structure PP-1W-2 the three pair interactions contribute to the stability of the complex, again with two strong cation interactions of similar magnitude plus a weak water···phenol interaction of -2.7 kcal mol^{-1}. The stronger interaction of the cation with phenol and the contribution of water···phenol contact make this structure more stable than the complexes with benzene. In PP-1W-1 the behavior is similar, the only difference due to the O-H···π contact instead of an O-H···O one. This reflects on a water···phenol interaction of -1.7 kcal mol^{-1} which combined with the slightly larger repulsive three-body term is the origin of the stability difference between these structures. PP-1W-3 and PP-1W-4 behave similarly to benzene complexes with large interactions of water and phenol with the cation and residual water···phenol interactions. The interaction of the cation with phenol is stronger than in benzene and independent of the orientation of the N-H group to the ring or to the hydroxyl oxygen. PP-1W-5 and PP-1W-6 are similar to benzene second most stable structure. However as phenol offers two locations for favorable interactions, the cation interacts more strongly than with benzene, as also does the water molecule. Finally PP-1W-7 and PP-1W-8 are less stable because the water molecule interacts by means of the hydroxyl group of phenol. Therefore, there is a strong cation···phenol interaction and a weak cation···water contact. However, the O-H···O hydrogen bond contributes with -6 kcal mol^{-1} to the stability of the complex.

Indole complexes show a similar behavior. All structures except PI-1W-7 and PI-1W-8 show strong cation···water interactions rounding -15 kcal mol^{-1}, whereas the interaction of the cation with the indole unit amounts to more than -20 kcal mol^{-1} in the most stable complexes, excepting PI-1W-5 and PI-1W-6. Also, the interaction of water with indole amounts to -2 kcal mol^{-1} in the most stable complexes whereas the N-H···O interaction contributes with -5.5 kcal mol^{-1} in PI-1W-7 and PI-1W-8. Overall, three-body effects are moderate in all the complexes considered, roughly spanning between -2 to 2 kcal mol^{-1}. Stabilizing contributions are associated to structures where the water molecule is intercalated between the cation and the aromatic ring or hydrogen bonded to the OH and NH groups of phenol and indole, respectively.

Figure 6.4 shows the most stable minima found for complexes containing 2 and 3 water molecules. Already in the most stable complexes with one water molecule, the N-H groups of pyrrolidinium cation participate in contacts with water or the aromatic molecule. Therefore, the second water molecule is incorporated to the complex by establishing hydrogen bonds with the previous water molecule, the N-H or O-H groups in indole and phenol, or displacing the aromatic molecule from the N-H···π contact.

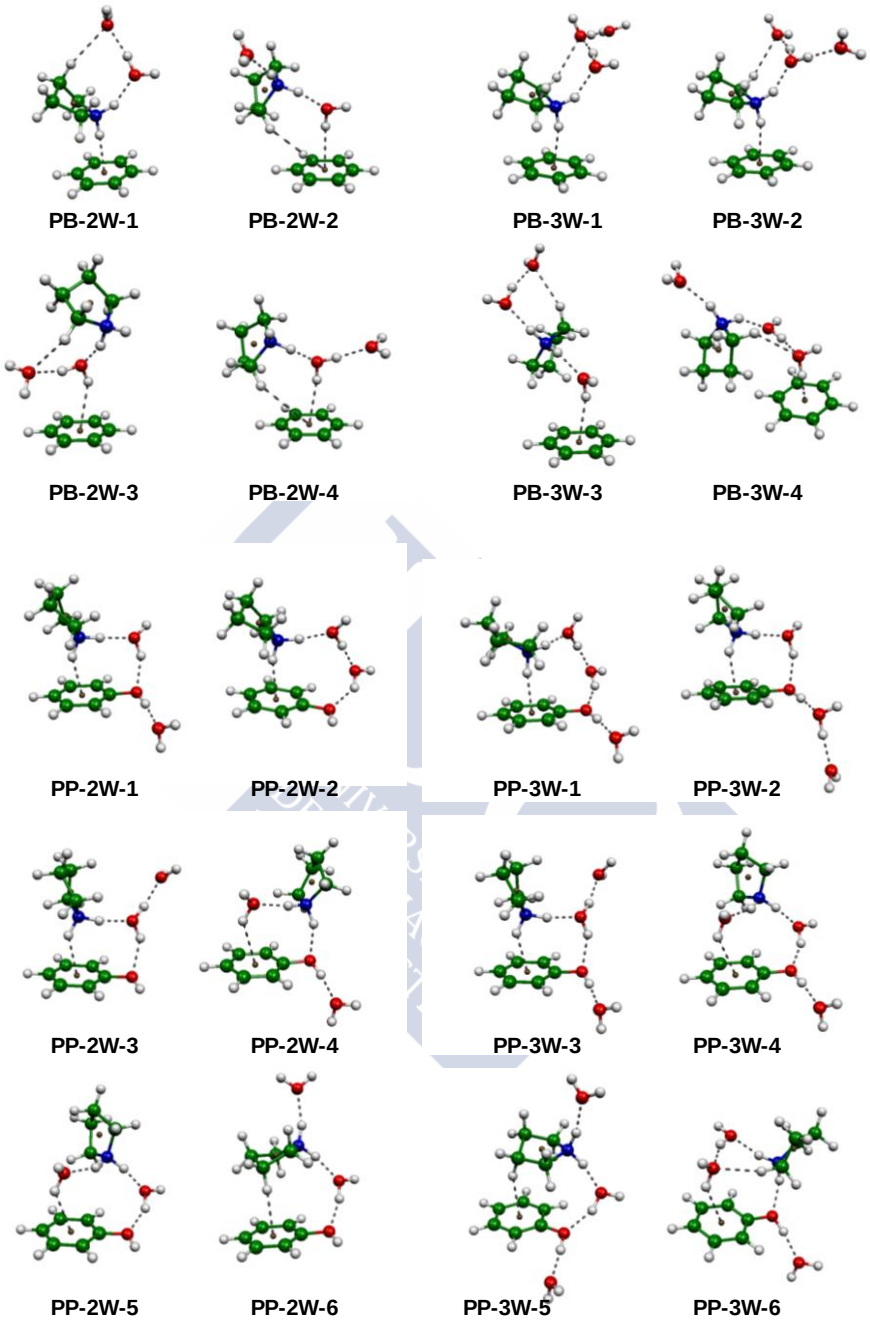


Figure 6.4a. Selected most stable minima for complexes formed including two and three water molecules.

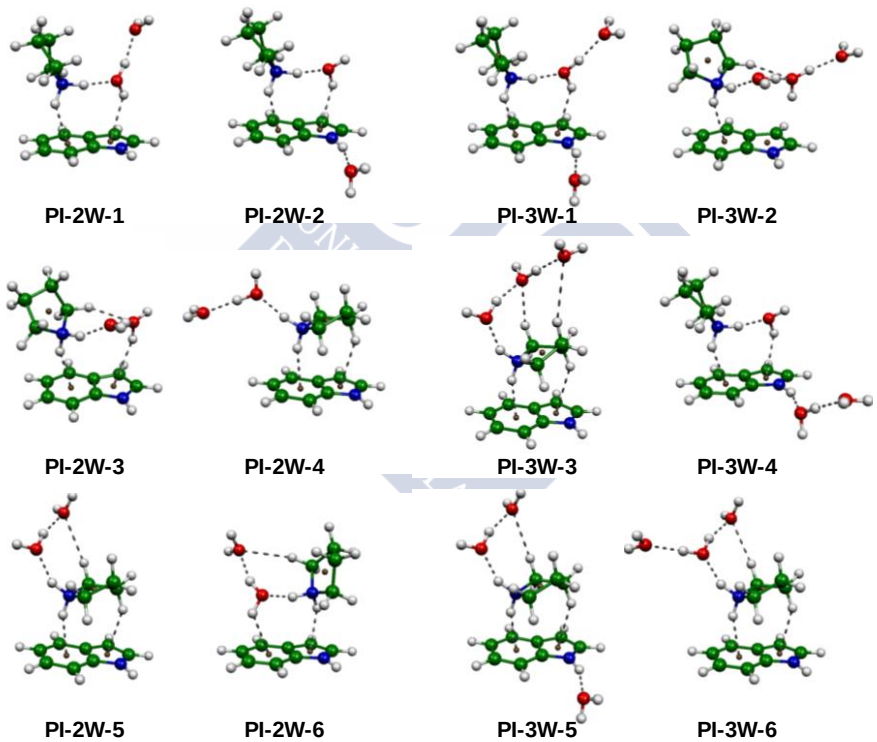


Figure 6.4b. Selected most stable minima for complexes formed including two and three water molecules.

These possibilities give rise to a great variety of minima, often with similar complexation energies. Those shown in Figure 6.4 are the most stable ones as calculated with the SCSN-MP2/aug-cc-pVDZ level. Table 6.4 lists the values for complexation energies in these complexes.

As in previous sections, the behavior of benzene complexes is the simplest, since it only establishes stable interactions by means of the aromatic cloud. Therefore, in the most stable minimum PB-2W-1, with complexation energy of -40.3 kcal mol^{-1}, the cation interacts directly with the benzene cloud, whereas water molecules hydrate pyrrolidinium cation. That is, starting from PB-1W-1, the second water molecule forms a hydrogen bond with the previous one while establishes secondary contacts with the CH groups of the cation. The rest of the minima are more than 3 kcal mol^{-1} less stable, because water has to be intercalated between the cation and the benzene ring, forming O-H$\cdots\pi$ hydrogen bonds. The same behavior is observed when the third water molecule is included. In the most stable structures PB-3W-1 and PB-3W-2, with complexation energies rounding -50 kcal mol^{-1}, the water molecule is added to the hydrogen bond network and does not interact directly with the cation.

As expected, phenol complexes show a larger variety of structures, due to participation of the hydroxyl group as part of the hydrogen bond network. Usually, the most stable structures show a cyclic arrangement as commented above, as in PP-2W-1 and PP-2W-3. However, other possibilities for hydrogen bonding also appear with similar stability, as in PP-2W-2, where the second water molecule incorporates to a chainlike arrangement of O-H$\cdots$O hydrogen bonds. All these minima show similar complexation energies which are comprised in barely 1 kcal mol^{-1}, reaching -46 kcal mol^{-1} for the most stable ones. It is worth noting that several of these minima show participation of the hydroxyl group of phenol, as already observed in similar systems.[25, 26] Also, structures appear where the cation does not interact directly with the aromatic molecule, as in PP-2W-5. In this case, the aromatic molecule interacts with the water molecules already bound to the cation. Complexes with three water molecules show a similar behavior, with a variety of minima with similar complexation energies differing in the structures of the hydrogen bond network, and reaching -55 kcal mol^{-1}. In any case, there seems that in most stable clusters, the three water molecules and the hydroxyl group of phenol tend to be aligned in a chainlike series of hydrogen bonds. Overall, the results indicate that already with two and especially three water molecules, the stability of the complex is not largely affected by whether a water molecule forms a hydrogen bond directly with the cation, with any of the previous water molecules, or with the hydroxyl group of phenol. All these possibilities produce minima with similar stability, thus showing the complexity of the potential surface of these clusters.

Table 6.4. Complexation energies (kcal mol^{-1}) for pyrrolidinium complexes containing two and three water molecules.

	2 H_2O			3 H_2O	
	M06-2X 6-31+G*	**SCSN-MP2 AVDZ**		**M06-2X 6-31+G***	**SCSN-MP2 AVDZ**
PB-2W-1	-43.41	-40.30	**PB-3W-1**	-55.39	-50.11
PB-2W-2	-41.02	-37.33	**PB-3W-2**	-54.42	-49.65
PB-2W-3	-41.02	-37.29	**PB-3W-3**	-53.77	-47.99
PB-2W-4	-38.59	-35.43	**PB-3W-4**	-53.20	-46.60
PP-2W-1	-48.96	-46.27	**PP-3W-1**	-59.25	-55.61
PP-2W-2	-48.43	-45.98	**PP-3W-2**	-59.50	-55.10
PP-2W-3	-48.75	-45.74	**PP-3W-3**	-60.32	-54.88
PP-2W-4	-48.24	-45.62	**PP-3W-4**	-58.59	-54.81
PP-2W-5	-48.81	-45.52	**PP-3W-5**	-59.01	-54.70
PP-2W-6	-47.56	-44.23	**PP-3W-6**	-59.13	-54.67
PI-2W-1	-49.07	-45.52	**PI-3W-1**	-60.58	-55.59
PI-2W-2	-47.18	-43.58	**PI-3W-2**	-60.42	-55.35
PI-2W-3	-47.09	-43.37	**PI-3W-3**	-60.03	-54.94
PI-2W-4	-47.37	-43.21	**PI-3W-4**	-60.99	-54.80
PI-2W-5	-47.84	-42.93	**PI-3W-5**	-59.51	-53.78
PI-2W-6	-46.40	-42.63	**PI-3W-6**	-59.86	-53.28

Indole complexes behave similarly. With two water molecules the most stable minimum presents a complexation energy of -45.5 kcal mol^{-1} and a cyclic hydrogen bond pattern π(phenyl)···H-N-H···O-H···π(pyrrol), and the second water molecule is hydrogen bonded to the previous one. The rest of minima in Figure 643 show complexation energies around -43 kcal mol^{-1}, differing in the hydrogen bond arrangement. With three water molecules the complexation energy reaches -55 kcal mol^{-1} in several structures. As in the case of phenol complexes, several of the most stable minima show participation of the N-H group of indole in the hydrogen bond network.

Table 6.5. Complexation energies (kcal mol^{-1}) for most stable complexes as water molecules are included.

	M06-2X/6-31+G*			SCSN-MP2/AVDZ		
	Benzene	**Phenol**	**Indole**	**Benzene**	**Phenol**	**Indole**
0H_2O	-15.83	-17.79	-22.09	-15.97	-16.89	-21.70
1H_2O	-30.62	-35.43	-37.67	-29.24	-33.14	-36.09
2H_2O	-43.41	-48.96	-49.07	-40.30	-46.27	-45.52
3H_2O	-55.39	-60.32	-60.99	-50.11	-55.61	-55.59

Table 6.5 summarizes the results obtained for the most stable complexes found at the different microhydration levels. As commented above, without water molecules, indole complexes are more stable than phenol ones (around 6 kcal mol^{-1}) and these more than benzene (1 kcal mol^{-1}). When the first water molecule is included, the energy gap increases relative to benzene, so indole monohydrated complex becomes more stable by 3 kcal mol^{-1} than phenol, and phenol by 4 kcal mol^{-1} to benzene. This is a consequence of the formation of the cyclic hydrogen bond network. The presence of the hydroxyl group in phenol and the pyrrol ring in indole allows for this water molecule to establish O-H···X favorable contacts which cannot be formed with benzene. By including more water molecules the stability difference between phenol and benzene complexes stabilizes around 5-6 kcal mol^{-1}, whereas the stability difference between phenol and indole comes to zero. Therefore, even when pyrrolidinium establishes the strongest interaction with indole, the formation of more favorable hydrogen bonds by participation of the hydroxyl group of phenol results in similarly stable complexes for both phenol and indole, once a couple of water molecules are present.

6.4. Conclusions

The interaction between pyrrolidinium cation and aromatic species present in the side chain of amino acids has been studied with a variety of computational methods. Several minima have been found for any of the aromatic molecules considered. Benzene shows the simplest behavior since the interaction can only take place by means of the aromatic cloud. In phenol and indole complexes, however, the presence of the two aromatic rings in indole and the hydroxyl group in phenol makes possible the formation of a larger variety of structures. The most stable complexes

show pyrrolidinium interacting with one of its N-H groups with the phenyl aromatic rings or, in the case of phenol, with the hydroxyl oxygen. Besides, the cation tends to adopt a parallel orientation with respect to phenol and indole in order to establish favorable interactions by means of its CH groups with the pyrrol ring or the phenyl ring in indole and phenol, respectively. The results indicate that pyrrolidinium interacts more strongly with indole (-21.9 kcal mol^{-1}) than with phenol (-17.5 kcal mol^{-1}) and benzene (-16.1 kcal mol^{-1}). The different computational methods tested provide similar results in most cases, but the results indicate that the best balance between accuracy and computational effort is provided by the M06-2X/6-31+G* (especially in non-hydrated clusters) and the SCSN-MP2/AVDZ levels of calculation.

SAPT(DFT) results indicate that the interaction is mainly controlled by electrostatics, though both induction and dispersion also contribute significantly to the stability of the complexes. The interaction with phenol is stronger than with benzene mainly as a consequence of larger electrostatic contributions. In indole complexes, all stabilizing terms contribute to a larger extent to the interaction energy than in benzene ones. Besides the increase in electrostatics and induction, the most striking effect is the increase in dispersion as a consequence of the larger size of indole.

Including water molecules produces a great variety of minima, in many cases showing similar complexation energies. Whereas for benzene complexes the pattern is pretty simple, with water molecules hydrogen bonded among themselves and to one of the N-H groups of pyrrolidinium, the behavior is more complex in phenol and indole clusters. A cyclic pattern of hydrogen bonds involving the two favorable sites of phenol or indole is often found in the most stable structures. It is worth noting that in several of the minima the hydroxyl group of phenol and the N-H group of indole participate as hydrogen donors in the hydrogen bond network. In any case, including more water molecules gives rise to a large variety of minima with similar interaction energies, showing that different hydrogen bond patterns are equally stable. Also, the inclusion of water molecules removes the stability differences between phenol and indole clusters, which already become equally stable with two water molecules, though around 5 kcal mol^{-1} more stable than benzene complexes. In summary, the presence of a small number of water molecules deeply affects the characteristics of these complexes, promoting important structural and energetic changes that could help understanding the behavior of these important cation$\cdots\pi$ interactions.

6.5. References

(1) P. Hobza, R. Zaradnik, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and the biodisciplines*, Elsevier, Amsterdam, **1988**.

(2) H.-J. Schneider, A. K. Yatsimirski, *Principles and methods in supramolecular chemistry*, Wiley, Chichester, **2000**.

(3) J.-M. Lehn, *Supramolecular chemistry: concepts and perspectives*, VCH, Weinheim, **1995**.

(4) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.

(5) K. E. Riley, P. Hobza. *Acc. Chem. Res.* **2013**, *46*, 927-936.

(6) M. L. Waters. *Biopolymers (Peptide Science)* **2004**, *76*, 435-445.

(7) I. Kaplan, *Intermolecular Interactions: Intermolecular Interactions: Physical Picture , Computational Methods*, John Wiley & Sons, Chichester, **2006**.

(8) A. J. Stone, *The theory of intermolecular forces*, Oxford University Press, Oxford, **2013**.

(9) K. E. Riley, P. Hobza. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3-17.

(10) E. G. Hohenstein, C. D. Sherrill. *WIREs Comput. Mol. Sci.* **2012**, *2*, 304-326.

(11) C. D. Sherrill. *Rev. Comput. Chem.* **2009**, *26*, 1-38.

(12) L. M. Salonen, M. Ellermann, F. F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-4842.

(13) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.

(14) J. P. Gallivan, D. A. Dougherty. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459-9464.

(15) J. C. Ma, D. A. Dougherty. *Chem. Rev.* **1997**, *97*, 1303-1324.

(16) N. Zacharias, D. A. Dougherty. *Trends Pharmacol. Sci.* **2002**, *23*, 281-287.

(17) D. A. Dougherty. *J. Nutr.* **2007**, *137*, 1504S-1508S.

(18) D. A. Dougherty. *Acc. Chem. Res.* **2013**, *46*, 885-893.

(19) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-2138.

(20) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.

(21) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.

(22) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, 870-874.

(23) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.

(24) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.

(25) A. Campo-Cacharrón, E. Cabaleiro-Lago, J. Rodríguez-Otero. *Theor. Chem. Acc.* **2012**, *131*, 1-13.

(26) A. A. Rodriguez-Sanz, J. Carrazana-Garcia, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *J. Mol. Model.* **2013**, *19*, 1985-1994.

(27) R. M. Hughes, M. L. Benshoff, M. L. Waters. *Chem. Eur. J.* **2007**, *13*, 5753-5764.
(28) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 13586-13591.
(29) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2005**, *127*, 6518-6519.
(30) R. M. Hughes, M. L. Waters. *J. Am. Chem. Soc.* **2006**, *128*, 12735-12742.
(31) B. W. Berry, M. M. Elvekrog, C. Tommos. *J. Am. Chem. Soc.* **2007**, *129*, 5308-5309.
(32) P. E. Mason, J. W. Brady, G. W. Neilson, C. E. Dempsey. *Biophys. J.* **2007**, *93*, L04-L06.
(33) P. E. Mason, C. E. Dempsey, G. W. Neilson, S. R. Kline, J. W. Brady. *J. Am. Chem. Soc.* **2009**, *131*, 16689-16696.
(34) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.
(35) F. Leavens, C. D. Churchill, S. Wang, S. D. Wetmore. *J. Phys. Chem. B* **2011**, *115*, 10990-11003.
(36) T. L. Greaves, C. J. Drummond. *Chem. Rev.* **2007**, *108*, 206-237.
(37) Y. Zhao, D. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
(38) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.
(39) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.
(40) T. Helgaker, W. Klopper, H. Koch, J. Noga. *J. Chem. Phys.* **1997**, *106*, 9639-9646.
(41) S. Grimme. *J. Chem. Phys.* **2003**, *118*, 9095-9102.
(42) J. G. Hill, J. A. Platts. *J. Chem. Theory Comput.* **2007**, *3*, 80-85.
(43) K. E. Riley, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2011**, *13*, 21121-21125.
(44) B. Jeziorski, R. Moszynski, K. Szalewicz. *Chem. Rev.* **1994**, *94*, 1887-1930.
(45) A. Hesselmann, G. Jansen, M. Schütz. *J. Chem. Phys.* **2005**, *122*, 014103/1-014103/17.
(46) A. J. Misquitta, K. Szalewicz. *J. Chem. Phys.* **2005**, *122*, 214109/1-214109/19.
(47) H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Kö, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, D. Peng, K. P. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang. *Molpro version 2010.1, a package of ab initio programs, see http://www.molpro.net (2010).*
(48) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. J. A. Montgomery, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A.

Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, J. A. Pople, Gaussian 09, Revision B.01, Gaussian, Inc., *Wallingford, CT* **2009**.

(49) *TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.*

(50) F. Weigend, A. Köhn, C. Hättig. *J. Chem. Phys.* **2002**, *116*, 3175-3183.

(51) F. Weigend. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285-4291.

7

Cation···π interactions between guanidinium and aromatic amino acids

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

7.1. Introduction

Non covalent intermolecular interactions play a crucial role in order to understand the structural characteristics and reactivity in biological systems.[1-3] Among the different kinds of intermolecular interactions established in proteins and other natural systems, the contacts involving aromatic units occupy an outstanding place.[4-8] It is believed that aromatic units introduce unique characteristics which make them differ from aliphatic interactions, and therefore can lead to specificity on structure and function.[4-6, 9]

If an aromatic molecule or fragment participates in a non covalent interaction, it most often establishes one or more of the following types of contacts: π···π interaction, X-H···π interaction or ion··π interaction.[4, 9] π···π stacking interactions are known to play a significant role in many fields, and more specifically, in protein folding and stability.[4, 5, 8] On the other hand, X-H···π interactions represent hydrogen bonds formed between O-H, N-H and C-H groups with the aromatic rings.[4, 6, 10] Even though both π···π and X-H···π interactions are usually weak interactions contributing just 0.5 kcal mol^{-1}, the simultaneous formation of a large number of these contacts can deeply affect the characteristics of the system.[4, 6, 10]

On the other hand, the interaction between a polarizable π cloud and a charged species is a strong interaction in the gas phase.[7, 11-13] Environmental effects such as solvent[7, 14-19] or other interacting groups[20, 21] can affect this kind of contacts significantly, though it has been shown that even in solvent their contribution can be larger than ion pairs of salt bridges.[22] Despite the recent interest aroused by the interaction between an anion and an electron-deficient aromatic molecule,[12, 23-26] in the so-called anion···π interaction, most contacts observed in biological systems correspond to cation···π interactions.[4, 7, 11, 22, 27, 28] Many works have been devoted to study the characteristics of the cation···π interaction, which is dominated by large electrostatic interactions together with important polarization contributions, which can be as large, or even larger than electrostatics.[7, 9, 29-31]

Focusing in proteins, it becomes clear that the presence of amino acids such as phenylalanine, tyrosine and tryptophan, carrying an aromatic unit in their side chain, makes them amenable for establishing cation···π interactions.[4, 5, 7, 8, 27, 28] This fact, together with the presence of other amino acids having a side chain that can be cationic depending on the pH of the medium, reveals the important role this kind of interaction can play in protein stability and structure. Several studies have already shown that cation···π interactions are widespread in proteins,[32] even in their inner parts, though different degrees of exposure of cation···π contacts have been observed.[7, 22, 27]

Many studies have been devoted to determine the characteristics of cation···π contacts in natural systems and to measure their intensity. Most often, complexes formed by aromatic amino acids and alkali cations have been considered, both experimentally and theoretically,[33-41] though other studies deal with alkaline-earth cations or transition metals as copper or aluminium.[36, 42-46] As a result of these works, the intensity of the interaction between the series of alkali cations and the aromatic amino acids has been determined, indicating that it increases going from phenylalanine to tyrosine and to tryptophan, though the energy differences are not large.[33, 34, 38] Also, Wu and McMahon have shown that the interaction of sodium and ammonium cation exhibits a noticeable enhancement in aromatic amino acids as compared to non-aromatic ones, thus pointing out the relevant role of the aromatic side chain in the stabilization of the complex.[40]

In any case, a cation···π contact frequently found in proteins has been scarcely considered, corresponding to the contact between arginine cationic side chain and the aromatic amino acids. In fact, the most common cation···π contact between cationic and aromatic amino acids corresponds to the interaction of guanidinium cation of arginine with tryptophan.[27] Guanidinium interaction with aromatic species has been already studied, showing that guanidinium cation interacts preferentially in a T-shaped structure with the NH_2 groups pointing to the aromatic rings.[16, 29, 47, 48] Previous studies have shown that solvation plays an important role and dramatically modulates the strength of the interaction.[14-16, 37, 42, 48-51] In fact, the presence of a small number of water molecules can change the orientation of guanidinium cation with respect to benzene, so the parallel orientation will be more favorable.[16, 29] This effect, together with other contacts with electron-donating groups nearby and other environmental effects can explain the preference for parallel structures observed in crystals.[20] Also, results indicate that guanidinium interacts more strongly with indole than with phenol and benzene, in agreement with the size effects pointed out by Sastry.[52]

However, most studies of this kind have been performed employing the aromatic residue on the side chain, thus avoiding the treatment of the full amino acid. In such a way the complex search on the conformational space of the amino acid can be avoided. At least in the gas phase, the presence of the carboxylic and amino groups of the neutral amino acid also act as electron donating groups and therefore can interact with the cation, thus competing with the aromatic cloud. Several studies with simple cations show that in fact the most stable minima of amino acid···cation complexes correspond to structures with the cation interacting with both the carboxylic and amino group, and the aromatic unit.[33, 34, 38, 40, 43, 45] However, a more complex cation as guanidinium can introduce particular effects which can result in a different behavior. Guanidinium interacts with electron-rich groups by forming hydrogen bonds with the NH_2 groups. Also, the planar structure of guanidinium makes it more prone to form stacked structures where dispersion plays a major role, so the behavior of guanidinium···amino acid complex could

be different from that observed with alkali cations. Therefore, in the present work, a detailed study of the interaction of guanidinium cation with phenylalanine, tyrosine and tryptophan (in neutral and zwitterionic form) is performed by using different computational techniques. The intensity of the interaction is measured in each case, and a thorough analysis is carried out in order to point out the factors controlling the interaction. This is to our knowledge the first time aromatic amino acid interactions with a complex cation have been studied in detail, quantifying the role played by the different contributions to the interaction. Though an approximate picture of the interaction could be qualitatively obtained by using approximate methods, quantitative results obtained with reliable methods are needed for grasping the subtleties controlling these interactions, which constitute the bricks upon which biomolecular interactions are constructed. The results obtained would help understanding the nature and characteristics of contacts involving arginine and aromatic amino acids, revealing the key factors controlling complex structure and stability.

7.2. Computational details

In order to guarantee a proper selection of starting geometries for optimization, the following procedure was employed. First, a Multiple Minimum Monte Carlo conformational search has been carried out employing the MMFF force field with the MacroModel program.[53] All structures with relative energies below 10 kcal mol^{-1} with respect to the most stable one have been retained for subsequent refinement. This procedure has been carried out for isolated phenylalanine, tyrosine and tryptophan and for their complexes with guanidinium cation. Both the neutral and zwitterionic forms of the amino acids have been considered in separate searches. The set of structures resulting from these conformational searches have been optimized employing the M06-2X/6-31+G* level of calculation,[54] which showed good performance in previous work.[51] Once a stationary point has been located, a frequency calculation has been performed in order to ensure that it corresponds to a minimum in the potential energy surface of the system.

The interaction energies of the dimers were obtained by applying the supermolecule method, using the counterpoise procedure to avoid basis set superposition error.[55, 56] Therefore, interaction energies are obtained as:

$$\Delta E_{int} = E_{AB}^{complex}(AB) - E_{A}^{complex}(AB) - E_{B}^{complex}(AB)\,. \qquad \text{(eq. 7.1)}$$

where terms in parentheses indicate the basis set and superscripts the geometry employed in the calculation. The energy associated to changes in the geometry of the fragments forming the dimer during complex formation has been included as a deformation energy term obtained as:

$$E_{def} = E^{com}(A) + E_B^{com}(B) - E^{iso}(A) - E_B^{iso}(B). \qquad \text{(eq. 7.2)}$$

Therefore, the final complexation energy is given by:

$$\Delta E_{comp} = \Delta E_{int} + E_{def} . \qquad \text{(eq. 7.3)}$$

Besides obtaining complexation energies with M06-2X, a dispersion-corrected B3LYP functional has been employed. Thus, an empirical dispersion term is added to the standard B3LYP functional as suggested by Grimme.[57-60] As recommended by this author, the newest D3 dispersion correction together with a Becke-Johnson (BJ) damping function has been employed.[59, 60] Also, the B2PLYP functional, including a mixture of DFT and the MP2 correlation has been used.[61] These calculations have been carried out employing the cc-pVTZ basis set.

MP2 single point calculations have been performed for the minima located employing the aug-cc-pVDZ and the aug-cc-pVTZ basis set. These values have been employed in order to estimate the MP2 complexation energy at the complete basis set limit (CBS) following an extrapolation procedure.[62]

$$\Delta E_{corr,MP2}^{CBS} = \frac{X^3}{X^3-(X-1)^3}\Delta E_{corr,MP2}^{AVXZ} - \frac{(X-1)^3}{X^3-(X-1)^3}\Delta E_{corr,MP2}^{AV(X-1)Z}; X = 3 \qquad \text{(eq. 7.4)}$$

The MP2 complexation energy to basis limit is then estimated as:

$$\Delta E_{MP2}^{CBS} = \Delta E_{HF}^{AVTZ} + \Delta E_{corr,MP2}^{CBS}. \qquad \text{(eq. 7.5)}$$

Two empirical variants of MP2 have also been considered; namely SCS-MP2 and SCSN-MP2.[63, 64] In these methods the contributions to correlation energies from same-spin and different-spin electron pairs are empirically scaled in order to improve the performance of the native MP2. In SCS-MP2 the scaling factors are 1.20 and 0.33 for opposite-spin and same-spin components,[63] whereas in SCSN-MP2 the factors are 0.00 and 1.76.[64] In both cases, extrapolation to the basis set limit has been performed.

