

**UNIVERSIDAD DE SANTIAGO DE COMPOSTELA
DEPARTAMENTO DE BIOQUÍMICA Y BIOLOGÍA
MOLECULAR**



**ESTRUCTURAS CRISTALOGRÁFICAS DE
PROTEÍNAS DE VIRUS ANIMALES
IMPLICADAS EN LA INTERACCIÓN VIRUS-
HOSPEDADOR**

TESIS DOCTORAL

PABLO GUARDADO CALVO

2010

Mark Johan van Raaij, Científico titular del Instituto de Biología Molecular de Barcelona (IBMB-CSIC).

José Manuel Martínez Costas, Profesor titular del Departamento de Bioquímica y Biología Molecular de la Universidad de Santiago de Compostela.

INFORMAN:

Que la presente Tesis Doctoral titulada "Estructuras cristalográficas de proteínas de virus animales implicadas en la interacción virus-hospedador", que presenta Pablo Guardado Calvo para optar al grado de Doctor en Biología, ha sido realizada en este Departamento bajo la dirección del Doctor Mark Johan van Raaij y el Doctor José Manuel Martínez Costas y, considerando que se haya concluída, autorizan su presentación para que sea evaluada por el Tribunal correspondiente.

Y para que así conste, firmo el presente informe:

Fdo: Mark Johan van Raaij
Costas

Fdo: José Manuel Martínez

Fdo: Pablo Guardado Calvo

Santiago de Compostela, 22 de Enero de 2010

El trabajo recogido en la presente memoria ha sido realizado en el Departamento de Bioquímica y Biología Molecular de la Universidad de Santiago de Compostela, bajo la dirección de los doctores Mark Johan van Raaij y José Manuel Martínez Costas.

AGRADECIMIENTOS

En primer lugar me gustaría agradecer a Ana y a Yaguito todo su amor y su apoyo y a mis padres y hermanas su apoyo constante a lo largo de los años. Sin todos ellos este trabajo hubiese sido mucho más difícil.

Un agradecimiento especial al Dr. Mark van Raaij, por haberme aceptado en su grupo y haberme dirigido la tesis. También me gustaría agradecer al Dr. Jose Manuel Martínez Costas por haber sido mi codirector de tesis, al Dr. Javier Benavente Martínez por haberme dejado usar su laboratorio durante tanto tiempo y a Lois Hermo Parrado que me ha enseñado mucho de lo que se del trabajo de laboratorio.

También me gustaría agradecer a la gente con la que he compartido tanto tiempo en el departamento. A Rubén que me ha invitado a tantos cafés durante todos estos años. A Félix e Iria. A la gente con la que he compartido espacio en el laboratorio de Javier durante todos estos años: Javi, Alberto, Irene, Rebeca, Leticia, Iria, Salvador, Marcelo, Patricia, Lorena y Lisa. A la gente del laboratorio de Mark: Jose, Sergio, Patricia F, Carmela y Antonio. A la gente del CACTUS que tanto me ha ayudado en la recogida de datos: Antonio, Bruno, Oscar, Inés y Guillermo. A la gente que trabaja en la facultad como Rosa, Carmela, Noya y un largo etcétera.

ÍNDICE

Capítulo 1 - INTRODUCCIÓN	1
1.1 - Los reovirus	3
1.1.1 - La familia <i>Reoviridae</i>. El género <i>Orthoreovirus</i>. El reovirus aviar	3
1.1.2 - El genoma de los reovirus aviares	4
1.1.3 - Las proteínas de los reovirus aviares	5
1.1.3.1 - La proteína λ A	6
1.1.3.2 - La proteína λ B	6
1.1.3.3 - La proteína λ C	6
1.1.3.4 - La proteína μ A	7
1.1.3.5 - La proteína μ B	7
1.1.3.6 - La proteína μ NS	7
1.1.3.7 - La proteína σ A	7
1.1.3.8 - La proteína σ B	8
1.1.3.9 - La proteína σ C	8
1.1.3.10 - La proteína p10	9
1.1.3.11 - La proteína p17	10
1.1.3.12 - La proteína σ NS	10
1.1.4 - La partícula reoviral	10
1.1.5 - El ciclo replicativo de los reovirus	13
1.1.5.1 - Penetración y pérdida de la cápside externa	14
1.1.5.2 - Expresión de los genes virales	15
1.1.5.3 - Morfogénesis y salida	16
1.1.6 - Interacciones virus-célula hospedadora	17
1.1.6.1 - Fusión celular	17
1.1.6.2 - Resistencia al interferón	17
1.2 - Los adenovirus	18
1.2.1 - La familia <i>Adenoviridae</i>	18
1.2.2 - Clasificación de la familia <i>Adenoviridae</i>	19
1.2.3 - Estructura del virión	20
1.2.4 - El genoma de los adenovirus	23
1.2.5 - La entrada de los adenovirus en la célula hospedadora	24
1.2.6 - El uso de los adenovirus en terapia génica	26
1.2.6.1 - Coxsackievirus-adenovirus receptor (CAR)	27
1.2.6.2 - CD46	28
1.2.6.3 - Ácido siálico	29
1.2.6.4 - Otros receptores primarios	29
1.2.6.5 - Integrinas	29
1.2.7 - El adenovirus aviar de tipo 1	30
1.2.8 - El adenovirus porcino de tipo 4 (cepa NADC-1)	32
1.2.9 - Objetivos	35
Capítulo 2 - MATERIALES Y MÉTODOS	37
2.1 - Materiales	39
2.1.1 - Materiales biológicos	39
2.1.1.1 - Bacterias	39
2.1.1.2 - Plásmidos	39
2.1.2 - Materiales no biológicos	39
2.1.2.1 - Disoluciones y tampones	39
2.1.2.2 - Medios de cultivo	40
2.1.2.3 - Cebadores	40
2.2 - Métodos	41
2.2.1 - Manipulación de células bacterianas	41
2.2.1.1 - Obtención de bacterias competentes y transformación	41
2.2.2 - Manipulación ácidos nucleicos	42
2.2.2.1 - Purificación de DNA de plásmidos	42
2.2.2.2 - Reacción en cadena de la polimerasa (PCR)	42
2.2.2.3 - Aislamiento de fragmentos de DNA y su purificación	42

2.2.2.4 - Tratamiento de DNA con endonucleasas	43
2.2.3 - Manipulación de proteínas	43
2.2.3.1 - Análisis por electroforesis SDS-PAGE	43
2.2.3.2 - Cromatografía de afinidad	43
2.2.3.3 - Cromatografía de intercambio aniónico	44
2.2.3.4 - Cromatografía de exclusión molecular	44
2.2.4 - Métodos cristalográficos	44
2.2.4.1 - Crecimiento de cristales	44
2.2.4.2 - Recogida de datos de difracción	46
2.2.4.3 - Procesado de los datos de difracción	46
2.2.4.4 - Resolución de estructuras	46
2.2.4.5 - Refinamiento de estructuras	48
2.3 - Estrategias de expresión, purificación, cristalización y resolución estructural	50
2.3.1 - La proteína σC	50
2.3.1.1 - Construcción de los vectores de expresión	50
2.3.1.2 - Expresión y purificación de σ C(27-326), σ C(49-326), σ C(117-326) y σ C(151-326)	51
2.3.1.3 - Cristalización de σ C(117-326) y σ C(151-326)	52
2.3.2 - La proteína σA	52
2.3.2.1 - Expresión y purificación de la proteína σ A	52
2.3.2.2 - Cristalización de la proteína σ A	54
2.3.2.4 - Experimentos de ultracentrifugación analítica	54
2.3.2.5 - Experimentos de microscopía electrónica	55
2.3.3 - El dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1	55
2.3.3.1 - Construcción de los vectores de expresión	55
2.3.3.2 - Expresión y purificación del dominio de cabeza (aminoácidos 579-793)	56
2.3.3.3 - Cristalización del dominio de cabeza	56
2.3.4 - Los dominios de la fibra del adenovirus porcino de tipo 4	57
2.3.4.1 - Construcción de los vectores de expresión	57
2.3.4.2 - Expresión y purificación del dominio de cabeza (116-291)	57
2.3.4.3 - Expresión y purificación de los dominios de shaft y cabeza (40-291)	58
2.3.4.4 - Expresión y purificación del dominio de galectina (393-703)	58
2.3.4.5 - Cristalización del dominio de cabeza	59
2.3.4.6 - Cristalización del dominio de shaft y cabeza	59
2.3.4.7 - Cristalización del dominio de galectina y de los complejos galectina-azúcar	59
Capítulo 3 - LA PROTEÍNA σC	61
3.1 - Purificación y cristalización	63
3.1.1 - Construcción de los vectores de expresión	63
3.1.2 - Expresión y purificación de mutantes de delección N-terminal de la proteína σ C	63
3.1.3 - Cristalización de mutantes de delección N-terminal de la proteína σ C	64
3.2 - Determinación estructural de los dominios C-terminal de la proteína σC	65
3.2.1 - Determinación estructural de σ C(151-326)	65
3.2.2 - Determinación estructural de σ C(117-326)	67
3.3 - Estructura general de la proteína σC	67
3.4 - La estructura de los dominios C-terminales de σC	68
3.4.1 - El dominio de cabeza	69
3.4.2 - El dominio fibroso	70
3.5 - Comparación estructural y funcional con la proteína σ1 y otras fibras virales	73
3.6 - Estabilidad y plegamiento de σ C	76
3.7 - Conclusión	78
Capítulo 4 - LA PROTEÍNA σA	79
4.1 - Purificación y cristalización	81
4.1.1 - Expresión y purificación	81
4.1.2 - Cristalización	82
4.2 - Determinación estructural	82

4.3 - La estructura y topología de σA	85
4.4 - Comparación estructural y funcional con otras proteínas	88
4.5 - Estudio de la interacción de la proteína σA con dsRNA	89
4.6 - Estudio de las interacciones entre las moléculas de σA en el cristal	90
4.7 - Estudio de la estequiometría de la interacción σA-dsRNA	94
4.7.1 - Estudios de ultracentrifugación analítica	95
4.7.2 - Estudios de microscopía electrónica	96
4.8 - Conclusión	97
Capítulo 5 - LA FIBRA LARGA DEL ADENOVIRUS AVIAR 1	99
5.1 - Purificación y cristalización	101
5.1.1 - Construcción de los vectores de expresión	101
5.1.2 - Expresión, purificación y cristalización del dominio de cabeza	102
5.2 - Determinación estructural de dominio de cabeza de la fibra larga del adenovirus CELO	102
5.3 - La estructura del dominio de cabeza	103
5.4 - Estabilidad y plegamiento del dominio de cabeza	106
5.5 - Comparación con otras proteínas de estructura conocida	106
5.6 - La capacidad de unión al receptor	109
5.7 - Conclusión	112
Capítulo 6 - LA FIBRA DEL ADENOVIRUS PORCINO 4	113
6.1 - Purificación y cristalización de los dominios de la fibra del adenovirus porcino NADC-1	115
6.1.1 - Construcción de los vectores de expresión	115
6.2.2 - Expresión y purificación del dominio de cabeza (116-291) y el dominio fibroso (40-291)	116
6.2.3 - Expresión y purificación del dominio de galectina (393-703)	117
6.2.4 - Cristalización del dominio de cabeza	117
6.2.5 - Cristalización del dominio de shaft y cabeza	117
6.2.6 - Cristalización del dominio de galectina y de los complejos galectina-azúcar	117
6.3 - Determinación estructural de los dominios de cabeza y de galectina	118
6.3.1 - Determinación estructural del dominio de cabeza	118
6.3.2 - Determinación estructural del dominio de galectina	119
6.4 - La estructura del dominio de cabeza y su capacidad de unirse a receptores celulares	121
6.5 - La estructura del dominio de galectina	123
6.6 - El dominio de galectina es capaz de interactuar con carbohidratos	125
6.7 - Las estructuras de los complejos galectina-carbohidrato	126
6.8 - Experimentos de resonancia de superficie de plasmones (SPR)	130
6.9 - Comparación con otras galectinas repetidas en tándem	133
6.10 - Conclusión	136
Capítulo 7 - DISCUSIÓN	137
7.1- Las fibras de reovirus y adenovirus son similares	139
7.2- Las fibras del reovirus de mamífero y aviar presentan diferencias significativas	140
7.2.1- La regulación del estado trimérico del dominio de cabeza	140
7.2.2 - La interfaz trimérica en los dominios de cabeza de adenovirus	143
7.2.3 - Las posibles regiones bisagra	144
7.2.4 - Las probables regiones de interacción con el receptor celular en σC	146
7.3- Las fibras de los adenovirus aviar y porcino unen receptores desconocidos	147
7.3.1 - El receptor celular del adenovirus aviar de tipo 1	147
7.3.2 - El receptor celular del adenovirus porcino de tipo 4	149
7.4 - La primera estructura de una galectina repetida en tándem	150
7.5- Implicaciones en enfermedades porcinas y humanas	151
7.6 - Los dominios de σA tienen funciones diferenciadas	152
7.7 - La proteína σA en los reovirus fusogénicos	153
7.8 - Conclusión	156

REFERENCIAS	157
ARTÍCULOS PUBLICADOS	181

Capítulo 1

INTRODUCCIÓN

Capítulo 1

Introducción

1.1 - Los reovirus

1.1.1 - La familia *Reoviridae*. El género *Orthoreovirus*. El reovirus aviar

La familia *Reoviridae* es, de las familias de virus que tienen genomas de dsRNA, la que tiene un rango de hospedadores más amplio. El genoma de los miembros de la familia *Reoviridae* es segmentado y puede estar compuesto de 10, 11 o 12 segmentos de dsRNA, y cada segmento puede codificar entre 1 y 3 proteínas en sólo una de sus cadenas. Todos los miembros de esta familia sintetizan sus mRNAs mediante el uso de factores virales localizados dentro de sus partículas icosaédricas.

El género *Orthoreovirus* incluye al miembro prototipo de la familia *Reoviridae*, el reovirus de mamífero que ha sido profundamente estudiado tanto a nivel bioquímico como estructural. El otro virus modelo de esta familia es el reovirus aviar. Los reovirus aviares y de mamífero son los dos grupos más importantes del género *Orthoreovirus* y, aunque comparten muchas características morfológicas y fisicoquímicas, presentan algunas diferencias importantes tanto a nivel biológico como serológico (Rosenberger *et al.*, 1989; Bodelon *et al.*, 2001; Duncan & Sullivan, 1998; Schnitzer, 1985; Spandidos & Graham, 1976a). Quizás las diferencias más características son que los reovirus aviares presentan actividad fusogénica (y los de mamífero no) mientras que los reovirus aviares carecen de la actividad hemoaglutinante característica de sus homólogos de mamífero (Nibert & Shiff, 2001). Dentro del género *Orthoreovirus* se encuentran los reovirus de mamífero no fusogénicos ya mencionados y tres especies de reovirus de mamífero fusogénicos (Nelson Bay virus, Baboon reovirus y reovirus Melaka) y varios reovirus fusogénicos aislados de reptiles (Mertens, 2004).

El análisis filogenético de la secuencia de la proteína σC de reovirus aviar ha permitido dividir a los reovirus aviares en cinco grupos diferenciados (Kant *et al.*, 2003). Sin embargo, no se ha podido establecer una correlación clara entre estos grupos y la patogenicidad que algunos de ellos presentan. El alto grado de

variabilidad antigénica que presentan los reovirus aviares ha impedido desarrollar una clasificación basada en serotipos.

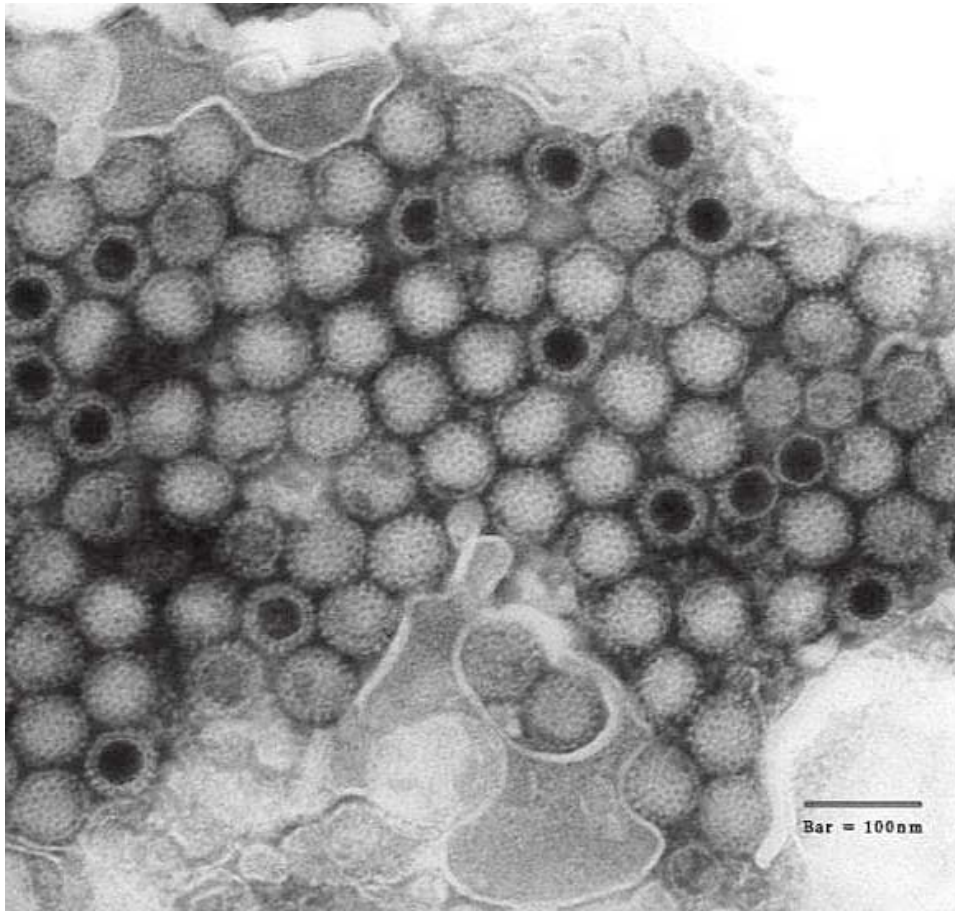


Figura 1.1 Fotos de microscopia de reovirus de mamífero. Obtenida de ICTVdB (<http://www.ncbi.nlm.nih.gov/ICTVdb/>)

1.1.2 - El genoma de los reovirus aviares

Los reovirus aviares poseen un genoma compuesto por 10 segmentos de dsRNA, los cuales están divididos en tres clases de acuerdo a su movilidad electroforética; tres segmentos de la clase L (large), tres de la clase M (medium) y cuatro de la clase S (small) (Spandidos & Graham, 1976a). A excepción del segmento S1, que codifica para tres proteínas, todos los demás segmentos codifican para una única proteína.

El mRNA viral está codificado por la cadena positiva de los segmentos genómicos. Esta cadena contiene un cap de tipo 1 en su extremo 5', mientras que la cadena negativa contiene un grupo pirofosfato en la misma posición. (Martínez-Costas *et al.*, 1995). Todos los miembros de la familia *Reoviridae* poseen en los extremos de todos sus genes unas pequeñas secuencias muy conservadas (Patton &

Spencer, 2000), que podrían actuar como señales de transcripción, replicación y/o encapsidación de los transcritos virales. La coinfección de células aviares con dos cepas reovirales distintas, produce una progenie viral que puede contener segmentos genómicos de ambos virus parentales. Este fenómeno, denominado *reassorting*, ha sido muy utilizado para caracterizar las propiedades fenotípicas de los segmentos genómicos virales.

Tabla I. Los genes del reovirus aviar y las proteínas que codifican

dsRNA	Proteína	Equivalente en reovirus de mamífero	Localización	Función
L1	λ A	λ 1	Core	Proteína principal del core
L2	λ B	λ 3	Core	RNA polimerasa
L3	λ C	λ 2	Core	Guaniltransferasa
M1	μ A	μ 2	Core	Cofactor de la RNA polimerasa
M2	μ B	μ 1	Cápside externa	
M3	μ NS	μ NS	No estructural	Formación de las factorías virales
S1	σ C	σ 1	Cápside externa	Unión al receptor
	p17	σ 1s (1)	No estructural	Unión a dsDNA
	p10	(2)	No estructural	Fusógeno
S2	σ A	σ 2	Core	Unión a dsRNA
S3	σ B	σ 3	Cápside externa	
S4	σ NS	σ NS	No estructural	Unión a ssRNA

(1) La proteína σ 1s no es la equivalente de p17 del reovirus aviar
 (2) La proteína p10 no tiene ningún equivalente en los reovirus de mamífero

1.1.3 - Las proteínas de los reovirus aviares

El genoma del reovirus aviar codifica para 12 proteínas, 8 proteínas estructurales y 4 no estructurales. Las proteínas no estructurales no se incorporan al virión pero se expresan en las células infectadas (Martínez-Costas *et al.*, 1997; Varela y Benavente, 1994). Las proteínas codificadas por los genes de la clase L se denominan lambda (λ), las codificadas por los de la clase M, mu (μ) y las codificadas por los de la clase S, sigma (σ) (Tabla I). A cada proteína se le asigna un sufijo alfabético en orden inverso a su movilidad electroforética. El virión contiene al menos 10 proteínas estructurales diferentes, 8 de las cuales (λ A, λ B, λ C, μ A, μ B, σ A, σ B y σ C) son productos primarios de traducción de los mRNAs, mientras que las otras dos, μ BN y μ BC, se originan por procesamiento proteolítico del precursor μ B (Varela *et al.*, 1996). Los genes M3 y S4 codifican para dos proteínas no estructurales mayoritarias, denominadas μ NS y σ NS respectivamente (Schnitzer *et al.*, 1982; Varela & Benavente, 1994). El gen S1, que es tricistrónico, codifica en sus dos primeras pautas

de lectura (ORF) para otras dos proteínas no estructurales, denominadas p10 y p17 (Bodelon *et al.*, 2001; Shmulevitz *et al.*, 2002).

Aunque serían necesarios estudios en profundidad sobre los mecanismos que utilizan los ribosomas para traducir cada uno de los tres cistrones del gen S1, es posible inferir algunos detalles de este mecanismo a partir de un análisis detallado de la secuencia del mRNA s1 y del conocimiento del mecanismo de traducción de otros transcritos policistrónicos. Cuando se examinan las secuencias de nucleótidos que rodean a los codones de iniciación de cada uno de los tres cistrones se observa que el codón de iniciación del primer cistrón está situado en un contexto muy débil, mientras que los contextos de los codones de iniciación del segundo y tercer cistrón son fuertes (Kozak, 2002). De esta forma y teniendo en cuenta que el mRNA s1 tiene un cap en su extremo 5' es posible especular que los dos primeros cistrones son traducidos mediante un mecanismo de *leaky scanning* (Kozak, 1991) lo que explicaría su baja eficiencia de traducción, el primer cistrón debido a su contexto débil y el segundo debido a la lejanía del extremo 5'. Por otro lado el codón de iniciación del tercer cistrón está demasiado alejado del cap por lo que la probabilidad de que el ribosoma lo alcance mediante *leaking scanning* es baja, para la traducción de este cistrón se ha propuesto que el ribosoma pueda anclarse en una secuencia interna sin necesidad de un cap (Jackson & Kaminski, 1995).

1.1.3.1 - La proteína λA

La proteína λA es la proteína viral de mayor tamaño. Es la encargada de formar la cápside interna del reovirión.

1.1.3.2 - La proteína λB

Por su tamaño, número de copias en la partícula viral y localización en el interior del core, probablemente pudiera tratarse de la RNA polimerasa viral (Martínez-Costas *et al.*, 1997), aunque hasta la fecha no existe evidencia experimental que soporte esta hipótesis.

1.1.3.3 - La proteína λC

Aunque estructuralmente forma parte de la cápside interna se extiende hasta la cápside externa (Martínez-Costas *et al.*, 1997; Zhang *et al.*, 2005). En el contexto del

reovirión es un pentámero localizado en los ejes de simetría 5 de la cápside. Es la proteína responsable de la adición de la caperuza 5' del RNA gracias a su actividad guanililtransferasa que reside en un fragmento amino terminal de 42 kDa.

1.1.3.4 - La proteína μA

Es una proteína estructural minoritaria de la cápside interna (Martínez-Costas *et al.*, 1997). No existen estudios sobre sus propiedades o su función. Su homóloga en reovirus de mamífero ($\mu 2$) interacciona con microtúbulos y con la proteína no estructural μNS , y se cree que es la proteína que media la interacción entre las factorías virales y los microtúbulos (Parker *et al.*, 2002). Sin embargo, en el reovirus aviar las factorías virales no aparecen asociadas a microtúbulos por lo que probablemente μA no sea capaz de interaccionar con ellos.

1.1.3.5 - La proteína μB

La proteína μB es una proteína miristoilada (Varela *et al.*, 1996) que sufre un proceso proteolítico mediado por un factor viral (Varela & Benavente, 1994) que la escinde en dos fragmentos μBC y μBN . Ambos fragmentos son componentes estructurales de la cápside externa del reovirión (Martínez-Costas *et al.*, 1997). El fragmento μBC parece desempeñar algún papel en la internalización del virus.

1.1.3.6 - La proteína μNS

La proteína μNS es el mínimo factor viral necesario para la formación de las factorías virales en células infectadas, por lo que juega un papel importante en las primeras etapas de la morfogénesis viral (Tourís-Otero *et al.*, 2004). La proteína μNS sufre un proceso proteolítico durante la infección, que genera dos fragmentos denominados μNSN (15 kDa) y μNSC (55 kDa). Hasta la fecha el significado biológico de este procesamiento es desconocido.

1.1.3.7 - La proteína σA

La proteína σA es un componente estructural de la cápside interna (Martínez-Costas *et al.*, 1997) que desempeña un papel estructural importante en la morfogénesis del reovirus aviar, ya que actúa estabilizando la cápside interna y promoviendo la formación de la cápside externa. Se ha demostrado que es capaz de

unir dsRNA *in vitro* independientemente de su secuencia con alta afinidad (Tourís-Otero *et al.*, 2005). Algunos estudios han mostrado que su capacidad para interactuar con el dsRNA depende de su conformación nativa puesto que pierde esta capacidad cuando se desnaturaliza. Además de su papel estructural, existen evidencias que apuntan a que esta proteína es la responsable de la resistencia del reovirus aviar al interferón, y que de alguna manera esta actividad anti-interferón está relacionada con la inhibición de la activación de la proteína PKR. Dado que el sustrato natural de la PKR es el dsRNA, probablemente la función anti-interferón de σA esté relacionada con su capacidad de interactuar con dsRNA (Martínez-Costas *et al.*, 2000; González-López *et al.*, 2003).

Su homóloga en reovirus de mamífero ($\sigma 2$) desempeña un papel estructural similar (Reinisch *et al.*, 2000) aunque es incapaz de interactuar con dsRNA. Además, la resistencia que el reovirus de mamífero presenta a la acción antiviral del interferón parece estar relacionada con la proteína de la cápside externa $\sigma 3$ (Yue & Shatkin, 1997). De esta forma la proteína σA parece ser homóloga estructuralmente de $\sigma 2$ pero análoga funcionalmente de la proteína $\sigma 3$.

1.1.3.8 - La proteína σB

La proteína σB es un componente estructural de la cápside externa del reovirión (Martínez-Costas *et al.*, 1995; Varela *et al.*, 1996). Al contrario que su homóloga de mamífero ($\sigma 3$) la proteína σB no une dsRNA en solución, lo que sugiere que no tiene actividad anti-interferón (Tourís-Otero *et al.*, 2005). Se asocia rápidamente en el citoplasma con μB y μBC formando un complejo ternario con cantidades estequiométricamente iguales de las tres proteínas. Este complejo preformado es el que se incorpora al reovirión (Tourís-Otero *et al.*, 2004a,b).

1.1.3.9 - La proteína σC

La proteína σC está codificada por la tercera pauta de lectura del gen S1 y es la proteína responsable de reconocer el receptor celular (todavía desconocido). La proteína σC parece ser la única proteína viral presente en extractos solubles que es capaz de adherirse a monocapas de células aviares (Martínez-Costas *et al.*, 1997). Además, es capaz de competir con el reoviriones aviares purificados bloqueando su entrada en células permisivas (Grande *et al.*, 2000). La proteína σC es un

homotrímico en su estado nativo, y su capacidad para unirse al receptor esta asociada a esta forma oligomérica (Grande *et al.*, 2002). Mediante experimentos de saturación de la unión, realizados utilizando σC expresada en bacterias, se ha estimado que cada célula CEF contiene 2.2×10^5 moléculas receptoras para la proteína σC (Grande *et al.*, 2000). Tiene un 20% de identidad de secuencia con su homólogo de mamífero $\sigma 1$, la cual se une a JAM-A (*Junction Adhesión Molecule A*). Ambas proteínas no se unen al mismo receptor ya que los reovirus de mamífero no son capaces de unirse a células aviares (Barton *et al.*, 2001).

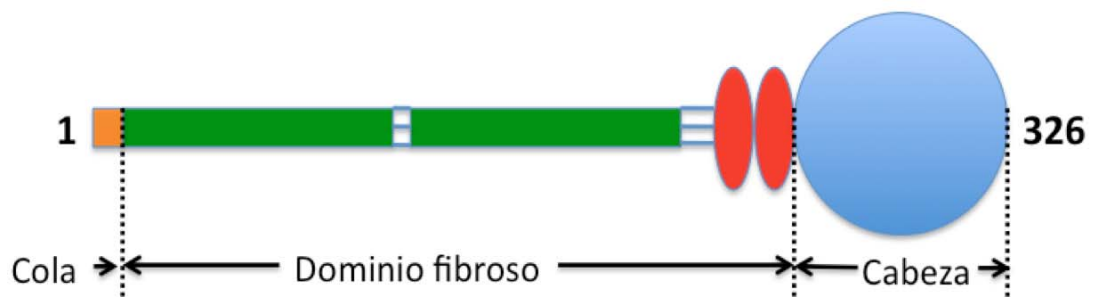


Figura 1.2 Diagrama en donde se muestran las regiones de σC fácilmente distinguibles en su secuencia. La pequeña región N-terminal responsable de la interacción con λC (cola) se muestra en naranja, las dos regiones probablemente plegadas formando un triple *coiled coil* se muestran en verde, las dos probables repeticiones en triple β espiral en rojo y el dominio globular de cabeza en azul.

La proteína σC está compuesta fundamentalmente por un dominio fibroso y un dominio de cabeza globular. En el dominio fibroso se pueden distinguir a su vez dos regiones, una región más C-terminal probablemente plegada en repeticiones en triple β espiral y una larga región central cuya secuencia sugiere la existencia de dos regiones plegadas formando un triple *coiled coil* (Figura 1.2), por último el extremo más N-terminal de la fibra que es el encargado de anclar σC a su base, un pentámero de la proteína λC .

1.1.3.10 - La proteína p10

La proteína p10 está codificada por la primera pauta de lectura del gen S1. Es una proteína de membrana con un dominio transmembrana central que separa un dominio citoplasmático y un dominio extracelular de tamaños similares (Shmulevitz & Duncan, 2000). La expresión de esta proteína en células transfectadas induce una

fusión célula-célula lo que sugiere que p10 juega un papel importante en el fenotipo fusogénico que muestran los reovirus aviares (Bodelon *et al.*, 2001).

1.1.3.11 - La proteína p17

La proteína p17 está codificada por la segunda pauta de lectura del gen S1 y no tiene ninguna homología de secuencia con proteínas virales o celulares conocidas. Tiene capacidad de unir DNA y se acumula en el núcleo de células infectadas y transfectadas (Costas *et al.*, 2005).

1.1.3.12 - La proteína σ NS

La proteína σ NS se acumula en las factorías virales gracias a su asociación con μ NS, lo que la convierte en candidata a jugar un papel importante en el empaquetamiento del RNA y su replicación, aunque sería necesario un estudio más detallado para confirmar esta hipótesis. Tiene la capacidad de unir ssRNA de una manera independiente de secuencia *in vitro* (Tourís-Otero *et al.*, 2005; Yin & Lee, 1998). Además, se encuentra formando parte de complejos ribonucleoproteicos en el citoplasma de las células infectadas, lo que sugiere que existe una asociación con ssRNA *in vivo* (Tourís-Otero *et al.*, 2005).

1.1.4 - La partícula reoviral

Los reoviriones aviares son partículas icosaédricas no envueltas de unos 85 nm de diámetro externo (Spandidos & Graham, 1976; Zhang *et al.*, 2005). Contienen un genoma formado por 10 segmentos de dsRNA encapsidados dentro de dos cápsides icosaédricas concéntricas. La cápside interna está formada por los polipéptidos λ A, λ B, μ A y σ A y la cápside externa por μ B, σ B y σ C. La proteína λ C forma parte de ambas cápsides (Figura 1.3).

La estructura de las partículas del reovirus aviar ha sido estudiada por microscopía electrónica y reconstrucción de imágenes a una resolución efectiva de 14.6 Å (Zhang *et al.*, 2005) utilizando la cepa 138 (Figura 1.4). Los datos derivados de este trabajo así como los estudios de difracción de rayos X llevados a cabo con partículas del reovirus de mamífero (Reinisch *et al.*, 2000) han permitido tener una información detallada sobre la organización de las proteínas estructurales en la

partícula viral. Tanto la estructura como la organización de las proteínas en el reovirión aviar es muy similar a la observada en el reovirus de mamífero (Zhang *et al.*, 2003). De esta manera el diámetro del reovirión es de 85.7 nm, mientras que el diámetro de la cápside interna es de 53.6 nm. Dentro de la cápside interna se aprecian anillos concéntricos que probablemente representen los segmentos de dsRNA empaquetados, también dentro de esta cápside y localizados alrededor de los ejes pentaméricos se encuentran los complejos de transcripción, formados por la RNA polimerasa λB y su cofactor μA . La cápside interna *per se* está formada por 120 copias de la proteína λA , y sobre su superficie se sitúan 150 copias de la proteína σA repartidas en tres posiciones distintas, 60 alrededor de los ejes pentaméricos, 60 alrededor de los ejes ternarios y las restantes 30 alrededor de los ejes binarios del icosaedro. Además, en los mapas de densidad electrónica se aprecia la presencia de 12 torres pentaméricas, cada una de ellas organizadas alrededor de uno de los ejes pentaméricos y compuestas por 5 copias de la proteína λC (Martínez-Costas *et al.*, 1997).

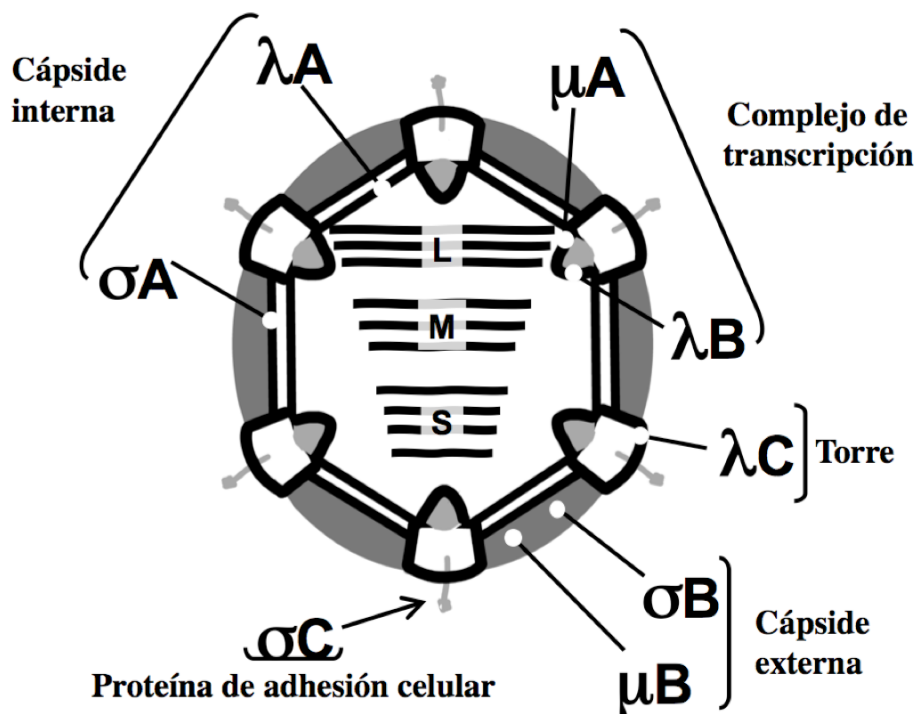


Figura 1.3 Estructura del reovirión aviar. Modificado de Benavente & Martínez-Costas, 2007

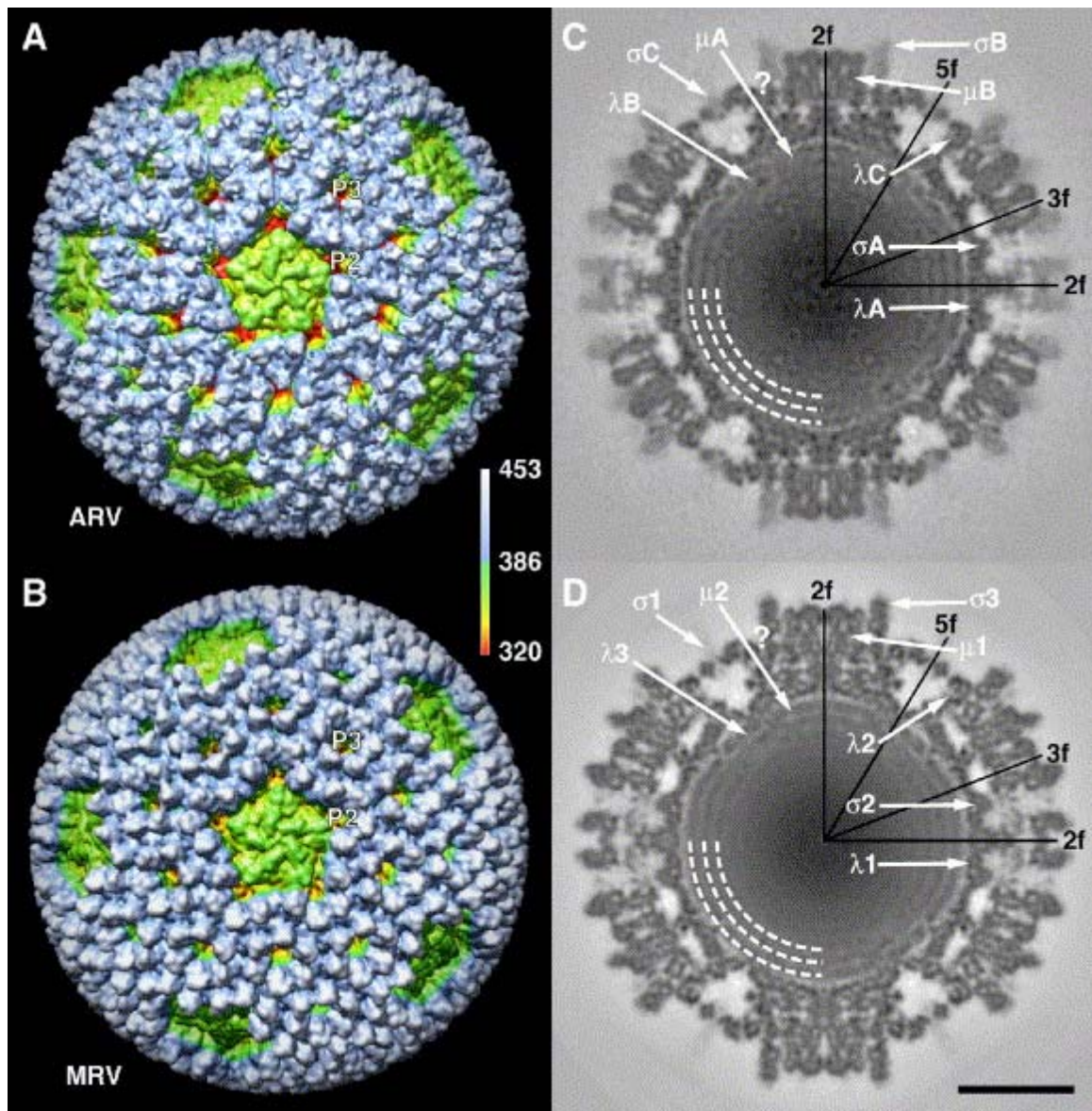


Figura 1.4 Estructuras de criomicroscopía electrónica del reovirus aviar 138 (AVR) y el reovirus de mamífero de tipo 3 (MRV). En (A) y (B) se muestran las superficies del reovirus aviar 138 y del reovirus de mamífero de tipo 3, respectivamente. Los colores representan distancias radiales medidas desde el centro de cada partícula en Å. Se muestra la localización de un canal P2 y de un canal P3. En (C) y (D) se muestran secciones del reovirus aviar 138 y el reovirus de mamífero 3, respectivamente. Se muestran las posiciones de los ejes de simetría binarios (2f), ternarios (3f) y de orden 5 (5f). La posición de cada una de las proteínas estructurales aparece etiquetada.

Con respecto a la cápside externa, su base está formada por trímeros de la proteína μ BC que están localizados sobre las moléculas de σ A y establecen contactos con la torres formadas por λ C. Situadas sobre μ B se encuentran 600 copias de la proteína σ B, que a su vez están organizadas en grupos de cuatro o seis moléculas.

Estas agrupaciones se sitúan alrededor de dos tipos de canales que discurren a lo largo de la cápside externa, de tal forma que sobre la superficie de la cápside externa nos encontramos con 60 canales P3 rodeado cada uno de ellos por seis monómeros de σB y 60 canales P2 rodeados por cuatro monómeros de σB . La principal diferencia estructural entre los reovirus aviáres y de mamífero es que el interior de estos canales en los reovirus de mamífero alberga un dominio C-terminal de la proteína $\mu 1$ mientras que en los reovirus aviáres los canales parecen estar vacíos.

Existe una característica importante observada en los reoviriones aviáres pero no en los de mamífero que consiste en que los reoviriones aviáres pierden la cápside externa cuando se incuban a 40 °C (Grande & Benavente, 2000). Si tenemos en cuenta que la temperatura corporal de las gallinas es de 39.5 °C, la liberación de los polipéptidos de la cápside externa en respuesta a la temperatura del hospedador nativo puede tener un efecto positivo en las primeras etapas de la replicación viral en el hospedador aviar, facilitando la penetración del virus a nivel intestinal y/o promoviendo la descapsidación de los reoviriones dentro de los endosomas.

Las partículas reovirales se pueden transformar *in vitro* en dos tipos de subpartículas virales, las partículas subvirales intermedias (ISVPs) y los cores. En las ISVPs faltan las proteínas de la cápside externa σB y μB pero conservan la proteína σC , al contrario que los cores que sólo contienen las proteínas λA , λB , λC , μA y σA . Las ISVPs son infecciosas y transcripcionalmente activas mientras que los cores, aunque son transcripcionalmente activos, no son infecciosos probablemente porque han perdido la proteína que reconoce el receptor celular, σC .

1.1.5 - El ciclo replicativo de los reovirus

El ciclo replicativo de los reovirus se puede dividir en tres fases diferenciadas: 1) Penetración y pérdida de la cápside externa, 2) Expresión de sus genes y 3) Morfogénesis y salida. En la figura 3 se muestra un diagrama en donde se muestra el ciclo replicativo propuesto para los reovirus aviáres.

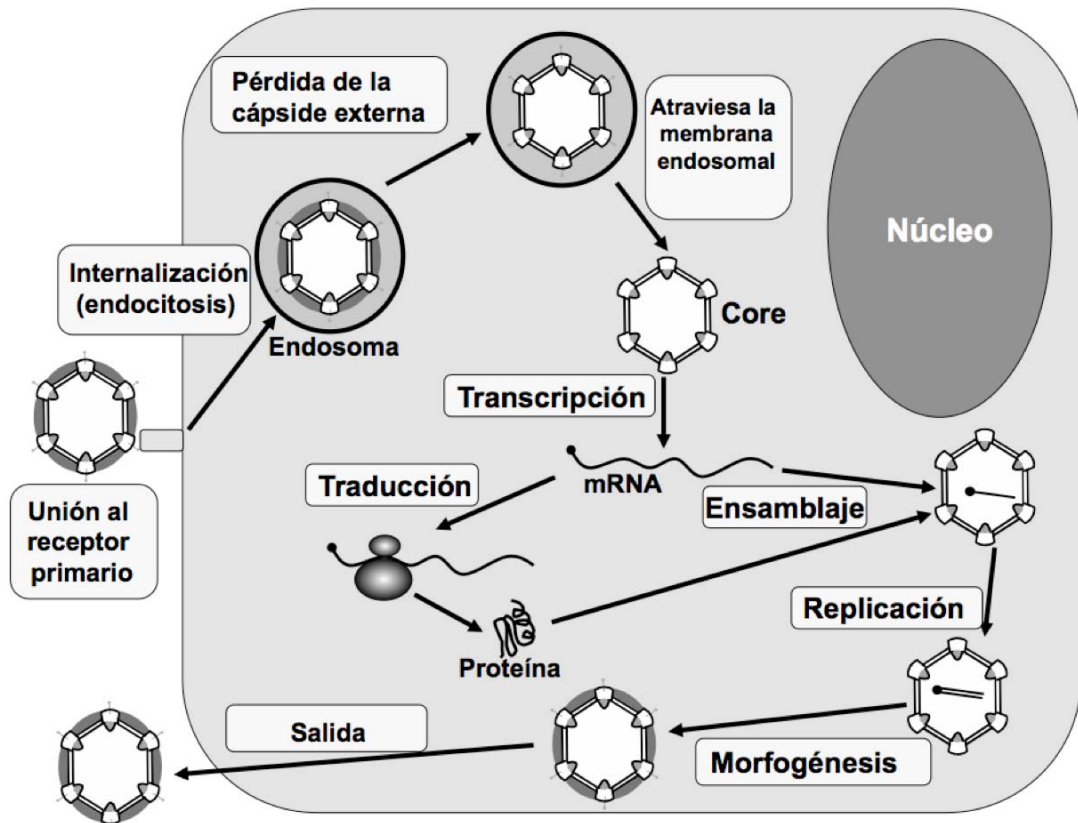


Figura 1.5 Modelo del ciclo replicativo del reovirus aviar (Benavente & Martínez-Costas, 2007)

1.1.5.1 - Penetración y pérdida de la cápside externa

La adhesión extracelular del reovirus aviar a células de cultivo esta mediada por la proteína de la cápside externa σC , que se une a receptores de la superficie celular (Grande *et al.*, 2000; Shapouri *et al.*, 1996). Aunque no se conoce la naturaleza exacta del receptor celular, la adhesión del reovirus aviar a células en cultivo sólo puede ser inhibida tras un pretratamiento de las células de cultivo con proteasas, mientras que un pretratamiento con glicosidasas, lipasas, EDTA o periodato sódico no inhibe esta adhesión. Esto sugiere que el receptor celular debe de tener naturaleza proteica. Además, el hecho de que los reovirus aviares sean capaces en adherirse y replicarse no sólo en células aviares sino también en células de mamífero sugiere que su receptor es una proteína ubicua de la superficie celular (Robertson & Wilcox, 1986). Por otro lado, el hecho de que los reovirus de mamífero no sean capaces de adherirse a células aviares sugiere que ambos virus utilizan receptores celulares distintos (Barton *et al.*, 2001).

La observación de las primeras etapas de la infección de células por el reovirus aviar mediante microscopía electrónica muestra como el virus penetra en el interior de los fibroblastos embrionarios de pollo mediante endocitosis mediada por receptor. La pérdida de la cápside externa en los endosomas es un proceso dependiente de un procesamiento proteolítico que tiene lugar sobre la proteína μB , en este procesamiento se generan dos polipéptidos específicos, δ y δ' (Duncan, 1996). Además de este procesamiento proteolítico es necesaria la acidificación endosomal para la descapsidación del reovirión aviar y su posterior replicación (Duncan, 1996; Labrada *et al.*, 2002).

1.1.5.2 - Expresión de los genes virales

La expresión de los genes virales empieza con la síntesis de los 10 mRNA virales. Este proceso es catalizado por una RNA polimerasa dependiente de RNA, que usa la cadena negativa de los segmentos genómicos como molde para la síntesis de los mRNA. Esta RNA polimerasa está asociada a los cores reovirales ya que tanto estos como las ISVPs son transcripcionalmente activos (Martínez-Costas *et al.*, 1995). La transcripción del genoma viral no parece ser un proceso temporalmente regulado (Benavente & Shatkin, 1988; Mallo *et al.*, 1991).

Los transcritos contienen una caperuza del tipo 1 en su extremo 5' y carecen de una cola de poliadeninas en su extremo 3' (Martínez-Costas *et al.*, 1995). Además contienen dos pequeñas regiones que no se traducen en sus extremos 3' y 5'. Los mRNA virales se producen dentro de la cápside interna y adquieren su caperuza 5' al atravesar las torres formadas por la proteína λC (Zhang *et al.*, 2005). Estos mRNA virales tienen una doble función, son usados para traducir las proteínas virales y como moldes para replicar el genoma viral.

La síntesis de los polipéptidos virales se detecta a tiempos tempranos de infección y con el tiempo casi todas las proteínas sintetizadas por la célula son de origen viral. Mientras todos los mRNA se producen en concentraciones similares en las células infectadas, existe una considerable variación en las cantidades de las distintas proteínas, siendo μB , σB y σNS las más abundantes y λB , μA y las tres proteínas codificadas por el gen S1 (p10, p17 y σC) las menos abundantes. Este hecho sugiere que la regulación de la expresión génica viral ocurre a un nivel traduccional, probablemente relacionada con características conformacionales de los mRNA

virales. Hasta la fecha se desconocen los mecanismos con los que el virus produce el *shut-off* de las proteínas celulares, aunque probablemente esté causado por la sobreproducción de los transcritos virales o por las altas eficiencias de traducción de estos en comparación con los mRNA celulares.

1.1.5.3 - Morfogénesis y salida

Todos los miembros de la familia *Reoviridae* se ensamblan dentro de inclusiones citosólicas denominadas inclusiones virales, factorías virales o viroplasmatas. Estas estructuras, que carecen de membranas u orgánulos celulares, contienen proteínas estructurales y no estructurales, así como partículas virales total o parcialmente ensambladas (Fields *et al.*, 1971; Rhim *et al.*, 1962; Silverstein & Schur, 1970). Se han podido observar mediante microscopía electrónica grandes inclusiones perinucleares en células infectadas por el reovirus aviar (Xu *et al.*, 2004). Mediante microscopía de inmunofluorescencia en células infectadas se ha demostrado que las factorías virales del reovirus aviar tienen una morfología globular y que, al contrario de las factorías virales del reovirus de mamífero, no están asociadas a microtúbulos (Tourís-Otero *et al.*, 2004a,b). Mediante experimentos de expresión de las distintas proteínas virales en células transfectadas se ha demostrado que la única proteína necesaria y suficiente para la formación de las factorías virales es μ NS, y que esta proteína media el reclutamiento de σ NS y λ A dentro de las inclusiones. Hasta la fecha no se han probado otras proteínas (Tourís-Otero *et al.*, 2004a,b). Aún así y dado el hecho de que todas las proteínas virales se encuentran localizadas en inclusiones globulares citoplasmáticas en células infectadas, parece que el reclutamiento de las proteínas virales a las factorías es un proceso selectivo regulado temporalmente, en el que algunas proteínas son inicialmente arrastradas a la inclusión mediante interacciones con μ NS mientras que las restantes son importadas posteriormente mediante interacciones con μ NS, las proteínas anteriormente reclutadas o factores todavía desconocidos.

Hasta la fecha no se han realizado estudios sobre los mecanismos moleculares que permiten la liberación de las partículas reovirales aviares de las células infectadas, pero parece probable que la proteína no estructural p10 juegue un importante papel en este proceso, debido a sus actividad permeabilizante y de fusógeno celular.

1.1.6 - Interacciones virus-célula hospedadora

1.1.6.1 - Fusión celular

Los reovirus aviares junto con el reovirus Nelson Bay, los reovirus de mono, de reptiles y los aquareovirus son los únicos ejemplos de virus no envueltos que inducen fusión entre las células infectadas (Duncan *et al.*, 2004; Lupiani *et al.*, 1995). La formación de sincitios no parece un proceso esencial para la replicación del reovirus aviar puesto que su inhibición no implica una disminución significativa en la producción de virus (Duncan *et al.*, 1996). Sin embargo, este proceso si puede tener importancia en la replicación del reovirus en animales infectados, ya que puede facilitar la diseminación del virus y permitir la dispersión viral a células vecinas sin estar expuesto al sistema inmune del hospedador.

Al contrario que los virus envueltos, los reovirus aviares inducen la formación de sincitios desde el interior de las células infectadas, lo que sugiere que la fusión es producida por las proteínas virales que se sintetizan durante la infección y que este proceso no está relacionado con la penetración del virus (Duncan *et al.*, 1996). Diversos estudios han relacionado la capacidad de formar sincitios con la expresión de la proteína p10 (Bodelon *et al.*, 2001; Shmulevitz & Duncan, 2000).

1.1.6.2 - Resistencia al interferón

Los interferones son citoquinas reconocidas por receptores específicos situados en las membranas celulares, la interacciones de estos con sus receptores dispara diversas rutas de señalización celular que culminan con la expresión de más de 30 proteínas, algunas de ellas juegan papeles importantes en la lucha contra las infecciones virales (Goodbourn *et al.*, 2000; Haller *et al.*, 2006; Sen, 2001). Tanto la 2-5A sintetasa (2'-5'-oligoadenylate synthetase) como la PKR (*Protein Kinase dsRNA Dependent*) son proteínas de esta clase y juegan papeles importantes en la regulación de la síntesis de proteínas en células infectadas (Clemens, 1997). La expresión de ambas proteínas se incrementa debido a la acción del interferón, pero permanecen de forma inactiva hasta que se activan tras la interacción con dsRNA viral, que se genera en la mayoría de las infecciones virales. La activación de la PKR por el dsRNA produce la fosforilación de un factor de transcripción (EIF-2), que finalmente causa la inhibición de la síntesis de proteínas en las células. La activación de la 2-5A sintetasa por dsRNA produce la activación de la endoribonucleasa L, lo que causa de

degradación indiscriminada de RNA y por tanto la inhibición intracelular de la síntesis de proteínas.

Los reovirus aviares son muy resistentes a la acción del interferón (Ellis *et al.*, 1983), lo que sugiere que el virus dispone de estrategias anti-interferón. Un estudio posterior (Martínez-Costas *et al.*, 2000) ha mostrado que el reovirus aviar expresa un factor viral que secuestra el dsRNA, impidiendo de esta manera la activación de la PKR y la 2-5A sintetasa. Evidencias indirectas apuntan a que este factor viral es la proteína σA , una proteína capaz de unir dsRNA de una manera independiente de secuencia.

1.2 - Los adenovirus

1.2.1 - La familia Adenoviridae

Los adenovirus son una causa frecuente de infecciones del tracto respiratorio superior donde causan diversos tipos de infecciones entre los que se incluyen resfriados. Fueron aislados por primera vez en 1953 (Hilleman, 1954) de tejidos adenoideos humanos. Estudios epidemiológicos han confirmado que los adenovirus han sido los responsables de un gran número de síndromes respiratorios agudos en reclutas (Gray *et al.*, 2000). Aunque no son los agentes etiológicos del resfriado común, los adenovirus son los responsables del 5 al 10% de las afecciones respiratorias en niños y también han sido asociados con conjuntivitis epidémicas (Jawetz, 1959) y gastroenteritis infantil (Mautner, 1995).

Se han descrito más de 100 miembros del grupo de los Adenovirus que infectan un amplio rango de hospedadores, desde mamíferos y reptiles a aves. Todos estos virus contienen un genoma lineal de DNA bicatenario dentro de una cápside icosaédrica de 70-100 nm de diámetro. En 1962 se demostró que el adenovirus humano de tipo 12 era capaz de generar tumores en ratones (Trentin, 1962), siendo ésta la primera vez que se demostraba que un virus humano tenía capacidad oncogénica. Sin embargo, tanto las evidencias epidemiológicas que apuntan a que no existe una relación entre los adenovirus y enfermedades oncológicas, como que no ha sido posible detectar ácidos nucleicos adenovirales en tumores (Green *et al.*, 1980), apuntan a que esta capacidad oncogénica no es una causa común de tumores en humanos. Aun así, la capacidad que presentan estos virus para inducir tumores en animales y para transformar células de cultivo los ha convertido en un importante

modelo en el estudio de la oncogénesis. Además han demostrado ser un excelente sistema experimental, debido a que se puede propagar fácilmente hasta alcanzar altos títulos y a que su genoma es fácilmente manipulable, por lo que se han podido estudiar las funciones de todos sus genes mediante análisis mutacionales. Los estudios llevados a cabo con adenovirus en células infectadas han llevado a aumentar nuestro conocimiento sobre expresión, regulación génica y replicación de DNA. Quizás la contribución más importante de los adenovirus a la biología moderna ha sido el descubrimiento del *splicing* del mRNA. Hoy en día el uso de vectores adenovirales en terapia génica está siendo profundamente investigado (Sack & Herzog, 2009).

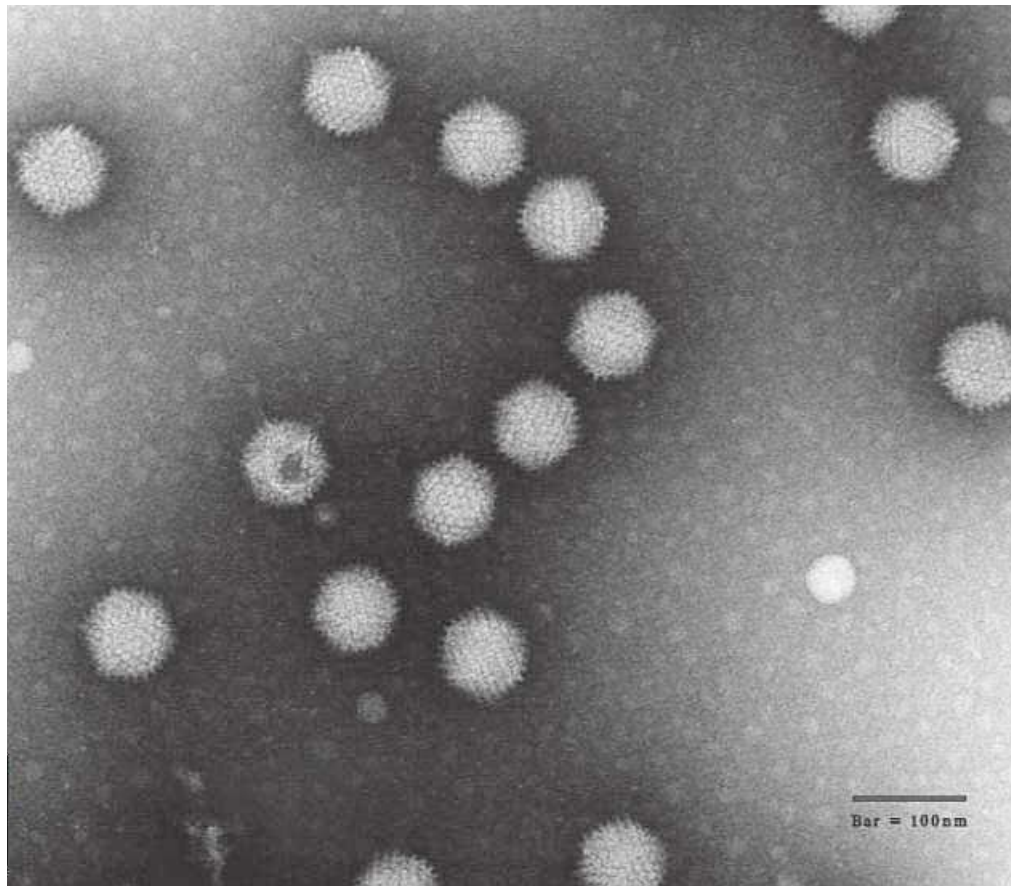


Figura 1.6 Fotos de microscopía de adenovirus humanos de la especie C. Obtenida de ICTVdB (<http://www.ncbi.nlm.nih.gov/ICTVdb/>)

1.2.2 - Clasificación de la familia Adenoviridae

Los adenovirus constituyen la familia *Adenoviridae*, que a su vez se puede dividir en cuatro géneros, *Mastadenovirus*, *Aviadenovirus*, *Atadenovirus* y *Siadenovirus* (Davidson *et al.*, 2003). El género *Aviadenovirus* está restringido a virus

que infectan aves. *Mastadenovirus* incluye adenovirus que infectan mamíferos, por lo que incluye a todos los serotipos humanos. *Atadenovirus* está caracterizado por el inusual alto contenido A+T e infectan marsupiales, aves y mamíferos. Por último, el género *Siadenovirus* incluye virus que han sido aislado de reptiles y aves. Aunque existe reactividad antigénica cruzada entre los miembros dentro de cada género, debido fundamentalmente a la presencia de epítomos conservados localizados en la proteína que forma el capsómero hexamérico (hexon) de la cápside (Nevins, 2001; Figura 1.7), no existe un antígeno conocido común para todos los adenovirus.

Hasta la fecha se han catalogado 51 serotipos distintos de adenovirus humanos basándose en su resistencia a ser neutralizados por antisueros de otros serotipos conocidos de adenovirus. Los epítomos reconocidos están localizados principalmente en el hexón y en la porción globular C-terminal de la fibra (Norrby, 1959). Los distintos serotipos pueden ser clasificados en seis subgrupos diferentes (Rosen, 1960 y Tabla II).

Tabla II. Clasificación de los Adenovirus humanos (subgrupos del Género Mastadenovirus)			
Especie	Serotipos	Potencial oncogénico	
		Tumores en animales	Transformación de células en cultivo
A	12, 18, 31	Alto	+
B	3, 7, 11, 14, 16, 21, 34, 35, 50	Moderado	+
C	1, 2, 5, 6	Bajo	+
D	8, 9, 10, 13, 15, 17, 19, 20	Bajo	+
E	4	Bajo	+
F	40, 41	Desconocido	+

1.2.3 - Estructura del virión

Los adenovirus son partículas icosaédricas de 70-100 nm de diámetro con una o dos fibras proyectándose desde los vértices de la partícula (Figura 1.7). Estas partículas contienen DNA (13% de la masa), proteínas (87% de la masa), no tienen lípidos y solo cantidades traza de azúcares debido a que la fibra está O-glicosilada por una N-acetil-glucosamina (Rux *et al.*, 2004). La mayoría de los estudios estructurales en adenovirus han sido realizados en los serotipos humanos 2 y 5 (hAd2 y hAd5). Se

han caracterizado 11 proteínas. Los principales componentes de la cápside son 240 copias del capsómero hexamérico que componen la mayoría de la superficie de la cápside y 12 copias del capsómero pentamérico situadas en los vértices de la misma. Se ha resuelto la estructura del capsómero hexamérico, que es un trímero del polipéptido II, de hAd2 y hAd5 (Rux *et al.*, 2003). Doce hexones forman cada una de las 20 caras del icosaedro. El tamaño de cada uno de estos hexones varía entre los distintos serotipos, siendo el mayor de los descritos hasta el momento el del hAd2 con 967 aminoácidos. Comparando las secuencias de los hexones de distintos serotipos se han descrito hasta nueve regiones hipervariables en cada molécula del hexón. Se sabe que estas regiones hipervariables están localizadas en la parte superior de la estructura y seis de ellas se han podido situar con precisión en regiones con estructura α helicoidal (Saban *et al.*, 2006). Se sabe que al menos una de estas regiones es responsable de la mayor parte de la actividad neutralizante sobre el virus (Crawford-Mikszá & Schnurr, 1996).

Los capsómeros pentaméricos, también denominados pentones, son un complejo covalente formado por dos proteínas, una proteína homopentamérica situada en la base del capsómero y una fibra trimérica que se proyecta hacia fuera (Figura 1.8). El monómero de la proteína que forma la base pentamérica del penton (del adenovirus humano 5) consiste en dos dominios diferenciados (Zubieta *et al.*, 2005). En uno de estos dominios (el superior) se encuentra un motivo RGD importante para la internalización del virus. La mayoría de los adenovirus humanos contienen una única fibra a excepción de los Ad40 y Ad41 (Kidd *et al.*, 1993) que contiene dos fibras en cada pentón. Esta característica inusual también está presente en el adenovirus aviar CELO (Chiocca *et al.*, 1996). Debido a que la fibra es la encargada de interactuar con el receptor celular, la incorporación de dos fibras presumiblemente aumentará el rango de tipos celulares que estos virus podrán unir. La estructura cristalina de la proteína de la base de hAd2 (un pentámero del polipéptido III), en complejo con péptidos del extremo N-terminal de la fibra, muestra como los motivos FNPVYPY de la fibra se unen en la interfaz entre monómeros del pentámero (Zubieta *et al.*, 2005). Dichos motivos están altamente conservados entre distintos adenovirus. Del análisis de esta estructura se concluye que la proteína pentamérica de la base del pentón puede albergar hasta cinco sitios distintos de interacción con el extremo amino terminal de la fibra trimérica. De este modo hay dos

posibles modos de interacción, o bien los tres extremos de la fibra ocupan posiciones adyacentes relacionadas entre sí por 72° , o bien dos extremos ocupan posiciones adyacentes y el otro extremo ocuparía una posición alterna girada 144° . Esta última organización parece más probable ya que se aproxima más a la simetría ternaria de la fibra.

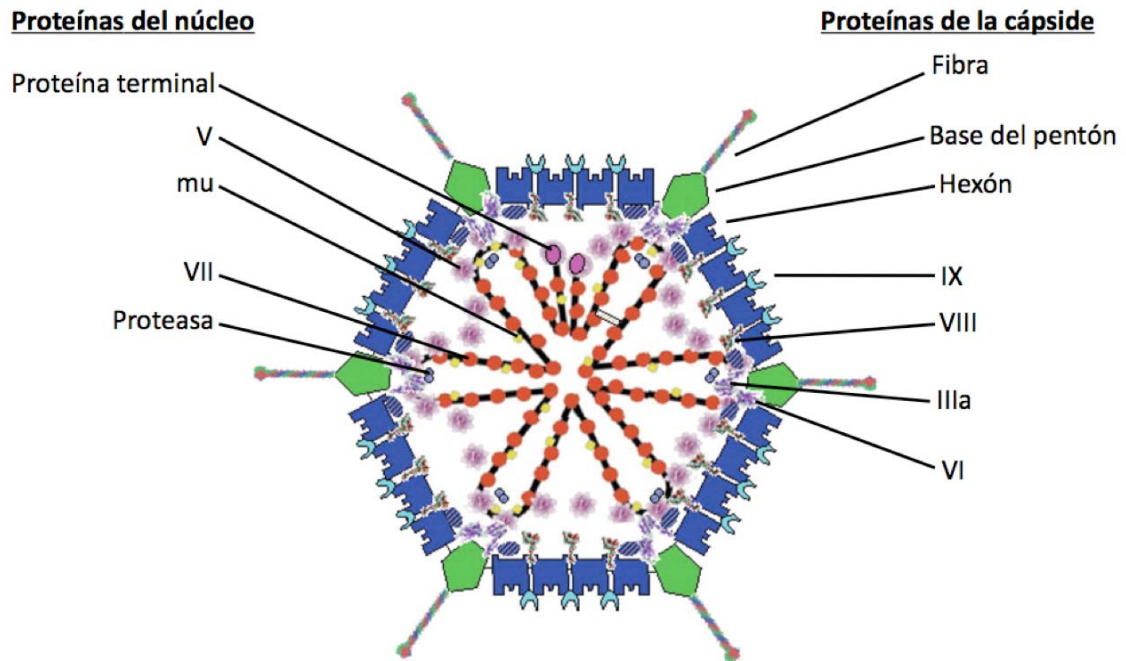


Figura 1.7 Descripción esquemática de la estructura de los adenovirus basada en estudios de criomicroscopía electrónica. La disposición de las proteínas del núcleo así como del DNA es puramente conjetural. Modificado de Russell (2009).

La fibra de los adenovirus está compuesta por tres dominios claramente diferenciados: un dominio N-terminal que es el responsable de la unión a la proteína pentamérica de la base del pentón, una región fibrosa central flexible, que es importante en el proceso infeccioso, y un dominio globular C-terminal que es el responsable de la interacción con el receptor celular primario (Wu *et al.*, 2003). El conocimiento generado a partir de las estructuras cristalinas del dominio globular de hAd5, hAd2 y hAd12 (Xia *et al.*, 1994; van Raaij *et al.*, 1999; Bewley *et al.*, 1999) y de la región fibrosa del hAd2 (van Raaij *et al.*, 1999) ha producido el modelo estructural en uso actualmente para las fibras adenovirales. La región fibrosa está compuesta por repeticiones de un motivo de unos 15 residuos que forma un

plegamiento único denominado triple- β -espiral (van Raaij *et al.*, 1999). La longitud de esta región fibrosa varía entre los distintos serotipos humanos, desde 5 repeticiones en hAd3 hasta 21 repeticiones en hAd2 y hAd5 (Rux *et al.*, 2004; San Martín *et al.*, 2003).

Existen otras proteínas minoritarias de la cápside que actúan estabilizando las interacciones entre los hexones en la partícula viral (proteína VIII) o que juegan papeles importantes en el ciclo viral (proteína I). La localización de estas proteínas minoritarias ha sido inferida de estudios de criomicroscopía electrónica (Saban *et al.*, 2005; San Martín *et al.*, 2008) y se representan en la figura 1.7.

El núcleo del virión está constituido por cinco proteínas y el genoma viral. Los polipéptidos V, VII y XX son proteínas básicas que interactúan con el DNA (Anderson *et al.*, 1989; Hosokawa *et al.*, 1976, Russell *et al.*, 1968) y probablemente actúen condensándolo dentro del núcleo. La cuarta proteína del núcleo es la denominada proteína terminal (671 aminoácidos) que está covalentemente unida a los extremos 5' del DNA (Rekosh *et al.*, 1977). Por último se encuentran 10 copias de la cisteína-proteasa viral p23, importante en el procesamiento de varios precursores de proteínas del virión durante el ensamblaje, desensamblaje y la salida del endosoma durante el proceso de infección (Russell, 2009)

1.2.4 - El genoma de los adenovirus

El primer genoma secuenciado de un adenovirus fue el del hAd2 (Roberts *et al.*, 1984). En la actualidad se conocen las secuencias de los genomas de 8 adenovirus humanos y 15 no humanos (Davison *et al.*, 2003). Analizando estas secuencias se descubrió que los genomas de los distintos adenovirus tienen repeticiones terminales invertidas que varían en tamaño desde 36 pares de bases hasta 200 pares de bases y que funcionan como orígenes de replicación (Davison *et al.*, 2003). Las secuencias de DNA que codifican para la mayoría de las proteínas estructurales del virión, las proteínas relacionadas con la replicación del DNA viral y aquellas que intervienen en el ensamblaje viral están mayoritariamente conservadas en todos los genomas de adenovirus. Por el contrario, las secuencias de DNA de las regiones de expresión temprana 1, 2 y 4 así como la región que codifica el VA RNA, (small virus-associated RNA) se encuentran en todos los adenovirus aislados en mamíferos, pero no están conservados en los otros géneros de *Adenoviridae* (Davison *et al.*, 2003). Las

proteínas codificadas por estas regiones tempranas así como el VA RNA regulan diversos procesos virales entre los que se incluyen la transcripción del genoma viral, el procesamiento de mRNA y su posterior exportación del núcleo.

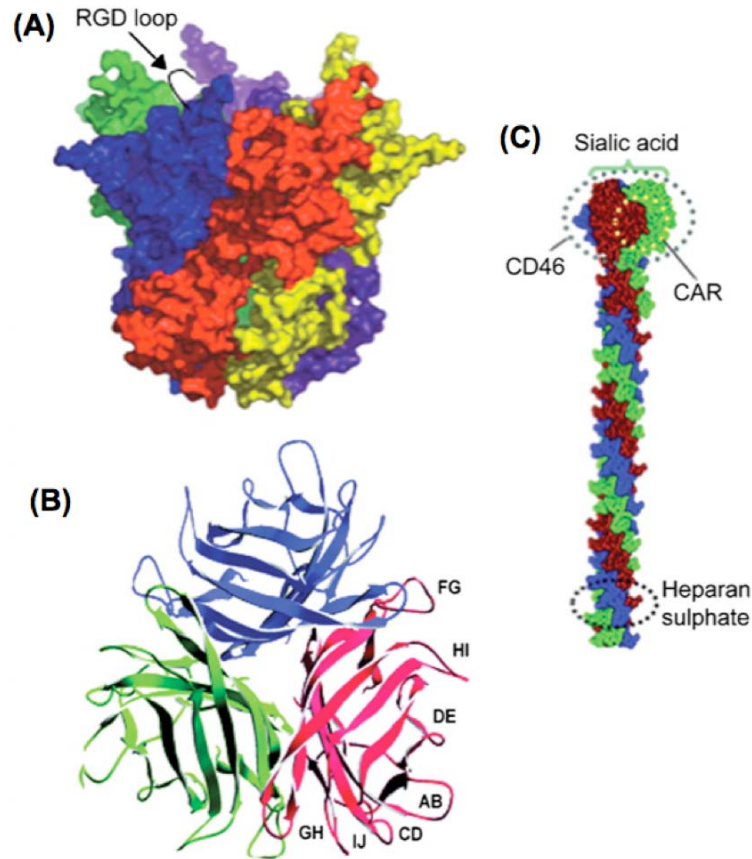


Figura 1.8 Estructura del capsómero pentamérico. (A) Estructura de la proteína pentamérica que forma la base del capsómero. Se indica la localización del bucle que contiene la secuencia RGD. (B) Estructura del dominio globular de cabeza de la fibra adenoviral de hAd35 (C) Los sitios de unión al receptor de las fibras adenovirales. Modificada de Russel (2009)

1.2.5 - La entrada de los adenovirus en la célula hospedadora

El ciclo replicativo de los adenovirus comienza con la adsorción o unión extracelular del virus a los receptores celulares. Este evento está mediado por las interacciones de gran afinidad que se establecen entre el dominio globular de la fibra (el dominio de cabeza) y su receptor celular primario. La subsiguiente internalización de la partícula viral por endocitosis mediada por clatrina se produce tras la interacción secundaria entre un motivo RGD (Arg-Gly-Asp) de la proteína pentamérica de la base de los pentones con integrinas α_v (Wickham *et al.*, 1993). Este motivo RGD está muy

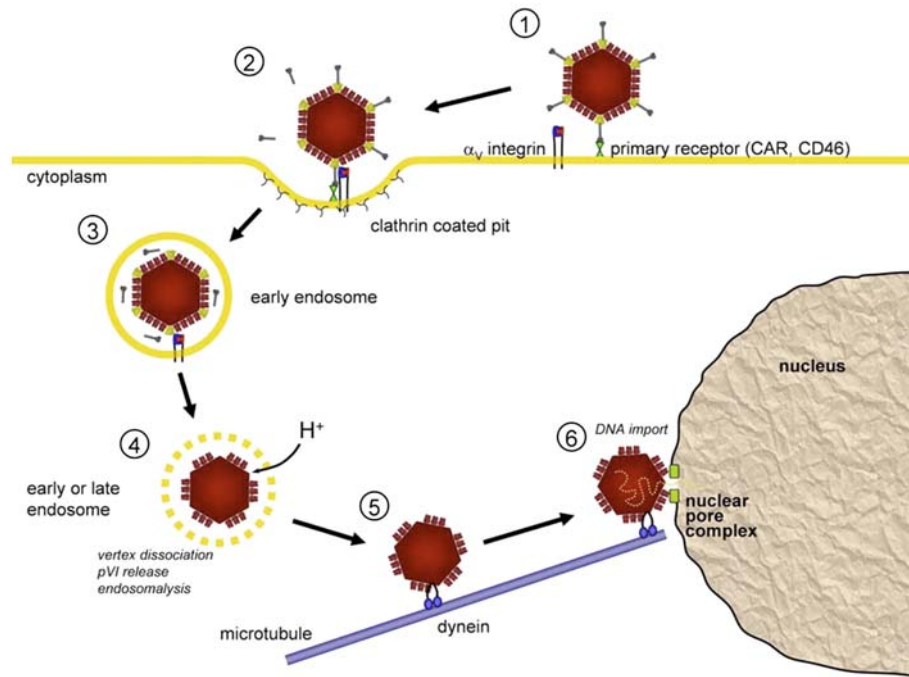


Figura 1.9 La ruta de entrada de los adenovirus comprende cuatro etapas claramente diferenciadas (1) la unión de la partícula viral a su receptor primario, (2) la internalización en endosomas recubiertos por clatrina, (3) formación de la subpartícula viral y su liberación en el endosoma tardío (4), (5) transporte de esta subpartícula por motores de dineína y por último la liberación del DNA en el núcleo de la célula (6). Figura obtenida de Nemerow *et al.* (2009).

conservado en los adenovirus aunque recientemente se ha descrito un adenovirus (el adenovirus de ratón de tipo 1) en el que no aparece en la secuencia de la proteína de la base del pentón sino en el dominio de cabeza de la fibra (Raman *et al.*, 2009). Aún así, parece jugar un papel crucial en la internalización a través de interacciones con integrinas α_v . Durante la etapa de internalización se produce un desensamblaje parcial de la partícula viral en el que las proteínas de los vértices del icosaedro son liberadas en el ambiente ácido del interior del endosoma. La liberación de la proteína VI produce la lisis de la membrana endosomal (Wiethoff *et al.*, 2005), un proceso en el cual la subpartícula viral formada en el endosoma alcanza el citoplasma de la célula. Tras esta etapa, la subpartícula se asocia con motores de dineína y es transportada al núcleo a lo largo de los microtúbulos (Suomalainen *et al.*, 1999). El proceso de entrada dentro de la célula culmina con la liberación del ácido nucleico viral y su tránsito a través del complejo del poro nuclear hasta alcanzar el núcleo celular (Figura 1.9)

1.2.6 - El uso de los adenovirus en terapia génica.

Los adenovirus han generado un gran interés como vectores de terapia génica. El adenovirus humano 5 ha sido extensamente estudiado y es el serotipo humano más usado en aplicaciones de terapia génica junto con el hAd2. Ambos han mostrado ser vehículos eficientes en células *in vitro* aunque, desafortunadamente, el uso de estos vectores *in vivo* ha sufrido serios reveses. En primer lugar, las mayoría de los humanos han desarrollado anticuerpos contra adenovirus de la especie C (hAd5 y hAd2) en edades tempranas, por lo que en muchos pacientes los vectores basados en estos serotipos son rápidamente neutralizados (Smith *et al.*, 1993). En segundo lugar, la mayoría de los vectores adenovirales administrados por el riego circulatorio terminan retenidos en el hígado y no alcanzan las células o tejidos diana (Smith *et al.*, 1993). En tercer lugar, los intentos de solventar estos problemas incrementando la dosis de vector ha resultado ser inadecuado y algunas veces tóxico, incluso letal (Raper *et al.*, 2003). A pesar de todo esto, los adenovirus siguen siendo los vectores más usados en ensayos de terapia génica en humanos y en terapia contra el cáncer (<http://www.wiley.co.uk/genmed/clinical/>) y una prueba de ello es que a fecha de septiembre de 2008, casi el 25% de los 1472 ensayos clínicos de terapia génica aprobados en el mundo utilizan vectores adenovirales, la mayoría de ellos encuadrados en terapia génica contra el cáncer (<http://www.wiley.co.uk/genetherapy/clinical/>). Las estrategias que se han desarrollado para evitar los problemas previamente mencionados están basadas en el uso de adenovirus humanos de especies diferentes a la C (Abrahamsen *et al.*, 1997), usar adenovirus humanos quiméricos (Mastrangeli *et al.*, 1996) o usar adenovirus no humanos (Nguyen *et al.*, 1999) para evitar la respuesta inmune. Por otro lado se han desarrollado vectores adenovirales que no quedan retenidos en el hígado y vectores que han sido diseñados para redirigirlos contra células o tejidos de interés. Para alcanzar este último objetivo se han realizado grandes esfuerzos para identificar las moléculas y los mecanismos que median la unión de los adenovirus a las células de interés. El receptor celular primario de los adenovirus humanos de las especies A, C, D E y F es CAR (coxsackie and adenovirus receptor), el de las especies B es CD46 mientras que el receptor de algunos adenovirus de la especie D es el ácido siálico (Arnberg, 2009; Tabla III).

Tabla III. Receptores de adenovirus humanos y su tropismo

Especie	Serotipos	Tropismo natural	Receptores principales
A	12, 18, 31	Gastrointestinal	CAR
B	3, 7, 16, 21, 50	Respiratorio	CD46
	11, 14, 34, 35	Renal	CD46
C	1, 2, 5, 6	Respiratorio	CAR, Integrinas
D	8,9,10,13,15,17,19	Ocular	CAR, Ácido siálico
E	4	Respiratorio, Ocular	CAR
F	40, 41	Gastrointestinal	CAR

1.2.6.1 - *Coxsackievirus-adenovirus receptor (CAR)*

El receptor CAR es una glicoproteína transmembrana de unos 46 kDa que fue inicialmente identificada porque se unía con gran afinidad a las fibras del coxsackievirus B y los serotipos 2 y 5 de los adenovirus humanos (Bergelson *et al.*, 1997; Tomko *et al.*, 1997). Tiene dos dominios extracelulares (distal-D1 y proximal-D2) que pertenecen a la superfamilia de la inmunoglobulinas, una región hidrofóbica transmembrana y un dominio citoplasmático (Chretien *et al.*, 1998). El dominio D1 por sí mismo es suficiente para la interacción con el dominio globular de la fibra adenoviral (Freimuth *et al.*, 1999). Además de los serotipos 2 y 5 de adenovirus humanos, varios serotipos de las especies A, D, E y F reconocen CAR (Roelvink *et al.*, 1998). Algunos adenovirus no humanos, como los adenovirus caninos parecen ser capaces de interactuar con CAR, mientras que adenovirus bovinos, porcinos y ovinos parecen entrar de una manera independiente de CAR (Bangari & Mittal, 2005; Glasgow *et al.*, 2004). Se han descrito proteínas homólogas a CAR humano en varias especies, incluyendo ratones, ratas, perros, cerdos y en el pez cebra; en todos estos casos con una gran nivel de homología (Bergelson *et al.*, 1998; Fechner *et al.*, 1999; Tomko *et al.*, 1997).

A pesar de la gran variación que existe entre las secuencias de los dominios de cabeza de las distintas fibras de los adenovirus que unen CAR (29-66%), los residuos implicados en la interacción con éste están bien conservados. Por otro lado, en aquellos adenovirus que no unen CAR (serotipos 3, 7, 19, 30 y 35) estos residuos o bien no están conservados o bien la conformación de algunos bucles clave impiden la interacción con el receptor (Law & Davison, 2002).

Muchas líneas cancerosas son refractarias a ser transducidas por vectores adenovirales que interactúan con CAR (Kim *et al.*, 2002). Los niveles de expresión

de CAR en tejidos cancerosos están inversamente correlacionados con la agresividad del tumor, y la inducción de la expresión de CAR en células cancerosas altamente tumorigénicas tiene un efecto supresor del tumor. Esta observación pone de manifiesto la necesidad de desarrollar vectores adenovirales que entren de una manera independiente de CAR para ser aptos en terapia génica de tumores.

1.2.6.2 - CD46

El receptor CD46 es una glicoproteína transmembrana cuya función biológica principal es evitar la activación del complemento en el tejido autólogo (Liszewski *et al.*, 2005). CD46 está compuesto por: i) un dominio extracelular N-terminal que a su vez está formando por cuatro módulos diferenciados (SCR I, SCR II, SCR III, SCR IV) y por 1-3 dominios ricos en motivos STP (Ser-Thr-Pro), ii) una pequeña región de función desconocida, iii) un dominio transmembrana, y iv) un dominio citoplasmático C-terminal (Russell, 2004). CD46 funciona como un "imán de patógenos" ya que no sólo es el receptor de los adenovirus de la especie B sino que también funciona como receptor de otros patógenos humanos como el herpesvirus 6, el virus de la diarrea bovina, *Streptococcus pyogenes*, *Neisseria gonorrhoeae*, *N. meningitidis* y *Helicobacter pylori*. Cada uno de ellos reconoce una parte diferente del ectodominio de CD46.

CD46 es el receptor de la mayoría de los serotipos humanos de los adenovirus de la especie B (3, 7, 16, 21, 50, 11, 14, 34 y 35). Se ha demostrado que el dominio globular de estos serotipos interaccionan con el dominio SCR II, aunque el dominio SCR I también juega un papel importante en la interacción al mantener al dominio SCR II en una conformación favorable a la unión. Estudios cristalográficos con los dominios globulares de hAd11 y hAd35 han revelado que ambos dominios usan las mismas regiones para interaccionar con CD46, los bucles HI, DG e IJ (Persson *et al.*, 2007; Pache *et al.*, 2008).

Vectores de terapia génica derivados de adenovirus de la especie B o vectores adenovirales quiméricos con fibras de la especie B pueden transducir eficientemente tipos celulares que son refractarios a los vectores tradicionales derivados de hAd5 incluyendo células cancerosas malignas, células madre hematopoyéticas y linfocitos.

1.2.6.3 - Ácido siálico

El ácido siálico es un derivado del ácido neuroamínico que generalmente se encuentra presente en gangliósidos y glicoproteínas. El virus de la influenza, los rotavirus, coronavirus y poliomavirus lo utilizan como receptor celular, aunque difieren profundamente en su modo de interacción (Dormitzer *et al.*, 2002; Stehle & Harrison, 1997; Weis *et al.*, 1988). Algunos miembros de la especie D (8, 19a, 37) han sido frecuentemente asociados con queratoconjuntivitis epidémica (Rekhter *et al.*, 1998) y se ha demostrado que estos serotipos utilizan el ácido siálico como receptor celular (Arnberg *et al.*, 2000a,b; 2002). Estudios cristalográficos en adenovirus humanos de la especie D (Ad19 y Ad37) han mostrado que el sitio de unión al ácido siálico se encuentra localizado en la parte superior del trímero y que no está superpuesto con el sitio de unión a CAR (Burmeister *et al.*, 2004).

Se han desarrollado vectores basados en adenovirus que unen ácido siálico y vectores quimera con fibras de estos adenovirus que han demostrado un tropismo mayor e infectan células hematopoyéticas que generalmente son refractarias a la transducción por vectores que utilizan CAR como receptor.

1.2.6.4 - Otros receptores primarios

Han sido descritos otros receptores primarios. CD80 y CD86 que son proteínas de membrana que pertenecen a la superfamilia de la inmunoglobulinas, o los proteoglicanos (PGs) que son glicoproteínas con una o más cadenas de glucosaminoglicanos.

1.2.6.5 - Integrinas

Las integrinas son heterodímeros asociados no covalentemente y que funcionan como moléculas de adhesión celular. Están formados por dos subunidades (α y β) que juegan un papel crítico en muchas funciones celulares como la adhesión, migración, crecimiento y diferenciación celular (Luo *et al.*, 2007; Stewart & Nemerow, 2007). Existen 18 subunidades α diferentes y 8 subunidades β , que pueden formar más de 20 complejos α/β distintos (Stewart & Nemerow, 2007). Múltiples tipos de integrinas ($\alpha v \beta 3$, $\alpha v \beta 5$, $\alpha v \beta 1$, $\alpha 3 \beta 1$, $\alpha 5 \beta 1$) son capaces de actuar como receptores secundarios de adenovirus, a través de motivos RGD y LDV situados en los bucles expuestos de la proteína pentamérica de la base del pentón. En general esta

interacción es de baja afinidad por lo que es necesaria la interacción primaria de alta afinidad del dominio globular de cabeza. La interacción adenovirus-integrina promueve una cascada de señalización que culmina con la polimerización de actina, el reordenamiento del citoesqueleto y el incremento de la internalización de los adenovirus mediante endocitosis mediada por receptores (Nemerow & Stewart, 1999). Como se ha citado previamente, aunque este motivo RGD está altamente conservado en la proteína de la base del pentón de los adenovirus, también es posible encontrarlo en la fibra de algunos adenovirus como el adenovirus de ratón 1 o el adenovirus porcino 4 (Raman *et al.*, 2009, Kleiboeker, 1995)

Para intentar aumentar la transducción de adenovirus en células deficientes de otros receptores primarios, se ha utilizado la estrategia de incorporar motivos RGD en los dominios de cabeza de la fibra adenoviral, lo que ha llevado al incremento de la transducción de varios de estos vectores en células endoteliales, fibroblastos, células tumorales y dendríticas. Todas ellas expresan una alta cantidad de integrinas (Okada *et al.*, 2001; Staba *et al.*, 2000).

1.2.7 - El adenovirus aviar de tipo 1

El adenovirus aviar CELO (chicken embryo letal orphan), también conocido como adenovirus aviar de tipo 1 o A, parece ser relativamente benigno. Esta aparente inocuidad ha despertado interés ante la posibilidad de utilizarlo en los seres humanos como un vector para terapia génica (Kelleher & Vos, 1994; Stevenson *et al.*, 2006) o como un vehículo para quimioterapia (Logunov *et al.*, 2004; Shashkova *et al.*, 2005). También se ha estudiado su potencial como vehículo de vacunación en aves (Francois *et al.*, 2004). Al contrario de sus homólogos de mamíferos, cada vértice contiene insertadas dos proteínas fibrosas triméricas en la proteína pentamérica de la base del pentón: la fibra corta (fibra 2) y la fibra larga (fibra 1). La fibra corta sale directamente de la base sin desviarse, mientras que la fibra larga sale formando un ángulo (Hess *et al.*, 1995) (Figura 1.10). Algunos estudios han mostrado que la fibra corta juega un papel clave en el proceso de infección en las células de pollo (Tan *et al.*, 2001), mientras que la fibra larga no es esencial, lo que la convierte en un diana atractiva para su modificación.

Cada monómero de la fibra larga contiene 793 residuos, en los que destacan dos regiones ricas en glicinas (Hess *et al.*, 1995). La primera región está compuesta por doce glicinas consecutivas situadas cerca del extremo N-terminal (residuos 51-

62). Esta región presumiblemente aporta la flexibilidad necesaria para que las fibras largas y cortas se adapten en una misma proteína pentamérica de la base del pentón y puede ser la responsable del hecho de que la fibra larga emerja de la proteína pentamérica de la base casi de una manera paralela a la superficie viral. La segunda región está compuesta por una secuencia de cuatro glicinas consecutivas situadas cerca del extremo C-terminal (residuos 576-579), posiblemente formando una región flexible entre la región rica en repeticiones en triple β espiral y el dominio de la cabeza. Los estudios cristalográficos realizados con la proteína de la base del pentón de adenovirus humano en presencia de péptidos de su fibra muestran que en cada una de ellas hay cinco sitios de unión disponibles para la fibra (Zubieta *et al.*, 2005), lo que implicaría que el extremo N terminal de uno de los monómeros de una de las fibras no se podría unir a la proteína pentamérica. El escenario más probable parece ser que la fibra corta ocupe tres de los sitios de unión y que la fibra larga ocupe los dos restantes.

Los dominios de cabeza de las fibras larga y corta presenten una homología relativamente baja (menos de un 20% de identidad), y en consonancia con esta baja homología parecen unirse a receptores distintos. Cuando las secuencias de ambas fibras se alinean, se puede identificar una región altamente homóloga cerca del extremo N-terminal, compuesta por unos 15 residuos que estarían implicados en la interacción de la fibra con la proteína pentamérica de la base del pentón. También existe cierta homología entre las regiones fibrosas de ambas fibras.

