

Universidad Autónoma de Madrid

Facultad de Ciencias

Departamento de Biología Molecular

Análisis estructural de CD5 y DDX3, dos proteínas
de la célula huésped implicadas en la infección por
el virus de la hepatitis C

Memoria presentada por **Bernardo Rodamilans Ramos** para optar al grado
de Doctor en Ciencias.

Mayo, 2007

Esta tesis ha sido realizada en el Centro Nacional de Investigaciones
Oncológicas (CNIO) bajo la dirección del **Dr. Guillermo Montoya Blanco**

Agradecimientos

Quiero agradecer al Dr. Guillermo Montoya por haberme permitido realizar este trabajo en este laboratorio, por su dirección y su apoyo.

A todos los compañeros del laboratorio, de los que he aprendido y los que me han ayudado. A Lissy y a Inés con las que empecé este camino, a Bego, con la que he disfrutado de más de una cerveza, a Sonia, siempre con una serpiente en el bolso y dispuesta a reír por una costilla, a Javi, que no solo nada 2000m sino que salvó la superdex a tiempo, a Jaska por sus jaskarrillos y su queso, a MJ, que nos llevó a ser aironmanes, a June, que a estas alturas ya habrá pasado, a Vladimir por sus seminarios, a Pablo, por su PWM y su conversación, a Javi (nuevo), que aunque todavía no lo hemos comprobado, dice ser 6º de España a braza, a David, gran cocinero que vino para dos días y se quedó seis meses y a Antonio, por terminar la tesis por mi, a pesar de nuestra falta de entendimiento.

También quiero agradecer a todos los compañeros del laboratorio de NMR (los que están y los que no) siempre dispuestos a tomarse algo conmigo aunque no beban lo mismo (different people, different needs). A los del laboratorio de Jero. Juntos, no sólo solucionamos los problemas del Tekan sino que pudimos conocer más a fondo a Costar. Gracias por su apoyo, no sólo científico, sino también moral y anímico. Del mismo modo, al resto de personas del CNIO con las que he tratado y que me han ayudado en estos años.

Gracias a todos vosotros que fuera del laboratorio habéis compartido conmigo las risas y las penas. A Ana, que me ha soportado incansablemente todos estos años a pesar de que soy un pesado. A mis compañeros de piso y amigos, David, Jon y Guillermo y al resto de personas y amigos que han colaborado desde su buen humor y compañía a que termine con éxito esta tesis.

Gracias a mi familia que siempre me ha prestado su apoyo y comprensión.

A mis padres y hermanos.

Gracias.

Summary

Hepatitis C virus is a major human pathogen that affects more than 170 million people worldwide. 80% of the patients that suffer this disease, develop chronic liver damage that can lead to hepatocarcinoma. The work carried out during this Ph.D. Thesis was aimed to the analysis of nucleocapsid like particles (NLPs) produced *in vitro* and their interactions inside and outside the cell.

The human proteins involved in the study of HCV were DDX3, an RNA helicase that directly interacts with protein C of the nucleocapsid of HCV, and CD5, a lymphocyte receptor related to the lymphotropism of HCV during the infection.

The structure of the helicase domain of DDX3 was determined at a resolution of 1.9Å. Comparison of this structure with the homologous protein from *D. melanogaster* Vasa that was previously solved, shed light into the structure of the ATPase domain and the binding site to ARN.

The structure of the third extracellular domain of CD5 was solved at a resolution of 2.2Å. Detailed analysis of the structure allowed to model the whole receptor and to model the interaction between the closely related CD6 and its binding partner ALCAM/CD166.

Abreviaturas

5'UTR	Región no codificante en el extremo 5'
ACN	Acetonitrilo
ATP	Adenosina trifosfato
ARN	Ácido ribonucleico
ccVHC	Partículas virales derivadas de cultivos celulares
C-terminal	Extremo carboxilo terminal
DC	Dicroísmo circular
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin
DDX3 _{ATP}	DDX3 173-406
DDX3 _{hel}	DDX3 407-578
DDX3 _{pdr}	DDX3 164-579
DDX3 _{pf}	DDX3 164-600
DDX3 _{RS}	DDX3 164-623
DI	Dominio I
DII	Dominio II
DIII	Dominio III
DLS	Dispersión dinámica de luz
<i>D.melanogaster</i>	<i>Drosophila melanogaster</i>
D.O.	Densidad óptica
DSS	2,2-dimetil-2-silapentano-5-sulfonato sódico
DTT	Ditiotreitol
<i>E.coli</i>	<i>Escherichia coli</i>

EDTA	Ácido etilendiaminotetraacético
ESRF	European Synchrotron Radiation Facilities
HPLC	Cromatografía líquida de alta presión
IgM	Inmunoglobulinas M
IPTG	Isopropil- β -D-tiogalactopiranosido
IRES	Región interna de entrada del ribosoma
LDH	Lactato deshidrogenasa
LEL	Lazo extracelular grande
MAb	Anticuerpo monoclonal
NADH	Nicotinamida adenina dinucleótido
NLPs	Partículas similares a nucleocápsida
N-terminal	Extremo amino terminal
ORF	Origen de lectura abierta
PBS	Tampón salino fosfato
PCR	Reacción en cadena de la polimerasa
PEP	Fosfoenolpiruvato
PK	Piruvato kinasa
PONDR	Predictor of Naturally Disorder Regions
ppVHC	Pseudo-partículas de VHC
Proteína C	Proteína del core
PVDF	Difluoruro de polivinilo
RPM	Revoluciones por minuto
RMN	Resonancia magnética nuclear
r.m.s.	Raíz media cuadrática
r.m.s.d.	Desviación de la raíz media cuadrática

RT-PCR	Retrotranscriptasa - PCR
SBF	Suero bovino fetal
SEL	Lazo extracelular pequeño
SIDA	Síndrome de inmunodeficiencia adquirida
SLS	Swiss Light Source
SR-BI	Scavenger Receptor Class B type I
SRCR	Scavenger Receptor Cysteine Rich
TEV	Virus del jaspeado del tabaco
TLS	Translation/Libration/Screw
T_m	Temperatura media de desnaturalización
TMV	Virus del mosaico del tabaco
TNFR-1	Tumor Necrosis Factor Receptor-1
VIH	Virus de inmunodeficiencia humana
VHB	Virus de la hepatitis B
VHC	Virus de la hepatitis C
VHCc124	Proteína C 1-124

ÍNDICE

<u>Introducción</u>	1
1.1. La enfermedad	3
1.2. El virus	5
1.3. Entrada del virus en la célula	7
1.4. La proteína C y su interacción con diferentes dianas celulares	9
1.4.1. Interacción con factores nucleares de transcripción	10
1.4.2. Interacción con el sistema de señalización de apoptosis	11
1.4.3. Interacción con el sistema de traducción celular	11
1.4.3.1. Descripción de las ARN helicasas	12
1.4.3.2. Detalles estructurales de DDX3	12
1.4.3.3. Función de DDX3	13
1.5. Interacciones del VHC a nivel del sistema inmune	14
1.5.1. CD5, descripción y función.	17
1.5.1.1. La superfamilia SRCR	17
1.5.1.2. Datos estructurales previos de CD5	18
1.5.1.3. Datos funcionales de CD5	20
1.6. Tratamientos contra la enfermedad	20
<u>Objetivos</u>	23
<u>Materiales y Métodos</u>	27
2.1. Proteína C de VHC	29
2.1.1 Clonaje y expresión de la proteína C	29
2.1.2. Purificación de VHCc124	29
2.1.3. Reconstitución <i>in vitro</i> de VHCc124	30
2.1.4. Dicroísmo Circular	30

2.1.5. Espectroscopía de fluorescencia	31
2.2. DDX3	31
2.2.1. Clonaje y expresión de DDX3	31
2.2.2. Clonaje y expresión de DDX3 _{pf}	31
2.2.3. Clonaje y expresión de DDX3 _{pdv}	32
2.2.4. Clonaje y expresión de DDX3 _{hel}	33
2.2.5. Purificación de DDX3	33
2.2.6. Purificación de DDX3 _{pf}	33
2.2.7. Purificación de DDX3 _{pdv}	34
2.2.8. Purificación de DDX3 _{hel}	34
2.2.9. Dicroísmo Circular	35
2.2.10. Proteólisis limitada de DDX3	35
2.2.11. Ensayo de actividad helicasa dependiente de la hidrólisis de ATP	36
2.2.12. Dispersión dinámica de luz	36
2.2.13. Digestión proteolítica con la proteasa TEV	37
2.2.14. Cristalización de DDX3 _{hel}	37
2.2.15. Recogida y reducción de datos	37
2.2.16. Determinación y refinado de la estructura	38
2.3. CD5	38
2.3.1. Clonaje de CD5DIII	38
2.3.2. Expresión de CD5DIII	39
2.3.3. Purificación del anticuerpo monoclonal	39
2.3.4. Purificación de CD5DIII	40
2.3.5. Espectrometría de masas	40
2.3.6. Resonancia magnética nuclear	41
2.3.7. Cristalización de CD5DIII	41
2.3.8. Ensayo de desplazamiento de banda con átomos pesados	42

2.3.9. Recogida y reducción de datos	42
2.3.10. Determinación y refinado de la estructura	43
<u>Resultados</u>	45
3.1. VHCc124	47
3.1.1 Expresión y purificación de VHCc124	47
3.1.2. Estudios biofísicos de VHCc124	49
3.1.2.1 VHCc124 no puede renaturalizarse tras su resuspensión en 8M urea	49
3.1.2.2. VHCc124 presenta estructura cuando se purifica de forma nativa	49
3.2. DDX3	52
3.2.1. Determinación de dominios de DDX3	52
3.2.2. Expresión y purificación de DDX3	54
3.2.3. Estudios biofísicos y bioquímicos de DDX3	55
3.2.3.1. DDX3 está plegada y presenta un estado homogéneo de oligomerización	55
3.2.3.2. DDX3 posee actividad helicasa asociada a la hidrólisis de ATP	56
3.2.4. Determinación mediante proteólisis limitada de los dominios flexibles de DDX3	58
3.2.5. Expresión y purificación de DDX3 _{pf}	60
3.2.6. Estudios biofísicos y bioquímicos de DDX3 _{pf}	61
3.1.6.1. DDX3 _{pf} está plegada, presenta un estado homogéneo de oligomerización y conserva la actividad helicasa asociada a la hidrólisis de ATP	61
3.2.7. Expresión y purificación de DDX3 _{pdr}	63
3.2.8. Estudios biofísicos y bioquímicos de DDX3 _{pdr}	65
3.2.8.1. DDX3 _{pdr} está plegada, presenta un estado homogéneo de oligomerización y conserva la actividad helicasa asociada a la hidrólisis de ATP	65
3.2.9. Expresión y purificación de DDX3 _{hel}	67

3.2.10. Estudios biofísicos de DDX3 _{hel}	70
3.2.10.1. DDX3 _{hel} está plegada y presenta un estado homogéneo de oligomerización	70
3.2.11. Cristalización de DDX3 _{hel}	71
3.2.12. Resolución de la estructura de DDX3 _{hel}	73
3.2.13. Descripción general de la estructura de DDX3 _{hel}	74
3.3. CD5	80
3.3.1. Expresión y purificación de CD5DIII	80
3.3.2. Análisis biofísicos de CD5DIII	81
3.3.2.1. CD5DIII mantiene intactos los puentes disulfuro descritos para este tipo de dominios SRCR	81
3.3.2.2. CD5DIII está plegado correctamente	84
3.3.3. Cristalización de CD5DIII	85
3.3.4. Resolución de la estructura de CD5DIII	86
3.3.5. Descripción general de la estructura de CD5DIII	90
3.3.6. Comparación entre los dominios SRCR de los grupos A y B	93
3.3.7. Orientación de la región extracelular de CD5 con respecto a la membrana plasmática	96
<u>Discusión</u>	99
4.1. La proteína C como posible diana a la hora de desarrollar una terapia efectiva contra la enfermedad	102
4.2. Intentos de cristalización de NLPs	102
4.2.1. ¿Es posible la renaturalización de la proteína C una vez desnaturalizada con 8M urea?	102
4.2.2. La purificación de VHCc124 en su forma nativa supone una mejor alternativa para la obtención de NLPs	103
4.3. ¿Cómo se puede avanzar en el desarrollo del trabajo con VHCc124?	104
4.4. DDX3 como diana a la hora de estudiar la hepatitis C	104

4.5. Estrategia de cristalización de DDX3	104
4.6. Los extremos amino y carboxilo terminal de la proteína son los responsables de la falta de solubilidad	105
4.7. El dominio helicasa de DDX3 es la parte más soluble de la proteína mientras que el dominio ATPasa es insoluble	106
4.8. La alta homología de DDX3 _{hel} con Vasa permite realizar un modelo de los dominios ATPasa y helicasa de DDX3	107
4.9. Hacia la resolución de DDX3	107
4.10. Estrategia de cristalización de CD5	109
4.11. La señal anómala de los azufres de CD5DIII es suficiente para resolver el problema de las fases	109
4.12. CD5DIII sirve de base para describir un modelo de la interacción entre CD6 y ALCAM	110
4.13. ¿Cómo se puede avanzar en el desarrollo del trabajo con CD5?	111
<u>Conclusiones</u>	113
<u>Bibliografía</u>	117
<u>Apéndice</u>	131

Índice de figuras

Introducción

Figura 1.1. Esquema simplificado de la progresión de la hepatitis C	4
Figura 1.2. Representación esquemática del ARN que codifica la poliproteína precursora del VHC	6
Figura 1.3. Estructuras cristalográficas de tres proteínas del VHC	7
Figura 1.4. Estructura cristalográfica del LEL de CD81	8
Figura 1.5. Representación esquemática de la replicación de un virus de ARN positivo adaptado a las proteínas del VHC	9
Figura 1.6. Representación esquemática de DDX3	13
Figura 1.7. Alineamiento de las secuencias de proteínas representativas en un amplio espectro de organismos	15
Figura 1.8. Representación esquemática de los miembros de la superfamilia SRCR	18
Figura 1.9. Representación esquemática del receptor CD5	19

Resultados

Figura 3.1. Expresión y purificación de VHCc124	48
Figura 3.2. Espectros de DC de la proteína VHCc124 purificada de forma desnaturalizada y en estado nativo	50
Figura 3.3. Espectros de Fluorescencia de VHCc124	51
Figura 3.4. Esquema de las diferentes construcciones de DDX3 diseñadas en este estudio	53
Figura 3.5. Expresión y purificación de DDX3	54
Figura 3.6. Análisis de dicroísmo circular de DDX3	55
Figura 3.7. Análisis de DLS de DDX3	56
Figura 3.8. Esquema de la reacción llevada a cabo durante el ensayo de la actividad ATPasa de DDX3	57

Figura 3.9. Actividad helicasa de DDX3 dependiente de la hidrólisis de ATP	58
Figura 3.10. Ensayo de proteólisis controlada de DDX3	59
Figura 3.11. Expresión y purificación de DDX3 _{pf}	60
Figura 3.12. Análisis de dicroísmo circular de DDX3 _{pf}	61
Figura 3.13. Análisis de DLS de DDX3 _{pf}	62
Figura 3.14. Actividad helicasa de DDX3 _{pf} dependiente de la hidrólisis de ATP	62
Figura 3.15. Expresión y purificación de DDX3 _{pd}	64
Figura 3.16. Análisis de dicroísmo circular de DDX3 _{pd}	65
Figura 3.17. Análisis de DLS de DDX3 _{pd}	66
Figura 3.18. Actividad helicasa de DDX3 _{pd} dependiente de la hidrólisis de ATP	67
Figura 3.19. Expresión y ensayo de purificación de DDX3 _{ATP} y DDX3 _{hel}	68
Figura 3.20. Expresión y purificación de DDX3 _{hel}	69
Figura 3.21. Análisis de dicroísmo circular de DDX3 _{hel} antes (a) y después (b) de congelar la muestra a 77K	70
Figura 3.22. Análisis de DLS de DDX3 _{hel}	71
Figura 3.23. Proceso de cristalización de DDX3 _{hel} e imagen de difracción	72
Figura 3.24. Diagrama de lazos de la proteína DDX3 _{hel}	75
Figura 3.25. Superposición de DDX3 _{hel} y el dominio helicasa de la proteína Vasa	76
Figura 3.26. Potencial electrostático de DDX3 _{hel} sobre el modelo obtenido	77
Figura 3.27. Diagrama de lazos de los dominios ATPasa y helicasa de DDX3 con los motivos conservados y el ARN en el punto de unión	79
Figura 3.28. Mapa 2fofc a 1σ de la estructura DDX3 _{hel}	77
Figura 3.29. Expresión y purificación de CD5DIII	81

Figura 3.30. Espectro de los fragmentos ESI-TRAP mostrando radios m/z monoisotrópicos correspondientes a los dobletes de péptidos trípticos que tenían los puentes disulfuro	82
Figura 3.31. Espectro de MALDI-TOF de la digestión tríptica de CD5DIII	83
Figura 3.32. Espectro de RMN de CD5DIII	84
Figura 3.33. Cristales de CD5DIII	85
Figura 3.34. Patrón de difracción de cristales tetragonales de CD5DIII	85
Figura 3.35. Ensayo de desplazamiento de CD5DIII tras la unión a diferentes átomos pesados	86
Figura 3.36. Análisis probabilística de la Normal de las diferencias anómalas observadas en base a los datos de difracción obtenidos a 1.8Å de longitud de onda	88
Figura 3.37. Diagrama de lazos de CD5DIII	91
Figura 3.38. Vista en estéreo de la estructura de CD5DIII	91
Figura 3.39. Mapa 2FoFc a 1σ de la densidad electrónica de una región de CD5DIII	92
Figura 3.40. Mapa de potencial electrostático de CD5DIII	93
Figura 3.41. Alineamiento de CD5DIII comparado con los miembros del grupo A M2bp y hepsina	94
Figura 3.42. Comparación de las estructuras de los dominios SRCR del grupo A y B	95
Figura 3.43. Modelado del receptor entero de CD5	97

Discusión

Figura 4.1. Modelo de la interacción CD6-ALCAM/CD166	111
--	-----

Índice de tablas

Introducción

Tabla 1.1. Datos por región sobre la incidencia total y porcentual del VHC	5
---	----------

Resultados

Tabla 3.1. Resumen de la actividad helicasa de DDX3, DDX3 _{pf} y DDX3 _{pd}	67
Tabla 3.2. Datos experimentales de las construcciones realizadas a partir de DDX3	73
Tabla 3.3. Estadísticas de recolección de datos y refinado	74
Tabla 3.4 Estadísticas de recolección de datos y refinado	89



INTRODUCCIÓN

1.1. La enfermedad

La hepatitis C es una enfermedad causada por un virus de la familia *Flaviviridae* denominado virus de la hepatitis C (VHC) que afecta en el mundo a unos 170 millones de personas de acuerdo a cifras de la organización mundial de la salud (www.who.int). Esta cifra representa el 3% de la población mundial y cada año se ve aumentada en 3-4 millones de personas ([tabla 1.1](#)). La cifra de personas contagiadas por VHC varía según las diferentes regiones y puede ir desde un 1 o un 2 % en países como España (Calleja 2001), Francia (Dubois 1997) o Estados Unidos (Williams 1999) hasta cifras tan elevadas como el 12 o el 20% en algunas zonas de Italia (Guadagnino 1997) o Egipto (Frank 2000).

La hepatitis C es una enfermedad principalmente de transmisión sanguínea. Se estima que el uso de drogas intravenosas es la principal fuente de contagio causando aproximadamente un 60% de los casos en Estados Unidos y Francia (Williams 1999). Otros factores de riesgo para la infección son las transfusiones no analizadas de donantes de sangre y procedimientos que requieren una penetración de la piel, como inyecciones no estériles, tatuajes y otras formas de decoración corporal (Conrad 2006). Aunque la hepatitis C raramente se transmite sexualmente (Cohen 1999), parece ser que esta causa, junto con transmisiones del tipo perinatal, por hemodiálisis o transmisiones por causas ocupacionales son las responsables del 10% de las infecciones. Más intrigante aún es el hecho de que en un 20-40% de los casos no hay una fuente reconocible de infección (Karmochkine 2006).

La infección por VHC da lugar a una hepatitis asintomática después de un período de incubación de entre 3 y 12 semanas. Sólo alrededor de un 30-40% de los pacientes presentan síntomas (www.who.int) y sólo en un porcentaje ínfimo de esos pacientes, que generalmente presentan otros factores agravantes, la enfermedad resulta fulminante (Moradpour 2001). Uno de los datos clínicamente más relevantes es que la mayor parte de los pacientes infectados (55-85%) no pueden eliminar el virus y se convierten así en portadores crónicos que mantienen una persistente presencia del virus en el suero sanguíneo (Hoffnagle 2002). El intervalo existente entre la infección y el diagnóstico es tremendamente largo y varía entre 10 y 18 años (Tong 1995). Esto es debido principalmente a que una gran parte de las personas crónicamente infectadas con VHC tienen una hepatitis C leve que, si no se combina con otros factores agravantes, permanece estable durante décadas con una progresión muy lenta de la fibrosis. De cualquier modo, la fibrosis que desarrollan estos pacientes, aunque lenta, puede desembocar en cirrosis en un 20% de los casos y posteriormente desarrollarse hasta convertirse en hepatocarcinoma con una incidencia anual del 1-4% (Diedrich 2006) siendo así la infección por VHC la principal causa de transplante hepático. En la [figura 1.1](#) se puede apreciar un esquema de la historia natural de la infección por hepatitis C.

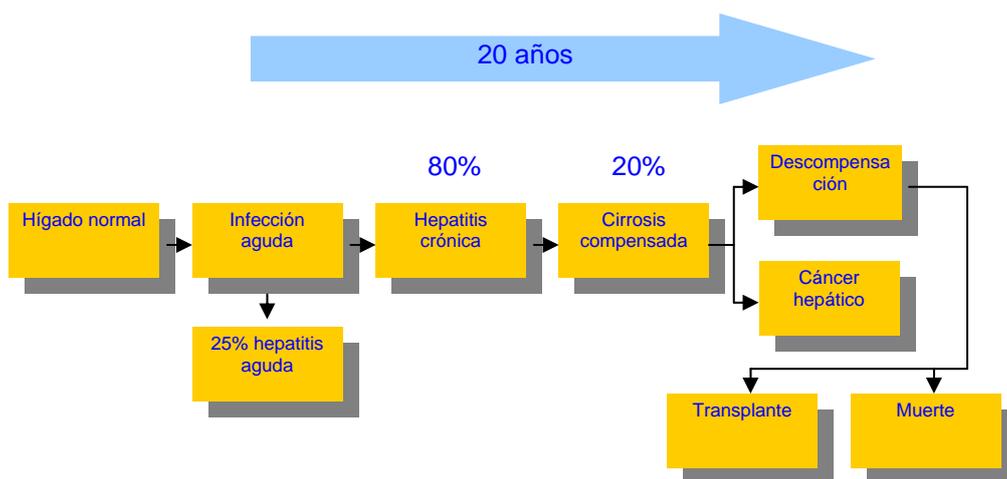


Figura 1.1. Esquema simplificado de la progresión de la hepatitis C. La enfermedad se vuelve crónica en un 80% de los casos y un 20% desarrolla cirrosis que en determinados casos puede llevar al cáncer de hígado o a la necesidad de un transplante para evitar el fatal desenlace. El esquema ha sido tomado de (Soza 2006)

Tabla 1.1. Datos por región sobre la incidencia total y porcentual del VHC

REGIÓN	POBLACIÓN TOTAL (millones)	POBLACIÓN INFECTADA (millones)	PORCENTAJES POBLACIONALES	PAISES CON DATOS NO DISPONIBLES
AFRICA	602	31.9	5.3	12
AMERICA	785	13.1	1.7	7
MEDITERRANEO ESTE	466	21.3	4.6	7
EUROPA	858	8.9	1.03	19
SURESTE ASIATICO	1500	32.3	2.15	3
PACIFICO OCCIDENTAL	1600	62.2	3.9	11
TOTAL	5811	169.7	3.1	57

Datos basados en Weekly Epidemiological Record. N°49, 10 Diciembre 1999, WHO.

1.2. El virus

El virus de la hepatitis C fue identificado por primera vez en 1989 (Kuo 1989), pertenece a la familia *Flaviviridae*, que incluye entre otros al virus de la fiebre amarilla y a los pestivirus animales y ha sido clasificado dentro del género *Hepacivirus* (Van Regenmortel 2000). Es un virus esférico de unos 50nm de diámetro con una envuelta glicoproteica que contiene lípidos. Tiene un ARN positivo de cadena sencilla de unos 9.6Kb. Su genoma incluye una región no codificante en el extremo 5' (5'UTR) esencial para la traducción, que actúa como región interna de entrada del ribosoma (IRES). El IRES dirige la traducción desde un único origen de lectura abierta (ORF) codificando una poliproteína de 3010 amino ácidos (Lukavsky 2000). Ésta, sufre cortes post-transduccionales por proteasas tanto de la célula huésped como del propio virus dando lugar a las proteínas estructurales y no-estructurales maduras (Grakoui 1993; Bartenschlager 1994; Huarte 2004; Soza 2006) (Figura 1.2).

Al igual que en otros miembros de la familia *Flaviviridae*, el amino terminal de la poliproteína se compone de las proteínas estructurales, conformándose el resto de la poliproteína con proteínas no-estructurales (Hijikata 1991; Ralston 1993). Las proteínas estructurales son la proteína del *core* que forma la nucleocápsida viral (proteína C) (Takahishi 1992; Yasui 1998), las glicoproteínas de la envuelta E1 y E2 (Hijikata 1991; Ralston 1993) y la proteína p7, un canal iónico cuya función no está del todo definida (Soza 2006). Las proteínas no-estructurales incluyen la autoproteasa NS2-3 y la serin-proteasa NS3, que son esenciales para el procesamiento de la poliproteína precursora, una ARN helicasa localizada en la región del carboxilo terminal de NS3. Además, incluyen también el polipéptido NS4A, que es un cofactor de NS3, las proteínas NS4B y NS5A, cuya función no está todavía descrita, y NS5B, una ARN polimerasa dependiente de ARN (Yao 1999).

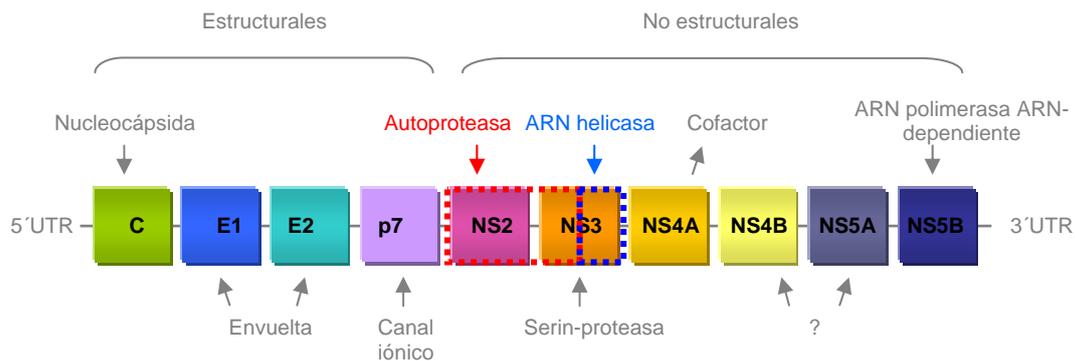


Figura 1.2. Representación esquemática del ARN que codifica la poliproteína precursora del VHC. Cada caja representa una proteína. La función se marca en gris excepto en el caso de la autoproteasa N2-3 y la ARN helicasa correspondiente al C-terminal de NS3 que se marcan en rojo y azul respectivamente. Ambas proteínas se señalan además por una caja punteada.

En la [figura 1.3](#) se pueden apreciar la estructura de algunas proteínas del VHC que ya han sido resueltas mediante la cristalografía de rayos X.

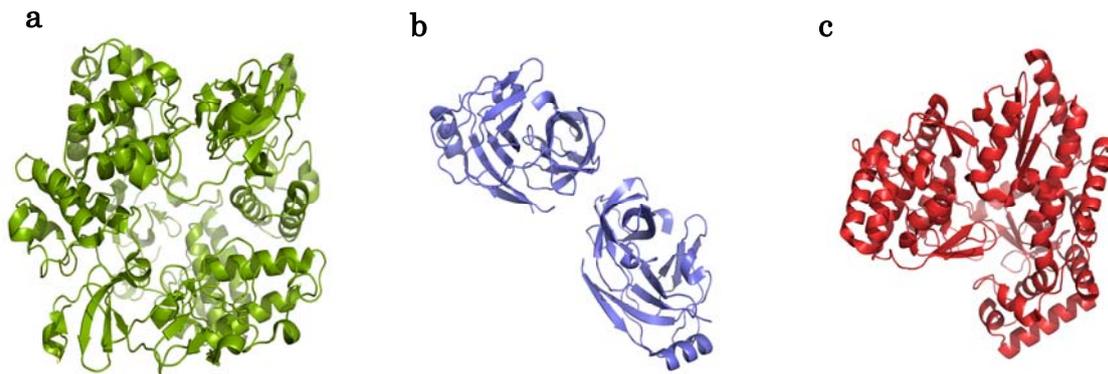


Figura 1.3. Estructuras cristalográficas de tres proteínas del VHC. Las estructuras corresponden a proteínas no-estructurales del virus; a) corresponde al dominio catalítico de NS2 (PDB ID: 2HD0), b) corresponde a la proteasa NS3 (PDB ID: 1W3C) y c) a la polimerasa NS5B (PDB ID: 1YV2).

1.3. Entrada del virus en la célula

Desde su descubrimiento, el principal cuello de botella para desentrañar el ciclo vital del VHC ha sido la imposibilidad de desarrollar un sistema fiable de cultivos celulares para la propagación del virus, sumada a la ausencia de un modelo animal satisfactorio. A pesar de todo, gracias a aproximaciones diferentes como la vacunación de chimpancés, estudios estructurales de productos aislados, estudios de unión con proteínas recombinantes de la envuelta y ensayos de infectividad, se ha conseguido revelar alguno de los mecanismos de entrada del virus en la célula. Sin embargo, los resultados de estos estudios no siempre han sido consistentes señalando así un mecanismo complejo de entrada del virus en la célula que implicaría a más de una proteína del huésped (Diedrich 2006).

Dentro de los estudios que se centran en la entrada del virus al interior de la célula, se ha observado que las glicoproteínas E1 y E2 del VHC confieren a las partículas retrovirales especificidad por las células hepáticas (Bartosch 2003; Drummer 2003; Hsu 2003). Formas truncadas solubles de E2 se unen a células humanas y han sido usadas para identificar interacciones con varias moléculas de superficie como CD81 (Pileri 1998), SR-BI (Scavenger Receptor Class B type I) (Scarselli 2002), DC-SIGN (Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin) (Gardner 2003;

Lozach 2003; Pohlmann 2003) o glicosaminoglicanos (Germi 2002; Germi 2002; Barth 2003).

De estas moléculas, CD81 parece ser indispensable para la entrada del virus en la célula huésped. CD81 es una proteína transmembrana que pertenece a la superfamilia de las tetraspaninas, proteínas que se caracterizan estructuralmente por la presencia de cuatro regiones hidrofóbicas transmembrana y dos dominios extracelulares, conocidos como lazos extracelulares grande (figura 1.4) y pequeño (LEL y SEL respectivamente). Está presente en la mayor parte de los tejidos humanos con la excepción de las plaquetas y los glóbulos rojos y se le han atribuido numerosas funciones incluyendo la adhesión celular, funciones de movilidad, metástasis y activación celular (Levy 1998).



Figura 1.4. Estructura cristalográfica del dominio LEL de CD81. Este dominio extracelular del receptor ha sido resuelto y presenta una estructura dimérica basada en la asociación de residuos conservados de tetraspanina (PDB ID: 1IV5).

La interacción entre la proteína E2 del VHC y CD81 ha sido ampliamente estudiada y los sitios de unión han sido descritos (Flint 1999; Higginbottom 2000; Yagnik 2000). De los ensayos de infectividad con pseudo-partículas de VHC (ppVHC) y partículas virales derivadas de cultivos celulares (ccVHC) se desprende que CD81 es necesaria aunque no suficiente por sí sola para la

entrada en la célula. Así, no se ha identificado ninguna línea celular CD81-negativa que haya podido ser significativamente infectada con ppVHC. Esto no significa sin embargo que todas las líneas CD81-positivas pueden ser infectadas (Bartosch 2003; Hsu 2003; McKeating 2004; Zhang 2004; Lavillette 2005). La expresión de CD81 también es requisito indispensable para la infección con ccVHC (Lindenbach 2005; Wakita 2005; Zhong 2005). Sin embargo, de los datos publicados se desprende que CD81 necesita un cofactor endocítico para promover la entrada en la célula del VHC ya que a pesar de que permite la unión de partículas virales no es, por sí sola, capaz de facilitar la entrada del virus en la célula (Tan 2003). En la [figura 1.5](#) se representa un esquema del ciclo de replicación de los virus de ARN positivos que sirve de modelo para ilustrar el ciclo del VHC dentro de la célula.

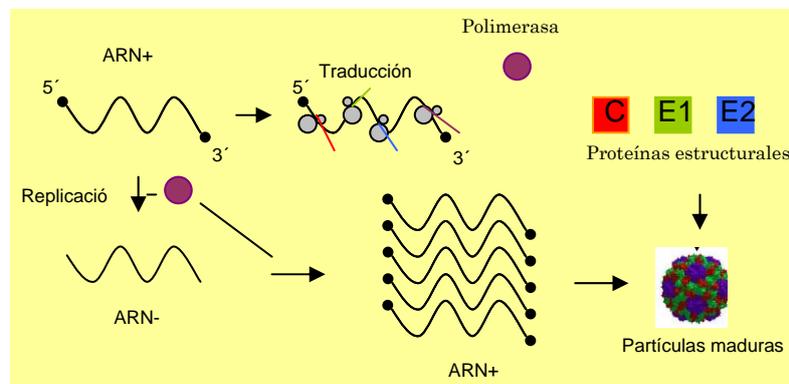


Figura 1.5. Representación esquemática de la replicación de un virus de ARN positivo adaptado a las proteínas del VHC. Se han representado las tres proteínas estructurales que formarían la envuelta y nucleocápsida del virus así como la proteína NS5B que actúa de RNA polimerasa. Esquema basado en el trabajo de Antonio Mas López del Centro Regional de Investigaciones Biomédicas, Facultad de Medicina-UCLM.

1.4. La proteína C y su interacción con diferentes dianas celulares.

Una vez que las células son infectadas por el virus, las proteínas virales interactúan con otras proteínas de la célula huésped modulando su fisiología. En concreto la proteína C es responsable de diversas interacciones con la maquinaria celular durante distintas etapas de la infección. Por ejemplo, la proteína C, no solo desarrolla una función estructural, sino que participa activamente en la transformación celular que origina hepatocarcinoma en algunos pacientes infectados por VHC (Jin 2000).

La proteína C es una de las más conservadas e inmunogénicas de todos los componentes proteicos del VHC (Lorenzo 2001). Está formada por los aminoácidos 1-191 del extremo amino terminal de la poliproteína del virus. Tras su procesamiento en el retículo endoplásmico se pierde la parte hidrófoba del carboxilo terminal, dando lugar a la forma madura de 179 aminoácidos. Esta proteína se ensambla dando lugar a nucleocápsidas de 25-30nm de diámetro (Majeau, Gagne et al. 2004). Los mecanismos que controlan el ensamblaje de la nucleocápsida así como su función se desconocen. Esto es debido principalmente al cuello de botella antes mencionado (sección 1.3). Estas carencias han llevado a buscar sistemas de expresión *in vitro* de partículas similares a nucleocápsida (NLPs) que permitan conocer detalles sobre la estructura del virus, su desarrollo y su modo de actuación. Se ha descrito la formación de NLPs cuando la proteína C es expresada en *E. coli* (Kunkel 2001; Lorenzo 2001), células de insecto (Baumert 1998), levaduras (Falcon 1999) y sistemas de mamíferos (Shimizu 1996; Blanchard 2002). A pesar de los intentos desarrollados con estos diferentes sistemas, el éxito a la hora de resolver la estructura del VHC no ha sido fructífero ni por microscopía electrónica ni por cristalografía de rayos X.

1.4.1. Interacción con factores nucleares de transcripción

La proteína C juega un importante papel durante la infección con VHC en el interior de la célula. Una variante truncada de la proteína C (1-174) deja al descubierto la presencia de una señal de localización nuclear en esta proteína, la cual consecuentemente puede ser transportada al núcleo, donde interacciona con distintos factores nucleares alterando la expresión génica. Una de estas proteínas es un factor de transcripción denominado LZIP que pertenece a la familia de factores de transcripción con dominios bzip que llevan a cabo su unión a ADN como dímeros. LZIP está implicado en mecanismos de supresión tumoral y la proteína C impide su dimerización y lo secuestra en el citoplasma, impidiendo su localización nuclear y potenciando la transformación celular. La interacción de este factor de transcripción con la proteína C reflejaría el mecanismo que conduce al desarrollo de hepatocarcinoma en las fases terminales de la enfermedad (Jin, Wang et al. 2000).

1.4.2. Interacción con el sistema de señalización de apoptosis

Otra diana celular de la proteína C es el TNFR-1 (Tumor Necrosis Factor Receptor-1) interaccionando de este modo con el sistema de señalización de apoptosis. Los receptores del tipo TNFR representan una familia de proteínas de gran importancia por su función de control en los procesos de muerte celular regulada o apoptosis mediante la transmisión de sus señales vía interacciones proteína-proteína, las cuales convergen bien en una señal de muerte o de supervivencia. Las interacciones de la proteína C con estos receptores parecen ser parte de la estrategia programada del virus durante los distintos estadios de la infección (Inghirami 1991).

1.4.3. Interacción con el sistema de traducción celular

Tal y como se ha descrito en la literatura (Mamiya 1999; Owsianka 1999), una ARN helicasa de la familia de las proteínas DEAD box, denominada DDX3 cuya función es desplegar el ARNm para iniciar la traducción interacciona con la proteína C del VHC. La proteína C interacciona con DDX3 y esta asociación de las dos proteínas forma un complejo que supone la supresión de la actividad helicasa y la consiguiente reducción de traducción de ARNm celular. El ARN del virus puede ser desenrollado por su propia ARN helicasa.

Estos datos revelan algunos de los mecanismos mediante los cuales el virus puede inhibir la traducción de proteínas en las células infectadas sin afectar la suya propia primando la producción de las proteínas víricas frente a las celulares.

1.4.3.1. Descripción de las ARN helicasas

Las ARN helicasas regulan prácticamente todos los procesos en los que una manipulación de ARN es requerida desde la transcripción y la traducción, pasando por la exportación del núcleo, o la degradación del ARN (de la Cruz 1999). Estas proteínas catalizan la modificación de estructuras complejas de ARN, tales como la ruptura de la estructura secundaria, la separación de hebras y la disociación de las interacciones ARN-proteína (Schwer 2001). De esta manera, las helicasas son unas de las proteínas más numerosas con representantes en los tres dominios (de la Cruz 1999), y a pesar de que no son esenciales en todos los casos, sí que juegan un papel muy importante en el metabolismo del ARN en general (Rocak 2004).

1.4.3.2. Detalles estructurales de DDX3

DDX3, también llamada CAP-Rf (You 1999) o DBX (Lahn 1997), es una proteína perteneciente a la familia de proteínas DEAD box. Está compuesta por los nueve motivos característicos de esta familia incluyendo el motivo Asp, Glu, Ala, Asp (DEAD en el código de una letra) que da nombre a la familia. Tiene dos regiones diferenciadas que contienen todos los motivos conservados. La primera región incluye el motivo Q (178-206) (Cordin 2004) y un dominio de unión e hidrólisis de ATP (209-401) que incluye los motivos I, II (Walker A y B respectivamente), Ia, Ib y el motivo III, que se ha propuesto como responsable de cambios conformacionales que permiten la hidrólisis del ATP previa unión del ARN. La segunda región contiene el dominio helicasa (412-573) e incluye los motivos IV y V que probablemente están implicados en la unión de ARN junto con los motivos Ia y Ib de la primera región. También presente en esta segunda región está el motivo VI que se piensa está implicado en la unión de ATP (Rocak 2004). (Figura 1.6). Recientemente se ha demostrado que la actividad helicasa de DDX3 es dependiente de la hidrólisis de ATP y viceversa (Yedavalli 2004) confirmando así esta interacción funcional entre ambas regiones.

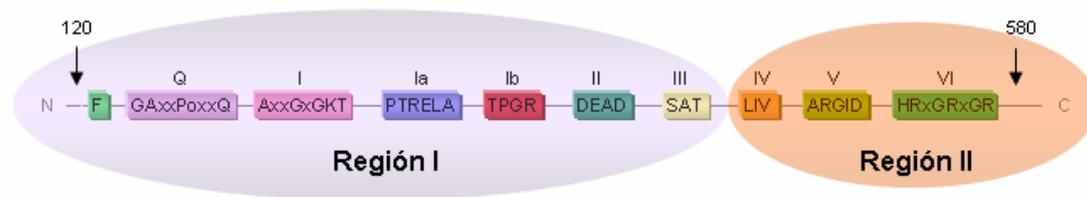


Figura 1.6. Representación esquemática de DDX3. Los motivos característicos de la familia de proteínas DEAD box así como sus dos regiones principales están coloreados en el esquema. Se señalan también los aminoácidos que están representados en la [figura 1.7](#).

Junto a estos datos estructurales es de destacar que se han resuelto por cristalografía de rayos X varias proteínas homólogas en numerosas especies con alta similitud en su estructura primaria. Un alineamiento de secuencia basado en la estructura de las proteínas revela los motivos conservados entre las diferentes proteínas homólogas de las distintas especies ([figura 1.7](#)).

1.4.3.3. Función de DDX3

Los datos acerca de la función específica que DDX3 desarrolla en el interior de la célula no son todavía concluyentes. Se ha descrito cómo esta enzima está implicada en diversos procesos. Uno de ellos es corte y empalme de ARNm mediante la colocación con factores de corte y empalme (Zhou 2002). También se ha publicado su relación con el transporte de ARNm ya que está presente en gránulos de transporte de ARN (Kanai 2004). Recientemente se ha descrito también cómo DDX3 actúa como factor antitumoral mediante la sobreexpresión de p21waf1/cip1 (Chao 2006). Además de ser importante en procesos proliferativos, DDX3 está directamente implicada en enfermedades de gran relevancia para la salud pública como el síndrome de inmunodeficiencia adquirida (SIDA). Durante la infección por VIH, DDX3 se sobreexpresa (Krishnan 2004) y es requerido para la exportación del núcleo celular de HIV-1 Rev-RRE (Yedavalli 2004).

Junto a todos estos datos funcionales está por supuesto la ya mencionada interacción de DDX3 con la proteína C del VHC. Durante la infección, la expresión de DDX3 se reduce (Chang 2006) y la proteína C del VHC

interacciona directamente con DDX3 suprimiendo su capacidad de desenrollar ARN (Mamiya 1999; Owsianka 1999; You 1999).

1.5. Interacciones del VHC a nivel del sistema inmune

Además de la relación directa del VHC con las células hepáticas y su interacción con diversas proteínas en el interior de la célula, el virus desarrolla también un linfotropismo que puede ser en parte el responsable de las múltiples manifestaciones extra-hepáticas relacionadas con el sistema inmune que se dan durante la infección con VHC (Zuckerman 2003). Además de la bien establecida asociación entre la infección con VHC y la crioglobulinemia mixta (Angello 1992; Misiani 1992; Marcellin 1993), el VHC está implicado en numerosos desordenes de carácter autoinmune como por ejemplo el síndrome Sjögren, la hepatitis autoinmune de tipo 2 y algunas neuropatías periféricas relacionadas con el sistema inmune (Cacoub 1999; Heckmann 1999). Es característica también la presencia del factor reumatoide en el suero de los pacientes infectados con VHC (Clifford 1995; Hadziyannis 1997).

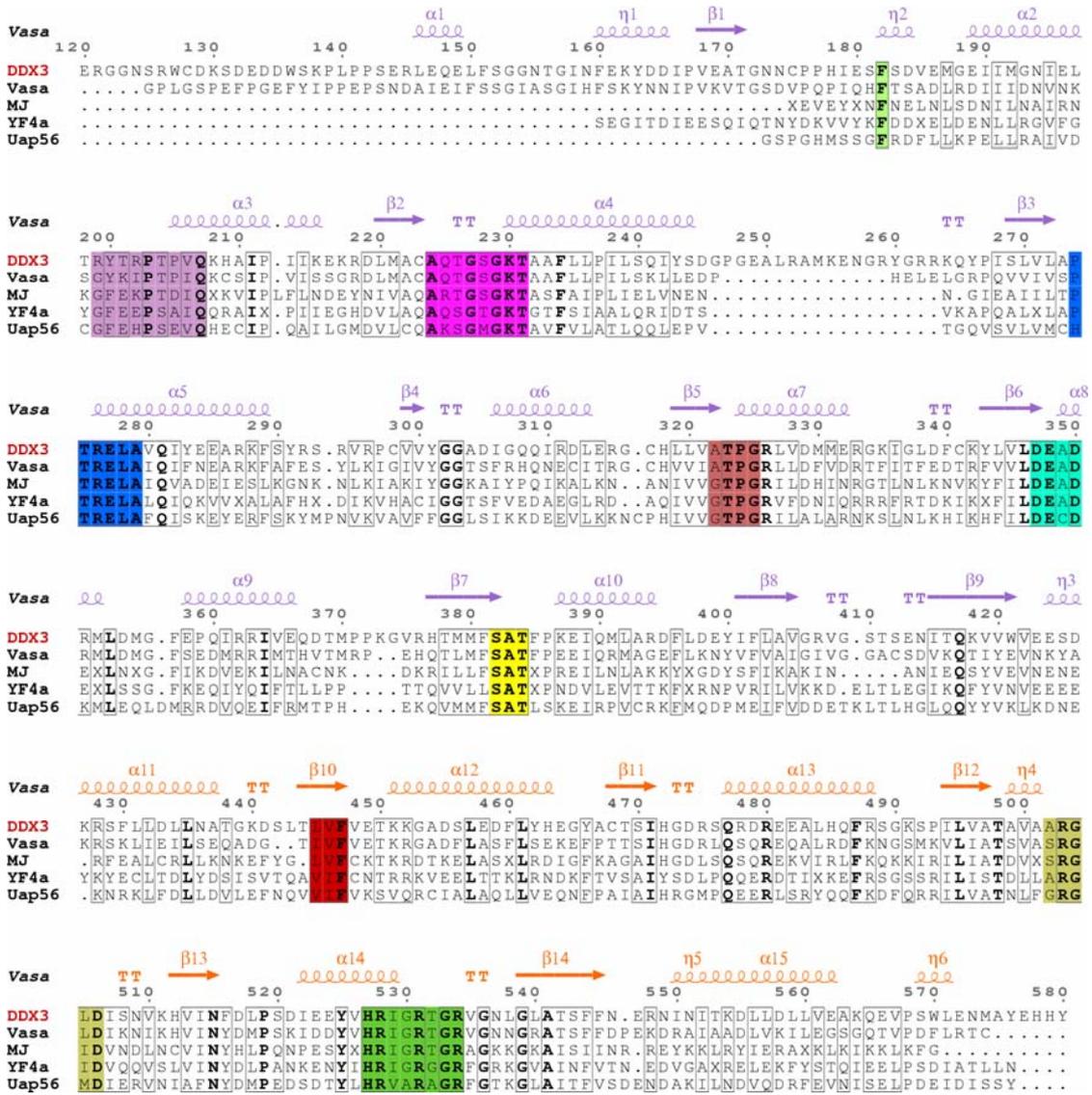


Figura 1.7. Alineamiento de las secuencias de proteínas representativas en un amplio espectro de organismos. Están coloreados los motivos característicos de las proteínas de la familia DEAD. La estructura secundaria de la proteína Vasa de *Drosophila melanogaster* está situada en la parte superior para dar una idea de la conformación secundaria de estas proteínas. Las regiones de este grupo se reflejan en el cambio de color de esta estructura secundaria. En negrita se marcan los residuos conservados. Las cajas grises marcan residuos que aunque diferentes, conservan el carácter del amino ácido. La proteína de *Methanococcus jannaschii* se ha nombrado como MJ para simplificar el esquema.

El mecanismo por el que el virus causa estos síndromes autoinmunes no se comprende del todo. Una posibilidad es que el VHC inhiba la apoptosis de células B o active de manera inespecífica las células B, posiblemente mediante la unión al receptor CD81 expresado en estas células (Pileri 1998; Petracca 2000). Estudios recientes señalan la posibilidad de que el virus interactúe con el receptor CD81 de las células dendríticas alterando su capacidad migratoria y reduciendo así la capacidad de la respuesta inmune (Nattermann 2006).