With the extrapolation procedure it is assumed that errors due to basis set incompleteness are taking care of. However, there still remains the error associated to the method employed for

electron correlation. In order to correct for this error and to obtain benchmark values in these systems, the complexation energies have been estimated at the CCSD(T)/CBS level of calculation following a procedure often employed in literature. The MP2/CBS values are corrected with the difference between CCSD(T) and MP2 results with a moderate-size basis set.[2, 65, 66] In the present case, the CCSD(T)/CBS complexation energy is obtained as:

$$\Delta E^{CBS}_{CCSD(T)} = \Delta E^{CBS}_{MP2} + \left(\Delta E^{aVDZ}_{CCSD(T)} - \Delta E^{aVDZ}_{MP2} \right). \qquad \text{(eq. 7.6)}$$

Another, cheaper possibility for correcting the deficiencies on the N-electron model comes from using MP3 results obtained with a small basis set and adequately scaled in order to reproduce CCSD(T) values as proposed by Hobza in the MP2.X method:[67]

$$\Delta E^{CBS}_{MP2.X} = \Delta E^{CBS}_{MP2} + 0.86 \left(\Delta E^{6-31G^*}_{MP3} - \Delta E^{6-31G^*}_{MP2} \right). \qquad \text{(eq. 7.7)}$$

In order to obtain more information about the characteristics of the interaction, the interaction energies have been decomposed by applying the Symmetry Adapted Perturbation Theory (SAPT) method with intramonomer correlation effects described at the DFT level (SAPT(DFT)).[68] Density fitted DFT-SAPT calculations were carried out, providing information on the individual physical components of the interaction energy.[69, 70] For these calculations the PBE0 functional was used, involving a shift parameter obtained as the sum of the ionization potential and the energy of the highest occupied molecular orbital. Orbital energies and ionization potentials have been obtained by using the PBE0 functional with the aug-cc-pVDZ basis set. The DFT-SAPT calculations were performed with the aug-cc-pVDZ basis set, employing the cc-pVTZ/JKFIT for Hartree–Fock and aug-cc-pVDZ/MP2FIT for the second-order dispersion terms. In order to provide a more accurate description, the dispersion contribution has been scaled by a factor of 1.193 as suggested by Hobza.[71]

SAPT calculations have been performed with Molpro 2010.1.[72] M06-2X and MP3 calculations have been done with Gaussian09,[73] whereas MP2 and CCSD(T) calculations have been performed with Turbomole 6.3.[74] B3LYP-D and B2PLYP calculations have been done with Orca 3.0.1,[75, 76] taking advantage of the RIJCOSX approach.[77] Also, the resolution of the identity approach has been employed in MP2 calculations both for the HF and correlation energies. That is; RI-JK-MP2 calculations have been performed using the aug-cc-pVXZ auxiliary basis set for correlation and the def2-TZVPP auxiliary basis set for both coulomb and exchange in the calculation of HF energies.[78, 79]

7.3. Results

7.3.1. Isolated amino acids

Since the conformational flexibility of amino acids is essential for protein folding, conformational studies of amino acids in the gas phase by means of spectroscopic techniques have frequently been the subject of researchers. Aromatic amino acids have been studied by means of theoretical and experimental methods, as well as combination of both approaches, in order to clarify their structures. As a consequence, sets of stable conformers have been identified for phenylalanine,[80-86] tyrosine,[87-93] and tryptophan[93-98] usually within a quite narrow energy gap. Even though the goal of the present work is the study of the cation···π interaction in amino acids, a conformational search has been carried out for the isolated amino acids with the same procedure employed in the study of the dimers in order to obtain their most stable structures. A large number of minima has been located, Figure 7.1 showing a selection comprising the most stable ones together with their relative energies as obtained at the M06-2X/6-31+G* and MP2/CBS levels of calculation. In the case of tyrosine the structures are grouped by pairs differing on the orientation of the hydroxyl group of phenol, and the most stable one of the pair is shown in Figure 7.1.[87-89] Also, zwitterionic stable structures are not found for any of the amino acids in the gas phase, in agreement with previous results.

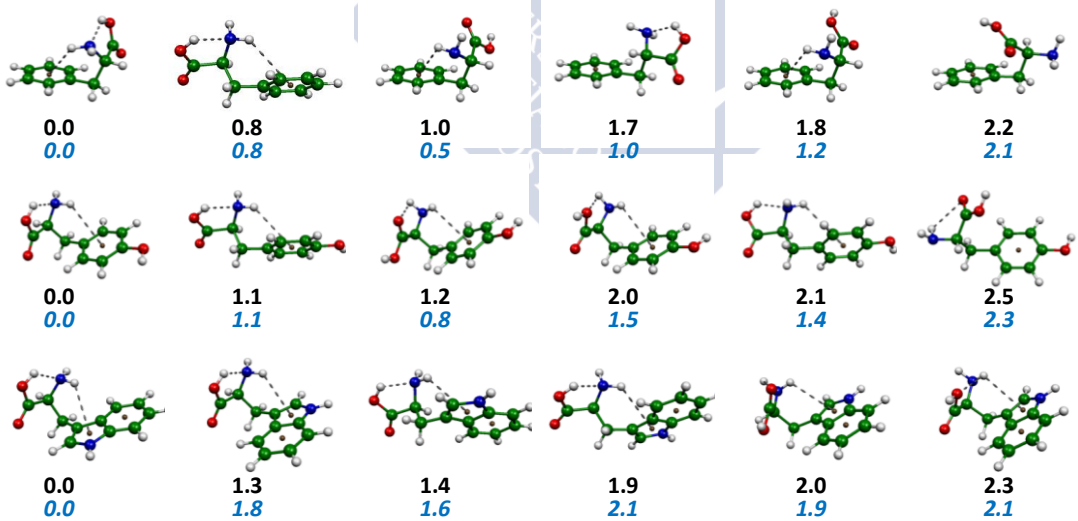


Figure 7.1. Selected most stable minima found for the isolated amino acids. Numbers correspond to relative energies in kcal mol^{-1}. M06-2X/6-31+G* values in blue italics; MP2/CBS ones in black.

The six minima shown for each amino acid in Figure 7.1 are comprised within an energy window around 2.5 kcal mol^{-1} (inclusion of zero point energy (ZPE) or entropy does not change this picture significantly, see Appendix 4). In all cases, the most stable conformation obtained for the three amino acids corresponds to a structure showing an intramolecular O-H···N hydrogen bond, whereas a secondary contact between one of the amino hydrogen atoms and the aromatic cloud is also formed. Other conformers correspond to different arrangements of the hydrogen bond contacts or to a different set of torsion angles affecting to the amino acid group. In fact, the second most stable minima are similar to the most stable one, but the structure is more extended, with the carboxylic group further away from the aromatic rings. This arrangement involves an energy penalty of 0.8-1.3 kcal mol^{-1}, the largest gap being observed for tryptophan. The results obtained are similar overall to those shown by other authors in their extensive conformational studies of aromatic amino acids,[83, 85, 87-89, 96] the minima in Figure 7.1 corresponding mostly to conformers identified in previous work.

In any case, as commented above, the purpose of this work is not to perform an analysis of amino acid conformers, and the most important information to be taken is the energy of the isolated amino acid. This quantity is crucial in the calculation of deformation energies which, as it will be shown below, can largely affect to the energetics of these systems. Therefore, since all amino acids exhibit the same most stable structure and both methods of calculation lead to similar results, these structures will be employed for the calculation of the deformation energies of the complexes.

7.3.2. Phenylalanine-guanidinium complexes

Figure 7.2 shows the most stable structures obtained for the complexes formed by phenylalanine and guanidinium cation. A selection has been performed and only the most stable minima are shown, though other less stable structures have been located. The minima are shown in order of decreasing stability within the type of structure, so zwitterionic structures are shown last.

First, it can be observed that only in two structures a contact between the cation and the aromatic phenyl ring is formed. These contacts are established at a N-H···X (to the center) distance of around 2.44 Å in Phe-A and 2.77 Å in Phe-B.The main difference between these two structures Phe-A and Phe-B, the two most stable ones among neutral amino acid complexes, relies on the different contact established by guanidinium with the amino and acid groups of phenylalanine.

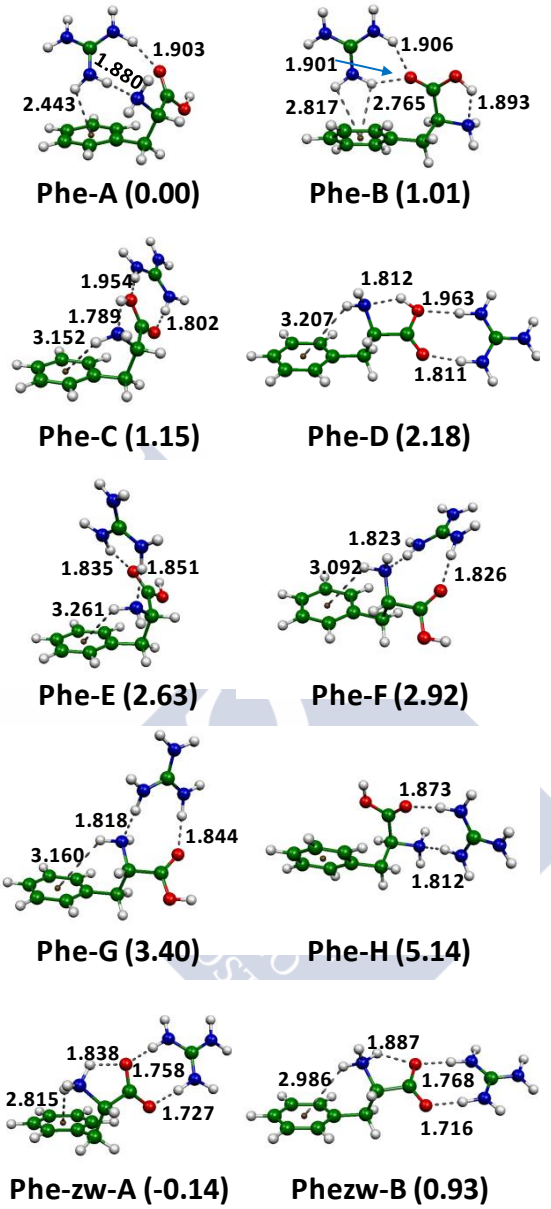


Figure 7.2. Selected most stable minima found for complexes formed by guanidinium cation and phenylalanine as obtained at the M06-2X/6-31+G* level of calculation. Selected distances in Å. Relative energies (kcal mol^{-1}) at the CCSD(T)/CBS level in parentheses.

Thus, in Phe-A, guanidinium forms two hydrogen bonds with the carboxyl oxygen and the nitrogen lone pair, whereas in Phe-B both hydrogen bonds are formed with the carboxyl oxygen thus reducing the strength of the interaction, though an intramolecular O-H···N hydrogen bond is also present. The distances of these contacts are pretty short as corresponds to strong hydrogen bonds formed by a cation and an electron rich group, being around 1.9 Å in all cases.

Overall, several aspects have to be taken into account in order to understand the stability of these complexes, as it will be detailed below. First, the number of favorable contacts between the cation and the electron-rich regions of the amino acid has to be considered. As commented above, the most stable structures exhibit contacts with three of these regions. On the other hand, in order to establish one or several of these hydrogen bonds, the geometry of the amino acid has to change in many cases. In structures such as Phe-A, Phe-E or Phe-F, the guanidinium cation forms a hydrogen bond with the NH_2 group of phenylalanine. To do so, the intramolecular O-H···N hydrogen bond observed in the most stable minima of the isolated amino acids has to be broken, introducing an energy penalization in order to form these structures. Finally, changes on the conformation of the amino acid are also observed in order to allow favorable contacts. The rotations around the distinct dihedral angles, departing from the most stable structure of the amino acid are also a penalization in the stability of these complexes. This is important in folded structures as Phe-A or Phe-B, being less relevant in more extended minima such as Phe-D. In summary, the final stability will be a balance of these different factors, but taking into account the structures in Figure 7.2, there seems that simultaneous formation of hydrogen bonds with both the C=O and NH_2 groups is favored over other kinds of hydrogen bond patterns, though double hydrogen bond to the two oxygen atoms of the carboxylic group is competitive since it allows keeping the N-H···O intramolecular hydrogen bond. Finally, a couple of minima have been found with the amino acid in zwitterionic form. In these structures, guanidinium strongly interacts with the two oxygen atoms of the carboxylate group (hydrogen bonds shorter than 1.8 Å), whereas the ammonium group interacts with the phenyl ring, with N-H···X contacts below 3 Å, somewhat shorter than other minima showing N-H···X contact as Phe-C or Phe-D.

Table 7.1 lists the values obtained for the complexation energies of the complexes shown in Figure 7.2, as obtained with a variety of methods. The reference values are, as commented above, the CCSD(T)/CBS estimated results. It can be observed that the most stable structure among the non-zwitterionic minima corresponds to Phe-A, with a complexation energy of -29.4 kcal mol^{-1}, which represents a very strong interaction, as expected taking into account the nature of the system considered. Closely following are Phe-B and Phe-C, which are isoenergetic, reaching complexation energies of -28.4 and -28.2 kcal mol^{-1}. Both structures keep the O-H···N hydrogen bond, so even though the intermolecular contact is less favorable, the global stability is quite close to that of Phe-A. Table 7.1 also shows the deformation energies obtained at the MP2/CBS

level of calculation for the different minima shown in Figure 7.2. It can be observed that this contribution, representing the changes in geometry of the fragments in the complex relative to the isolated geometry, is large in many cases. It is worth noting that almost all the deformation contribution comes from changes on the structure of the amino acid, guanidinium contributing with a small fraction of the total value (0.5-1 kcal mol^{-1}). In minima such as Phe-A, Phe-E, Phe-F, Phe-G and Phe-H the O-H···N intramolecular bond is broken, so all these structures show large deformation contributions. The lack of this hydrogen bond together with the conformational changes of the amino acid is the main source of the deformation energy values observed in Table 7.1. Therefore, the interaction in these structures is strong but counterbalanced to some extent by these deformation effects. On the other hand, Phe-C and Phe-D keep the intramolecular hydrogen bond and correspond to structures quite similar to those most stable for the isolated amino acid, so the contribution of deformation is much smaller. Therefore, even when the interaction between guanidinium and the amino acid is less favorable in these cases, the small deformation energy promotes these structures to be among the most stable ones. It has to be taken into account that the differences in deformation energy amount up to 6.3 kcal mol^{-1} and are therefore of the same magnitude as the energy differences among minima.

As regards the two zwitterionic structures, they are very stable (in fact the most stable structure at the CCSD(T)/CBS level is Phe-zw-A). This is in agreement with previous results where the presence of a cation allows the formation of stable zwitterionic structures in amino acid complexes even in the gas phase.[99-101] In any case, the energy pattern in zwitterionic structures is different. In this case, a huge interaction is established between the positively charged cation and the negative carboxylate group, thus leading to very large interaction energies. However, this huge interaction is counterbalanced by large deformation energies needed to break the O-H hydrogen bond to transfer the proton to the amino group. Thus, deformation energies of around 18-21 kcal mol^{-1} are observed in these structures. The balance of these two large contributions results in the final complexation energies observed. The small energy differences among the different minima suggest that the contribution from zero point energy (ZPE) correction can be of importance in these systems. As it can be observed in the last column of Table 7.1, the inclusion of ZPE introduces changes of about 0.5-1.5 kcal mol^{-1} in the energies of the complexes, but the order of stability remains almost unchanged (slight variations are only observed in structures with very similar complexation energies). Similar effects from ZPE corrections are also observed in complexes with the other amino acids. Another related aspect refers to entropy contributions to complexation. As listed in Appendix 4, including entropy at 298 K introduces several changes on the stability of the complexes, though the global picture does not change significantly. Also, entropy corrections should be taken with some prevention taking into account the level of calculation employed and the use of harmonic approximation.

Table 7.1. Complexation energies (kcal mol-1) for the complexes in Figure 7.2.

	M06-2X		B3LYP-D3	B2PLYP	MP2						MP2.X	SAPT/ scaled	CCSD(T)	
	6-31+G*	cc-PVTZ	cc-pVTZ	cc-pVTZ	AVDZ	SCSN/ AVDZ	SCS/ CBS	SCSN/ CBS	CBS	E_{def} MP2/CBS	6-31G*		CBS	CBS + ZPE[b]
Phe-A	-30.52	-30.26	-30.47	-28.66	-29.33	-29.52	-26.94	-30.38	-30.90	7.50	-29.54	-29.33	-29.40	-28.23
Phe-B	-28.97	-28.80	-29.80	-28.92	-27.63	-28.44	-25.67	-29.42	-29.25	6.29	-28.73	-28.75	-28.39	-27.34
Phe-C	-27.87	-27.70	-29.13	-29.57	-27.32	-28.01	-25.05	-28.54	-28.30	1.00	-28.20	-28.01	-28.25	-27.46
Phe-D	-27.01	-26.94	-28.23	-29.42	-26.22	-27.13	-24.45	-27.72	-27.27	1.94	-27.17	-27.10	-27.22	-26.48
Phe-E	-27.20	-27.08	-27.70	-27.80	-26.39	-26.91	-24.24	-27.34	-27.32	5.71	-26.74	-27.03	-26.77	-25.60
Phe-F	-26.86	-26.70	-27.70	-28.11	-25.88	-26.91	-24.26	-27.51	-27.08	5.62	-26.29	-26.72	-26.48	-25.67
Phe-G	-26.86	-26.70	-27.39	-27.99	-25.45	-26.55	-23.88	-27.06	-26.58	5.95	-25.91	-26.36	-26.00	-25.07
Phe-H	-24.40	-24.35	-25.33	-25.69	-23.95	-24.57	-21.99	-24.95	-24.69	7.29	-24.09	-24.31	-24.26	-23.40
Phe-zw-A	-28.61	-25.24	-28.78	-29.28	-28.80	-28.51	-25.50	-28.80	-29.66	18.52	-29.90	-27.84	-29.54	-28.47
Phe-zw-B	-27.89	-26.15	-29.66	-29.42	-27.65	-27.68	-24.90	-28.04	-28.61	20.96	-28.87	-26.84	-28.47	-27.08
MAE[a]	0.57	0.96	0.88	1.15	0.62	0.38	2.80	0.74	0.48	-	0.19	0.50	0.00	-

[a] Mean absolute error.

[b] Zero point energy correction in the M06-2X/6-31+G* level.

Table 7.1 also shows the performance of other methods of calculation describing the interaction in these systems. Considering the different MP2 variants extrapolated to basis limit, it is observed that all perform reasonably well, excepting the SCS-MP2 method which underestimates the interaction in these systems. A similarly good behavior is obtained with the cheaper DFT methods. M06-2X, B3LYP-D3 and B2PLYP perform similarly, though using the M06-2X/6-31+G* seems to provide slightly better results. Also, it can be observed that due to error compensation MP2 results obtained with the medium-sized AVDZ basis set give a description of similar quality as those extrapolated to basis limit. SAPT(DFT) scaled results are also in agreement with the reference ones. The best results are obtained with the MP2.X method, as expected since it is the second highest level employed and it could constitute an interesting approach when CCSD(T) calculations are out of reach. In fact the MP2.X results match those of CCSD(T), the order of stability being the same. Other methods, though producing good results show several differences in the order of stability of the complexes. It can be observed in Table 7.1 that the main difference among methods is related to the zwitterionic structures, and only MP2.X is able to reproduce the results obtained at the CCSD(T) level of calculation. In fact, the quality of SAPT(DFT) results improves to a large extent if zwitterionic structures are not considered (Mean Absolute Error (MAE)) of 0.21 kcal mol^{-1}; with the rest of methods there is no significant improvement). The discrepancy is not a problem of basis incompleteness, since MP2/CBS also fails to describe it properly. On the other hand, the problem is related to the quality of the *N*-electron model, and only when corrections taking this into account are included (MP2.X and CCSD(T)) the results are correct. In fact, MP2/CBS describes quite accurately the zwitterionic structures, the problem being the overestimation of the interaction in the other minima. This can be checked considering that the CCSD(T) or MP2.X corrections to the MP2/CBS values are among the smallest in zwitterions, Phe-C and Phe-D, corresponding to the most extended structures. Though relatively small, these corrections introduce a difference of about 1.5 kcal mol^{-1} between Phe-A and Phe-zw-A, which reverses the order of stability obtained with MP2/CBS. Therefore, the differences seem to be related to the well-known problem of the MP2 method overestimating interactions with aromatic systems.

7.3.3. Tyrosine- guanidinium complexes

Figure 7.3 shows the selected most stable structures obtained for tyrosine complexes with guanidinium cation. As expected, tyrosine shows a more complex behavior due to the presence of the hydroxyl group. As observed in previous work, the minima appear by pairs with similar stability differing on the orientation of the O-H group of tyrosine pointing to opposite sides of the ring.[87, 89]

For brevity, Figure 7.3 shows the selected most stable minima avoiding those that differ only by the position of the OH group. As shown in Figure 7.3, the presence of the hydroxyl group offers a new point for favorable interaction with guanidinium cation, so most of the structures exhibit a contact between the NH_2 groups of guanidinium and the hydroxyl oxygen. Tyr-A is similar to Phe-A, with two hydrogen bonds between guanidinium and the carboxyl and NH_2 groups at around 2.0 Å. Also, a contact to the π cloud is observed at 2.5 Å. However, a new hydrogen bond is formed with the hydroxyl group of phenol at 2.2 Å. In order to form this contact guanidinium is oriented more parallel to the ring than in the complex with Phe.

Tyr-A is the only structure among those shown in Figure 7.3, where the intramolecular O-H···N hydrogen bond is broken. Also, except the more extended structures Tyr-D and Tyr-F (also Tyr-G) all structures show a contact to the hydroxyl group. Even though this hydrogen bond is quite long it introduces extra stabilization into the complex, so the most stable structures show this motif. The distance is around 2.2 Å in all cases. When guanidinium does not interact with the ring and the hydroxyl group, this is partially compensated by secondary N-H···X interactions at around 3 Å, both in zwitterionic and neutral species. As in Phe complexes, the zwitterionic structures clearly correspond to other neutral ones, so Tyr-zw-A is like Tyr-D and Tyr-zw-B like Tyr-F, the only difference being the proton transferred to the ammonium group in zwitterionic minima.

Table 7.2 lists the complexation energies obtained for these species. Focusing again in CCSD(T)/CBS values, it can be appreciated that Tyr complexes are slightly more tightly bound than Phe ones, so the complexation energy of the most stable neutral species reaches -29.8 kcal mol^{-1}, just a barely 0.4 kcal mol^{-1} with respect to Phe most stable neutral species.[34, 38] Tyr-B and Tyr-C are very close in energy (closer than in Phe complexes) being just 0.31-0.34 kcal mol^{-1} less stable than Tyr-A. Afterwards, there is a gap of around 0.93 kcal mol^{-1} to the next most stable minimum. As in Phe complexes, zwitterionic structures are relevant, so Tyr-zw-A is isoenergetic with Tyr-A. The interaction in zwitterionic structures is stronger by large, but due to the cost of proton transfer from neutral species, there is an energetic cost of deformation which partially compensates this effect. In fact, deformation is large in all structures, even larger than in Phe complexes because the molecule has to be more distorted in order to interact with the hydroxyl group. Only the extended structures Tyr-D and Tyr-F show small values for the deformation energy.

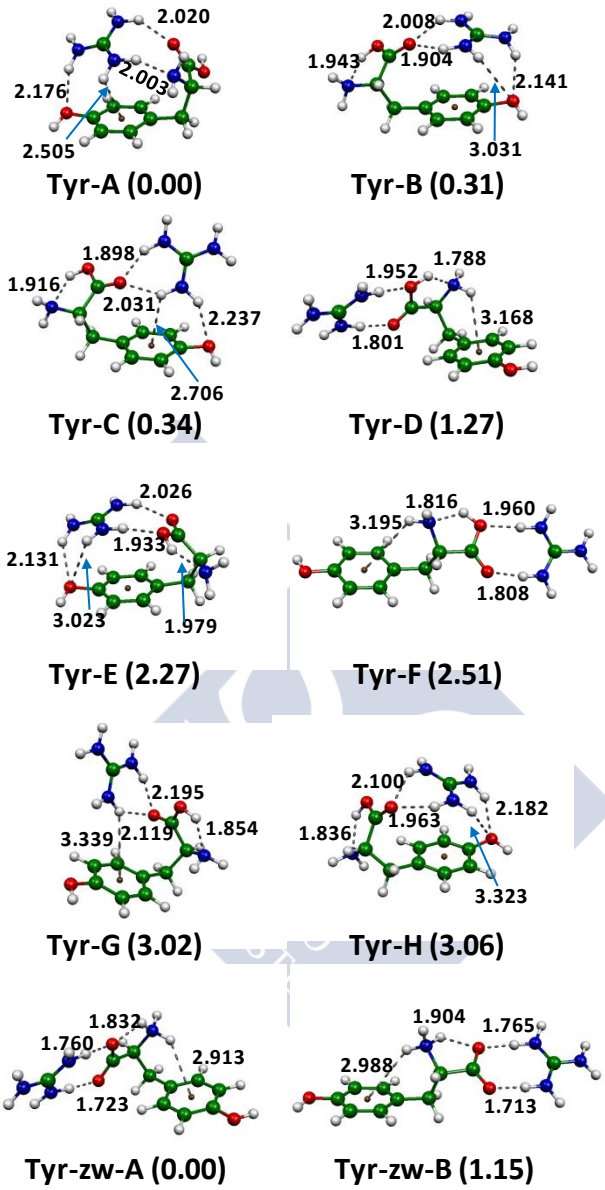


Figure 7.3. Selected most stable minima found for complexes formed by guanidinium cation and tyrosine as obtained at the M06-2X/6-31+G* level of calculation. Selected distances in Å. Relative energies (kcal mol^{-1}) at the CCSD(T)/CBS level in parentheses.

Table 7.2. Complexation energies (kcal mol-1) for the complexes in Figure 7.3.

	M06-2X		B3LYP-D3	B2PLYP	MP2						MP2.X	SAPT/ scaled	CCSD(T)	
	6-31+G*	cc-PVTZ	cc-pVTZ	cc-pVTZ	AVDZ	SCSN/ AVDZ	SCS/ CBS	SCSN/ CBS	CBS	E_{def} MP2/CBS	6-31G*		CBS	CBS + ZPE[b]
Tyr-A	-31.69	-31.12	-30.47	-27.37	-29.64	-29.30	-26.72	-30.26	-31.48	10.25	-29.78	-29.52	-29.76	-28.61
Tyr-B	-31.02	-30.16	-30.43	-27.82	-28.85	-28.70	-26.08	-29.78	-30.71	7.07	-29.57	-29.61	-29.45	-28.11
Tyr-C	-29.83	-29.40	-30.19	-29.33	-28.75	-28.99	-26.27	-29.80	-30.16	7.77	-29.59	-29.49	-29.42	-28.51
Tyr-D	-28.06	-27.99	-29.59	-30.09	-27.58	-28.25	-25.14	-28.82	-28.59	1.41	-28.39	-28.06	-28.49	-27.87
Tyr-E	-28.78	-27.96	-28.54	-25.12	-27.25	-26.10	-23.78	-26.98	-28.85	8.27	-27.49	-27.15	-27.49	-26.31
Tyr-F	-26.94	-27.03	-28.42	-29.64	-26.20	-27.15	-24.31	-27.80	-27.29	2.25	-27.10	-27.17	-27.25	-26.67
Tyr-G	-27.17	-27.20	-27.75	-26.86	-25.91	-26.46	-23.35	-27.13	-27.22	9.18	-26.72	-26.65	-26.74	-26.20
Tyr-H	-28.82	-28.20	-27.61	-24.09	-26.48	-25.62	-22.97	-26.43	-28.11	3.49	-26.77	-26.84	-26.70	-25.48
Tyr-zw-A	-28.75	-26.43	-30.14	-30.02	-29.04	-28.78	-25.57	-29.13	-29.92	18.88	-30.04	-28.25	-29.76	-28.70
Tyr-zw-B	-27.94	-25.45	-29.11	-29.64	-27.77	-27.82	-24.86	-28.23	-28.75	21.44	-28.92	-27.20	-28.61	-28.08
MAE[a]	1.03	1.22	0.81	1.51	0.62	0.65	3.47	0.43	0.74	-	0.12	0.45	0.00	-

[a] Mean absolute error.

[b] Zero point energy correction in the M06-2X/6-31+G* level.

As regards the different methods considered, most of them underestimate the relative stability of the zwitterionic structures except MP2.X which becomes closer to the reference values at the CCSD(T)/CBS level. In any case, even MP2.X slightly departs from CCSD(T), giving somewhat more stabilized zwitterions, which become the most stable structure by 0.3 kcal mol^{-1} over Tyr-A. In Tyr complexes, the deviations observed for M06-2X are somewhat larger, but this is probably related with overestimation of the N-H···O hydrogen bonds. The rest of methods behave similarly as before, with the best behavior given by SAPT(DFT), especially if zwitterionic structures are not considered. As before, MP2/AVDZ systematically underestimates the strength of the interaction, whereas when extrapolated to basis limit the values are overestimated.

7.3.4. Tryptophan-guanidinium complexes

Figure 7.4 shows the selected most stable structures found for tryptophan complexes with guanidinium. The behavior is similar to that observed in the previous cases, and even more similar to Phe complexes. Most minima keep the intramolecular O-H···N hydrogen bond, except in two of the selected structures in Figure 7.4. One of these structures is Trp-A, which is similar to Phe-A, and allows the formation of the hydrogen bonds to carboxyl and amino group whereas guanidinium also interacts with the phenyl ring in the indole moiety. As in the case of Tyr complexes, most structures show a guanidinium···π contact despite the energy cost needed to fold the amino acid. It is worth noting that due to the structure of tryptophan, the cation···π contact is more easily formed, so the NH···X distance to the center of the phenyl ring is considerably shorter than in Phe complexes, decreasing to around 2.3-2.5 Å. The larger amount of structures showing cation···π contact allows thinking that the guanidinium···indole contact is more favorable, so most stable structures show this motif.

As regards complexation energies listed in Table 7.3, the trends are similar to those shown before. However, up to four structures are found in barely 0.5 kcal mol^{-1}, so from Trp-A to Trp-D the differences on whether guanidinium interacts with the carboxyl or NH group, or the different orientations relative to the ring, hardly have an effect on the energy of the system. Of course, there are several factors affecting the values, such as the deformation energy, but it becomes clear that the extended aromatic cloud of Trp allows for a larger variety of similarly stable structures. In Trp complexes, however, zwitterionic structures are always less stable than the most stable neutral minimum, even at the MP2.X or CCSD(T) levels of calculation. In any case, the values obtained with any of the other methods suggest even larger energy differences which, as before, are reduced after a better description of electron correlation is included. M06-2X and MP2-based methods tend to overestimate the interaction when a cation···π contact is formed with an extended aromatic system, though SCSN-MP2 gives quite a good performance. As in previous cases, SAPT(DFT) results are among the best obtained. Considering results overall, the best

balance between accuracy and computational cost would probably be that of M06-2X, B3LYP-D3 and SCSN-MP2/aug-cc-pVDZ.

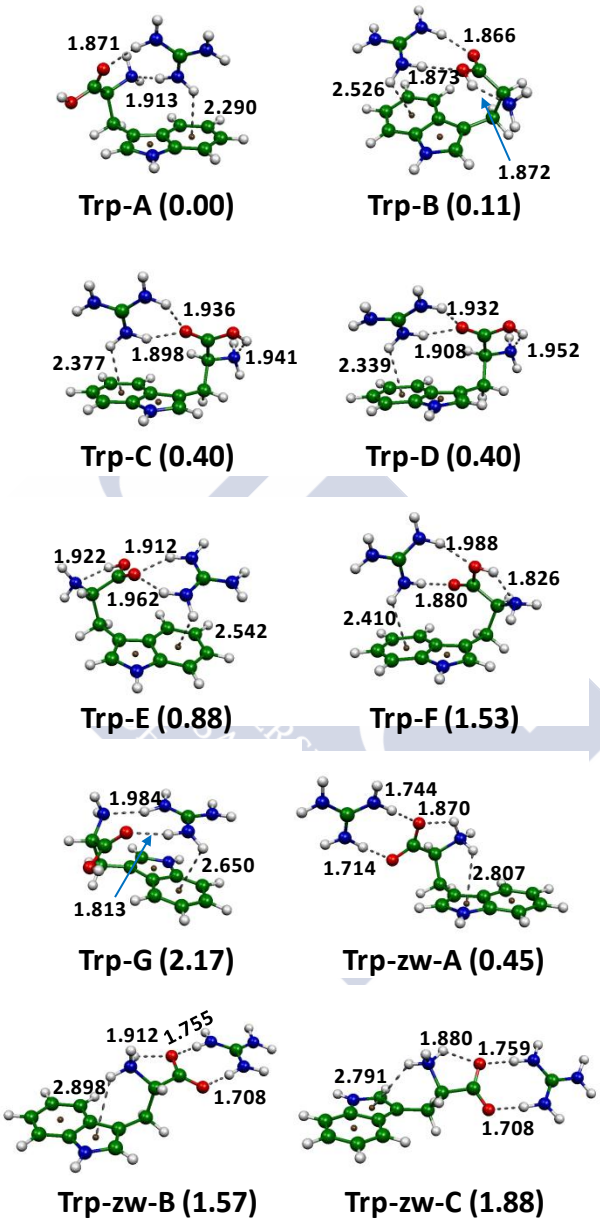


Figure 7.4. Selected most stable minima found for complexes formed by guanidinium cation and tryptophan as obtained at the M06-2X/6-31+G* level of calculation. Selected distances in Å. Relative energies (kcal mol^{-1}) at the CCSD(T)/CBS level in parentheses.