El dominio central de la fibra larga (residuos 72-572) es rico en repeticiones características de un plegamiento denominado triple- β -espiral (Chiocca *et al.*, 1996; van Raaij *et al.*, 1999b). Estas repeticiones se caracterizan por la presencia de una secuencia consenso $XX\varphi X\varphi X\varphi X-1-X*\varphi X\varphi XX-2$, donde X es cualquier aminoácido, φ generalmente es un aminoácido hidrofóbico y * es típicamente una prolina o glicina, aunque otros aminoácidos pequeños podrían ser tolerados en esta posición. En la posición 1 de esta secuencia consenso, se podría acomodar la inserción de unos pocos aminoácidos; mientras que en la posición 2, serían posibles inserciones más largas que se acomodarían como bucles.

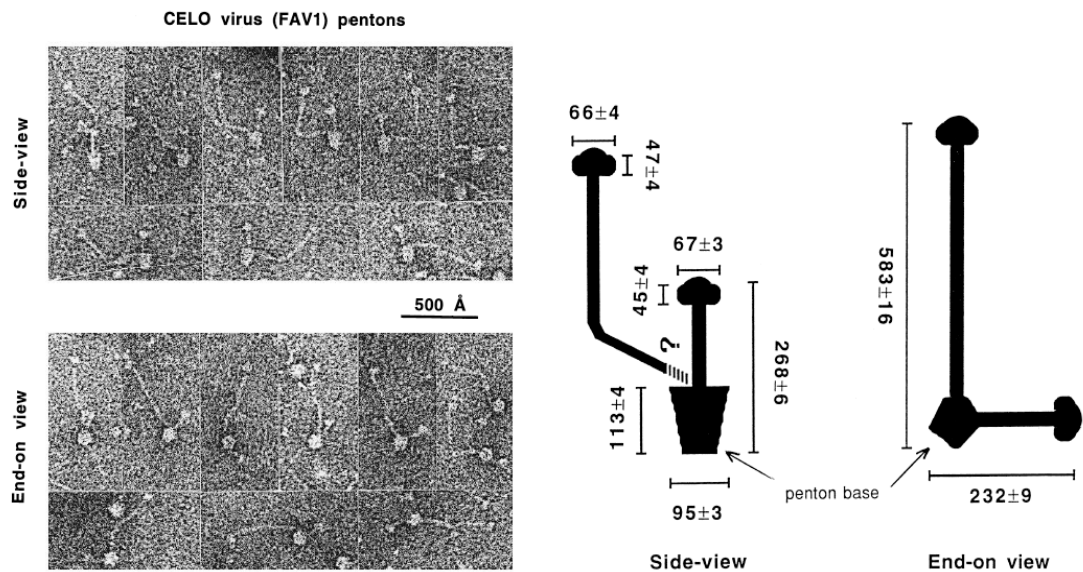


Figura 1.10 Fotografías de microscopía electrónica de las proteínas que componen los vértices de las cápsides del adenovirus aviar de tipo 1 (CELO). Se pueden observar dos perspectivas distintas, en una de ellas la proteína pentamérica está de lado (side-view) mientras que en la otra esta apoyada sobre su base (end-on view). También se muestran esquemas de ambas perspectivas en donde se indican las dimensiones de cada componente. Extraído de Hess *et al.* (1995)

El dominio de cabeza está compuesto por los residuos 580-793 y es comparable en tamaño a los dominios de cabeza de algunos adenovirus humanos cuyas estructuras han sido resueltas: el adenovirus humano de tipo 5 (Xia *et al.*, 1994), tipo 2 (van Raaij *et al.*, 1999a), tipo 12 (Bewley *et al.*, 1999), tipo 3 (Durmort *et al.*, 2001), los tipos 37 y 19p (Burmeister *et al.*, 2004) y la fibra corta del tipo 41 (Seiradake & Cusack, 2005). Sin embargo, la homología de secuencia entre el dominio de cabeza de la fibra larga del adenovirus aviar y los dominios de cabeza antes mencionados es limitada (14-17%).

1.2.8 - El adenovirus porcino de tipo 4 (cepa NADC-1)

Los adenovirus porcinos fueron aislados por primera vez por Haig *et al.* (1964) y como el adenovirus aviar CELO, están considerados como poco patogénicos. Se clasifican dentro del género *Mastadenovirus* y tienen un genoma de dsDNA de aproximadamente 34 kbp. Un análisis de las secuencias de los serotipos 1-5 ha mostrado que los serotipos 1-3 están muy relacionados mientras que los serotipos 4 y 5 están poco relacionados con este grupo y entre ellos. Los cinco serotipos son distintos de los adenovirus humanos, de ratón, caninos, bovinos o aviares (Hammond

et al., 2005). La cepa NADC-1 de adenovirus porcino fue aislada por primera vez de las amígdalas de una cerda adulta en 1972 y se cree que es una cepa de un adenovirus porcino de tipo 4 (Kleiboeker *et al.*, 1993). La fibra de este virus tiene 703 residuos y su arquitectura es única entre los adenovirus ya que, además de los dominios clásicos de las fibras adenovirales, contiene dos dominios extra situados en el extremo C-terminal. Su dominio más N-terminal (residuos 1-37) es el responsable del anclaje de la fibra a la proteína pentamérica de la base y es similar tanto en tamaño como en secuencia con otros dominios equivalentes de otras fibras adenovirales. El dominio fibroso (residuos 38-120) contiene hasta 6 repeticiones en triple β espiral (van Raaij *et al.*, 1999b). Por último, el dominio de cabeza (residuos 121-287) es parecido a otros dominios de cabeza de fibras adenovirales. En el extremo C-terminal del dominio de cabeza encontramos un dominio (residuos 288-392) que contiene una secuencia RGD así como una secuencia rica en alaninas y glutamatos, lo que sugiere que este dominio podría estar relacionado con la interacción con integrinas. En el extremo más C-terminal de la fibra, encontramos un posible dominio de galectinas (residuos 393-681) compuesto por dos dominios de unión a carbohidratos (CRD) repetidos en tándem. Estos posibles dominios CRD son similares a los hallados en la familia de las lectinas de unión a β -galactósidos (Cooper & Barondes, 1999; Cooper, 2002). Este dominio de galectinas no sólo es único entre las fibras adenovirales sino que, junto al del LCDV-1 (lymphocystis disease virus 1), son los únicos dominios de este tipo descritos hasta el momento en virus. Existe una gran probabilidad de que este dominio sea un transgen que ha evolucionado a partir de un gen porcino (Cooper, 2002). El dominio de galectina de la fibra del adenovirus porcino de tipo 4 está compuesto por los dos dominios CRD unidos por un péptido de 23 residuos rico en prolina, y su posición en el extremo más C-terminal de la fibra sugiere que la cepa NADC-1 del adenovirus porcino 4 puede utilizar azúcares de la superficie celular como receptores primarios.

La familia de las galectinas es una familia de lectinas muy versátil, con representantes en mamíferos, anfibios, peces, nematodos, esponjas, hongos y virus (Cooper *et al.*, 1997; Cooper, 2002). Los miembros de esta familia están caracterizados por poseer una secuencia de aminoácidos consenso que les proporciona la capacidad de interactuar con β -galactósidos, sin la necesidad de iones metálicos ni agentes reductores (Barondes *et al.*, 1994). Los dominios CRD están compuestos

por unos 130 aminoácidos plegados en un β -sándwich, que a su vez está formado por dos láminas- β con una topología de tipo *jelly-roll*. Las cadenas β se designan F1 a F5 y S1 a S6 (Lobsanov *et al.*, 1993). El bolsillo de interacción con carbohidratos siempre está localizado en la lámina- β S.

Existen tres familias de galectinas basándose en la organización de sus dominios (Hirabayashi *et al.*, 1993). Las galectinas prototipo (galectinas-1, 2, 5, 7, 10, 11, 13, 14 y 15) consisten en un único dominio CRD más una pequeña secuencia N-terminal y generalmente se encuentran formando homodímeros en solución. Las galectinas de tipo quimérico (galectina 3) están compuestas por un único dominio CRD y un largo dominio N-terminal rico en prolinas y glicinas. Por último, las galectinas repetidas en tándem (galectinas-4, 6, 8, 9, y 12) contienen dos dominios CRD distintos unidos por un linker de longitud variable. Hasta la fecha se han resuelto las estructuras de varias galectinas prototipo como la galectina-1 bovina (Liao *et al.*, 1994) y la galectina-2 humana (Lobsanov *et al.*, 1993), también se ha resuelto la estructura de una galectina quimérica, la galectina-3 humana (Seetharaman *et al.*, 1998) y varios dominios CRD aislados de galectinas repetidas en tándem como los dominios N-terminales de la galectina-9 humana y de ratón (Nagae *et al.*, 2006; 2008; 2009). En las galectinas prototipo diméricas, como las galectinas-1, -2 y -7, los dominios CRD están relacionados entre sí por un eje de rotación binario, de tal forma que los sitios de interacción con el azúcar están localizados en lados opuestos del dímero. Hasta la fecha, la organización que ambos dominios CRD adquieren en las galectinas repetidas en tándem es desconocida, ya que sólo se han resuelto sus dominios CRD de manera independiente.

Se han resuelto algunas estructuras de galectinas en complejo con azúcares y se ha descrito el sitio canónico de interacción con detalle. Este bolsillo de interacción está compuesto por ocho residuos altamente conservados localizados en tres cadenas β adyacentes (S4, S5 y S6). La sustitución de uno o más de estos residuos puede eliminar la capacidad de la galectina para interactuar con β galactósidos, aunque se han descrito algunas galectinas en donde no están presente los ocho residuos conservados que pueden interactuar con otros tipos de azúcares como manosa (Swaminathan *et al.*, 1999) o quito-oligosacáridos (Walti *et al.*, 2008). Otros residuos menos conservados, situados en las cadenas β S2 y S3, también pueden estar implicados en el reconocimiento de azúcares específicos.

1.2.9 - Objetivos

En este trabajo nos propusimos caracterizar estructuralmente diversas proteínas del reovirus aviar S1133, el adenovirus aviar 1 y el adenovirus porcino 4. Para ello nos planteamos los siguientes objetivos:

1. Clonar, expresar, purificar, cristalizar y resolver la estructura de la fibra del reovirus aviar (σC). Idealmente la fibra completa o, en el caso de no ser posible, el dominio de interacción con el receptor celular. Una vez resuelta la estructura, analizarla para intentar averiguar que regiones están implicadas en la interacción con el receptor e intentar explicar la gran estabilidad que presenta la forma oligomérica de esta proteína. La estructura también nos permitirá realizar comparaciones con la fibra del reovirus de mamífero ($\sigma 1$).
2. Expresar, purificar, cristalizar y resolver la estructura de la proteína de unión a dsRNA del reovirus aviar (σA) y a partir de su estructura intentar explicar sus papeles en la biología del reovirus aviar. Fundamentalmente describir que regiones están implicadas en la interacción con la proteínas de la cápside (su función estructural) y que regiones están implicadas en su interacción con el dsRNA (una característica que parece estar relacionada con su actividad antiinterferón).
3. Clonar, expresar, purificar, cristalizar y resolver la estructura de la fibra larga del adenovirus aviar 1. Idealmente la fibra completa o, en el caso de no ser posible, el dominio de interacción con el receptor celular. Una vez resuelta la estructura, analizarla para determinar a que receptor clásico de adenovirus puede unirse. Hasta la fecha no se ha resuelto ninguna estructura de una fibra de adenovirus aviar por lo que se realizaría una comparación con las estructuras conocidas de adenovirus humanos.
4. Clonar, expresar, purificar, cristalizar y resolver la estructura de la fibra del adenovirus porcino 4. Idealmente la fibra completa o, en el caso de no ser posible, los posibles dominios de interacción con el receptor celular. Una vez resuelta la estructura, analizarla para determinar cual es el posible receptor celular. Además, si se logra resolver la estructura del dominio de galectina, se realizaría un análisis más profundo de esta estructura pues supondría la primera estructura de un galectina repetida en tándem. Se procurará estudiar los azúcares que este dominio es capaz de unir y, en caso de disponer de la estructura, se estudiarían los complejos galectina-azúcar.

Capítulo 2

MATERIALES Y MÉTODOS

Capítulo 2

Materiales y métodos

2.1 - Materiales

2.1.1 - Materiales biológicos

2.1.1.1 - Bacterias

XL1-blue: Es una cepa de *Escherichia coli*: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac(F'proAB lacIqZDM15 Tn10, Tet^r)*. Obtenida de Stratagene.

BL21(DE3): Es una cepa de *E. coli* B: *F' dcm ompT hsdS(r_B⁻ m_B⁻) gal λ(DE3)*. Obtenida de Promega

JM109(DE3): Es una cepa de *E. coli*: *endA1, recA1, gyrA96, thi, hsdR17 (r_k⁻, m_k⁺), relA1, supE44, λ⁻, Δ(lac-proAB), [F', traD36, proAB, lacI^qZΔM15], IDE3*. Obtenida de Promega

2.1.1.2 - Plásmidos

pET23d(+) (Invitrogen): vector de expresión en bacterias. Expresa la proteína de interés con una cola de purificación N terminal compuesta por seis histidinas y una secuencia T7 (MGSSHHHHHSSGLVPRGSHMASMTGGQQRI; Novagen; el tag T7 es un péptido de 11 aminoácidos para el cual existen anticuerpos monoclonales disponibles que se pueden unir a resinas) para facilitar su purificación mediante cromatografía de afinidad.

pET28c(+) (Invitrogen): vector de expresión en bacterias. Expresa la proteína de interés con una cola de purificación N terminal compuesta por seis histidinas y una secuencia T7, para facilitar su purificación mediante cromatografía de afinidad.

pMAL-c2x (New England Biolabs): vector de expresión en bacterias. Expresa la proteína de interés con una proteína de fusión (MBP) en su extremo N-terminal para facilitar su purificación mediante cromatografía de afinidad.

2.1.2 - Materiales no biológicos

2.1.2.1 - Disoluciones y tampones

Medio LB (Luria-Bertani): 10 g/l triptona, 5 g/l extracto de levadura, 10 g/l NaCl, pH 7.5

PBS: 137 mM NaCl, 2.7 mM KCl, 8 mM Na₂PO₄ y 1.5 mM KH₂PO₄.

Tampón de electroforesis para SDS-PAGE (Tris-glicina SDS): 25 mM Tris-HCl, pH 8.3, 192 mM glicina y 0.1 % SDS.

Tampón MBP: 20 mM TrisHCl, pH 7.5, 200 mM NaCl, 1 mM EDTA

Tampón ME: 10 mM MES-NaOH pH 6.5, 1 mM EDTA

Tampón TAE 1X: 40mM TrisHCl, 1.1% ácido acético y 0.001% EDTA, pH 8.3

Tampón TE: 10 mM TrisHCl 8.5, 1 mM EDTA 8.0

Tampón de muestra para DNA (6X): 0.25% azul de bromofenol, 0.25% xilenocianol y 30% glicerol.

Tampón de muestra de Laemmli 1X: 50 mM TrisHCl, pH 6.8, 2% SDS, 2.5% β- mercaptoetanol, 90 mM DTT, 10% glicerol y 0.024% azul de bromofenol.

TFB I: 30 mM acetato potásico, 100 mM RbCl, 10 mM CaCl₂, 50 mM MgCl₂ y 15% glicerol, pH 5.5.

TFB II: 10 mM MOPS, pH 6.5, 75 mM CaCl₂, 10 mM RbCl y 15% glicerol

2.1.2.2 - Medios de cultivo

Las bacterias se crecieron en medio de LB. Para el crecimiento de bacterias transformadas con plásmidos, el medio se suplementó con los antibióticos ampicilina (100 µg/ml) (Sigma), tetraciclina (10 µg/ml) (Sigma) o kanamicina (50 µg/ml) (Sigma), según el método de selección para cada plásmido. Los cultivos de colonias en medio sólido se realizaron en el mismo medio con 1.5% bacto-agar (Difco) y antibióticos a las concentraciones indicadas anteriormente.

2.1.2.3 - Cebadores

La secuencia de los cebadores usados, así como su finalidad, se resumen a continuación:

Secuencia	Enzima	Observaciones
Cebadores usados en el clonaje de los fragmentos de la proteína σC		
5'-GATCCCAAGCTTTTAGGTGTCGATGCCGGT-3'	HindIII	σC(326)-REV
5'-CGCGGATCCGCTCCATTTCCGATATATCC-3'	BamHI	σC(49)-FOR
5'-CGCGGATCCATGGCGATTTGACGCCGATC-3'	BamHI	σC(27)-FOR
5'-CGCGGATCCATGCAGACTTCCGTGCAGAA-3'	BamHI	σC(117)-FOR

Secuencia	Enzima	Observaciones
Cebadores usados en el clonaje de los fragmentos de la fibra larga del adenovirus aviar de tipo 1		
5'-CCGGAATTCGAGGAAGTTCCACACCCGAGGTG-3'	EcoR1	AShaft(579)-FOR
5'-CATGCCATGGCAGGCGGAGTGAACCTCTCTC-3'	NcoI	AShaft(61)-FOR
5'-CGCGGATCCGCTGTTTCGCTAACGCGAGCGCACC-3'	BamH1	AShaft(576)-REV
5'-CATGCCATGGCACTCTGCCTCAACCTGGACGAG-3'	NcoI	AShaft(195)-FOR
5'-CATGCCATGGCACTCGAGCTCAACACGGACCCG-3'	NcoI	AShaft(398)-FOR
5'-CGCGGATCCGACTCGAGGTCAACACGGACCCG-3'	BamH1	AShaft(398)-FOR
5'-GATCCCAAGCTTCATTGATAGTACCCAGATAAG-3'	HindIII	AShaft(792)-REV
5'-CCGGAATTCCTCGAGGTCAACACGGACCCG-3'	EcoR1	AShaft(398)-FOR
5'-CGCGGATCCTCCACACCCGAGCTGGCCACC-3'	BamH1	AShaft(576)-FOR
5'-CCGGAATTCCTGGTCAACCTGAACTATCAGTC-3'	EcoR1	AShaft(430)-FOR
5'-CCGGAATTCCTGTGACGGTAAACACGATGGA-3'	EcoR1	AShaft(460)-FOR
5'-CCGGAATTCCTGCTGACGCTGCTACTAATCCC-3'	EcoR1	AShaft(490)-FOR
5'-CCGGAATTCCTGTGGGGTTGAATGTAGACGGC-3'	EcoR1	AShaft(520)-FOR
5'-CCGGAATTCCTGCTACAGATCCGTAAGGAAGCC-3'	EcoR1	AShaft(537)-FOR
5'-CCGGAATTCCTCGAGGTCAACACGGACCCG-3'	EcoR1	AShaft(398)MBP-FOR
Cebadores usados en el clonaje de los dominios de la fibra del adenovirus porcino de tipo 4		
5'-CGCGGATCCTAAGCTTGAACATTGCAAACCCT-3'	BamH1	pAd4(40)-FOR
5'-CGCGGATCCTAGTCTATCCACCCCTGTGGACG-3'	BamH1	pAd4(116)-FOR
5'-ATAAGAATGCGGCCGCTTATTACAGTATCTG-3'	NotI	pAd4(703)-REV
5'-ATAGTTTAGCGGCCGCTTAGTCTGCAAGGTTGT-3'	NotI	pAd4(392)-REV
5'-CGCGGATCCCCTTTGTCCCCCTACCCGAC-3'	BamH1	pAd4(393)-FOR
5'-ATAAGAATGCGGCCGCTTAATCCTGACCCAGGTAGCC-3'	NotI	pAd4(291)-REV
5'-CGCGGATCCCAGCTCCCCTCCTTCTGAGA-3'	BamH1	pAd4(292)-FOR
5'-ATAAGAATGCGGCCGCTTATGCCAGTTGCCAAAGTA-3'	NotI	pAd4(313)-REV
5'-ATAAGAATGCGGCCGCTTAGACCCTATTGTATAGGGC-3'	NotI	pAd4(371)-REV

2.2 - Métodos

2.2.1 - Manipulación de células bacterianas

2.2.1.1 - Obtención de bacterias competentes y transformación

Para la obtención de células competentes, se creció la cepa bacteriana en medio LB a 37°C durante toda la noche y el cultivo se diluyó cien veces en el mismo medio precalentado. Las bacterias se crecieron hasta alcanzar un valor de absorbancia a 600 nm de 0.3, se incubaron 5 minutos en hielo y se recogieron por centrifugación a 2500 g durante 10 min a 4° C. Seguidamente se resuspendieron en un medio de su volumen original en tampón TFB I previamente enfriado, se incubaron en hielo 5 minutos y se centrifugaron a 3000 rpm durante 10 min a 4° C. Este segundo *pellet* de células se resuspendió en 1/25 de su volumen original en TFB II previamente enfriado. Las células se congelaron utilizando hielo seco y se guardaron a -70° C hasta su utilización.

Para su transformación, las bacterias competentes se descongelaron en hielo y

una alícuota de 100 μ l se mezcló con el DNA transformante (1-10 ng). Después de incubar en hielo durante 30 minutos, se realizó un choque térmico a 42° C durante 45 segundos. Tras mantenerlas de nuevo en hielo durante 2 minutos, se añadió 1 ml de medio LB y se incubaron a 37° C durante 1-6 h con agitación. Por último, se recogieron por centrifugación y se esparcieron sobre una placa de LB-agar suplementada con el antibiótico correspondiente.

2.2.2 - Manipulación ácidos nucleicos

2.2.2.1 - Purificación de DNA de plásmidos

La purificación de DNA a pequeña escala para el análisis de transformantes, se realizó usando el sistema *DNA WizardTM minipreps* (Promega) a partir de 5 ml de cultivo bacteriano. Para la obtención de DNA a mayor escala se utilizaron 100 ml de cultivo bacteriano, a partir de los cuales se aisló el plásmido usando el sistema *PureYieldTM Plasmid Midiprep System* (Promega), siguiendo las instrucciones de la casa comercial.

2.2.2.2 - Reacción en cadena de la polimerasa (PCR)

Todas las reacciones se realizaron con el kit *Platinum[®] Pfx DNA Polymerase* (Invitrogen), el cual incluye 50 mM MgSO₄, tampón de amplificación 10X y solución Enhancer 10X, que aumenta la especificidad de los cebadores. Las reacciones se llevaron a cabo manteniendo constantes las concentraciones de los reactivos y los ciclos de amplificación, a excepción de la temperatura de hibridación que varió dependiendo del oligonucleótido empleado para iniciar cada reacción. En todos los casos la reacción de amplificación se realizó con 2.5 unidades de Platinum[®] Pfx DNA Polymerase (Invitrogen), 1X de tampón de amplificación, 1X de solución Enhancer e incluyó 1 mM MgSO₄, 0,4 mM dNTPs, 0,2 μ M de cada cebador y 1 ng de plásmido.

2.2.2.3 - Aislamiento de fragmentos de DNA y su purificación

Para aislar y purificar fragmentos de DNA procedentes de reacciones de amplificación por PCR o de digestiones con enzimas de restricción, primero se sometieron las muestras a electroforesis en geles de 0.8% de agarosa en 1X TAE en presencia de 0.5 μ g/ml de bromuro de etidio (Sambrook *et al.*, 1989). Posteriormente se cortó la banda correspondiente al ácido nucleico de interés con una cuchilla, y se

transfirió a un tubo eppendorf. El DNA se purificó mediante el kit comercial *GFX PCR DNA and Gel Band Purification Kit* (Amersham Biosciences), siguiendo las instrucciones del proveedor.

2.2.2.4 - Tratamiento de DNA con endonucleasas

Las digestiones de DNA con enzimas de restricción (Promega) se llevaron a cabo según las instrucciones del proveedor, en los tampones suministrados y a la temperatura óptima para cada caso. Las reacciones de ligación se realizaron a temperatura ambiente durante 12 h con la DNA ligasa del fago T4 (Promega).

2.2.3 - Manipulación de proteínas

2.2.3.1 - Análisis por electroforesis SDS-PAGE

La electroforesis en geles de SDS-poliacrilamida (SDS-PAGE) se llevó a cabo tal y como describe Laemmli (1970). Las concentraciones de poliacrilamida empleadas para el gel separador fueron entre 10-15% dependiendo del tamaño de la proteína y para el gel concentrador del 4%. Después de la electroforesis las proteínas se fijaron y tiñeron con azul de Coomassie al 0.25% en 33% metanol y 10% ácido acético y se destiñeron en la misma disolución sin colorante.

2.2.3.2 - Cromatografía de afinidad

La cromatografía de afinidad separa proteínas basándose en la afinidad de una proteína a un determinado ligando. El método de basa en que las proteínas que no presentan ninguna afinidad por el ligando inmovilizado en la resina no interaccionan y son eliminadas en el lavado. Las proteínas retenidas son eluidas mediante cambios de pH o el uso de un tampón con un compuesto que compita con la interacción entre la proteína y el ligando de la resina.

Las proteínas que fueron clonadas en el vector de expresión pET28c(+) fueron purificadas en su estado nativo utilizando una resina de Ni-NTA (Qiagen) que tiene una elevada afinidad por la cola de seis histidinas que presentaban las proteínas en su extremo N-terminal. Las proteínas se eluyeron utilizando concentraciones crecientes de imidazol según el protocolo del fabricante.

Las proteínas que fueron clonadas en el vector de expresión pMAL-c2x fueron purificadas en su estado nativo utilizando una resina de amilosa (New England Biolabs) gracias a que la proteína de fusión MBP (*Maltose-Binding Protein*)

interaccionaba con la resina con gran afinidad. Las proteínas se eluyeron utilizando un tampón con maltosa.

2.2.3.3 - Cromatografía de intercambio iónico

La cromatografía de intercambio iónico separa proteínas basándose en diferencias en la carga de su superficie. El método se basa en la formación de interacciones iónicas entre los grupos polares cargados de la superficie de las proteínas y el gel de intercambio iónico. De esta forma las proteínas neutras al pH utilizado no interaccionan con la columna y son eliminadas en el lavado. Las proteínas que interaccionan con la columna son eluidas con un tampón que o bien tenga una mayor fuerza iónica o un pH diferente.

Las proteínas descritas en esta tesis fueron purificadas mediante cromatografía de intercambio aniónico utilizando columnas UnoQ y UnoS (Biorad, Madrid, España). Las muestras se dializaron en tampón de cromatografía (TE o ME) y se eluyeron en un gradiente lineal 0-1 M NaCl en este mismo tampón.

2.2.3.4 – Cromatografía de exclusión molecular

La cromatografía de exclusión molecular separa proteínas basándose en diferencias en su tamaño. El método se basa en la utilización de una resina compuesta por partículas esféricas con poros de un tamaño adecuado a través de las cuales las proteínas difunden a una velocidad proporcional a su tamaño.

Algunas de las proteínas descritas en esta tesis fueron purificadas mediante cromatografía de exclusión molecular utilizando columnas Sephacryl S-100 y S-200 (GE Healthcare) usando como tampón para la cromatografía 10 mM TrisHCl 8.5, 1 mM EDTA 8.0, 150 mM NaCl.

2.2.4 - Métodos cristalográficos

2.2.4.1 - Crecimiento de cristales

El objetivo de la cristalización de proteínas es obtener cristales bien ordenados y de un tamaño adecuado (>0.05 mm; idealmente de unos 0.3 mm). Estos cristales se producen a partir de una disolución sobresaturada de proteína mediante la adición ordenada de moléculas de soluto a un núcleo inicial de cristalización. De esta forma cuando realizamos un experimento de cristalización lo que pretendemos es ir incrementando lentamente la concentración de la proteína hasta obtener una solución

sobresaturada que genere unos pocos núcleos de cristalización sobre los que crezcan unos pocos cristales. Si se supera en exceso el punto de saturación o si se hace demasiado rápido la proteína precipita como un agregado amorfo que no es apto para estudios de difracción.

La proteína ha de cumplir cinco requisitos esenciales para ser considerada apta para cristalizar: Debemos disponer de ella en una elevada concentración (5-20 mg/ml), tiene que estar pura (>95%), tiene que ser homogénea, estable y debemos disponer de una cantidad suficiente (>5 mg). Una vez se ha obtenido la proteína existen varias estrategias de cristalización. La más común se basa en alcanzar la sobresaturación mediante difusión de vapor. Una pequeña gota compuesta por partes iguales de solución de proteína y de cristalización se mantiene dentro de cámara hermética formada por un reservorio que contiene un volumen mucho mayor de solución de cristalización. En estas condiciones la gota pierde agua poco a poco hasta que las concentraciones en la gota y el reservorio son iguales. Existen dos modalidades de esta técnica, en gota colgante y en gota sentada. Otra estrategia es la basada en la diálisis, la solución de proteína se introduce en un pequeño botón de diálisis que a su vez se sumerge en la solución de cristalización lo que permite que la proteína alcance gradualmente la condición de sobresaturación.

Todas las proteínas descritas en este trabajo han sido cristalizadas utilizando el método de difusión de vapor en gota sentada. En esta técnica la gota, una mezcla de la solución de cristalización y la solución de proteína, se asienta sobre una elevación cóncava situada por encima del nivel de la solución de cristalización. Todos estos ensayos de cristalización se realizaron en placas de 96 pocillos CompactClover (Jena Biosciences, Jena, Alemania). Para localizar las condiciones iniciales de cristalización se utilizaron condiciones comerciales como el "Crystal Screen 1" (Hampton Research), "Crystal Screen 2" (Hampton Research), "Crystallization basic kit for proteins" (Sigma), "PEG/Ion Screen 1" (Jena Biosciences) y "PEG/Ion Screen 2" (Jena Biosciences).

Una vez localizadas unas condiciones iniciales prometedoras se optimizaron variando la temperatura, el pH, el tampón usado, las concentraciones de precipitante y de sal y buscando nuevos aditivos que finalmente mejorasen el tamaño y/o el límite máximo de difracción de los cristales. Los cristales obtenidos en el proceso de optimización se probaron a temperatura ambiente utilizando un difractómetro Bruker-Nonius con un detector Kappa-CCD para determinar sus parámetros de celdilla, grupo

espacial y límite de resolución. Una vez optimizados los cristales se procedió a diseñar condiciones de crioprotección. Generalmente los cristales fueron crioprotegidos en una solución con la misma composición que la solución de cristalización a la que se le añadió un 20-25% de glicerol. En el caso de los cristales de los complejos galectina-azúcar decidimos crioproteger los cristales con un 35% PEG (w/v) 3350, puesto que el glicerol puede ocupar el sitio de reconocimiento de la galactosa. Los cristales se sumergieron en la solución crioprotectora y posteriormente se congelaron rápidamente en nitrógeno líquido. Para su transporte a Grenoble (ESRF) y Hamburgo (DESY) se utilizaron contenedores Dewar.

2.2.4.2 - Recogida de datos de difracción

Los conjuntos de datos fueron tomados en las líneas X11 (DESY, Hamburgo), BM16 e ID23-4 (ESRF, Grenoble) o con un difractómetro Bruker-Nonius (Edificio CACTUS, USC)

2.2.4.3 - Procesado de los datos de difracción

El procesado de los datos de difracción es el proceso por el cual a partir de las imágenes de difracción obtenemos una lista de reflexiones, cada una con sus intensidades y sus desviaciones estándar. Fundamentalmente este proceso se puede dividir en tres etapas, en una primera etapa se determina la orientación del cristal, los parámetros de celdilla y su posible grupo espacial. En una segunda etapa se genera una lista de reflexiones y se integran todas las imágenes, es decir a cada *spot* de cada una de las imágenes se le asigna un índice de Miller y una intensidad/desviación estándar. Finalmente en una tercera etapa se escalan las reflexiones medidas en cada imagen y se mezclan.

Este proceso se puede llevar a cabo utilizando distintos programas siendo los más conocidos el Denzo/Scalepack (Otwinowski & Minor, 1997) dentro del paquete HKL2000, el MOSFLM/Scala (Leslie, 2006; Collaborative Computational Project, Number 4, 1994) dentro del paquete CCP4 o el paquete XDS (Kabsch, 1993).

2.2.4.4 - Resolución de estructuras

El objetivo de un estudio de difracción de rayos X es determinar la densidad electrónica $\rho(xyz)$ para cada punto de la celdilla unidad y a partir de esta poder construir un modelo de la proteína cristalizada.

La función densidad electrónica, que representa la transformada de Fourier entre el espacio real, $\rho(xyz)$ y el recíproco $F(hkl)$ viene definida por:

$$\rho(xyz) = \frac{1}{V} \sum_h \sum_k \sum_l |F(hkl)| \cos 2\pi(hx + ky + lz - \phi(hkl))$$

En donde V es el volumen de la celdilla, $|F(hkl)|$ es el módulo del factor de estructura hkl y $\phi(hkl)$ su fase.

Los módulos de los factores de estructura que representan las amplitudes de los haces difractados se obtienen directamente a partir de las intensidades medidas en el detector. En cambio las fases no se pueden obtener experimentalmente, por lo que hay que recurrir a distintos métodos en lo que se ha denominado el problema de la fase.

No existe ninguna relación formal entre las amplitudes y las fases a excepción de la que existe en la función de densidad electrónica. Esto implica que si podemos tener algún conocimiento *a priori* de la estructura o la densidad electrónica podríamos obtener un valor para las fases, esta es la base de todos los métodos para resolver el problema de la fase.

Los métodos directos están basados en la positividad y atomicidad de la densidad electrónica. Estas propiedades producen que existan relaciones entre las fases de los factores de estructura de tal forma que si conocemos las fases de algunas reflexiones podemos deducir las del resto. Aunque esta técnica es muy utilizada en la resolución de estructuras de moléculas pequeñas, en proteínas su uso generalmente está limitado a la resolución de subestructuras ya que son necesarios conjuntos de datos a una resolución muy alta ($<1.2 \text{ \AA}$). El número de átomos también representa una limitación, actualmente este límite está alrededor de los 1000 átomos.

Cuando se dispone de una proteína parecida ($>25\%$ de identidad de secuencia por lo que probablemente tenga una estructura muy similar) de estructura conocida, esta se puede utilizar como si fuera la proteína que va a determinarse y sirve como un primer modelo que posteriormente será refinado. Esta técnica se denomina reemplazamiento molecular (MR).

Existen diversas técnicas para resolver el problema de la fase que están basadas en la introducción de átomos pesados en el cristal. Todas ellas están basadas fundamentalmente en dos métodos. Podemos introducir átomos altamente dispersores y utilizar el método del Reemplazo Isomorfo Simple/Múltiple (técnicas SIR/MIR) o se pueden introducir átomos dispersores anómalos y utilizar el método de la

Dispersión Anómala Simple/Múltiple (técnicas SAD/MAD).

Los métodos de reemplazamiento isomorfo se basan en introducir en el cristal átomos que sean grandes dispersores de los rayos X (es decir átomos con números atómicos elevados) sin que estos distorsionen la estructura cristalina. Estos átomos pueden llegar a modificar sensiblemente la intensidad de las reflexiones en el cristal derivado frente al nativo. Estas diferencias de intensidad se utilizan para calcular un mapa de Patterson a partir del cual podemos determinar las posiciones de estos átomos pesados dentro de la celdilla unidad. A partir de éstas podemos calcular unos factores de estructura calculados (módulos y fases). Después calculamos un mapa de densidad electrónica utilizando los módulos de los factores de estructura observados y las fases calculadas sobre el que podemos empezar a construir nuestro modelo.

Los métodos de dispersión anómala se basan en introducir en el cristal átomos pesados y utilizar rayos X a una frecuencia próxima a la frecuencia natural de oscilación de los electrones de estos átomos pesados. Bajo estas condiciones se produce la denominada dispersión anómala que produce diferencias en la intensidades entre las denominadas reflexiones de Friedel (parejas de reflexiones que en condiciones normales deberían tener la misma amplitud e idénticas fases, pero con signos opuestos). A partir de estas diferencias se puede hacer una primera estimación de las fases con las que se puede calcular un mapa de densidad electrónica sobre el que construir el modelo.

2.2.4.5 - Refinamiento de estructuras

El refinamiento de una estructura es el proceso por el cual se ajustan los parámetros de la estructura, es decir las coordenadas de cada átomo (xyz) y sus factores térmicos (B) para optimizar el acuerdo entre las observaciones y la predicción a partir del modelo. Existen varios métodos para realizar esta optimización, el de los mínimos cuadrados, *Simulated Annealing* o los métodos de máxima verosimilitud (*Maximum Likelihood*). Este último está relacionado con el de los mínimos cuadrados y que usa métodos probabilísticos para estimar el mejor modelo de acuerdo con las variaciones experimentales.

Los factores térmicos B son una estimación de la vibración de cada átomo alrededor de su posición central. Dependiendo del ratio parámetros/observaciones, se pueden modelar los factores térmicos isotrópicamente, es decir cada átomo vibra igualmente en todas las direcciones lo que supone un parámetro por átomo a refinar, o

se pueden modelar anisotrópicamente, en donde cada átomo vibra de manera distinta a lo largo de cada una de las tres direcciones del espacio, para modelar los factores térmicos de una manera anisotrópica se necesitan seis parámetros por átomo. De esta manera aunque los átomos de los cristales que difractan a menos de 1.5 Å de resolución vibren de una manera anisotrópica, sus factores térmicos son generalmente refinados isotrópicamente debido a la carencia de observaciones, una solución a este problema consiste en la inclusión de los denominados grupos de TLS (translation/libration/screw) en el refinamiento. Utilizando este modelo de refinamiento, el movimiento de un grupo de átomos se describe mediante el uso de tres matrices, una describe la vibración translacional del grupo, una describe la libración de este grupo sobre un punto fijo y una tercera que describe la translación y la libración en su conjunto. De esta forma se necesitan 20 parámetros para describir la vibración de los átomos de cada grupo de TLS, lo que implica una disminución del número de parámetros a refinar con respecto al modelo anisotrópico. En este modelo se asume que cada grupo de TLS se mueve como un cuerpo rígido.

Las estructuras descritas en esta tesis fueron resueltas utilizando las técnicas MAD (Dispersión Anómala Múltiple) para σ C(151-326), SIRAS (Reemplazamiento Isomorfo Simple con Dispersión Anómala) en la determinación de la estructura del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1 y MR (Reemplazamiento Molecular) en las estructuras de σ C(117-326), σ A, los dominios de cabeza y galectina de la fibra del adenovirus porcino de tipo 4 así como las estructuras de los complejos galectina-azúcar. Tras resolver el problema de la fase utilizando las metodologías previamente citadas, los modelos iniciales se trazaron automáticamente utilizando el programa ARP-WARP (Morris *et al.*, 2003) o manualmente utilizando los programas O (Jones *et al.*, 1991) y/o COOT (Emsley *et al.*, 2004). El refinamiento fue realizado utilizando el programa REFMAC (Murshudov *et al.*, 1997) utilizando métodos de máxima verosimilitud. En los dominios de la fibra del adenovirus porcino de tipo 4 se incluyeron grupos de TLS durante el refinamiento. Las moléculas de agua fueron modeladas usando el programa ARP (Lamzin & Wilson, 1993). La validación final del modelo se realizó utilizando los programas WHATCHECK (Vriend, 1990), MolProbity (Davies *et al.*, 2007) y MOLEMAN (Kleywegt *et al.*, 2001).

Tabla 2.I - Códigos PDB de las estructuras resueltas en este trabajo

Estructura	Tipo (grupo espacial)	Código PDB
<i>Estructuras de la fibra σC del reovirus aviar</i>		
$\sigma C(151-326)$	I (P6 ₃ 22)	2BT8
$\sigma C(151-326)$	I _c (P321)	2BT7
$\sigma C(151-326)$	I _z (P321)	2BSF
$\sigma C(117-326)$	P (P321)	2JJL
$\sigma C(117-326)$	C (C121)	2VRS
<i>Estructura de la proteína σA del reovirus aviar</i>		
σA	P1	2VAK
<i>Estructura del dominio de cabeza de la fibra larga del reovirus aviar 1</i>		
fAd1-cabeza	C121	2IUM
fAd1-cabeza	P2 ₁	2IUN
<i>Estructura del dominio de cabeza de la fibra del adenovirus porcino 4</i>		
pAd4-cabeza	P2 ₁ 2 ₁ 2 ₁	2WST
<i>Estructura del dominio de galectina de la fibra del adenovirus porcino 4</i>		
pAd4-galectina	Sin ligando (C121)	2WSU
pAd4-galectina	Lactosa (C121)	2WSV
pAd4-galectina	N-acetil-lactosamina (C121)	2WT0
pAd4-galectina	Lacto-N-neotetraosa (C121)	2WT1
pAd4-galectina	tri-(N-acetil-lactosamina) (C121)	2WT2

2.3 - Estrategias de expresión, purificación, cristalización y resolución estructural

2.3.1 - La proteína σC

2.3.1.1 - Construcción de los vectores de expresión.

En el transcurso de este trabajo se han clonado tres versiones diferentes de la proteína σC ; 27-326, 49-326 y 117-326 (clonada por Xose Lois Hermo Parrado). Todas ellas han sido clonadas en el vector de expresión pET28c(+) (Novagen, Darmstadt, Alemania) entre los sitios BamHI-HindIII. Los plásmidos resultantes pET28- σC 27-326, pET28- σC 49-326 y pET28- σC 117-326 codifican el mencionado fragmento fusionado con un cola de purificación N-terminal compuesta por seis

histidinas y un secuencia T7. Las secuencias de los insertos fueron confirmadas por secuenciación de ADN (Sistemas Genómicos, Valencia, España).

2.3.1.2 - Expresión y purificación de $\sigma C(27-326)$, $\sigma C(49-326)$, $\sigma C(117-326)$ y $\sigma C(151-326)$.

El plásmido pET28- $\sigma C117-326$ se usó para transformar cepas de *Escherichia coli* BL21(DE3) (Promega, Madrid, España). Se crecieron cuatro cultivos de 0.7 litros aeróbicamente a 37 °C hasta que alcanzaron una densidad óptica de 0.6-0.8 medida a 600 nm. Los cultivos fueron enfriados a 25 °C tras lo cual se indujo la expresión añadiendo 1 mM IPTG durante 4-5 horas a 25 °C. Tras ese tiempo se centrifugaron las células y se resuspendieron en 40 ml de tampón de resuspensión (PBS con 0.1% (w/v) Tween-20) y se congelaron a -20 °C. Las bacterias se lisaron pasándolas dos veces a través de un emulsificador (emulsificador C5 Avestin, Avestin Europe GmbH, Mannheim, Alemania). Después de eliminar el material insoluble por centrifugación, se añadieron 3 ml de resina de Ni-NTA (Qiagen, Valencia, CA, EE.UU). La suspensión se incubó durante 1 hora a 4 °C y se vertió en una columna vacía. La resina se lavó con tampón PBS y elución se realizó con un gradiente de imidazol pH 7 en tampón PBS (se eluyó en 10, 20, 50, 100, 200, 500 y 1000 mM imidazol). La proteína eluía a partir de 200 mM. El procedimiento fue repetido hasta que se vio que no quedaba mas proteína retenida en la columna.

Para la obtención del dominio $\sigma C(151-326)$, la proteína parcialmente purificada fue incubada con 1 mg ml⁻¹ de tripsina durante 30 minutos a 37 °C. A continuación se añadieron 1.5 volúmenes de sulfato de amonio saturado, el precipitado se redisolvió en tampón TE, y se dializó toda la noche en tampón TE. Posteriormente la muestra se pasó a través de una columna Uno-Q6 (Biorad, Barcelona, España), donde la proteína eluía justo al comienzo del gradiente lineal de cloruro sódico 0-1 M en tampón TE. La proteína fue concentrada utilizando concentradores Centricon (Millipore, Madrid, España), y se realizaron tres lavados con tampón TE para eliminar impurezas de bajo peso molecular.

Para la obtención del dominio $\sigma C(117-326)$, la proteína parcialmente purificada tras la purificación con la resina de Ni-NTA se dializó toda la noche en tampón TE, posteriormente se incubó con 0.2 mg ml⁻¹ de tripsina durante 45 minutos a 37 °C, y la reacción se interrumpió añadiendo 1 mM PMSF. La muestra se pasó a

través de una columna Uno-Q6. La proteína eluía justo al comienzo del gradiente lineal de cloruro sódico 0-1 M en tampón TE. A continuación se concentró utilizando concentradores Centricon y se realizaron tres lavados con tampón TE para eliminar impurezas de bajo peso molecular. La expresión y la purificación de los dominios σ C(27-326) y σ C(49-326) se llevó a cabo de la misma manera que hemos descrito para σ C(117-326) con la excepción de que se digerían con 1 mg ml⁻¹ de tripsina.

2.3.1.3 - Cristalización de σ C(117-326) y σ C(151-326).

Los ensayos de cristalización de las cuatro muestras se llevaron a cabo utilizando la técnica de difusión de vapor en gota sentada con 0.1-0.15 ml de solución en el reservorio y 2-5 μ l de solución de proteína mezcladas con 2-5 μ l de solución de reservorio. La proteína σ C(151-326) cristalizó en varias condiciones: 1.5 M sulfato de amonio, 0.1 M Tris-HCl 8.5, 12% glicerol (cristales de tipo I), su posterior optimización dio como resultado cristales crecidos en presencia de 10 mM cloruro de zinc (cristales de tipo IIz) y 10 mM cloruro de cadmio (cristales de tipo IIc). Gracias a la presencia de glicerol en la condición de cristalización, los cristales fueron crioprotectados sin necesidad de diseñar ninguna condición crioprotectora. La proteína σ C(117-326) cristalizó en 0.6-0.75 mM sulfato de amonio, 0.1 M Tris-HCl 8.4, 25% glicerol, 50 mM sulfato de zinc, donde tras un día aparecieron cristales con forma de barra y sección hexagonal (cristales de tipo P). Durante la recogida de datos la condición se abrió para poder coger uno de estos cristales con sección hexagonal. Tras unos días aparecieron unos nuevos cristales en esta misma condición con forma de prisma regular oblicuo (cristales de tipo C). Los cristales de σ C(27-326) aparecieron utilizando 27% PEG 4000, 0.1M Tris-HCl 8.5, 0.2M LiSO₄.

2.3.2 - La proteína σ A

2.3.2.1 - Expresión y purificación de la proteína σ A

La proteína σ A del reovirus aviar S1133 fue expresada utilizando el vector de expresión pMALCS2 fusionada a MBP como había sido descrito previamente (Gonzalez-Lopez *et al.*, 2003). La secuencia del inserto fue confirmada mediante secuenciación (Sistemas Genómicos, Valencia, España).

Para su expresión en gran escala, la cepa XL1-Blue de *Escherichia coli* fue transformada con el vector pMALCS2- σ A. Cuatro cultivos de 0.7 litros se dejaron

crecer a 37 °C hasta que se alcanzó una densidad óptica de 0.6 medida a 600 nm en medio LB. En ese momento se indujo la expresión añadiendo 1 mM IPTG y se dejó a 37 °C durante 3 horas. Las células se centrifugaron y se resuspendieron en 40 ml de tampón de lisis (20 mM Tris-HCl pH 7.3, 200 mM NaCl, 1 mM DTT), al que se le añadió una mezcla de inhibidores de proteasas específico para extractos de bacterias (Sigma-Aldrich, Madrid, España). y se congelaron a -20 °C. El precipitado celular fue lisado con un emulsificador (Avestin Emulsifier C5, Avestin Europe GMBH, Mannheim, Alemania). Después de eliminar el material insoluble mediante centrifugación (20000 x g, 1 hora, 4 °C) se añadió cloruro sódico y PEI (polietilenimina) hasta unas concentraciones finales de 1 M y 0.1% (w/v) respectivamente y se incubó a 4 °C durante 1 h para precipitar los ácidos nucleicos. Después de un segundo paso de centrifugación (20000 g, 1 hora, 277 K), se le añadió ribonucleasa A de origen bovino (Sigma-Aldrich, Madrid, España) hasta una concentración final de unos 30 µg ml⁻¹ y se incubó durante 30 minutos a temperatura ambiente para degradar el RNA. Posteriormente se le añadieron 3 ml de resina de amilosa (New England Biolabs, Ipswich MA, USA), se mezcló suavemente y se incubó 10 minutos a temperatura ambiente y se decantó en una columna. La resina fue lavada tres veces con 30 ml de tampón MBP y se eluyó dos veces con 10 ml de tampón MBP suplementado con 10 mM de maltosa. Para eliminar la proteína de fusión, se le añadió cloruro de calcio hasta una concentración final 1 mM y factor Xa hasta una concentración final de 1 µg ml⁻¹ y se incubó la solución durante 20 horas a temperatura ambiente. Tras dializar la mezcla toda la noche a temperatura ambiente en tampón TE, se sometió a cromatografía de interacción iónica con una columna UnoQ6. La proteína σ A eluía en la fracción no retenida, mientras que la proteína MBP y otras impurezas quedaban retenidas en la columna. La proteína fue concentrada con concentradores Centricon haciendo tres lavados con tampón TE para eliminar impurezas de bajo peso molecular. Desde la etapa en la que σ A fue separada de MBP (digestión con el factor Xa), todo el proceso se llevó a cabo a temperatura ambiente, ya que σ A tiene la tendencia a precipitar en frío en ausencia de sal (aunque de manera reversible).

2.3.2.2 - *Cristalización de la proteína σA*

La cristalización fue llevada a cabo mediante experimentos de difusión de vapor en gota sentada a 278 K usando un volumen de 0.1-0.15 ml en el reservorio y gotas que consistían en 2-5 μl de solución de proteína (14-30 mg ml^{-1} en tampón TE) más 2-5 μl de solución de reservorio, siempre se mezclaron volúmenes iguales de ambos. Se obtuvieron cristales después de 6-12 meses en pH 5.5-6.5, 278 K usando como tampón 100 mM citrato sódico, fosfato sódico o MES y como precipitante entre 0.1 y 0.6 M sulfato amónico. Los cristales fueron crioprotectados en la solución de cristalización a la que se le suplementó 25% glicerol.

2.3.2.3 - *Experimentos de ultracentrifugación analítica*

Los análisis de velocidad de sedimentación fueron llevados a cabo utilizando una ultracentrífuga Beckman Optima XL-A (Centro de Investigaciones Biológicas, Madrid) equipada con un sistema óptico de detección UV-VIS y un rotor An-50Ti. La proteína σA fue preparada como ha sido descrito previamente; un RNA autocplementario de secuencia 5'-ACACUGUGAUCAGCAGGCCAGCUGAUCACAGUGU-3' fue obtenido de Nedken (Barcelona, España). El RNA de doble cadena fue preparado utilizando este oligonucleótido calentándolo primero y dejándolo enfriar lentamente. En todos los experimentos se utilizó un tampón PBS suplementado con 2 mM de cloruro de magnesio. Se centrifugaron muestras de 0.4 ml de σA y dsRNA a diferentes ratios a 40000 rpm y 20 °C. La concentración de σA en todas las muestras con exceso de proteína fue de 4.9 μM , mientras que el dsRNA fue añadido a concentraciones 0.0, 1.0, 2.0 y 3.0 μM lo que dio lugar a unos ratios σA -dsRNA de 1:0.0, 1:0.21, 1:0.41 y 1:0.62 respectivamente. La concentración de dsRNA en todas las muestras con exceso de dsRNA fue de 1.77 μM mientras que la proteína σA fue añadida a concentraciones 0.0, 1.71, 0.58, 0.18 y 0.058 μM lo que llevo a unos ratios σA -dsRNA de 0:1.0, 1:1.0, 1:3.1, 1:9.3 y 1:31 respectivamente. Para el cálculo de las concentraciones, se utilizó el coeficiente de extinción molar de σA y el dsRNA calculado a partir de su secuencia (95590 $\text{M}^{-1}\text{cm}^{-1}$ a 280 nm en el caso de la proteína σA y 347100 $\text{M}^{-1}\text{cm}^{-1}$ a 260 nm para el dsRNA). Los experimentos fueron monitorizados a 280 nm (experimento 1) y 243 nm (experimento 2), y los datos fueron analizados utilizando la distribución

continua $c(s)$ de las soluciones de la ecuación de Lamm con el programa SEDFIT (Schuck, 2000).

Los valores teóricos del coeficiente de sedimentación $s_{20,w}$ para los distintos complejos fueron estimados utilizando el programa HYDROPRO (García de la Torre *et al.*, 2000). Los complejos fueron construidos utilizando una, dos o tres copias de la cadena G de la estructura de σA y un modelo de la forma A canónica del dsRNA con la secuencia que utilizamos. Las moléculas de σA fueron colocadas en el centro del dsRNA. El volumen específico parcial de σA (0.728 ml/g) fue calculado a partir de su secuencia de aminoácidos directamente por el programa HYDROPRO. Para el dsRNA se utilizó un valor de 0.55 ml/g. El volumen específico parcial de los distintos complejos se calculó promediando los valores antes citados, lo que llevo a volúmenes específicos parciales de 0.669, 0.692 y 0.712 ml/g para los complejos σA -dsRNA 1:1, 2:1 y 3:1 respectivamente. Como se recomendaba en el manual de HYDROPRO, los valores de los radios atómicos efectivos para la proteína, el dsRNA y los complejos proteína-dsRNA fueron 3.2, 2.8 y 3.0 respectivamente.

2.3.2.4 - Experimentos de microscopía electrónica

Para el experimento con proteína sólo se utilizó a una concentración de 0.05 mg/ml y para los experimentos con dsRNA se mezcló σA a 0.05 mg/ml con dsRNA de 36 pares de bases a 1.8 μM (ratio σA -dsRNA de 15:1). Todos los experimentos se realizaron en tampón PBS suplementado con 2 mM de cloruro de magnesio. Las muestras fueron colocadas sobre una superficie de carbono/mica y se realizó una tinción negativa utilizando 1% (w/v) de silicotungstato de sodio (pH 7.0). Las fotografías de microscopía electrónica fueron tomadas utilizando un microscopio JEOL 120 EX II a 100 kV y utilizando una magnificación nominal de 40000.

2.3.3 - El dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1

2.3.3.1 - Construcción de los vectores de expresión

Un fragmento de ADN que codificaba para los residuos 579-793 (el dominio de cabeza) de la fibra larga del adenovirus aviar (UniProt Q64787) fue producida por PCR y clonada en el vector de expresión pET28c(+) (Novagen, Darmstadt, Alemania) entre los sitios BamHI-HindIII. El plásmido resultante pCELOlongfib579-793 codifica para el mencionado fragmento fusionado con una cola de purificación N-

terminal compuesto por seis histidinas y una secuencia T7. La secuencia fue confirmada por secuenciación de ADN (Sistemas Genómicos, Valencia, España).

2.3.3.2 - Expresión y purificación del dominio de cabeza (aminoácidos 579-793)

El plásmido pCELOlongfib579-793 fue transformado en cepas de *Escherichia coli* JM109 (DE3) (Promega, Madrid, España). Se crecieron cuatro cultivos de 0.7 litros aeróbicamente a 37 °C hasta que se alcanzó una densidad óptica de 0.6-0.8 medida a 600 nm. Se indujo la expresión añadiendo 1 mM IPTG durante 3 horas. Tras ese tiempo se centrifugaron las células y se resuspendieron en 40 ml de tampón PBS frío y se congelaron a -20 °C. Las bacterias fueron lisadas en un emulsificador (emulsificador C5 Avestin, Avestin Europe GmbH, Mannheim, Alemania). Después de eliminar el material insoluble por centrifugación, se añadieron 3 ml de resina de Ni-NTA (Qiagen, Valencia, CA, EE.UU.). La suspensión se incubó durante 1 hora a 4 °C y se vertió en una columna vacía. La resina se lavó con tampón PBS y elución se realizó con un gradiente de imidazol pH 7 en tampón PBS (se eluyó en 0.01, 0.05, 0.125, 0.25, 0.5 y 1.0 M imidazol). La proteína eluyó en concentraciones de imidazol de entre 0.125 y 0.5 M. Las fracciones entre 0.25 y 0.5 M de imidazol contenían la proteína prácticamente pura, la fracción no retenida así como las fracciones hasta 0.125 M imidazol se combinaron y el procedimiento se repitió de nuevo hasta que ya no eluía más proteína.

Para su posterior purificación, las fracciones que contenían la proteína fueron dializadas contra tampón ME y se pasaron a través de una columna Uno-S6 de intercambio catiónico (Bio-Rad, Madrid, España) equilibrada con el mismo tampón. La proteína eluía en varios picos entre 0.2 y 0.4 M de cloruro sódico cuando se usaba un gradiente lineal de 0.0-1.0 M NaCl en tampón ME. Se agruparon todas las fracciones que contenían la proteína pura a juzgar por electroforesis desnaturante. La proteína se concentró a 23 mg ml⁻¹ utilizando concentradores Centricon (Millipore, Madrid, España), se lavaron tres veces con el tampón TE para eliminar las impurezas provenientes de moléculas pequeñas.

2.3.3.3 - Cristalización del dominio de cabeza

La cristalización se llevó a cabo utilizando la técnica de difusión de vapor en gota sentada con 0.2 ml de solución en el reservorio y 2 µl de solución de proteína

mezcladas con 2 μ l de solución de reservorio. La proteína cristalizó usando 10-12% (w/v) PEG 10000, 0.1 M HEPES-NaOH pH 4.7-6.1 y 0-10% (v/v) de glicerol incubando las placas de cristalización a 5 °C. Para la recogida de datos, los cristales se crioprotegieron en una solución de cristalización que contenía un 25% (v/v) de glicerol, y se congelaron en una corriente de nitrógeno.

2.3.4 - Los dominios de la fibra del adenovirus porcino de tipo 4

2.3.4.1 - Construcción de los vectores de expresión

Nuestros colaboradores en Birmingham (EEUU) obtuvieron la cepa NADC-1 de adenovirus porcino (catálogo número 030-PDV) de los laboratorios del servicio nacional de veterinaria del USDA Animal and Plant Health Inspection Service (Ames, Iowa). El gen de la fibra de NADC-1 fue amplificado directamente desde el stock viral usando la técnica de PCR y clonado en el plásmido pKan-NMOvPAd4, que fue el material de partida con el que se empezó el trabajo. Se clonaron los fragmentos 40-703, 116-703, 40-291, 40-317, 40-392, 116-291, 291-392, 317-703, 371-703, 393-703 en el vector pET28c(+) (Novagen/Merck, Nottingham, England) entre los sitios BamHI-NotI. En todas las construcciones se usó el codón de iniciación del vector de expresión y se insertó un codón de terminación después del residuos 291, 317 o 392 según correspondiera. Los plásmidos resultantes codifican para los dominios de la fibra correspondientes, además de una cola de purificación N-terminal que consiste en seis residuos consecutivos de histidinas y una secuencia T7. Las secuencias de estos insertos han sido confirmadas por secuenciación de DNA (Sistemas Genómicos, Valencia, España). Finalmente se lograron cristalizar los dominios 40-291, 116-291 y 393-703 y se resolvieron las estructuras del dominio de cabeza (116-291) y galectina (393-703).

2.3.4.2 - Expresión y purificación del dominio de cabeza (116-291)

Para expresar el dominio de cabeza se transformó la cepa BL21(DE3) de *Escherichia coli* con el plásmido pET28c-Fib116-291. Se crecieron cuatro cultivos de 0.7 litros cada uno a 37 °C hasta que alcanzaron una densidad óptica de 0.6-0.8 medida a 600 nm, y se indujo la expresión de la proteína con 1 mM IPTG. Tras 3 horas a 37 °C los cultivos se centrifugaron y los precipitados se resuspendieron en 40 ml de tampón de resuspensión (50 mM TrisHCl pH 8.5, 1 M cloruro sódico) y se congelaron a -20 °C. Las bacterias se lisaron utilizando un emulsificador (Avestin

EmulsiFlex C5, Avestin Europe GMBH, Mannheim, Alemania). Después de eliminar el material insoluble por centrifugación (20 minutos a 20000 g), se añadieron 4 ml de resina de Ni-NTA (Qiagen, Valencia, California, USA). La suspensión fue incubada durante 1 hora a 4 °C tras lo que se decantó en una columna vacía, se lavó con tampón de resuspensión y la proteína se eluyó con un gradiente de imidazol pH 7.0 (10, 20, 50, 100, 200, 400, 600, 800 y 1000 mM imidazol). La proteína eluía a partir de 600 mM imidazol.

2.3.4.3 - Expresión y purificación de los dominios de shaft y cabeza (40-291)

Para expresar los dominios de shaft y cabeza se transformó la cepa BL21 (DE3) de *Escherichia coli* con el plásmido pET28c-Fib40-291. Se crecieron cuatro cultivos de 0.7 litros cada uno a 37 °C hasta que alcanzaron una densidad óptica de 0.6-0.8 medida a 600 nm, y se indujo la expresión de la proteína con 1 mM IPTG. La proteína se purificó con el mismo protocolo que ha sido descrito para el dominio de cabeza.

2.3.4.4 - Expresión y purificación del dominio de galectina (393-703)

Para expresar el dominio de galectina se transformó la cepa BL21 (DE3) de *Escherichia coli* con el plásmido pET28c-Fib393-703. Se crecieron cuatro cultivos de 0.7 litros cada uno a 37 °C hasta que alcanzaron una densidad óptica de 0.6-0.8 medida a 600 nm, y se indujo la expresión de la proteína con 1 mM IPTG. Tras 16 horas a 25 °C los cultivos se centrifugaron y los precipitados se resuspendieron en 40 ml de tampón de resuspensión (50 mM TrisHCl pH 8.5, 1 M cloruro sódico) y se congelaron a -20 °C. Las bacterias se lisaron utilizando un emulsificador. Después de eliminar el material insoluble por centrifugación (20 minutos a 20000 g), se añadieron 4 ml de resina de Ni-NTA. La suspensión fue incubada durante 1 hora a 4 °C, se lavó con tampón de resuspensión y la proteína se eluyó con un gradiente de imidazol pH 7.0 (10, 20, 50, 100, 200, 400, 600, 800 y 1000 mM imidazol). El dominio de galectina eluía entre las fracciones de 50 y 500 mM. La proteína eluída se dializó en 10 mM Tris-HCl pH 8.5, 1 mM EDTA y se cargó en una columna Uno-Q6 (Biorad, Madrid, España). La proteína no se unía a la columna y eluía en las fracciones no retenidas. El contenido de estas fracciones se concentró hasta 1.0 ml utilizando concentradores Centricon (Millipore, Madrid, España) y se cargó en una columna de

exclusión molecular Sephacryl S100 (GE Healthcare, Alcobendas, España) previamente equilibrada con tampón TE conteniendo 150 mM cloruro sódico.

2.3.4.5 - *Cristalización del dominio de cabeza*

Los experimentos de cristalización se han llevado a cabo utilizando la técnica de gota sentada. Se usaron 150 μl de solución de cristalización en el reservorio y 4-5 μl de solución con proteína mezclados con 4-5 μl de solución de cristalización a partes iguales. La proteína pAd4(116-291) se concentró hasta 40-60 mg ml^{-1} utilizando concentradores Centricon y se lavó con tampón TE hasta 3 veces para eliminar impurezas. La solución de cristalización consistió en 1 M sulfato de litio suplementado con polietilenglicol, se obtuvieron cristales equivalentes utilizando un rango 1-10% (w/v) de PEG 8000.

2.3.4.6 - *Cristalización del dominio de shaft y cabeza*

Los experimentos de cristalización se han llevado a cabo utilizando la técnica de gota sentada. Se usaron 150 μl de solución de cristalización en el reservorio y 4-5 μl de solución con proteína mezclados con 4-5 μl de solución de cristalización a partes iguales. La proteína pAd4(40-291) se concentró hasta 40-60 mg ml^{-1} y se lavó con tampón TE hasta 3 veces para eliminar impurezas. La solución de cristalización consistió en 15% (w/v) PEG 4000, 0.1M citrato sódico pH 6.0 y 0.1M sulfato de amonio. Para la recogida de datos los cristales de este dominio fueron trasferidos a una solución con la misma composición que la solución de cristalización más un 20% (v/v) de glicerol. Los cristales difractaron hasta 3.1 Å de resolución en la línea X11 del DESY, desafortunadamente aunque ha sido posible resolver su estructura no ha sido posible trazar el dominio shaft.

2.3.4.7 - *Cristalización del dominio de galectina y de los complejos galectina-azúcar*

Del mismo modo que el dominio de cabeza, todos los experimentos de cristalización del dominio de galectina fueron llevados a cabo utilizando la técnica de gota sentada. Se usaron 150 μl de solución de cristalización en el reservorio y 4 μl de solución con proteína mezclados con 4 μl de solución de cristalización a partes iguales. La proteína pAd4(393-703) se concentró hasta 10-20 mg ml^{-1} utilizando concentradores Centricon, lavando tres veces con tampón TE 8.0. Los cristales del

dominio de galectina se obtuvieron utilizando como solución de cristalización 25% (w/v) PEG 3350, 500 mM nitrato de sodio, 2 mM ácido siálico (este último fue utilizado como un aditivo aunque sepamos *a priori* que no es un ligando de la galectina). Para la recogida de datos, los cristales del dominio de galectina se transfirieron a una solución con la misma composición que la solución de cristalización más un 20% (v/v) de glicerol. Para resolver las estructuras del dominio de galectina con cuatro azúcares diferentes (lactosa, *N*-acetil-lactosamina, lacto-*N*-neotetraosa y tri(*N*-acetil-lactosamina)), en primer lugar se intentó incubar cristales nativos con soluciones que contenían estos azúcares, pero esta aproximación no tuvo el éxito esperado, por lo que procedimos a buscar condiciones de cristalización en presencia de los azúcares. La lactosa utilizada en los experimentos de cristalización se obtuvo de Sigma-Aldrich Química, S.A. (Madrid, España), la *N*-acetil-lactosamina y la lacto-*N*-neotetraosa se obtuvieron de Dextra Laboratories (Reading, UK) y la 2-aminoethyl-tri(*N*-acetil-lactosamina) se obtuvo del Glycan Array Synthesis core D of the Consortium for Funcional Glycomics (The Scripps Research Institute, Department of Molecular Biology, La Jolla, CA). Los cristales del complejo con lactosa se obtuvieron usando 28% (w/v) PEG 3350, 100 mM nitrato de sodio y 5 mM DTT. Como solución de cristalización, se añadió 70 mM de lactosa en la solución de proteína. Los cristales aparecieron en 1-2 semanas y fueron crioprottegidos en 35% (w/v) PEG 3350, 100 mM nitrato de sodio, 1 mM DTT, 70 mM lactosa. Los cristales del complejo con *N*-acetil-lactosamina se obtuvieron usando 35% (w/v) PEG 3350, 500 mM nitrato de sodio y 5 mM DTT. Como solución de cristalización, se añadió 40 mM de *N*-acetil-lactosamina en la solución de proteína. Los cristales aparecieron en 1-2 semanas y fueron congelados directamente. Los cristales del complejo con lacto-*N*-neotetraosa fueron obtenidos usando 27% (w/v) PEG 3350, 200 mM nitrato de litio y 5 mM DTT. Como solución de cristalización, se añadió 5 mM de lacto-*N*-neotetraosa en la solución de proteína. Los cristales aparecieron en 1-2 semanas y fueron congelados directamente. Los cristales del complejo con tri-(*N*-acetil-lactosamina) se obtuvieron usando 28% (w/v) PEG 3350, 300 mM nitrato de litio y 5 mM DTT. Como solución de cristalización se añadió 5 mM de tri-(*N*-acetil-lactosamina) en la solución de proteína. Los cristales aparecieron en 1-2 semanas y fueron crioprottegidos en 35% (w/v) PEG 3350, 300 mM nitrato de litio, 1 mM DTT, 5 mM tri-(*N*-acetil-lactosamina).

Capítulo 3

LA PROTEÍNA σ C

Capítulo 3

Estructura de la proteína σ C

La proteína σ C es una proteína estructural que forma parte de la cápside externa del reovirus aviar. Es la encargada de reconocer el receptor celular del reovirus aviar, hasta la fecha todavía desconocido. La proteína σ C es un homotrímero en su estado nativo y sólo puede interactuar con su receptor en esta forma trimérica (Grande *et al.*, 2002).

3.1 - Purificación y cristalización

3.1.1 - Construcción de los vectores de expresión.

En el transcurso de este trabajo se han clonado tres diferentes mutantes de delección N-terminal de la proteína σ C; 27-326, 49-326 y 117-326. Los plásmidos resultantes pET28- σ C27-326, pET28- σ C49-326 y pET28- σ C117-326 codifican el fragmento comprendido entre los residuos indicados, fusionado con una cola de purificación N terminal que comprende seis histidinas y una secuencia T7.

3.1.2 - Expresión y purificación de mutantes de delección N-terminal de la proteína σ C.

Las proteínas σ C117-326 y 151-326 fueron obtenidas de cultivos de *Escherichia coli* transformados con el plásmido pET28- σ C117-326. Para obtener la proteína soluble y correctamente plegada, fue necesario llevar a cabo la expresión a 25 °C. Después de la lisis de células y la eliminación del material insoluble por centrifugación, la proteína se purificó por cromatografía de afinidad, utilizando una resina con cationes bivalentes de níquel inmovilizados. La proteína eluyó a altas concentraciones de imidazol (0.2-0.5 M), lo que resultaba coherente con una proteína trimérica que contenía tres colas de histidinas. Para la obtención de los dominios σ C(151-326) y σ C(117-326), la proteína parcialmente purificada fue tratada con tripsina, utilizando altas concentraciones de tripsina para obtener σ C(151-326) y concentraciones más bajas para obtener σ C(117-326), ver Capítulo 2. Las proteínas tratadas con tripsina fueron purificadas por cromatografía de intercambio aniónico.

Los dominios $\sigma C(27-326)$ y $\sigma C(49-326)$ se expresaron utilizando los plásmidos pET28- $\sigma C27-326$ y pET28- $\sigma C49-326$. La purificación estos dominios se llevó a cabo de la misma manera que hemos descrito para $\sigma C(117-326)$ con la excepción de que se digirieron con 1 mg ml^{-1} de tripsina.

3.1.3 - Cristalización de los mutantes de delección N-terminal de σC

La proteína $\sigma C(151-326)$ fue cristalizada utilizando el método de difusión de vapor (gotas sentadas). Se obtuvieron cristales en varias condiciones. Aquellos cristales crecidos en 1.5 M sulfato de amonio, 0.1 M Tris-HCl 8.5, 12% glicerol (cristales I, Fig. 3.1A) difractaban hasta 3.0 Å. Su posterior optimización dio como resultado cristales crecidos en presencia de 10 mM cloruro de zinc (cristales IIz, Fig. 3.1B) que difractaban hasta 2.1 Å y 10 mM cloruro de cadmio (cristales IIc, Fig. 3.1C) que difractaban hasta 2.3 Å. Gracias a la presencia de glicerol en la condición de cristalización, los cristales fueron congelados sin necesidad de diseñar ninguna condición especial. La proteína $\sigma C(117-326)$ cristalizó en 0.6-0.75 mM sulfato de amonio, 0.1 M Tris-HCl 8.4, 25% glicerol, 50 mM sulfato de zinc. Tras un día aparecieron cristales con forma de barra y sección hexagonal que difractaban hasta 2.3 Å (cristales P, no mostrados). Tras unos días aparecieron unos nuevos cristales en esta misma condición con forma de prisma regular oblicuo, que difractaban hasta 1.7 Å (cristales C, no mostrados). No se obtuvieron cristales de la proteína $\sigma C(49-326)$. La proteína $\sigma C(27-326)$ cristalizó en 27% PEG 4000 (w/v), 0.1 M Tris-HCl 8.5, 200 mM LiSO_4 , pero los cristales sólo difractaban hasta 9 Å, por lo que no se intentó resolver la estructura.

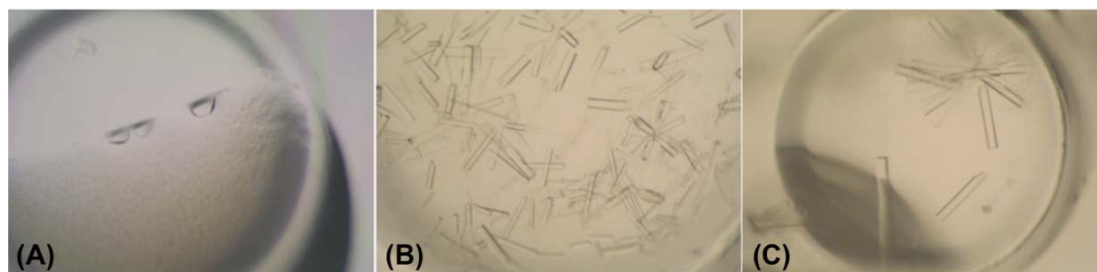


Figura 3.1. Cristales de la proteína $\sigma C(151-326)$. **(A)** Cristales de tipo I que pertenecen al grupo espacial $P6_322$ **(B)** Cristales de tipo IIz **(C)** Cristales de tipo IIc. Los cristales de tipo II pertenecen al grupo espacial $P321$.

3.2 - Determinación estructural de los dominios C-terminal de la proteína σC

3.2.1 - Determinación estructural de $\sigma C(151-326)$

Para resolver la estructura de esta región disponíamos de tres tipos de cristales. Los cristales de tipo I pertenecían al grupo espacial $P6_322$, mientras que los cristales del tipo II pertenecían al grupo espacial $P321$. En una primera instancia se intentó resolver la estructura por reemplazamiento molecular usando $\sigma 1$ como modelo, pero no fue posible. Para poder obtener las fases experimentalmente se derivó un cristal del tipo IIc con cloruro de metilmercurio y se recogieron tres conjuntos de datos a diferentes longitudes de onda ($\lambda_1 = 1.0077 \text{ \AA}$, $\lambda_2 = 0.91838 \text{ \AA}$, $\lambda_3 = 1.00927 \text{ \AA}$). Utilizando la técnica MAD, identificamos cuatro átomos pesados en la celdilla unidad, tres de ellos correspondían a átomos de mercurio unidos a la Cys222 y el cuarto a un átomo de cadmio. Se obtuvo un mapa de densidad electrónica fácilmente interpretable en donde se pudieron trazar 160 residuos automáticamente, aunque parte de ellos resultaron estar fuera de fase. Este modelo preliminar se utilizó para resolver mediante reemplazamiento molecular los cristales IIz, que difractaban a mayor resolución (2.1 \AA). Se trazó un modelo automáticamente que posteriormente se completó a mano. Este modelo fue utilizado posteriormente para resolver las estructuras de los cristales I y IIc mediante reemplazamiento molecular, los ajustes necesarios en estos modelos fueron realizados a mano. Las estructuras refinadas muestran una buena correspondencia con los datos, tienen una buena geometría y presentan pocos residuos en regiones desfavorables del diagrama de Ramachandran (Tabla 3.I). Entre estos están la Val186, situada en un giro β que tiene una densidad electrónica clara, y la His255 que está situada en un bucle con una densidad electrónica no demasiado clara en el dominio globular de cabeza.

Cuando superponemos las tres estructuras entre sí muestran unas desviaciones cuadráticas medias de solo $0.5-0.6 \text{ \AA}$ entre las posiciones de todos sus carbonos α . Existen ligeras diferencias entre los residuos N-terminales en el bucle EF y en algunos residuos del bucle DE (residuos 254-258). Si en vez de los monómeros superponemos los trímeros, las desviaciones cuadráticas medias son similares (alrededor de 0.6 \AA) lo que sugiere que el trímero de $\sigma C(151-326)$ es estructuralmente rígido, sin regiones bisagra entre el dominio fibroso y el dominio globular de cabeza.

Tabla 3.I. Estadísticas cristalográficas

	$\sigma C(156-326)$			$\sigma C(117326)$	
	IIz	IIc	I	C	P
Fuente de radiación	BM16 (ESRF)	BM16 (ESRF)	BM16 (ESRF)	BM16 (ESRF)	BM16 (ESRF)
Temperatura	100 K	100 K	100 K	100 K	100 K
Longitud de onda (Å)	1.2819 Å	0.9796 Å	0.9756 Å	0.8266 Å	0.9840 Å
Detector	165 mm MAR-CCD	165 mm MAR-CCD	165 mm MAR-CCD	ADSC Q210r CCD	ADSC Q210r CCD
Distancia detector-cristal (mm)	105.0	180.0	301.3	178.5	178.0
Rango de oscilación (°)	1	0.5	1	1	1
Grupo espacial	P321	P321	P6 ₃ 22	C2	P321
Parámetros de celdilla a, b, c (Å)	74.78, 74.78, 74.64	73.08, 73.08, 69.93	75.60, 75.60, 243.14	142.8, 54.0, 121.3	77.7, 77.7, 121.4
Parámetros de celdilla α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 101.8, 90	90, 90, 120
Mosaicidad (°)	0.47	0.70	0.78	0.30	0.22
Reflexiones observadas ^a	14231(2019)	9277 (1005)	8895 (1256)	90954 (9036)	19318 (1858)
Rango de resolución (Å)	30-2.1 (2.21-2.10)	37-2.35 (2.45-2.32)	30-3.0 (3.2-3.0)	20.0-1.75 (1.81-1.75)	50.0-2.30 (2.38-2.30)
Multiplicidad	7.9 (8.0)	10.5 (10.0)	8.8 (9.2)	3.3 (3.2)	20.0 (20.3)
Completitud (%)	98.7 (97.7)	95.8 (71.6)	99.8 (100)	99.2 (99.2)	99.0 (98.7)
R _{symm} (%) ^b	7.1 (26.4)	7.6 (33.7)	7.7 (26.2)	5.8 (38.4)	6.8 (42.8)
Estadísticas de phasing (cristales IIz derivados con cloruro de metilmercurio)					
Longitud de onda (Å)	1.0077	0.91838	1.00927		
Rango de resolución (Å)	20-2.2 (2.34-2.20)	20-2.2 (2.34-2.20)	20-2.35 (2.48-2.35)		
Multiplicidad	10.7 (10.9)	10.6 (10.8)	10.0 (10.1)		
Completitud (%)	99.8 (100.0)	99.9 (100.0)	99.8 (100.0)		
R _{symm} (%) ^b	7.1 (28.1)	7.4 (27.4)	7.2 (29.9)		
<I/ σ (I)>	7.3 (2.5)	7.3 (2.7)	8.1 (2.5)		
Rango de resolución usado en SHELDX (Å)	20-2.23 (2.31-2.23)				
Número de reflexiones	11469 (1132)				
Número de átomos pesados	4				
Contraste (SHELXE: original/invertido)	0.667/0.084				
Figura de mérito (SHELXE)	0.600 (0.598)				
Estadísticas de refinamiento^c					
Rango de resolución	30.0-2.1 (2.21-2.10)	37.0-2.35 (2.41-2.35)	30.0-3.0 (3.2-3.0)	18.43-1.75 (1.87-1.75)	25.00-2.30 (2.46-2.30)
Moléculas de proteína / ASU	1	1	1	3	1
Numero de reflexiones usadas ^a	12815 (950)	8348 (541)	8003 (556)	89213 (15852)	17714 (3116)
Numero de reflexiones de Rfree	1416 (88)	929 (67)	885 (70)	1738 (329)	1514 (268)
R-factor ^d	16.1 (18.0)	18.3 (25.0)	19.8 (27.0)	18.0 (23.4)	18.1 (21.6)
R-free ^e	22.3 (25.0)	24.3 (36.0)	24.9 (34.0)	21.8 (27.8)	22.0 (30.1)
Modelo final					
Número de átomos					
proteína	1326	1285	1260	4649	1551
iones (Cl/Zn ⁺² /SO ₄ ⁻¹)	- / 3 / 3	0 / 2 (Cd ⁺²) / 1	0 / 0 / 0	2 / 7 / 5	2 / 3 / 5
glicerol/agua	0 / 164	0 / 120	0 / 44	18 / 975	0 / 202
Factores térmicos promedio (Å²)					
proteína	24.6	38.8	44.6	24.6	37.5
iones (Cl/Zn ⁺² /SO ₄ ⁻¹)	- / 45.7 / 61.3	- / 55.7 / 63.6	- / - / -	26.8 / 28.7 / 45.7	49.1 / 37.8 / 69.1
glicerol/agua	- / 33.0	- / 40.7	- / 35.9	40.1 / 42.3	- / 45.5
Geometría (desviaciones cuadráticas medias)					
Longitud de enlace (Å) ^f	0.015	0.014	0.014	0.017	0.014
Angulos de enlace (°) ^g	1.800	1.600	1.700	1.566	1.509
Estadísticas de Ramachandran ^f	97.4 / 99.4	97.3 / 98.7	97.2 / 99.3	97.7 / 100	97.1 / 100
Código en el PDB	2BSF	2BT7	2BT8	2VRS	2JIL

Se usó un cristal para resolver cada estructura; se valores entre paréntesis se refieren a la capa de mayor resolución.

^aNo se descartó ninguna reflexión.

^bR_{symm} = $\sum_n \sum_h |I_{h^*} - \langle I_n \rangle| / \sum_n \sum_h I_{h^*}$

^cResultados de REFMAC (Murshudov *et al.*, 1997).

^dR-factor = $\sum ||F_{obs}(hkl)| - |F_{calc}(hkl)|| / \sum |F_{obs}(hkl)|$.

^eDe acuerdo con Brunger (1997).

^fDe acuerdo con el programa RAMPAGE (Lovell *et al.*, 2003). Los porcentajes indican residuos en las regiones favorables, permitidas y desfavorable, respectivamente.

3.2.2 - Determinación estructural de $\sigma C(117-326)$

Para resolver la estructura de $\sigma C(117-326)$ se disponía de dos tipos de cristales, cristales de tipo C (grupo espacial C2) y tipo P (grupo espacial P321). Los cristales de tipo C, fueron resueltos mediante reemplazamiento molecular usando como modelo la estructura del monómero de $\sigma C(151-326)$. El modelo final está compuesto por 620 residuos (residuos 122-326 para la cadena A y B, residuos 117-326 para la cadena C) y las moléculas de solvente especificadas en la tabla 3.I. La cadena C de este modelo fue utilizada para resolver la estructura del cristal de tipo P, mediante reemplazamiento molecular. En el cristal P, la unidad asimétrica consiste en un único monómero (aminoácidos 120-326) de tal forma que la unidad biológica, el trímero, se forma gracias a las operaciones de simetría del grupo espacial. Cuando superponemos el monómero del cristal P sobre los monómeros A, B y C del cristal C obtenemos unas desviaciones cuadráticas medias de sólo 0.4-0.6 Å. Cuando superponemos los trímeros, la desviación cuadrática media es de sólo 0.7 Å, lo que significa que no existen diferencias significativas entre ellos, por lo que el análisis que sigue se hizo únicamente con el modelo derivado del cristal C. Las estructuras refinadas muestran una buena correspondencia con los datos, tienen una buena geometría y no tienen residuos en las regiones desfavorables del diagrama de Ramachandran (Tabla 3.I)

3.3 - Estructura general de la proteína σC

Los estudios de microscopía electrónica realizados con la proteína $\sigma C(115-326)$ muestran que ésta es una proteína de unos 80 Å de longitud y una región globular de unos 50 Å de ancho. El análisis de la secuencia realizado con el programa COILS sugiere que los residuos 53-106 y 126-154 están plegados formando un triple *coiled coil* (Costas, 2004; Lupas *et al.*, 1991). La estructura del dominio $\sigma C(117-326)$ muestra que esta región tiene una longitud de unos 90 Å y una región globular de unos 55 Å de ancho, medidas que guardan una buena concordancia con las previamente citadas para esta proteína.

3.4 - La estructura de los dominios C-terminales de σC

La estructura del fragmento C-terminal de σC puede ser dividido en cuatro partes claramente diferenciadas (Figura 3.3): un barril β globular que compone el dominio de cabeza (residuos 196-326), y una región fibrosa que puede ser subdividida en otros tres subdominios. El primer subdominio (residuos 160-195) contiene dos repeticiones en triple- β -espiral (Mitraki *et al.*, 2002), el segundo subdominio está compuesto por una región bisagra (residuos 155-159) que conecta el primer subdominio con un tercero que está plegado formando un triple *coiled coil* (residuos 117-154).

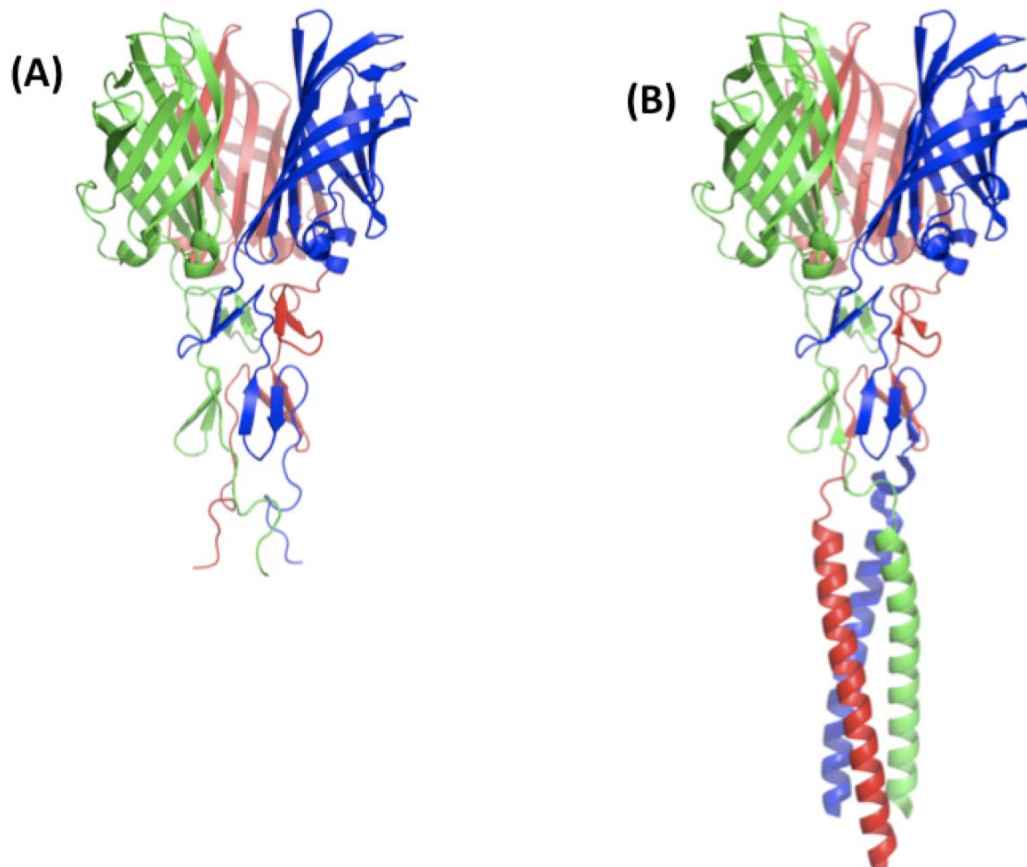


Figura 3.2. Estructuras cristalinas del cristal IIz de $\sigma C(151-326)$ (A) y el cristal C de $\sigma C(117-326)$ (B).

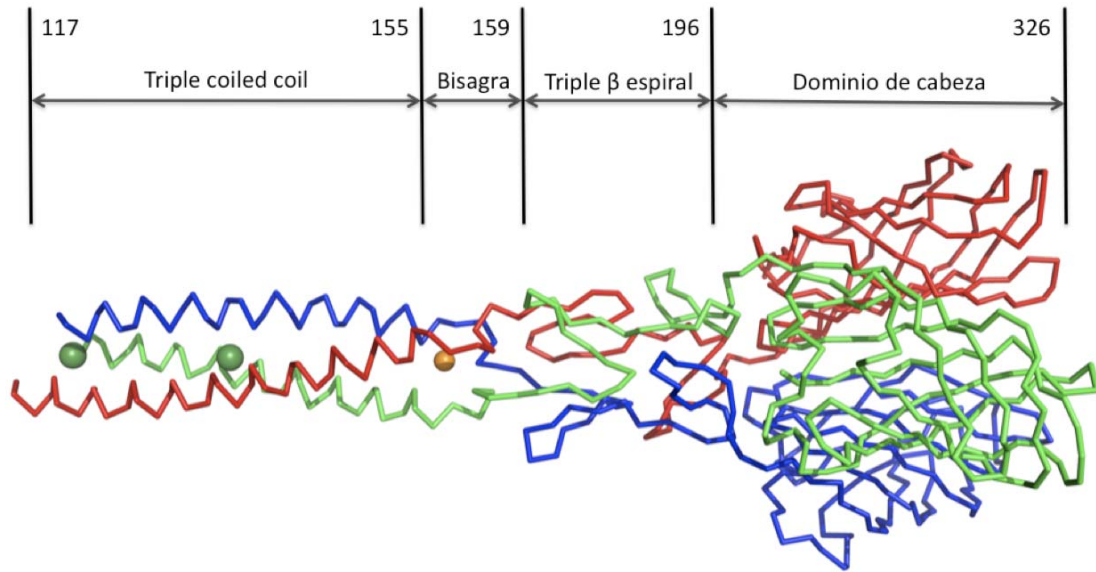


Figura 3.3. Estructura de la proteína $\sigma C(117-326)$ en donde se muestran los cuatro dominios diferenciados. El átomo de zinc en la región bisagra se muestra en naranja, los átomos de cloro en el centro del triple *coiled coil* están indicados como esferas verdes.

3.4.1 - El dominio de cabeza

El dominio globular de cabeza de cada cadena forma un barril β con la misma topología que el dominio globular de $\sigma 1$ (Chappell *et al.*, 2002). El barril β esta compuesto por ocho cadenas β que conforman una lámina β continua en la que la cadena G interacciona con la cadena D (Figura 3.4). A su vez esta lámina β puede ser subdividida en dos motivos *Greek key* compuestos por las láminas β DABC y HEFG (Chappell *et al.*, 2002). Al igual que en la estructura de $\sigma 1$, las cadenas β C y H contienen discontinuidades, aunque estas no son debidas a residuos de prolina como ocurre en $\sigma 1$. La mayoría de los bucles entre cadenas β son cortos, a excepción del bucle DE que contiene 22 aminoácidos y esta compuesto por una pequeña cadena (D') que interacciona con la cadena A y dos hélices α . En $\sigma 1$ la región de unión al receptor esta localizada principalmente en el bucle DE (Chappell *et al.*, 2002).

Un alineamiento llevado a cabo con todas las secuencias conocidas de la proteína σC de reovirus aviares revela que la región más N-terminal (aminoácidos 1-50) muestra un grado elevado de conservación (19 residuos idénticos de 50), la región de triple *coiled coil* tiene un grado de conservación menor (3 residuos idénticos de 155; aunque las posiciones hidrofóbicas se conservan y probablemente su capacidad para formar un triple *coiled coil*). La región C-terminal (aminoácidos 156-326)

también muestra un grado elevado de conservación (60 residuos idénticos de 171). Este patrón de conservación es consistente con un modelo en el que la región más N-terminal estaría implicada en la unión a λC , mientras que la región plegada en triple *coiled coil* jugaría un papel estructural alejando a la región globular del virus y facilitando de esta manera la interacción con el receptor. Es interesante recalcar que la región que comprende los residuos 140-160 está relativamente conservada, incluso cuando tenemos en cuenta las fibras de los reovirus Nelson Bay (Shmulevitz *et al.*, 2002), muscovy duck (Kuntz-Simon *et al.*, 2002) y el reovirus de reptil (Duncan *et al.*, 2004), por lo tanto esta región podría jugar un importante papel estructural como se discutirá posteriormente.