Todos estos datos apuntan hacia CD81 como principal diana a la hora de elaborar drogas o diseñar terapias contra la infección. Sin embargo, y a pesar de estar descritas las regiones de interacción entre E2 y CD81 (Flint 1999; Higginbottom 2000; Yagnik 2000), un estudio reciente (Flint 2006) muestra la dificultad actual para identificar posiciones específicas de amino ácidos en CD81 que sean críticas para la unión con ppVHC o ccVHC. Debido a esta plasticidad, el estudio apunta a que el diseño de pequeñas moléculas que inhiban la interacción entre CD81 y E2 puede resultar ineficaz a la hora de tratar de bloquear la entrada del virus en las células. Por otro lado, se ha descrito en la literatura la existencia de una relación directa entre la expresión del receptor CD5 y la sobreexpresión de CD81 que puede abrir una interesante vía de estudio de este receptor.

En las células B, CD81 forma parte de un complejo de señalización que incluye a CD19 y CD21 (Levy 1998). Un entrecruzamiento con estos complejos mediante el uso de anticuerpos contra CD81 o CD19 disminuye la barrera para la activación y proliferación de células B. Además, está descrito cómo una subpoblación de células B CD5+ está sobreexpresada en pacientes con hepatitis C crónica (Curry 2000). Estas células se caracterizan por la producción de IgM de baja afinidad con presencia de factor reumatoide, surgen en la ontogenia temprana y se considera que representan la unión entre la respuesta inmune innata y la adquirida (Tsuji 1997). La producción de autoanticuerpos acompañada de una elevada frecuencia de artritis reumatoide y del síndrome Sjögren ha implicado a estas células en el desarrollo de enfermedades autoinmunes (Dauphinee 1988), (Plater Z. 1985).

El número de células B CD5+ también se incrementa en pacientes con crioglobulinemia mixta (Pietrogrande 1995) y estas células son productoras de inmunoglobulinas M (IgM) monoclonales y policlonales en los folículos linfáticos hepáticos de pacientes infectados con VHC (Monteverde 1997). Se ha descrito también cómo durante una infección crónica de hepatitis C se produce una sobreexpresión de CD81 en células B periféricas así como una sobreexpansión de la subpoblación de células B CD5+ que se correlaciona con la carga de ARN vírico y se relacionan con el desarrollo de autoinmunidad ligada a VHC (Zuckerman 2002). No sólo esto, sino que se ha demostrado que en pacientes a los que se les ha aplicado un tratamiento de manera exitosa hubo un significativo descenso de las células B CD81 periféricas así como una desaparición de la expansión de células B CD5+ seguida de la desaparición del factor reumatoide y crioglobulinas, así como una disminución en la producción de autoanticuerpos (Zuckerman 2003).

1.5.1. CD5, descripción y función.

1.5.1.1. La superfamilia SRCR (Scavenger Receptor Cysteine Rich)

El dominio SRCR es un módulo proteico antiguo y altamente conservado que define una superfamilia de proteínas bien solubles o de receptores de membrana expresados por células hematopoyéticas y no-hematopoyéticas en estadios embrionarios y/o en adulto (Resnick 1994; Sarrias 2004). La existencia de dos tipos de dominios SRCR permite la división de la superfamilia en dos grupos. Los miembros del grupo A contienen seis cisteínas en sus dominios y están codificados por no menos de dos exones, mientras que los miembros del grupo B contienen ocho cisteínas y están codificadas por un único exón. Esta disposición diferente en la organización de exones e intrones indica que ambos grupos A y B han evolucionado a partir de distintos genes ancestrales. De cualquier modo, en ambos casos todas las cisteínas están implicadas en la formación de puentes disulfuro. Los miembros del grupo A son generalmente proteínas formadas por multidominios mosaicos conteniendo un único dominio SRCR asociado a otros dominios funcionales. Por el contrario, los miembros del grupo B están compuestos en su mayoría por repeticiones en tandem de dominios SRCR (Resnick 1994; Sarrias 2004) ([figura 1.8](#)).

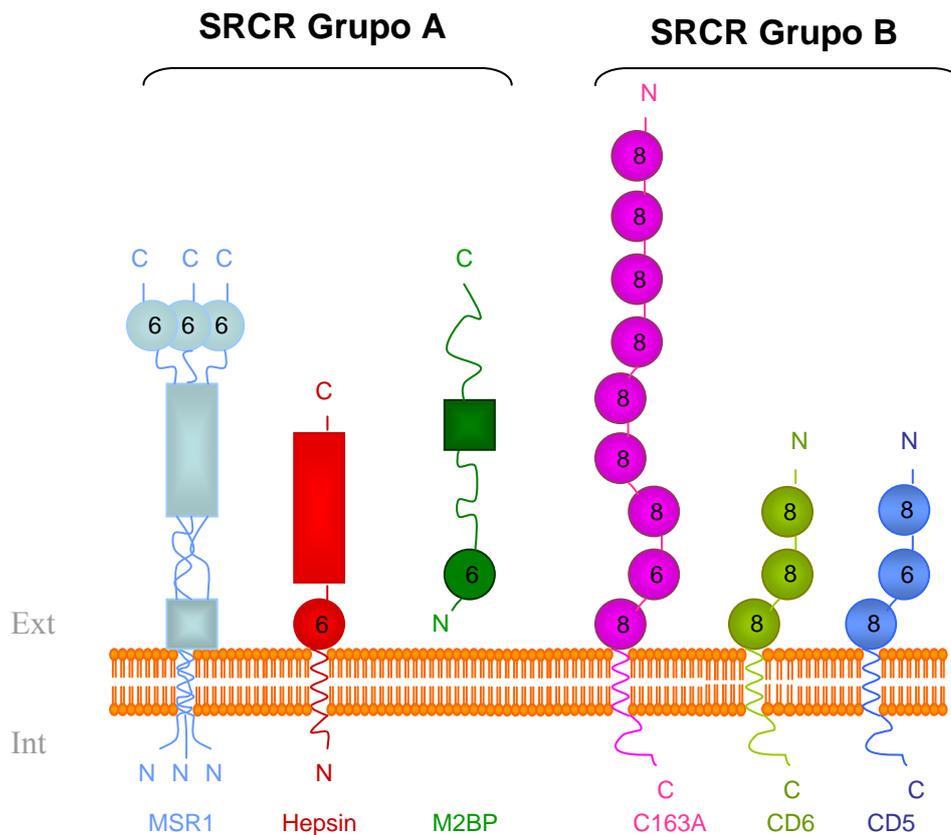


Figura 1.8. Representación esquemática de los miembros de la superfamilia SRCR. Los dominios SRCR están representados por círculos que marcan en su interior el número de cisteínas de cada dominio. Los dominios que no son SRCR se señalan con cuadrados o rectángulos. La membrana plasmática se representa en naranja.

Miembros de estos grupos se encuentran en diferentes especies animales, desde invertebrados primitivos a los vertebrados más evolucionados. A pesar de que no se ha encontrado una función común que defina a los miembros de la superfamilia SRCR numerosas evidencias unidas al alto grado de conservación estructural y filogenético sugieren un papel importante en funciones básicas de homeostasis incluyendo la defensa del sistema inmune innato.

1.5.1.2. Datos estructurales previos de CD5

CD5 es una proteína transmembrana de tipo I de 67 kDa presente en timocitos, linfocitos T y un subgrupo de células B denominado B1a (Hardy 1994). Se encuentra también de forma libre en el suero sanguíneo

seguramente como resultado de un procesamiento proteolítico a raíz de la activación linfocitaria (Calvo 1999). Pertenece a la superfamilia de receptores con dominios extracelulares (Resnick 1994; Aruffo 1997).

Consta de una región citosólica de 94 amino ácidos que carece de actividad catalítica intrínseca, pero que posee sitios potenciales de fosforilación para kinasas intracelulares así como otros motivos estructurales que sugieren un papel de señalización celular (Vila 2000). Diferentes evidencias señalan su capacidad para interactuar con proteínas de señalización (Calvo 1998; Dennehy 1998) y del citoesqueleto (Alberola-Ila 1993; Vila 2001). Esta región citosólica está constitutivamente fosforilada (Calvo 1998; Raman 1998) y se hiperfosforila después de una activación linfocitaria (Burgess 1992; Alberola-Ila 1993). De esta manera, se ha descrito la fosforilación *in vivo* de CD5 por parte de la protein kinasa C (Griffith 1989; Lozano 1990; Vila 2001), la casein kinasa II (Calvo 1998; Raman 1998) y la protein tirosin kinasa (Davies 1992; Raab 1994).

CD5 presenta una región transmembrana de 22 amino ácidos formada por una única hélice y una región extracelular compuesta por tres dominios SRCR del grupo B. El dominio I (CD5DI) y el dominio III (CD5DIII) son prácticamente idénticos y poseen las ocho cisteínas clásicas de los receptores SRCR del grupo B. El dominio II (CD5DII) sin embargo, presenta seis cisteínas al igual que los miembros del grupo A, pero tiene el puente disulfuro C1-C4 característico del grupo B. Esto hace de DII un subtipo único dentro de los dominios del grupo B (figura 1.9).

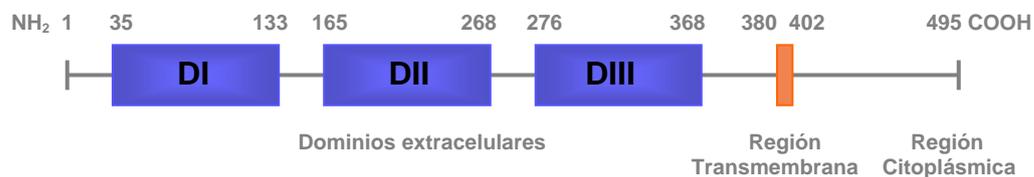


Figura 1.9. Representación esquemática del receptor CD5. Los tres dominios SRCR se muestran en azul señalados como DI, DII y DIII. La región transmembrana está representada en naranja. Se indican también el amino ácido donde empieza y acaba cada región o dominio.

1.5.1.3. Datos funcionales de CD5

En general, ha resultado difícil identificar ligandos de la superfamilia SRCR a pesar de sus importantes funciones biológicas. Algunos autores han propuesto un papel de CD5 en interacciones entre células T y B (Van de Velde 1996; Bikah 1998) dadas las relaciones descritas previamente entre CD5 y distintas proteínas presentes en la superficie de las células B como por ejemplo CD72 (Van de Velde 1991; Luo 1992), gp40-80 (Biancone 1996; Bikah 1998), secuencias de la región base de Ig Vh (Pospisil 1996; Pospisil 2000) y gp150 (Calvo 1999). Sin embargo, la verdadera naturaleza del ligando de CD5 es todavía motivo de controversia y resulta difícil asignar una función clara a este receptor.

1.6. Tratamientos contra la enfermedad

En la actualidad el tratamiento utilizado para combatir la hepatitis C es una combinación de interferón-alfa y ribavirina. El interferón constituye una potente citoquina con acción antiproliferativa, inmunomoduladora y antiviral inhibiendo la síntesis de proteínas virales, induciendo la destrucción del ARN viral y facilitando la eliminación de las células infectadas por activación de las células “natural killer”. La ribavirina es un análogo sintético de nucleósidos, con actividad antiviral *in vitro* contra virus ADN y ARN cuyo mecanismo de acción todavía no está del todo claro. Puede que actúe inhibiendo la síntesis del ácido nucleico viral, inhibiendo la producción de IL-4 e incluso inhibiendo la respuesta pro-inflamatoria desencadenada por la infección viral (Calleja 2001).

Los efectos secundarios del tratamiento son diversos y son la causa principal del incumplimiento de la terapia. En general son dependientes de la dosis y suelen desaparecer al suspender la medicación. La mayoría de los efectos adversos son leves, van desde síntomas pseudo-gripales, insomnio, mareo, depresión, irritabilidad o cambios de humor, hasta náuseas, diarrea, dolor abdominal, erupciones cutáneas, prurito o tos. Estos efectos se presentan sobre todo al iniciar el tratamiento. En algunos casos sin embargo, los efectos secundarios son severos pudiendo ocasionar psicosis, delirio, convulsiones, mielosupresión, retinopatía, anemia hemolítica, depresión mayor e incluso suicidio. Esto hace que alrededor de un 20% de pacientes tenga que abandonar el tratamiento (Calleja 2001). Unidos a estos

efectos, están los elevados costes del tratamiento ya que la relación coste-eficacia y coste-ahorro es muy superior a otros procedimientos aceptados por la comunidad médica como el tratamiento de la hipertensión arterial o el despistaje de cáncer de mama o colon (Bennett 1997).

Todos estos factores relacionados con el actual tratamiento como son la eficacia, el elevado coste y los efectos secundarios hacen necesaria la búsqueda de terapias alternativas o complementarias que permitan abordar la enfermedad con mayores garantías de éxito. De este modo, no sólo es necesario abordar el ataque del virus a las células hepáticas sino también el ataque que produce al sistema inmune. Una comprensión mayor de cómo actúa el virus, no sólo dentro de la célula interaccionando específicamente con proteínas huésped sino también antes de su entrada en la célula nos proporcionará herramientas mejores para seguir avanzando en el camino de la investigación de una terapia más adecuada para paliar o incluso eliminar los efectos y la causa de la enfermedad.



OBJETIVOS

La Hepatitis C es una enfermedad que carece de un tratamiento eficaz. Las terapias actuales se centran en una combinación de interferón y ribavirina. El desarrollo de nuevos tratamientos que puedan ayudar a paliar esta enfermedad necesita una previa comprensión a nivel molecular de las distintas etapas de la infección. En esta tesis doctoral se ha intentado ampliar este conocimiento mediante el estudio estructural del virus de la Hepatitis C y sus interacciones con diferentes proteínas celulares. Esta es una aproximación nueva ya que las terapias que se han intentado desarrollar se han dirigido a actuar sobre otras proteínas del virus como NS3 y NS5B.

La proteína C que forma la nucleocápsida del virus interacciona con varias proteínas (LZIP, TNFR-1, DDX3) implicadas en procesos esenciales para la célula (supresión tumoral, apoptosis, traducción). Estas interacciones reflejan distintos estadios del proceso de infección. Además, el primer caso parece ser responsable del origen de hepatocarcinoma por VHC. Por otro lado, el virus interacciona con el receptor CD81 para entrar en la célula. Este fenómeno está acompañado de una sobreexpresión del receptor CD5 con efectos, todavía no muy claros, en el sistema inmune, pero que parecen favorecer la infección.

En el presente trabajo se propusieron los siguientes objetivos:

1. Producir de forma recombinante NLPs a partir de la proteína C del VHC con el objetivo de cristalizarlas para estudiar en detalle su estructura molecular.
2. Clonar, purificar y cristalizar DDX3, LZIP y TNFR-1 tratando de resolver su estructura atómica mediante difracción de rayos X.
3. Estudiar la interacción de la proteína C con DDX3, LZIP y TNFR-1.
4. Estudiar la región extracelular del receptor CD5 y clonar, purificar y cristalizar al menos uno de los dominios que permita realizar un modelado del receptor entero.

MATERIALES

2.1. Proteína C de VHC

2.1.1 Clonaje y expresión de la proteína C

El gen de la proteína C se clonó a partir del suero de un paciente infectado con el VHC utilizando RT-PCR. Para producir la proteína de 124 aminoácidos (VHCc124) se usó como molde un plásmido pQE30 (Qiagen) que contenía 1-179 aminoácidos de la proteína C. VHCc124 fue amplificada mediante PCR usando como cebadores directo y reverso 5'-GGGAAATCCATATGAGCACGAATCCTAAACCTCAAAGAAAA-3' y 5'-GGAATTCCTCAATCGATGACCTTACCCAGATTGCGCGACCTA-3', respectivamente. El producto de PCR fue digerido usando las enzimas de restricción NdeI y EcoRI y fue clonado en el vector de expresión pET30a (Novagen). Células de *E. coli* BL21 (DE3) fueron transformadas con el plásmido recombinante y crecidas en medio 2xYT a 37° C en presencia de kanamicina (50mg/ml). La sobreexpresión de VHCc124 se inició con la adición de 0.5mM isopropil-β-D-tiogalactopiranosido (IPTG) cuando la densidad óptica (D.O.) del cultivo a 600nm alcanzó 0.4. La inducción se llevó a cabo durante la noche a 16° C. Las células fueron recogidas por centrifugación y los precipitados obtenidos fueron congelados a -80° C.

2.1.2. Purificación de VHCc124

La purificación de VHCc124 se llevó a cabo siguiendo dos protocolos distintos. En el primer caso se siguió el protocolo descrito por Kunkel et al (Kunkel 2001) desnaturalizando la proteína con 8M urea y replegándola posteriormente mediante diálisis extensiva en un tampón sin urea (20mM Tris-HCl pH7.0, 100mM NaCl). Las células se rompieron con un tampón de ruptura con 25mM Na₂HPO₄ pH7.0, 8M urea, 250mM NaCl, 2mM EDTA, 1mM ditiotreitól (DTT) seguido de 9 minutos de sonicación y 20 minutos de centrifugación a 40000RPM en un rotor Beckman 45 Ti E-4082 (186.000g). Posteriormente se utilizó una columna de intercambio catiónico SP sefarosa (Amersham Biosciences) previamente equilibrada en un tampón con 250mM

HEPES pH 7.0, 8M urea, 250mM NaCl. La proteína se eluyó mediante un gradiente lineal hasta 1.5M NaCl. Las fracciones que contenían la proteína se aplicaron a una columna de fase reversa en cromatografía líquida de alta presión (HPLC). Una vez cargadas, la proteína se eluyó con un gradiente lineal de 20mM fosfato sódico pH 3.0 hasta 20mM fosfato sódico con 90% metanol pH 3.0. Las fracciones de proteína se dializaron toda la noche frente al mismo tampón sin urea. La proteína purificada se concentró por ultrafiltración mediante filtros con un tamaño de poro de 10kDa (Amicon).

VHCc124 fue también purificada de células de *E. coli* siguiendo el protocolo de Lorenzo et al (Lorenzo 2001). Las células fueron resuspendidas en 10mM Tris-HCl pH 8.0, 6mM EDTA y se rompieron usando un Emulsiflex-C5 de Avestin. Después de centrifugar a 186.000g durante 20 minutos, se resuspendió la fracción insoluble en un tampón con 1% Tritón X-100, 10mM EDTA, 300mM NaCl, 10mM β -mercaptoetanol, 10mM Tris-HCl pH 8.0 y 2M urea. Las fracciones que contenían la proteína se cargaron en una columna de intercambio catiónico SP sefarosa (Amersham Biosciences) previamente equilibrada con 250mM HEPES pH 7.0, 250mM NaCl. Finalmente, la proteína fue eluída con un gradiente lineal hasta 1.5M NaCl. Usando los mismos tampones se pasaron las fracciones con la proteína por una columna de intercambio iónico QFF (Amersham Biosciences). La proteína purificada se concentró por ultrafiltración usando filtros con un tamaño de poro de 30 kDa (Amicon).

2.1.3. Reconstitución *in vitro* de VHCc124

Para renaturalizar VHCc124, la proteína se dializó frente a un tampón con 1.7mM acetato magnésico, 100mM acetato potásico, 25mM HEPES pH 7.4, Así mismo, las reacciones de ensamblado se llevaron a cabo incubando en condiciones libres de ARNasas VHCc124 con ARNt^{phe} durante 10 minutos a 30° C, seguidos de 15 minutos en hielo. Las relaciones molares de proteína : ARNt^{phe} usadas fueron de 5:1, 10:1 y 100:1.

2.1.4. Dicroísmo Circular

Las medidas de dicroísmo circular (DC) se llevaron a cabo en un espectrofotómetro Jasco J-810 a 20° C, usando cubetas de 0.2cm de paso óptico. La concentración de las muestras de VHCc124 fue de ~0.5mg/ml

exceptuando las medidas hechas a los NLPs que variaban según la muestra fuera 5:1, 10:1 o 100:1. La desnaturalización térmica de VHCc124 en condiciones nativas se realizó entre 20 y 80° C a una longitud de onda de 212nm.

2.1.5. Espectroscopía de fluorescencia

Los espectros de fluorescencia se llevaron a cabo usando una cubeta de 1cm de paso de luz en un espectrofluorímetro PTI (Photon Technology International). La concentración de la proteína fue de 2.8mg/ml. La fluorescencia intrínseca del triptófano se monitorizó usando una longitud de onda de excitación de 295nm y una longitud de onda de emisión entre 300nm y 400nm.

2.2. DDX3

2.2.1. Clonaje y expresión de DDX3

El gen de la proteína DDX3 se clonó a partir de una librería de cDNA humano preparada a partir de células de testículo cedido amablemente por el Dr. M. Way (Cancer UK). El gen de la proteína DDX3 se subclonó en un vector de expresión pET28a (Novagen) mediante amplificación por PCR de un vector TOPO con el gen de la proteína original. Los cebadores usados directo y reverso fueron 5'-TATTGCTCTCATGAGTCATGTGGCAGTGGAAA-3' y 5'-CGCGGATCCTCAATGGTGATGGTGATGGTGGTTACCCACCCAGTCAACCCC-3' respectivamente. Se utilizaron las enzimas de restricción BspHI, BamHI y NcoI para llevar a cabo el clonaje. El cebador reverso llevaba incluida una secuencia de 6 histidinas para la posterior purificación de la proteína. Las células de la cepa BL21 (DE3) de *E coli* fueron transformadas con el vector pET28a-DDX3 y las bacterias se cultivaron en medio LB con kanamicina (50mg/ml). La expresión de la proteína se llevo a cabo tras añadir 0.5mM IPTG cuando la D.O. del cultivo a 600nm alcanzó 0.5. La inducción se mantuvo durante 3 horas a 37° C.

2.2.2. Clonaje y expresión de DDX3_{pf}

DDX3_{pf} (164-600) fue subclonado y sobreexpresado a partir del gen de la proteína silvestre. Se utilizó un vector pETM-11 (EMBL-Heidelberg) con un

marcador de kanamicina. El cDNA se insertó en el vector mediante las enzimas de restricción BspHI, BamHI y NcoI. La amplificación por PCR se realizó partiendo del vector pET28a-DDX3 y los cebadores directo y reverso usados fueron 5'-TATTGCTCTCATGAGTGATGACATTCCAGTTGAGGCA-3' y 5'-CGCGGATCCCTATTTTTTCGAACTGCGGGTGGCTCCAAGCGCTTCTGGC ACCAAACCCTCCAC-3' respectivamente. El vector contiene en el extremo 5' una secuencia que codifica 6 histidinas seguidas de un sitio de corte para la proteasa del virus del jaspeado del tabaco (TEV). El cebador reverso se aprovechó para introducir una secuencia de 8 aminoácidos denominada STREP (WSHPQFEK). Las células de la cepa BL21 (DE3) de *E. coli* fueron transformadas con el vector pETM11-DDX3_{pf}. Las bacterias fueron cultivadas en medio LB con kanamicina (50mg/ml). La expresión de la proteína se llevó a cabo tras añadir 0.5mM IPTG cuando la D.O. del cultivo a 600nm llegó a 0.5. La inducción se realizó durante 3 horas a 37° C. Las células fueron recogidas por centrifugación y los precipitados obtenidos fueron congelados a -80° C.

2.2.3. Clonaje y expresión de DDX3_{pdR}

DDX3_{pdR} (169-579) fue subclonado mediante amplificación por PCR en un vector pColdIII (Takara) usando NdeI y BamHI como enzimas de restricción. Los cebadores directo y reverso fueron 5'-GGGAATTCCATATGTGGAGCCACCCGCAGTTCGAAAAAGAAAACCTTAT ATTTCCAAGGAGAGGCAACAGGCAACAACACTGT-3' y 5'-AAAACATGGCTTATGAACACCACCATCACCATCACTGAGGATCCGCG-3' respectivamente. Una secuencia STREP seguida de un sitio de corte TEV fue introducida en el extremo 5'. En el extremo 3' se añadieron cuatro histidinas, que formaron una secuencia de seis sumadas a las dos histidinas naturales en las que termina la construcción. Las células de la cepa BL21(DE3) fueron transformadas con el vector pColdIII-DDX3_{pdR}. Los cultivos bacterianos se crecieron a 37° C y se indujeron con 1mM IPTG cuando la D.O. del cultivo a 600nm alcanzó 0.5-0.8. La temperatura del cultivo se redujo entonces a 15° C y se mantuvo así toda la noche. A la mañana siguiente se recolectaron las células por centrifugación y los precipitados obtenidos fueron congelados a -80° C.

centrifugación en 20 minutos. Posteriormente se cargó la proteína en una columna de afinidad con resina de TALON (BD Biosciences Clontech) igual que la utilizada para la purificación de DDX3. La resina, cargada con cobalto, se equilibró con tampón A. La elución de la muestra se realizó con 150mM de imidazol. Las fracciones que contenían la proteína se cargaron en una columna de afinidad con resina STREPtactin superflow de 5ml (IBA) equilibrada previamente con un tampón 100mM Tris-HCl pH 8.0, 150mM NaCl, 1mM EDTA. La elución se realizó con 2.5mM de destiobiotina. La muestra con la proteína se dializó de manera extensiva frente a un tampón con 50mM Tris-HCl pH 8.0, 500mM (NH₄)₂SO₄ y se cargó en una columna de afinidad de HisTrap (Amersham Biosciences) de 1ml.

2.2.7. Purificación de DDX3_{pd}

La purificación de DDX3_{pd} se realizó a partir de un precipitado correspondiente a 4.5L de cultivo. Las células se resuspendieron en 100ml de tampón B (50mM Tris, 1M NaCl pH 8.0), y la muestra se sonicó a 50w durante 20 minutos a 4° C. Los restos celulares se centrifugaron a 186.000g. La fracción soluble se cargó en una columna con resina de TALON (BD Biosciences Clontech) previamente equilibrada con tampón B conectada a un equipo Äkta prime. La muestra se eluyó con un gradiente hasta 150mM imidazol. Las fracciones que contenían la proteína se cargaron en una columna de STREPtactin superflow de 5ml (IBA) equilibrada con 50mM Tris, 1M NaCl, 1mM EDTA, 1mM DTT, pH 8.0. La elución de la muestra se realizó en un solo paso con un tampón con 2.5mM de destiobiotina. Se realizó entonces el corte proteolítico de la muestra con TEV a 4° C durante la noche. La muestra se cargó en una columna de filtración en gel Hi Load 26/60 superdex 200 (Amersham-Biosciences) previamente equilibrada con 10mM Hepes, 1M NaCl, 1mM DTT, 1mM EDTA, 5mM MgCl₂, pH8.0. Las fracciones que contenían la proteína se concentraron por ultrafiltración usando filtros Amicon de un diámetro de poro de 30kDa. La proteína se concentró hasta 10mg/ml.

2.2.8. Purificación de DDX3_{hel}

DDX3_{hel} se purificó a partir de un precipitado celular correspondiente a 4L de cultivo. Las células se resuspendieron en tampón C (50mM Tris, 500mM (NH₄)₂SO₄, pH 8.0) y se rompieron aplicando 10 minutos de sonicación a

50w y a 4° C. Los restos celulares se separaron por ultracentrifugación durante 20 minutos a 186.000g. La fracción soluble con la proteína fue purificada usando una columna de afinidad de 5ml His-Trap (Amersham-Biosciences). La muestra se cargó en la columna previamente equilibrada con tampón C y se eluyó usando un gradiente escalonado de imidazol hasta 500mM. La proteína DDX3_{hel} fue sometida a corte proteolítico con la proteasa TEV a 4° C durante la noche. La muestra fue cargada de nuevo en una columna de afinidad de 5ml His-Trap (Amersham-Biosciences) y la proteína sin la secuencia de histidinas eluyó de la columna sin unirse a la resina. Posteriormente fue cargada en una columna de filtración en gel Hi Load 26/60 superdex 75 (Amersham Biosciences) previamente equilibrada con 10mM Hepes, 500mM (NH₄)₂SO₄, pH 8.0. La muestra fue concentrada hasta 10mg/ml por ultrafiltración usando filtros de un diámetro de poro de 10kDa (Amicon). Una vez concentrada, la proteína se congeló en nitrógeno líquido y se almacenó a -80° C.

2.2.9. Dicroísmo Circular (DC)

Las medidas de dicroísmo circular fueron tomadas en un espectrofotómetro Jasco J-810 a 20° C, usando cubetas de 0.2cm de paso óptico. La concentración de DDX3 fue 0.4mg/ml. La concentración de DDX3_{pf}, DDX3_{pdr} y DDX3_{hel} fue de 1mg/ml. La absorbancia de las soluciones en las que estaban las diversas muestras fue medida y sustraída de cada uno de los espectros. Los espectros se promediaron a partir de 10 medidas tomadas entre 250nm y 208nm para DDX3 y DDX3_{hel}, 300nm y 210nm para DDX3_{pf} y 250nm y 211nm para DDX3_{pdr}. El ancho de banda fue de 1 nm, la velocidad del barrido fue de 100 nm min⁻¹ y el tiempo de respuesta fue de 2 segundos. Las desnaturalizaciones térmicas se llevaron a cabo a 220nm variando la temperatura de 20° C a 80° C. En el caso de DDX3_{hel}, se repitió el ensayo de DC usando proteína que había sido previamente congelada en nitrógeno líquido.

2.2.10. Proteólisis limitada de DDX3

Con el objetivo de eliminar partes flexibles de la construcción, la proteína DDX3 se sometió a proteólisis limitada con las endoproteasas tripsina, quimotripsina, elastasa, V8, subtilisina y Asp-N. La concentración de cada proteasa fue de 0.1mg/ml y la relación proteína: proteasa fue de 4 a 1. La

concentración de DDX3 fue de 0.4mg/ml y la digestión se llevó a cabo a temperatura ambiente durante toda la noche. Se tomaron muestras que se analizaron mediante SDS-PAGE en geles del 10%, las bandas se tiñeron con azul de Coomasie y se analizaron mediante espectrometría de masas para la determinación del extremo C-terminal por el método MALDI-TOF. Además, los fragmentos derivados de la digestión de DDX3 se transfirieron a membranas de difluoruro de polivinilo (PVDF) (Millipore) para la secuenciación del extremo N-terminal mediante degradación de Edman.

2.2.11. Ensayo de actividad helicasa dependiente de la hidrólisis de ATP

La actividad ATPasa de la proteína fue monitorizada de forma continua acoplado la reacción a la reducción de NAD⁺. Este método se basa en la hidrólisis del ATP asociada a la oxidación del NADH que pierde su absorbancia a 338nm a medida que ocurre este fenómeno. El ensayo se realizó a 37° C en un volumen de reacción de 1ml utilizando una cubeta de 1cm de paso óptico. La reacción se llevó a cabo con 20mM Tris-HCl pH 8.0, 50mM KCl, 5mM MgCl₂, 1mM DTT, 1mM ATP, 300µM NADH, 2mM fosfoenolpiruvato y 3 unidades/ml de piruvato kinasa y lactato deshidrogenasa. La concentración de la proteína fue 150nM. En una de las reacciones se utilizó ARN (50µg/ml) como sustrato obtenido a partir de extracto total de fibroblastos de ratón. El ensayo se realizó para DDX3, DDX3_{pf} y DDX3_{pdr} bajo las mismas condiciones experimentales. La única variación fue en el tiempo de la reacción, que en el caso de DDX3 fue de 3600 segundos, 6000 segundos para DDX3_{pf} y 1500 segundos para DDX3_{pdr}.

2.2.12. Dispersión dinámica de luz (DLS)

Para evaluar su monodispersidad, las diferentes construcciones se analizaron mediante DLS usando una cubeta de 1cm de paso óptico en un DLS de Protein Solutions de DynaPro. DDX3 y DDX3_{pf} se midieron a una concentración de 1mg/ml. DDX3_{pdr} y DDX3_{hel} se midieron a una concentración de 10mg/ml. Las muestras se centrifugaron previamente 5 minutos a 16.100g para eliminar partículas de polvo. Para analizar los resultados se promediaron 30 medidas tomadas cada 10 segundos a una temperatura de 20° C excepto en el caso de DDX3_{hel} que la temperatura fue de 16° C.

2.2.13. Digestión proteolítica con la proteasa TEV

Las condiciones de corte con la proteasa TEV fueron las mismas para DDX3_{pdr} y para DDX3_{hel}. Se usó proteasa TEV a una concentración de 1mg/ml en un tampón de 10mM Tris, 300mM NaCl, 1mM EDTA, 1mM DTT, 25% Glicerol pH8.0. El corte se realizó a 4° C durante toda la noche en tubos Eppendorf de 1.5ml girando en una noria. La relación proteína:TEV fue de 10:1.

2.2.14. Cristalización de DDX3_{hel}

Los ensayos de cristalización se llevaron a cabo mediante un robot Cartesian Mycrosys SQ (Genomic Solutions) utilizando el método de la gota sentada con gotas de 0.1µl de proteína más 0.1µl de líquido del reservorio y un volumen de reservorio de 70µl en placas Grenier. Inicialmente, se obtuvieron pequeñas esferulitas después de 48 horas a 16° C en 0.1M HEPES pH 7.5, 2.5M (NH₄)₂SO₄, condición 10 del MbClass II screen (Nextal-Qiagen). El refinado de las condiciones de cristalización se llevó a cabo a la misma temperatura usando el método de la gota colgante en bandejas Limbro. El volumen del reservorio fue de 500µl y las gotas fueron de 1µl + 1µl. Después de varias rondas de optimización usando diferentes concentraciones de precipitante, variaciones de pH y un cribado de aditivos de Hampton Research, aparecieron cristales en 2M (NH₄)₂SO₄, 0.1M Imidazol pH 6.4 más 5mM espermina tetra-HCl en gotas de 2µl. A pesar de la gran cantidad de cristales que crecieron en la gota, estos especímenes pudieron ser recogidos usando un crioloop. El tamaño medio de los cristales obtenidos fue de 0.2x0.01x0.01mm.

2.2.15. Recogida y reducción de datos

Los cristales de DDX3_{hel} fueron crioprottegidos usando un tampón con 2M (NH₄)₂SO₄, 0.1M imidazol pH 6.4 y 20% de glicerol. Los cristales fueron transferidos de la gota original a una gota de crioprotector. Se mantuvieron en esa solución alrededor de cinco segundos y se transfirieron entonces a nitrógeno líquido. A pesar de que algunos sets de datos fueron obtenidos en casa usando un generador FR-591 de Bruker, los cristales sólo difractaron a 3.8Å. Así, el resto de los grupos de datos de DDX3_{hel} fueron recogidos usando radiación de sincrotrón en las líneas ID14-1 e ID23-1 del European Synchrotron Radiation Facility (ESRF) en Grenoble y la línea X06SA del

Swiss Light Source (SLS), en Villigen. Los datos de difracción fueron recogidos con un detector ADSC-Q210 en la línea ID14-1, un detector ADSC-Q315r en la línea ID23-1 en el ESRF, y un detector MAR225 CCD en el SLS. Los mejores grupos de datos fueron recogidos usando una $\Delta\varphi = 1^\circ$ y una longitud de onda de 0.9 Å en ID23-1. El procesamiento y el escalado de los datos se llevó a cabo con el programa HKL2000 (Otwinowski 1997).

2.2.16. Determinación y refinado de la estructura

La estructura se resolvió usando el método del reemplazo molecular con el programa MOLREP (Murshudov 1997). El modelo utilizado fue el dominio helicasa de la proteína Vasa de *D. melanogaster* (PDB ID:2DB3). Los aminoácidos que eran diferentes a la secuencia de DDX3 fueron mutados a Ala. ARP/wARP (Morris 2003) y REFMAC5 (Murshudov 1997) se aplicaron para llevar a cabo la construcción automática del modelo y posterior refinado (PDB ID:2JGN).

2.3. CD5

2.3.1. Clonaje de CD5DIII

La forma recombinante de la proteína humana CD5DIII fue expresada usando un sistema de expresión episómica en células embrionarias de hígado humano (HEK 293-EBNA). Estas células expresan de forma constitutiva la proteína viral de Epstein Barr EBNA-1, permitiendo una replicación episómica del vector pCEP-Pu (donado por el Dr. Sasaki y el Dr. Timpl del instituto de Bioquímica Max Planck, Alemania) (Kohfeldt 1997). El cDNA que codificaba la secuencia correspondiente a CD5DIII fue obtenida mediante amplificación por PCR del plásmido pH[®].T1T2 (Calvo 1999) usando como cebadores directo y reverso 5'-CCCCGCTAGCTTTCCAGCCCAAGGTGCAG-3' y 5'-GTGGATCCTAATCCTGGCATGTGACAAAC-3' respectivamente. El cebador directo incorporaba el sitio de restricción NheI mientras que el cebador reverso incorporaba un codon stop seguido del sitio de restricción BamHI. Los productos amplificados de la PCR fueron clonados en el vector pCEP-Pu, en línea con el péptido señal BM-40 para la secreción de la proteína recombinante al medio extracelular. Esto resultó en la adición de los aminoácidos Ala-Pro-Leu-Ala en el extremo N-terminal de la proteína.

2.3.2. Expresión de CD5DIII

Las construcciones resultantes fueron expresadas en células HEK 293-EBNA tal y como está descrito previamente (Kohfeldt 1997). Las células HEK 293-EBNA en platos de cultivo fueron transfectadas usando lipofectamina y crecidas en monocapa usando D-MEM:F-12(1:1)+Glutamax (Gibco) suplementado con suero bovino fetal (SBF) (10%), (Sigma), 100 unidades/ml 100 µg/ml de penicilina-estreptomicina (Gibco), 250 µg/ml de geneticina (Gibco) y 2.5 µg/ml de puromicina (Sigma). Para la producción a gran escala, las células, una vez alcanzada la fase de crecimiento exponencial, fueron tripsinizadas y finalmente plateadas en frascos de cultivo Costar (162 cm²). El medio se intercambiaba cada 48 horas y cuando la superficie del frasco estaba cubierta en un 80% se añadía nuevo medio sin SBF. 3-4 días después de este cambio se recolectaba la proteína expresada. Cada dos días el medio era recogido y juntado en tubos Falcon de 50 ml, centrifugado a 1000g durante 5 minutos y los restos celulares eran retirados. El sobrenadante limpio era transferido a un nuevo tubo Falcon, congelado en nitrógeno líquido y almacenado a -80° C. Al cabo de 3 o 4 semanas, alrededor de 2.5-3L de sobrenadante con 200-300 µg/L de proteína eran recolectados. En este punto las células empezaban a desprenderse de la superficie y el cultivo se detenía.

2.3.3. Purificación del anticuerpo monoclonal

CD5DIII fue purificada a 4° C por cromatografía de afinidad usando un anticuerpo monoclonal de ratón (83-C4 Mab) específico para CD5DIII (Calvo 1999). Alrededor de 4L de sobrenadante del hibridoma productor de 83-C4 Mab fueron precipitados añadiendo 50% (NH₄)₂SO₄ y agitando suavemente en hielo durante 30 minutos. Posteriormente, se centrifugó la mezcla a 6000g durante 30 minutos. El precipitado con el anticuerpo fue resuspendido en 50 ml de PBS (0.137M NaCl, 0.01M Na₂HPO₄·2H₂O, 2.7mM KCl, 2mM KH₂PO₄) y extensamente dializado frente a PBS para eliminar los restos de (NH₄)₂SO₄. Después de esta diálisis la muestra fue filtrada con filtros de 0.22 µm (Millipore).

83-C4 Mab fue purificado usando una columna de 5ml HiTrap de proteína G (Amersham Biosciences) conectada a un Äkta prime. La columna fue

equilibrada previamente con PBS y la muestra cargada en la columna. Después de un lavado extensivo con PBS, 83-C4 Mab fue eluido usando 0.1M glicina pH 2.6. Los tubos en el colector contenían 200 µl de 1M Tris pH 9.0 para tamponar la muestra y evitar la desnaturalización. Las fracciones con el anticuerpo se dializaron frente a 0.1M NaHCO₃ pH 8.3, 0.5M NaCl. El rendimiento fue de 25mg de 83-C4 Mab por litro de sobrenadante. Una vez que el anticuerpo estuvo purificado, se cargó en una resina de CNBr-Sepharosa 4B (Amersham Biosciences) previamente activada con varios lavados de 1mM HCl, y equilibrada con la solución de copulado (0.1M NaHCO₃ pH 8.3, 0.5M NaCl). Se incubó la resina con el anticuerpo toda la noche a 4° C usando una noria. Finalmente, se retiró el exceso de 83-C4 Mab lavándolo con el tampón de copulado y se bloquearon los grupos activos restantes mediante incubación de la resina cargada con una solución 0.1M Tris-HCl, pH 8.0 durante 2 horas. Con la resina se empaquetó entonces una columna que se conectó a un Äkta prime para el proceso de purificación.

2.3.4. Purificación de CD5DIII

Para la purificación se usó sobrenadante correspondiente a 2.5-3L de cultivo de HEK-EBNA-293-CD5DIII. Este volumen fue concentrado a 4° C en una celda Amicon usando unas membranas de 10kDa de tamaño de poro hasta una cantidad de 100-200ml. Entonces fue cargado en una columna de 10ml de 83-C4 Mab previamente equilibrada con PBS. La muestra fue recirculada 10 veces para mejorar la unión y posteriormente la columna se lavó con 10 volúmenes de PBS más 0.5M NaCl y 1% NP40 y 10 volúmenes de PBS. La proteína se eluyó con 3.5M MgCl₂ en PBS. Todos los pasos cromatográficos se llevaron a cabo a un flujo constante de 1.5ml/min. La cantidad de proteína purificada dependía del sobrenadante recolectado y oscilaba entre 200µg y 500µg de 2.5-3L de cultivo. Finalmente, el tampón de elución se intercambiò a 0.01M HEPES, 0.2M NaCl, pH 7.4 usando una columna de desalado (Amersham Biosciences). La muestra con la proteína se concentró entonces hasta 4mg/ml usando un centricon de 10kDa de poro a 4° C.

2.3.5. Espectrometría de masas

La proteína purificada fue cargada en un gel SDS-PAGE del 10%. Se extrajo la banda correspondiente a CD5DIII, se lavó, y se redujo la proteína con 20mM DTT durante 30 minutos a 37° C, se alquiló y seguidamente se digirió

con tripsina (Promega) durante la noche (Shevchenko 1996). Los fragmentos de la digestión fueron analizados por doble espectrometría de masas usando HPLC-MS/MS junto a una trampa iónica ESI o MALDI-TOF. Los análisis de HPLC-MS/MS se llevaron a cabo usando un sistema nano-HPLC Ultimate 3000 (LC-Packings) acoplado a un espectrómetro de masas de trampa iónica lineal LTQ (Thermo Finnigan) equipado con una fuente de nano-ESI. Los fragmentos de la digestión triptica se separaron usando una columna de fase reversa C18 PepMap100 (75 μ m \times 15cm) mediante un gradiente de cuatro pasos en 60 minutos: 1-60% ACN durante 40 minutos, 60-90% ACN durante 5 minutos, 5 minutos de 90% ACN y 1% ACN en 0.1% de ácido fórmico durante 5 minutos. Los péptidos ionizados se detectaron en un escaner completo de masa frente a carga (m/z) entre 400 y 200 m/z. Los espectros de MS2 fueron adquiridos al 35% de energía de colisión para la disociación colisión-inducida. Los datos fueron analizados con el algoritmo TurboSequestTM desarrollado por BioWorks 3.1 software (Thermo Finnigan) por contraste con la base de datos de la secuencia de CD5DIII en formato FASTA. Todos los fragmentos del espectro fueron revisados de forma manual para comprobar que todos los péptidos coincidieran con los picos principales. Los péptidos en el rango de m/z de 800-3000 se tomaron con un 4800 MALDI-TOF (Applied Biosystems) usando ácido α -ciano-4-hidroxicinámico como matriz.

2.3.6. Resonancia magnética nuclear (RMN)

La muestra de RMN, con un volumen de 300 μ L (99% H_2O , 1% 2H_2O), contenía 43 μ M CD5DIII, PBS pH 7.1, y 15 μ M DSS (2,2-dimetil-2-silapentano-5-sulfonato sódico), usado como referencia interna para los desplazamientos químicos de 1H . Los datos se adquirieron a 298 K en un equipo Bruker AVANCE 600 equipado con una criosonda, con un tamaño de 8kb y usando presaturación de agua durante el tiempo de relajación (1 segundo). El espectro fue procesado con XWINNMR (Bruker) realizándose un relleno de ceros hasta 32kb y una multiplicación exponencial (de 2Hz de ensanchamiento de línea) antes de la transformada de Fourier.

2.3.7. Cristalización de CD5DIII

Los ensayos de cristalización se llevaron a cabo usando un robot Cartesian Mycrosys SQ (Genomic Solutions) por el método de la gota sentada, con

nanogotas de 0.05 μl . Un total de 1000 condiciones de diferentes cribados de Hampton Research, Emerald Biosciences y Qiagen fueron probadas usando un único bloque de proteína. Después de 48 horas se obtuvieron pequeñas agujas en la condición 17 del wizard I (30% PEG 8000, 0.1M Na Acetate, pH4.5, 0.2M Li_2SO_4). Tres bloques más de proteína fueron necesarios para llevar a cabo el proceso de optimización de las condiciones de cristalización. Después de varias rondas de refinamiento se obtuvieron cristales con la calidad suficiente para desarrollar experimentos de difracción de rayos X. Las condiciones finales de cristalización fueron 15-25% PEG 8000, 0.1M Na Acetate pH 4.4, 0.2M Li_2SO_4 en el reservorio mientras la gota (colgada o sentada) sólo contenía el tampón en el que se encontraba la proteína 0.01M HEPES, 0.2M NaCl, pH 7.4

2.3.8. Ensayo de desplazamiento de banda con átomos pesados

El ensayo se realizó tal y como está descrito en la literatura (Boggon 2000). La proteína, a una concentración de 0.5mg/ml, se incubó con las distintas sales de átomos pesados en las concentraciones en que éstas serían utilizadas. Se prepararon geles nativos del 10% y se corrieron durante 10 horas primero y después de observar la baja movilidad en gel de CD5DIII se aumentó el tiempo de electroforesis hasta un total de 20 horas.

2.3.9. Recogida y reducción de datos

Todos los datos fueron recogidos a temperaturas criogénicas usando radiación de sincrotrón a 100K. Los cristales de CD5DIII fueron montados y crioprottegidos usando un 30% de glicerol. Los datos de difracción fueron recolectados usando radiación de sincrotrón en la línea X06SA en el SLS (Villigen). Los datos de difracción fueron recogidos en un detector MAR225 CCD. La recolección de datos de SAD de azufre se llevó a cabo usando la estrategia de rayo inverso para maximizar las diferencias anómalas recolectando pares de Bijvoet intactos. Se usó una longitud de onda de 1.8Å para optimizar la señal anómala del azufre. El procesamiento y el escalado de los datos se realizó con el programa HKL2000 (Otwinowski 1997).

2.3.10. Determinación y refinado de la estructura

Las intensidades reducidas se usaron para buscar la subestructura de los azufres. Durante este proceso se aplicó un corte en la resolución a 4.5Å. Así, los cuatro disulfuros fueron tratados como superátomos de azufre. Las posiciones iniciales de los azufres fueron encontradas con SnB (Weeks 1999) y SHELX. Las posiciones de los azufres se introdujeron en SHARP (de la Fortelle 1997). Después de realizar el aplanado de solvente (solvent flattening) con SOLOMON (Abrahams 1996) el mapa inicial a 2.8Å fue usado para construcción automática del modelo usando MAID (Levitt 2001). El modelo de CD5DIII fue reconstruido y refinado hasta 2.5Å. Un grupo de datos de más alta resolución a 2.2Å fue usado para el refinado. La estructura de CD5DIII resuelta en el grupo espacial tetragonal (PDB ID: 2OTT) se usó como modelo en MOLREP (Vagin 2000) para resolver la estructura del cristal trigonal por reemplazo molecular (PDB ID: 2JA4). ARP/wARP (Morris 2003) y REFMAC5 (Murshudov 1997) se aplicaron para hacer construcción automática del modelo y posterior refinado.