Table 7.3. Complexation energies (kcal mol-1) for the complexes in Figure 7.4.

	M06-2X		B3LYP-D3	B2PLYP	MP2						MP2.X	SAPT/ scaled	CCSD(T)	
	6-31+G*	cc-PVTZ	cc-pVTZ	cc-pVTZ	AVDZ	SCSN/ AVDZ	SCS/ CBS	SCSN/ CBS	CBS	E_{def} MP2/CBS	6-31G*		CBS	CBS + ZPE[b]
Trp-A	-32.74	-32.82	-33.20	-31.43	-32.17	-32.39	-29.33	-33.32	-33.87	8.63	-32.10	-31.72	-32.07	-30.95
Trp-B	-32.53	-31.93	-32.91	-29.92	-31.98	-31.02	-28.06	-31.79	-33.44	5.90	-32.00	-31.57	-31.96	-30.43
Trp-C	-31.74	-31.72	-32.89	-31.60	-31.17	-31.74	-28.59	-32.86	-33.08	6.79	-31.93	-31.74	-31.67	-30.69
Trp-D	-31.86	-31.52	-32.91	-31.31	-31.14	-31.84	-28.66	-32.96	-32.96	6.72	-31.93	-31.67	-31.67	-30.40
Trp-E	-32.15	-31.76	-32.50	-30.16	-30.83	-30.78	-27.84	-31.88	-32.70	7.60	-31.41	-31.48	-31.19	-29.80
Trp-F	-30.66	-30.33	-31.52	-30.07	-30.31	-29.92	-26.96	-30.74	-31.81	5.11	-30.62	-30.09	-30.54	-29.52
Trp-G	-31.81	-31.24	-30.98	-27.06	-30.52	-29.68	-26.72	-30.54	-32.17	10.95	-29.92	-29.95	-29.90	-28.39
Trp-zw-A	-30.50	-27.94	-31.72	-31.84	-30.95	-30.62	-27.20	-30.88	-31.76	18.07	-31.93	-29.90	-31.62	-30.16
Trp-zw-B	-29.66	-27.01	-30.88	-31.48	-29.80	-29.59	-26.48	-29.88	-30.64	21.32	-30.81	-29.09	-30.50	-29.28
Trp-zw-C	-29.33	-22.87	-26.39	-23.30	-29.28	-29.45	-26.29	-29.88	-30.35	21.10	-30.50	-28.68	-30.19	-28.56
MAE[a]	0.74	1.31	1.22	1.72	0.45	0.55	3.51	0.72	1.15	-	0.19	0.62	0.00	-

[a] Mean absolute error.

[b] Zero point energy correction in the M06-2X/6-31+G* level.

7.3.5. SAPT(DFT) analysis

As commented in Computational Details, SAPT(DFT) allows partitioning the total interaction energy into components associated to recognizable physical phenomena. Figure 7.5 shows the results for the SAPT(DFT) decomposition in selected complexes of guanidinium with the three aromatic amino acids considered in this work. In the following, four contributions will be considered: electrostatic (E_{ele}= E_{1pol}), repulsion (E_{rep}= E_{1exch}), induction (E_{ind}= $E_{2ind-exch}$ + E_{2ind} + δHF), and dispersion (E_{disp}= $E_{2disp-exch}$ + E_{2disp}).

Also included is the deformation energy obtained at the MP2/CBS level of calculation, as an indicative of the role played by geometry changes on the energetics of the systems. This final term is included because SAPT(DFT) only gives information about the interaction at a given geometry. The complexes selected correspond to typical structures: the most stable minimum, which in all cases shows a guanidinium···π contact, the most stable zwitterionic structure, and a structure without the cation···π contact. It has to be indicated that in the case of Trp complexes no extended structure is shown in Figure 7.4, so a less stable, extended structure has been taken from Appendix 4 to do this analysis. SAPT(DFT) results for all minima can be found in Appendix 4.

It can be observed in Figure 7.5 that in the Phe complex exhibiting cation···π contact, the leading stabilizing contribution comes from electrostatics, as expected taking into account that a cation is interacting with electron-rich regions of the amino acid. This large electrostatic contribution is matched by a large repulsion contribution. On the other hand, induction contributes significantly to the stability of the complex, giving -17.8 kcal mol^{-1}. Dispersion, though being the smallest of the attractive contributions, still amounts to an important -13.9 kcal mol^{-1}. Therefore, the total interaction energy comes from the joined effect of three important stabilizing contributions against the repulsive wall. However, as commented above, the deformation contribution has an important role in the energetics of these systems, amounting to 7.5 kcal mol^{-1} in Phe-A complex.

In the extended minimum, the combination of first-order contributions (repulsion plus electrostatics) favors the extended minimum by 5.1 kcal mol^{-1}, being reinforced by a significantly smaller deformation contribution. Despite of this, the cation···π structure is more stable than the extended one as a consequence of larger contributions from dispersion and induction, which are capable of overcoming the unfavorable balance of first-order terms and especially deformation. As a consequence of these opposite factors the differences in stability reduce to a barely 1.3 kcal mol^{-1} favoring the cation···π minimum.

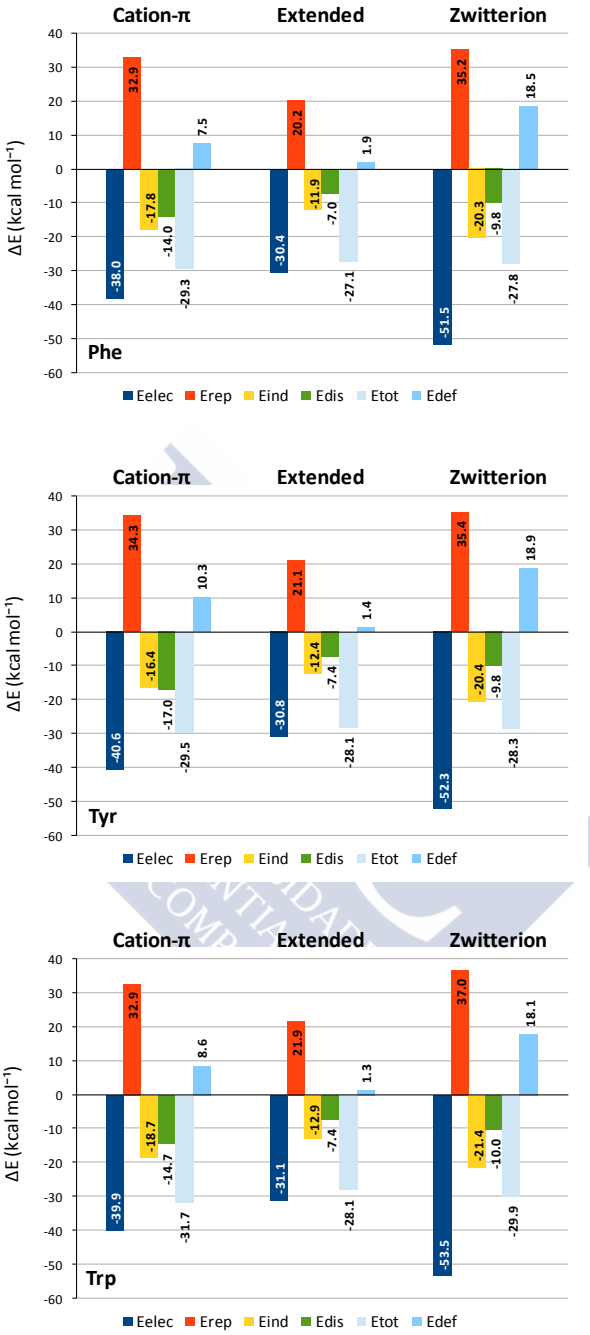


Figure 7.5. Contributions to the interaction energy (kcal mol^{-1}) for selected structures of the complexes formed by guanidinium and aromatic amino acids as obtained with the SAPT(DFT) calculations. Deformation energies included at the MP2/CBS level of calculation.

In the case of the zwitterionic structure, the components of the interaction energy show a different pattern. Electrostatic contribution is even larger, reaching -51.5 kcal mol^{-1}, whereas repulsion shows values similar to those found for cation···π structures. Such a large electrostatic interaction is also accompanied by a large induction contribution, whereas dispersion is moderate. The key factor in the energy balance of zwitterionic minima is the large deformation contribution which amounts to 18.5 kcal mol^{-1}, and partially cancels out the large electrostatic contribution, making the zwitterionic structures less stable than the cation···π ones.

In tyrosine complexes the behavior is similar. The extended structure is favored by smaller deformation energies together with more attractive electrostatic + repulsion terms, but the difference is recovered and the stability reversed due to larger dispersion and induction contributions in the cation···π complex. The same is observed in Trp complexes, though the combination of electrostatic + repulsion + deformation favors the extended structure by 9.5 kcal mol^{-1}, the smallest among the different amino acids. This value, combined with the larger induction and dispersion contributions, makes the extended structure to be 3.6 kcal mol^{-1} less stable than the cation···π one, whereas in Phe and Tyr the difference amounts to only 1.3 and 1.4 kcal mol^{-1}, respectively. This explains why in Trp complexes the extended structures are not among the most stable ones.

Comparing similar structures with different amino acids the following aspects can be highlighted. Considering cation···π structures, going from Phe to Tyr and Trp, the contributions from electrostatics and dispersion increase. This, together with an increase in deformation, especially in Tyr, makes the stability of Tyr complex almost the same as with Phe, whereas increases by -2.4 kcal mol^{-1} are observed in Trp. In extended structures, the complexes show similar stability, with an increase of -1 kcal mol^{-1} in Tyr, as a consequence of small increments in all contributions and a decrease in deformation contribution. In Trp complex, only induction increases but it is matched by similar increments in repulsion. Finally, zwitterionic Tyr complex shows similar stability as Phe one, whereas Trp one is around -2 kcal mol^{-1} more stable, mainly as a consequence of larger electrostatic and induction contributions. Therefore, Phe and Tyr complexes show similar stability, whereas Trp ones are the most stable ones, due to small increases on several of the contributions to the interaction.

7.3.6. NCI analysis

Figure 7.6 shows the Non covalent Interaction (NCI) index plots for the selected amino acids commented above. NCI is an index based on the analysis of the reduced density gradient which can be employed to visualize both favorable and unfavorable interactions. These interactions can be graphically displayed as a plot of the product of the second eigenvalue of the hessian of the

density times the density, mapped onto an isosurface of reduced density gradient. As shown in Figure 7.6, (sign(λ_2) ρ) is plotted onto an isosurface of reduced density gradient of value 0.5 a.u. The color scale ranges from 0.02 a.u. for unfavorable interactions (red) to -0.02 a.u. for favorable interactions (blue).

It can be observed that there is a distinct interaction pattern for each of the structures. The cation···π minimum exhibits two well-developed hydrogen bonds, corresponding to N-H···O and N-H···N contacts, being the latter the strongest. The interaction pattern in both the extended and the zwitterionic forms is very similar, the most important fact being the intramolecular N-H···O hydrogen bond which gives stability to the amino acid as commented above.

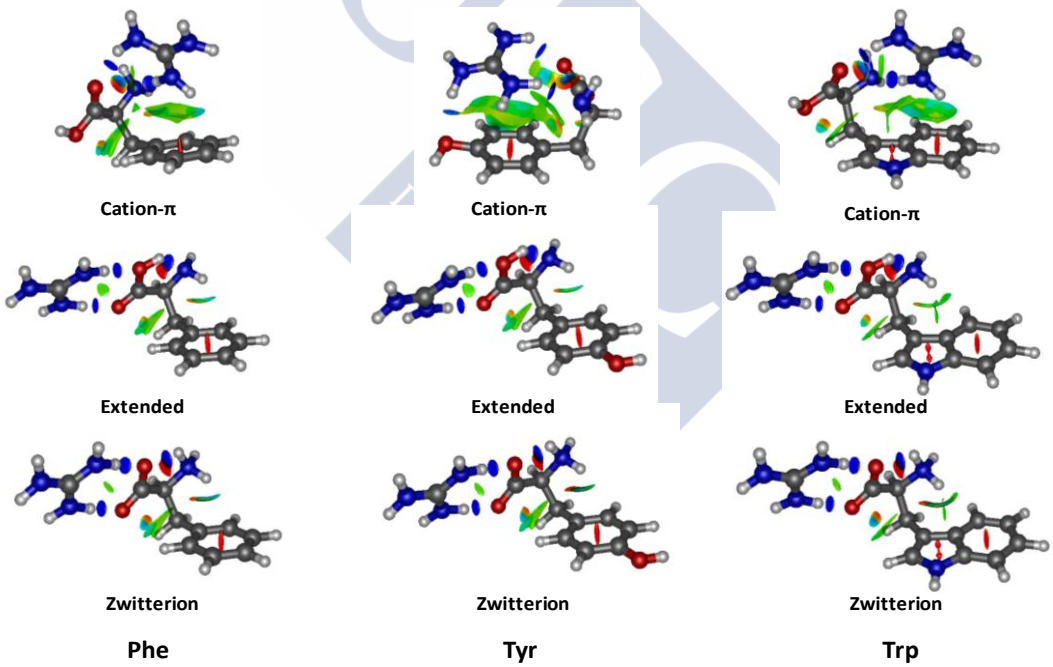


Figure 7.6. NCI plots for selected minima of guanidinium with the aromatic amino acids studied in this work. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to -0.02 a.u. (blue).

However, this is an intramolecular contact and its effect upon the complexation energy comes by means of the deformation energy of the cation···π minima, where this intramolecular contact has to be broken. Two N-H···O hydrogen bonds are also formed and, though difficult to appreciate, in the zwitterionic form the contacts are stronger as a consequence of the direct contact of the carboxylate group and the cation. Besides, there is a weak contact between the amino group and the aromatic clouds of the amino acids. Also, weak contacts are observed between the carboxyl oxygens and C-H groups nearby. In cation···π minima, there is also a broad region of favorable weak interaction corresponding to the contact of the cation with the aromatic cloud of benzene, phenol and indole. In tyrosine, an extra hydrogen bond is displayed between the cation and the hydroxyl group of phenol. Therefore, the interaction of guanidinium with the carboxylic/amino group of the amino acids is quite similar in all cases, the differences being a consequence of the favorable contact of the cation with the aromatic rings, which has to overcome the breaking of the intramolecular N-H···O hydrogen bond. As commented above, the balance of stronger interactions with the aromatic rings, partially cancelled out by deformation of the amino acid structure controls the final order of stability. In fact, as broader regions of the aromatic cloud (and the hydroxyl group) are participating on the stabilizing interactions, the cation···π minimum becomes more and more stable with respect to extended forms.

7.4. Conclusions

The characteristics of the interaction between the guanidinium cation and phenylalanine, tyrosine and tryptophan have been computationally studied with a set of different computational methods. Benchmark values for the complexation energy of the systems studied have been estimated with the CCSD(T) method extrapolated to the complete basis set limit. The results show that the interaction for the complexes formed with the amino acids in neutral form amounts to -29.40, -29.76 and -32.07 kcal mol^{-1} for Phe, Tyr and Trp respectively. In all cases, these structures correspond to minima showing cation···π contacts. However, complexes with the zwitterionic form of the amino acid can be as stable as the most stable minima found with the neutral forms.

As regards method performance it has been found that overall all methods tested give a pretty reasonable description of the interaction. MP2.X gives the best results when compared to CCSD(T)/CBS, followed by the SAPT(DFT) when dispersion is empirically scaled. Overall, probably the M06-2X, B3LYP-D and the SCSN-MP2/aug-cc-pVDZ are the best choices between accuracy and computational costs. In any case, the relative stability between neutral and zwitterionic dimers is only properly described with the MP2.X and CSSD(T) methods, the failures of other methods being related with overestimation of the interaction in cation···π contacts. These results show the importance of using high-level calculations in order to properly describe the characteristics of this kind of complexes.

SAPT(DFT) partitioning of the complexation energies reveals the intrinsic different nature of the interaction depending on the conformation of the amino acid. Electrostatic and repulsion favor the formation of dimers with the amino acid adopting an extended structure which is also favored by significantly smaller deformation energies. However, dimers showing cation···π interactions are more stable as a consequence of larger dispersion and induction contributions, which are able to overcome the cost of deformation and thus to revert the stability order. In the case of zwitterionic complexes, the interaction is strongest, but partially cancelled out by the large deformation cost associated to proton transfer.

In summary, guanidinium interacts with virtually the same intensity, both in neutral and zwitterionic form, with Phe and Tyr. Interaction with Trp is somewhat stronger, with a slight preference for the neutral form. Folded structures are favored both by increased induction and dispersion contributions associated to the cation···π contact, which are capable of overcoming the cost of deformation associated to the structural change of the amino acids. A proper description of induction (usually neglected in common force fields) and dispersion (usually underrated) is crucial in order to correctly describe the interaction in these systems. Only by properly quantifying all contributions to the interaction and the cost coming from conformational changes of the amino acid the relative stability and structure of these systems can be understood.

7.5. References

(1) P. Hobza, R. Zaradnik, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and the biodisciplines*, Elsevier, Amsterdam, **1988**.
(2) K. E. Riley, P. Hobza. *Acc. Chem. Res.* **2013**, *46*, 927-936.
(3) H.-J. Schneider, A. K. Yatsimirski, *Principles and methods in supramolecular chemistry*, Wiley, Chichester, **2000**.
(4) L. M. Salonen, M. Ellermann, F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-42.
(5) M. L. Waters. *Biopolymers (Peptide Science)* **2004**, *76*, 435-445.
(6) M. Nishio. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873-900.
(7) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-38.
(8) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.
(9) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.
(10) O. Takahashi, Y. Kohno, M. Nishio. *Chem. Rev.* **2010**, *110*, 6049-6076.
(11) D. A. Dougherty. *Acc. Chem. Res.* **2013**, *46*, 885-93.
(12) A. Frontera, D. Quiñonero, P. M. Deyà. *WIREs Comput. Mol. Sci.* **2011**, *1*, 440-459.
(13) B. L. Schottel, H. T. Chifotides, K. R. Dunbar. *Chem. Soc. Rev.* **2008**, *37*, 68-83.

(14) A. Campo-Cacharron, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Theor. Chem. Acc.* **2012**, *131*, 1-13.
(15) A. A. Rodriguez-Sanz, J. Carrazana-Garcia, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *J. Mol. Model.* **2013**, *19*, 1985-1994.
(16) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.
(17) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.
(18) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.
(19) D. Y. Kim, N. J. Singh, J. W. Lee, K. S. Kim. *J. Chem. Theory Comput.* **2008**, *4*, 1162-1169.
(20) J. A. Carrazana-Garcia, J. Rodriguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2012**, *116*, 5860-5871.
(21) D. Kim, E. C. Lee, K. S. Kim, P. Tarakeshwar. *J. Phys. Chem. A* **2007**, *111*, 7980-6.
(22) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.
(23) P. Ballester. *Acc. Chem. Res.* **2013**, *46*, 874-884.
(24) A. Campo-Cacharrón, E. M. Cabaleiro-Lago, J. Rodríguez-Otero. *ChemPhysChem* **2012**, *13*, 570-577.
(25) H. T. Chifotides, K. R. Dunbar. *Acc. Chem. Res.* **2013**, *46*, 894-906.
(26) P. Garcia-Novo, A. Campo-Cacharron, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Phys. Chem. Chem. Phys.* **2012**, *14*, 104-112.
(27) J. P. Gallivan, D. A. Dougherty. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459-64.
(28) N. Zacharias, D. A. Dougherty. *Trends Pharmacol. Sci.* **2002**, *23*, 281-287.
(29) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.
(30) I. Soteras, M. Orozco, F. J. Luque. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2616-2624.
(31) J. A. Carrazana-García, J. s. Rodríguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2011**, *115*, 2774-2782.
(32) A. S. Reddy, G. M. Sastry, G. N. Sastry. *Proteins: Structure, Function, and Bioinformatics* **2007**, *67*, 1179-1184.
(33) P. B. Armentrout, M. Citir, Y. Chen, M. T. Rodgers. *J. Phys. Chem. A* **2012**, *116*, 11823-11832.
(34) P. B. Armentrout, B. Yang, M. T. Rodgers. *J. Phys. Chem. B* **2013**, *117*, 3771-3781.
(35) F. Costanzo, R. G. Della Valle. *J. Phys. Chem. B* **2008**, *112*, 12783-12789.
(36) N. C. Polfer, J. Oomens, R. C. Dunbar. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2744-2751.
(37) M. Remko, S. Soralova. *JBIC, J. Biol. Inorg. Chem.* **2012**, *17*, 621-630.
(38) C. Ruan, M. T. Rodgers. *J. Am. Chem. Soc.* **2004**, *126*, 14600-14610.

(39) P. Wang, G. Ohanessian, C. Wesdemiotis. *Int. J. Mass spectrom.* **2008**, *269*, 34-45.
(40) R. Wu, T. B. McMahon. *J. Am. Chem. Soc.* **2008**, *130*, 12554-12555.
(41) S. J. Ye, A. A. Clark, P. B. Armentrout. *J. Phys. Chem. B* **2008**, *112*, 10291-10302.
(42) M. Remko, D. Fitz, R. Broer, B. M. Rode. *J. Mol. Model.* **2011**, *17*, 3117-3128.
(43) E. Rezabal, T. Marino, J. M. Mercero, N. Russo, J. M. Ugalde. *Inorg. Chem.* **2007**, *46*, 6413-6419.
(44) E. Rezabal, T. Marino, J. M. Mercero, N. Russo, J. M. Ugalde. *J. Chem. Theory Comput.* **2007**, *3*, 1830-1836.
(45) A. Rimola, L. Rodriguez-Santiago, M. Sodupe. *J. Phys. Chem. B* **2006**, *110*, 24189-24199.
(46) A. Rimola, M. Sodupe, J. Tortajada, L. Rodriguez-Santiago. *Int. J. Mass spectrom.* **2006**, *257*, 60-69.
(47) H. Minoux, C. Chipot. *J. Am. Chem. Soc.* **1999**, *121*, 10366-10372.
(48) A. Rodríguez-Sanz, E. Cabaleiro-Lago, J. Rodríguez-Otero. *J. Mol. Model.* **2014**, *20*, 1-10.
(49) J. S. Rao, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2009**, *113*, 7225-7236.
(50) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.
(51) A. A. Rodriguez-Sanz, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Org. Biomol. Chem.* **2014**, *12*, 2938-2949.
(52) D. Vijay, G. N. Sastry. *Phys. Chem. Chem. Phys.* **2008**, *10*, 582-590.
(53) *Suite 2012: MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2012.*
(54) Y. Zhao, D. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
(55) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.
(56) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.
(57) S. Grimme. *J. Comput. Chem.* **2004**, *25*, 1463-1473.
(58) S. Grimme. *J. Comput. Chem.* **2006**, *27*, 1787-1799.
(59) S. Grimme, J. Antony, S. Ehrlich, H. Krieg. *J. Chem. Phys.* **2010**, *132*, 154104/1-154104/19.
(60) S. Grimme, S. Ehrlich, L. Goerigk. *J. Comput. Chem.* **2011**, *32*, 1456-1465.
(61) S. Grimme. *J. Chem. Phys.* **2006**, *124*, 034108/1-034108/16.
(62) T. Helgaker, W. Klopper, H. Koch, J. Noga. *J. Chem. Phys.* **1997**, *106*, 9639-9646.
(63) S. Grimme. *J. Chem. Phys.* **2003**, *118*, 9095-9102.
(64) J. G. Hill, J. A. Platts. *J. Chem. Theory Comput.* **2007**, *3*, 80-85.
(65) C. D. Sherrill. *Rev. Comput. Chem.* **2009**, *26*, 1-38.
(66) K. E. Riley, P. Hobza. *WIREs Comput. Mol. Sci.* **2011**, *1*, 3-17.
(67) K. E. Riley, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2011**, *13*, 21121-21125.
(68) B. Jeziorski, R. Moszynski, K. Szalewicz. *Chem. Rev.* **1994**, *94*, 1887-1930.

(69) A. Hesselmann, G. Jansen, M. Schütz. *J. Chem. Phys.* **2005**, *122*, 014103/1-014103/17.

(70) A. J. Misquitta, K. Szalewicz. *J. Chem. Phys.* **2005**, *122*, 214109/1-214109/19.

(71) J. Řezáč, P. Hobza. *J. Chem. Theory Comput.* **2011**, *7*, 685-689.

(72) H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Kö, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, D. Peng, K. P. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, *Molpro version 2010.1, a package of ab initio programs, see http://www.molpro.net,(2010)*.

(73) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. M. J. A, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox. Gaussian 09, Revision B.01, Gaussian, Inc., *Wallingford, CT* **2009**.

(74) *TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.*

(75) F. Neese. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73-78.

(76) F. Neese, *ORCA, an ab initio, DFT and semiempirical SCF-MO package, Version 3.0.1*, **2014**.

(77) F. Neese, F. Wennmohs, A. Hansen, U. Becker. *Chem. Phys.* **2009**, *356*, 98-109.

(78) F. Weigend. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285-4291.

(79) F. Weigend, A. Köhn, C. Hättig. *J. Chem. Phys.* **2002**, *116*, 3175-3183.

(80) L. C. Snoek, E. G. Robertson, R. T. Kroemer, J. P. Simons. *Chem. Phys. Lett.* **2000**, *321*, 49-56.

(81) Y. Lee, J. Jung, B. Kim, P. Butz, L. C. Snoek, R. T. Kroemer, J. P. Simons. *J. Phys. Chem. A* **2004**, *108*, 69-73.

(82) T. Hashimoto, Y. Takasu, Y. Yamada, T. Ebata. *Chem. Phys. Lett.* **2006**, *421*, 227-231.

(83) G. von Helden, I. Compagnon, M. N. Blom, M. Frankowski, U. Erlekam, J. Oomens, B. Brauer, R. B. Gerber, G. Meijer. *Phys. Chem. Chem. Phys.* **2008**, *10*, 1248-1256.

(84) T. Ebata, T. Hashimoto, T. Ito, Y. Inokuchi, F. Altunsu, B. Brutschy, P. Tarakeshwar. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4783-4791.

(85) U. Purushotham, D. Vijay, G. Narahari Sastry. *J. Comput. Chem.* **2012**, *33*, 44-59.

(86) W. Zhang, V. Carravetta, O. Plekan, V. Feyer, R. Richter, M. Coreno, K. C. Prince. *J. Chem. Phys.* **2009**, *131*, 035103/1-035103/11.

(87) U. Purushotham, G. Narahari Sastry. *Theor. Chem. Acc.* **2012**, *131*, 1-14.

(88) Y. Shimozono, K. Yamada, S.-i. Ishiuchi, K. Tsukiyama, M. Fujii. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5163-5175.

(89) M. Zhang, Z. Huang, Z. Lin. *J. Chem. Phys.* **2005**, *122*, 134313/1-134313/7.

(90) A. Abo-Riziq, L. Grace, B. Crews, M. P. Callahan, T. van Mourik, M. S. d. Vries. *J. Phys. Chem. A* **2011**, *115*, 6077-6087.

(91) Y. Inokuchi, Y. Kobayashi, T. Ito, T. Ebata. *J. Phys. Chem. A* **2007**, *111*, 3209-3215.

(92) L. I. Grace, R. Cohen, T. M. Dunn, D. M. Lubman, M. S. de Vries. *J. Mol. Spectrosc.* **2002**, *215*, 204-219.

(93) A. Lindinger, J. P. Toennies, A. F. Vilesov. *J. Chem. Phys.* **1999**, *110*, 1429-1436.

(94) L. C. Snoek, R. T. Kroemer, M. R. Hockridge, J. P. Simons. *Phys. Chem. Chem. Phys.* **2001**, *3*, 1819-1826.

(95) K. Y. Baek, Y. Fujimura, M. Hayashi, S. H. Lin, S. K. Kim. *J. Phys. Chem. A* **2011**, *115*, 9658-9668.

(96) U. Purushotham, G. N. Sastry. *J. Comput. Chem.* **2014**, *35*, 595-610.

(97) B. Hernández, F. Pflüger, A. Adenier, S. G. Kruglik, M. Ghomi. *J. Phys. Chem. B* **2010**, *114*, 15319-15330.

(98) M. E. Sanz, C. Cabezas, S. Mata, J. L. Alonso. *J. Chem. Phys.* **2014**, *140*, 204308/1-204308/6.

(99) R. Wu, T. B. McMahon. *J. Am. Chem. Soc.* **2008**, *130*, 3065-3078.

(100) R. R. Julian, M. F. Jarrold. *J. Phys. Chem. A* **2004**, *108*, 10861-10864.

(101) T. Wyttenbach, M. Witt, M. T. Bowers. *J. Am. Chem. Soc.* **2000**, *122*, 3458-3464.

8

Cation···π interactions between imidazolium and aromatic amino acids

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

8.1. Introduction

Within every living organism, proteins carry out specific essential functions, so much that without them life would be non-viable. These molecular machines are the responsible of obtaining chemical energy and also of fabricating useful compounds for the organism itself. This high degree of specificity can be obtained with a certain number of molecules that in the case of proteins corresponds to twenty standard amino acids only differing on their side chains.[1] Each side chain is highly specific, exhibiting different characteristics or interacting in a different manner with other surrounding molecules. As a result, non covalent interactions involved in proteins play a crucial role, so changing the strength or nature of the interactions can induce changes on functionality.[2-5]

Among the different kinds of amino acids, the subset constituted by those carrying an aromatic group on their side chain exhibits specific characteristics that can be assigned to interactions involving the aromatic moieties.[5-9] As a consequence, phenylalanine, tyrosine, tryptophan and histidine are deeply involved in different recognition processes partially controlled by interactions with the aromatic units on their side chains.

Non covalent interactions involving aromatic units have been the subject of different studies which have allowed the identification of the different kinds of interaction an aromatic ring can establish.[5, 7, 10] Thus, an aromatic unit can be involved in one or more of the following interaction types: π···π; XH···π and ion···π. π···π stacking interactions play an important role in protein stability and function and, even though they are usually weak interactions, the combination of several stacking interactions can deeply affect the characteristics of the systems.[5, 7] The same could be said of the hydrogen bond-like interactions between X-H groups and aromatic rings, mostly involving OH, NH and CH bonds.[8, 10] Finally, the interaction between aromatic rings and charged species is controlled by the combination of electrostatic and induction contributions in cation···π contacts,[9, 11-14] whereas the role of induction is less significant when anions are involved.[14, 15] In any case, the interaction between a cation and a polarizable π cloud is usually strong in the gas phase, being a contact frequently observed in biological systems.[16-20] However, the strength of the cation···π interactions can be modulated by changes on the environment or on the nature of the aromatic ring involved.

Cation···π contacts do exist in a great variety of environments, and with very different degrees of exposure to solvent molecules.[9, 21] Therefore, a given cation···π contact can be buried into a hydrophobic pocket, thus showing a large interaction energy, or it can be fully exposed to the bulk solvent. In this latter case, the presence of solvent molecules competing with the aromatic unit in order to interact with the cation can weaken significantly the strength of the

cation···π contact.[22-33] On the other hand, this effect can also be induced by electron-donor groups or anions nearby.[15, 34] Finally, substitution on the aromatic ring also changes the characteristics of the cation···π interactions. A great deal of effort has been devoted to determine the effect of substituents onto the characteristics of the various types of interactions involving π clouds.[35-37] Recently, it has been suggested that most of substituent effects can be related to the local effect of the substituent, thus reducing the role of conjugation in changing the properties of the electron cloud in aromatic systems. Accordingly, the main effect of electron-donor or electron-acceptor substituents is the modulation of the molecular electrostatic potential over the aromatic ring.[35, 38, 39] The changes observed can be mostly described by the direct long-range electrostatic effect associated to the substituent unit.[40, 41] Also, there seems that substituents also locally modify dispersion interactions in substituted rings.[35, 41]

The presence of aromatic rings in the side chain of aromatic amino acids makes them amenable for establishing cation···π interactions. This is even more evident considering that another subset of amino acids can be in cationic form depending on the pH of the medium. Several studies have shown therefore that cation···π interactions involving side chains of aromatic and cationic amino acids are relatively common in proteins.[9, 16-20] Wu and McMahon have also shown that the presence of the aromatic side chain results on an enhancement of the interaction compared to non-aromatic amino acids, thus highlighting the role of the aromatic units on the interaction.[42] Many other studies have been devoted to determine the characteristics of the cation···π interaction, mainly in complexes formed with alkali cations,[43-50] though studies employing alkali-earth and transition-metal cations are available.[27, 51-54]

However, studies involving more complex cations are scarce.[28, 29, 55-57] Very recently, the authors have carried out different studies on the cation···π interactions involving more complex cations such as methylammonium,[25] pyrrolidinium[24] and guanidinium,[23, 31] also estimating the role of solvent molecules on the interaction. The results indicate that the interaction is still controlled by electrostatics, but with an increasing role of dispersion interactions, especially in structures showing stacked cation and aromatic ring. Also, the results obtained indicate that for aromatic amino acid···guanidinium complexes stacked structures are favored due to larger induction and dispersion interactions associated to the cation···π contact, which have to overcome the effort of folding the amino acid.[58]

Following this line of work, in the present work a study is performed on the interaction between aromatic amino acids and imidazolium cation ($C_3N_2H_5^+$). Imidazolium cation is a five membered planar aromatic cycle which can be formed by protonation of the amino acid histidine. There is some ambiguity on the protonation state of histidine depending on pH, so this amino acid can be found both in protonated and neutral form, this being one of the reasons why

histidine has not been often considered in studies on the interaction of aromatic amino acids with cations. The interaction of imidazolium with the DNA base pairs has been the subject of a research from Wetmore,[55] showing that both stacked and T-shaped structures are relevant in order to understand the characteristics of these systems. Besides being part of histidine, imidazolium cations are also important because the imidazolium ring constitutes the basic unit on most cations of ionic liquids. In the present work a detailed study of the interaction between imidazolium (Imz) cation and phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp) and histidine (His) is performed by using computational methods. The results obtained will help to shed light onto the behavior and properties of imidazolium···aromatic amino acid interactions.