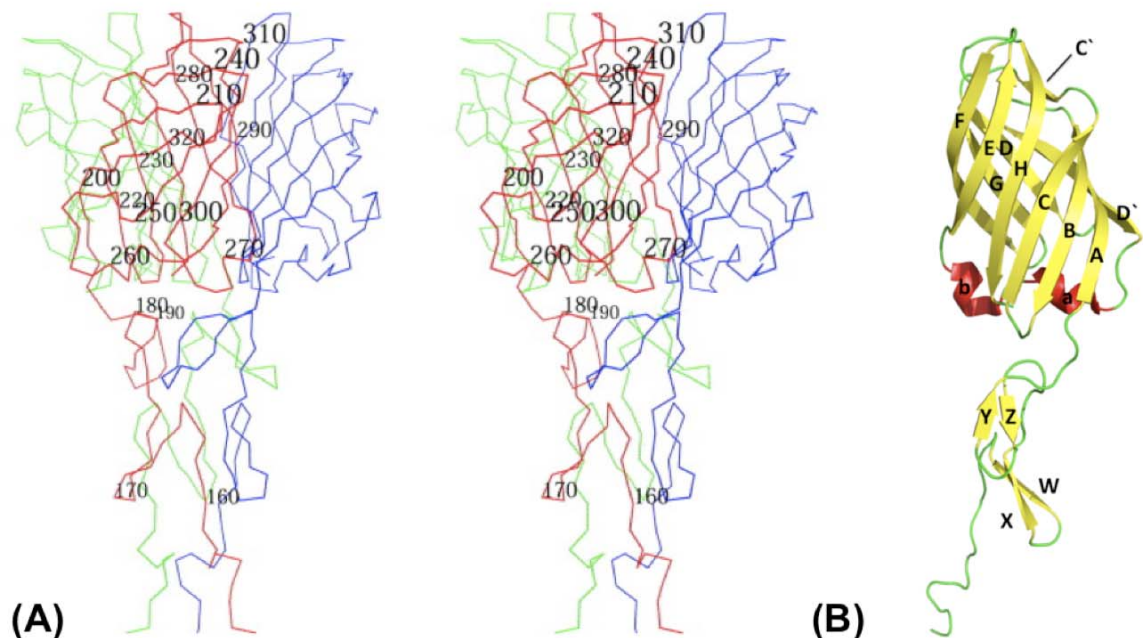


Figura 3.4. Estructura de $\sigma C(151-326)$. (A) Vista estereoscópica de los carbonos α conectados por líneas. Los tres monómeros se han coloreado en rojo, azul y verde. Los residuos se han numerado cada diez. (B) Diagrama en donde se etiquetan los elementos de estructura secundaria.

3.4.2 - El dominio fibroso

El primer subdominio de la región fibrosa (residuos 156-195) contiene dos repeticiones en triple- β -espiral. La primera (residuos 160-176) es del tipo prolina (van Raaij *et al.*, 1999; Chappell *et al.*, 2002) y contiene un giro β de tipo II, con la Pro165 en conformación *cis*. La segunda (residuos 177-191) es bastante irregular, en ella los

residuos 179-182 forman un giro β de tipo VIb (Hutchinson *et al.*, 1994). Las posiciones de los residuos hidrofóbicos no están estrictamente conservadas en esta repetición, la cadena β situada inmediatamente después del giro β solo tiene un residuo hidrofóbico (Cys182) contribuyendo al núcleo hidrofóbico. Algo similar ocurre con la cadena β que esta inmediatamente después del bucle compuesto por los residuos 184-187, en este caso el residuo hidrofóbico es la Leu189. La pérdida de residuos hidrofóbicos en estas cadenas β es compensada por la Phe181 situada en el giro β cuya cadena lateral se proyecta hacia el centro de la fibra, contribuyendo de esta manera al núcleo hidrofóbico.

Conectando este primer subdominio y la región plegada en un triple *coiled coil* encontramos una región (residuos 155-159) en la que los residuos presentan unos factores térmicos elevados, lo que sugiere que presentan una elevada movilidad y que esta región podría actuar de bisagra. En ella las tres His158 están coordinando un catión divalente de zinc a través de sus átomos NE2 (Figura 3.5). La coordinación tetragonal del zinc se completa con la interacción con una molécula de agua.

Por último, hacia el extremo N-terminal encontramos una región plegada formando un triple *coiled coil* (residuos 117-154). Esta región está formada por 5.5 repeticiones heptapeptídicas, con un "*pitch length*" de 166 Å. El radio del *coiled coil* es de 6.2 Å y el número promedio de residuos por vuelta es de 3.6 (Strelkov & Burkhard, 2002), valores comparables con los ya descritos para *coiled coils* triméricos (Fraser & MacRae, 1973; Tao *et al.*, 1997). Un análisis de los residuos en las posiciones a y d del *coiled coil* (Leu152, Val149, Leu145, Ile142, Asn138, Ile135, Leu131, Ile128, Asn124, Val121 y Leu117) muestra que los cinco residuos que ocupan la posición a son β ramificados (dos valinas y tres isoleucinas) mientras que los seis residuos que ocupan la posición d no lo son (cuatro leucinas y dos asparaginas). Se ha propuesto que las estructuras plegadas en *coiled coil* con residuos β ramificados en la posición a y leucinas en la posición d son más estables como dímeros que como trímeros (Harbury *et al.*, 1993), aunque otros *coiled coils* paralelos triples, como el de la fibritina (Tao *et al.*, 1997) o la proteína de unión a manosa humana (Sheriff *et al.*, 1994), contienen residuos β ramificados en posición a y leucinas en posición d. En la región no cristalizada del triple *coiled coil* (residuos 53-106), de los siete residuos en posición a cuatro son valinas y tres son leucinas, mientras que de los ocho residuos en posición d hay cuatro leucinas, dos isoleucinas,

una valina y una metionina, es decir hay tanto residuos β ramificados como leucinas tanto en posición a como en posición d, una configuración más favorable con la formación de trímeros.

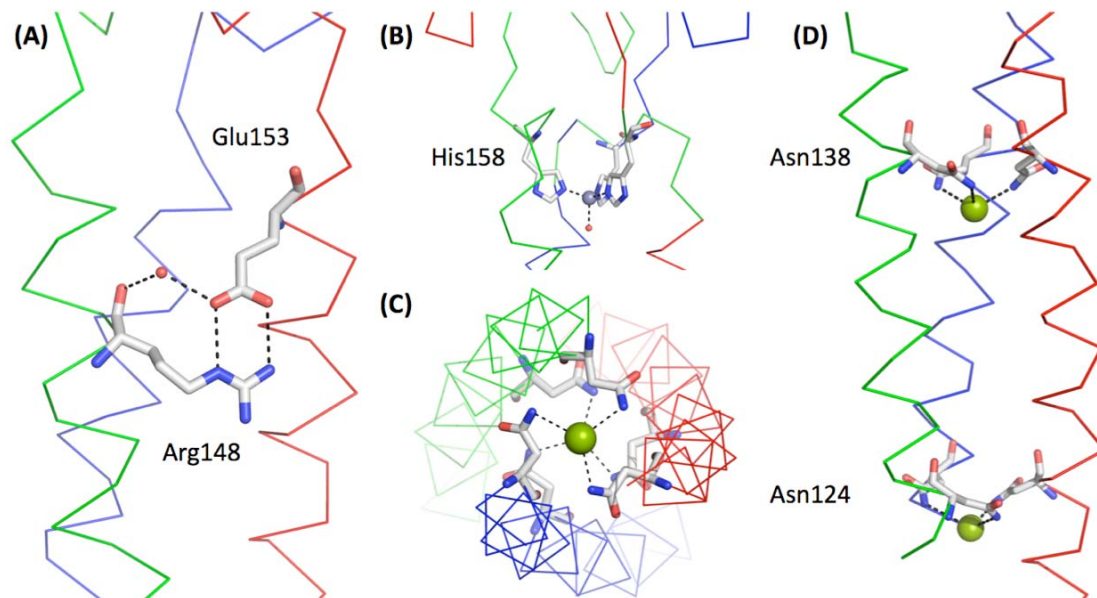


Figura 3.5. Elementos estructurales del dominio fibroso de σC . **(A)** Motivo de trimerización que consiste en la Arg148, el Glu153 y una molécula de agua. Sólo se muestra uno de los tres puentes salinos. **(B)** El dominio bisagra contiene un átomo de zinc (gris) coordinado por tres histidinas (His158) con una distancia de coordinación de 2.0-2.1 Å **(C y D)** Vista sagital y lateral de los átomos de cloro (verde) que se encuentran en el interior del triple *coiled coil* coordinados por asparraginas (Asn138 y Asn124).

La estructura de esta región también revela la existencia de un motivo de trimerización (Meier *et al.*, 2002; Kammerer *et al.*, 2005) compuesto por una red de puentes salinos intercatenarios formados por contactos entre la Arg148 en la posición g de una cadena y el Glu153 en la posición e de la cadena adyacente (Figura 3.5 A). El potencial de este dominio de trimerización sobre el efecto de los residuos hidrofóbicos de las posiciones a y d ha sido demostrado en *coiled coils* triples pequeños (Burkhard *et al.*, 2000; Kammerer *et al.*, 2005). Es interesante recalcar que el análisis de la secuencia revela otro posible motivo de trimerización en el que participarían la Arg36 y el Glu41, lo que sugeriría que el triple *coiled coil* se extendería mas allá del residuo 53. La estructura revela que a mayores de las interacciones hidrofóbicas existentes en el núcleo hidrofóbico del *coiled coil* y de las interacciones polares en su superficie, las asparraginas situadas en la posición d

(Asn124 y Asn138) están coordinando un ión cloro central (Figura 3.5 C y D). Los iones cloro fueron identificados como tales basándose en evidencias cristalográficas, tales como sus factores térmicos (similares a los de los residuos que lo rodean), las distancias de coordinación y la similitud con otros *coiled coils* triples virales. Se ha propuesto que interacciones de esta clase puede favorecer que el *coiled coil* se pliegue con el registro adecuado (Oakley & Kim, 1998), así como que las asparaginas en la posición d favorecen el estado trimérico (Tripet *et al.*, 2000). Un ión cloro central coordinado por residuos de asparaginas parece ser una característica común en *coiled coils* paralelos triples de proteínas virales de fusión, fibras y adhesinas bacterianas (Fass *et al.*, 1996; Tao *et al.*, 1997; Malashkevich *et al.*, 1999; Renard *et al.*, 2005; Connors *et al.*, 2008) y ha sido usado en el diseño *de novo* de *coiled coils* triméricos (Lumb & Kim, 1995).

3.5 - Comparación estructural y funcional con la proteína $\sigma 1$ y otras fibras virales

Un alineamiento de las secuencias de la proteína σC y $\sigma 1$ (Chappell *et al.*, 2002), revela una identidad de secuencia baja. La identidad de secuencia es mayor para los primeros 21 residuos del extremo N-terminal, que probablemente estén implicados en la interacción con la proteína λC de la cápside interna del reovirus. La proteína λC comparte un 27% de identidad de secuencia con su homóloga en reovirus de mamífero, $\lambda 2$, y probablemente adopten una conformación similar. De este modo λC y σC interaccionarían del mismo modo en el que lo hacen $\lambda 2$ y $\sigma 1$. Tanto la región fibrosa que esta plegada con repeticiones en triple β espiral como el dominio globular de cabeza tienen una baja similitud de secuencia pero sus estructuras cristalinas han resultado ser idénticas tanto en sus dimensiones como en su topología. Cabe destacar que en la región fibrosa de σC sólo existen dos repeticiones en triple β espiral, mientras que en la de $\sigma 1$ existen ocho de estas repeticiones, lo que hace de $\sigma 1$ una fibra más larga. Las regiones plegadas formando un triple *coiled coil* tienen una longitud similar.

Cuando superponemos los residuos estructuralmente equivalentes de los dominios globulares de los monómeros de σC y $\sigma 1$ obtenemos unas desviaciones cuadráticas medias de 2.5 Å para los 127 carbonos α alineados (los dominios globulares de σC y $\sigma 1$ contienen 133 y 142 residuos, respectivamente). Por lo que ambos dominios globulares comparten no sólo la misma topología sino que los bucles

que conectan las cadenas β son muy similares, la mayoría de ellos cortos y un bucle DE largo.

Cuando lo que superponemos son los trímeros, tras el alineamiento de 3 x 127 carbonos α obtenemos una desviación cuadrática media de 5.3 Å. Es evidente que el trímero no se superpone tan bien como lo hacen los monómeros independientemente, y esto es debido a la diferente orientación de los monómeros en el trímero. Aunque esta diferente orientación podría ser debida a los diferentes contactos cristalinos, merece la pena recordar que no hemos detectado ninguna diferencia en la orientación de los monómeros dentro del trímero entre nuestras cinco formas cristalinas de σC . La superposición muestra que los monómeros de σC se disponen de una manera más abierta que los monómeros de $\sigma 1$. En ambos casos, el interior del trímero contiene una cavidad con moléculas de agua, pero en el caso de σC esta cavidad tiene una forma cónica, siendo más ancha en la parte superior que en la inferior, mientras que en el caso de $\sigma 1$ esta cavidad tiene una forma más o menos cilíndrica. Esta diferente orientación produce que los contactos entre monómeros de la proteína σC se produzcan principalmente en la región fibrosa y en la parte inferior del dominio globular, en contraposición a $\sigma 1$ cuyos contactos se producen a lo largo de todo el dominio de cabeza. Sin embargo, cuando calculamos el área de la superficie de cada monómero que está participando en los contactos entre monómeros vemos que ésta es mayor en σC de lo que lo es en $\sigma 1$ (≈ 700 versus ≈ 450 Å²). La gran complementariedad de superficie que existe entre los monómeros y el alto grado de conservación que se observa en las interacciones monómero-monómero en las diferentes formas cristalinas resueltas, nos hace suponer que esta disposición abierta que adquiere el trímero de σC no es debida a los contactos cristalinos, sino que existe en solución, y que por lo tanto podría ser importante para la interacción de σC con su receptor celular.

En el dominio globular de cabeza de $\sigma 1$, existe una agrupación de contactos Asp-Asp que podrían ser potencialmente inestables. Estos contactos no existen en σC y en cambio existen dos anillos de puentes salinos en el centro del dominio globular de cabeza. La pequeña superficie de interacción entre los monómeros en el dominio de cabeza así como la agrupación de aspartatos que se forma entre los monómeros de $\sigma 1$ ha llevado a pensar que la naturaleza trimérica de esta proteína podría ser temporal, aunque sólo esta sea capaz de unirse a su receptor (Schelling *et al.*, 2007).

En este sentido la agrupación de aspartatos funcionaría como un interruptor mediado por pH que regularía el estado de oligomerización del dominio globular de $\sigma 1$ (Schelling *et al.*, 2007). En el caso de σC , la superficie que mantiene el trímero parece ser más estable ya que tiene una superficie de interacción monómero-monómero. Por otro lado, no podemos predecir el resultado que tendría un cambio de pH sobre los puentes salinos que mantienen estable el trímero, y por lo tanto el efecto que este cambio tendría sobre la estabilidad de la forma trimérica. En este sentido vale la pena recordar que la proteína ha sido cristalizada a pH 8.5, un valor no demasiado alejado del pH fisiológico.

Además de presentar similitud con la proteína $\sigma 1$, la proteína σC también muestra similitud con las fibras de adenovirus (van Raaij *et al.*, 1999) y con el *spike* del bacteriofago PRD1 (Merckel *et al.*, 2005). Ambas proteínas son triméricas y tienen repeticiones en triple β espiral en su dominio fibroso: 21 en el caso del adenovirus de tipo 2 y 2 en el caso de la fibra de PRD1. En ambos casos sus dominios globulares de cabeza están plegados formando barriles β , que sin embargo tienen una topología diferente. Se ha propuesto (Stehle & Dermody, 2003) que tanto los reovirus aviares como de mamífero comparten un virus ancestro común con un genoma de dsRNA, el cual estaría a su vez evolutivamente relacionado con virus de dsDNA como los adenovirus y el bacteriofago PRD1 (Chappell *et al.*, 2002; Merckel *et al.*, 2005), aunque la baja similitud que existen entre estas secuencias lleva a pensar que se puede tratar de un proceso de convergencia evolutiva. Una característica que todas estas fibras comparten es la existencia de regiones bisagra, que probablemente jueguen un papel importante en el proceso de infección. Los adenovirus tienen regiones bisagra en la parte más N-terminal del dominio fibroso así como entre el dominio fibroso y el dominio globular (van Raaij *et al.*, 1999), la proteína P5 de PRD1 también tiene un región bisagra entre el dominio fibroso y el globular (Merckel *et al.*, 2005) y la proteína $\sigma 1$ del reovirus de mamífero tiene una región bisagra en la parte C-terminal del dominio fibroso (Chappell *et al.*, 2002). La proteína σC , como se ha discutido previamente, tiene una región bisagra justo en el extremo C-terminal de la región plegada en triple *coiled coil* y quizás otra alrededor de la His110 (ver Discusión).

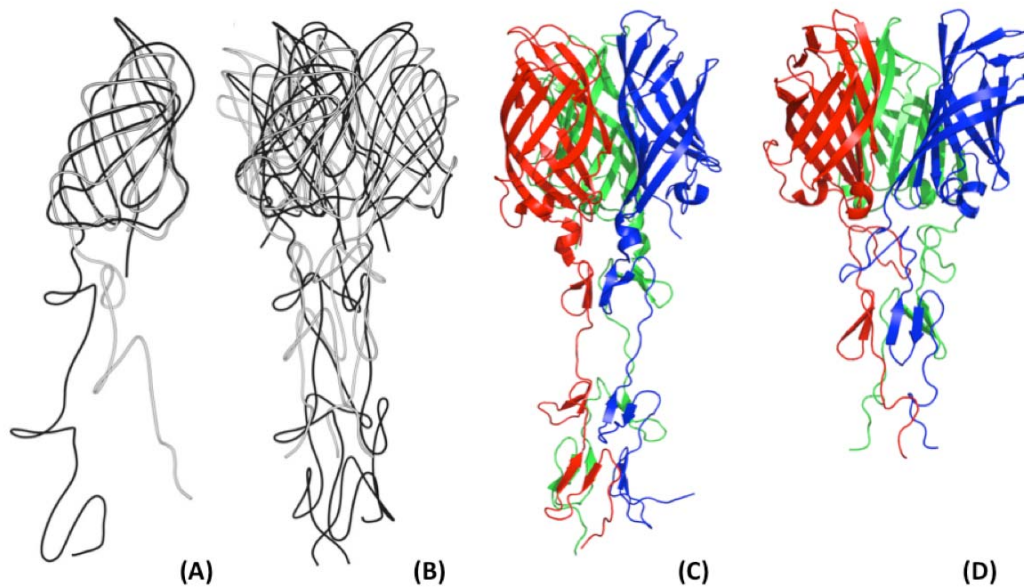


Figura 3.6 Comparación de las estructuras de las fibras del reovirus de mamífero ($\sigma 1$ -en negro) y aviar (σC -en gris). **(A)** Superposición de los monómeros y **(B)** Superposición de los trímeros. Las superposiciones fueron realizadas utilizando el programa LSQKAB (Collaborative Computational Project Number 4, 1994) utilizando únicamente los dominios globulares de cabeza de ambos dominios. **(C)** Estructura del trímero de $\sigma 1$ (1KKE) y **(D)** Estructura del trímero de σC (2BSF)

La única otra fibra viral de estructura conocida con un plegamiento β estructurado es la fibra corta del bacteriofago T4 (van Raaij *et al.*, 2001), aunque en este plegamiento no aparecen repeticiones en triple β espiral como en el caso de la fibra σC y el dominio de cabeza está compuesto por tres monómeros entrelazados (Thomassen *et al.*, 2003), mas que por tres monómeros plegados independientemente como en el caso de las fibras antes descritas. Una característica común para todas estas fibras virales es que el extremo C-terminal de la fibra esta plegada hacia el interior de la estructura, de tal modo que no es accesible al solvente, lo que podría tener importancia al otorgarle a estas fibras resistencia a exo-proteasas que eliminen residuos comenzando por el extremo C-terminal.

3.6 - Estabilidad y plegamiento de σC

Si miramos los monómeros por separado, cada uno de ellos tiene una superficie de $13 \times 10^3 \text{ \AA}^2$, de la cual un 15 % es parte de la superficie de interacción entre monómeros. Esta gran cantidad de superficie de interacción explica, al menos en

parte, la gran estabilidad del trímero de σC , algo que queda patente por el hecho de que la proteína no se disocia en sus monómeros en un gel desnaturante de poliacrilamida con SDS a no ser que sea previamente hervida en un buffer que contenga SDS (Grande *et al.*, 2000; Grande *et al.*, 2002; van Raaij *et al.*, 2005). La estructura confirma que no existen puentes disulfuro en la proteína, como previamente se había descrito (Grande *et al.*, 2002). El tipo de interacciones que estabilizan el trímero son de naturaleza mixta, el núcleo central de las repeticiones en triple β espiral sólo contiene interacciones hidrofóbicas, y la región de triple *coiled coil* esta estabilizada tanto por interacciones hidrofóbicas como por puentes salinos intercatenarios. En el dominio globular encontramos tanto interacciones hidrofóbicas como polares incluyendo dos probables puentes salinos (His223-Asp325 y Asp277-Arg321). Las evidencias bioquímicas (Grande *et al.*, 2002) sugieren que la contribución de las interacciones hidrofóbicas es mayor que la de las polares ya que se ha visto que los detergentes son capaces de desestabilizar el trímero de σC mientras que las sales lo estabilizan. En $\sigma 1$ y las fibras de adenovirus las interacciones que estabilizan el trímero son muy similares, generalmente interacciones hidrofóbicas en la región fibrosa e interacciones de tipo mixto en la globular.

El plegamiento de σC podría suceder de dos maneras. Podría empezar con la interacción entre los barriles β del extremo C-terminal, lo que conllevaría a la formación de un trímero en el dominio de cabeza, tras lo cual la región fibrosa se plegaría empezando por la región con repeticiones en triple β espiral. De este modo el dominio de cabeza y/o el dominio fibroso asegurarían que los tres monómeros estén en pauta. Esta ruta de plegamiento sería similar a la propuesta para la fibras virales de adenovirus (Mitraki *et al.*, 1999). En la fibra $\sigma 1$ el dominio globular de cabeza es incapaz de trimerizar por sí mismo sin la ayuda de al menos una repetición en triple β espiral (Schelling *et al.*, 2007), aunque como hemos visto previamente, el dominio de cabeza de σC es más estable que el de $\sigma 1$. Una ruta alternativa sería aquella en la cual el dominio plegado en triple *coiled coil* sería el primero en plegarse, de tal forma que los residuos en posición d en el núcleo del triple *coiled coil* asegurarían que este plegamiento se produjese con el registro adecuado, y a partir de ahí se plegaría el resto de la proteína.

3.7 - Conclusión

Las estructuras resueltas aportan información a nivel molecular de las interacciones que son necesarias para formar y estabilizar la fibra y sugiere regiones que podrían estar implicadas en la unión al receptor. Los datos estructurales podrán ser usados para profundizar en la primeras etapas del ciclo de replicación viral y puede dar lugar al diseño de aplicaciones terapéuticas dirigidas a enfermedades aviarias, bien a través del diseño de inhibidores de la interacción fibra-receptor celular o bien basadas en la generación de fibras quiméricas artificiales basadas en la estructura de σC y que podrían ser útiles como vectores de vacunación.

Capítulo 4

LA PROTEÍNA σ A

Capítulo 4

Estructura de la proteína de la cápside interna σA

La proteína σA es una proteína estructural que forma parte de la cápside interna del reovirus aviar. Tiene dos funciones fundamentales en la biología del reovirus aviar. Por un lado desempeña una función estructural, anclando entre si las unidades de λA y promoviendo la formación de la cápside externa durante la morfogénesis, de una manera similar a como lo hace su equivalente en reovirus de mamífero, $\sigma 2$. Además de su papel estructural, la capacidad que tiene la proteína σA para unir dsRNA parece estar relacionada con la resistencia que el reovirus aviar presenta al interferón, inhibiendo a la proteína PKR, interacción que ocurriría en el citoplasma celular. En este sentido, la proteína σA parece realizar una función análoga a la proteína $\sigma 3$ del reovirus de mamífero (la equivalente a σB) que también tiene la capacidad de unir dsRNA (Yue *et al.*, 1997), aunque no lo hace con tanta afinidad como σA (Martínez-Costas *et al.*, 2000).

4.1 - Purificación y cristalización

4.1.1 - Expresión y purificación

La proteína σA del reovirus aviar S1133 fue expresada en el vector de expresión pMALCs2 fusionada a MBP (Maltose Binding Protein) como había sido descrito previamente (González-López *et al.*, 2003). La proteína fusionada con MBP fue obtenida de cultivos de *Escherichia coli* transformados con el plásmido PMALCs2- σA . La proteína se obtuvo soluble tras inducir su expresión a 37 °C. Después de la lisis de células y la eliminación del material insoluble por centrifugación, la proteína se purificó por cromatografía de afinidad, utilizando una resina de amilosa. Finalmente, la proteína se eluyó utilizando un tampón con maltosa.

Para la obtención de la proteína σA pura, la proteína MBP- σA se incubó con 1 $\mu\text{g ml}^{-1}$ de factor Xa. Las proteínas σA y MBP fueron posteriormente purificadas mediante cromatografía de intercambio aniónico. Los detalles de la purificación se encuentran en Materiales y Métodos (Sección 2.3.2.1).

4.1.2 - Cristalización

La proteína fue cristalizada mediante el método de difusión de vapor a 5 °C. Se obtuvieron cristales de un tamaño adecuado en 6-12 meses utilizando como solución de cristalización 0.1 M citrato sódico/MES/fosfato sódico 5.5-6.5 y 0.1-0.6 M sulfato amónico. La solución de crioprotección consistía en la solución de cristalización más un 25% de glicerol.



Figura 4.1 Cristales de la proteína σA del reovirus aviar. **(a)** Típica condición de cristalización en donde aparecen múltiples cristales pequeños. **(b)** Un único cristal de gran tamaño (0.3 x 0.3 x 0.05 mm).

4.2 - Determinación estructural

En los patrones de difracción obtenidos de estos cristales se aprecia la presencia de múltiples redes. Se recogieron varios set de datos de distintos cristales hasta conseguir un set de datos con unas estadísticas razonables (menos maclado) que difractaba hasta 2.3 Å utilizando radiación sincrotrón (Tabla 4.I). Los cristales fueron indexados y su grupo espacial resultó ser P1. Un análisis de los coeficientes de Matthews sugiere que en cada celdilla unidad hay entre 6 y 18 moléculas (lo que se correspondería con una fracción de solvente del 81 y 43% respectivamente). Para obtener las fases se utilizó la técnica del reemplazamiento molecular, en donde se usó un modelo basado en la estructura de la proteína $\sigma 2$ del reovirus de mamífero (Reinisch *et al.*, 2000) y en la secuencia de la proteína σA . Finalmente, la estructura se resolvió revelando que la celdilla unidad contiene 12 moléculas y un 62% de solvente.

Las funciones de autorotación (Figura 4.2) muestran varios picos en la secciones $\chi = 180^\circ$ y $\chi = 50^\circ$. En la sección $\chi = 50^\circ$ aparece un único pico con

coordenadas $\theta = 49^\circ$, $\varphi = -90^\circ$, mientras que en la sección $\chi = 180^\circ$ aparecen una serie de picos. La existencia de estos picos indica que existen varios ejes binarios de simetría no cristalográfica. Sus coordenadas sugieren que estos ejes binarios son perpendiculares al eje de simetría no cristalográfica que aparece en la sección $\chi = 50^\circ$. El valor máximo del pico de coordenadas $\theta = 49^\circ$, $\varphi = -90^\circ$ se localiza en las secciones $\chi = 49.7-50.2^\circ$, lo que se desvía más de un grado de lo que sería el valor ideal para un eje de simetría de orden 7 ($\chi \approx 51.4^\circ$). Por otro lado un eje de simetría de este orden, perpendicular a los ejes de simetría binarios localizados, no sería consistente con la solución del reemplazamiento molecular ya que ésta contiene sólo 12 moléculas en la celdilla unidad, en vez de las 14 que serían necesarias.

La estructura de la proteína se resolvió mediante reemplazamiento molecular, en donde se usó un modelo basado en la estructura de la proteína $\sigma 2$ del core del reovirus de mamífero (Reinisch *et al.*, 2000) en el cual los residuos conservados se mantuvieron, mientras que los residuos no conservados se cortaron hasta el carbono γ . El proceso de reemplazamiento molecular fue realizado utilizando MOLREP (Vagin *et al.*, 2000), en donde los factores términos fueron recalculados en función de su superficie accesible al solvente.

El modelo final está compuesto por 12 moléculas, que ocupan un 38% de la celdilla unidad (con un coeficiente de Matthews de $3.2 \text{ \AA}^3 \text{ Da}^{-1}$ (Matthews, 1968)). El refinamiento final fue realizado sin utilizar restricciones basadas en ejes de simetría no cristalográfica y las estadísticas del refinamiento se muestran en la Tabla 4.I. El modelo final consiste en los residuos Met1-Gln37 y Ile51-Ala416 para cada una de las 12 moléculas, 12 iones sulfato y 3003 moléculas de agua. Algunos monómeros también contienen un pequeño vestigio de la cola de purificación tras su escisión con el factor X_a (ISEFGST), mientras que en otros monómeros ésta no es visible. La conformación de este fragmento no será discutida y no ha sido incluida en los análisis posteriores.

Tabla 4.I - Estadísticas cristalográficas

		σA
	Parámetros	Valores
Fuente de radiación		ID23-1 (ESRF)
Temperatura		100 K
Longitud de onda (Å)		0.954
Detector		MAR Mosaic
Distancia detector-cristal (mm)		251
Mosaicidad (°)		0.6
Grupo espacial		P1
Parámetros de celdilla a, b, c (Å)		103.2, 129.9, 144.0
Parámetros de celdilla α, β, γ (°)		93.8, 105.1, 98.2
Reflexiones observadas ^a		282973 (32868)
Rango de resolución (Å)		30-2.34 (2.47-2.34)
Multiplicidad		2.0 (1.9)
Complejidad (%)		94.3 (74.8)
$\langle I/\sigma(I) \rangle$		7.7 (2.2)
R_{symm} (%) ^b		10.5 (30.0)
Estadísticas de refinamiento ^c		
Rango de resolución		30.0-2.34 (2.47-2.34)
Moléculas de proteína / ASU		12
Numero de reflexiones usadas		281587 (32494)
Numero de reflexiones de Rfree		1385 (119)
R-factor ^d		21.0 (26.0)
R-free ^e		27.1 (34.0)
Modelo final		
Número de átomos		
proteína		38161
sulfato		60
agua		3003
Factores térmicos promedio (Å ²)		
proteína		37.2
iones (Cl/Zn ²⁺ /SO ₄ ⁻¹)		60.4
glicerol/agua		37.2
Geometría (desviaciones cuadráticas medias)		
Longitud de enlace (Å)		0.014
Ángulos de enlace (°)		1.4
Estadísticas de Ramachandran ^f		99.5 / 99.9
Código en el PDB		2VAK

^aNo se descartó ninguna reflexión.

$$^b R_{\text{symm}} = \frac{\sum_h \sum_l |I_{hl}| \langle I_{hl} \rangle}{\sum_h \sum_l |I_{hl}|}$$

^cResultados de REFMAC (Murshudov *et al.*, 1997).

$$^d R\text{-factor} = \frac{\sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)||}{\sum |F_{\text{obs}}(hkl)|}$$

^eDe acuerdo con Brunger (1997).

^fDe acuerdo con el programa RAMPAGE (Lovell *et al.*, 2003). Los porcentajes indican residuos en las regiones favorables, permitidas y desfavorable, respectivamente.

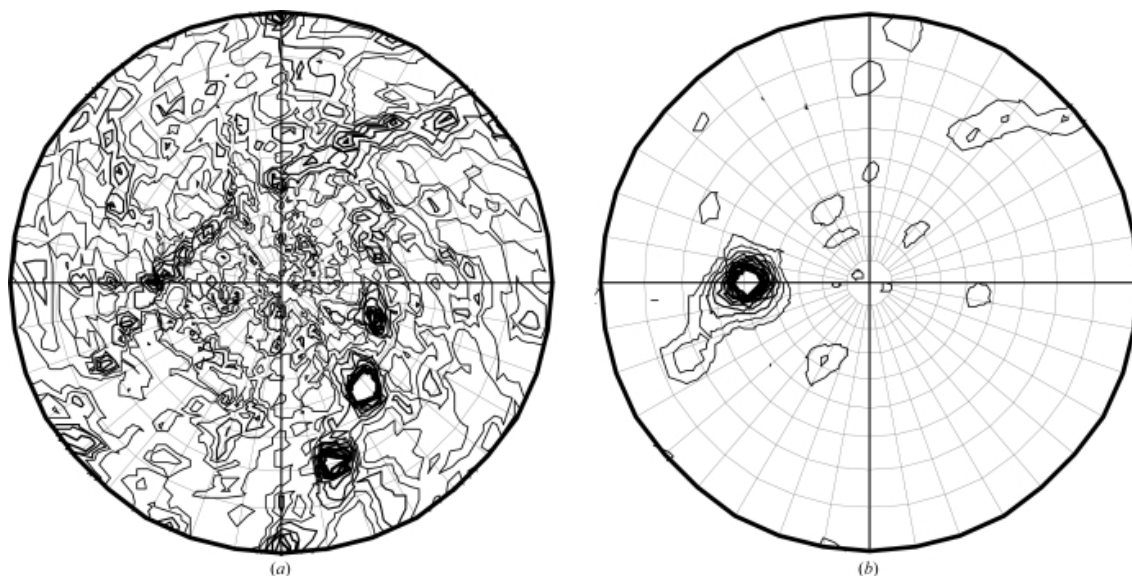


Figura 4.2 Funciones de autorotación calculadas con el programa MOLREP (Vagin & Teplyakov, 2000) a partir de un set de datos completos de los cristales de la proteína σA . (a) Sección $\chi = 180^\circ$ (b) Sección $\chi = 50^\circ$. Se puede apreciar la presencia de múltiples ejes de simetría no cristalográfica binarios y un único eje de simetría no cristalográfica de orden mayor.

4.3 - La estructura y topología de σA

Cada monómero de σA tiene una forma piramidal (Figura 4.3A) con una base aproximadamente rectangular de unos 65-70 Å de longitud y 45-50 Å de ancho, en donde se localiza el posible sitio de unión a la proteína mayoritaria de la cápside interna, λA . Su altura es de aproximadamente 45 Å y en su vértice estaría localizado el sitio de unión a la proteína μB en la partícula reoviral madura. Las doce copias, aún siendo cristalográficamente independientes, son prácticamente idénticas, con unas desviaciones cuadráticas medias entre sus carbonos α en el rango 0.2-0.7 Å (Figura 4.3B). Cada monómero contiene doce hélices α , una lámina β mixta formada por tres cadenas β , y tres láminas β antiparalelas formada por dos cadenas β cada una de ellas (Figura 4.4).

En las doce moléculas presentes en la unidad asimétrica, la región situada entre los residuos 38 y 50 no presenta una densidad electrónica interpretable que se haya podido modelizar. En la proteína $\sigma 2$ la región equivalente estaría entre los aminoácidos 39 y 52 y estaría plegada formando una hélice α en una de las copias de $\sigma 2$ (residuos 39-46) y como una región desestructurada en las otras dos copias

(Reinisch *et al.*, 2000). Esta hélice α forma parte del sitio de unión a $\lambda 1$ en la cápside interna del reovirus de mamífero, lo que nos hace pensar que esta región de la proteína σA se estructuraría tras la interacción con λA formando una hélice α . El hecho de que este bucle sea flexible y pueda adoptar varias conformaciones sugiere que puede ser usado por la proteína para adaptarse a los distintos entornos de unión de la cápside interna.

Tanto el extremo amino como el carboxilo están orientados hacia el vértice de la pirámide y son accesibles al solvente (aunque en el virión estarían tapados por μB). Sin embargo, la cadena- βA , que es el primer elemento de estructura secundaria de la estructura, está escondida en el interior de la estructura. Tanto la Phe8 como Phe9 de esta cadena β están interaccionando tanto con aminoácidos hidrofóbicos como aromáticos. Se pueden distinguir tres dominios en la estructura: un dominio más N-terminal que estaría formado por las hélices- α 1, 3, 4, la posible hélice- α 2 y las láminas- β BC y AD. Este primer dominio estaría empaquetado al lado de un segundo dominio central con la misma topología y que estaría formado por las hélices- α 5, 6, 7 y 8, y las láminas- β GF y HE. Por último, ambos dominios estarían empaquetados con un tercer dominio C-terminal topológicamente distinto y que estaría formado por las hélices- α 9, 10 y 11 y la cadena- β I (que forma parte de la lámina- β HEI). El ápice de la pirámide estaría formado por el bucle entre la hélice- α 8 y la cadena- β I.

Las regiones de σA que están en contacto con λA y μB se pueden predecir a partir de la estructura cristalina del core del reovirus de mamífero [PDB 1EJ6 ((Reinisch *et al.*, 2000)], el modelo de la partícula viral del reovirus de mamífero construido a partir de datos de criomicroscopía electrónica [PDB 2CSE (Zhang *et al.*, 2005b)] y los estudios de criomicroscopía electrónica del reovirus aviar (Zhang *et al.*, 2005a). El lugar de interacción con μB comprende principalmente residuos del tercer dominio mientras que los sitios de interacción con λA estarían casi exclusivamente en los dominios 1 y 2. Los dominios 1 y 2 presentan los mismos elementos de estructura secundaria en la superficie de interacción con λA , lo que podría implicar que ha existido una duplicación de dominios a partir de la cual cada uno habría evolucionado de manera independiente para interaccionar con distintas regiones de λA .

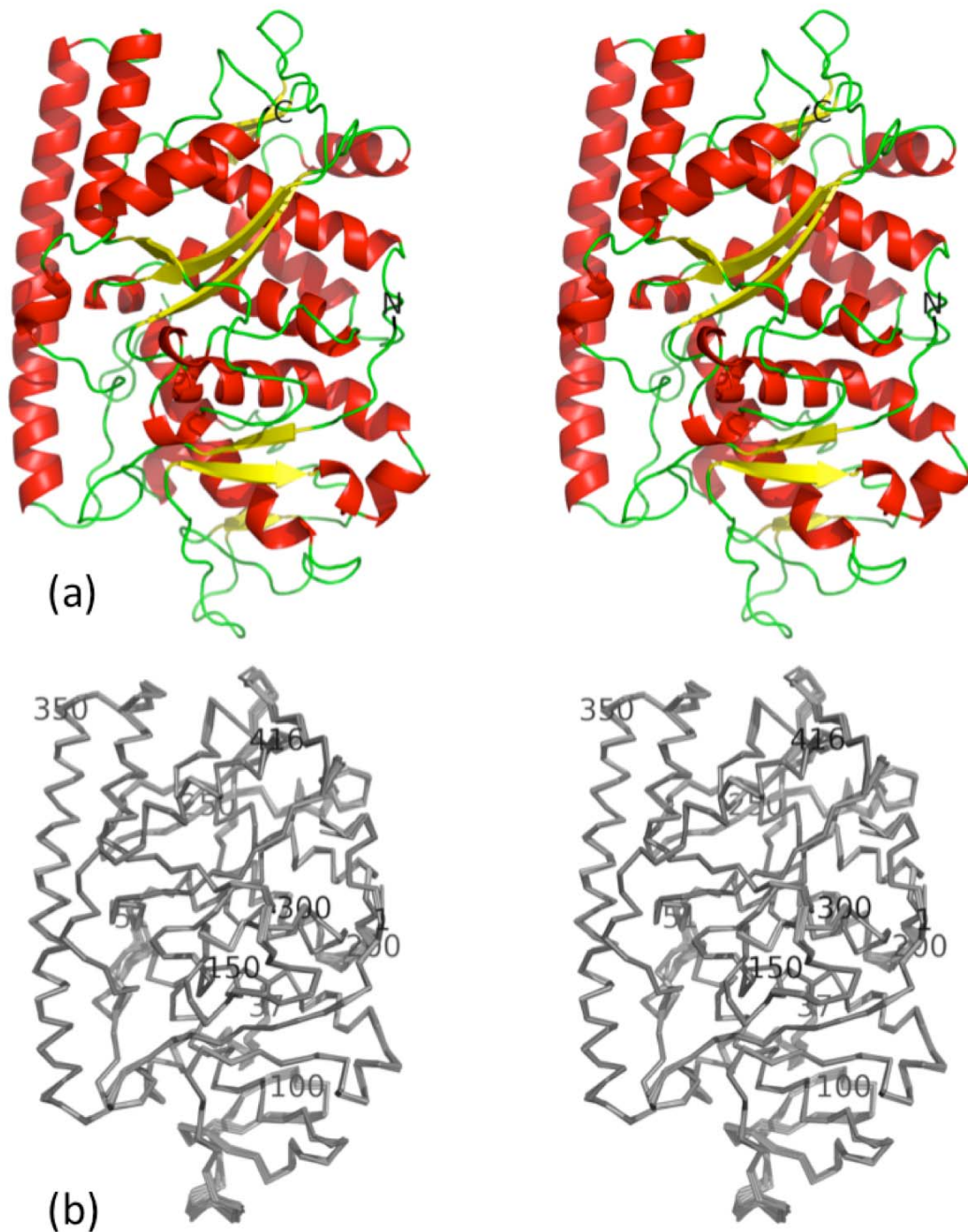


Figura 4.3 Estructura de la proteína σA del reovirus aviar. **(a)** Vista estereoscópica de la parte superior de la proteína. Para realizar la figura se usó el monómero con el factor térmico promedio menor, el monómero G. Los extremos N y C-terminales están marcados en la figura. **(b)** Vista estereoscópica en donde se muestra la superposición de las 12 moléculas presentes en la celdilla unidad. Se indican la metionina amino terminal, la alanina C-terminal, y un residuo de cada 50, así como los residuos que están justo antes y después de la región desordenada (Gln37 y Ile61).

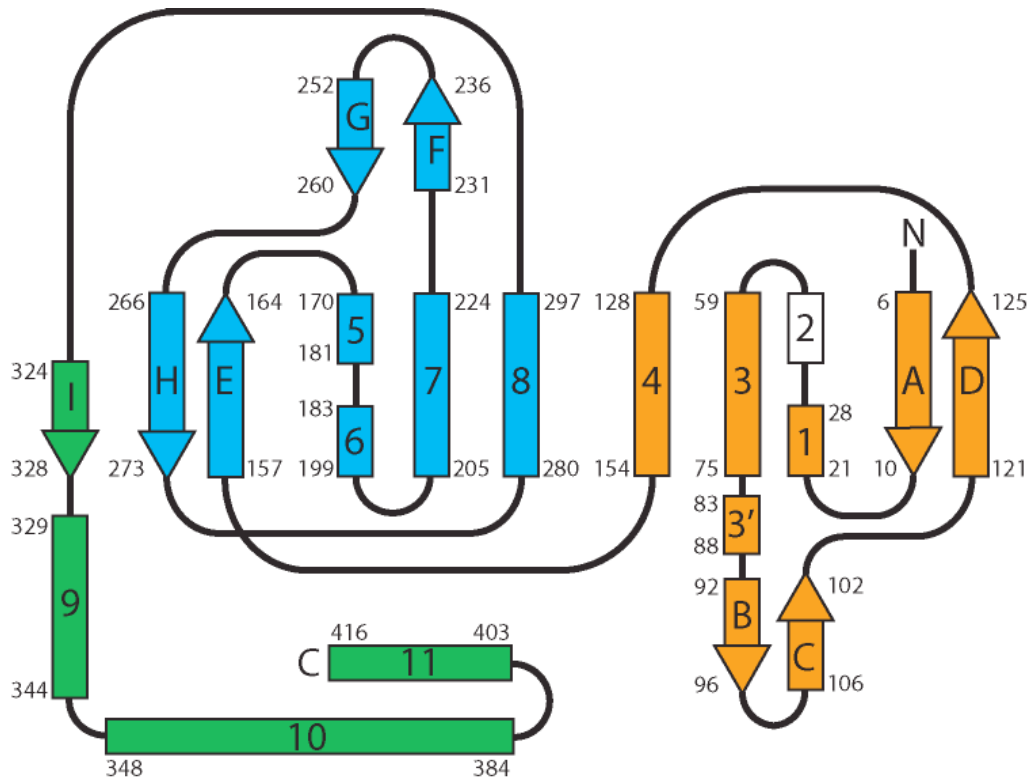


Figura 4.4 Diagrama topológico de la estructura de σA . Las hélices- α se muestran como rectángulos etiquetados con números, las cadenas- β se muestran como flechas etiquetadas con letras. Los dos dominios estructuralmente homólogos son mostrados en naranja y azul, mientras que el dominio carboxi-terminal se muestra en verde. La hélice- α 2 (en blanco) no aparece en la estructura pero su estructura se ha inferido tanto por la estructura de la proteína homóloga $\sigma 2$ de reovirus de mamífero como de la homología con el dominio central. Se ha etiquetado cada elemento de estructura secundaria con el residuo inicial y el final.

4.4 - Comparación estructural y funcional con otras proteínas

Usando el servidor DALI y el servicio SSM del EBI se realizó una búsqueda de estructuras con similitud estructural tanto con σA entera como por cada uno de sus dominios por separado. Dicha búsqueda reveló que la única proteína de estructura conocida con una homología estructural destacable fue la proteína $\sigma 2$ de reovirus de mamífero con que comparte un 29% de identidad de secuencia. La topología general de ambas proteínas es idéntica y sólo se detectan pequeñas diferencias en determinadas regiones que podrían ser atribuidas a su flexibilidad inherente, puesto que la mayoría de estas regiones también varían entre las doce copias de σA en la celdilla unidad. La única diferencia notable entre σA y $\sigma 2$ se encuentra en la región C-terminal de la hélice 10 (aminoácidos 374-388 en la estructura de σA) que está claramente distorsionada en σA .

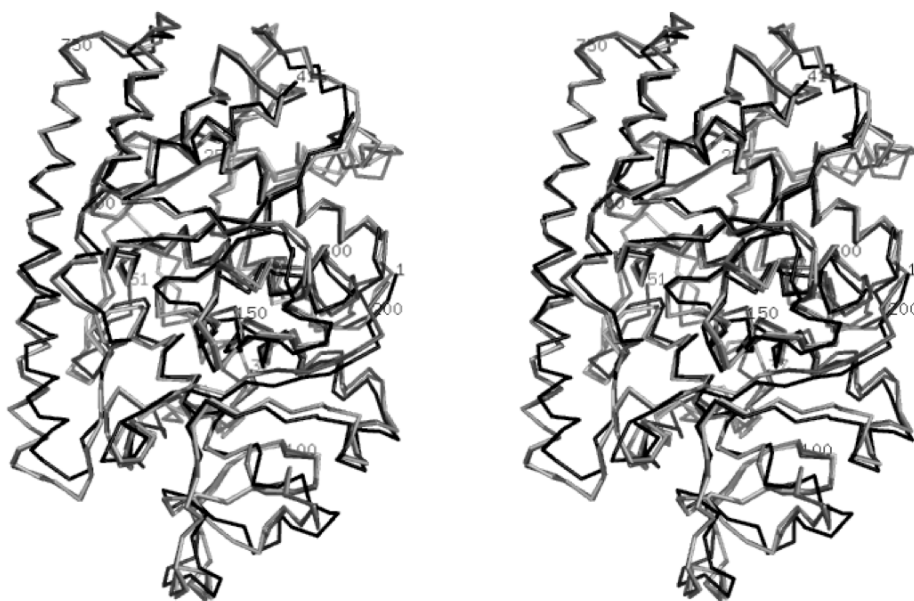


Figura 4.5 Vista estereoscópica en donde se muestra la superposición de las cadenas principales de las proteínas σA sobre $\sigma 2$. La cadena G de la proteína σA ha sido superpuesta sobre las cadenas D y E de la proteína $\sigma 2$ [PDB 1EJ6 (Reinisch *et al.*, 2000)]. La desviación cuadrática media de sus carbonos α es de 1.6 Å (Krissinel *et al.*, 2004).

4.5 - Estudio de la interacción de la proteína σA con dsRNA

Con respecto a su capacidad de interactuar con dsRNA, σA es funcionalmente, aunque no estructuralmente, análoga a la proteína $\sigma 3$ del reovirus de mamífero. Ambas proteínas unen dsRNA y han sido implicadas en la inhibición de la PKR (Gonzalez-Lopez *et al.*, 2003; Imani *et al.*, 1988). De la misma manera que σA , la capacidad de $\sigma 3$ de unir dsRNA está relacionada con su capacidad para inhibir la activación de la PKR compitiendo por su sustrato, el dsRNA. Ha sido propuesto que la proteína $\sigma 3$ se una al dsRNA en forma dimérica y que múltiples dímeros se alinearían a lo largo de la molécula de dsRNA (Olland *et al.*, 2001).

En un mapa de potencial electrostático de la proteína σA , se pueden apreciar varias regiones cargadas positivamente en la superficie de la proteína. Una de estas regiones contiene, entre otros residuos, la Arg155 y la Arg273 (Figura 4.6a). Cuando se mutan independientemente cualquiera de estos dos residuos por alanina y se analiza la capacidad de la proteína mutada de unir dsRNA mediante un ensayo de *gel shift*, se ve que ambos mutantes carecen de la capacidad de unir dsRNA (Figura 4.6b), mientras que si las mutaciones se realizan en otras regiones cargadas electropositivamente (Arg134, Arg135, Arg389 y Arg390) la capacidad de unión a

dsRNA no se ve alterada. (Estos experimentos de mutagénesis dirigida así como los ensayos de unión fueron realizados por Lorena Vázquez-Iglesias). Las proteínas mutadas son solubles y no hay indicaciones de que ninguna de las mutaciones afecten a la estabilidad de la proteína o que tengan algún impacto en su estructura. Otra indicación a favor de que esta región puede jugar un importante papel en la unión a dsRNA es la presencia de un ión sulfato cerca de la Gln305, en la misma región que las Arg155 y Arg273. Este sulfato está presente en las doce copias de la proteína en la celdilla unidad y puede estar mimetizando la interacción con un fosfato de la cadena de dsRNA.

En los ensayos de *gel shift* usando la proteína σA nativa, se observan complejos con un alto peso molecular que apenas son capaces de entrar en el gel separador (Figura 4.6b, panel central). Esto puede sugerir que o bien los complejos σA -dsRNA son de un gran tamaño o que los complejos resultantes pueden estar sin carga al pH utilizado en el ensayo (8.0). Esto último encajaría con un modelo en el cual, varias moléculas de σA (punto isoeléctrico teórico 8.6) cubrirían la mayoría de la superficie de dsRNA, enmascarando las cargas negativas de los fosfatos y produciendo un complejo σA -dsRNA sin carga. Los experimentos de unión a dsRNA indican que la proteína σA es capaz de unir un dsRNA de al menos 21 bp (Figura 4.6b, panel de la izquierda). También se aprecia que, conforme aumenta la longitud de las moléculas de dsRNA se necesita una concentración menor de σA para unirlas, lo que indica la existencia de una interacción cooperativa.

4.6 - Estudio de las interacciones entre las moléculas de σA en el cristal

Se ha estudiado el estado oligomérico de la proteína σA en solución mediante microscopía electrónica (Figura 4.8) y ultracentrifugación analítica (Tabla 4.II). Todos los resultados apuntan a que σA es monomérica en solución. Sin embargo, a 4 °C y en ausencia de sal, la proteína tiende a precipitar reversiblemente, lo que indicaría que tiene tendencia a formar agregados de orden superior.

En la estructura las doce moléculas independientes se empaquetan en lo que a primera vista parece ser una doble hélice (Figura 4.7a). Cuando las interacciones son analizadas usando el servidor PISA (Krissinel *et al.*, 2007) se observan tres tipos de interacciones con potencial relevancia biológica, que denominamos “cola-cola”, “cabeza-cola” y “cabeza-cabeza” según qué superficies están participando en la

interacción. La cola esta compuesta por residuos del bucle N-terminal, las cadenas- β A y D, el bucle que conecta a esta última con la hélice- α 4, el bucle entre las cadenas- β G y H y el bucle entre la hélice- α 8 y la cadena- β I. La cabeza esta compuesta por los aminoácidos del bucle entre la posible hélice- α 2 y la hélice- α 3, por la hélice- α 10 y el bucle entre las hélices- α 10 y 11. La interacción “cola-cola” tiene una superficie de interacción de 690 \AA^2 , una ganancia de energía de solvatación de $12.4 \text{ kcal mol}^{-1}$, tres puentes de hidrógeno y un P-value de 0.033. La interacción “cabeza-cola” tiene una superficie de interacción de 467 \AA^2 , una ganancia de energía de solvatación de $7.4 \text{ kcal mol}^{-1}$, dos puentes de hidrógeno y un P-value de 0.054, mientras que la interacción “cabeza-cabeza” tiene una superficie de interacción de 373 \AA^2 , una ganancia de energía de solvatación de $4.7 \text{ kcal mol}^{-1}$, dos puentes de hidrógeno y un P-value de 0.26. Ninguna de las interacciones observadas entre moléculas de σA en el cristal son similares a las observadas entre moléculas $\sigma 2$ en la estructura del *core* del reovirus de mamífero.

La interacción “cabeza-cola” es la única interacción que relaciona a las doce moléculas de σA de la celdilla unidad. Esta interacción divide las doce moléculas en cuatro trímeros, de tal forma que en cada uno de estos trímeros se producirían dos interacciones “cabeza-cola”. Cada trímero esta relacionado con su opuesto en la doble hélice por una rotación de aproximadamente 180° . Si superponemos ambos trímeros vemos un ajuste muy bueno (desviaciones cuadráticas medias de sus carbonos α de 0.4 \AA) lo que sugiere que los trímeros se comportan como un cuerpo rígido. Los monómeros de cada trímero están relacionados entre sí por una rotación de aproximadamente 50° alrededor de un eje que pasaría por el centro de la doble hélice, lo que explicaría el pico que se registra en la sección $\chi = 50^\circ$ de función de auto rotación. Es interesante resaltar que la forma A del dsRNA tiene una rotación helicoidal de alrededor de 50° .

Cada uno de los hexámeros así formado se relaciona con el otro hexámero de la unidad asimétrica por una rotación de aproximadamente 180° . En este caso el ajuste entre ambos hexámeros no es demasiado bueno (desviación cuadrática media de 1.6 \AA).

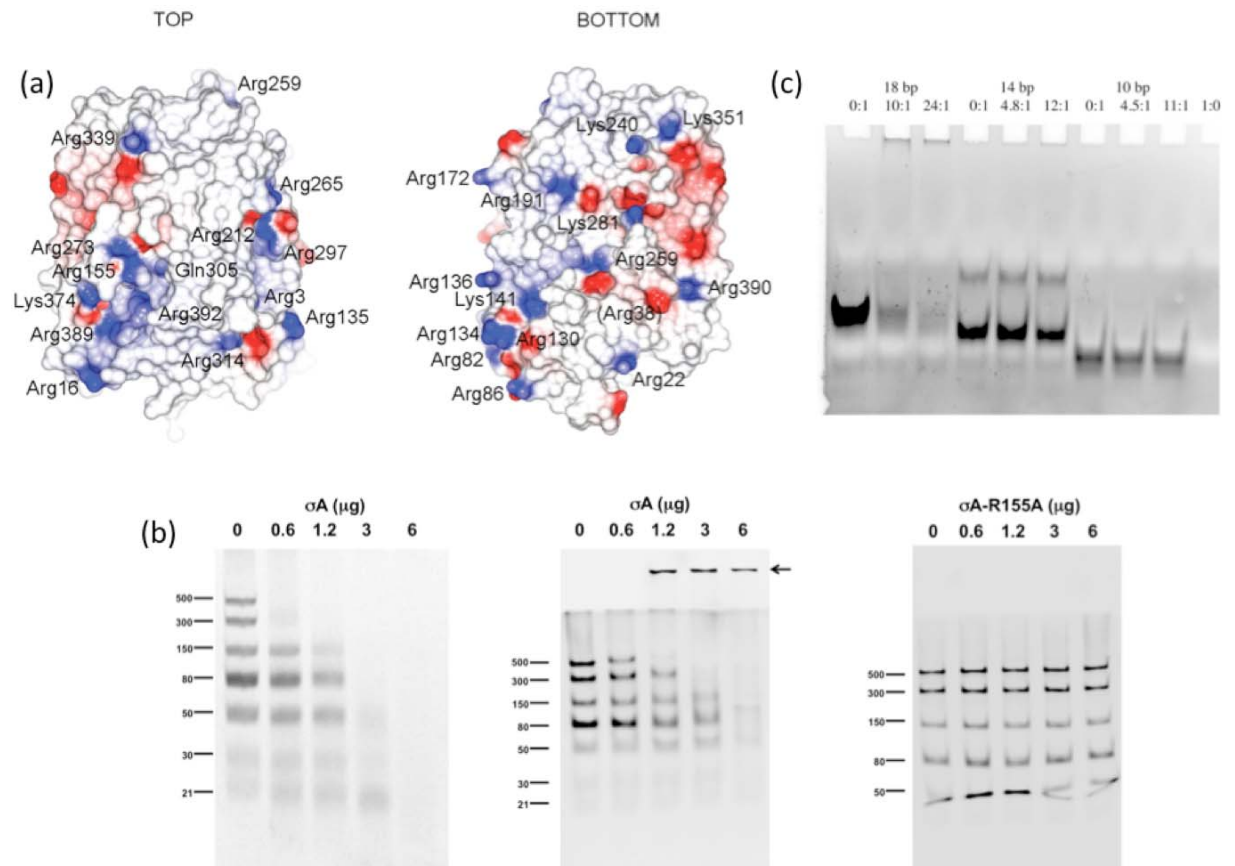


Figura 4.6 Unión de la proteína σA a dsRNA. **(a)** Mapa de potencial electrostático de la proteína σA . En la figura de la izquierda la orientación es la misma que en la figura 4.3a; la orientación de la figura de la derecha ha sido rotada 180° alrededor de un eje vertical centrado en el ápice de la pirámide. Los residuos cargados positivamente de la superficie han sido etiquetados. También se muestra la localización de la Gln305, la Arg38 se muestra entre paréntesis ya que solo esta parcialmente expuesta. **(b)** Experimentos de *gel shift*. Los ensayos fueron realizados usando la proteína σA a una concentración de 0.6 mg/ml en buffer TE (10 mM TrisHCl pH 8.5, 1 mM EDTA) y un marcador de peso molecular de dsRNA que consiste en siete tipos distintos de dsRNA de diferentes longitudes (500, 300, 150, 80, 50, 30 y 21 pares de bases) a 0.5 mg/ml. Se incubaron cantidades crecientes de σA (0, 0.6, 1.2, 3.0 y 6 μg) con 0.5 μg de marcador de dsRNA en buffer TE suplementado con 150 mM de NaCl durante 10 minutos en hielo. Posteriormente las muestras fueron analizadas en un gel no desnaturalizante. (Izquierda) Se usó un gel al 10% de poliacrilamida en buffer TBE. (Centro y derecha) Fueron usados geles discontinuos como describió Laemmli (Laemmli, 1970) con un gel separador de acrilamida al 10% y la omisión de SDS de la muestra, el gel y el buffer. En los paneles de la izquierda y el centro se utilizó σA nativa, mientras que el panel de la derecha muestra un ensayo de *gel shift* realizado con el mutante R155A. Se obtuvieron resultados idénticos con el mutante R273A. La flecha en el panel central indica los complejos de alto peso molecular que se observan. (Experimentos realizados por Lorena Vazquez-Iglesias) **(c)** La interacción de la proteína σA con fragmentos más pequeños de dsRNA se estudió mediante experimentos de *gel shift*. En los tres primeros carriles se utilizó un dsRNA de 18 pares de bases, en los tres siguientes uno de 14 pares de bases y en los carriles 7 a 9 un dsRNA de 10 pares de bases. En el último carril se cargó la proteína σA sola. Los ratios σA /dsRNA usados se muestran sobre cada carril.

Tomando todo en su conjunto, se podría proponer un modelo en el que el trímero actuaría como una unidad biológica, o al menos la interacción “cabeza-cola” sería la interacción biológica con la que las moléculas de σA se dispondrían alrededor del dsRNA, bien formando una hélice simple o bien una hélice doble. Estos modelos podrían explicar la interacción cooperativa entre la proteína σA y el dsRNA así como los grandes complejos observados en los experimentos de *gel shift*. Las regiones cargadas electropositivamente que se han discutido anteriormente en donde se localizan las argininas responsables de la interacción con el dsRNA están situadas en el interior de la hélice. El ancho de la cavidad en el interior de la doble hélice que forman las moléculas de σA en el interior del cristal varía entre 11 y 35 Å, siendo siempre mayor de 20 Å en una de las dimensiones. La sección de la forma A del dsRNA tiene unas dimensiones de 10 x 20 Å. Por lo tanto, las dimensiones de la forma A del dsRNA podrían hacer posible que éste encajase dentro de la doble hélice, aunque serían necesarios algunos ajustes menores en el dsRNA o en la doble hélice que forman las moléculas de σA .

La distancia entre las argininas 155 y 273 de la misma molécula es de 8.5 Å, mientras que la distancia entre argininas equivalentes de moléculas contiguas es de 39-40 Å. En un modelo (Figura 4.7b) construido a partir de los trímeros observados en la estructura cristalina y la forma A canónica del RNA bicatenario, la cadena lateral de una de las argininas podría interaccionar con un grupo fosfato a un lado del surco menor mientras la cadena lateral de la otra arginina interaccionaría con un grupo fosfato del otro lado del surco (la distancia entre grupos fosfatos a ambos lados del surco menor es de 10 Å). La siguiente molécula podría hacer la misma interacción con grupos fosfato a ambos lados del surco menor 13 pares de bases más allá (la distancia entre fosfatos separados 13 pares de bases es de 38 Å). Para confirmar este modelo serían necesarios estudios estructurales del complejo σA -dsRNA. También se podría intentar mutar residuos en la superficie de interacción entre los monómeros, aunque una inspección de esta superficie no muestra ninguna mutación obvia que pueda afectar seriamente a la superficie sin comprometer la integridad estructural de la proteína.

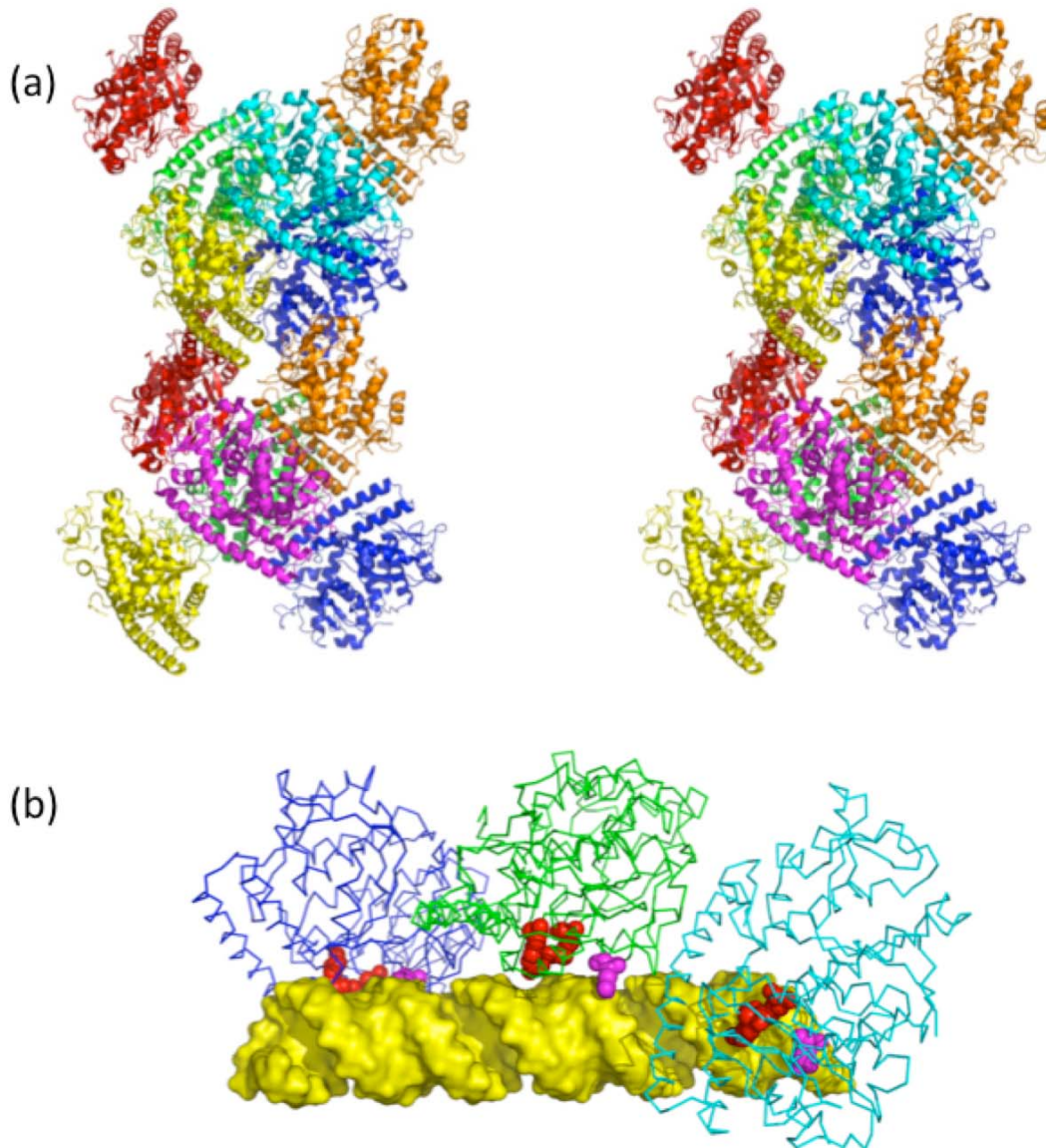


Figura 4.7 Visión estereoscópica de la unidad asimétrica del cristal y modelo de interacción con la forma A de dsRNA. **(a)** Visión estereoscópica de la doble hélice que forman las moléculas de σA en la celdilla unidad del cristal. **(b)** Modelo de interacción de un trímero de moléculas de σA con la forma A de dsRNA con una longitud de 36 pares de bases (amarillo). Los residuos implicados en esta interacción (Arg155 y Arg273) se muestran en rojo, la Gln305 se muestra en magenta.

4.7 - Estudio de la estequiometría de la interacción σA -dsRNA

Para obtener evidencias experimentales independientes que nos permitieran obtener evidencias que apoyasen el modelo de interacción previamente propuesto, y que nos permitiese discriminar entre un modelo en donde la proteína σA se organizase en una hélice sencilla o en una doble, decidimos realizar experimentos de ultracentrifugación analítica con mezclas dsRNA- σA y microscopía electrónica de transmisión de muestras de dsRNA- σA teñidas.

Tabla 4.II. Resultados de velocidad de sedimentación

Componente o complejo	Calculado	$S_{20,w}$ (10^{-13}) (%)									
		Experimento 1 (exceso σA)					Experimento 2 (exceso dsRNA)				
		σA	dsRNA 36bp	Mezcla			dsRNA 36bp	Mezcla			
σA	3.9	3.6 (100)	NO ^a	1/0.21 3.6 (71)	1/0.41 3.7 (57)	1/0.62 3.5 (35)	NO	1/1.0 NO	1/3.1 NO	1/9.3 NO	1/31 NO
dsRNA	3.4	NO	3.4(100)	NO	NO	NO	3.6 (100)	3.2 (62)	3.2 (75)	3.2 (77)	3.2 (75)
Complejo											
1:1	5.6	NO	NO	5.8 (7)	6.0 (11)	5.4 (16)	NO	5.2 (21)	4.4 (19)	4.3 (23)	4.2 (25)
2:1	7.3	NO	NO	7.6 (9)	8.0 (12)	7.4 (18)	NO	6.9 (15)	5.7 (6)	NO	NO
3:1	8.8	NO	NO	9.0 (12)	9.4 (13)	8.8 (25)	NO	NO	NO	NO	NO

4.7.1 - Estudios de ultracentrifugación analítica

Los experimentos de ultracentrifugación analítica fueron realizados en el CIB como ha sido descrito en Materiales y Métodos. Tanto con la proteína σA como con el dsRNA de 36 pares de bases sin mezclar se obtuvo un único pico con un coeficiente de sedimentación adecuado, confirmando que ambas muestras eran monodispersas y del tamaño adecuado. Cuando se le añadió un exceso de proteína a diferentes cantidades de dsRNA (experimento 1), se obtuvieron cuatro picos reproducibles. Los coeficientes de sedimentación de estos picos son compatibles con la forma monomérica de σA y tres complejos σA :dsRNA con estequiometrías 1:1, 2:1 y 3:1 respectivamente. En el experimento 2, se añadió un exceso de dsRNA a la proteína, el dsRNA sólo dio un único pico compatible con un dsRNA de 36 bp mientras en las mezclas se observaron dos o tres picos de los cuales uno claramente pertenecía al dsRNA sin acomplejar. Los picos pertenecientes a los complejos tenían coeficientes de sedimentación menores de lo esperado para complejos σA :dsRNA del tipo 1:1 y 2:1 lo que sugiere que existe un equilibrio entre el dsRNA libre y los diferentes complejos formados. Tomando los datos de ultracentrifugación analítica en su conjunto estos parecen sugerir que a un dsRNA de 36 pares de bases se pueden unir hasta un máximo de tres moléculas de σA , lo que sería compatible con nuestro

modelo de unión en donde las moléculas de σA cubren el dsRNA formando una hélice sencilla (Figura 4.7b).

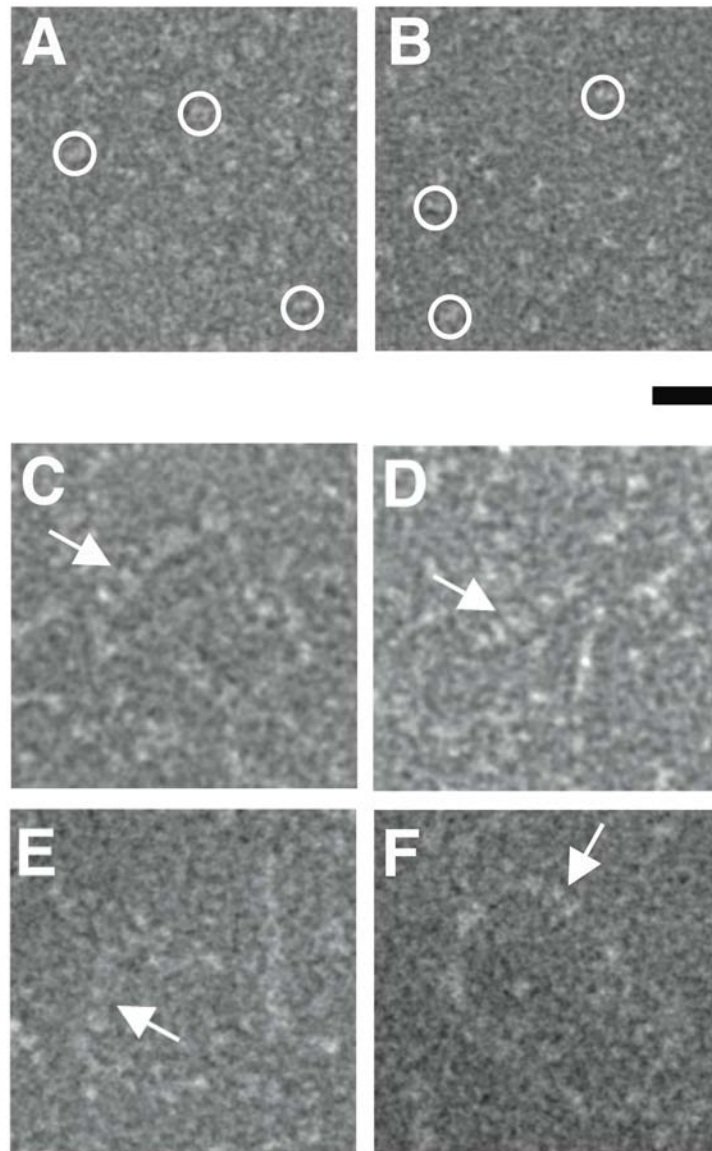


Figura 4.8. Fotografía de microscopía electrónica de transmisión de mezclas σA -dsRNA. (A y B) Microscopía electrónica de transmisión de muestras de σA sólo, en donde se aprecia que es monomérica. Las partículas del tamaño adecuado están indicadas al ser rodeadas con un círculo. (C-F) Posibles complejos σA -dsRNA indicados con flechas. La barra son 20 nm.

4.7.2 - Estudios de microscopía electrónica

Los experimentos de microscopía electrónica (Figura 4.8) también muestran múltiples moléculas de σA unidas al dsRNA (Figura 4.8, paneles c-f) mientras que σA aparece monodispersa en ausencia de dsRNA (Figura 4.8, paneles a-b). El ancho de los complejos observados varía entre 54 y 98 Å y no nos permite distinguir entre

un modelo de hélice sencilla (que tendría un diámetro de alrededor de 60 Å) y un modelo de doble hélice (con un diámetro aproximado de unos 100 Å). En realidad, la unión podría ser mixta, con regiones de dsRNA cubiertas por una hélice sencilla de σA y regiones cubiertas por una doble hélice, dependiendo de la estructura local del dsRNA. Las fotografías de microscopía electrónica muestran unas estructuras con una longitud claramente mayor a la esperada para una molécula de dsRNA de 36 pares de bases, lo que sugiere que múltiples complejos σA -dsRNA podrían alinearse, presumiblemente a través de interacciones σA - σA .

4.8 - Conclusión

Los datos estructurales de la proteína σA en conjunto con los experimentos de *gel shift*, velocidad de sedimentación y microscopía electrónica nos han llevado a proponer un modelo de interacción σA -dsRNA que podría explicar tanto la cooperatividad de esta unión como el mecanismo por el cual esta proteína es capaz de impedir la activación de la proteína PKR. Además el análisis de la estructura ha revelado la existencia de tres dominios diferenciados. Los dos primeros dominios del extremo N terminal, que tienen la misma topología, están involucrados en la interacción con λA y son los responsables de la estabilización de la cápside interna. El dominio más C-terminal, que tiene una topología diferente, parece estar involucrado en la interacción con μB y por lo tanto con la formación de la cápside externa durante la morfogénesis viral.

Capítulo 5

LA FIBRA LARGA DEL ADENOVIRUS AVIAR 1

Capítulo 5

Estructura del dominio C-terminal de cabeza de la fibra larga del adenovirus aviar de tipo 1

El adenovirus aviar CELO (Chicken Embryo Letal Orphan), también conocido como adenovirus aviar de tipo 1 o de tipo A, es un adenovirus relativamente benigno que ha despertado interés para su uso en terapia génica (Kelleher & Vos, 1994; Stevenson *et al.*, 2006). Una de sus características más destacadas es que cada uno de los vértices de su cápside contiene dos proteínas fibrosas triméricas insertadas en la proteína pentamérica de la base del pentón, una fibra corta (fibra 2) y una larga (fibra 1) (Hess *et al.*, 1995). Se ha mostrado que mientras la fibra corta juega un papel clave en el proceso de infección en las células de pollo (Tan *et al.*, 2001), la fibra larga no es esencial, lo que la convierte en un diana atractiva para su modificación.

Un análisis de la secuencia de la fibra larga revela que entre sus 793 residuos existen dos regiones ricas en glicinas (Hess *et al.*, 1995) que probablemente funcionen como regiones bisagra, una cerca del extremo N-terminal (residuos 51-62) y la otra situada entre el dominio fibroso y el dominio globular (residuos 577-579). El dominio de cabeza está compuesto por los residuos 579-793 y es comparable en tamaño a los dominios de cabeza de algunos adenovirus humanos cuyas estructuras han sido resueltas, aunque su homología de secuencia es limitada (14-17%).

5.1 - Purificación y cristalización

5.1.1 - Construcción de los vectores de expresión

Se clonó un fragmento de ADN que codificaba los residuos 579-793 (el dominio de cabeza) de la fibra larga del adenovirus aviar en un vector de expresión como ha sido descrito en Materiales y Métodos. Además del dominio de cabeza se clonaron los fragmentos 62-793 y 398-793, que resultaron insolubles cuando se intentaron purificar. También se clonaron los fragmentos 398-793, 430-793, 460-793, 490-793, 520-793 y 537-793, todos ellos sin la región bisagra de 3 glicinas (residuos 577-579). Todos resultaron insolubles cuando se intentaron purificar. El fragmento 398-793 sin las tres glicinas también se clonó en pMALcX2: expresaba una proteína soluble pero cuando se cortaba la proteína de fusión MBP la proteína adenoviral

precipitaba. Finalmente se clonaron los fragmentos 61-576, 195-576 y 398-576 fusionados a un dominio de trimerización (foldon) en su extremo C-terminal. Las construcciones (61-576)foldon y (195-576)foldon fueron insolubles, mientras que la construcción (398-576)foldon era soluble y trimérica. Aunque se purificó, no fue posible cristalizarla.

5.1.2 - Expresión, purificación y cristalización del dominio de cabeza

Para expresar el dominio de cabeza (aminoácidos 579-793) de la fibra larga del adenovirus aviar, se transformaron bacterias con el plásmido pCELOlongfib579-793, estas se crecieron y usaron para expresar la proteína como ha sido descrito en Materiales y Métodos. Tras la lisis de las bacterias, la proteína se purificó de la fracción soluble por cromatografía de afinidad por níquel e intercambio catiónico. La proteína se concentró a 23 mg ml^{-1} utilizando concentradores Centricon (Millipore, Madrid, España). La cristalización se llevó utilizando la técnica de difusión de vapor en gota sentada. La proteína cristalizó usando PEG 10000 como precipitante a pH 4.7-6.1 a 5°C .

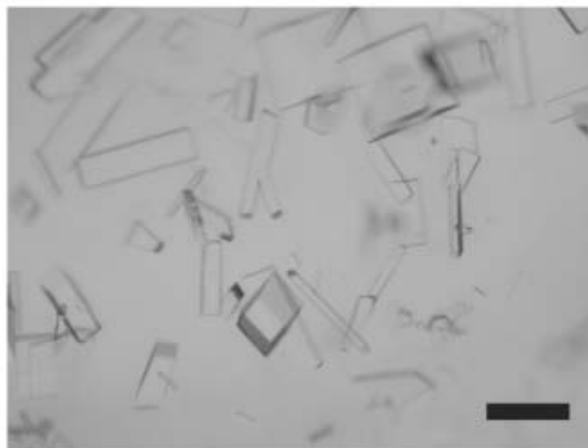


Figura 5.1. Cristales de el dominio de cabeza. La barra mide aproximadamente 0.25 mm

5.2 - Determinación estructural de dominio de cabeza de la fibra larga del adenovirus CELO

Los cristales se incubaron en presencia de distintas concentraciones de cloruro de metilmercurio. Debido a la falta de isomorfismo entre los cristales derivados con mercurio y los cristales nativos no fue posible resolver la estructura utilizando los datos nativos. Para resolver este problema los cristales derivados con bajas

concentraciones de cloruro de metilmercurio fueron tratados como pseudo-nativos, mientras que los cristales derivados en presencia de una alta concentración de cloruro de metilmercurio fueron usados como derivados. De esta forma se localizaron ocho átomos pesados utilizando el programa SHELXD (Schneider & Sheldrick, 2002) de los cuales seis estaban relacionados por un eje de simetría ternario. Este operador de simetría fue utilizado para mejorar las fases, tras lo cual se obtuvo un mapa de fácil interpretación. Usando este mapa, se modelaron automáticamente 613 residuos y el modelo resultante se utilizó para resolver la estructura a alta resolución mediante reemplazamiento molecular. El modelo final contiene los residuos 583-793 de cada uno de los tres monómeros del trímero y no se observa ninguna incoherencia entre la densidad electrónica y la secuencia de aminoácidos. En el mapa no se observa densidad electrónica para los tres residuos del extremo N-terminal (579-582) y la cola de purificación. Presumiblemente esta región es flexible, lo que permite a estos aminoácidos adoptar conformaciones diferentes en el cristal. Las estructuras refinadas se ajustan bien a los datos, con una buena geometría y pocos residuos en las regiones desfavorables del diagrama de Ramachandran (Tabla 5.I). Sólo dos residuos se encuentran en regiones generosamente permitidas, la Asn694 del monómero B y el Asp705 del monómero C. El primero tiene una densidad electrónica clara que inequívocamente apoya la conformación del modelo y el segundo es parte de un bucle relativamente desordenado con una densidad electrónica mal definida (para un resumen de las estadísticas de los datos, véase la Tabla 5.I)

5.3 - La estructura del dominio de cabeza

Cada monómero del dominio de cabeza está plegado formando un sándwich β con una topología similar a la de otros dominios de cabeza de adenovirus humanos previamente descritos (Figura 5.2). Cada monómero está compuesto por ocho cadenas β , que a su vez se subdividen en dos láminas β (ABCJ y GHID) utilizando la nomenclatura descrita en Xia *et al.* (1994). La mayoría de los bucles entre las distintas cadenas β son cortos, con la excepción del bucle DG (residuos 659-715) en el que está presente una pequeña cadena β E (residuos 672-677), que interacciona con la cadena β A. Tanto la cadena β A como la G son discontinuas. La cadena β J contiene una discontinuidad causada por dos prolinas sucesivas (Pro782 y Pro783). Cuando se superponen, los tres monómeros exhiben una desviación cuadrática media entre sus

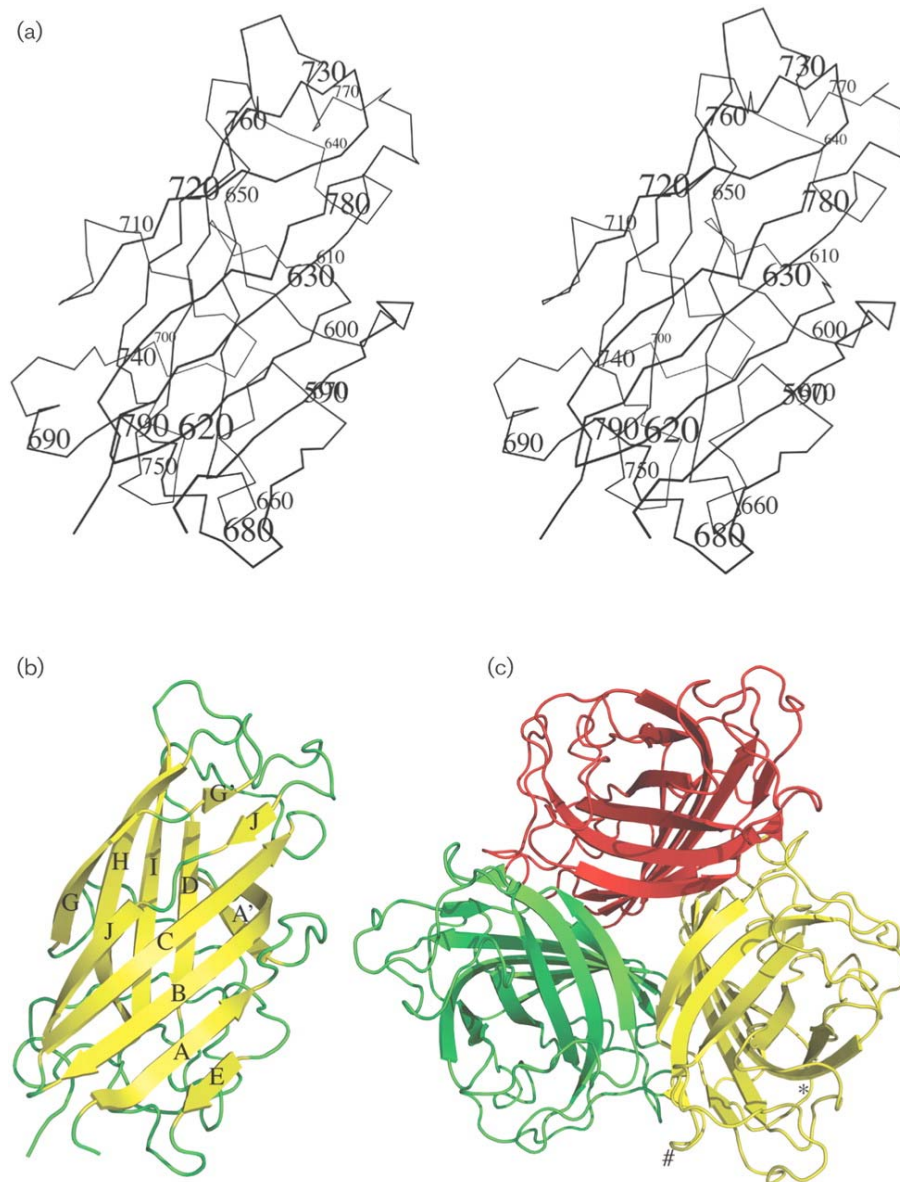


Figura 5.2. Estructura del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1. **(a)** Vista estereoscópica del monómero visto de lado. Los carbonos α están conectados por líneas y han sido etiquetados cada 10 residuos. **(b)** Estructura del monómero, las cadenas β están representadas en amarillo y los bucles en verde. Las cadenas β han sido nombradas de acuerdo con Xia *et al.* (1994). **(c)** Vista desde arriba del trímero. Los monómeros han sido representados en rojo, verde y amarillo.

carbonos α de tan sólo 0.2-0.3 Å. Las principales diferencias entre los monómeros se localizan en los residuos 641-642 y 771. Estos residuos, que son parte de los bucles CD e IJ, están situados en la parte superior del trímero y están involucrados en los contactos cristalinos entre los monómeros A y C, pero no en el caso del monómero B, lo que explica las diferencias de conformación. Los contactos que se producen entre los monómeros están principalmente localizados en las regiones N-terminales de las

cadenas β A y C, así como las regiones C-terminales de las cadenas β B y J y los bucles entre las cadenas A y A' (particularmente los residuos Asn593 y Leu594), así como una pequeña región del bucle DG (Pro689-Thr695).

Tabla 5.I - Estadísticas cristalográficas

	Cristal nativo	Cristales derivados	
		Baja concentración	Alta concentración
Fuente de radiación	BM16 (ESRF)	Cu-K α (Bruker-Nonius FR591)	Cu-K α (Bruker-Nonius FR591)
Temperatura	100 K	100 K	100 K
Longitud de onda (Å)	0.9799	1.5418	1.5418
Detector	165 mm MAR-CCD	KappaCCD-2000	KappaCCD-2000
Grupo espacial	C 1 2 1	C 1 2 1	C 1 2 1
Parámetros de celdilla a, b, c (Å)	216.8, 58.7, 57.6	216.5, 59.2, 57.5	216.8, 59.1, 57.6
Parámetros de celdilla α , β , γ (°)	90.0, 101.5, 90.0	90.0, 101.3, 90.0	90.0, 101.3, 90.0
Reflexiones observadas ^a	88294 (6836)	35555 (4586)	36315 (3624)
Rango de resolución (Å)	25.0-1.6 (1.66-1.60)	43.0-2.2 (2.3-2.2)	43.0-2.2 (2.28-2.19)
Multiplicidad	3.1 (2.5)	12.2 (5.2)	4.5 (2.3)
Compleitud (%)	93.8 (73.0)	99.6 (96.8)	97.9 (85.7)
$\langle I/\sigma(I) \rangle$	20.9 (2.3)	27.0 (8.2)	10.4 (2.8)
R_{sym} (%) ^b	4.9 (37.8)	7.3 (20.0)	18.1 (42.3)
Estadísticas de phasing			
Rango de resolución (Å)		43.0-2.3 (2.39-2.30)	
Número de reflexiones usadas		31459 (2900)	
Número de átomos pesados		8	
Phasing power (reflexiones céntricas/acéntricas)		1.08 / 0.90	
R-cullis (reflexiones céntricas/acéntricas)		0.82 / 0.71	
Estadísticas de refinamiento^c			
Rango de resolución	16.0-1.6 (1.69-1.60)		
Moléculas de proteína / ASU	3		
Número de reflexiones usadas	85836 (9941)		
Número de reflexiones de Rfree	2034 (230)		
R-factor ^d	18.0 (31.0)		
R-free ^e	21.3 (31.0)		
Modelo final			
Número de átomos proteína	4797		
agua	857		
Factores térmicos promedio (Å²)			
proteína	23.44		
aguas	39.74		
Geometría (desviaciones cuadráticas medias)			
Longitud de enlace (Å)	0.015		
Ángulos de enlace (°)	1.618		
Estadísticas de Ramachandran ^f	99.6 / 100.0		
Código en el PDB	2IUM		

^aNo se descartó ninguna reflexión.

^b $R_{\text{sym}} = \sum_b \sum_i |I_{h_i} - \langle I_{h_i} \rangle| / \sum_b \sum_i I_{h_i}$

^cResultados de REFMAC (Murshudov *et al.*, 1997).

^dR-factor = $\sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|| / \sum |F_{\text{obs}}(hkl)|$.

^eDe acuerdo con Brunger (1997).

^fDe acuerdo con el programa RAMPAGE (Lovell *et al.*, 2003). Los porcentajes indican residuos en las regiones favorables, permitidas y desfavorable, respectivamente.

5.4 - Estabilidad y plegamiento del dominio de cabeza

La superficie de cada monómero tiene un área de unos $9.6 \times 10^3 \text{ \AA}^2$, de la cual un 23% forma parte de la superficie de interacción entre monómeros ($2.2 \times 10^3 \text{ \AA}^2$). Cuando se comparan las energías de solvatación teórica del trímero con las del monómero, se prevé que en la formación de trímero se produzca una ganancia de energía de solvatación de unos $150\text{-}200 \text{ kJ mol}^{-1}$ (Henrick & Thornton, 1998). La gran superficie de interacción entre monómeros (Jones y Thornton, 1996) así como esta gran ganancia de energía de solvatación que se produce en la formación del trímero, explican al menos en parte la extraordinaria estabilidad del dominio de cabeza de esta fibra, como queda de manifiesto por el hecho de que ésta no se disocie en sus monómeros en un gel de poliacrilamida en condiciones desnaturalizantes con SDS, a menos que previamente se haya hervido en un tampón en presencia de SDS (Guardado-Calvo *et al.*, 2006). La superficie de interacción es de carácter mixto, puesto que en ella intervienen tanto contactos polares como hidrofóbicos. En las superficies de trimerización de los dominios de cabeza de adenovirus humanos (van Raaij *et al.*, 1999a), $\sigma 1$ (Chappell *et al.*, 2002), σC (Capítulo 3; Guardado-Calvo *et al.*, 2005) y la proteína p5 de PRD1 (Merckel *et al.*, 2005), también se observa una superficie de interacción de naturaleza mixta. Por otra parte, el área de estas superficies en los dominios de cabeza de la fibras previamente citadas es aproximadamente la misma (en los adenovirus humanos esta área es de unos $2 \times 10^3 \text{ \AA}^2$ o el 20% de la superficie disponible).