RESULTADOS

3.1. VHCc124

3.1.1 Expresión y purificación de VHCc124

El gen de VHC que codifica la proteína C y que compone la nucleocápsida del virus se clonó mediante RT-PCR en el plásmido pQE30 a partir del suero de un paciente. A partir de esta construcción y como paso previo a la formación de NLPs, se subclonó y purificó un fragmento de la proteína C correspondiente a los aminoácidos 1-124 (VHCc124). Una vez sobreexpresada VHCc124 en células de *E. coli* (figura 3.1) la purificación se llevó a cabo de dos maneras distintas: a) de forma desnaturizante al desarrollar el proceso en presencia de 8M urea (Kunkel 2001) y b) al no conseguir resultados satisfactorios, reproduciendo el protocolo de purificación de la proteína en condiciones nativas (Lorenzo 2001).

a) La purificación de la proteína desnaturizada se realizó usando un tampón de ruptura celular con 25mM Na₂HPO₄ pH7.0, 8M Urea, 250mM NaCl, 2mM EDTA, 1mM DTT. Después de una centrifugación a 186.000g se cargó la fracción soluble en una columna de intercambio catiónico con resina SP sefarosa. A continuación, las fracciones eluidas de VHCc124 se cargaron en una columna de fase reversa montada en un sistema de HPLC (figura 3.1). La proteína asilada fue posteriormente concentrada por ultrafiltración con filtros de un tamaño de poro de 10kDa.

b) La purificación de la proteína en estado nativo se realizó resuspendiendo las células en un tampón 10mM Tris-HCl pH 8.0, 6mM EDTA. Las células se lisaron por presión usando un Emulsiflex-C5. La muestra fue centrifugada a 186.000g. Después de observar que la mayor parte de la proteína permanecía en el precipitado, éste se lavó con 1% Tritón X-100, 10mM EDTA, 300mM NaCl, 10mM β -mercaptoetanol, 10mM Tris-HCl pH 8.0 y 2M urea. La muestra se volvió a centrifugar y la fracción soluble fue cargada en la misma columna de intercambio catiónico utilizada anteriormente para la purificación con 8M urea (figura 3.1). La muestra fue eluída con un gradiente hasta 1.5M NaCl y las fracciones que contenían la proteína VHCc124 se cargaron en una columna de intercambio aniónico QFF de 5ml (figura 3.1). La proteína purificada fue concentrada por ultrafiltración usando filtros con un diámetro de poro de 30kDa intentando concentrar los NLPs ya formados.

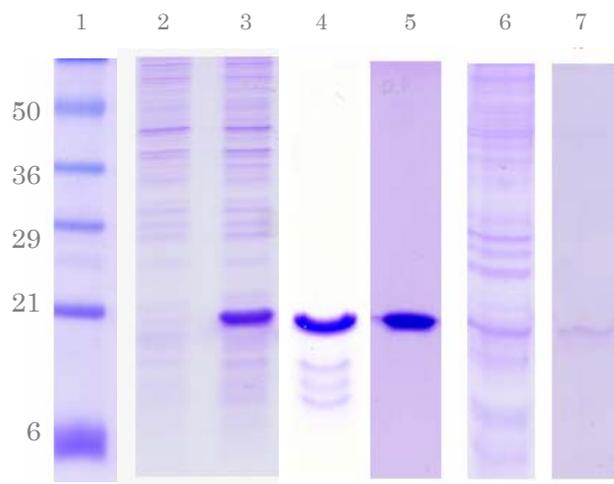


Figura 3.1. Expresión y purificación de VHCc124. Las calles 2 y 3 corresponden a muestras del cultivo antes y después de la inducción con IPTG. Las dos calles siguientes muestran la purificación de la proteína en 8M urea, después de la columna de SP Sefarosa (calle 4) y después de cargar las fracciones en una columna de fase reversa en un HPLC (calle 5). Las últimas calles corresponden a la proteína purificada en su forma nativa. Primero se utilizó una columna de SP Sefarosa (calle 6) y posteriormente la muestra se cargó en una columna de intercambio aniónico QFF (calle 7). La calle 1 señala los marcadores de peso molecular.

3.1.2. Estudios biofísicos de VHCc124

3.1.2.1 VHCc124 no puede renaturalizarse tras su resuspensión en 8M urea.

Una vez que la proteína fue aislada en su forma desnaturalizada resuspendiéndola en 8M urea, se intentó renaturalizar mediante diálisis extensiva frente a un tampón 20mM Tris-HCl pH 7.0, 100mM NaCl durante la noche. Después de este proceso se analizó la muestra por espectroscopía de DC en el ultravioleta lejano. El espectro de DC no muestra estructura secundaria en VHCc124 después de la diálisis (figura 3.2a, verde). Siguiendo el mismo procedimiento se dializó la muestra, ésta vez frente a tampón de ensamblaje (1.7mM acetato magnésico, 100mM acetato potásico, 25mM HEPES pH 7.4), pero los resultados de DC fueron los mismos, indicando la ausencia de replegamiento (figura 3.2a, roja). También se probó a ensamblar la nucleocápsida añadiendo ARNt^{phe} pero no se detectaron cambios en el espectro de DC (figura 3.2a, azul).

A estos datos hay que añadir el espectro de fluorescencia realizado a la muestra de VHCc124 con ARNt^{phe} en proporción de 100:1 que corrobora el estado de desnaturalización de la muestra. La proteína contiene 5 residuos de triptófano localizados en las posiciones 76, 83, 93, 96 y 107 conservados en todas las secuencias de las distintas variantes de la proteína C (Bukh 1994). Se conoce (Nolandt 1997) que esta región rica en triptófanos es la principal región encargada del ensamblaje y se supone que los residuos no deben estar expuestos al medio. Sin embargo, el $\lambda_{\text{máx}}$ de emisión de los triptófanos de la proteína fue de 355nm indicando la exposición al medio de los mismos. Como control negativo del experimento se utilizó la proteína antes del tratamiento de replegamiento (figura 3.3).

3.1.2.2. VHCc124 presenta estructura cuando se purifica de forma nativa

En contraste con los datos anteriores, los resultados obtenidos utilizando la construcción VHCc124 purificada bajo condiciones no desnaturalizantes revelan que la proteína tiene estructura secundaria. En este caso el análisis del espectro de DC en el ultravioleta lejano de la proteína, muestra un mínimo a 210nm (figura 3.2b) en contraste con el espectro de la proteína purificada en 8M urea. Un posterior análisis por desnaturalización térmica muestra un cambio en la elipticidad de la proteína medida a 212nm a una temperatura media de desnaturalización (T_m) de 52° C (figura 3.2c). La

curva de desnaturalización es, no obstante, poco cooperativa lo que puede indicar un plegamiento sólo parcial de la proteína.

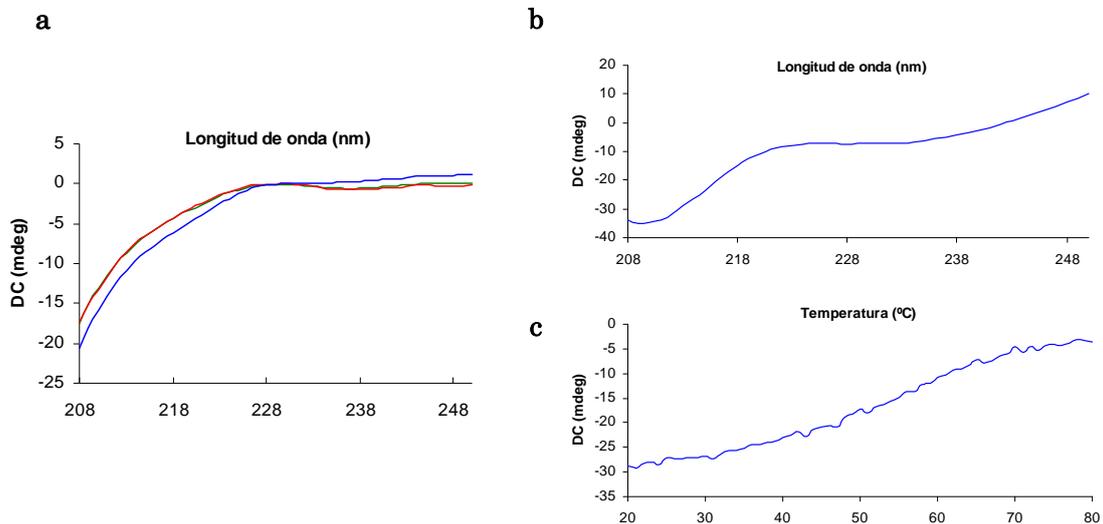


Figura 3.2. Espectros de DC de la proteína VHCc124 purificada de forma desnaturalizada y en estado nativo. a) Espectro de DC de la proteína VHCc124 en diferentes etapas de la formación de NLPs, renaturalizada (verde), en tampón de ensamblaje (roja) y en tampón de ensamblaje con ARNt^{phe} en proporción 100:1 (azul). b) Espectro de DC de la proteína VHCc124 purificada de forma nativa. c) Desnaturalización térmica de la proteína de 20° C a 80° C a una longitud de onda de 212nm que presenta una T_m de 52° C.

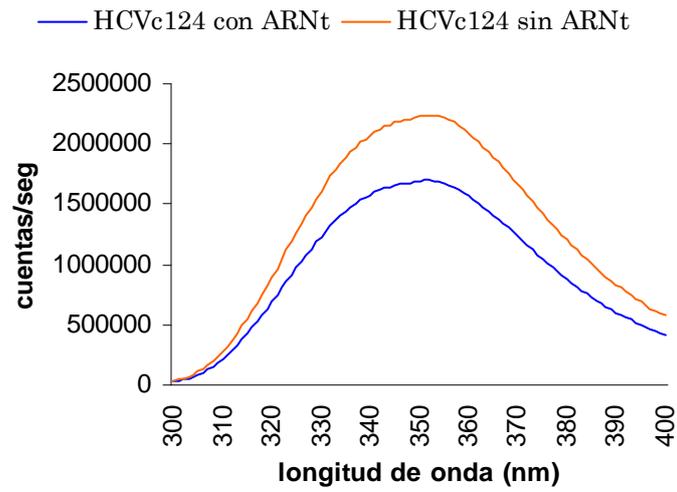


Figura 3.3. Espectros de Fluorescencia de VHCc124. Se realizó un barrido de 300nm a 400nm y se observó la máxima emisión a 355nm tanto para la proteína antes de renaturalizarla (naranja) como después de renaturalizada (azul).

3.2. DDX3

3.2.1. Determinación de dominios de DDX3

Antes de abordar el subclonaje, expresión y purificación de DDX3, se llevó a cabo un estudio *in-silico* de la proteína para obtener la mayor cantidad posible de información sobre los distintos dominios de la enzima. Así, en caso de no poder cristalizar la proteína silvestre, poder abordar la cristalización de sus diversos dominios para conseguir información estructural de DDX3. De esta manera, se realizó un alineamiento de la secuencia de DDX3 con las secuencias correspondientes a las estructuras resueltas que presentan una mayor homología (introducción, [figura 1.5](#)) y donde se pueden observar las dos regiones descritas que contienen todos los motivos funcionales de la familia de las proteínas DEAD box.

En el límite de ambas regiones se encuentra un grupo de aminoácidos de unión ausentes en el caso de *M. jannaschii* (409-412). También es de destacar las regiones en los extremos N- y C-terminal de la proteína DDX3 que están ausentes en el resto de las estructuras resueltas. En el extremo C-terminal se halla una región rica en Arg y Ser comprendida entre los aminoácidos 582-632 que se asemeja mucho a los motivos RS de proteínas como ASF/SF2 y SC-35, conocidas por mediar interacciones proteína-proteína (Wu 1993; Kohtz 1994). Además, los aminoácidos 585-622 están descritos como la zona de interacción de DDX3 con la proteína C del VHC (Owsianka 1999).

Se realizó un análisis de las regiones desordenadas o más flexibles de la proteína usando el software Predictor of Naturally Disorder Regions (PONDR) (Radivojac 2003). Analizando todos estos datos en su conjunto, los alineamientos de secuencia, los motivos funcionales y las regiones desordenadas, se plantearon varias opciones interesantes para producir diferentes construcciones de DDX3. Los fragmentos de DDX3 que se subclonaron o se plantearon subclonar están representados en la [figura 3.4](#). La primera construcción que se diseñó fue DDX3_{pdr} (169-579) y está basada en los resultados estrictos del programa de predicción PONDR. Esta construcción incluye todos los motivos conservados de las proteínas DEAD box. Se planteó también el subclonaje de DDX3_{ATP}, que constituye la zona amino terminal (173-406) del núcleo de la proteína (180-580), y de DDX3_{hel}

(407-578), el carboxilo terminal de esta región que incluye los aminoácidos (409-412) que enlazan ambas zonas.



Figura 3.4. Esquema de las diferentes construcciones de DDX3 diseñadas en este estudio. En la parte superior de la proteína DDX3 se señalan los motivos de las proteínas DEAD box. Se puede observar como DDX3_{RS}, DDX3_{pf} y DDX3_{pdr} contienen todos estos motivos mientras que DDX3_{ATP} contiene los motivos F, Q, I, Ia, Ib, II y III y DDX3_{hel} los restantes IV, V y VI.

A raíz de los experimentos que se realizaron con la proteína silvestre y debido a los problemas de solubilidad de la misma, se realizó un ensayo de proteólisis controlada que dio como resultado el fragmento DDX3_{pf} (164-600). Este fragmento es similar al fragmento previamente propuesto por el programa PONDR, lo que corroboraba los datos experimentales. A la vez que se consideró este clonaje, se planteó la posibilidad de incluir, a pesar de que pudieramos añadir desorden en el extremo C-terminal, la región que pertenece a la zona descrita anteriormente como responsable de la interacción con la proteína C (585-622) para intentar formar el complejo, generando así la construcción DDX3_{RS} (164-623).

3.2.2. Expresión y purificación de DDX3

Para poder realizar los estudios bioquímicos y la posterior cristalización de DDX3 (74kDa) la proteína se fusionó con una secuencia de seis histidinas en el extremo C-terminal que nos permitió su purificación mediante una columna de afinidad. La proteína fue sobreexpresada en células de *E. coli* de la cepa BL21 (DE3). Tras lisar las células por sonicación en un tampón 50mM Tris pH 7.5, 300mM NaCl, y centrifugar la muestra a 186.000g, alrededor del 100% de la proteína permanecía en la fracción insoluble. Este primer contratiempo se solventó sustituyendo los 300mM NaCl del tampón de ruptura por 500mM $(\text{NH}_4)_2\text{SO}_4$ (Story 2001). De esta forma, cerca del 20% de la proteína quedó en la fracción soluble. Esta muestra se purificó con una columna de afinidad de TALON seguida de una columna de filtración en gel Hi Load 26/60 superdex 200, obteniendo proteína monomérica y pura (figura 3.5) consiguiendo un rendimiento de alrededor de 1.5mg/L.

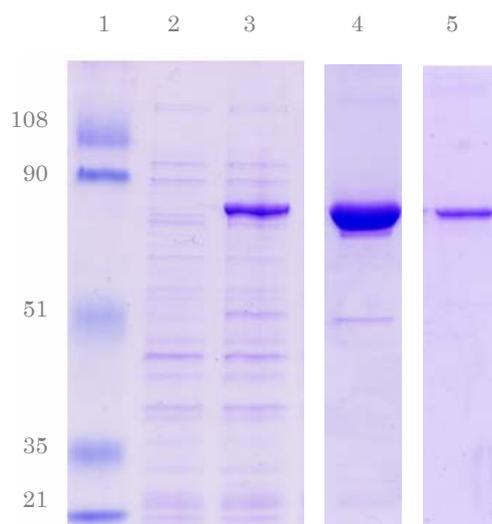


Figura 3.5. Expresión y purificación de DDX3. Las calles 2 y 3 corresponden a una muestra de los cultivos celulares antes y después de la inducción de la proteína con IPTG. La calle cuatro corresponde a la muestra de proteína después de cargar el sobrenadante en una columna de afinidad y la calle 5 corresponde a la muestra de proteína que eluye de la columna de filtración en gel. Los marcadores de pesos moleculares se señalan en la calle 1.

3.2.3. Estudios biofísicos y bioquímicos de DDX3

3.2.3.1. DDX3 está plegada y presenta un estado homogéneo de oligomerización.

Tras obtener una muestra de proteína de alta pureza se realizaron los estudios necesarios para obtener información acerca del plegamiento de la proteína y la homogeneidad de la muestra, como paso previo a su cristalización. La estructura secundaria de la proteína DDX3 se analizó mediante un espectro de DC en el ultravioleta lejano y un estudio de desnaturalización térmica reveló la presencia de estructura terciaria. El análisis del espectro de DC muestra mínimos a 220nm y a 212nm que indican una clara presencia de estructura secundaria (figura 3.6a). La desnaturalización térmica reveló una curva sigmoidea con un cambio en la elipticidad de la proteína a una T_m de 55° C (figura 3.6b).

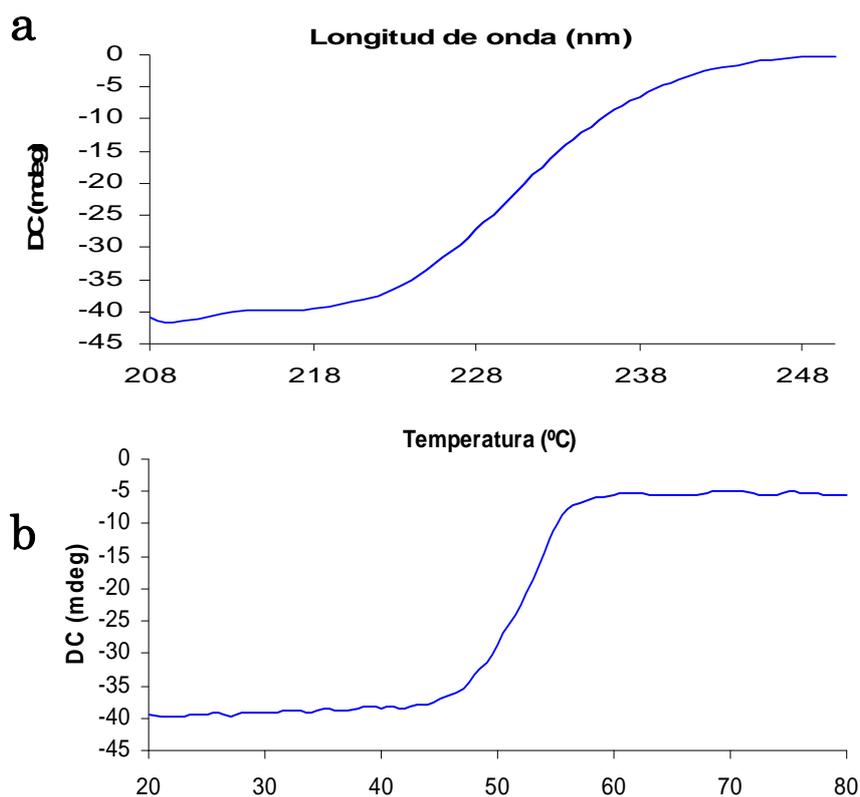
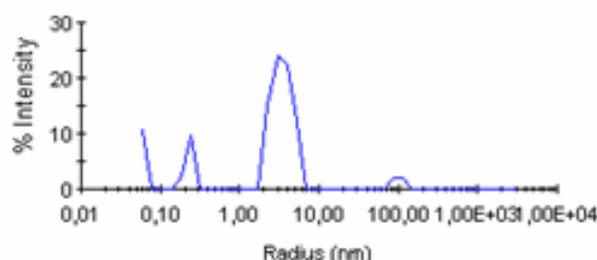


Figura 3.6. Análisis de dicroísmo circular de DDX3. a) Espectro de la proteína DDX3 entre 208nm y 250nm que reveló mínimos a 220nm y 210nm indicando una presencia de estructura secundaria. b) Desnaturalización térmica de la proteína DDX3 que muestra una clara curva sigmoidea indicativa de una transición cooperativa con una T_m de 55° C.

El estado de oligomerización de la muestra y su monodispersidad se estudió mediante un análisis de DLS. Este análisis de DDX3 muestra un único estado de oligomerización con un índice de polidispersidad inferior al 30% (27.9 %) caracterizado por una sola especie de peso molecular de 63kDa que se ajusta razonablemente a la forma monomérica (74kDa) de la helicasa (figura 3.7).



	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
Pico 1	0.1	0.0	0	10.7	97.9
Pico 2	0.2	10.2	0	12.0	2.1
Pico 3	3.5	27.9	63	73.3	0.0

Figura 3.7. Análisis de DLS de DDX3. La muestra da una señal con un pico mayoritario de un tamaño de 63kDa y un índice de polidispersidad de 27.9%.

3.2.3.2. DDX3 posee actividad helicasa asociada a la hidrólisis de ATP

El desplegamiento de ARN por parte de DDX3 es dependiente de la hidrólisis de ATP (Yedavalli 2004). Mediante la medida de la cinética de hidrólisis de ATP se monitorizó la actividad helicasa del enzima de una manera indirecta, ya que la hidrólisis de ATP se da cuando el ARN se une a la proteína, siendo la hidrólisis del nucleótido trifosfato prácticamente nula cuando el ácido nucleico está ausente. Para medir la actividad ATPasa de la proteína se siguió el protocolo descrito en la literatura (Iost 1999), acoplado la hidrólisis de ATP a la transformación del piruvato en lactato en presencia de las enzimas piruvato kinasa (PK) y lactato deshidrogenasa (LDH). A partir de fosfoenolpiruvato (PEP), comienza la reacción que aprovecha la

hidrólisis de ATP y la presencia de PK para convertir el PEP en piruvato. Éste a su vez en presencia de LDH y NADH se transforma en lactato, oxidando el NADH a NAD⁺ (figura 3.8).

Esta oxidación se puede monitorizar con un espectrofotómetro ya que el NADH absorbe la luz a 338nm y el NAD⁺ no. Así, se observa un descenso en la absorbancia de la muestra a 338nm a medida que transcurre la reacción a 37° C pero sólo cuando la proteína está en presencia de ARN. Cuando no se añade ARN a la reacción, la hidrólisis de ATP y por tanto la oxidación del NADH es prácticamente nula. Mediante análisis cinéticos llevados a cabo a concentraciones saturantes de ARN determinamos la V_{max} en 0.35 μmol de ATP $\text{min}^{-1} \text{mg}^{-1}$, que corresponde a 25.8 μmol de ATP $\text{min}^{-1} \mu\text{mol}^{-1}$ de DDX3 (figura 3.9).

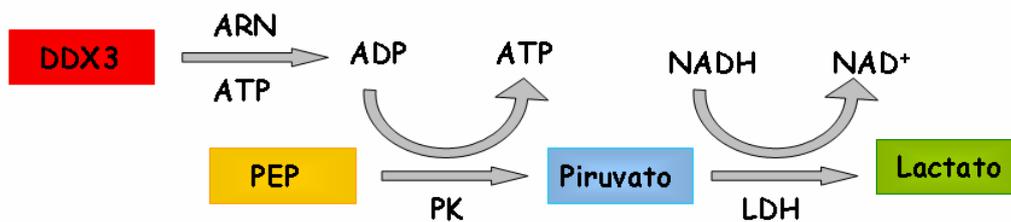


Figura 3.8. Esquema de la reacción llevada a cabo durante el ensayo de la actividad ATPasa de DDX3. La reacción comienza con la hidrólisis del ATP en presencia de ARN que produce ADP que sirve de detonante para la reacción. Como resultado se produce la oxidación del NADH que se monitoriza en un espectrofotómetro.

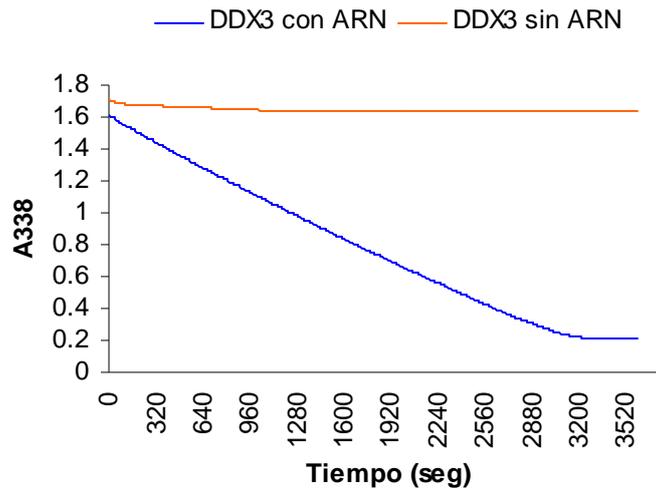


Figura 3.9. Actividad helicasa de DDX3 dependiente de la hidrólisis de ATP. La actividad helicasa de DDX3 fue analizada midiendo con un espectrofotómetro la cinética de la actividad ATPasa de la proteína en presencia y ausencia de ARN. La oxidación del NADH, directamente proporcional a la hidrólisis de ATP fue monitorizada de manera constante midiendo la absorbancia de la muestra a 338nm. La reacción se llevó a cabo a 37° C. La curva naranja corresponde a la muestra de DDX3 sin ARN. La curva azul corresponde a la muestra de DDX3 con extracto de ARN

3.2.4. Determinación mediante proteólisis limitada de los dominios flexibles de DDX3

Una vez purificada la proteína DDX3 se intentó solventar el problema de solubilidad para poder concentrar la muestra a más de 1mg/ml. Para ello se utilizaron detergentes como NP40 o n-octilglucósido y se probó a añadir un análogo no hidrolizable de ATP (AMP-PNP) hasta una concentración de 2mM. Ninguna de las pruebas resultó satisfactoria y la concentración de la muestra no llegó a superar 1.2mg/ml. A pesar de estas limitaciones se realizó un cribado usando alrededor de 1000 condiciones de cristalización pero el análisis no deparó resultados positivos.

Antes de decidirnos a trabajar con los fragmentos más pequeños y carentes de actividad helicasa, DDX3_{ATP} y DDX3_{hel}, optamos por realizar un estudio de proteólisis limitada de la proteína tratando de encontrar un fragmento de DDX3 que fuera más estable y que conservara la actividad helicasa. Para obtener este fragmento (DDX3_{pf}) resistente a proteasas se realizó un estudio de proteólisis controlada de DDX3 (figura 3.10) con tripsina, quimotripsina, elastasa, V8, subtilisina y Asp-N, que dio como resultado un fragmento

mayoritario de unos 55kDa (fragmento 2), otro un poco mayor de unos 60kDa (fragmento 1) y otro un poco más pequeño de unos 50kDa (fragmento 3). Los tres fragmentos fueron analizados por espectrometría de masas y degradación de Edman para determinar la secuencia exacta de los aminoácidos que formaban los fragmentos. El estudio de espectrometría de masas reveló que el extremo C-terminal de los tres fragmentos estaba en el aminoácido 600 de DDX3. El análisis por degradación de Edman del extremo N-terminal mostró un fragmento que empezaba en un residuo de serina en la posición 86 de la proteína (fragmento 1) y otro que empezaba en un residuo de aspártico en la posición 164 (fragmento 3). Del fragmento mayoritario (fragmento 2) no se pudo determinar el extremo N-terminal posiblemente por no ser un fragmento homogéneo. Este fragmento mayoritario debería tener su extremo N-terminal entre los aminoácidos 86 y 164 de DDX3.

Basándonos en el análisis previo del alineamiento de secuencias (figura 1.7) observamos cómo el fragmento de DDX3, con los amino ácidos 162-600, contiene todos los motivos conservados y teóricamente debería conservar su actividad helicasa dependiente de la hidrólisis de ATP.

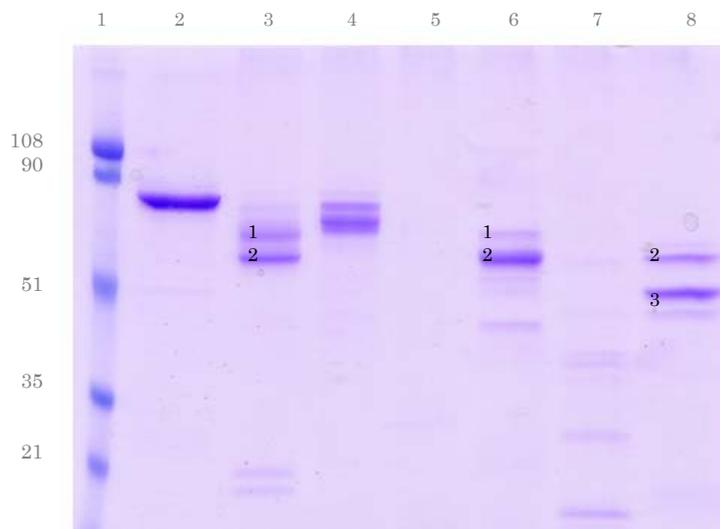


Figura 3.10. Ensayo de proteólisis controlada de DDX3. La muestra de DDX3 (calle 2) fue digerida toda la noche a temperatura ambiente con tripsina, elastasa, subtilisina, chimotripsina, V8 y Asp-N (calles 3-8 respectivamente) y analizada mediante un gel SDS-PAGE al 10%. Los marcadores de pesos moleculares vienen indicados en la calle 1. Los fragmentos descritos en el texto como 1, 2 y 3 están marcados en el gel.

3.2.5. Expresión y purificación de DDX3_{pf}

A la proteína DDX3_{pf} (51kDa) se le añadió en el clonaje, mediante los cebadores directo y reverso utilizados en la amplificación por PCR dos secuencias, una de 6 histidinas en el extremo N-terminal de la proteína y una secuencia STREP en el extremo C-terminal, para optimizar la purificación de DDX3_{pf} y evitar cualquier posible degradación. DDX3_{pf} se sobreexpresó en células de *E. coli* de la cepa BL21 (DE3). Utilizando el mismo tampón de resuspensión que para la proteína silvestre se solubilizó alrededor del 30% de la muestra. El lisado se cargó en una columna de afinidad TALON. Posteriormente se cargaron las fracciones que contenían la proteína en una columna de afinidad de STREPtactin dejando una muestra con la proteína pura al 99% según la estimación mediante un gel SDS-PAGE teñido con azul de Coomassie (figura 3.11). Tras una diálisis extensiva para eliminar el EDTA del tampón de la muestra, la proteína fue cargada de nuevo en una columna de afinidad, esta vez una HisTrap de 1ml. Esta columna nos permitió concentrar la proteína a 1.5mg/ml y a una pureza estimada mediante un gel de SDS- PAGE teñido con plata del 99% (figura 3.11).

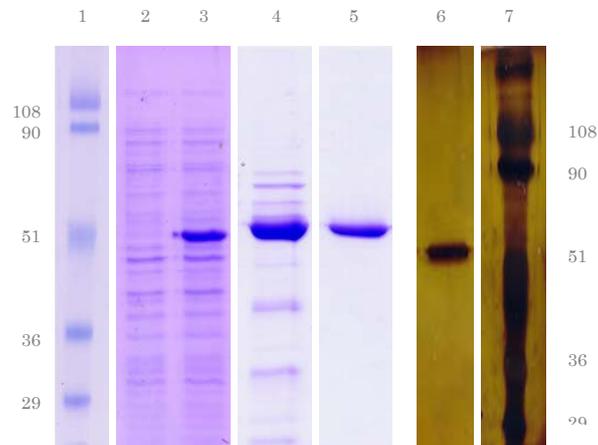


Figura 3.11. Expresión y purificación de DDX3_{pf}. Las calles 2 y 3 corresponden a una muestra de los cultivos celulares antes y después de la inducción de la proteína con IPTG. Las dos calles siguientes corresponden a la purificación de la proteína, la muestra después de la columna de afinidad de TALON (calle 4) y después de la columna de STREPtactin (calle 5). Tras dializar la muestra toda la noche se pasó por una columna de HisTrap y se analizó su pureza con un gel SDS-PAGE teñido con plata (calle 6). Las calles 1 y 7 corresponden a los marcadores de pesos moleculares.

3.2.6. Estudios biofísicos y bioquímicos de DDX3_{pf}

3.1.6.1. DDX3_{pf} está plegada, presenta un estado homogéneo de oligomerización y conserva la actividad helicasa asociada a la hidrólisis de ATP

El análisis por medio de DC (figura 3.12a) y la desnaturalización térmica (figura 3.12b) de DDX3_{pf} muestran que la proteína conserva su estructura secundaria de la misma manera que la proteína completa DDX3. El estudio de la muestra proteica muestra un único estado de oligomerización correspondiente a una especie monomérica de 64kDa, con un bajo índice de polidispersidad, inferior al 15% (figura 3.13). La actividad helicasa de la proteína fue ensayada igual que en el caso de la proteína silvestre monitorizando la oxidación del NADH con un espectrofotómetro. Se observó que la actividad se mantenía dependiente de ARN y que la cinética de la reacción arrojaba unos valores de V_{max} de 0.14 μmol de ATP min⁻¹ mg⁻¹, que corresponde a 7.2 μmol de ATP min⁻¹ μmol⁻¹ de DDX3_{pf} (figura 3.14).

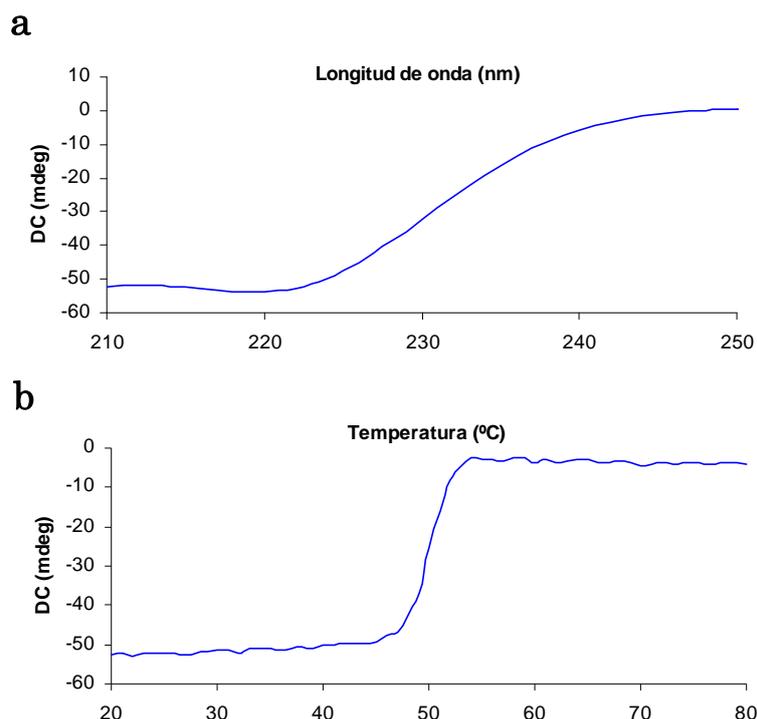
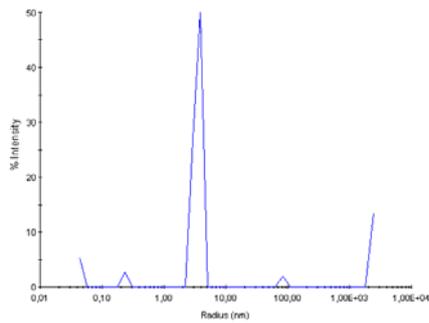


Figura 3.12. Análisis de dicroísmo circular de DDX3_{pf}. a) El espectro de DC se realizó de 250nm a 208nm y muestra una clara presencia de estructura secundaria con un mínimo a 220nm. b) La desnaturalización térmica de la proteína se realizó variando la temperatura entre 20° C y 80° C. La longitud de onda se mantuvo constante a 220nm y el resultado muestra una curva altamente cooperativa, con una T_m de 53° C.



	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
Pico 1	0.0	0.0	0	5.2	99.7
Pico 2	0.2	0.0	0	2.6	0.3
Pico 3	3.5	12.7	64	76.8	0.0

Figura 3.13. Análisis de DLS de DDX3_{pf}. La proteína muestra un índice de polidispersidad muy bajo resaltando su carácter de molécula unitaria dentro de la muestra y su falta de agregación. El peso molecular estimado es de 64kDa.

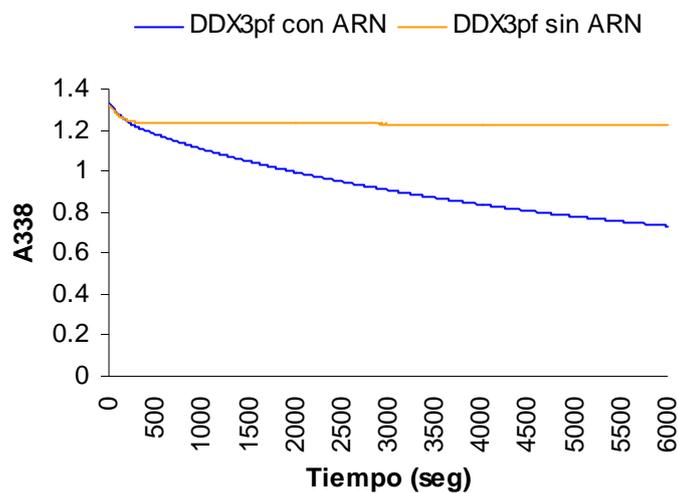


Figura 3.14. Actividad helicasa de DDX3_{pf} dependiente de la hidrólisis de ATP. La actividad helicasa de DDX3_{pf} fue medida de igual manera que la actividad de DDX3. La cinética de la actividad ATPasa se midió con un espectrofotómetro en presencia (azul) y ausencia (naranja) de ARN. La oxidación del NADH, directamente proporcional a la hidrólisis de ATP, fue monitorizada de manera constante midiendo la absorbancia de la muestra a 338nm. La reacción se llevó a cabo a 37° C.

DDX3_{pf} planteó los mismos problemas de solubilidad que la proteína silvestre. Todos los ensayos llevados a cabo hasta este momento se habían realizado a una concentración que no superaba 1.5mg/ml. Cuando se intentó concentrar más la muestra, la proteína precipitaba haciendo imposible aumentar su concentración. A pesar de todo se llevaron a cabo numerosos cribados de cristalización que, al igual que en el caso de DDX3, no depararon resultados positivos. Mientras se realizaron estos ensayos se pasó a trabajar con el fragmento definido por PONDR, DDX3_{pd}, que mantiene prácticamente intacto el extremo N-terminal de la proteína (164 de DDX3_{pf} frente a 169 de DDX3_{pd}) pero reduce en 21 aminoácidos el extremo C-terminal (600 de DDX3_{pf} frente a 579 de DDX3_{pd}). De este modo se intentaron reducir las regiones supuestamente desordenadas causantes de este problema de agregación, presente tanto en DDX3 como en el fragmento DDX3_{pf}.

3.2.7. Expresión y purificación de DDX3_{pd}

Al igual que para el caso de DDX3_{pf} y dado que la purificación de este fragmento se realizó con éxito tanto en rendimiento como en pureza, el esquema utilizado para expresar y purificar DDX3_{pd} (49kDa) fue similar. Se le añadieron dos secuencias a la proteína para facilitar su purificación, aunque en esta ocasión se cambiaron de posición. La secuencia de STREP se añadió en el extremo N-terminal y las seis histidinas se situaron en el extremo C-terminal de la proteína. Como el extremo C-terminal de la construcción terminaba en 2 histidinas sólo fue necesario añadirle 4 más. Al igual que en los casos anteriores, en esta purificación se probó primero a purificar la proteína con un tampón de resuspensión que contenía sulfato amónico y los resultados obtenidos fueron similares a los conseguidos anteriormente para DDX3 y DDX3_{pf} en lo que a solubilidad se refiere. Sin embargo, utilizando un tampón con una alta concentración de cloruro sódico en vez de sulfato amónico la proteína permaneció en solución a concentraciones elevadas. Así, el sulfato amónico se sustituyó en el tampón por NaCl a una concentración 1M y se llevó a cabo la purificación.

Un precipitado correspondiente a 4.5L de cultivo se resuspendió en 50mM Tris, 1M NaCl pH8.0. Se rompieron las células por sonicación, y después de ultracentrifugar la muestra, la fracción soluble con la proteína se cargó en una columna de afinidad con resina TALON. Las fracciones que contenían

DDX3_{pdr} se cargaron en una columna de streptactin. La secuencia de STREP fue eliminada tras incubar la muestra con la proteasa TEV durante una noche. Finalmente se cargó la muestra en una columna de filtración en gel Hi Load 26/60 superdex 200 para separar DDX3_{pdr} de la proteasa TEV y dejar la proteína en el tampón adecuado para los análisis biofísicos y los ensayos de cristalización (figura 3.15). El rendimiento aproximado del proceso de purificación fue de 1mg de proteína pura por litro de cultivo. La muestra se concentró con éxito hasta 10mg/ml y se llevaron a cabo numerosos ensayos de cristalización cubriendo alrededor de 1000 condiciones diferentes.

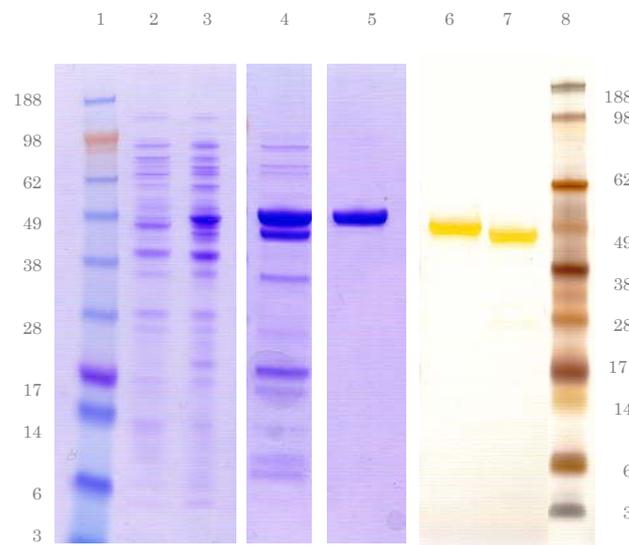


Figura 3.15. Expresión y purificación de DDX3_{pdr}. Gel de SDS-PAGE del 10% teñido con coomassie (calles 1-5) y con plata (calles 6-8). Las calles 2 y 3 corresponden a una muestra de los cultivos celulares antes y después de la inducción de la proteína con IPTG. La calle 4 es la muestra después de la primera columna de afinidad. Las calles 5 y 6 corresponden a la muestra después de la columna de afinidad de streptactin teñidas con Coomassie y plata respectivamente. La calle 7 es la proteína tras el procesamiento proteolítico con TEV y una columna de filtración en gel. Los marcadores se muestran en la calles 1 y 8.

3.2.8. Estudios biofísicos y bioquímicos de DDX3_{pdr}

3.2.8.1. DDX3_{pdr} está plegada, presenta un estado homogéneo de oligomerización y conserva la actividad helicasa asociada a la hidrólisis de ATP

Al igual que en los casos anteriores, el análisis por medio de DC (figura 3.16a) y la desnaturalización térmica (figura 3.16b) de DDX3_{pdr} muestran que la proteína conserva su estructura secundaria. El estudio de DLS muestra un único estado de oligomerización correspondiente a una especie monomérica de 66kDa, con un bajo índice de polidispersidad, 6.1% (figura 3.17).

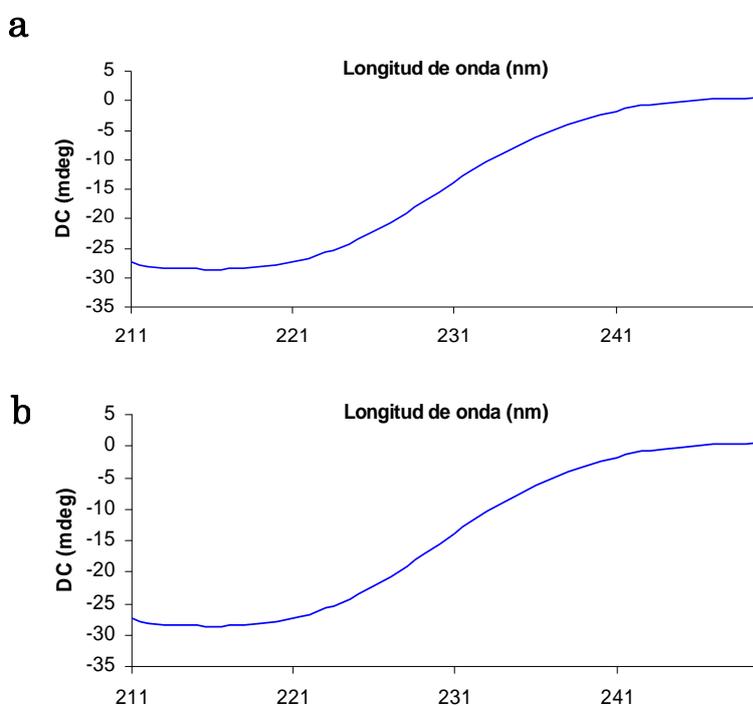
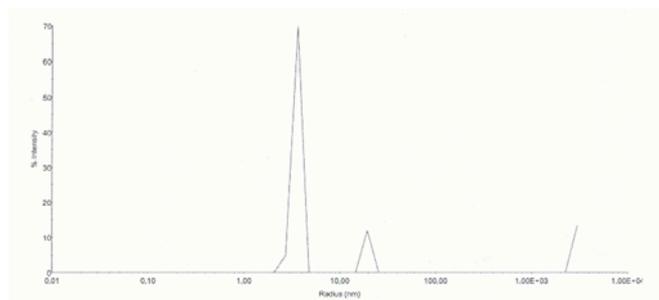


Figura 3.16. Análisis de dicroísmo circular de DDX3_{pdr}. a) El espectro de DC se realizó de 250nm a 208nm y muestra que la proteína presenta estructura secundaria. b) La desnaturalización térmica de la proteína se realizó variando la temperatura entre 20° C y 80° C. La longitud de onda se mantuvo constante a 220nm y el resultado muestra una curva altamente cooperativa, con una T_m de 53° C. Ambas medidas, tanto del espectro de DC, como de la desnaturalización térmica, se llevaron a cabo en una cubeta de 0,2cm de paso de luz.



	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
Pico 1	3.6	6.1	66	74.7	99.9
Pico 2	19.5	0.0	3498	11.9	0.1
Pico 3	2991.8	0.0	457280000	13.4	0.0

Figura 3.17. Análisis de DLS de DDX3_{pdr}. La proteína muestra un índice de polidispersidad muy bajo (6.1%) resaltando su carácter monomérico (66kDa) dentro de la muestra y su falta de agregación.

La actividad helicasa de la proteína fue comprobada con el mismo ensayo que el realizado para la proteína silvestre, monitorizando la oxidación del NADH espectrofotométricamente. Se observó que la actividad se mantenía dependiente de ARN y que la cinética de la reacción arrojaba unos valores de V_{max} de $0.38 \mu\text{mol de ATP min}^{-1} \text{mg}^{-1}$, que corresponde a $18.7 \mu\text{mol de ATP min}^{-1} \mu\text{mol}^{-1}$ de DDX3_{pdr} (figura 3.18). Un resumen de la actividad enzimática de DDX3, DDX3_{pf} y DDX3_{pdr} se puede observar en la tabla 3.1.

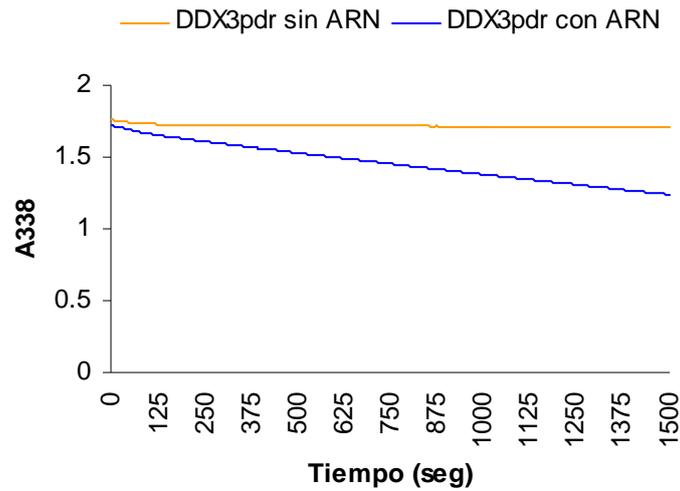


Figura 3.18. Actividad helicasa de DDX3_{pdr} dependiente de la hidrólisis de ATP. La actividad helicasa de DDX3_{pdr} fue medida de igual manera que la actividad de DDX3 y DDX3_{pf}. Se midió con un espectrofotómetro la cinética de la actividad ATPasa de la proteína en presencia y ausencia de ARN. La oxidación del NADH, directamente proporcional a la hidrólisis de ATP, fue monitorizada de manera constante midiendo la absorbancia de la muestra a 338nm. La reacción se llevó a cabo a 37° C. En naranja se indica la cinética de la reacción sin ARN. En azul se muestra la cinética de la reacción con extracto de ARN.