8.2. Computational details

Considering whole amino acids poses a problem related to their conformational flexibility, which gives rise to a large amount of different minima. Therefore, a Multiple Minimum Monte Carlo conformational search has been carried out with the Macromodel program.[59] The MMFF force field was employed in the calculations, and all structures with relative energies below 10 kcal mol^{-1} respect to the most stable one were kept for further refinement. This procedure has been carried out for complexes of Imz with His, Phe, Tyr, and Trp, in both their neutral and zwitterionic forms. As a result of these conformational searches, a large number of minima has been found which were re-optimized at the M06-2X/6-31+G* level of calculation.[60] Once a stationary point was located, a frequency calculation is performed in order to ensure that it corresponds to a proper minimum. The topology of the electron density of the most relevant structures has been analyzed within the framework of the Quantum Theory of Atoms in Molecules (QTAIM) by using the AIM2000 package,[61, 62] and by using the Non covalent Interaction Index (NCI), which allows visualizing the favorable and unfavorable interactions based on the analysis of the reduced density gradient.[63, 64]

The interaction energies of the dimers were obtained by applying the supermolecule method, using the counterpoise procedure to avoid basis set superposition error.[65, 66] Therefore, interaction energies are obtained as:

$$\Delta E_{int} = E_{AB}^{complex}(AB) - E_{A}^{complex}(AB) - E_{B}^{complex}(AB), \qquad \text{(eq. 8.1)}$$

with terms in parentheses indicating the basis set and superscripts the geometry employed in the calculation. The energy associated to changes in the geometry of the fragments forming the dimer during complex formation has been included as a deformation energy term obtained as:

$$E_{def} = E^{complex}(A) + E_{B}^{complex}(B) - E^{isolated}(A) - E_{B}^{isolated}(B). \qquad \text{(eq. 8.2)}$$

Therefore, the final complexation energy is given by:

$$\Delta \mathrm{E}_{\mathrm{comp}} = \Delta E_{int} + E_{def}\,. \qquad \text{(eq. 8.3)}$$

Following this procedure, complexation energies were obtained at the M06-2X/6-31+G* level. However, in previous work it has been shown that even though M06-2X provides fairly good results, a proper description of the stabilities in this kind of systems requires reliable methods capable of appropriately describing all factors involved in the interaction. Therefore, complexation energies have been obtained by using the so-called MP2.X method proposed by Hobza.[67, 68]

First, MP2 values extrapolated to basis limit have to be obtained. Thus, an extrapolation to complete basis set limit is performed at the MP2 level by using the Halkier/Helgaker procedure with the aug-cc-pVTZ and aug-cc-pVQZ basis sets:[69, 70]

$$\Delta E^{X}_{corr,MP2} = \Delta E^{CBS}_{corr,MP2} + A\,X^{-3}. \qquad \text{(eq. 8.4)}$$

and the correlation term is combined with estimated HF energies extrapolated to complete basis set according to Karton-Martin two-point scheme:[71]

$$\Delta E^{X}_{HF} = \Delta E^{CBS}_{HF} + A(X+1)\exp\left(-9\sqrt{X}\right). \qquad \text{(eq. 8.5)}$$

With this extrapolation procedure it is assumed that errors due to basis set incompleteness are taking care of. However, there still remains the error associated to the method employed for electron correlation. The deficiencies on the *N*-electron model have been corrected by using MP3 results obtained with a small basis set and adequately scaled in order to reproduce CCSD(T) values as proposed by Hobza in the MP2.X method:

$$\Delta E^{CBS}_{MP2.X} = \Delta E^{CBS}_{MP2} + C\left(\Delta E^{smallbasis}_{MP3} - \Delta E^{smallbasis}_{MP2}\right). \qquad \text{(eq. 8.6)}$$

The empirical coefficient C depends on the basis set employed (in this work 6-31G*; C = 0.86).[67, 68]

Also, in order to obtain more information about the characteristics of the interaction, the interaction energies have been decomposed by applying the SAPT method with intramonomer correlation effects described at the DFT level (SAPT(DFT)).[72-74] For these calculations the PBE0 functional was used, involving a shift parameter obtained as the sum of the ionization potential and the energy of the highest occupied molecular orbital. Orbital energies and ionization potentials have been obtained by using the PBE0 functional with the aug-cc-pVDZ basis set. The

DFT-SAPT calculations were performed with the aug-cc-pVDZ basis set, employing density fitting with the corresponding basis sets. In order to provide a more accurate description, the dispersion contribution has been scaled by a factor of 1.193 as suggested by Rezak and Hobza.[75]

Finally, solvent effects upon the stability of the complexes studied have been estimated by using polarizable continuum model (PCM) calculations at the M06-2X/6-31+G* level.[76] The structures of the complexes have been re-optimized in the presence of water and their complexation energies estimated as indicated above, employing the basis set superposition error correction from a gas phase calculation with the same geometry.

All SAPT calculations have been performed with Molpro 2010.1.[77] M06-2X calculations have been carried out with Gaussian09,[78] whereas MP2 and MP3 calculations have been performed with Orca3.0.1.[79] In order to save computational time in MP2 calculations, the resolution of the identity approach has been employed both for the HF and correlation energies.

8.3. Results

The behavior of Phe, Tyr and Trp complexes is quite similar, whereas His behaves in a different way, so His complexes will be discussed separately from the rest of aromatic amino acids.

After the conformational search and subsequent re-optimization at the M06-2X/6-31+G* level of calculation, a large amount of possible minima has been found for each of the amino acids considered. Therefore, this set of minima has been reduced by selecting only a set constituted by the most stable ones as obtained at the M06-2X/6-31+G* level of calculation. Only this selected set is subjected to further higher level calculations, so a concern may arise as to whether significant structures are being skipped by this procedure. However, using an energy window of 10 kcal mol^{-1} it can be expected that most important minima are already included in the conformational search employing the MMFF force field. Besides, previous work has shown the good performance of the M06-2X/6-31+G* level of calculation in describing the relative stabilities of these complexes.[58] Finally, the subset of selected minima covers all structures within 2-3 kcal mol^{-1} above the global minima, so it can be considered that all structures accessible at room temperature are already taken into account.

8.3.1. Phe, Tyr and Trp complexes

Figure 8.1 shows the most stable minima found for complexes formed by Imz cation and Phe. Other minima found are listed in Appendix 5. Figure 8.1 shows the most stable minima sorted according to their relative stability as obtained with the MP2.X method. The MP2.X values together with results obtained with other computation levels are listed in Table 8.1. As in our

previous work on guanidinium complexes,[58] the different methods reproduce fairly well the order of stability of the different complexes, though several differences arise. The largest differences correspond to the relative stability of the zwitterionic/neutral complexes. The inclusion of the MP3 correction destabilizes in a larger extent neutral stacked complexes, so zwitterionic minima are favored.

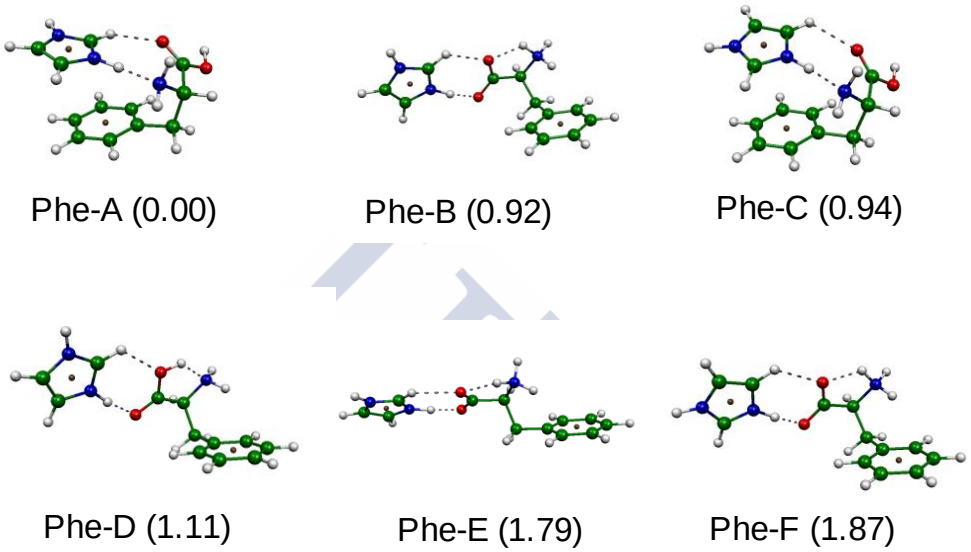


Figure 8.1. Most stable minima found for the complexes formed by Imz cation and Phe. Complexation energies at the MP2.X level in parentheses (kcal mol^{-1}).

The most stable structure corresponds to a minimum with the Imz cation oriented parallel over the phenyl ring forming a stacked parallel complex (Phe-A). The stability of this complex reaches -29.2 kcal mol^{-1} as obtained at the MP2.X level. Around 1 kcal mol^{-1} less stable three minima are obtained, corresponding to another arrangement of the parallel stacked structure (Phe-C) and to complexes with the amino acid in an extended conformation and the cation interacting with the carboxylic group. The difference between these two extended minima relies on the zwitterionic (Phe-B) or neutral form of the amino acid (Phe-D). In both parallel minima two hydrogen bonds are observed, corresponding to NH···N and CH···O contacts, but in Phe-A the CH group is the one between the nitrogens in the ring, whereas in Phe-C it corresponds to a different CH group of the cation.

Table 8.1. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the Imz···Phe complex. Last column lists deformation energies at the MP2/CBS level.

	M06-2X/6-31+G*	MP2/CBS	MP2.X	SAPT	$E_{def,MP2/CBS}$
Phe-A	-30.50	-31.90	-29.24	-29.48	8.23
Phe-B	-27.07	-28.38	-28.31	-26.47	20.07
Phe-C	-29.51	-31.14	-28.29	-28.57	8.29
Phe-D	-27.54	-28.48	-28.13	-27.92	1.68
Phe-E	-26.54	-27.47	-27.45	-25.68	22.55
Phe-F	-25.94	-27.76	-27.37	-25.83	20.74
Phe-G	-26.78	-27.61	-27.25	-27.21	2.40
Phe-H	-26.41	-27.42	-26.98	-27.06	2.56
Phe-I	-27.44	-27.96	-26.68	-27.29	7.05
Phe-J	-25.50	-26.88	-26.55	-25.10	23.53
Phe-K	-27.41	-27.45	-26.54	-26.90	7.40
Phe-L	-26.56	-27.64	-26.44	-26.22	5.49

Other different minima have been found corresponding to different conformations of the amino acid and orientations of the Imz ring. Overall, most structures can be classified as belonging to four general types, which are also found in complexes with other amino acids. Therefore, minima have been found corresponding to parallel stacked arrangements of the cation over the aromatic ring (P), such as Phe-A or Phe-C. Structures with the amino acid in extended conformation both in zwitterionic (Z) or neutral (E) form are also located, with the zwitterionic minima usually being slightly more stable than the neutral ones. Finally, minima with the Imz cation interacting perpendicularly (T) to the aromatic ring are also found. This kind of structures is not among the most stable minima of Phe···Imz, the most closely related one (Phe-L) appearing 2.8 kcal mol^{-1} above the most stable minimum.

It is worth noting the role played by conformational changes of the amino acids in the stability of the complexes. As observed in Table 8.1, the energy cost of changing the amino acid structure as to adopt the geometry on the complex is normally large, reaching around 8 kcal mol^{-1} in the case of the parallel stacked complexes. As expected, extended minima show smaller deformation energies around 2 kcal mol^{-1}, whereas in the case of zwitterions, the cost of proton transfer is

huge, reaching 20 kcal mol^{-1}. As indicated in previous work,[58] the contribution of deformation is essential in order to understand the stability in these complexes. The same behavior is observed in Tyr and Trp complexes, with stacked minima being penalized by conformational changes.

The main features of these minima can be observed by means of their molecular graphs as obtained by the QTAIM analysis of their electron density.[61, 62] These molecular graphs are shown in Figure 8.2, together with values of the density (times 10^3) at the bond critical points (BCP) of non covalent contacts.

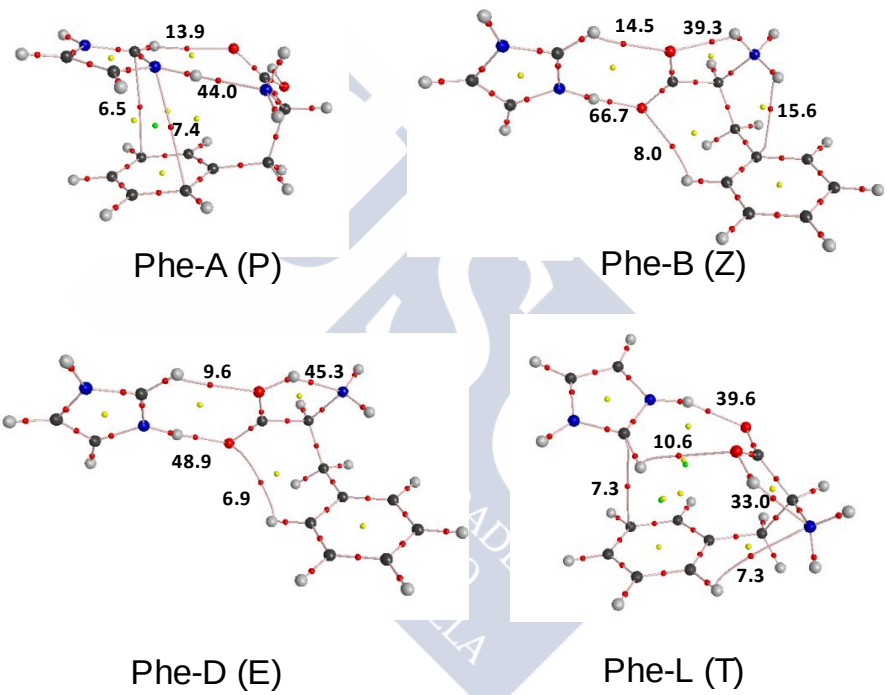


Figure 8.2. Molecular graphs obtained for selected complexes of Imz and Phe. Numbers correspond to electron density at the bond critical point (a.u. /10^{-3}). Bond critical points in red, ring critical points in yellow and cage critical points in green.

The different stabilizing non covalent interactions can be easily seen. In Phe-A, corresponding to a parallel stacked minima, there are two hydrogen bonds formed between the Imz cation and the carboxylic/amino group. The CH···O hydrogen bond shows a density at the BCP of 0.014 a.u., whereas the NH···N contact displays a density reaching 0.044 a.u., thus showing the greater strength of this latter hydrogen bond. Besides these hydrogen bonds, two bond paths connecting

the Imz ring and the phenyl ring are observed, with densities around 0.006 a.u.. These two bond paths are related to the favorable stacking interaction of the rings in parallel minima. As regards extended minima, in zwitterionic Phe-B the interaction is governed by two strong hydrogen bonds to the carboxylate group (again the CH···O weaker than the NH···O one). These are the only intermolecular contacts, but there are several favorable intramolecular contacts between different regions of the amino acid. The strongest one corresponds to the intramolecular hydrogen bond between the ammonium and carboxylate group (density 0.039 a.u.). Also, a weak interaction between the carboxylate oxygen and hydrogen atoms in the phenyl ring is observed. Finally, there is a bond path related to the interaction between the ammonium group and the phenyl ring, with density at the BCP around 0.016 a.u.. The pattern obtained for the neutral form Phe-D is almost identical, the only differences being the lack of bond path from the amino group to the phenyl ring, and the location of the hydrogen atom in the OH···N hydrogen bond. However, as indicated by the NCI plot (see Appendix 5), there is a favorable interacting region between the amino group and the ring, so the pattern is almost identical in zwitterionic and neutral forms. T-shaped minima occupy an intermediate place between the other minima considered. There are two hydrogen bonds to the carboxylic oxygens as in extended minima, whereas an intramolecular O-H···N hydrogen bond helps stabilizing the structure. The most remarkable feature is the interaction of the Imz ring with the aromatic ring by means of a CH group.

When the complex is formed with Tyr or Trp, the behavior quite resembles that observed for Phe complexes, but several differences arise. Figure 8.3 shows the most stable minima found for complexes with Tyr and Trp, whereas Table 8.2 lists the complexation energy values obtained with the different methods. It can be observed that in the case of Tyr complexes, there are several structures that only differ on the orientation of the hydrogen of the hydroxyl group of phenol, thus leading to minima with pretty similar stabilities. As observed in other cases,[80, 81] Tyr complexes are grouped by pairs differing in the position of the hydroxyl hydrogen. Besides this feature of Tyr, the minima resemble those found for Phe, most structures belonging to one of the four groups considered above. However, in the case of Tyr complexes, true T-shaped minima are found. The presence of the hydroxyl oxygen makes possible for the cation to interact by means of a hydrogen bond with the hydroxyl group while forming other hydrogen bonds with the carboxyl/amino groups. Thus, Tyr-C is a typical T-shaped minimum, with the Imz cation perpendicular to the phenyl ring of the amino acid.

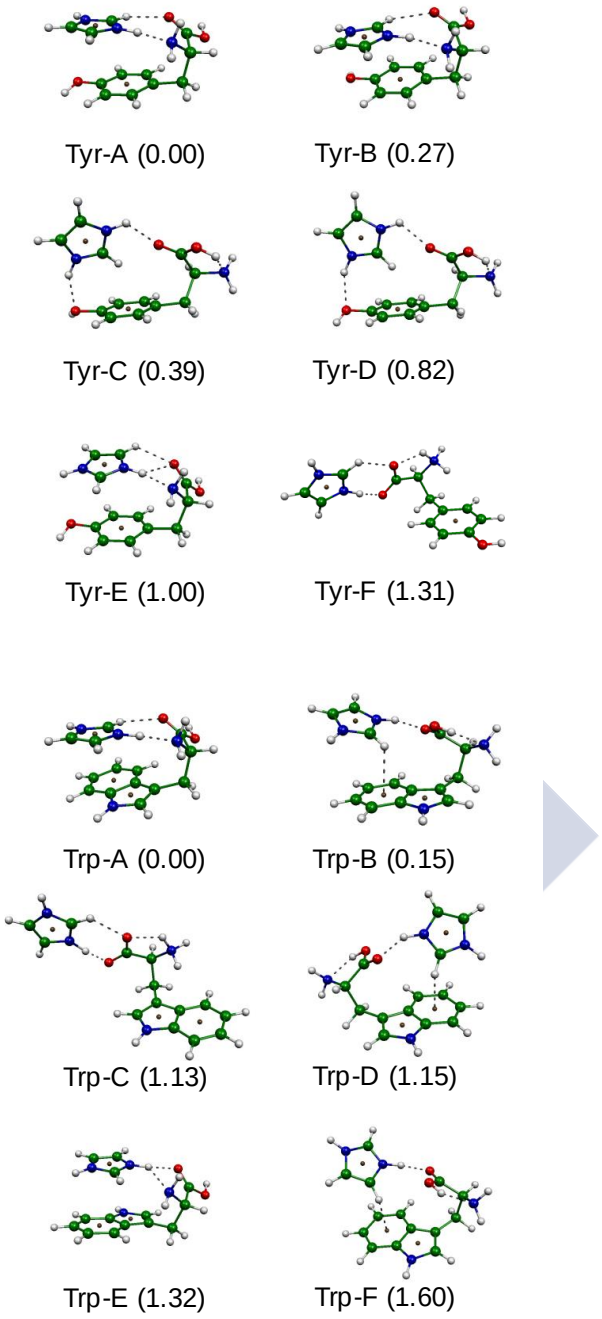


Figure 8.3. Most stable minima found for the complexes formed by Imz cation with Tyr and Trp. Complexation energies at the MP2.X level in parentheses (kcal mol^{-1}).

Table 8.2. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of Imz cation and Tyr and Trp complexes. Last column lists deformation energies at the MP2/CBS level.

	M06-2X/6-31+G*	MP2/CBS	MP2.X	SAPT	$E_{def,MP2/CBS}$
Tyr-A	-31.38	-32.90	-29.77	-29.95	9.31
Tyr-B	-31.28	-32.68	-29.50	-29.52	9.78
Tyr-C	-30.23	-30.64	-29.38	-29.04	7.02
Tyr-D	-29.85	-30.14	-28.94	-28.58	6.99
Tyr-E	-30.66	-32.26	-28.77	-28.92	9.98
Tyr-F	-27.26	-28.64	-28.46	-26.90	20.54
Tyr-G	-27.69	-28.73	-28.27	-27.93	2.10
Tyr-H	-27.10	-28.46	-28.27	-26.71	20.20
Tyr-I	-29.32	-30.24	-28.24	-27.84	7.11
Tyr-J	-27.21	-28.42	-27.86	-27.64	2.25
Tyr-K	-27.15	-28.18	-27.72	-27.38	1.81
Tyr-L	-26.81	-27.71	-27.68	-25.70	19.81
Trp-A	-33.03	-35.28	-31.42	-31.60	10.65
Trp-B	-31.36	-33.03	-31.27	-30.98	5.66
Trp-C	-29.02	-30.44	-30.30	-28.49	20.18
Trp-D	-29.61	-31.71	-30.27	-30.15	7.28
Trp-E	-31.53	-33.57	-30.10	-30.01	9.26
Trp-F	-29.45	-31.68	-29.82	-29.62	5.74
Trp-G	-28.74	-29.90	-29.60	-27.92	20.98
Trp-H	-29.37	-31.95	-29.48	-29.28	5.92
Trp-I	-28.29	-29.49	-29.36	-27.86	23.59
Trp-J	-27.80	-29.72	-29.25	-27.80	21.26
Trp-K	-30.51	-33.21	-28.95	-29.05	10.91
Trp-L	-30.03	-32.39	-28.60	-28.85	11.95

As in Phe complexes, the most stable minimum shows a parallel stacked structure, with complexation energy of -29.8 kcal mol^{-1}, very similar to that found in Phe. Barely 0.4 kcal mol^{-1} less stable is Tyr-C, with the cation perpendicular to the ring. Extended structures are around 1.3 kcal mol^{-1} higher in energy as in Phe complexes. The behavior of Trp complexes is even more similar to that found for Phe ones. The most stable minimum again corresponds to a parallel stacked structure with complexation energy reaching -31.4 kcal mol^{-1}, so complexes with Trp are more stable than with any of the other two amino acids. However, the T-like structure Trp-B is only 0.2 kcal mol^{-1} above the most stable structure. Contrary to Phe T complex, this structure is quite stable with Trp as a consequence of the larger aromatic cloud of Trp allowing an easier contact with the cation.

Figure 8.4 shows the molecular graphs for prototypical complexes formed with Trp and Tyr. No neutral extended minima are shown since the behavior is similar to zwitterionic ones, and are usually less stable. The characteristics of the molecular graphs in P minima are similar to those discussed above for Phe complexes. Besides the two hydrogen bonds to the carboxyl oxygen and amino nitrogen, up to three paths connecting atoms in the indole ring are observed, corresponding to the stacking of the planar cation over indole ring. The only different behaviour is observed for T complexes, specifically in Tyr ones. Tyr complex is typically T-shaped, so the cation forms two hydrogen bonds with the carbonyl oxygen and the hydroxyl oxygen by means of the two NH groups, with density at the BCP around 0.02 a.u.. Besides, the central CH group exhibits two paths connecting the carbonyl oxygen and the carbons in the phenyl ring. These features can be also observed by using the NCI plots, as displayed in Appendix 5.[63, 64]

8.3.2. His complexes

In the present section the results obtained for His complexes are discussed. The behavior of His complexes is further complicated by the possibility of tautomerism between different forms of His depending on the nitrogen atom on the imidazole ring which carries a hydrogen atom.[43] Very recently, the structure of gas phase His has been elucidated by Alonso et al.,[82] with the result of the most stable tautomer being the one with the hydrogen atom on the nitrogen further from the asymmetric carbon atom (**A** in Figure 8.5, N_τ-H). The energy difference between the two most stable structures amounts to around 0.7 kcal mol^{-1} at the MP2/CBS level in agreement with the experimental estimates. Therefore, conformational searches have been carried out as outlined above employing both forms of His, in neutral and zwitterionic form. As a consequence the number of minima increases significantly, leading to different structures than those observed for other aromatic amino acids.

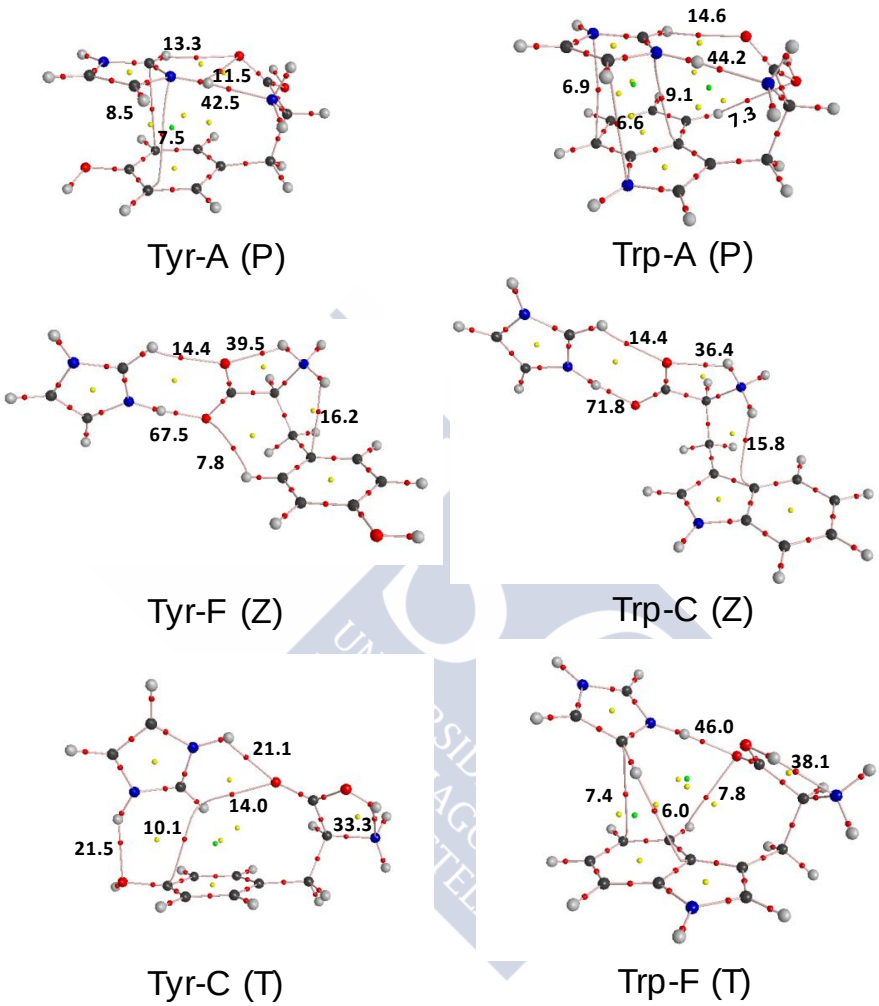
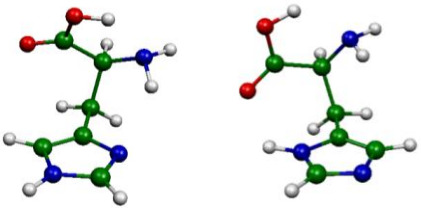


Figure 8.4. Molecular graphs obtained for selected complexes of Imz with Tyr and Trp. Numbers correspond to electron density at the bond critical point (a.u. /10^{-3}). Bond critical points in red, ring critical points in yellow and cage critical points in green.



A (0.00) B (0.68)

Figure 8.5. Two most stable minima of His in the gas phase. Relative energies at the MP2/CBS level in parentheses (kcal mol^{-1}).

Figure 8.6 shows a subset of the most stable minima found for complexes of His and Imz cation, with complexation energy values listed in Table 8.3. The most striking effect, considering the minima of the different complexes studied in this work, is that His behaves completely different from the other aromatic amino acids. As observed in Figure 8.6, there are no minima displaying cation···π interactions, the non covalent interaction being governed by different sets of hydrogen bonds. Another relevant aspect when considering the minima of His···Imz complexes is that most stable minima correspond to zwitterionic forms of the amino acid. In fact, the most stable neutral complex is quite high in energy, being 4.6 kcal mol^{-1} less stable than His-A. Also, there are plenty of structures where a proton transfer between the cation and the amino acid has taken place, as it can be observed in His-D. Considering that the nature of the cation and the aromatic ring is similar, both being based in imidazole, the transfer or even sharing of one proton between rings is feasible. This proton transfer can also happen between Imz and the carboxylate group, leading to another subfamily of structures.

It is also worth noting that even though the energy difference between the two most stable tautomers of His is quite small, most stable complexes with Imz are formed with tautomer **A** (Figure 8.5, N_τ-H). The most stable complex formed with tautomer **B** (Figure 8.5, N_π-H) and without a proton being transferred to the amino acid is 8 kcal mol^{-1} less stable than His-A. The reason for this behavior can be observed in the minima of Figure 8.6. It is clearly seen that in all minima in Figure 8.6, and in many others, there are hydrogen bonds formed with the lone pair of a nitrogen atom of imidazole. The position of this lone pair on tautomer **A** (N_π) is perfectly placed so hydrogen bonds can be easily formed with the ammonium group when the amino acid is in zwitterionic form (or even neutral).

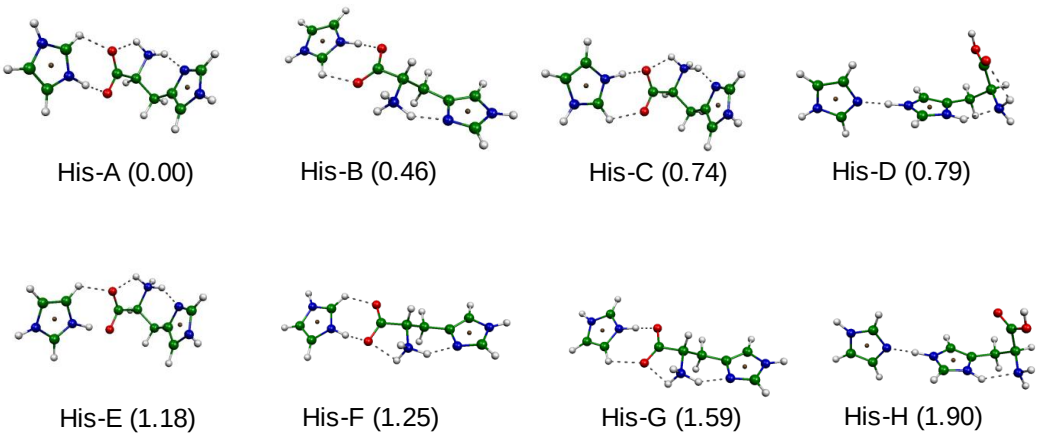


Figure 8.6. Most stable minima found for the complexes formed by Imz cation and His. Complexation energies at the MP2.X level in parentheses (kcal mol^{-1}).

Table 8.3. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the Imz···His complex. Last column lists deformation energies at the MP2/CBS level.