5.5 - Comparación con otras proteínas de estructura conocida

La superposición de los residuos estructuralmente equivalentes de los monómeros del dominio de cabeza de la fibra larga del adenovirus aviar con los monómeros de los dominios de cabeza de distintos adenovirus humanos nos ha permitido observar (Figura 5.3A) que no sólo todos los dominios de cabeza de todos estos adenovirus tienen la misma topología, sino que los bucles que conectan las cadenas β son muy similares, siendo todos cortos a excepción del bucle DG. La discontinuidad en la cadena βA también está conservada en el adenovirus humano de tipo 2. Cuando se superpusieron los trímeros del dominio de cabeza de la fibra larga del adenovirus aviar y el del adenovirus humano de tipo 2 (3×153 carbonos α

estructuralmente equivalentes, Figura 5.3B), se pudo observar que la orientación relativa de los monómeros en ambos trímeros resultó ser muy similar.

En comparación con el dominio de cabeza del adenovirus humano de tipo 2, el dominio de cabeza de la fibra aviar contiene 16 residuos extra, que se encuentran localizados en el bucle AB, la cadena β G, y los bucles GH e IJ (Figura 5.3C). Además, los bucles CD y HI situados en la parte superior del trímero se orientan de manera diferente, y el bucle DG tiene una conformación distinta. Estas variaciones dan lugar a lo que probablemente sean diferencias funcionalmente significativas en la superficie externa del trímero, mientras que su estructura interna mantiene un alto grado de similitud. Otra diferencia significativa se encuentra en la conformación del bucle DG, que en la cabeza del adenovirus humano de tipo 2 tiene más residuos en la superficie de interacción monómero-monómero (a la izquierda en la Figura 5.3A y marcado con una 'I' en la Figura 5.3B), mientras que en el dominio de cabeza de la fibra del adenovirus aviar este bucle interacciona más con la lámina β GHIDA' (a la derecha en la Figura 5.3A e indicado con una 'S' en la Figura 5.3B). Por último, un análisis de las superficies electrostáticas (Figura 5.4) de los dominios de cabeza de los adenovirus humanos 5 y 12 nos muestra que estas guardan más parecido entre sí que con el dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1, una diferencia que también puede tener connotaciones funcionales.

Una búsqueda de estructuras similares al dominio de cabeza de la fibra larga del adenovirus aviar muestra que esta guarda similitud estructural con fibras de reovirus (Chappell *et al.*, 2002; Guardado-Calvo *et al.*, 2005) y fibras del fago *Lactobacillus* (Spinelli *et al.*, 2006a,b). Del mismo modo que los dominios de cabeza de adenovirus, los dominios de cabeza de estas fibras están compuestos principalmente de láminas β , aunque a diferencia de los adenovirus estas forman barriles β circulares en vez de sándwiches β . Los dominios de cabeza de las fibras de reovirus tienen la misma topología que el dominio de cabeza del adenovirus aviar.

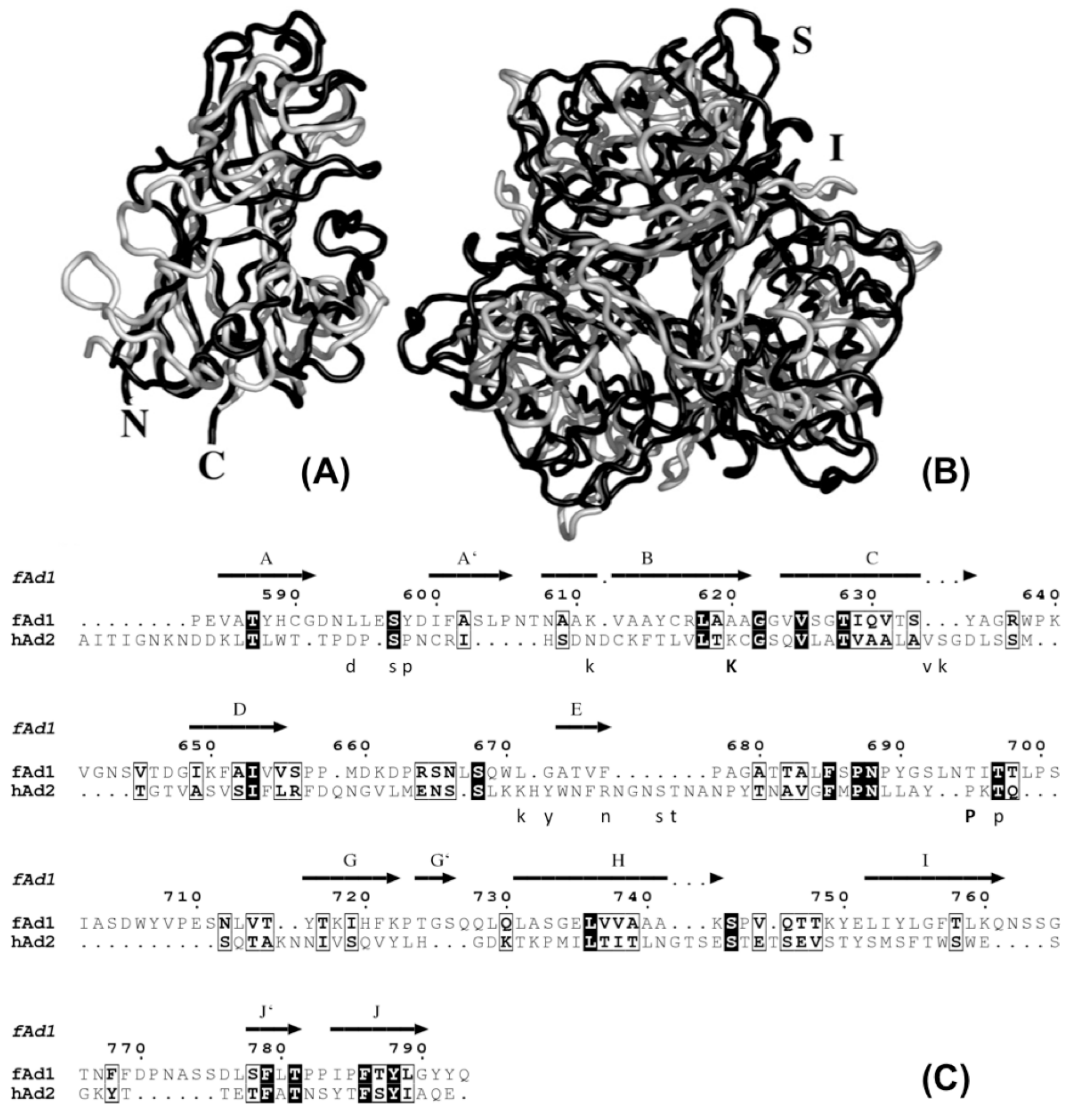


Figura 5.3. Comparación de los dominios de cabeza de las fibras del adenovirus aviar de tipo 1 y el adenovirus humano de tipo 2. **(A)** Monómeros superpuestos. Se muestran las cadenas principales de los dominios de cabeza de las fibras del adenovirus aviar de tipo 1 (negro) y el adenovirus humano de tipo 2 (gris). Se indican los extremos N y C-terminales del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1. **(B)** Comparación de la estructura de los trímeros de ambos adenovirus, los colores se corresponden con los indicados en (A). Se indica la superficie de interacción entre dos monómeros (I) así como la posición de la lámina externa (S). **(C)** Alineamiento basado en la estructura de los dominios de cabeza de la fibra larga del adenovirus aviar de tipo 1 (fAd1) y el adenovirus humano de tipo 2 (hAd2). Se indican los residuos idénticos con una cuadro negro y las cadenas β del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1. En la línea inferior, se destacan los residuos implicados en la unión a CAR, los absolutamente conservados en mayúsculas y los mayoritariamente conservados en minúsculas (de acuerdo con Law & Davidson, 2005)

5.6 - La capacidad de unión al receptor

El adenovirus aviar de tipo 1 es capaz de infectar diferencialmente células CHO (Chinese Hamster Ovary) deficientes en CAR y células CHO expresando CAR humano, lo que sugiere que este adenovirus tiene la capacidad de unirse a CAR humano y, por extensión también al aviar. Estudios basados en deleciones de la fibra larga del adenovirus aviar han relacionado a esta fibra larga con la unión a CAR, mientras que el receptor para la fibra corta todavía no ha sido identificado (Tan *et al.*, 2001). En el transcurso de nuestro estudio, tratamos de medir la interacción del dominio de cabeza de la fibra larga del adenovirus aviar con el dominio 1 de CAR; dominio que se ha demostrado necesario y suficiente para la unión de gran afinidad de los dominios de cabeza de algunos adenovirus humanos (Freimuth *et al.*, 1999; Lortat-Jacob *et al.*, 2001).

Para determinar si existía interacción entre el dominio de cabeza de la fibra larga del adenovirus aviar con el dominio 1 de CAR se realizaron experimentos de exclusión molecular y de SPR (surface plasmon resonance). Los experimentos de exclusión molecular muestran que tanto el dominio de cabeza de la fibra larga del adenovirus aviar como el dominio 1 de CAR migran de manera independiente en una columna de exclusión molecular, lo que indica que no interaccionan en solución. Para confirmar este resultado se realizó un experimento de SPR, para ello se colaboró con Hugues Lortat-Jacob (IBS, Grenoble). El dominio D1 de CAR se inmovilizó en un chip y el dominio de cabeza de la fibra larga del adenovirus 1 se usó como analito. No fue posible detectar ninguna interacción, lo que estaba en concordancia con la falta de interacción del experimento de exclusión molecular. Esta falta de interacción sugiere que, o bien que necesite del dominio 2 para establecer una interacción eficiente a CAR, o bien que el dominio fibroso de la fibra larga contenga elementos necesarios para la interacción con CAR. La otra alternativa es que CAR no sea el receptor del adenovirus aviar. Cabe señalar que estos experimentos se llevaron a cabo con el dominio 1 de CAR humano, no de pollo (NCBI RefSeq n° XW-416681, Pruitt *et al.*, 2005), debido al interés en el uso del adenovirus aviar como una terapia génica en humanos o como vector antitumoral. Sin embargo, cabe resaltar que el dominio homólogo en pollos es un 61% idéntico al dominio 1 de CAR humano, y que los residuos de CAR humano que se sabe interaccionan con el dominio de cabeza de adenovirus humanos (Bewley *et al.*, 1999, Law & Davidson, 2005) están totalmente

conservados (Glu56, Leu73, Ser75, Tyr80, Tyr83, Lys121 y Lys123) o son similares (Asp68/Glu y Val70/Thr) en su homólogo de pollo.

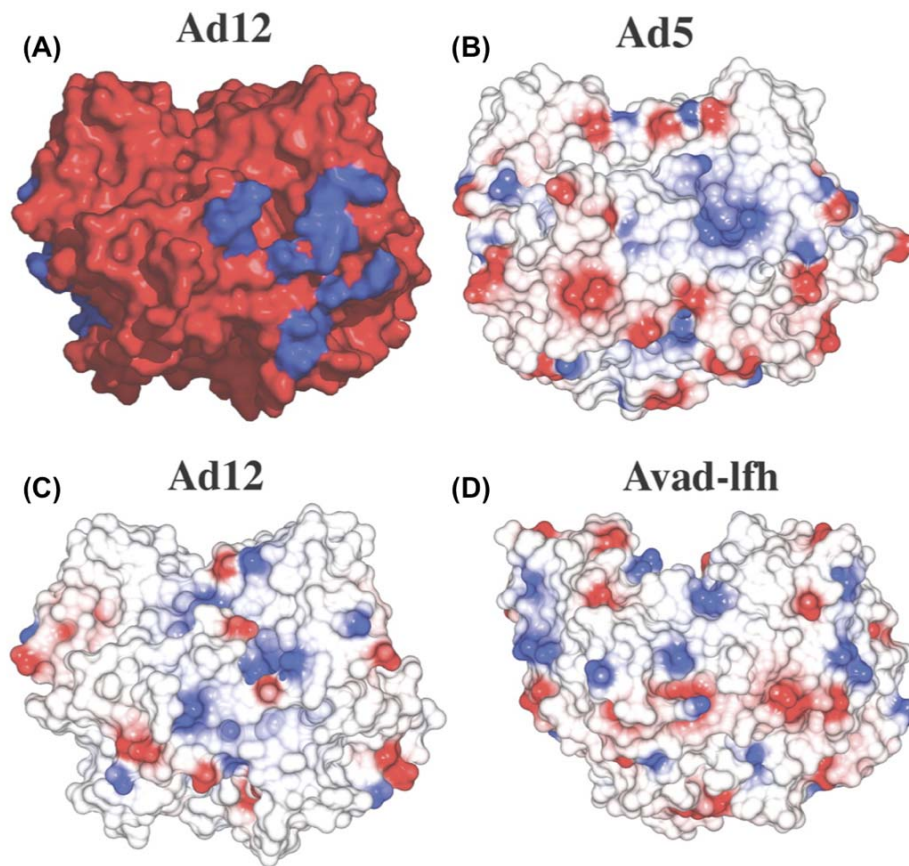


Figura 5.4. Propiedades de la superficie de los dominios de cabeza del adenovirus aviar de tipo 1 y de otros adenovirus humanos. Todos los paneles muestran vistas laterales con la misma orientación. **(A)** Representación de la superficie del dominio de cabeza del adenovirus humano de tipo 12, los superficie implicada en la unión a CAR se muestran en azul (Bewley *et al.*, 1999). **(B-D)** Superficie de potenciales electrostáticos de los adenovirus humanos de tipo 5 **(B)**, de tipo 12 **(C)** y el dominio de cabeza del adenovirus aviar de tipo 1 **(D)**.

La ubicación del sitio de unión a CAR se ha determinado a partir de estudios de mutagénesis dirigida de la fibra del adenovirus humano de tipo 5 (Kirby *et al.*, 1999; Roelvink *et al.*, 1999) y de los datos estructurales de los complejos con el adenovirus humano de tipo 12 (Bewley *et al.*, 1999) y tipo 37 (Seiradake *et al.*, 2006). Este sitio de interacción se encuentra ubicado a un lado de la molécula y en él participan residuos de los bucles AB, CD y DG (Law & Davidson, 2005). Cuando se realiza un alineamiento estructural entre el dominio de cabeza de la fibra larga del adenovirus aviar y los dominios de cabeza de aquellos adenovirus humanos que unen CAR, se aprecia que pocos de los residuos implicados en la interacción con CAR

están conservados en la fibra del adenovirus aviar. Además, los dos residuos que están absolutamente conservados entre los adenovirus que unen CAR (una lisina en la cadena β B y una prolina en el bucle DG) no están conservados en la fibra larga del adenovirus aviar. Aún más, la superficie equivalente es significativamente diferente tanto en su forma como en su carga (Figura 5.4 y 5.5). Cabe señalar que el dominio de cabeza del adenovirus canino de tipo 2 se une a CAR, tanto en experimentos de SPR como cristalográficos (Seiradake *et al.*, 2006), aunque no mantenga los residuos absolutamente conservados. Aún así, la estructura del bucle DG en este adenovirus canino es mucho más similar a la de los adenovirus humanos que unen CAR de lo que lo es la del adenovirus aviar.

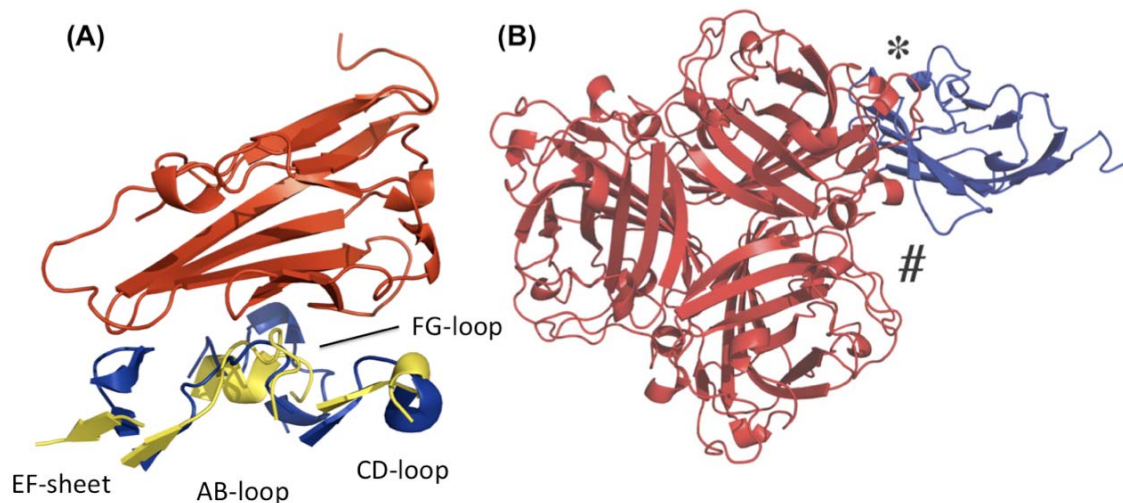


Figura 5.5. Superposición del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1 sobre el dominio de cabeza del adenovirus humano de tipo 12 en complejo con el dominio 1 de CAR (Bewley *et al.*, 1999). **(A)** Comparación estructural entre los dominios de cabeza de la fibra larga del adenovirus aviar 1 (amarillo) y el del adenovirus humano 12 (azul) en complejo con el dominio D1 de CAR (rojo) (Bewley *et al.*, 1999). Se han señalado los bucles AB-, CD-, y FG- así como la pequeña lámina EF. **(b)** Vista superior del trímero del dominio de cabeza de la fibra larga del adenovirus aviar 1. Sólo se muestra uno de los tres dominios 1 de CAR que pueden unirse simultáneamente a la fibra del adenovirus humano de tipo 12. Los asteriscos marcan la región en donde la conformación del bucle DG del dominio de cabeza del adenovirus aviar resulta incompatible con la unión a CAR, mientras que la almohadilla indica el lugar en donde el dominio de cabeza del adenovirus humano de tipo 12 contacta CAR mientras que la fibra del adenovirus aviar no lo hace.

La superposición del dominio de cabeza de la fibra larga del adenovirus aviar sobre el dominio de cabeza del adenovirus humano de tipo 12 en complejo con el dominio 1 de CAR (Figura 5.5) muestra como la diferente orientación del bucle DG en el adenovirus aviar resulta incompatible con la unión al dominio 1 de CAR de la misma manera en que lo hace el adenovirus humano de tipo 12 (o cualquier otro adenovirus humano cuya estructura en complejo con CAR se haya descrito), lo que sugiere que o bien el receptor CAR se une de una manera distinta a la fibra larga del adenovirus aviar, como se ha discutido previamente, o que CAR no es el receptor real de esta fibra. Serían necesarios experimentos de mutagénesis dirigida, *cross-linking* o estudios de cocrystalización con CAR (aviar) o un receptor alternativo para determinar con exactitud los residuos de la fibra larga de este adenovirus aviar que están implicados en la interacción con el receptor.

5.7 - Conclusión

En este trabajo hemos resuelto la estructura del dominio de unión al receptor de la fibra larga del adenovirus aviar de tipo 1. La estructura proporciona información a nivel molecular de las interacciones que estabilizan el dominio de cabeza y revela que hay importantes diferencias a nivel estructural, y probablemente funcional, en lo que se ha demostrado ser el sitio de unión al receptor en fibras de adenovirus humanos previamente descritas. Los datos estructurales presentados aquí podrán ser usados en estudios destinados a comprender las primeras etapas de la replicación de estos adenovirus aviares y pueden servir para generar aplicaciones terapéuticas destinadas a enfermedades aviares y/o humanas, facilitando el diseño de fibras quiméricas artificiales basadas en esta estructura, que podrían ser usadas como vacunas o vectores de terapia génica. Esta información también podría ser usada en el desarrollo de péptidos antivirales que podrían bloquear la unión al receptor por inhibición competitiva.

Capítulo 6

LA FIBRA DEL ADENOVIRUS PORCINO 4

Capítulo 6

Estructura de los dominios de cabeza y de galectina del adenovirus porcino NADC-1

Los adenovirus porcinos están considerados como unos virus poco patogénicos, lo que los convierte en candidatos potenciales a ser vectores útiles en terapia génica. Se cree que la cepa NADC-1, que fue aislada por primera vez de las amígdalas de una cerda adulta, es una cepa de un adenovirus porcino de tipo 4 (Kleiboeker *et al.*, 1993). La fibra de esta cepa de adenovirus tiene 703 residuos y una arquitectura inusual entre los adenovirus ya que, además de los dominios clásicos de la fibras adenovirales, contiene dos dominios extra situados en el extremo C-terminal. El primero de estos contiene un motivo RGD, por lo que potencialmente podría ser un dominio de interacción con integrinas. El segundo de estos dominios está situado en el extremo más C-terminal de la fibra y está compuesto por dos dominios de unión a carbohidratos (CRD) repetidos en tándem. Estos posibles dominios CRD son similares a aquellos hallados en la familia de las galectinas (Cooper & Barondes, 1999; Cooper, 2002), por lo que potencialmente podría interactuar con β -galactósidos.

Existen tres familias de galectinas basándose en la organización de sus dominios (Hirabayashi *et al.*, 1993). Aquellas que contienen un único dominio CRD (galectinas prototipo), aquellas compuestas por un único dominio CRD y un largo dominio N-terminal (galectinas quiméricas) y aquellas compuestas por dos dominios CRD distintos (galectinas repetidas en tándem). Hasta la fecha no se ha resuelto la estructura de ninguna galectina repetida en tándem, por lo que se desconoce la organización de sus dominios.

6.1 - Purificación y cristalización de los dominios de la fibra del adenovirus porcino NADC-1

6.1.1 - Construcción de los vectores de expresión

Los fragmentos de ADN que codificaban para los distintos dominios fueron producidos por PCR a partir del plásmido pKan-NMOvPAD4. Se clonaron los fragmentos 40-703, 116-703, 40-291, 40-317, 40-392, 116-291, 291-392, 317-703,

371-703, 393-703 (Figura 6.1) en un vector de expresión como se describe en Materiales y Métodos. Todos ellos resultaron solubles aunque los fragmentos 40-703 y 116-703 no se expresaban en cantidades suficientes como para poder realizar con ellos experimentos de cristalización. Finalmente sólo se lograron cristalizar los dominios 40-291, 116-291 y 393-703 y se resolvieron las estructuras del dominio de cabeza (116-291) y galectina (393-703). Aunque se logró resolver la estructura del dominio 40-291 no se apreciaba densidad electrónica para el dominio fibroso, probablemente porque es flexible.

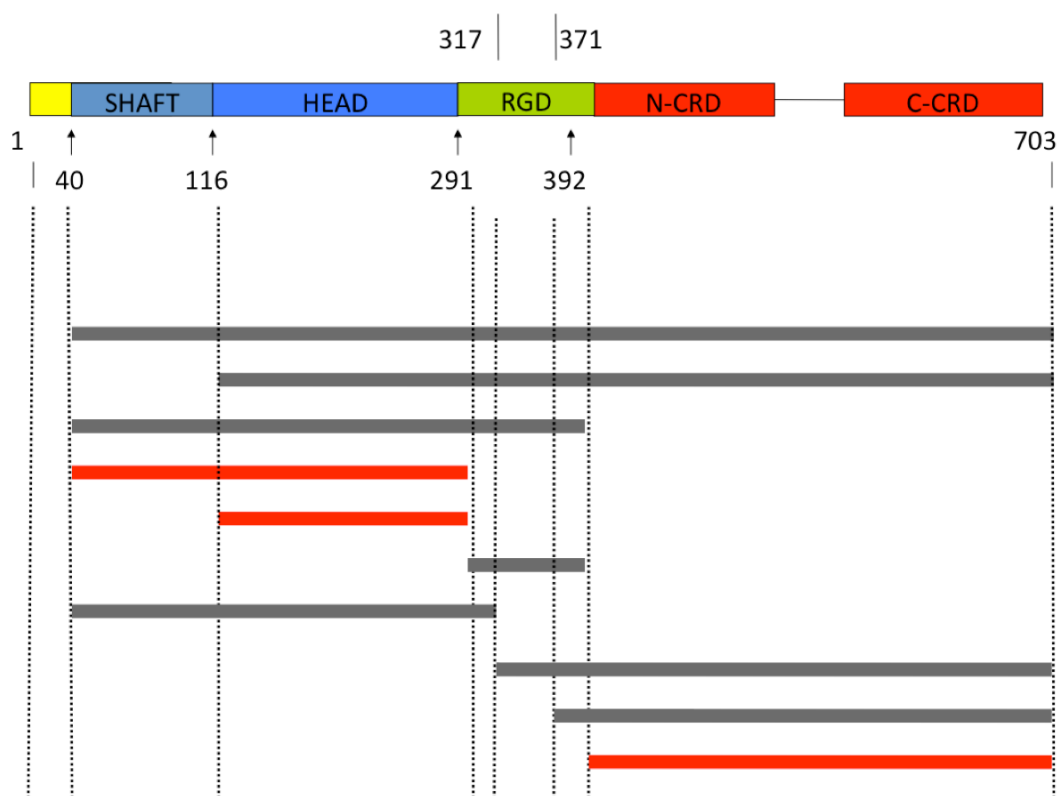


Figura 6.1 Diagrama mostrando los distintos dominios de la fibra del adenovirus porcino de tipo 4 en donde se muestran los dominios que han sido clonados y purificados. En rojo aparecen señalados aquellos dominios que se han logrado cristalizar.

6.1.2 - Expresión y purificación del dominio de cabeza (116-291) y el dominio fibroso y cabeza (40-291)

Para expresar el dominio de cabeza (residuos 116-291) y los dominios de fibroso y cabeza (40-291) los plásmidos pET28c-Fib116-291 y pET28c-Fib40-291 se usaron para transformar bacterias, y se expresaron como se describe en Materiales y Métodos. Tras la recogida de bacterias por centrifugación y lisis, ambas proteínas fueron purificadas de la fracción soluble por cromatografía de afinidad por níquel.

Ambas eluían a altas concentraciones de imidazol (200-800 mM), lo que es consistente con una proteína trimérica con tres colas de histidinas. La purificación se completó mediante cromatografía de intercambio catiónico en ambos casos.

6.1.3 - Expresión y purificación del dominio de galectina (393-703)

Para expresar el dominio de galectina (residuos 393-703) el plásmido pET28c-Fib393-703 se usó para transformar bacterias y se expresó como se describe en Materiales y Métodos. Tras la recogida de bacterias por centrifugación y lisis, la proteína fue purificada de la fracción soluble por cromatografía de afinidad por níquel, intercambio catiónico y exclusión molecular.

6.1.4 - Cristalización del dominio de cabeza

La proteína pAd4Fib(116-291) concentrada se cristalizó utilizando sulfato de litio y polietilenglicol 8000 como precipitante. Aunque se intentaron congelar estos cristales usando diversas soluciones crioprotectoras, finalmente resultó imposible. Sin embargo, fue posible recoger un set de datos completos a temperatura ambiente utilizando un ánodo rotante, hasta un resolución de 3.2 Å. Las estadísticas de los datos recogidos de este cristal aparecen resumidas en la Tabla 6.I

6.1.5 - Cristalización del dominio de fibroso y cabeza

La proteína pAd4Fib(40-291) concentrada se cristalizó utilizando 15% PEG 4000 (w/v), 0.1M citrato sódico, 0.1M sulfato amónico. Los cristales se transfirieron a una solución con la misma composición que la solución de cristalización más un 20% (v/v) de glicerol. Los cristales difractaron hasta 3.1 Å de resolución en la línea X11 del DESY. La estructura se resolvió mediante reemplazamiento molecular utilizando la estructura del dominio de cabeza como modelo. Desafortunadamente, el dominio fibroso no era visible en el mapa de densidad electrónica.

6.1.6 - Cristalización del dominio de galectina y de los complejos galectina-azúcar

La proteína pAd4Fib(393-703) concentrada se cristalizó utilizando 25% (w/v) PEG 3350, 500 mM nitrato de sodio, 2 mM ácido siálico. Los cristales fueron trasferidos a una solución crioprotectora y se midieron en la línea X11 (DESY, Hamburgo) en donde difractaron hasta 1.9 Å de resolución. Los cristales del complejo con lactosa fueron obtenidos usando 28% (w/v) PEG 3350, 100 mM nitrato de sodio y

5 mM DTT como solución de cristalización, añadiendo 70 mM de lactosa en la solución de proteína. Los cristales fueron crioprottegidos en 35% (w/v) PEG 3350, 100 mM nitrato de sodio, 1 mM DTT, 70 mM lactosa. Los cristales del complejo con *N*-acetillactosamina se obtuvieron usando 35% (w/v) PEG 3350, 500 mM nitrato de sodio y 5 mM DTT como solución de cristalización, añadiendo 40 mM de *N*-acetillactosamina en la solución de proteína. Los cristales fueron congelados directamente. Los cristales del complejo con lacto-*N*-neotetraosa se obtuvieron usando 27% (w/v) PEG 3350, 200 mM nitrato de litio y 5 mM DTT como solución de cristalización, añadiendo 5 mM de lacto-*N*-neotetraosa en la solución de proteína. Los cristales fueron congelados directamente. Los cristales del complejo con tri-(*N*-acetillactosamina) se obtuvieron usando 28% (w/v) PEG 3350, 300 mM nitrato de litio y 5 mM DTT como solución de cristalización, añadiendo 5 mM de tri-(*N*-acetillactosamina) en la solución de proteína. Los cristales fueron crioprottegidos en 35% (w/v) PEG 3350, 300 mM nitrato de litio, 1 mM DTT, 5 mM tri-(*N*-acetillactosamina). Las estadísticas de los datos recogidos de estos cristales aparecen resumidas en la Tabla 6.I.

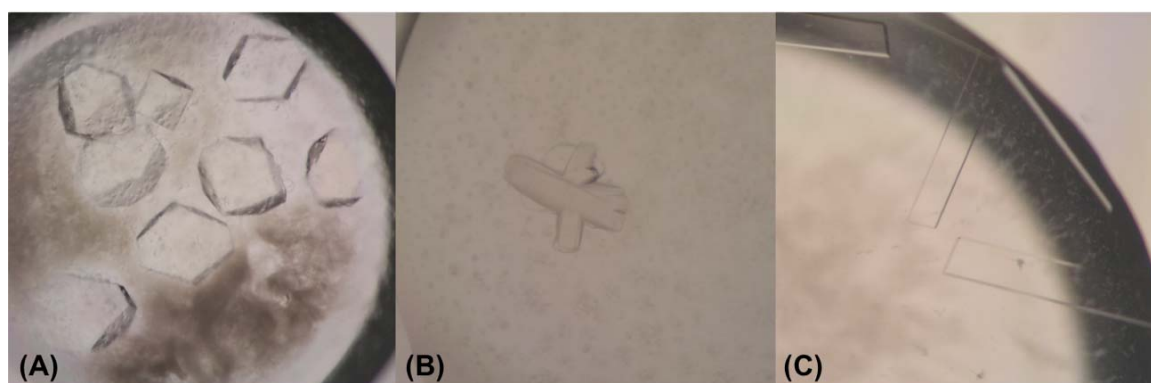


Figura 6.2 Cristales de el dominio de cabeza (A), el dominio fibroso y cabeza (B) y del dominio de galectina (C).

6.2 - Determinación estructural de los dominios de cabeza y de galectina

6.2.1 - Determinación estructural del dominio de cabeza

La estructura del dominio de cabeza se resolvió mediante reemplazamiento molecular usando como modelo la estructura de la cabeza del adenovirus humano de tipo 5 (1KNB). La solución mostró la existencia de dos trímeros en la unidad asimétrica. Durante el refinamiento se utilizó la simetría no cristalográfica presente. A pesar de que la resolución de los datos era moderada (3.2 Å), el alto contenido de

solvente así como la existencia de simetría no cristalográfica, contribuyeron a que los mapas de densidad electrónica generados fuesen de buena calidad. En la parte final del refinamiento se añadieron seis grupos de TLS (translation/libration/screw), uno por cada cadena presente en la unidad asimétrica. El modelo final consistía en los residuos 116-291 de cada una de las seis cadenas, más cuatro aguas adicionales. No se observó densidad electrónica adicional para la cola de purificación N-terminal. Las estructuras refinadas se correspondían bien con los datos, tenían una geometría aceptable y sólo un residuo en una región desfavorable del diagrama de Ramachandran. Las estadísticas de refinamiento de esta estructura aparecen resumidas en la Tabla 6.I

6.2.2 - Determinación estructural del dominio de galectina

Para resolver las estructuras del dominio de galectina en complejo con los azúcares se utilizó la cadena A de la estructura previamente resuelta como modelo de reemplazamiento molecular. El modelo en complejo con lactosa consiste en una única cadena polipeptídica en donde se han podido modelar los residuos 386-684, 184 moléculas de agua, 2 iones nitrato y 2 moléculas de lactosa. Dos residuos han sido modelados con conformaciones alternativas. La molécula de lactosa unida al dominio CRD N-terminal tiene una densidad clara, mientras que la lactosa unida al dominio C-terminal tiene densidad para el anillo de galactosa y algunas nubes de densidad electrónica que se corresponden con el anillo de glucosa, esta molécula ha sido refinada finalmente con una ocupancia 0.5. El modelo en complejo con *N*-acetillactosamina consiste en los residuos 386-685, 163 moléculas de agua, 4 iones nitrato y 1 molécula de *N*-acetillactosamina unida al dominio N-terminal. Un residuo ha sido modelado con conformación alternativa. La Pro540 es el único residuo situado en una región desfavorable del diagrama de Ramachandran. El modelo en complejo con lacto-*N*-neotetraosa consiste en los residuos 386-684, 145 moléculas de agua, 3 iones nitrato y 1 molécula de lacto-*N*-neotetraosa. Finalmente, el modelo en complejo con tri-(*N*-acetillactosamina) consiste en dos cadenas con los residuos 386-685, 129 moléculas de agua, 6 iones nitrato y 1 molécula de tri-(*N*-acetillactosamina). Las estadísticas de refinamiento de estas estructuras aparecen resumidas en la Tabla 6.I

Tabla 6.I. Estadísticas de los datos, el refinamiento de la estructura y su validación.

	Dominio de cabeza			Dominio de galectina		
		Libre	Lactosa	N-acetilactosamina	Lacto-Nneotetraosa	tri(N-acetilactosamine)
Fuente de radiación	Cu-K α (Bruker-Nonius FR591)	ID14-2 (ESRF)	BM30A (ESRF)	EMBL X11 (DESY)	BM30A (ESRF)	EMBL X11 (DESY)
Temperatura	298 K	100 K	100 K	100 K	100 K	100 K
Longitud de onda (Å)	1.5418	0.9390	0.9797	0.8150	0.9797	0.8150
Detector	KappaCCD-2000	ADSC Q4 CCD	ADSC Q315r CCD	MAR555 flatpanel	ADSC Q315r CCD	MAR555 flatpanel
Distancia detector-cristal (mm)	90.0	301.1	291.9	300	292.3	300.0
Rango de oscilación (°)	0.6	0.8	1.0	1.0	1.0	1.0
Número de imágenes	166	272	176	136	180	180
Grupo espacial	P2 ₁ 2 ₁ 2 ₁	C2	C2	C2	C2	C2
Parámetros de celdilla a, b, c (Å)	125.7, 145.4, 147.6	167.6, 77.3, 94.1	114.6, 38.8, 70.2	108.5, 43.7, 63.9	95.4, 38.8, 84.4	176.1, 38.3, 86.5
Parámetros de celdilla α, β, γ (°)	90.0, 90.0, 90.0	90.0, 101.5, 90.0	90.0, 93.5, 90.0	90.0, 105.8, 90.0	90.0, 102.2, 90.0	90.0, 92.1, 90.0
Mosaicidad (°)	0.36	0.68	0.80	0.90	0.68	0.87
Observaciones únicas^a	45035 (4367)	89027 (12129)	21094 (3019)	22025 (3229)	23751 (3155)	20441 (2926)
Rango de resolución (Å)	40.0-3.2 (3.32-3.2)	40.0-1.9 (2.0-1.9)	40.0-2.0 (2.1-2.0)	35.0-1.9 (2.0-1.9)	40.0-1.9 (2.0-1.9)	40.0-2.5 (2.6-2.5)
Multiplicidad	3.8 (3.8)	4.1 (3.8)	3.5 (3.5)	2.7 (2.7)	3.4 (2.7)	3.6 (3.6)
Complejidad (%)	98.9 (97.5)	96.1 (90.1)	99.6 (99.3)	98.0 (98.5)	98.5 (91.7)	99.9 (99.8)
<I/σ(I)>	17.5 (2.9)	10.0 (3.7)	10.4 (3.3)	5.9 (2.7)	12.6 (3.2)	8.8 (3.1)
R_{sym} (%)^b	7.2 (30.7)	7.5 (35.2)	6.9 (32.2)	9.8 (31.3)	5.8 (30.3)	9.5 (43.2)
Estadísticas de refinamiento^c						
Rango de resoluciónⁿ	29.2-3.2 (3.37-3.20)	35-1.9 (2.00-1.90)	40-2.0 (2.11-2.00)	35-1.91 (2.01-1.91)	40-1.9 (2.00-1.90)	40-2.5 (2.63-2.50)
Moléculas de proteína / ASU	6	4	1	1	1	2
Numero de reflexiones usadas^a	40566 (5769)	86627 (11703)	20006 (2873)	20895 (3071)	22538 (2995)	19313 (2789)
Numero de reflexiones de Rfree	2026 (218)	2124 (293)	1087 (140)	1126 (148)	1213 (153)	1123 (129)
R-factor^d	0.173 (0.264)	0.206 (0.298)	0.183 (0.244)	0.182 (0.270)	0.187 (0.275)	0.199 (0.324)
R-free^e	0.201 (0.299)	0.259 (0.327)	0.228 (0.296)	0.238 (0.337)	0.223 (0.310)	0.268 (0.355)
Modelo final						
Número de átomos						
proteína	7980	9556	2368	2360	2347	4708
ligandos/iones	0	84	54	42	60	176
agua	4	625	184	163	145	129
Factores térmicos promedio (Å²)						
proteína	65.4	38.5	33.5	33.1	40.4	31.9
nitrato /glicerol		42.1	28.3	31.2	28.5	27.8
Azúcares			32.0	22.0	34.1	33.5
aguas	58.1	44.3	38.3	37.5	41.8	28.7
Geometría. Desviaciones cuadráticas medias.						
Longitud de enlace (Å)^f	0.009	0.010	0.009	0.012	0.011	0.009
Angulos de enlace (°)^f	1.3	1.3	1.3	1.3	1.4	1.3
Estadísticas de Ramachandran^f	93.9 / 6.0 / 0.1	97.0 / 2.4 / 0.6	96.6 / 3.4 / 0.0	97.7 / 2.0 / 0.3	98.0 / 2.0 / 0.0	97.5 / 2.5 / 0.0
Código en el PDB	2WST	2WSU	2WSV	2WT0	2WT1	2WT2

Se usó un cristal para resolver cada estructura; se valores entre paréntesis se refieren a la capa de mayor resolución.

^aNo se descartó ninguna reflexión.

$$^b R_{\text{sym}} = \frac{\sum_h \sum_k |I_{hk} - \langle I_h \rangle|}{\sum_h \sum_k I_{hk}}$$

^cResultados de REFMAC (Murshudov *et al.*, 1997).

$$^d R\text{-factor} = \frac{\sum |F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|}{\sum |F_{\text{obs}}(hkl)|}$$

^eDe acuerdo con Brunger (1997).

^fDe acuerdo con el programa RAMPAGE (Lovell *et al.*, 2003). Los porcentajes indican residuos en las regiones favorables, permitidas y desfavorable, respectivamente.

6.3 - La estructura del dominio de cabeza y su capacidad de unirse a receptores celulares

El dominio de cabeza de cada monómero es una estructura globular plegada formando un sándwich β antiparalelo con una topología similar a la descrita para los dominios de cabeza de fibras adenovirales humanas (Figura 6.3). El sándwich β esta compuesto por ocho cadenas β , las cuales pueden ser divididas en dos láminas β (ABCJ y GHID usando la nomenclatura propuesta para el adenovirus humano de tipo 5; Xia *et al.*, 1994). El bucle DG (residuos 180-231) incluye las cadenas β E (residuos 197-200) y F (residuos 203-205), que interaccionan con la cadena β A extendiendo de este modo la lámina ABCJ. Además de estas dos cadenas β , el bucle DG contiene dos pequeñas hélices alfa. Una búsqueda de estructuras similares (Krissinel & Henrick, 2004) muestra que los homólogos más cercanos estructuralmente son el adenovirus humano de tipo 5 (Xia *et al.*, 1994), de tipo 2 (van Raaij *et al.*, 1999a), de tipo 12 (Bewley *et al.*, 1999) y los tipos 37 y 19p (Burmeister *et al.*, 2004). Las diferencias más notables se dan en los bucles CD e IJ, situados en la parte superior del trímero, que son notablemente más cortos en el dominio de cabeza del adenovirus porcino de tipo 4, y en el bucle HI, situado a un lado del trímero, que es más corto en el adenovirus porcino de lo que lo es en los adenovirus humanos de tipo 2 y 5.

Muchos adenovirus humanos utilizan CAR (coxsackievirus and adenovirus receptor) como receptor primario (Bergelson *et al.*, 1997; Roelvink *et al.*, 1998; Law & Davidson, 2005). Los análisis estructurales llevados a cabo con los complejos adenovirus humano de tipo 12 – CAR (Bewley *et al.*, 1999) y adenovirus humano de tipo 37 – CAR (Seiradake *et al.*, 2006) han mostrado que la superficie de interacción entre estas dos moléculas esta compuesta principalmente por los residuos Asp415, Pro417, Pro418 y Leu426 del bucle AB, Val450 y Lys451 del bucle CD, Gln487, Gln494, Ser497 y Val498 del la lámina β EF y Pro517, Pro519, Asn520 y Glu523 del bucle FG (la numeración corresponde al adenovirus humano de tipo 12; Bewley *et al.*, 1999). Estos residuos están mayoritariamente conservados en la fibras de aquellos serotipos que usan CAR como receptor primario. También se ha sugerido que las fibras que usan CAR comparten una conformación específica del bucle AB (Seiradake *et al.*, 2005). La figura 6.4A muestra la estructura de los bucles implicados en la interacción entre el dominio de cabeza del adenovirus humano de tipo 12 (azul)

y CAR (rojo). Se han superpuesto las regiones estructuralmente equivalentes del dominio de cabeza de la fibra del adenovirus porcino de tipo 4 (amarillo). El bucle AB del adenovirus porcino tiene una conformación similar al de los adenovirus humanos que interaccionan con CAR. Sin embargo, sólo uno de los residuos implicados en esta interacción, la Pro517, está conservado en la fibra del adenovirus porcino. Además, no hemos sido capaces de detectar ninguna interacción entre el dominio de cabeza del adenovirus porcino y el dominio D1 de CAR ni por cromatografía exclusión molecular ni en experimentos de SPR.

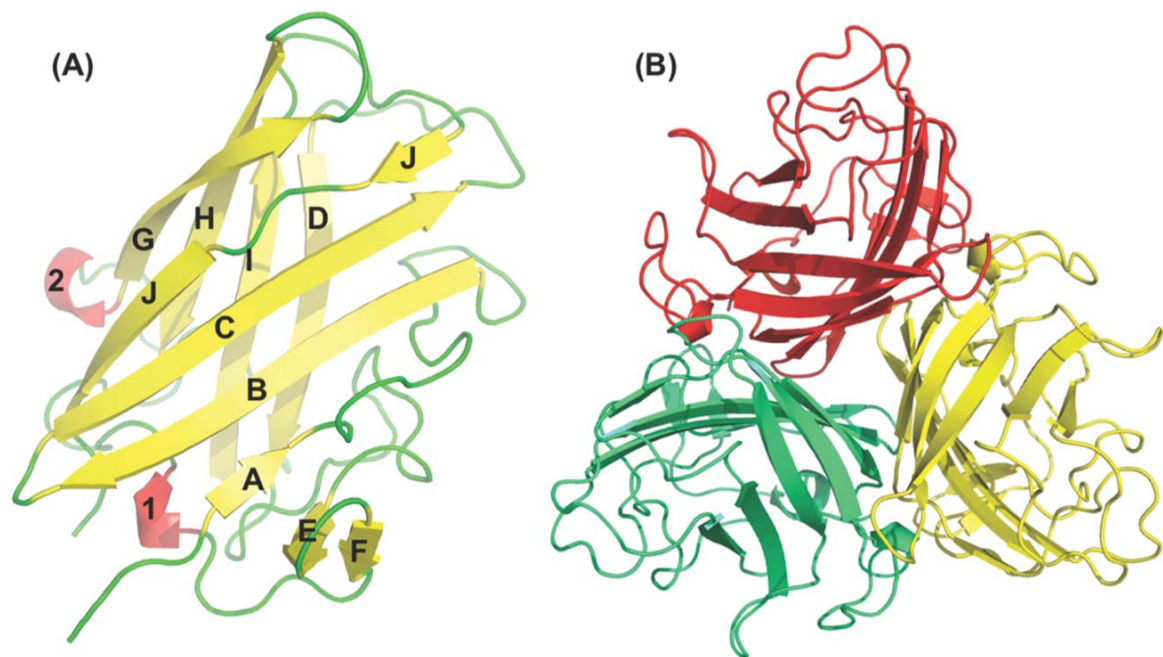


Figura 6.3 Estructura del dominio de cabeza del adenovirus porcino de tipo 4 (cepa NADC-1). **(A)** Monómero del dominio de cabeza en el que las cadenas β han sido coloreadas en amarillo, las hélices en rojo y los bucles en verde. Las cadenas β han sido nombradas de acuerdo con Xia *et al.* (1994). **(B)** El trímero del dominio de cabeza visto desde su parte superior. Los monómeros han sido coloreados en rojo, verde y amarillo.

Otro receptor primario para adenovirus humanos es CD46, mayoritariamente usado por los adenovirus de las especies B y algunos de la D (incluyendo el tipo 37) (Segerman *et al.*, 2003; Gaggar *et al.*, 2003; Sirena *et al.*, 2004; Wu *et al.*, 2004; Marttila *et al.*, 2005). La estructura del complejo formado por el dominio de cabeza del adenovirus humano de tipo 11 y dos repeticiones extracelulares de CD46 (Persson *et al.*, 2007) muestra que el dominio de cabeza utiliza tres regiones separadas para interaccionar con este receptor, los bucles DG- HI- e IJ. Una superposición de los

dominios de cabeza del adenovirus porcino de tipo 4 y el adenovirus humano de tipo 11 muestra que la conformación de los bucles DG y HI son diferentes y que el bucle IJ es más corto en el adenovirus porcino de tipo 4 (Figura 6.4B). Además, los residuos implicados en esta interacción en el adenovirus humano 11, la Asn283, Asp284, Arg280 y Gln305 no están conservados en el dominio de cabeza del adenovirus porcino 4. Experimentos llevados a cabo con un vector de adenovirus compuesto por un adenovirus humano de tipo 5 modificado con la fibra del adenovirus porcino de tipo 4, ha demostrado que es capaz de entrar en células tumorales de glioma a través de receptores independientes de CAR y CD46 (Paul *et al.*, 2008).

Finalmente, se ha demostrado que algunos glicoconjugados de la superficie celular que contienen ácido siálico unido por enlaces $\alpha(2,3)$ o $\alpha(2,6)$ funcionan como receptores celulares para los adenovirus humanos de tipo 37 y 19p (Burmeister *et al.*, 2004). El dominio de cabeza de estos dos serotipos reconocen el ácido siálico en una región electropositiva situada en la parte superior del trímero (Figura 6.4C, panel de la izquierda). En el panel de la derecha mostramos un mapa de potenciales electroestáticos de la superficie del dominio de cabeza del adenovirus porcino 4 en el que podemos ver que esta región electropositiva no existe lo que sugiere que este dominio de cabeza no puede interactuar con ácido siálico de la misma forma que los adenovirus humanos de tipo 37 y 19p. Tomadas juntas, todas las observaciones previamente citadas sugieren que este dominio de cabeza es incapaz de interactuar con ninguno de los receptores celulares adenovirales anteriormente mencionados.

6.4 - La estructura del dominio de galectina

La estructura del dominio de galectina (Figura 6.5) muestra dos dominios CRD unidos por un péptido rico en prolinas e interactuando entre ellos a través de una nueva superficie de interacción. Ambos CRDs están orientados de tal manera que sus sitios de unión a azúcares están enfrentados, de esta forma, al menos en potencia, ambos dominios podrían interactuar simultáneamente con un mismo carbohidrato. Ambos CRDs están plegados

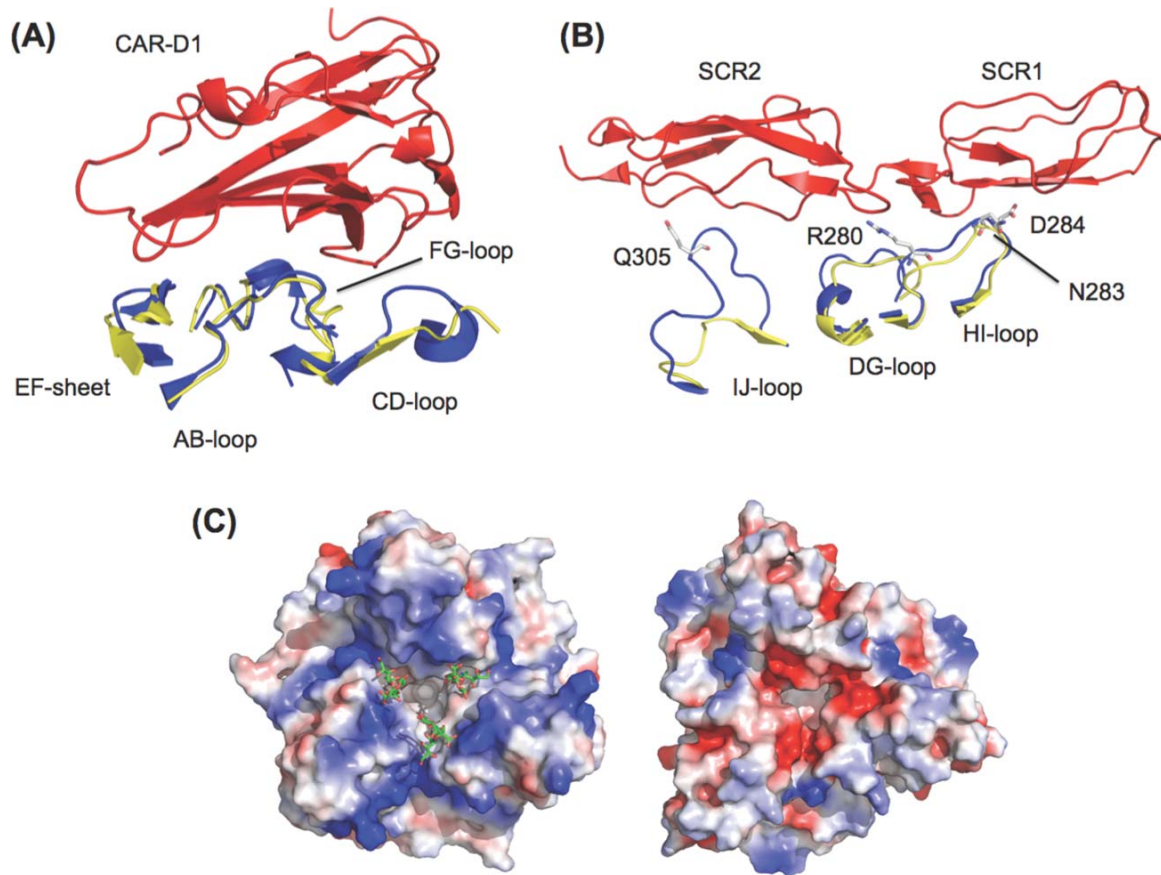


Figura 6.4. El dominio de cabeza del adenovirus porcino de tipo 4 probablemente no es capaz de interactuar con CAR, CD46 ni ácido siálico. **(A)** Comparación estructural entre los dominios de cabeza de las fibras de adenovirus porcino 4 (amarillo) y el del adenovirus humano 12 (azul) en complejo con el dominio D1 de CAR (rojo). Se han señalado los bucles AB-, CD-, y FG- así como la pequeña lámina EF. **(B)** Comparación estructural entre los dominios de cabeza de la fibras de adenovirus porcino 4 (amarillo) y el del adenovirus humano 11 (azul) en complejo con las dos primeras repeticiones consenso de CD46. Se han señalado los bucles DG-, HI-, e IJ- así como los residuos involucrados en la interacción con CD46 en la fibra del adenovirus humano de tipo 11. **(C)** Mapa de potenciales electrostáticos de superficie del dominio de cabeza de la fibra del adenovirus humano 37 (izquierda) en complejo con sialil-lactosa y el dominio de cabeza del adenovirus porcino 4 (derecha).

formando un sándwich β con una topología de tipo *jelly-roll* que es típica de lectinas vegetales y otras galectinas (Kasai *et al.*, 1996). Cada dominio contiene dos láminas β , denominadas F y S, compuestas por cinco y seis cadenas β respectivamente. El sitio de unión a carbohidratos está localizado en la lámina β de tipo S. Se han designado las cadenas β F1-5 y S1-6 y se ha añadido el prefijo n o c para distinguir los elementos de los dominios N- o C-terminal respectivamente. Además de estas cadenas β existe una hélice 3_{10} (hélice 2) situada en el bucle cF5-cS2 (Figura 6.5).

El péptido que une ambos dominios cruza a lo largo de la lámina β nF. Está compuesto por 23 residuos, es rico en prolinas y tiene una pequeña hélice 3_{10} (hélice 1). En la forma cristalina libre de ligando ha sido posible trazar una pequeña cola C-terminal en las cadenas B y C, que forma una pequeña cadena β que interacciona con la cadena β nS6 de una molécula vecina. En todas las formas cristalinas ha sido posible modelar un pequeño vestigio de la cola de purificación N-terminal que no será tenido en cuenta de aquí en adelante.

Un alineamiento estructural entre ambos dominios CRD muestra una desviación cuadrática media entre sus carbonos α de 1.4 Å con una identidad de secuencia del 39%. Las principales diferencias están localizadas en el bucle S3-S4, que es más largo en el dominio C-terminal, el bucle S4-S5 y la presencia de la hélice 3_{10} en el bucle cS2-cF5, que no está presente en el dominio N-terminal.

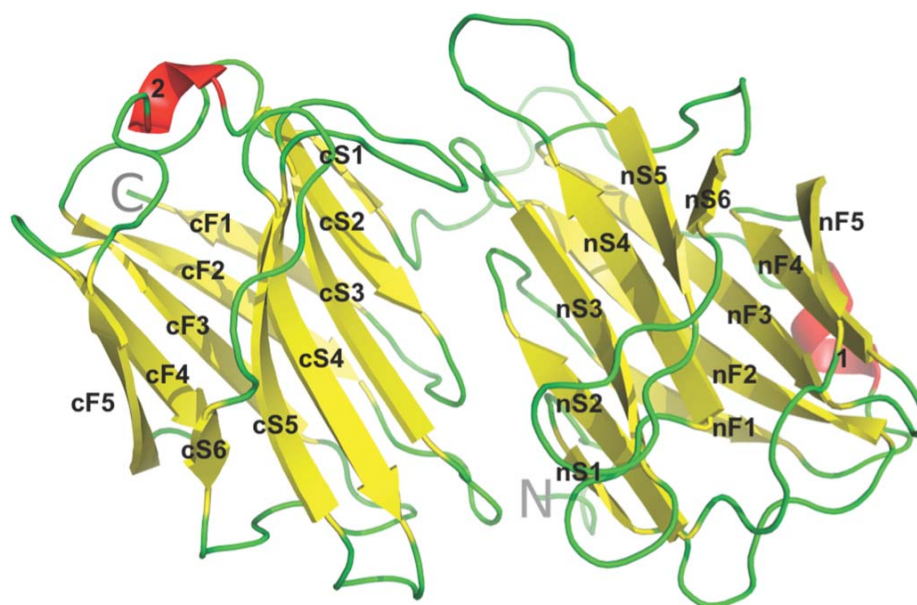


Figura 6.5 Estructura del dominio de galectina del adenovirus porcino de tipo 4 (cepa NADC-1). Las cadenas β han sido coloreadas en amarillo, las hélices en rojo y los bucles en verde. Las cadenas β han sido nombradas de acuerdo con Lobsanov *et al.* (1993). Los extremos N- y C-terminal están marcados.

6.5 - El dominio de galectina es capaz de interactuar con carbohidratos

Para determinar con qué carbohidratos es capaz de interactuar el dominio de galectinas se estableció una colaboración con el grupo de David Curiel y Joel Glasgow. Se utilizó un *microarray* con azúcares desarrollado por el “*Consortium for Functional Glycomics*” (Emory University School of Medicine, Atlanta GA, U.S.A.). En este *microarray* se evaluaron 406 perfiles de unión a carbohidratos distintos de

manera simultánea. Como muestra se utilizó el adenovirus 5 humano modificado con la fibra porcina (Ad5Luc1-PK) y como control, el adenovirus 5 humano (Ad5Luc1). En la figura 6.6 se muestra los perfiles de unión a carbohidratos que se obtuvieron. Se puede observar que el control Ad5Luc1 no muestra ninguna unión específica en el *microarray* (panel superior). Por el contrario, el vector Ad5Luc1-PK (panel inferior) reconoce específicamente 5 carbohidratos estructuralmente relacionados: Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc (tri(*N*-acetillactosamina)), GlcNAc α 1-4Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc β 1-3Gal β 1-4GlcNAc, Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc, (lacto-*N*-neotetraosa), Gal α 1-4Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc and Gal β 1-4GlcNAc β 1-3Gal β 1-3GlcNAc. Todos estos azúcares contienen al menos una unidad de *N*-acetillactosamina (Gal β 1-4GlcNAc). Estos datos muestran que la fibra de la cepa NADC-1 del adenovirus porcino 4 une específicamente cinco carbohidratos. Como el vector Ad5Luc1-PK contiene los residuos 43-703 de la fibra no podemos concluir que el dominio de galectina sea el responsable de mediar estas interacciones específicas, aunque el grado de conservación de algunos residuos implicados en la interacción con β galactósidos en otras galectinas hace de esta opción algo probable. Para determinar si este dominio es capaz de interactuar con los azúcares utilizamos dos aproximaciones experimentales distintas, SPR (surface plasmon resonance) y la cristalización de los complejos proteína-azúcar.

6.6 - Las estructuras de los complejos galectina-carbohidrato

Para demostrar que existe interacción entre el dominio de galectina y los carbohidratos identificados en el *microarray*, cocrystalizamos este dominio con cuatro azúcares diferentes: lactosa, *N*-acetillactosamina, lacto-*N*-neotetraosa y tri(*N*-acetillactosamina) y resolvimos su estructura mediante reemplazamiento molecular (Figura 6.7; ver la Tabla 6.I en donde figuran las estadísticas de los datos, el refinamiento y la validación). Las moléculas de todas las formas cristalinas son virtualmente idénticas con desviaciones cuadráticas medias entre sus carbonos α de 0.4-1.1 Å. Las principales diferencias se encuentran en el péptido que une ambos dominios, lo que junto con los elevados factores térmicos de algunas regiones en este péptido, sugiere que puede actuar como una región bisagra entre ambos CRDs.

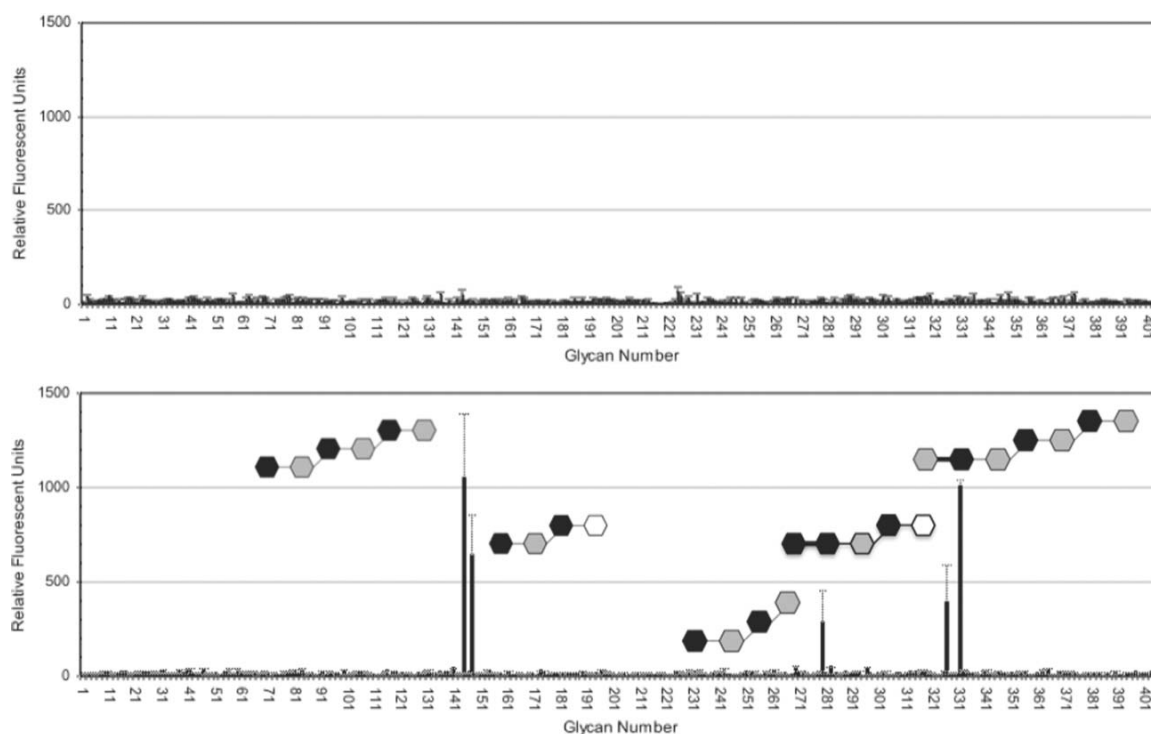


Figura 6.6. Perfil de unión a azúcares que mostraron los vectores adenovirales Ad5Luc1 y Ad5Luc1-PK en el *microarray*. El panel superior muestra el perfil de unión de un vector basado en el adenovirus humano 5 mientras que el panel inferior muestra el mismo vector modificado con la fibra del adenovirus porcino 4 (cepa NADC-1) en donde los dominios de cabeza, RGD y galectina han reemplazado a la fibra del Ad5. En el eje y se muestra el promedio de las unidades relativas de fluorescencia (RFU) de seis medidas de cada azúcar mientras que en el eje de las x el número de cada azúcar en el *array*. Los azúcares que dieron una señal clara se representan esquemáticamente como sigue: la galactosa se representa con un hexágono negro, la glucosa con uno blanco y la *N*-acetilglucosamina con uno gris. El carbono anomérico está situado en el vértice derecho y el resto se numeran en el sentido de las agujas del reloj. Las barras delgadas y gruesas representan enlaces glicosídicos alfa y beta, respectivamente.

Los análisis estructurales de otros complejos entre galectinas y lactosa y sus derivados han mostrado cuáles son los residuos implicados en la interacción con carbohidratos. En el dominio N-terminal los residuos His435, Asn437, Arg439, Val449, Asn451, Trp459, Glu462 y Arg464 componen el sitio canónico de unión a carbohidratos (ver alineamiento en la Figura 6.11). En el dominio C-terminal solo 5 de estos residuos están conservados, la valina ha sido reemplazada por la Ala607, y las argininas han sido reemplazadas por la Gly597 y la Thr621, respectivamente. En concordancia con esto, sólo el complejo co-cristalizado con altas concentraciones de

lactosa (70 mM), muestra el azúcar unido en el dominio C-terminal con una ocupancia de 0.5.

Las estructuras de los complejos con lactosa y *N*-acetillactosamina muestran que ambos azúcares se unen al sitio canónico de reconocimiento de carbohidratos del dominio N-terminal (Figuras 6.7A y B). En ambos azúcares, existe una interacción de apilamiento entre la cadena lateral del Trp459 y el anillo de la galactosa. Las cadenas laterales de la His435, Asn437 y Arg439 forman puentes de hidrógeno con el oxígeno O4 de la galactosa, la Arg439 con el O5 y las cadenas laterales de la Asn451 y el Glu462 con el oxígeno O6 de la galactosa. La glucosa (en la lactosa) y la *N*-acetilglucosamina (en la *N*-acetillactosamina) ocupan la misma posición y hacen contactos similares. El oxígeno O3 forma puentes de hidrógeno con las cadenas laterales de la Arg439, Glu462 y Arg464 y el oxígeno O4 con la cadena lateral de la Arg439. En el complejo con lactosa se forma un puente de hidrógeno entre el oxígeno O2 de la glucosa y la cadena lateral de la Arg464 mientras que en el complejo con *N*-acetillactosamina los oxígenos O3 y O4 de la galactosa forman un puente de hidrógeno con la cadena lateral de la Tyr585 y el oxígeno O2 interacciona con la cadena lateral del Asp586. Ambos residuos están localizados en el bucle cS3-cS4. Proponemos que este sitio de interacción común a lactosa y *N*-acetillactosamina funciona como un sitio primario de reconocimiento.

La estructura del complejo con lacto-*N*-neotetraosa (Figura 6.7C) muestra que el residuo de lactosa está ocupando el sitio de reconocimiento primario haciendo las mismas interacciones que la lactosa que han sido previamente descritas. Sin embargo, el residuo de *N*-acetillactosamina está localizado en un nuevo sitio de unión en el cual el oxígeno O6 de la *N*-acetilglucosamina forma un puente de hidrógeno con la cadena lateral del Asp424 y la cadena lateral de la Phe574 forma una interacción por apilamiento con el anillo de la galactosa.

La estructura del complejo con tri-*(N*-acetillactosamina) (Figura 6.7D) muestra que el residuo de *N*-acetillactosamina situado en el extremo reductor (es decir los residuos 5 y 6 del azúcar) ocupan el sitio de reconocimiento primario y, con la excepción de la interacción con la cadena lateral del Asp586, se observan las mismas interacciones observadas para la *N*-acetillactosamina sola (Figura 6.7A). La unidad de *N*-acetillactosamina central (residuos 3 y 4) ocupan la misma posición que la unidad de *N*-acetillactosamina en la lacto-*N*-neotetraosa y, además de las interacciones con la Phe574 y Asp424 descritas previamente, el oxígeno O4 de la *N*-acetilglucosamina y

el O3 de la galactosa forman un puente de hidrógeno con la cadena lateral de la Arg453. Finalmente, la unidad de *N*-acetillactosamina situada en el extremo no reductor (residuos 1 y 2) está localizada en un nuevo sitio de unión formado exclusivamente por residuos del dominio C-terminal. Allí, los oxígenos O2 y O3 de la galactosa forman puentes de hidrógeno con la cadena lateral de la Arg629 y el oxígeno carbonílico de la Ala572, Glu573 y Phe598 forman un puente de hidrógeno con el oxígeno O2 de la galactosa.

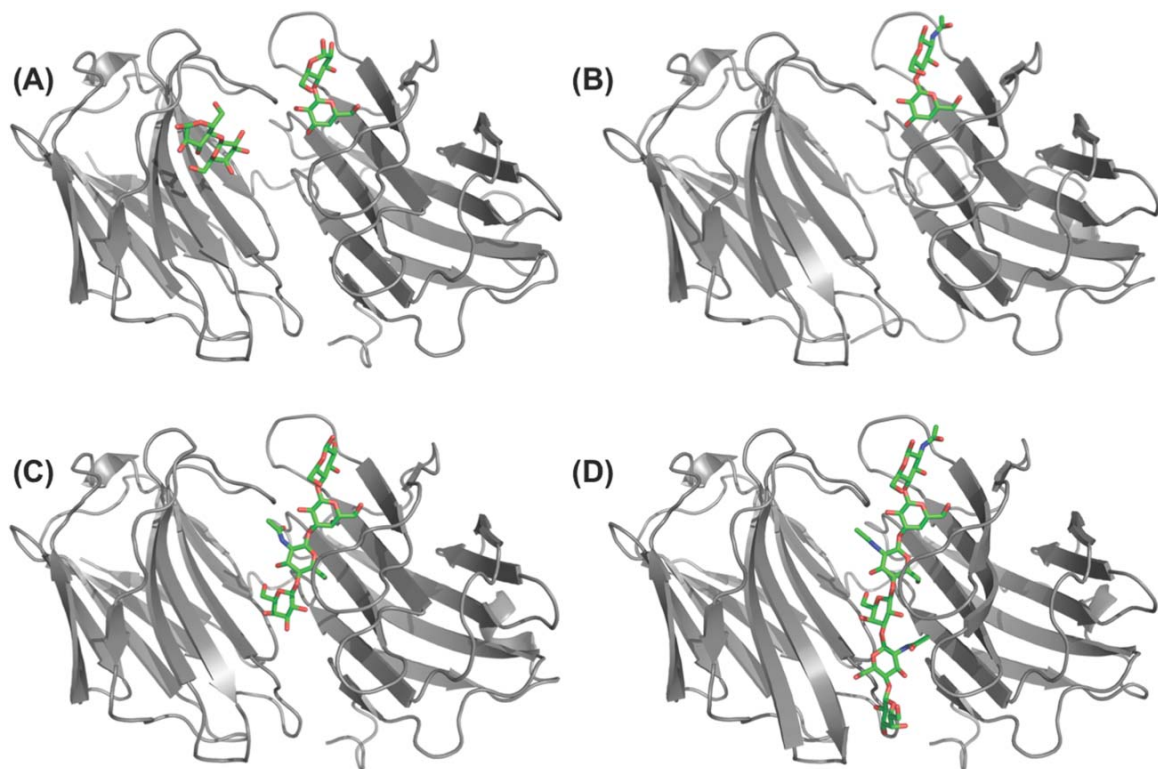


Figura 6.7 Estructuras del dominio de galectina unido a lactosa (A), *N*-acetillactosamina (B), lacto-*N*-neotetraosa (C) y tri-*(N*-acetillactosamina) (D). Sólo se muestra la cadena A en la estructura en complejo con tri-*(N*-acetillactosamina).

La molécula de tri-*(N*-acetillactosamina) rellena todo el surco que se forma entre ambos dominios (Figura 6.8B). La baja afinidad de unión que muestra el sitio de interacción canónica del dominio C-terminal así como la presencia del bucle largo cS3-cS4 impide que esta interacción pueda ocurrir en el sentido opuesto. Las estructuras presentadas en este trabajo sugieren que la fibra del aislado NADC-1 del adenovirus porcino de tipo 4 no puede hacer interacciones adicionales con polímeros de *N*-acetillactosamina más largos. Por lo tanto, si se detectase una afinidad mayor por estos polímeros, tendría que estar relacionada con contribuciones de avidéz

(Lortat-Jacob *et al.*, 2001). Sin embargo, los carbohidratos no lineales que hemos identificado en el *microarray* podrían tener interacciones adicionales a las aquí descritas, en las posiblemente participaría el dominio C-terminal.

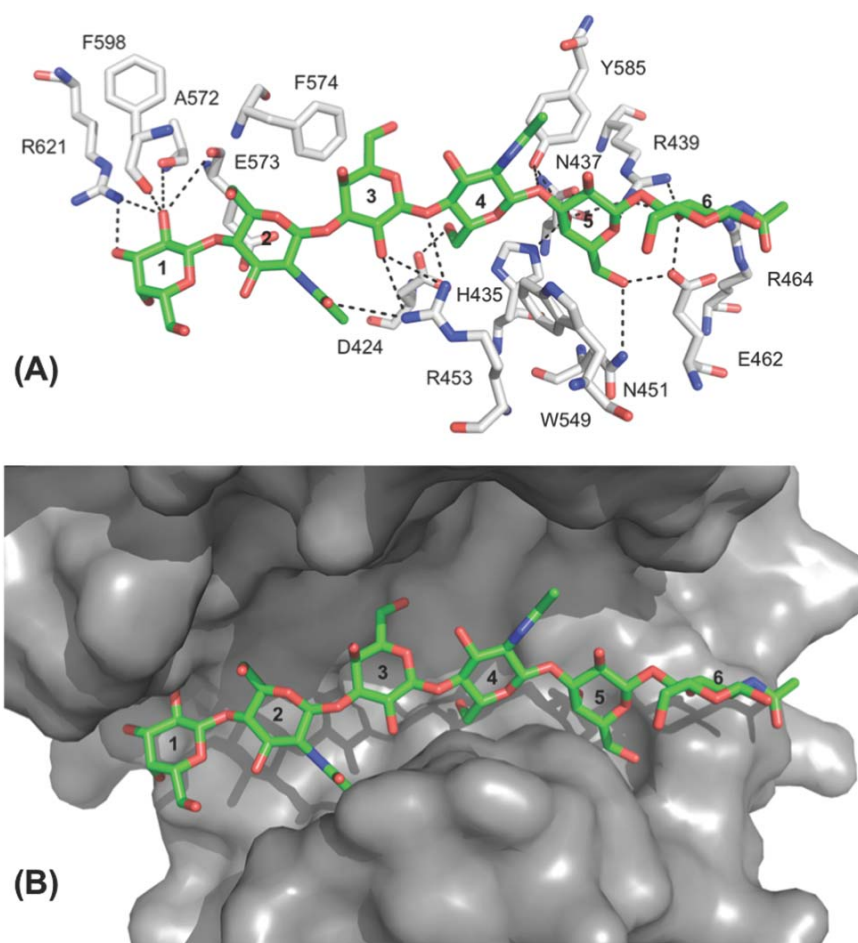


Figura 6.8 El sitio de interacción con la tri-(*N*-acetilglucosamina). (A) Representación de los residuos que interactúan con la tri-(*N*-acetilglucosamina). (B) Representación de la superficie del bolsillo de interacción. En el azúcar los carbonos se representan en verde, los nitrógenos en azul y los oxígenos en rojo.

6.7 - Experimentos de resonancia de superficie de plasmones (SPR)

En colaboración con Eva Muñoz del departamento de Química Orgánica de la Universidad de Santiago de Compostela se estimó la afinidad del dominio de galectina por los siguientes 5 azúcares: lactosa, *N*-acetilglucosamina, lacto-*N*-neotetraosa, di-(*N*-acetilglucosamina) y tri-(*N*-acetilglucosamina) mediante resonancia de superficie de plasmones (SPR: Piliarik *et al.*, 2009). Los resultados se resumen en la Tabla 6.II. Tanto la lactosa como la *N*-acetilglucosamina se unían muy débilmente y sólo se pudo estimar el orden de magnitud de la constante de disociación. Aún así, la

afinidad de la galectina por la *N*-acetillactosamina parece ser unas diez veces mayor que por la lactosa. La estructura del complejo galectina-*N*-acetillactosamina muestra que este azúcar hace dos interacciones adicionales con la Tyr585 y Asp586, lo que podría explicar esta mayor afinidad. Debido a la baja afinidad que la proteína presentaba por estos dos disacáridos no se pudo hacer una estimación de la estequiometría de la unión. Del mismo modo, esta baja afinidad también explicaría el por qué no se han podido detectar estos azúcares en el *microarray*.

Los tetrasacáridos lacto-*N*-neotetraosa y di-(*N*-acetillactosamina) se unían a la galectina con afinidades similares, con una constante de disociación de unos 300 μM , 30 veces mayor que la estimada para la *N*-acetillactosamina, y con una estequiometría estimada de 1:1. La estructura con lacto-*N*-neotetraosa muestra que este azúcar hace dos interacciones adicionales a las que hace la lactosa, un puente de hidrógeno con el Asp424 y una interacción de apilamiento con la Phe574. Aunque no tenemos una estructura con la di-(*N*-acetillactosamina) es probable que también haga esta dos interacciones adicionales.

Los datos de unión para el hexasacárido tri-(*N*-acetillactosamina) se ajustan a un modelo en el que el azúcar puede interactuar de dos modos distintos con la proteína. Uno de los modos de interacción, con una constante de disociación de 20 μM y una estequiometría 1:1, correspondería a una interacción en la que los seis residuos del azúcar estarían interactuando con la proteína; esto es, el complejo observado en la estructura cristalina. El segundo modo de unión, con una constante de disociación de 380 μM y una estequiometría 1:1, se podría corresponder con un modo de unión en el cual sólo los cuatro residuos del extremo no reductor del azúcar interactuarían con la galectina, de modo parecido a lo que lo haría la di-(*N*-acetillactosamina) y de ahí que tengan afinidades similares. Una forma de interacción similar se ha demostrado en el dominio N-terminal de la galectina humana 9 (Nagae *et al.*, 2009).

Aunque las afinidades que presenta la galectina monomérica hacia los azúcares son bastante débiles, debemos recordar que la fibra del adenovirus es trimérica y que por lo tanto contiene tres dominios de galectina independientes. Además, cada partícula viral contiene doce fibras triméricas. La presencia de 36 dominios de galectinas independientes en cada partícula viral probablemente

producirá que la afinidad de estas partículas por los oligosacáridos previamente descritos sea mucho mayor (Lortat-Jacob *et al.*, 2001; Lundquist & Toone, 2002).

Tabla 6.II. Constantes de afinidad obtenidas por SPR

Carbohidrato	K_d
lactosa	~100 mM
<i>N</i> -acetillactosamina	~10 mM
lacto- <i>N</i> -neotetraosa	303 μ M \pm 4
di(<i>N</i> -acetillactosamina)	309 μ M \pm 9
tri(<i>N</i> -acetillactosamina)	20 μ M \pm 3; 380 μ M \pm 40 ^a

^aSe detectaron dos modos de unión con dos afinidades distintas.

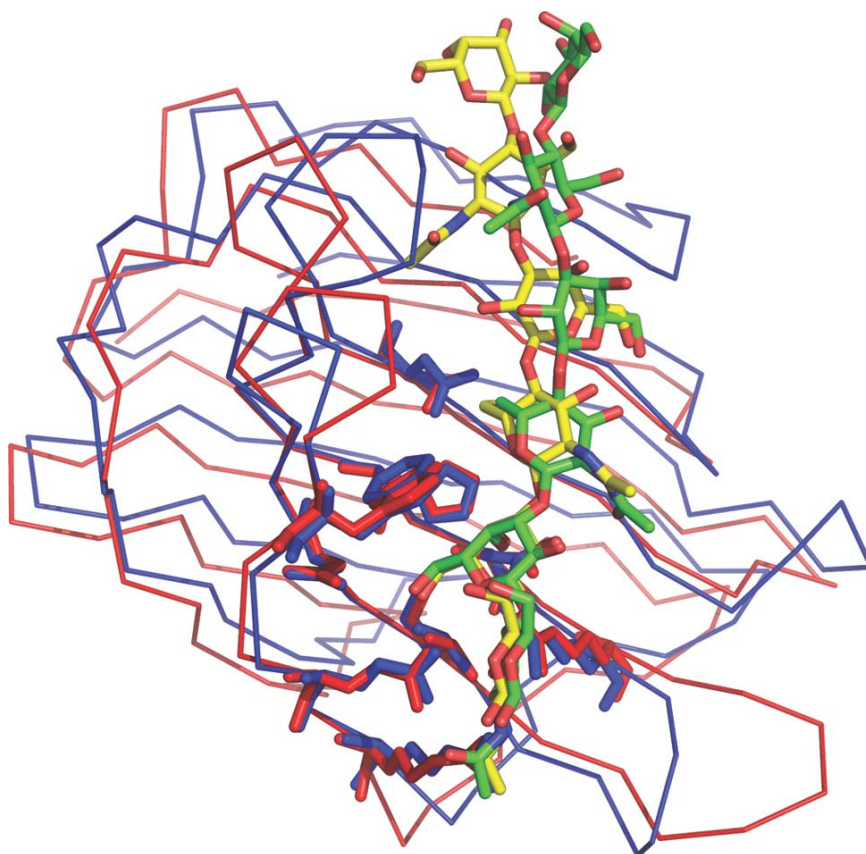


Figura 6.9 Superposición de los dominios N-terminal del dominio de galectina del adenovirus porcino 4 (la proteína en rojo y el azúcar en amarillo) y la galectina humana 9 (la proteína en azul y el azúcar en verde). Los átomos de oxígeno y nitrógeno del azúcar están coloreados en rojo y azul, respectivamente. Se representan los residuos que interaccionan con el azúcar usados en la superposición.

6.8 - Comparación con otras galectinas repetidas en tándem

Un alineamiento de múltiples secuencias de galectinas repetidas en tándem (Figura 6.11) muestra que la mayoría de los residuos implicados en la interacción con la tri-(*N*-acetillactosamina) en nuestra estructura están conservados tanto en la galectina 9 humana como en la porcina. Una superposición de la estructuras de los dominios N-terminales de ambas estructuras basada en estos residuos conservados muestra que ambos dominios comparten un mismo modo de interacción con la tri-(*N*-acetillactosamina) (Figura 6.9). Por lo tanto, la conformación extendida que el hexasacárido adopta en complejo con el dominio N-terminal de la galectin 9 humana sería compatible con un modelo de organización similar al observado en el dominio de galectina del adenovirus porcino de tipo 4. Potencialmente, todas las galectinas repetidas en tándem podrían estar organizadas de tal modo que ambos dominios estén posicionados frente a frente, como previamente se ha descrito. Este modelo de organización podría explicar que, aunque tanto la di- como la tri-(*N*-acetillactosamina) hacen las mismas interacciones con el dominio N-terminal de la galectina 9 (Nagae *et al.*, 2009), experimentos de cromatografía de afinidad indican que la tri-(*N*-acetillactosamina) tiene una afinidad mayor por la galectina 9 que contiene ambos dominios.

Hasta donde sabemos, esta estructura representa la primera estructura de una galectina repetida en tándem intacta. Previamente se ha sugerido que esta clase de galectinas no se pueden cristalizar fácilmente debido al desorden introducido por el péptido que une ambos dominios CRD (Cummings *et al.*, 2009). Quizás el péptido relativamente corto que presenta este dominio de galectina nos ha permitido salvar esta dificultad. Las estructuras resueltas han permitido realizar un análisis detallado de las interacciones presentes en la interfaz de interacción entre ambos dominios. Esta interfaz utiliza el 12% de la superficie de cada dominio (830-900 Å). En ella se forman puentes salinos (entre Asp514-His670 y Arg508-Glu573), tres puentes de hidrógeno e interacciones de van der Waals. Los residuos involucrados en la formación de esta interfaz son los mismos en todas las estructuras resueltas a excepción de la presencia de la Tyr585 en las estructuras que contienen *N*-acetillactosamina en el sitio de interacción primario.

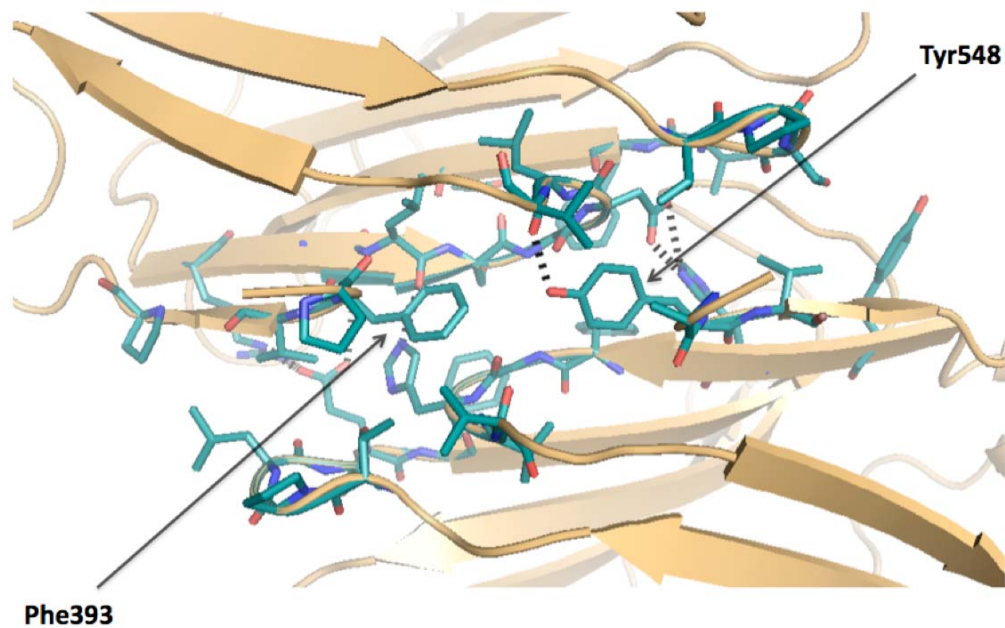


Figura 6.10 Vista en detalle de los residuos que componen la interfaz entre los dos dominios CRD, se indican los residuos aromáticos Phe393 y Tyr548 que están conservados en otras galectinas repetidas en tándem.

Los residuos implicados en esta interfaz pueden ser divididos en dos regiones. La primera región comprende los residuos situados en el bucle nS4-nS5 y el final de la cadena β nS4 que están interaccionando con residuos situados en el bucle cS3-cS4. La segunda región comprende los residuos situados al comienzo de la cadena β nS1, el bucle nF2-nS3, el final de la cadena β nS2 y el bucle nS2-nF1 que están interaccionando con los residuos estructuralmente equivalentes del dominio C-terminal. Dos residuos aromáticos situados justo al comienzo de la cadena β S1 de cada dominio forman un núcleo hidrofóbico que puede tener gran importancia para la estabilidad de la interfaz previamente descrita. Estos dos residuos aromáticos están universalmente conservados en todas las galectinas repetidas en tándem (Figura 6.11) y están ausentes en las galectinas prototípicas. Este patrón de conservación sugiere que otras galectinas repetidas en tándem podrían en potencia adoptar el mismo modo de organización al descrito aquí para el dominio de galectinas de la fibra del adenovirus porcino de tipo 4.

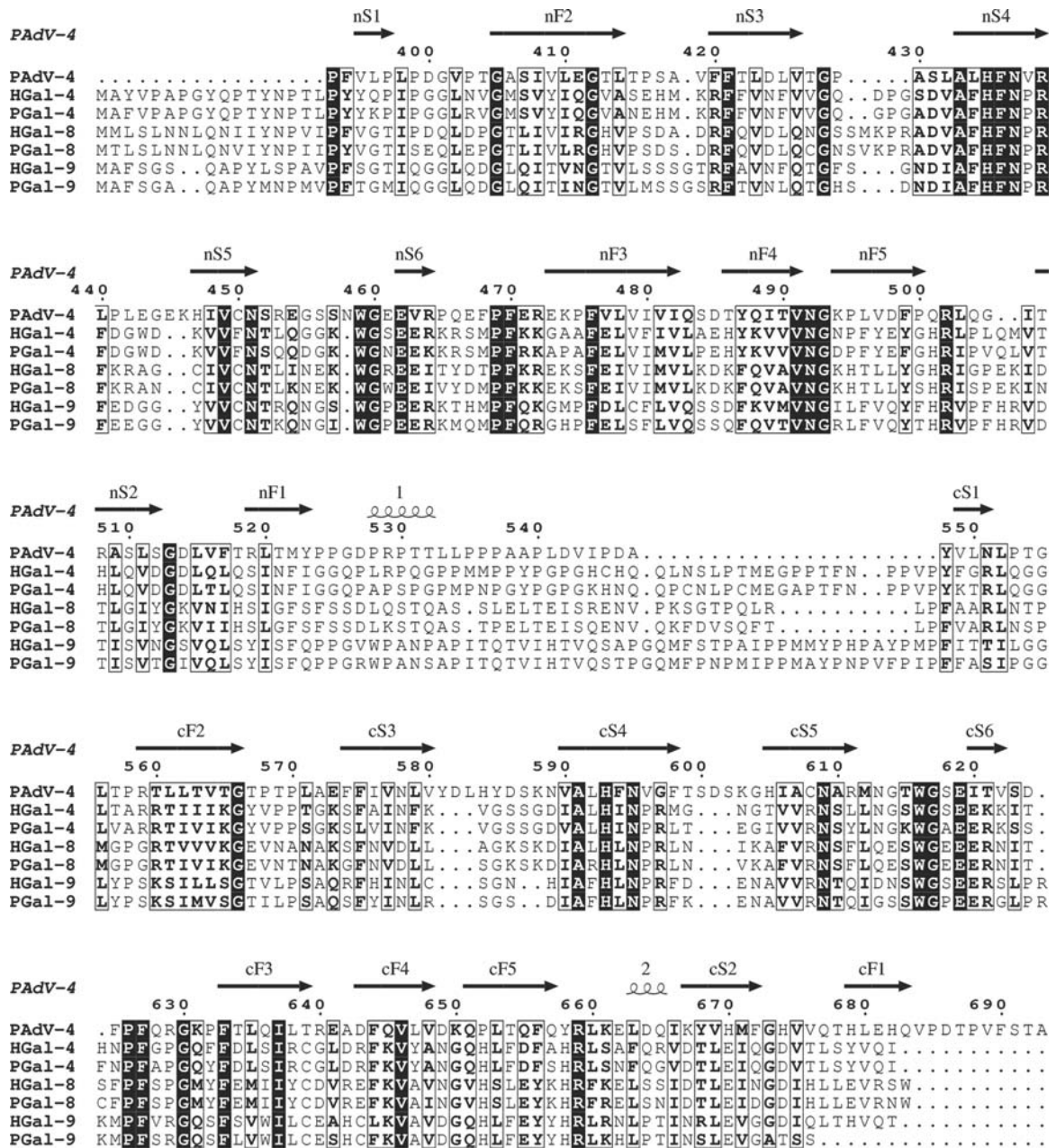


Figura 6.11 Alineamiento entre múltiples secuencias de galectinas repetidas en tándem. El dominio de galectina del adenovirus porcino 4 (PAdV-4; UNIPROT-code Q83467), galectina humana-4 (HGal-4; P56470), galectina porcina-4 (PGal-4; Q29058), galectina humana-8 (HGal-8; O00214), galectina porcina-8 (PGal-8; NP_001136299.1), galectina humana-9 (HGal-9; O00182) y galectina porcina-9 (PGal-9; Q9XSM8). Se muestran los residuos idénticos y similares con cajas negras y blancas, respectivamente. Se indican los elementos de estructura secundaria del dominio de galectina de la fibra del adenovirus porcino 4.

6.9 - Conclusión

En este trabajo hemos resuelto la estructura del dominio de cabeza y galectina de la fibra del aislado NADC-1 del adenovirus porcino de tipo 4. La estructura del dominio de cabeza sugiere que determinadas características a nivel estructural impiden que este dominio se una a ningún receptor adenoviral clásico. La estructura del dominio de galectina no sólo representa la primera estructura de una galectina viral sino la primera de una galectina repetida en tándem en la que ambos dominios están presentes. Los resultados presentados en este trabajo podrán ser usados para profundizar en la primeras etapas del ciclo replicativo de estos virus y podría llevar al desarrollo de aplicaciones terapéuticas, bien dirigidas a prevenir la interacción fibra-receptor y por lo tanto impidiendo la infección, o bien explotando las propiedades de unión a receptores únicas de esta fibra en el diseño de vectores de vacunación.

Capítulo 7

DISCUSIÓN

Capítulo 7

Discusión

Los reovirus y los adenovirus han despertado un gran interés debido a sus grandes aplicaciones biotecnológicas. Los adenovirus en el ámbito de la terapia génica y los reovirus en la terapia contra el cáncer. Hasta la fecha se han realizado numerosos estudios estructurales de proteínas de reovirus de mamífero y de adenovirus humanos. En este trabajo nos propusimos extender estos trabajos a representantes aviares (reovirus y adenovirus aviar) y porcinos (adenovirus porcino) con el objetivo de ampliar el conocimiento de la biología de estos virus en otros hospedadores y de esta manera poder ampliar y mejorar sus aplicaciones biotecnológicas.

