



RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes

Noguera, Martín Ezequiel

Estudio de las propiedades estructurales y mecanismo de plegado de la superfamilia de módulos SEA



Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Sin Obra Derivada 2.5
<https://creativecommons.org/licenses/by-nc-nd/2.5/ar/>

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:

Noguera, M. E. (2012). *Estudio de las propiedades estructurales y mecanismo de plegado de la superfamilia de módulos SEA. (Tesis de doctorado). Universidad Nacional de Quilmes, Bernal, Argentina. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes*
<http://ridaa.unq.edu.ar/handle/20.500.11807/141>

Puede encontrar éste y otros documentos en: <https://ridaa.unq.edu.ar>

Estudio de las propiedades estructurales y mecanismo de plegado de la superfamilia de módulos SEA

TESIS DOCTORAL

Martín Ezequiel Noguera

mnoguera@unq.edu.ar

Resumen

Al examinar la estructura tridimensional de proteínas relativamente grandes, se observan regiones donde los átomos se encuentran más densamente empaquetados, separadas por otras regiones de menor densidad. Estas regiones de mayor empaquetamiento se conocen como dominios proteicos. Los dominios de una misma familia guardan similitud estructural, y por lo general también una relación funcional y evolutiva. De esta manera, cuando se identifica un dominio en una proteína en estudio, se pueden inferir muchos aspectos de su función al revisar lo que se conoce previamente de ese dominio en otras proteínas. En esta tesis, se estudiaron en detalle las propiedades estructurales de cuatro representantes de la familia de dominios SEA, un grupo muy abundante y poco caracterizado de dominios presentes en proteínas funcionalmente diversas y de gran importancia biológica. Los cuatro dominios aislados, que abarcan un amplio rango de similitud de secuencia, se obtuvieron por expresión heteróloga y sus estructuras secundaria, terciaria y cuaternaria fueron estudiadas mediante métodos espectroscópicos e hidrodinámicos. También se estudiaron sus estabilidades termodinámicas mediante desnaturalización térmica y química. Se encontró que el estado de agregación y la estabilidad conformacional son variables entre los distintos dominios SEA, y no parecen reflejar aspectos funcionales comunes a toda la familia, sino más bien particularidades fisiológicas de cada miembro. También se determinó la estructura tridimensional de uno de estos dominios, proveniente del receptor fagrina, mediante difracción de rayos x. No se encontraron diferencias significativas en las estructuras obtenidas en dos condiciones de pH extremas, correspondientes a su ciclo de vida durante el proceso de secreción celular. Del análisis de las secuencias y de la información estructural disponible, se puede concluir que el plegado característico de los dominios SEA puede alcanzarse por medio de secuencias muy diferentes, dando lugar a una amplia variabilidad en las propiedades de su superficie, lo que resulta en un andamiaje robusto capaz de exhibir una gran diversidad funcional.

ÍNDICE

AGRADECIMIENTOS

PUBLICACIONES

ABREVIATURAS

Residuos de aminoácidos

1. Introducción

- 1.1. Dominios y módulos proteicos
- 1.2. La familia de dominios SEA
- 1.3. Características de algunas proteínas que contienen dominios SEA
- 1.4. Estructura y función de las mucinas de superficie celular
- 1.5. La familia de proteínas tirosina fosfatasa tipo receptor
 - 1.5.1 Las PTPR ICA512 y fgrina
 - 1.5.2. Dominio extracelular maduro de ICA512
 - 1.5.3. Dominio extracelular maduro de fgrina
- 1.6. Autoproteólisis catalizada por estrés conformacional
- 1.7. Objetivos

2. Propiedades estructurales de los dominios SEA

- 2.1. Introducción
- 2.2. Estructura primaria y definición de los dominios
- 2.3. Expresión heteróloga de las proteínas recombinantes
- 2.4. Estado de agregación
- 2.5. Flexibilidad conformacional explorada por proteólisis controlada
- 2.6. Estructura secundaria
- 2.7. Estructura terciaria
- 2.8. Estabilidad conformacional de los dominios SEA
 - 2.8.1. Desplegado por temperatura