	M06-2X/6-31+G*	MP2/CBS	MP2.X	SAPT	$E_{def,MP2/CBS}$
His-A	-34.35	-34.90	-34.78	-33.64	19.36
His-B	-34.52	-34.46	-34.32	-33.22	21.18
His-C	-33.85	-34.06	-34.04	-32.49	18.48
His-D	-32.34	-32.87	-33.99	-44.46	79.67
His-E	-33.02	-34.07	-33.60	-32.88	21.11
His-F	-33.99	-33.56	-33.53	-32.02	20.00
His-G	-33.30	-33.65	-33.19	-32.47	22.92
His-H	-31.34	-31.72	-32.88	-43.42	82.80
His-I	-31.54	-31.78	--32.68	-43.06	78.81
His-J	-32.05	-32.80	-32.38	-31.09	19.52
His-K	-31.07	-30.20	-32.16	-36.87	89.81
His-L	-30.19	-29.26	-31.14	-35.12	87.34

On the other hand, the lone pair (N_τ) in tautomer **B** is further away of the amino/carboxylic group, with one or two XH groups in between. Therefore, the formation of hydrogen bonds with N_τ is hindered and in consequence most complexes formed with tautomer **B** are less stable. In any case, it can also be observed that minima as His-D or His-H, where a proton transfer to the ring has taken place, are formed in principle from tautomer **B**, so once the imidazole ring of the amino acid is protonated the N_π-H group forms hydrogen bonds with the amino nitrogen.

Therefore, it can be stated that, overall, no parallel stacked or cation···π complexes are formed between Imz and His, a totally different behavior as compared with the other aromatic amino acids. The most stable tautomer of His forms more stable complexes since it favors the formation of hydrogen bonds between the nitrogen in the ring and the carboxylic/amino group. Also, the chain-like disposition of the hydrogen bonds in the most stable minima (see for example His-A of His-C) also favors the formation of zwitterionic structures, which are by far the most stable ones even in the gas phase, reaching -34.8 kcal mol^{-1}, the strongest among the four aromatic amino acids, an order of stability similar to that found in Na^+ complexes.[43, 48] While Phe, Tyr and Trp behave as expected, increasing the strength of the interaction as the size of the aromatic unit grows, the smaller His departs from this trend since no cation···π contact is formed but a series of stronger hydrogen bonds.

The main features of the different kinds of complexes can be observed in Figure 8.7. The most stable minimum His-A shows a series of strong hydrogen bonds with high densities at the BCP. The NH···O hydrogen bond between Imz and the carboxyl oxygen is in fact very strong (density at the BCP 0.076 a.u.) suggesting proton sharing to some extent, in accordance with the elongation on N-H bond distance. His-D and His-K are minima with a proton transferred to the amino acid. In His-D the proton is transferred to the imidazole ring, whereas in His-K the process is achieved with participation of the carboxylate group. His-P is the most stable minima found with non-zwitterionic amino acid. Its features resemble those observed for extended minima in the other amino acids. There are hydrogen bonds between the carboxylic oxygens and the Imz cation, plus a strong hydrogen bond between the carboxylic proton and the lone pair on the imidazole ring. His-R is another minimum with neutral His, with the Imz cation interacting with the carboxylic oxygen while simultaneously forming a N-H···N hydrogen bond with the lone pair in the imidazole ring (N_π). A closely related structure is His-AL, where in order to form the N-H···N hydrogen bond (N_τ) the cation is oriented perpendicular to the imidazole ring, in a similar way as observed for the T-shaped minimum in Tyr complexes.

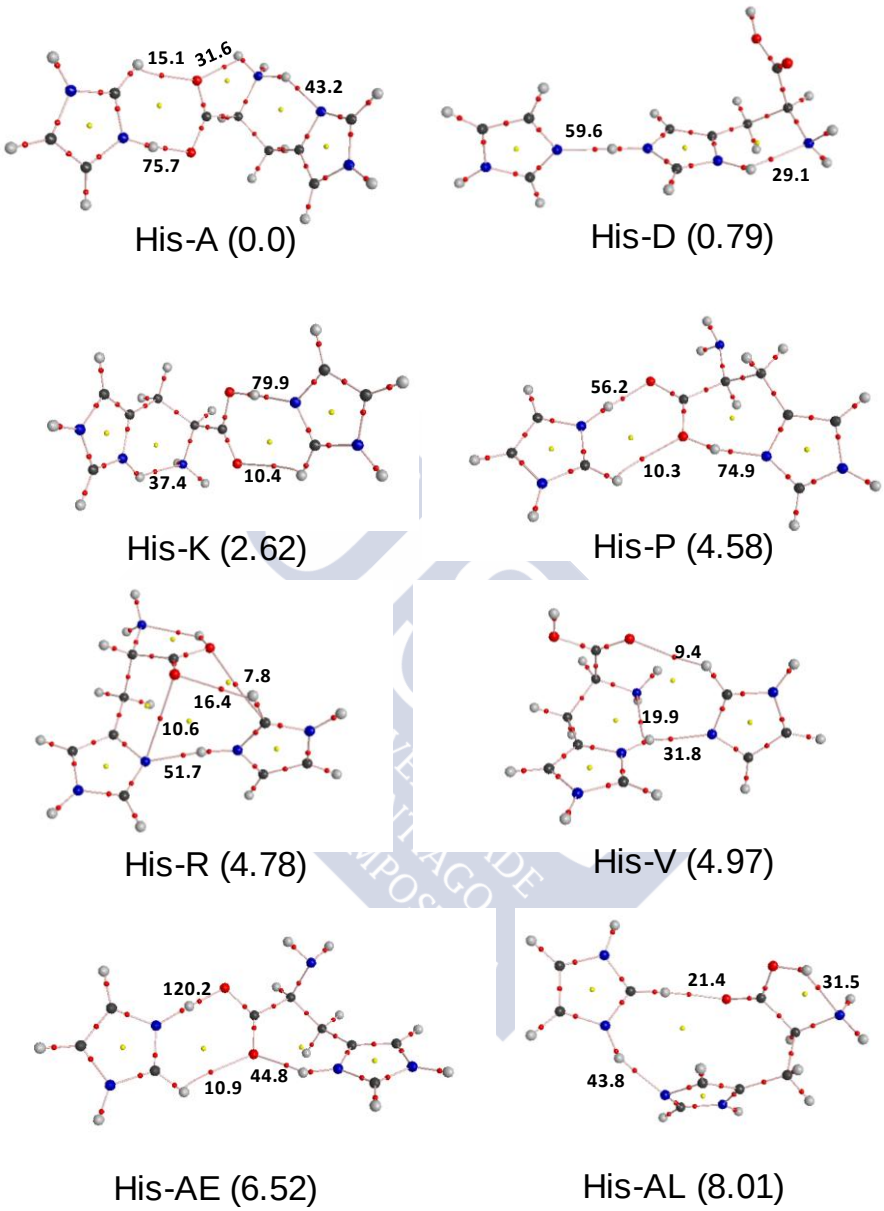


Figure 8.7. Molecular graphs obtained for selected complexes of Imz and His. Numbers correspond to electron density at the bond critical point (a.u. $/10^{-3}$). Relative energies at the MP2.X level in parentheses (kcal mol^{-1}). Bond critical points in red, ring critical points in yellow and cage critical points in green.

8.3.3. SAPT(DFT)

Figure 8.8 shows the results for the SAPT(DFT) decomposition in selected complexes of Imz with the aromatic amino acids considered in this work. Four contributions will be considered: electrostatic (E_{ele}= E_{1pol}), repulsion (E_{rep}= E_{1exch}), induction (E_{ind}= $E_{2ind-exch}$ + E_{2ind} + δ_{HF}), and dispersion (E_{disp}= $E_{2disp-exch}$ + E_{2disp}). Also included is the deformation energy obtained at the MP2/CBS level of calculation, as an indicative of the role played by geometry changes on the energy of the systems. This final term is included because SAPT(DFT) only gives information about the interaction at a given geometry. Also, since SAPT(DFT) requires identifying which atoms belong to each fragment, His complexes with proton transfer are not presented.

Globally it can be observed that, as expected, the interaction is governed by large electrostatic contributions, due to the presence of the cation, together with large repulsion terms consequence of the proximity of the two fragments. Apart from these terms, the following most important term comes from induction, as a consequence of the effect of a cation which, even though it is not especially polarizing due to its complex structure and delocalization of the positive charge, it is still capable of deforming significantly the charge distribution of the amino acid. Dispersion effects are usually smaller, though still significant, as a consequence of the larger size of the cation as compared, for example, with alkali cations, where dispersion effects are negligible.[12, 83] Even though the amino acids are quite flexible, so the geometry can change quite easily in order to prepare for a better interaction with the cation, deformation effects can be large as already pointed out in guanidinium complexes.[58] Since deformation effects are larger than the relative energy differences among minima, they are crucial on determining the final stability order of the different structures of the complexes.

Considering the different structures, the results are similar for Phe, Tyr and Trp complexes. Parallel minima show large electrostatic contributions due to the simultaneous interaction of the cation with the carboxylic group and the aromatic ring, matched with an equally large repulsion term. Dispersion and induction contributions are the largest among neutral minima as a consequence of the cation···π contact, overcoming the deformation cost in order to fold the amino acid. The favorable balance between these opposite terms is the main responsible of the parallel stacked structures being the most stable minima found. The electrostatic contribution decreases in T-shaped structures despite the cation also interacting with both the ring and the carboxylic/amino group, even when a hydrogen bond is formed with the hydroxyl group in Tyr complex. In extended minima the electrostatic contribution decreases even further because now the cation cannot interact with the aromatic ring, but only with the carboxylic/amino group of the amino acid. Finally, in complexes with the zwitterionic amino acid, the electrostatic contribution is the largest, as it takes place between the cation and the anionic parts of the amino acid.

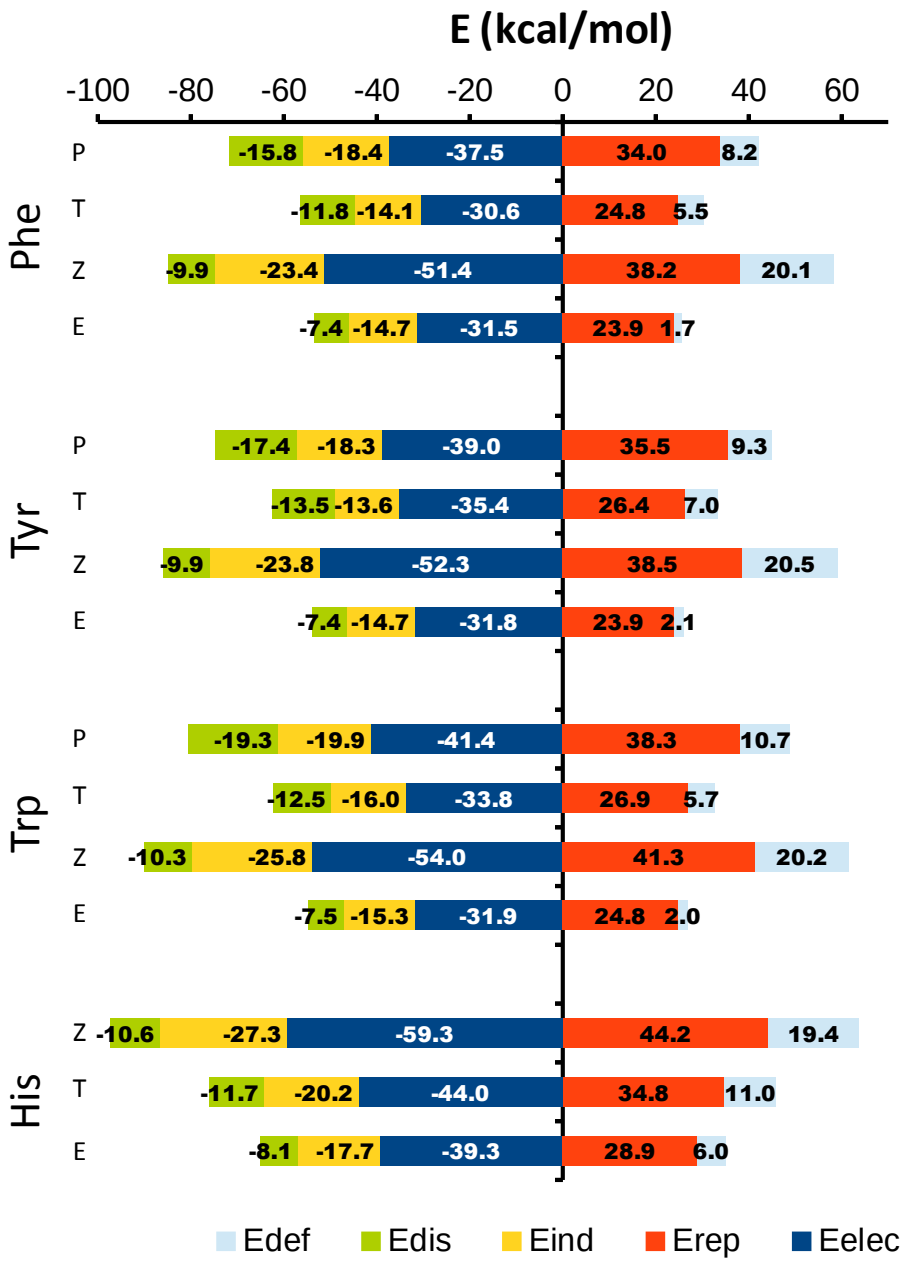


Figure 8.8. SAPT(DFT) partitioning of the complexation energy in typical minima. The deformation contribution is obtained at the MP2/CBS level.

Considering the differences among complexes of the same kind formed with the different amino acids, it can be observed that in all structures the pattern is similar. The behavior of extended and zwitterionic minima is similar in all amino acids, since the aromatic rings do not directly participate on the interaction, which is controlled by the two hydrogen bonds formed with the carboxylic/carboxylate oxygens. In parallel minima the aromatic rings participate on the interaction, so all contributions increase their intensity as the size of the aromatic ring increases, though differences between Phe and Tyr complexes are negligible in some cases. In summary, the preference for cation···π minima is directly related to larger contributions in all stabilizing terms, especially dispersion and induction, which overcome the repulsive and deformation barriers.

In His complexes, the behavior is different due to the different structures of the minima found. In any case, the most stable minima corresponds to a zwitterionic complex, so the energy contributions display a similar pattern as that observed in the rest of zwitterionic minima. Therefore, there are large electrostatic and induction contributions, which are in fact larger than in any of the other complexes. This large electrostatic and induction contributions could be related with the chain-like hydrogen bond network formed in His-A, making the complexation process the most favorable among the different aromatic amino acids.

8.3.4. Solvent effects

In the previous sections, the behavior of the complexes formed by Imz and the aromatic amino acids has been analyzed in the gas phase. However, it is evident that the presence of the solvent can affect the properties of the complexes. In order to test the effect of water upon the characteristics of the complexes, calculations have been carried out modeling the solvent by means of a polarizable continuum model. This way, the effect of the bulk solvent upon the properties of the complexes can be estimated, though short-range specific effects are not properly described. Figure 8.9 shows the most stable structures found for complexes in water with each of the amino acids considered in this work.

The most stable structures still correspond to those observed in the gas phase. Phe most stable minimum is a parallel stacked structure, though extended zwitterionic ones are very close in energy. It has to be taken into account that M06-2X/6-31+G* penalizes zwitterionic structures as already observed in the gas phase results. Therefore it could be expected the most stable minima to correspond to zwitterionic extended structures, as it is the case of Tyr and Trp complexes. Therefore, no cation···π contact would survive in bulk solvent, perhaps with the exception of Trp complexes (the most stable zwitterion is not fully extended, partially allowing contact with the aromatic moiety).

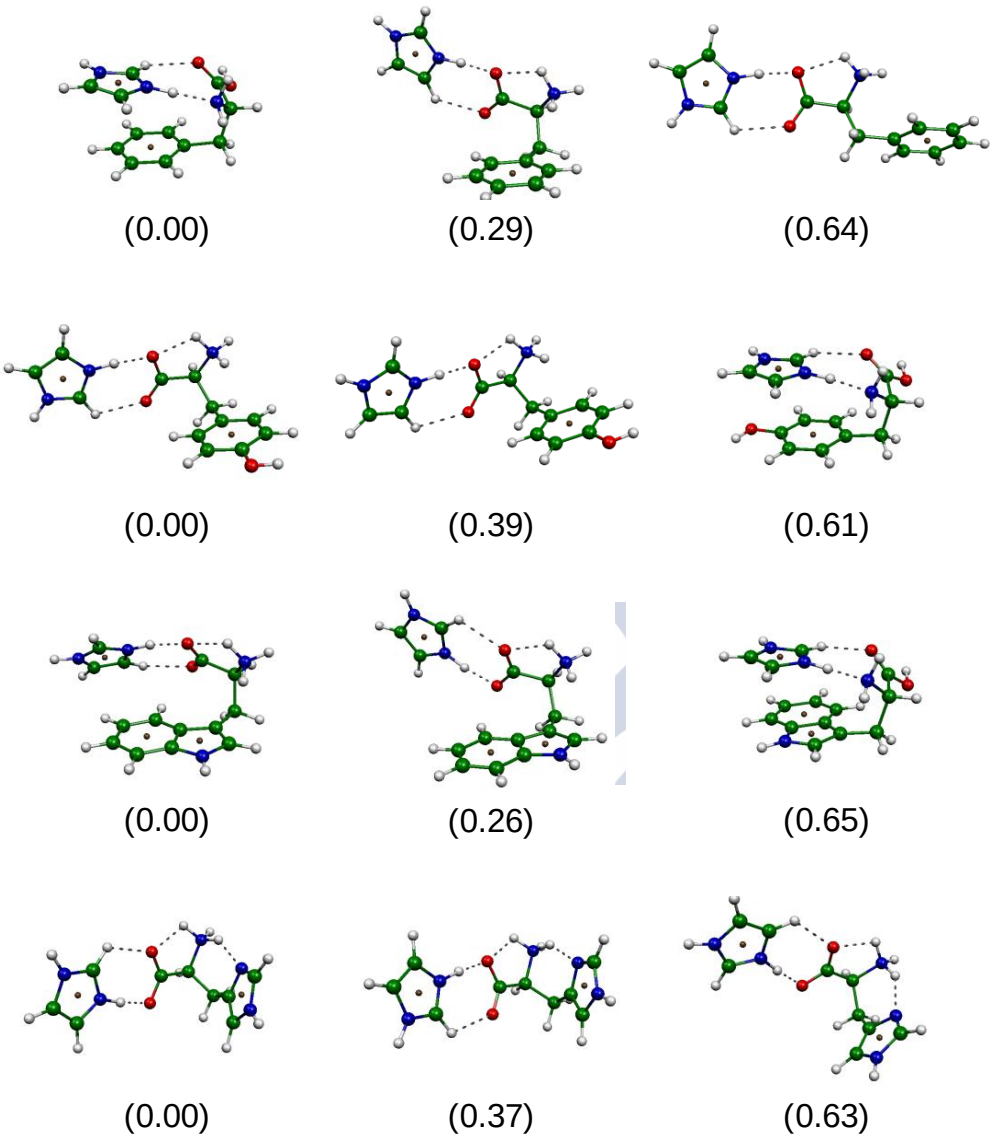


Figure 8.9. Most stable minima found in water as modeled with PCM. In parentheses relative complexation energies in kcal mol^{-1} at the M062X/6-31+G* level. Complexation energies for the most stable minima amount to -10.6 kcal mol^{-1}, -11.5 kcal mol^{-1}, -12.4 kcal mol^{-1},and -13.1 kcal mol^{-1} for Phe, Tyr, trp and His, respectively.

.

In any case, depending on the degree of exposure to solvent, the system would evolve from parallel stacked cation···π minima to zwitterionic extended ones in order to place the cation in the bulk of the solvent. His most stable complexes, as in the gas phase, correspond to zwitterionic structures held by a series of hydrogen bonds, so neither in the gas phase or bulk solvent His will form cation···π contacts

8.4. Conclusions

A detailed computational study has been carried out in order to reveal the characteristics of the interaction between Imz cation and aromatic amino acids. Phe, Tyr and Trp most stable minima correspond to structures with the Imz cation hydrogen bonded to the carboxylic/amino group of the amino acid, simultaneously establishing a cation···π contact with the aromatic units of the amino acids. On the other hand, in His···Imz most stable complexes the amino acid adopts a zwitterionic form, with the cation hydrogen bonded to the carboxylate group, and His in an extended conformation forming a series of intramolecular hydrogen bonds. The complexation energies obtained at the MP2.X level of calculation amount to -29.2, -29.8, -31.4, and -34.8 kcal mol^{-1} for Phe, Tyr, Trp, and His, respectively. The interaction is stronger as the size of the aromatic unit increases as a consequence of cation···π contacts, with the exception of His where the interaction is strongest due to hydrogen bonds involving the nitrogen atoms in the imidazole ring.

Most minima found can be classified as belonging to one of the following four prototypes: parallel stacked, extended zwitterion, extended neutral and T-Shaped. These structures are quite close in energy in Phe, Tyr and Trp complexes, while for His, the extended zwitterionic structures are by far the most stable ones. No parallel stacked minima or cation···π contact has been observed among the most stable minima found for His complexes. Including the solvent does not seem to introduce large changes, apart from favoring even more zwiterionic minima.

The analysis of the different contributions to the interaction reveals that it is controlled by a combination of electrostatic and induction contributions. As already observed in guanidinium complexes, the parallel stacked minima are the most stable ones in Phe, Tyr and Trp complexes due to large contributions from dispersion and induction capable of overcoming the energy cost for folding the amino acid. Complexes formed with the amino acid as zwitterion show the largest interaction, but this is cancelled to a great extent by the energy cost of proton transfer. In the case of His complexes, the large interaction is capable of overcoming the cost of proton transfer, so zwitterionic complexes are more stable than any of them in neutral form. As a consequence, there are no cation···π contacts in His complexes, so the role of dispersion is smaller and the complexes are more clearly controlled by induction and electrostatics.

In any case, it is worth noting the crucial role conformational changes play in the final stability of these complexes. Amino acids tend to adopt extended conformations, thus favoring extended minima. The formation of parallel stacked complexes leads to larger interactions which have to compensate for the energy cost needed for folding the structure. The same happens in zwitterionic complexes, with huge cation···carboxylate interactions overcoming the large cost of proton transfer.

Overall, the results clearly indicate that whereas Phe, Tyr and Trp complexes can show parallel structures in competition with similarly stable zwitterionic ones, His only shows zwitterionic minima, with stability even larger than any of the other aromatic amino acids despite lacking participation of the π cloud in the interaction.

8.5. References

(1) A. V. Finkelstein, O. B. Ptitsyn, *Protein Physics: A Course of Lectures*, Academic Press, Amsterdam, **2002**.

(2) P. Hobza, R. Zaradnik, *Intermolecular complexes: the role of van der Waals systems in physical chemistry and the biodisciplines*, Elsevier, Amsterdam, **1988**.

(3) P. Hobza, K. Müller-Dethlefs, *Non covalent Interactions*, The Royal Society of Chemistry, Cambridge, **2009**.

(4) H.-J. Schneider, A. K. Yatsimirski, *Principles and methods in supramolecular chemistry*, Wiley, Chichester, **2000**.

(5) L. M. Salonen, M. Ellermann, F. Diederich. *Angew. Chem. Int. Ed.* **2011**, *50*, 4808-42.

(6) E. A. Meyer, R. K. Castellano, F. Diederich. *Angew. Chem. Int. Ed.* **2003**, *42*, 1210-1250.

(7) M. L. Waters. *Biopolymers (Peptide Science)* **2004**, *76*, 435-445.

(8) M. Nishio. *Phys. Chem. Chem. Phys.* **2011**, *13*, 13873-900.

(9) A. S. Mahadevi, G. N. Sastry. *Chem. Rev.* **2013**, *113*, 2100-38.

(10) S. Tsuzuki, T. Uchimaru. *Curr. Org. Chem.* **2006**, *10*, 745-762.

(11) D. A. Dougherty. *Acc. Chem. Res.* **2013**, *46*, 885-93.

(12) I. Soteras, M. Orozco, F. J. Luque. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2616-2624.

(13) J. A. Carrazana-Garcia, J. Rodriguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2011**, *115*, 2774-2782.

(14) A. Campo-Cacharrón, E. M. Cabaleiro-Lago, J. Rodríguez-Otero. *J. Comput. Chem.* **2014**, *35*, 1533-1544.

(15) D. Kim, E. C. Lee, K. S. Kim, P. Tarakeshwar. *J. Phys. Chem. A* **2007**, *111*, 7980-6.

(16) J. P. Gallivan, D. A. Dougherty. *Proceedings of the National Academy of Sciences of the United States of America* **1999**, *96*, 9459-64.

(17) N. Zacharias, D. A. Dougherty. *Trends Pharmacol. Sci.* **2002**, *23*, 281-7.
(18) J. C. Ma, D. A. Dougherty. *Chem. Rev.* **1997**, *97*, 1303-1324.
(19) N. S. Scrutton, A. R. Raine. *Biochem. J* **1996**, *319*, 1-8.
(20) D. A. Dougherty. *J. Nutr.* **2007**, *137*, 1504S-1508S.
(21) J. P. Gallivan, D. A. Dougherty. *J. Am. Chem. Soc.* **2000**, *122*, 870-874.
(22) A. Campo-Cacharron, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Theor. Chem. Acc.* **2012**, *131*, 1-13.
(23) A. Rodríguez-Sanz, E. M. Cabaleiro-Lago, R.-O. J. *J. Mol. Model.* **2014**, *20*, 2209/1-2209/10.
(24) A. A. Rodriguez-Sanz, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Org. Biomol. Chem.* **2014**, *12*, 2938-2949.
(25) A. A. Rodriguez-Sanz, J. Carrazana-Garcia, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *J. Mol. Model.* **2013**, *19*, 1985-1994.
(26) M. Remko, S. Soralova. *J. Biol. Inorg. Chem.* **2012**, *17*, 621-630.
(27) M. Remko, D. Fitz, R. Broer, B. M. Rode. *J. Mol. Model.* **2011**, *17*, 3117-3128.
(28) Y. Xu, J. Shen, W. Zhu, X. Luo, K. Chen, H. Jiang. *J. Phys. Chem. B* **2005**, *109*, 5945-5949.
(29) C. Adamo, G. Berthier, R. Savinelli. *Theor. Chem. Acc.* **2004**, *111*, 176-181.
(30) N. J. Singh, S. K. Min, D. Y. Kim, K. S. Kim. *J. Chem. Theory Comput.* **2009**, *5*, 515-529.
(31) E. M. Cabaleiro-Lago, J. Rodríguez-Otero, Á. Peña-Gallego. *J. Chem. Phys.* **2011**, *135*, 214301/1-214301/9.
(32) J. S. Rao, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2009**, *113*, 7225-7236.
(33) A. S. Reddy, H. Zipse, G. N. Sastry. *J. Phys. Chem. B* **2007**, *111*, 11546-11553.
(34) J. A. Carrazana-Garcia, J. Rodriguez-Otero, E. M. Cabaleiro-Lago. *J. Phys. Chem. B* **2012**, *116*, 5860-5871.
(35) S. E. Wheeler. *Acc. Chem. Res.* **2013**, *46*, 1029-38.
(36) C. A. Hunter, K. R. Lawson, J. Perkins, C. J. Urch. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651.
(37) C. A. Hunter, J. K. M. Sanders. *J. Am. Chem. Soc.* **1990**, *112*, 5525.
(38) S. E. Wheeler, K. N. Houk. *J. Phys. Chem. A* **2010**, *114*, 8658-64.
(39) S. E. Wheeler, K. N. Houk. *J. Am. Chem. Soc.* **2009**, *131*, 3126-7.
(40) A. Bauzá, P. M. Deyà, A. Frontera, D. Quiñonero. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1322-6.
(41) R. M. Parrish, C. D. Sherrill. *J. Am. Chem. Soc.* **2014**, *136*, 17386-17389.
(42) R. Wu, T. B. McMahon. *J. Am. Chem. Soc.* **2008**, *130*, 12554-12555.

(43) P. B. Armentrout, M. Citir, Y. Chen, M. T. Rodgers. *J. Phys. Chem. A* **2012**, *116*, 11823-11832.
(44) P. B. Armentrout, B. Yang, M. T. Rodgers. *J. Phys. Chem. B* **2013**, *117*, 3771-3781.
(45) F. Costanzo, R. G. Della Valle. *J. Phys. Chem. B* **2008**, *112*, 12783-12789.
(46) N. C. Polfer, J. Oomens, R. C. Dunbar. *Phys. Chem. Chem. Phys.* **2006**, *8*, 2744-2751.
(47) M. Remko, S. Soralova. *J. Biol. Inorg. Chem.* **2012**, *17*, 621-30.
(48) C. Ruan, M. T. Rodgers. *J. Am. Chem. Soc.* **2004**, *126*, 14600-14610.
(49) P. Wang, G. Ohanessian, C. Wesdemiotis. *Int. J. Mass spectrom.* **2008**, *269*, 34-45.
(50) S. J. Ye, A. A. Clark, P. B. Armentrout. *J. Phys. Chem. B* **2008**, *112*, 10291-10302.
(51) E. Rezabal, T. Marino, J. M. Mercero, N. Russo, J. M. Ugalde. *Inorg. Chem.* **2007**, *46*, 6413-6419.
(52) T. Marino, J. M. Mercero, N. Russo, J. M. Ugalde. *J. Chem. Theory Comput.* **2007**, *3*, 1830-1836.
(53) A. Rimola, L. Rodríguez-Santiago, M. Sodupe. *J. Phys. Chem. B* **2006**, *110*, 24189-24199.
(54) A. Rimola, M. Sodupe, J. Tortajada, L. Rodríguez-Santiago. *Int. J. Mass spectrom.* **2006**, *257*, 60-69.
(55) C. D. M. Churchill, S. D. Wetmore. *J. Phys. Chem. B* **2009**, *113*, 16046-16058.
(56) F. Blanco, B. Kelly, I. Alkorta, I. Rozas, J. Elguero. *Chem. Phys. Lett.* **2011**, *511*, 129-134.
(57) F. Blanco, B. Kelly, G. Sánchez-Sanz, C. Trujillo, I. Alkorta, J. Elguero, I. Rozas. *J. Phys. Chem. B* **2013**, *117*, 11608-11616.
(58) A. A. Rodriguez-Sanz, E. M. Cabaleiro-Lago, J. Rodriguez-Otero. *Phys. Chem. Chem. Phys.* **2014**, *16*, 22499-22512.
(59) *Suite 2012: MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2012.*
(60) Y. Zhao, D. Truhlar. *Theor. Chem. Acc.* **2008**, *120*, 215-241.
(61) R. F. W. Bader, *Atoms in Molecules: A Quantum Theory*, Clarendon Press, Oxford **1990**.
(62) F. Biegler–König, J. Schönbohm, D. Bayles. *J. Comput. Chem.* **2001**, *22*, 545-559.
(63) J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Piquemal, D. N. Beratan, W. Yang. *J. Chem. Theory Comput.* **2011**, *7*, 625-632.
(64) E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, W. Yang. *J. Am. Chem. Soc.* **2010**, *132*, 6498-6506.
(65) S. F. Boys, F. Bernardi. *Mol. Phys.* **1970**, *19*, 553-566.
(66) G. Chałasiński, M. M. Szczęśniak. *Chem. Rev.* **2000**, *100*, 4227-4252.
(67) R. Sedlak, K. E. Riley, J. Řezáč, M. Pitoňák, P. Hobza. *ChemPhysChem* **2013**, *14*, 698-707.
(68) K. E. Riley, J. Rezac, P. Hobza. *Phys. Chem. Chem. Phys.* **2011**, *13*, 21121-21125.

(69) A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, A. K. Wilson. *Chem. Phys. Lett.* **1998**, *286*, 243-252.

(70) T. Helgaker, W. Klopper, H. Koch, J. Noga. *J. Chem. Phys.* **1997**, *106*, 9639-9646.

(71) A. Karton, J. L. Martin. *Theor. Chem. Acc.* **2006**, *115*, 330-333.

(72) G. Jansen. *WIREs Comput. Mol. Sci.* **2014**, *4*, 127-144.

(73) K. Szalewicz. *WIREs Comput. Mol. Sci.* **2012**, *2*, 254-272.

(74) B. Jeziorski, R. Moszynski, K. Szalewicz. *Chem. Rev.* **1994**, *94*, 1887-1930.

(75) J. Řezáč, P. Hobza. *J. Chem. Theory Comput.* **2011**, *7*, 685-689.

(76) J. Tomasi, B. Mennucci, R. Cammi. *Chem. Rev.* **2005**, *105*, 2999-3094.

(77) H. J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, Y. Liu, A. W. Lloyd, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklaß, D. P. O'Neill, P. Palmieri, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, A. Wolf, *Molpro version 2010.1, a package of ab initio programs, see http://www.molpro.net (2010)*.