En esta tesis doctoral aparecen caracterizadas las estructuras de la proteína σC (la fibra del reovirus aviar), la proteína σA (la proteína responsable de la actividad anti-interferón del reovirus aviar), el dominio de cabeza de la fibra larga del adenovirus aviar 1 y dos dominios de la fibra del adenovirus porcino 4. Sus estructuras han permitido ahondar en procesos clave de la biología de los virus previamente citadas como se discutirá a continuación.

7.1- Las fibras de reovirus y adenovirus son similares

Aunque los reovirus y los adenovirus pertenecen a familias de virus distintas y tienen pocas propiedades en común, los estudios realizados en la fibra del reovirus de mamífero (Chappell *et al.*, 2002) y el adenovirus humano 2 (van Raaij *et al.*, 1999) pusieron de manifiesto unas similitudes inesperadas entre ambas. Entre estas semejanzas cabe destacar que ambas comparten una única organización de sus dominios, que son proteínas triméricas que se anclan a proteínas pentaméricas por su extremo N-terminal y que ambas interactúan con receptores similares desde el punto de vista estructural y funcional. Estos parecidos no sólo podrían sugerir la existencia de un proceso de convergencia evolutiva, sino la existencia de un mecanismo común de adhesión y entrada celular, un proceso poco caracterizado en virus no envueltos.

Tanto la fibra del reovirus de mamífero $\sigma 1$ como la fibra de adenovirus humano 2 son fibras triméricas alargadas con una región fibrosa compuesta por repeticiones en triple β espiral, aunque en el caso de los reovirus esta región fibrosa β estructurada coexiste con una región plegada formando un triple *coiled coil*, como se ha descrito en el Capítulo 3. La única otra fibra de estructura conocida que contiene repeticiones en triple β espiral es la fibra de fago PRD1: en este caso la región β estructurada coexiste con una región con un plegamiento de tipo colágeno. Tanto en la fibra de los reovirus como en la fibra de los adenovirus, el dominio globular de cabeza situado en el extremo C-terminal está formado por ocho cadenas β antiparalelas. Estas constituyen un barril β en el caso de la fibra de reovirus y un sándwich β en la de adenovirus, ambas con una topología única y exclusiva de estos dominios. En las fibras de ambos virus existen regiones bisagra entre el dominio de cabeza y el dominio fibroso (Fraser *et al.*, 1990; van Raaij *et al.*, 1999). Estas regiones han resultado tener una importancia crítica en la interacción de las fibras de adenovirus con su receptor (Chiu *et al.*, 2001) y aunque todavía no se ha definido su relevancia biológica en reovirus, probablemente jueguen un papel importante en la interacción fibra-receptor y/o en la maduración de la partícula viral. Con respecto a sus receptores celulares, tanto el reovirus de mamífero como algunos adenovirus se unen a receptores que pertenecen a la familia de la inmunoglobulinas. En ambos casos el dominio más N-terminal del receptor es reconocido por el dominio más C-terminal de la fibra (Stehle & Dermody, 2003).

7.2- Las fibras del reovirus de mamífero y aviar presentan diferencias significativas

En esta tesis (Capítulo 3) se ha caracterizado estructuralmente por primera vez la fibra de un reovirus aviar. Su estructura en términos generales es similar a su homóloga de reovirus de mamífero, aunque un análisis detallado revela algunas diferencias fundamentales.

7.2.1- Regulación del estado trimérico del dominio de cabeza

Diversos estudios realizados con la proteína $\sigma 1$ del reovirus de mamífero sugieren que el estado trimérico de la estructura cristalina es transitorio y que se produce un cambio conformacional durante la pérdida de la cápside externa en el pH ácido del endosoma. Entre estas evidencias se encuentran los experimentos de

microscopía electrónica realizados con el reovirus de mamífero (Furlong *et al.*, 1988) que muestran que $\sigma 1$ adopta una conformación compacta en el virión mientras que en las ISVPs adopta una conformación extendida. También, los estudios con virus PI mutantes (virus de infectividad persistente), que muestran cómo mutaciones localizadas en $\sigma 1$ que alteran la estabilidad del trímero (Wilson *et al.*, 1996) mejorando su capacidad infectiva. Estas evidencias sugieren que la forma monomérica puede jugar un papel importante en el desensamblaje de la cápside externa y la replicación del virus.

En el contexto de la importancia que un cambio conformacional de $\sigma 1$ puede tener en la biología del reovirus de mamífero, se han señalado dos regiones que pueden ser cruciales en este proceso. Una región bisagra de cuatro residuos que conecta la segunda y tercera repetición en triple β espiral y que permitiría que el dominio globular se mueva de una manera independiente del dominio fibroso (Chappell *et al.*, 2002) y un agrupación de aspartatos localizada en la interfaz trimérica, justo en la base del dominio de cabeza (Chappell *et al.*, 2002; Schelling *et al.*, 2007). Se ha propuesto (Schelling *et al.*, 2007) que el estado de protonación de uno de estos aspartatos podría jugar un papel importante en la regulación del estado trimérico de $\sigma 1$, al actuar como un interruptor mediado por pH. Así, en un ambiente ácido los grupos carboxilo de los aspartatos estarían protonados estableciendo dos puentes de hidrógeno y contribuyendo de esta manera a la estabilidad del trímero. En un ambiente neutro o básico, estos grupos carboxilo estarían desprotonados de tal forma que la acumulación de tres cargas negativas en esta interfaz sería altamente desfavorable y este dominio tendría tendencia a estar en forma monomérica (Schelling *et al.*, 2007). Hasta la fecha sólo se ha encontrado una agrupación de aspartatos similar en la proteína G del virus de la estomatitis vesicular (VSV) (Roche *et al.*, 2006). Ambos virus utilizan la vía endosomal para entrar en la célula hospedadora.

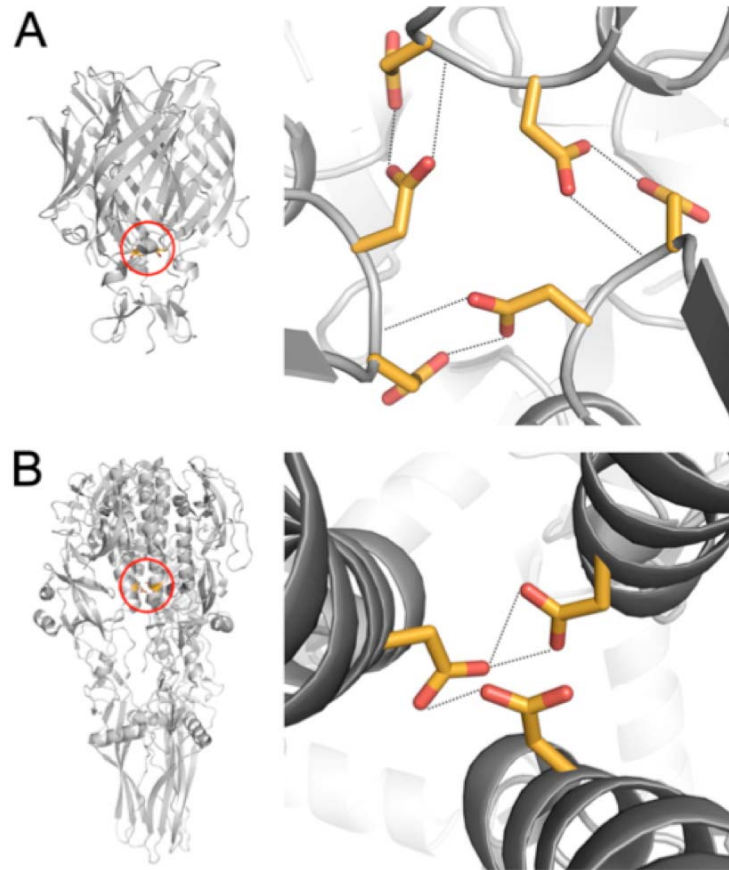


Figura 7.1 Comparación de las agrupaciones de aspartatos halladas en la fibra $\sigma 1$ del reovirus de mamífero (A) y la proteína G de VSV (B). A la izquierda se muestra una vista del trímero completo en donde la localización de las agrupaciones se ha indicado con un círculo rojo. A la derecha se muestra una vista detallada de las agrupaciones de aspartatos. Figura extraída de Schelling *et al.* (2007)

Hasta la fecha no hay evidencias que sugieran que la proteína σC pueda sufrir cambios conformacionales similares durante la maduración de la partícula viral en la vía endosomal. El análisis de la estructura muestra que, del mismo modo que ocurre en $\sigma 1$, la gran estabilidad del trímero se puede explicar por la gran superficie de interacción que existe entre los monómeros. Sin embargo, y a diferencia de $\sigma 1$, la orientación que adquieren los monómeros de σC en el trímero es más abierta y la superficie de contacto entre ellos es mayor. En esta interfaz encontramos tanto interacciones hidrofóbicas como polares. No encontramos la agrupación de aspartatos que aparecen en $\sigma 1$, aunque sí dos anillos de puentes salinos en el centro del dominio globular. En el triple *coiled coil* encontramos un motivo de trimerización compuesto por un puente salino intercatenario en el extremo más C-terminal y otro probablemente en su extremo más N-terminal, así como dos iones cloro coordinados por asparaginas en su centro. No podemos predecir el efecto que tendría un cambio de

pH sobre la estabilidad de los motivos antes citado ni por tanto la estabilidad del trímero a un pH distinto al usado en los experimentos de cristalización (8.5).

Por lo tanto una de las grandes diferencias que encontramos entre la fibra del reovirus de mamífero y aviar que esta última carece de un detector de pH similar al propuesto para $\sigma 1$. Aunque no está claro si los motivos de trimerización descritos en la proteína σC podrían tener una función similar a esta agrupación de aspartatos de $\sigma 1$, es posible que el comportamiento de la fibra del reovirus σC en respuesta a un cambio de pH sea distinto al observado en $\sigma 1$, lo que podría tener implicaciones importantes en la biología en ambos virus.

7.2.2 - La interfaz trimérica en los dominios de cabeza de adenovirus

Mientras que únicamente se conoce la estructura del dominio de cabeza de dos fibras reovirales, se ha resuelto la estructura de muchos más dominios globulares de cabeza de adenovirus. Si se realiza un análisis de las interacciones presentes en la interfaz trimérica en los dominios de cabeza de estos, se pueden distinguir tres clases de interfaces (Figura 7.2 y Tabla 7.I). (1) Aquellas formadas fundamentalmente por residuos hidrofóbicos, en donde se encuentran representantes de los serotipos humanos 37, 11, 7, 14, 19, 2 y 5, además del adenovirus porcino 4. (2) Aquellas formadas por tres o seis cadenas laterales de asparaginas, como las de los dominios globulares de los adenovirus humanos 3, 12, 16, 35 y 41. (3) Aquellas en las que están presentes tres cadenas laterales de glutamato que podrían ser potencialmente desestabilizadoras a pHs elevados, cuyos únicos representantes son los dominios de cabeza de las fibras larga y corta del adenovirus aviar de tipo 1.

Si la agrupación de glutamatos que se encuentra en la interfaz trimérica de los dominios de cabeza del adenovirus aviar tuviese una funcionalidad similar a la agrupación de aspartatos de $\sigma 1$, el trímero sólo debería de ser estable a pHs ácidos (la pKa de un glutamato en agua es de 4) y debería desestabilizarse a pHs más elevados. En este sentido cabe destacar que ambos dominios han sido cristalizados a pHs relativamente bajos (5.5 y 5.6).

Tabla 7.I - Interacciones entre monómeros de los dominios de cabeza en adenovirus

Virus	PDB	Interfaz	PS ⁽¹⁾	pH ⁽³⁾	Receptor
hAd37 (D)	2WGT	1	2	7.5	CD46 / CAR
hAd11 (B2)	3EXV	1	6	7.0	CD46
hAd7 (B1)	3EXW	1	3	7.6	CD46
hAd14 (B2)	3F0Y	1	6	9.0	CD46
hAd19p (D)	1UXB	1	2	7.5	CD46 / CAR
hAd2 (C)	1QHV	1	6	4.0	CAR
hAd5 (C)	1KNB	1	1	Desconocido	CAR
pAd4	2WST	1	2	8.5	Desconocido
hAd3 (B1)	1H7Z	2*	1	7.0	CD46
hAd12 (A)	1NOB	2	2	7.0	CAR
hAd16 (B1)	3CNC	2	4	7.5	CD46
hAd35 (B2)	3QLK	2	1	8.0	CD46
hAd41 (sf) ⁽²⁾	2BZV	2	8	8.0	CAR
(F)	2BZU			5.0	
fAd1 (sf) ⁽²⁾	2VTW	3	0	5.5	Desconocido
fAd1 (lf) ⁽²⁾	2IUM	3	0	5.6	Desconocido

(1) Número máximo de puentes salinos que se pueden formar en la interfaz monómero-monómero.

(2) Fibra corta (sf) y fibra larga (lf)

(3) pH al que ha sido cristalizado

7.2.3 - Las posibles regiones bisagra

Las regiones bisagra son motivos estructurales importantes para el correcto funcionamiento de muchas fibras virales como se ha citado previamente (Capítulo 3). La fibra $\sigma 1$ del reovirus de mamífero tiene dos regiones bisagra, la primera situada entre la segunda y tercera repetición en triple β espiral (Chappell *et al.*, 2002) y la segunda localizada aproximadamente en el centro de la fibra, alrededor de la región de transición entre el *coiled coil* y la triple β espiral más N-terminal (Chappell *et al.*, 2002).

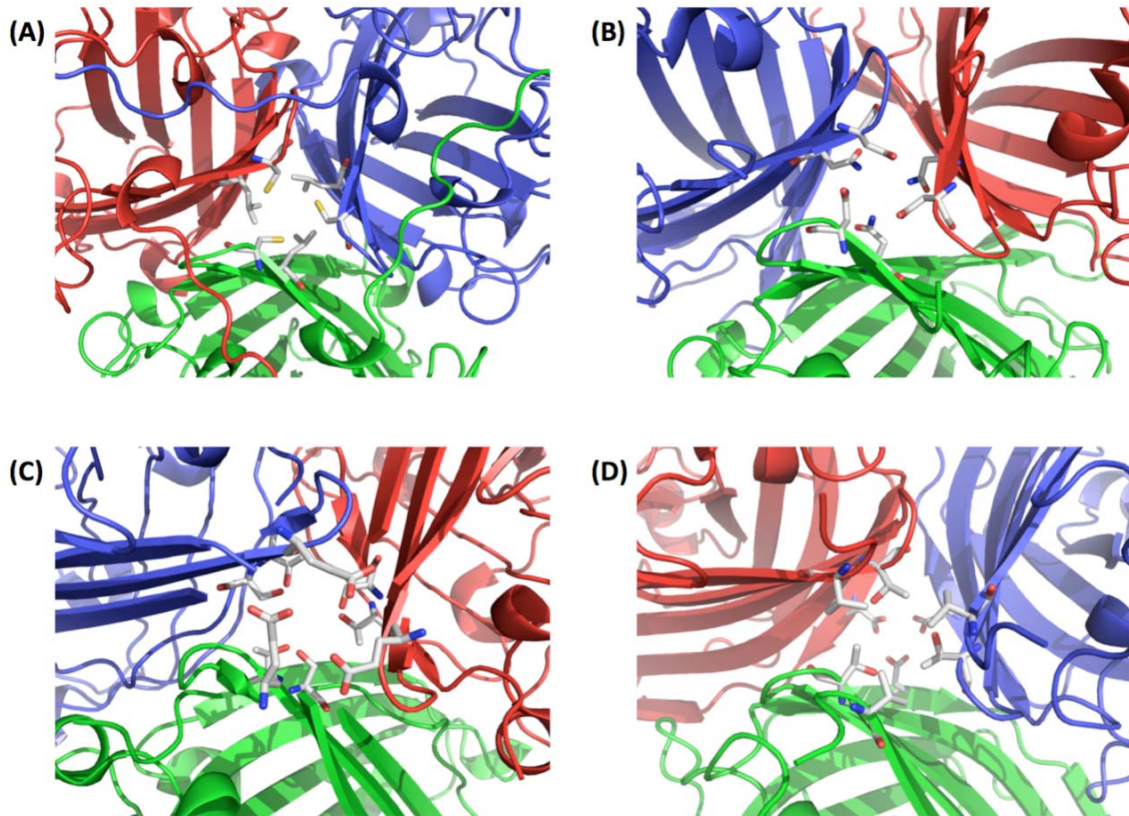


Figura 7.2 Interfaz trimérica del dominio de cabeza del adenovirus humano de tipo 2 (A), adenovirus humano de tipo 3 (B), la fibra larga (C) y la fibra corta (D) del adenovirus aviar de tipo 1.

En el caso de σC , la región que se pliega formando triple β espirales es corta y rígida, sin indicios de regiones bisagra (Capítulo 3). Sin embargo, la región de transición entre esta región β estructurada y el triple *coiled coil* está compuesta por una región con factores térmicos elevados en la que tres histidinas están coordinando un catión divalente de zinc. Es común encontrar átomos de Zn^{+2} formando parte de los sitios catalíticos de determinadas enzimas. La esfera de coordinación del Zn^{+2} en la estructura de la anhidrasa carbónica (Auld, 2005) es similar a la descrita previamente en la región bisagra de σC . Sin embargo, en la condición de cristalización utilizamos sulfato de zinc por lo que no podemos asegurar que este átomo sea un ligando natural de σC . Aunque se podría especular que el Zn^{+2} , o quizás otro catión metálico divalente, juegue un papel estructural estabilizando la conformación extendida de σC en la partícula subviral que se forma tras la pérdida parcial de la cápside externa (Benavente & Martínez-Costas, 2007). En este sentido vale la pena resaltar que la presencia de este catión depende del estado de protonación de las histidinas y por lo tanto del pH. Un análisis de la secuencia muestra otra posible región bisagra alrededor de la His110, que estaría localizada entre las dos regiones

plegadas en triple *coiled coil* predichas por el programa COILS, ya que el salto entre estas dos regiones se produciría entre los residuos 106-114.

7.2.3 - Las probables regiones de interacción con el receptor celular en σC

La región 156-326 mantiene su capacidad para inhibir competitivamente la entrada del reovirus aviar (Guardado-Calvo *et al.*, 2005), y por lo tanto su capacidad para interactuar con el receptor, lo que sugiere que en esta región que comprende el dominio globular de cabeza y dos repeticiones en triple β espiral están localizados todos los residuos necesarios para la interacción con el receptor. Para intentar especular sobre qué regiones podrían estar involucradas en esta interacción realizamos una comparación con la estructura de su homólogo de mamífero ($\sigma 1$) y analizamos el grado de conservación de los residuos de la superficie.

La región globular de la fibra del reovirus de mamífero $\sigma 1$ es capaz de interactuar con su receptor celular primario, JAM, a través del bucle DE (Chappell *et al.*, 2002). En la estructura del dominio de cabeza de σC el bucle DE es el único bucle largo del dominio globular y ocupa una posición similar a la observada en la estructura de $\sigma 1$, lo que sugiere que podría estar involucrado en la interacción con el receptor celular de σC . Un análisis de los residuos de este bucle muestra que el Asp259, Arg262, Leu263, Pro265, Gly268, Ph269, Gln270 y la Ala272 están totalmente conservados entre todas las cepas de reovirus aviar, aunque de todos ellos únicamente el Asp259 está totalmente expuesto al solvente. Tanto la similitud estructural con $\sigma 1$ como el alto grado de conservación de los residuos que componen este bucle hacen de él un buen candidato para ser el sitio de reconocimiento del receptor celular, aunque para que la mayoría de los residuos conservados puedan interactuar con el receptor sería necesario un cambio de conformación del bucle que exponga a estos residuos conservados al solvente.

Cuando localizamos los residuos universalmente conservados que además estén expuestos al solvente (Figura 7.3) en la estructura de $\sigma C(151-326)$, vemos que estos aparecen dispersos por toda su superficie a excepción de dos grupos: el primero está localizado en el límite entre la región globular y la fibrosa (justo debajo del bucle DE), y el segundo está situado entre dos monómeros. Ambas regiones son posibles candidatos a formar parte del sitio de reconocimiento del receptor celular. Si la región de unión al receptor estuviese situada en esta región entre los monómeros (Figura

7.3B) podría ser una interacción similar a la descrita para la interacción entre el dominio de cabeza de la fibra de adenovirus y su receptor primario CAR (Bewley *et al.*, 1999). Para poder determinar con precisión la localización de los residuos implicados en la interacción con el receptor, sería necesario o bien realizar estudios de mutagénesis dirigida o bien, en el momento en que se conozca el receptor celular para esta proteína, experimentos de *cross-linking* o co-cristalización.

7.3- Las fibras de los adenovirus aviar y porcino unen receptores desconocidos

7.3.1 - El receptor celular del adenovirus aviar de tipo 1

El adenovirus aviar de tipo 1 es capaz de infectar diferencialmente células CHO deficientes en CAR y células CHO que expresan CAR humano, lo que sugiere que al menos una de sus dos fibras es capaz de interaccionar con CAR humano. Los estudios basados en deleciones de la fibra larga (Tan *et al.*, 2001) han mostrado que esta es prescindible para la infección y responsable de la interacción con CAR humano, mientras que la fibra corta juega un papel clave en la infección y se une a un receptor que no se ha identificado hasta el momento. Como se ha discutido con anterioridad (Capítulo 1), el dominio D1 de CAR humano es necesario y suficiente para la interacción con los dominios de cabeza de algunos adenovirus humanos (Freimuth *et al.*, 1999; Lortat-Jacob *et al.*, 2001). Sorprendentemente no ha sido posible detectar ninguna interacción entre el dominio de cabeza de la fibra larga del adenovirus aviar 1 y este dominio de CAR ni por cromatografía de exclusión molecular ni por SPR (Capítulo 5).

El estudio del grado de conservación de los residuos habitualmente implicados en la interacción con el dominio D1 de CAR (Law & Davidson, 2005) ha revelado que pocos de los residuos implicados en la interacción con el dominio D1 de CAR humano están conservados en la fibra larga del adenovirus aviar. Así mismo, el análisis de la conformación de los bucles que generalmente participan en esta interacción muestra que la conformación del bucle DG en la fibra larga del adenovirus aviar es distinta e incompatible con una unión similar a la descrita entre los dominios de cabeza de adenovirus humanos que unen CAR y el dominio D1 de CAR (Bewley *et al.*, 1999; Seiradake *et al.*, 2006). Los datos experimentales presentados en este trabajo así como los análisis de la secuencia y la estructura del dominio de cabeza de la fibra larga del adenovirus aviar de tipo 1, muestran que ésta no es capaz de interaccionar con el dominio D1 de CAR humano de la misma manera que lo hacen otros adenovirus humanos. Esto sugiere que, o bien la fibra larga necesita del dominio D2 de CAR para establecer una interacción eficiente, o bien el dominio fibroso de esta fibra contiene los elementos necesarios para la interacción con CAR. Una tercera posibilidad que no podemos descartar es que esta fibra no sea capaz de unir CAR humano.

7.3.2 - El receptor celular del adenovirus porcino de tipo 4

El receptor celular del adenovirus porcino de tipo 4 no ha sido identificado todavía. Un vector compuesto por un adenovirus 5 humano en el que su fibra ha sido reemplazada por la fibra del aislado NADC-1 del adenovirus porcino de tipo 4 es capaz de entrar en células tumorales de glioma independientemente de CAR y CD46 (Paul *et al.*, 2008). Los análisis estructurales y de secuencia del dominio de cabeza (Capítulo 6) sugieren que éste no puede interaccionar con los receptores clásicos de adenovirus (CAR, CD46 y ácido siálico). Sin embargo, un análisis de la secuencia situada hacia el extremo C-terminal de la fibra muestra la existencia de una secuencia RGD y un dominio de galectina compuesto por dos dominios CRD. Otras fibras de adenovirus no tienen estos dominios C-terminales.

En muchos otros adenovirus, el motivo RGD se encuentra presente en el bucle flexible de la proteína pentamérica de la base, donde se ancla la fibra adenoviral. Se ha demostrado que esta secuencia RGD es la responsable de la interacción con integrinas, que actúan como receptores secundarios (Wickham *et al.*, 1993; Lindert *et al.*, 2009). Es posible que el adenovirus porcino de tipo 4 tenga la capacidad de

interaccionar con integrinas via este motivo RGD presente en su fibra (la secuencia de la proteína pentamérica de este virus no se conoce, por lo que no sabemos si tiene un motivo RGD). Otros adenovirus, como el adenovirus de ratón de tipo 1, también contienen un motivo RGD en su fibra (en este caso en el dominio de cabeza), que juega un papel importante en la infección probablemente interaccionando con integrinas α_v .

Con respecto a la cuestión del receptor primario, los experimentos de SPR y los experimentos de cristalografía demuestran que el dominio de galectina de la fibra del adenovirus porcino de tipo 4 es capaz de unir azúcares que contengan unidades de *N*-acetil-lactosamina y también reconoce unidades de lactosa aunque con menos afinidad. El dominio de galectina es capaz de unir tres unidades de *N*-acetil-lactosamina linealmente conectadas al mismo tiempo. De los carbohidratos que hemos estudiado, hemos encontrado que la tri-*(N*-acetil-lactosamina) es la que tiene mayor afinidad. La localización de este dominio de galectina en el extremo más C-terminal de la fibra sugiere que este dominio es proyectado lejos de la partícula viral haciendo posible que interaccione con un receptor celular. Por lo tanto, es posible que azúcares que contengan unidades de *N*-acetil-lactosamina en su estructura funcionen como receptores primarios para este virus. Las poly-*N*-acetillactosaminas pueden formar parte de glicoproteína y glicolípidos y generalmente funcionan como un esqueleto en el cual se producen modificaciones adicionales que comúnmente son específicas de cada tipo celular (Sasaki *et al.*, 1997). También cabe la posibilidad de que este dominio una con mayor afinidad algún carbohidrato complejo que no forme parte del *glycan array*.

7.4 - La primera estructura de una galectina repetida en tándem

Los estudios estructurales llevados a cabo con otras galectinas han mostrado el modelo de organización de algunas galectinas prototípicas. En estos modelos, las galectinas se autoasocian formando dímeros en los cuales los sitios canónicos de unión a carbohidratos se sitúan en lados opuesto del dímero haciendo difícil que ambos sitios pueden unirse cooperativamente a un mismo carbohidrato (Barondes *et al.*, 1994; Figura 7.4). El dímero de galectina prototípica ha sido usada como paradigma para explicar la organización de las galectinas repetidas en tándem (Zick *et al.*, 2004). Sin embargo, no puede explicar las evidencias experimentales que sugieren

que existe una interacción cooperativa entre los dos dominios y que el péptido que conecta ambos dominios es importante para su correcto funcionamiento (Levy *et al.*, 2006). Además, experimentos de *glycan array* llevados a cabo con otras galectinas repetidas en tándem apuntan hacia la necesidad de ambos dominios para reconocer azúcares específicos. Algunos estudios con la galectina 8 humana han mostrado que, aunque la especificidad de los dos dominios parece aditiva en solución, ambos actúan sinérgicamente al interactuar con azúcares unidos a superficies (Carlsson *et al.*, 2007).

En este trabajo hemos resuelto la estructura de la primera galectina repetida en tándem en donde están presentes ambos dominios de unión a carbohidratos. El modo en que ambos dominios están organizados no se parece al modo de organización de las galectinas prototípicas, sino que los dominios están orientados de tal forma que ambos pueden interactuar cooperativamente con un único azúcar (Figura 7.4). Además, la estructura en complejo con tri-(*N*-acetillactosamina) nos ha confirmado esto. Por lo tanto, la estructura que presentamos en este trabajo puede representar un nuevo paradigma en la organización de las galectinas repetidas en tándem y explicar la importancia que tiene el péptido que une los dominios CRD en la organización de ambos dominios.

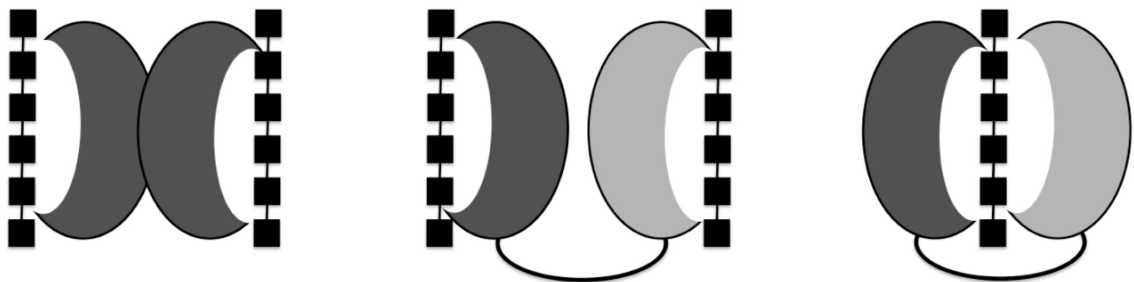


Figura 7.4 La organización de los dominios CRD en las galectinas. Organización en galectinas prototipo diméricas (izquierda), el actual modelo que se utiliza en galectinas repetidas en tandem (centro) y nuestro nuevo modelo obtenido a partir de las estructuras descritas en este trabajo.

7.5- Implicaciones en enfermedades porcinas y humanas

A pesar de que la mayoría de las infecciones causadas por adenovirus en cerdos tienen un carácter subclínico, el adenovirus porcino de tipo 4 ha sido asociado con encefalitis (Edington *et al.*, 1972). Los resultados presentados en este trabajo podrán ser usados para profundizar en la primeras etapas del ciclo replicativo de estos

virus y podría llevar al desarrollo de aplicaciones terapéuticas, bien dirigidas a prevenir la interacción fibra-receptor y por lo tanto impidiendo la infección, o bien explotando las propiedades de unión a receptores únicas de esta fibra en el diseño de vectores de vacunación.

El hecho de que el dominio de galectina sea el único dominio de esta clase de la familia *Adenoviridae* abre las puertas a su uso en el desarrollo de vectores adenovirales en posibles aplicación médicas. En un trabajo reciente se ha descrito que la incorporación de la fibra del adenovirus porcino 4 en vectores adenovirales ha permitido a estos infectar células que de otra manera son incapaces (Paul *et al.*, 2008). Estos vectores modificados son capaces de infectar células tumorales de glioma que carecen de receptores clásicos de adenovirus como CAR o CD46. La capacidad de infectar a estas células podría estar relacionada con el reconocimiento de azúcares de la superficie celular mediado por el dominio de galectina. En este sentido, las estructuras de los complejos galectina-azúcar han identificado los residuos clave en esta interacción, lo que abre la posibilidad a generar mutantes puntuales con distinta especificidad y de esta manera redirigir los vectores adenovirales a células que expresen en su superficie carbohidratos específicos de enfermedades humanas (Hakomori, 1984).

7.6 - Los dominios de σA tienen funciones diferenciadas

La proteína σA tiene dos funciones claramente diferenciadas en la biología del reovirus aviar. Por un lado, ejerce un papel estructural anclando la unidades de λA entre sí y de esta manera estabiliza la cápside interna e interacciona con la proteína μB promoviendo la correcta formación de la cápside externa durante la morfogénesis viral. Por otro lado es la proteína encargada de inhibir el sistema antiviral de las células promoviendo la inhibición de la proteína PKR. El análisis estructural de los monómeros de la proteína σA ha revelado la existencia de tres dominios estructuralmente diferenciados (Capítulo 4), los dos primeros tienen la misma topología y están plegados contra un tercer dominio con una topología distinta. Estos dominios están orientados en la partícula reoviral de tal forma que el tercer dominio es el responsable de la interacción con μB y por lo tanto de promover la formación de la cápside externa mientras que los dos primeros dominios serían los responsables de interaccionar con λA estabilizando la cápside interna (Figura 7.5).

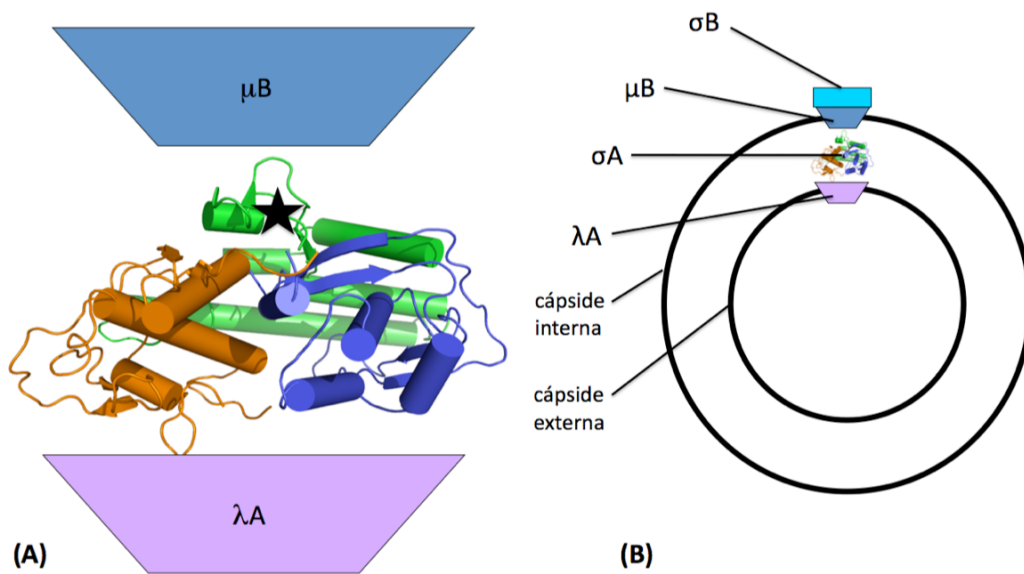


Figura 7.5 Los dos dominios N-terminales topológicamente iguales (azul y naranja) y el tercer dominio (verde) son los responsables de la interacción con λA y μB respectivamente (A). El sitio de interacción con el dsRNA se representa con una estrella negra. En (B) se presenta una visión esquemática de la orientación de σA en el contexto del reovirión maduro.

7.7 - La proteína σA en los reovirus fusogénicos

Los virus del género *Orthoreovirus* han sido subdivididos en dos subgrupos, fusogénicos y no fusogénicos, basándose en su capacidad de producir fusiones celulares y formar sincitios. El arquetipo de reovirus de mamífero es un reovirus no fusogénico, aunque tiene la capacidad de formar poros de manera específica en las membranas celulares, lo que constituye un paso esencial en la ruta de penetración de este virus a través de la membrana plasmática (Agosto *et al.*, 2006). El resto de los componentes del género, lo que incluye los reovirus aviares (Benavente *et al.*, 2007), reovirus de reptiles (Duncan *et al.*, 2004), reovirus de babuino (Dawe *et al.*, 2002), y otros reovirus de mamífero como el reovirus Nelson Bay (Duncan, 1999) y el reovirus Palau (Pritchard *et al.*, 2006) son fusogénicos. Se ha atribuido una serie de enfermedades a las infecciones con ortoreovirus en mamíferos, y recientemente ha sido publicado que el reovirus Melaka, que tiene su origen en los murciélagos y que es un reovirus de mamífero fusogénico, ha sido relacionado con una enfermedad respiratoria aguda en humanos (Chua *et al.*, 2007). Sin embargo, las infecciones con reovirus no fusogénicos son relativamente benignas, con muy raros casos en los que causan afecciones respiratorias superiores suaves o enteritis en niños.

Un análisis de las secuencias y de la capacidad de unión a dsRNA ha sido llevado a cabo en la proteína σB (y sus equivalentes) de reovirus fusogénicos. Estos estudios (Duncan, 1999; Wang *et al.*, 1996; Yin *et al.*, 2000) han mostrado que esta proteína, al contrario que su homólogo $\sigma 3$ de reovirus no fusogénicos, no tiene capacidad de unir dsRNA. Dado el hecho de que la inhibición de la PKR en el reovirus aviar S1133 viene producido por σA (Gonzalez-Lopez *et al.*, 2003), hicimos un análisis de las secuencias de sus homólogos en reovirus fusogénicos y no fusogénicos para ver si los residuos implicados en la interacción estaban conservados o no. De este análisis se desprende que la Arg273 está totalmente conservada en reovirus fusogénicos (tanto de mamíferos como aviares) mientras que en los reovirus no fusogénicos ha sido reemplazada por una treonina (Figura 7.6). La Arg155 está totalmente conservada en reovirus fusogénicos aviares y ha sido reemplazada por una glutamina en los reovirus fusogénicos de mamífero y por una glicina en los no fusogénicos. Este patrón de conservación sugiere que las proteínas homólogas de σA en reovirus fusogénicos mantienen su capacidad para unir dsRNA (y por lo tanto, en potencia, para inhibir la PKR). En cambio, en los reovirus no fusogénicos, las proteínas $\sigma 2$ no parecen tener esta capacidad, y parece que es la proteína $\sigma 3$ la encargada de desempeñar esta función. Recientemente ha sido descrito (Vazquez-Iglesias *et al.*, 2009) que las argininas implicadas en la interacción con el dsRNA también están implicadas en la localización nucleolar de esta proteína. Estas importantes diferencias podrían influenciar profundamente en la interacción célula-hospedador y podría estar relacionado y/o haber evolucionado junto con el modo sincitial de propagación de los reovirus fusogénicos y por lo tanto estar relacionado con su patogenicidad.

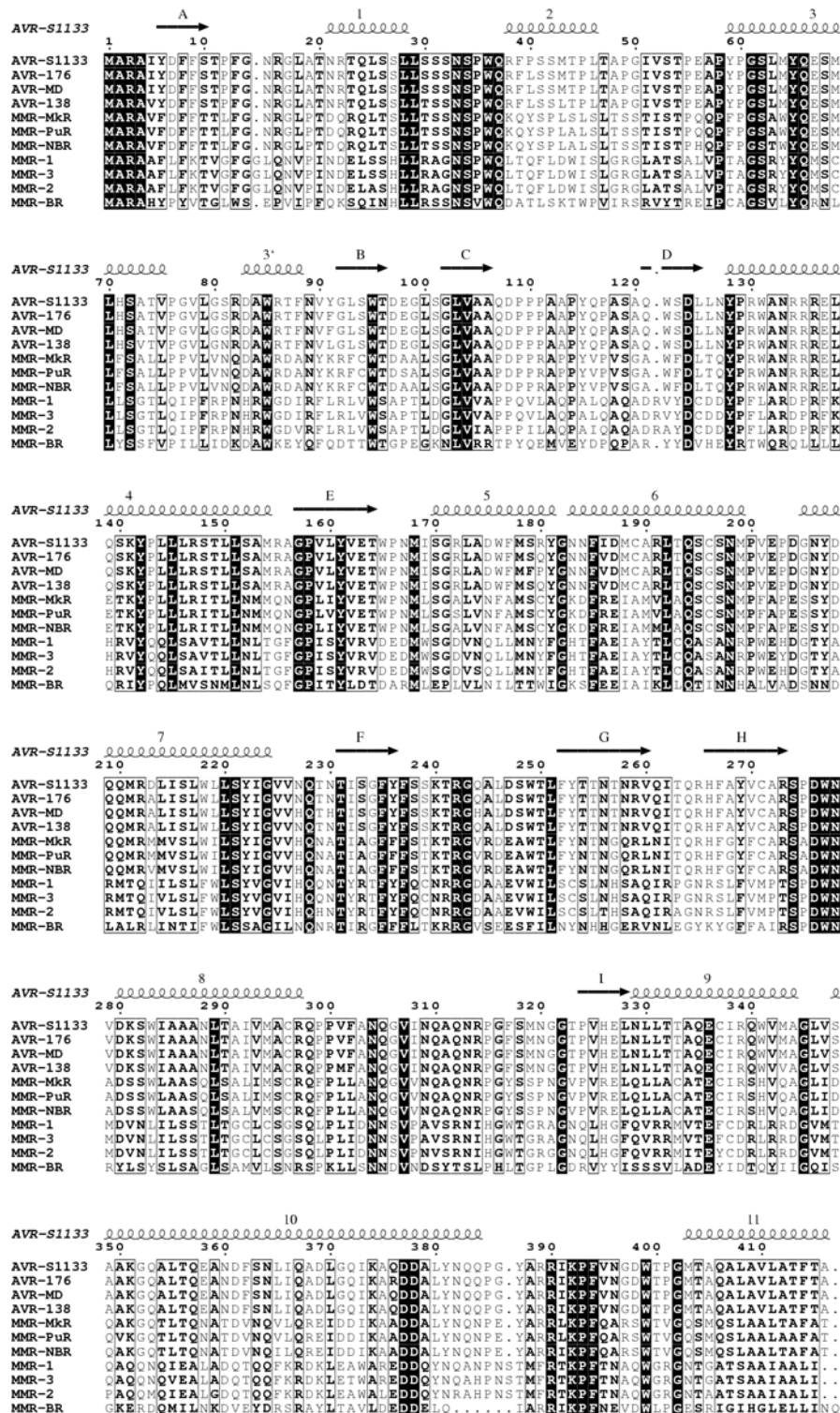


Figura 7.6 Alineamiento de las secuencias de σA de diversos reovirus. Se han usado las secuencias de los reovirus aviares fusogénicos S1133 (AVR-S1133), 176 (AVR-176), 138 (AVR-138) y Muscovy Duck (AVR-MD). Los reovirus de mamífero fusogénicos Melaka (MMR-MkR), Pulau (MMR-PuR), Nelson Bay (MMR-NBR) y Baboon (MMR-BR). Los reovirus de mamífero no fusogénicos de tipo 1 (MMR-1), 2 (MMR-2) y 3 (MMR-3).

7.8 Conclusión

En este trabajo se ha logrado caracterizar estructuralmente diversas proteínas del reovirus aviar S1133, el adenovirus aviar 1 y el adenovirus porcino 4.

1. Se cristalizaron dos fragmentos de la fibra del reovirus aviar S1133. A partir de la información estructural obtenida se han podido describir dos regiones potenciales de interacción con el receptor. Serían necesarios más estudios para determinar si realmente son las encargadas de esta interacción. Se han descrito posibles motivos de trimerización tanto en la región fibrosa como en la globular que difieren de su homóloga de mamífero $\sigma 1$.
2. Se ha cristalizado y resuelto la estructura de la proteína σA . A partir de su estructura se han podido determinar que residuos están implicados en la interacción con el dsRNA y se ha propuesto un modelo por el cual se explica su actividad anti-interferón.
3. Se ha cristalizado el dominio globular de cabeza de la fibra larga del adenovirus aviar 1. A partir de su estructura se ha determinado que este dominio no puede interaccionar con los receptores clásicos de adenovirus al menos de la misma forma en que lo hacen sus homólogos humanos previamente caracterizados.
4. Se han resuelto las estructuras del dominio globular de cabeza y de galectina del adenovirus porcino 4. La estructura del dominio globular de cabeza muestra que este no es capaz de interaccionar con los receptores clásicos de adenovirus. Un análisis del dominio de galectina muestra que este puede unir poli-*N*-acetillactosaminas. Se ha cocrystalizado con diversos azúcares y se han descrito los residuos implicados en la interacción.

REFERENCIAS

Abrahamsen K, Kong HL, Mastrangeli A, Brough D, Lizonova A, Crystal RG, Falck-Pedersen E. (1997) Construction of an adenovirus type 7a E1A- vector. *J Virol.* **71**:8946-51

Agosto, M. A., T. Ivanovic, and M. L. Nibert. (2006) Mammalian reovirus, a nonfusogenic nonenveloped virus, forms size-selective pores in a model membrane. *Proc. Natl. Acad. Sci. USA* **103**:16496-16501

Auld, D. S. (2005) Zinc enzymes. In *Encyclopedia of Inorganic Chemistry*, 2nd edn, vol. IX, pp. 5885–5927. Edited by R. B. King. Chichester, UK: John Wiley

Anderson CW, Young ME, Flint SJ. (1989) Characterization of the adenovirus 2 virion protein, mu. *Virology* **172**:506-12

Arnberg N (2009) Adenovirus receptors: implications for tropism, treatment and targeting. *Rev Med Virol* **19**: 165-178

Arnberg N, Kidd AH, Edlund K, Olfat F, Wadell G (2000) Initial interactions of subgenus D adenoviruses with A549 cellular receptors: sialic acid versus alpha(v) integrins. *J Virol* **74**: 7691-7693

Arnberg N, K. Edlund, A.H. Kidd and G. Wadell (2000) Adenovirus type 37 uses sialic acid as a cellular receptor. *J. Virol.* **74**:42–48

Arnberg, N., Kidd, A. H., Edlund, K., Nilsson, J., Pring-Akerblom, P. & Wadell, G. (2002) Adenovirus type 37 binds to cell surface sialic acid through a charge-dependent interaction. *Virology* **302**:33–43

Bachtarzi H, Stevenson M, Fisher K (2008) Cancer gene therapy with targeted adenoviruses. *Expert Opin Drug Deliv* **5**: 1231-1240

Bangari DS, Mittal SK. (2005) Porcine adenovirus serotype 3 internalization is independent of CAR and alphavbeta3 or alphavbeta5 integrin. *Virology* **332**:157-66

Barondes SH, Cooper DNW, Gitt MA, Leffler H (1994). Galectins. Structure and function of a large family of animal lectins. *J Biol Chem* **269**: 20807-20810

Barton ES, J.C. Forrest, J.L. Connolly, J.D. Chappell, Y. Liu, F.J. Schnell, A. Nusrat, C.A. Parkos and T.S. Dermody (2001) Junction adhesion molecule is a receptor for reovirus, *Cell* **104**:441–451

Benavente J & A.J. Shatkin (1988) Avian reovirus mRNAs are nonfunctional in infected mouse L cells: translational basis for virus host-range restriction. *Proc. Natl. Acad. Sci. U.S.A.* **85**:4257–4261

Benavente, J. & Martinez-Costas, J. (2007) Avian reovirus: structure and biology. *Virus Res* **123**, 105–119

Bergelson JM, J.A. Cunningham, G. Droguett, E.A. Kurt-Jones, A. Krithivas, J.S. Hong, M.S. Horwitz, R.L. Crowell and R.W. Finberg (1997) Isolation of a

common receptor for Coxsackie B viruses and adenoviruses 2 and 5, *Science* **275**:1320–1323

Bewley MC, Springer K, Zhang YB, Freimuth P, Flanagan JM (1999) Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**: 1579-1583

Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong CH, Paulson JC (2004) Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc Natl Acad Sci USA* **101**: 17033–17038.

Bodelon G, L. Labrada, J. Martínez-Costas and J. Benavente (2001) The avian reovirus genome segment S1 is a functionally tricistronic gene that expresses one structural and two nonstructural proteins in infected cells *Virology* **290**:181–191

Burmeister WP, Guilligay D, Cusack S, Wadell G, Arnberg N (2004) Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. *J Virol* **78**: 7727-7736

Burkhard, P., Meier, M. & Lustig, A. (2000) Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci* **9**, 2294–2301

Brünger AT (1997) Free R value: cross-validation in crystallography. *Methods Enzymol* **277**:366-96

Chang, K. Y., and A. Ramos. (2005) The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J.* **272**:2109-2117

Chappell, J. D., Protá, A. E., Dermody, T. S. & Stehle, T. (2002) Crystal structure of reovirus attachment protein $\sigma 1$ reveals evolutionary relationship to adenovirus fiber. *EMBO J* **21**:1–11

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* **31**; 3497-500

Chrétien L, Richard DE, Poirier SN, Poitras M, Guillemette G. (1998) Bovine adrenal glomerulosa cells express such a low level of functional B2 receptors that bradykinin does not significantly increase their aldosterone production. *J Endocrinol* **156**:449-60

Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. & Cotten, M. (1996) The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J Virol* **70**:2939–2949

Chua, K. B., G. Cramer, A. Hyatt, M. Yu, M. R. Tompong, J. Rosli, J. McEachern, S. Cramer, V. Kumarasamy, B. T. Eaton, and L. F. Wang. (2007) A

previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proc. Natl. Acad. Sci. USA* **104**:11424-11429

Clemens M (1997) PKR—a protein kinase regulated by double-stranded RNA. *Int. J. Biochem. Cell Biol.* **29**:945–949

Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763

Connors, R., Hill, D. J., Borodina, E., Agnew, C., Daniell, S. J., Burton, N. M., Sessions, R. B., Clarke, A. R., Catto, L. E. & other authors (2008) The *Moraxella* adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. *EMBO J* **27**:1779–1789

Cooper DN, Boulianne RP, Charlton S, Farrell EM, Sucher A, Lu BC (1997) Fungal galectins, sequence and specificity of two isolectins from *Coprinus cinereus*. *J Biol Chem* **272**: 1514-1521

Cooper DN, Barondes SH (1999) God must love galectins; He made so many of them. *Glycobiology* **9**: 979-984

Cooper DN (2002) Galectinomics: finding themes in complexity. *Biochim Biophys Acta* **1572**: 209-231

Costas C, J. Martínez-Costas, G. Bodelón and J. Benavente (2005) The second open reading frame of the avian reovirus S1 gene encodes a transcription-dependent and CRM1-independent nucleocytoplasmic shuttling protein. *J. Virol.* **79**:2141–2150

Cowen, B., Calnek, B. W., Menendez, N. A. & Ball, R. F. (1978) Avian adenoviruses: effect on egg production, shell quality, and feed consumption. *Avian Dis* **22**:459–470

Cowtan, K. & Main, P. (1998) Miscellaneous algorithms for density modification. *Acta Crystallogr D Biol Crystallogr* **54**:487–493

Crawford-Miksza, L. & Schnurr, D. P. (1996) Analysis of 15 adenovirus hexon proteins reveals the location and structure of seven hypervariable regions containing serotype-specific residues. *J Virol* **70**:1836–1844

Cummings RD, Liu F (2009) S-type Lectins (Galectins). In *Essentials of Glycobiology (Second Edition)*, Varki A, Cummings RD, Esko JD, Freeze HH, Hart GW, Marth J (ed) pp. 475-479. Cold Spring Harbor Laboratory Press

Davidson BL, Bohn MC. (1997) Recombinant adenovirus: a gene transfer vector for study and treatment of CNS diseases. *Exp Neurol.* **144**:125-30

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, Richardson DC (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**: W375-W383

Davison AJ, Benko M, Harrach B. (2003) Genetic content and evolution of adenoviruses. *J Gen Virol* **84**:2895-908

Dawe, S., and R. Duncan. (2002) The S4 genome segment of baboon reovirus is bicistronic and encodes a novel fusion-associated small transmembrane protein. *J. Virol.* **76**:2131-2140

Dormitzer PR, Sun ZY, Blixt O, Paulson JC, Wagner G, Harrison SC. (2002) Specificity and affinity of sialic acid binding by the rhesus rotavirus VP8* core. *J Virol.* **76**:10512-7

Duncan, R. (1999) Extensive sequence divergence and phylogenetic relationships between the fusogenic and nonfusogenic orthoreoviruses: a species proposal. *Virology* **260**:316-328

Duncan R (1996) The low pH-dependent entry of avian reovirus is accompanied by two specific cleavages of the major outer capsid protein μ 2C. *Virology* **219**:179–189

Duncan R, J. Corcoran, J. Shou and D. Stoltz (2004) Reptilian reovirus: a new fusogenic *Orthoreovirus* species, *Virology* **319**:131–140

Duncan R & K. Sullivan (1998) Characterization of two avian reoviruses that exhibit strain-specific quantitative differences in their syncytium-inducing and pathogenic capabilities. *Virology* **250**:263–272

Durmort, C., Stehlin, C., Schoehn, G., Mitraki, A., Drouet, E., Cusack, S. & Burmeister, W. P. (2001) Structure of the fiber head of Ad3, a non-CAR-binding serotype of adenovirus. *Virology* **285**:302–312

Edington N, Kasza L, Christofinis GJ (1972) Meningo-encephalitis in gnotobiotic pigs inoculated intranasally and orally with porcine adenovirus 4. *Res Vet Sci* **13**: 289-291

El Bakkouri M, Seiradake E, Cusack S, Ruigrok RW, Schoehn G (2008) Structure of the C-terminal head domain of the fowl adenovirus type 1 short fibre. *Virology* **378**: 169-176

Ellis MN, C.S. Eidson, J. Brown and S.H. Kleven (1983) Studies of interferon induction and interferon sensitivity of avian reoviruses, *Avian Dis.* **27**:927–936

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132

Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model* **15**:132–134

Evans PR (1999) Some notes on choices in data collection, *Acta Crystallog. sect. D* **55**:1771–1772

Fabry, C. M., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2005) A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J* **24**:1645–1654

Fass, D., Harrison, S. C. & Kim, P. S. (1996) Retrovirus envelope domain at 1.7 Å resolution. *Nat Struct Biol* **3**, 465–469

Fechner, H., Haack, A., Wang, H., Wang, X., Eizema, K., Pauschinger, M., Schoemaker, R., Veghel, R., Houtsmuller, A. & other authors (1999) Expression of coxsackie adenovirus receptor and alpha_v-integrin does not correlate with adenovector targeting in vivo indicating anatomical vector barriers. *Gene Ther* **6**:1520–1535

Fields BN , C.S. Raine and S.G. Baum (1971) Temperature-sensitive mutants of reovirus type 3: defects in viral maturation as studied by immunofluorescence and electron microscopy, *Virology* **43**:569–578

Francois, A., Chevalier, C., Delmas, B., Eterradossi, N., Toquin, D., Rivallan, G. & Langlois, P. (2004) Avian adenovirus CELO recombinants expressing VP2 of infectious bursal disease virus induce protection against bursal disease in chickens. *Vaccine* **22**:2351–2360

Fraser, R. D. B. & MacRae, T. P. (1973) *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*. New York, USA: Academic Press.

Freimuth, P., Springer, K., Berard, C., Hainfeld, J., Bewley, M. & Flanagan, J. (1999) Coxsackievirus and adenovirus receptor amino-terminal immunoglobulin V-related domain binds adenovirus type 2 and fiber knob from adenovirus type 12. *J Virol* **73**:1392–1398

Gaggar A, Shayakhmetov DM, Lieber A (2003) CD46 is a cellular receptor for group B adenoviruses. *Nat Med* **9**: 1408-1412

Garcia de la Torre, J., M. L. Huertas, and B. B. Carrasco. (2000) Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**:719-730

Gouet P, Courcelle E, Stuart DI, Métoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**: 305-308

González-López C, J. Martínez-Costas, M. Esteban and J. Benavente (2003) Evidence that avian reovirus σA protein is an inhibitor of the double-stranded RNA-dependent protein kinase. *J. Gen. Virol.* **84**:1629–1639

Goodbourn S, L. Didcock and R.E. Randall (2000) Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. *J. Gen. Virol.* **81**:2341–2364

- Glasgow JN, Kremer EJ, Hemminki A, Siegal GP, Douglas JT, Curiel DT.** (2004) An adenovirus vector with a chimeric fiber derived from canine adenovirus type 2 displays novel tropism. *Virology* **324**:103-16
- Graham FL, Prevec L** (1995) Methods for construction of adenovirus vectors. *Mol Biotechnol* **3**: 207-220
- Grande A and J. Benavente** (2000) Optimal conditions for the growth, purification and storage of the avian reovirus S1133. *J. Virol. Methods* **85**:43–54
- Grande A, C. Costas and J. Benavente** (2002) Subunit composition and conformational stability of the oligomeric form of the avian reovirus cell-attachment protein σ_C . *J. Gen. Virol.* **83**:131–139
- Grande A, E. Rodriguez, C. Costas, E. Everitt and J. Benavente** (2000) Oligomerization and cell-binding properties of the avian reovirus cell-attachment protein sigmaC. *Virology* **274**:367–377
- Green MR, Roeder RG.** (1980) Definition of a novel promoter for the major adenovirus-associated virus mRNA. *Cell* **22**:231-42.
- Guardado-Calvo, P., Fox, G. C., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martinez-Costas, J. M., Benavente, J. & van Raaij, M. J.** (2005) Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre σ_C . *J Mol Biol* **354**:137–149
- Guardado-Calvo, P., Llamas-Saiz, A. L., Langlois, P. & van Raaij, M. J.** (2006) Crystallization of the C-terminal head domain of the avian adenovirus CELO long fibre. *Acta Crystallograph Sect F Struct Biol Cryst Commun* **62**:449–452
- Guardado Calvo P, G.C. Fox, X.L. Hermo Parrado, A.L. Llamas-Saiz, C. Costas, J. Martínez-Costas, J. Benavente and M.J. van Raaij** (2005) Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J. Mol. Biol.* **354**:137–149
- Guardado-Calvo P, Llamas-Saiz AL, Fox GC, Langlois P, van Raaij MJ** (2007) Structure of the C-terminal head domain of the fowl adenovirus type 1 long fiber. *J Gen Virol* **88**: 2407-2416
- Guglielmi, K. M., Johnson, E. M., Stehle, T. & Dermody, T. S.** (2006) Attachment and cell entry of mammalian orthoreovirus. *Curr Top Microbiol Immunol* **309**:1–38
- Haig DA, Clarke MC, Pereira MS** (1964) Isolation of an adenovirus from a pig. *J Comp Pathol* **74**: 81-84
- Hakomori S** (1984) Tumor-associated carbohydrate antigens. *Annu Rev Immunol* **2**: 103-126
- Haller O, G. Kochs and F. Weber** (2006) The interferon response circuit: Induction and suppression by pathogenic viruses. *Virology* **344**:119–130

- Hammond JM, Johnson MA** (2005) Porcine adenovirus as a delivery system for swine vaccines and immunotherapeutics. *Vet J* **169**: 17-27
- Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T.** (1993) A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**:1401–1407
- Henrick, K. & Thornton, J. M.** (1998) PQS: a protein quaternary structure file server. *Trends Biochem Sci* **23**:358–361
- Hess, M., Cuzange, A., Ruigrok, R. W., Chroboczek, J. & Jacrot, B.** (1995) The avian adenovirus penton: two fibres and one base. *J Mol Biol* **252**:379–385
- Hermo-Parrado, X. L., P. Guardado-Calvo, A. L. Llamas-Saiz, G. C. Fox, L. Vazquez-Iglesias, J. Martinez-Costas, J. Benavente, and M. J. van Raaij.** (2007) Crystallization of the avian reovirus double-stranded RNA-binding and core protein sigmaA. *Acta Crystallogr. F* **63**:426-429
- Hilleman MR, Werner JH.** (1954) Recovery of new agent from patients with acute respiratory illness. *Proc Soc Exp Biol Med* **85**:183-8
- Hirabayashi J, Kasai K** (1993) The family of metazoan metal-independent beta-galactoside-binding lectins: structure, function and molecular evolution. *Glycobiology* **3**: 297-304
- Hirabayashi J, Hashidate T, Arata Y, Nishi N, Nakamura T, Hirashima M, Urashima T, Oka T, Futai M, Muller WE, Yagi F, Kasai K** (2002) Oligosaccharide specificity of galectins: a search by frontal affinity chromatography. *Biochim Biophys Acta* **1572**: 232-254
- Holm L and C. Sander** (1998) Touring protein fold space with Dali/FSSP. *Nucl. Acids Res.* **26**:316–319
- Hosokawa K, Sung MT.** (1976) Isolation and characterization of an extremely basic protein from adenovirus type 5. *J Virol* **17**:924-34
- Hutchinson G and J.M. Thornton** (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**:2207–2216
- Hsiao J, J. Martinez-Costas, J. Benavente and V.N. Vakharia** (2002) Cloning, expression, and characterization of avian reovirus guanylyltransferase, *Virology* **296**:288–299
- Imani, F., and B. L. Jacobs.** (1988) Inhibitory activity for the interferon-induced protein kinase is associated with the reovirus serotype 1 sigma 3 protein. *Proc. Natl. Acad. Sci. USA* **85**:7887-7891
- Jackson RJ & A. Kaminski** (1995) Internal initiation of translation in eukaryotes: the picornavirus paradigm and beyond. *RNA* **1**:985–1000

Jawetz E, Hanna L, Sonne M, Thygeson P. (1959) A laboratory infection with adenovirus type 8; laboratory and epidemiologic observations. *Am J Hyg.* **69**:13-20

Jones, R. C. (2000) Avian reovirus infections. *Rev. Sci. Technol.* **19**:614-625

Jones, S. & Thornton, J. M. (1996) Principles of protein–protein interactions. *Proc Natl Acad Sci U S A* **93**:13–20

Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr A* **47**:110–119

Kammerer, R. A., Kostrewa, D., Zurdo, J., Detken, A., Garcia-Echevarria, C., Green, J. D., Muller, S. A., Meier, B. H., Winkler, F. K. & Dobson, C. M. (2004) Exploring amyloid formation by a *de novo* design. *Proc Natl Acad Sci U S A* **101**:4435–4440

Kammerer, R. A., Kostrewa, D., Progius, P., Honnappa, S., Avila, D., Lustig, A., Winkler, F. K., Pieters, J. & Steinmetz, M. O. (2005) A conserved trimerization motif controls the topology of short coiled coils. *Proc Natl Acad Sci U S A* **102**:13891–13896

Kant A, F. Balk, L. Born, D. van Roozelaar, J. Heijmans, A. Gielkins and A. ter Huurne. (2003) Classification of Dutch and German avian reoviruses by sequencing the σ C protein, *Vet. Res.* **34**:203–212

Kasai K, Hirabayashi J (1996) Galectins: a family of animal lectins that decipher glycocodes. *J Biochem* **119**: 1-8

Kelleher, Z. T. & Vos, J. M. (1994) Long-term episomal gene delivery in human lymphoid cells using human and avian adenoviral-assisted transfection. *Biotechniques* **17**:1110–1117

Kidd, A. H., Chroboczek, J., Cusack, S. & Ruigrok, R. W. (1993) Adenovirus type 40 virions contain two distinct fibres. *Virology* **192**:73–84

Kim M, Zinn KR, Barnett BG, Sumerel LA, Krasnykh V, Curiel DT, Douglas JT. (2002) The therapeutic efficacy of adenoviral vectors for cancer gene therapy is limited by a low level of primary adenovirus receptors on tumour cells. *Eur J Cancer* **38**:1917-26

Kirby, I., Davison, E., Bevil, A. J., Soh, C. P., Wickham, T. J., Roelvink, P. W., Kovsdi, I., Sutton, B. J. & Santis, G. (1999) Mutations in the DG loop of adenovirus type 5 fiber knob protein abolish high-affinity binding to its cellular receptor CAR. *J Virol* **73**:9508–9514

Kleiboeker SB, Seal BS, Mengeling WL (1993) Genomic cloning and restriction site mapping of a porcine adenovirus isolate: demonstration of genomic stability in porcine adenovirus. *Arch Virol* **133**: 357-368

Kleiboeker SB (1995) Sequence analysis of the fiber genomic region of a porcine adenovirus predicts a novel fiber protein. *Virus Res* **39**: 299-309

Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001) Around O. In *International Tables for Crystallography, vol. F. Crystallography of Biological Macromolecules*, pp. 353–356. Edited by M. G. Rossmann & E. Arnold. Dordrecht, NL: Kluwer Academic Publishers

Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**:19867–19870

Kozak M (2002) Pushing the limits of the scanning mechanism for initiation of translation. *Gene* **299**:1–34

Krasnykh V, Belousova N, Korokhov N, Mikheeva G, Curiel DT (2001). Genetic targeting of an adenovirus vector via replacement of the fiber protein with the phage T4 fibritin. *J Virol* **75**: 4176-83

Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**:946–950

Krissinel E, Henrick K (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst D* **60**: 2256-2268

Krissinel E, Henrick K (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**: 774-797

Kuntz-Simon G, G. Le Gall-Recule, C. de Boisseson and V. Jestin (2002) Muscovy duck reovirus sigmaC protein is atypically encoded by the smallest genome segment. *J. Gen. Virol.* **83**:1189–1200

Labrada L, G. Bodelon, J. Vinuela and J. Benavente (2002) Avian reoviruses cause apoptosis in cultured cells: viral uncoating but not viral gene expression is required for apoptosis induction. *J. Virol.* **76**:7932–7941

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680-685

Lamzin, V. S. & Wilson, K. S. (1993) Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr* **49**:129–149

Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst* **26**: 283-291

Laver, W. G., Youngusband, H. B. & Wrigley, N. G. (1971) Purification and properties of chick embryo lethal orphan virus (an avian adenovirus). *Virology* **45**:598–614

- Law LK, Davidson BL.** (2002) Adenovirus serotype 30 fiber does not mediate transduction via the coxsackie-adenovirus receptor. *J Virol.* **76**:656-61
- Law LK, Davidson BL** (2005) What does it take to bind CAR? *Mol Ther* **12**: 599-609
- Leslie AG** (1999) Integration of macromolecular diffraction data, *Acta Crystallog. sect. D* **55**:1696–1702
- Levy Y, Auslender S, Eisenstein M, Vidavski RR, Ronen D, Bershadsky AD, Zick Y** (2006) It depends on the hinge: a structure-functional analysis of galectin-8, a tandem-repeat type lectin. *Glycobiology* **16**: 463-476
- Li E, Brown SL, Stupack DG, Puente XS, Cheresch DA, Nemerow GR** (2001) Integrin alpha v beta 1 is an adenovirus coreceptor. *J Virol* **75**: 5405-5409
- Liao DI, Kapadia G, Ahmed H, Vasta GR, Herzberg O** (1994) Structure of S-lectin, a developmentally regulated vertebrate -galactoside-binding protein. *Proc Natl Acad Sci USA* **91**: 1428-1432
- Liao M, Z. Xie, J. Liu, X. Deng, Y. Pang, Z. Xie and X. Tang** (2004) Cloning and sequence analysis of sigma3 protein gene of avian reovirus isolates. *Chin. J. Prev. Vet. Med.* **26**:324–327
- Liemann S, K. Chandran, T.S. Baker, M.L. Nibert and S.C. Harrison** (2002) Structure of the reovirus membrane-penetration protein, Mu1, in a complex with its protector protein, Sigma3. *Cell* **108**:283–295
- Liu HJ, P.-Y. Lin, J.-W. Lee and W.-L. Shih** (2005) Retardation of cell growth by avian reovirus p17 through activation of p53 pathway, *Biochem. Biophys. Res. Commun.* **336**:709–715
- Liszewski MK, Kemper C, Price JD, Atkinson JP.** (2005) Emerging roles and new functions of CD46. *Springer Semin Immunopathol.* **27**:345-58
- Lobsanov YD, Gitt MA, Leffler H, Barondes SH, Rini JM** (1993) X-ray crystal structure of the human dimeric S-Lac lectin L-14-II, in complex with lactose at 2.9 Å resolution. *J Biol Chem* **268**: 27034-27038
- Logunov, D. Y., Ilyinskaya, G. V., Cherenova, L. V., Verhovskaya, L. V., Shmarov, M. M., Chumakov, P. M., Kopnin, B. P. & Naroditsky, B. S.** (2004) Restoration of p53 tumor-suppressor activity in human tumor cells *in vitro* and in their xenografts *in vivo* by recombinant avian adenovirus CELO-p53. *Gene Ther* **11**:79–84
- Lortat-Jacob H, Chouin E, Cusack S, van Raaij MJ** (2001) Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor-ligand interactions. *J Biol Chem* **276**: 9009-9015

- Lovell SC, Davis IW, Arendall WB 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC** (2003) Structure validation by C-alpha geometry: phi, psi and C-beta deviation. *Proteins* **50**: 437-50
- Lumb, K. J. & Kim, P. S.** (1995) A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**:8642–8648
- Lupas, A., Van Dyke, M. & Stock, J.** (1991) Predicting coiled coils from protein sequences. *Science* **252**:1162–1164
- Lupiani B, K. Subramanian and S.K. Samal** (1995) Aquareoviruses. *Annu. Rev. Fish Dis.* **5**:175–208
- Lupyan D, Leo-Macias A, Ortiz AR** (2005) A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics* **21**: 3255-3263
- Luo BHV, C.V. Carman and T.A. Springer** (2007) Structural basis of integrin regulation and signaling. *Ann. Rev. Immunol.* **25**:619–647
- Maizel Jr JV, White DO, Scharff MD** (1968) The polypeptides of adenovirus. I. Evidence for multiple protein components in the virion and a comparison of types 2, 7A, and 12. *Virology* **36**: 115-125
- Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milholle, M. A., Pang, J. X. & Kim, P. S.** (1999) Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9 Å resolution. *Proc Natl Acad Sci U S A* **96**:2662–2667
- Mallo M, J. Martínez-Costas and J. Benavente** (1991) Avian reovirus S1133 can replicate in mouse L cells: effect of pH and cell attachment status on viral infection, *J. Virol.* **65**:5499–5505
- Martínez-Costas J, R. Varela and J. Benavente** (1995) Endogenous enzymatic activities of the avian reovirus S1133: identification of the viral capping enzyme. *Virology* **206**:1017–1026
- Martínez-Costas J, C. González-López, V.N. Vakharia and J. Benavente** (2000) Possible involvement of the double-stranded RNA-binding core protein sigmaA in the resistance of avian reovirus to interferon. *J. Virol.* **74**:1124–1131
- Martínez-Costas J, A. Grande, R. Varela, C. García-Martínez and J. Benavente** (1997) Protein architecture of avian reovirus S1133 and identification of the cell attachment protein *J. Virol.* **71**:59–64
- Marttila M, Persson D, Gustafsson D, Liszewski MK, Atkinson JP, Wadell G, Arnberg N** (2005) CD46 is a cellular receptor for all species B adenoviruses except types 3 and 7. *J Virol* **79**: 14429-14436
- Mastrangeli A, Harvey BG, Yao J, Wolff G, Kovesdi I, Crystal RG, Falck-Pedersen E.** (1996) "Sero-switch" adenovirus-mediated in vivo gene transfer:

circumvention of anti-adenovirus humoral immune defenses against repeat adenovirus vector administration by changing the adenovirus serotype. *Hum Gene Ther.* **7**:79-87

Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**:491-497

Meier, M., Lustig, A., Aebi, U. & Burkhard, P. (2002) Removing an interhelical salt bridge abolishes coiled-coil formation in a *de novo* designed peptide. *J Struct Biol* **137**:65–72

Mertens P (2004) The dsRNA viruses *Virus Res.* **101**:3–13

McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**:458–464

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674

McCracken, R. M. & Adair, B. M. (1993) Avian adenoviruses. In *Viral Infections of Vertebrates*, vol. **3**:123–144

Merckel, M. C., Huiskonen, J. T., Bamford, D. H., Goldman, A. & Tuma, R. (2005) The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* **18**:161–170

Mitraki, A., Miller, S. & van Raaij, M. J. (2002) Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. *J Struct Biol* **137**:236–247

Mitraki A, A. Barge, J. Chroboczek, J.P. Andrieu, J. Gagnon and R.W. Ruigruik (1999) Unfolding studies of human adenovirus type 2 fibre trimers. Evidence for a stable domain, *Eur. J. Biochem.* **264**:599–606

Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003) ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**:229–244

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255

Nagae M, Nishi N, Murata T, Usui T, Nakamura T, Wakatsuki S, Kato R (2006) Crystal structure of the galectin-9 N-terminal carbohydrate recognition domain from *Mus musculus* reveals the basic mechanism of carbohydrate recognition. *J Biol Chem* **281**: 35884-35893

Nagae M, Nishi N, Nakamura-Tsuruta S, Hirabayashi J, Wakatsuki S, Kato R (2008) Structural analysis of the human galectin-9 N-terminal carbohydrate recognition domain reveals unexpected properties that differ from the mouse orthologue. *J Mol Biol* **375**: 119-135

- Nagae M, Nishi N, Murata T, Usui T, Nakamura T, Wakatsuki S, Kato R** (2009) Structural analysis of the recognition mechanism of poly-N-acetyllactosamine by the human galectin-9 N-terminal carbohydrate recognition domain. *Glycobiology* **19**: 112-117
- Nanduri, S., B. W. Carpick, Y. Yang, B. R. Williams, and J. Qin.** (1998) Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.* **17**:5458-5465
- Navaza, J.** (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**:1367–1372
- Nemerow GR, Pache L, Reddy V, Stewart PL** (2009) Insights into adenovirus host cell interactions from structural studies. *Virology* **384**: 380-388
- Nemerow GR, Stewart PL.** (1999) Role of alpha(v) integrins in adenovirus cell entry and gene delivery. *Microbiol Mol Biol Rev* **63**:725-34
- Nevins JR.** The Rb/E2F pathway and cancer. (2001) *Hum Mol Genet.* **10**:699-703.
- Nibert ML and Schiff LA.** (2001) Reoviruses and their replication. In: D.M. Knipe and P.M. Hooley, Editors, *Fields Virology* (fourth ed.), Lippincott Williams & Wilkins, Philadelphia, PA 1679–1728
- Nicklin SA, Baker AH** (2008) Efficient vascular endothelial gene transfer following intravenous adenovirus delivery. *Mol Ther* **16**: 1904-1905
- Oakley, M. G. & Kim, P. S.** (1998) A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* **37**:12603–12610
- Okada N, Saito T, Masunaga Y, Tsukada Y, Nakagawa S, Mizuguchi H, Mori K, Okada Y, Fujita T, Hayakawa T, Mayumi T, Yamamoto A.** (2001) Efficient antigen gene transduction using Arg-Gly-Asp fiber-mutant adenovirus vectors can potentiate antitumor vaccine efficacy and maturation of murine dendritic cells. *Cancer Res.* **61**:7913-9
- Olland AM, J. Jane-Valbuena, L.A. Schiff, M.L. Nibert and S.C. Harrison** (2001) Structure of the reovirus outer capsid and dsRNA-binding protein sigma3 at 1.8 Å resolution, *EMBO J.* **20**:979–989
- Otwinowski, Z. & Minor, W.** (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**:307–326
- Pache L, S. Venkataraman, V.S. Reddy and G.R. Nemerow** (2008) Structural variations in species B adenovirus fibers impact CD46 association. *J. Virol.* **82**:7923–7931
- Paul CP, Everts M, Glasgow JN, Dent P, Fisher PB, Ulasov IV, Lesniak MS, Stoff-Khalili MA, Roth JC, Preuss MA, Dirven CM, Lamfers ML, Siegal GP,**

Zhu ZB, Curiel DT (2008) Characterization of infectivity of knob-modified adenoviral vectors in glioma. *Cancer Biol Ther* **7**: 786-793

Papanikolopoulou, K., Teixeira, S., Belrhali, H., Forsyth, V. T., Mitraki, A. & van Raaij, M. J. (2004) Adenovirus fibre shaft sequences fold into the native triple β -spiral fold when N-terminally fused to the bacteriophage T4 fibritin foldon trimerisation motif. *J Mol Biol* **342**:219–227

Parker JSL, T.J. Broering, J. Kim, D.E. Higgins and M.L. Nibert (2002) Reovirus core protein $\mu 2$ determines the filamentous morphology of viral inclusion bodies by interacting with and stabilizing microtubules. *J. Virol.* **76**:4483–4496

Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**:458–463

Persson BD, Reiter DM, Marttila M, Mei YF, Casasnovas JM, Arnberg N, Stehle T (2007) Adenovirus type 11 binding alters the conformation of its receptor CD46. *Nat Struct Mol Biol* **14**: 164-166

Piliarik M, Vaisocherova H, Homola J (2009) Surface plasmon resonance biosensing. *Methods Mol Biol* **503**: 65-88

Potterton L, McNicholas S, Krissinel E, Gruber J, Cowtan K, Emsley P, Murshudov GN, Cohen S, Perrakis A, Noble M (2004) Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **60**: 2288-2294

Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallog. sect. D* **55**:1690–1695

Pritchard, L. I., K. B. Chua, D. Cummins, A. Hyatt, G. Cramer, B. T. Eaton, and L. F. Wang. (2006) Pulau virus; a new member of the Nelson Bay orthoreovirus species isolated from fruit bats in Malaysia. *Arch. Virol.* **151**:229-239

Pruitt, K. D., Tatusova, T. & Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**:D501–D504

Raman S, Hsu TH, Ashley SL, Spindler KR (2009) Usage of integrin and heparan sulfate as receptors for mouse adenovirus type 1. *J Virol* **83**: 2831-2838

Raper SE, N. Chirmule, F.S. Lee, N.A. Wivel, A. Bagg, G. -P. Gao, J.M. Wilson and M.L. Batshaw. (2003) Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Molec. Genet. Metab.* **80**:148–158

Read RJ (1986) Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst* **42**: 140-149

Reinisch KM, M.L. Nibert and S.C. Harrison (2000) Structure of the reovirus core at 3.6 Å resolution, *Nature* **404**:960–967

- Rekosh, D. M., Russell, W. C., Bellet, A. J. & Robinson, A. J.** (1977) Identification of a protein linked to the ends of adenovirus DNA. *Cell* **11**:283–295
- Rekhter MD, Simari RD, Work CW, Nabel GJ, Nabel EG, Gordon D.** (1998) Gene transfer into normal and atherosclerotic human blood vessels. *Circ Res.* **82**:1243-52
- Renard, M., Varela, P. F., Letzelter, C., Duquerroy, S., Rey, F. A. & Heidmann, T.** (2005) Crystal structure of a pivotal domain of human syncytin-2, a 40 million years old endogenous retrovirus fusogenic envelope gene captured by primates. *J Mol Biol* **352**:1029–1034
- Rhim JS, L.E. Jordan and H.D. Mayor** (1962) Cytochemical, fluorescent-antibody and electron microscopic studies on the growth of reovirus (ECHO 10) in tissue culture, *Virology* **17**:342–355
- Richardson, J. S.** (1981) The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**:167–339
- Roberts RJ, O'Neill KE, Yen CT.** (1984) DNA sequences from the adenovirus 2 genome. *J Biol Chem* **259**:13968-75
- Robertson MD & G.E. Wilcox** (1986) Avian reovirus. *Vet. Bull.* **56**:759–766
- Roelvink PW, Lizonova A, Lee JG, Li Y, Bergelson JM, Finberg RW, Brough DE, Kovesdi I, Wickham TJ** (1998) The coxsackievirus-adenovirus receptor protein can function as a cellular attachment protein for adenovirus serotypes from subgroups A, C, D, E, and F. *J Virol* **72**: 7909-7915
- Roelvink, P. W., Mi Lee, G., Einfeld, D. A., Kovesdi, I. & Wickham, T. J.** (1999) Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. *Science* **286**:1568–1571
- Rosenberger JK, F.J. Sterner, S. Botts, K.P. Lee and A. Margolin** (1989) In vitro and in vivo characterization of avian reoviruses. I. Pathogenicity and antigenic relatedness of several avian reovirus isolates, *Avian Dis.* **33**:535–544
- Rosen L.** (1960) A hemagglutination-inhibition technique for typing adenoviruses. *Am J Hyg.* **71**:120-8
- Russell IA, Royds JA, Braithwaite AW.** (2004) Exploitation of cell cycle and cell death controls by adenoviruses: the road to a productive infection. *Prog Mol Subcell Biol.* **36**:207-43
- Rux, J. J. & Burnett, R. M.** (2004) Adenovirus structure. *Hum Gene Ther* **15**:1167–1176

Rux, J. J., Kuser, P. R. & Burnett, R. M. (2003) Structural and phylogenetic analysis of adenovirus hexons by use of high-resolution X-ray crystallographic, molecular modeling, and sequence-based methods. *J Virol* **77**:9553–9566

Saban, S. D., Nepomuceno, R. R., Gritton, L. D., Nemerow, G. R. & Stewart, P. L. (2005) CryoEM structure at 9 Å resolution of an adenovirus vector targeted to hematopoietic cells. *J Mol Biol* **349**:526–537

Saban, S. D., Silvestry, M., Nemerow, G. R. & Stewart, P. L. (2006) Visualization of α -helices in a 6 Å resolution cryoelectron microscopy structure of adenovirus allows refinement of capsid protein assignments. *J Virol* **80**:12049–12059

San Martín C, Burnett RM. (2003) Structural studies on adenoviruses *Curr Top Microbiol Immunol* **272**:57-94

San Martín C, Glasgow JN, Borovjagin A, Beatty MS, Kashentseva EA, Curiel DT, Marabini R, Dmitriev IP. (2008) Localization of the N-terminus of minor coat protein IIIa in the adenovirus capsid. *J Mol Biol* **383**:923-34.

Salsman J, D. Top, J. Boutilier and R. Duncan (2005) Extensive syncytium formation mediated by the reovirus FAST proteins triggers apoptosis-induced membrane instability. *J. Virol.* **79**:8090–8100

Sasaki K, Kurata-Miura K, Ujita M, Angata K, Nakagawa S, Sekine S, Nishi T, Fukuda M (1997) Expression cloning of cDNA encoding a human beta-1,3-N-acetylglucosaminyltransferase that is essential for poly-N-acetyllactosamine synthesis. *Proc Natl Acad Sci U S A.* **94**: 14294-14299

Schneider, T. R. & Sheldrick, G. M. (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**:1772–1779

Schnitzer TJ (1985) Protein coding assignment of the S genes of the avian reovirus S1133. *Virology* **141**:167–170

Schnitzer TJ, T. Ramos and V. Gouvea (1982) Avian reovirus polypeptides: analysis of intracellular virus-specified products, virions, top component, and cores *J. Virol.* **43**:1006–1014

Schuck, P. (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**:1606-1619

Segerman A, Atkinson JP, Marttila M, Dennerquist V, Wadell G, Arnberg N (2003) Adenovirus type 11 uses CD46 as a cellular receptor. *J Virol* **77**: 9183-9191

Seiradake E, Cusack S (2005) Crystal structure of enteric adenovirus serotype 41 short fiber head. *J Virol* **79**: 14088-14094

Seiradake E, Lortat-Jacob H, Billet O, Kremer EJ, Cusack S (2006) Structural and mutational analysis of human Ad37 and canine adenovirus 2 fiber heads in

complex with the D1 domain of coxsackie and adenovirus receptor. *J Biol Chem* **281**: 33704-33716

Seiradake E, Henaff D, Wodrich H, Billet O, Perreau M, Hippert C, Mennechet F, Schoehn G, Lortat-Jacob H, Dreja H, Ibanes S, Kalatzis V, Wang JP, Finberg RW, Cusack S, Kremer EJ (2009) The cell adhesion molecule "CAR" and sialic acid on human erythrocytes influence adenovirus in vivo biodistribution. *PLoS Pathog* **5**: e1000277

Seetharaman J, Kanigsberg A, Slaaby R, Leffler H, Barondes SH, Rini JM (1998) X-ray crystal structure of the human galectin-3 carbohydrate recognition domain at 2.1 Å resolution. *J Biol Chem* **273**: 13047-13052

Sen GN (2001) Viruses and interferons. *Ann. Rev. Microbiol.* **55**:255–281

Shapouri, M. R. S., Kane, M., Letarte, M., Bergeron, J., Arella, M. & Salim, A. (1995) Cloning, sequencing and expression of the S1 gene of avian reovirus. *J Gen Virol* **76**:1515–1520

Shapouri MR, M. Arella and A. Silim (1996) Evidence for the multimeric nature and cell binding ability of avian reovirus sigma 3 protein, *J. Gen. Virol.* **77**:1203–1210

Shashkova, E. V., Cherenova, L. V., Kazansky, D. B. & Doronin, K. (2005) Avian adenovirus vector CELO-TK displays anticancer activity in human cancer cells and suppresses established murine melanoma tumors. *Cancer Gene Ther* **12**:617–626

Sheldrick GM (2002) Macromolecular phasing with SHELXE, *Z. Kristallog.* **217**:644–650

Sheriff, S., Chang, C. Y. & Ezekowitz, R. A. (1994) Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α -helical coiled-coil. *Nat Struct Biol* **1**:789–794

Short JJ, Vasu C, Holterman MJ, Curiel DT, Pereboev A (2006) Members of adenovirus species B utilize CD80 and CD86 as cellular attachment receptors. *Virus Res* **122**: 144-153

Shih WL, H.W. Hsu, M.H. Liao, L.H. Lee and H.J. Liu (2004) Avian reovirus σ C protein induces apoptosis in cultured cells. *Virology* **321**:65–74

Shmulevitz M, Z. Yameen, S. Dawe, J. Shou, D. O'Hara, I. Holmes and R. Duncan (2002) Sequential partially overlapping gene arrangement in the tricistronic S1 genome segments of avian reovirus and Nelson Bay reovirus: implications for translation initiation *J. Virol.* **76**:609–618

Shmulevitz M, R.F. Epand, R.M. Epand and R. Duncan (2004) Structural and functional properties of an unusual internal fusion peptide in a nonenveloped virus membrane fusion protein, *J. Virol.* **78**:2808–2818

Shmulevitz M, J. Corcoran, J. Salsman and R. Duncan (2004) Cell–cell fusion induced by the avian reovirus membrane fusion protein is regulated by protein degradation, *J. Virol.* **78**:5996–6004

Shmulevitz M & R. Duncan (2000) A new class of fusion-associated small transmembrane (FAST) proteins encoded by the non-enveloped fusogenic reoviruses. *EMBO J.* **19**:902–912

Silverstein SC & P.H. Schur (1970) Immunofluorescent localization of double-stranded RNA in reovirus-infected cells. *Virology* **41**:564–566

Sirena D, Lilienfeld B, Eisenhut M, Kälin S, Boucke K, Beerli RR, Vogt L, Ruedl C, Bachmann MF, Greber UF, Hemmi S (2004) The human membrane cofactor CD46 is a receptor for species B adenovirus serotype 3. *J Virol* **78**: 4454-4462

Songserm, T., van Roozelaar, D., Kant, A., Pol, J., Pijpers, A. & ter Huurne, A. (2003) Enteropathogenicity of Dutch and German avian reoviruses in SPF white leghorn chickens and broilers. *Vet Res* **34**:285–295

Spandidos DA & Graham AF (1976) Physical and chemical characterization of an avian reovirus, *J. Virol.* **19**:968–976

Spinelli, S., Desmyter, A., Verrips, C. T., de Haard, H. J., Moineau, S. & Cambillau, C. (2006a) Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat Struct Mol Biol* **13**:85–89

Spinelli, S., Campanacci, V., Blangy, S., Moineau, S., Tegoni, M. & Cambillau, C. (2006b) Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J Biol Chem* **281**:14256–14262

Stewart PL and G.R. Nemerow (2007) Cell integrins: commonly used receptors for diverse viral pathogens. *Trends Microbiol.* **15**:500–507

Smith TA, Mehaffey MG, Kayda DB, Saunders JM, Yei S, Trapnell BC, McClelland A, Kaleko M. (1993) Adenovirus mediated expression of therapeutic plasma levels of human factor IX in mice. *Nat Genet* **5**:397-402

Strelkov, S. V. & Burkhard, P. (2002) Analysis of α -helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol* **137**:54–64

Swaminathan GJ, Leonidas DD, Savage MP, Ackerman SJ, Acharya KR (1999) Selective recognition of mannose by the human eosinophil Charcot-Leyden crystal protein (galectin-10): a crystallographic study at 1.8 Å resolution. *Biochemistry* **38**: 13837-13843

Staba MJ, Wickham TJ, Kovesdi I, Hallahan DE. (2000) Modifications of the fiber in adenovirus vectors increase tropism for malignant glioma models. *Cancer Gene Ther.* **7**:13-9

Stehle T, Harrison SC. (1997) High-resolution structure of a polyomavirus VP1-oligosaccharide complex: implications for assembly and receptor binding. *EMBO J.* **16**:5139-48

Stevenson, M., Boos, E., Herbert, C., Hale, A., Green, N., Lyons, M., Chandler, L., Ulbrich, K., van Rooijen, N. & other authors (2006) Chick embryo lethal orphan virus can be polymer-coated and retargeted to infect mammalian cells. *Gene Ther* **13**:356–368

Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T. & other authors (2007) DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **448**:501–505

Tan, P. K., Michou, A. I., Bergelson, J. M. & Cotten, M. (2001) Defining CAR as a cellular receptor for the avian adenovirus CELO using a genetic analysis of the two viral fibre proteins. *J Gen Virol* **82**:1465–1472

Tao Y, D.L. Farsetta, M.L. Nibert and S.C. Harrison (2002) RNA synthesis in a cage, structural studies of reovirus polymerase lambda3. *Cell* **111**:733–745

Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997) Structure of bacteriophage T4 fibrin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**:789–798

Thacker EE, Timares L, Matthews QL (2009) Strategies to overcome host immunity to adenovirus vectors in vaccine development. *Expert Rev Vaccines* **8**: 761-777

Thomassen E, G. Gielen, M. Schütz, G. Schoehn, J.P. Abrahams, S. Miller and M.J. van Raaij (2003) The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**:361–373

Tourís-Otero F, M. Cortez-San Martín, J. Martínez-Costas and J. Benavente (2004a) Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of σ NS and λ A to μ NS inclusions. *J. Mol. Biol.* **341**:361–374

Tourís-Otero F, J. Martínez-Costas, V.N. Vakharia and J. Benavente (2004b) Avian reovirus nonstructural protein μ NS forms viroplasm-like inclusions and recruits σ NS to these structures, *Virology* **319**:94–106

Tourís-Otero F, J. Martínez-Costas, V.N. Vakharia and J. Benavente (2005) Characterization of the nucleic acid-binding activity of the avian reovirus nonstructural protein σ NS. *J. Gen. Virol.* **86**:1159–1169

Tomko RP, Xu R, Philipson L. (1997) HCAR and MCAR: the human and mouse cellular receptors for subgroup C adenoviruses and group B coxsackieviruses. *Proc Natl Acad Sci USA* **94**: 3352-3356

Trentin JJ, Yabe Y, Taylor G. (1962) The quest for human cancer viruses. *Science* **137**:835-41.

Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T. & Hodges, R. S. (2000) Effects of side-chain characteristics of stability and oligomerization state of a *de novo*-designed model coiled coil: 20 amino acid substitutions in position d. *J Mol Biol* **300**:377-402

Vagin A, Teplyakov A (2000) An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56**: 1622-1624

Varela R and J. Benavente (1994) Protein coding assignment of avian reovirus strain S1133 *J. Virol.* **68**:6775-6777

Varela R, J. Martínez-Costas, M. Mallo and J. Benavente (1996) Intracellular posttranslational modifications of S1133 avian reovirus proteins *J. Virol.* **70**:2974-2981

van Raaij, M. J., Hermo Parrado, X. L., Guardado Calvo, P., Fox, G. C., Llamas-Saiz, A. L., Costas, C., Martinez-Costas, J. & Benavente, J. (2005) Crystallization of the C-terminal globular domain of avian reovirus fibre. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**:651-654

van Raaij MJ, Louis N, Chroboczek J, Cusack S (1999a) Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. *Virology* **262**: 333-343

van Raaij MJ, Mitraki A, Lavigne G, Cusack S (1999b) A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**: 935-938

van Raaij MJ, G. Schoehn, M.R. Burda and S. Miller (2001) Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**:1137-1146

Wälti MA, Walser PJ, Thore S, Grünler A, Bednar M, Künzler M, Aebi M (2008) Structural basis for chitotetraose coordination by CGL3, a novel galectin-related protein from *Coprinopsis cinerea*. *J Mol Biol* **379**: 146-159

Wickham TJ, Mathias P, Cheresch DA, Nemerow GR (1993) Integrins alpha v beta 3 and alpha v beta 5 promote adenovirus internalization but not virus attachment. *Cell* **73**: 309-319

Wiener AS, Unger LJ, Cohen L, Feldman J (1956) Type-specific cold auto-antibodies as a cause of acquired hemolytic anemia and hemolytic transfusion reactions: biologic test with bovine red cells. *Ann Intern Med.* **44**: 221-240

Wiethoff, C. M., Wodrich, H., Gerace, L. & Nemerow, G. R. (2005) Adenovirus protein VI mediates membrane disruption following capsid disassembly. *J Virol* **79**:1992–2000

Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**:52–56

Wang, Q., J. Bergeron, T. Mabrouk, and G. Lemay. (1996) Site-directed mutagenesis of the double-stranded RNA binding domain of bacterially-expressed sigma 3 reovirus protein. *Virus Res.* **41**:141-151

Weis W, Brown JH, Cusack S, Paulson JC, Skehel JJ, Wiley DC. (1988) Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid. *Nature* **333**:426-31

Wu E, Fernandez J, Fleck SK, Von Seggern DJ, Huang S, Nemerow GR (2001) A 50-kDa membrane protein mediates sialic acid-independent binding and infection of conjunctival cells by adenovirus type 37. *Virology* **279**: 78-89

Wu, E., Pache, L., Von Seggern, D. J., Mullen, T. M., Mikyas, Y., Stewart, P. L. & Nemerow, G. R. (2003) Flexibility of the adenovirus fiber is required for efficient receptor interaction. *J Virol* **77**:7225–7235

Wu E, Trauger SA, Pache L, Mullen TM, von Seggern DJ, Siuzdak G, Nemerow GR (2004) Membrane cofactor protein is a receptor for adenoviruses associated with epidemic keratoconjunctivitis. *J Virol* **78**: 3897-3905

Xia D, Henry LJ, Gerard RD, Deisenhofer J (1994) Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 Å resolution. *Structure* **2**: 1259-1270

Xu W, M.K. Patrick, P.R. Hazelton and K.M. Coombs (2004) Avian reovirus temperature-sensitive mutant *tsA12* has a lesion in major core protein σA and is defective in assembly. *J. Virol.* **78**:11142–11151

Yin HS & L.H. Lee (1998) Identification and characterization of RNA-binding activities of avian reovirus non-structural protein σNS . *J. Gen. Virol.* **79**:1411–1413

Yin, H. S., J. H. Shien, and L. H. Lee. (2000) Synthesis in *Escherichia coli* of avian reovirus core protein σA and its dsRNA-binding activity. *Virology* **266**:33-41

Yue, Z., and A. J. Shatkin. (1997) Double-stranded RNA-dependent protein kinase (PKR) is regulated by reovirus structural proteins. *Virology* **234**:364-371

Zick Y, Eisenstein M, Goren RA, Hadari YR, Levy Y, Ronen D (2004) Role of galectin-8 as a modulator of cell adhesion and cell growth. *Glycoconj J* **19**: 517-526

Zhang X, J. Tang, S.B. Walker, D. O'Hara, M.L. Nibert, R. Duncan and T.S. Baker. (2005) Structure of avian *Orthoreovirus* virion by electron cryomicroscopy and image reconstruction. *Virology* **343**:25–35

Zhang X, S.B. Walker, P.R. Chipman, M.L. Nibert and T.S. Baker (2003) Reovirus polymerase lambda 3 localized by cryo-electron microscopy of virions at a resolution of 7.6 Å. *Nat. Struct. Biol.* **10**:1011–1018

Zhang, X., Y. Ji, L. Zhang, S. C. Harrison, D. C. Marinescu, M. L. Nibert, and T. S. Baker. (2005) Features of reovirus outer capsid protein mu1 revealed by electron cryomicroscopy and image reconstruction of the virion at 7.0-Ångstrom resolution. *Structure* **13**:1545-1557

Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005) The structure of the human adenovirus 2 penton. *Mol Cell* **17**:121–135

ARTÍCULOS PUBLICADOS

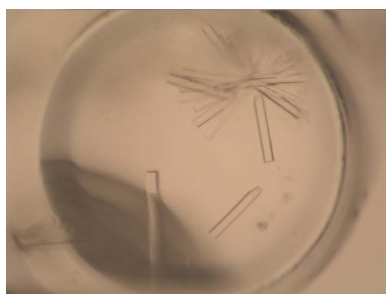
Mark J. van Raaij,^{a,b*}
X. Lois Hermo Parrado,^a
Pablo Guardado Calvo,^a
Gavin C. Fox,^c Antonio L.
Llamas-Saiz,^b Celina Costas,^{a,‡}
José Martínez-Costas^a and Javier
Benavente^a

^aDepartamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, ^bUnidad de Difracción de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, and ^cSpanish CRG Beamline BM16, European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, BP 220, F-38043 Grenoble, France

‡ Current address: Cancer Research UK Molecular Oncology Unit, Institute of Cancer, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, England.

Correspondence e-mail: vanraaij@usc.es

Received 1 April 2005
Accepted 27 May 2005
Online 15 June 2005



© 2005 International Union of Crystallography
All rights reserved

Crystallization of the C-terminal globular domain of avian reovirus fibre

Avian reovirus fibre, a homotrimer of the σ C protein, is responsible for primary host-cell attachment. Using the protease trypsin, a C-terminal σ C fragment containing amino acids 156–326 has been generated which was subsequently purified and crystallized. Two different crystal forms were obtained, one grown in the absence of divalent cations and belonging to space group $P6_322$ (unit-cell parameters $a = 75.6$, $c = 243.1$ Å) and one grown in the presence of either zinc or cadmium sulfate and belonging to space group $P321$ (unit-cell parameters $a = 74.7$, $c = 74.5$ Å and $a = 73.1$, $c = 69.9$ Å for the Zn^{II} - and Cd^{II} -grown crystals, respectively). The first crystal form diffracted synchrotron radiation to 3.0 Å resolution and the second form to 2.2–2.3 Å. Its closest related structure, the C-terminal fragment of mammalian reovirus fibre, has only 18% sequence identity and molecular-replacement attempts were unsuccessful. Therefore, a search is under way for suitable heavy-atom derivatives and attempts are being made to grow protein crystals containing selenomethionine instead of methionine.

1. Introduction

Avian reovirus, a non-enveloped virus and a member of the *Orthoreovirus* genus of the *Reoviridae* family, is an important poultry pathogen (Rosenberger *et al.*, 1989). The virus has a dsRNA genome consisting of ten segments encased by a double concentric icosahedral capsid (Spandidos & Graham, 1976). The inner capsid is made up of the λ A, λ C and σ A proteins, which enclose up to 12 copies of the RNA-dependent RNA polymerase complex (consisting of λ B and μ A), together forming the 'core'. Proteins μ B, σ B and σ C are components of the outer shell (Martínez-Costas *et al.*, 1997). The virus also encodes several non-structural proteins, involved in cell fusion (p10), virus assembly (muNS and σ NS) or as yet unknown functions (p17) (Varela *et al.*, 1996; Bodelón *et al.*, 2001; Touris-Otero *et al.*, 2004; Costas *et al.*, 2005).

The minor outer capsid protein, σ C, is the avian reovirus fibre, the avian counterpart of mammalian reovirus σ 1. σ C is an elongated homotrimer and is responsible for primary host-cell attachment (Grande *et al.*, 2000, 2002). Avian reovirus σ C has 20% sequence identity to mammalian reovirus σ 1. The most similar region is at the N-terminus, which is presumably implicated in the interaction with λ C (the equivalent of mammalian reovirus λ 2). The region of σ C comprising residues 50–156 is predicted to form a triple coiled-coil structure (Costas Iglesias, 2004), while the C-terminal part (residues 156–326) probably forms a globular receptor-binding domain.

The mammalian reovirus receptor has recently been determined to be a junction adhesion molecule, JAM (Barton *et al.*, 2001). However, σ C probably binds a different as yet unknown receptor, since the C-terminal region of the two proteins have only 18% sequence identity and since mammalian reoviruses do not attach to avian cells (Barton *et al.*, 2001). The structure of a C-terminal fragment of mammalian reovirus σ 1 containing the receptor-binding domain and part of the elongated shaft has been solved (Chappell *et al.*, 2002). The receptor-binding domain has a β -barrel fold, while the part of the elongated shaft contains several triple β -spiral repeats comparable to those of adenovirus fibre (van Raaij *et al.*, 1999).

Using proteolysis, we have now identified a crystallizable C-terminal fragment of σ C. Resolution of its structure should hopefully shed light on the differences between the fibres of avian and mammalian reoviruses and provide clues about receptor binding.

2. Methods

A DNA-fragment encoding σ C residues 117–326 was produced by the polymerase chain reaction and cloned into the expression vector pET28c+ (Novagen, Darmstadt, Germany). The resultant plasmid pET28- σ C117-326 encodes the mentioned σ C fragment fused to an

N-terminal purification tag containing six consecutive histidine residues and a T7 tag (the N-terminal 11 residues of the bacteriophage T7 gene 10 product). The sequence of the insert was confirmed by DNA-sequence analysis (Sistemas Genómicos, Valencia, Spain).

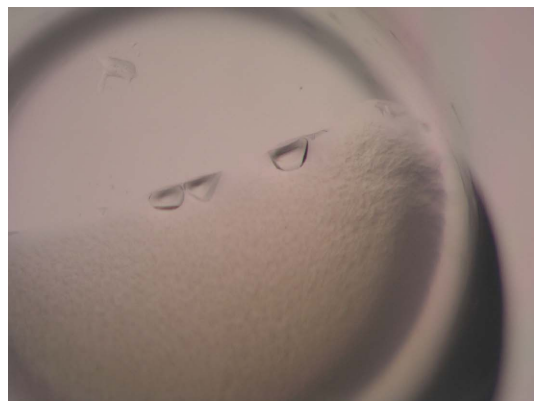
For expression, *Escherichia coli* strain BL21(DE3) was freshly transformed with the plasmid and four 1 l cultures were grown aerobically at 310 K to an optical density of 0.6–0.8 measured at 600 nm. The cultures were cooled to below 298 K, after which expression was induced by adding 1 mM isopropyl-thio- β -D-galactopyranoside and was allowed to continue for 4–5 h at 298 K. Harvested cells were resuspended in 40 ml cold resuspension buffer [4.29 mM disodium hydrogen phosphate, 1.47 mM potassium dihydrogen phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride, 0.1% (w/v) Tween-20] and frozen at 253 K. Bacteria were lysed by a double pass through an emulsifier (Avestin emulsifier C5, Avestin Europe GmbH, Mannheim, Germany). After removing insoluble material, 3 ml Ni-NTA resin (Qiagen, Valencia CA, USA) was added. The suspension was incubated for 1 h at 277 K and poured into an empty column. The resin was washed with PBS buffer (resuspension buffer without Tween-20) and elution was performed with a step gradient of imidazole pH 7 in PBS buffer (10, 20, 50, 100, 200, 500 mM imidazole). His-T7-tagged σ C eluted at high imidazole concentrations (200–500 mM), consistent with the protein being a trimer and containing three 6 \times His tags. This procedure was repeated up to four times until no more His-T7-tagged σ C could be captured from the solution.

The partially purified protein was incubated with 1 mg ml⁻¹ trypsin for 30 min at 310 K, precipitated by adding 1.5 volumes of saturated ammonium sulfate, redissolved in TE buffer (10 mM Tris-HCl pH 8.5, 1 mM EDTA), dialysed against the same buffer and applied onto a 1 ml Uno-Q column (Biorad, Barcelona, Spain). Pure σ C eluted right at the start of a linear 0–1 M sodium chloride gradient in TE buffer. The protein was concentrated to between 15 and 20 mg ml⁻¹ using Centricon concentrators (Millipore, Madrid, Spain), incorporating three washes with TE buffer to eliminate small-molecule impurities.

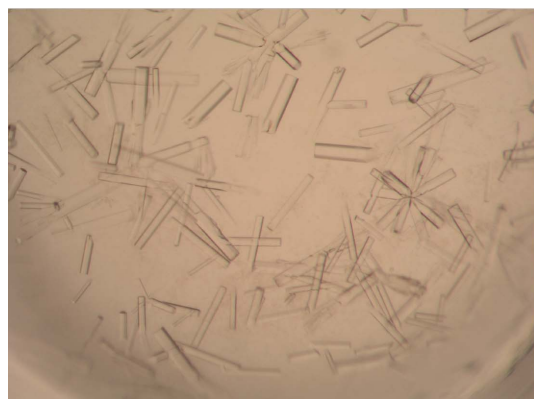
Crystallization took place by vapour diffusion in sitting-drop CompactClover plates (Jena Biosciences, Jena, Germany), with 0.1–0.15 ml reservoirs and drops of 2–5 μ l protein solution mixed with 2–5 μ l reservoir solution. For N-terminal sequence analysis, protein from the crystallization drop was taken up in 60 mM Tris-HCl pH 6.6, 2% (w/v) sodium dodecylsulfate, 5% (v/v) β -mercaptoethanol, 10% (v/v) glycerol, 0.05% bromophenol blue and heated and unheated samples were analysed by SDS-PAGE, transferred to polyvinylidene fluoride membrane and subjected to Edman degradation. Crystallographic data were obtained using *MOSFLM* (Powell, 1999; Leslie, 1999) and scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994).

3. Results and discussion

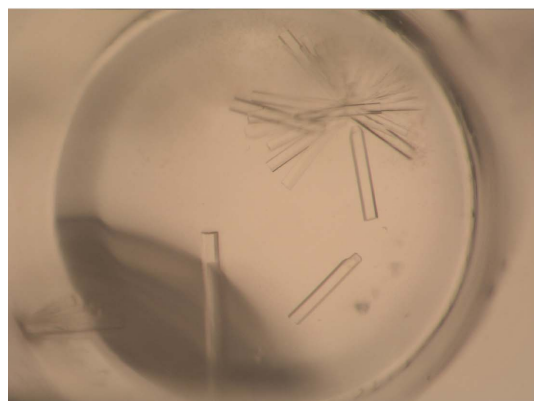
Previously, σ C had been cloned and expressed using the pIL plasmid (Grande *et al.*, 2000, 2002); however, we subsequently found it difficult to reliably obtain significant quantities of soluble and correctly folded protein. To overcome this problem, the protein was cloned into the expression vector pET21d+ (Novagen, Darmstadt, Germany), which expressed soluble trimeric σ C with an N-terminal T7 tag, but still in relatively low yield. Trypsin digestion generated a major C-terminal stable fragment of this protein containing residues 115–326 (Costas Iglesias, 2004). Electron microscopy showed this



(a)



(b)



(c)

Figure 1

Crystals of the avian reovirus fibre C-terminal globular domain. (a) Irregular prisms belonging to space group *P6₃22*. (b) Bar-shaped crystals of space group *P321* grown in the presence of zinc sulfate. (c) Bar-shaped crystals of space group *P321* grown in the presence of cadmium sulfate. The crystals are up to 0.15 mm wide (a), 0.1 mm long (b) and 0.4 mm long (c), respectively.

Table 1

Crystallographic data statistics.

Values in parentheses are for the highest resolution bin.

	Form I	Form II (Zn ²⁺)	Form II (Cd ²⁺)
Crystal habit	Irregular prisms/lenses	Hexagonal/trigonal bars	Hexagonal/trigonal bars
Crystal size (mm)	0.15 × 0.15 × 0.05	0.025 × 0.025 × 0.1	0.1 × 0.1 × 0.4
Likely space group	<i>P</i> 6 ₃ 22	<i>P</i> 321	<i>P</i> 321
Unit-cell parameters (Å)	<i>a</i> = 75.6, <i>c</i> = 243.1	<i>a</i> = 74.7, <i>c</i> = 74.5	<i>a</i> = 73.1, <i>c</i> = 69.9
No. of molecules per AU†	2	1	1
No. of molecules per cell‡	24	6	6
Matthews coefficient† (Å ³ Da ⁻¹)	2.7	3.3	2.9
Beamline (at ESRF)	ID23-1	Spanish CRG BM16	Spanish CRG BM16
Wavelength (Å)	0.976	1.282	0.980
Detector	MAR Mosaic	165 mm MAR CCD	165 mm MAR CCD
No. of observed reflections‡	8895 (1256)	12298 (1724)	9252 (736)
Resolution range (Å)	30.0–3.0 (3.16–3.00)	20.0–2.2 (2.32–2.20)	30.0–2.3 (2.42–2.32)
Multiplicity	8.8 (9.2)	10.5 (10.6)	10.6 (9.8)
Completeness (%)	99.8 (100.0)	98.3 (97.4)	95.8 (65.5)
$\langle I/\sigma(I) \rangle$	8.2 (2.9)	9.1 (2.8)	8.0 (2.5)
R_{sym}^{\S} (%)	7.7 (26.2)	6.8 (26.5)	6.7 (28.3)

† Estimation based on likely solvent content. ‡ No σ cutoff or other restrictions were used for inclusion of observed reflections. § $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity of that reflection.

fragment to be globular, with a short vestige of stalk (results not shown).

We have now cloned a very similar fragment (amino acids 117–326, see §2) into the related expression vector pET28c+ and obtained soluble trimeric protein with an N-terminal His-T7 tag. Using low concentrations of trypsin, the expression tag could be removed and the protein was further purified using strong anion chromatography (results not shown). However, extensive trials failed to yield crystals. Therefore, we treated the protein with high concentrations of trypsin (1 mg ml⁻¹, see §2) and obtained a smaller fragment which could be crystallized (Fig. 1). Although we have not yet obtained electron-microscope images for this fragment, it is likely that the protease treatment removes the short remainder of the stalk, leaving the globular domain. It is also likely that this domain retains receptor-binding activity, although this has not yet been tested. Yields of purified concentrated protein were between 5 and 10 mg from 4 l culture (assuming the absorbance at 280 nm of a 1 mg ml⁻¹ σ C solution to be 1.0).

Crystals were obtained using various conditions, but only those grown using 1.5 M ammonium sulfate, 0.1 M Tris-HCl pH 8.5, 12% glycerol as reservoir solution yielded interpretable X-ray diffraction patterns, to around 3.0 Å (crystal form I). Optimization of the conditions using divalent-cation additives yielded other related crystal forms that diffracted X-rays to higher resolution (2.2–2.3 Å; crystal forms II and III grown in the presence of 10 mM zinc chloride and cadmium chloride, respectively). Crystallographic data were collected at the European Synchrotron Radiation Facility (ESRF) on the Spanish CRG beamline BM16 and on beamline ID23-1. Data from crystals of form I can be scaled in the hexagonal space group *P*6₃22 and show systematic absences along the reciprocal axis *l* consistent with this space group. These crystals are most likely to contain two independent monomers in each asymmetric unit (each forming a trimer arising from crystallographic symmetry). Although the resolution is limited, the data quality is good to 3.0 Å (see Table 1). The crystals of form II belong to the trigonal space group *P*321 (no systematic absences), with one monomer in the asymmetric unit, and diffract X-rays to 2.2–2.3 Å resolution. The length of the *c* axis varies between crystals, a property that may make different crystals non-isomorphous and structure solution more difficult. This property appears to be correlated with the nature of the divalent cation used in the crystallization medium (Table 1), although we have not yet tested enough crystals to be certain.

SDS-PAGE combined with N-terminal sequence analysis showed that the crystals contained amino acids 156–326 of σ C, although we cannot exclude the removal of a few residues from the C-terminus. Taking advantage of the extraordinary stability of the σ C trimer, we compared the electrophoretic mobility of heated and unheated samples of dissolved crystals in the presence of sodium dodecyl-sulfate. The results showed that the σ C ‘head’ domain, similar to the full-length protein, retains its trimeric structure when the sample is not boiled.

We were not successful in molecular-replacement attempts using the equivalent domain of the mammalian reovirus fibre (PDB code 1kke; Chappell *et al.*, 2002), presumably owing to the low sequence identity (18%). Similarly, zinc SAD and MAD structure-solution attempts were not successful, possibly because the crystals do not contain ordered zinc ions or because there is only one ordered zinc ion per σ C trimer on the threefold crystallographic symmetry axis, as has been reported for the structure of the receptor-binding domain of the bacteriophage short tail fibre (PDB code 1ocv; Thomassen *et al.*, 2003). We are currently trying to obtain heavy-atom derivatives, which will hopefully allow us to solve the structure. Alternatively, we will attempt to grow crystals of protein containing selenomethionine instead of methionine. The structure will reveal whether the avian reovirus σ C head domain has structural homology to that of mammalian reovirus σ 1, despite its low sequence homology. We also hope to obtain other information from the structure, such as which regions may be involved in receptor binding.

We thank Salvador Blanco Turnes for technical assistance, Stefan Müller (Profos AG, Regensburg, Germany) for providing crystallization plates, Javier Varela (CSIC-CIB, Madrid, Spain) for N-terminal sequence analysis and Dave Hall (ESRF, France) for help with data collection. This research was sponsored by research grants BMC2002-02436 (MJvR) and BMC2001-2839 (JB) from the Spanish Ministry of Science and Technology and PGIDIT03PXIC20307PN (MJvR) and PGIDIT02PXIC20301PN (JB) from the Xunta de Galicia.

References

- Barton, E. S., Forrest, J. C., Connolly, J. L., Liu, Y., Schnell, F. J., Nusrat, A., Parkos, C. A. & Dermody, T. S. (2001). *Cell*, **104**, 441–451.

- Bodelón, G., Labrada, L., Martínez-Costas, J. & Benavente, J. (2001). *Virology*, **290**, 181–191.
- Chappell, J. D., Prota, A. E., Dermody, T. S. & Stehle, T. (2002). *EMBO J.* **21**, 1–11.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Costas Iglesias, C. (2004). PhD thesis, Universidad de Santiago de Compostela, Spain.
- Costas, C., Martínez-Costas, J., Bodelón, G. & Benavente, J. (2005). *J. Virol.* **79**, 2141–2150.
- Grande, A., Costas, C. & Benavente, J. (2002). *J. Gen. Virol.* **83**, 131–139.
- Grande, A., Rodríguez, E., Costas, C., Everitt, E. & Benavente, J. (2000). *Virology*, **274**, 367–377.
- Leslie, A. G. W. (1999). *Acta Cryst.* **D55**, 1696–1702.
- Martínez-Costas, J., Grande, A., Varela, R., García-Martínez, C. & Benavente, J. (1997). *J. Virol.* **71**, 59–64.
- Powell, H. R. (1999). *Acta Cryst.* **D55**, 1690–1695.
- Raaij, M. J. van, Mitraki, A., Lavigne, G. & Cusack, S. (1999). *Nature (London)*, **401**, 935–938.
- Rosenberger, J. K., Sterner, F. J., Botts, S., Lee, K. P. & Margolin, A. (1989). *Avian Dis.* **33**, 535–544.
- Spandidos, D. A. & Graham, A. F. (1976). *J. Virol.* **19**, 968–976.
- Thomassen, E., Gielen, G., Schutz, M., Schoehn, G., Abrahams, J. P., Miller, S. & van Raaij, M. J. (2003). *J. Mol. Biol.* **331**, 361–373.
- Touris-Otero, F., Martínez-Costas, J., Vakharia, V. N. & Benavente, J. (2004). *Virology*, **319**, 94–106.
- Varela, R., Martínez-Costas, J., Mallo, M. & Benavente, J. (1996). *J. Virol.* **70**, 2974–2981.

Structure of the Carboxy-terminal Receptor-binding Domain of Avian Reovirus Fibre SigmaC

Pablo Guardado Calvo¹, Gavin C. Fox², X. Lois Hermo Parrado¹
 Antonio L. Llamas-Saiz³, Celina Costas^{1†}
 José Martínez-Costas¹, Javier Benavente¹ and Mark J. van Raaij^{1,3*}

¹Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur E-15782 Santiago de Compostela, Spain

²Spanish CRG Beamline BM16 European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP 220, F-38043 Grenoble, France

³Unidad de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido Edificio CACTUS, Universidad de Santiago de Compostela Campus Sur, E-15782 Santiago de Compostela, Spain

Avian reovirus fibre, a homo-trimer of the sigmaC protein, is responsible for primary host cell attachment. The protein expressed in bacteria forms elongated fibres comprised of a carboxy-terminal globular head domain and a slender shaft, and partial proteolysis yielded a carboxy-terminal protease-stable domain that was amenable to crystallisation. Here, we show that this fragment retains receptor-binding capability and report its structure, solved using two-wavelength anomalous diffraction and refined using data collected from three different crystal forms at 2.1 Å, 2.35 Å and 3.0 Å resolution. The carboxy-terminal globular domain has a beta-barrel fold with the same overall topology as the mammalian reovirus fibre (sigma1). However, the monomers of the sigmaC trimer show a more splayed-out arrangement than in the sigma1 structure. Also resolved are two triple beta-spiral repeats of the shaft or stalk domain. The presence in the sequence of heptad repeats amino-terminal to these triple beta-spiral repeats suggests that the unresolved portion of the shaft domain contains a triple alpha-helical coiled-coil structure. Implications for the function and stability of the sigmaC protein are discussed.

© 2005 Elsevier Ltd. All rights reserved.

*Corresponding author

Keywords: orthoreovirus; avian reovirus fibre; triple beta-spiral; beta-barrel; receptor-binding

Introduction

Avian reovirus, a member of the *Orthoreovirus* genus of the Reoviridae family, is an important poultry pathogen.¹ This non-enveloped double-stranded RNA virus is composed of a double concentric icosahedral capsid.² The inner “core” contains the proteins lambdaA, lambdaB, lambdaC (the guanylyltransferase³), muA and sigmaA (a double-stranded RNA-binding protein involved in interferon resistance^{4,5}), plus ten double-

stranded RNA genome segments.⁶ The outer shell contains the proteins sigmaB, sigmaC and muB. Avian reoviruses also express several non-structural proteins: p10, an integral membrane protein with fusogenic and permeabilising activities;^{7–10} muNS and sigmaNS, involved in virus assembly;^{11–13} and p17, a nuclear protein of as yet unknown function.¹⁴

Crystallographic studies have so far been limited to mammalian reoviruses, of which several structures have been solved, notably amongst these the “core”¹⁵ and the RNA polymerase.¹⁶ The structure of the double-stranded RNA-binding protein sigma3 has also been solved, both independently¹⁷ and in complex with the membrane-penetration protein mu1;¹⁸ their avian reovirus equivalents are sigmaB and muB, respectively. Finally, the structure of the carboxy-terminal 206 residues of mammalian reovirus fibre (sigma1) revealed a triple beta-spiral shaft domain and a globular triple beta-barrel receptor-binding domain.¹⁹

† Present address: C. Costas, Cancer Research UK Molecular Oncology Unit, Institute of Cancer, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, UK.

Abbreviations used: BSA, bovine serum albumin; r.m.s., root-mean-square; S1133, strain 1133 of avian reovirus.

E-mail address of the corresponding author: vanraaij@usc.es

The avian reovirus fibre is formed by the minor outer capsid protein sigmaC; like sigma1, it is an elongated homotrimer, which is responsible for extracellular host cell attachment.^{20,21} It has 20% sequence identity with mammalian reovirus sigma1. The mammalian reovirus receptor has been identified as junction adhesion molecule A (JAM-A).²² However, sigmaC probably binds a different, as yet unknown, receptor, because mammalian reoviruses do not attach to avian cells.²² Previously, we have reported the generation of a crystallisable carboxy-terminal fragment of sigmaC.²³ Here, we show its structure, solved using two-wavelength anomalous diffraction analysis of a mercury derivative and refined against data collected from three different crystal forms.

Results and Discussion

Inhibition of avian reovirus infection of chicken embryo fibroblasts

We have previously identified a carboxy-terminal, protease-stable domain of the avian reovirus fibre.²³ Generating this stable domain using trypsin was reproducible, although sometimes a fragment starting at Ser151 (the residue immediately following Lys150) was produced, and other times a fragment starting at Ala156; amino-terminal sequence analysis did not reveal evidence for mixtures of the two. Apparently, in some cases, the presence of an additional protease impurity in the sigmaC or trypsin preparation removes some extra residues from the amino terminus. The carboxy terminus is unaffected by proteolysis, as is evident from the crystal structures (see below).

By analogy with related proteins the carboxy-terminal domain was predicted to play a key role in receptor binding, therefore we investigated whether the stable fragment could inhibit infection of chicken embryo fibroblasts by avian reovirus S1133. Chicken embryo fibroblasts were incubated in the presence of increasing amounts of the sigmaC fragment, after which avian reovirus was added and the intracellular production of the non-structural viral protein muNS was monitored by immunoblotting. Figure 1 shows that incubation of cells with increasing concentrations of the sigmaC fragment causes reduction of viral gene expression, as demonstrated by diminished intracellular concentration of the non-structural protein muNS. This indicates that the carboxy-terminal sigmaC domain inhibits an early step of avian reovirus infection, presumably extracellular viral attachment.

Structure solution

Crystals belonging to two different space groups were obtained, $P6_322$ (form I) and $P321$ (form II).²³ The crystals of form II can be further subdivided into two classes with different c -axes parameters (around 75 Å for the IIz-form and around 70 Å for

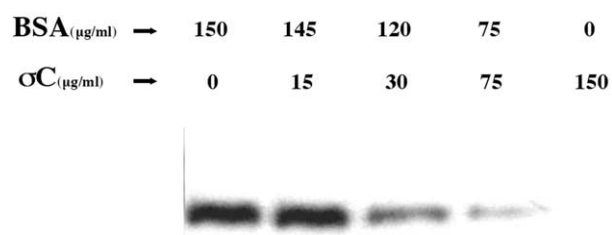


Figure 1. SigmaC Δ 1-150 inhibits infection of chicken embryo fibroblasts by avian reovirus. Monolayers of chicken embryo fibroblasts in five wells were washed, overlaid with medium and placed on ice. Then, 150 $\mu\text{g/ml}$ of protein was added to each well, containing increasing amounts of sigmaC and decreasing amounts of BSA, as shown. After 15 min of incubation, one plaque-forming unit/cell of avian reovirus S1133 was added to each well and allowed to adsorb. After adsorption, the cells were washed with medium and incubated at 37 °C for 8 h to allow primary infection only. Lysates of infected cells were analysed for the cellular production of the viral non-structural protein muNS.¹¹

the IIc-form). The growth of form IIz or IIc appears to be dependent on the presence of divalent zinc or cadmium cations, respectively, although we have not tested a sufficiently large number of crystals to be conclusive on this observation. Interestingly, IIz-form crystals derivatised with mercury compounds appear to convert to form IIc, suggesting that divalent mercury cations can mimic the effect of divalent cadmium ions.

A multi-wavelength anomalous diffraction dataset was collected on a crystal of form IIc derivatised with methyl mercury chloride. The best results were obtained by scaling together the peak and high-energy remote wavelengths and omitting the inflection point dataset. The inflection point dataset had slightly worse merging statistics, completeness and signal-to-noise ratio, and was therefore cut to 2.35 Å resolution. A possible explanation is radiation damage, as the inflection point dataset was collected after the other two; although we did not investigate this further. We identified four heavy-atom sites, of which three turned out to correspond to alternative positions of the mercury-modified Cys222 and the fourth an ordered cadmium ion.

After phasing and solvent flattening, a readily interpretable map was obtained in which 160 residues were automatically traced, although partially out-of-phase with respect to side-chain docking. This model was used as input in a molecular replacement with the highest resolution dataset obtained (form IIz, 2.1 Å resolution), and automatic tracing produced a largely complete protein model, which was finished by hand. This model was then used to solve the structures of the two other crystal forms by molecular replacement, followed by manual adjustment where necessary, and removing residues for which the electron density was unconvincing. The refined structures show good

correspondence to the data, have good geometry and have few residues in unlikely regions of the Ramachandran plot (Table 1). One amino acid residue, Val186, in a beta-turn of the stalk, is at the same position in a “forbidden region” in all three structures; however, it has convincing electron density in all three structures. His255 of the IIC structure is also in a forbidden region; it is located in

a less well-resolved surface loop of the globular head domain. Other amino acid residues (Leu152, His255 and Ile256 in form IIC, Ile256 and Ser284 in form IIC and Asn208, Asp283 and Ser284 in form I) are in generously allowed regions of the Ramachandran plot and near the amino terminus or in surface loops of the structure for which the density is less than optimal.

Table 1. Crystallographic data and refinement statistics

A. Derivative data collection			
Beam-line	BM16 European Synchrotron Radiation Facility (ESRF)		
Detector	165 mm MAR-CCD		
Space group	P321		
Cell parameters (<i>a,b/c</i> , Å) ^a	74.66/69.91		
Mosaic spread (deg.)	0.74		
	Peak	Remote	Inflection
Wavelength (Å)	1.00770	0.91838	1.00927
Resolution (Å)	20–2.22 (2.34–2.22) ^b	20–2.22 (2.34–2.22)	20.0–2.35 (2.48–2.35)
Observed reflections ^c	11,470 (1649)	11,477 (1649)	9713 (1393)
Multiplicity	10.7 (10.9)	10.6 (10.8)	10.0 (10.1)
Completeness (%)	99.8 (100.0)	99.9 (100.0)	99.8 (100.0)
<i>R</i> _{sym} (%) ^d	7.1 (28.1)	7.4 (27.4)	7.2 (29.9)
<i>I</i> / <i>σ</i> (<i>I</i>)	7.3 (2.5)	7.3 (2.7)	8.1 (2.5)
Rejected measurements (%) ^e	0.04	0.04	0.06
B. Phasing			
Resolution range used in SHELXD (Å)	20.0–2.23 (2.31–2.23)		
No. of reflections	11,469 (1132)		
No. of heavy atom sites (SHELXD)	4		
Contrast (SHELXE: original/inverted)	0.667/0.084		
Figure of merit (SHELXE)	0.600 (0.598)		
Map correlation coefficient (SHELXE)	0.827 (0.841)		
Pseudo-free correlation coefficient (SHELXE)	0.667		
C. Native data and refinement statistics			
Crystal form	IIC	IIC	I
Beam-line (ESRF)	BM16	BM16	ID23-1
Detector (MAR-CCD)	165 mm	165 mm	Mosaic 225 mm
Wavelength (Å)	1.28191	0.97959	0.97565
Space group	P321	P321	P6 ₃ 22
Cell parameters (<i>a,b/c</i> , Å)	74.78/74.64	73.08/69.93	75.60/243.14
Mosaic spread (deg.)	0.47	0.70	0.78
Resolution range (Å)	30–2.1 (2.21–2.10)	37–2.35 (2.45–2.32)	30–3.0 (3.2–3.0)
Observed reflections ^c	14,231 (2019)	9277 (1005)	8895 (1256)
Multiplicity	7.9 (8.0)	10.5 (10.0)	8.8 (9.2)
Completeness (%)	98.7 (97.7)	95.8 (71.6)	99.8 (100.0)
<i>R</i> _{sym} (%) ^d	7.2 (28.8)	7.7 (33.7)	7.7 (26.2)
<i>I</i> / <i>σ</i> (<i>I</i>)	7.3 (2.6)	7.1 (2.1)	8.2 (2.9)
Rejected measurements (%) ^e	0.13	0.06	1.06
Resolution range used, refinement (Å)	30–2.1 (2.16–2.10)	37–2.35 (2.41–2.35)	30–3.0 (3.1–3.0)
No. of reflections used, refinement	12,815 (950)	8348 (541)	8003 (556)
No. of reflections used, <i>R</i> -free	1416 (88)	929 (67)	885 (70)
<i>R</i> -factor ^f	0.161 (0.18)	0.183 (0.25)	0.198 (0.27)
<i>R</i> -free	0.223 (0.25)	0.243 (0.36)	0.249 (0.34)
No. of protein/water atoms	1326/164	1285/120	1260/44
Ions	3 SO ₄ ²⁻ /Zn ²⁺	1 SO ₄ ²⁻ /2 Cd ²⁺	–
Average <i>B</i> -value protein/solvent (Å ²)	24.6/33.0	38.8/40.7	44.6/35.9
Average <i>B</i> -value sulphate/metal (Å ²)	61.3/45.7	63.6/55.7	–/–
Ramachandran statistics (%) ^g	88.4/9.0/1.9/0.6	87.9/9.4/1.3/1.3	84.8/12.4/2.1/0.7
r.m.s. deviations (bonds, Å/angles, °) ^h	0.015/1.8	0.014/1.6	0.014/1.7

^a Ångström (0.1 nm).

^b Values in parentheses are for the highest resolution bin, where applicable.

^c No *σ* cut-off was used for inclusion of observed reflections.

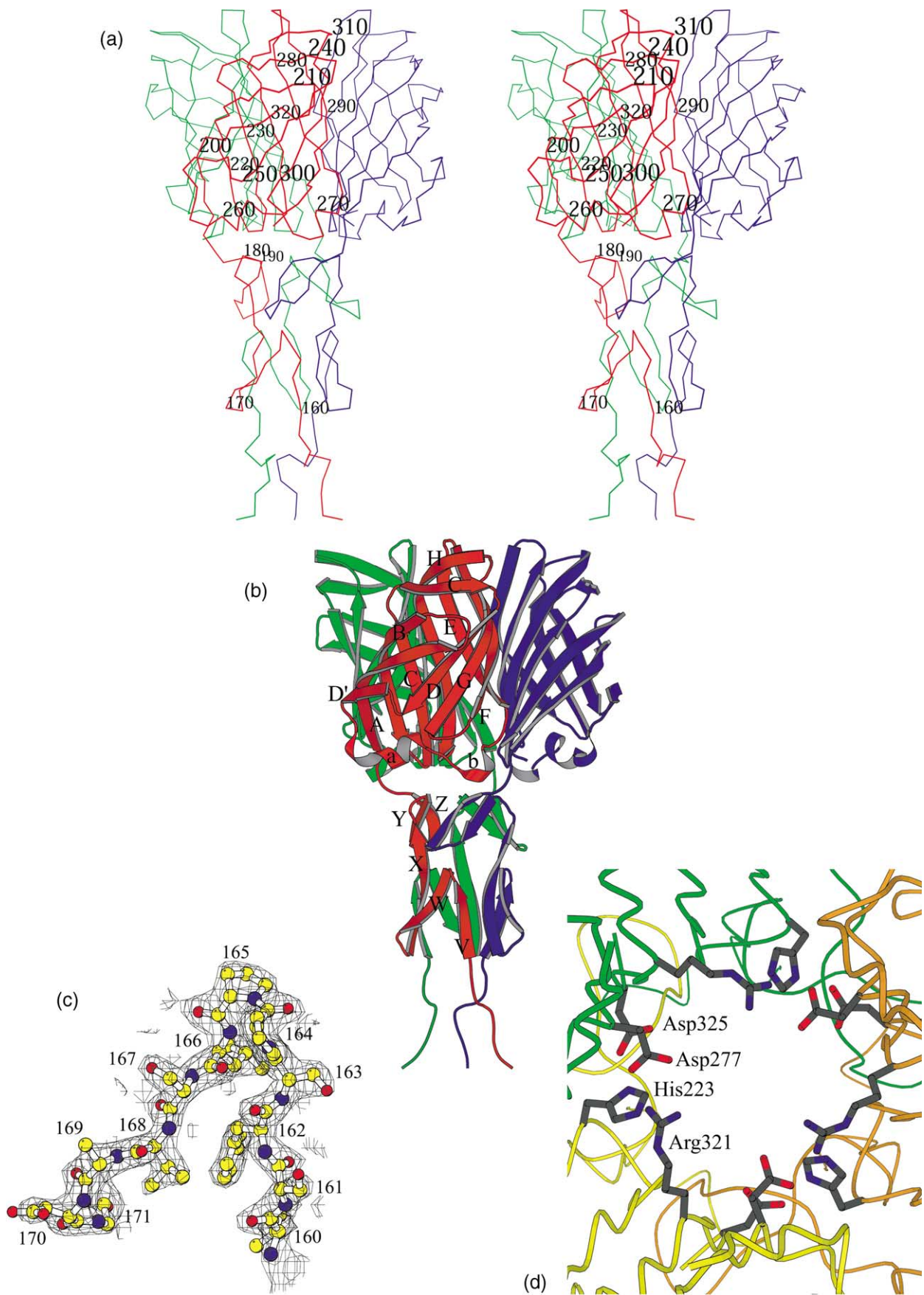
^d $R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.

^e Measurements with intensities differing more than six standard deviations from the mean were rejected (3.5 standard deviations in the case of the P6₃22 dataset).

^f $R = \sum ||F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)|| / \sum |F_{\text{obs}}(hkl)|$.

^g According to the program PROCHECK.⁴⁹ The percentages are indicated of residues in the most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

^h Estimates provided by the program REFMAC.⁴⁶



When superimposed, the three structures exhibit root-mean-square (r.m.s.) differences of only 0.5–0.6 Å between all C^α positions. Slight but significant differences are apparent in the conformations of amino acids amino-terminal to residue 160, in the EF-loop on the top of the trimer and in a region of the DE-loop (amino acid residues 254–258). If instead of monomers, the trimeric biological units are superimposed for the three structures, they superimpose equally well, with r.m.s. differences of around 0.6 Å, suggesting that the sigmaCΔ1-150 trimer is structurally rigid, with no evidence of a hinge region between the shaft and head domains. In contrast, the adenovirus fibre does contain a hinge region between the shaft and head domains,²⁴ while sigma1 has a hinge region between the last and second-to-last beta triple-spiral repeats of the shaft.¹⁹

Overview of the structure

Electron micrographs revealed sigmaCΔ1-114 to be a globular protein containing a short vestige of stalk with axial dimensions of 80 Å long and 50 Å wide at the widest point.²⁵ The crystal structure of the somewhat smaller sigmaCΔ1-150 fragment substantiates these data (Figure 2). The structure can be divided into two parts, a globular beta-barrel or head domain (residues 196–326) and a short stalk (residues 151–195) containing two triple beta-spiral repeats. With regards to secondary structure elements, both domains contain almost exclusively beta-strands, with the exception of two short helices in the long DE-loop of the head domain.

The carboxy-terminal globular head domain

The head domain of each monomer (amino acid residues 196–326) forms a beta-barrel with the same topology as the sigma1 head domain.¹⁹ The beta-barrel contains eight beta-strands, forming a “continuous” circular DABCHEFG-sheet in which the G-strand interacts with the D-strand. The sheet can be further subdivided into two Greek key motifs, beta-sheets DABC and HEFG.¹⁹ The C and H-strands contain kinks, although not caused by proline residues like in sigma1. Most loops between the beta-strands are short, with the exception of the DE-loop, which contains 22 residues. In our crystal structures, this loop contains a short beta-strand (D'), interacting with the A-strand and two alpha-helices (Pro257–Val264 and Ser266–Phe269). In sigma1, the receptor-binding residues have been proposed to be located mainly in the DE-loop.¹⁹

Alignment of all known avian reovirus sigmaC sequences (identical amino acid residues shown in bold type in Figure 3) revealed that the very amino-terminal residues 1–50 show a high degree of conservation (19 of 50 residues or 38% identity), while the heptad repeat region (51–156) has fewer identical residues (only 3 of 155 residues or 2%), although generally maintaining the hydrophobic a and d positions, and presumably, the capacity to form a triple coiled-coil structure. The carboxy-terminal domain (amino acid residues 156–326) also has a high sequence identity (60 out of 171 or 35%). This is consistent with a model wherein the amino-terminal sequence is involved in binding lambdaC while the heptad repeat region plays a structural role, projecting the conserved carboxy-terminal receptor-binding domain away from the body of the virus, providing “reach”.

The observation that the DE-loop is the only long loop in the head domain, and the fact that the DE-loop has been proposed to be involved in receptor binding for sigma1,¹⁹ suggests that this loop may also contribute receptor-binding residues in sigmaC. Of the DE-loop residues, Asp259, Arg262, Leu263, Pro265, Gly268, Phe269, Gln270 and Ala272 are fully conserved among all strains of avian reovirus with a known sigmaC sequence (Figure 3). However, of these residues, only Asp259 is fully solvent-exposed and for the other conserved amino acid residues to be involved in receptor binding some structural rearrangement of the loop would be necessary.

In Figure 4, all universally conserved and solvent-exposed residues of the carboxy-terminal domain are shown. As can be seen from the Figure, conserved residues are scattered on the surface apart from perhaps those located in a region on the border of the stalk and the head domains and in a cleft between two monomers. If the receptor-binding region would be the inter-monomer cleft, this would be analogous to the case of adenovirus.²⁶ Site-directed mutagenesis experiments and, once the receptor for avian reovirus had been identified, cross-linking or co-crystallisation studies will be needed to identify the sigmaC-residues responsible for receptor-binding.

The stalk

The short vestige of the stalk present in the crystal structure contains two beta-spiral repeats (amino acid residues 160–191). Several extra residues (151–156) are resolved in the IIz crystal form, but at this stage it is still unclear whether they represent an

Figure 2. Crystal structure of avian reovirus sigmaCΔ1-150. (a) Stereo view of C^α atoms connected by lines. The three monomers are coloured red, green and blue, and every tenth residue of the red monomer is numbered. (b) Ribbon diagram in the same orientation and colouring scheme as in (a). In the red monomer, secondary structure elements are identified with upper case letters for beta-strands and lower case letters for helices. (c) The first triple beta-spiral repeat consisting of amino acid residues 160–171 in ball-and-stick representation covered by the final 2F_o–F_c electron density map contoured at a σ level of 1.0. (d) Top view of the sigmaC trimer. Inter-monomer salt bridges are shown. Monomer backbones are coloured yellow, orange and green. Figures 2, 4 and 5 were prepared using Bobscript,⁵⁰ a modified version of MOLSCRIPT.⁵¹

```

          10          20          30          40          50          60
sC 1133  |          |          |          |          | a d a |
          *   *   *   *   *   *   *   *   *   *   *   *
s1 reo3  MD---PRLREEVRLIIALTSNNGVSLSK--GLESRVSALEKTSQIHSDTILRITQGLDD
          |          |          |          |          a a | a
          10          20          30          40          50

          70          80          90          100          110          120
sC 1133  d          a d | a d a | d a d | a d a d          d a d |
          ISA-----NLQDMTHLDDVTANLDGLRRTVTALQDSVSI LSTNVTDLTNRSSAHAAI LSSLQTT
          *           *           *           *** *           * * * * *
s1 reo3  ANKRIIALEQSRDDLVASVSDAQLAISRLLESSIGALQTVVNGLDSSVTQLGARVGQLETGLADVRVD
          d a d          d a d | a d a | d a d | a d a | d a d |
          60          70          80          90          100          110          120

          130          140          150
sC 1133  a          a | d a d | a d a | d a
          V DGNSTAI SNL KSDISSNGLA ITDLQDRVKSLESTASHG-----
          *           * *           * **
s1 reo3  HDNLVARVDTAERNIGSLTTELSTLTLRVTSIQADFESRISTLERTAVTSAGAPLSIRNNRMTMGLND
          a d a d a d a d | a d a |          sssssss ssss sssssss|
          130          140          150          160          170          180          190

sC 1133  -----
s1 reo3  GLTLSGNLAI RLPNGTGLNIQNGGLQFRFNTDQFQIVNNNLT LKTTVFDSINSRIGATEQSYVASAVTP
          ssss sssssss ssss sssssss ssss | sssssss | ssss | sssssss |
          200          210          220          230          240          250          260

          160          170          180          190          200          210
sC 1133  SSSS SSSSS | SSSSSSSS | SSSS SSSSS SSSSSSSSSSSSS |
          -----LSFSPPLSVA--DGVVSLDMDP-----YFCSQRVSLTSYS-AEAQLMQFRWMARGTN-GS
          * * * * *
s1 reo3  LRLNSSTKVLDMLIDSSTLEINSSGQLTVRSTSPNLRYPADVSGGIGMSPNYRFRQSMWIGIVSYSGSGL
          SSSS SSSSSSS SSSS SSSSSSS SSSS SSSS HHHH SSSSSSSSSSSSS |S
          270          280          290          300          310          320          330

          220          230          240          250          260          270          280
sC 1133  SSSSSSSSSSS SSSSSSSSS SSSSSSSSSSSSSSS SSSS HHHHHHHH HHHH | SSSSS-SSS
          SDTIDMTVNAHCHGRRTDYMSSTGNLTVTSNVVLLTFDLSDITHIPSDLARLVPSAGFQAASFPVD-VSF
          * * * * *
s1 reo3  NWRVQVNSDIFIVDDYIHICLPAPF-DGFSIADGGDLNLFVVTGLLPPLL TGDTEPAFHNDVVTYGAQTVAI
          SSSSSSSSSSSSS SSSSSS S-SSSS | SSSSSSSS HHHH | SSSSSSSS
          340          350          360          370          380          390          400

          290          300          310          320
sC 1133  SS SSSSSSSSSSS SSSSSSSSS SSSSSSSSSSSSSSS
          TRD----SATHAYQAYGVYSSSRVFTITFPPT-GGDGTANIRSLTVRTGIDT 326
          * * *
s1 reo3  GLSSGGTPQYMSKNLWVEQWQDGVLRRLRVEGGGSITHSNSKWPAMTVSYPRSFT 455
          SSSS SSSSSSSSSSS SSSSSSSSS SSSS SSSSSSSS
          410          420          430          440          450

```

Figure 3. Structure-based sequence alignment of avian reovirus S1133 sigmaC and mammalian reovirus type 3 Dearing sigma1. The sigmaC structure was superimposed onto the sigma1 structure¹⁹ using the DALI server³³ and equivalent residues of the structure were aligned. The amino-terminal domains (sigmaC 1–150 and sigma1 1–249), for which no structure is known, were aligned by hand. Known secondary structure elements are indicated with H for helices and S for beta-strands; predicted beta-strands of the putative five extra triple beta-spiral repeats¹⁹ with s; heptad repeats are indicated with the letters a and d for the first and fourth hydrophobic positions, respectively. Every tenth amino acid is numbered, with the corresponding residue under or above the (last) zero of the number. Stars indicate amino acid residues that are conserved between avian reovirus S1133 sigmaC and mammalian reovirus strain Dearing; those conserved in avian reovirus strains are in bold. Avian reovirus strains used for alignment were S1133 (Uniprot accession number Q51238);⁵² 138 (Q9J1B3) and 176 (Q9J1B6);²⁷ GuangxiR1 (Q4I515);⁵³ GEL1298M (Q8QV12), GEI0997M (Q8QV17), GEL0697M (Q8QV13), GEL13a98M (Q8QV11), GEL0597M (Q8QV14), NLA1396T (Q8QV09), NLI1296M

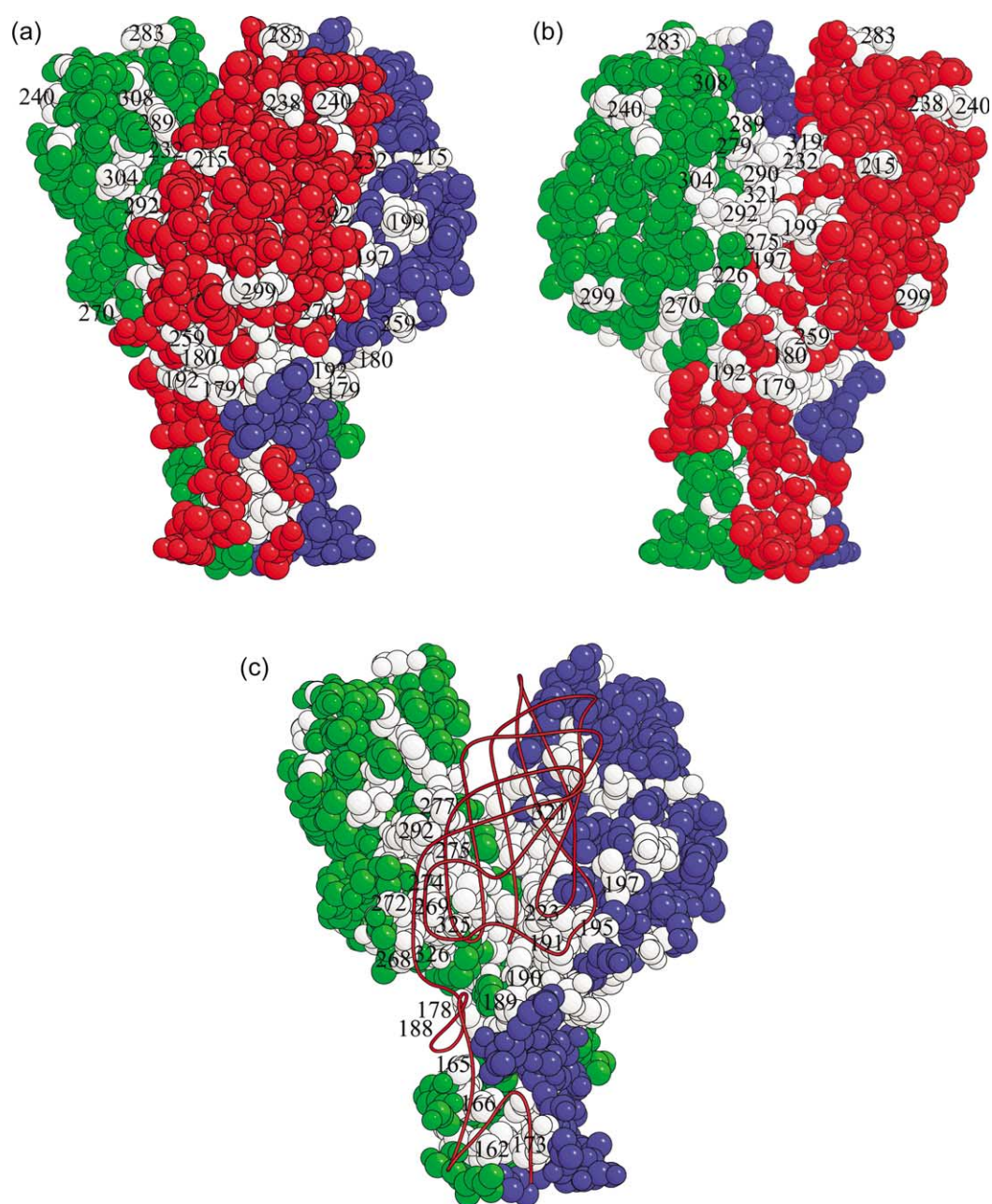


Figure 4. Mapping of conserved residues onto the structure of sigmaC. (a) and (b) The structure of residues 160–326 of crystal form IIz are shown in space-filled representation and the locations of solvent-accessible residues (as judged by the program SURFACE⁴⁰) conserved in all known avian orthoreovirus strains are indicated. The view in (b) is rotated 60° around the vertical trimer axis as compared to (a). (c) Conserved amino acid residues in the trimer interface. Monomer A is shown as a coil, B and C in space-filled representation. The view is the same as in (a). The locations of residues conserved in all known avian orthoreovirus strains contacting another monomer are indicated (contact less than 4 Å as calculated by the program CONTACT⁴⁰). The location of Asp277 is also shown, although in some avian reovirus strains it is replaced by a Glu, presumably still able to form a salt-bridge with Arg321.

actual physiological conformation, as they are involved in a lattice contact, coordinating a zinc ion. Furthermore, residues 157–159 have somewhat different conformations in the three crystal forms, suggesting that the observed conformations are

influenced by lattice contacts in this region. The relative disorder of residues 151–159 could help explain why a species starting at amino acid residue 156 was observed after trypsin digestion. A contaminating protease may have removed

(Q8QV07), GEL0397T (Q8QV15), GEL0196T (Q8QV16), NLI0298M (Q8QV08), GEI1097M (Q8QV18) and GEL13b98M (Q8QV10);⁵⁴ 750505 (Q91LM3), 919 (Q91LM4), 601SI (Q91LM5), T6 (Q91CG7), S1733 (O12287), 601G (Q9DPB1), 1017-1 (Q9DPB2), 918 (Q9DPB3), 916 (Q9DPB4), R2/TW (Q9DPB5), Somerville4 (Q83931) and RAM-1 (Q84144).

residues 151–155; although it is not clear why Thr155 should be especially labile. Interestingly, the region comprising amino acid residues 140–160 is relatively conserved in sigmaC, even when taking into account Nelson Bay,²⁷ muscovy duck²⁸ and reptilian²⁹ reovirus sigmaC proteins, and therefore may play some structural role, perhaps as a hinge region important for conversion of the reovirion into an infectious particle, a process which uncovers the reovirus fibre, or in the infection itself, either in receptor access or subsequent internalisation.

The first triple beta-spiral repeat (residues 160–176) is of the proline-type,^{24,19} and contains a type II beta-turn, with Pro165 adopting the *cis*-conformation (Figure 2(c)). The second triple beta-spiral repeat (amino acid residues 177–191) is rather irregular, with residues 179–182 forming a type VIb beta-turn.³⁰ The positions of the hydrophobic residues are not strictly conserved in this repeat, the beta-strand immediately following the beta-turn has only one hydrophobic (Cys182) residue contributing to the hydrophobic core, similar to the beta-strand following the solvent-exposed loop consisting of amino acid residues 184–187 (Leu189 in this case). The loss of hydrophobic amino acids in the beta-strands is compensated by Phe181 of the beta-turn, which projects into the centre, capping the longitudinal hydrophobic core of the stalk.

Stability and folding of sigmaC

Considering the region comprising amino acid residues 157–326 and ignoring residues 151–156, which may adopt non-physiological conformations as described above, the surface area of a monomer is around $9.2 \times 10^3 \text{ \AA}^2$, of which 24% is buried upon trimer formation (2.2×10^3 – $2.3 \times 10^3 \text{ \AA}^2$). The stalk domain, being intertwined, buries $1.2 \times 10^3 \text{ \AA}^2$ of its monomer surface into the trimer (33%), while the head domain only contributes $0.7 \times 10^3 \text{ \AA}^2$ (11%). This relatively large buried surface area at least partly explains the extraordinary stability of trimeric sigmaC, as revealed by the fact that it does not dissociate into monomers in SDS-PAGE, unless previously boiled in SDS-containing buffer.^{20,21,23} Grande *et al.*²¹ found that there is no disulphide bond formation in the sigmaC trimer. Our structure confirms the absence of disulphide bonds in the C-terminal construct, and this fragment contains the only two cysteine residues. The surface buried in the trimer is of mixed nature, the central core of the triple beta-spiral contains only hydrophobic residues, while the contacts between two adjoining monomers of the head domain are both hydrophobic and polar and include possible salt bridges (for example His223 with Asp325 of one neighbouring monomer and Asp277 with Arg321 of the other neighbouring monomer; see Figure 2(d)). The biochemical results²¹ suggest that the hydrophobic effect is larger than the effect of polar interactions and salt-bridges, because detergents were shown to destabilise the sigmaC trimer, while salts actually stabilised it. In sigma1¹⁹ and adeno-

virus,^{31,21} the trimer contacts are very similar, with a central hydrophobic core in the shaft domain and more limited, mixed contacts in the head domain.

SigmaC folding may begin with a monomeric carboxy-terminal beta-barrel. Interaction of three beta-barrels could then lead to trimer formation. The stalk region would then “zip up”, starting with the short triple beta-spiral, to form the intact fibre. The head and/or stalk domains could act as signals to ensure correct registration. This folding pathway would be the same as that proposed for adenovirus fibre.³² Alternatively, trimerisation could start with the N-terminal alpha-helical coiled coil domain. In this case, residues of the amino terminus or of the coiled-coil would have to ensure correct registration.

Comparison with mammalian reovirus fibre sigma1

A structure-based sequence alignment of avian reovirus sigmaC with mammalian reovirus type 3 sigma1 (the only other reovirus fibre for which a partial structure is known¹⁹) reveals low sequence identity (Figure 3), explaining the difficulties encountered in attempts to solve the structure by molecular replacement using sigma1 as a model. Sequence identity is highest for the amino-terminal 21 residues, which are likely to play a role in the interaction with lambdaC of the reovirus core. LambdaC shares 27% sequence identity with mammalian reovirus type-3 lambda2, and is likely to adopt a similar conformation; lambdaC and sigmaC very probably interact in a similar fashion as their mammalian homologues lambda2 and sigma1. Both the triple beta-spiral repeat region and the beta-barrel head domain do not show appreciable sequence identity between sigmaC and sigma1, but their crystal structures show identical topology and very similar dimensions (see below).

Sequence inspection (Figure 3) did not reveal the presence of more beta triple-spiral repeats amino-terminal to the two present in our crystal structure, instead heptad repeats are apparent between residues 51–156, suggesting this part forms a triple coiled-coil. In sigma1 a total of eight triple beta-spiral repeats have been inferred¹⁹, making mammalian reovirus sigma1 a longer fibrous protein. Sigma1 also has a hinge inserted just amino-terminal of the C-terminal beta-spiral repeat, while in sigmaC the two triple beta-spiral repeats are packed closely together and a possible hinge region is located around amino acid residues 150–155, as discussed above. The predicted triple coiled-coil domains in the two proteins are similar in length, with mammalian reovirus type 3 containing one more predicted heptad repeat than sigmaC. Superposition of the structurally equivalent residues of the head domain as identified by the DALI³³ server leads to r.m.s. differences of 2.5 Å for 127 superimposed C α atoms when the individual monomers are superimposed (Figure 5(a)); the head domains of sigmaC and sigma1 contain 133

and 142 amino acid residues, respectively). Not only does the sigmaC head domain have the same topology as the sigma1 head domain, but also the loops connecting the beta-strands are similar, with mostly very short loops and one long DE-loop.

Figure 5(b) shows a superimposition of the sigma1 trimer onto sigmaC; superimposition of 3×127 head C^α atoms leads to an r.m.s. difference of

5.3 Å. Clearly the trimers do not superimpose as well as the monomers, due to the different orientation of the monomers in the trimer. Although, in principle, this could be due to different crystal contacts, we did not detect differences in orientations between our three different sigmaC crystal forms. The superimposition shows the three head domains in the trimer are more “splayed out”



Figure 5. Comparison of the structure of the mammalian reovirus fibre, sigma1, with that of sigmaC. (a) Superimposition of a sigma1 monomer¹⁹ onto sigmaC (crystal form IIz). (b) Superimposition of a sigma1 trimer onto a sigmaC trimer generated from crystal form IIz. SigmaC is shown in white, sigma1 in black. Superimpositions were done using the program LSQKAB,⁴⁰ using only the sigmaC and sigma1 head domains.

in sigmaC than in sigma1. In both cases, the interior of the trimer contains a solvent filled space, but in the case of sigmaC this space is cone-shaped and wider at the top than at the bottom, rather than more-or-less cylindrical as in sigma1. This means trimer interactions do not extend all the way to the top of the trimer like in sigma1, but are present only in the stalk and the bottom half of the head domain. However, the monomer surface involved in trimer contacts is larger for sigmaC than that reported for sigma1¹⁹ (~700 *versus* ~450 Å²). There does not appear to be a single structural reason for the sigmaC trimer being more splayed-out than sigma1, rather it is a sum of many separate interactions (charged, polar and hydrophobic) making the interaction surfaces more compatible in this orientation. The different orientation may well be important for receptor interaction in the avian system.

In sigma1, trimer contacts involve potentially unstable Asp-Asp side-chain interactions; in sigmaC these contacts are absent, instead there are two rings of salt-bridges in the centre of the head domain (see Figure 2(d)). The relatively small contact area in the sigma1 head trimer and the Asp-Asp contacts led to the proposal that the trimeric nature of the sigma1 head domain may be transient, and the head domains may be monomeric in the reovirion (trimerising upon conversion to infectious particles) and/or after receptor binding¹⁹. The receptor-binding region proposed for sigma1 is consistent with this as it involves residues from a single monomer. In the case of sigmaC, the trimer interface is predicted to be more stable and the receptor-binding site may involve an intermonomer cleft (as discussed above); so perhaps sigmaC stays trimeric even after receptor binding. On the other hand, we cannot rule out the effect a change in pH may have on the possible salt-bridges and thus on the structure. In this context, it is worth noting that the protein was crystallised at pH 8.5, not too far from physiological values.

Comparison with other viral fibre proteins

Apart from sigma1, a DALI³³ search also turned up similarity with adenovirus fibre,²⁴ which in turn is structurally related to the spike of bacteriophage PRD1.³⁴ Both these proteins are trimeric and have triple beta-spiral repeats in the shaft domain, 21 in the case of adenovirus type 2 and two in the case of the PRD1 spike protein. Both also contain beta-barrels in their head domains; however, they are topologically different. Presumably, avian and mammalian reoviruses share a common double-stranded RNA-virus ancestor; this ancestor in turn has been proposed to be evolutionarily related to double-stranded DNA-viruses like adenovirus and PRD1.^{19,34} A feature that all these viral fibres share is the presence of hinges, which are probably important in the infection process. Thus, adenovirus fibre has hinges in the amino-terminal part of

the shaft domain and between the shaft and head domains,²⁴ PRD1 P5 also has it between the stalk and the head domain³⁴ and sigma1 in the carboxy-terminal part of the stalk.¹⁹ The structure of the sigmaC carboxy-terminal fragment does not show the presence of hinges, although they may be located in the unsolved amino-terminal part of sigmaC.

The only other beta-structured viral fibre shaft structure known so far, the T4 short tail fibre, has a different fold, lacking triple beta-spiral repeats³⁵ and with a head domain consisting of three intertwined monomers,³⁶ rather than composed of individually folded monomeric domains. A common feature of all these viral fibres is that their carboxy-terminal beta-strand is folded into the centre of the structure, a feature presumably important for conferring resistance towards cleavage by proteases that operate in the carboxy to amino-terminal direction.

Conclusion

We have solved the structure of a major, receptor-binding domain of the avian reovirus fibre, sigmaC. The structure provides insights at the molecular level into the complex interactions within the homo-trimer required to form an intact tail fibre and suggests regions that may be involved in receptor binding. The structural data will be of use in elucidating the early steps of the avian reovirus replication cycle and may lead to therapeutic applications for avian diseases, through the rational design of receptor-binding inhibitors or the generation of artificial, chimeric fibres based on the sigmaC structure, which would be useful as vaccination vectors.

Materials and Methods

Protein production

A trimeric sigmaC fragment containing amino acid residues 151–326 was produced as described.²³ Briefly, sigmaC (amino acid residues 117–326) was expressed, partially purified and treated with trypsin to produce a further truncated protein, which was then purified by quaternary ammonium anion-exchange chromatography (Uno-Q; Biorad, Madrid, Spain). SDS-PAGE combined with amino-terminal sequence analysis revealed that the protein contains sigmaC amino acid residues 151–326 or 156–326.

Inhibition of avian reovirus infection by the carboxy-terminal sigmaC fragment

Monolayers of chick embryo fibroblasts growing in 12-well plates were washed with cold PBS (137 mM sodium chloride, 2.7 mM potassium chloride, 8 mM disodium hydrogen phosphate, 15 mM potassium dihydrogen phosphate), overlaid with 0.5 ml of 199 medium (Invitrogen Gibco Cell Culture Systems, Barcelona, Spain)

containing 5 mg/ml of bovine serum albumin (BSA) and placed on ice for 15 min. Then, an additional 150 µg/ml of protein was added to each well, containing increasing amounts of sigmaC and decreasing amounts of BSA, as shown in Figure 1. After 15 min of incubation, one plaque-forming unit/cell of avian reovirus S1133 was added to each well and allowed to adsorb for 30 min at 4 °C. The cells were then washed extensively with cold 199 medium containing 10% (v/v) newborn calf serum (Invitrogen Gibco Cell Culture Systems, Barcelona, Spain) and finally overlaid with pre-warmed 199 medium containing 2% newborn calf serum and incubated at 37 °C for 8 h to allow primary infection only. Infected cells were lysed in 50 mM Tris-HCl (pH 7.5), 150 mM sodium chloride, 1% (v/v) Nonidet P-40 (Roche Diagnostics GmbH, Mannheim, Germany), 0.5% (w/v) sodium deoxycholate, 0.1% (w/v) SDS, and the lysates were analysed for the production of the non-structural viral protein muNS, by Western blotting as described.¹² Production and use of primary antibodies against the non-structural viral protein muNS have been described.¹¹

Crystallographic data collection and processing

Crystallogenes and crystal mounting have been described.²³ Native data were collected on beam-lines BM16 and ID23-1 at the European Synchrotron Radiation Facility. Data were obtained from three crystals, one of crystal form I, belonging to space group $P6_322$ and two of form II (one grown in the presence of 10 mM zinc sulphate, hereinafter denoted IIz, and one in the presence of 10 mM cadmium sulphate, hereinafter denoted IIc). Reflections were integrated with the programme MOSFLM,^{37,38} scaled using SCALA³⁹ and further processed using programs from the Collaborative Computational Project Number 4.⁴⁰ For data statistics, see Table 1. For collection of derivative data, crystals of form IIc were soaked for several hours in methyl mercury chloride (a few grains of methyl mercury chloride powder were added to the reservoir and then 1 µl of the reservoir was mixed with the drop containing the crystals). Datasets at three different wavelengths were measured at the Spanish CRG BM16 beam-line at the European Synchrotron Radiation Facility. The derivative data were processed as described above; measurement parameters and statistics can also be found in Table 1.

Structure solution and refinement

Datasets were scaled together using XPREP (Bruker Nonius B.V., Delft, The Netherlands). Four putative mercury sites were identified by the SHELXD program⁴¹ and phases were refined using SHELXE,⁴² using data between 20 Å and 3.1 Å. Further solvent flattening and histogram matching was done with the program DM⁴³ and automated building of a partial model proceeded using ARP-WARP⁴⁴ and data between 20 Å and 2.22 Å. A model containing 160 residues including amino acid side-chains was obtained, but partly out-of-phase. This model was used for molecular replacement into higher resolution native data obtained from a crystal of form IIz. The model was then input into the ARP-WARP auto-trace mode using the data to 2.1 Å resolution, resulting in a model containing 159 residues, all correctly docked into the sequence. Adjustment of the model and addition of 17 extra amino acid residues was done with O.⁴⁵ Refinement was done using the REFMAC program.⁴⁶ Water molecules were built using ARP.⁴⁷ The final IIz model

contains 176 amino acid residues, three sulphate ions, a divalent zinc ion and 164 water molecules.

Simultaneously, the structures of form IIc and I were solved and refined. To this end, a model partially refined against the form IIz data was used in molecular replacement using the AMoRe program.⁴⁸ Models were adjusted in O and refined as described above. The final model for crystal form IIc contains 170 amino acid residues, one sulphate ion, two divalent cadmium ions and 120 water molecules; the final model for crystal form I 167 amino acid residues and 44 water molecules. Statistics can be found in Table 1.

Database codes

Coordinates and structure factors are available from the Protein Structure Database under codes 2BSF, 2BT7 and 2BT8 for crystal forms IIz, IIc and I, respectively.

Acknowledgements

We thank Guy Schoehn (Laboratoire de Virologie Moléculaire et Structurale, Grenoble, France) for electron microscopy and Dave Hall (European Synchrotron Radiation Facility, Grenoble, France) for help with data collection on crystal form I. This research was funded by research grants BMC2002-02436 (to M.J.v.R.) and BMC2001-2839 (to J.B.) from the Spanish Ministry of Science and Technology and PGIDIT03PXIC20307PN (to M.J.v.R.) and PGIDIT02PXIC20301PN (to J.B.) from the Xunta de Galicia. Our grants from the Spanish Ministry of Science and Technology were co-financed by the European Union FEDER programme for 70% of their value. M.J.v.R. is supported by a "Ramón y Cajal" contract of the Spanish Ministry of Education and Science.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2005.09.034

References

1. Songserm, T., van Roozelaar, D., Kant, A., Pol, J., Pijpers, A. & ter Huurne, A. (2003). Enteropathogenicity of Dutch and German avian reoviruses in SPF white leghorn chickens and broilers. *Vet. Res.* **34**, 285–295.
2. Spandidos, D.A. & Graham, A.F. (1976). Physical and chemical characterization of an avian reovirus. *J. Virol.* **19**, 968–976.
3. Hsiao, J., Martinez-Costas, J., Benavente, J. & Vakharia, V.N. (2002). Cloning, expression, and characterization of avian reovirus guanylyltransferase. *Virology*, **296**, 288–299.
4. Martinez-Costas, J., Gonzalez-Lopez, C., Vakharia, V.N. & Benavente, J. (2000). Possible involvement of

- the double-stranded RNA-binding core protein sigmaA in the resistance of avian reovirus to interferon. *J. Virol.* **74**, 1124–1131.
5. Gonzalez-Lopez, C., Martinez-Costas, J., Esteban, M. & Benavente, J. (2003). Evidence that avian reovirus sigmaA protein is an inhibitor of the double-stranded RNA-dependent protein kinase. *J. Gen. Virol.* **84**, 1629–1639.
 6. Martinez-Costas, J., Grande, A., Varela, R., Garcia-Martinez, C. & Benavente, J. (1997). Protein architecture of avian reovirus S1133 and identification of the cell attachment protein. *J. Virol.* **71**, 59–64.
 7. Bodelon, G., Labrada, L., Martinez-Costas, J. & Benavente, J. (2002). Modification of late membrane permeability in avian reovirus-infected cells: viroporin activity of the S1-encoded non-structural p10 protein. *J. Biol. Chem.* **277**, 17789–17796.
 8. Shmulevitz, M., Eband, R.F., Eband, R.M. & Duncan, R. (2004). Structural and functional properties of an unusual internal fusion peptide in a nonenveloped virus membrane fusion protein. *J. Virol.* **78**, 2808–2818.
 9. Shmulevitz, M., Corcoran, J., Salsman, J. & Duncan, R. (2004). Cell–cell fusion induced by the avian reovirus membrane fusion protein is regulated by protein degradation. *J. Virol.* **78**, 5996–6004.
 10. Salsman, J., Top, D., Boutilier, J. & Duncan, R. (2005). Extensive syncytium formation mediated by the reovirus FAST proteins triggers apoptosis-induced membrane instability. *J. Virol.* **79**, 8090–8100.
 11. Touris-Otero, F., Martinez-Costas, J., Vakharia, V.N. & Benavente, J. (2004). Avian reovirus nonstructural protein muNS forms viroplasm-like inclusions and recruits protein sigmaNS to these structures. *Virology*, **319**, 94–106.
 12. Touris-Otero, F., Cortez-San Martin, M., Martinez-Costas, J. & Benavente, J. (2004). Avian reovirus morphogenesis occurs within viral factories and begins with the selective recruitment of sigmaNS and lambdaA to muNS inclusions. *J. Mol. Biol.* **341**, 361–374.
 13. Touris-Otero, F., Martinez-Costas, J., Vakharia, V.N. & Benavente, J. (2005). Characterization of the nucleic acid-binding activity of the avian reovirus non-structural protein sigma NS. *J. Gen. Virol.* **86**, 1159–1169.
 14. Costas, C., Martinez-Costas, J., Bodelon, G. & Benavente, J. (2005). The second open reading frame of the avian reovirus S1 gene encodes a transcription-dependent and CRM1-independent nucleocytoplasmic shuttling protein. *J. Virol.* **79**, 2141–2150.
 15. Reinisch, K.M., Nibert, M.L. & Harrison, S.C. (2000). Structure of the reovirus core at 3.6 Å resolution. *Nature*, **404**, 960–967.
 16. Tao, Y., Farsetta, D.L., Nibert, M.L. & Harrison, S.C. (2002). RNA synthesis in a cage, structural studies of reovirus polymerase lambda3. *Cell*, **111**, 733–745.
 17. Olland, A.M., Jane-Valbuena, J., Schiff, L.A., Nibert, M.L. & Harrison, S.C. (2001). Structure of the reovirus outer capsid and dsRNA-binding protein sigma3 at 1.8 Å resolution. *EMBO J.* **20**, 979–989.
 18. Liemann, S., Chandran, K., Baker, T.S., Nibert, M.L. & Harrison, S.C. (2002). Structure of the reovirus membrane-penetration protein, Mu1, in a complex with its protector protein, Sigma3. *Cell*, **108**, 283–295.
 19. Chappell, J.D., Protá, A.E., Dermody, T.S. & Stehle, T. (2002). Crystal structure of reovirus attachment protein sigma1 reveals evolutionary relationship to adenovirus fiber. *EMBO J.* **21**, 1–11.
 20. Grande, A., Rodriguez, E., Costas, C., Everitt, E. & Benavente, J. (2000). Oligomerization and cell-binding properties of the avian reovirus cell-attachment protein sigmaC. *Virology*, **274**, 367–377.
 21. Grande, A., Costas, C. & Benavente, J. (2002). Subunit composition and conformational stability of the oligomeric form of the avian reovirus cell-attachment protein sigmaC. *J. Gen. Virol.* **83**, 131–139.
 22. Barton, E.S., Forrest, J.C., Connolly, J.L., Liu, Y., Schnell, F.J., Nusrat, A. *et al.* (2001). Junction adhesion molecule is a receptor for reovirus. *Cell*, **104**, 441–451.
 23. van Raaij, M.J., Hermo Parrado, X.L., Guardado Calvo, P., Fox, G.C., Llamas-Saiz, A.L., Costas, C. *et al.* (2005). Crystallization of the C-terminal globular domain of avian reovirus fibre. *Acta Crystallog. sect. F*, **61**, 651–654.
 24. van Raaij, M.J., Mitraki, A., Lavigne, G. & Cusack, S. (1999). A triple beta-spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature*, **401**, 935–938.
 25. Costas Iglesias, C. (2004). Caracterización de las proteínas sigmaC y p17 del reovirus aviar S1133. PhD thesis, Universidad de Santiago de Compostela.
 26. Bewley, M.C., Springer, K., Zhang, Y.B., Freimuth, P. & Flanagan, J.M. (1999). Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science*, **286**, 1579–1583.
 27. Shmulevitz, M., Yameen, Z., Dawe, S., Shou, J., O'Hara, D., Holmes, I. & Duncan, R. (2002). Sequential partially overlapping gene arrangement in the tricistronic S1 genome segments of avian reovirus and Nelson Bay reovirus: implications for translation initiation. *J. Virol.* **76**, 609–618.
 28. Kuntz-Simon, G., Le Gall-Recule, G., de Boisseson, C. & Jestin, V. (2002). Muscovy duck reovirus sigmaC protein is atypically encoded by the smallest genome segment. *J. Gen. Virol.* **83**, 1189–1200.
 29. Duncan, R., Corcoran, J., Shou, J. & Stoltz, D. (2004). Reptilian reovirus: a new fusogenic orthoreovirus species. *Virology*, **319**, 131–140.
 30. Hutchinson, G. & Thornton, J.M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**, 2207–2216.
 31. van Raaij, M.J., Louis, N., Chroboczek, J. & Cusack, S. (1999). Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. *Virology*, **262**, 333–343.
 32. Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J.P., Gagnon, J. & Ruigruik, R.W. (1999). Unfolding studies of human adenovirus type 2 fibre trimers. Evidence for a stable domain. *Eur. J. Biochem.* **264**, 599–606.
 33. Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucl. Acids Res.* **26**, 316–319.
 34. Merckel, M.C., Huisken, J.T., Bamford, D.H., Goldman, A. & Tuma, R. (2005). The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol. Cell.* **18**, 161–170.
 35. van Raaij, M.J., Schoehn, G., Burda, M.R. & Miller, S. (2001). Crystal structure of a heat and protease-stable part of the bacteriophage T4 short tail fibre. *J. Mol. Biol.* **314**, 1137–1146.
 36. Thomassen, E., Gielen, G., Schütz, M., Schoehn, G., Abrahams, J.P., Miller, S. & van Raaij, M.J. (2003). The structure of the receptor-binding domain of the bacteriophage T4 short tail fibre reveals a knitted trimeric metal-binding fold. *J. Mol. Biol.* **331**, 361–373.

37. Powell, H.R. (1999). The Rossmann Fourier auto-indexing algorithm in MOSFLM. *Acta Crystallog. sect. D*, **55**, 1690–1695.
38. Leslie, A.G. (1999). Integration of macromolecular diffraction data. *Acta Crystallog. sect. D*, **55**, 1696–1702.
39. Evans, P.R. (1999). Some notes on choices in data collection. *Acta Crystallog. sect. D*, **55**, 1771–1772.
40. Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
41. Schneider, T.R. & Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallog. sect. D*, **58**, 1772–1779.
42. Sheldrick, G.M. (2002). Macromolecular phasing with SHELXE. *Z. Kristallog.* **217**, 644–650.
43. Cowtan, K. & Main, P. (1998). Miscellaneous algorithms for density modification. *Acta Crystallog. sect. D*, **54**, 487–493.
44. Morris, R.J., Perrakis, A. & Lamzin, V.S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244.
45. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
46. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
47. Lamzin, V.S. & Wilson, K.S. (1997). Automated refinement for protein crystallography. *Methods Enzymol.* **277**, 269–305.
48. Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallog. sect. D*, **57**, 1367–1372.
49. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
50. Esnouf, R.M. (1999). Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallog. sect. D*, **55**, 938–940.
51. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
52. Bodelon, G., Labrada, L., Martinez-Costas, J. & Benavente, J. (2001). The avian reovirus genome segment S1 is a functionally tricistronic gene that expresses one structural and two nonstructural proteins in infected cells. *Virology*, **290**, 181–191.
53. Liao, M., Xie, Z., Liu, J., Deng, X., Pang, Y., Xie, Z. & Tang, X. (2004). Cloning and sequence analysis of sigma3 protein gene of avian reovirus isolates. *Chin. J. Prev. Vet. Med.* **26**, 324–327.
54. Kant, A., Balk, F., Born, L., van Roozelaar, D., Heijmans, J., Gielkens, A. & ter Huurne, A. (2003). Classification of Dutch and German avian reoviruses by sequencing the sigma C protein. *Vet. Res.* **34**, 203–212.

Edited by I. Wilson

(Received 22 July 2005; received in revised form 9 September 2005; accepted 9 September 2005)
Available online 30 September 2005

Short Communication

Crystallographic structure of the α -helical triple coiled-coil domain of avian reovirus S1133 fibre

Pablo Guardado-Calvo,¹ Gavin C. Fox,² Antonio L. Llamas-Saiz³
and Mark J. van Raaij^{1,4}

Correspondence

Mark J. van Raaij
mark.vanraaij@usc.es

¹Departamento de Bioquímica e Bioloxía Molecular, Facultade de Farmacia, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain

²Spanish CRG beamline BM16, European Synchrotron Radiation Facility, 6 rue Jules Horowitz, F-38043 Grenoble, France

³Unidade de Raios X, Laboratorio Integral de Dinámica e Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Campus Sur, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain

⁴Instituto de Biología Molecular de Barcelona-CSIC, c/Josep-Samitier 1-5, E-08028 Barcelona, Spain

Avian reovirus fibre, a homo-trimer of the σ C protein, is a minor component of the avian reovirus outer capsid. It is anchored via a short N-terminal sequence to the inner capsid λ C pentamer, and its protruding globular C-terminal domain is responsible for primary host cell attachment. We have previously solved the structure of a receptor-binding fragment in which residues 160–191 form a triple β -spiral and 196–326 a β -barrel head domain. Here we have expressed, purified and crystallized a major σ C fragment comprising residues 117–326. Its structure, which was solved by molecular replacement using the previously determined receptor-binding domain structure and refined to 1.75 Å (0.175 nm) resolution, reveals an α -helical triple coiled-coil connected to the previously solved structure by a zinc-ion-containing linker. The coiled-coil domain contains two chloride ion binding sites, as well as specific trimerization and registration sequences. The linker may act as a functionally important hinge.

Received 3 November 2008

Accepted 4 December 2008

Avian reovirus, a member of the *Orthoreovirus* genus of the *Reoviridae* family, is an important poultry pathogen (Songserm *et al.*, 2003; Benavente & Martinez-Costas, 2007). This non-enveloped double-stranded RNA virus is composed of a double concentric icosahedral capsid (Spandidos & Graham, 1976), enclosing ten double-stranded RNA genome segments and several copies of the RNA polymerase complex λ B/ μ A. The inner capsid is composed of an icosahedric λ A shell covered by σ A 'staples' and with λ C pentamers at its vertices (Zhang *et al.*, 2005). The outer shell contains the proteins σ B, σ C and μ B.

Avian reovirus fibre is formed by the protein σ C. It is an elongated homotrimer, anchored into the λ C pentamer via its N terminus. Its shaft domain is predicted to contain an α -helical coiled-coil (Shapouri *et al.*, 1995). σ C is responsible for primary host-cell attachment (Shapouri *et al.*, 1996; Grande *et al.*, 2000), presumably through its C-terminal globular domain, and has 20% sequence identity to its functionally equivalent mammalian reovirus protein, σ 1. The mammalian reovirus receptor has been identified as junction adhesion molecule A (JAM-A; Barton *et al.*, 2001). However, σ C probably binds a different, as yet unknown, receptor, because mammalian reoviruses do not

attach to avian cells. Previously, we have crystallized a carboxy-terminal fragment of σ C containing amino acids 151–326 and shown that it retains receptor-binding properties. Its structure revealed two triple β -spiral repeats (amino acids 160–191) and an eight-stranded circular β -barrel head domain (Guardado-Calvo *et al.*, 2005). Here we report the crystallization and structure solution of an extended recombinant avian reovirus S1133 σ C fragment containing amino acids 117–326, refined against data collected to 1.75 Å (0.175 nm) resolution. Our results show the shaft domain has a mixed α -helical coiled-coil and triple β -spiral domain, separated by a flexible linker containing a divalent zinc cation.

For expression, *Escherichia coli* strain JM109 (DE3) transformed with the plasmid pET28-sigmaC117-326 was grown as described (van Raaij *et al.*, 2005). Cells harvested from 4 l of culture were resuspended in 40 ml cold resuspension buffer (10 mM Tris/HCl 8.0, 300 mM sodium chloride), frozen at -20 °C and lysed by a double pass through an Avestin C5 emulsifier (Avestin). After removing insoluble material, 3 ml Ni-nitriloacetic acid resin (Qiagen) was added. The suspension was incubated for 1 h on ice and poured into an empty column. The resin

was washed with resuspension buffer; elution was performed with a step gradient of imidazole pH 7 in resuspension buffer. His-T7-tagged σ C117–326 eluted at imidazole concentrations of 200–1000 mM, was dialysed overnight against TE buffer (10 mM Tris/HCl pH 8.5, 1 mM EDTA pH 8.0) and incubated with $0.2 \mu\text{g ml}^{-1}$ trypsin for 45 min at 37°C . PMSF was added to 1 mM to stop the reaction and the protein was applied onto a 6 ml Uno-Q column (Bio-Rad). σ C117–326 eluted at the start of a linear 0–1 M sodium chloride gradient in TE buffer. The protein was concentrated to 11.5 mg ml^{-1} using Centricon concentrators (Millipore).

Crystallization took place by vapour diffusion in sitting-drop plates with 0.8 ml reservoirs and drops of $5 \mu\text{l}$ protein solution mixed with $5 \mu\text{l}$ reservoir solution. Bar-shaped crystals with hexagonal cross-sections belonging to the trigonal space group *P*321 appeared after 1 day in solutions equilibrated against 0.6–0.75 M ammonium sulphate, 0.1 M Tris/HCl pH 8.4, 25% glycerol and 50 mM zinc sulphate. Rectangular prism with parallelogram base-shaped crystals of the monoclinic space group *C*2 were harvested from a solution containing *P*321 crystals, which had been opened, left open for a while and had then been reclosed, thus partially drying out. Crystals were mounted in cryo-loops and kept at 100 K during data collection. Complete datasets were collected from a crystal of each crystal form, to 2.3 \AA (0.23 nm) for the *P*321 crystal form and to 1.75 \AA (0.175 nm) for the *C*2 crystal form; see Table 1. Reflections were integrated and scaled with the program HKL2000 (Otwinowski & Minor, 1997) and further processed using programs from the Collaborative Computational Project Number 4 (1994).

The structure of the *C*2 crystal form was solved by molecular replacement using protein structure database entry 2BSF (Guardado-Calvo *et al.*, 2005) as a search model in the program MOLREP (Vagin & Teplyakov, 2000). Three monomers were located in the asymmetrical unit, forming the biological trimer. Automatic model building using ARP-WARP (Morris *et al.*, 2003) combined with manual adjustment using the programs O (Jones *et al.*, 1991) and COOT (Emsley & Cowtan, 2004) led to a final model containing 620 residues (amino acids 122–326 for chains A and B; 117–326 for chain C) and the solvent molecules specified in Table 1. Refinement was done using the REFMAC program (Murshudov *et al.*, 1997); reflections selected in thin resolution shells were set apart for calculation of *R*-free (Table 1). Water molecules were built using ARP (Lamzin & Wilson, 1993). Model validation was performed with WHATCHECK (Vriend, 1990), MolProbity (Davis *et al.*, 2007) and MOLEMAN (Kleywegt *et al.*, 2001). The most complete chain of this model, chain C, was used to solve the structure of the *P*321 crystal form by molecular replacement. The coordinates have been deposited in the protein structure database (<http://www.rcsb.org>) under accession code 2VRS for the *C*2 crystal form and 2JJL for the *P*321 crystal form; structure factors are also available. In the *P*321 crystal, the asymmetrical unit consists of a

single monomer and the biological trimer is formed by crystallographic symmetry. When the *P*321 monomer is superimposed on monomers A, B and C of the *C*2 trimer, root mean square differences (r.m.s.d.) of only 0.4–0.6 \AA are observed. If instead the trimeric biological units are superimposed, they superimpose equally well, with an r.m.s.d. of under 0.7 \AA , revealing no significant differences. Therefore only the *C*2 structure will be discussed further. The refined structure shows good correspondence to the data, has good geometry and no residues in very unfavourable regions (>99.8%) of the Ramachandran plot (Table 1).

Electron micrographs revealed σ C115–326 to be a globular protein containing an 80 \AA long stalk and 50 \AA wide at the widest point of the head domain; while structure prediction with the program COILS, a window size of 28 residues and taking 0.15 as probability cut-off, suggests amino acids 53–106 and 126–154 are in a coiled-coil conformation (Costas, 2004; Lupas *et al.*, 1991). The crystal structure of σ C117–326 fragment, with a stalk 90 \AA long and 55 \AA wide at the head domain, largely substantiates these data (Fig. 1). When the overall structure is contemplated, a clear division between shaft (amino acids 117–191) and head domains (residues 196–326) is observed. The shaft domain can be further subdivided (Fig. 1b) into an α -helical triple coiled-coil (amino acids 117–154), a linker region (residues 155–159) and two repeats of a triple β -spiral (Mitraki *et al.*, 2002). The head and triple β -spiral domains have been described before (Guardado-Calvo *et al.*, 2005).

The α -helical region is an uninterrupted coiled-coil structure 53 \AA (5.3 nm) long (Fig. 1b). It is composed of 5.5 heptad repeats, with a pitch length of 166 \AA (16.6 nm), a coiled-coil radius of 6.2 \AA (0.62 nm) and an average number of 3.6 residues per turn (Strelkov & Burkhard, 2002), which is comparable with average values for trimeric coiled-coils (Fraser & MacRae, 1973; Tao *et al.*, 1997). Analysis of buried residues within the coiled-coil, at positions d and a in the heptad repeat (Leu152, Val149, Leu145, Ile142, Asn138, Ile135, Leu131, Ile128, Asn124, Val121 and Leu117) reveals all five residues at position a are β -branched (two valines and three isoleucines), while the six residues at the d position are not (four leucines and two asparagines). It has been proposed that coiled-coil structures with β -branched residues at position a and leucines at position d are more stable as dimers rather than trimers (Harbury *et al.*, 1993), but the parallel trimeric coiled-coils of fibritin (Tao *et al.*, 1997) and human mannose-binding protein (Sheriff *et al.*, 1994) also contain many β -branched residues at position a and leucines at position d. In the other section of predicted coiled-coil region (residues 53–106), of the seven residues at position a, four are valines and three leucines; while of the eight residues at position d, there are four leucines, two isoleucines, a valine and a methionine, i.e. both β -branched amino acids and leucines at positions a and d.