2.8.2. Desplegado inducido por agentes químicos

2.8.2.1. Estudios de desplegado en el equilibrio a pH 7.0

2.8.2.2. Desplegado químico de MPE ICA512 a pH 4.8

2.9. Proteólisis *in vitro* de MPE fognina

3. Estructuras y secuencias de los dominios SEA

3.1. Introducción

3.2. Determinación de la estructura tridimensional de MPE fognina por difracción de rayos x

3.2.1. Cristalogénesis y resolución de la estructura

3.2.2. Estructura tridimensional de MPE fognina

3.3. Comparación de las estructuras y secuencias de MPE ICA512 y MPE fognina

3.4. Patrones de residuos conservados en las secuencias y estructuras de la región MPE de las RPTP

3.5. Comparación de las estructuras 3D de los SEA

3.6. Residuos conservados en la familia de dominios SEA

4. Discusión general y conclusiones

4.1. Características generales de los dominios SEA

4.2. Autoproteólisis de los dominios SEA

4.3. Propiedades estructurales de los dominios SEA

4.4. Conclusiones

4.5. Proyecciones de esta tesis y futuros experimentos

5. Materiales y métodos

5.1. Generalidades

5.1.1. Electroforesis de proteínas

5.1.2. Espectrometría de masas y secuenciación N-terminal

5.1.3. Determinación de la concentración de proteínas

5.1.4. Concentración y almacenamiento de proteínas

5.2. Ingeniería genética

- 5.2.1. Generalidades
- 5.2.2. Clonado del dominio M1SEA
- 5.2.2. Clonado del dominio M16SEA
- 5.2.3. Clonado del dominio MPE ICA512
- 5.2.4. Clonado del dominio MPE fogrina
- 5.2.5. Preparación y transformación de bacterias competentes
- 5.3. Expresión y purificación de proteínas
 - 5.3.1. Dominio M16SEA
 - 5.3.2. Dominio M1SEA
 - 5.3.3. Dominio MPE ICA512
 - 5.3.4. Dominio MPE fogrina
- 5.4. Reactividad química de tioles
- 5.5. Modificación química con agentes entrecruzantes
- 5.6. Digestión controlada con tripsina
- 5.7. Comportamiento hidrodinámico
 - 5.7.1. Exclusión molecular (SEC)
 - 5.7.2. Dispersión de luz estática (SLS)
- 5.8. Dicroísmo circular
 - 5.8.1. CD en el UV lejano
 - 5.8.2. CD en el UV cercano
- 5.9. Espectroscopía de fluorescencia
 - 5.9.1. Fluorescencia total de tirosinas
 - 5.9.2. Unión de ANS
- 5.10. Desnaturalización por temperatura
- 5.11. Desnaturalización química
- 5.12. Análisis de datos de equilibrio de desplegado
 - 5.12.1. Generalidades

- 5.12.2. Modelo de desplegado unimolecular de dos estados
- 5.12.3. Modelo de desplegado unimolecular de tres estados
- 5.12.4. Modelo de desplegado dimérico de dos estados
- 5.12.5. Modelos de desplegado diméricos de tres estados
- 5.13. Detección de proteasas contaminantes
- 5.14. Determinación de la estructura tridimensional de proteínas por difracción de rayos x
 - 5.14.1. Cristalización de proteínas
 - 5.14.2. Recolección de datos y procesamiento
 - 5.14.3. Resolución de la estructura, construcción del modelo y refinamiento
- 5.15. Métodos bioinformáticos

Bibliografía

Apéndices

Apéndice 1: Caracterización estructural de IA-2b⁴⁸⁶⁻⁶¹⁴

Apéndice 2: Alineamiento de secuencias de la región MPE de las RPTP

Apéndice 3: Alineamiento de los dominios SEA

AGRADECIMIENTOS

CONFIANZA. Es la palabra clave para describir gran parte del período que termina con la presentación de esta tesis.

La confianza es muy importante en muchos aspectos de la vida, en la ciencia es esencial. Así que quiero agradecer especialmente a todos aquellos que fueron importantes para que este trabajo pudiera realizarse, a quienes, de alguna forma u otra, confiaron.

A mi director de tesis, Mario Ermácora, que me abrió las puertas de su laboratorio para realizar este trabajo, permitiéndome avanzar con libertad. Por trasmitirme su conocimiento, su enorme entusiasmo por la ciencia, y el valor del profesionalismo.

A mis compañeros del LEPP (& LIE-IR) que me han dado más de una mano en la realización de los experimentos. Quiero agradecer especialmente a Evangelina, Valeria y Laura, con quienes compartimos muchos momentos, más allá del laboratorio. Gracias Evange, porque además de enseñarme y ayudarme con los experimentos, me enseñaste mucho con tu profesionalismo y eso me ayudó a crecer. Gracias Vale, por estar siempre dispuesta a ayudarme con los equipos, por tus planes para facilitarme las cosas, y por catalizarme en la toma de decisiones. Y gracias Lau, por tantos favores en el labo, por tu apoyo y aliento para que pudiera avanzar tanto en lo profesional como en lo personal.

A Raúl, por las charlas que me ayudaron a distenderme en momentos de larga y tensa purificación, y también a aprender del mundo académico. A Mauricio, por tus opiniones y humor exquisito. A Fede PdB, Noelia, Alejandro, Gabriela, Juan Pablo, Leandro, Erica y Ramiro, por los múltiples favores y/u opiniones y/o enseñanzas y sobre todo por la buena onda.

A Leopoldo, junto a quien di mis primeros pasos en el mundo experimental proteico. A "los pibes" del LEPP que reciclaron la energía del labo, haciendo las jornadas más amenas: Julián, Fede-Veiga, Ivana, Alejo, Iván y Luciano.

A Edgardo Poskus, por abrirme las puertas de su laboratorio para hacer una buena parte de este trabajo, y extendiendo este agradecimiento a la gente del LIE, en especial a Alberto, a todo el IDEHU y a la División Endocrinología del Hospital de Clínicas.