(78) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. M. J. A, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox Gaussian 09, Revision B.01, Gaussian, Inc., *Wallingford, CT* **2009**.

(79) F. Neese. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73-78.

(80) U. Purushotham, G. Narahari Sastry. *Theor. Chem. Acc.* **2012**, *131*, 1-14.

(81) M. Zhang, Z. Huang, Z. Lin. *J. Chem. Phys.* **2005**, *122*, 134313/1-134313/7.

(82) C. Bermúdez, S. Mata, C. Cabezas, J. L. Alonso. *Angew. Chem. Int. Ed.* **2014**, *53*, 11015-11018.

(83) D. Kim, S. Hu, P. Tarakeshwar, K. S. Kim, J. M. Lisy. *J. Phys. Chem. A* **2003**, *107*, 1228-1238.

9

Conclusions

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Complexes showing cation···π interactions have been computationally studied in order to understand the characteristics and factors controlling the interactions between cations and aromatic units present in the side chain of aromatic amino acids.

Two different kinds of studies have been carried out: the first type (chapters 4 to 6) employs model aromatic systems in order to study the characteristics of the cation···π interaction and how it is affected by the presence of a small number of water molecules nearby; other studies (chapters 7 and 8) are also devoted to the characteristics of the cation···π interaction, but employing the whole amino acid structure, thus introducing other factors as conformational changes or competition with other electron donor groups. The main conclusions reached in these studies are listed below.

Model aromatic systems and microhydration effects

- ✓ Non-hydrated complexes formed by phenol with ammonium and methylammonium cations show similar minima, with the cation hydrogen bonded to the phenyl ring or the hydroxyl group. The interaction with the first water molecule included in the complex is especially strong, forming a cyclic hydrogen-bonded pattern with the water molecule interacting with the cation while forming a O-H···X hydrogen bond to the phenyl ring or the hydroxyl group of phenol. This cyclic pattern is always present as more water molecules are included in the complex.

- ✓ As water molecules are included, the energy differences among the different favorable interaction sites diminish, so minima with different structural patterns show similar complexation energies. Already with three water molecules the hydroxyl group of phenol starts acting as donor in several of the most stable minima.

- o Guanidinium cation preferentially forms T-shaped complexes with benzene, phenol, and indole. The strength of the interaction increases as the size of the aromatic systems grows.

- o The interaction is electrostatically governed, though there are important contributions from induction and dispersion. Indole complexes are more stable due to larger dispersion contributions, whereas both phenol and indole show larger electrostatic stabilization than benzene.

- Only when three water molecules are included there are minima showing participation of the hydroxyl and NH groups as donors, as well as minima where the cation does not interact directly with the aromatic molecule and being oriented parallel to the ring. Therefore, at least three water molecules are needed in order to promote an structural change on the guanidinium···π contact.

- The stabilization obtained as one new water molecule is included into the complex remains fairly constant. As a consequence, the differences of stability among different aromatic moieties are controlled by the guanidinium···aromatic interaction. Only when the third water is included there is a significant decrease since the water molecule has to be located in the second solvation shell.

- Pyrrolidinium complexes with benzene, phenol and indole show a variety of different structural arrangements, the most stable ones with the cation stacked over the ring and its N-H group interacting with the aromatic cloud.

- The cation···aromatic complexes have complexation energies of -16.1 kcal mol^{-1}, -17.4 kcal mol^{-1} and -21.9 kcal mol^{-1} at the CCSD(T)/CBS level for benzene, phenol and indole, respectively. Cheaper methods as M06-2X/6-31+G* and SCSN/aug-cc-pVDZ also provide a pretty good description of these systems.

- The interaction is controlled by electrostatics and induction contributions, but also with relevant contributions from dispersion, especially in stacked complexes. Phenol complexes are more stable than benzene ones due to larger electrostatic contributions, as a consequence of the participation of the hydroxyl group. Indole complexes are the most stable ones due to larger contributions from all stabilizing effects, but mainly to larger dispersion contributions.

- Phenol and indole hydrated complexes usually show a cyclic pattern where besides the direct pyrrolidinium···π interaction, a water molecule forms hydrogen bonds to the hydroxyl group or to the aromatic clouds, thus introducing further stabilization into the complexes. This cyclic pattern usually survives as more water molecules are included into the complex. In benzene complexes, though, water molecules only interact among themselves and directly with the cation.

- The stability differences between phenol and indole complexes are already removed with the inclusion of three water molecules, though both forming more stable complexes than benzene.

Cation complexes with full amino acids

- Guanidinium complexes with phenylalanine, tyrosine and tryptophan show complexation energies reaching -29.4 kcal mol^{-1}, -29.8 kcal mol^{-1} and -32.1 kcal mol^{-1}, respectively. Therefore, Phe and Tyr interact with similar strength while Trp complex is significantly more stable.

- Most stable minima formed with the amino acids correspond to cation$\cdots\pi$ structures with the neutral forms of the amino acids. In these minima, guanidinium cation forms two hydrogen bonds to the amino and carboxylic groups of the amino acids. However, extended complexes formed with the amino acids in zwitterionic form are equally stable in Phe and Tyr complexes.

- A series of different methods has been tested, with MP2.X giving the best results. Cheaper methods as M06-2X/6-31+G* give a fairly good description of the interaction but underestimate the relative stability of zwitterionic structures.

- The preference for cation$\cdots\pi$ complexes comes from a complex balance of several contributions. Folded structures are favored both by increased induction and dispersion contributions associated to the cation$\cdots\pi$ contact, which are capable of overcoming the cost of deformation associated to the structural change of the amino acids.

- Imidazolium cation complexes with phenylalanine, tyrosine, tryptophan and histidine show complexation energies reaching -29.2 kcal mol^{-1}, -29.8 kcal mol^{-1}, -31.4 kcal mol^{-1}, and -34.8 kcal mol^{-1}. Therefore, the interaction strength is similar to that found for guanidinium complexes.

- Most stable minima with Phe, Tyr, and Trp correspond to structures with the cation forming a double hydrogen bond to the amino/carboxylic group, while simultaneously interacting with the aromatic cloud by means of stacking interactions. His most stable

complexes, on the other hand, behave differently, with the amino acid in extended zwitterionic structure and interacting with the cation by means of two hydrogen bonds with the carboxylate group. Also, intramolecular hydrogen bonds with the lone pair of the nitrogen atom in the imidazole ring contribute to stabilize His zwitterionic minima.

- The interaction is stronger as the size of the aromatic moiety grows in Phe, Tyr and Trp as a consequence of the contribution of the cation···π contact. His departs from this trend forming the most stable complexes despite lacking cation···π contacts.

- Solvent effects as modeled by a continuum model do not affect significantly to the trends observed in the gas phase. The main effect is a greater relative stabilization of zwitterionic minima.

- Overall, there seems that Phe, Tyr and Trp complexes would appear as cation···π structures or extended zwitterionic ones depending on the degree of exposure to the solvent, while in the case of His complexes the results predict that cation···π structures are not feasible, complexes always corresponding to extended zwitterionic minima.

Appendices

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix 1

"Microhydration study: complexes between NH_4^+, $CH_3NH_3^+$ and aromatic species"

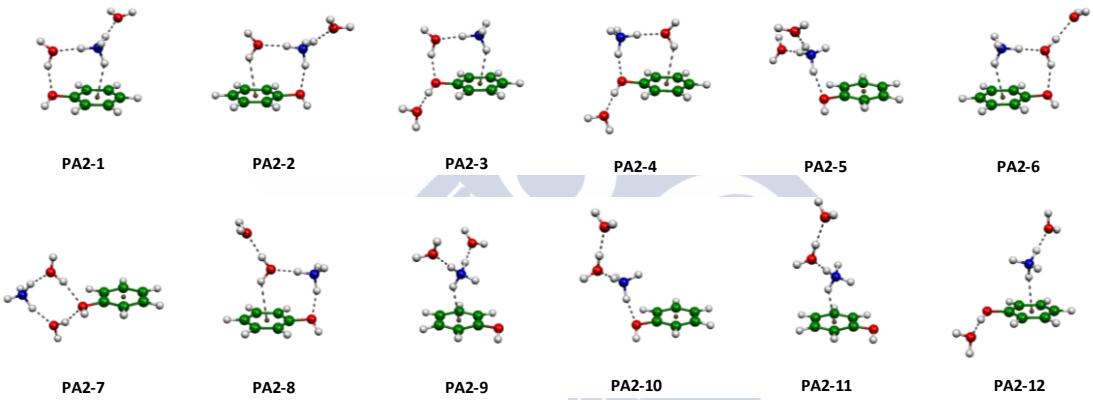


Figure A1.1. Selected most stable minima for the complexes formed by ammonium with phenol in the presence of two water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

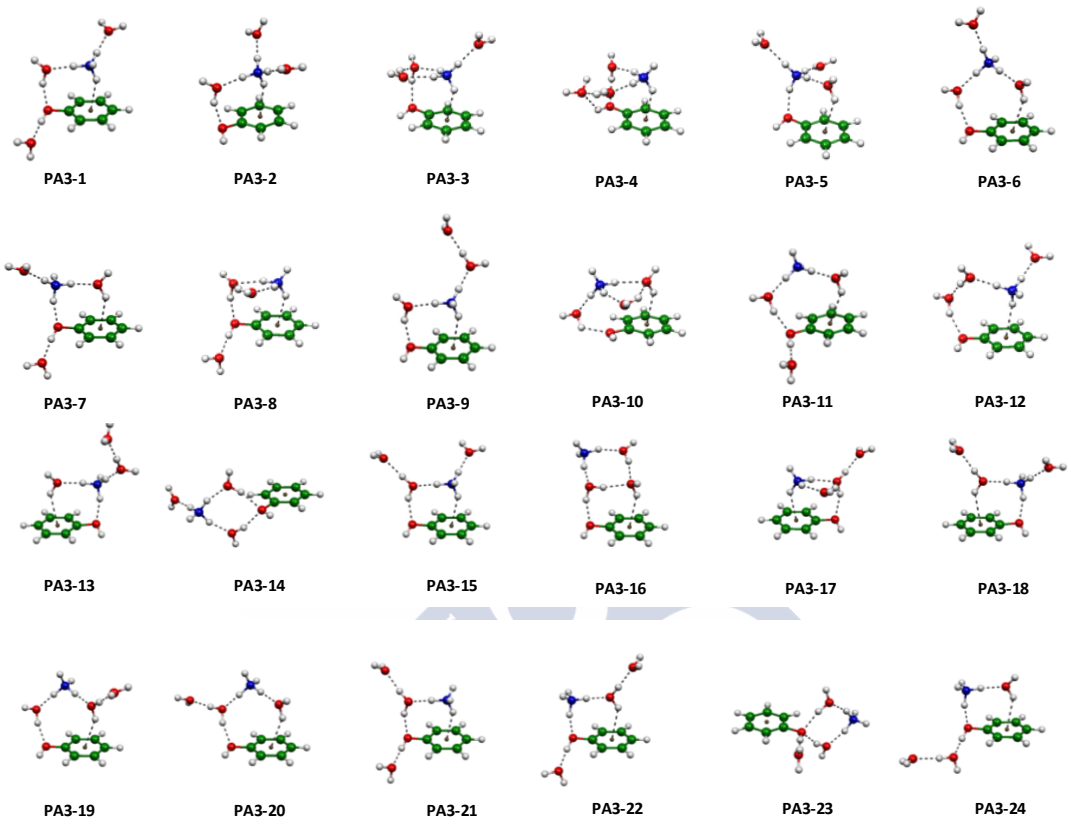


Figure A1.2. Selected most stable minima for the complexes formed by ammonium with phenol in the presence of three water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

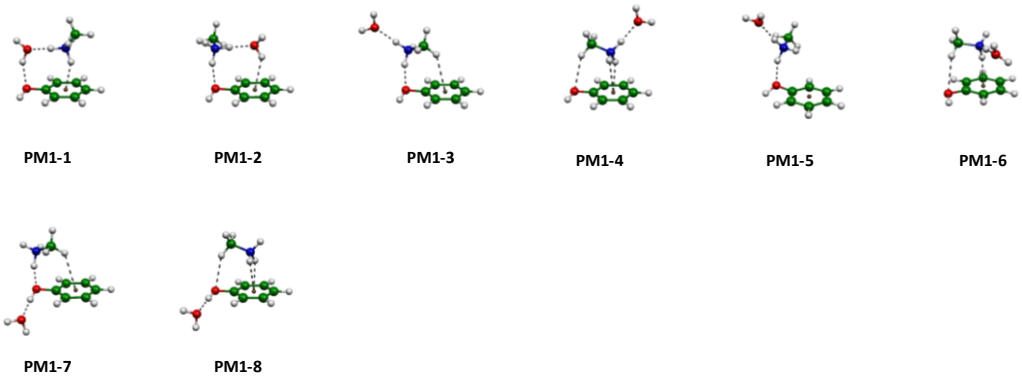


Figure A1.3. Selected most stable minima for the complexes formed by methylammonium with phenol in the presence of one water molecule as obtained at the MP2/6-31+G(2d,p) level of calculation.

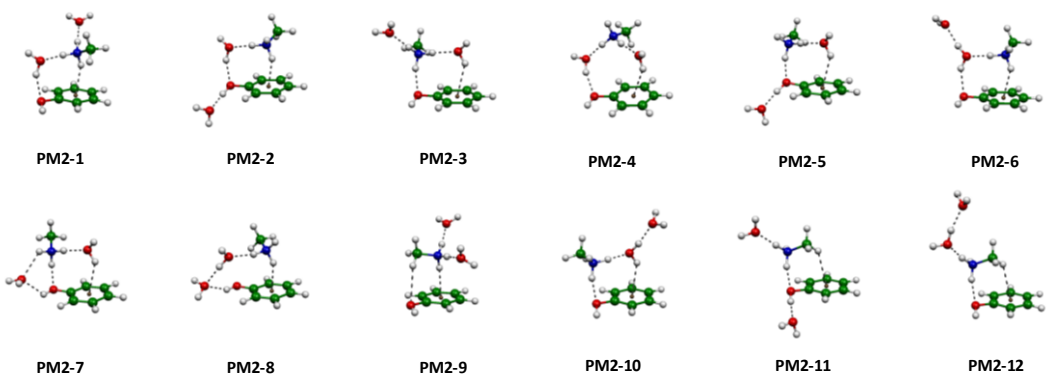


Figure A1.4. Selected most stable minima for the complexes formed by methylammonium with phenol in the presence of two water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

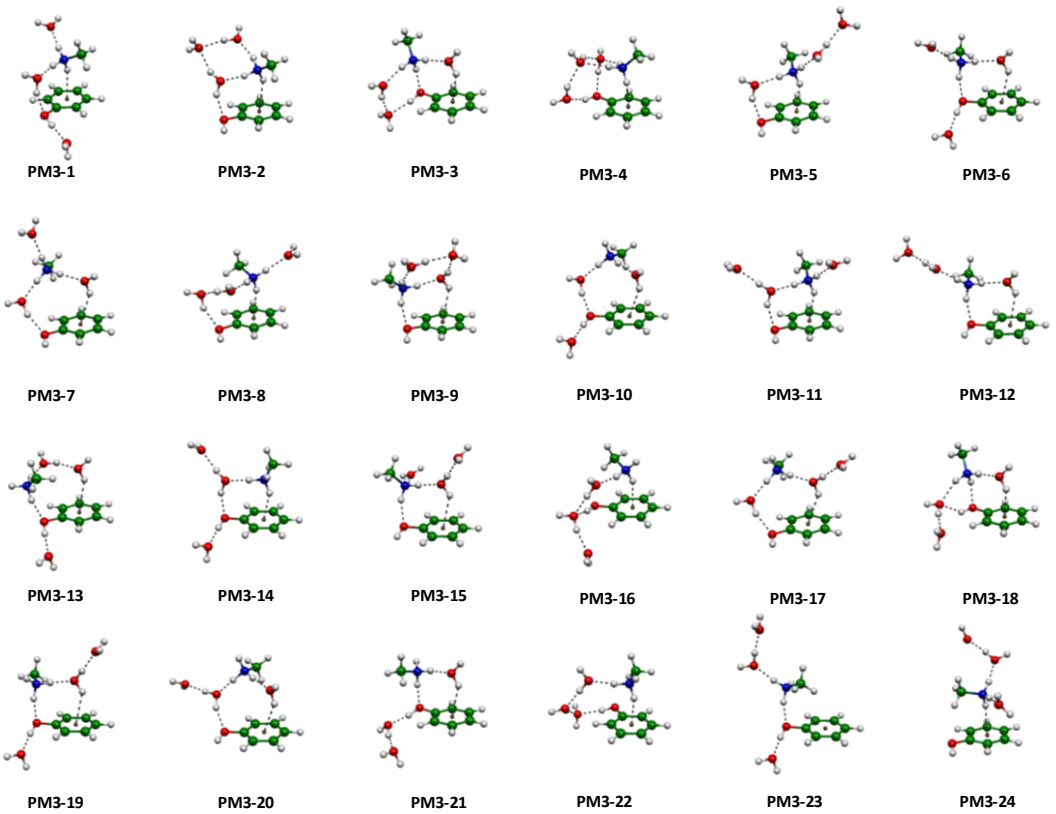


Figure A1.5. Selected most stable minima for the complexes formed by methylammonium with phenol in the presence of three water molecules as obtained at the MP2/6-31+G(2d,p) level of calculation.

Table A1.1. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the ammonium clusters containing zero, one and two water molecules obtained at the MP2/6-31+G(2d,p) level.

	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hydr}
PA0-1	-19.88	-18.87	-18.98	-19.88
PA0-2	-18.31	-17.34	-17.22	-18.31
PA0-3	-18.28	-17.27	-17.17	-18.28
PA1-1	-37.77	-34.37	-35.17	-18.56
PA1-2	-37.30	-34.30	-34.91	-18.03
PA1-3	-35.44	-32.81	-33.02	-16.07
PA1-4	-34.30	-31.88	-31.84	-14.89
PA1-5	-32.96	-30.22	-30.71	-13.09
PA1-6	-30.28	-27.29	-27.48	-10.34
PA2-1	-51.18	-46.38	-47.08	-15.83
PA2-2	-50.57	-46.01	-46.61	-15.16
PA2-3	-49.59	-44.06	-45.31	-13.81
PA2-4	-49.43	-44.45	-45.44	-13.56
PA2-5	-48.76	-44.59	-44.80	-13.31
PA2-6	-48.31	-43.13	-44.28	-13.25
PA2-7	-48.18	-43.22	-44.21	-12.84
PA2-8	-47.84	-42.92	-43.96	-12.71
PA2-9	-47.64	-43.72	-43.70	-12.22
PA2-10	-46.52	-41.79	-42.52	-11.33
PA2-11	-45.47	-39.64	-41.05	-10.31
PA2-12	-45.04	-40.59	-40.85	-9.06

Table A1.2. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the ammonium clusters containing three water molecules obtained at the MP2/6-31+G(2d,p) level.

	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hydr}
PA3-1	-63.22	-56.35	-57.41	-13.31
PA3-2	-63.06	-56.79	-57.42	-13.59
PA3-3	-62.91	-55.65	-56.94	-13.60
PA3-4	-62.62	-53.84	-56.25	-13.67
PA3-5	-62.55	-56.43	-57.00	-13.04
PA3-6	-62.30	-55.83	-56.75	-12.98
PA3-7	-62.10	-55.53	-56.49	-12.10
PA3-8	-61.90	-54.10	-55.82	-12.14
PA3-9	-61.71	-54.69	-55.96	-12.46
PA3-10	-61.19	-53.59	-55.26	-11.91
PA3-11	-61.07	-53.92	-55.43	-11.31
PA3-12	-61.05	-54.25	-55.37	-11.83
PA3-13	-61.01	-54.31	-55.40	-11.71
PA3-14	-60.98	-54.52	-55.36	-11.49
PA3-15	-60.98	-54.25	-55.31	-11.71
PA3-16	-60.58	-52.39	-54.58	-12.16
PA3-17	-60.51	-52.87	-54.49	-11.47
PA3-18	-60.41	-53.97	-54.89	-11.09
PA3-19	-60.01	-53.11	-54.50	-10.92
PA3-20	-59.83	-52.98	-54.32	-10.75
PA3-21	-59.80	-52.58	-54.08	-10.16
PA3-22	-59.67	-52.75	-54.14	-9.91
PA3-23	-59.44	-52.40	-54.35	-9.59
PA3-24	-59.25	-52.09	-53.62	-9.43

Table A1.3. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the methylammonium clusters containing zero, one and two water molecules obtained at the MP2/6-31+G(2d,p) level.

	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hydr}
PM0-1	-18.18	-17.25	-17.06	-18.18
PM0-2	-17.84	-17.04	-16.74	-17.84
PM0-3	-17.73	-16.86	-16.57	-17.73
PM1-1	-35.26	-29.00	-29.35	-18.03
PM1-2	-34.17	-28.20	-28.38	-16.95
PM1-3	-32.70	-27.12	-26.88	-15.36
PM1-4	-32.38	-26.94	-26.59	-15.04
PM1-5	-32.28	-26.80	-26.47	-14.93
PM1-6	-31.75	-26.17	-25.98	-14.45
PM1-7	-30.76	-24.73	-24.91	-12.82
PM1-8	-29.57	-23.45	-23.48	-11.59
PM2-1	-48.14	-40.50	-40.72	-16.00
PM2-2	-47.20	-38.81	-39.58	-14.55
PM2-3	-46.97	-39.70	-39.67	-14.80
PM2-4	-46.48	-38.56	-39.12	-14.56
PM2-5	-46.29	-38.37	-38.88	-13.61
PM2-6	-45.72	-37.58	-38.28	-13.86
PM2-7	-45.30	-37.43	-37.83	-12.94
PM2-8	-45.18	-36.73	-37.45	-13.00
PM2-9	-44.70	-37.88	-37.48	-12.48
PM2-10	-44.68	-36.87	-37.37	-12.77
PM2-11	-44.40	-36.95	-36.98	-11.57
PM2-12	-43.93	-36.15	-36.49	-11.95

Table A1.4. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the methylammonium clusters containing three water molecules obtained at the MP2/6-31+G(2d,p) level.

	$\Delta E_{complex}$	ΔE_{ZPE}	ΔH^{298}	ΔE_{hydr}
PM3-1	-59.18	-49.47	-50.03	-13.54
PM3-2	-58.82	-49.07	-49.75	-14.43
PM3-3	-58.27	-48.17	-49.05	-13.14
PM3-4	-58.14	-46.35	-48.29	-13.30
PM3-5	-58.02	-48.36	-48.97	-13.10
PM3-6	-57.88	-48.68	-48.93	-12.19
PM3-7	-57.86	-48.72	-48.99	-12.99
PM3-8	-57.86	-47.90	-48.70	-13.07
PM3-9	-57.70	-48.52	-48.81	-13.28
PM3-10	-57.28	-47.41	-48.26	-11.82
PM3-11	-57.14	-47.69	-48.16	-12.22
PM3-12	-56.93	-47.52	-47.97	-11.98
PM3-13	-56.59	-47.02	-47.50	-11.20
PM3-14	-56.56	-46.32	-47.40	-11.11
PM3-15	-55.97	-46.69	-47.09	-11.04
PM3-16	-55.96	-45.21	-46.49	-11.05
PM3-17	-55.95	-46.27	-47.05	-11.16
PM3-18	-55.90	-45.70	-46.62	-10.85
PM3-19	-55.83	-46.07	-46.86	-10.30
PM3-20	-55.81	-46.09	-46.89	-11.05
PM3-21	-55.80	-45.67	-46.68	-10.29
PM3-22	-55.38	-44.72	-45.99	-10.40
PM3-23	-54.78	-45.08	-45.68	-9.21
PM3-24	-54.69	-43.76	-45.28	-9.73

Appendix 2

"Microhydration study: complexes between guanidinium cation and aromatic species"

Table A2.1. SAPT(DFT) energy decomposition in kcal mol^{-1} for non-hydrated complexes formed by guanidinium cation and benzene, phenol and indole.

	E_{elec}	E_{exch}	E_{ind}	$E_{exch-ind}$	E_{disp}	$E_{exch-disp}$	δ_{HF}
BG0-1	-9.958	11.312	-9.804	3.793	-7.176	0.883	-2.406
PG0-1	-15.850	16.440	-12.301	4.846	-8.195	1.071	-3.034
PG0-2	-17.866	17.502	-12.292	5.013	-7.133	1.067	-2.131
PG0-3	-10.742	11.542	-10.257	3.969	-7.348	0.897	-2.575
PG0-4	-13.768	15.022	-9.954	5.635	-10.305	1.489	-1.368
IG0-1	-15.601	16.252	-13.360	5.380	-10.165	1.308	-3.584
IG0-2	-13.854	13.569	-12.051	4.882	-8.548	1.111	-2.917
IG0-3	-15.006	15.911	-11.284	6.103	-11.938	1.730	-1.740

In the manuscript $E_{rep}= E_{exch}$; $E_{ind}=E_{exch-ind}+E_{ind}+\delta_{HF}$; $E_{disp}=E_{exch-disp}+E_{disp}$

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix 3

"Microhydration study: complexes between pyrrolidinium cation and aromatic amino acids"

Table A3.1. SAPT(DFT) energy decomposition in kcal mol^{-1} for non-hydrated complexes formed by pyrrolidinium cation and benzene, phenol and indole.

	E_{elec}	E_{exch}	E_{ind}	$E_{exch-ind}$	E_{disp}	$E_{exch-ind}$	δ_{HF}
PB-1	-10.70	12.37	-10.94	4.38	-8.99	1.10	-2.32
PB-2	-11.18	12.24	-11.02	4.44	-8.04	1.00	-2.55
PP-1	-15.39	15.63	-12.06	5.08	-8.33	1.09	-2.46
PP-2	-12.01	13.47	-11.33	4.68	-10.03	1.23	-2.26
PP-3	-17.66	19.34	-12.84	5.64	-11.01	1.47	-3.13
PP-4	-12.25	14.12	-11.43	4.95	-10.78	1.34	-2.28
PP-5	-15.40	15.43	-12.40	5.00	-7.14	0.96	-2.62
PP-6	-11.24	12.48	-11.31	4.49	-9.07	1.10	-2.39
PI-1	-14.65	15.82	-13.40	5.74	-11.62	1.46	-3.06
PI-2	-15.43	16.44	-13.25	5.57	-12.78	1.62	-2.79
PI-3	-14.76	17.28	-13.61	5.89	-12.79	1.65	-3.33
PI-4	-15.46	16.67	-13.14	5.52	-12.78	1.63	-3.00
PI-5	-15.39	14.55	-13.51	5.38	-9.65	1.22	-3.11

In the manuscript $E_{rep}= E_{exch}$; $E_{ind}=E_{exch-ind}+E_{ind}+\delta_{HF}$; $E_{disp}=E_{exch-disp}+E_{disp}$

Table A3.2. Pair energy contributions to monohydrated complexes as obtained with M062X/6-31+G*.

	Pyrr···π	**Pyrr···H_2O**	**H_2O···π**	**3-body**
PB-1W-1	-15.92	-17.20	0.44	1.62
PB-1W-2	-5.81	-17.09	-2.69	-1.85
PP-1W-1	-16.95	-17.04	-2.16	1.05
PP-1W-2	-16.55	-16.99	-3.38	0.47
PP-1W-3	-19.62	-17.25	0.54	1.68
PP-1W-4	-17.62	-17.23	0.47	1.61
PP-1W-5	-10.56	-17.18	-3.95	-1.66
PP-1W-6	-10.13	-17.13	-2.49	-1.20
PP-1W-7	-19.47	-5.59	-7.22	-1.88
PP-1W-8	-17.51	-4.08	-7.64	-1.85
PI-1W-1	-20.47	-16.83	-2.33	1.18
PI-1W-2	-18.59	-16.56	-2.58	0.91
PI-1W-3	-22.32	-17.22	0.62	1.93
PI-1W-4	-21.04	-17.26	0.52	1.97
PI-1W-5	-12.10	-16.90	-2.83	-1.55
PI-1W-6	-12.39	-17.19	-2.67	-1.49
PI-1W-7	-22.57	-3.80	-6.04	-1.30
PI-1W-8	-21.11	-3.65	-6.02	-1.47

Appendix 4

"Cation···π interactions between guanidinium and aromatic amino acids"

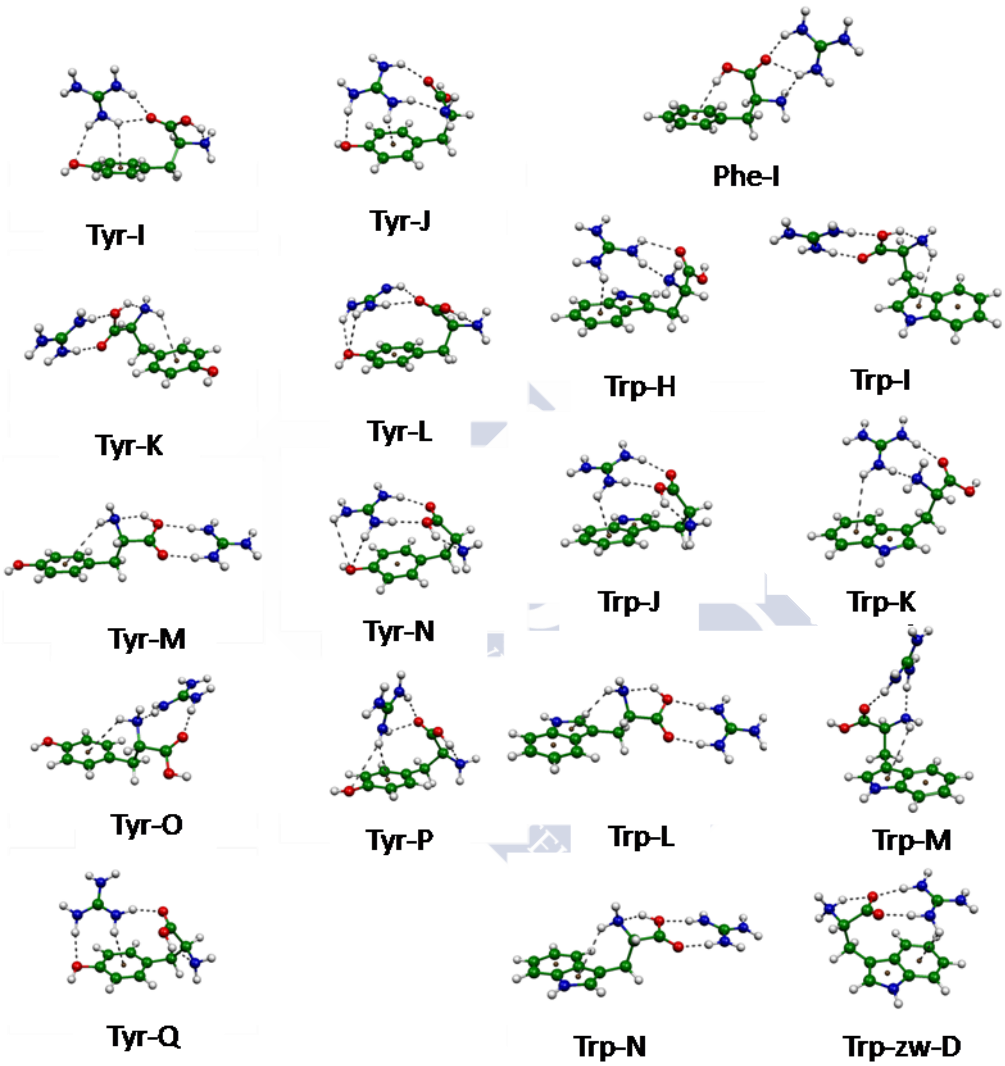


Figure A4.1. Other minima not included in the text.

Table A4.1. Complexation energies (kcal mol^{-1}) for phenylalanine complexes with guanidinium obtained with different methods.