Table 1. Data collection and refinement statistics

Data processing*	Space group	
	C2	P321
Cell parameters	a=142.8 Å†, b=54.0 Å, c=121.3 Å, β=101.8°	a=b=77.7 Å, c=121.4 Å
Beamline	Spanish beamline BM16 (European Synchrotron Radiation Facility)	Spanish beamline BM16 (European Synchrotron Radiation Facility)
Wavelength	0.8266 Å	0.9840 Å
Detector	ADSC Q210r CCD	ADSC Q210r CCD
Crystal-to-detector distance	178.5 mm	178.0 mm
Observed reflections‡	90954 (9036)§	19318 (1858)
Resolution range	20.00–1.75 Å (1.81–1.75 Å)	50.00–2.30 Å (2.38–2.30 Å)
Wilson B	21.7 Å ²	44.1 Å ²
Multiplicity	3.3 (3.2)	20.0 (20.3)
Completeness	0.992 (0.992)	0.990 (0.987)
R _{merge}	0.058 (0.384)	0.068 (0.428)
Refinement 		
Reflections used in refinement‡	89213 (15852)	17714 (3116)
Reflections used for R-free	1738 (329)	1514 (268)
Resolution range	18.43–1.75 Å (1.87–1.75 Å)	25.00–2.30 Å (2.46–2.30 Å)
R-factor¶	0.180 (0.234)	0.181 (0.216)
R-free#	0.218 (0.278)	0.220 (0.301)
r.m.s.d. (bonds / angles)	0.017 Å / 1.6°	0.014 Å / 1.5°
Final model		
Protein / Cl ⁻ / Zn ²⁺ / SO ₄ ²⁻ / glycerol / water atoms	4649 / 2 / 7 / 5 / 18 / 975 /	1551 / 2 / 3 / 5 / 0 / 202
Average B protein / Cl ⁻ / Zn ²⁺ / SO ₄ ²⁻ / glycerol / water	24.6 Å ² / 26.8 Å ² / 28.7 Å ² / 45.7 Å ² / 40.1 Å ² / 42.3 Å ²	37.5 Å ² / 49.1 Å ² / 37.8 Å ² / 69.1 Å ² / - / 45.5 Å ²
Ramachandran statistics**	97.7 % / 100.0 %	97.1 % / 100.0 %

*Results from SCALEPACK (Otwinowski & Minor, 1997).

†One Ångstrom (Å) is 0.1 nm.

‡No σ cut-off or other restrictions were used for inclusion of reflections.

§Values in parentheses are for the highest resolution bin, where applicable.

||Results from REFMAC (Murshudov *et al.*, 1997).

¶R-factor = $\sum ||F_{obs}(hkl)| - F_{calc}(hkl)|| / \sum |F_{obs}(hkl)|$.

#According to Brunger (1997).

**According to the program MolProbity (Davis *et al.*, 2007). The percentages indicated are for residues in favoured and total allowed regions, respectively.

The structure reveals a trimerization motif (Meier *et al.*, 2002; Kammerer *et al.*, 2005), composed of a network of interhelical salt bridges formed by contacts between Arg148 at position g of one chain and Glu153 at position e of the adjacent chain (Fig. 2a). The potential of the trimerization motif to dominate the effect of the hydrophobic core residues has been already demonstrated in short coiled-coils (Burkhard *et al.*, 2000; Kammerer *et al.*, 2004). Interestingly, analysis of the sequence reveals another putative trimerization motif containing Arg36 and Glu41, suggesting the triple-helical coiled-coil extends to the N terminus further than the predicted residue 53. In addition to hydrophobic interactions occurring in the coiled-coil core and polar interactions at the surface, the structure reveals two asparagines situated at position d (Asn124 and Asn138), stabilized by central chloride ions (Fig. 2b). The

electron density features were identified as chlorides based on crystallographic evidence (refinement as chlorides leads to good agreement with the crystallographic data and temperature factors very similar to the surrounding residues), suitability of the coordination distances and similarity to other viral coiled-coils (see below). On the other side of the chloride ions are the hydrophobic side chains of Val121 and Ile135, respectively. It has been postulated that these central polar interactions favour correct 'registration' (Oakley & Kim, 1998); asparagines at position d also favour the trimeric state (Tripet *et al.*, 2000). A central chloride ion coordinated by asparagine residues appears to be a quite common feature in parallel trimeric coiled-coils of viral fusion proteins, fibres and bacterial adhesins (see for example Fass *et al.*, 1996; Tao *et al.*, 1997; Malashkevich *et al.*, 1999; Renard *et al.*, 2005;

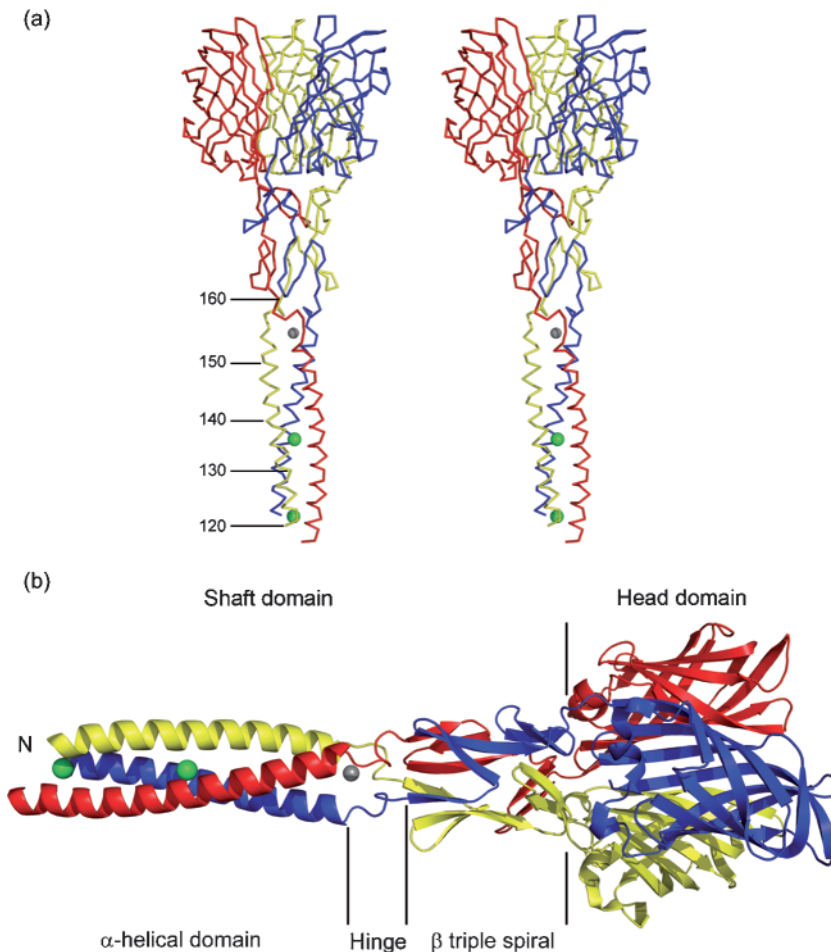


Fig. 1. Overall structure of avian reovirus S1133 fibre σ C117–326. (a) Stereoview of C- α s connected by lines. The monomers are coloured blue, yellow and red, and some of the residues of the yellow monomer are numbered. Green balls, chloride anions; grey ball, zinc cation. (b) Ribbon diagram; secondary structure is shown as corkscrews (α -helices) or arrows (β -strands). The boundaries between the α -helical, hinge and β triple spiral-domains of the shaft domain and head domain are indicated. Figures were prepared using PyMOL (DeLano Scientific; <http://www.pymol.org>).

Conners *et al.*, 2008), and it has been used in *de novo* design of trimeric coiled-coils (Lumb & Kim, 1995). As discussed before (Guardado-Calvo *et al.*, 2005), σ C folding may begin at its monomeric C-terminal β -barrel. Interaction among three β -barrels could then lead to trimer formation. The stalk region would then ‘zip up’, starting with the short triple β -spiral, to form the intact fibre. The head and/or stalk domains could act as signals to ensure correct registration. Alternatively, trimerization could start at the registration domains in the coiled-coil domain identified here.

Amino acids 155–159 link the α -helical coiled-coil to the triple β -spiral (Fig. 2c). Thr155 is in an extended conformation, with residues 155–159 forming two nested β -turns (Ala156 and Ser157 form a type I, His158 and Gly159 a type II β -turn; Richardson, 1981). The three His158 bind a divalent zinc cation through their NE2 atoms (the other six zinc ions in the structural model are on the protein surface and are not likely to have a biological role). The fourth ligand in the distorted-tetragonal coordination environment is a water molecule. Zinc often contributes to catalytic sites and a coordination sphere similar to the one observed here is present in carbonic anhydrase (Auld, 2005). However, as zinc was

added in the crystallization buffer, we cannot be sure that σ C contains zinc in its natural state. However, it is tempting to speculate that zinc, or perhaps another divalent metal cation, has a structural role in stabilizing the extended σ C conformation in the intermediate subviral particle obtained upon reovirion partial uncoating, when outer capsid proteins σ B and μ B are lost (Benavente & Martinez-Costas, 2007). If σ C is folded away in the intact reovirion (as its analogue σ 1 in mammalian reovirus is proposed to be; Guglielmi *et al.*, 2006), its zinc-binding motif may be disordered and function as a hinge region. The presence of the metal ion may be dependent on the protonation state of its histidine residues, and thus on pH. A similar hinge region may be present around His110, which is near to its trypsin cleavage site (see above). This would be consistent with the COILS prediction, where there is a gap in likely coiled-coil formation from amino acid 106 onwards (see above). Residues up to around amino acid 30 probably interact with λ C at the pentameric vertices of the avian reovirus core, and are likely to be unstructured in the isolated protein. It is currently not clear how σ C incorporates into the virus; apart from a small stub of density emerging from the λ C pentamer identified by electron microscopy, no electron density has been assigned

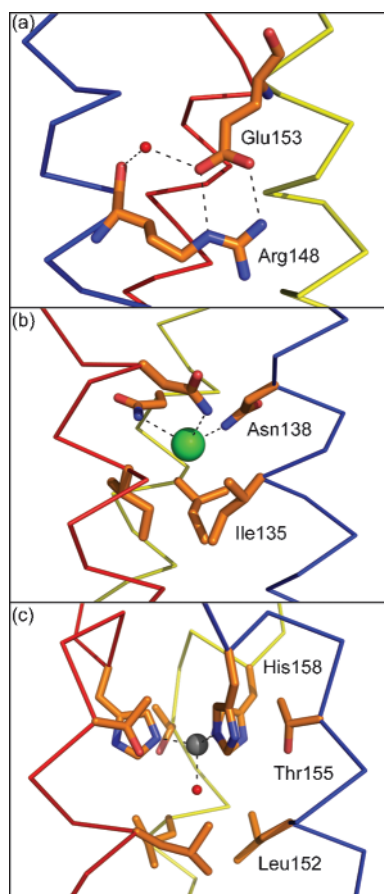


Fig. 2. Structural elements of the σ C fibrous domain. (a) Trimerization motif consisting of Arg148, Glu153 and a water molecule. Only one of the three salt bridges is shown. (b) Coordination of a chloride ion (green ball) by ND2 atoms of Asn138 of each chain [distances just over 3.1 Å (0.31 nm)]. Three isoleucines form a hydrophobic pocket at the other side of the chloride ion. (c) The zinc-containing linker sequence. The three His158 bind a zinc cation (grey) through their NE2 atoms, with coordination distances of 2.0–2.1 Å (0.20–0.21 nm). The fourth ligand is a water molecule, at 2.3 Å (0.23 nm).

to σ C (Zhang *et al.*, 2005). This may be due to σ C adopting different possible conformations and/or to σ C not following the icosahedral symmetry imposed in the electron microscopy reconstruction.

In summary, we have for the first time experimentally determined the structure of the α -helical coiled-coil region of a reovirus fibre and shown that the linker between the α -helical and β -structured parts of the σ C protein may bind a divalent zinc cation. Our structural data suggest the coiled-coil is important for in-register trimerization of the avian reovirus fibre and suggest the presence of hinge regions around residues 110 and 158, which may be important for receptor-binding or subsequent approach of the infectious viral particle to the cell. They may also be important for the accommodation of the σ C trimer in the virus particle.

Acknowledgements

We thank Patricia Ferraces-Casais and Rebeca Menaya-Vargas for technical assistance, Laurent Terradot-Piot and the Partnership for Structural Biology (Grenoble, France) for use of crystallization facilities and Javier Benavente, José Martínez-Costas and David Auld for discussions. This research was sponsored by research grants BFU2005-02974 and BFU2005-24982-E and by a pre-doctoral FPU fellowship to P.G.-C. from the Spanish Ministry of Education and Science. This work was also supported by funds from the European Commission under contracts ERAS-CT-2003-980409 (as part of the European Science Foundation EUROCORES Programme EuroSCOPE) and NMP4-CT-2006-033256 (BeNatural coordinated project).

References

- Auld, D. S. (2005). Zinc enzymes. In *Encyclopedia of Inorganic Chemistry*, 2nd edn, vol. IX, pp. 5885–5927. Edited by R. B. King. Chichester, UK: John Wiley.
- Barton, E. S., Forrest, J. C., Connolly, J. L., Chappell, J. D., Liu, Y., Schnell, F. J., Nusrat, A., Parkos, C. A. & Dermody, T. S. (2001). Junction adhesion molecule is a receptor for reovirus. *Cell* **104**, 441–451.
- Benavente, J. & Martínez-Costas, J. (2007). Avian reovirus: structure and biology. *Virus Res* **123**, 105–119.
- Brunger, A. T. (1997). Free R value: cross-validation in crystallography. *Methods Enzymol* **277**, 366–396.
- Burkhard, P., Meier, M. & Lustig, A. (2000). Design of a minimal protein oligomerization domain by a structural approach. *Protein Sci* **9**, 2294–2301.
- Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.
- Conners, R., Hill, D. J., Borodina, E., Agnew, C., Daniell, S. J., Burton, N. M., Sessions, R. B., Clarke, A. R., Catto, L. E. & other authors (2008). The *Moraxella* adhesin UspA1 binds to its human CEACAM1 receptor by a deformable trimeric coiled-coil. *EMBO J* **27**, 1779–1789.
- Costas, C. (2004). Caracterización de las proteínas σ C y p17 del reovirus aviar S1133 (Characterisation of the proteins σ C and p17 of avian reovirus S1133). PhD thesis, Universidade de Santiago de Compostela (in Spanish).
- Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., III, Snoeyink, J. & other authors (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375–W383.
- Emsley, P. & Cowtan, K. (2004). Model building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
- Fass, D., Harrison, S. C. & Kim, P. S. (1996). Retrovirus envelope domain at 1.7 Å resolution. *Nat Struct Biol* **3**, 465–469.
- Fraser, R. D. B. & MacRae, T. P. (1973). *Conformation in Fibrous Proteins and Related Synthetic Polypeptides*. New York, USA: Academic Press.
- Grande, A., Rodriguez, E., Costas, C., Everitt, E. & Benavente, J. (2000). Oligomerization and cell-binding properties of the avian reovirus cell-attachment protein σ C. *Virology* **274**, 367–377.
- Guardado-Calvo, P., Fox, G. C., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J., Benavente, J. & van Raaij, M. J. (2005). Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J Mol Biol* **354**, 137–149.

- Guglielmi, K. M., Johnson, E. M., Stehle, T. & Dermody, T. S. (2006). Attachment and cell entry of mammalian orthoreovirus. *Curr Top Microbiol Immunol* **309**, 1–38.
- Harbury, P. B., Zhang, T., Kim, P. S. & Alber, T. (1993). A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* **47**, 110–119.
- Kammerer, R. A., Kostrewa, D., Zurdo, J., Detken, A., Garcia-Echevarria, C., Green, J. D., Muller, S. A., Meier, B. H., Winkler, F. K. & Dobson, C. M. (2004). Exploring amyloid formation by a *de novo* design. *Proc Natl Acad Sci U S A* **101**, 4435–4440.
- Kammerer, R. A., Kostrewa, D., Progiass, P., Honnappa, S., Avila, D., Lustig, A., Winkler, F. K., Pieters, J. & Steinmetz, M. O. (2005). A conserved trimerization motif controls the topology of short coiled coils. *Proc Natl Acad Sci U S A* **102**, 13891–13896.
- Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001). Around O. In *International Tables for Crystallography, vol. F. Crystallography of Biological Macromolecules*, pp. 353–356. Edited by M. G. Rossmann & E. Arnold. Dordrecht, NL: Kluwer Academic Publishers.
- Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr* **49**, 129–149.
- Lumb, K. J. & Kim, P. S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. *Biochemistry* **34**, 8642–8648.
- Lupas, A., Van Dyke, M. & Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162–1164.
- Malashkevich, V. N., Schneider, B. J., McNally, M. L., Milholle, M. A., Pang, J. X. & Kim, P. S. (1999). Core structure of the envelope glycoprotein GP2 from Ebola virus at 1.9-Å resolution. *Proc Natl Acad Sci U S A* **96**, 2662–2667.
- Meier, M., Lustig, A., Aebi, U. & Burkhard, P. (2002). Removing an interhelical salt bridge abolishes coiled-coil formation in a *de novo* designed peptide. *J Struct Biol* **137**, 65–72.
- Mitraki, A., Miller, S. & van Raaij, M. J. (2002). Review: conformation and folding of novel beta-structural elements in viral fiber proteins: the triple beta-spiral and triple beta-helix. *J Struct Biol* **137**, 236–247.
- Morris, R. J., Perrakis, A. & Lamzin, V. S. (2003). ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol* **374**, 229–244.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
- Oakley, M. G. & Kim, P. S. (1998). A buried polar interaction can direct the relative orientation of helices in a coiled coil. *Biochemistry* **37**, 12603–12610.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.
- Renard, M., Varela, P. F., Letzelter, C., Duquerroy, S., Rey, F. A. & Heidmann, T. (2005). Crystal structure of a pivotal domain of human syncytin-2, a 40 million years old endogenous retrovirus fusogenic envelope gene captured by primates. *J Mol Biol* **352**, 1029–1034.
- Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv Protein Chem* **34**, 167–339.
- Shapouri, M. R. S., Kane, M., Letarte, M., Bergeron, J., Arella, M. & Salim, A. (1995). Cloning, sequencing and expression of the S1 gene of avian reovirus. *J Gen Virol* **76**, 1515–1520.
- Shapouri, M. R. S., Arella, M. & Salim, A. (1996). Evidence for the multimeric nature and cell binding ability of avian reovirus σ 3 protein. *J Gen Virol* **77**, 1203–1210.
- Sheriff, S., Chang, C. Y. & Ezekowitz, R. A. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α -helical coiled-coil. *Nat Struct Biol* **1**, 789–794.
- Songserm, T., van Roozelaar, D., Kant, A., Pol, J., Pijpers, A. & ter Huurne, A. (2003). Enteropathogenicity of Dutch and German avian reoviruses in SPF white leghorn chickens and broilers. *Vet Res* **34**, 285–295.
- Spandidos, D. A. & Graham, A. F. (1976). Physical and chemical characterization of an avian reovirus. *J Virol* **19**, 968–976.
- Strelkov, S. V. & Burkhard, P. (2002). Analysis of α -helical coiled coils with the program TWISTER reveals a structural mechanism for stutter compensation. *J Struct Biol* **137**, 54–64.
- Tao, Y., Strelkov, S. V., Mesyanzhinov, V. V. & Rossmann, M. G. (1997). Structure of bacteriophage T4 fibritin: a segmented coiled coil and the role of the C-terminal domain. *Structure* **5**, 789–798.
- Tripet, B., Wagschal, K., Lavigne, P., Mant, C. T. & Hodges, R. S. (2000). Effects of side-chain characteristics of stability and oligomerization state of a *de novo*-designed model coiled coil: 20 amino acid substitutions in position d. *J Mol Biol* **300**, 377–402.
- Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallogr D Biol Crystallogr* **56**, 1622–1624.
- van Raaij, M. J., Hermo Parrado, X. L., Guardado Calvo, P., Fox, G. C., Llamas-Saiz, A. L., Costas, C., Martinez-Costas, J. & Benavente, J. (2005). Crystallization of the C-terminal globular domain of avian reovirus fibre. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **61**, 651–654.
- Vriend, G. (1990). WHAT IF: a molecular modeling and drug design program. *J Mol Graph* **8**, 52–56.
- Zhang, X., Tang, J., Walker, S. B., O'Hara, D., Nibert, M. L., Duncan, R. & Baker, T. S. (2005). Structure of avian orthoreovirus virion by electron cryomicroscopy and image reconstruction. *Virology* **343**, 25–35.

X. Lois Hermo-Parrado,^a Pablo Guardado-Calvo,^a Antonio L. Llamas-Saiz,^b Gavin C. Fox,^c Lorena Vazquez-Iglesias,^a José Martínez-Costas,^a Javier Benavente^a and Mark J. van Raaij^{a,b,*}

^aDepartamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain,

^bUnidad de Difracción de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, and ^cSpanish CRG Beamline BM16, European Synchrotron Radiation Facility (ESRF), 6 Rue Jules Horowitz, BP 220, F-38043 Grenoble, France

Correspondence e-mail: vanraaij@usc.es

Received 16 March 2007
 Accepted 10 April 2007

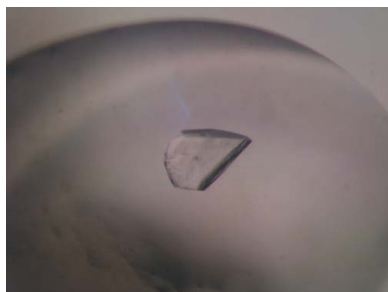
Crystallization of the avian reovirus double-stranded RNA-binding and core protein σA

The avian reovirus protein σA plays a dual role: it is a structural protein forming part of the transcriptionally active core, but it has also been implicated in the resistance of the virus to interferon by strongly binding double-stranded RNA and thus inhibiting the double-stranded RNA-dependent protein kinase. The σA protein has been crystallized from solutions containing ammonium sulfate at pH values around 6. Crystals belonging to space group *P1*, with unit-cell parameters $a = 103.2$, $b = 129.9$, $c = 144.0$ Å, $\alpha = 93.8$, $\beta = 105.1$, $\gamma = 98.2^\circ$ were grown and a complete data set has been collected to 2.3 Å resolution. The self-rotation function suggests that σA may form symmetric arrangements in the crystals.

1. Introduction

Avian reovirus is a non-enveloped virus and is a member of the *Orthoreovirus* genus of the *Reoviridae* family (Benavente & Martínez-Costas, 2007). The reovirus has been linked to various diseases in birds that account for considerable economic losses in the poultry industry (Jones, 2000). The virus has a double-stranded RNA genome consisting of ten segments which are encased by a double concentric icosahedral capsid (Spandidos & Graham, 1976). The inner capsid is made up of the λA , λC and σA proteins, which encapsulate up to 12 copies of the RNA-dependent RNA polymerase complex (consisting of λB and μA) together forming the viral 'core'. Proteins μB , σB and σC are components of the outer shell (Martínez-Costas *et al.*, 1997), with σC being responsible for primary host-cell attachment (Grande *et al.*, 2000; Guardado-Calvo *et al.*, 2005). The virus also encodes four nonstructural proteins which are involved in cell fusion (p10), viral assembly (μNS and σNS) or in as yet undetermined functions (p17) (Varela *et al.*, 1996; Bodelon *et al.*, 2001; Touris-Otero *et al.*, 2004; Costas *et al.*, 2005).

The avian reovirus protein σA plays a dual role. It functions primarily as a structural protein that forms part of the transcriptionally active core. In addition, it has also been implicated in the resistance of the virus to interferon by binding double-stranded RNA and thus inhibiting the double-stranded RNA-dependent protein kinase (Martínez-Costas *et al.*, 2000; Gonzalez-Lopez *et al.*, 2003). The structure of the mammalian reovirus core has been solved by X-ray crystallography (Reinisch *et al.*, 2000). In this structure, the structural equivalent of avian reovirus σA , $\sigma 2$, is present in 150 copies per core in three different locations ($\sigma 2$ -i, $\sigma 2$ -ii and $\sigma 2$ -iii) and acts as a core-stabilizing clamp. The protein interacts extensively with $\lambda 1$ (the mammalian reovirus equivalent of λA) at two different binding sites (the binding sites for $\sigma 2$ -ii and $\sigma 2$ -iii are similar). There are only sparse contacts between $\sigma 2$ monomers and between $\sigma 2$ and the other core protein $\lambda 2$ (the equivalent of avian reovirus λC). In agreement with the mammalian reovirus 'core' structure, cryo-electron microscopy of avian reovirions confirmed that they contain 150 σA 'nodules' contacting mainly λA , with minor contacts formed with the λC turrets and other σA molecules (Zhang *et al.*, 2005). They also contact the μB proteins of the outer capsid (equivalent to $\mu 1$ in mammalian reovirus). However, with respect to double-stranded RNA binding, σA appears to play an analogous role to the mammalian reovirus protein $\sigma 3$, as this protein (the structural



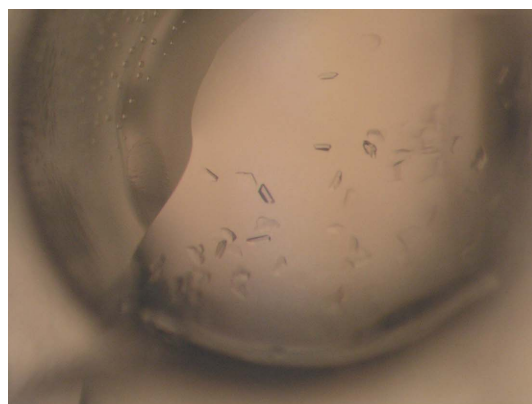
equivalent of avian reovirus σ B) exhibits double-stranded RNA-binding activity (Yue & Shatkin, 1997), albeit not as strongly as avian reovirus σ A (Martinez-Costas *et al.*, 2000).

In our ongoing efforts to obtain structural and functional information about the protein, we have expressed and purified σ A in milligram amounts, crystallized it and collected a complete data set to 2.3 Å resolution.

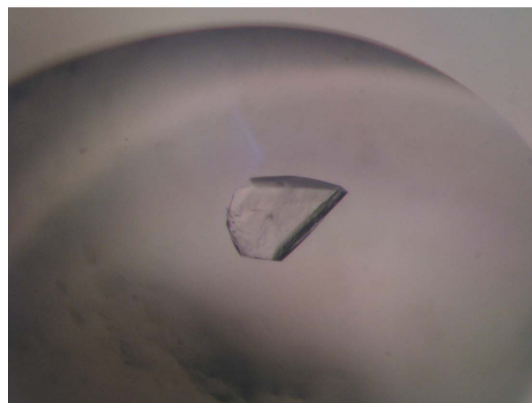
2. Methods

The avian reovirus S1133 protein σ A was expressed as a maltose-binding-protein fusion using the expression vector pMALCS2 described in Gonzalez-Lopez *et al.* (2003). Prior to proceeding with milligram-scale expression and purification for crystallization experiments, the sequence of the insert encoding σ A was confirmed by DNA-sequence analysis (Sistemas Genomicos, Valencia, Spain).

For large-scale expression, the *Escherichia coli* strain XL1-Blue was freshly transformed with the pMALCS2 vector and four 750 ml cultures were grown aerobically at 310 K to an optical density of 0.6 measured at 600 nm in LB medium. Expression was induced by adding 1 mM isopropyl β -D-thiogalactopyranoside and allowed to proceed for 3 h. Harvested cells were resuspended in 40 ml lysis buffer (20 mM Tris-HCl pH 7.3, 200 mM sodium chloride, 1 mM dithiothreitol) containing a protease-inhibitor cocktail specific for bacterial cell extracts (Sigma-Aldrich, Madrid, Spain) and the suspension was frozen at 253 K. The cellular pellet was lysed by a double pass through an emulsifier (Avestin Emulsifier C5, Avestin



(a)



(b)

Figure 1

Crystals of the avian reovirus σ A protein. (a) A typical crystallization drop containing multiple small crystals. (b) A single large crystal; this crystal is about 0.3 mm in width and 0.05 mm thick.

Table 1

Statistics of crystallographic data collected from a single σ A crystal.

Values in parentheses are for the highest resolution bin (where applicable).

Beamline	ESRF ID23-1
Wavelength (Å)	0.954
Detector	MAR Mosaic
Crystal-to-detector distance (mm)	251
Space group	P1
Unit-cell parameters (Å, °)	$a = 103.2, b = 129.9, c = 144.0,$ $\alpha = 93.8, \beta = 105.1, \gamma = 98.2$
Mosaicity (°)	0.6
No. of observed reflections†	282973 (32868)
Resolution range (Å)	30.0–2.34 (2.47–2.34)
Multiplicity	2.0 (1.9)
Completeness (%)	94.3 (74.8)
$\langle I/\sigma(I) \rangle$	7.7 (2.2)
$R_{\text{sym}}^{\ddagger}$ (%)	10.5 (30.0)

† No σ -cutoff or other restrictions were used for the inclusion of observed reflections. $\ddagger R_{\text{sym}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.

Europe GmbH, Mannheim, Germany). After removal of the insoluble material by centrifugation, sodium chloride and poly(ethyleneimine) were added to final concentrations of 1 M and 0.1% (w/v), respectively, and the suspension was incubated at 277 K for 1 h to precipitate nucleic acids. Following a second centrifugation step, bovine ribonuclease A (>70 Kunitz units per milligram; Sigma-Aldrich, Madrid, Spain) was added to the supernatant to a final concentration of around 30 $\mu\text{g ml}^{-1}$ and incubated for 30 min at room temperature. 3 ml suspended amylose resin (New England Biolabs, Ipswich MA, USA) was then added and gently mixed and the suspension was left for 10 min at room temperature before being poured into an empty column. The resin was washed three times with three column volumes of MBP buffer (20 mM Tris-HCl pH 7.5, 200 mM sodium chloride, 1 mM EDTA) and elution was performed by washing twice with 10 ml 10 mM maltose in MPB buffer. Calcium chloride and protease factor Xa (New England Biolabs, Ipswich MA, USA) were added to final concentrations of 1 $\mu\text{g ml}^{-1}$ and 1 mM, respectively. Incubation was carried out at room temperature for 20 h. The digested protein was dialysed overnight against TE buffer (10 mM Tris-HCl pH 8.5, 1 mM EDTA) and applied onto a 1 ml Uno-Q column (Biorad, Barcelona, Spain). Pure σ A eluted in the flowthrough, while maltose-binding protein and other impurities were retained on the column. The protein was concentrated with Centricon concentrators (Millipore, Madrid, Spain), using three washes with TE buffer to eliminate small-molecule impurities.

Crystallization was carried out by sitting-drop vapour diffusion at 278 K in CompactClover plates (Jena Biosciences, Jena, Germany), with 0.1–0.15 ml reservoirs and drops consisting of 2–5 μl protein solution (14–30 mg ml^{-1} protein in TE buffer) mixed with 2–5 μl reservoir solution, using equal volumes of both. Diffraction images were processed using *MOSFLM* (Leslie, 2006) and the crystallographic data were scaled using *SCALA* (Collaborative Computational Project, Number 4, 1994). Self-rotation functions were calculated using *MOLREP* (Vagin & Teplyakov, 2000).

3. Results and discussion

Recombinant σ A was expressed as a fusion protein with maltose-binding protein. After lysis, excess nucleic acids were removed by precipitation and the preparation was extensively treated with RNase A to remove all traces of bound RNA (σ A binds double-stranded RNA very tightly). The fusion protein was subsequently partially

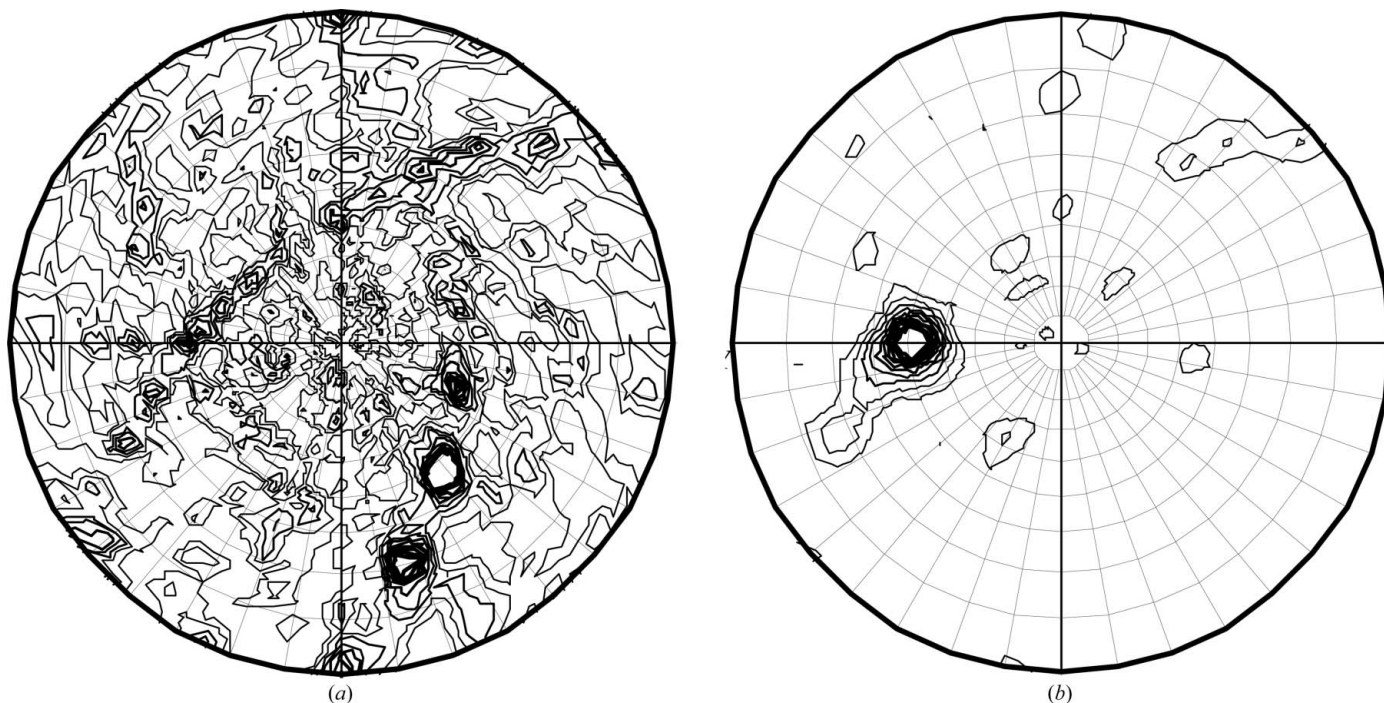


Figure 2 Self-rotation function calculated using the program *MOLREP* (Vagin & Teplyakov, 2000) from a complete σA data set. (a) $\chi = 180^\circ$ section; (b) $\chi = 50^\circ$ section, indicating the presence of multiple noncrystallographic twofolds and a single noncrystallographic higher symmetry axis.

purified using amylose resin. Upon cleavage with a specific protease, σA was separated from maltose-binding protein and other impurities by anion-exchange chromatography; pure σA eluted in the flow-through fraction. From the protease-cleavage step onwards, the procedures were performed at room temperature, since σA was found to precipitate in the cold in the absence of salt (albeit reversibly). This tendency to precipitate reversibly is an indication that the protein may form higher order aggregates in solution and the temperature-dependence of this aggregation suggests that polar interactions may predominate in the oligomerization process.

In vapour-diffusion experiments, sizable crystals were obtained after 6–12 months at pH 5.5–6.5 and 278 K using 100 mM sodium citrate, sodium phosphate or 2-(*N*-morpholino)ethanesulfonic acid–NaOH as a buffer and 0.1–0.6 M ammonium sulfate as precipitant (Fig. 1). For data collection, crystals were briefly incubated in mother liquor containing 25% (*v/v*) glycerol and then flash-frozen in liquid nitrogen.

A high prevalence of multiple lattices was observed in the diffraction patterns obtained from these crystals. However, fairly complete data sets with reasonable statistics were collected from several crystals using synchrotron radiation, in the best case yielding a data set to 2.3 Å resolution (Table 1). For each crystal, 180 1° oscillation-angle images were collected, with exposure times of 1–2 s for each image, using the insertion-device beamline ID23-1 at the European Synchrotron Radiation Facility. Data sets from different crystals were not merged, as the crystals were found to be insufficiently isomorphous to each other. The crystals were assigned to space group *P1* and analysis of the Matthews coefficient suggested the presence of between six and 18 protein molecules in the asymmetric unit (with a solvent fraction of between 81 and 43%, respectively). We have recently obtained a convincing molecular-replacement solution containing 12 copies of a model based on the mammalian reovirus core protein $\sigma 2$, which would give 62% solvent content.

Self-rotation functions calculated from the different data sets consistently show peaks in the $\chi = 180$ and 50° sections (Fig. 2). The $\chi = 50^\circ$ section shows a single peak at $\theta = 49^\circ$ and $\varphi = -90^\circ$, while the more noisy $\chi = 180^\circ$ section shows two rows of peaks of different intensities, indicating multiple twofold rotation axes perpendicular to a single higher symmetry axis. The $\theta = 49^\circ$, $\varphi = -90^\circ$ peak is also present at $\chi = 100^\circ$, but not at $\chi = 150^\circ$, suggesting that this higher symmetry axis is not a complete sevenfold symmetry axis. The peak maximum is at a χ value of $49.7\text{--}50.2^\circ$, deviating by more than a degree from exact sevenfold symmetry. An exact complete sevenfold symmetry axis with perpendicular twofold axes would also not be consistent with the molecular-replacement solution, which contains only 12 molecules, not 14. Apparently, multiple σA molecules assemble into a regularly ordered assembly, which is likely to have importance for its function in sequestering double-stranded RNA. Structure solution, refinement and analysis are in progress and will be reported elsewhere.

We thank José Antonio Trillo, Salvador Blanco-Turnes and Rebeca Menaya-Vargas for technical assistance in the laboratory. Dave Hall (beamline ID23-1) helped with data collection at the European Synchrotron Radiation Facility. This research was sponsored by research grants BMC2002-02436, BFU2005-02974 and BFU2005-24982-E from the Spanish Ministry of Science and Technology and PGIDIT03PXIC20307PN from the Xunta de Galicia. As part of the European Science Foundation EUROCORES Programme EuroSCOPE, the work was also supported by funds from the European Commission's Sixth Framework Programme under contract ERAS-CT-2003-980409.

References

Benavente, J. & Martinez-Costas, J. (2007). *Virus Res.* **123**, 105–119.

- Bodelon, G., Labrada, L., Martínez-Costas, J. & Benavente, J. (2001). *Virology*, **290**, 181–191.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Costas, C., Martínez-Costas, J., Bodelon, G. & Benavente, J. (2005). *J. Virol.* **79**, 2141–2150.
- Gonzalez-Lopez, C., Martínez-Costas, J., Esteban, M. & Benavente, J. (2003). *J. Gen. Virol.* **84**, 1629–1639.
- Grande, A., Rodríguez, E., Costas, C., Everitt, E. & Benavente, J. (2000). *Virology*, **274**, 367–377.
- Guardado-Calvo, P., Fox, G. C., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J., Benavente, J. & van Raaij, M. J. (2005). *J. Mol. Biol.* **354**, 137–149.
- Jones, R. C. (2000). *Rev. Sci. Tech.* **19**, 614–625.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- Martínez-Costas, J., González-López, C., Vakharia, V. N. & Benavente, J. (2000). *J. Virol.* **74**, 1124–1131.
- Martínez-Costas, J., Grande, A., Varela, R., García-Martínez, C. & Benavente, J. (1997). *J. Virol.* **71**, 59–64.
- Reinisch, K. M., Nibert, M. L. & Harrison, S. C. (2000). *Nature (London)*, 960–967.
- Spandidos, D. A. & Graham, A. F. (1976). *J. Virol.* **19**, 968–976.
- Touris-Otero, F., Martínez-Costas, J., Vakharia, V. N. & Benavente, J. (2004). *Virology*, **319**, 94–106.
- Vagin, A. & Teplyakov, A. (2000). *Acta Cryst.* **D56**, 1622–1624.
- Varela, R., Martínez-Costas, J., Mallo, M. & Benavente, J. (1996). *J. Virol.* **70**, 2974–2981.
- Yue, Z. & Shatkin, A. J. (1997). *Virology*, **234**, 364–371.
- Zhang, X., Tang, J., Walker, S. B., O'Hara, D., Nibert, M. L., Duncan, R. & Baker, T. S. (2005). *Virology*, **343**, 25–35.

Crystal Structure of the Avian Reovirus Inner Capsid Protein σA^{∇}

Pablo Guardado-Calvo,¹ Lorena Vazquez-Iglesias,¹ José Martínez-Costas,¹ Antonio L. Llamas-Saiz,² Guy Schoehn,^{3,4} Gavin C. Fox,⁵ X. Lois Hermo-Parrado,¹ Javier Benavente,¹ and Mark J. van Raaij^{1,6*}

Departamento de Bioquímica e Biología Molecular, Facultad de Farmacia,¹ and Unidade de Difracción de Raios X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS,² Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain; Unit for Virus Host Cell Interaction, UMR 5233, Université Joseph Fourier, EMBL, CNRS,³ and Institut de Biologie Structurale Jean-Pierre Ebel, UMR5075 CEA-CNRS-UJF,⁴ Grenoble, France; Spanish CRG Beam-Line BM16, European Synchrotron Radiation Facility (ESRF), 6 rue Jules Horowitz, BP 220, F-38043 Grenoble, France⁵; and Instituto de Biología Molecular de Barcelona (CSIC-Parc Científic), c/Josep-Samitier 1-5, E-08028 Barcelona, Spain⁶

Received 3 April 2008/Accepted 3 September 2008

Avian reovirus, an important avian pathogen, expresses eight structural and four nonstructural proteins. The structural σA protein is a major component of the inner capsid, clamping together λA building blocks. σA has also been implicated in the resistance of avian reovirus to the antiviral action of interferon by strongly binding double-stranded RNA in the host cell cytoplasm and thus inhibiting activation of the double-stranded RNA-dependent protein kinase. We have solved the structure of bacterially expressed σA by molecular replacement and refined it using data to 2.3-Å resolution. Twelve σA molecules are present in the P1 unit cell, arranged as two short double helical hexamers. A positively charged patch is apparent on the surface of σA on the inside of this helix and mutation of either of two key arginine residues (Arg155 and Arg273) within this patch abolishes double-stranded RNA binding. The structural data, together with gel shift assay, electron microscopy, and sedimentation velocity centrifugation results, provide evidence for cooperative binding of σA to double-stranded RNA. The minimal length of double-stranded RNA required for σA binding was observed to be 14 to 18 bp.

Avian reovirus belongs to the *Orthoreovirus* genus of the *Reoviridae* family. It has been linked to various diseases in birds, which generate considerable economic losses in the poultry industry (17). The virus has a double concentric icosahedral capsid (2). The inner capsid is made up of the λA , λC , and σA proteins, which encapsulate a double-stranded RNA (dsRNA) genome consisting of 10 segments and up to 12 copies of the RNA-dependent RNA polymerase. The RNA polymerase consists of proteins λB and μA . Proteins μB , σB , and σC are components of the outer shell. The virus also encodes four nonstructural proteins (2), which are involved in cell fusion (p10), viral assembly (μNS and σNS), or as-yet-undetermined functions (p17). Up to now, only the crystal structure of the primary cell attachment protein σC has been reported (14).

The object of our present study, the protein σA , has two functions in avian reovirus biology. On one hand, it plays a structural role, forming part of the inner capsid. In this respect it is the structural equivalent of mammalian reovirus $\sigma 2$. The structure of the mammalian reovirus inner capsid has been solved by X-ray crystallography (33). $\sigma 2$ is present in 150 copies per core, in three different locations ($\sigma 2$ -i, $\sigma 2$ -ii, and $\sigma 2$ -iii), acting as a stabilizing clamp on the outside of the inner capsid. The protein interacts extensively with $\lambda 1$ (the mammalian reovirus equivalent of λA) at two different binding sites (the bind-

ing sites on $\lambda 1$ for $\sigma 2$ -ii and $\sigma 2$ -iii are similar). There are only sparse contacts between $\sigma 2$ monomers and between $\sigma 2$ and the other core protein $\lambda 2$ (the equivalent of avian reovirus λC). Cryo-electron microscopy of avian reovirions confirmed that avian reovirus contains 150 σA “nodules” contacting mainly λA , with minor contacts formed with the λC turrets and other σA molecules (40). σA also contacts the μB proteins of the outer capsid (equivalent to $\mu 1$ in mammalian reovirus).

Apart from its structural role in the virus, σA has been shown to strongly bind dsRNA, and in this way it has been implicated in the virus’s resistance to interferon by preventing the activation of the dsRNA-dependent protein kinase (PKR) (13, 25). This interaction would presumably take place in the host cell cytoplasm (outside the virus). Here, σA appears to play a role analogous to that of the mammalian reovirus outer capsid protein $\sigma 3$, as this protein (the structural equivalent of avian reovirus σB) also exhibits dsRNA binding activity (39), albeit not as strongly as avian reovirus σA (25). Here we report the crystal structure of recombinant σA and show that it self-assembles in a helical fashion. Mutational studies support the hypothesis that dsRNA may bind to the inside of this helix.

MATERIALS AND METHODS

For structural studies, the avian reovirus S1133 protein σA was expressed as a maltose binding protein fusion, cleaved using the protease factor Xa, purified, and crystallized as described previously (15). Data collection, processing, and scaling were performed as described in the same paper. Molecular replacement was performed using a model based on the structure of mammalian reovirus $\sigma 2$ (33). To construct the model used in molecular replacement, alignment of the σA and $\sigma 2$ sequences was performed with CLUSTALW (4); this alignment and the $\sigma 2$ structure were input into the program CHAINSAW (6), which retains conserved amino acids but prunes nonconserved residues to the C-gamma atom.

* Corresponding author. Mailing address: Departamento de Bioquímica e Biología Molecular, Facultad de Farmacia, Universidade de Santiago de Compostela, E-15782 Santiago de Compostela, Spain. Phone and fax: 34 981 599157. E-mail: mark.vanraaij@usc.es.

[∇] Published ahead of print on 17 September 2008.

Molecular replacement itself was performed with MOLREP (35), using the option that adapts the search model by recalculating the temperature factor of each atom as a function of its surface area. Model adjustment was done using COOT (11) and O (18), refinement was performed with REFMAC (27), and water molecules were built using ARP (23). Model validation was performed with PROCHECK (24), WHATCHECK (36), MOLPROBITY (7), and MOLEMAN (19).

Site-directed mutagenesis was performed using the QuikChange kit (Stratagene, Cultek, Madrid, Spain), following the manufacturer's instructions. σ A and mutated σ A were purified as described previously (15). Gel shift assays were performed using σ A at 0.6 mg/ml in TE buffer (10 mM Tris-HCl, pH 8.5, 1 mM EDTA [disodium EDTA]) and a dsRNA ladder which contains seven dsRNAs of different lengths (500, 300, 150, 80, 50, 30, and 21 bp) at 0.5 mg/ml (New England Biolabs, Izasa S.A., Barcelona, Spain). Increasing amounts of σ A (0, 0.6, 1.2, 3.0, and 6.0 μ g) were incubated with 0.5 μ g of dsRNA ladder in TE buffer-150 mM sodium chloride for 10 min on ice, and then samples were analyzed by non-denaturing polyacrylamide gel electrophoresis (see the legend to Fig. 3B for details). The dsRNA was stained with ethidium bromide and visualized under UV light. To determine the minimum dsRNA length that σ A can bind, gel shift assays were performed using σ A in phosphate-buffered saline (PBS) buffer (137 mM sodium chloride, 2.7 mM potassium chloride, 8 mM disodium hydrogen phosphate, 15 mM potassium dihydrogen phosphate, pH 7.4), supplemented with 2 mM magnesium chloride and three dsRNAs of different lengths (18, 14, and 10 bp). Experiments were performed in which increasing concentrations of σ A (0, 16.0, and 38 μ M) were added to fixed concentrations of dsRNA (1.6 μ M of 18-bp dsRNA, 3.3 μ M of 14-bp dsRNA, and 3.6 μ M of 10-bp dsRNA) in a final volume of 12.2 μ l leading to σ A-to-dsRNA ratios indicated in the figures. Mixtures were incubated for 10 min at room temperature and analyzed on a 15% polyacrylamide gel in TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA, pH 8.0). The dsRNA was visualized as described above.

Sedimentation velocity analysis was carried out with a Beckman Optima XL-A analytical ultracentrifuge equipped with absorbance optics, using an An-50Ti rotor and standard 12-mm optical path double-sector centerpieces of Epon-charcoal. σ A protein was prepared as described before; an RNA oligonucleotide with the sequence 5'-ACA CUG UGA UCA GCA GGC GCC AGC UGA UCA CAG UGU-3' was obtained from Nedken (Barcelona, Spain). dsRNA was prepared from this self-complementary oligonucleotide by heating and slow cooling. We used PBS buffer for all experiments, supplemented with 2 mM magnesium chloride. Samples of 0.4 ml with σ A and dsRNA at different ratios were centrifuged at 40,000 rpm and 20°C. The concentration of σ A in samples with excess protein was 4.9 μ M, while dsRNA was added at 0.0, 1.0, 2.0, and 3.0 μ M (duplex concentration), leading to σ A-to-dsRNA ratios of 1:0.0, 1:0.21, 1:0.41, and 1:0.62, respectively. The concentration of dsRNA in samples with excess dsRNA was 1.77 μ M (duplex concentration), while σ A was added at 0.0, 1.71, 0.58, 0.18, and 0.058 μ M, leading to σ A-to-dsRNA ratios of 0:1.0, 1:1.0, 1:3.1, 1:9.3, and 1:31, respectively. For calculation of the concentrations, the calculated molar extinction coefficients of σ A and the dsRNA based on their sequence were used (95,590 M⁻¹ cm⁻¹ at 280 nm for σ A, 347,100 M⁻¹ cm⁻¹ at 260 nm for the dsRNA). Radial absorbance scans were taken at 280 nm (experiment 1) or 243 nm (experiment 2), and data were analyzed using the c(s) continuous distribution of Lamm equation solutions with the software SEDFIT (34).

Theoretical values of the sedimentation coefficient $s_{20,w}$ were estimated with the HYDROPRO program (12), using models composed of one, two, or three copies of chain G of our σ A structure and a model of canonical A-form dsRNA of the sequence used. σ A molecules were placed in the center of the RNA. The partial specific volume for σ A (0.728 ml/g) was calculated from its amino acid composition automatically by HYDROPRO. For dsRNA a value of 0.55 ml/g was used. For the different complexes, weight-averaged values were employed, giving 0.669, 0.692, and 0.792 for the putative 1:1, 2:1, and 3:1 σ A-dsRNA complexes, respectively. As recommended in the HYDROPRO manual, values of 3.2, 2.8, and 3.0 were used for the effective atomic radii of protein, dsRNA, and protein/dsRNA samples, respectively.

For electron microscopy, σ A alone at 0.05 mg/ml and mixtures of σ A with dsRNA (0.05 mg/ml σ A; 1.8 μ M 36-bp dsRNA; molar ratio, 15:1) in PBS buffer plus 2 mM magnesium chloride were applied to the clean side of carbon on mica (carbon/mica interface) and negatively stained with 1% (wt/vol) sodium silicotungstate (pH 7.0). Micrographs were taken under low-dose conditions with a JEOL 120 EX II microscope at 100 kV and a nominal magnification of \times 40,000.

Figures 1A and B, 2A, and 4 were prepared using Pymol (Pymol Molecular Graphics System, DeLano Scientific, San Carlos, CA). Figure 3A was prepared using the program CCP4 mg (31).

TABLE 1. Refinement statistics

Parameter	Value ^a
Space group.....	P1
Cell parameters (a, b, c, α , β , γ).....	103.2 Å, 129.9 Å, 144.0 Å, 93.8°, 105.1°, 98.2°
Resolution range	30.0–2.34 Å (2.47–2.34 Å)
No. of reflections used in refinement ^b	281,587 (32,494)
No. of reflections used for R-free	1,385 (119)
R-factor ^c	0.210 (0.26)
R-free	0.271 (0.34)
No. of protein/sulfate/water atoms.....	38,161/60/3,003
Avg B protein/sulfate/water	37.2 Å ² /60.4 Å ² /37.2 Å ²
Ramachandran statistics ^d	91.0%/8.5%/0.4%/0.1%
RMSD ^e (bonds/angles).....	0.014 Å/1.4°

^a Values in parentheses are for the highest-resolution bin, where applicable.

^b No sigma cutoff or other restrictions were used for inclusion of observed reflections.

^c $R = \sum |F_{\text{obs}}(\text{hkl})| - |F_{\text{calc}}(\text{hkl})| / \sum |F_{\text{obs}}(\text{hkl})|$.

^d According to the program PROCHECK (24). The percentages are indicative of residues in the most favored, additionally allowed, generously allowed, and disallowed regions of the Ramachandran plot, respectively.

^e Estimates provided by the program REFMAC (27).

Protein structure accession number. The coordinates have been deposited in the protein structure database (<http://www.rcsb.org>) under accession code 2VAK; the structure factors are also available.

RESULTS

Bacterial expression, purification, and crystallization of the σ A protein have been described previously (15). Molecular replacement yielded a solution with 12 copies of a model based on the mammalian reovirus core protein σ 2. The σ A molecules occupy 38% of the unit cell (Matthews coefficient 3.2 [26]). The final refinement was performed without using noncrystallographic symmetry restraints; statistics are shown in Table 1. The final model contains amino acids Met1-Gln37 and Ile51-Ala416 of σ A for each of the 12 chains, 12 sulfate ions, and 3,003 water molecules. Some monomers also contain a vestige of the expression linker segment remaining after factor Xa protease cleavage (Ile-Ser-Glu-Phe-Gly-Ser-Thr), whereas in others this segment is partially or completely disordered. The conformation of the expression tail will not be further discussed here and is not included in the analyses described below.

Structure and topology of σ A. The overall topology of the σ A monomer is roughly tetrahedral in form with an approximately rectangular base (Fig. 1). The base is relatively planar, 65 to 70 Å long and 45 to 50 Å wide, and encompasses the proposed site of interaction with the main core protein λ A (see below). The distance from the base to the apex, which is expected to form the interaction site with μ B in the mature virion, is around 45 Å. The 12 crystallographically independent copies in the unit cell are virtually identical (root mean square deviation [RMSD] of C-alpha atoms after superimposition of 0.2 to 0.7 Å; Fig. 1B). Each σ A monomer contains 12 alpha-helices, one mixed three-stranded beta-sheet, and three anti-parallel two-stranded beta-sheets (Fig. 1C).

There is a gap where the protein chain is disordered in all 12 copies present in the crystal between helices 1 and 3. No interpretable electron density was observed for σ A residues 38 to 50 in this region. In σ 2, the structurally equivalent amino

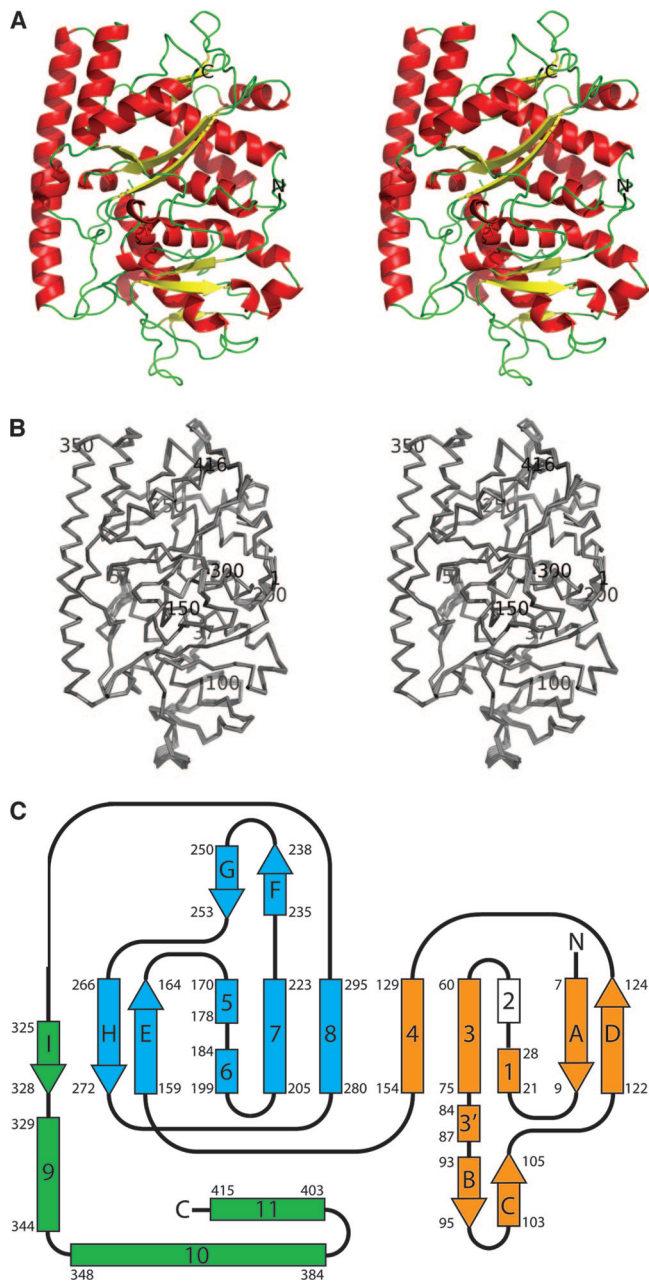


FIG. 1. Structure of the avian reovirus σA protein. (A) Stereo view from the top of the σA pyramid (ribbon representation). The monomer with the lowest average temperature factor, G, was used to make this figure. The amino and carboxy termini of the protein are labeled. (B) Stereo view of main chain superimposition of all 12 σA copies in the asymmetric unit of the crystal. The amino-terminal Met, carboxy-terminal Ala, and every 50th residue are labeled, as are the amino acids flanking the disordered loop (Gln37 and Ile61). (C) Topology of the σA protein chain. Alpha-helices are shown as rectangles labeled with numbers, and beta-strands are shown as arrows labeled with letters. The two structurally homologous domains 1 and 2 are shown in orange and blue, and the carboxy-terminal domain 3 is shown in green. Helix 2 (in white) is not observed in our structure but is inferred from the structure of mammalian reovirus $\sigma 2$ (see text). Starting and ending residues of the secondary structure elements are labeled.

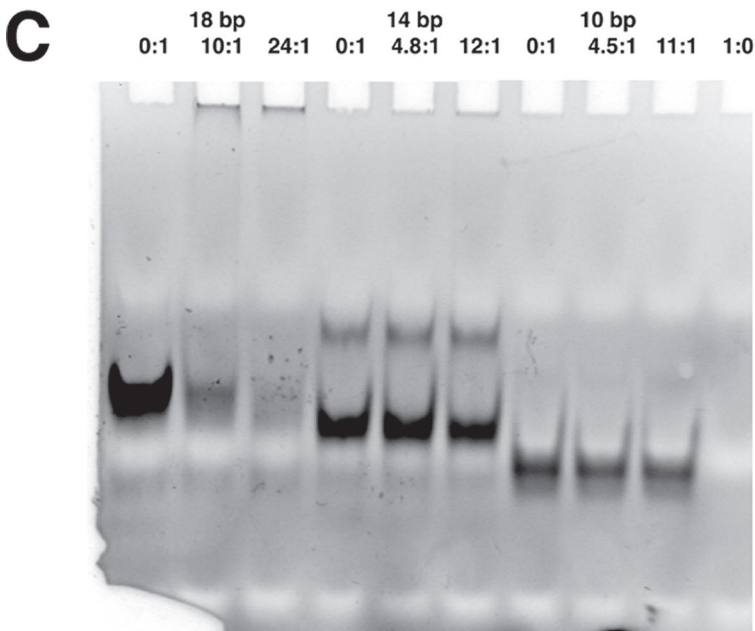
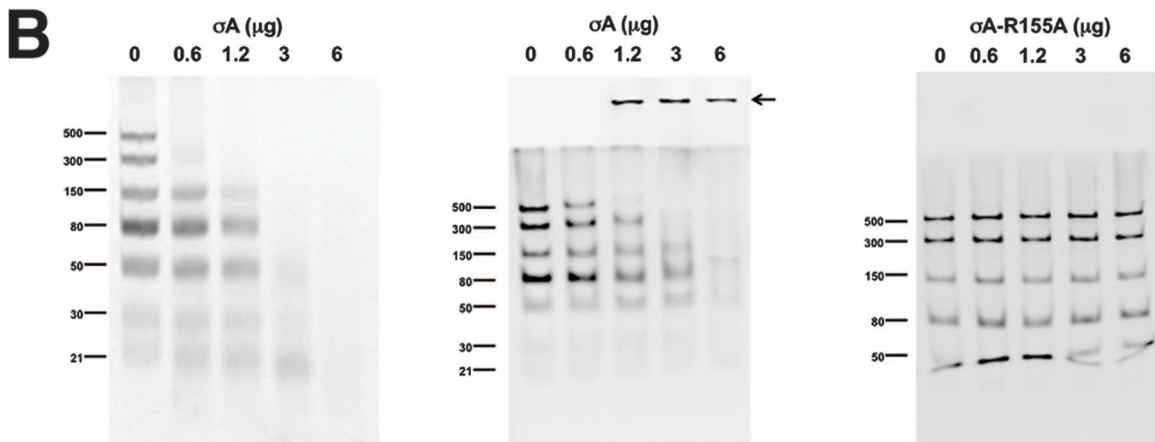
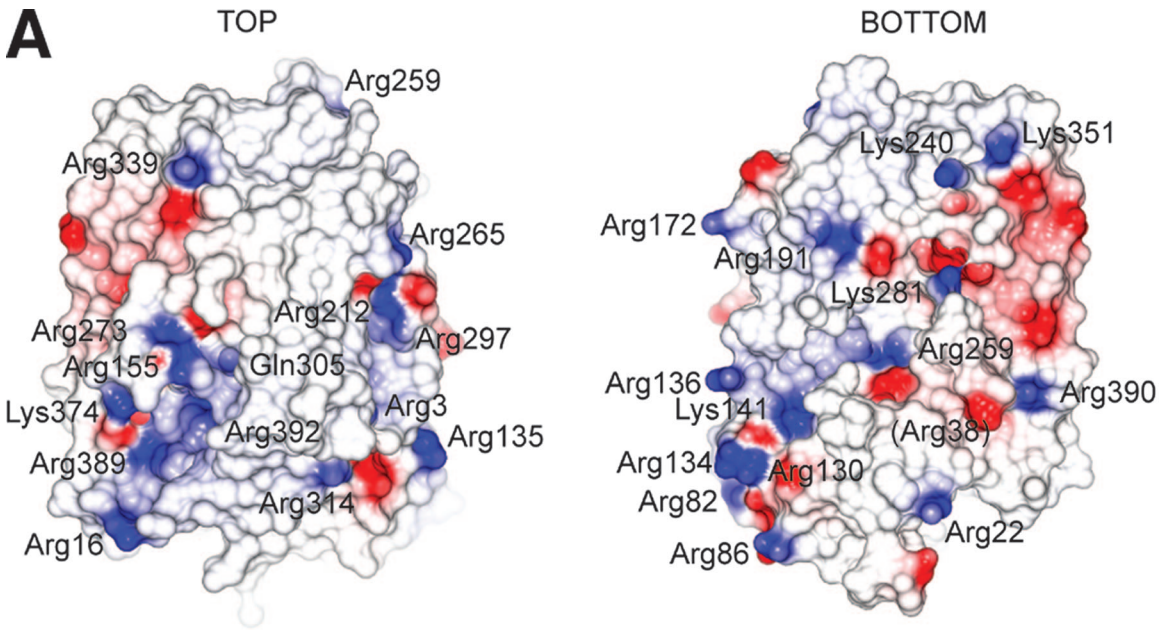
acids are 39 to 52 (Fig. 2). In the previously solved mammalian reovirus core structure, this region contains an alpha-helix in one of the $\sigma 2$ copies (residues 39 to 46) which is melted in the other two (33). This region contacts $\lambda 1$ in the mammalian reovirus core (33), so it is conceivable that the σA alpha-helix becomes ordered only upon binding to λA . We have designated the putative equivalent helix in σA as helix 2. The fact that this loop is flexible and can adopt various conformations suggests that it may be used by the protein to adapt to the different binding environments that the λA shell presents.

The N and C termini are both oriented toward the apex of the molecule and are solvent accessible in this context (although in the virion they are likely to be buried in the interface with μB). The first secondary structure element in the N-terminal region, beta-strand A, however, is buried in the core of the molecule. Phe-8 and Phe-9 in this strand both pack against surrounding hydrophobic and aromatic residues. Helices 1, 3, and 4 (and the putative helix 2 [see below]), together with beta-sheets B-C and A-D, form a discrete subdomain, which is packed against another subdomain with the same topology (helices 5, 6, 7, and 8 and sheets G-F and H-E). These two domains are packed against a third domain formed by helices 9, 10, and 11 and beta-strand I (which lies alongside strand E, forming the H-E-I beta-sheet motif). The apex of the pyramid is largely formed by the loop between helix 8 and strand I.

The regions of σA contacting λA and μB can be predicted from the previously published crystal structure of the mammalian reovirus core (PDB code 1EJ6 [33]), the published pseudoatomic model of the entire mammalian reovirus obtained by fitting known crystal structures into cryo-electron microscopic density (PDB code 2CSE [41]), and cryo-electron microscopic studies of avian reovirus (40). The interaction site with μB mainly comprises residues contained within this third domain, while the proposed λA binding site is almost exclusively composed of residues from domains 1 and 2. Domains 1 and 2 present the same secondary structural elements on the surface at the λA binding site, potentially indicating domain duplication, where the two domains have evolved to interact with different facets of the λA component in the inner shell.

Structural and functional comparison with other proteins. Analysis of the whole structure, or domains 1, 2, or 3 independently, by three-dimensional secondary structure matching (20) did not reveal any close structural homologues for σA , other than mammalian reovirus $\sigma 2$, with which it shares 29% sequence identity. The overall topologies of these proteins are very similar; however, there are small differences in some surface residues, which may be related to inherent flexibility in these regions (Fig. 2A). Most of the surface residues that differ between σA and $\sigma 2$ vary in position between the 12 independent σA monomers in the unit cell (Fig. 1B and 2A). One notable difference between σA and $\sigma 2$ occurs at the carboxy terminus of helix 10 (residues 374 to 388 in the σA structure), which is markedly distorted in σA .

dsRNA binding of σA . With regard to dsRNA binding, σA is functionally, though not structurally, analogous to mammalian reovirus $\sigma 3$. σA and $\sigma 3$ both bind dsRNA and have been implicated in PKR inhibition (13, 16). As with σA , the ability of $\sigma 3$ to bind dsRNA has been linked to its capacity to inhibit dsRNA-dependent activation of PKR through competitive



binding. It has been proposed that $\sigma 3$ binds dsRNA as a dimer and that multiple dimers may line up along the RNA molecule with their twofold axes aligned along the RNA local dyads (30).

When the electrostatic potential is superimposed onto the surface of the σA monomer, a positively charged region, containing among other residues Arg155 and Arg273, is notable (Fig. 3A). We mutated these residues independently to alanine and examined the effects on dsRNA binding by gel shift assays (Fig. 3B). Both mutations abolished the dsRNA binding properties of σA (shown for the Arg155-to-Ala mutation), whereas mutation of Arg134, Arg135, Arg389, or Arg390 had no significant effect on dsRNA binding. Both mutant proteins were soluble, and there were no indications that either of the mutations affected the protein stability or had any structural impact. A further indication that this region may play a role in RNA binding comes from the binding of a sulfate ion at or near Gln305, which is close to Arg155 and Arg273. This sulfate is present in each of the 12 monomers in our structure and may mimic the binding of phosphate in the RNA backbone at this site.

In the gel shift assays using native σA , high-molecular-weight complexes were observed that barely enter the stacking gel (Fig. 3B, middle panel). This suggests either that σA and dsRNA may form large complexes or that the resulting complex may be relatively uncharged at the pH employed here (8.0). The latter is in keeping with a model where σA molecules (theoretical pI 8.6) cover the bulk of the surface of the dsRNA, thus masking negative charges arising from the phosphate backbone. The dsRNA-binding experiments also indicate that σA is able to bind RNA down to a minimum length of at least 21 bp (Fig. 3B, left panel). However, as dsRNA size increases, complexes form at lower σA concentrations, which suggests that the binding is cooperative.

To identify the minimal length of dsRNA bound by σA , gel shift assays were performed with 10-bp, 14-bp, and 18-bp dsRNAs (Fig. 3C). This showed that σA can bind 18-bp dsRNA and also 14-bp dsRNA, albeit less efficiently, while 10-bp dsRNA was not bound by σA . In contrast, $\sigma 3$ has been reported to require a minimum length of 32 to 45 bp for efficient RNA binding (39). We investigated whether the putative dsRNA binding site is structurally related to the dsRNA binding motif (dsRBM) identified in several other proteins (3), including PKR (28, 29). The archetypal dsRBM consists of a contiguous alpha-beta-beta-beta-alpha fold, in which the two helices are arranged on one side, packed against a three-

stranded antiparallel beta-sheet. The residues implicated in dsRNA interaction are located in loop 2 (between beta-strands 1 and 2) and loop 4 (between beta-strand 3 and alpha-helix 2). The dsRNA binding site of σA is somewhat similar and is made up from a discontinuous alpha-beta-beta-beta-alpha subdomain, in which helix 11 and helix 4 and/or helix 8 lie on opposite sides of a three-stranded mixed beta-sheet. The arginines implicated in dsRNA interaction are located in the loop between alpha-helix 4 and beta-strand E and in the loop between alpha-helix 8 and beta-strand H. However, these loops are not conserved and have a different connectivity in the dsRBM, suggesting that the structural similarity is superficial, and hence that there is no direct evolutionary relationship between the σA dsRNA-binding region and the dsRBM.

Interactions between σA molecules in the crystal. σA is monomeric in salt solutions as judged by size-exclusion chromatography (not shown), sedimentation velocity ultracentrifugation, and electron microscopy (see below), which indicated an absence of higher-order aggregates. However, at 4°C, and in the absence of salt, σA precipitates reversibly, indicating that the protein has a tendency to form higher-order aggregates.

In the crystal structure, 12 σA molecules pack together into what at first view appears to be a double-helical assembly (Fig. 4A). When the interactions are analyzed on the basis of buried surface area using the PISA server (21), three types of interactions with potential biological relevance are observed. We have divided these possible interactions into three classes, the “tail-to-tail,” “head-to-tail,” and “head-to-head” interactions. The tail consists of residues from the N-terminal loop, strand A, strand D, the loop to helix 4, the loop between strands G and H, and the loop between helix 8 and strand I, while the “head” consists of amino acids from the loop between putative helix 2 and helix 3, from helix 10, and the loop between helices 10 and 11. The tail-to-tail interaction has a buried surface of 690 Å², a solvation energy gain of 12.4 kcal/mol, three hydrogen bonds, and a *P* value of 0.033; the head-to-tail interaction has a buried surface of 467 Å², a solvation energy gain of 7.4 kcal/mol, six hydrogen bonds, and a *P* value of 0.054, while the head-to-head interaction buries a surface of 373 Å² and has a solvation energy gain of 4.7 kcal/mol, two hydrogen bonds, and a calculated *P* value of 0.26. None of the interactions observed among σA molecules in our crystal structure are similar to the previously observed interactions among $\sigma 2$ molecules in the mammalian reovirus core.

The head-to-tail interaction is the only one involving all 12 σA molecules of the asymmetric unit. It joins the 12 molecules

FIG. 3. dsRNA binding of the σA protein. (A) Electrostatic surface of σA . On the left the orientation is as in Fig. 1A; on the right the view has been rotated 180° around a vertical axis centered on the apex. Positively charged residues on the molecular surface are labeled. The location of Gln305 is also shown, whereas Arg38 is shown in parentheses, as it is barely surface exposed. (B) dsRNA gel shift assays. The assays were performed using σA at 0.6 mg/ml in TE buffer (10 mM Tris-HCl, pH 8.5, 1 mM EDTA) and a dsRNA ladder which contains seven types of dsRNA of different lengths (500, 300, 150, 80, 50, 30, and 21 bp) at 0.5 mg/ml. Increasing amounts of σA (0, 0.6, 1.2, 3.0, and 6.0 μ g) were incubated with 0.5 μ g of dsRNA ladder in TE buffer-150 mM NaCl for 10 min on ice. Then, samples were analyzed by nondenaturing gel electrophoresis. (Left) A 10% polyacrylamide gel in TBE buffer was used. (Middle and right) Discontinuous gels as described by Laemmli (22) were used, with a 10% polyacrylamide separating gel and omission of sodium dodecyl sulfate from the sample, the gel, and the running buffer (13). Nonmutated σA protein was used in the left and middle panels. The right panel shows a gel shift assay performed with the Arg155-to-Ala mutant. Identical results were obtained with the Arg273-to-Ala mutant (data not shown). An arrow in the middle panel indicates the high-molecular-weight complexes observed. (C) Binding of σA to small dsRNA fragments as observed by gel shift assay. In the first three lanes an 18-bp dsRNA was used, in the next three lanes a 14-bp dsRNA was used, and in lanes 7 to 9 a 10-bp dsRNA was used. In lane 10 σA alone was loaded. σA /dsRNA ratios are indicated above the lanes.

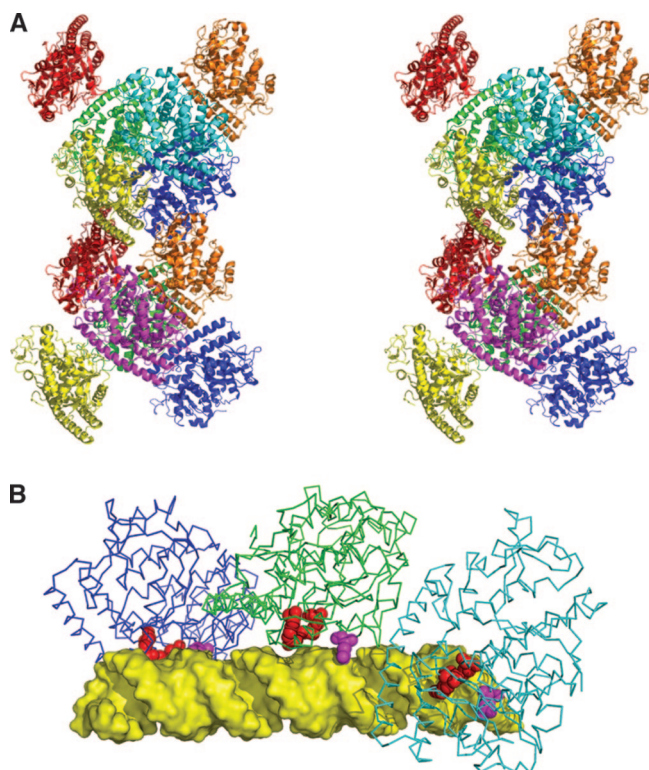


FIG. 4. Assembly of σA molecules in the crystal and model of dsRNA binding. (A) Stereogram of the dodecameric assembly of σA in the crystallographic asymmetric unit. Monomers are shown in cartoon format and colored differently (two yellow, two orange, two red, two green, two blue, one magenta, and one cyan). (B) Model of a σA trimer interacting with a 36-bp canonical A-form dsRNA (yellow). Residues implicated in RNA binding, Arg155 and Arg273, are shown in red; Gln305 is shown in magenta.

into four trimers. Each trimer is related to its corresponding opposite in a double-helical hexamer by a rotation of about 180° . The fit between trimers is very good (RMSD after superposition of 0.4 \AA), suggesting that the trimers act as a rigid body. The monomers of the trimer are related by a rotation of approximately 50° , around an axis running through the center of the double helix, which explains the previously reported peak in the self-rotation function at a chi angle of 50° (15). Interestingly, we note that canonical A-form dsRNA also has a helical rotation of around 50° .

The double-helical hexamer is related to the other hexamer in the asymmetric unit by a rotation of around 180° . In this case, the fit between the two hexamers is not so good (RMSD after superposition of 1.6 \AA). It is tempting to propose the trimer as a biological unit, arranged around dsRNA as a single-stranded or double-stranded assembly. These models could explain the cooperative binding of σA to dsRNA as well as the large complexes observed in the gel shift assays (Fig. 3B, middle panel) (38). The previously discussed positively charged patch maps to the inside of the helix (Fig. 4B). The width of the cavity inside the double-helical σA assembly varies from 11 to 35 \AA , with the cross section being at least 20 \AA wide in one dimension, while the diameter of A-form dsRNA is about 10 by 20 \AA in cross section. In other words, A-form dsRNA would probably fit, although some minor structural adjustments in the dsRNA or in the σA assembly may be necessary.

The distance between Arg155 and Arg273 in the same molecule is 8.5 \AA , while the distances to their counterparts in the next molecule of the trimer are 39 to 40 \AA . In a model constructed with canonical A-form RNA and the crystallographic trimer, one arginine side chain could bind a phosphate on one side of the minor groove and the other could bind a phosphate on the other side of the minor groove (distances between phosphates, 10 \AA). The next molecule could make the same interaction with phosphates 13 bp further along the RNA chain (distance between phosphates, 38 \AA) (Fig. 4B). A structural analysis of the σA -dsRNA complex will be needed to confirm this hypothesis. Mutation of residues involved at the putative protein-protein interface might also be attempted. However, inspection of this interface did not suggest any obvious target residues that, when mutated, would significantly affect protein-protein binding, without having a potential serious structural impact.

Stoichiometry of σA binding to dsRNA. To obtain independent evidence for our dsRNA binding model and in an attempt to discriminate between our putative single and double σA helices, we performed sedimentation velocity centrifugation experiments on σA -dsRNA complexes and transmission electron microscopy on negatively stained σA -dsRNA samples.

In sedimentation velocity centrifugation experiments, a single peak was obtained in control experiments containing σA or dsRNA alone, indicating that both are monodisperse. The estimated molecular masses were those expected for σA monomers and dsRNA duplexes (Table 2). When excess protein was mixed with different amounts of dsRNA (experiment

TABLE 2. Sedimentation velocity centrifugation results and analysis of σA /36-bp dsRNA mixtures

Component or complex	$s_{20,w}$ (10^{-13} s) (%)										
	Calculated	Expt 1 (excess σA)					Expt 2 (excess dsRNA)				
		σA	36-mer dsRNA	Mixture			36-mer dsRNA	Mixture			
			1/0.21	1/0.41	1/0.62		1/1.0	1/3.1	1/9.3	1/31	
σA	3.9	3.6 (100)	NO ^a	3.6 (71)	3.7 (57)	3.5 (35)	NO	NO	NO	NO	
dsRNA	3.4	NO	3.4 (100)	NO	NO	NO	3.6 (100)	3.2 (62)	3.2 (75)	3.2 (77)	
Complex											
1:1	5.6	NO	NO	5.8 (7)	6.0 (11)	5.4 (16)	NO	5.2 (21)	4.4 (19)	4.3 (23)	
2:1	7.3	NO	NO	7.6 (9)	8.0 (12)	7.4 (18)	NO	6.9 (15)	5.7 (6)	NO	
3:1	8.8	NO	NO	9.0 (12)	9.4 (13)	8.8 (25)	NO	NO	NO	NO	

^a NO, not observed.

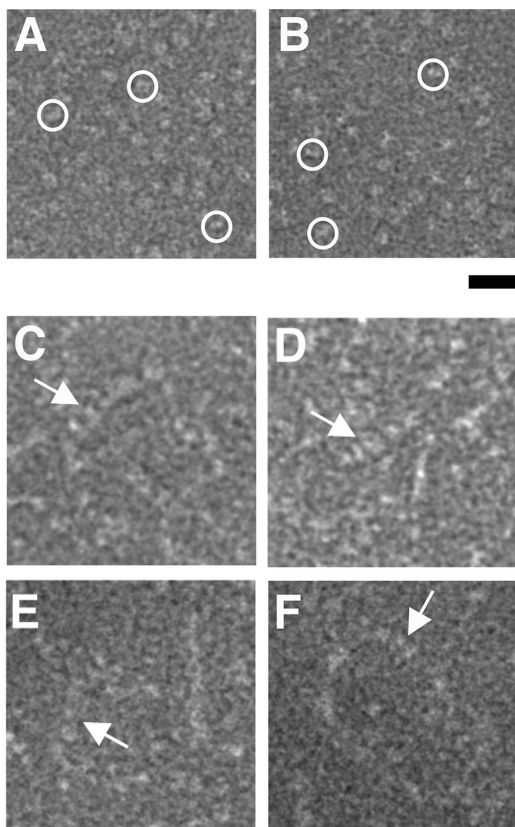


FIG. 5. Transmission electron microscopy of σ A-dsRNA mixtures. (A and B) σ A alone, which appears monomeric. Particles with approximate sizes of monomeric σ A are shown by circles. (C to F) Putative σ A-dsRNA complexes, indicated by arrows. Bar, 20 nm.

1), four reproducible peaks were obtained. Sedimentation coefficients obtained for these peaks are compatible with monomeric σ A and three different complexes of σ A/dsRNA with stoichiometries of 1:1, 2:1, and 3:1 (Table 2). In experiment 2 (Table 2), excess dsRNA was added to protein. Again, dsRNA of 36 bp gave a monodisperse peak, while in the mixtures two or three peaks were observed, of which one clearly corresponds to uncomplexed dsRNA. The complex peaks have smaller sedimentation coefficients than expected for 1:1 and 2:1 σ A/dsRNA complexes, indicating that they are probably in equilibrium with free dsRNA and 1:1 σ A/dsRNA complex, respectively. Taken together, the centrifugation experiments suggest a maximum of three σ A molecules binding to a 36-bp dsRNA (Fig. 4B), which would be compatible with a single helix of σ A molecules covering dsRNA.

Electron micrographs (Fig. 5) also show multiple σ A molecules binding to dsRNA (Fig. 5C to F), while σ A alone appears monomeric (Fig. 5A and B). The width of the observed complexes varies from 54 to 98 Å and does not permit distinction between a single helix (expected diameter, 60 Å) or double helices (expected diameter, 100 Å) of σ A binding to the dsRNA. In fact, binding may be plastic, with regions of dsRNA covered by single-helical σ A and others covered by double-helical σ A, depending on the local dsRNA structure. Curiously, the structures formed are clearly longer than the 36-bp dsRNA used, suggesting that multiple dsRNA duplexes cov-

ered by σ A line up, presumably via σ A- σ A interactions. Future structural studies by cryo-electron microscopy and/or co-crystallization studies will hopefully shed more light on the non-sequence-specific, cooperative mode of dsRNA binding by RNA.

DISCUSSION

Orthoreoviruses can be functionally classified into two subgroups, fusogenic and nonfusogenic, based on their ability to induce cell-cell fusion and form syncytia. The archetypal mammalian orthoreoviruses are nonfusogenic, although they do form specific pores in the membrane, proposed to be part of the membrane penetration pathway during cell entry (1). The remaining members of the genus, including avian orthoreoviruses (2), reptilian orthoreoviruses (10), baboon orthoreoviruses (8), and other mammalian orthoreoviruses like Nelson Bay reovirus (9) and Pulau reovirus (32), are fusogenic. A series of diseases have been attributed to orthoreovirus infection in animals, and recently it has been reported that Melaka reovirus of bat origin, another fusogenic mammalian reovirus, is associated with an acute respiratory disease in humans (5). However, infections with nonfusogenic orthoreoviruses are generally benign, with very rare cases of mild upper respiratory tract illness or enteritis occurring in infants and children.

Sequence analysis and RNA binding assays carried out on the major outer capsid protein, σ B (and equivalents) from fusogenic reoviruses, have revealed that this protein, in contrast to its nonfusogenic counterpart σ 3, does not bind dsRNA (9, 37, 38). As PKR downregulation in avian reovirus S1133 is caused by σ A (13), we analyzed the sequence of the minor inner capsid proteins of fusogenic and nonfusogenic reoviruses to see if the residues implicated in dsRNA binding are conserved among them. This analysis revealed that Arg273 is invariant among fusogenic reoviruses (both avian and mammalian), whereas in nonfusogenic reoviruses it is replaced by Thr. Arg155 is invariant among avian reoviruses but is replaced by Gln in mammalian fusogenic reoviruses and by Gly in mammalian nonfusogenic reovirus. This pattern of conservation suggests that the fusogenic counterparts of the σ A protein retain their ability to bind dsRNA (and by extension the potential to downregulate PKR). In nonfusogenic reoviruses, the σ 2 proteins, which are structural homologues of σ A, do not appear to bind dsRNA. Among this viral subclass RNA binding is linked to the outer capsid protein σ 3. These differences in function may profoundly influence host cell interactions and could be related to and/or have evolved together with the syncytial mode of propagation of the fusogenic viruses and are thus intimately tied to their pathogenicity.

In conclusion, the structure, gel shift, sedimentation velocity, and electron microscopy data presented here suggest a cooperative mode of σ A binding to dsRNA, which may be of general relevance to all fusogenic reoviruses, thereby contributing to a better understanding of their biology. This improved understanding may in turn lead to better strategies to combat these viruses.

ACKNOWLEDGMENTS

This research was sponsored by research grants BFU2005-02974 and BFU2005-24982-E (awarded to M.J.V.) and by a predoctoral FPU

fellowship to P.G.-C., all from the Spanish Ministry of Education and Science, and by research grant PGIDIT04BTF203003PR from the Xunta de Galicia (awarded to J.M.-C.). As part of the European Science Foundation EUROCORES Programme EuroSCOPE, the work was also supported by funds from the European Commission under contract ERAS-CT-2003-980409.

We thank Carlos Alfonso and Germán Rivas of the CSIC (Centro de Investigaciones Biológicas, Madrid, Spain) for help with the analytical ultracentrifugation experiments and Patricia Ferraces Casais and Rebeca Menaya Vargas for technical assistance.

REFERENCES

- Agosto, M. A., T. Ivanovic, and M. L. Nibert. 2006. Mammalian reovirus, a nonfusogenic nonenveloped virus, forms size-selective pores in a model membrane. *Proc. Natl. Acad. Sci. USA* **103**:16496–16501.
- Benavente, J., and J. Martínez-Costas. 2007. Avian reovirus: structure and biology. *Virus Res.* **123**:105–119.
- Chang, K. Y., and A. Ramos. 2005. The double-stranded RNA-binding motif, a versatile macromolecular docking platform. *FEBS J.* **272**:2109–2117.
- Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson. 2003. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **31**:3497–3500.
- Chua, K. B., G. Cramer, A. Hyatt, M. Yu, M. R. Tompang, J. Rosli, J. McEachern, S. Cramer, V. Kumarasamy, B. T. Eaton, and L. F. Wang. 2007. A previously unknown reovirus of bat origin is associated with an acute respiratory disease in humans. *Proc. Natl. Acad. Sci. USA* **104**:11424–11429.
- Collaborative Computational Project Number 4. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**:760–763.
- Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall III, J. Snoeyink, J. S. Richardson, and D. C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**:W375–W383.
- Dawe, S., and R. Duncan. 2002. The S4 genome segment of baboon reovirus is bicistronic and encodes a novel fusion-associated small transmembrane protein. *J. Virol.* **76**:2131–2140.
- Duncan, R. 1999. Extensive sequence divergence and phylogenetic relationships between the fusogenic and nonfusogenic orthoreoviruses: a species proposal. *Virology* **260**:316–328.
- Duncan, R., J. Corcoran, J. Shou, and D. Stoltz. 2004. Reptilian reovirus: a new fusogenic orthoreovirus species. *Virology* **319**:131–140.
- Emsley, P., and K. Cowtan. 2004. Model building tools for molecular graphics. *Acta Crystallogr. D* **60**:2126–2132.
- García de la Torre, J., M. L. Huertas, and B. B. Carrasco. 2000. Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. *Biophys. J.* **78**:719–730.
- Gonzalez-Lopez, C., J. Martínez-Costas, M. Esteban, and J. Benavente. 2003. Evidence that avian reovirus sigmaA protein is an inhibitor of the double-stranded RNA-dependent protein kinase. *J. Gen. Virol.* **84**:1629–1639.
- Guardado-Calvo, P., G. C. Fox, X. L. Hermo-Parrado, A. L. Llamas-Saiz, C. Costas, J. Martínez-Costas, J. Benavente, and M. J. van Raaij. 2005. Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre sigmaC. *J. Mol. Biol.* **354**:137–149.
- Herme-Parrado, X. L., P. Guardado-Calvo, A. L. Llamas-Saiz, G. C. Fox, L. Vazquez-Iglesias, J. Martínez-Costas, J. Benavente, and M. J. van Raaij. 2007. Crystallization of the avian reovirus double-stranded RNA-binding and core protein sigmaA. *Acta Crystallogr. F* **63**:426–429.
- Imani, F., and B. L. Jacobs. 1988. Inhibitory activity for the interferon-induced protein kinase is associated with the reovirus serotype 1 sigma 3 protein. *Proc. Natl. Acad. Sci. USA* **85**:7887–7891.
- Jones, R. C. 2000. Avian reovirus infections. *Rev. Sci. Technol.* **19**:614–625.
- Jones, T. A., J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**:110–119.
- Kleywegt, G. J., J. Y. Zou, M. Kjeldgaard, and T. A. Jones. 2001. Around O, p. 353–356 and 366–367. *In* M. G. Rossmann and E. Arnold (ed.), *International tables for crystallography*, vol. F. Crystallography of biological macromolecules. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Krissinel, E., and K. Henrick. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**:2256–2268.
- Krissinel, E., and K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**:774–797.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**:680–685.
- Lamzin, V. S., and K. S. Wilson. 1993. Automated refinement of protein models. *Acta Crystallogr. D* **49**:129–149.
- Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**:283–291.
- Martínez-Costas, J., C. González-López, V. N. Vakharia, and J. Benavente. 2000. Possible involvement of the double-stranded RNA-binding core protein sigmaA in the resistance of avian reovirus to interferon. *J. Virol.* **74**:1124–1131.
- Matthews, B. W. 1968. Solvent content of protein crystals. *J. Mol. Biol.* **33**:491–497.
- Murshudov, G. N., A. A. Vagin, and E. J. Dodson. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**:240–255.
- Nanduri, S., B. W. Carpick, Y. Yang, B. R. Williams, and J. Qin. 1998. Structure of the double-stranded RNA-binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA-mediated activation. *EMBO J.* **17**:5458–5465.
- Nanduri, S., F. Rahman, B. R. Williams, and J. Qin. 2000. A dynamically tuned double-stranded RNA binding mechanism for the activation of antiviral kinase PKR. *EMBO J.* **19**:5567–5574.
- Olland, A. M., J. Jané-Valbuena, L. A. Schiff, M. L. Nibert, and S. C. Harrison. 2001. Structure of the reovirus outer capsid and dsRNA-binding protein sigma3 at 1.8 Å resolution. *EMBO J.* **20**:979–989.
- Potterton, L., S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis, and M. Noble. 2004. Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60**:2288–2294.
- Pritchard, L. I., K. B. Chua, D. Cummins, A. Hyatt, G. Cramer, B. T. Eaton, and L. F. Wang. 2006. Pulau virus; a new member of the Nelson Bay orthoreovirus species isolated from fruit bats in Malaysia. *Arch. Virol.* **151**:229–239.
- Reinisch, K. M., M. L. Nibert, and S. C. Harrison. 2000. Structure of the reovirus core at 3.6 Å resolution. *Nature* **404**:960–967.
- Schuck, P. 2000. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**:1606–1619.
- Vagin, A., and A. Teplyakov. 2000. An approach to multi-copy search in molecular replacement. *Acta Crystallogr. D* **56**:1622–1624.
- Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**:52–56.
- Wang, Q., J. Bergeron, T. Mabrouk, and G. Lemay. 1996. Site-directed mutagenesis of the double-stranded RNA binding domain of bacterially-expressed sigma 3 reovirus protein. *Virus Res.* **41**:141–151.
- Yin, H. S., J. H. Shien, and L. H. Lee. 2000. Synthesis in *Escherichia coli* of avian reovirus core protein sigmaA and its dsRNA-binding activity. *Virology* **266**:33–41.
- Yue, Z., and A. J. Shatkin. 1997. Double-stranded RNA-dependent protein kinase (PKR) is regulated by reovirus structural proteins. *Virology* **234**:364–371.
- Zhang, X., J. Tang, S. B. Walker, D. O'Hara, M. L. Nibert, R. Duncan, and T. S. Baker. 2005. Structure of avian orthoreovirus virion by electron cryomicroscopy and image reconstruction. *Virology* **343**:25–35.
- Zhang, X., Y. Ji, L. Zhang, S. C. Harrison, D. C. Marinescu, M. L. Nibert, and T. S. Baker. 2005. Features of reovirus outer capsid protein mu1 revealed by electron cryomicroscopy and image reconstruction of the virion at 7.0-Å resolution. *Structure* **13**:1545–1557.