Tabla 3.1. Resumen de la actividad helicasa de DDX3, DDX3_{pf} y DDX3_{pdr}

	$\mu\text{mol ATP} \times \text{min}^{-1} \times \text{mg}^{-1} \text{ proteína}$	$\mu\text{mol ATP} \times \text{min}^{-1} \times \text{ }^{-1} \mu\text{mol proteína}$
DDX3	0.35	25.8
DDX3_{pf}	0.14	7.2
DDX3_{pdr}	0.38	18.7

3.2.9. Expresión y purificación de DDX3_{hel}

La expresión del fragmento DDX3_{hel} se llevó a cabo en paralelo con la expresión de DDX3_{ATP}. Ambas proteínas recombinantes tenían en el extremo N-terminal una secuencia de seis histidinas para su purificación. Antes de realizar la purificación de ambas proteínas a gran escala se realizaron tanto pruebas de expresión como pruebas preliminares de

purificación usando para ello unas columnas con resina de afinidad para histidinas Ni-NTA Spin Columns. Para esta purificación preliminar se utilizaron dos tampones distintos, uno con 1M NaCl y otro con 0.5M $(\text{NH}_4)_2\text{SO}_4$, ambos con 50mM Tris, pH 8.0. DDX3_{hel} fue la única construcción soluble usando el tampón con 0.5M $(\text{NH}_4)_2\text{SO}_4$ (figura 3.19). Por lo tanto, procedimos a la purificación de este fragmento aprovechando la ya mencionada secuencia de histidinas en el extremo N-terminal de la proteína seguida de un sitio de corte TEV para la posterior eliminación de dicha secuencia.

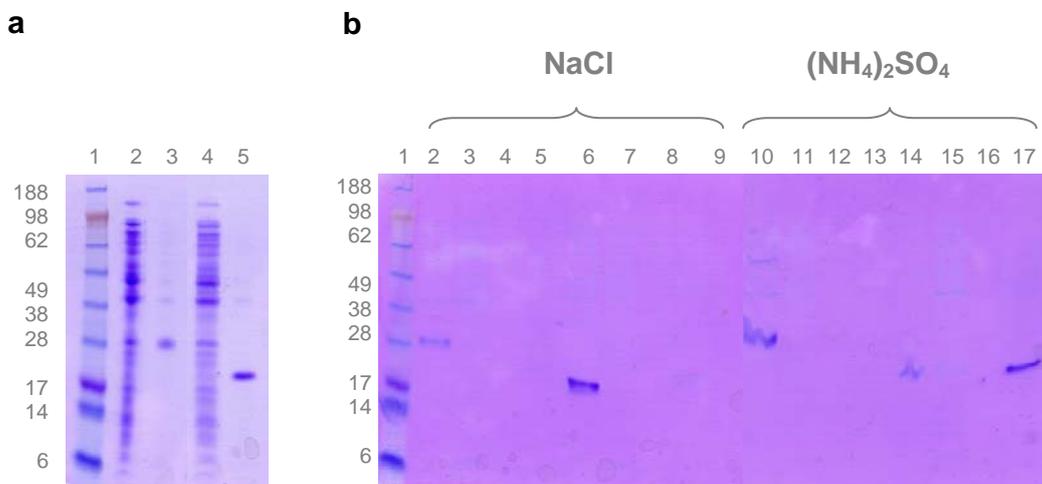


Figura 3.19. Expresión y ensayo de purificación de DDX3_{ATP} y DDX3_{hel}. (a) Expresión de DDX3_{ATP} y DDX3_{hel}. Las calles 2 y 3 corresponden a DDX3_{ATP} antes y después de la inducción con IPTG. Las calles 4 y 5 corresponden a DDX3_{hel} antes y después de la inducción celular. La calle 1 corresponde a los marcadores de peso molecular. (b) Las calles 2-9 son las purificaciones con NaCl y las calles 10-17 son las purificaciones con $(\text{NH}_4)_2\text{SO}_4$. Las calles 2-5 corresponden a DDX3_{ATP}. La calle 2 corresponde al precipitado, la calle 3 es la muestra que pasa por la columna, la calle cuatro corresponde al lavado con imidazol y la calle 5 es la elución de la proteína. Este orden (precipitado-sobranante-lavado-elución) se mantiene en las calles 6-9 que corresponden a la proteína DDX3_{hel}. También se mantiene el orden en las calles siguientes. Las calles 10-13 corresponden a la proteína DDX3_{ATP} y las calles 14-17 corresponden a DDX3_{hel}.

La purificación de DDX3_{hel} se realizó usando una columna His Trap de 5ml. La sobreexpresión de este fragmento fue muy elevada. 2L de cultivo fueron suficientes para obtener una saturación completa de la columna, consiguiendo aproximadamente 200mg de proteína pura en un solo paso de purificación. Posteriormente se realizó la digestión de la proteína con TEV y se volvió a cargar la muestra en la columna de afinidad, retirando así de la muestra tanto la proteasa TEV, que se asocia a la resina con su propia secuencia de histidinas, como DDX3_{hel} sin cortar. Para obtener una mayor pureza de la muestra y conservar la proteína en el tampón adecuado para los ensayos de cristalización se cargaron las fracciones de DDX3_{hel} en una columna de filtración en gel Hi Load 26/60 Superdex75 y se concentró la muestra por ultrafiltración hasta 10mg/ml (figura 3.20).

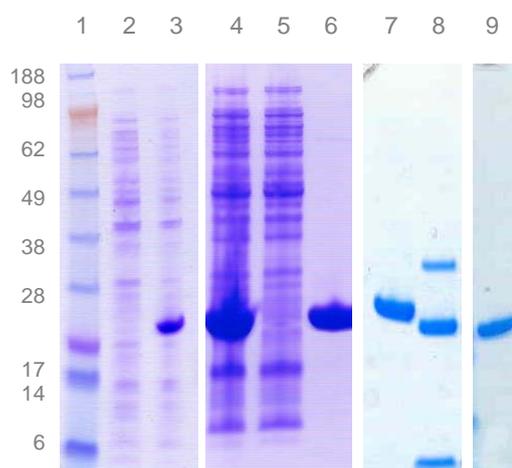


Figura 3.20. Expresión y purificación de DDX3_{hel}. Las calles 2 y 3 corresponden a una muestra de los cultivos celulares antes y después de la inducción de la proteína con IPTG. La calle 4 corresponde a la muestra cargada en la columna de HisTrap. La calle 5 es el conjunto de proteínas que no presentaron afinidad por la resina y la calle 6 corresponde a la fracción de DDX3_{hel} eluída de la columna de afinidad. La calle 7 es la muestra como control negativo del corte proteolítico con TEV cuyo resultado se ve en la calle 8. La calle 9 corresponde a la muestra de proteína una vez cargada en la columna de filtración en gel y concentrada.

3.2.10. Estudios biofísicos de DDX3_{hel}

3.2.10.1. DDX3_{hel} está plegada y presenta un estado homogéneo de oligomerización.

El análisis por medio de DC y la desnaturalización térmica (figura 3.21a) de DDX3_{hel} muestran que la proteína conserva su estructura secundaria al igual que DDX3. La T_m de este fragmento es 10° C más elevada que en el caso de DDX3 mostrando así una estructura más compacta que la proteína silvestre.

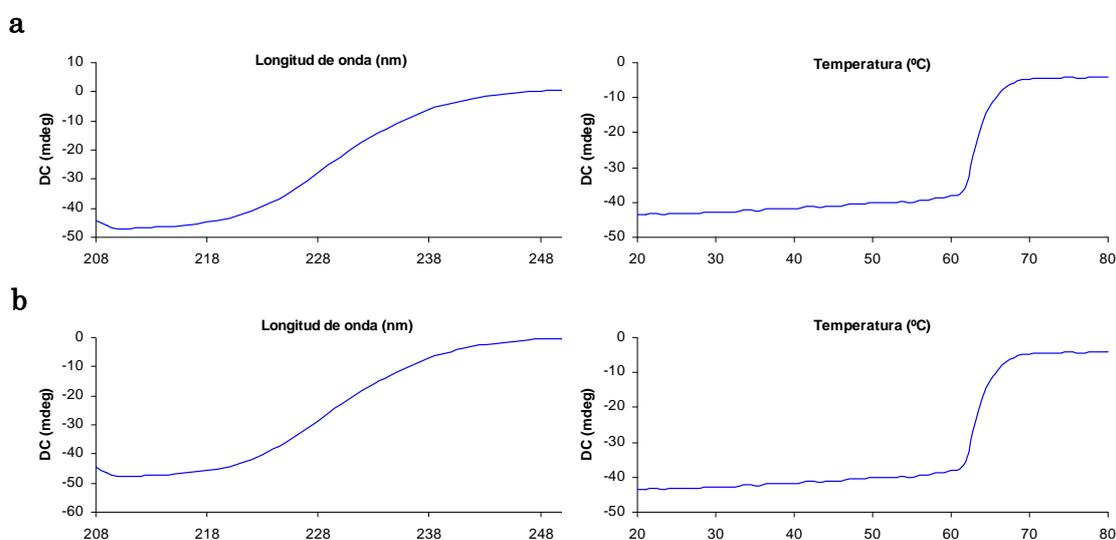
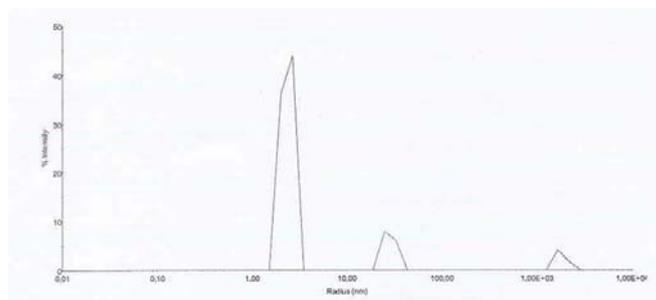


Figura 3.21. Análisis de dicroísmo circular de DDX3_{hel} antes (a) y después (b) de congelar la muestra a 77K. En los gráficos de la izquierda se observan mínimos a 220nm y 210nm señalando la presencia de estructura secundaria. La T_m de la desnaturalización térmica es de 65° C.

El estudio de DDX3_{hel} por DLS muestra un único estado de oligomerización correspondiente a una especie monomérica de unos 25kDa, con un bajo índice de polidispersidad, inferior al 15% (figura 3.22).



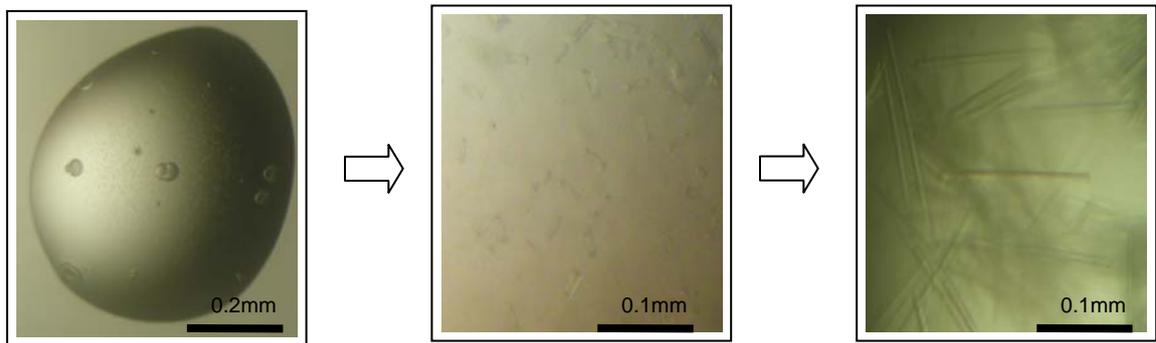
	R (nm)	%Pd	MW-R (kDa)	%Int	%Mass
Pico 1	2.4	13.7	25	80.3	100.0
Pico 2	28.4	14.0	8468	14.0	0.0
Pico 3	1806.3	13.4	1404150000	5.8	0.0

Figura 3.22. Análisis de DLS de DDX3_{hel}. Un pico mayoritario de 25kDa y un índice bajo de polidispersidad (13.7%) señalan una muestra homogénea y monomérica.

3.2.11. Cristalización de DDX3_{hel}

Se realizaron unos primeros ensayos haciendo una criba de alrededor de 1000 condiciones de cristalización. Algunas de los ensayos fueron muy prometedores presentando pequeñas esferulitas sobre las que refinar las condiciones. Después de varias rondas de optimización variando las concentraciones de precipitante, solución tampón y rango de pH se consiguió obtener cristales de tamaño y calidad suficientes para llevar a cabo experimentos de difracción (figura 3.23).

a



b

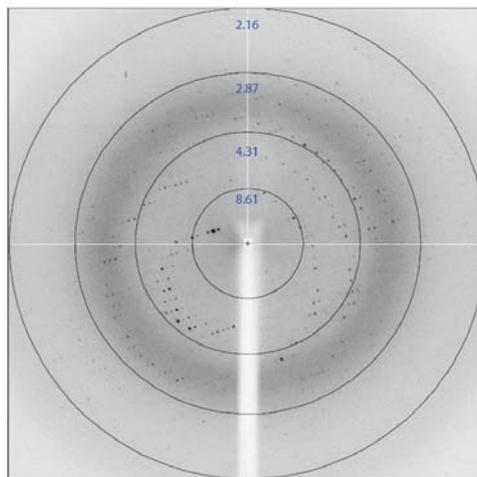


Figura 3.23. Proceso de cristalización de DDX3_{hel} e imagen de difracción. a) En el panel de la izquierda se muestran las esferulitas iniciales. En el panel central se muestran pequeños cristales, todavía de tamaño insuficiente para realizar experimentos de difracción. En el panel de la derecha se observan los cristales ya mejorados que sirvieron para resolver la estructura cristalográfica y con los que se obtuvo la imagen de difracción mostrada en (b). Los círculos indican los límites de resolución en ångstroms (Å).

La [tabla 3.2](#) muestra un resumen de los datos obtenidos con las diferentes construcciones de DDX3.

Tabla 3.2. Datos experimentales de las construcciones realizadas a partir de DDX3

Proteína	Expresión	Solubilidad NaCl	Solubilidad (NH ₄) ₂ SO ₄	[] mg ml ⁻¹	DC	DLS	Actividad	Cristal
DDX3	+	-	+	1.5	+	+	+	-
DDX3 _{pf}	+	-	+	1.5	+	+	+	-
DDX3 _{pdr}	+	+	-	10	+	+	+	-
DDX3 _{ATP}	+	-	-	-	-	-	-	-
DDX3 _{hel}	+	-	+	10	+	+	-	+

3.2.12. Resolución de la estructura de DDX3_{hel}

Para resolver la estructura atómica de DDX3_{hel} se recogieron varios grupos de datos usando radiación sincrotrón a una longitud de onda de 0.9 Å. El que presentó mejores estadísticas se utilizó para resolver el problema de fases. Para ello se utilizó el método del reemplazo molecular utilizando el dominio helicasa de la proteína Vasa de *D. Melanogaster* que presentaba una identidad de secuencia de acuerdo al programa DALI (Holm 2000) del 53% con respecto a DDX3_{hel}. Los aminoácidos diferentes fueron sustituidos en la secuencia de Vasa por alaninas para el primer intento de solución de fases. En el refinado final se aplicaron restricciones TLS. Un resumen de los datos cristalográficos está recogido en la [tabla 3.3](#).

Tabla 3.3. Estadísticas de recolección de datos y refinado

Recolección de datos^a	
Entorno	Detector ADSC-Q315r, ESRF, línea ID23-1
Longitud de onda	0.9 Å
Dimensiones de la celda (Å, °), grupo espacial P2 ₁	a=44.43 Å, b= 61.03 Å, c= 89.11 Å, β= 101.67°.
Resolución (Å)	36.1-1.9
Reflexiones únicas	36670
Multiplicidad media	3.1 (2.1)
% Reflexiones totales	93.2 (72.2)
R_{merge} ^b	0.081 (0.332)
$\langle I/\sigma I \rangle$	11.6 (2.5)
Refinado	
Número de reflexiones (%)	32363 (93.01)
Rango de resolución (Å)	36.1 (1.9)
R-factor/R-free (%)	19.8/25.5
Número de átomos de proteína (Medio B, Å ³) ^c	3725 (26.45)
Número de moléculas de agua (Medio B, Å ³) ^c	193 (39.14)
r.m.s. longitud de enlace (Å)	0.019
r.m.s. ángulo de enlace (°)	1.682
Inválidos del gráfico de Ramachandran (número) ^d	0

^a Los valores de la celda de máxima resolución se dan en paréntesis

^b $R_{merge} = \sum_{\eta} \sum_i |I_{\eta,i} - \langle I_{\eta} \rangle| / \sum_{\eta} \sum_i I_{\eta,i}$

^c Calculado con MOLEMAN

^d Calculado con PROCHECK

3.2.13. Descripción general de la estructura de DDX3_{hel}

La estructura de DDX3_{hel} conserva la estructura de otros dominios helicasa resueltos anteriormente. Tiene una estructura de 6 hojas β paralelas formando una lámina β que gira sobre si misma 90° y que se encuentra flanqueada de manera casi simétrica por 6 α hélices completando así una estructura altamente compacta tal y como se aprecia en la [figura 3.24](#). Las hojas β1, β2 y β3 se conectan entre sí intercalándose con las hélices α1, α2 y α3. Las hojas β4 y β5 continúan, unidas por un loop, hasta la hélice α4, la hoja β6 y finalmente las hélices α5 y α6 que marca el extremo C-terminal de la construcción. El primer y último residuo observados son Gly409 y Tyr576, no pudiendo definir por falta de densidad electrónica los residuos Arg407 y Val408 en el extremo N-terminal y Glu577 e His578 en el extremo C-terminal, que sí estaban presentes en la construcción de la proteína. Además, nuestro modelo presenta dos pequeños loops entre las regiones

Gly439-Ser442 y Ala501-Ile507 que no pudieron ser modelados por falta de densidad electrónica.

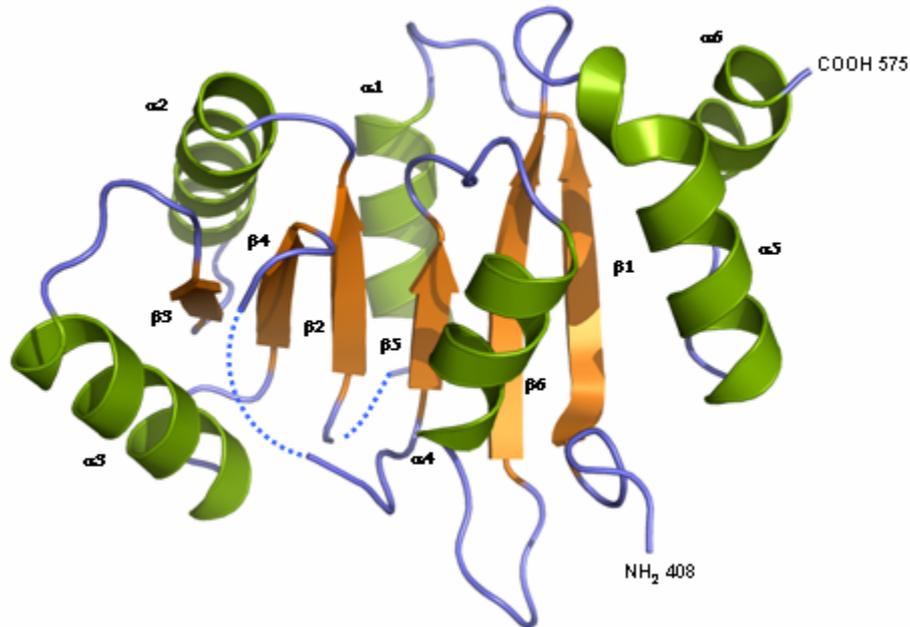


Figura 3.24. Diagrama de lazos de la proteína DDX3_{hel}. Se observan las seis hojas β paralelas formando la lámina central (naranja) rodeada por seis hélices α (verde) formando una estructura compacta. Punteados se señalan dos loops que no se pudieron construir por la ausencia de densidad electrónica.

La estructura de DDX3_{hel} es muy similar a la estructura de la helicasa de la familia DEAD box, Vasa (figura 3.25). Los elementos secundarios están esencialmente conservados siendo la desviación cuadrática media (r.m.s.d.) de los C α de una con respecto a la otra de 1.4Å según el programa DALI (Holm 2000). Las diferencias que se observan son variaciones en los loops que conectan α 1 con β 2 y α 4 con β 6.

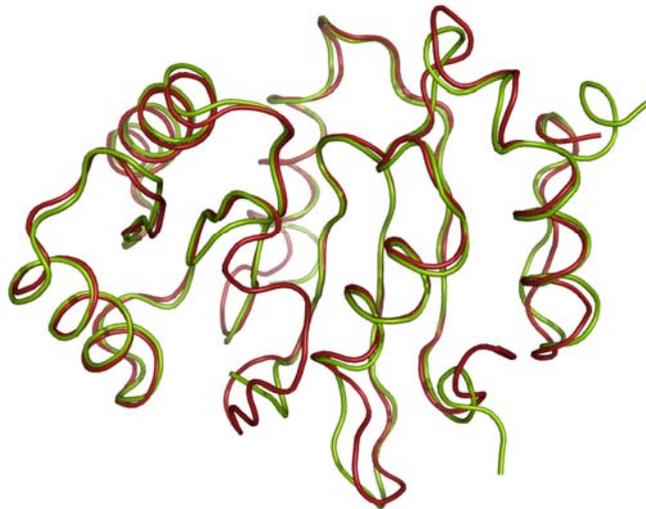


Figura 3.25. Superposición de DDX3_{hel} y el dominio helicasa de la proteína Vasa. La estructura de ambas proteínas es muy similar. En verde está representada DDX3_{hel} y en rojo Vasa. Pequeñas variaciones se observan en los extremos N- y C-terminal de ambas proteínas así como en regiones de las proteínas que carecen de estructura secundaria.

Tomando como base la proteína Vasa, por medio del programa Swiss-Model (Guex 1997; Schwede 2003) realizamos un modelo de ambos dominios de DDX3. Posteriormente sustituimos el dominio helicasa modelado por nuestra estructura y alineamos el ARN resuelto en Vasa al lugar donde debería estar en nuestro modelo. La parte inferior de la lámina β central del dominio helicasa corresponde a la región de contacto con el dominio ATPasa de la proteína. En un lateral de esta zona, tal y como se aprecia en la [figura 3.26d](#), hay una región básica que corresponde al sitio de unión del ARN a la helicasa ([figura 3.27](#)). Otra región en la cara opuesta de la lámina β también mantiene un claro carácter básico ([figura 3.26b](#)), presentando regiones ácidas los extremos perpendiculares de la lámina β ([figura 3.26a y 3.26c](#)).

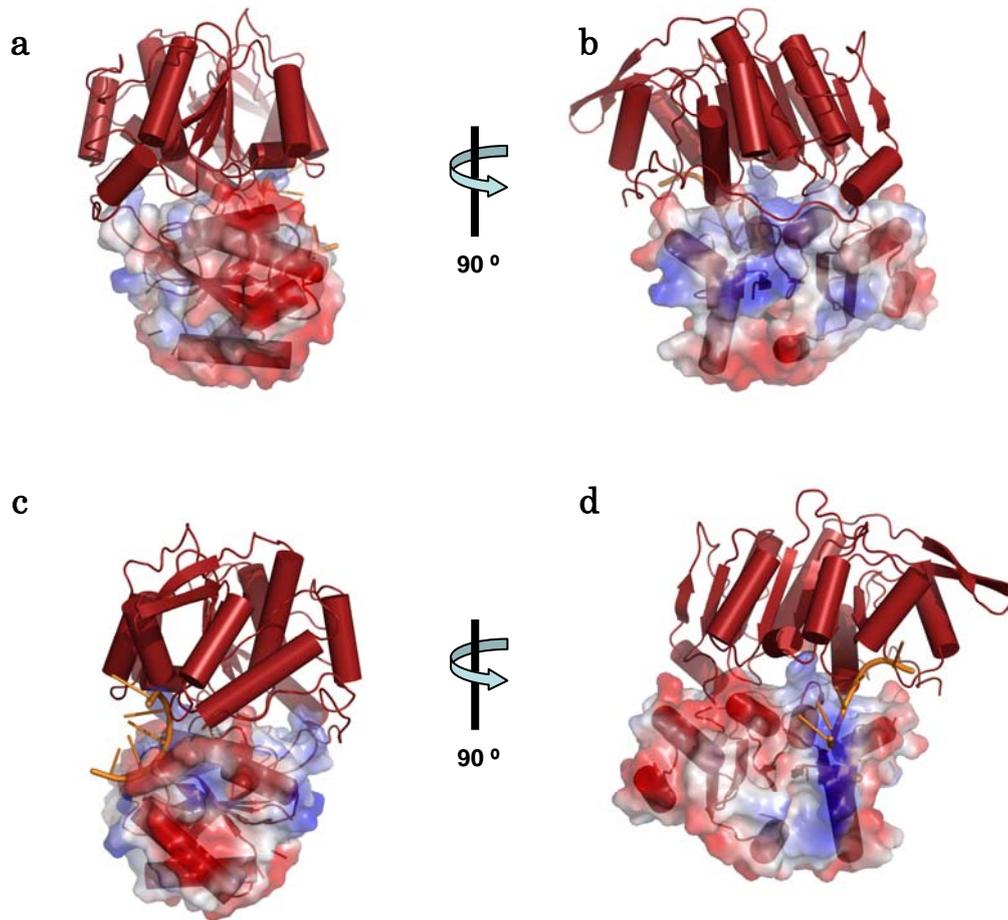


Figura 3.26. Potencial electrostático de DDX3_{hel} sobre el modelo obtenido. Se representan cuatro figuras que cubren 360° de la proteína. Las dos caras que se sitúan en el eje de giro de la lámina β (a y c) se componen principalmente de residuos ácidos mientras que las caras sobre las que se extiende la lámina β (b y d) son esencialmente básicas. La región básica que se observa en (d) es el lugar de unión de la proteína al ARN (naranja).

DDX3_{hel} contiene los motivos IV, V y VI tal y como se aprecia en la [figura 3.27](#) se aprecian los motivos dentro de la estructura y su disposición con respecto a la situación del ARN. Se aprecia como los dominios Ia y Ib interaccionan directamente con el ARN tal y como estaba descrito (Rocak 2004) relacionándose claramente con el motivo V del dominio helicasa. Sin embargo, la implicación del dominio IV en esta unión vendría dada por la

interacción de este motivo con el motivo V. El motivo VI está íntimamente ligado a los motivos de hidrólisis de ATP I y III corroborando este modelo los datos que indican que este motivo desempeña un papel importante en la unión del enzima a ATP.

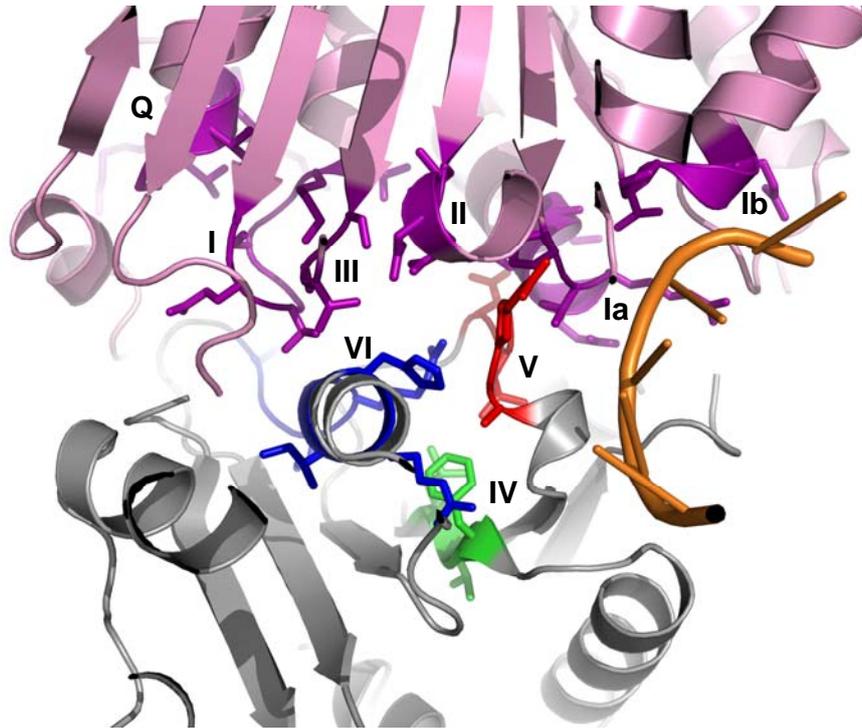


Figura 3.27. Diagrama de lazos del modelo de los dominios ATPasa y helicasa de DDX3 con los motivos conservados y el ARN en el punto de unión. Sobre el modelo realizado con Swiss-Model se han marcado los motivos conservados tanto en el dominio ATPasa (violeta) como en el dominio helicasa (verde, rojo y azul). El ARN se muestra en naranja.

Un detalle de los motivos IV, V y VI de DDX3_{hel} se aprecia en el mapa de densidad electrónica de la [figura 3.28](#) donde se observa más claramente la interrelación de los tres motivos.

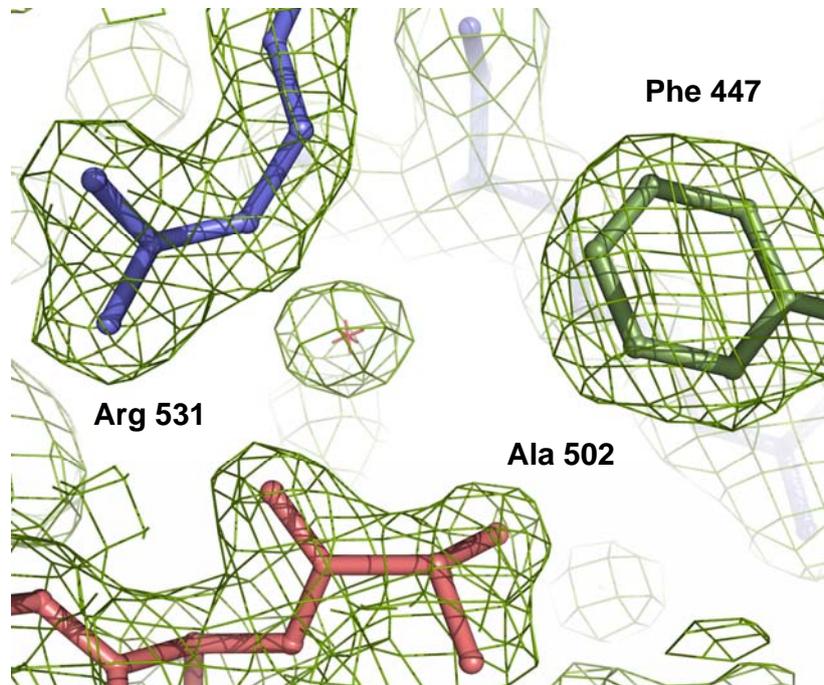


Figura 3.28. Mapa 2fofc a 1σ de la estructura DDX3_{hel}. Los motivos IV, V y VI se han coloreado en verde, rojo y azul respectivamente. Un aminoácido de cada motivo se marca como referencia de secuencia.

3.3. CD5

3.3.1. Expresión y purificación de CD5DIII

El estudio del receptor CD5 se centró en la región extracelular. Esta región se compone, como ya está descrito en la introducción, de tres dominios SRCR pertenecientes al grupo B. Los dominios I y II presentan N- y O-glicosilaciones, que plantean una dificultad añadida a la hora de abordar su cristalización. Como un primer intento se decidió, dada la complejidad del receptor entero, abordar el análisis de uno de los dominios de la región extracelular. El dominio III, carente de glicosilaciones, se reveló como el principal candidato para obtener proteína pura que pudiera ser cristalizada (McAlister 1998).

La expresión de proteínas humanas en sistemas de expresión heterólogos como *E. coli* se hace a menudo imposible debido a la ausencia de la maquinaria necesaria para el correcto plegamiento de la proteína. Así, la expresión de CD5DIII, dada la cantidad de puentes disulfuro que tienen que emparejarse, se realizó en células humanas que contienen esta maquinaria para obtener la proteína correctamente plegada. A pesar de todo, antes de abordar este sistema de expresión, sin duda el más complejo de los disponibles, se probó a expresar la proteína en otros sistemas más sencillos pero todos ellos fallaron, bien por ser insoluble (bacteria), bien por tener un plegamiento incorrecto (levadura) o incluso por tener un rendimiento prácticamente nulo (sistema de baculovirus en células de insecto).

Una vez expresada, para abordar el proceso de purificación de la proteína recombinante humana CD5DIII, dada la escasa cantidad de proteína de la que podíamos disponer, fue indispensable el uso de una columna de alta afinidad. En este caso se usó una columna con un anticuerpo monoclonal frente a este dominio de la proteína (83-C4 MAb). Esto nos permitió obtener en un solo paso proteína de una pureza de 99% estimado en Coomassie (figura 3.29). La muestra de proteína se concentró hasta 4mg/ml.

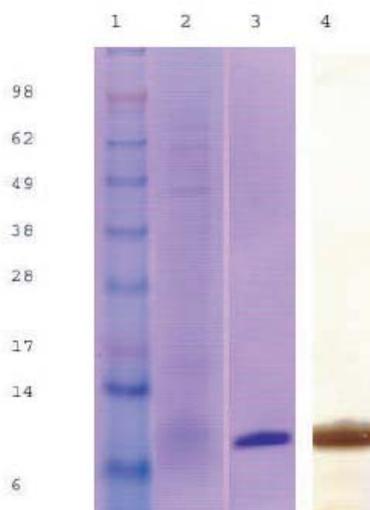


Figura 3.29. Expresión y purificación de CD5DIII. Geles de SDS-PAGE del 10% teñidos con azul de Coomasie (calles 1-3) o plata (calle 4). La calle 1 son los marcadores de peso molecular. En la calle 2 está el medio extracelular del cultivo de células HEK-293-EBNA-CD5DIII cargado en la columna con 83-C4 MAb, La calle 3 es la proteína una vez eluída de la columna y la calle 4, la proteína ya concentrada.

3.3.2. Análisis biofísicos de CD5DIII

Antes de llevar a cabo los ensayos de cristalización y una vez que tuvimos una muestra de proteína pura, a una concentración de 4mg/ml se realizaron análisis de espectrometría de masas y resonancia magnética nuclear que nos aportaron datos sobre el plegamiento de la estructura.

3.3.2.1. *CD5DIII recombinante mantiene intactos los puentes disulfuro descritos para este tipo de dominios SRCR.*

Para comprobar la correcta formación de puentes disulfuro dentro de la proteína se realizó un análisis por espectrometría de masas utilizando una muestra de CD5DIII purificada. Para ello se realizaron dos pruebas. En una se digirió la proteína con tripsina y se rompieron los puentes disulfuro con DTT. Esta muestra se usó como control negativo del análisis. En la segunda prueba, la proteína se digirió pero los fragmentos se dejaron intactos, de tal manera que su peso debería corresponder a la suma de los fragmentos unidos por puentes disulfuro. El resultado mostró que los puentes disulfuro estaban emparejados tal y como se describieron previamente en la literatura (Vila 2000). La [figura 3.30](#) muestra los puentes disulfuro hallados por el método del electrospray.

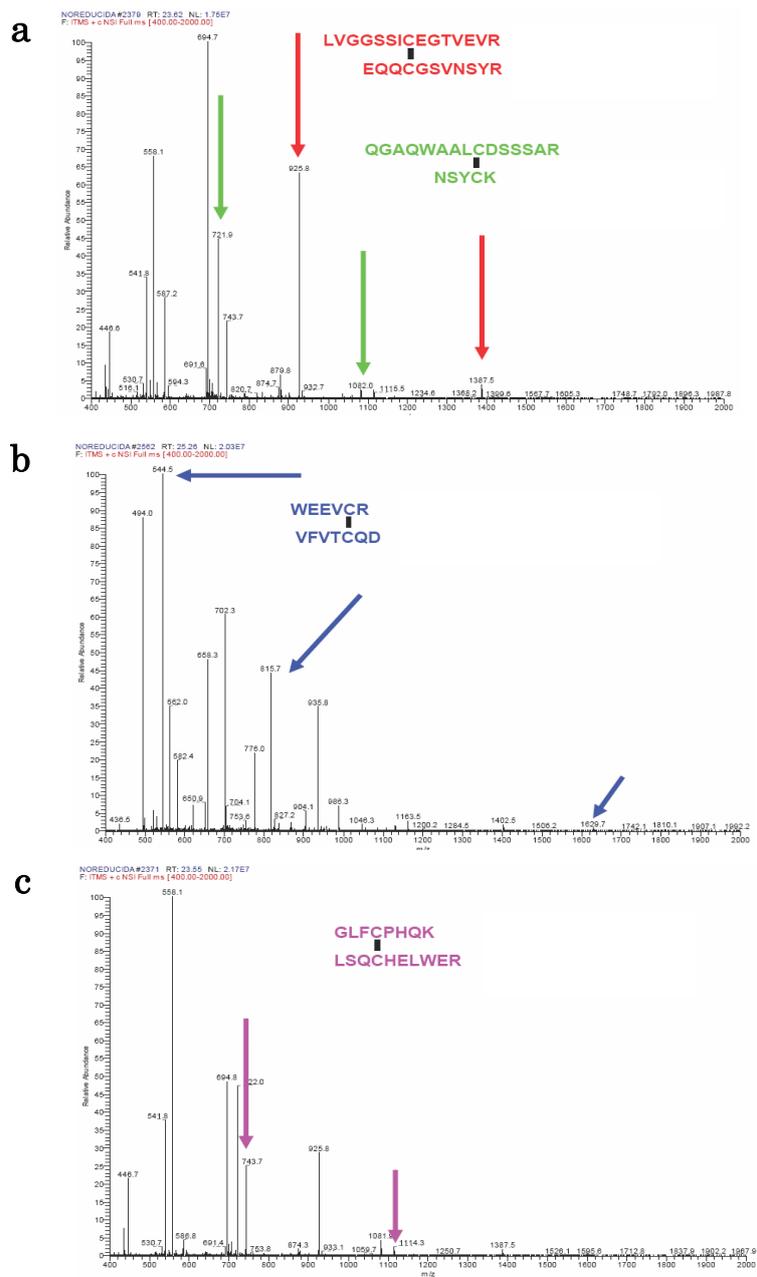


Figura 3.30. Espectro de los fragmentos ESI-TRAP mostrando radios m/z monoisotrópicos correspondientes a los dobletes de péptidos trípticos que tenían los puentes disulfuro. En (a) se detectaron dos dobletes diferentes de péptidos trípticos que co-eluyeron en el HPLC. Las flechas rojas corresponden a especies triple ($m/z = 925.8$) y doblemente ($m/z = 1387.5$) cargadas de el enlace disulfuro Cys285-Cys321 (C1-C4). Las flechas verdes corresponden a especies triple ($m/z = 721.9$) y doblemente ($m/z = 1082.0$) cargadas correspondientes al puente disulfuro Cys301-Cys360 (C2-C7). En (b) las flechas azules muestran las especies triplemente ($m/z = 544.5$), doblemente ($m/z = 815.7$) y de forma única ($m/z = 1629.7$) cargadas correspondientes al disulfuro Cys316-Cys367 (C3-C8). Las flechas violetas marcan en (c) las especies triple ($m/z = 743.7$) y doblemente ($m/z = 1114.3$) cargadas correspondientes al disulfuro Cys342-Cys350 (C5-C6)

La [figura 3.31](#) confirma el resultado obteniendo los puentes disulfuro por el método de MALDI-TOF.

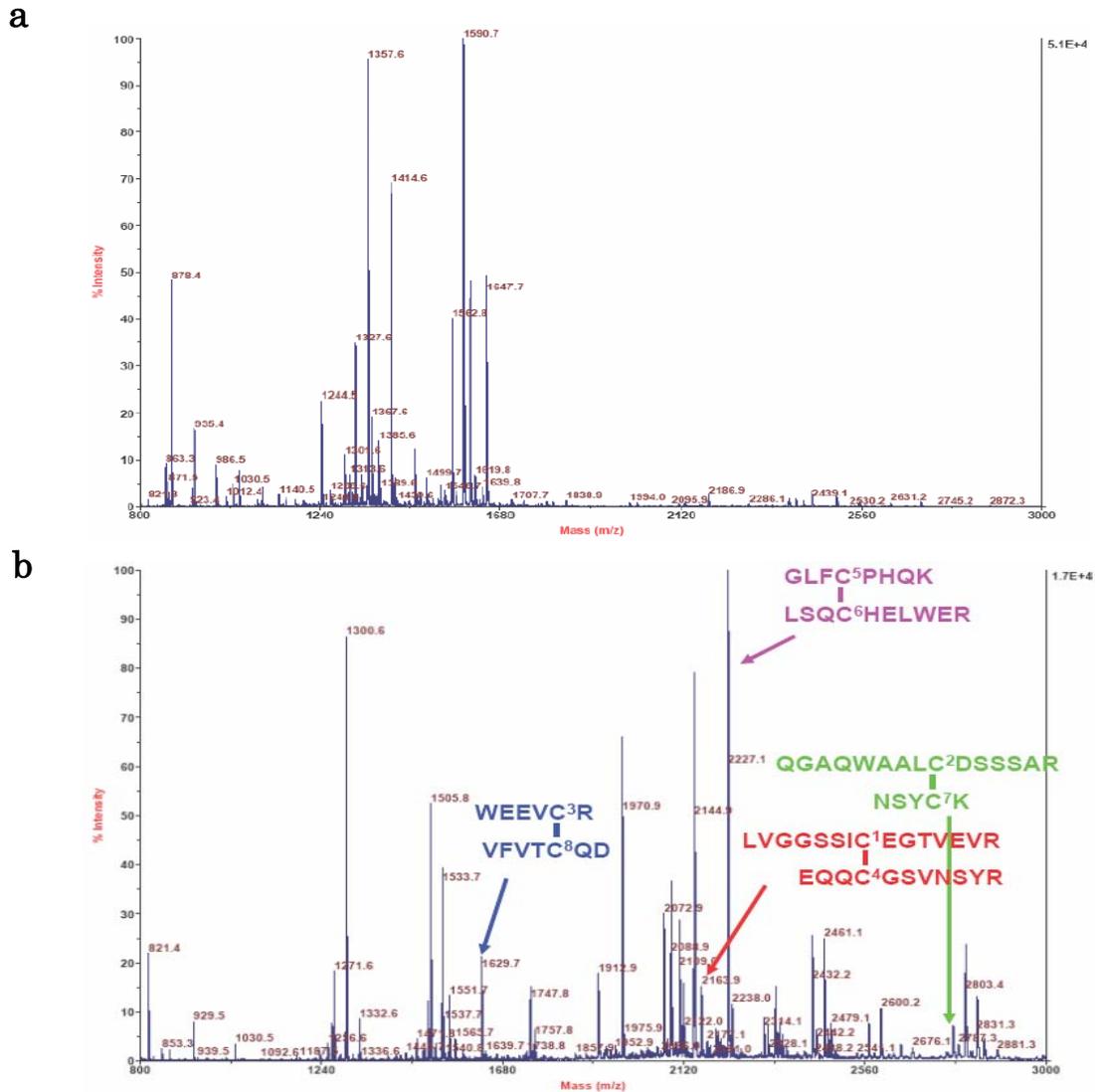


Figura 3.31. Espectro de MALDI-TOF de la digestión trípica de CD5DIII. El espectro en (a) incluyen los pasos de alquilación y reducción y sirve como control negativo. El espectro (b) no incluye estos pasos. Las flechas en (b) muestran los radios m/z de carga única correspondientes a los cuatro puentes disulfuro. La flecha roja ($m/z = 2773.3$) es Cys285-Cys321 (C1-C4), la verde ($m/z = 2161.9$) el disulfuro Cys301-Cys360 (C2-C7), la flecha azul ($m/z = 1629.7$) corresponde a Cys316-Cys367 (C3-C8) y la flecha violeta representa el disulfuro Cys342-Cys350 (C5-C6)

3.3.2.2. CD5DIII está plegado correctamente.

La muestra se sometió a un análisis por resonancia magnética nuclear para comprobar que la proteína estaba plegada. La dispersión de las señales en un espectro ^1H es típica de proteínas con estructura terciaria definida, mientras que una baja dispersión es indicativa de proteínas desplegadas. El espectro de la [figura 3.32](#) muestra señales dispersas indicando que CD5DIII está plegado. La existencia de dos señales separadas para los triptófanos de la proteína muestra que ambos se encuentran en entornos diferentes.

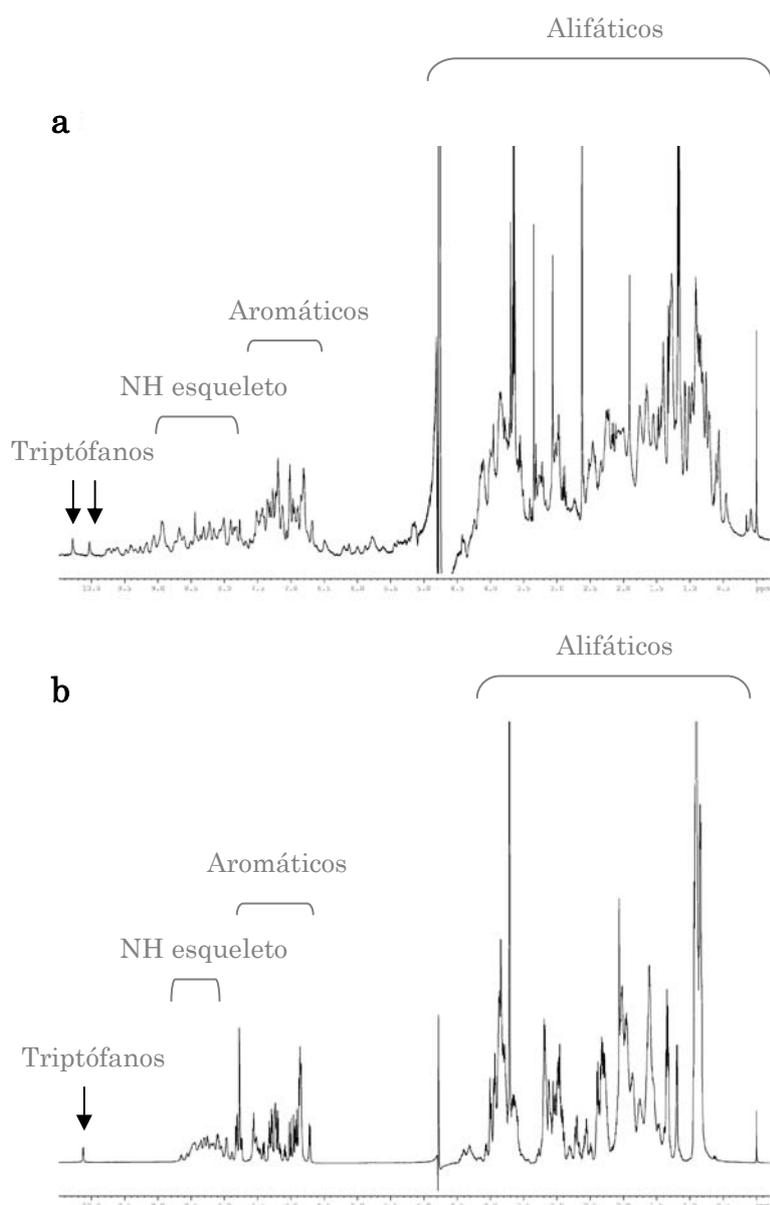


Figura 3.32. Espectro de RMN de CD5DIII. (a) Espectro de CD5DIII con señales dispersas y resaltando los triptófanos de la proteína. (b) Espectro de una proteína desplegada de peso molecular similar con baja dispersión de señal.

3.3.3. Cristalización de CD5DIII

Tras el análisis biofísico se realizaron numerosos cribados de cristalización mediante el método de la gota sentada tratando de ajustar la escasa cantidad de proteína obtenida a las necesidades de realizar un amplio estudio de las condiciones de cristalización. Así, se pusieron gotas de 0.1 μ l de proteína y 0.1 μ l de reservorio que dieron unos cristales iniciales que pudieron ser optimizados hasta obtener, como se muestra en la [figura 3.33](#), unos cristales de tamaño y calidad suficientes para llevar a cabo experimentos de difracción ([figura 3.34](#)).

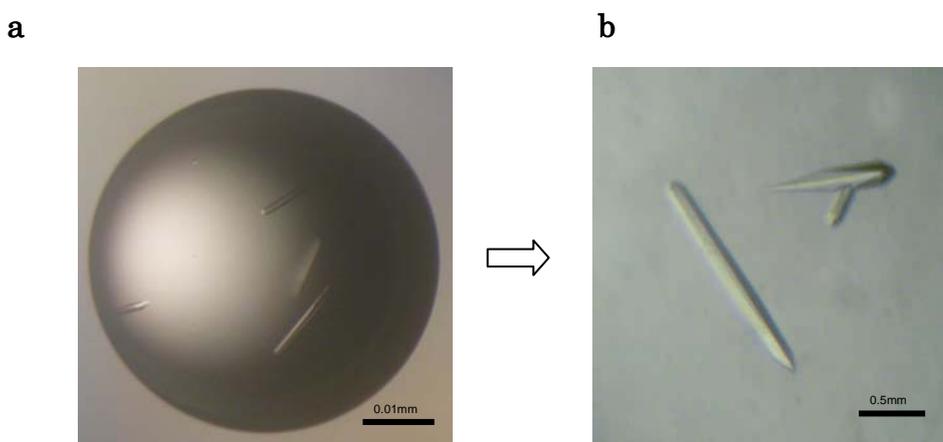


Figura 3.33. Cristales de CD5DIII. (a) Cristales crecidos a 16° C con el método de la gota sentada obtenidos a partir de las condiciones primitivas tras los que se obtuvieron (b) los cristales tetragonales optimizados. Estos cristales difractaron hasta 2.5 Å.