A mi consejero de tesis, Claudio Valverde, por su interés y ayuda.

A la Universidad Nacional de Quilmes, en especial al Departamento de Ciencia y Tecnología. Un agradecimiento especial para la gente de portería de la Universidad, por su comprensión y predisposición para soportar mis hábitos nocturnos de trabajo.

Al CONICET por apoyarme con una beca doctoral.

Y ahora sí, agradezco el apoyo más importante que he tenido en este tiempo, y en toda mi vida. A mi madre, Carmen, por su paciencia y ejemplo de lucha y perseverancia. A mi hermana Laura, por su entusiasmo y comprensión.

Y en memoria de mi padre, Jorge, a quien siento en mi corazón.

PUBLICACIONES

Los resultados de esta tesis han sido parcialmente publicados en el siguiente artículo:

Martín E. Noguera, María E. Primo, Laura N. F. Sosa, Valeria A. Risso, Edgardo Poskus, and Mario R. Ermácora. Biophysical characterization of the membrane-proximal ectodomain of the receptor-type protein-tyrosine phosphatase phogrin. *Protein Pept Lett.* 2012 Sep 25. (en prensa)

Además, se encuentra en preparación el artículo:

Martín E. Noguera, María E. Primo, Jean Jakoncic, Edgardo Poskus, and Mario R. Ermácora. Crystal structure and comparative structural analysis of the mature ectodomain of the receptor-type protein-tyrosine phosphatase phogrin.

Las estructuras cristalográficas resultas en esta tesis se depositaron en la base de datos PDB con los siguientes códigos: 4HTI y 4HTJ.

ABREVIATURAS

ADP	parámetro de desplazamiento atómico
ANS	(1,8) anilinaftalensulfonato
BSA	albúmina sérica bovina
CD	dicroísmo circular
CI	cuerpos de inclusión (fracción insoluble)
C_P	capacidad calorífica a presión constante
DNA	ácido desoxirribonucleico
dNTP	desoxiribonucleótidos trifosfato
DO	densidad óptica
DTNB	ácido 5,5' ditiobis (2 nitrobenzóico)
DTPA	ácido dietilen triamino pentaacético
DTT	ditiotreitol, ($\pm$)-threo-1,4-dimercapto-2,3-butanediol
EDTA	ácido etilendiaminotetracético
ϵ	coeficiente de extinción molar
F_n	fracción nativa
F_u	fracción desplegada
FPLC	cromatografía rápida líquida de proteínas y péptidos
GdmCl	cloruro de guanidinio
GS	gránulo de secreción
HPLC	cromatografía líquida de alta performance
I	estado intermediario
ID	código de identificación de una base de datos
IPTG	β -D isopropil tiogalactopiranósido
K_d	constante de disociación
K	constante de equilibrio
LB	medio de cultivo Luria Bertani

min	minutos
MPE	ectodominio proximal a la membrana
N	estado nativo
ON	<i>over night</i> , toda la noche
PCR	reacción en cadena de la polimerasa
PDB	Protein Data Bank
PM	peso molecular
PMSF	fluoruro de fenilmetilsulfonilo
RMN	resonancia magnética nuclear
rpm	revoluciones por minuto
RNA	ácido ribonucleico
R_s	radio de Stokes
s	segundos
SDS	dodecil sulfato de sodio
SDS PAGE	electroforesis en geles de poliacrilamida en presencia de SDS
SEC	cromatografía de exclusión molecular
TCA	ácido tricloroacético
U	estado completamente desplegado
UV	ultravioleta
ΔG	variación en la energía libre para el proceso de desplegado
ΔH	variación en la entalpía para el proceso de desplegado

Residuos de aminoácidos

Ala	A	alanina
Arg	R	arginina
Asn	N	asparagina
Asp	D	ácido aspártico

Cys	C	cisteína
Gln	Q	glutamina
Glu	E	ácido glutámico
Gly	G	glicina
His	H	histidina
Ile	I	isoleucina
Leu	L	leucina
Lys	K	lisina
Met	M	metionina
Phe	F	fenilalanina
Pro	P	prolina
Ser	S	serina
Thr	T	treonina
Trp	W	triptófano
Tyr	Y	tirosina
Val	V	valina

1. INTRODUCCIÓN

1.1. Dominios y módulos proteicos

Las proteínas son macromoléculas biológicas que llevan a cabo una enorme variedad de funciones con elevada eficiencia y especificidad, lo que es resultado de las propiedades de su estructura tridimensional. Desde un punto de vista químico, las proteínas son heteropolímeros formados por la condensación de los veinte aminoácidos naturales. Cada proteína tiene el potencial de adoptar un número casi ilimitado de estructuras tridimensionales o conformaciones. Sin embargo, en condiciones biológicamente relevantes, la mayor parte de las proteínas naturales¹ adopta una estructura tridimensional compacta y específica, conocida como el estado nativo. Un principio fundamental de la biología estructural establece que la conformación proteica nativa es determinada por la secuencia de aminoácidos, aunque todavía no existe una teoría satisfactoria que explique este principio (Lattman y Rose 1993; Baldwin y Rose 1999; Fitzkee, Fleming *et al.* 2005).