	6-31+G*	CBS					Aug-cc-pVDZ				cc-pVTZ			
	M062X	**MP2**	**SCS-MP2**	**SCSN-MP2**	**MP2.X**	**CCSD(T)**	**SAPT**	**MP2**	**SCS-MP2**	**SCSN-MP2**	**M062X**	**B3LYP-D3**	**B2PLYP-D3**	**B2PLYP**
Phe-A	-30.5	-30.9	-26.9	-30.4	-29.5	-29.4	-29.3	-29.3	-25.7	-29.5	-30.3	-30.5	-32.3	-28.7
Phe-B	-29.0	-29.3	-25.7	-29.4	-28.7	-28.4	-28.8	-27.6	-24.4	-28.4	-28.8	-29.8	-31.5	-28.9
Phe-C	-27.9	-28.3	-25.0	-28.5	-28.2	-28.3	-28.0	-27.3	-24.4	-28.0	-27.7	-29.1	-31.1	-29.6
Phe-D	-27.0	-27.3	-24.5	-27.7	-27.2	-27.2	-27.1	-26.2	-23.7	-27.1	-26.9	-28.2	-30.3	-29.4
Phe-E	-27.2	-27.3	-24.2	-27.3	-26.7	-26.8	-27.0	-26.4	-23.6	-26.9	-27.1	-27.7	-29.9	-27.8
Phe-F	-26.9	-27.1	-24.3	-27.5	-26.3	-26.5	-26.7	-25.9	-23.4	-26.9	-26.7	-27.7	-29.8	-28.1
Phe-G	-26.9	-26.6	-23.9	-27.1	-25.9	-26.0	-26.4	-25.5	-23.1	-26.6	-26.7	-27.4	-29.3	-28.0
Phe-H	-24.4	-24.7	-22.0	-25.0	-24.1	-24.3	-24.3	-23.9	-21.6	-24.6	-24.4	-25.3	-27.4	-25.7
Phe-I	-23.7	-23.5	-20.7	-23.8	-23.3	-23.2	-23.1	-22.6	-20.1	-23.4	-23.6	-23.5	-25.7	-24.0
Phe-zw-A	-28.6	-29.7	-25.5	-28.8	-29.9	-29.5	-27.8	-28.8	-25.0	-28.5	-26.1	-29.7	-31.3	-29.4
Phe-zw-B	-27.9	-28.6	-24.9	-28.0	-28.9	-28.5	-26.8	-27.7	-24.4	-27.7	-25.2	-28.8	-30.5	-29.3
Sum. Squares	**17.8**	**18.2**	**366.7**	**27.0**	**2.1**	**0.0**	**25.0**	**21.0**	**579.7**	**10.1**	**100.7**	**44.4**	**378.3**	**76.8**
Standard Dev.	**0.6**	**0.6**	**2.8**	**0.8**	**0.2**	**0.0**	**0.7**	**0.7**	**3.5**	**0.5**	**1.5**	**1.0**	**2.9**	**1.3**
Average Dev.	**-0.2**	**-0.5**	**2.8**	**-0.5**	**-0.1**	**0.0**	**0.2**	**0.6**	**3.5**	**0.0**	**0.4**	**-0.9**	**-2.8**	**-1.0**
MAE	**0.5**	**0.5**	**2.8**	**0.7**	**0.2**	**0.0**	**0.5**	**0.6**	**3.5**	**0.4**	**1.0**	**0.9**	**2.8**	**1.1**
MAE-NozW	**0.5**	**0.5**	**2.5**	**0.7**	**0.1**	**0.0**	**0.2**	**0.6**	**3.3**	**0.2**	**0.4**	**1.0**	**3.0**	**1.3**

Table A4.2. Complexation energies (kcal mol^{-1}) for tyrosine complexes with guanidinium obtained with different methods.

	6-31+G*	CBS					Aug-cc-pVDZ				cc-pVTZ			
	M062X	MP2	SCS-MP2	SCSN-MP2	MP2.X	CCSD(T)	SAPT	MP2	SCS-MP2	SCSN-MP2	M062X	B3LYP-D3	B2PLYP-D3	B2PLYP
Tyr-A	-31.7	-31.5	-26.7	-30.3	-29.8	-29.7	-29.5	-29.6	-25.4	-29.3	-31.1	-30.5	-32.2	-27.4
Tyr-B	-31.0	-30.7	-26.1	-29.8	-29.6	-29.4	-29.6	-28.8	-24.7	-28.7	-30.2	-30.4	-32.1	-27.8
Tyr-C	-29.8	-30.2	-26.3	-29.8	-29.6	-29.4	-29.5	-28.8	-25.3	-29.0	-29.4	-30.2	-32.2	-29.3
Tyr-D	-28.1	-28.6	-25.1	-28.8	-28.4	-28.5	-28.1	-27.6	-24.5	-28.3	-28.0	-29.6	-31.7	-30.1
Tyr-E	-28.8	-28.8	-23.8	-27.0	-27.5	-27.5	-27.2	-27.2	-22.7	-26.1	-28.0	-28.5	-30.1	-25.1
Tyr-F	-26.9	-27.3	-24.3	-27.8	-27.1	-27.2	-27.2	-26.2	-23.6	-27.2	-27.0	-28.4	-30.5	-29.6
Tyr-G	-27.2	-27.2	-23.4	-27.1	-26.7	-26.7	-26.6	-25.9	-22.5	-26.5	-27.2	-27.7	-29.6	-26.9
Tyr-H	-28.8	-28.1	-23.0	-26.4	-26.8	-26.7	-26.8	-26.5	-21.9	-25.6	-28.2	-27.6	-29.1	-24.1
Tyr-O	-27.0	-27.2	-24.2	-27.7	-26.3	-26.6	-26.6	-26.1	-23.4	-27.1	-26.9	-28.1	-30.1	-28.4
Tyr-Q	-26.9	-27.2	-22.5	-25.5	-26.0	-26.0	-25.6	-25.9	-21.7	-24.9	-26.3	-26.8	-28.5	-24.3
Tyr-P	-27.2	-26.9	-22.5	-26.1	-26.0	-25.9	-25.9	-25.6	-21.7	-25.4	-26.9	-27.0	-28.7	-24.9
Tyr-zw-A	-28.8	-29.9	-25.6	-29.1	-30.0	-29.8	-28.3	-29.0	-25.1	-28.8	-26.4	-30.1	-31.9	-30.0
Tyr-zw-B	-27.9	-28.8	-24.9	-28.2	-28.9	-28.6	-27.2	-27.8	-24.3	-27.8	-25.5	-29.1	-30.9	-29.6
Tyr-zw-E	-21.3	-22.8	-18.0	-21.0	-22.4	-22.0	-21.3	-21.6	-17.3	-20.3	-19.0	-22.2	-24.0	-21.1
Tyr-I	-29.7	-30.0	-26.1	-29.8	-29.4	-29.2	-29.4	-28.5	-25.1	-29.0	-29.3	-30.1	-32.0	-29.3
Tyr-J	-30.9	-30.6	-25.9	-29.4	-28.9	-28.9	-28.7	-28.8	-24.6	-28.5	-30.2	-29.5	-31.4	-26.6
Tyr-K	-27.5	-28.0	-24.6	-28.3	-27.8	-27.9	-27.5	-27.0	-23.9	-27.7	-27.4	-29.0	-31.0	-29.5
Tyr-L	-28.8	-28.5	-24.1	-27.7	-27.5	-27.4	-27.7	-26.7	-22.7	-26.6	-28.0	-28.2	-29.7	-25.6
Tyr-M	-26.9	-27.3	-24.3	-27.8	-27.1	-27.2	-27.2	-26.2	-23.6	-27.2	-27.0	-28.5	-30.7	-29.8
Tyr-N	-27.2	-27.9	-23.0	-26.2	-26.6	-26.7	-26.4	-26.4	-22.0	-25.5	-26.7	-27.7	-29.2	-24.5
Tyr-zw-C	-28.6	-29.8	-25.4	-29.0	-29.9	-29.6	-28.1	-28.9	-25.0	-28.7	-26.3	-29.9	-31.6	-29.8
Tyr-zw-D	-27.8	-28.7	-24.9	-28.2	-28.9	-28.6	-27.2	-27.7	-24.3	-27.8	-25.4	-29.1	-30.9	-29.6
Tyr-zw-F	-22.3	-22.6	-17.0	-19.8	-21.6	-21.2	-20.4	-21.1	-16.0	-19.0	-18.9	-21.1	-22.7	-18.1
Sum. Squares	**113.2**	**86.9**	**1162.6**	**35.3**	**3.9**	**0.0**	**44.6**	**39.7**	**1831.1**	**79.5**	**272.2**	**74.7**	**677.8**	**301.0**
Standard Dev.	**1.1**	**1.0**	**3.5**	**0.6**	**0.2**	**0.0**	**0.7**	**0.6**	**4.4**	**0.9**	**1.7**	**0.9**	**2.7**	**1.8**
Average Dev.	**-0.4**	**-0.8**	**3.4**	**0.0**	**-0.1**	**0.0**	**0.4**	**0.6**	**4.3**	**0.7**	**0.5**	**-0.8**	**-2.6**	**0.4**
MAE	**0.9**	**0.8**	**3.4**	**0.5**	**0.2**	**0.0**	**0.5**	**0.6**	**4.3**	**0.7**	**1.2**	**0.8**	**2.6**	**1.5**
MAE-NozW	**1.0**	**0.9**	**3.3**	**0.5**	**0.1**	**0.0**	**0.2**	**0.5**	**4.2**	**0.6**	**0.6**	**1.0**	**2.8**	**1.7**

Table A4.3. Complexation energies (kcal mol^{-1}) for tryptophan complexes with guanidinium obtained with different methods.

	6-31+G*	CBS					Aug-cc-pVDZ				cc-pVTZ			
	M062X	MP2	SCS-MP2	SCSN-MP2	MP2.X	CCSD(T)	SAPT	MP2	SCS-MP2	SCSN-MP2	M062X	B3LYP-D3	B2PLYP-D3	B2PLYP
Trp-A	-32.7	-33.9	-29.3	-33.3	-32.1	-32.1	-31.7	-32.2	-28.1	-32.4	-32.8	-33.2	-35.4	-31.4
Trp-B	-32.5	-33.4	-28.1	-31.8	-32.0	-32.0	-31.6	-32.0	-27.1	-31.0	-31.9	-32.9	-34.8	-29.9
Trp-C	-31.9	-33.0	-28.6	-33.0	-31.9	-31.7	-31.7	-31.1	-27.2	-31.8	-31.7	-32.9	-34.8	-31.6
Trp-D	-31.7	-33.1	-28.6	-32.9	-31.9	-31.7	-31.7	-31.2	-27.1	-31.7	-31.5	-32.9	-34.8	-31.3
Trp-E	-32.1	-32.7	-27.8	-31.9	-31.4	-31.2	-31.5	-30.8	-26.4	-30.8	-31.8	-32.5	-34.4	-30.2
Trp-F	-30.7	-31.8	-26.9	-30.7	-30.6	-30.5	-30.1	-30.3	-25.9	-29.9	-30.3	-31.5	-33.7	-30.1
Trp-G	-31.8	-32.2	-26.7	-30.5	-29.9	-29.9	-29.9	-30.5	-25.5	-29.7	-31.2	-31.0	-32.9	-27.1
Trp-H	-30.6	-31.4	-26.5	-29.9	-29.1	-29.1	-28.9	-29.5	-25.1	-29.0	-30.2	-29.9	-32.2	-27.4
Trp-I	-28.1	-28.8	-25.2	-28.9	-28.6	-28.7	-28.1	-27.7	-24.5	-28.3	-27.9	-29.7	-32.0	-30.5
Trp-J	-28.3	-30.1	-24.7	-28.3	-28.4	-28.4	-28.0	-28.6	-23.6	-27.5	-28.0	-30.1	-31.9	-26.8
Trp-K	-28.6	-29.4	-25.7	-29.2	-28.1	-28.3	-28.3	-27.8	-24.5	-28.4	-28.3	-30.0	-32.0	-29.6
Trp-L	-27.7	-28.2	-25.0	-28.8	-28.0	-28.1	-27.8	-27.0	-24.2	-28.1	-27.7	-29.3	-31.5	-30.7
Trp-M	-28.0	-28.3	-24.9	-28.3	-27.7	-27.7	-27.8	-27.4	-24.4	-27.9	-28.0	-28.9	-31.4	-29.3
Trp-N	-27.2	-27.8	-24.5	-28.1	-27.6	-27.7	-27.5	-26.7	-23.8	-27.4	-27.2	-28.9	-31.2	-30.3
Trp-zw-A	-30.5	-31.8	-27.2	-30.9	-31.9	-31.6	-29.9	-31.0	-26.8	-30.6	-27.9	-31.7	-33.7	-31.8
Trp-zw-B	-29.7	-30.6	-26.5	-29.9	-30.8	-30.5	-29.1	-29.8	-26.1	-29.6	-27.0	-30.9	-32.9	-31.5
Trp-zw-C	-29.3	-30.4	-26.3	-29.9	-30.5	-30.2	-28.7	-29.3	-25.6	-29.4	-22.9	-26.4	-28.1	-23.3
Trp-zw-D	-25.8	-27.2	-21.3	-24.6	-26.2	-25.8	-24.8	-25.8	-20.3	-23.8	-22.9	-26.4	-28.1	-23.3
Sum. Squares	**47.8**	**126.6**	**896.6**	**44.6**	**3.1**	**0.0**	**39.5**	**28.3**	**1460.8**	**38.4**	**387.4**	**151.0**	**720.9**	**394.1**
Standard Dev.	**0.8**	**1.3**	**3.5**	**0.8**	**0.2**	**0.0**	**0.7**	**0.6**	**4.4**	**0.7**	**2.3**	**1.4**	**3.1**	**2.3**
Average Dev.	**-0.1**	**-1.0**	**3.4**	**-0.3**	**-0.1**	**0.0**	**0.4**	**0.4**	**4.4**	**0.4**	**0.9**	**-0.8**	**-2.8**	**0.5**
MAE	**0.6**	**1.0**	**3.4**	**0.7**	**0.2**	**0.0**	**0.5**	**0.5**	**4.4**	**0.5**	**1.3**	**1.2**	**3.1**	**1.7**
MAE-NozW	**0.6**	**1.2**	**3.2**	**0.7**	**0.1**	**0.0**	**0.2**	**0.5**	**4.3**	**0.3**	**0.5**	**1.2**	**3.3**	**1.5**

Table A4.4. SAPT(DFT) partitioning (kcal mol^{-1}) for complexes with phenylalanine.

	E_{def}	E_{tot}	E_{ele}	E_{exch}	E_{ind}	$E_{exch-ind}$	E_{disp}	$E_{exch-dis}$	δHF
Phe-A	7.5	-29.3	-38.0	32.9	-22.8	10.8	-13.9	2.2	-5.8
Phe-B	6.3	-28.8	-34.8	25.1	-17.8	7.1	-10.9	1.6	-3.7
Phe-C	1.0	-28.0	-30.4	21.0	-15.0	6.4	-7.3	1.2	-3.6
Phe-D	1.9	-27.1	-30.4	20.3	-14.6	6.1	-7.0	1.1	-3.4
Phe-E	5.7	-27.0	-36.4	30.2	-21.4	10.2	-9.8	1.8	-5.7
Phe-F	5.6	-26.7	-36.0	32.2	-23.1	11.0	-10.4	1.9	-6.2
Phe-G	5.9	-26.4	-36.5	31.3	-22.2	10.6	-9.7	1.8	-6.0
Phe-H	7.3	-24.3	-34.6	29.4	-21.0	9.7	-9.6	1.7	-5.6
Phe-I	10.5	-23.1	-34.9	23.6	-17.0	7.1	-8.6	1.4	-3.7
Phe-zw-A	18.5	-27.8	-51.5	35.2	-25.1	12.0	-10.2	2.0	-7.1
Phe-zw-B	21.0	-26.8	-52.9	35.4	-25.6	12.1	-10.1	2.0	-7.1

Table A4.5. SAPT(DFT) partitioning (kcal mol^{-1}) for complexes with tyrosine.

	E_{def}	E_{tot}	E_{ele}	E_{exch}	E_{ind}	$E_{exch-ind}$	E_{disp}	$E_{exch-dis}$	δHF
Tyr-A	10.3	-29.5	-40.6	34.3	-23.6	11.8	-17.0	2.7	-4.6
Tyr-B	7.8	-29.6	-38.8	30.9	-20.8	10.1	-15.3	2.4	-3.4
Tyr-C	7.1	-29.5	-35.7	24.1	-17.6	7.2	-11.0	1.6	-3.3
Tyr-D	1.4	-28.1	-30.8	21.1	-15.1	6.4	-7.4	1.2	-3.6
Tyr-E	8.3	-27.2	-36.5	30.4	-20.6	10.6	-15.6	2.4	-3.6
Tyr-F	2.2	-27.2	-30.7	20.4	-14.7	6.2	-7.0	1.2	-3.5
Tyr-G	3.5	-26.6	-29.0	20.7	-16.7	7.6	-9.5	1.4	-3.0
Tyr-H	9.2	-26.8	-35.8	28.4	-19.7	10.0	-15.6	2.4	-3.0
Tyr-O	6.1	-26.7	-36.3	32.2	-23.2	10.9	-10.4	1.9	-6.3
Tyr-Q	7.5	-25.6	-31.8	25.6	-18.5	8.6	-13.1	1.9	-3.6
Tyr-P	6.8	-25.9	-31.6	23.7	-17.8	8.5	-12.1	1.8	-3.2
Tyr-zw-A	18.9	-28.3	-52.3	35.4	-25.3	12.0	-10.3	2.0	-7.2
Tyr-zw-B	21.4	-27.2	-53.8	35.8	-25.9	12.3	-10.2	2.1	-7.2
Tyr-zw-E	25.9	-21.3	-47.9	30.6	-22.5	9.6	-12.2	2.0	-4.8
Tyr-I	6.5	-29.4	-34.7	23.4	-17.2	6.8	-10.8	1.6	-3.3
Tyr-J	10.0	-28.7	-39.3	33.6	-23.1	11.6	-16.7	2.6	-4.6
Tyr-K	1.1	-27.5	-30.1	21.2	-15.1	6.4	-7.3	1.2	-3.7
Tyr-L	8.2	-27.7	-36.0	28.3	-19.3	9.1	-14.6	2.2	-3.1
Tyr-M	2.2	-27.2	-30.7	20.4	-14.7	6.1	-7.0	1.1	-3.5
Tyr-N	6.7	-26.4	-33.0	28.1	-19.4	9.6	-14.2	2.1	-4.0
Tyr-zw-C	18.5	-28.1	-51.8	35.7	-25.5	12.1	-10.3	2.0	-7.2
Tyr-zw-D	21.4	-27.2	-53.7	35.7	-25.8	12.2	-10.2	2.1	-7.2
Tyr-zw-F	29.1	-20.4	-52.5	37.1	-26.2	13.2	-16.8	2.9	-4.5

Table A4.6. SAPT(DFT) partitioning (kcal mol^{-1}) for complexes with tryptophan.

	E_{def}	E_{tot}	E_{ele}	E_{exch}	E_{ind}	$E_{exch-ind}$	E_{disp}	$E_{exch-dis}$	δHF
Trp-A	8.6	-31.7	-39.9	32.9	-23.5	10.7	-14.6	2.3	-5.9
Trp-B	5.9	-31.6	-36.6	29.3	-21.3	10.3	-14.8	2.3	-4.3
Trp-C	6.7	-31.7	-36.7	26.6	-19.1	7.4	-12.4	1.8	-4.0
Trp-D	6.8	-31.7	-37.0	27.5	-19.6	8.0	-13.1	1.9	-4.0
Trp-E	7.6	-31.5	-38.0	28.7	-21.1	9.8	-14.5	2.2	-3.8
Trp-F	5.1	-30.1	-32.9	24.7	-18.5	8.0	-12.0	1.7	-4.2
Trp-G	10.9	-29.9	-39.6	34.4	-23.8	11.9	-18.2	2.8	-5.4
Trp-H	9.4	-28.9	-37.8	33.2	-23.5	11.8	-16.3	2.5	-5.5
Trp-I	1.3	-28.1	-31.1	21.9	-15.7	6.7	-7.4	1.2	-3.9
Trp-J	8.7	-28.0	-35.3	29.7	-20.8	9.7	-15.0	2.2	-4.6
Trp-K	9.5	-28.3	-40.8	35.1	-25.5	12.3	-12.6	2.2	-6.4
Trp-L	2.5	-27.8	-31.6	20.7	-15.0	6.3	-7.1	1.2	-3.5
Trp-M	7.0	-27.8	-38.7	31.6	-22.5	10.7	-10.1	1.8	-6.1
Trp-N	3.1	-27.5	-32.0	21.0	-15.2	6.4	-7.2	1.2	-3.6
Trp-zw-A	18.1	-29.9	-53.5	37.0	-26.5	12.7	-10.5	2.1	-7.6
Trp-zw-B	21.3	-29.1	-55.9	36.8	-26.6	12.7	-10.4	2.1	-7.5
Trp-zw-C	21.1	-28.7	-55.1	36.5	-26.4	12.5	-10.4	2.1	-7.4
Trp-zw-D	29.0	-24.8	-55.7	39.7	-28.8	14.5	-16.8	2.9	-6.9

Appendix 5

"Cation···π interactions between imidazolium and aromatic amino acids"

Table A5.1. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the imidazolium···phenylalanine complex.

	M06-2X/ 6-31+G*	**MP2/CBS**	**MP2.X**	**SAPT**	$\mathbf{E_{def,MP2/CBS}}$
Phe-A	-30.50	-31.90	-29.24	-29.48	8.23
Phe-B	-27.07	-28.38	-28.31	-26.47	20.07
Phe-C	-29.51	-31.14	-28.29	-28.57	8.29
Phe-D	-27.54	-28.48	-28.13	-27.92	1.68
Phe-E	-26.54	-27.47	-27.45	-25.68	22.55
Phe-F	-25.94	-27.76	-27.37	-25.83	20.74
Phe-G	-26.78	-27.61	-27.25	-27.21	2.40
Phe-H	-26.41	-27.42	-26.98	-27.06	2.56
Phe-I	-27.44	-27.96	-26.68	-27.29	7.05
Phe-J	-25.50	-26.88	-26.55	-25.10	23.53
Phe-K	-27.41	-27.45	-26.54	-26.90	7.40
Phe-L	-26.56	-27.64	-26.44	-26.22	5.49
Phe-M	-26.83	-27.22	-26.43	-26.31	5.79
Phe-N	-27.08	-26.88	-25.85	-26.28	7.78
Phe-O	-24.82	-26.14	-25.83	-23.81	20.06
Phe-P	-25.99	-26.70	-25.67	-25.63	6.49
Phe-Q	-25.30	-25.11	-25.22	-22.96	23.22
Phe-R	-25.92	-26.42	-25.17	-25.65	7.19
Phe-S	-24.04	-25.04	-24.76	-22.72	22.25
Phe-T	-24.72	-25.32	-24.45	-24.08	3.85

Table A5.2. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the imidazolium···tyrosine complex.

	M06-2X/ 6-31+G*	**MP2/CBS**	**MP2.X**	**SAPT**	$\mathbf{E_{def,MP2/CBS}}$
Tyr-A	-31.38	-32.90	-29.77	-29.95	9.31
Tyr-B	-31.28	-32.68	-29.50	-29.52	9.78
Tyr-C	-30.23	-30.64	-29.38	-29.04	7.02
Tyr-D	-29.85	-30.14	-28.94	-28.58	6.99
Tyr-E	-30.66	-32.26	-28.77	-28.92	9.98
Tyr-F	-27.26	-28.64	-28.46	-26.90	20.54
Tyr-G	-27.69	-28.73	-28.27	-27.93	2.10
Tyr-H	-27.10	-28.46	-28.27	-26.71	20.20
Tyr-I	-29.32	-30.24	-28.24	-27.84	7.11
Tyr-J	-27.21	-28.42	-27.86	-27.64	2.25
Tyr-K	-27.15	-28.18	-27.72	-27.38	1.81
Tyr-L	-26.81	-27.71	-27.68	-25.70	19.81
Tyr-M	-26.57	-27.64	-27.52	-26.08	23.05
Tyr-N	-27.94	-28.46	-27.48	-27.49	5.72
Tyr-O	-26.51	-27.56	-27.45	-26.00	23.04
Tyr-P	-25.97	-27.80	-27.29	-26.06	21.07
Tyr-Q	-28.40	-28.91	-27.19	-26.77	8.00
Tyr-R	-26.67	-27.62	-27.18	-27.24	2.78
Tyr-S	-28.29	-29.24	-27.17	-26.79	7.35
Tyr-T	-28.47	-28.70	-27.07	-26.62	8.62

Table A5.3. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the imidazolium···tryptophan complex.

	M06-2X/ 6-31+G*	MP2/CBS	MP2.X	SAPT	$E_{def,MP2/CBS}$
Trp-A	-33.03	-35.28	-31.42	-31.60	10.65
Trp-B	-31.36	-33.03	-31.27	-30.98	5.66
Trp-C	-29.02	-30.44	-30.30	-28.49	20.18
Trp-D	-29.61	-31.71	-30.27	-30.15	7.28
Trp-E	-31.53	-33.57	-30.10	-30.01	9.26
Trp-F	-29.45	-31.68	-29.82	-29.62	5.74
Trp-G	-28.74	-29.90	-29.60	-27.92	20.98
Trp-H	-29.37	-31.95	-29.48	-29.28	5.92
Trp-I	-28.29	-29.49	-29.36	-27.86	23.59
Trp-J	-27.80	-29.72	-29.25	-27.80	21.26
Trp-K	-30.51	-33.21	-28.95	-29.05	10.91
Trp-L	-30.03	-32.39	-28.60	-28.85	11.95
Trp-M	-29.18	-30.30	-28.53	-28.99	9.57
Trp-N	-27.87	-28.85	-28.47	-27.94	2.03
Trp-O	-26.55	-29.11	-28.47	-27.20	22.06
Trp-P	-27.20	-28.82	-28.39	-27.20	24.64
Trp-Q	-28.01	-30.16	-28.25	-27.92	5.31
Trp-R	-27.46	-28.13	-28.15	-26.10	22.04
Trp-S	-27.01	-28.54	-28.08	-26.91	24.69
Trp-T	-28.06	-30.04	-28.06	-27.75	8.37

Table A5.4a. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the imidazolium···histidine complex.

	M06-2X/ 6-31+G*	MP2/CBS	MP2.X	SAPT	$E_{def,MP2/CBS}$
His-A	-34.35	-34.90	-34.78	-33.64	19.36
His-B	-34.52	-34.46	-34.32	-33.22	21.18
His-C	-33.85	-34.06	-34.04	-32.49	18.48
His-D	-32.34	-32.87	-33.99	-44.46	79.67
His-E	-33.02	-34.07	-33.60	-32.88	21.11
His-F	-33.99	-33.56	-33.53	-32.02	20.00
His-G	-33.30	-33.65	-33.19	-32.47	22.92
His-H	-31.34	-31.72	-32.88	-43.42	82.80
His-I	-31.54	-31.78	-32.68	-43.06	78.81
His-J	-32.05	-32.80	-32.38	-31.09	19.52
His-K	-31.07	-30.20	-32.16	-36.87	89.81
His-L	-30.19	-29.26	-31.14	-35.12	87.34
His-M	-29.20	-29.62	-30.70	-41.06	80.89
His-N	-28.49	-29.62	-30.57	-40.87	74.55
His-O	-29.71	-28.36	-30.39	-34.48	95.91
His-P	-30.52	-30.57	-30.21	-30.18	6.05
His-Q	-28.62	-28.44	-30.10	-34.26	79.17
His-R	-30.69	-31.21	-30.00	-30.14	11.01

Table A5.4b. Complexation energies (kcal mol^{-1}) obtained for the most stable minima of the imidazolium···histidine complex.

	M06-2X/ 6-31+G*	**MP2/CBS**	**MP2.X**	**SAPT**	**$E_{def,MP2/CBS}$**
His-S	-31.22	-30.82	-29.93	-30.38	8.45
His-T	-30.77	-30.45	-29.89	-30.03	7.12
His-U	-30.07	-30.32	-29.83	-29.99	6.38
His-V	-27.04	-26.69	-29.81	-43.20	122.61
His-W	-30.52	-30.82	-29.70	-29.95	8.41
His-X	-30.45	-29.99	-29.62	-29.85	6.19
His-Y	-27.91	-27.59	-29.08	-33.28	77.58
His-Z	-28.95	-29.32	-28.94	-29.04	9.80
His-AA	-29.56	-29.81	-28.89	-29.31	7.33
His-AB	-28.11	-26.63	-28.71	-33.32	100.73
His-AC	-26.75	-27.64	-28.49	-40.57	88.08
His-AD	-29.73	-29.41	-28.47	-28.54	14.47
His-AE	-28.27	-28.44	-28.26	-27.54	37.41
His-AF	-29.41	-28.32	-28.20	-26.54	5.26
His-AG	-28.77	-27.92	-27.71	-25.54	6.95
His-AH	-25.78	-26.96	-27.43	-24.54	75.71
His-AI	-25.66	-26.97	-27.42	-23.54	80.72

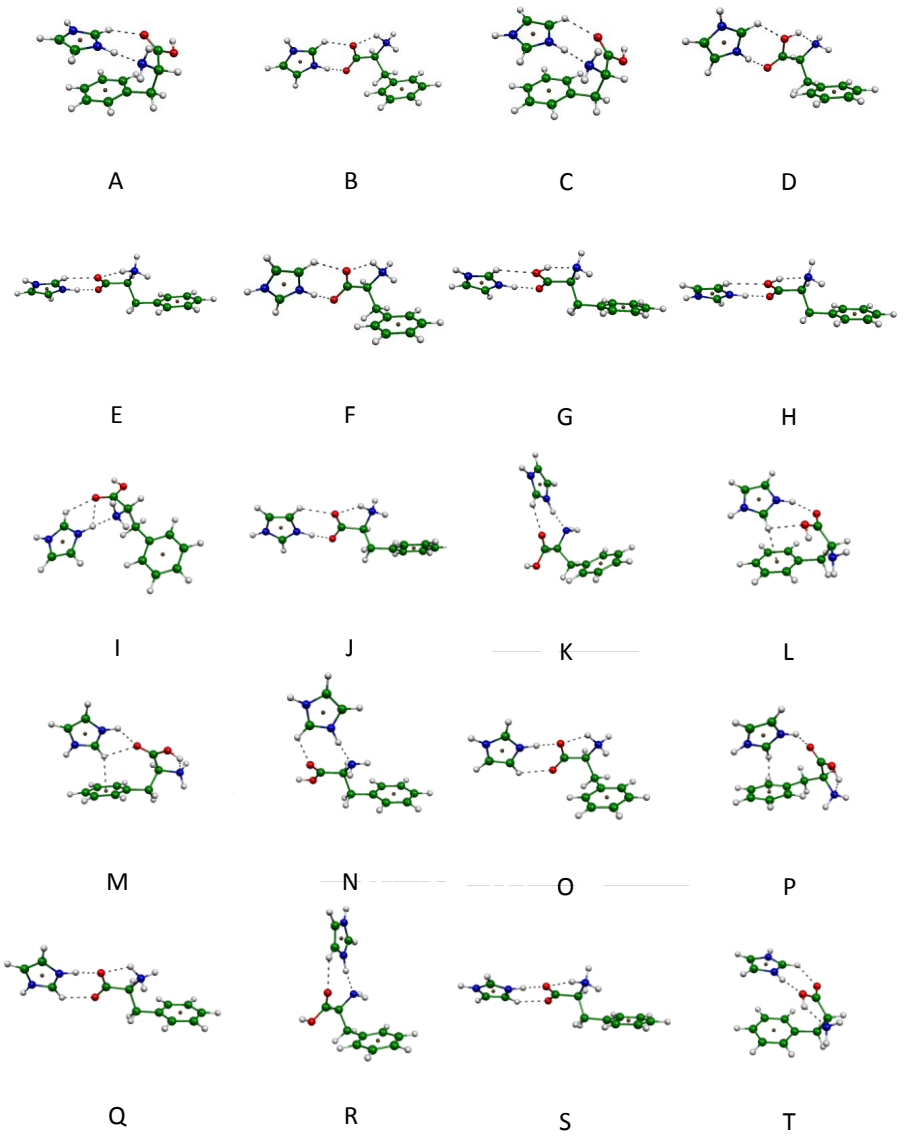


Figure A5.1. Structures of the most stable minima of Imz···Phe complex as obtained at the M06-2X/6-31+G* level.