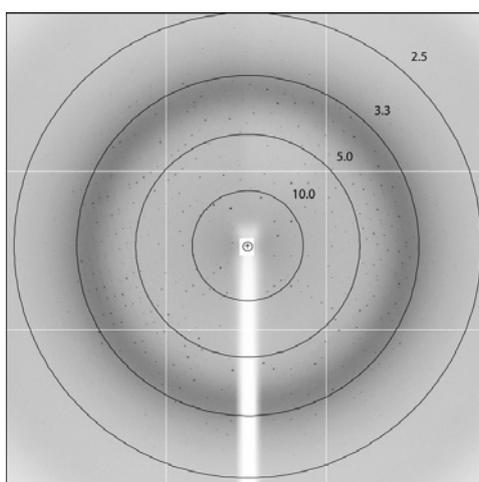


Figura 3.34. Patrón de difracción de cristales tetragonales de CD5DIII. Los anillos marcan los límites de resolución expresados en Å.

3.3.4. Resolución de la estructura de CD5DIII

Para resolver la estructura se intentaron dos aproximaciones. La primera fue el uso de átomos pesados previo análisis de la interacción del compuesto con el átomo pesado y la proteína, tal y como está descrito en la literatura (Boggon 2000). Los resultados según se observa en la [figura 3.35](#) no son claramente positivos. A pesar de todo, se probó a adicionar los átomos pesados a las gotas con los cristales de CD5DIII. Se mantuvieron los cristales en la gota con átomos pesados un tiempo variable que iba desde 30 minutos hasta 2 días. Aunque exteriormente el cristal parecía permanecer intacto, los patrones de difracción obtenidos mostraban un claro daño interno obteniendo datos de muy baja calidad que no pudieron ser procesados.

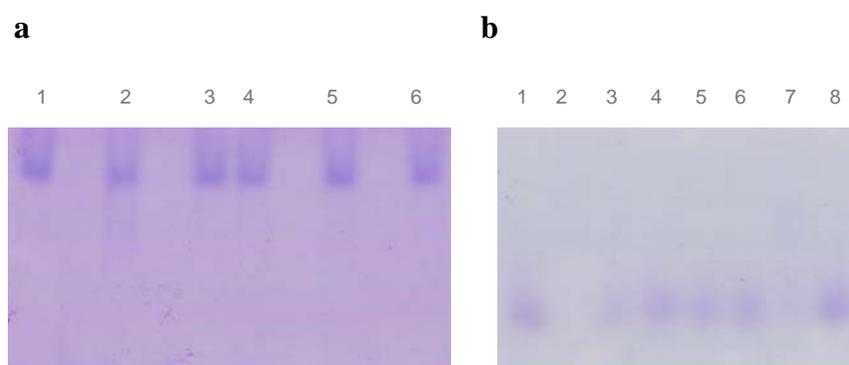


Figura 3.35. Ensayo de desplazamiento de CD5DIII tras la unión a diferentes átomos pesados. (a) y (b) son dos geles del 10% de SDS PAGE, (a) ha estado en marcha 10 horas mientras que (b) ha estado en marcha 20 horas. En (a), 1 corresponde a la proteína sin mezclar con ningún metal pesado, las siguientes calles corresponden a la proteína incubada con PtCl_4K_2 , HgCl_2 , $\text{C}_2\text{H}_7\text{HgO}_4\text{P}$, $\text{KAu}(\text{CN})_2$ y $(\text{CH}_3)_3\text{Pb}(\text{CH}_3\text{O})$, respectivamente. En (b), 1 corresponde a la proteína sin mezclar con átomos pesados y las siguientes calles contienen la proteína incubada con K_2IrCl_6 , AgNO_3 , $\text{C}_9\text{H}_9\text{HgNaO}_2\text{S}$, $\text{C}_{13}\text{H}_{17}\text{HgNO}_6$, $\text{K}_2\text{Pt}(\text{CN})_4$, K_2HgI_4 y $\text{C}_{17}\text{H}_{29}\text{GdN}_4\text{O}_7$, respectivamente.

La segunda aproximación trataba de aprovechar la presencia, en esta proteína, de ocho cisteínas. Esto, para un total de 104 aminoácidos supone una señal anómala de 1.6% estimada de acuerdo a Hendrikson et al (Hendrickson 1981):

$$\Delta F_{anom}/F = \sqrt{(2N_A/N_T)} \cdot f''/Z_{eff}$$

Siendo N_A el número de átomos con señal anómala, N_T el número total de átomos y $Z_{eff} = 6.7$ el número efectivo de electrones.

Así, se calculó que la señal anómala producida por los azufres de la proteína es suficiente para la resolución de las fases ya que el límite inferior está estimado en $\Delta F/F = 0.6\%$ (Wang 1985). Otros factores a tener en cuenta para considerar la resolución de la estructura por SAD de azufre es tener unos datos de difracción de alta calidad. Para ello se requiere una elevada redundancia, combinada con una cuidada optimización de los parámetros a la hora de la recogida de datos (Wagner 2004). La obtención de cristales tetragonales de un grupo espacial p4122 o p4322 que presentan una alta simetría facilita la obtención de unos datos de difracción de alta calidad. Estructuras previamente resueltas que cristalizaron en el mismo grupo espacial que CD5DIII y con un contenido similar de azufre se habían resuelto mediante el uso de esta técnica (Debreczeni 2003; Wagner 2004).

Utilizando radiación de sincrotrón a 1.8\AA se maximizó la señal anómala del azufre. También se probó con 1.75\AA pero no se consiguieron medir las diferencias anómalas adecuadamente, siendo ese pequeño cambio en la señal el que permitió finalmente resolver el problema de las fases. Para maximizar las diferencias de la señal anómala se recogieron datos usando una estrategia de rayo inverso (inverse beam) para conservar intactas las diferencias entre los pares de Bijvoet. Después de integrar y escalar los datos se obtuvo una muestra de la señal anómala obtenida que queda reflejada en la [figura 3.36](#).

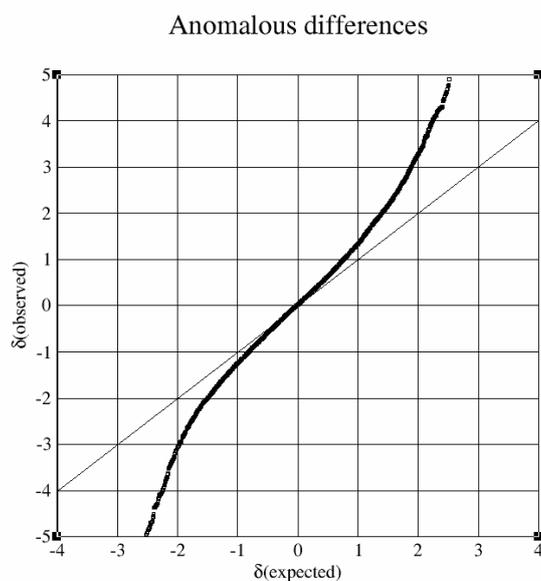


Figura 3.36. Análisis probabilístico de la distribución Gaussiana de la Normal de las diferencias anómalas observadas en base a los datos de difracción obtenidos a 1.8Å de longitud de onda. Esta gráfica representa la desviación $\delta = (I - \langle I \rangle) / \sigma$ frente a lo que se esperaría de una distribución Gaussiana (Normal). Así, si los datos siguieran una distribución Gaussiana de los errores con el σ estimado todos los puntos de la gráfica se ajustarían a la diagonal con la mayor parte de los puntos en el centro y pocos en los extremos. En este caso, las diferencias observadas se separan de manera clara de las esperadas mostrando señal anómala.

Finalmente, la estructura fue determinada a 2.8Å de resolución tal y como se describe en materiales y métodos ([sección 2.3.10](#)), refinada a 2.5Å en el mismo grupo espacial tetragonal y posteriormente resuelta por reemplazo molecular a 2.2Å en un grupo espacial trigonal ([tabla 3.4](#)).

Tabla 3.4. Estadísticas de recolección de datos y refinado

Recolección de datos ^a	Tetragonal S-SAD	Tetragonal	Trigonal
Número de cristales	1	1	1
Temperatura (K)	100	100	100
Entorno	MAR225, SLS, Línea PX	ADSC-Q315, ESRF, Línea ID-14-4	MAR225, SLS, Línea PX
Longitud de onda (Å)	1.80	0.97	0.93
Grupo espacial	P4 ₁ 22	P4 ₁ 22	P3 ₂ 2 ₁
Dimensiones de la celda (Å, °),	a=b=74.420, c=128.824, α,β,γ=90°	a=b=74.161, c=128.992, α,β,γ=90°	a=b=62.93, c=72.56, α,β=90°, γ=120°
Nº moléculas en la unidad asimétrica	2	2	1
Resolución (Å)	55.3-2.8	50-2.5	50-2.2
Reflexiones únicas	9432	12892	12098
Multiplicidad media	52 (43)	8.9(9.1)	5.4(4.1)
% Reflexiones totales	100 (99.9)	99.6(98.9)	96.9(88.3)
R _{merge} ^b	0.09(0.25)	0.07(0.34)	0.05(0.24)
<I/σ(I)>	10.7(3.1)	6.9(2.1)	15.3(2.2)
B factor Wilson plot	56	57	50
Estadísticas de faseado de S-SAD			
Nº S-S encontrados	5/8		
FOM (SHARP)	0.35		
RCullis (SHARP)	0.8		
Poder de faseado	1.06		
Modificación de la densidad			
FOM (Solomon)	0.85		
Refinado			
Nº Reflexiones		12350	7899
Rango de resolución (Å)		40/2.5	20/2.2
R-factor / R-free (%)		22/25.52	19.32/24.75
Nº átomos de proteína (media B, Å ²) ^c		1382 48.53	789 44.56
Nº moléculas de agua (media B, Å ²) ^c		80 54.70	58 47.53
r.m.s longitud de enlace (Å)		0.021	0.018
r.m.s. ángulo de enlace (°)		2.120	2.156
Inválidos del gráfico de Ramachandran (número) ^d		0	0

^a Los valores de la celda de máxima resolución se dan en paréntesis.

^b $R_{merge} = \sum_{\eta} \sum_i |I_{\eta,i} - \langle I_{\eta} \rangle| / \sum_{\eta} \sum_i I_{\eta,i}$

^c Calculado con MOLEMAN.

^d Calculado con PROCHECK.

3.3.5. Descripción general de la estructura de CD5DIII

El primer y el último residuo observados en la densidad electrónica son Ala269 y Asp369, una cadena polipeptídica ligeramente más larga que la definición canónica del dominio SRCR del grupo B. El dominio CD5DIII adopta un plegamiento compacto adquiriendo una forma de corazón de dimensiones aproximadas de 28x38x35Å. El núcleo de la estructura está formado por la asociación de la hélice $\alpha 1$ con una lámina β antiparalela curvada que incluye las hojas $\beta 1$, $\beta 2$, $\beta 7$ y $\beta 4$. Las tres hojas β adicionales ($\beta 3$, $\beta 5$ y $\beta 6$) forman otra lamina β antiparalela en la parte inferior del dominio central. La hélice $\alpha 2$ conecta el final de la hoja $\beta 4$ a $\beta 5$ mientras que la hélice $\alpha 3$ conecta $\beta 5$ y $\beta 6$ (figura 3.37). El núcleo central hidrofóbico está formado por la hélice $\alpha 1$ asociada con la cara cóncava de la hoja β más grande y la parte superior de la más pequeña. Así, esta parte central de CD5DIII forma una cavidad triangular formada por las hojas $\beta 2$, $\beta 7$ y $\beta 3$ ocupada en su interior por una molécula de agua que forma puentes de hidrógeno con las tres hojas (Val289, Leu300 y Val363).

Hay ocho cisteínas en CD5DIII SRCR enlazadas formando puentes disulfuro: Cys285 y Cys321 forman un puente que une el extremo N-terminal de $\beta 2$ con el C-terminal de $\alpha 1$; el disulfuro Cys316-Cys367 conecta la hélice $\alpha 1$ a la hoja $\beta 7$ en el medio de la hoja antiparalela; Cys342-Cys350 une la hélice $\alpha 3$ y finalmente Cys301-Cys360 forma un puente disulfuro que une el final de la hoja $\beta 3$ con la zona amino terminal de la hoja final $\beta 7$ (figuras 3.37, 3.38 y 3.39).

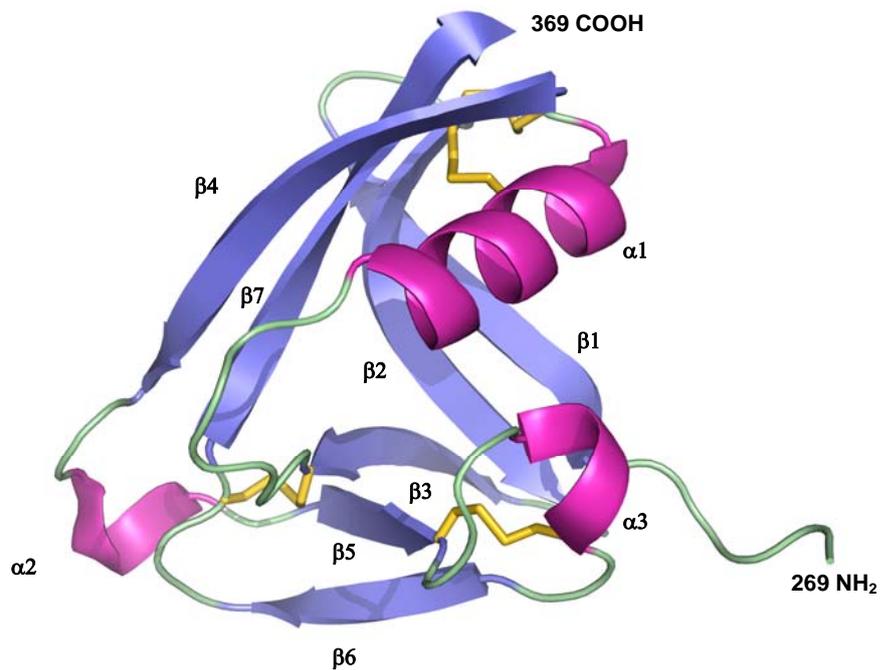


Figura 3.37. Diagrama de lazos de CD5DIII. Las hojas β se señalan en azul y en rosa las hélices α . Los cuatro puentes disulfuro están marcados en naranja.

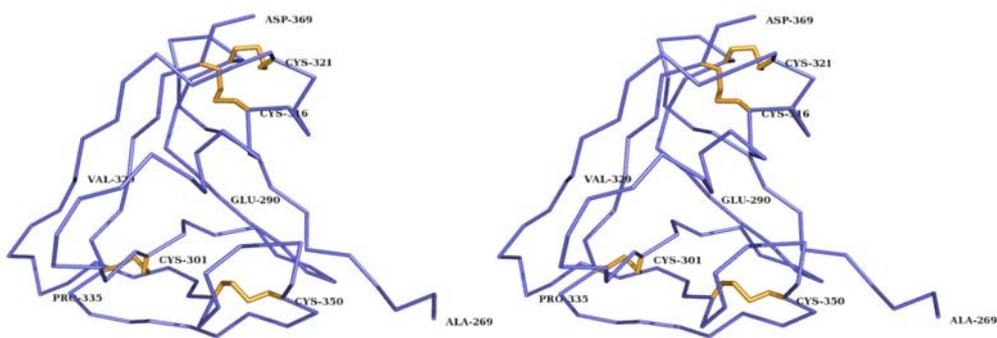


Figura 3.38. Vista en estéreo de la estructura de CD5DIII. Los puentes disulfuro se marcan en naranja y una cisteína de cada puente está marcada para aclarar su posición dentro de la estructura.

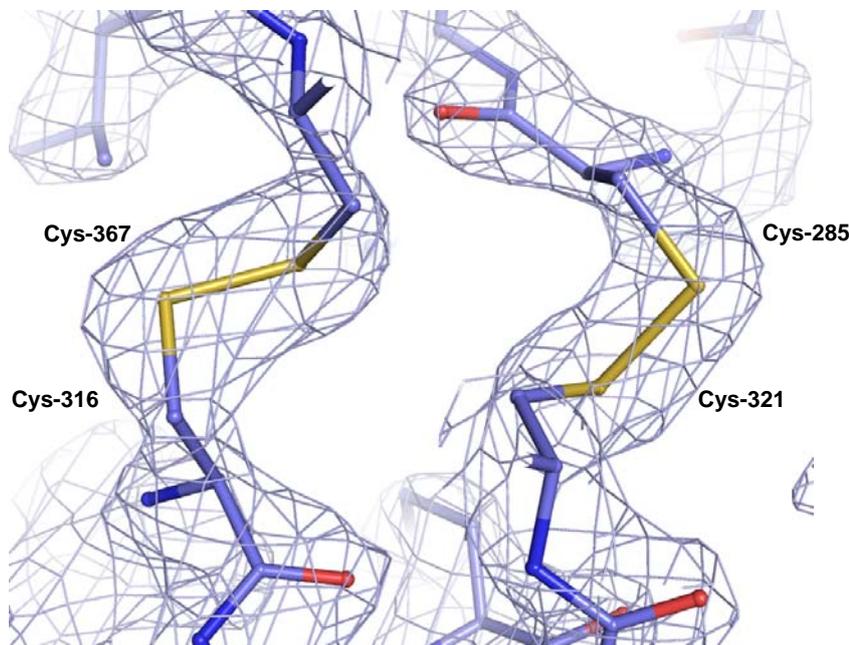


Figura 3.39. Mapa 2FoFc a 1σ de la densidad electrónica de una región de CD5DIII. Se muestra en naranja el característico puente disulfuro C1-C4 (Cys285 y Cys321) de los dominios SRCR del grupo B y el puente disulfuro C3-C8 (Cys316 – Cys367)

Las regiones expuestas al solvente de CD5DIII muestran una ordenación especial en la superficie de la proteína. Algunos residuos (Lys273, Arg307, Arg328, Lys346, Lys361, Arg356) están organizados formando una tira básica que corre a lo largo de la superficie desde el extremo N-terminal sobre $\alpha 1$ y la parte de atrás de las cuatro hojas β antiparalelas y termina en la cavidad formada por las tres hojas β antiparalelas y la hélice $\alpha 2$. Los residuos ácidos se agrupan en lugares opuestos de la molécula. La primera región ácida la forma el loop que une $\beta 1$ y $\beta 2$ junto con el final de la $\beta 7$, que conecta el dominio extracelular a la hélice transmembrana (Glu286, Asp369). También se extiende a lo largo de una cavidad formada por el amino terminal de $\beta 1$ y un lado de la hélice $\alpha 1$ (Glu314 y Glu318). La segunda región se localiza en la parte posterior de la hélice $\alpha 2$ junto con los residuos del loop que conectan la hélice a $\beta 4$ y $\beta 5$ (Asp31 y Asp334) (figura 3.40).

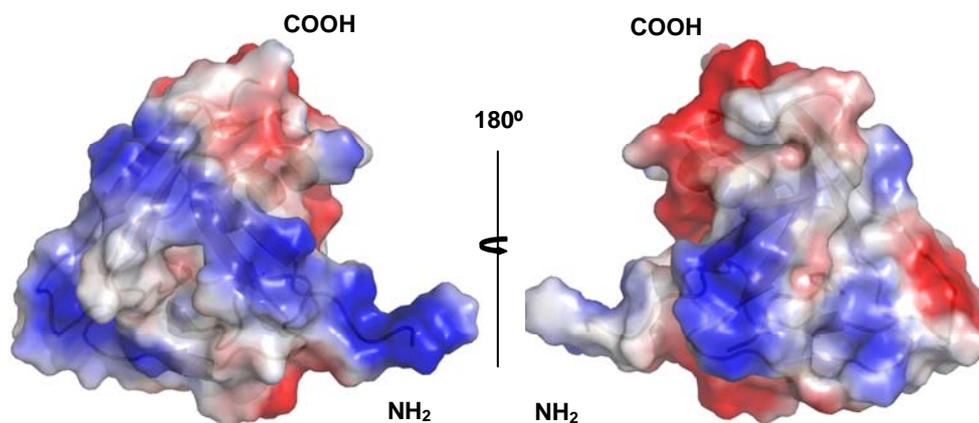


Figura 3.40. Mapa de potencial electrostático de CD5DIII. Las áreas cargadas positivamente se muestran en azul y las que están cargadas negativamente se muestran en rojo.

3.3.6. Comparación entre los dominios SRCR de los grupos A y B

Las principales diferencias entre los miembros del grupo A y B dentro de la superfamilia SRCR reside en el número de cisteínas implicadas en los puentes disulfuro así como en la composición de sus dominios (Sarrias 2004). Hasta la fecha, la información estructural acerca de ambos grupos era más bien escasa. M2bp (Hohenester 1999) y el dominio SRCR de la hepsina (Somoza 2003) son los únicos representantes del grupo A cuyas estructuras están resueltas. CD5DIII es la primera estructura que aporta información acerca del grupo B contribuyendo así a un mayor entendimiento de esta superfamilia. Una comparación detallada de las tres estructuras revela que a pesar de que las diferencias de r.m.s.d. son significativas (CD5DIII-M2BP: 2.4Å, para 87 átomos C α que coinciden y CD5DIII-hepsin: 2.6Å, para 79 átomos C α que coinciden), comparten un plegamiento similar conservando la mayor parte de los elementos de estructura secundaria (figuras 3.41 y 3.42). Las principales diferencias se observan en los loops conectores (figura 3.42b y 3.42c). La superposición de las estructuras muestra que las primeras cuatro hojas β y la hélice α siguen un camino paralelo. De ese punto en adelante las estructuras divergen, convergiendo de nuevo junto a la última hoja β . Es interesante remarcar que la posición de la molécula de agua antes mencionada que une las hojas β 2, β 7 y β 3 interaccionando con los átomos de

desordenado. El segundo loop presenta mayores diferencias (figura 3.42c). Las tres estructuras tienen un loop largo que protruye desde una de las caras de la lámina β y termina de cara a la hélice α opuesta. Esta disposición forma una depresión de aproximadamente 18\AA de diámetro con un hueco de 5\AA entre los residuos que forman la punta de los loops y la α hélice. La supresión parcial de la punta del loop de CD5DIII con respecto a las otras dos estructuras se compensa con la presencia de residuos más grandes de tal manera que la distancia de apertura se conserva.

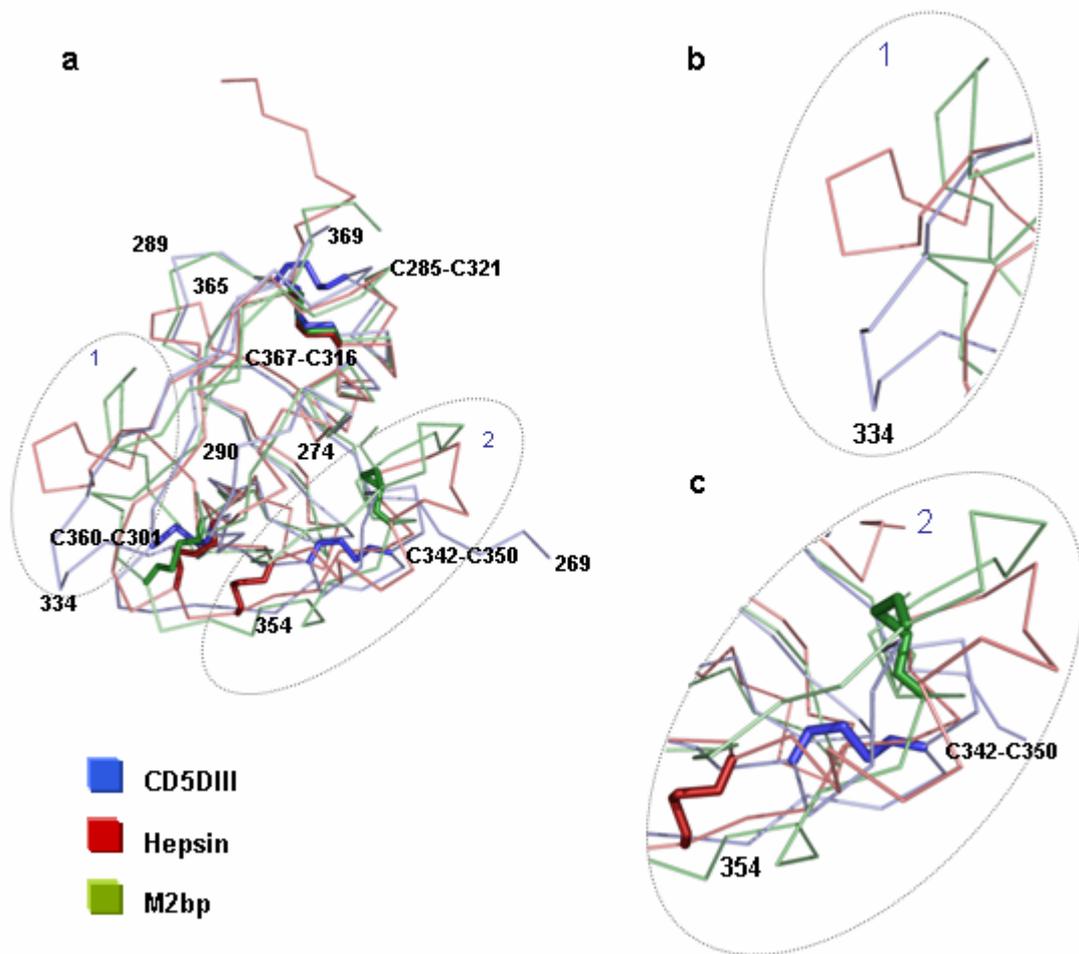


Figura 3.42. Comparación de las estructuras de los dominios SRCR del grupo A y B. (a) Superposición de los Ca de la estructura de CD5DIII con las estructuras de los dominios SRCR del grupo A de la hepsina y M2bp. (b) y (c) muestran en detalle las dos regiones que presentan mayores diferencias en sus loops.

En lo que a los puentes disulfuro se refiere, en la figura se puede observar las principales diferencias ocasionadas por las posiciones de los mismos. Sólo uno de los cuatro puentes que estabilizan la estructura de CD5DIII (C316-C367) conserva su posición en la hepsina y M2bp. A pesar de que los disulfuros C301-C360 y C342-C350 están presentes en ambas moléculas del grupo A sus posiciones no están conservadas en el grupo B. El cuarto puente disulfuro, exclusivo del grupo B (C285-C321) que se puede apreciar en la figura 41 delimita la posición del carboxilo terminal que une el dominio extracelular con la membrana.

3.3.7. Orientación de la región extracelular de CD5 con respecto a la membrana plasmática.

La estructura cristalina sugiere la forma en la que el receptor podría estar orientado con respecto a la membrana plasmática. El residuo del carboxilo terminal Asp369 se encuentra a tan solo ocho residuos de distancia de la previamente predicha región transmembrana. Esta pequeña separación impone fuertes restricciones espaciales para definir la relación de la región extracelular con la hélice transmembrana. De esta manera, la hoja 87 de CD5DIII, que contiene al carboxilo terminal, debería estar orientada prácticamente perpendicular al plano de la membrana (figura 3.43). La hélice transmembrana propuesta la forman 22 amino ácidos (figura 1.7); consecuentemente la longitud indica que CD5 contiene una única hélice transmembrana que ancla el receptor. La estructura resuelta también incluye residuos que pertenecen al linker y al amino terminal de CD5DII indicando que el resto de la zona extracelular de CD5 se situaría a un lado del dominio III, dejando la parte superior de este dominio expuesta. Así, considerando las dimensiones de CD5DIII, la región extracelular del receptor debería sobresalir unos 90-100 Å de la membrana plasmática. Esta organización es similar a la forma en la que se presenta CD5DIII en la unidad asimétrica de la forma cristalina tetragonal.

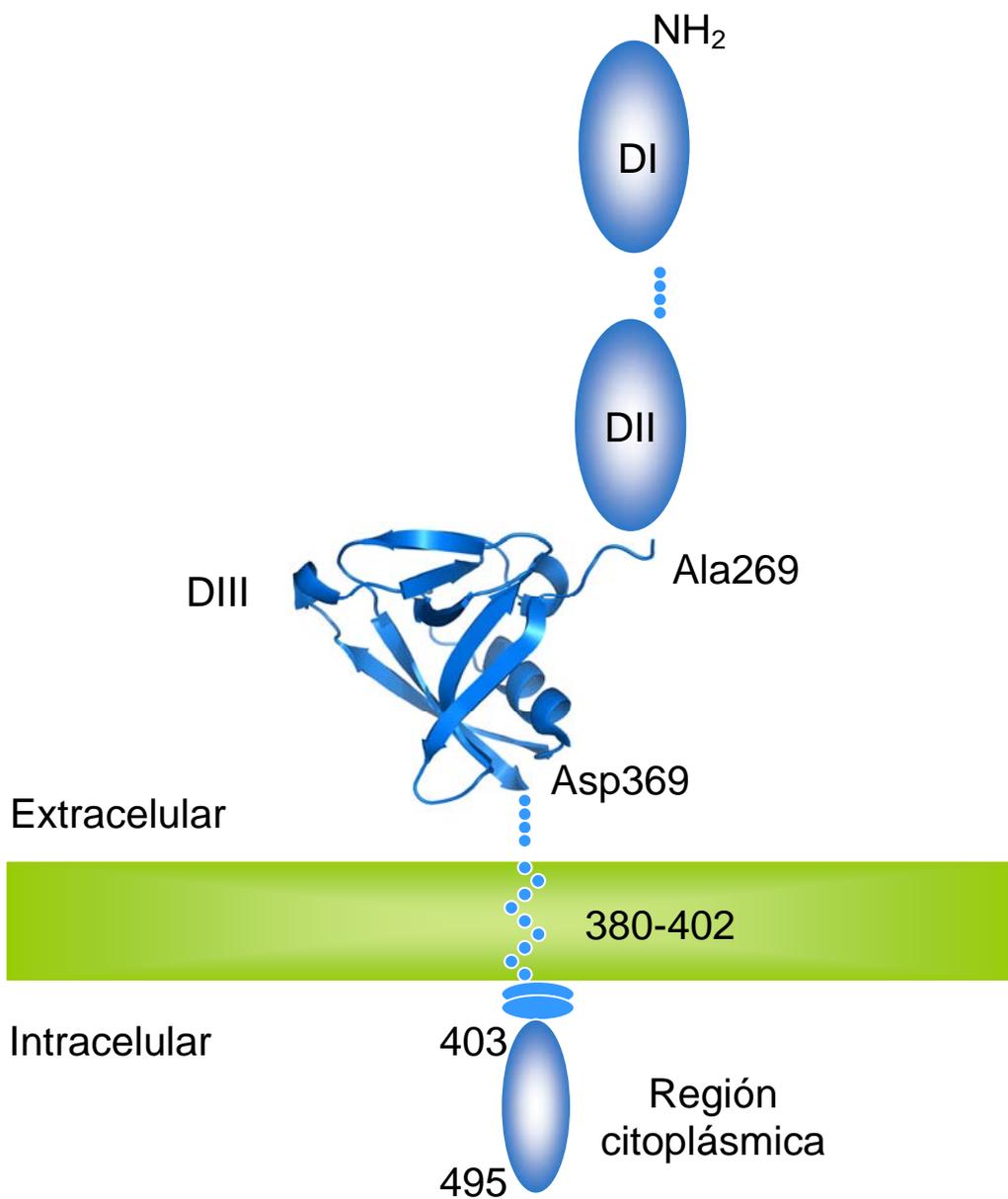


Figura 3.43. Modelado del receptor entero de CD5. El dominio III ha sido resuelto por cristalografía de rayos X. El modelo se ha realizado teniendo en cuenta que el dominio I es de gran similitud y teniendo en cuenta las glicosilaciones presentes entre los dominios I y II, que como está descrito confieren rigidez a la estructura.



DISCUSIÓN

En esta tesis doctoral se ha intentado abordar el estudio estructural de las interacciones del VHC con proteínas del huésped intentando aportar datos concretos sobre la interacción directa o indirecta con algunas proteínas celulares. Para la realización de este trabajo se han llevado a cabo estudios en paralelo con la proteína C del VHC, con la helicasa DDX3 y con el receptor CD5 con resultados diversos que se han descrito previamente y que discutiremos a continuación.

4.1. La proteína C como posible diana a la hora de desarrollar una terapia efectiva contra la enfermedad

El VHC es un virus cuyo genoma está compuesto por una única hebra de ARN positivo protegido por una envuelta formada principalmente por las proteínas E1 y E2 y una nucleocápsida compuesta por unidades de la proteína C (Lukavsky 2000). La proteína C directamente interacciona con proteínas celulares ocasionando trastornos a diversos niveles (Inghirami 1991; Mamiya 1999; Owsianka 1999; Jin 2000). La formación *in vitro* de partículas similares a nucleocápsidas (NLPs) puede ser una herramienta importante para estudiar cómo interacciona la proteína C a nivel molecular. Un conocimiento detallado de la estructura atómica de esta proteína una vez que está ensamblada formando nucleocápsidas puede aportar claves importantes sobre la infección e incluso servir para el desarrollo de alguna molécula que impida dichas interacciones y funcione como fármaco para tratar la enfermedad.

4.2. Intentos de cristalización de NLPs

4.2.1. ¿Es posible la renaturalización de la proteína C una vez desnaturalizada con 8M urea?

En el laboratorio se han tratado de reproducir partículas similares a nucleocápsidas (NLPs) siguiendo dos protocolos distintos. En ambos casos una forma truncada de la proteína C formada por los 124 primeros aminoácidos (VHCc124) se expresaba en células de *E. coli*. En el protocolo descrito por Kunkel et al (Kunkel 2001) se realizaba la purificación de la proteína expresada de forma desnaturalizada usando un tampón con 8M urea. Posteriormente, la proteína se renaturalizaba mediante una diálisis extensiva en un tampón sin urea. Tras intercambiar de nuevo el tampón por uno de ensamblado, se formaban NLPs, previa incubación de la muestra con ARN^{t^{phe}}. Los NLPs resultaban visibles por microscopía electrónica. El ARN^{t^{phe}} sustituiría en este caso al ARN genómico del virus que se asocia con las proteínas básicas de la nucleocápsida para formar la misma.

La expresión y purificación de la proteína se llevó a cabo de manera satisfactoria pero, a pesar de seguir los pasos descritos, VHCc124 permaneció desplegada después de pasar por el proceso de diálisis y no presentó ningún tipo de plegamiento ni después del intercambio al tampón

de ensamblado, ni una vez realizada la incubación de la proteína con el ARN^{t_{phe}}. Los análisis realizados para comprobar el plegamiento fueron espectros de dicroísmo circular en el ultravioleta lejano que nos permitieron analizar la existencia de estructura secundaria. El espectro de DC de la proteína después de la diálisis no presentó estructura secundaria. Un espectro prácticamente idéntico se pudo observar al analizar la muestra de proteína una vez dializada frente al tampón de ensamblado o incubada con el ARN^{t_{phe}}. También se realizó un estudio de la fluorescencia de los triptófanos de la proteína y el resultado en este caso de la muestra con los NLPs ya formados indicó un desplegamiento de VHCc124 ya que los triptófanos que se suponían responsables del ensamblaje estaban expuestos al medio.

El correcto plegamiento de la proteína C y la presencia del ARN^{t_{phe}} puede que sea suficiente a la hora de formar NLPs *in Vitro*, tal y como sucede en el caso del TMV (Fraenkel-Conrat 1955) o del VHB (Birnbaum 1990), pero el replegamiento de la proteína C ha resultado en este caso una barrera insalvable. Es posible que las divergencias en las secuencias de VHCc124 debido a la variabilidad en las muestras de los pacientes hayan podido causar estas discrepancias entre los datos obtenidos y los datos presentados por Kunkel et al.

4.2.2. La purificación de VHCc124 en su forma nativa supone una mejor alternativa para la obtención de NLPs

En contraste con el primer intento de purificación de la proteína C y de formación de NLPs a partir de VHCc124 desnaturalizada, los resultados obtenidos siguiendo el protocolo de Lorenzo et al (Lorenzo 2001) en el que VHCc124 se purificaba de manera soluble, han sido satisfactorios. Si bien todavía no se ha podido comprobar la presencia de NLPs por microscopía electrónica, sí se han obtenido resultados como los espectros de DC que avalan la idea de que hemos conseguido purificar proteína C con estructura secundaria y, al menos parcialmente, terciaria. Dado que en este caso no se ha desnaturalizado la proteína C, ésta ha podido unirse de manera espontánea a su propio ARNm y ensamblarse formando nucleocápsidas al igual que ocurre, por ejemplo, con el VHB (Birnbaum 1990).

4.3. ¿Cómo se puede avanzar en el desarrollo del trabajo con VHCc124?

Al contrario de lo que sucede con el VHB en el que se consiguen mediante un protocolo, relativamente sencillo, expresar en células de *E. coli* una gran cantidad de partículas homogéneas (Crowther 1994), en el VHC hay un alto grado de heterogeneidad de los NLPs que se han descrito en los estudios realizados previamente (Kunkel 2001). Así, a pesar de ser un trabajo científicamente relevante, su dificultad puede dilatar en el tiempo la resolución de una estructura completa de la nucleocápsida del virus.

Otra línea de actuación es estudiar las interacciones de la proteína C con diversas dianas celulares como DDX3. Así, se puede trabajar con la proteína C silvestre o incluso con el péptido correspondiente a los aminoácidos de interacción entre VHCc124 y DDX3. De esta manera, se puede intentar la co-cristalización de ambas proteínas y la resolución de este complejo revelaría datos importantes para avanzar en el entendimiento de la forma de actuación del virus y la manera en la que éste interacciona con proteínas en el interior de la célula.

4.4. DDX3 como diana a la hora de estudiar la hepatitis C.

DDX3 es una ARN helicasa que desempeña un papel en la modificación del ARN celular a diversos niveles (Zhou 2002; Kanai 2004). Además, es una proteína implicada en enfermedades de gran relevancia para la salud pública como son la hepatitis C (Mamiya 1999; Owsianka 1999), el SIDA (Yedavalli 2004) o el cáncer (Chao 2006). Como parte de este estudio, se ha tratado de resolver la estructura atómica de esta proteína mediante técnicas de difracción de rayos X para desentrañar su función y su modo de interacción con la proteína C del VHC.

4.5. Estrategia de cristalización de DDX3

Inicialmente se llevó a cabo un estudio *in silico* de DDX3 para obtener la mayor cantidad de información sobre posibles construcciones para cristalizar. El análisis comparativo de DDX3 con la secuencia de otras proteínas homólogas previamente resueltas (figura 1.7) y los datos acerca de las regiones desordenadas sugeridos por el programa PONDR presentaron dos regiones a priori desordenadas en los extremos N- y C-terminal de la

proteína. Estas regiones más flexibles estaban ausentes en todas las proteínas homólogas cuya estructura había sido definida. Así, se planteó el clonaje de diversas construcciones que están detalladas en la sección de resultados 3.2.1 (figura 3.4).

Los resultados preliminares obtenidos con DDX3 confirmaban el estudio inicial presentando unos problemas de solubilidad seguramente derivados de la presencia de estas regiones flexibles en sus extremos. La proteína en un tampón de cloruro sódico permanecía insoluble en su casi totalidad. Basándonos en los datos de la literatura (Story 2001) que describían el uso de altas concentraciones de sulfato amónico para solubilizar el homólogo de DDX3 en *M. janaaschii*, se consiguió solventar el problema de forma parcial. A partir de este punto la purificación fue razonablemente bien alcanzando una pureza en la muestra de proteína de más del 95%. Estos datos sumados a los estudios realizados de dicroísmo circular y dispersión dinámica de luz presentaron una muestra homogénea viable para la cristalización. El análisis positivo del estudio de la actividad ATPasa de la proteína dependiente de ARN suponía que la proteína no sólo estaba plegada sino que conservaba su actividad. El hecho de no haber podido solventar totalmente el problema de la solubilidad de DDX3 se tradujo en una incapacidad para aumentar la concentración de la muestra a más de 1.5mg/ml. Esta concentración resultó insuficiente para obtener resultados positivos en los ensayos de cristalización en los que las gotas o estaban claras o presentaban proteína desnaturalizada.

4.6. Los extremos amino y carboxilo terminal de la proteína son los responsables de la falta de solubilidad

Se realizó un ensayo de proteólisis limitada para tratar de obtener un fragmento de la proteína que fuera activo y resultara estable al suprimir las regiones supuestamente más desordenadas de la proteína. El fragmento DDX3_{pf} (164-600) se obtuvo a partir del análisis por espectrometría de masas y degradación de Edman de la banda mayoritaria resultante del ensayo de proteólisis limitada, el cual permitió determinar los extremos N- y C-terminal de la nueva construcción. Los problemas de solubilidad no se solucionaron con esta nueva construcción. La pureza que se obtuvo de la muestra fue incluso mayor que en el caso de DDX3 y el análisis de DLS señalaba una muestra más homogénea. Además, DDX3_{pf} conservaba su

actividad helicasa dependiente de la hidrólisis de ARN. Sin embargo, la solubilidad de la proteína seguía siendo reducida y la concentración de la muestra resultó inviable por encima de 1.5mg/ml. Se llevaron a cabo, a pesar de todo, ensayos de cristalización pero los resultados fueron idénticos a los obtenidos con DDX3, gotas principalmente claras o muestra desnaturalizada.

A pesar de que DDX3_{pf} reducía enormemente los extremos N- y C-terminal de DDX3, contenía 5 aminoácidos en el N-terminal y 21 en el C-terminal que no estaban presentes en ninguna de las estructuras homólogas resueltas y que el programa PONDR había predicho como regiones desordenadas. Se llevó a cabo el clonaje del fragmento recombinante DDX3_{pd}r (169-579) que reducía DDX3 al centro funcional de la misma. La muestra con DDX3_{pd}r funcionó de manera satisfactoria en lo que a la expresión y la purificación se refiere sustituyendo en este caso el tampón de resuspensión con sulfato amónico por un tampón de resuspensión con una alta concentración de cloruro sódico. Esta solución, no sólo permitió solubilizar la proteína de manera parcial, sino que permitió concentrar la muestra hasta 10mg/ml y realizar así los ensayos de cristalización. Los datos de DC, DLS y el ensayo de actividad de la muestra unidos a la desaparición del problema de solubilidad invitan a pensar que se obtengan resultados positivos de los cribados de cristalización.

4.7. El dominio helicasa de DDX3 es la parte más soluble de la proteína mientras que el dominio ATPasa es insoluble.

DDX3_{pd}r se compone de dos dominios, el dominio ATPasa, DDX3_{ATP} (173-406), y el dominio helicasa, DDX3_{hel} (407-578). Los ensayos que se realizaron en paralelo de las pruebas de expresión de DDX3_{ATP} y DDX3_{hel} mostraron un comportamiento diferente de ambos dominios. Ambas construcciones se expresaron de manera satisfactoria en células de *E. coli*. Dado que un tampón con 1M de NaCl en sustitución del sulfato amónico había sido suficiente para solventar el problema de solubilidad de la construcción DDX3_{pd}r, se probó el efecto de ambos tampones en las proteínas recombinantes a la hora de realizar los ensayos preliminares de solubilidad. El dominio helicasa resultó insoluble en cloruro sódico pero permaneció altamente soluble en una solución con sulfato amónico similar a las

utilizadas para la purificación de las construcciones DDX3 y DDX3_{pf}. El dominio ATPasa resultó insoluble en ambos tampones.

Es interesante destacar que el espectro de DC de DDX3_{hel} muestra una T_m 10° C más alta que la T_m que mostraban todas las construcciones anteriores. Esto implica que esta región de DDX3 es más compacta que, por ejemplo, los dos dominios juntos (DDX3_{pdv}).

4.8. La alta homología de DDX3_{hel} con Vasa permite realizar un modelo de los dominios ATPasa y helicasa de DDX3.

Los ensayos de cristalización con DDX3_{hel} dieron resultados positivos y gracias a estructuras homólogas previamente resueltas (Vasa de *D. melanogaster*) pudimos resolver la estructura de DDX3_{hel} usando el método de reemplazo molecular para resolver el problema de fases. Una vez resuelta esta estructura se observó que la estructura terciaria era prácticamente idéntica a la estructura del dominio helicasa de Vasa. Esto permitió diseñar un modelo del dominio helicasa y ATPasa de DDX3 usando para ello el programa Swiss-Model. La parte correspondiente a DDX3_{hel} en el modelo se sustituyó por la estructura resuelta. Este modelo modificado muestra como hay una interacción clara entre los motivos de uno y otro dominio pero sin embargo no ayuda demasiado al esclarecimiento del funcionamiento real de la enzima. El ARN cristalizado junto con la proteína Vasa se ha situado en en la zona de DDX3 donde debería encontrarse y se aprecia la interacción de los motivos del dominio ATPasa Ia y Ib, así como de los motivos del dominio helicasa IV y V. Sin embargo, es necesaria la cristalización de la proteína silvestre para poder esclarecer realmente la manera en que actúa la helicasa y como las regiones N- y C-terminal de la proteína influyen a la hora de desempeñar su actividad enzimática.

4.9. Hacia la resolución de DDX3

Teniendo en cuenta la capacidad del fragmento helicasa para “solubilizar” partes más flexibles pegadas a él, sería interesante realizar el subclonaje de un fragmento que incluyera este dominio, más la región de la proteína que se une a la proteína C del VHC. En un principio se planteó una construcción similar (DDX3_{RS}) pero dado que la construcción DDX3_{pf}, que carecía de estos aminoácidos, presentó grandes problemas de solubilidad, no se llevó a cabo

ningún estudio con dicha construcción. A tenor de los resultados obtenidos, una construcción formada por la región helicasa que incluyera los aminoácidos 579-623 parece más viable para su cristalización. La resolución de esta construcción arrojaría datos interesantes relacionados con la interacción de la proteína DDX3 y la proteína C del VHC. Además, se podrían realizar ensayos de unión de este fragmento con la proteína C previamente purificada y en caso de ser positivos, tal y como está descrito en la literatura (Mamiya 1999; Owsianka 1999), plantear la co-cristalización de ambos fragmentos. La superación del problema de fases a la hora de resolver la estructura no debería entramar especial dificultad puesto que dispondríamos ya de la estructura de la región helicasa con lo que el reemplazo molecular sería algo, a priori, rutinario.

No hay que olvidar tampoco que los resultados obtenidos con DDX3_{pdv} son prometedores y resolver la estructura de esta construcción arrojará más luz sobre la estructura de DDX3 y su función. Además, sería interesante tratar de mejorar la solubilidad de DDX3 mediante el uso de nuevas técnicas descritas recientemente (Berrow 2006) que hablan del empleo de amplios cribados usando distintos tampones y aditivos que puedan ayudar a la solubilidad de la proteína y así tratar de conseguir resolver la estructura de DDX3 silvestre. Estos cribados se pueden optimizar con el uso de técnicas de fluorescencia (Lo 2004; Ericsson 2006) que permitirían realizar el análisis de los compuestos de la criba de una manera rápida y eficaz.

Los trabajos realizados con DDX3 representan una línea directa de lo que puede suponer la resolución de la interacción del VHC con una proteína celular. En otra línea, algo más alejada de estas interacciones del VHC con dianas celulares se enmarcan los estudios realizados con el receptor CD5. Estos estudios son interesantes no sólo desde un punto de vista estructural, ya que es la primera vez que se resuelve la estructura molecular de una proteína perteneciente al grupo B de los dominios SRCR sino también desde la relación de ésta con el VHC, ya que es una proteína que parece, según se ha descrito anteriormente, y tal y como está detallado en la introducción, estar implicada en la infección del VHC, y más concretamente en la entrada del virus en la célula aunque de una forma desconocida hasta ahora.

4.10. Estrategia de cristalización de CD5

Como ya se ha comentado en los resultados, abordar directamente el estudio de todo el receptor hubiera supuesto un comienzo demasiado ambicioso debido principalmente a que los tres dominios extracelulares, incluidos los dominios I y II que presentan N- y O-glicosilaciones aportan una heterogeneidad y una pérdida de compacidad que dificultan enormemente el proceso de cristalización y de resolución de la estructura. Así, el estudio se centró en el dominio III, carente de glicosilaciones y enormemente similar al dominio I del receptor, que se podría utilizar como molde para la construcción de un modelo del receptor entero.