Cuando las proteínas se estudian desde un punto de vista estructural, funcional, o evolutivo, es conveniente definir una unidad de organización fundamental llamada dominio proteico. Las proteínas pequeñas, de menos de 100-150 residuos, se consideran formadas por un único dominio. Este parece ser el límite para un empaquetado eficiente de la cadena polipeptídica con las correspondientes cadenas laterales. Las proteínas más grandes, suelen exhibir partes menos empaquetadas que separan regiones más densas. El examen de las regiones más densas lleva a reconocerlas como dominios que se agrupan unos con otros para formar la proteína completa.

Desde un punto de vista estructural, los dominios se definen como subunidades de la proteína capaces de plegarse independientemente. De esta manera, en la práctica, un *dominio estructural* se identifica como una subestructura compacta y estable de la estructura tridimensional de la proteína, y que en general puede aislarse experimentalmente del resto de la cadena polipeptídica.

Desde un punto de vista funcional, se encuentra a menudo que un segmento de la proteína está asociado a una función específica, por ejemplo, un dominio de unión de nucleótidos de guanina o un dominio de unión a ARN, y estas regiones también suelen definirse como dominios. Estos *dominios funcionales* suelen identificarse experimentalmente separando distintos segmentos del resto de la proteína y ensayando su actividad biológica.

En ocasiones, cuando se comparan las secuencias de un grupo de proteínas, suelen identificarse regiones conservadas y muy probablemente relacionadas desde un punto de vista evolutivo, aun cuando las proteínas bajo estudio son funcionalmente diversas. Estas

¹ Aquí, el adjetivo natural se utiliza en el sentido de existente en la naturaleza en contraposición a las proteínas diseñadas y sintetizadas por el hombre.

regiones conservadas también suelen definirse como dominios, o *módulos*, tienen en general una función conservada, y se encuentran en diferentes contextos dentro de proteínas con más de un dominio.

Existen argumentos tanto fisicoquímicos como evolutivos que justifican la subdivisión en dominios. El tamaño reducido, de aproximadamente 50-120 residuos, facilita su plegado eficiente en el estado nativo. Por otro lado, una proteína con una función compleja puede construirse a partir del ensamblado de módulos funcionales preexistentes, lo que permite a los sistemas biológicos evolucionar más rápido (Hartwell, Hopfield *et al.* 1999), en lugar de tener que reinventar las mismas funciones en nuevos dominios. Algunos dominios ocurren con más frecuencia y en diversas combinaciones, por lo que se los considera “promiscuos”, y esto es un reflejo de su potencial de ser útiles en diversos contextos. Aunque en general la función de un dominio suele preservarse en el curso de la evolución, también puede cambiar como resultado de alteraciones en la secuencia, que si son seleccionadas, pueden resultar en nuevas funciones, incluso con una modificación sustancial de la estructura original.

La importancia de los dominios proteicos se puede reconocer en el hecho de que más de las dos terceras partes de todas las proteínas de organismos eucariotas identificadas en estudios de genomas completos corresponden a proteínas multidominio (Teichmann, Chothia *et al.* 1999), y también es ilustrado por el hecho de que la mayoría de los genes asociados a diversas enfermedades humanas, que han sido identificados mediante técnicas de genética inversa, corresponden también a grandes proteínas con más de un dominio (Mushegian, Bassett *et al.* 1997). Generalmente no es posible establecer la función exacta de una proteína modular simplemente por el reconocimiento de uno de sus dominios como miembro de una familia, aunque esto sí permite el diseño de estudios más detallados con ese fin. A propósito de ello, el objetivo general de este trabajo de tesis es contribuir al conocimiento de la relación estructura-función de una familia de dominios extracelulares abundantes y poco caracterizados, conocida como dominios SEA, a través del estudio de cuatro miembros representativos de esta familia, encontrados en las mucinas MUC1 y Muc16, y en los receptores tirosina fosfatasa ICA512 y fognina.

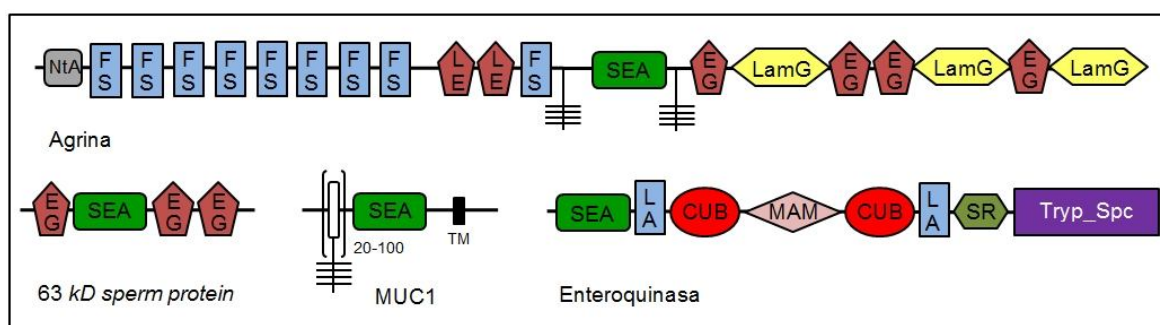
1.2. La familia de dominios SEA

La familia SEA está constituida por dominios globulares de 100-150 residuos de aminoácidos, presentes en proteínas de origen eucariota. Su nombre proviene de las primeras proteínas identificadas como miembros de la familia: *sea urchin sperm protein*, *enterokinase* y *agrin* (Bork y Patthy 1995). Suelen presentarse acompañados de otros módulos distintos, y en ocasiones se encuentra un gran número de dominios SEA en tándem. También tienen en común su presencia en proteínas profusamente glicosiladas. Algunos