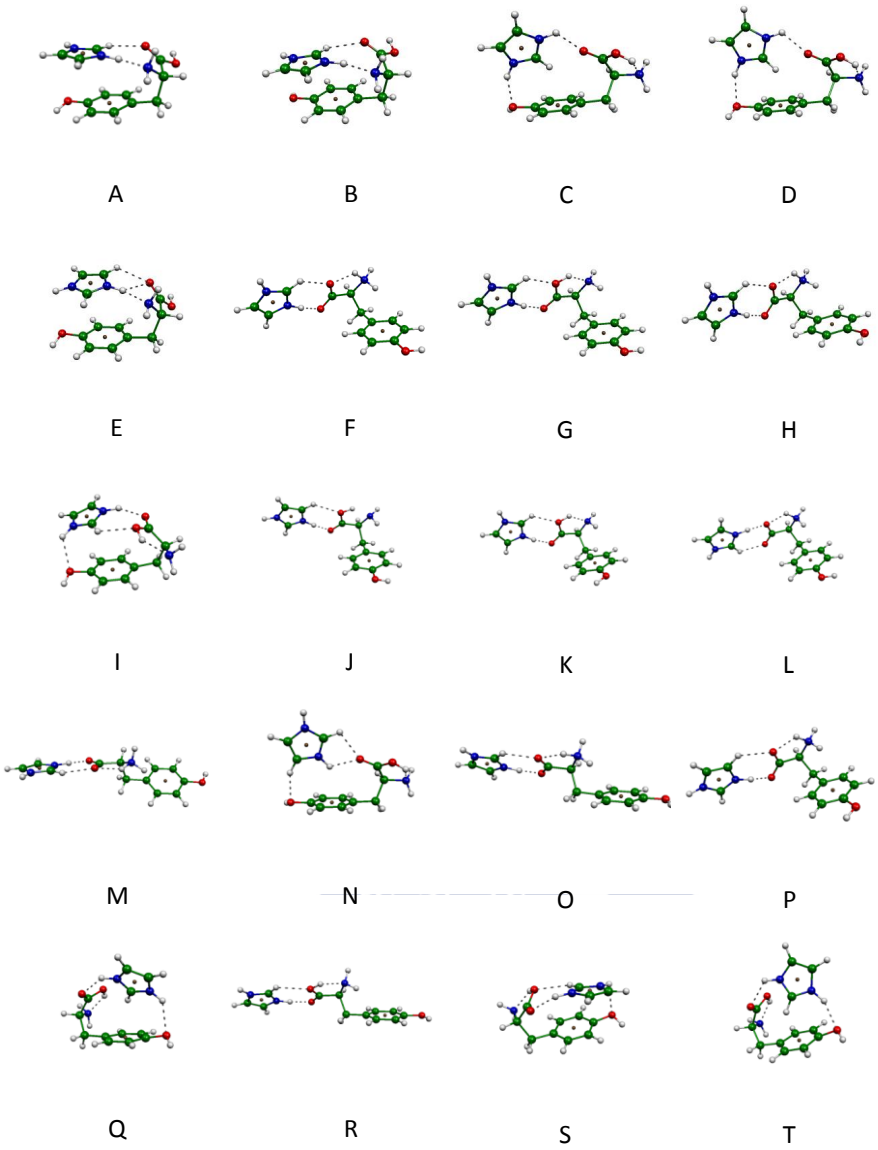


Figure A5.2. Structures of the most stable minima of Imz···Tyr complex as obtained at the M06-2X/6-31+G* level.

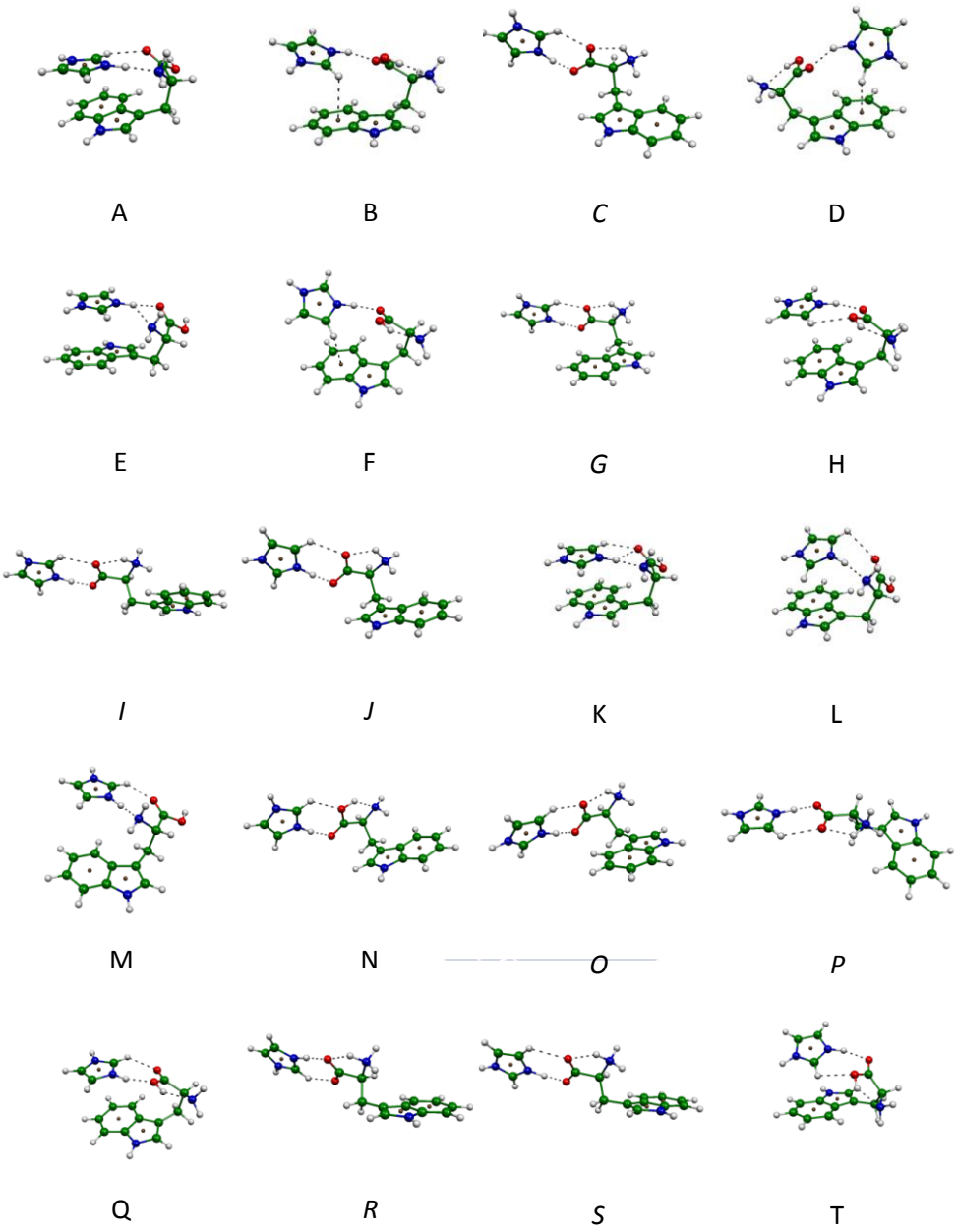


Figure A5.3. Structures of the most stable minima of Imz···Trp complex as obtained at the M06-2X/6-31+G* level.

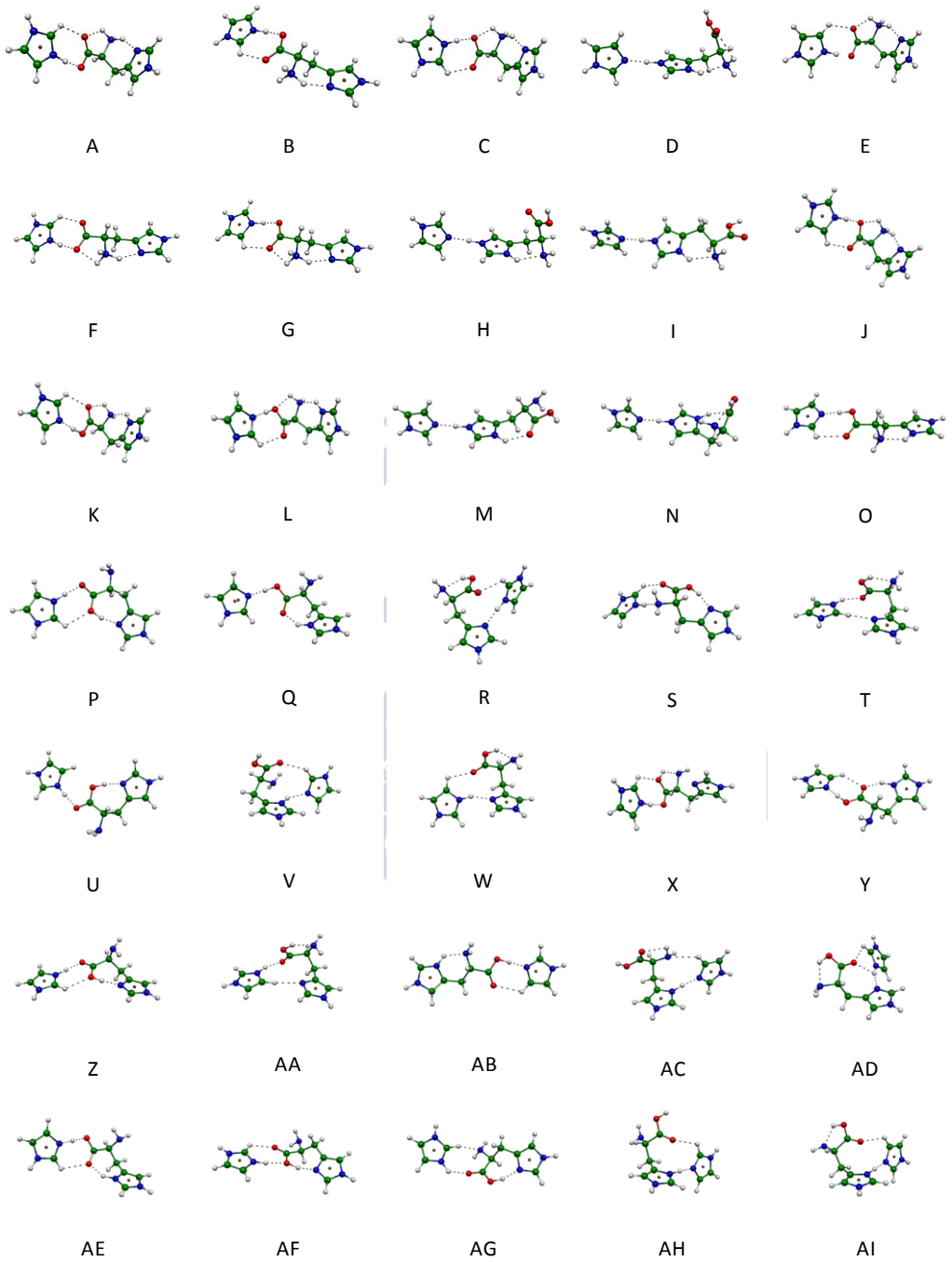
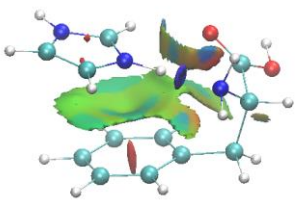
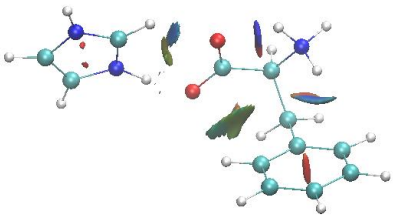


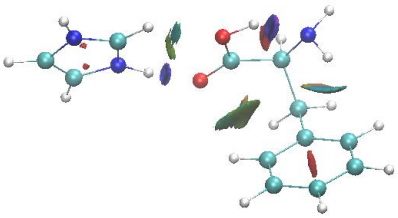
Figure A5.4. Structures of the most stable minima of Imz···His complex as obtained at the M06-2X/6-31+G* level.



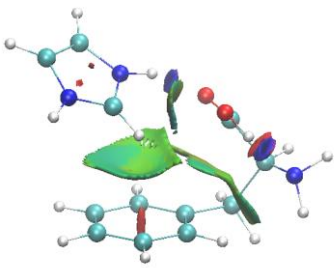
Phe-A (P)



Phe-B (Z)



Phe-D (E)



Phe-L (T)

Figure A5.5. Non covalent interaction (NCI) plots for selected complexes of Imz with Phe. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to 0.02 a.u. (blue).

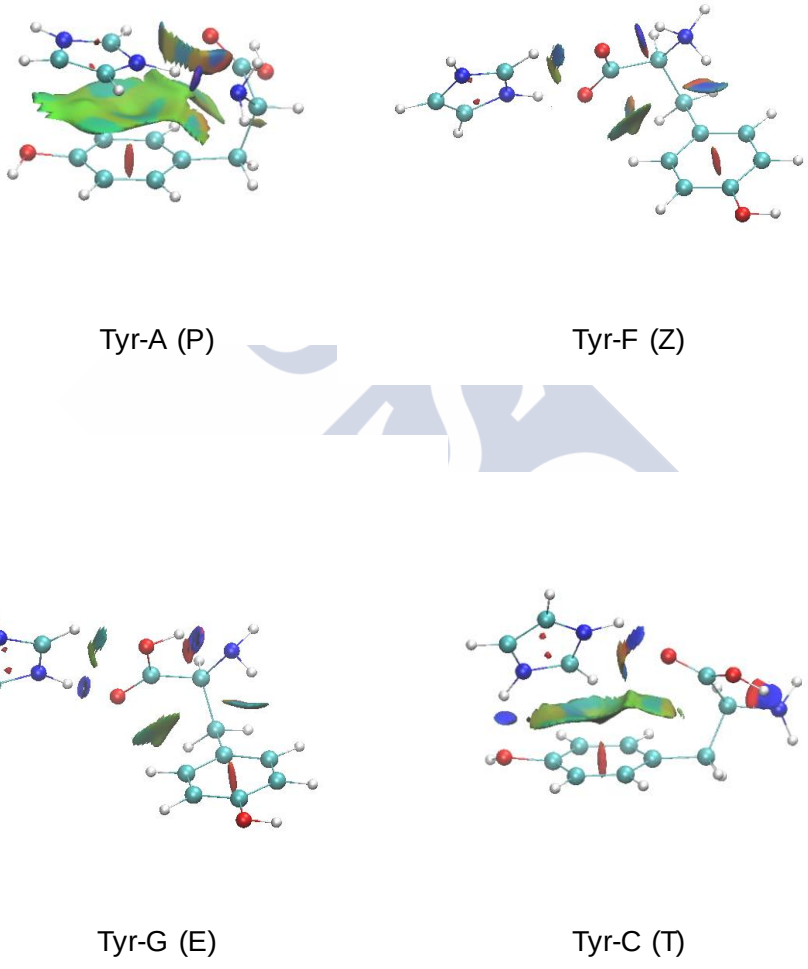


Figure A5.6. Non covalent interaction (NCI) plots for selected complexes of Imz with Tyr. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to 0.02 a.u. (blue).

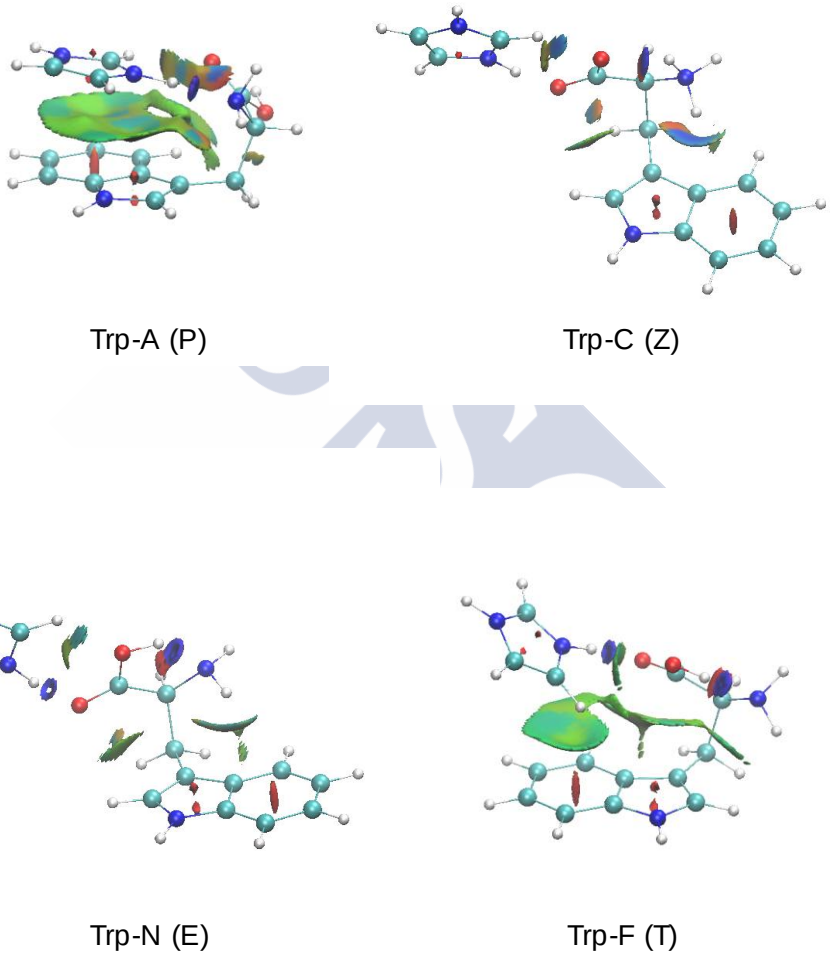


Figure A5.7. Non covalent interaction (NCI) plots for selected complexes of Imz with Trp. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to 0.02 a.u. (blue).

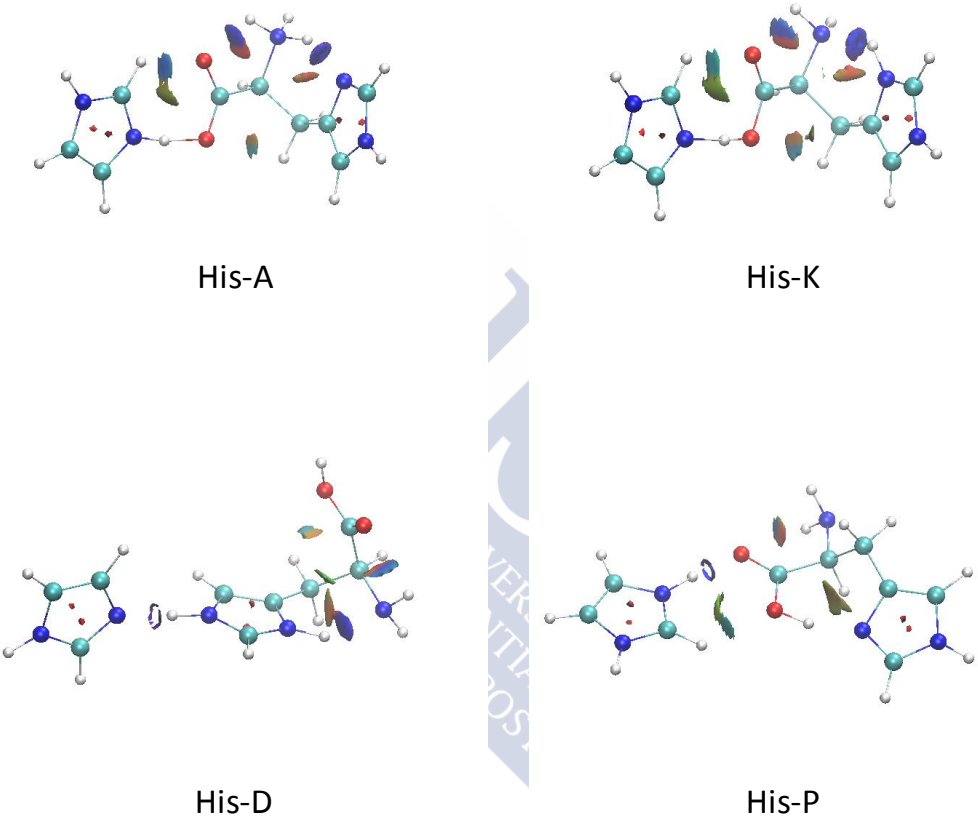


Figure A5.8a. Non covalent interaction (NCI) plots for selected complexes of Imz with His. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to 0.02 a.u. (blue).

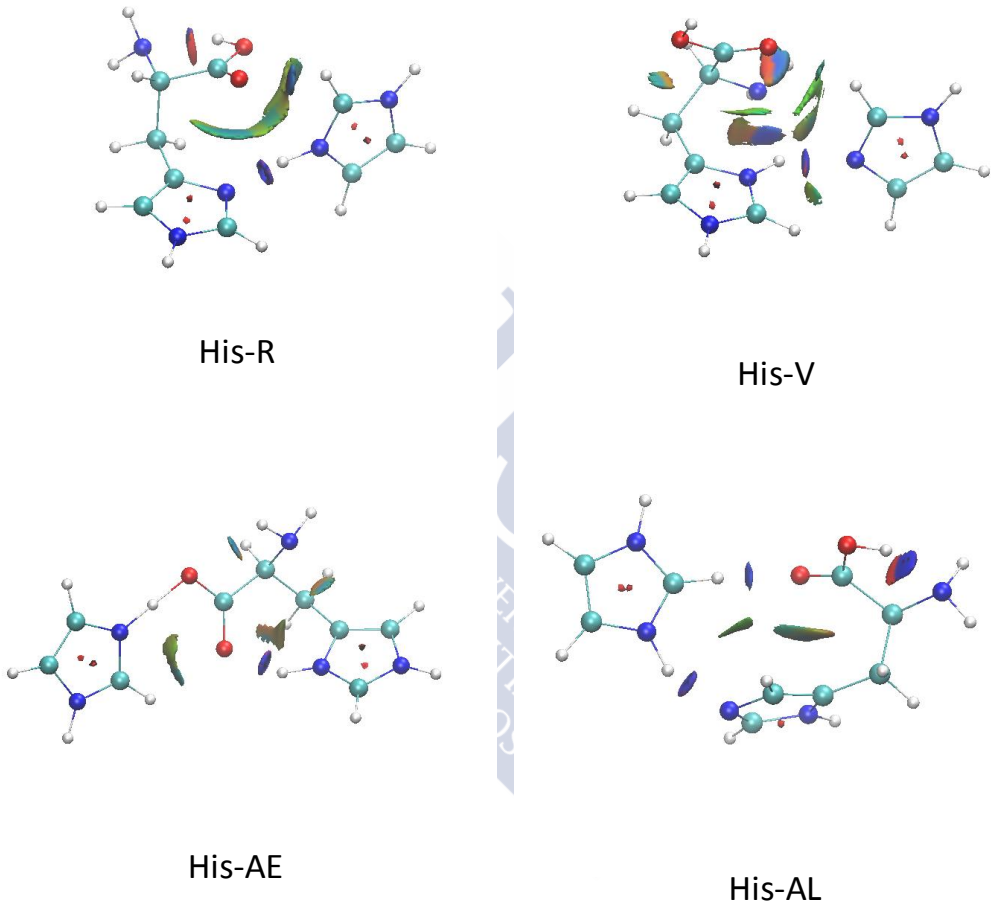


Figure A5.8b. Non covalent interaction (NCI) plots for selected complexes of Imz with His. The reduced gradient density isosurface amounts to 0.5 a.u. and the color scale runs from 0.02 a.u. (red) to 0.02 a.u. (blue).

Table A5.5. SAPT(DFT) contributions in mE_h to the interaction energy in Imz···Phe complexes as obtained employing PBE0 with the aug-cc-pVDZ basis set.

	E_{ele}	E_{rep}	E_{ind}	$E_{exch-ind}$	E_{dis}	$E_{exch-dis}$	δ_{HF}
Phe-A	-59.84	54.14	-38.58	20.10	-25.04	3.97	-10.78
Phe-B	-81.96	60.91	-45.64	22.80	-16.50	3.29	-14.52
Phe-C	-58.59	53.17	-38.02	19.91	-24.72	3.91	-10.39
Phe-D	-50.13	38.09	-28.64	13.32	-11.86	2.01	-8.04
Phe-E	-84.56	61.66	-46.80	23.21	-16.38	3.31	-14.78
Phe-F	-81.52	63.80	-47.47	23.20	-16.61	3.29	-16.33
Phe-G	-50.22	37.42	-28.27	13.09	-11.44	1.96	-7.89
Phe-H	-50.08	37.83	-28.54	13.11	-11.27	1.93	-8.37
Phe-I	-59.46	53.71	-40.19	19.78	-17.02	3.02	-11.85
Phe-J	-84.77	65.65	-49.43	23.95	-16.58	3.33	-17.08
Phe-K	-59.12	48.94	-36.62	17.76	-15.41	2.76	-10.52
Phe-L	-48.77	39.54	-29.64	14.26	-18.46	2.74	-7.15
Phe-M	-47.91	33.80	-25.56	10.98	-17.30	2.46	-4.77
Phe-N	-59.60	51.50	-38.21	18.65	-15.66	2.84	-11.32
Phe-O	-75.86	55.07	-41.54	20.51	-15.69	3.05	-13.02
Phe-P	-47.12	34.70	-26.95	11.79	-16.93	2.37	-6.23
Phe-Q	-80.21	54.38	-41.92	21.06	-15.64	3.10	-11.95
Phe-R	-55.89	46.07	-34.70	16.79	-14.88	2.64	-10.00
Phe-S	-77.41	54.64	-41.81	20.59	-15.53	3.05	-12.79
Phe-T	-41.85	29.86	-23.29	10.83	-15.23	2.13	-4.43

Table A5.6. SAPT(DFT) contributions in mE_h to the interaction energy in Imz···Tyr complexes as obtained employing PBE0 with the aug-cc-pVDZ basis set.

	E_{ele}	E_{rep}	E_{ind}	$E_{exch-ind}$	E_{dis}	$E_{exch-dis}$	δ_{HF}
Tyr-A	-62.22	56.64	-39.64	21.22	-27.64	4.36	-10.81
Tyr-B	-63.62	57.91	-40.89	22.49	-28.29	4.55	-10.19
Tyr-C	-56.34	42.03	-29.87	13.43	-21.02	3.04	-5.27
Tyr-D	-54.70	41.28	-28.96	12.36	-20.83	2.93	-5.31
Tyr-E	-63.46	58.31	-40.64	22.65	-28.80	4.63	-10.01
Tyr-F	-83.29	61.43	-46.08	22.87	-16.56	3.30	-14.71
Tyr-G	-50.68	38.12	-28.65	13.25	-11.92	2.01	-8.07
Tyr-H	-82.68	62.28	-46.63	23.18	-16.67	3.33	-15.00
Tyr-I	-55.99	46.04	-33.38	17.89	-23.59	3.58	-6.38
Tyr-J	-50.19	38.41	-28.91	13.25	-11.69	1.96	-8.58
Tyr-K	-49.49	38.39	-28.79	13.37	-11.96	2.02	-8.14
Tyr-L	-78.81	53.84	-41.12	20.57	-15.72	3.07	-11.91
Tyr-M	-86.08	62.52	-47.40	23.44	-16.51	3.34	-15.06
Tyr-N	-49.09	31.64	-23.76	9.72	-16.90	2.25	-3.99
Tyr-O	-85.92	62.44	-47.34	23.40	-16.50	3.34	-15.04
Tyr-P	-82.48	65.58	-48.82	23.69	-16.78	3.34	-17.04
Tyr-Q	-54.03	43.72	-31.31	15.44	-22.71	3.33	-6.11
Tyr-R	-50.84	37.72	-28.49	13.15	-11.53	1.97	-7.99
Tyr-S	-53.76	45.48	-32.40	16.99	-23.72	3.53	-6.62
Tyr-T	-52.81	40.76	-30.40	14.51	-22.31	3.16	-5.37

Table A5.7. SAPT(DFT) contributions in mE_h to the interaction energy in Imz···Trp complexes as obtained employing PBE0 with the aug-cc-pVDZ basis set.

	E_{ele}	E_{rep}	E_{ind}	$E_{exch-ind}$	E_{dis}	$E_{exch-dis}$	δ_{HF}
Trp-A	-65.96	60.99	-42.83	22.69	-30.50	4.77	-11.53
Trp-B	-55.41	43.61	-33.60	16.49	-21.22	3.22	-7.99
Trp-C	-85.98	65.88	-49.50	24.63	-17.23	3.50	-16.21
Trp-D	-56.26	42.20	-32.95	15.10	-19.48	2.92	-7.99
Trp-E	-58.57	50.86	-37.45	20.12	-28.88	4.39	-8.32
Trp-F	-53.89	42.89	-31.90	15.23	-19.57	2.93	-8.85
Trp-G	-48.75	35.62	-27.58	14.10	-20.93	3.03	-4.99
Trp-H	-54.37	45.32	-33.57	17.50	-22.70	3.47	-8.03
Trp-I	-90.41	66.76	-50.48	24.99	-17.10	3.50	-16.63
Trp-J	-86.14	70.15	-52.26	25.37	-17.51	3.54	-18.62
Trp-K	-62.57	59.84	-41.42	22.19	-30.44	4.76	-11.09
Trp-L	-66.39	62.52	-44.15	24.02	-29.96	4.77	-10.98
Trp-M	-64.96	59.39	-43.98	21.41	-20.46	3.45	-13.02
Trp-N	-50.88	39.45	-29.77	13.86	-12.15	2.08	-8.45
Trp-O	-85.15	68.24	-51.15	24.69	-17.88	3.47	-17.95
Trp-P	-90.59	70.81	-53.09	25.67	-17.34	3.54	-18.95
Trp-Q	-48.75	35.62	-27.58	14.10	-20.93	3.03	-4.99
Trp-R	-83.52	55.94	-43.18	21.73	-15.97	3.20	-12.43
Trp-S	-90.13	71.04	-53.27	25.68	-17.35	3.53	-19.06
Trp-T	-52.94	43.29	-32.57	15.50	-22.27	3.18	-8.04

Table A5.8. SAPT(DFT) contributions in mE_h to the interaction energy in Imz···His complexes as obtained employing PBE0 with the aug-cc-pVDZ basis set.

	E_{ele}	E_{rep}	E_{ind}	$E_{exch-ind}$	E_{dis}	$E_{exch-dis}$	δ_{HF}
His-A	-94.54	70.49	-52.35	26.51	-17.89	3.74	-17.68
His-B	-96.76	71.01	-52.98	26.70	-17.77	3.74	-17.92
His-C	-90.39	63.42	-47.60	24.42	-17.10	3.53	-14.88
His-D	-109.76	170.50	-171.33	49.54	-30.58	4.14	-105.23
His-E	-95.60	77.40	-57.00	27.94	-18.50	3.85	-21.32
His-F	-91.79	62.35	-47.33	24.28	-16.88	3.50	-14.44
His-G	-97.86	77.60	-57.34	28.01	-18.35	3.84	-21.37
His-H	-112.07	170.47	-171.78	49.51	-30.60	4.15	-105.73
His-I	-108.00	170.22	-170.23	49.42	-30.46	4.12	-104.18
His-J	-89.23	66.78	-49.58	24.79	-17.31	3.55	-16.99
His-K	-148.93	176.99	-169.25	63.47	-32.33	6.53	-93.37
His-L	-147.24	174.92	-164.47	62.44	-31.82	6.39	-90.46
His-M	-109.32	169.96	-168.75	48.99	-30.39	4.14	-103.90
His-N	-103.94	169.41	-165.84	48.88	-30.20	4.10	-101.29
His-O	-156.38	176.03	-167.64	62.88	-32.01	6.49	-92.25
His-P	-62.66	46.03	-34.38	16.54	-13.24	2.42	-10.33
His-Q	-139.31	170.00	-156.05	59.49	-30.41	6.00	-85.78
His-R	-70.09	55.48	-40.37	19.76	-18.92	3.24	-11.64
His-S	-66.02	47.35	-35.53	17.32	-15.75	2.82	-9.58
His-T	-61.29	40.42	-30.35	14.33	-15.10	2.49	-7.26
His-U	-62.70	46.89	-34.91	16.64	-13.18	2.41	-11.03
His-V	-130.42	191.47	-218.49	66.19	-41.17	5.70	-130.67
His-W	-60.92	39.60	-30.31	13.27	-15.12	2.44	-7.66
His-X	-61.70	43.45	-32.30	15.35	-12.83	2.28	-9.65
His-Y	-135.25	168.14	-153.80	58.23	-29.93	5.84	-85.27
His-Z	-67.20	47.31	-35.25	16.99	-13.28	2.46	-10.84
His-AA	-60.69	42.75	-31.40	14.77	-15.11	2.48	-8.75
His-AB	-152.48	174.68	-173.13	63.39	-31.78	6.29	-95.65
His-AC	-111.02	174.52	-176.11	51.93	-35.23	4.78	-108.01
His-AD	-71.13	50.91	-37.23	16.55	-16.88	2.86	-10.91
His-AE	-113.49	106.76	-79.49	37.06	-22.30	4.73	-34.13
His-AF	-54.73	31.80	-25.31	12.20	-11.73	2.01	-5.19
His-AG	-58.13	34.80	-25.34	11.47	-12.97	2.16	-5.61
His-AH	-113.56	172.19	-157.49	48.64	-34.19	4.63	-94.28
His-AI	-114.69	174.22	-163.96	50.43	-34.92	4.74	-98.80