En contraste con la expresión de DDX3, que se realizó en células de *E. coli*, la expresión de CD5 se llevó a cabo en células humanas para obtener el dominio correctamente plegado. Esto complicó tanto el proceso de purificación como el de cristalización ya que la cantidad de proteína que se producía en 1L de sobrenadante de cultivos celulares apenas alcanzaba los 150µg. La reducida cantidad de proteína hizo indispensable utilizar una columna de afinidad con un anticuerpo monoclonal frente a CD5 que permitiera la purificación de CD5DIII en un solo paso. A pesar de trabajar con cantidades tan pequeñas, gracias al uso de aparatos Äkta prime con los que se pudo monitorizar detalladamente la presencia de proteína en las muestras, se consiguió obtener alrededor de 400µg de proteína pura de cada 2-3L de sobrenadante. La proteína se concentró hasta 4mg/ml, obteniendo alrededor de 100µl de muestra con cada purificación. Los microlitros obtenidos se aprovecharon al máximo gracias al uso de un robot de cristalización que permitió poner gotas de un tamaño inferior a 100nl de volumen. Así, se realizó un cribado de más de mil condiciones que dieron resultados positivos sobre los que trabajar, ya a escala de microlitros. Después de optimizar las condiciones se consiguieron cristales de calidad que usamos para los experimentos de difracción.

4.11. La señal anómala de los azufres de CD5DIII es suficiente para resolver el problema de las fases.

El hecho de trabajar en el límite de la producción presentaba la imposibilidad de usar selenio-metionina para tratar de resolver el problema de fases ya que, aunque las células hubieran podido sobrevivir al

tratamiento con átomos pesados, la reducción de la expresión de CD5DIII hubiera hecho imposible su purificación. Así, tal y como está descrito en los resultados en la sección 3.3.4, se planteó en un principio el uso de sales de átomos pesados para tratar de resolver el problema de fases. Siguiendo el protocolo descrito por Boggon et al (Boggon 2000) se realizó un cribado de sales que pudieran interactuar con la proteína probando las que presentaban algún resultado que pudiera ser positivo y que pudieran, en teoría, interactuar con la proteína dada su composición de aminoácidos. A pesar de realizar variaciones en la concentración y el tiempo de exposición, los experimentos de difracción con los cristales tratados con átomos pesados dieron como resultado patrones de difracción improcesables debido a la pérdida de orden dentro de la red cristalina.

Se barajó entonces la posibilidad de resolver el problema de fases usando los propios azufres de la proteína mediante la realización de un experimento de SAD. Gracias a la obtención de cristales de alta calidad y al uso de radiación de sincrotrón a una longitud de onda adecuada se consiguió resolver con éxito la estructura de CD5DIII. En base a ella se modeló la estructura del receptor tal y como se describe en la sección de resultados 3.3.8.

4.12. CD5DIII sirve de base para describir un modelo de la interacción entre CD6 y ALCAM.

El desconocimiento de un ligando concreto para el receptor de CD5 hizo que no se pudiera modelar una interacción receptor-ligando del mismo, pero este dominio sí se pudo utilizar para modelar la interacción de CD6 con su ligando natural ALCAM/CD166, una proteína transmembrana de tipo 1 perteneciente a la superfamilia de inmunoglobulinas. CD6 es, al igual que CD5, un receptor presente en timocitos, linfocitos T y células B-1a, perteneciente al grupo B de la familia SRCR (Hardy 1994; Hardy 1994).

Estudios previos de mutagénesis dirigida señalaban los aminoácidos implicados en la interacción CD6-ALCAM (Bodian 1997; Skonier 1997), pero debido a la falta de información estructural acerca de los miembros del grupo B de la familia SRCR no se había podido realizar un modelo detallado de esta interacción. La alta homología entre CD5DIII y CD6DIII permitió

realizar un modelo fiable y altamente descriptivo de cómo puede ser la interacción entre CD6 y su ligando (Rodamilans 2007) (Figura 4.1).

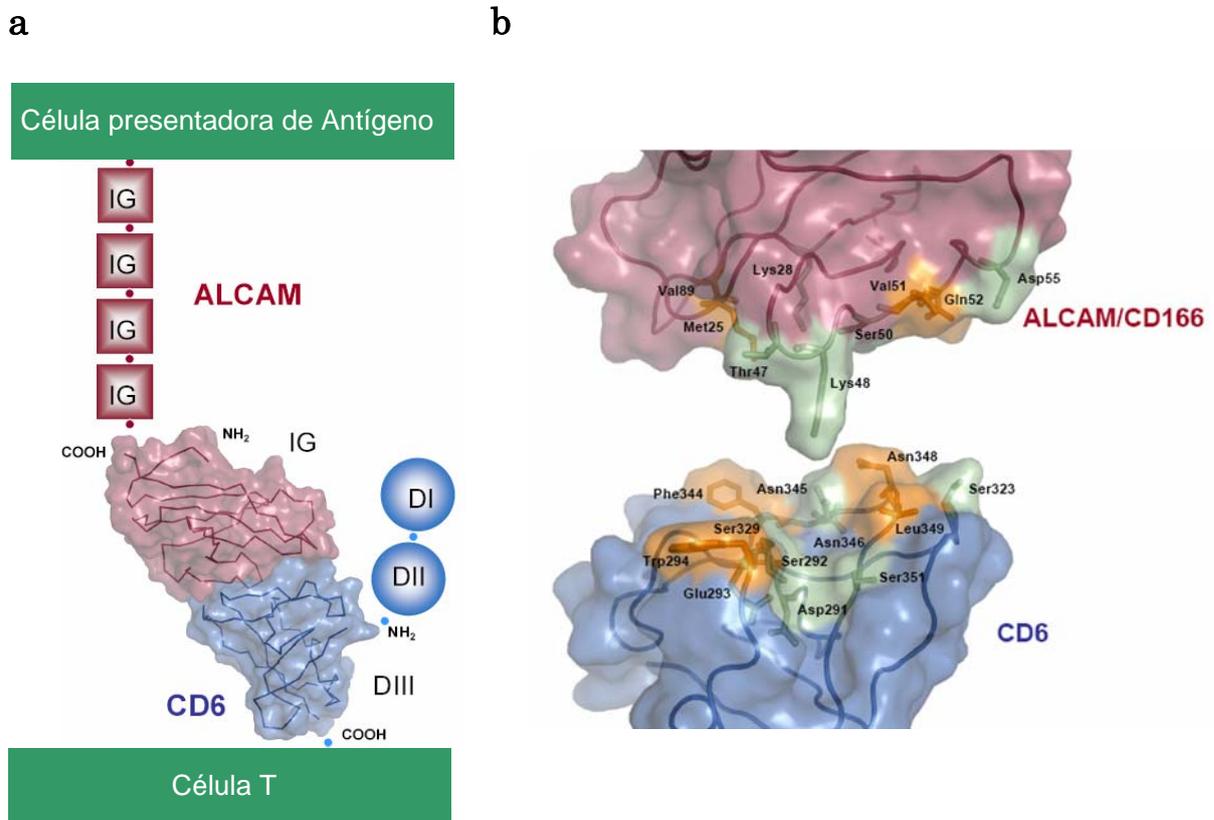


Figura 4.1. Modelo de la interacción CD6-ALCAM/CD166. (a) Detalle de la interacción entre ambos receptores celulares. (b) Interacción detallada a nivel molecular de los aminoácidos que intervienen en la misma.

4.13. ¿Cómo se puede avanzar en el desarrollo del trabajo con CD5?

Una vez resuelta esta primera etapa con el dominio III del receptor se plantean diversas posibilidades. Una de ellas es tratar de cristalizar toda la región extracelular del receptor. En caso de conseguir cristales, la resolución del problema de las fases se facilitaría mucho gracias a que esta vez habría disponible un modelo, la estructura del dominio III, con el que realizar el reemplazo molecular. Otra de las posibilidades es buscar el ligando natural de este receptor y tratar de modelar esta interacción al igual que se hizo para CD6/ALCAM e incluso realizar estudios estructurales tratando de formar un complejo receptor-ligando que pudiera ser cristalizado y resuelto.



CONCLUSIONES

1. Los protocolos descritos para la formación *in vitro* de NLPs a partir de la proteína C recombinante del VHC no se han podido reproducir. Por un lado, se ha intentado sin éxito el replegamiento de la proteína C una vez desnaturalizada. Por otro lado, la purificación de NLPs en su forma nativa no ha aportado resultados claros. Por tanto, nuevas estrategias deberán ser desarrolladas para desentrañar la estructura de la nucleocápsida.
2. La proteína DDX3 se ha conseguido expresar y purificar así como dos fragmentos de la misma, DDX3_{pf} (164-600) y DDX3_{ptr} (169-579). Todas las construcciones conservan la actividad helicasa característica de la enzima.
3. La estructura de DDX3_{hel} se ha resuelto por medio de cristalografía de rayos-X. Este dominio revela un plegamiento muy compacto similar al dominio helicasa de otras estructuras homólogas. De esta manera se ha podido desarrollar un modelo de los dominios ATPasa y helicasa de DDX3 en base a la proteína Vasa.
4. El dominio III de CD5 se ha resuelto por medio de cristalografía de rayos X. Representa el primer ejemplo de un dominio SRCR del grupo B y nos ha ayudado a definir la arquitectura del receptor y a realizar un modelo de la interacción de una proteína de la misma familia, CD6, con su ligando natural ALCAM/CD166.

- Abrahams, J. P., and Leslie, A. G. (1996). "Methods used in the structure determination of bovine mitochondrial F1 ATPase." Acta Crystallogr D Biol Crystallogr **52**(Pt 1): 30-42.
- Alberola-Ila, J., Places, L., Fabregat, V., Vives, J., and Lozano, F. (1993). "Different mechanisms regulate the monoclonal antibody-induced modulation of CD2, CD3, and CD5 in human lymphocytes." Cell Immunol **147**(2): 247-55.
- Alberola-Ila, J., Places, L., Lozano, F., and Vives, J. (1993). "Association of an activation inducible serine kinase activity with CD5." J Immunol **151**(9): 4423-30.
- Angello, V., R. T. Chung, and L. M. Kaplan. (1992). "A role for hepatitis C virus infection in type II cryoglobulinemia." N. Engl. J. Med. **327**: 1490-1495.
- Aruffo, A., Bowen, M. A., Patel, D. D., Haynes, B. F., Starling, G. C., Gebe, J. A., and Bajorath, J. (1997). "CD6-ligand interactions: a paradigm for SRCR domain function?" Immunol Today **18**(10): 498-504.
- Bartenschlager, R., Ahlborn-Laake, L., Mous, J. and Jacobsen, H. (1994). "kinetic and structural analyses of hepatitis C virus polyprotein processing." J. Virol. **68**: 5045-5055.
- Barth, H., Schafer, C., Adah, MI., Zhang, F., Linhardt, RJ., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, TH., Depla, E., Von Weizsacker F, Blum HE, Baumert TF. (2003). "Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate." J Biol Chem **278**(42): 41003-12.
- Bartosch, B., J. Dubuisson, and F. L. Cosset. (2003). "Infectious hepatitis C virus pseudo-particles containing functional E1-E2 envelope protein complexes." J Exp. Med. **197**: 633-642.
- Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A. & Cosset, F.L. (2003). "Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-BI scavenger receptor." J Biol Chem **278**: 41624-41630.
- Baumert, T. F., Ito, S., Wong, D. T., and Liang, T. J. (1998). "Hepatitis C virus structural proteins assemble into viruslike particles in insect cells." J Virol **72**(5): 3827-36.
- Bennett, W., Inoue, Y., Beck, JR., Wong, JB., Pauker, SG., Davis, GI. (1997). "Estimates of the cost-effectiveness of a single course of interferon-alfa 2b in patients with histologically mild chronic hepatitis C." Ann Intern Med **127**: 855-865.
- Berrow, N. S., Bussow K, Coutard B, Diprose J, Ekberg M, Folkers GE, Levy N, Lieu V, Owens RJ, Peleg Y, Pinaglia C, Quevillon-Cheruel S, Salim L, Scheich C, Vincentelli R, Busso D. (2006). "Recombinant protein expression and solubility screening in Escherichia coli: a comparative study." Acta Cryst D **62**(10): 1218-26.

- Biancone, L., Bowen, M. A., Lim, A., Aruffo, A., Andres, G., and Stamenkovic, I. (1996). "Identification of a novel inducible cell-surface ligand of CD5 on activated lymphocytes." *J Exp Med* **184**(3): 811-9.
- Bikah, G., Lynd, F. M., Aruffo, A. A., Ledbetter, J. A., and Bondada, S. (1998). "A role for CD5 in cognate interactions between T cells and B cells, and identification of a novel ligand for CD5." *Int Immunol* **10**(8): 1185-96.
- Birnbaum, F., and Nassal, M. (1990). "Hepatitis B Virus Nucleocapsid Assembly: Primary Structure Requirements in the Core Protein." *Journal of Virology* **64**(7): 3319-3330.
- Blanchard, E., Brand, D., Trassard, S., Goudeau, A., Roingeard, P. (2002). "Hepatitis C virus-like particle morphogenesis." *J Virol* **76**(8): 4073-9.
- Bodian, D. L., Skonier J. E., Bowen, M. A., Neubauer, M., Siadak, A. W., Aruffo, A., and Bajorath, J. (1997). "Identification of residues in CD6 which are critical for ligand binding." *Biochemistry* **36**(9): 2637-41.
- Boggon, T. J. a. S., L. (2000). "Screening for phasing atoms in protein crystallography." *Structure* **8**: R143-R149.
- Bukh, J., Purcell, R. H., and Miller, R. H. (1994). "Sequence analysis of the core gene of 14 hepatitis C virus genotypes." *Proc Natl Acad Sci U S A* **91**(17): 8239-43.
- Burgess, K. E., Yamamoto, M., Prasad, K. V., and Rudd, C. E. (1992). "CD5 acts as a tyrosine kinase substrate within a receptor complex comprising T-cell receptor zeta chain/CD3 and protein-tyrosine kinases p56lck and p59fyn." *Proc Natl Acad Sci U S A* **89**(19): 9311-5.
- Cacoub, P., Poynard, T, Ghillani, P., et al. (1999). "Extrahepatic manifestations of chronic hepatitis C. MULTIVIRC Group. Multidepartment virus C." *Arthritis Rheum* **42**: 2204-2212.
- Calvo, J., Padilla, O., Places, L., Vigorito, E., Vila, J. M., Vilella, R., Mila, J., Vives, J., Bowen, M. A., Lozano, F. (1999). "Relevance of individual CD5 extracellular domains on antibody recognition, glycosylation and co-mitogenic signalling." *Tissue Antigens* **54**(1): 16-26.
- Calvo, J., Places, L., Padilla, O., Vila, J. M., Vives, J., Bowen, M. A., and Lozano, F. (1999). "Interaction of recombinant and natural soluble CD5 forms with an alternative cell surface ligand." *Eur J Immunol* **29**(7): 2119-29.
- Calvo, J., Places, L., Espinosa, G., Padilla, O., Vila, J. M., Villamor, N., Ingelmo, M., Gallart, T., Vives, J., Font, J., and Lozano, F. (1999). "Identification of a natural soluble form of human CD5." *Tissue Antigens* **54**(2): 128-37.
- Calvo, J., Vilda, J. M., Places, L., Simarro, M., Padilla, O., Andreu, D., Campbell, K. S., Aussel, C., Lozano, F. (1998). "Human CD5 signaling and constitutive phosphorylation of C-terminal serine residues by casein kinase II." *J Immunol* **161**(11): 6022-9.
- Calleja, J., Martinez, JL and Albillos, A. (2001). "Tratamiento de la hepatitis crónica por virus C." *Inf ter Sist Nac Salud* **25**: 69-77.

- Clifford, B. D., D. Donahue, L. Smith, E. Cable, B. Luttig, M. Manns, and H. L. Bonkovsky. (1995). "High prevalence of serological markers of autoimmunity in patients with chronic hepatitis C." *Hepatology* **21**: 613–619.
- Cohen, J. (1999). "The scientific challenge of hepatitis C." *Science* **285**(5424): 26-30.
- Conrad, S., Garrett, L.E., Cooksley, W. G.E., Dunne, M.P., and Macdonald, G.A. (2006). "Living with chronic hepatitis C means "you just haven't got a normal life anymore"." *Chronic Illness* **2**: 121-131.
- Cordin, O., , Tanner, N.K., Doere, M., Linder, P. and Banroques, J (2004). "The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity." *The EMBO journal* **23**: 2478-2487.
- Crowther, F. A., Kiselev, I.A. , Ottcher, B.B., Berriman, J. A., Borisova, G. P., Ose, S. V. and Pumpens, P. (1994). "Three-Dimensional Structure of Hepatitis B Virus Core Particles Determined by Electron Cryomicroscopy." *Cell* **77**: 943-950.
- Curry, M. P., Golden-Mason, L., Nolan, N., Parfrey, N. A., Hegarty, J. E., and O'Farrelly, C. (2000). "Expansion of peripheral blood CD5+ B cells is associated with mild disease in chronic hepatitis C virus infection." *J Hepatol* **32**(1): 121-5.
- Chang, P. C., Chi, C. W., Chau, G. Y., Li, F. Y., Tsai, Y. H., Wu, J. C., and Wu Lee, Y. H. (2006). "DDX3, a DEAD box RNA helicase, is deregulated in hepatitis virus-associated hepatocellular carcinoma and is involved in cell growth control." *Oncogene* **25**(14): 1991-2003.
- Chao, C. H., Chen, C. M., Cheng, P. L., Shih, J. W., Tsou, A. P. and Lee, Y. H. (2006). "DDX3, a DEAD box RNA helicase with tumor growth-suppressive property and transcriptional regulation activity of the p21waf1/cip1 promoter, is a candidate tumor suppressor." *Cancer Research* **66**(13): 6579-88.
- Dauphinee, M., Tovar, Z., and Talal, N. (1988). "B cells expressing CD5 are increased in Sjogren's syndrome." *Arthritis Rheum* **31**(5): 642-7.
- Davies, A. A., Ley, S. C., and Crumpton, M. J. (1992). "CD5 is phosphorylated on tyrosine after stimulation of the T-cell antigen receptor complex." *Proc Natl Acad Sci U S A* **89**(14): 6368-72.
- de la Cruz, J., Kressler, D., and Linder, P. (1999). "Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families." *Trends Biochem Sci* **24**(5): 192-8.
- de la Fortelle, E., and Bricogne, G. (1997). *Macromolecular Crystallography*. New York, Academic Press.
- Debreczeni, J. E., Bunkóczi, G., Ma, Q., Blaser, H. and Sheldrick, M. (2003). "In-house measurements of the sulfur anomalous signal and its use for phasing." *Acta Cryst D* **59**: 688-696.
- Dennehy, K. M., Broszeit, R., Ferris, W. F., and Beyers, A. D. (1998). "Thymocyte activation induces the association of the proto-

- oncoprotein c-cbl and ras GTPase-activating protein with CD5." Eur J Immunol **28**(5): 1617-25.
- Diedrich, G. (2006). "How does hepatitis C virus enter cells?" FEBS journal **273**: 3871-3885.
- Drummer, H. E., A. Maerz, and P. Poubourios (2003). "Cell surface expression of functional hepatitis C virus E1 and E2 glycoproteins." FEBS Lett. **546**: 385-390.
- Dubois, F., Désenclos, J.C., Mariotte, N., Goudeau, A. (1997). "Hepatitis C in a French population-based survey, 1994: seroprevalence, frequency of viremia, genotype distribution, and risk factors. The Collaborative Study Group." Hepatology **25**: 1490-1496.
- Ericsson, U., Hallberg BM, Detitta GT, Dekker N, Nordlund P. (2006). "Thermofluor-based high-throughput stability optimization of proteins for structural studies." Anal Biochem. **357**(2): 289-98.
- Falcon, V., Garcia, C., de la Rosa, M. C., Menendez, I., Seoane, J., and Grillo, J. M. (1999). "Ultrastructural and immunocytochemical evidences of core-particle formation in the methylotrophic *Pichia pastoris* yeast when expressing HCV structural proteins (core-E1)." Tissue Cell **31**(2): 117-25.
- Flint, M., Maidens C, Loomis-Price LD, Shotton C, Dubuisson J, Monk P, Higginbottom A, Levy S & McKeating JA (1999). "Characterization of hepatitis C virus E2 glycoprotein interaction with a putative cellular receptor, CD81." J Virol **73**: 6235-6244.
- Flint, M., von Hahn, T., Zhang, J., Farquhar, M., Jones, C.T., Balfe, P., Rice, C.M., and McKeating JA (2006). "Diverse CD81 Proteins Support Hepatitis C Virus Infection." J Virol **80**(22): 11331-11342.
- Fraenkel-Conrat, H. a. W., R. C. (1955). "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components." PNAS **41**(10): 690-698.
- Frank, C., Mohamed MK, Strickland GT et al. (2000). "The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt." Lancet **355**: 887-891.
- Gardner, J. P., R. J. Durso, R. R. Arrigale, G. P. Donovan, P. J. Maddon, T., Dragic, and W. C. Olson. (2003). "L-SIGN (CD 209L) is a liver-specific capture receptor for hepatitis C virus." Proc. Natl. Acad. Sci. USA **100**: 4498-4503.
- Germi, R., Crance, J.M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W., Zarski, J.P., Drouet, E. (2002). "Heparan sulfate-mediated binding of infectious dengue virus type 2 and yellow fever virus." Virology **292**(1): 162-8.
- Germi, R., Crance, J.M., Garin, D., Guimet, J., Lortat-Jacob, H., Ruigrok, R.W., Zarski, J.P., Drouet, E. (2002). "Cellular glycosaminoglycans and low density lipoprotein receptor are involved in hepatitis C virus adsorption." J Med Virol **68**(2): 206-15.

- Grakoui, A., Wychowski, C., Lin, C., Feinstone, SM. and Rice, CM. (1993). "Expression and identification of hepatitis C virus polyprotein cleavage products." J. Virol. **67**: 1385-1395.
- Griffith, L. C., Schulman, H., and Tagawa, M. (1989). "In vivo and in vitro phosphorylation of murine lymphocyte differentiation antigen CD5." Biochem Biophys Res Commun **159**(2): 536-41.
- Guadagnino, V., Stroffolini T, Rapicetta M et al. (1997). "Prevalence, risk factors, and genotype distribution of hepatitis C virus infection in the general population: a community-based survey in southern Italy." Hepatology **26**: 1006–1011.
- Guex, N. a. P., M. C. (1997). "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling." Electrophoresis **18**: 2714-2723.
- Hadziyannis, S. J. (1997). "The spectrum of extrahepatic manifestations in hepatitis C virus infection." J. Viral Hepat. **4**: 9–28.
- Hardy, R. R., and Hayakawa, K. (1994). "CD5 B cells, a fetal B cell lineage." Adv Immunol **55**: 297-339.
- Hardy, R. R., Carmack, C. E., Li, Y. S., and Hayakawa, K. (1994). "Distinctive developmental origins and specificities of murine CD5+ B cells." Immunol Rev **137**: 91-118.
- Heckmann, J., Hayser, C., Heuss, D., et al. (1999). "Neurological manifestations of chronic hepatitis C." J Neurol **246**: 486-491.
- Hendrickson, W. A. a. T., M.M. (1981). "Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur " Nature **290**(107-113).
- Higginbottom, A., Quinn ER, Kuo CC, Flint M, Wilson LH, Bianchi E, Nicosia A, Monk PN, McKeating JA & Levy S (2000). " Identification of amino acid residues in CD81 critical for interaction with hepatitis C virus envelope glycoprotein E2." J Virol **74**: 3642–3649.
- Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohna, K. (1991). "Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis." Proc Natl Acad Sci USA **88**: 5547-5551.
- Hoffnagle, J. H. (2002). "Course and outcome of Hepatitis C." Hepatology **35**: S21-S29.
- Hohenester, E., Sasaki, T., and Timpl, R. (1999). "Crystal structure of a scavenger receptor cysteine-rich domain sheds light on an ancient superfamily." Nat Struct Biol **6**(3): 228-32.
- Holm, L. a. P., J. (2000). "DaliLite workbench for protein structure comparison." Bioinformatics **16**: 566-567.
- Hsu, M., J. Zhang, M. Flint, C. Logvinoff, C. Cheng-Mayer, C. M. Rice, and J. A. McKeating. (2003). "Hepatitis C virus glycoproteins mediate pH-dependent cell entry of pseudotyped retroviral particles." Proc. Natl. Acad. Sci. USA **100**: 7271–7276.

- Huarte, M. P., Casi, M. A. (2004). "Virology, diagnostic tests, epidemiology and transmission mechanisms of Hepatitis C virus." An.Sist.Sanit.Navar **27**(2): 41-50.
- Inghirami, G., Foitl, D. R., Sabichi, A., Zhu, B. Y., Knowles, D. M. (1991). "Autoantibody-associated cross-reactive idiotype-bearing human B lymphocytes: distribution and characterization, including Ig VH gene and CD5 antigen expression." Blood **78**(6): 1503-15.
- Iost, I., Dreyfus, M., and Linder, P. (1999). "Ded1p, a DEAD-box protein required for translation initiation in *Saccharomyces cerevisiae*, is an RNA helicase." J Biol Chem **274**(25): 17677-83.
- Jin, D. Y., H. L. Wang, et al. (2000). "Hepatitis C virus core protein-induced loss of LZIP function correlates with cellular transformation." Embo J **19**(4): 729-40.
- Jin, D. Y., Wang, H. L., Zhou, Y., Chun, A. C., Kibler, K. V., Hou, Y. D., Kung, H., Jeang, K. T. (2000). "Hepatitis C virus core protein-induced loss of LZIP function correlates with cellular transformation." Embo J **19**(4): 729-40.
- Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). "Kinesin transports RNA: isolation and characterization of an RNA-transporting granule.[see comment]." Neuron **43**(4): 513-25.
- Karmochkine, M., Carrat, F., Dos Santos, O., Cacoub, P. and Raguin, G. (2006). "A Case-Control Study of Risk Factors for Hepatitis C Infection in Patients With Unexplained Routes of Infection." J Viral Hepat **13**(11): 775-782.
- Kohfeldt, E., Maurer, P., Vannahme, C., and Timpl, R. (1997). "Properties of the extracellular calcium binding module of the proteoglycan testican." FEBS Lett **414**(3): 557-61.
- Kohtz, J. D., Jamison, S.F., Will, C.L. Zuo, P., Luhrmann, R., Garcia-Blanco, M.A. and Manley, J.L. (1994). "Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors." Nature **368**: 119-124.
- Krishnan, V., and Zeichner, S. L. (2004). "Alterations in the expression of DEAD-box and other RNA binding proteins during HIV-1 replication." Retrovirology **1**(1): 42.
- Kunkel, M. L., M. Rijnbrand, R. Lemon, S. M. Watowich, S. J. (2001). "Self-assembly of nucleocapsid-like particles from recombinant hepatitis C virus core protein." J Virol **75**(5): 2119-29.
- Kuo, G. C., QL., Alter HJ et al. (1989). "An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis." Science **244**: 362-364.
- Lahn, B., and Page, DC. (1997). "Functional coherence of the human Y chromosome." Science **278**: 675-80.
- Lavillette, D., Tarr AW, Voisset C, Donot P, Bartosch B, Bain C, Patel AH, Dubuisson J, Ball JK & Cosset FL (2005). "Characterization of host-range and cell entry properties of the major genotypes and subtypes of hepatitis C virus. ." Hepatology **41**: 265-274.

- Levitt, D. G. (2001). "A new software routine that automates the fitting of protein X-ray crystallographic electron-density maps." Acta Crystallogr D Biol Crystallogr **57**(Pt 7): 1013-9.
- Levy, S., Todd SC & Maecker HT (1998). "CD81 (TAPA-1): a molecule involved in signal transduction and cell adhesion in the immune system." Annu Rev Immunol **16**: 89–109.
- Lindenbach, B., Evans MJ, Syder AJ, Wolk B, Tellinghuisen TL, Liu CC, Maruyama T, Hynes RO, Burton DR, McKeating JA et al. (2005). "Complete replication of hepatitis C virus in cell culture." Science **309**: 623–626.
- Lo, M., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004). "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery." Anal Biochem. **332**(1): 153-9.
- Lorenzo, L. J. D.-C., S. Falcon, V. Acosta-Rivero, N. Gonzalez, E. de la Rosa, M. C. Menendez, I. Morales, J. (2001). "Assembly of truncated HCV core antigen into virus-like particles in *Escherichia coli*." Biochem Biophys Res Commun **281**(4): 962-5.
- Lozach, P. Y., H. Lortat-Jacob, A. de Lacroix de Lavalette, I. Staropoli, S. Foug, A. Amara, C. Houles, F. Fieschi, O. Schwartz, J. L. Virelizier, F. Arenzana-Seisdedos, and R. Altmeyer. (2003). "DC-SIGN and L-SIGN are high affinity binding receptors for hepatitis C virus glycoprotein E2." J. Biol. Chem **278**: 20358–20366.
- Lozano, F., Alberola-Ila, J., Places, L., Gallart, T., and Vives, J. (1990). "Phosphorylation-mediated changes in the electrophoretic mobility of CD5 molecules." Eur J Biochem **193**(2): 469-77.
- Lukavsky, P. J., G. A. Otto, et al. (2000). "Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function." Nat Struct Biol **12**: 1105-10.
- Luo, W., Van de Velde, H., von Hoegen, I., Parnes, J. R., and Thielemans, K. (1992). "Ly-1 (CD5), a membrane glycoprotein of mouse T lymphocytes and a subset of B cells, is a natural ligand of the B cell surface protein Lyb-2 (CD72)." J Immunol **148**(6): 1630-4.
- Majeau, N., V. Gagne, et al. (2004). "The N-terminal half of the core protein of hepatitis C virus is sufficient for nucleocapsid formation." J Gen Virol **85**(Pt 4): 971-81.
- Mamiya, N., Worman, H. J. (1999). "Hepatitis C virus core protein binds to a DEAD box RNA helicase." J Biol Chem **274**(22): 15751-6.
- Mamiya, N. a. H. W. (1999). "Hepatitis C Virus Core protein Binds to a DEAD Box RNA Helicase." The Journal of Biological Chemistry **274**(28): 15751-15756.
- Marcellin, P., V. Descamps, M. Marinot-Peignoux, D. Larzul, L. Xu, N., Boyer, B. N. Pham, B. Cricks, L. Guillevin, and S. Belaich. (1993). "Cryoglobulinemia with vasculitis associated with hepatitis C virus infection." Gastroenterology **104**: 272–277.

- McAlister, M. S., Brown, M. H., Willis, A. C., Rudd, P. M., Harvey, D. J., Aplin, R., Shotton, D. M., Dwek, R. A., Barclay, A. N., and Driscoll, P. C. (1998). "Structural analysis of the CD5 antigen-expression, disulphide bond analysis and physical characterisation of CD5 scavenger receptor superfamily domain 1." Eur J Biochem **257**(1): 131-41.
- McKeating, J., Zhang, LQ., Logvinoff, C., Flint, M., Zhang, J., Yu, L., Butera, D., Ho, DD., Dustin, LB., Rice, CM. et al. (2004). "Diverse hepatitis C virus glycoproteins mediate viral infection in a CD81-dependent manner." J Virol **78**: 8496–8505.
- Misiani, R., P. Bellavita, D. Fenili, G. Borelli, D. Marchesi, M. Massazza, G., Vendramin, B. Comotti, E. Zanzi, and G. Scudeller. (1992). "Hepatitis C infection in patients with essential mixed cryoglobulinemia." Ann. Intern. Med **117**: 573–577.
- Monteverde, A., M. Ballare, and S. Pileri. (1997). "Hepatic lymphoid aggregates in chronic hepatitis C and mixed cryoglobulinemia. Springer Semin." Immunopathol **19**: 99–110.
- Moradpour, D., Cerny, A., Heim, M and Blum, H.E. (2001). "Hepatitis C: an update." Swiss med Wkly **131**: 291-298.
- Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003). "ARP/wARP and automatic interpretation of protein electron density maps." Methods Enzymol **374**: 229-44.
- Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.
- Nattermann, J., Zimmermann, H., Iwan, A., von Lilienfeld-Toal, M., Leifeld, L., Nischalke, HD., Langhans, B., Sauerbruch, T, and Spengler, U. (2006). "Hepatitis C Virus E2 and CD81 Interaction May Be Associated With Altered Trafficking of Dendritic Cells in Chronic Hepatitis C." Hepatology(44): 945-954.
- Nolandt, O., Kern, V., Muller, H., Pfaff, E., Theilmann, L., Welker, R., and Krausslich, H. G. (1997). "Analysis of hepatitis C virus core protein interaction domains." J Gen Virol **78 (Pt 6)**: 1331-40.
- Otwinowski, Z., and Minor, W. (1997). "" Processing of X-ray Diffraction Data Collected in Oscillation Mode "." Methods in Enzymology **276**(Macromolecular Crystallography, part A): 307-326.
- Owsianka, A. M., and Patel, A. H. (1999). "Hepatitis C virus core protein interacts with a human DEAD box protein DDX3." Virology **257**(2): 330-40.
- Petracca, R. F., F. Galli, G., et al. (2000). "Structure-function analysis of hepatitis C virus envelope-CD81 binding." J Virol **74**: 4824-4830.
- Pietrogrande, M., M. Corona, S. Milani, A. Rosti, M. Ramella, and G. Tordato. (1995). "Relationship between rheumatoid factor and the immune response against hepatitis C virus in essential mixed cryoglobulinemia." Exp. Rheumatol. **13**(13): 109–113.

- Pileri, P., Y. Uematsu, S. Campagnoli, G. Galli, F. Falugi, R. Petracca, A. J., Weiner, M. Houghton, D. Rosa, G. Grandi, and S. Abrignani. (1998). "Binding of hepatitis C virus to CD81. ." Science **282**: 938–941.
- Plater Z., M., R. N., Lam, K., Kennedy, T. D., and G. Janossy. (1985). "A rheumatoid arthritis B cell subset expresses a phenotype similar to that in chronic lymphocytic leukemia." Arthritis Rheum. **28**: 971–976.
- Pohlmann, S., J. Zhang, F. Baribaud, Z. Chen, G. J. Leslie, G. Lin, A. Granelli-Piperno, R. W. Doms, C. M. Rice, and J. A. McKeating. (2003). "Hepatitis C virus glycoproteins interact with DC-SIGN and DC-SIGNR." J. Virol. **77**: 4070–4080.
- Pospisil, R., Fitts, M. G., and Mage, R. G. (1996). "CD5 is a potential selecting ligand for B cell surface immunoglobulin framework region sequences." J Exp Med **184**(4): 1279-84.
- Pospisil, R., Silverman, G. J., Marti, G. E., Aruffo, A., Bowen, M. A., and Mage, R. G. (2000). "CD5 is A potential selecting ligand for B-cell surface immunoglobulin: a possible role in maintenance and selective expansion of normal and malignant B cells." Leuk Lymphoma **36**(3-4): 353-65.
- Raab, M., Yamamoto, M., and Rudd, C. E. (1994). "The T-cell antigen CD5 acts as a receptor and substrate for the protein-tyrosine kinase p56lck." Mol Cell Biol **14**(5): 2862-70.
- Radivojac, P., Obradovic, Z., Brown, C.J., Dunker, A.K. (2003). "Prediction of boundaries between intrinsically ordered and disordered protein regions." Pac. Symp. Biocomput. **8**: 216-227.
- Ralston, R., Thudium, K., Berger, K., Kuo, C., Gervase, B., Hall, J., Selby, M., Kuo, G., Houghton, M., and Choo, Q.-L. (1993). "Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses." J. Virol **67**: 6753-6761.
- Raman, C., Kuo, A., Deshane, J., Litchfield, D. W., and Kimberly, R. P. (1998). "Regulation of casein kinase 2 by direct interaction with cell surface receptor CD5." J Biol Chem **273**(30): 19183-9.
- Resnick, D., Pearson, A., and Krieger, M. (1994). "The SRCR superfamily: a family reminiscent of the Ig superfamily." Trends Biochem Sci **19**(1): 5-8.
- Rocak, S., and Linder, P. (2004). "DEAD-box proteins: the driving forces behind RNA metabolism." Nat Rev Mol Cell Biol **5**(3): 232-41.
- Rocak, S., and Linder, P. (2004). "DEAD-box proteins: the driving forces behind RNA metabolism." Nature Reviews Molecular Cell Biology **5**(3): 232-41.
- Rodamilans, B., Muñoz, I.G., Bragado-Nilsson, E., Sarrias, M.R., Padilla, O., Blanco, F.J., Lozano, F. & Montoya, G. (2007). "Crystal structure of the third extracellular domain of CD5 reveals the fold of a group B scavenger cysteine-rich receptor domain." J. Biol. Chem.
- Sarrias, M. R., Gronlund, J., Padilla, O., Madsen, J., Holmskov, U., and Lozano, F. (2004). "The Scavenger Receptor Cysteine-Rich (SRCR)

- domain: an ancient and highly conserved protein module of the innate immune system." Crit Rev Immunol **24**(1): 1-37.
- Scarselli, E., H. Ansuini, R. Cerino, R. M. Roccasecca, S. Acali, G. Filocamo, C. Traboni, A. Nicosia, R. Cortese, and A. Vitelli. (2002). "The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus." EMBO J **21**: 5017–5025.
- Schwede, T., Kopp, J., Guex, N., and Peitsch, MC. (2003). "SWISS-MODEL: an automated protein homology-modeling server." Nucleic Acids Research **31**: 3381-3385.
- Schwer, B. (2001). "A new twist on RNA helicases: DExH/D box proteins as RNAPases." nature structural biology **8**(2): 113-116.
- Shevchenko, A. e. a. (1996). "Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels. ." Analytical Chemistry. **68**: 850-858.
- Shimizu, Y. K., Feinstone, S. M., Kohara, M., Purcell, R. H., and Yoshikura, H. (1996). "Hepatitis C virus: detection of intracellular virus particles by electron microscopy." Hepatology **23**(2): 205-9.
- Skonier, J. E., Bodian, D. L., Emswiler, J., Bowen, M. A., Aruffo, A., and Bajorath, J. (1997). "Mutational analysis of the CD6 ligand binding domain." Protein Eng **10**(8): 943-7.
- Somoza, J. R., Ho, J. D., Luong, C., Ghate, M., Sprengeler, P. A., Mortara, K., Shrader, W. D., Sperandio, D. Chan, H., McGrath, M. E., Katz, B. A. (2003). "The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain." Structure **11**(9): 1123-31.
- Soza, A. R., López-Lastra, M. (2006). "Hepatitis C in Chile: Burden of the disease." Rev Med Chile **134**: 777-788.
- Story, R. M., Li, H. and Abelson, J.N. (2001). "Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii*." PNAS **98**(4): 1465-1470.
- Takahishi, K., Kishimoto, S., Yoshizawa, H., Okamoto, H., Yoshikawa, A., and Mishiro, A. (1992). "p26 protein and 33-nm particle associated with nucleocapsid of hepatitis C virus recovered from the circulation of infected hosts." Virology **191**: 431-434.
- Tan, Y., Lim SP, Ng P, Goh PY, Lim SG, Tan YH & Hong W (2003). "CD81 engineered with endocytotic signals mediates HCV cell entry: implications for receptor usage by HCV in vivo." Virology **308**: 250–269.
- Tong, M., El-Farra, N.S., Reikes, A.R and Co, R.L. (1995). "Clinical outcomes after transfusion associated hepatitis C." N Engl J Med **332**: 1463-6.
- Tsuji, R. F., G. P. Geba, Y. Wang, K. Kawamoto, L. A. Matis, and P. W. Askenase. (1997). "Required early complement activation in contact sensitivity with generation of local C5-dependent chemotactic activity, and late T cell interferon gamma: a possible initiating role of B cells. ." J. Exp. Med. **186**: 1015–1026.

- Vagin, A., and Teplyakov, A. (2000). "An approach to multi-copy search in molecular replacement." Acta Crystallogr D Biol Crystallogr **56**(Pt 12): 1622-4.
- Van de Velde, H., and Thielemans, K. (1996). "Native soluble CD5 delivers a costimulatory signal to resting human B lymphocytes." Cell Immunol **172**(1): 84-91.
- Van de Velde, H., von Hoegen, I., Luo, W., Parnes, J. R., and Thielemans, K. (1991). "The B-cell surface protein CD72/Lyb-2 is the ligand for CD5." Nature **351**(6328): 662-5.
- Van Regenmortel, M., Fauquet, Cm., Bishop, DHL., Carstens, EB., Estes, MK., Lemon, SM., et al. (2000). "Virus Taxonomy." The VIIth Report of the International Committee on Taxonomy of Viruses. San Diego: Academic Press.
- Vila, J., Padilla, O., Arman, M., Gimferrer, I and .Lozano, F. (2000). "The scavenger receptor cysteine-rich superfamily (SRCR-SF). Structure and function of group B members." Inmunología **19**(4).
- Vila, J. M., Calvo, J., Places, L., Padilla, O., Arman, M., Gimferrer, I., Aussel, C., Vives, J., and Lozano, F. (2001). "Role of two conserved cytoplasmic threonine residues (T410 and T412) in CD5 signaling." J Immunol **166**(1): 396-402.
- Wagner, A., Pieren, M., Schulze-Briese, C., Ballmer-Hofer, K. and Prota, A.E. (2004). "Sulfur-SAD with low resolution data." PSI-Scientific Report **7**: 86-87.
- Wakita, T., Pietschmann T, Kato T, Date T, Miyamoto M, Zhao Z, Murthy K, Habermann A, Krausslich HG, Mizokami M et al. (2005). "Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. ." Nat Med **11**: 791–796.
- Wang, B. C. (1985). "Resolution of Phase Ambiguity in Macromolecular Crystallography." Methods in Enzymology **115** 90-112.
- Weeks, C. M., and Miller, R. (1999). "The design and implementation of SnB v2.0." J. Appl. Cryst **32**: 120-124.
- Williams, I. (1999). "Epidemiology of hepatitis C in the United States." Am J Med **107**(6B): 2S–9S.
- Wu, J. Y. a. M., T. (1993). "Specific interactions between proteins implicated in splice site selection and regulated artemative splicing." Cell **75**: 1061-1070.
- Yagnik, A., Lahm, A., Meola, A., Roccasecca, RM., Ercole, BB., Nicosia, A. & Tramontano, A. (2000). "A model for the hepatitis C virus envelope glycoprotein E2." Proteins **40**: 355–366.
- Yao, N., Reichert P, Taremi SS, Prosis WW, Weber PC. (1999). "Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase." Structure Fold Des **7**: 1353-63.
- Yasui, K., Wakita, T., Tsukiyama-Kohara, K., Funahashi, S-I., Ichikawa, M., Kajita, T., Moradpour, D., Wands, J. R., and Kohara, M. (1998).

- "The native form and maturation process of hepatitis C virus core protein." J. Virol **72**: 6048-6055.
- Yedavalli, V., Neuveut, C., YH. Chi, L. Kleiman and KT. Jeang (2004). "Requirement of DDX3 DEAD Box RNA Helicase fo HIV-1 Rev-RRE Export Function." Cell **119**: 381-392.
- You, L., Chen, CM., Yeh, TS., Tsai, TY., Mai, RT., Lin, CH., and Lee, YH. (1999). "Hepatitis C Virus Core protein Interacts with Cellular Putative RNA helicase." Journal of Virology **73**(4): 2841-2853.
- Zhang, J., Randall, G., Higginbottom, A., Monk, P., Rice, CM. & McKeating, JA. (2004). "CD81 is required for hepatitis C virus glycoprotein-mediated viral infection." J Virol **78**: 1448–1455.
- Zhong, J., Gastaminza P, Cheng G, Kapadia S, Kato T, Burton DR, Wieland SF, Uprichard SL, Wakita T & Chisari FV (2005). "Robust hepatitis C virus infection in vitro." Proc Natl Acad Sci USA **102**: 9294–9299.
- Zhou, Z., LJ. Licklider, SP. Gygi and R. Reed (2002). "Comprehensive proteomic analysis of the human spliceosome." Nature **419**: 182-185.
- Zuckerman, E., Kessel, A., Slobodin, G., Sabo, E., Yeshurun, D., and Toubi, E. (2003). "Antiviral treatment down-regulates peripheral B-cell CD81 expression and CD5 expansion in chronic hepatitis C virus infection." J Virol **77**(19): 10432-6.
- Zuckerman, E., Slobodin, G., Kessel, A., Sabo, E., Yeshurun, D., Halas, K., and Toubi, E. (2002). "Peripheral B-cell CD5 expansion and CD81 overexpression and their association with disease severity and autoimmune markers in chronic hepatitis C virus infection." Clin Exp Immunol **128**(2): 353-8.

APÉNDICE

Publicaciones y manuscritos enviados del autor de la presente memoria de tesis doctoral:

Rodamilans, B., Muñoz, I.G., Bragado-Nilsson, E., Sarrias, M.R., Padilla, O., Blanco, F.J., Lozano, F. & Montoya, G. (2007). "Crystal structure of the third extracellular domain of CD5 reveals the fold of a group B scavenger cysteine-rich receptor domain." J. Biol. Chem. **282**(17):12669-77

Rodamilans, B., Ibanez, S., Bragado-Nilsson, E., Sarrias, MR., Lozano, F., Blanco, FJ., & Montoya, G. (2007). "Expression, purification and crystallization of human CD5 domain III, a nano-scale crystallization example." J. Struct. Biol.

Rodamilans, B., & Montoya, G. (2007). "Expression, purification, crystallization and preliminary X-ray diffraction analysis of the DDX3 RNA helicase domain." Acta Crist. F **63**:283-6.

Rodamilans, B., & Montoya, G. (2007). "Structure of the helicase domain of the tumour supresor DDX3." In preparation.

Crystal Structure of the Third Extracellular Domain of CD5 Reveals the Fold of a Group B Scavenger Cysteine-rich Receptor Domain^{*[5]}

Received for publication, December 21, 2006, and in revised form, February 21, 2007. Published, JBC Papers in Press, February 23, 2007, DOI 10.1074/jbc.M611699200

Bernardo Rodamilans^{†1}, Inés G. Muñoz[‡], Elisabeth Bragado-Nilsson[‡], María Rosa Sarrías[§], Olga Padilla[§], Francisco J. Blanco[¶], Francisco Lozano^{§||}, and Guillermo Montoya^{‡2}

From the Spanish National Cancer Center (CNIO), Structural Biology and Biocomputing Program, [‡]Macromolecular Crystallography Group and [¶]NMR Group, c/Melchor Fernandez Almagro 3, 28029 Madrid, Spain, [§]Department of Immunology, Hospital Clínic of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Villarroel 170, 08036 Barcelona, Spain, and ^{||}Department of Cell Biology and Pathology, Faculty of Medicine, University of Barcelona, Casanova 143, 08036 Barcelona, Spain

Scavenger receptor cysteine-rich (SRCR) domains are ancient protein modules widely found among cell surface and secreted proteins of the innate and adaptive immune system, where they mediate ligand binding. We have solved the crystal structure at 2.2 Å of resolution of the SRCR CD5 domain III, a human lymphocyte receptor involved in the modulation of antigen specific receptor-mediated T cell activation and differentiation signals. The first structure of a member of a group B SRCR domain reveals the fold of this ancient protein module into a central core formed by two antiparallel β -sheets and one α -helix, illustrating the conserved core at the protein level of genes coding for group A and B members of the SRCR superfamily. The novel SRCR group B structure permits the interpretation of site-directed mutagenesis data on the binding of activated leukocyte cell adhesion molecule (ALCAM/CD166) binding to CD6, a closely related lymphocyte receptor homologue to CD5.

The SRCR³ domain is an ancient and highly conserved protein module that defines a superfamily of either soluble or membrane-bound receptors expressed by hematopoietic and non-hematopoietic cells, at either embryonic or adult stages (1, 2). The existence of two types of SRCR domains allows the division of the SRCR superfamily in two groups. Members of group A contain six cysteine residues and are encoded by no

less than two exons, whereas those of group B contain eight cysteines and are encoded by a single exon. This different exon-intron organization indicates that group A and B members might have evolved from different ancestral genes. In both cases all the cysteines are engaged in intradomain disulfide bridges. Group A members are usually present as multidomain mosaic proteins containing a single SRCR domain associated to other functional domains. On the contrary, group B members are mostly composed of tandem repeats of SRCR domains (1, 2). Members of these groups are found in different animal species, from low invertebrates to high vertebrates. Although no common function has been defined for SRCR superfamily members, several evidences in conjunction with the high degree of structural and phylogenetic conservation of the SRCR family suggest an important role in basic homeostatic functions including innate immune defense.

Human CD5 and CD6 are lymphocyte surface receptors composed of group B SRCR domains. The contiguous location and the conserved structure of their genes suggest that they may have evolved from duplication of a primordial gene. They act as accessory molecules expressed in thymocytes, mature peripheral T lymphocytes, and a subpopulation of peripheral B lymphocytes (B-1a cells). B-1a cells have been related to the production of auto-antibodies, and this subpopulation is expanded in certain autoimmune diseases and in B-cell chronic lymphocytic leukemias (3, 4).