ejemplos de proteínas conteniendo dominios SEA se ilustran en la Figura 1.1. Se trata de dominios de localización exclusivamente extracelular, presentes tanto en proteínas solubles secretadas, como en proteínas ancladas a la membrana celular, e incluso también en componentes de la matriz extracelular. A pesar de su abundancia, y de estar presentes en proteínas funcionalmente diversas, actualmente se desconoce la función exacta de estos dominios. Partiendo del conocimiento de la función de las proteínas en las que se encuentra presente, y de las características comunes mencionadas, se ha propuesto que los dominios SEA participan en procesos de adhesión entre células y con la matriz extracelular, y también en rutas de señalización intracelular (Bork y Patthy 1995; Wreschner, McGuckin *et al.* 2002).

Una particularidad de un gran número de dominios SEA, es que naturalmente son procesados proteolíticamente para dar un heterodímero, y se ha demostrado en varios casos que esto ocurre mediante un mecanismo autoproteolítico (Macao, Johansson *et al.* 2006). Experimentos con los dominios SEA de MUC1, distroglicano y del receptor GPR116, avalan la conclusión de que la autoproteólisis es resultado de un estrés conformacional actuando sobre un enlace peptídico específico (Akhavan, Crivelli *et al.* 2008; Johansson, Macao *et al.* 2008; Sandberg, Johansson *et al.* 2008). Los fragmentos generados permanecen fuertemente asociados, y se ha propuesto que la disociación de este complejo puede ser el modo de iniciar rutas de señalización celular (Wreschner, McGuckin *et al.* 2002).

La familia de dominios SEA está formada actualmente por un total de 711 secuencias alineadas en la base de datos Pfam (PF01390), pertenecientes a 68 especies y distribuidas en 142 arquitecturas diferentes² (Sonnhammer, Eddy *et al.* 1998). El grado de similitud de estas secuencias es variable, pero generalmente bajo. Aun así, se encuentran dos elementos conservados: la secuencia de corte autoproteolítico, que es en general G↓SVVV, y un par de residuos de cisteína. Notablemente, este par de cisteínas está ausente en los dominios que poseen la secuencia de corte, por esta razón, se ha propuesto que mientras algunos miembros de la familia SEA han evolucionado para tener una elevada estabilidad termodinámica, otros lo han hecho para formar un dímero por autoproteólisis (Macao, Johansson *et al.* 2006).



² El término arquitectura se define aquí como la colección de dominios presentes en la proteína.

Figura 1.1. Ejemplos representativos de proteínas modulares que contienen dominios SEA. NtA, dominio amino-terminal de Agrina; FS, dominio *follistatin-like*; LE, dominio *laminin* EGF-like; SEA, dominio SEA; EG, dominio EGF-like; LamG, dominio *laminin* G; LA, dominio LDL receptor class A; CUB, dominio CUB; MAM, dominio MAM; SR, dominio *Scavenger receptor cysteine-rich*; Tryp-Spc, dominio tipo tripsina; TM, región transmembrana. Los sitios de O-glicosilación se muestran con símbolos tipo antena. Adaptado de Bork y Patthy, 1995.

En cuanto a las estructuras tridimensionales, se han determinado por RMN los dominios SEA de las mucinas Muc16 (murina) y MUC1 (humana) (Maeda, Inoue *et al.* 2004; Macao, Johansson *et al.* 2006), y se obtuvo por cristalografía de rayos x la estructura del dominio SEA de la proteasa de transmembrana AsP (murina, PDB ID: 2e7v, no publicado). A estos resultados, se suma la determinación, en nuestro grupo de investigación, de la estructura del dominio extracelular maduro del receptor ICA512 (Primo, Klinke *et al.* 2008), y también de una región del ectodominio del receptor fgrina, que se describe en esta tesis. Estas estructuras poseen un plegado denominado *ferredoxin-like* que, de acuerdo a los elementos de estructura secundaria que lo constituyen, puede describirse como $\beta\alpha\beta\alpha\beta$, y está caracterizado por tener una lámina β antiparalela, cubierta en una de sus caras por dos hélices alfa, como se ilustra en la Figura 1.2. El tipo de plegado *ferredoxin-like* es muy abundante y abarca a un gran número de familias de dominios.