Pablo Guardado Calvo,^a
Antonio L. Llamas-Saiz,^b Patrick
Langlois^c and Mark J. van
Raaij^{a,b,*}

^aDepartamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain,

^bUnidad de Difracción de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain, and ^cAgence Française de Sécurité Sanitaire des Aliments, Unité Génétique Virale et Bioécurité, Site Les Croix, BP 53, F-22440 Ploufragan, France

Correspondence e-mail: vanraaij@usc.es

Received 7 March 2006

Accepted 3 April 2006

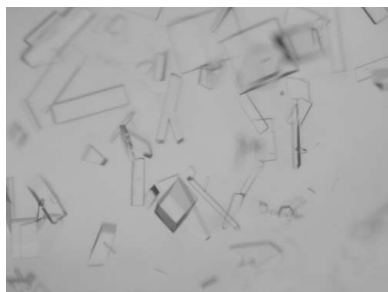
Crystallization of the C-terminal head domain of the avian adenovirus CELO long fibre

Avian adenovirus CELO contains two different fibres: fibre 1, the long fibre, and fibre 2, the short fibre. The short fibre is responsible for binding to an unknown avian receptor and is essential for infection of birds. The long fibre is not essential, but is known to bind the coxsackievirus and adenovirus receptor protein. Both trimeric fibres are attached to the same penton base, of which each icosahedral virus contains 12 copies. The short fibre extends straight outwards, while the long fibre emerges at an angle. The carboxy-terminal amino acids 579–793 of the avian adenovirus long fibre have been expressed with an amino-terminal hexahistidine tag and the expressed trimeric protein has been purified by nickel-affinity chromatography and crystallized. Crystals were grown at low pH using PEG 10 000 as precipitant and belonged to space group *C2*. The crystals diffracted rotating-anode Cu *K* α radiation to at least 1.9 Å resolution and a complete data set was collected from a single crystal to 2.2 Å resolution. Unit-cell parameters were $a = 216.5$, $b = 59.2$, $c = 57.5$ Å, $\beta = 101.3^\circ$, suggesting one trimer per asymmetric unit and a solvent content of 46%. The long fibre head does not have significant sequence homology to any other protein of known structure and molecular-replacement attempts with known fibre-head structures were unsuccessful. However, a map calculated using SIRAS phasing shows a clear trimer with a shape similar to known adenovirus fibre-head structures. Structure solution is in progress.

1. Introduction

Avian adenovirus CELO (chick embryo lethal orphan virus) or fowl adenovirus type 1 (FAV-1; FAV-1) is a non-enveloped virus (Laver *et al.*, 1971; McCracken & Adair, 1993). It infects poultry but is not associated with serious pathologies or economic losses; when experimentally introduced in chickens, it does not cause any evident disease (Cowen *et al.*, 1978). However, this lack of danger has drawn interest in using CELO virus as a gene therapy (Kelleher & Vos, 1994; Stevenson *et al.*, 2006) and anti-cancer vehicle (Logunov *et al.*, 2004; Shashkova *et al.*, 2005) in humans and as a vaccination vehicle for birds (Francois *et al.*, 2004). The virus has a dsDNA genome of 44 kbp encased by an icosahedral capsid (Chiocca *et al.*, 1996). In analogy to the human adenoviruses, the capsid is made up of the hexon and penton proteins, with further minor proteins involved in stabilizing the capsid (Fabry *et al.*, 2005); the hexons form the faces and the pentons the vertices. Each pentameric vertex contains two fibre proteins, the short fibre (fibre 2) emerging straight and the long fibre (fibre 1) at an angle (Hess *et al.*, 1995). Transfection studies have shown that the long fibre binds the coxsackie and adenovirus receptor protein CAR, while the short fibre probably binds an unknown avian receptor (Tan *et al.*, 2001). The short fibre is essential for infection of chicken cells, while the long fibre is not, making it an attractive target for modification.

The long-fibre gene encodes 793 amino acids and has two regions consisting of multiple glycine residues (Hess *et al.*, 1995). The 12 consecutive glycines near the amino-terminus (Gly51–Gly62) are presumably important for flexibility; this flexibility may be necessary for accommodating the long fibre as one of two fibres attached to the same penton and may be responsible for the fact that the long fibre emerges from the viral surface at an angle. Crystallographic studies of the human adenovirus penton in the presence of fibre peptides



© 2006 International Union of Crystallography
All rights reserved

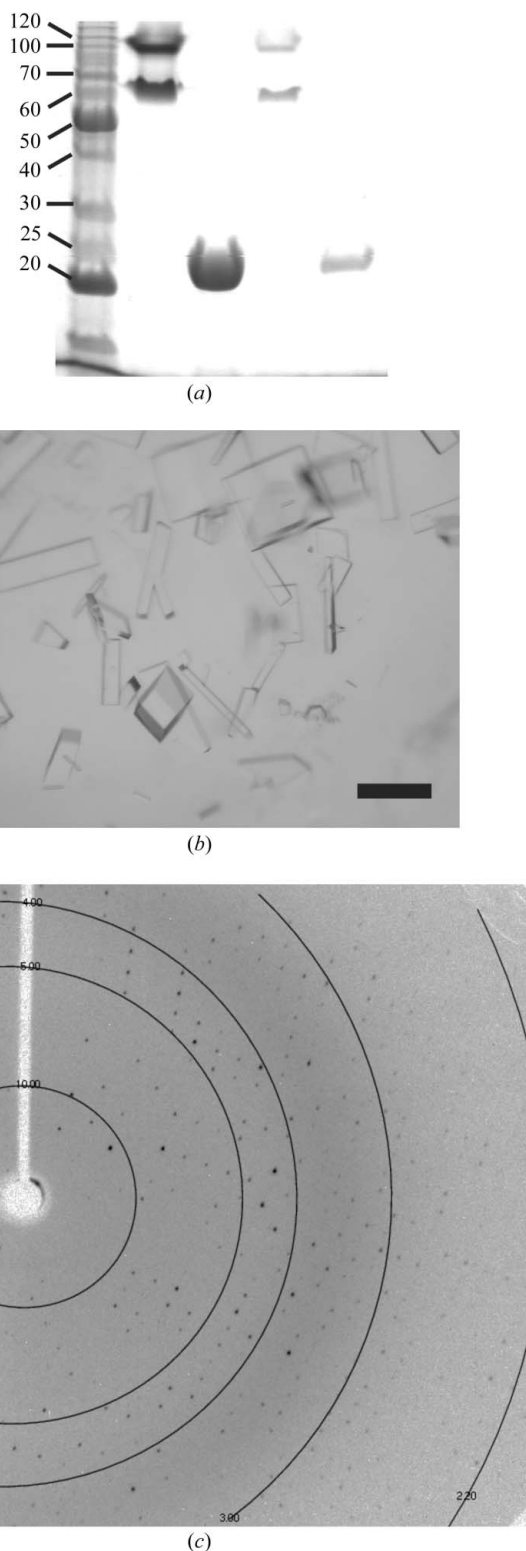


Figure 1
 (a) SDS-PAGE of purified fibre-head protein. Lanes from left to right: molecular-weight markers with their sizes indicated in kDa, high-concentration fibre head without boiling, high-concentration fibre head boiled, low-concentration fibre head without boiling and low-concentration fibre head boiled. (b) Crystals of the avian adenovirus long-fibre head. In conditions containing high-molecular-weight PEG and at low pH crystals of diverse shape grow. All crystals analysed belonged to space group C2 with nearly identical unit-cell parameters. For more details, see §2. The size bar corresponds to around 0.25 mm. (c) Diffraction image. A 1° rotation diffraction image is shown with resolution rings superimposed. Exposure time was 90 s; Cu K α radiation from a rotating-anode source was used.

suggest five fibre-binding sites are available in each penton (Zubieta *et al.*, 2005); therefore, it is possible that one of the amino-termini of the long or short fibres does not bind to the penton. It is possible that the short fibre occupies three binding sites and the long fibre only two. The central domain of the long fibre (residues 72–572) contains repeats compatible with the triple β -spiral fold (Chiocca *et al.*, 1996; van Raaij, Mitraki *et al.*, 1999). These repeats are characterized by the consensus sequence $XX\phi X\phi X\phi X-1-X^*\phi X\phi X X-2$, where X is any amino acid, ϕ is generally hydrophobic and $*$ is generally a proline or glycine, although other small amino acids are also tolerated in this position. At position 1, a central β -turn, insertions of a few amino acids are permitted; at position 2 more extensive insertions are possible as surface loops. The central domain forms the slim central fibre shaft. A second sequence of four consecutive glycines is present near the carboxy-terminal end (amino acids 576–579), possibly forming a flexible region linking the triple β -spiral shaft domain to the head domain.

Here, we report the expression, purification and crystallization of a carboxy-terminal fragment of the avian adenovirus long fibre consisting of amino acids 579–793 containing an amino-terminal hexahistidine and T7 purification tag.

2. Methods

A DNA fragment encoding residues 579–793 of the avian adenovirus long fibre (UniProtKB/TrEMBL accession No. Q64787; Hess *et al.*, 1995) was produced by the polymerase chain reaction and cloned into the expression vector pET28c+ (Novagen, Darmstadt, Germany). The resultant plasmid pET28-CELOlongfib579-793 encodes the mentioned long-fibre fragment fused to an N-terminal purification tag containing a hexahistidine tag and a T7 tag (the N-terminal 11 residues of the bacteriophage T7 gene 10 product). The sequence of the insert was confirmed by DNA-sequence analysis (Sistemas Genómicos, Valencia, Spain).

For expression, *Escherichia coli* strain JM109(DE3) (Promega, Madrid, Spain) was freshly transformed with the plasmid and four 0.7 l cultures were grown aerobically at 310 K to an optical density of 0.6–0.8 measured at 600 nm. Expression was induced by adding 1 mM IPTG (isopropyl thio- β -D-galactopyranoside) and was allowed to continue for 3 h. Harvested cells were resuspended in 40 ml cold PBS buffer (4.29 mM disodium hydrogen phosphate, 1.47 mM potassium dihydrogen phosphate, 2.7 mM potassium chloride, 137 mM sodium chloride) and frozen at 253 K. Bacteria were lysed by a double pass through an emulsifier (Avestin emulsifier C5, Avestin Europe GmbH, Mannheim, Germany). After removing insoluble material by centrifugation, 3 ml Ni-NTA resin (Qiagen, Valencia, CA, USA) was added. The suspension was incubated for 1 h at 277 K and poured into an empty column. The resin was washed with PBS buffer and elution was performed with a step gradient of imidazole pH 7 in PBS buffer (steps of 0.01, 0.05, 0.125, 0.25, 0.5 and 1.0 M imidazole). The protein eluted at imidazole concentrations of 0.125–0.5 M. The 0.25 and 0.5 M imidazole fractions contained virtually pure protein; the flowthrough and fractions up to 0.125 M imidazole were pooled and the procedure was repeated, again keeping the 0.25 and 0.5 M imidazole fractions.

For crystallization, the pooled fractions were dialysed against ME buffer (10 mM MES-NaOH pH 6.5, 1 mM EDTA) and applied onto an Uno-S6 cation-exchange column (Biorad, Madrid, Spain) equilibrated in the same buffer. The protein eluted in several peaks between 0.2 and 0.4 M NaCl in a linear gradient of 0.0–1.0 M NaCl in ME buffer. The fractions all contained pure protein as judged by

Table 1

Crystallographic data statistics.

Values in parentheses are for the highest resolution bin.

Space group	C2
Unit-cell parameters (Å, °)	$a = 216.45$, $b = 59.19$, $c = 57.49$, $\beta = 101.3$
No. of trimers per ASU [†]	1
Matthews coefficient [†] (Å ³ Da ⁻¹)	2.28
Resolution range (Å)	45.0–2.2 (2.3–2.2)
No. of unique reflections [‡]	36555 (4586)
Multiplicity	12.2 (5.2)
Completeness (%)	99.6 (96.8)
$I/\sigma(I)$	27.0 (8.2)
R_{sym} [§] (%)	7.4 (20.4)

[†] Estimation based on likely solvent content. [‡] No sigma cutoff or other restrictions were used for inclusion of observed reflections. [§] $R_{\text{sym}} = \frac{\sum_h \sum_i |I_{hi} - \langle I_h \rangle|}{\sum_h \sum_i I_{hi}}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.

denaturing gel electrophoresis and were thus pooled. The protein was concentrated to 23 mg ml⁻¹ using Centricon concentrators (Millipore, Madrid, Spain), incorporating three washes with TE buffer to eliminate small-molecule impurities.

Crystallization took place by vapour diffusion in sitting-drop CompactClover plates (Jena Biosciences, Jena, Germany) with 0.2 ml reservoirs and drops consisting of 2 µl protein solution mixed with 2 µl reservoir solution. Reservoirs contained 10–12% (w/v) PEG 10 000, 0.1 M HEPES–NaOH pH 4.7–6.1 and 0–10% (v/v) glycerol; crystallization plates were incubated at 278 K. For data collection, crystals were passed through reservoir solution containing 25% (v/v) glycerol, flash-frozen to 100 K in a nitrogen stream and maintained at 100 K. Crystallographic data were integrated and scaled using the *HKL2000* package (Otwinowski & Minor, 1997). Self-rotation functions were calculated using *GLRF* (Tong & Rossmann, 1997).

3. Results and discussion

Structural information about the head domains of several human adenovirus serotypes is available. The first to be solved was the type 5 fibre head (Xia *et al.*, 1994), followed by the type 2 (van Raaij, Louis *et al.*, 1999), type 12 (Bewley *et al.*, 1999), type 3 (Durmort *et al.*, 2001), types 37 and 19p (Burmeister *et al.*, 2004) and the type 41 short-fibre head (Seiradake & Cusack, 2005). They all have identical topology between serotypes and each monomer contains a β -sandwich. The intact heads consist of three β -sandwiches, while the orientation of the monomers in the trimers is also very similar between serotypes.

Both avian short- and long-fibre heads have very limited sequence homology (less than 20% identity) with human adenovirus fibre heads, although it is of course still possible and perhaps likely that they have a similar structure. Furthermore, the long-fibre head probably binds the same coxsackie and adenovirus receptor protein (Tan *et al.*, 2001) as do many human adenovirus fibre heads (groups A and C–F; Roelvink *et al.*, 1998). In order to determine the structure of the head of avian adenovirus long fibre and obtain more information on its function, we have cloned amino acids 579–793, the putative globular head domain, into the bacterial expression vector pET28c+ and obtained soluble protein with an N-terminal His-T7 tag.

The protein could be purified by nickel-affinity chromatography and was shown to be trimeric by gel electrophoresis. Taking advantage of the extraordinary stability of the adenovirus fibre trimers, we compared the electrophoretic mobility of heated and unheated samples avian adenovirus long-fibre head in the presence of sodium

dodecylsulfate (Fig. 1*a*). The results showed that the avian adenovirus long-fibre head domain retains its trimeric structure when the sample is not boiled, although two discrete trimeric bands are observed, one probably corresponding to a compact trimer and a slower migrating band corresponding to a partially unfolded 'hydra', similar to the case of native human adenovirus type 2 fibre (Mitraki *et al.*, 1999). Avian reovirus fibre expressed with the same expression and purification tag (van Raaij *et al.*, 2005; Guardado Calvo *et al.*, 2005) as that employed in this work shows the same behaviour (unpublished results).

For crystallization, the protein was further purified by cation-exchange chromatography. Yields of purified concentrated protein were around 3 mg from 2.8 l culture (assuming that the absorbance at 280 nm of a 1 mg ml⁻¹ protein solution is 1.3, using the theoretical extinction coefficient from the amino-acid sequence of the expressed protein). Crystals were obtained under acidic conditions in the presence of high-molecular-weight PEG (Fig. 1*b*). Using a Bruker–Nonius FR591 rotating-anode generator equipped with a kappa-CCD2000 detector, we were able to obtain diffraction to 1.9 Å resolution (Fig. 1*c*) and a complete highly redundant data set to 2.2 Å resolution was collected from one crystal (0.28 × 0.13 × 0.05 mm) flash-frozen at 100 K (Table 1). A total of 1473 1° oscillation images were collected in 12 runs. Data could be scaled in the monoclinic space group C2; the crystals are most likely to contain one trimer in each asymmetric unit, which would give a solvent content of 46% and a Matthews coefficient of 2.3 Å³ Da⁻¹ (Matthews, 1968).

The self-rotation function is consistent with threefold non-crystallographic symmetry; two clear peaks are present in the $\kappa = 120^\circ$ section, interrelated by the crystallographic twofold axis. The calculated coordinates of the self-rotation peaks are $\varphi = 77.45^\circ$, $\psi = 56.80^\circ$ and the corresponding twofold-related peak $\varphi = 102.55^\circ$, $\psi = 123.20^\circ$, using polar angles and the *XZK* convention (Fitzgerald, 1988). We were not successful in molecular-replacement attempts using the known structures of human adenovirus head domains, presumably owing to their low sequence identity (less than 20%).

However, a map calculated at 2.3 Å using SIRAS phasing on a mercury derivative shows a clear trimer with a shape similar to known adenovirus fibre-head structures. Structure solution and refinement are in progress; details will be published elsewhere. The structure should reveal the extent of structure homology of the avian adenovirus long-fibre head domain with other adenovirus fibre heads. We also hope to obtain other clues from the structure, such as which regions may be involved in receptor binding.

We thank Guy Schoehn for electron microscopy of the purified protein, Xosé Lois Hermo Parrado and Rebeca Menaya Vargas for technical assistance, Stefan Miller (Profos AG, Regensburg, Germany) for providing crystallization plates and José Martínez-Costas and Javier Benavente for sharing laboratory facilities. This research was sponsored by research grants BMC2005-02974 from the Spanish Ministry of Education and Science and PGIDIT03P-XIC20307PN from the Xunta de Galicia; both the grants and the X-ray apparatus on which the data were collected were co-financed by the European Regional Development Fund programme.

References

- Bewley, M. C., Springer, K., Zhang, Y. B., Freimuth, P. & Flanagan, J. M. (1999). *Science*, **286**, 1579–1583.
- Burmeister, W. P., Guilligay, D., Cusack, S., Wadell, G. & Arnberg, N. (2004). *J. Virol.* **78**, 7727–7736.
- Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. & Cotten, M. (1996). *J. Virol.* **70**, 2939–2949.

- Cowen, B., Calnek, B. W., Menendez, N. A. & Ball, R. F. (1978). *Avian Dis.* **22**, 459–470.
- Durmort, C., Stehlin, C., Schoehn, G., Mitraki, A., Drouet, E., Cusack, S. & Burmeister, W. P. (2001). *Virology*, **285**, 302–312.
- Fabry, C. M., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2005). *EMBO J.* **24**, 1645–1654.
- Fitzgerald, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273–278.
- Francois, A., Chevalier, C., Delmas, B., Etteradossi, N., Toquin, D., Rivallan, G. & Langlois, P. (2004). *Vaccine*, **22**, 2351–2360.
- Guardado Calvo, P., Fox, G. C., Hermo Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J., Benavente, J. & van Raaij, M. J. (2005). *J. Mol. Biol.* **354**, 137–149.
- Hess, M., Cuzange, A., Ruigrok, R. W., Chroboczek, J. & Jacrot, B. (1995). *J. Mol. Biol.* **252**, 379–385.
- Kelleher, Z. T. & Vos, J. M. (1994). *Biotechniques*, **17**, 1110–1117.
- Laver, W. G., Younghusband, H. B. & Wrigley, N. G. (1971). *Virology*, **45**, 598–614.
- Logunov, D. Y., Ilyinskaya, G. V., Cherenova, L. V., Verhovskaya, L. V., Shmarov, M. M., Chumakov, P. M., Kopnin, B. P. & Naroditsky, B. S. (2004). *Gene Ther.* **11**, 79–84.
- McCracken, R. M. & Adair, B. M. (1993). *Viral Infections of Vertebrates*, edited by J. B. McFerran & M. S. McNulty, Vol. 3, pp. 123–144. Amsterdam: Elsevier.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Mitraki, A., Barge, A., Chroboczek, J., Andrieu, J. P., Gagnon, J. & Ruigrok, R. W. (1999). *Eur. J. Biochem.* **264**, 599–606.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Raaij, M. J. van, Louis, N., Chroboczek, J. & Cusack, S. (1999). *Virology*, **262**, 333–343.
- Raaij, M. J. van, Mitraki, A., Lavigne, G. & Cusack, S. (1999). *Nature (London)*, **401**, 935–938.
- Raaij, M. J. van, Hermo Parrado, X. L., Guardado Calvo, P. G., Fox, G. C., Llamas-Saiz, A. L., Costas, C., Martínez-Costas, J. & Benavente, J. (2005). *Acta Cryst. F* **61**, 651–654.
- Roelvink, P. W., Lizonova, A., Lee, J. G., Li, Y., Bergelson, J. M., Finberg, R. W., Brough, D. E., Kovesdi, I. & Wickham, T. J. (1998). *J. Virol.* **72**, 7909–7915.
- Seiradake, E. & Cusack, S. (2005). *J. Virol.* **79**, 14088–14094.
- Shashkova, E. V., Cherenova, L. V., Kazansky, D. B. & Doronin, K. (2005). *Cancer Gene Ther.* **12**, 617–626.
- Stevenson, M., Boos, E., Herbert, C., Hale, A., Green, N., Lyons, M., Chandler, L., Ulbrich, K., van Rooijen, N., Mautner, V., Fisher, K. & Seymour, L. (2006). *Gene Ther.* **13**, 356–368.
- Tan, P. K., Michou, A. I., Bergelson, J. M. & Cotten, M. (2001). *J. Gen. Virol.* **82**, 1465–1472.
- Tong, L. & Rossmann, M. G. (1997). *Methods Enzymol.* **276**, 594–611.
- Xia, D., Henry, L. J., Gerard, R. D. & Deisenhofer, J. (1994). *Structure*, **2**, 1259–1270.
- Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). *Mol. Cell*, **17**, 121–135.

Structure of the C-terminal head domain of the fowl adenovirus type 1 long fiber

Pablo Guardado-Calvo,¹ Antonio L. Llamas-Saiz,² Gavin C. Fox,³ Patrick Langlois⁴ and Mark J. van Raaij^{1,2}

Correspondence
Mark J. van Raaij
vanraaij@usc.es

¹Departamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain

²Unidad de Rayos X, Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Campus Sur, E-15782 Santiago de Compostela, Spain

³Spanish CRG Beam Line BM16, European Synchrotron Radiation Facility, 6 rue Jules Horowitz, BP 220, F-38043 Grenoble, France

⁴Agence Francaise de Sécurité Sanitaire des Aliments, Unité Génétique Virale et Biosecurité, Site Les Croix, BP 53, F-22440 Ploufragan, France

Avian adenovirus CELO (chicken embryo lethal orphan virus, fowl adenovirus type 1) incorporates two different homotrimeric fiber proteins extending from the same penton base: a long fiber (designated fiber 1) and a short fiber (designated fiber 2). The short fibers extend straight outwards from the viral vertices, whilst the long fibers emerge at an angle. In contrast to the short fiber, which binds an unknown avian receptor and has been shown to be essential to the invasiveness of this virus, the long fiber appears to be unnecessary for infection in birds. Both fibers contain a short N-terminal virus-binding peptide, a slender shaft domain and a globular C-terminal head domain; the head domain, by analogy with human adenoviruses, is likely to be involved mainly in receptor binding. This study reports the high-resolution crystal structure of the head domain of the long fiber, solved using single isomorphous replacement (using anomalous signal) and refined against data at 1.6 Å (0.16 nm) resolution. The C-terminal globular head domain had an anti-parallel β -sandwich fold formed by two four-stranded β -sheets with the same overall topology as human adenovirus fiber heads. The presence in the sequence of characteristic repeats N-terminal to the head domain suggests that the shaft domain contains a triple β -spiral structure. Implications of the structure for the function and stability of the avian adenovirus long fiber protein are discussed; notably, the structure suggests a different mode of binding to the coxsackievirus and adenovirus receptor from that proposed for the human adenovirus fiber heads.

Received 12 January 2007
Accepted 3 May 2007

INTRODUCTION

Avian adenovirus or chick embryo lethal orphan (CELO) virus (also known as fowl adenovirus type 1, species *Fowl adenovirus A* in the genus *Adenovirus* of the family *Adenoviridae*) is a large, non-enveloped, double-stranded (ds) DNA virus (Laver *et al.*, 1971; McCracken & Adair, 1993). The CELO virus appears to be relatively benign and has not been associated with serious pathogenicity or economic losses (Cowen *et al.*, 1978), nor does it give rise to any evident disease state when experimentally introduced into chickens. This apparent harmlessness has sparked interest in the possibility of using CELO virus in humans as a gene therapy vector (Kelleher & Vos, 1994; Stevenson *et al.*, 2006) or chemotherapy vehicle (Logunov *et al.*, 2004; Shashkova *et al.*, 2005). Studies have also been

undertaken to explore its potential as a vaccination vehicle in birds (Francois *et al.*, 2004). The 44 kb dsDNA genome of the virus is contained within a single icosahedral capsid layer (Chiocca *et al.*, 1996). As is the case for human adenoviruses (Fabry *et al.*, 2005; Saban *et al.*, 2006), the capsid vertices are comprised of penton proteins and the faces of hexon proteins, with further minor proteins incorporated as stabilizers. Each penton vertex contains a pentameric penton base and two trimeric fiber proteins, the short fiber (fiber 2) and the long fiber (fiber 1), with the short fiber emerging straight from the base and the long fiber at an angle (Hess *et al.*, 1995). Whilst the short fiber plays a key role in the infection process in chicken cells (Tan *et al.*, 2001), the long fiber is not essential and is thus an attractive target for modification.

The long fiber monomer contains 793 aa, with two notable glycine-rich regions (Hess *et al.*, 1995). Twelve consecutive glycines located near the N terminus (Gly⁵¹–Gly⁶²) presumably facilitate the flexibility that would be required to accommodate the long and short fibers on the same penton base and may be responsible for the fact that the long fiber emerges from the viral surface at an angle. A sequence of four consecutive glycines is present nearer the C-terminal end (aa 576–579), possibly forming a flexible linker region between the β triple-spiral shaft and the head domain. Crystallographic studies of the human adenovirus penton in the presence of fiber peptides show that five fiber-binding sites are available on each penton (Zubieta *et al.*, 2005), which would imply that the N terminus of one of the fiber monomers does not bind to the penton. The most likely scenario appears to be that the short fiber occupies three binding sites and the long fiber the remaining two.

The long and short fiber head domains exhibit relatively low homology (less than 20% identical amino acids), consistent with them binding to different receptors. When the entire fiber sequences are aligned, regions of homology can be identified near the N terminus and in the putative shaft regions. Close to the N terminus, a region of around 15 highly homologous amino acids may be involved in attaching the fiber to the penton base.

The central domain of the long fiber (aa 72–572) contains repeats compatible with a triple β -spiral fold (Chiocca *et al.*, 1996; van Raaij *et al.*, 1999b). These repeats are characterized by the presence of the consensus sequence XX ϕ X ϕ X ϕ X-1-X* ϕ X ϕ XX-2, where X is any amino acid, ϕ is generally hydrophobic and * is typically a proline or glycine, although other small amino acids are also tolerated in this position. At position 1 in the consensus sequence above, a central β -turn, insertions of a few amino acids are permitted; at position 2, more extensive insertions can be accommodated as surface loops. The central domain forms the slim central fiber shaft.

The head domain is predicted to comprise aa 580–793 and is comparable in size to those fiber head domains of human adenoviruses for which structures have been reported: human adenovirus type 5 fiber head (Xia *et al.*, 1994), type 2 (van Raaij *et al.*, 1999a), type 12 (Bewley *et al.*, 1999), type 3 (Durmort *et al.*, 2001), types 37 and 19p (Burmeister *et al.*, 2004) and the type 41 short fiber head (Seiradake & Cusack, 2005). However, sequence homology between the avian adenovirus long fiber head and these proteins is limited, ranging between 14 and 17% sequence identity.

Previously, we have reported the crystallization of the avian adenovirus long fiber head (Guardado-Calvo *et al.*, 2006). Here, we describe the structure refined against data collected to 1.6 Å (0.16 nm) resolution, revealing the detailed tertiary fold of the molecule. The implications of the structure for the stability and receptor-binding properties of the avian adenovirus long fiber are discussed.

METHODS

Avian adenovirus long fiber head was produced and crystallized as described previously (Guardado-Calvo *et al.*, 2006). The nucleotide sequence of the long fiber head as present in the expression vector was confirmed to be identical to the sequences reported by Hess *et al.* (1995) and Chiocca *et al.* (1996). Crystals were cryo-protected with 25% (v/v) glycerol in reservoir solution and flash-frozen in liquid nitrogen or directly in a nitrogen stream maintained at 100 K. Native data was collected on the Spanish CRG Beam Line BM16 at the European Synchrotron Radiation Facility, France. For collection of derivative data, crystals were soaked in reservoir solution containing methylmercury chloride, both at low concentration (0.25 mM) and at high concentration. For the high-concentration soak, approximately 1 mg methylmercury chloride powder was added to the reservoir (not all of the powder dissolved), the reservoir was allowed to equilibrate with the drop overnight, 1 μ l of the reservoir solution was mixed with the drop containing the crystals, and the soak was allowed to proceed for several hours. Datasets of the derivatized crystals were measured on a Bruker-Nonius FR591 rotating anode source equipped with a KappaCCD2000 detector at the copper K- α wavelength. Reflections were integrated with the HKL2000 program suite (Otwinowski & Minor, 1997). Heavy atom sites were identified by the SHELXD program (Schneider & Sheldrick, 2002) and phases were refined with MLPHARE (Collaborative Computational Project, Number 4, 1994). The non-crystallographic symmetry operator was identified using PROFESS (Collaborative Computational Project, Number 4, 1994). Further solvent flattening, threefold averaging and histogram matching were carried out with the program DM (Cowtan & Main, 1998) and automatic tracing was carried out using ARP-WARP (Perrakis *et al.*, 1999). Molecular replacement was performed with AMORE (Navaza, 2001) or PHASER (McCoy *et al.*, 2005) and model adjustment and manual addition of extra amino acids was done using O (Jones *et al.*, 1991). Refinement was performed with REFMAC (Collaborative Computational Project, Number 4, 1994; Murshudov *et al.*, 1997) and water molecules were built using ARP (Lamzin & Wilson, 1993).

The coordinates have been deposited in the protein structure database (<http://www.rcsb.org>) under accession code 2IUM; the structure factors, including those for the derivative data (accession code r2IUMsf), have also been deposited. Fig. 1(a) was prepared with BOBSRIPT (Esnouf, 1997), a modified version of MOLSCRIPT (Kraulis, 1991). Figs 1(b, c), 2(a, b), 3(a) and 4 were prepared using PYMOL (W. L. DeLano, The PyMOL Molecular Graphics System, 2002, DeLano Scientific, San Carlos, CA, USA; <http://www.pymol.org>). Fig. 3(b–d) was prepared using the program CCP4MG (Potterton *et al.*, 2004).

RESULTS AND DISCUSSION

Structure solution

Crystals of the avian adenovirus long fiber head (Guardado-Calvo *et al.*, 2006) were incubated in the presence of low and high concentrations of methylmercury chloride as described in Methods. Complete and redundant datasets were collected from crystals treated under different soaking conditions (for a summary of the data statistics, see Table 1). Due to problems of non-isomorphism between the mercury-soaked and native crystals, the low-concentration-soaked crystals were treated as the (pseudo)-native and the high concentration soaked crystals as the derivative during substructure solution and phasing. Eight heavy atom positions were identified, of which the occupancies

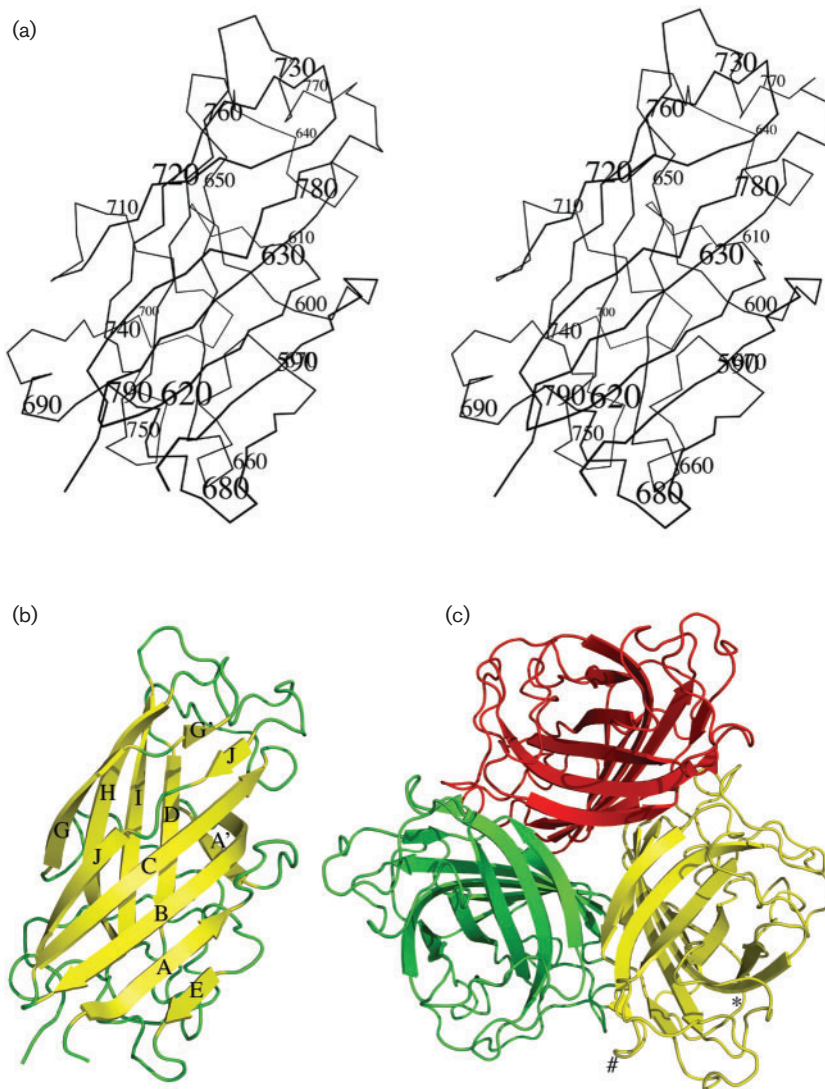


Fig. 1. Structure of avian adenovirus CELO long fiber head. (a) Stereo view of a monomer viewed from the side. C- α atoms are connected by lines; every tenth residue is numbered. (b) The monomer is depicted with β -strands coloured yellow and loops green. β -Strands were identified according to secondary structure designations based on Xia *et al.* (1994). (c) A view from the top of the trimer. Monomers are coloured red, green and yellow. The AB and DG loops of the yellow monomer are labelled with an asterisk and a hash sign, respectively.

and temperature factors were refined. The occupancy of one of the sites refined to zero; of the seven remaining, six were related by a threefold non-crystallographic symmetry operator. Although overall phasing power was relatively low (Table 1), at low resolution it reached values above 2.0. The resulting phases and the symmetry operator were used in simultaneous solvent flattening and averaging, after which a readily interpretable map was obtained. Into this map, 613 aa were automatically traced and the resulting model was used in molecular replacement with the high-resolution, native data. The resulting model was input as a free-atom model in ARP-WARP, allowing the automatic tracing and sequence assignment of 610 residues; a further 23 aa and 857 water atoms were manually modelled into clear electron density before the refinement was completed. The final model contained residues 583–793 of each of the three monomers in the trimer; no inconsistencies between electron density and the expected amino acid sequence were observed.

Density for the three N-terminal aa 579–582 (as well as the six-histidine and T7 expression and purification tag) was missing from the maps. Presumably, this region is flexible, allowing these amino acids to adopt different conformations in the crystal. The refined structures corresponded well to the data, with good geometry and few residues in unfavoured regions of the Ramachandran plot by the PROCHECK criteria (Laskowski *et al.*, 1993; Table 1). Only 2 aa were in ‘generously allowed’ regions, Asn⁶⁹⁴ of monomer B and Asp⁷⁰⁵ of monomer C, the first of which had clear unambiguous density supporting the modelled conformation and the second of which was part of a relatively disordered loop with poorly defined density.

Overview of the structure

The head domain of each monomer formed a compact β -sandwich with a topology similar to the human adenovirus fiber head domains (Fig. 1). The β -sandwich was made up

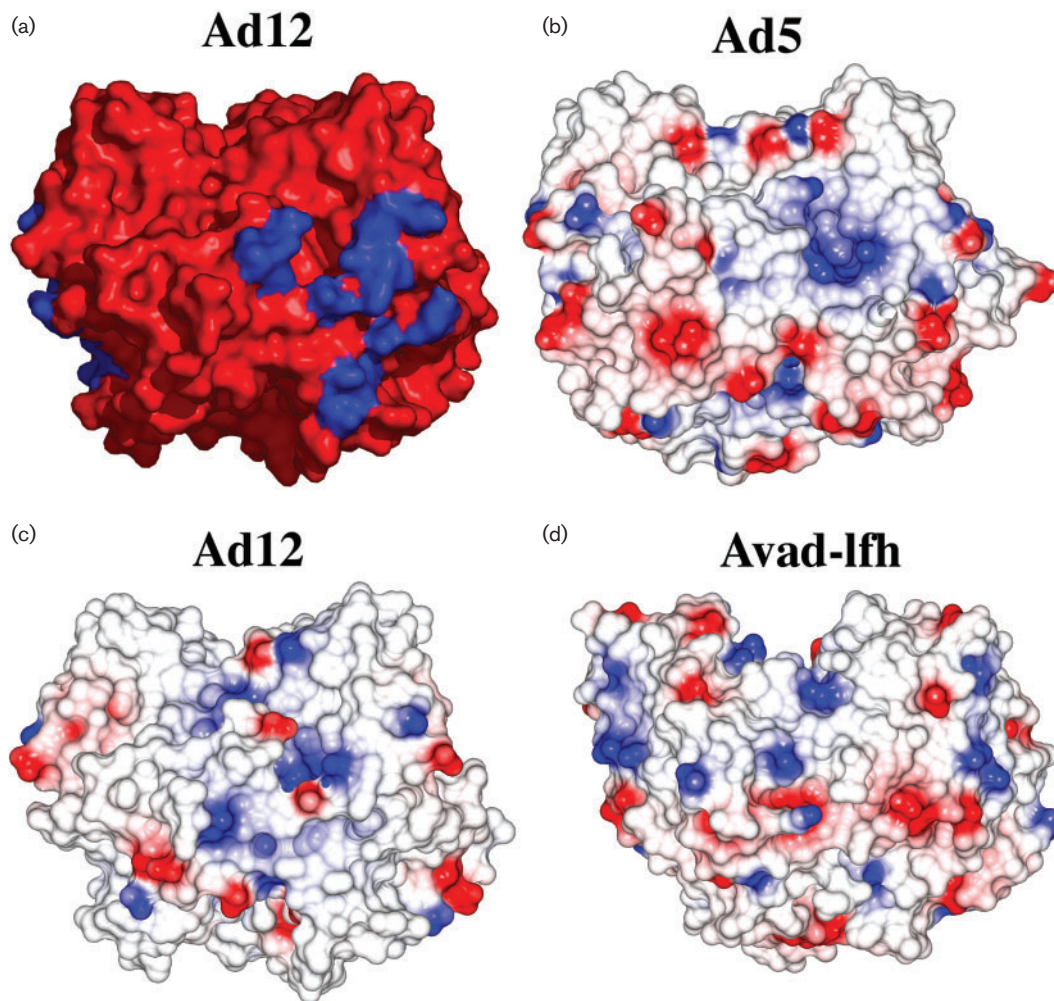


Fig. 3. Surface properties of avian adenovirus long fiber head and human adenovirus fiber heads. All panels are side views, maintaining the same orientation. (a) Surface representation of the human adenovirus type 12 fiber head in red with CAR-binding residues in blue (Bewley *et al.*, 1999). (b–d) Electrostatic potential surface of the human adenovirus type 5 fiber head (b), human adenovirus type 12 fiber head (c) and avian adenovirus long fiber head (d).

parts of these strands A' and G', so that the two sheets could be denoted as EABCJG' and GHIDA'. The J strand also contained a notable kink, caused by two successive prolines (Pro⁷⁸² and Pro⁷⁸³), but did not cross over between the sheets. The structure contained one *cis* peptide bond, between Phe⁶⁷⁶ and Pro⁶⁷⁷, Phe⁶⁷⁶ being the C-terminal residue of strand E. In human adenovirus fiber head domains of known structure, the equivalent region (which is involved in receptor binding) does not contain this feature.

When superimposed, the three monomers exhibited root mean square (r.m.s.) differences of only 0.2–0.3 Å (0.02–0.03 nm) between all C- α positions. Differences between monomers larger than 1 Å (0.1 nm) were limited to aa 641–642 and 771. These residues are part of the CD and IJ loops on the top surface of the trimer, respectively,

and are involved in forming crystal contacts between monomers A and C, but not in the case of monomer B, which explains the conformational differences. Trimer contacts involved numerous residues in the N-terminal section of strands A and C, the C-terminal section of strands B and J, and the loops between strands A and A' (particularly residues Asn⁵⁹³ and Leu⁵⁹⁴), as well as a short stretch of the DG loop (Pro⁶⁸⁹–Thr⁶⁹⁵).

Stability and folding

The surface area of each monomer was around $9.6 \times 10^3 \text{ \AA}^2$, of which 23% was buried upon trimer formation ($2.2 \times 10^3 \text{ \AA}^2$). When the theoretical solvation energies of the trimer were compared with those of the monomer, a solvation energy gain of 150–200 kJ mol⁻¹ was predicted (around 40 kcal mol⁻¹) on trimer formation (Henrick &

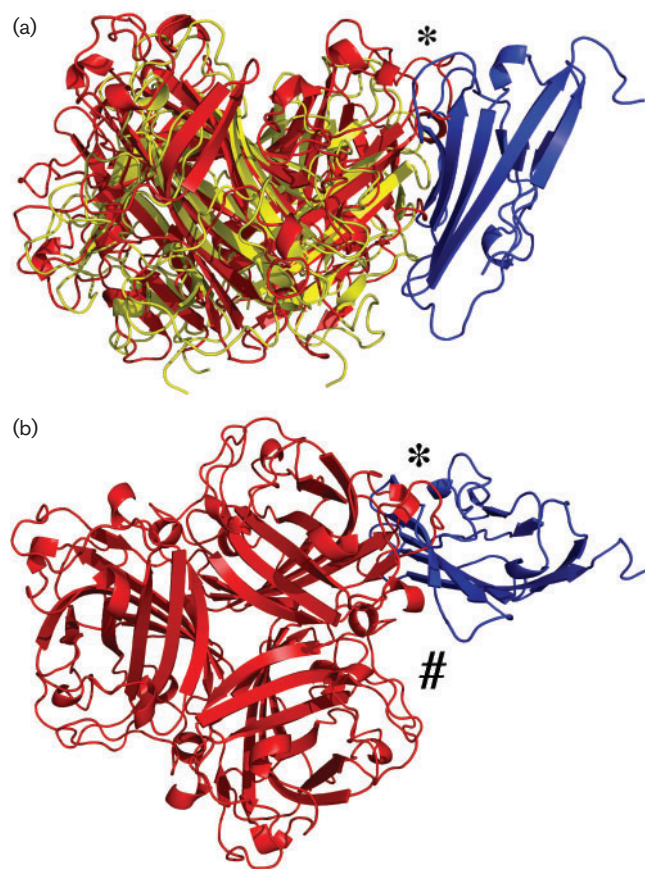


Fig. 4. Superposition of the avian adenovirus long fiber head structure onto the human adenovirus type 12-CAR domain 1 complex structure (Bewley *et al.*, 1999). (a) Side view with the human adenovirus type 12 fiber head structure in yellow, the CAR domain 1 structure in blue and the avian adenovirus long fiber head in red. (b) Top view using the same colour scheme as in (a). Only one of the three CAR domain 1s that could bind simultaneously to the human adenovirus type 12 fiber head is shown. Asterisks mark the region where the avian adenovirus long fiber head would overlap with CAR domain 1 and a hash sign marks the region where the human adenovirus fiber head contacts CAR domain 1 but the avian adenovirus long fiber head does not.

Thornton, 1998). This relatively large buried surface area (Jones & Thornton, 1996) and large solvation energy gain at least partly explain the extraordinary stability of the trimeric long fiber head, as revealed by the fact that it does not dissociate into monomers during SDS-PAGE, unless previously boiled in SDS-containing buffer (Guardado-Calvo *et al.*, 2006). The surface buried within the trimer interface is of mixed nature, containing both hydrophobic and polar contacts. In human adenovirus (van Raaij *et al.*, 1999a), $\sigma 1$ (Chappell *et al.*, 2002), σC (Guardado-Calvo *et al.*, 2005) and the PRD1 p5 spike protein (Merckel *et al.*, 2005), a similar range of mixed contacts is observed in the interface between the head domains. Moreover, the total surface area buried in the trimer correlates closely between

the avian and human adenoviruses (around $2 \times 10^3 \text{ \AA}^2$ or 20% of the available surface area).

Comparison of the structure with known protein structures

Superposition of the structurally equivalent residues from the avian adenovirus long fiber head and human adenovirus head domain monomers using the DALI server resulted in an overall r.m.s. difference of just over 3 Å (0.3 nm) (153 superposed C- α atoms when compared with human adenovirus type 2; Fig. 2a). Not only did all adenovirus head domains have the same topology, but the loops connecting the β -strands were also highly similar, with mainly short inter-connecting loops apart from one long DG loop. The kink in the A strand was also conserved in both types of fiber head, although the cross-over of the A strand to the other β -sheet did not occur in human adenovirus fiber heads. When trimers were superposed (3×153 structurally equivalent C- α atoms; Fig. 2b), the agreement was only slightly worse [r.m.s. differences of 4.2 Å (0.42 nm) when compared with human adenovirus type 2] and the relative orientation of the monomers in the trimer was very similar.

In comparison with the human adenovirus type 2 fiber head, the avian fiber head contained an additional 16 aa, which were located in the AB loop, the G strand, the GH loop and the IJ loop, all of which were longer in the avian long fiber head (Fig. 2c). Furthermore, the CD and HI loops on the top of the trimer were oriented differently, and the DG loop displayed a different conformation. These variations give rise to what are likely to be functionally significant differences in the external surfaces of the trimer unit, whilst maintaining a highly similar internal structural framework. One notable difference was in the conformation of the long DG loop, which in human adenovirus type 2 has more residues at the monomer-monomer interface (on the left in Fig. 2a and indicated with an 'I' in Fig. 2b), whilst in the avian adenovirus fiber head it covers more of the outer GHIDA' sheet surface (on the right in Fig. 2a and indicated with an 'S' in Fig. 2b). Inspection of the electrostatic surfaces (Fig. 3) revealed that human types 5 and 12 were more similar to each other than to the avian adenovirus long fiber head.

Apart from human adenovirus fiber heads, the search also indicated structural similarity with reovirus (Chappell *et al.*, 2002; Guardado-Calvo *et al.*, 2005) and lactobacillus phage fibers (Spinelli *et al.*, 2006a, b). These proteins contain β -sheet structures in their head domains; however, they are topologically different, forming circular β -barrels rather than β -sandwiches. When comparing the avian reovirus fiber head with the mammalian reovirus fiber head, differences in how the monomers self-associate in the trimer were observed, with the individual monomers of the avian reovirus fiber being more 'splayed out' than in their mammalian counterpart. It appeared that there was no such variability in quaternary structure between avian and human adenovirus fiber heads.

Table 1. Crystallographic data and refinement statistics

The crystallographic space group is C2. Values in parentheses are for the highest resolution bin, where applicable.

Data collection	Low-concentration derivative	High-concentration derivative	Native
Detector	KappaCCD2000	KappaCCD2000	165 mm MAR-CCD
Cell parameters (a, b, c) (Å)	216.5, 59.2, 57.5	216.8, 59.1, 57.6	216.8, 58.7, 57.6
Cell parameters (α , β , γ) (°)	90.0, 101.3, 90.0	90.0, 101.3, 90.0	90.0, 101.5, 90.0
Wavelength (Å)	1.5418	1.5418	0.9799
Resolution (Å)	43.0–2.2 (2.30–2.20)	43.0–2.2 (2.28–2.19)	25.0–1.6 (1.66–1.60)
Observed reflections*	36555 (4586)	36315 (3624)	88294 (6836)
Multiplicity	12.2 (5.2)	4.5 (2.3)	3.1 (2.5)
Completeness (%)	99.6 (96.8)	97.9 (85.7)	93.8 (73.0)
Rmerge (%)†	7.3 (20.0)	18.1 (42.3)	4.9 (37.8)
$\langle I/\sigma(I) \rangle$	27.0 (8.2)	10.4 (2.8)	20.9 (2.3)
Phasing (using high-concentration derivative as ‘derivative’ and low-concentration derivative as ‘native’)			
Resolution range used (Å)		43.0–2.3 (2.39–2.30)	
Number of reflections		31459 (2900)	
Number of heavy atom sites		8	
Phasing power (centric/acentric)‡		1.08/0.90	
Rcullis (centric/acentric)§		0.82/0.71	
Refinement			
Resolution range (Å)		16–1.6 (1.69–1.60)	
No. reflections used in refinement		85836 (9941)	
No. reflections used for R-free		2034 (230)	
R-factor		0.180 (0.31)	
R-free		0.213 (0.31)	
No. of protein/water atoms		4797/857	
Mean B protein/water atoms (Å ²)		23.44/39.74	
Ramachandran statistics (%)¶		91.2/8.4/0.4/0.0	
R.m.s. deviations [bonds (Å)/angles (°)]#		0.015/1.618	

*No sigma cut-off or other restrictions were used for inclusion of observed reflections.

† $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \langle I_h \rangle| / \sum_h \sum_i I_{hi}$, where I_{hi} is the intensity of the i th measurement of the same reflection and $\langle I_h \rangle$ is the mean observed intensity for that reflection.

‡Phasing power = $(|F_H|/|F_{PH}| - |F_P + F_H|)$, where F_H , F_{PH} and F_P are the heavy atom, derivative and native structure factors, respectively.

§ $R_{\text{cullis}} = (|F_{PH}| - |F_P + F_H|) / (|F_{PH}| + |F_P|)$.

|| $R = \sum |F_{\text{obs}}(hkl)| - |F_{\text{calc}}(hkl)| / \sum |F_{\text{obs}}(hkl)|$.

¶According to the program PROCHECK (Laskowski *et al.*, 1993). The percentages indicated are of residues in the most favoured, additionally allowed, generously allowed and disallowed regions of the Ramachandran plot, respectively.

#Estimates provided by the program REFMAC (Murshudov *et al.*, 1997).

Receptor binding

Experiments carried out to compare the effects of infecting coxsackie and adenovirus receptor (CAR)-deficient Chinese hamster ovary (CHO) cells with avian adenovirus, with the CHO cells expressing human CAR, suggest that avian adenovirus binds to human, and consequently avian, CAR. Subsequent fiber-deletion studies have implicated the long fiber in CAR binding, whilst the receptor for the short fiber remains to be identified (Tan *et al.*, 2001). Therefore, we tried to measure the binding of the purified avian adenovirus long fiber head domain with domain 1 of CAR; this domain was shown previously to be necessary and sufficient for high-affinity human adenovirus fiber head binding (Freimuth *et al.*, 1999; Lortat-Jacob *et al.*, 2001).

Our results from gel filtration (P. Guardado-Calvo & M. J. van Raaij, unpublished results) and surface plasmon resonance experiments (Hugues Lortat-Jacob, personal communication) did not show any interaction. This suggests either that, as is the case with coxsackievirus binding, domain 2 of CAR may be required for efficient binding of the avian adenovirus long fiber head, or that the long fiber shaft contains elements necessary for CAR binding. The other alternative is that CAR may not be a physiological receptor for avian adenovirus. It should be noted that these experiments were carried out with domain 1 of human CAR, not chicken CAR (NCBI RefSeq no. XW-416681; Pruitt *et al.*, 2005), presumably because the chicken CAR sequence was not known at the time and due to the interest in using avian adenovirus as a human gene therapy or

anti-cancer vehicle. However, domain 1 of the predicted CAR homologue is 61% identical to human CAR, and residues of human CAR known to interact with adenovirus fiber head (Bewley *et al.*, 1999; Law & Davidson, 2005) are identical (Glu⁵⁶, Leu⁷³, Ser⁷⁵, Tyr⁸⁰, Tyr⁸³, Lys¹²¹ and Lys¹²³) or similar (Asp⁶⁸/Glu and Val⁷⁰/Thr) in chicken CAR.

The location of the CAR-binding site on adenovirus fiber heads has been determined by mutation analysis for human adenovirus type 5 (Kirby *et al.*, 1999; Roelvink *et al.*, 1999) and from structural data for human adenovirus type 12 (Bewley *et al.*, 1999) and type 37 (Seiradake *et al.*, 2006); it is located on the side of the molecule and involves residues of the AB, CD and DG loops (Law & Davidson, 2005). Structural alignments of the avian adenovirus long fiber head with fiber heads from the CAR-binding human adenoviruses revealed that few of the amino acids involved in CAR binding were conserved in the long fiber head. Moreover, the 2 aa that were absolutely conserved (a lysine in the B strand and a proline in the DG loop) were lacking in the long fiber head. In fact, the equivalent surface was significantly different in shape and charge (Fig. 3). It should be noted that the canine adenovirus type 2 fiber head does bind CAR, both in surface plasmon resonance experiments and crystals (Seiradake *et al.*, 2006), although it does not contain the conserved lysine and proline residues. The structure of the CAR-binding region of the DG loop in the canine adenovirus type 2 fiber head is more similar to that of the human adenovirus fiber head than the avian adenovirus long fiber head.

Superposition of the avian adenovirus long fiber head structure onto that of the human adenovirus type 12–CAR domain 1 complex (Fig. 4) indicated that the different orientation of the DG loop was incompatible with CAR domain 1 binding in the same way that it binds to adenovirus type 12 (or any of the other adenovirus fiber head–CAR complexes analysed structurally so far), suggesting that either CAR binds in a different fashion to the avian adenovirus long fiber, as discussed above, or that CAR is not the real receptor. Site-directed mutagenesis experiments in conjunction with cross-linking or co-crystallization studies with (avian) CAR or an alternative receptor are needed to reveal the actual long fiber head residues responsible for receptor binding.

Conclusions and perspective

We have solved the structure of the receptor-binding domain of the avian adenovirus long fiber. The structure provides insights at the molecular level into the interactions required to form an intact fiber head and reveals what are likely to be functionally important differences at the putative receptor-binding site from the human adenovirus fiber head structures. The structural data will be of use in efforts focused on elucidating the early steps of the avian adenovirus replication cycle and may lead to therapeutic applications for avian (and human) diseases, by facilitating the design of artificial, chimeric fibers based on the

structure, which could be used as vaccination or gene therapy vectors, or as anti-adenoviral peptides, which could block receptor binding through competitive inhibition.

Future structural work will target the shaft domain and is likely to require the deletion or modification of the four-glycine hinge domain, as carried out by Merckel *et al.* (2005) for the bacteriophage PRD1 p5 spike protein; alternatively, fusion to the foldon trimerization domain may be employed (Papanikolopoulou *et al.*, 2004). Co-crystallization of the long fiber head with its receptor may also be undertaken once the receptor elements necessary for fiber head binding have been identified.

ACKNOWLEDGEMENTS

We thank Lois Hermo for technical assistance and Hugues Lortat-Jacob for surface plasmon resonance measurements. This research was funded by research grants BFU2005-02974 from the Spanish Ministry of Education and Science and PGIDIT03PXIC20307PN from the Xunta de Galicia. Both these grants, and the rotating anode diffractometer used, were co-financed by the European Union Regional Development Fund. P.G.-C. was financed by a fellowship from the Deputación da Coruña and M.J.v.R. by a 'Ramón y Cajal' fellowship of the Spanish Ministry of Education and Science.

REFERENCES

- Bewley, M. C., Springer, K., Zhang, Y. B., Freimuth, P. & Flanagan, J. M. (1999). Structural analysis of the mechanism of adenovirus binding to its human cellular receptor, CAR. *Science* **286**, 1579–1583.
- Burmeister, W. P., Guilligay, D., Cusack, S., Wadell, G. & Arnberg, N. (2004). Crystal structure of species D adenovirus fiber knobs and their sialic acid binding sites. *J Virol* **78**, 7727–7736.
- Chappell, J. D., Protá, A. E., Dermody, T. S. & Stehle, T. (2002). Crystal structure of reovirus attachment protein $\sigma 1$ reveals evolutionary relationship to adenovirus fiber. *EMBO J* **21**, 1–11.
- Chiocca, S., Kurzbauer, R., Schaffner, G., Baker, A., Mautner, V. & Cotten, M. (1996). The complete DNA sequence and genomic organization of the avian adenovirus CELO. *J Virol* **70**, 2939–2949.
- Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**, 760–763.
- Cowen, B., Calnek, B. W., Menendez, N. A. & Ball, R. F. (1978). Avian adenoviruses: effect on egg production, shell quality, and feed consumption. *Avian Dis* **22**, 459–470.
- Cowtan, K. & Main, P. (1998). Miscellaneous algorithms for density modification. *Acta Crystallogr D Biol Crystallogr* **54**, 487–493.
- Durmort, C., Stehlin, C., Schoehn, G., Mitraki, A., Drouet, E., Cusack, S. & Burmeister, W. P. (2001). Structure of the fiber head of Ad3, a non-CAR-binding serotype of adenovirus. *Virology* **285**, 302–312.
- Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Mol Graph Model* **15**, 132–134.
- Fabry, C. M., Rosa-Calatrava, M., Conway, J. F., Zubieta, C., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2005). A quasi-atomic model of human adenovirus type 5 capsid. *EMBO J* **24**, 1645–1654.
- Francois, A., Chevalier, C., Delmas, B., Etteradossi, N., Toquin, D., Rivallan, G. & Langlois, P. (2004). Avian adenovirus CELO recombinants expressing VP2 of infectious bursal disease virus

- induce protection against bursal disease in chickens. *Vaccine* **22**, 2351–2360.
- Freimuth, P., Springer, K., Berard, C., Hainfeld, J., Bewley, M. & Flanagan, J. (1999). Coxsackievirus and adenovirus receptor amino-terminal immunoglobulin V-related domain binds adenovirus type 2 and fiber knob from adenovirus type 12. *J Virol* **73**, 1392–1398.
- Guardado-Calvo, P., Fox, G. C., Hermo-Parrado, X. L., Llamas-Saiz, A. L., Costas, C., Martinez-Costas, J. M., Benavente, J. & van Raaij, M. J. (2005). Structure of the carboxy-terminal receptor-binding domain of avian reovirus fibre σ C. *J Mol Biol* **354**, 137–149.
- Guardado-Calvo, P., Llamas-Saiz, A. L., Langlois, P. & van Raaij, M. J. (2006). Crystallization of the C-terminal head domain of the avian adenovirus CELO long fibre. *Acta Crystallogr Sect F Struct Biol Cryst Commun* **62**, 449–452.
- Henrick, K. & Thornton, J. M. (1998). PQS: a protein quaternary structure file server. *Trends Biochem Sci* **23**, 358–361.
- Hess, M., Cuzange, A., Ruigrok, R. W., Chroboczek, J. & Jacrot, B. (1995). The avian adenovirus penton: two fibres and one base. *J Mol Biol* **252**, 379–385.
- Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc Natl Acad Sci U S A* **93**, 13–20.
- Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and location of errors in these models. *Acta Crystallogr A* **47**, 110–119.
- Kelleher, Z. T. & Vos, J. M. (1994). Long-term episomal gene delivery in human lymphoid cells using human and avian adenoviral-assisted transfection. *Biotechniques* **17**, 1110–1117.
- Kirby, I., Davison, E., Beavil, A. J., Soh, C. P., Wickham, T. J., Roelvink, P. W., Kovetski, I., Sutton, B. J. & Santis, G. (1999). Mutations in the DG loop of adenovirus type 5 fiber knob protein abolish high-affinity binding to its cellular receptor CAR. *J Virol* **73**, 9508–9514.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* **24**, 946–950.
- Lamzin, V. S. & Wilson, K. S. (1993). Automated refinement of protein models. *Acta Crystallogr D Biol Crystallogr* **49**, 129–149.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26**, 283–291.
- Laver, W. G., Younghusband, H. B. & Wrigley, N. G. (1971). Purification and properties of chick embryo lethal orphan virus (an avian adenovirus). *Virology* **45**, 598–614.
- Law, L. K. & Davidson, B. L. (2005). What does it take to bind CAR? *Mol Ther* **12**, 599–609.
- Logunov, D. Y., Ilyinskaya, G. V., Cherenova, L. V., Verhovskaya, L. V., Shmarov, M. M., Chumakov, P. M., Kopnin, B. P. & Naroditsky, B. S. (2004). Restoration of p53 tumor-suppressor activity in human tumor cells *in vitro* and in their xenografts *in vivo* by recombinant avian adenovirus CELO-p53. *Gene Ther* **11**, 79–84.
- Lortat-Jacob, H., Chouin, E., Cusack, S. & van Raaij, M. J. (2001). Kinetic analysis of adenovirus fiber binding to its receptor reveals an avidity mechanism for trimeric receptor–ligand interactions. *J Biol Chem* **276**, 9009–9015.
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* **61**, 458–464.
- McCracken, R. M. & Adair, B. M. (1993). Avian adenoviruses. In *Viral Infections of Vertebrates*, vol. 3, pp. 123–144. Edited by J. B. McFerran & M. S. McNulty. Amsterdam: Elsevier.
- Merckel, M. C., Huiskonen, J. T., Bamford, D. H., Goldman, A. & Tuma, R. (2005). The structure of the bacteriophage PRD1 spike sheds light on the evolution of viral capsid architecture. *Mol Cell* **18**, 161–170.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
- Navaza, J. (2001). Implementation of molecular replacement in AMoRe. *Acta Crystallogr D Biol Crystallogr* **57**, 1367–1372.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307–326.
- Papanikolopoulou, K., Teixeira, S., Belrhali, H., Forsyth, V. T., Mitraki, A. & van Raaij, M. J. (2004). Adenovirus fibre shaft sequences fold into the native triple β -spiral fold when N-terminally fused to the bacteriophage T4 fibrin foldon trimerisation motif. *J Mol Biol* **342**, 219–227.
- Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* **6**, 458–463.
- Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). Developments in the CCP4 molecular-graphics project. *Acta Crystallogr D Biol Crystallogr* **60**, 2288–2294.
- Pruitt, K. D., Tatusova, T. & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **33**, D501–D504.
- Roelvink, P. W., Mi Lee, G., Einfeld, D. A., Kovetski, I. & Wickham, T. J. (1999). Identification of a conserved receptor-binding site on the fiber proteins of CAR-recognizing adenoviridae. *Science* **286**, 1568–1571.
- Saban, S. D., Silvestry, M., Nemerow, G. R. & Stewart, P. L. (2006). Visualization of α -helices in a 6 Å resolution cryoelectron microscopy structure of adenovirus allows refinement of capsid protein assignments. *J Virol* **80**, 12049–12059.
- Schneider, T. R. & Sheldrick, G. M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**, 1772–1779.
- Seiradake, E. & Cusack, S. (2005). Crystal structure of enteric adenovirus serotype 41 short fiber head. *J Virol* **79**, 14088–14094.
- Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E. J. & Cusack, S. (2006). Structural and mutational analysis of human Ad37 and canine adenovirus 2 fibre heads in complex with the D1 domain of CAR. *J Biol Chem* **281**, 33704–33716.
- Shashkova, E. V., Cherenova, L. V., Kazansky, D. B. & Dronin, K. (2005). Avian adenovirus vector CELO-TK displays anticancer activity in human cancer cells and suppresses established murine melanoma tumors. *Cancer Gene Ther* **12**, 617–626.
- Spinelli, S., Desmyter, A., Verrips, C. T., de Haard, H. J., Moineau, S. & Cambillau, C. (2006a). Lactococcal bacteriophage p2 receptor-binding protein structure suggests a common ancestor gene with bacterial and mammalian viruses. *Nat Struct Mol Biol* **13**, 85–89.
- Spinelli, S., Campanacci, V., Blangy, S., Moineau, S., Tegoni, M. & Cambillau, C. (2006b). Modular structure of the receptor binding proteins of *Lactococcus lactis* phages. The RBP structure of the temperate phage TP901-1. *J Biol Chem* **281**, 14256–14262.
- Stevenson, M., Boos, E., Herbert, C., Hale, A., Green, N., Lyons, M., Chandler, L., Ulbrich, K., van Rooijen, N. & other authors (2006). Chick embryo lethal orphan virus can be polymer-coated and retargeted to infect mammalian cells. *Gene Ther* **13**, 356–368.
- Tan, P. K., Michou, A. I., Bergelson, J. M. & Cotten, M. (2001). Defining CAR as a cellular receptor for the avian adenovirus CELO using a genetic analysis of the two viral fibre proteins. *J Gen Virol* **82**, 1465–1472.

van Raaij, M. J., Louis, N., Chroboczek, J. & Cusack, S. (1999a). Structure of the human adenovirus serotype 2 fiber head domain at 1.5 Å resolution. *Virology* **262**, 333–343.

van Raaij, M. J., Mitraki, A., Lavigne, G. & Cusack, S. (1999b). A triple β -spiral in the adenovirus fibre shaft reveals a new structural motif for a fibrous protein. *Nature* **401**, 935–938.

Xia, D., Henry, L. J., Gerard, R. D. & Deisenhofer, J. (1994). Crystal structure of the receptor-binding domain of adenovirus type 5 fiber protein at 1.7 Å resolution. *Structure* **2**, 1259–1270.

Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). The structure of the human adenovirus 2 penton. *Mol Cell* **17**, 121–135.

Pablo Guardado-Calvo,^a
Antonio L. Llamas-Saiz,^b
Gavin C. Fox,^c Joel N. Glasgow^d
and Mark J. van Raaij^{a,e*}

^aDepartamento de Bioquímica y Biología Molecular, Facultad de Farmacia, Universidad de Santiago de Compostela, Spain, ^bUnidad de Rayos X (RIADT), Laboratorio Integral de Dinámica y Estructura de Biomoléculas José R. Carracido, Edificio CACTUS, Universidad de Santiago de Compostela, Spain, ^cLaboratoire des Proteines Membranaires, Institut de Biologie Structurale J. P. Ebel, 41 Rue Jules Horowitz, Grenoble, France, ^dDivisions of Cardiology and Human Gene Therapy, Departments of Medicine, Obstetrics and Gynecology, Pathology, Surgery and the Gene Therapy Center, University of Alabama at Birmingham, Birmingham, Alabama, USA, and ^eDepartamento de Biología Estructural, Instituto de Biología Molecular (IBMB-CSIC), Parc Científic, Baldiri Reixac 10, 08028 Barcelona, Spain

Correspondence e-mail:
mark.vanraaij@ibmb.csic.es

Received 21 July 2009
Accepted 21 September 2009



© 2009 International Union of Crystallography
All rights reserved

Crystallization of the head and galectin-like domains of porcine adenovirus isolate NADC-1 fibre

The porcine adenovirus NADC-1 isolate, a strain of porcine adenovirus type 4, has a fibre with an atypical architecture. In addition to a classical virus-attachment region, shaft and head domains, it contains an additional galectin-like domain C-terminal to the head domain and connected to the head domain by a long RGD-containing loop. The galectin-like domain contains two putative carbohydrate-recognition domains. The head and galectin-like domains have been independently crystallized. Diffraction data have been obtained to 3.2 Å resolution from crystals of the head domain and to 1.9 Å resolution from galectin-like domain crystals.

1. Introduction

Adenoviruses may find application as vectors for gene therapy and cancer therapy and as vaccination agents (for recent reviews, see Arnberg, 2009; Bachtarzi *et al.*, 2008; Thacker *et al.*, 2009). The family *Adenoviridae* contains over a hundred known serotypes, including 52 that infect humans. They are non-enveloped viruses with a linear double-stranded DNA genome that can infect all five major vertebrate classes. They have an icosahedral ($T = 25$) capsid consisting of three major proteins: the trimeric hexon, forming the facets of the particle, the pentameric penton base, which forms the vertices, and the trimeric fibre protein, which extends from the penton base at the vertex positions. Proteins IIIa, VI, VIII and IX reinforce the penton-hexon and hexon-hexon interactions (Nemerow *et al.*, 2009). The principal structural determinant of adenovirus tropism is its protruding rod-like fibre protein (reviewed in Nouredini & Curiel, 2005; Nicklin *et al.*, 2005). The distal tip of each fibre is composed of a globular knob domain, which serves as the major viral attachment site for a variety of cellular receptors such as CAR (Roelvink *et al.*, 1998), CD46/80/86 (Short *et al.*, 2006) and sialic acid (Arnberg *et al.*, 2000). An RGD motif localized in the hypervariable loop of the penton base of many adenoviruses binds to integrins (Wickham *et al.*, 1993; Li *et al.*, 2001), which act as secondary receptors, promoting viral internalization. Given the importance of the fibre, structural characterization of different fibre proteins from human and nonhuman adenoviruses has been critical in understanding adenovirus tropism and in developing new vectors with modified tropisms (Nicklin & Baker, 2008). Animal adenoviruses are of particular interest in this regard, as they may be less immunogenic to humans and have novel receptor-binding properties. Recently, structures of canine and fowl adenovirus fibre heads have been published (Seiradake *et al.*, 2006; Guardado-Calvo *et al.*, 2007; El Bakkouri *et al.*, 2008).

Porcine adenovirus was isolated for the first time by Haig *et al.* (1964), is classified within the genus *Mastadenovirus* and has a genome of approximately 34 kb. It is commonly regarded as a low-grade pathogen. Restriction-enzyme mapping and sequence analysis of serotypes 1–5 has shown that serotypes 1–3 are closely related to each other while serotypes 4 and 5 are more divergent. All of the serotypes are distinct from human, murine, canine, bovine and fowl adenoviruses (Hammond & Johnson, 2005). The porcine adenovirus NADC-1 isolate was isolated from a tonsillar swab of an adult sow in 1972 and is thought to be a strain of porcine adenovirus type 4 (Kleiboeker *et al.*, 1993). Its 703-residue fibre protein is unique

among adenovirus strains and contains, from the N-terminus to the C-terminus, a tail domain (residues 1–37) where the interaction sequence with the penton protein is found, a short shaft domain (residues 38–120) with up to six predicted triple β -spiral repeats (van Raaij *et al.*, 1999) and what was predicted to be a larger than average head domain (residues 121–703; Kleiboeker, 1995; Fig. 1). The first 167 amino acids of this domain (121–287) are homologous to other adenoviral head domains; therefore, we propose to designate this as the head domain and differentiate the sequence C-terminal to this as follows. The sequence adjacent to the head domain (288–392) contains an RGD sequence and alanine-rich and glutamate-rich sequences that are likely to be involved in interaction with integrins. Additionally, a domain composed of two carbohydrate-recognition domains (CRDs) repeated in tandem was predicted from sequence analysis (residues 393–681). This galectin-like domain is unique among adenovirus fibres; only one other galectin-like domain has been identified in viruses (in lymphocystis disease virus; Cooper, 2002). The porcine adenovirus galectin-like domain is composed of two carbohydrate-recognition domains linked by a 23-residue sequence rich in prolines. The position of the putative CRDs at the C-terminus of the fibre suggests that the NADC-1 virus may employ interactions with cell-surface galactosides as a method of cell attachment.

Here, we describe the expression, purification, crystallization and crystallographic data collection of two protein-construct crystals containing the head (116–291) and galectin-like domains (393–703) of porcine adenovirus NADC-1.

2. Methods

2.1. Construction of expression vectors

Porcine adenovirus NADC-1 (catalogue No. 030-PDV) was obtained from the National Veterinary Services Laboratories of the USDA Animal and Plant Health Inspection Service (Ames, Iowa). The NADC-1 fibre gene sequence was amplified directly from the viral stock using the polymerase chain reaction (PCR). DNA fragments encoding porcine adenovirus NADC-1 fibre (UniProt accession code Q83467) residues 116–291 (the predicted head domain; Fig. 1) and 393–703 (the predicted galectin-like domain) were produced by PCR from a plasmid (pKan-NMOvPad4) which encodes the shaft, head and galectin-like domains of the fibre gene and cloned into the expression vector pET28c(+) (Novagen/Merck, Nottingham, England) between the *Bam*HI and *Not*I restriction sites. The start codon of the expression vector pET28c(+) was used; a stop codon was inserted after residue 291 for the head-domain expression clone; the galectin-like domain, being the C-terminal domain, ends at residue 703, the last residue of the porcine adenovirus type 4 fibre. The resultant plasmids pET28c-pAd4Fib116–291 and pET28c-pAd4Fib393–703 encode the head and galectin-like domains, respectively, fused to an N-terminal purification tag containing six consecutive histidine residues and a T7 tag (MGSSH HHHHH SSGLV PRGSH MASMT GGQQG RI; Novagen; the T7 tag is an 11-amino-acid peptide encoded in the leader sequence of T7 bacteriophage gene10 for which specific monoclonal antibodies and antibody-containing chromatography resins are available). The sequences of all inserts were confirmed by DNA-sequence analysis (Sistemas Genómicos, Valencia, Spain).

2.2. Protein expression and purification

Escherichia coli strain BL21 (DE3) was freshly transformed with pET28c-pAd4Fib116–291 or pET28c-pAd4Fib393–703 and four 0.7 l

cultures were grown aerobically at 310 K to an optical density of 0.6–0.8 measured at 600 nm. In the case of the head domain, growth was allowed to continue for 3 h at 310 K after induction with 1 mM isopropyl β -D-1-thiogalactopyranoside. For expression of the galectin-like domain, cultures were cooled on ice before induction and growth was allowed to continue for 16 h at 289 K. Harvested cells were resuspended in 40 ml of a buffer containing 50 mM Tris–HCl pH 8.5, 1 M sodium chloride and frozen at 253 K. Bacteria were lysed by a double pass through an emulsifier (Avestin EmulsiFlex C5, Avestin Europe GmbH, Mannheim, Germany). After removing insoluble material by centrifugation (20 min at 20 000g), 4 ml Ni–NTA resin (Qiagen, Valencia, California, USA) was added. The suspension was incubated for 1 h at 277 K and poured into an empty column and the resin was washed using the same buffer. For purification of the head domain, elution was performed with a step gradient of imidazole pH 7.0 (10, 20, 50, 100, 200, 400, 600, 800 and 1000 mM imidazole). Pure histidine-T7-tagged head domain eluted at high imidazole concentrations (200–800 mM), consistent with the proteins being trimeric and thus containing three histidine tags.

For purification of the galectin domain, elution was performed with the following step gradient: 10, 20, 50, 100, 200, 500 and 1000 mM imidazole pH 7.0. Histidine-T7-tagged galectin-like domain eluted at imidazole concentrations of between 50 and 500 mM. The eluted protein was dialysed against 10 mM Tris–HCl pH 8.5, 1 mM EDTA (TE buffer pH 8.5) and loaded onto an Uno-Q6 anion-exchange

```

1 MKRSVPSDFN PVYPDYQPI SLMPAFYDNY GFHEGPGSGVL SLNIANPLGY
51 TPRKKLCLKL GEGLALDSGD HLRVQIPDMQ AQPPLLYQGH RLSLLFDADA
101 GFHLTEDGAL SLTKTLVYPT LWTGPAPEAN VTFSGENSPS GILRLCLSRF
151 GGTVIGTSLV QGSLTNPSTG QTLGMNLYFD ADGNVLSSESN LVRGSGWGMKD
201 QDTLVTPIAN GQYLMPLNTA YPRLIQTLTS SYIYTQAHLD HNSVVDIKI
251 GLNTDLRPTA AYGLSFTMTF TNSPPTSFGT DLVQFGYLGQ DSSPSFLREL
301 PLASEAGYFG KLAASEEMP APPEAQTDQ AAEPPAPAE AEAPAPAEAE
351 AEAEPFRKPP RGDLAALYNR VHS DTRAEDT PTSPELVTTL PDFFVLPLPD
401 GVP TGASIVL EGTLP S AVF FTLDLVTGPA SLALHFNVRL PLEGEKHIVC
451 NSREGSSNWG EEVRPQEFPP EREKPFVLVI VIQSDTYQIT VNGKPLVDFF
501 QRLQGITRAS LSGDLVFTRL TMYPPGDRP TTLPPPPAAP LDVIPDAYVL
551 NLPTGLTPRT LLTVTGTPTP LAEFFIVNLV YDLHYDSKNV ALHFNVGFTS
601 DSKGHIAACNA RMNGTWGSEI TVSDFFPQRG KPFTLQILTR EADFQVLVDK
651 QPLTQFQYRL KELDOIKYVH MFGHVQVTHL EHQPDPVPV STAGVSKVYP
701 QIL
    
```

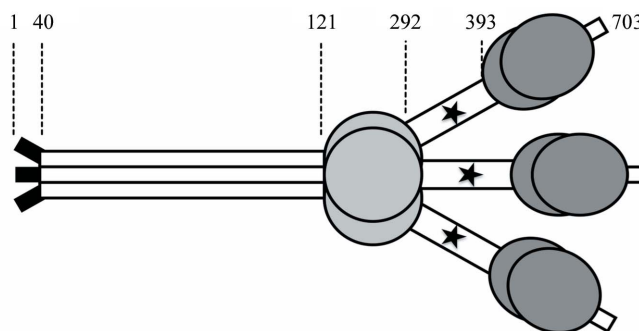


Figure 1

Predicted domain organization of porcine adenovirus type 4 NADC-1 strain fibre. The amino-acid sequence (single-letter code; above) and a schematic drawing (below) are shown. The predicted virus-binding tail (residues 1–37) is shown in black, the shaft domain (40–120) in white, the predicted head domain (121–287) in light grey, the RGD-containing domain (292–392) in white and the predicted galectin-like domain (393–681) in dark grey. The putative integrin-binding RGD sequence is indicated in bold in the sequence and with an asterisk in the drawing. The sequences expressed and crystallized (116–291 and 393–703) are underlined; they correspond to the head and galectin-like domains plus a few N-terminal and C-terminal amino acids. It should be noted that in the absence of electron-microscopy data the relative orientations and dimensions of the domains are unknown.

Table 1

Crystallographic data statistics of porcine adenovirus NADC-1 isolate fibre crystals.

Values in parentheses are for the highest resolution bin (where applicable).

	Head domain	Galectin-like domain
Radiation source	Cu K α (Bruker–Nonius FR591)	ID14-2 (ESRF)
Temperature (K)	295	100
Wavelength (\AA)	1.5418	0.9390
Detector	Kappa CCD-2000	ADSC Q4 CCD
Crystal-to-detector distance (mm)	90.0	301.1
Oscillation range ($^{\circ}$)	0.6	0.8
No. of images	166	272
Space group	$P2_12_12_1$	$C2$
Unit-cell parameters (\AA , $^{\circ}$)	$a = 125.7$, $b = 145.4$, $c = 147.6$, $\alpha = \beta = \gamma = 90.0$	$a = 167.6$, $b = 77.3$, $c = 94.1$, $\alpha = \gamma = 90.0$, $\beta = 101.5$
Mosaicity ($^{\circ}$)	0.36	0.68
Unique observations	45035 (4367)	89027 (12129)
Resolution range (\AA)	40–3.2 (3.32–3.2)	40–1.9 (2.0–1.9)
Multiplicity	3.8 (3.8)	4.1 (3.8)
Completeness (%)	98.9 (97.5)	96.1 (90.1)
$\langle I/\sigma(I) \rangle$	17.5 (2.9)	10.0 (3.7)
$R_{\text{merge}}^{\dagger}$ (%)	7.2 (30.7)	7.5 (35.2)
V_M^{\ddagger} ($\text{\AA}^3 \text{Da}^{-1}$)	6.0	2.2

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the i th measurement of the same reflection and $\langle I(hkl) \rangle$ is the mean observed intensity for that reflection. \ddagger According to Matthews (1968) and assuming two domain trimers and four copies of the galectin domain in the asymmetric unit.

column (Bio-Rad, Madrid, Spain). The protein did not bind and eluted in the flowthrough. Fractions containing galectin-domain protein were pooled, concentrated to 1.0 ml using Centricon concentrators (Milipore, Madrid, Spain) and loaded onto a Sephacryl S100 gel-filtration column (GE Healthcare, Alcobendas, Spain) equilibrated with TE buffer pH 8.5 containing 150 mM sodium chloride.

2.3. Crystallization and data collection

Crystallization took place by vapour diffusion in sitting-drop CompactClover plates (Jena Biosciences, Jena, Germany), with 0.15 ml reservoirs and drops consisting of 4–5 μl protein solution mixed with equal volumes of reservoir solution. The head-domain protein was concentrated to 40–60 mg ml $^{-1}$ using Centricon concentrators, incorporating three washes with TE buffer pH 8.0 to eliminate small-molecule impurities. 1 M lithium sulfate supplemented with polyethylene glycol was used as the reservoir solution for crystallization; equivalent crystals were obtained over the range 1–10% (w/v) polyethylene glycol 8000. Ligand-free galectin-domain protein was concentrated to 10–20 mg ml $^{-1}$ using Centricon concentrators, incorporating three washes with TE buffer pH 8.0. Galectin-like domain crystals were obtained using 25% (w/v) polyethylene glycol 3350, 500 mM sodium nitrate as reservoir solution and with 2 mM sialic acid added to the drop as an additive (although carbohydrate-array experiments showed that the galectin-like domain does not bind sialic acid; results to be published elsewhere). For data collection, galectin-like domain crystals were transferred to reservoir solution containing 20% (v/v) glycerol, keeping the concentrations of the precipitant components constant. For both head-domain and galectin-like domain protein preparations, the final buffer was TE buffer pH 8, although traces of imidazole or sodium chloride may remain, respectively. We did not check the final pH; owing to the low concentration of buffering agent the pH may have changed upon the addition of precipitant solution and/or cryo-protectant.

2.4. Computation

Prediction of structural disorder was performed with the programs *FOLDINDEX* (Prilusky *et al.*, 2005) and *RONN* (Yang *et al.*, 2005). Crystallographic data were integrated and scaled using the *HKL-2000* suite of programs (Otwinowski & Minor, 1997; head-domain crystals) or by a combination of *MOSFLM* (Leslie, 2006) and *SCALA* (Collaborative Computational Project, Number 4, 1994). Data-processing statistics are summarized in Table 1. Self-rotation functions were calculated with the program *XPREP* (Sheldrick & Schneider, 2001).

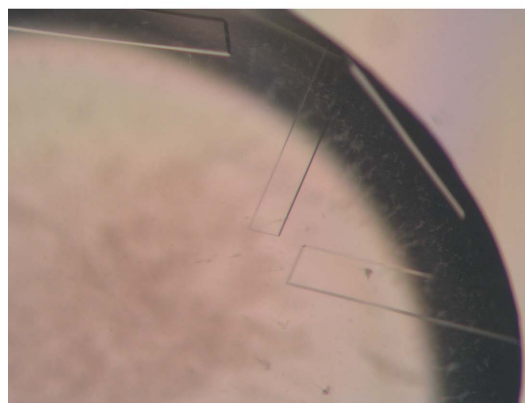
3. Results and discussion

The fibre of the porcine adenovirus NADC-1 isolate is predicted to be composed of an N-terminal virus-binding peptide that is likely to be disordered in the absence of the rest of the virus, a relatively short shaft region, a head domain and a predicted galectin-like domain (Fig. 1). *In silico* analysis predicts the region between the knob and the galectin-like domains (amino acids 292–392) to be flexible. Consistent with this, truncated versions of the protein containing residues 292–392 either did not form crystals or the crystals obtained did not diffract. The loop of the human adenovirus type 2 penton that contains the RGD sequence is also highly flexible and this flexibility may be related to its biological function (Zubieta *et al.*, 2005). We did obtain diffraction-quality crystals of the head domain and of the galectin-like domain independently.

For the head domain (amino acids 116–291), large crystals were obtained (Fig. 2a). A variety of different cryoprotection solutions



(a)



(b)

Figure 2

Crystals of (a) head-domain and (b) galectin-domain crystals, measuring approximately 0.5 \times 0.5 \times 0.5 and 0.4 \times 0.1 \times 0.02 mm, respectively.

were tried without success (oil-based methods were not tried). However, a complete data set was collected to 3.2 Å resolution at room temperature (Table 1). The self-rotation function (not shown) showed the presence of two independent threefold noncrystallographic axes, suggesting the presence of two trimers in the asymmetric unit and a solvent content of 79.3% (Matthews, 1968). Crystals of the galectin-like domain appeared after 1–2 weeks (Fig. 2*b*). A complete data set was collected to 1.9 Å resolution (Table 1). Four copies of the galectin domain are expected in the asymmetric unit.

Structure solution by molecular replacement has been performed; refinement and analysis are ongoing and will be reported elsewhere. The structures will provide insight into the receptor and carbohydrate ligand-binding properties of the porcine adenovirus NADC-1 isolate fibre. These structural data may also be useful for designing mutants that bind other sugars and chimeric adenoviruses targeting specific cell types (Paul *et al.*, 2008). If suitable cryo-electron microscopy images can be obtained, the high-resolution crystal structures will be fitted into EM maps to produce an overall picture of the relative organization of the domains in the intact fibre.

We thank Patricia Ferraces-Casais for excellent technical assistance. This research was sponsored by research grants BFU2005-02974, BFU2005-24982-E and BFU2008-01588 and by a predoctoral FPU fellowship to PG-C from the Spanish Ministry of Education and Science. This work was also supported by funds from the European Commission under contracts ERAS-CT-2003-980409 (as part of the European Science Foundation EUROCORES Programme EuroSCOPE) and NMP4-CT-2006-033256 (BeNatural coordinated project) and by the Xunta de Galicia (grant INCITE08E1R203013ES).

References

- Arnberg, N. (2009). *Rev. Med. Virol.* **19**, 165–178.
- Arnberg, N., Edlund, K., Kidd, A. H. & Wadell, G. (2000). *J. Virol.* **74**, 42–48.
- Bachtarzi, H., Stevenson, M. & Fisher, K. (2008). *Expert Opin. Drug Deliv.* **5**, 1231–1240.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cooper, D. N. W. (2002). *Biochim. Biophys. Acta*, **1572**, 209–231.
- El Bakkouri, M., Seiradake, E., Cusack, S., Ruigrok, R. W. & Schoehn, G. (2008). *Virology*, **378**, 169–176.
- Guardado-Calvo, P., Llamas-Saiz, A. L., Fox, G. C., Langlois, P. & van Raaij, M. J. (2007). *J. Gen. Virol.* **88**, 2407–2416.
- Haig, D. A., Clarke, M. C. & Pereira, M. S. (1964). *J. Comp. Pathol.* **74**, 81–84.
- Hammond, J. M. & Johnson, M. A. (2005). *Vet. J.* **169**, 17–27.
- Kleiboeker, S. B. (1995). *Virus Res.* **39**, 299–309.
- Kleiboeker, S. B., Seal, B. S. & Mengeling, W. L. (1993). *Arch. Virol.* **133**, 357–368.
- Leslie, A. G. W. (2006). *Acta Cryst.* **D62**, 48–57.
- Li, E., Brown, S. L., Stupack, D. G., Puente, X. S., Cheresch, D. A. & Nemerow, G. R. (2001). *J. Virol.* **75**, 5405–5409.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Nemerow, G. R., Pache, L., Reddy, V. & Stewart, P. L. (2009). *Virology*, **384**, 380–388.
- Nicklin, S. A. & Baker, A. H. (2008). *Mol. Ther.* **16**, 1904–1905.
- Nicklin, S. A., Wu, E., Nemerow, G. R. & Baker, A. H. (2005). *Mol. Ther.* **12**, 384–393.
- Noureddini, S. C. & Curiel, D. T. (2005). *Mol. Pharm.* **2**, 341–347.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Paul, C. P., Everts, M., Glasgow, J. N., Dent, P., Fisher, P. B., Ulasov, I. V., Lesniak, M. S., Stoff-Khalili, M. A., Roth, J. C., Preuss, M. A., Dirven, C. M., Lamfers, M. L., Siegal, G. P., Zhu, Z. B. & Curiel, D. T. (2008). *Cancer Biol. Ther.* **7**, 786–793.
- Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I. & Sussman, J. L. (2005). *Bioinformatics*, **21**, 3435–3438.
- Raaij, M. J. van, Mitraki, A., Lavigne, G. & Cusack, S. (1999). *Nature (London)*, **401**, 935–938.
- Roelvink, P. W., Lizonova, A., Lee, J. G., Li, Y., Bergelson, J. M., Finberg, R. W., Brough, D. E., Kovessdi, I. & Wickham, T. J. (1998). *J. Virol.* **72**, 7909–7915.
- Seiradake, E., Lortat-Jacob, H., Billet, O., Kremer, E. J. & Cusack, S. (2006). *J. Biol. Chem.* **281**, 33704–33716.
- Sheldrick, G. M. & Schneider, T. R. (2001). *Methods in Macromolecular Crystallography*, edited by D. Turk & L. Johnson, pp. 72–81. Amsterdam: IOS Press.
- Short, J. J., Vasu, C., Holterman, M. J., Curiel, D. T. & Pereboev, A. (2006). *Virus Res.* **122**, 144–153.
- Thacker, E. E., Timares, L. & Matthews, Q. L. (2009). *Expert Rev. Vaccines*, **8**, 761–777.
- Wickham, T. J., Mathias, P., Cheresch, D. A. & Nemerow, G. R. (1993). *Cell*, **73**, 309–319.
- Yang, Z., Thomson, R., McNeil, P. & Esnouf, R. M. (2005). *Bioinformatics*, **21**, 3369–3376.
- Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). *Mol. Cell*, **17**, 121–135.