Ligands of the SRCR superfamily have in general been difficult to identify despite their important biological function. Several reports have proposed a role for CD5 in cognate interactions between T and B cells (5) (6) from the reported interactions between CD5 and distinct B-cell surface proteins such as CD72 (7, 8), gp40–80 (6, 9), Ig Vh framework region sequences (10, 11), and gp150 (12). However, the ultimate nature of the CD5 ligand is still a controversial matter.

On the contrary it is widely accepted that CD6 specifically binds to ALCAM/CD166, a type I transmembrane protein belonging to the immunoglobulin superfamily (13, 14). ALCAM/CD166 extracellular region consists of five Ig-like domains, and its identification as a ligand of CD6 has triggered extensive efforts to study this interaction. Biochemical studies have identified domain III (DIII) of CD6 and the domain I (DI) of ALCAM/CD166 as responsible for the interaction between

^{*} Financial support was obtained through Ministerio de Educación y Ciencia Grants BFU-2005-02403, GEN2003-20642-C09-02 (to G. M.), BIO2003-02246 (to F. J. B.), SAF2004-3251 (to F. L.) as well as Ministerio de Sanidad y Consumo Grant FIS CP05/100 (to M. R. S.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

^[5] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental Figs. 1 and 2.

The atomic coordinates and structure factors (code 2OTT and 2JA4) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

¹ Supported by a Consejería de Educación of the Comunidad Autónoma de Madrid and Fondo Social Europeo predoctoral fellowship.

² To whom correspondence should be addressed. Tel.: 34-912246986; Fax: 34-912246976; E-mail: gmontoya@cnio.es.

³ The abbreviations used are: SRCR, scavenger receptor cysteine-rich; ALCAM, activated leukocyte cell adhesion molecule; HEK, human embryonic kidney; EBNA, Epstein-Barr virus nuclear antigen; PBS, phosphate-buffered saline; M2bp, Mac-2-binding protein; IgV, immunoglobulin variable; mAb, monoclonal antibody; DIII, domain III; SAD, single anomalous dispersion.

Crystal Structure of Human CD5 Domain III

these two molecules (15, 16). Although site-directed mutagenesis studies have allowed the mapping of the residues responsible for this interaction (17–19), no detailed structural model of the CD6DIII-ALCAM/CD166DI interaction has been proposed due to the lack of structural information of the SRCR group B family.

Here we present the crystallographic structure of the human CD5 receptor DIII, the first structure solved of a group B SRCR family member. The structure unravels the network of disulfide bridges that form this folding module. The architecture of the CD5 membrane receptor could be modeled based on its domain composition. Besides, the high homology between CD5DIII and CD6DIII permits the use of our structure as a template for a CD6DIII model which sheds light onto the CD6-ALCAM/CD166 interaction.

EXPERIMENTAL PROCEDURES

CD5 Domain III Expression—The recombinant form of the human CD5DIII was expressed using an episomal expression system in human embryonic kidney cells (HEK 293-EBNA). These cells constitutively express the Epstein-Barr viral protein EBNA-1, allowing episomal replication of the pCEP-Pu vector, a kind gift from Drs T. Sasaki and R. Timpl (Max Planck Institute for Biochemistry, Martinsried, Germany) (20). The cDNA sequence coding the human CD5DIII was obtained by PCR amplification (12) with the 5'-GCCCGCTAGCTTCCAGC-CCAAGGTGCAG-3' (forward), and 5'-GTGGATCCTA-ATCCTGGCATGTGACAAAC-3' (reverse) primers. The forward primer incorporated the NheI restriction site, whereas the reverse primer incorporated a stop codon followed by the BamHI restriction site (underlined). The amplified PCR products were cloned into a digested pCEP-Pu vector, in-frame with the BM-40 leader sequence.

The resulting constructs were expressed into HEK 293-EBNA cells as described before (20). Briefly, HEK 293-EBNA cells in culture dishes were transfected using Lipofectamine and then grown in monolayers using DMEM:F-12 (1:1) plus Glutamax (Invitrogen) supplemented with fetal bovine serum (10%; Sigma), penicillin-streptomycin (100 units/ml and 100 μ g/ml; Invitrogen), Geneticin (250 μ g/ml; Invitrogen), and puromycin (2.5 μ g/ml; Sigma). For large scale protein production, cells in exponential growth phase were trypsinized and plated finally in Costar culture flasks (162 cm²). The medium was exchanged every 48 h, and when 80% of the surface was covered, new media without fetal bovine serum was added. The expressed protein was collected in the culture media 3–4 days after this change.

Every 2 days the media was collected, pooled in 50-ml falcon tubes, and spun at 1000 \times g in a centrifuge for 5 min to discard cell debris. The clean supernatant was transferred into a new falcon tube, flash-frozen in liquid nitrogen, and stored at -80 °C. Around 2.5–3 liters of supernatant containing around 200–300 μ g/liter of protein were collected in 3–4 weeks. At that time cells started to detach from the surface, and the culture was stopped.

Protein Purification—CD5DIII was purified at 4 °C by affinity chromatography using a mouse monoclonal antibody (83-C4) specific to CD5DIII (21). Around 4 liters of supernatant from

the hybridoma-producing 83-C4 mAb were precipitated adding 50% (NH₄)₂SO₄ with gentle stirring for 30 min on ice. Subsequently, the mixture was centrifuged at 6000 \times g for 30 min. The pellet containing the antibody was resuspended in 50 ml of PBS (0.137 M NaCl, 0.01 M Na₂HPO₄·2H₂O, 2.7 mM KCl, 2 mM KH₂PO₄) and then extensively dialyzed against PBS to eliminate the residual (NH₄)₂SO₄. After dialysis the sample was filtered with a 0.22- μ m filter (Millipore). 83-C4 mAb was purified using a 5-ml HiTrap Protein G column (Amersham Biosciences) attached to an Äkta prime. The column was equilibrated with PBS, and the sample was loaded onto the column. After extensive washing with PBS, the 83-C4 mAb was eluted using 0.1 M glycine, pH 2.6. The tubes in the fraction collector contained 200 μ l of 1 M Tris, pH 9.0, to avoid denaturation. Fractions containing the antibody were pooled and dialyzed against 0.1 M NaHCO₃, pH 8.3, plus 0.5 M NaCl. The usual yield was around 100 mg of 83-C4 mAb. Once the antibody was isolated, it was loaded onto a CNBr-Sephacryl 4B resin (Amersham Biosciences) previously activated with several washes of 1 mM HCl and washed once with the coupling solution (0.1 M NaHCO₃, pH 8.3, 0.5 M NaCl), and then the resin was mixed with the ligand and incubated overnight at 4 °C using a rotating wheel. Finally the excess of 83-C4 mAb was washed out with coupling buffer, and the remaining active groups were blocked by incubation with 0.1 M Tris-HCl, pH 8.0, for 2 h. The resin was used to pack a column which was attached to an Äkta prime for the purification process. The supernatant from 2.5–3 liters of HEK 293-EBNA-CD5DIII culture was used for purification. This volume was concentrated at 4 °C in an Amicon cell to 100–200 ml and then loaded onto a 10-ml 83-C4 mAb column previously equilibrated with PBS. The sample was recirculated 10 times to improve binding. Subsequently, the column was washed with 10 volumes of PBS plus 0.5 M NaCl and 1% Nonidet P-40 and then with 10 volumes of PBS. Finally, the protein was eluted with 3.5 M MgCl₂ in PBS. All the chromatographic steps were performed at a 1.5 ml/min flow rate. The amount of purified protein depended on the collected supernatant and ranged from 200 to 500 μ g from 2.5–3 liters of culture. Finally the elution buffer was exchanged to 0.01 M HEPES, 0.2 M NaCl, pH 7.4, using a desalting column (Amersham Biosciences). The purified protein was concentrated to 4 mg/ml at 4 °C using a Centricon device with a 10-kDa cut-off. The purity of the sample after the isolation process was tested by SDS-PAGE using Coomassie and silver staining (22). The mass spectrometry analysis of the purified CD5DIII revealed the protein exact molecular weight. Additional experiments in the presence of a reducing agent (dithiothreitol) showed the existence of four disulfide bridges involving the eight cysteines (Cys-1—Cys-4, Cys-2—Cys-7, Cys-3—Cys-8, and Cys-5—Cys-6) in the recombinant protein (data not shown). These data together with the sharp and disperse signals observed in the one-dimensional ¹H NMR spectrum recorded in a Bruker 600 MHz spectrometer indicated that the SRCR domain folding was well defined (data not shown). The production of good quality CD5DIII in HEK 293-EBNA cells contrasts with other attempts to express CD5 domains in a heterologous system (*Pichia pastoris*) (23), when this recombinant protein was used for NMR studies did not yield well defined spectra.

TABLE 1
Data collection, structure solution, and refinement statistics

Data collection	Tetragonal	Tetragonal	Trigonal
	Sulfur-SAD		
Number of crystals	1	1	1
Temperature (K)	100	100	100
Environment	MAR225, SLS, PX beamline	ADSC-Q315, ESRF, ID-14-4	MAR225, SLS, PX beamline
Wavelength (Å)	1.80	0.97	0.93
Space group	P4 ₁ 22	P4 ₁ 22	P3 ₂ 2 ₁
Cell dimensions (Å, °)	$a = b = 74.420, c = 128.824,$ $\alpha\beta\gamma = 90^\circ$	$a = b = 74.161, c = 128.992,$ $\alpha\beta\gamma = 90^\circ$	$a = b = 62.93, c = 72.56, \alpha\beta = 90^\circ,$ $\gamma = 120^\circ$
No. mol in the asymmetric unit	2	2	1
Resolution (Å)	55.3–2.8	50–2.5	50–2.2
Unique reflections	9,432	12,892	12,098
Average multiplicity ^a	52 (43)	8.9 (9.1)	5.4 (4.1)
Completeness (%)	100 (99.9)	99.6 (98.9)	96.9 (88.3)
R_{merge}^b	0.09 (0.25)	0.07 (0.34)	0.05 (0.24)
$(I/\sigma(I))$	10.7 (3.1)	6.9 (2.1)	15.3 (2.2)
B-factor Wilson plot	56	57	50
Sulfur-SAD phasing statistics			
No. S-S found	5/8		
FOM (SHARP)	0.35		
RCullis (SHARP)	0.8		
Phasing power	1.06		
Density modification			
FOM (Solomon)	0.85		
Refinement			
No. reflections		12,350	7,899
Resolution range (Å)		40/2.5	20/2.2
R-factor/R-free (%)		22/25.52	19.32/24.75
No. protein atoms (average B, Å ²) ^c		1,382	789
		48.53	44.56
No. water molecules (Average B, Å ²) ^c		80	58
		54.70	47.53
Root mean square bond length (Å)		0.021	0.018
Root mean square bond angle (°)		2.120	2.156
Ramachandran plot outliers (number) ^d		0	0

^a Values in the highest resolution shell are given in parentheses.

^b $R_{\text{merge}} = \sum_i \sum_j |I_{h,i} - \langle I_h \rangle| / \sum_i \sum_j I_{h,i}$.

^c Calculated using MOLEMAN.

^d Calculated using PROCHECK.

Crystallization—Crystallization screenings were performed with a Cartesian Mycrosys SQ robot (Genomic Solutions) using the sitting drop method with nanodrops of 0.05 μl . A total of 1000 conditions from different screens from Hampton Research, Emerald Biosciences, and Qiagen were tested with one batch of the purified protein. Small needles were obtained after 48 h in condition 17 of wizard I (30% polyethylene glycol 8000, 0.1 M sodium acetate, pH 4.5, 0.2 M Li₂SO₄). Three protein batches were necessary for the optimization of the crystallization conditions. After several refinement rounds from the initial crystallization conditions, crystals suitable for x-ray diffraction experiments were obtained. The final crystallization conditions contained 15–25% polyethylene glycol 8000, 0.1 M sodium acetate, pH 4.4, 0.2 M Li₂SO₄ in the reservoir, whereas the drop (hanging or sitting) only contained the protein in buffer 0.01 M HEPES, 0.2 M NaCl, pH 7.4.

Data Collection—All data were collected at cryogenic temperatures using synchrotron radiation at 100 K. The CD5DIII crystals were mounted and cryoprotected using 30% glycerol. Diffraction data were collected using synchrotron radiation at the X06SA beamline at the Swiss light source (Villigen). Diffraction data were recorded on a Mar225 CCD detector. A sulfur SAD data collection was performed using an inverse beam strategy to maximize the anomalous differences collecting intact Bijvoet pairs. A 1.80-Å wavelength was used to optimize the S anomalous signal. The diffraction intensities were processed with HKL2000 (24).

Structure Determination and Refinement—Reduced intensities were used to search for the S substructure. A resolution cut-off of 4.5 Å was applied during the substructure solution. Thus, the four disulfides were treated as super-sulfur atoms. Initially five super-sulfur positions were found with SnB (25) and SHELXD (26). These positions were fed into SHARP (27). After solvent flattening with SOLOMON (28), the initial 2.8 Å map, showing all the disulfides, was used for automatic model building using MAID (29). The model of CD5DIII was rebuilt and refined to 2.5 Å (PDB code 2OTT). A high resolution data set of a trigonal crystal form at 2.2 Å was used for refinement. The CD5DIII structure solved in the tetragonal space group was used as model in MOLREP (30) to solve the trigonal crystal structure by molecular replacement. ARP/wARP (31) and REF-MAC5 (32) were applied for automatic model building and refinement (Table 1) (PDB code 2JA4).

CD6-ALCAM/CD166 Modeling—The structural model of the CD6-ALCAM/CD166 interaction was built using the docking program HADDOCK (33). For the application of the ambiguous distance restraints, the active residues were those that mutagenesis analysis had indicated important for complex formation (residues Glu-293, Tyr-327, Ser-329, Phe-344, Asn-346, Asn-348, Gln-352, and Ser-353 of CD6 and residues Phe-26, Lys-28, Phe-40, Phe-43, Lys-48, Asp-54, Met-87, Thr-90, and Glu-91 of ALCAM/CD166). The passive residues were selected as those in the vicinity of the active residues with a surface accessible area greater than 40% as determined by the program

Crystal Structure of Human CD5 Domain III

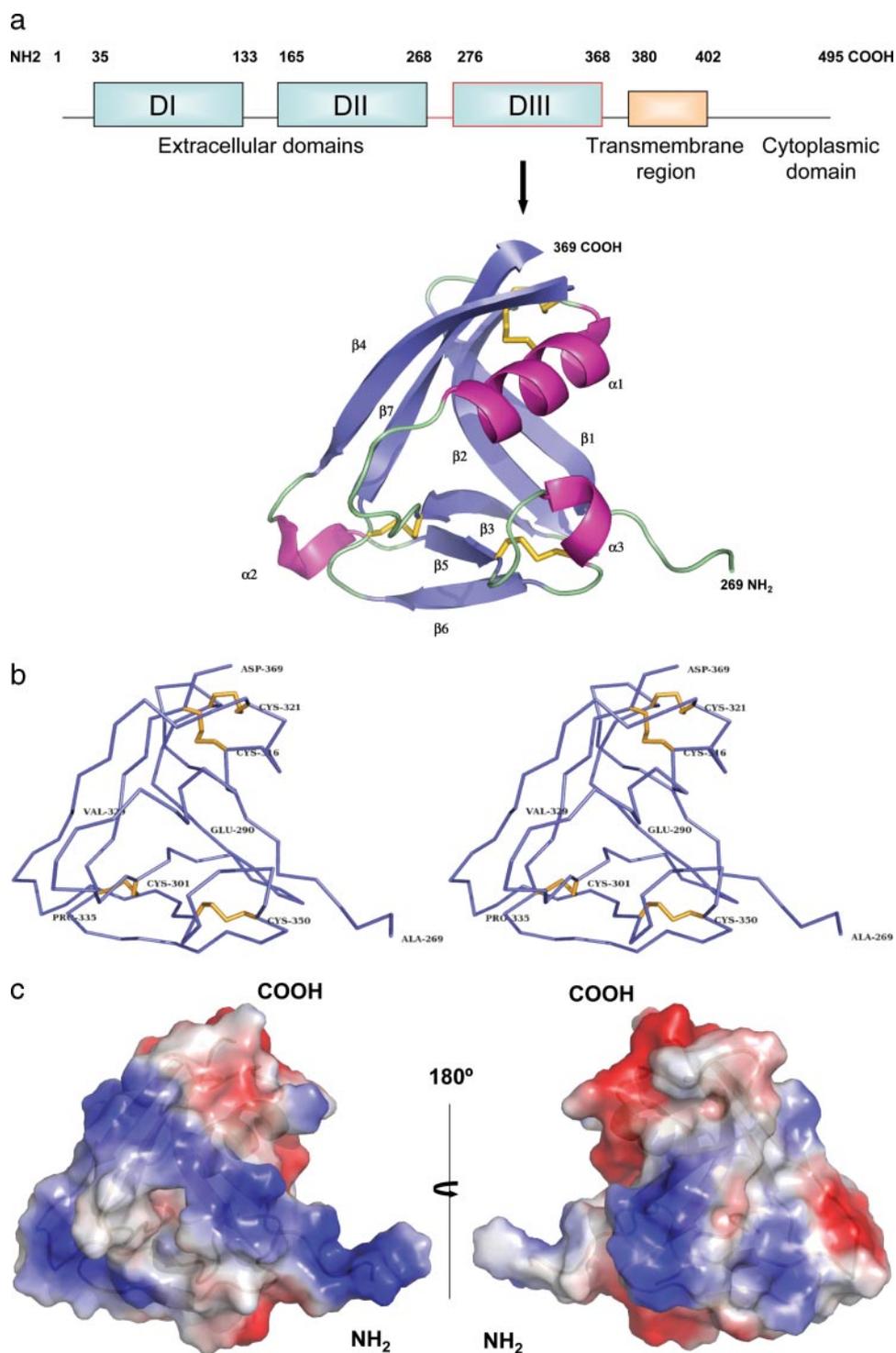


FIGURE 1. Human CD5DIII structure. *a*, upper panel, domain organization of the human CD5 receptor; the red border delimits the structure solved. Lower panel, $C\alpha$ ribbon representation of the CD5DIII structure; helical segments are shown in magenta, β -strands are in blue, and loop regions are in green. Helices are labeled from $\alpha 1$ to $\alpha 3$, and β -strands are from $\beta 1$ to $\beta 7$. The disulfide bridges are colored in orange. Residues Ala-269 at the N-terminus and Asp-369 at the C-terminus have been numbered for clarity. *b*, stereo diagram of the CD5DIII structure. The Ala-269 and Asp-369 represent the N- and C-terminal ends; one cysteine of each disulfide has been labeled to localize the disulfide bridges. *c*, electrostatic potential map of CD5 domain III showing the positive (blue) and negative (red) charged areas on CD5DIII surface.

WHATIF (34). Starting from the ALCAM/CD166 model (PDB code 1KCJ) and the modeled structure of CD6DIII, 100 randomly generated structures were subjected to rigid body mini-

mization. The 20 structures with the lowest energy were energy-minimized by allowing for the movement of the side chains of the segments 290–295, 323–332, and 339–352 of CD6 and 23–30, 38–56, and 85–98 of ALCAM/CD166. These structures showed a large degree of convergence (backbone root mean square deviation at the interface smaller than 1.5 Å). The lowest energy structure was further refined in explicit water. The model contained 98.9% of the residues in the allowed regions in the Ramachandran plot, indicating that proper geometry was satisfied during docking and the subsequent minimization.

RESULTS AND DISCUSSION

Overall Structure Description—The structure was determined using the sulfur SAD method at 2.8 Å of resolution, refined to 2.5 Å in the same tetragonal space group and subsequently solved by molecular replacement at 2.2 Å in a trigonal space group (Table 1, Fig. 1). The first and last residues observed in the electron density are Ala-269 and Asp-369, a slightly longer polypeptide chain than the defined canonical group B SRCR. The extra seven amino acids (Ala-269 to Asn-275) at the N terminus comprise a well ordered loop that connects domains II and III (Fig. 1*a*). The CD5DIII domain adopts a dense fold inside a heart shape of approximate dimensions 28 × 38 × 35 Å. The structure core is formed by the association of helix $\alpha 1$ with a curved four-stranded antiparallel β -sheet, which includes $\beta 1$, $\beta 2$, $\beta 7$, and $\beta 4$ (Fig. 1, *a* and *b*). The three additional β -strands ($\beta 3$, $\beta 5$, and $\beta 6$) form another antiparallel β -sheet at the bottom of the domain core. The helix $\alpha 2$ connects the end of $\beta 4$ to $\beta 5$, whereas $\alpha 3$ joins $\beta 5$ and $\beta 6$ (Fig. 1, *a* and *b*).

The central hydrophobic core is formed by the helix $\alpha 1$ in association with the concave face of the larger β sheet and the upper part of the smaller one. This central hub of the CD5DIII forms a triangular cavity composed by strands $\beta 2$, $\beta 7$, and $\beta 3$ occupied by a buried water molecule, which hydrogen-

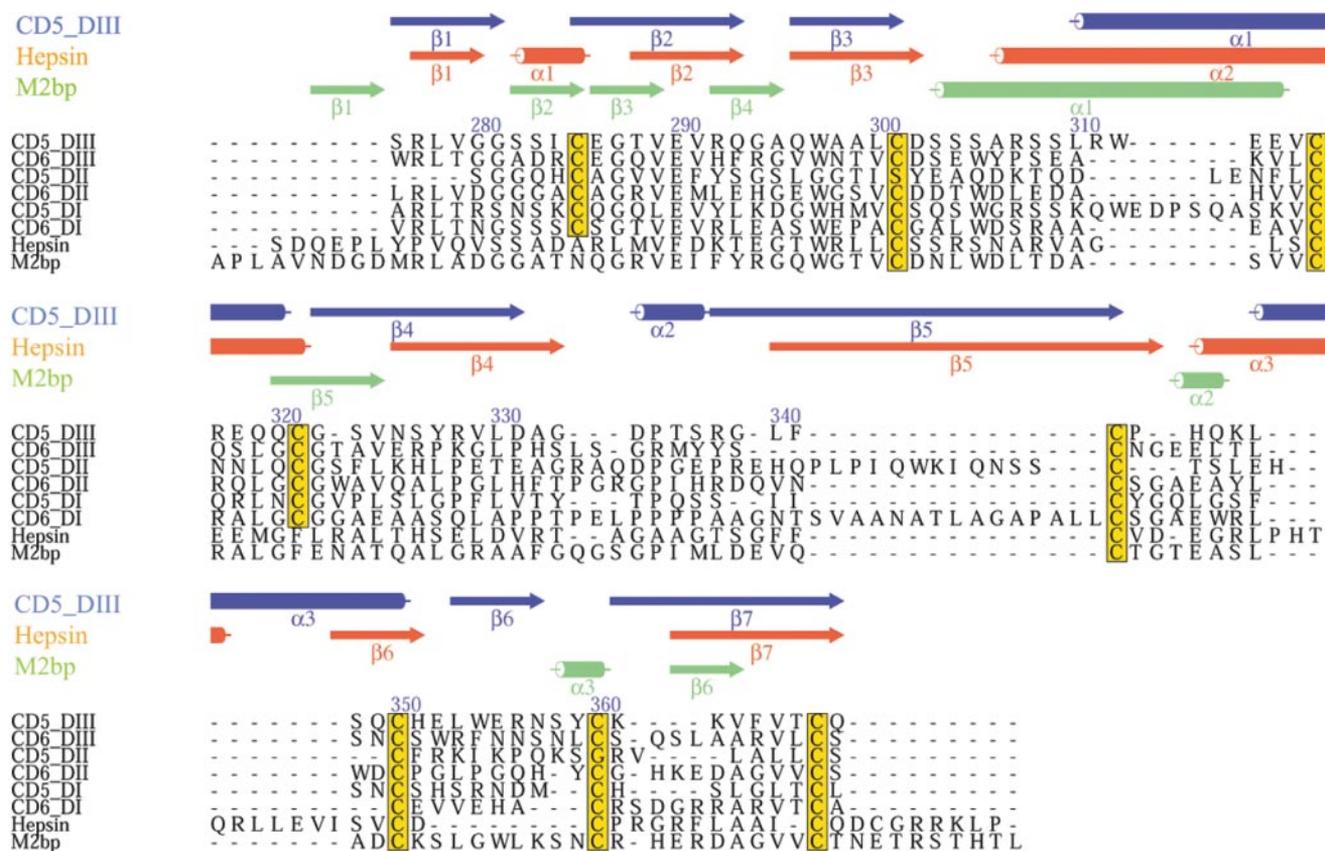


FIGURE 2. Structure-based sequence alignment of the SRCR group B domains of human CD5 and CD6 receptors compared with the group A members M2bp and hepsin. The secondary structures of CD5DIII, M2bp, and Hepsin are colored in blue green and orange, respectively. Secondary structure was defined using DSSP criteria (41), and the cysteine positions are highlighted in yellow.

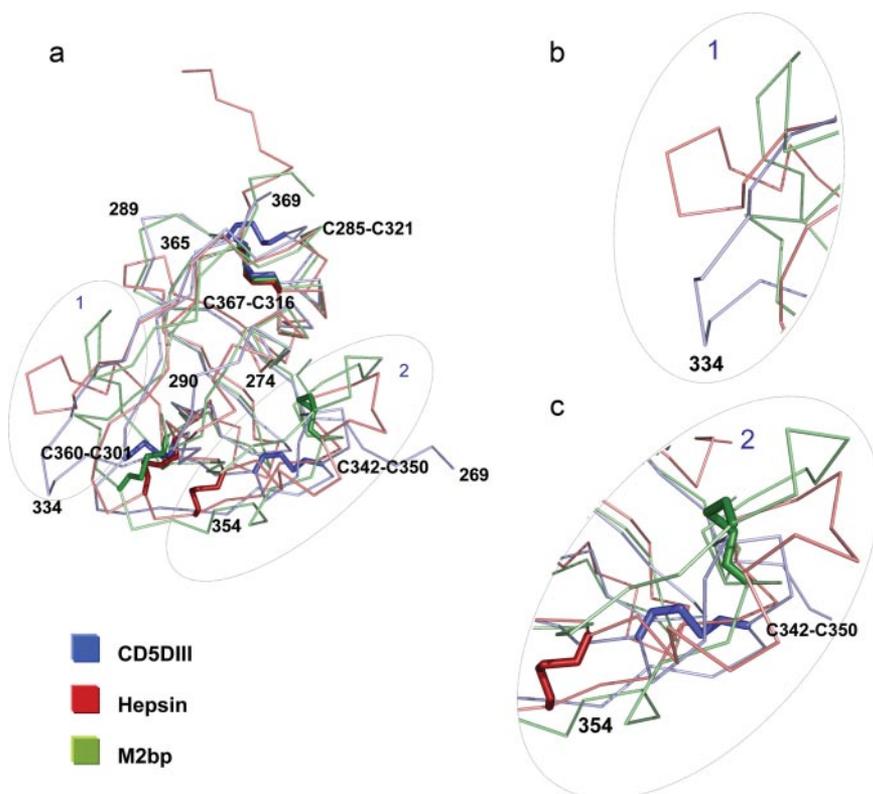
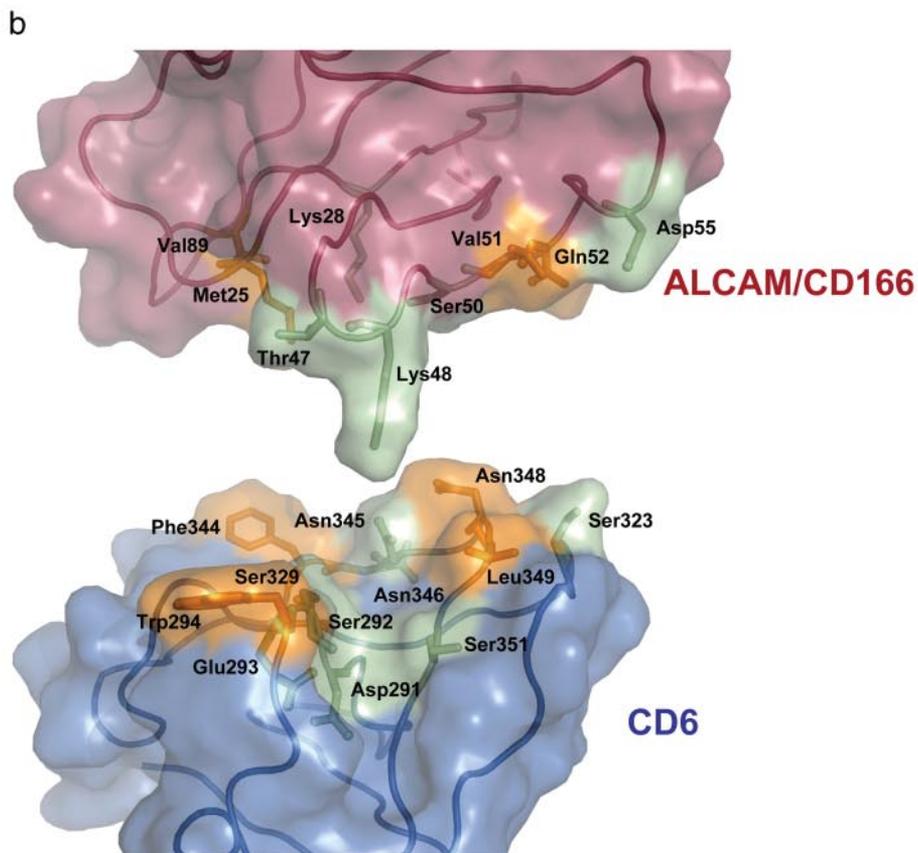
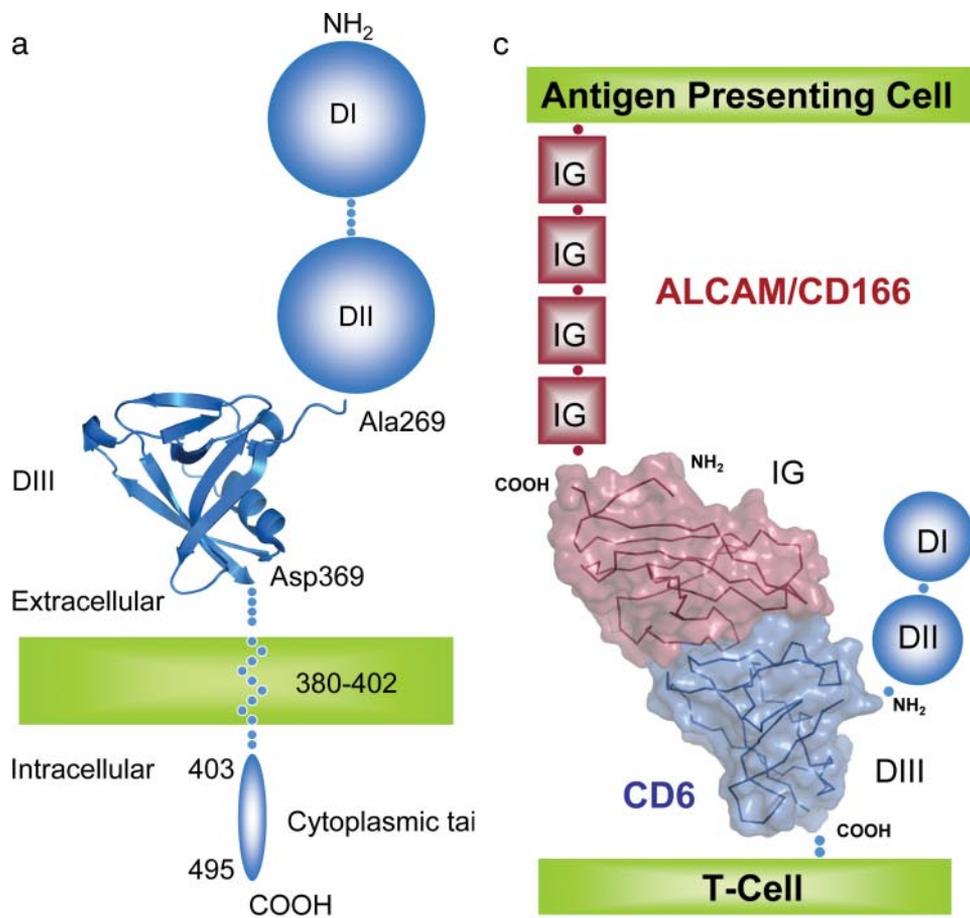


FIGURE 3. Comparison of SRCR domain group A and group B structures. *a*, C α superposition of the CD5DIII structure with SRCR group A domains of Hepsin and the M2bp. *Panel b* and *c* display detailed views of the two regions with major loop differences.

bonds to all three strands (Val-289, Leu-300, and Val-363 main chains). There are eight cysteines in CD5DIII SRCR engaged in disulfide bridges; Cys-285 and Cys-321 form a disulfide bridge that links the N terminus of $\beta 2$ with the C terminus of $\alpha 1$, Cys-316—Cys-367 disulfide connects the helix $\alpha 1$ to the strand $\beta 7$ located in the middle of the antiparallel β -sheet, the Cys-342—Cys-350 disulfide encloses the helix $\alpha 3$, and finally, Cys-301—Cys-360 bridge joins the end of the $\beta 3$ strand with the N-terminal side of the final strand $\beta 7$ (Fig. 1, *a* and *b*).

The solvent-exposed regions of CD5DIII show a special arrangement on the protein surface. Some residues (Lys-273, Arg-307, Arg-328, Lys-346, Lys-361, Arg-356) are organized, forming two basic parallel stripes running on the surface from the N terminus over $\alpha 1$ and the back side of the four-antiparallel-strand β -sheet and ending in the cavity formed by the three antiparallel β -sheet and $\alpha 2$ (Fig. 1c). The acidic residues are grouped

Crystal Structure of Human CD5 Domain III



in two opposite sites of the molecule. The first acidic area is formed by the loop that joins $\beta 1$ and $\beta 2$ together with the end of $\beta 7$, the strand that connects the extracellular domain to the transmembrane helix. It also extends through a cavity formed by the N terminus of $\beta 1$ and one side of the helix $\alpha 1$ (Glu-314 and Glu-318). The second area is located on the back side of $\alpha 2$ together with the loop residues that connect the helix to $\beta 4$ and $\beta 5$ (Asp-331 and Asp-334).

Comparison of the SRCR Family Group A and Group B Domains—The main differences between the members of the SRCR group A and B families reside both in the number of cysteines engaged in disulfide bridges (Fig. 2) and their domain composition (2). Nowadays there is scant structural information regarding the two groups. The structures of M2bp (35) and hepsin SRCR domain (36) are the only representatives of group A. Our structure provides the first structural information of group B and contributes to a better understanding of this ancient superfamily. The DALI server (37) was used to search the data base for structures with a fold similar to the CD5DIII structure. This search only retrieved the structures of M2bp and Hepsin. A close comparison of the three structures reveals that although the root mean square deviation differences are significant (CD5DIII-M2bp: 2.4 Å, for 87 matching C α -atoms; CD5DIII-Hepsin: 2.6 Å, for 79 matching C α -atoms), they have a similar fold with most of the secondary elements conserved. The main differences are observed at the connecting loops (Fig. 3). The superposition of the structures shows that the first four β -strands and α -helix follow a parallel track. From that point onward, the structures start to diverge, converging again along the last β -strand. Remarkably, the position of the water molecule that joins strands $\beta 2$, $\beta 7$, and $\beta 3$ by interacting with the main chain atoms from one residue of each strand is preserved in the three structures.

Some of the loops are shortened or simply not present in CD5 (Figs. 2 and 3); this combination gives a somewhat more compact structure since they are located at the surface of the three molecules. Fig. 3, *b* and *c*, show the two main differences, which are probably produced by the differences in the disulfides positions. The first is located at the back of the molecules where the loops form a tight turn with different degrees of bending depending on the molecule (Fig. 3*b*). M2bp is the more curved, Hepsin occupies an intermediate position, and CD5 is extended. All start at the end of the fourth β -strand, and whereas CD5 and Hepsin finish leading the fifth β -strand, Mb2p does it into a disordered loop. The second loop differences are more striking (Fig. 3*c*). The three structures present a long loop protruding from one face of the β -sheet that ends facing the opposite α -helix. This arrangement forms a depression of ~ 18 Å length with a small gap of 5 Å between residues forming the tip of their loops and the α -helix. The partial deletion of the tip of the CD5DIII loop with respect to the others is compensated by the presence of larger residues so that the opening distance is conserved.

Only one of the four disulfides stabilizing the structure of CD5DIII (Cys-367—Cys-316) conserves its position in Hepsin and M2bp. Although the Cys-360—Cys-301 and Cys-342—Cys-350 disulfides are present in M2bp and Hepsin, their positions are not equivalent between the group A and B SRCR domains. However, the fourth disulfide bridge, exclusive of the group B SRCR (Cys-285—Cys-321) constrains the position of the C terminus, which joins the extracellular domain with the membrane.

To date the single common characteristic among the different SRCR domain-containing proteins, independently of their classification as group A or B, is that they are all extracellular. This group of molecules seems to be as diverse as the Ig superfamily of proteins without a unifying function. This seems to be the situation in the SRCR superfamily where independently of the protein function both the group A and B domains have evolved toward a similar scaffold with small variations depending on the function.

Orientation of CD5 Extracellular Region with Respect to the Plasma Membrane—The crystal structure suggests how the extracellular part of the CD5 receptor could be oriented with respect to the plasma membrane. The C-terminal residue Asp-369 is only 8 residues away from the predicted transmembrane region of CD5. This short stretch imposes strong restraints on the spatial relation between the extracellular portion of CD5 and its transmembrane helix. Therefore, the strand $\beta 7$ of CD5DIII, which contains its C terminus, should be oriented nearly perpendicular to the membrane plane (Fig. 4*a*). The proposed membrane-spanning helix consists of 22 amino acids (Fig. 1*a*, upper panel); consequently, the length indicates that CD5 contains a single transmembrane helix that anchors the receptor. Our structure also includes residues that belong to the linker and the N terminus of CD5DII, indicating that the rest of the extracellular part of CD5 is located on one side of DIII, leaving the upper part exposed. Thus, considering the dimensions of CD5DIII, the extracellular region of the receptor should protrude from the plasma membrane around 90–100 Å. This organization is similar to the arrangement of the asymmetric unit in the tetragonal crystal form.

The CD6-ALCAM/CD166 Interaction—Previous studies of the CD6-ALCAM/CD166 interaction were based on the structural similarities of CD6DIII with M2bp, a SRCR group A member with unrelated biological function (19). Using our CD5DIII structure, a typical SRCR domain group B member, we have produced a structural model for CD6DIII (38) based on this structural and functional homologue (39) (Fig. 2). Experimental restraints based in antibody mapping and site-directed mutagenesis data in CD6 and ALCAM/CD166 (17, 18) were used to dock (33) the two proteins and model their interaction. The contact surface is highly complementary, and no bad contacts were detected (Fig. 4*b* and supplemental Figs. 1 and 2). The ALCAM/CD166 residues that were identified by site-directed mutagenesis were mainly found in the interface with

FIGURE 4. Model showing the position of CD5 extracellular region with respect to the plasma membrane and CD6-ALCAM/CD166 interaction model. *a*, arrangement of CD5 extracellular region with respect to the plasma membrane (see "Results and Discussion" for details). *b*, detailed view of the CD6DIII-ALCAM/CD166 interaction. CD6 and ALCAM surfaces are colored in blue and dark red, respectively. The residues involved in polar contacts are colored in green, and hydrophobic interactions are in orange (for a detailed scheme, see supplemental Fig. 1). The model shows a high degree of surface complementarity (see supplemental Fig. 2). *c*, molecular architecture of the CD6-ALCAM/CD166 model in a T-cell antigen-presenting cell interaction.

Crystal Structure of Human CD5 Domain III

CD6 (Phe-26, Phe-40, Lys-28, Phe-43, Lys-48, Asp-54) (Fig. 4b), whereas some of the CD6 residues were found out of the interaction area. This could be due to the fact that antibody epitopes overlapped with the CD6 surface involved in ALCAM/CD166 interaction imposing a large uncertainty in the identification of the residues. Some of them, such as Tyr-327 and Gln-352, seem to be important for protein folding, but they are located away from the interaction surface. Noteworthy are the contacts of ALCAM/CD166 Lys-28 and Asp-53 with the CD6 side chains of Glu-293, Asp-291, Ser-351, and the contacts of ALCAM/CD166 Lys-48 and Ser-50 with CD6 side and main chains of Asn-346 and Asn-345 (Fig. 4b). The rest of the ALCAM/CD166 and CD6 residues build the interaction surface by hydrophobic contacts except Ser-290 and Thr-47, which are located in the middle of the interface. The buried surface area is 695 \AA^2 , which is in the range of other typical protein-protein associations.

The orientation of CD6DIII with respect to the membrane (Fig. 4c) must be similar to its homologue CD5DIII. Therefore, the location of the extracellular domains, DII and DI, leaves the interaction area exposed to the extracellular media, allowing the interaction with the N-terminal IgV-like domain from ALCAM/CD166. The localization of CD6 and ALCAM/CD166 at the central supramolecular activation clusters formed at the immunological synapse indicates that the distance spanned by the interaction between CD6 and ALCAM/CD166 must be similar to that spanned by the T-cell antigen receptor-major histocompatibility complex (150 \AA) interaction (40). Otherwise, it would be physically impossible for the CD6-ALCAM/CD166 pair to redistribute and to co-localize with T-cell antigen receptor/CD3 (39). Our model of CD6-ALCAM/CD166 interaction, in which the ALCAM V₁ domain embraces the CD6DIII domain, is compatible with that assumption (39). The length of the CD6 receptor should be similar to CD5 (see previous section); thus, the interaction of CD6DIII with ALCAM/CD166 must occur in an intercellular distance around 150 \AA , and the position of the N and C termini of both molecules in the complex would link opposing cell surfaces.

CD5 and CD6 are closely related molecules involved in the modulation of antigen-receptor-induced T cell activation and differentiation. Based on the first SRCR group B domain structure, we modeled the CD6DIII-ALCAM/CD166 association addressing the molecular basis of the T-B or T-dendritic cell interactions. The high homology between CD5 and CD6 SRCR domains (Fig. 2) suggests that a similar mode of interaction could be the basis of the CD5DII-IgV_H association, which has been proposed to affect the maintenance and selective expansion of normal and malignant human B cells (11). The available structures of the SRCR domains in conjunction with the reported biological interactions indicate that this ancient scaffold can be adapted to associate with other extracellular domains.

Acknowledgments—We thank the European Synchrotron Radiation Facility (ID14-4) and Swiss light source (PX) biocrystallography beamline personnel for help during data collection and Takako Sasaki and Belén Suárez for advice and assistance with rsCD5 expression and purification. We also thank Ana Rojas and Tahl Zimmermann for help and advice with Swiss Modeler and HADDOCK programs.

REFERENCES

1. Resnick, D., Pearson, A., and Krieger, M. (1994) *Trends Biochem. Sci.* **19**, 5–8
2. Sarrias, M. R., Gronlund, J., Padilla, O., Madsen, J., Holmskov, U., and Lozano, F. (2004) *Crit. Rev. Immunol.* **24**, 1–37
3. Hardy, R. R., and Hayakawa, K. (1994) *Adv. Immunol.* **55**, 297–339
4. Hardy, R. R., Carmack, C. E., Li, Y. S., and Hayakawa, K. (1994) *Immunol. Rev.* **137**, 91–118
5. Van de Velde, H., and Thielemans, K. (1996) *Cell Immunol.* **172**, 84–91
6. Bikah, G., Lynd, F. M., Aruffo, A. A., Ledbetter, J. A., and Bondada, S. (1998) *Int. Immunol.* **10**, 1185–1196
7. Van de Velde, H., von Hoegen, I., Luo, W., Parnes, J. R., and Thielemans, K. (1991) *Nature* **351**, 662–665
8. Luo, W., Van de Velde, H., von Hoegen, I., Parnes, J. R., and Thielemans, K. (1992) *J. Immunol.* **148**, 1630–1634
9. Biancone, L., Bowen, M. A., Lim, A., Aruffo, A., Andres, G., and Stamenkovic, I. (1996) *J. Exp. Med.* **184**, 811–819
10. Pospisil, R., Fitts, M. G., and Mage, R. G. (1996) *J. Exp. Med.* **184**, 1279–1284
11. Pospisil, R., Silverman, G. J., Marti, G. E., Aruffo, A., Bowen, M. A., and Mage, R. G. (2000) *Leuk. Lymphoma* **36**, 353–365
12. Calvo, J., Places, L., Padilla, O., Vila, J. M., Vives, J., Bowen, M. A., and Lozano, F. (1999) *Eur. J. Immunol.* **29**, 2119–2129
13. Aruffo, A., Bowen, M. A., Patel, D. D., Haynes, B. F., Starling, G. C., Gebe, J. A., and Bajorath, J. (1997) *Immunol. Today* **18**, 498–504
14. Williams, A. F., and Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6**, 381–405
15. Bowen, M. A., Bajorath, J., Siadak, A. W., Modrell, B., Malacko, A. R., Marquardt, H., Nadler, S. G., and Aruffo, A. (1996) *J. Biol. Chem.* **271**, 17390–17396
16. van der Merwe, P. A., and Barclay, A. N. (1996) *Curr. Opin. Immunol.* **8**, 257–261
17. Bodian, D. L., Skonier, J. E., Bowen, M. A., Neubauer, M., Siadak, A. W., Aruffo, A., and Bajorath, J. (1997) *Biochemistry* **36**, 2637–2641
18. Skonier, J. E., Bodian, D. L., Emswiler, J., Bowen, M. A., Aruffo, A., and Bajorath, J. (1997) *Protein Eng.* **10**, 943–947
19. Bowen, M. A., Aruffo, A. A., and Bajorath, J. (2000) *Proteins* **40**, 420–428
20. Kohfeldt, E., Maurer, P., Vannahme, C., and Timpl, R. (1997) *FEBS Lett.* **414**, 557–561
21. Calvo, J., Places, L., Espinosa, G., Padilla, O., Vila, J. M., Villamor, N., Ingelmo, M., Gallart, T., Vives, J., Font, J., and Lozano, F. (1999) *Tissue Antigens* **54**, 128–137
22. Rodamilans, B., Ibañez, S., Bragado-Nilsson, E., Sarrias, M. R., Lozano, F., Blanco, F. J., and Montoya, G. (January 23, 2007) *J. Struct. Biol.* **10.1016/j.jsb.2007.01.003**
23. McAlister, M. S., Davis, B., Pfuhl, M., and Driscoll, P. C. (1998) *Protein Eng.* **11**, 847–853
24. Otwinowski, Z., and Minor, W. (1997) *Methods in Enzymology*, pp. 307–323, Academic Press, Inc., New York
25. Weeks, C. M., and Miller, R. (1999) *J. Appl. Crystallogr.* **32**, 120–124
26. Schneider, T. R., and Sheldrick, G. M. (2002) *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1772–1779
27. de la Fortelle, E., and Bricogne, G. (1997) *Macromolecular Crystallography*, pp. 472–494, Academic Press, Inc., New York
28. Abrahams, J. P., and Leslie, A. G. (1996) *Acta Crystallogr. D Biol. Crystallogr.* **52**, 30–42
29. Levitt, D. G. (2001) *Acta Crystallogr. D Biol. Crystallogr.* **57**, 1013–1019
30. Vagin, A., and Teplyakov, A. (2000) *Acta Crystallogr. D Biol. Crystallogr.* **56**, 1622–1624
31. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2003) *Methods Enzymol.* **374**, 229–244
32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255
33. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *J. Am. Chem. Soc.* **125**, 1731–1737
34. Vriend, G. (1990) *J. Mol. Graph.* **8**, 52–56
35. Hohenester, E., Sasaki, T., and Timpl, R. (1999) *Nat. Struct. Biol.* **6**,

- 228–232
36. Somoza, J. R., Ho, J. D., Luong, C., Ghate, M., Sprengeler, P. A., Mortara, K., Shrader, W. D., Sperandio, D., Chan, H., McGrath, M. E., and Katz, B. A. (2003) *Structure* **11**, 1123–1131
37. Holm, L., and Sander, C. (1994) *Nucleic Acids Res.* **22**, 3600–3609
38. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) *Nucleic Acids Res.* **31**, 3381–3385
39. Gimferrer, I., Farnos, M., Calvo, M., Mittelbrunn, M., Enrich, C., Sanchez-Madrid, F., Vives, J., and Lozano, F. (2003) *J. Biol. Chem.* **278**, 8564–8571
40. Dustin, M. L. (2002) *Arthritis Res.* **4**, 119–125
41. Kabsch, W., and Sander, C. (1983) *Biopolymers* **22**, 2577–2637





Crystallization Notes

Expression, purification and crystallization of human CD5 domain III, a nano-scale crystallization example

Bernardo Rodamilans ^a, Sonia Ibañez ^a, Elisabeth Bragado-Nilsson ^a, Maria Rosa Sarrias ^c, Francisco Lozano ^{c,d}, Francisco J. Blanco ^b, Guillermo Montoya ^{a,*}

^a Structural Biology and Biocomputing Programme, Macromolecular Crystallography Group, Spanish National Cancer Center (CNIO) c/Melchor Fdez. Almagro 3, 28029 Madrid, Spain

^b Structural Biology and Biocomputing Programme, Nuclear Magnetic Resonance Group, Spanish National Cancer Center (CNIO) c/Melchor Fdez. Almagro 3, 28029 Madrid, Spain

^c Department of Immunology, Hospital Clinic of Barcelona, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Villarroel 170, 08036 Barcelona, Spain

^d Department of Cell Biology and Pathology, Faculty of Medicine, University of Barcelona, Casanova 143, 08036 Barcelona, Spain

Received 26 September 2006; received in revised form 5 January 2007; accepted 16 January 2007

Abstract

The human lymphocyte receptor CD5, a key regulator of immune responses, is involved in the modulation of antigen specific receptor-mediated T cell activation and differentiation signals. CD5 is a membrane glycoprotein which belongs to the group B scavenger receptor cysteine-rich (SRCR) superfamily for which no structural information is available. The most conserved membrane-proximal SRCR domain of CD5 (domain III) has been expressed in HEK-EBNA-293 cells. Although the yield of the purified protein was at the level of micrograms, well diffracting crystals have been obtained. The crystals belong to a tetragonal space group P₄₁22 or P₄₃22. They contain two molecules per asymmetric unit and diffracted to 2.5 Å resolution using synchrotron radiation. The strategy shown here to produce, isolate and crystallize CD5 domain III can be used for other mammalian proteins difficult to produce for structural or other biophysical studies.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Crystallization; Post-translational modifications; X-ray; Receptors; Protein purification; Protein expression

CD5 is a lymphoid-specific accessory molecule providing either positive or negative signals necessary for proper T and B cell activation and differentiation. It is expressed on thymocytes from early stages of development, mature peripheral T lymphocytes, and a subpopulation of peripheral B lymphocytes (B-1a cells). B-1a cells have been related to the production of auto-antibodies, and this subpopulation is expanded in certain autoimmune diseases and in B-cell chronic lymphocytic leukemias (B-CLL) (Hardy et al., 1994; Hardy and Hayakawa, 1994). Several reports have proposed that CD5 interacts with ligands on cell sur-

faces during T-cell–B-cell interactions. In the light of these observations, a role for CD5 in cognate interactions between T and B cells (Van de Velde and Thielemans, 1996; Bikah et al., 1998) was deduced from the reported interactions between CD5 and distinct B-cell surface proteins such as CD72 (Van de Velde et al., 1991; Luo et al., 1992), gp40-80 (Bikah et al., 1998; Biancone et al., 1996), Ig Vh framework region sequences (Pospisil et al., 1996; Pospisil et al., 2000) and gp150 (Calvo et al., 1999b). Definitive characterization of these and other putative ligands for CD5 is still a matter of debate.