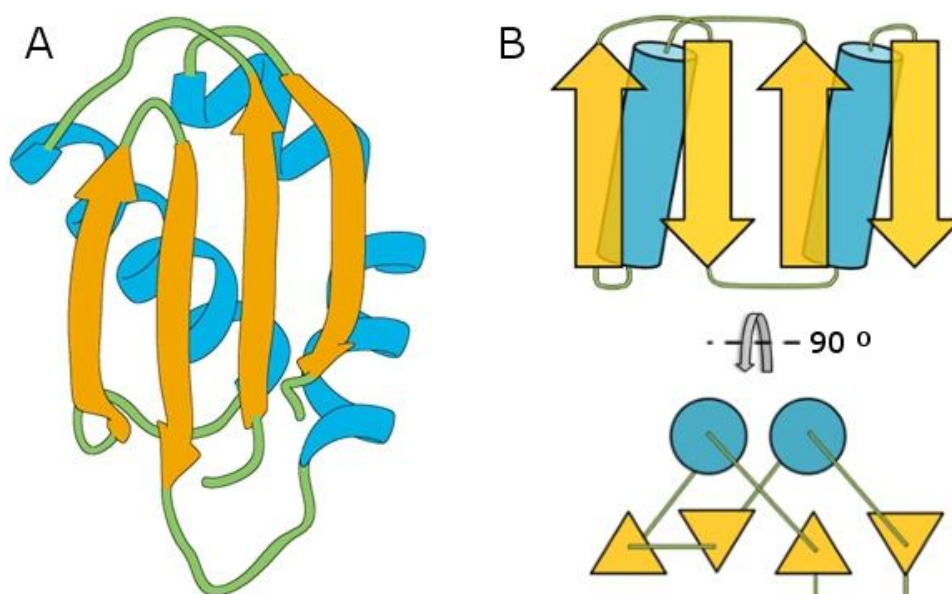


Figura 1.2. Estructura tridimensional de los dominios SEA. (A) Representación de cintas del ectodominio maduro de ICA512. (B) Diagrama topológico del plegado tipo ferredoxina mostrado en dos orientaciones, arriba: en la misma orientación que en A, abajo: una rotación de 90° en torno a un eje horizontal en el plano del papel. Nótese que el plegado comprende dos motivos $\beta\alpha\beta$ consecutivos, lo que resulta en una lámina β de cordones antiparalelos; en contraste con otro

arreglo frecuente en las proteínas donde el motivo $\beta\alpha\beta$ forma una serie $\beta\alpha[\beta\alpha]_n\beta\alpha$ que resulta en láminas de cordones paralelos.

1.3. Características de algunas proteínas que contienen dominios SEA

Como se mencionó anteriormente, poco es lo que se sabe acerca de la función exacta de los dominios SEA, a pesar de que están ampliamente distribuidos y se encuentran en proteínas de gran importancia biológica. Para intentar comprender la función biológica de un dominio, es útil repasar las funciones conocidas de aquellas proteínas que lo contienen. Las propiedades funcionales de los dominios SEA se han estudiado en mayor detalle en la familia de las mucinas de superficie celular, lo cual se describe en detalle en la Sección 1.4, en esta sección se repasan las características de otras proteínas destacadas que contienen este dominio.

De las secuencias depositadas en la base de datos Pfam³, la mayoría corresponde a proteínas que tienen segmentos transmembrana, y también es lo más común que presenten un solo SEA como único dominio (21% del total de secuencias), como es el caso de la mucina MUC1. La segunda arquitectura más frecuente corresponde a un dominio SEA seguido de uno tipo tripsina (8%). También es más habitual que exista solo un dominio SEA, o dos por molécula (94% y 4% de los casos, respectivamente), que pueden estar acompañados de un número variable de otros dominios. El dominio acompañante más frecuente es el de tipo tripsina (presente en un 21 % de las secuencias), y en segundo lugar, el dominio LDL-receptor-A (14%).

En cuanto a las proteínas que contienen SEA, un ejemplo destacado es perlecano, que es el proteoglicano más abundante de las membranas basales. Su porción proteica contiene un solo dominio SEA y una variedad de otros módulos, y tiene varias funciones como la unión de *nidogen* y *fibulin-2*, que son dos componentes de la matriz extracelular (ECM)(Battaglia, Mayer *et al.* 1992; Sasaki, Gohring *et al.* 1995). Por otro lado, su porción glucídica (glucosaminoglicanos) es responsable de la unión a otros componentes de la ECM, como fibronectina, colágeno IV, laminina-1 y bFGF (Iozzo, Cohen *et al.* 1994).

La proteína agrina es un proteoglicano de la membrana basal de la unión neuromuscular, que participa en los procesos de formación y mantenimiento de la sinapsis y de la diferenciación post-sináptica (Tsen, Halfter *et al.* 1995). Es responsable del agrupamiento de los receptores de acetilcolina y de otras proteínas de la unión neuromuscular (Nastuk y Fallon 1993; Patthy y Nikolics 1993). Posee una variedad de

³ Accedido durante el mes de marzo de 2012

dominios, incluyendo un único dominio SEA, que está altamente conservado entre distintas especies.