The CD5 receptor is a plasma membrane glycoprotein, which is conserved in all mammalian and avian species and belongs to the SRCR superfamily, of either soluble or membrane-bound receptors expressed by hematopoietic

* Corresponding author. Fax: +34 912246976.

E-mail address: gmontoya@cnio.es (G. Montoya).

and non-hematopoietic cells (Resnick et al., 1994; Sarrias et al., 2004). The SRCR superfamily can be divided in two groups. Members of group A contain SRCR domains with six cysteine residues and encoded by two exons, whereas those of group B contain eight cysteines and are encoded by a single exon. In both cases the cysteines form disulphide bridges. Group A members are usually present as multidomain proteins containing a single SRCR domain associated to other functional domains. On the contrary, group B members are quite often exclusively composed of repeats of the SRCR domain (Resnick et al., 1994; Sarrias et al., 2004). The extracellular region of CD5 contains three SRCR domains, whereas the cytoplasmic tail does not present any enzymatic activity and seems to be involved in several signal transduction pathways (Raab et al., 1994; Raman et al., 1998; Simarro et al., 1999; Bauch et al., 1998).

The crystallization of eukaryotic proteins is still not common. One of the main problems is their low production yield in heterologous expression systems with consequent difficulties in structure elucidation. This situation is more dramatic in the case of membrane and/or post-translationally modified proteins such as CD5. Here we describe the first crystals of a SRCR group B member and its expression, purification and crystallization protocol. To overcome the difficulties to produce this protein for structural studies we have expressed it in mammalian cells. The development of a monoclonal antibody (Mab) was essential to obtain the pure protein in a single step. The use of nanocrystallization facilities eased the screening of different crystallization conditions which gave the first hit that could be refined with other purification batches to yield good quality diffracting crystals.

1. CD5 domain III expression

The recombinant form of the human CD5-DIII was expressed using an episomal expression system in human embryonic kidney cells (HEK 293-EBNA). These cells constitutively express the Epstein Barr viral protein EBNA-1, allowing episomal replication of the pCEP-Pu vector, a kind gift from Drs. T. Sasaki and R. Timpl (Max Planck Institute for Biochemistry, Martinsried, Germany) (Kohfeldt et al., 1997). The cDNA sequence coding the human CD5-DIII was obtained by PCR amplification from the pH β . T1T2 plasmid (Calvo et al., 1999a) with the 5'-CCCCGCTAGCTTTCCAGCCCAAGGTGCAG-3' (forward), and 5'-GTGGATCCTAATCCTGGCATGTGACAAAC-3' (reverse) primers. The forward primer incorporated the *Nhe*I restriction site, while the reverse primer incorporated a stop codon followed by the *Bam*HI restriction site (underlined). The amplified PCR products were cloned into a digested pCEP-Pu vector, in frame with the BM-40 leader sequence. This resulted in the addition of amino acids Ala-Pro-Leu-Ala to its N-terminal sequence. The resulting constructs were expressed into HEK 293-EBNA cells as described before (Kohfeldt et al., 1997). Briefly, HEK-EBNA-293 cells in culture dishes were trans-

fecting using lipofectamine and then grown in monolayers using D-MEM:F-12(1:1)+Glutamax (Gibco) supplemented with fetal bovine serum (FBS) (10%), (Sigma), Penicillin–Streptomycin 100 U/ml and 100 μ g/ml (Gibco), Geneticin 250 μ g/ml (Gibco) and Puromycin 2.5 μ g/ml (Sigma). For large-scale protein production, cells in exponential growth phase were trypsinized and plated finally in Costar culture flasks (162 cm²). The media was exchanged every 48 h and when 80% of the surface was covered, new media without FBS was added. The expressed protein was collected in the culture media 3–4 days after this change. Every 2 days the media was collected and pooled in 50 ml falcon tubes, spun at 1000g in a centrifuge for 5 min to discard cell debris. The clean supernatant was transferred into a new falcon tube, flash-frozen in liquid nitrogen and stored at -80°C . Around 2.5–3 L of supernatant containing around 200–300 μ g/L of protein were collected in 3–4 weeks. At that time cells started to detach from the surface and the culture was stopped.

2. Protein purification

CD5-DIII was purified at 4°C by affinity chromatography using a mouse monoclonal antibody (83-C4) specific to CD5-DIII (Calvo et al., 1999a,b). Around 4 L of supernatant from the hybridoma producing 83-C4 Mab were precipitated adding 50% (NH₄)₂SO₄ with gentle stirring for 30 min on ice. Subsequently, the mixture was centrifuged at 6000g during 30 min. The pellet containing the antibody was resuspended in 50 ml of PBS (0.137 M NaCl, 0.01 M Na₂HPO₄–2H₂O, 2.7 mM KCl, 2 mM KH₂PO₄) and then extensively dialyzed against PBS to eliminate the residual (NH₄)₂SO₄. After dialysis the sample was filtered with a 0.22 μ m filter (Millipore). 83-C4 Mab was purified using a 5 ml HiTrap Protein G column (Amersham Biosciences) attached to an Äkta prime. The column was equilibrated with PBS, and the sample was loaded onto the column. After extensive washing with PBS the 83-C4 Mab was eluted using 0.1 M glycine pH 2.6. The tubes in the fraction collector contained 200 μ l of 1 M Tris, pH 9.0, to avoid denaturation. Fractions containing the antibody were pooled down and dialyzed against 0.1 M NaHCO₃ pH 8.3 plus 0.5 M NaCl. The usual yield was around 100 mg of 83-C4 Mab. Once the antibody was isolated, it was loaded onto a CNBr-Sepharose 4B resin (Amersham Biosciences) previously activated with several washes of 1 mM HCl, washed once with the coupling solution (0.1 M NaHCO₃, pH 8.3, 0.5 M NaCl) and then the resin was mixed with the ligand and incubated overnight at 4°C using a rotating wheel. Finally the excess of 83-C4 Mab was washed out with coupling buffer and the remaining active groups were blocked by incubation with 0.1 M Tris–HCl, pH 8.0, for 2 h. The resin was used to pack a column which was attached to an Äkta prime for the purification process.

The supernatant from 2.5 to 3 L of HEK-EBNA-293-CD5-DIII culture was used for purification. This volume was concentrated at 4°C in an Amicon cell to 100–200 ml

and then loaded onto a 10 ml 83-C4 Mab column previously equilibrated with PBS. The sample was recirculated 10 times to improve binding. Subsequently, the column was washed with 10 volumes of PBS plus 0.5 M NaCl and 1% NP40 and then with 10 volumes of PBS. Finally, the protein was eluted with 3.5 M MgCl₂ in PBS. All the chromatographic steps were performed at 1.5 ml/min flow rate. The amount of purified protein depended on the collected supernatant and ranged from 200 to 500 µg from 2.5 to 3 L of culture. Finally the elution buffer was exchanged to 0.01 M Hepes, 0.2 M NaCl, pH 7.4 using a desalting column (Amersham Biosciences). The purified protein was concentrated to 4 mg/ml at 4°C using a centricon device with a 10 kDa cut-off. The purity of the sample after the isolation process is shown in Fig. 1a. The mass spectrometry analysis of the purified CD5-DIII revealed the protein exact molecular weight. Additional experiments in the presence of a reducing agent (DTT) showed the existence of four disulphide bridges involving the eight cysteines (cys1-cys4 cys2-cys7 cys3-cys8 and cys5-cys6) in the recombinant protein (data not shown). These data together with the sharp and disperse signals observed in the one-dimensional ¹H NMR spectrum recorded in a Bruker 600 MHz spectrometer indicated that the SRCR domain folding was well defined (data not shown). The production of good quality CD5-DIII in HEK-293-EBNA cells contrast with other attempts to express CD5 domains in an heterologous system (*Pichia pastoris*) (McAlister et al., 1998), when this recombinant protein was used for NMR studies did not yield well defined spectra.

3. Crystallization

Crystallization screenings were performed with a Cartesian Mycosys SQ robot (Genomic Solutions) using the sitting drop method with nanodrops of 0.05 µl. A total of 1000 conditions from different screens from Hampton Research, Emerald Biosciences and Qiagen were tested with one batch of the purified protein. Small needles were obtained after 48 h in condition 17 of wizard I (30% PEG 8000, 0.1 M Na acetate, pH 4.5, 0.2 M LiSO₄). Three protein batches were necessary for the optimization of the crystallization conditions. After several refinement rounds from the initial crystallization conditions crystals suitable for X-ray diffraction experiments were obtained. The final crystallization conditions contained 15–25% PEG 8000, 0.1 M Na acetate, pH 4.4, 0.2 M LiSO₄ in the reservoir whereas the drop (hanging or sitting) only contained the protein in buffer 0.01 M Hepes, 0.2 M NaCl, pH 7.4. These crystals (Fig. 1b) diffracted to 3.2 Å in-house and to 2.5 Å using synchrotron radiation (Fig. 2).

4. Data collection

All data were collected at cryogenic temperatures using synchrotron radiation at 100 K. The CD5-DIII crystals were mounted and cryoprotected using the reservoir solution plus

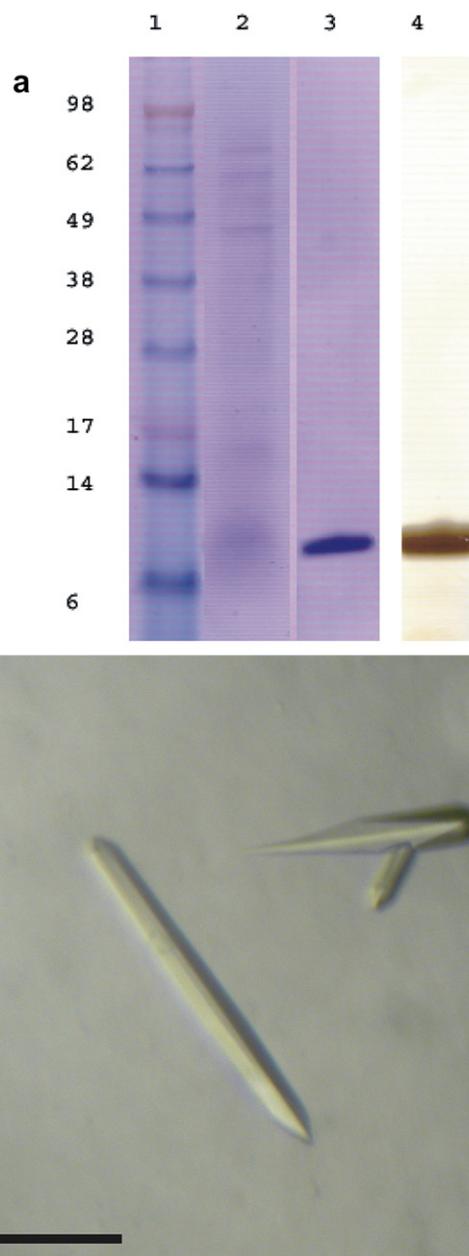


Fig. 1. (a) SDS-PAGE of the CD5-DIII purification, lane 1 molecular weight markers, lane 2 HEK-293-EBNA-CD5DIII cells culture media loaded in the Mb-DIII column, and lanes 3 and 4 purified CD5-DIII stained with Coomassie blue and silver, respectively. (b) Tetragonal crystals of CD5-DIII grown by the hanging drop methods at 16°C. These crystals behaved well in the cryo-buffer conditions described in the text and diffracted to 2.5 Å. The scale bar at the bottom of the picture represents 0.05 mm.

30% glycerol. The data sets were collected using synchrotron radiation at the ID14-4 beamline at the ESRF (Grenoble), and at the PX beamline at the SLS (Villigen). Diffraction data were recorded on an ADSC-Q315 or Mar225 CCD detectors depending on the beamline. The best data set (see Table 1) was collected using a $\Delta\lambda/\lambda = 0.036$; $= 1^\circ$ and a wavelength of 0.939 Å. Processing and scaling were accomplished with HKL2000 (Otwinowski and Minor, 1997). Statistics for the crystallographic data are summarized in Table 1.

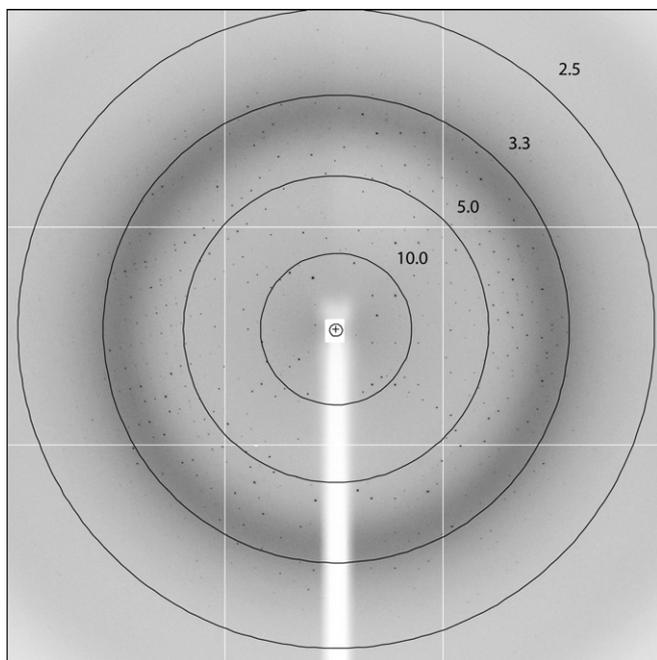


Fig. 2. Diffraction pattern of the native crystals using synchrotron radiation on ID14-4 beamline at the ESRF. Circles indicate the resolution in Å.

The recombinant human CD5-DIII (residues 269–370 from CD5) has a molecular weight of 12 kDa. Even though the expression level was low (Fig. 1a, lane 2) the isolation using the 83-C4 mAb affinity column allowed us to isolate enough pure and correctly folded protein (Fig. 1a, lanes 3 and 4) to perform initial crystallization tests with 0.05 µl drops. Although small needles were relatively easy to obtain using the nanodrop crystallization robot, it took several rounds of refinement to obtain crystals in hanging drops with 2 µl volume. Each of these refinement rounds were performed with a new protein batch. Finally, the CD5-DIII crystals diffracted to 2.5 Å resolution and several data sets were collected. The crystals belong to the tetragonal space group $P4_122$ or $P4_322$ with unit cell parameters $a = 74.42$ Å, $b = 74.42$ Å, $c = 128.08$ Å, $\alpha = \gamma = \beta = 90^\circ$. The Matthews coefficient and the self-rotation function (data not shown) suggested the presence of two protein molecules

per asymmetric unit ($VM = 3.7$ Å³ Da⁻¹) and a solvent content of 67.4%. The diffraction data was 99.6% complete, a multiplicity of 8.9 and an overall $I/\sigma(I)$ of 6.9 (see Table 1).

Probably due to the crystal size, data sets could only be collected at 3.2 Å using an in-house source at 110 K. Crystallization reproducibility was purification dependent and some batches did not yield crystals or were of very poor quality. In order to solve the structure several heavy atoms soaks were carried out, but so far the analysis of the data did not reveal a derivative. A possible strategy to solve the structure is to use the presence of the eight cysteines, and then employ the anomalous sulphur signal to solve the structure. These experiments are in progress.

The structure determination of biologically and medically important proteins is frequently hampered by low expression levels that compromise the isolation of these samples for subsequent biophysical studies. The use of small amounts of sample during crystallization trials is critical in these situations, and several groups have reported successful approaches in this important area of crystallization technology (Hosfield et al., 2003; Zheng et al., 2003; Li et al., 2005; Hazes and Price, 2005).

In this report, we present a strategy that makes use of a homologous expression system in conjunction with a highly efficient purification strategy, demonstrating that a homogeneous sample at the level of micrograms can be used to screen for initial crystallization conditions with success. This strategy can be used to produce several hundreds of micrograms of eukaryotic membrane or post-translationally modified proteins which rarely can be produced by heterologous expression systems due to the absence of their folding partners or modification systems. Although the protein production levels in mammalian expression systems are low and require long time, their use guarantee the expression of a mammalian protein in the presence of all the native folding machinery.

The use of high affinity purification systems is then necessary in order to collect and purify the expressed polypeptide in a single step. The use of Mabs against the target protein is a good option when they are available, and the protein is stable during the isolation process which requires high salt or low pH to elute it from the column. However,

Table 1
Data collection statistics

Resolution limits (Å)	$\langle I/\sigma(I) \rangle$	Completeness (%)	Multiplicity	Observed reflections	Unique reflections	R_{sym}
50.0–7.92	17.4	94.1	7.1	3235	454	0.036
7.92–5.60	12.5	99.9	8.2	6572	804	0.049
5.60–4.57	8.1	100.0	8.4	8490	1005	0.072
4.57–3.96	9.2	100.0	8.8	10,280	1167	0.064
3.96–3.54	10.4	100.0	9.0	11,761	1307	0.060
3.54–3.23	9.2	100.0	9.1	13,074	1435	0.072
3.23–2.99	6.9	100.0	9.3	14,499	1558	0.101
2.99–2.80	5.2	100.0	9.3	15,470	1658	0.141
2.80–2.64	3.1	100.0	9.1	16,000	1754	0.234
2.64–2.50	2.1	98.9	9.1	16,655	1840	0.340
Overall	6.9	99.6	8.9	116,036	12,982	0.076

$$R_{\text{sym}} = \frac{\sum_i \sum_j |I_{h,i} - \langle I_{h,j} \rangle|}{\sum_i \sum_j I_{h,i}}$$

very often, the required Mabs may not be available or the target protein is very labile. Then the target protein can be produced with a tag, which has available Mabs, (such as anti-Flag, or anti-HA) and the native protein could be eluted from the column by the use of peptides which resemble the epitope, or by the use of a specific protease (3C protease, TEV protease) after insertion of its cleavage site between the tag and the target protein such as in the TAP-tag method (Puig et al., 2001).

Not so long ago structural biology projects with this amount of sample were not feasible. In this work, we show that combining mammalian protein expression and affinity purification together with the nanodrop robotic facilities for crystallization make possible the structural determination of proteins expressed at very low levels.

Acknowledgments

We thank the ESRF (ID14-4) and SLS (PX) biocrystallography beamlines personnel for their help during data collection. B.R. thanks the Consejería de Educación de la Comunidad Autónoma Madrid and Fondo Social Europeo for a predoctoral fellowship. Financial support was obtained through Ministerio de Educación y Ciencia (BFU-2005-02403, GEN2003-20642-C09-02) grants to G.M, Instituto de Salud Carlos III (FIS CP05/100) to MRS, Ministerio de Educación (SAF 2004-3251) to FL and MEC (BIO 2003-02246) to FJB.

References

- Bauch, A., Campbell, K.S., Reth, M., 1998. Interaction of the CD5 cytoplasmic domain with the Ca²⁺/calmodulin-dependent kinase II δ . *Eur. J. Immunol.* 28, 2167–2177.
- Biancone, L., Bowen, M.A., Lim, A., Aruffo, A., Andres, G., Stamenkovic, I., 1996. Identification of a novel inducible cell-surface ligand of CD5 on activated lymphocytes. *J. Exp. Med.* 184, 811–819.
- Bikah, G., Lynd, F.M., Aruffo, A.A., Ledbetter, J.A., Bondada, S., 1998. A role for CD5 in cognate interactions between T cells and B cells, and identification of a novel ligand for CD5. *Int. Immunol.* 10, 1185–1196.
- Calvo, J., Padilla, O., Places, L., Vigorito, E., Vila, J.M., Vilella, R., Mila, J., Vives, J., Bowen, M.A., Lozano, F., 1999a. Relevance of individual CD5 extracellular domains on antibody recognition, glycosylation and co-mitogenic signalling. *Tissue Antigens* 54, 16–26.
- Calvo, J., Places, L., Padilla, O., Vila, J.M., Vives, J., Bowen, M.A., Lozano, F., 1999b. Interaction of recombinant and natural soluble CD5 forms with an alternative cell surface ligand. *Eur. J. Immunol.* 29, 2119–2129.
- Hardy, R.R., Carmack, C.E., Li, Y.S., Hayakawa, K., 1994. Distinctive developmental origins and specificities of murine CD5⁺B cells. *Immunol. Rev.* 137, 91–118.
- Hardy, R.R., Hayakawa, K., 1994. CD5 B cells, a fetal B cell lineage. *Adv. Immunol.* 55, 297–339.
- Hazes, B., Price, L., 2005. A nanovolume crystallization robot that creates its crystallization screens on-the-fly. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1165–1171.
- Hosfield, D., Palan, J., Hilgers, M., Scheibe, D., McRee, D.E., Stevens, R.C., 2003. A fully integrated protein crystallization platform for small-molecule drug discovery. *J. Struct. Biol.* 142, 207–217.
- Kohfeldt, E., Maurer, P., Vannahme, C., Timpl, R., 1997. Properties of the extracellular calcium binding module of the proteoglycan testican. *FEBS Lett.* 414, 557–561.
- Li, F., Robinson, H., Yeung, E.S., 2005. Automated high-throughput nanoliter-scale protein crystallization screening. *Anal. Bioanal. Chem.* 383, 1034–1041.
- Luo, W., Van de Velde, H., von Hoegen, I., Parnes, J.R., Thielemans, K., 1992. Ly-1 (CD5), a membrane glycoprotein of mouse T lymphocytes and a subset of B cells, is a natural ligand of the B cell surface protein Lyb-2 (CD72). *J. Immunol.* 148, 1630–1634.
- McAlister, M.S., Davis, B., Pfuhl, M., Driscoll, P.C., 1998. NMR analysis of the N-terminal SRCR domain of human CD5: engineering of a glycoprotein for superior characteristics in NMR experiments. *Protein Eng.* 11, 847–853.
- Otwinowski, Z., Minor W., 1997. Processing of X-ray diffraction data collected in oscillation mode. In: *Methods in Enzymology*, vol. 276, Academic Press, New York.
- Pospisil, R., Fitts, M.G., Mage, R.G., 1996. CD5 is a potential selecting ligand for B cell surface immunoglobulin framework region sequences. *J. Exp. Med.* 184, 1279–1284.
- Pospisil, R., Silverman, G.J., Marti, G.E., Aruffo, A., Bowen, M.A., Mage, R.G., 2000. CD5 is a potential selecting ligand for B-cell surface immunoglobulin: a possible role in maintenance and selective expansion of normal and malignant B cells. *Leuk. Lymphoma* 36, 353–365.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., Seraphin, B., 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24 (3), 218–229.
- Raab, M., Yamamoto, M., Rudd, C.E., 1994. The T-cell antigen CD5 acts as a receptor and substrate for the protein-tyrosine kinase p56lck. *Mol. Cell Biol.* 14, 2862–2870.
- Raman, C., Kuo, A., Deshane, J., Litchfield, D.W., Kimberly, R.P., 1998. Regulation of casein kinase 2 by direct interaction with cell surface receptor CD5. *J. Biol. Chem.* 273, 19183–19189.
- Resnick, D., Pearson, A., Krieger, M., 1994. The SRCR superfamily: a family reminiscent of the Ig superfamily. *Trends Biochem. Sci.* 19, 5–8.
- Sarrias, M.R., Gronlund, J., Padilla, O., Madsen, J., Holmskov, U., Lozano, F., 2004. The scavenger receptor cysteine-rich (SRCR) domain: an ancient and highly conserved protein module of the innate immune system. *Crit. Rev. Immunol.* 24, 1–37.
- Simarro, M., Calvo, J., Vila, J.M., Places, L., Padilla, O., Alberola-Ila, J., Vives, J., Lozano, F., 1999. Signaling through CD5 involves acidic sphingomyelinase, protein kinase C-zeta, mitogen-activated protein kinase kinase, and c-Jun NH2-terminal kinase. *J. Immunol.* 162, 5149–5155.
- Van de Velde, H., Thielemans, K., 1996. Native soluble CD5 delivers a costimulatory signal to resting human B lymphocytes. *Cell. Immunol.* 172, 84–91.
- Van de Velde, H., von Hoegen, I., Luo, W., Parnes, J.R., Thielemans, K., 1991. The B-cell surface protein CD72/Lyb-2 is the ligand for CD5. *Nature* 351, 662–665.
- Zheng, B., Roach, L.S., Ismagilov, R.F., 2003. Screening of protein crystallization conditions on a microfluidic chip using nanoliter-size droplets. *J. Am. Chem. Soc.* 125, 11170–11171.

Bernardo Rodamilans and
Guillermo Montoya*Structural Biology and Biocomputing
Programme, Spanish National Cancer Centre
(CNIO) Macromolecular Crystallography Group,
c/Melchor Fdez Almagro 3, 28029 Madrid,
Spain

Correspondence e-mail: gmontoya@cnio.es

Received 12 December 2006

Accepted 6 February 2007

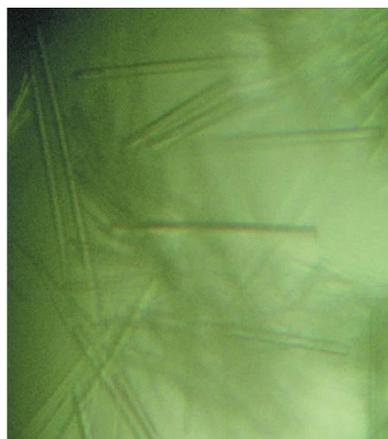
Expression, purification, crystallization and preliminary X-ray diffraction analysis of the DDX3 RNA helicase domain

DDX3 is a human RNA helicase that is involved in RNA processing and important human diseases. This enzyme belongs to the DEAD-box protein family, the members of which are characterized by the presence of nine conserved motifs including the Asp-Glu-Ala-Asp motif that defines the family. DDX3 has two distinct domains: an ATP-binding domain in the central region of the protein and a helicase domain in the carboxy-terminal region. The helicase domain of DDX3 was cloned and overexpressed in *Escherichia coli*. Crystallization experiments yielded crystals that were suitable for X-ray diffraction analysis. The final crystallization conditions were a reservoir solution consisting of 2 M ammonium sulfate, 0.1 M imidazole pH 6.4 plus 5 mM spermine tetrahydrochloride and a protein solution containing 10 mM HEPES, 500 mM ammonium sulfate pH 8.0. The crystals of the helicase domain belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 43.85$, $b = 60.72$, $c = 88.39$ Å, $\alpha = \gamma = 90$, $\beta = 101.02^\circ$, and contained three molecules per asymmetric unit. These crystals diffracted to a resolution limit of 2.2 Å using synchrotron radiation at the European Synchrotron Radiation Facility (ESRF) and the Swiss Light Source (SLS).

1. Introduction

Higher order RNA structures are formed by intramolecular and intermolecular base pairs that are functionally important. RNA helicases regulate practically all processes in which RNA manipulation is involved, from transcription and translation to nuclear export and RNA degradation (Cruz *et al.*, 1999). They catalyze the modification of higher order RNA structures, such as secondary-structure melting, strand separation and RNA-protein dissociation (Schwer, 2001). Thus, helicases are among the most numerous proteins, with representative members in all three kingdoms of life (Cruz *et al.*, 1999). Even though they are not essential in all cases, they do play a very important role in general RNA metabolism (Rocak & Linder, 2004).

DDX3, also called CAP-Rf (You *et al.*, 1999) or DBX (Lahn & Page, 1997), is a DEAD-box protein of 74 kDa whose precise functions inside the cell are not completely understood. DDX3 bears all nine motifs of the DEAD-box protein family including the characteristic Asp-Glu-Ala-Asp (DEAD box). It has two distinct regions that contain all the conserved motifs. The first region contains the Q-motif (amino acids 178–206; Cordin *et al.*, 2004) and an ATP-binding and hydrolysis domain (amino acids 209–401). This includes motifs I and II (Walker A and B, respectively), Ia, Ib and the motif III that links the ATP binding and hydrolysis to the conformational changes required for helicase activity. The second region consists of the helicase domain (amino acids 412–573) and comprises motifs IV and V. These motifs are probably involved in RNA binding together with motifs Ia and Ib from the first region. Also present in the second region is motif VI, which is believed to participate in ATP binding (Rocak & Linder, 2004). Moreover, the helicase activity of DDX3 has been shown to be dependent on ATP hydrolysis and *vice versa* (Yedavalli *et al.*, 2004).

© 2007 International Union of Crystallography
All rights reserved

This enzyme is involved in diverse processes such as mRNA splicing by co-localizing with splicing factors (Zhou *et al.*, 2002) and mRNA transport, as it is present in RNA-transporting granules (Kanai *et al.*, 2004). More recently, it has also been described as a tumour suppressor through up-regulation of p21waf1/cip1 (Chao *et al.*, 2006). In addition to its possible implication in proliferation processes, DDX3 is directly involved in other important human diseases that affect large percentages of the world population. This is the case in AIDS, in which DDX3 is upregulated (Krishnan & Zeichner, 2004) and is required for HIV-1 Rev-RRE export from the nucleus (Yedavalli *et al.*, 2004). In hepatitis C, DDX3 is down-regulated (Chang *et al.*, 2006) through a direct interaction with the HCV (hepatitis C virus) core protein during infection, suppressing its RNA-unwinding activity (Owsianka & Patel, 1999; Mamiya & Worman, 1999; You *et al.*, 1999).

Although the structures of other RNA helicases and/or their domains have been solved (Sengoku *et al.*, 2005; Caruthers *et al.*, 2000; Story *et al.*, 2001), there is scant structural information on human RNA helicases and none is available for DDX3. Even though the helicase domain is supposed to be well conserved among the different DEAD-box proteins, an accurate molecular description would be very useful in order to fully understand the molecular mechanisms of RNA unwinding and the interactions of DDX3 with other molecules.

Moreover, the fact that DDX3 is involved in different processes during the development of HIV and HCV viral infections and the possibility that this RNA helicase could function as a tumour suppressor reveal it to be a key regulator of important RNA processes. These crucial roles in RNA metabolism suggest it as a possible drug target in these diseases.

2. Materials and methods

2.1. Protein expression and purification

The cDNA sequence corresponding to residues 407–578 of human DDX3 (DDX3_{hel}) was amplified by PCR and cloned into a pCold

vector (Takara) using *Nde*I and *Bam*HI restriction sites. A 6×His tag followed by a tobacco etch virus (TEV) protease cleavage site were introduced at the 5' site. *Escherichia coli* BL21 (DE3) cells were transformed with the pColdIII-DDX3_{hel} vector. Cell cultures were grown at 310 K and induced with 1 mM IPTG when the OD reached 0.5–0.8. The temperature was then reduced to 288 K and the cultures were grown overnight. Cells were collected the next morning and the pellets were frozen at 193 K. Upon thawing, a pellet corresponding to 4 l culture was resuspended in 50 mM Tris, 500 mM ammonium sulfate pH 8.0 (buffer A). Cells were disrupted by sonication at 278 K and cell debris was removed by centrifugation. The protein was purified using a 5 ml His-Trap affinity column (Amersham Biosciences). The column was washed with buffer A and the protein was eluted with an imidazole step gradient. The eluted DDX3_{hel} was then subjected to proteolytic cleavage with TEV protease at 278 K overnight to remove the 6×His tag. The mixture was again loaded onto a 5 ml His-Trap affinity column (Amersham Biosciences) and the untagged protein eluted in the flowthrough. The sample was collected and loaded onto a gel-filtration column (Superdex 75, Amersham Biosciences) previously equilibrated with 10 mM HEPES, 500 mM ammonium sulfate pH 8.0. The protein was concentrated to 10 mg ml⁻¹ using an ultrafiltration cell (Amicon), flash-frozen in liquid nitrogen and stored at 193 K. The protein purity was analyzed by running a 10% SDS-PAGE gel (Fig. 1a) and protein homogeneity was evaluated using dynamic light scattering and mass spectrometry (data not shown). Circular-dichroism spectra and thermal denaturation showed that the protein conserved its secondary and tertiary structure before and after freezing (data not shown). The protein concentration was determined by the Bradford assay (Bradford, 1976; Bio-Rad) using BSA as a standard.

2.2. Crystallization

Crystallization screenings were performed with a Cartesian robot (Genomic Solutions) using the sitting-drop method with nanodrops of 0.1 µl protein solution plus 0.1 µl reservoir solution and a reservoir

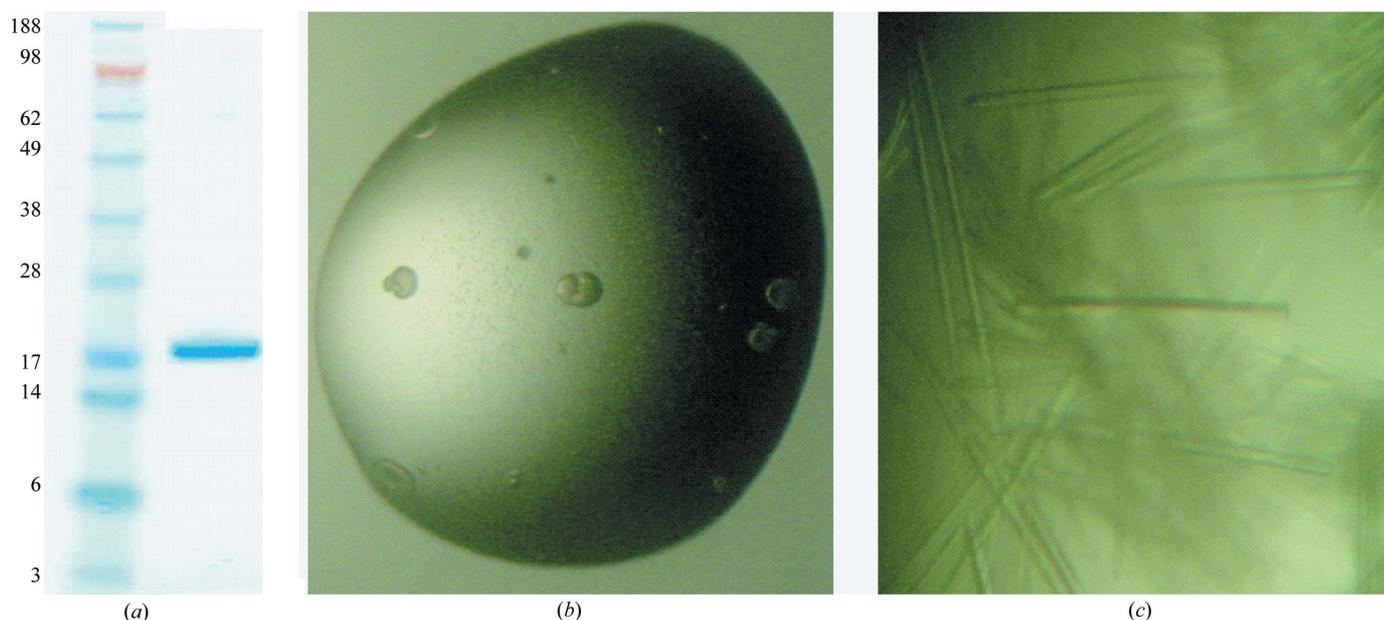


Figure 1 Purification and crystallization of the human DDX3 helicase domain. (a) SDS-PAGE of purified and concentrated DDX3_{hel}. Left lane, molecular-weight markers (kDa). (b) Initial spherulites of DDX3_{hel}. (c) Monoclinic crystals of DDX3_{hel} obtained from the initial spherulites (see text for details). Crystals of human DDX3_{hel} behaved well in the cryo-buffer conditions described in the text and diffracted to 2.2 Å. The pictures were taken at different magnification settings.

Table 1

Data-collection statistics of human DDX3 helicase domain.

Resolution limits (Å)	$\langle I/\sigma(I) \rangle$	Completeness (%)	Multiplicity	Observed reflections	Unique reflections	R_{sym}^\dagger
50.0–6.96	12.9	98.7	6.1	4712	776	0.042
6.95–4.92	11.1	99.2	7.2	9940	1363	0.055
4.91–4.02	10.6	99.4	7.5	13064	1752	0.057
4.01–3.48	8.7	99.4	7.5	15700	2081	0.073
3.47–3.11	6.0	99.4	7.6	17520	2314	0.110
3.10–2.84	3.5	99.3	6.8	17674	2582	0.177
2.83–2.63	2.4	99.3	5.6	15483	2784	0.244
2.62–2.46	1.8	99.2	4.8	14381	2968	0.294
2.45–2.32	1.5	99.1	4.3	13597	3133	0.356
2.31–2.20	1.6	99.0	4.0	13254	3331	0.388
Overall	6.2	99.0	5.9	135325	23084	0.091

$$^\dagger R_{\text{sym}} = \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$$

volume of 70 μl . Initially, small spherulites (Fig. 1*b*) were obtained after 48 h in 0.1 M HEPES pH 7.5, 2.5 M ammonium sulfate (condition No. 10 from the MbClass II suite screen; Nextal-Qiagen) at 289 K. Refinement of the crystallization conditions was performed at the same temperature using the hanging-drop method in Linbro plates. The reservoir volume was 500 μl and drops were 1 + 1 μl . After several rounds of optimization with different concentrations of precipitant, pH variations and the Additive Screen from Hampton Research, crystals appeared in 2 M ammonium sulfate, 0.1 M imidazole pH 6.4 plus 5 mM spermine tetrahydrochloride in 2 μl drops. Spermine was not critical for crystal formation, but the best data sets were obtained from crystals grown under these conditions. Although many crystals grew in the drop, these specimens could be mounted in a cryoloop. The quality and size (average dimensions $0.2 \times 0.01 \times 0.01$ mm) of the crystals were suitable for diffraction experiments (Fig. 1*c*).

2.3. Data collection and reduction

The DDX3_{hel} crystals were cryoprotected using a buffer consisting of 2 M ammonium sulfate, 0.1 M imidazole pH 6.4 and 20% glycerol. Crystals were transferred from the original drop to the cryoprotectant drop, incubated for 5 s and then flash-cooled in liquid nitrogen. Although several data sets were obtained in-house using an FR-591 Bruker generator, the crystals only diffracted to 3.8 Å. Therefore, the remainder of the data sets of DDX3_{hel} were collected using synchrotron radiation at the ID14-1 and the PX beamlines at the ESRF (Grenoble) and the SLS (Villigen), respectively. The diffraction data in Table 1 were recorded on an ADSC Q210 detector at ID14-1. The best data set was collected using $\Delta\varphi = 1^\circ$ and a wavelength of 0.9 Å. Processing and scaling were accomplished with *HKL-2000* (Otwinowski & Minor, 1997); the statistics of the crystallographic data are summarized in Table 1.

3. Results and discussion

The recombinant DDX3 helicase domain (amino acids 407–578; DDX3_{hel}) has a molecular weight of 20 kDa. Attempts to purify the full-length protein as well as constructs larger than DDX3_{hel} were initially conducted, but solubility problems led us to try this construct instead. DDX3_{hel} was expressed in *E. coli* with a 6 \times His tag at the N-terminus, with typical yields of 75 mg pure protein per litre of culture. The recombinant protein was subjected to His-tag affinity purification and the His tag was subsequently removed using TEV protease. The TEV protease was separated from the DDX3_{hel} fragment using another round of His-trap affinity chromatography and a gel-filtration purification step. The purified DDX3_{hel} was concen-

trated and used for crystallization assays. Although initial hits were just spherulites without crystalline form (Fig. 1*b*), several rounds of refinement allowed us to produce small needles. After optimization, crystals that were suitable in size and quality for diffraction experiments were obtained (Fig. 1*c*). Crystallization reproducibility was very high since only one batch of protein was used for the entire screening and optimization of the crystallization conditions.

The crystals were tested using a synchrotron-radiation source in order to obtain higher resolution data. Several native data sets were collected at 100 K on beamline ID14-1 and the PX beamline at the ESRF (Grenoble) and the SLS (Villigen), respectively. Using synchrotron radiation on these undulator-equipped beamlines, the crystals diffracted to 2.2 Å (Fig. 2). The statistics for a data set are given in Table 1.

The crystals belong to the monoclinic space group $P2_1$, with unit-cell parameters $a = 43.85$, $b = 60.72$, $c = 88.39$ Å, $\alpha = \gamma = 90$, $\beta = 101.02^\circ$. The Matthews coefficient and the self-rotation function (not shown) suggested the presence of three protein molecules per asymmetric unit ($V_M = 1.96$ Å³ Da⁻¹) and a solvent content of 37.13%. The collected diffraction data were 99% complete with a multiplicity of 5.9 and an overall $\langle I/\sigma(I) \rangle$ of 6.2 (see Table 1 for details). One possibility for solution of the structure of DDX3_{hel} is to use the molecular-replacement method, taking advantage of the previously described structures of Vasa (PDB code 2db3) from *Drosophila melanogaster*, eIF4A (PDB code 1fuk) from *Saccharomyces cerevisiae* and MjDEAD (PDB code 1hv8) from *Methanococcus jannaschii*, all of which are related to DDX3. In case of difficulties, heavy-atom derivatization or selenomethionine-derived protein could be produced to solve the structure using multiple isomorphous replacement (MIR) or multiple anomalous dispersion (MAD) experiments.

These are the first crystals reported of DDX3 RNA helicase. We believe that these studies will help to elucidate the molecular mechanisms of RNA unwinding and the important interactions of this enzyme with different proteins and RNA in the cell. Moreover, owing to the involvement of DDX3 in severe diseases such as AIDS,

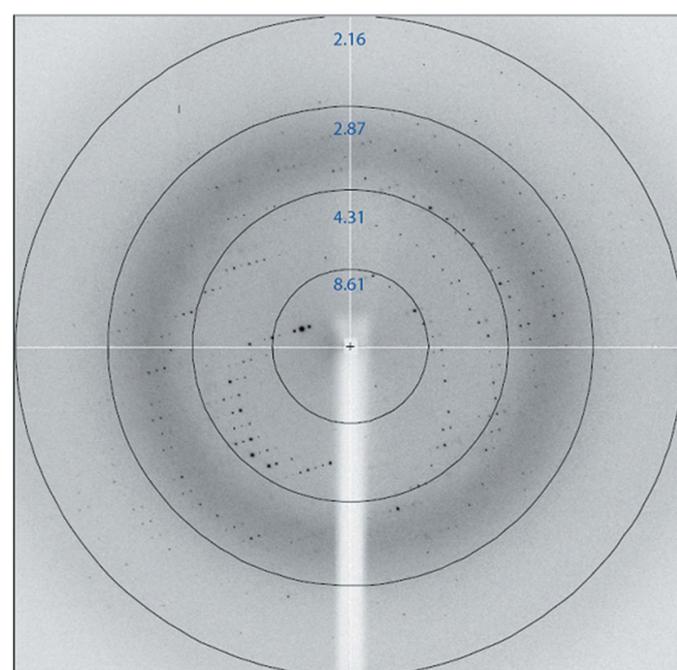


Figure 2 Diffraction pattern from the native crystals using synchrotron radiation at ID14-1 at the ESRF. Circles are labelled with the resolution limits in Å.

hepatitis C and cancer, high-resolution structural information on DDX3 could help in the design of drugs in the search for new therapies.

We would like to thank the personnel of the ESRF and SLS biocrystallography beamlines for their help during data collection. BR thanks the Consejería de Educación de la Comunidad Autónoma de Madrid (CAM) and Fondo Social Europeo (FSE) for a pre-doctoral fellowship. Financial support was obtained through Ministerio de Educación y Ciencia (BFU-2005-02403, GEN2003-20642-C09-02) grants to GM. We also thank Antonio Rosal for helpful technical assistance.

References

- Bradford, M. (1976). *Anal. Biochem.* **72**, 248–254.
- Caruthers, J. M., Johnson, E. R. & McKay, D. B. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 13080–13085.
- Chang, P. C., Chi, C. W., Chau, G. Y., Li, F. Y., Tsai, Y. H., Wu, J. C. & Wu Lee, Y. H. (2006). *Oncogene*, **25**, 1991–2003.
- Chao, C. H., Chen, C. M., Cheng, P. L., Shih, J. W., Tsou, A. P. & Lee, Y. H. (2006). *Cancer Res.* **66**, 6579–6588.
- Cordin, O., Tanner, N. K., Doere, M., Linder, P. & Banroques, J. (2004). *EMBO J.* **23**, 2478–2487.
- Cruz, J., Kressler, D. & Linder, P. (1999). *Trends Biochem. Sci.* **24**, 192–198.
- Kanai, Y., Dohmae, N. & Hirokawa, N. (2004). *Neuron*, **43**, 513–525.
- Krishnan, V. & Zeichner, S. L. (2004). *Retrovirology*, **1**, 42.
- Lahn, B. & Page, D. (1997). *Science*, **278**, 675–680.
- Mamiya, N. & Worman, H. (1999). *J. Biol. Chem.* **274**, 15751–15756.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Owsianka, A. & Patel, A. (1999). *Virology*, **257**, 330–340.
- Rocak, S. & Linder, P. (2004). *Nature Rev. Mol. Cell Biol.* **5**, 232–241.
- Schwer, B. (2001). *Nature Struct. Biol.* **8**, 113–116.
- Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. (2005). *Cell*, **125**, 287–300.
- Story, R. M., Li, H. & Abelson, J. N. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 1465–1470.
- Yedavalli, V., Neuveut, C., Chi, Y., Kleiman, L. & Jeang, K. (2004). *Cell*, **119**, 381–392.
- You, L., Chen, C., Yeh, T., Tsai, T., Mai, R., Lin, C. & Lee, Y. (1999). *J. Virol.* **73**, 2841–2853.
- Zhou, Z., Licklider, L., Gygi, S. & Reed, R. (2002). *Nature (London)*, **419**, 182–185.

Crystal Structure of the Helicase Domain of Tumour Suppressor DDX3

Bernardo Rodamilans & Guillermo Montoya*

Structural Biology and Biocomputing Programme, Spanish National Cancer Centre (CNIO)
Macromolecular Crystallography Group. c/Melchor Fdez. Almagro 3 28029 Madrid Spain.

*Corresponding author email: gmontoya@cniio.es