La proteína TMPRSS6 (matriptasa-2) es una serina proteasa que consta de una región intracitoplasmática, un segmento transmembrana y una región extracelular con un dominio SEA y una variedad de otros módulos, incluyendo un dominio tipo tripsina responsable de la actividad catalítica (Velasco, Cal *et al.* 2002). Esta proteína está relacionada con la regulación de los niveles de hierro a través de su acción proteolítica sobre otras proteínas involucradas en la detección de hierro (Du, She *et al.* 2008). De esta forma, un grupo de mutaciones dentro del dominio SEA de TMPRSS6, están asociadas a una forma de anemia por refractancia a hierro (IRIDA) (Finberg, Heeney *et al.* 2008; Melis, Cau *et al.* 2008; Altamura, D'Alessio *et al.* 2010).

La enzima enteroquinasa es la responsable de la conversión del tripsinógeno en tripsina, iniciando una cascada de reacciones proteolíticas que producen la activación de muchos zimógenos provenientes del fluido pancreático (Light y Janska 1989). Es una proteína altamente glicosilada, que interactúa con la membrana de las células intestinales. Posee una arquitectura altamente modular, incluyendo un único dominio SEA y un dominio catalítico tipo tripsina (Bork y Patthy 1995). Se postula que también está involucrada en una variedad de interacciones con otras macromoléculas de la superficie celular, que están mediadas por sus dominios modulares (Bork y Patthy 1995).

La proteasa transmembrana matriptasa es una proteína multidominio, que se ha detectado en la mayor parte de los tejidos epiteliales, y parece tener un rol fisiológico crucial en el funcionamiento de las barreras epidérmicas. Además, juega un rol importante en los procesos de invasión y metástasis de cáncer, a través de su acción proteolítica sobre los componentes de la ECM (Bugge, List *et al.* 2007). Esta proteasa contiene un único dominio SEA, entre otros dominios, que posee la secuencia de corte GSVI. La activación del zimógeno de matriptasa requiere de dos eventos de procesamiento proteolítico. El primero es la autoproteólisis dentro de la secuencia de corte del dominio SEA, luego de lo cual las dos subunidades permanecen asociadas en la superficie celular (Cho, Kim *et al.* 2001), y el evento siguiente es el corte justo al inicio del dominio catalítico, lo que resulta en su activación (Oberst, Williams *et al.* 2003). El procesamiento dentro del dominio SEA es un prerequisite para la activación, ya que los mutantes diseñados para interferir la autoproteólisis también afectan severamente la activación (Oberst, Williams *et al.* 2003).

Recientemente se han aportado un número de evidencias que indican que el distroglicano, una proteína integral de la membrana celular que une el citoesqueleto con la matriz extracelular, posee un dominio SEA (Akhavan, Crivelli *et al.* 2008). Esta glicoproteína, codificada por un único gen, se encuentra constitutivamente escindida en dos subunidades, la subunidad α que está altamente glicosilada y se une con alta afinidad a proteínas de la lámina

basal, como laminina, agrina y perlecano (Winder 2001), y la subunidad β , que posee una región extracelular, un segmento transmembrana y una porción intracelular que se une a distrofina (o utrofina en tejidos no musculares), que a su vez se une a actina (Huang, Poy *et al.* 2000). El dominio SEA comprende segmentos de las dos subunidades, ya que es el sitio donde ocurre el corte proteolítico, dentro de la secuencia de corte G↓SIVV. La inhibición del corte proteolítico mediante el remplazo de la serina del sitio de corte por alanina, condujo a un incremento en la afinidad por laminina, como resultado de una mayor glicosilación de la subunidad α (Akhavan, Crivelli *et al.* 2008). De esta forma, la proteólisis dentro del SEA puede ser un mecanismo de regulación de su actividad de unión a laminina.

En resumen, la función de los dominios SEA parece estar relacionada con la modulación de la actividad de otras regiones funcionales de las proteínas de las que forma parte. Esto puede ser a través de la modificación de sus propiedades adhesivas, como se observa en distroglicano, o la participación en el proceso de maduración de la proteína, como en la activación proteolítica de la proteasa matriptasa.

1.4. Estructura y función de las mucinas de superficie celular

Las mucinas son proteínas altamente glicosiladas producidas por todas las células epiteliales, aunque en cantidades variables dependiendo del tipo de célula y del tejido. Más del 50% de su masa está constituida por azúcares unidos a residuos de serina o treonina del esqueleto polipeptídico, por lo que son altamente hidrofílicas. Las mucinas se pueden clasificar en dos grandes familias: las mucinas secretadas (o formadoras de gel), que son exclusivamente extracelulares, y las mucinas de superficie celular (o unidas a membrana), que poseen un segmento transmembrana. Las mucinas secretadas contribuyen a la formación de las barreras de protección del tracto respiratorio, gastrointestinal y de otros órganos. Las mucinas de superficie celular también contribuyen a las barreras mucosas de protección de los epitelios, pero además juegan importantes roles en procesos de señalización celular a través de su pequeño segmento intracitoplasmático (Hattrup y Gendler 2008).

La familia de las mucinas está caracterizada por la presencia, en su segmento extracelular, de un número variable de repeticiones en tándem (VNTR), que consisten en secuencias ricas en residuos de prolina, treonina y serina, cuyo número suele estar sujeto a variaciones por polimorfismo genético (Fowler, Vinall *et al.* 2001). En estos residuos de serina y treonina se anclan los residuos de azúcares. Las mucinas de superficie celular están caracterizadas también por la presencia de al menos un dominio SEA, a excepción de la mucina MUC4. También suelen encontrarse otros módulos, tales como *epidermal growth*

factor-like (EGF), *Nidogen* (NIDO) y *von Willebrand factor type-D* (VWD), entre otros (Duraismy, Ramasamy *et al.* 2006).

Las mucinas MUC1, MUC4 y MUC16 se expresan de manera aberrante en varios tipos de cáncer. Se ha observado sobreexpresión de MUC1 en cáncer de mama, pulmón, próstata, páncreas, ovario y otros epitelios (Chauhan, Singh *et al.* 2006; Bafna, Kaur *et al.* 2010). En el proceso de transformación se observan altos niveles de MUC1 sobre toda la superficie celular y el citoplasma (Gendler 2001). MUC16 se sobreexpresa en cáncer de ovario, y se emplea ampliamente como un marcador para su diagnóstico clínico (donde recibe el nombre de antígeno CA125) (Bast, Feeney *et al.* 1981; Maggino y Gadducci 2000; Rustin, Marples *et al.* 2001).

La mucina humana MUC1 contiene un único dominio SEA, que se localiza adyacente a la región transmembrana (Hattrup y Gendler 2008). Este dominio es escindido tempranamente, en la secuencia $G^{1097}\downarrow S^{1098}VVV$, mientras se encuentra en el retículo endoplasmático, y se ha demostrado que esto ocurre a través de un mecanismo autoproteolítico (Ligtenberg, Kruijschaar *et al.* 1992; Parry, Silverman *et al.* 2001; Levitin, Stern *et al.* 2005). Una vez escindido, los fragmentos generados permanecen asociados de manera no covalente, presentándose como un heterodímero en la superficie celular. Se ha propuesto que el proceso de disociación y asociación de las subunidades generadas constituye un mecanismo de regulación en rutas de señalización celular (Wreschner, McGuckin *et al.* 2002).

La estructura del dominio SEA de MUC1 se determinó por RMN, y es una estructura compacta con las hélices α empaquetadas sobre la superficie de una hoja β antiparalela de cuatro hebras (Macao, Johansson *et al.* 2006), conforme a las características de los SEA. El dominio puede describirse como $\beta\alpha\beta\beta\alpha\beta$, con una superficie altamente polar, donde los residuos que se cree que están N-glicosilados en la proteína nativa se encuentran expuestos al solvente. Los residuos del sitio de corte están parcialmente ocultos, pero los extremos N y C generados por el corte proteolítico, se encuentran expuestos al solvente e integrados a la hoja β , preservando las interacciones por puente de hidrógeno. Los intentos de reconectar dichos extremos en simulaciones de modelado molecular generaron grandes distorsiones que afectaron a toda la estructura, lo que condujo a la hipótesis de que existe una base estructural para el mecanismo de autoproteólisis de los dominios SEA, asociada al estrés conformacional (esto se explica en detalle en la Sección 1.6).

La mucina humana MUC16 posee varios dominios SEA en su región extracelular, que se repiten 7, 12 o 60 veces, de acuerdo a la isoforma considerada (O'Brien, Beard *et al.* 2001; Bouanene y Miled 2010; Weiland, Martin *et al.* 2012). Las secuencias de los distintos dominios SEA de MUC16 son más similares entre sí que a las de los dominios SEA de otras proteínas, por esta razón se cree que la multiplicación de los dominios de MUC16 se produjo

con posterioridad a la aparición de esta proteína. Los cuatro dominios SEA más cercanos a la membrana difieren más entre sí que con el resto, y los del extremo N son los que presentan mayor similitud, lo que sugiere que estos últimos se duplicaron luego de que lo hicieran los cuatro más cercanos a la membrana (Maeda, Inoue *et al.* 2004). Es notable que sólo uno de los dominios SEA de MUC16, el segundo desde la membrana, posea un motivo compatible con la secuencia consenso de corte, y que por lo tanto podría sufrir autoproteólisis, aunque aun no ha sido reportado ningún procesamiento proteolítico en esta región.

Un homólogo murino de MUC16, clonado a partir de una biblioteca de cDNA neonatal, posee un único dominio SEA, cuya estructura tridimensional se determinó por RMN (Maeda, Inoue *et al.* 2004). La proteína completa posee un total de 258 aminoácidos, con una región transmembrana y un dominio SEA, que llamaremos M16SEA a lo largo de esta tesis, que posee alta similitud de secuencia con el SEA más cercano a la membrana plasmática del homólogo humano. Su estructura puede ser dividida en dos capas, una conteniendo la hoja beta antiparalela de cuatro cordones y otra con las dos hélices α . Esta molécula posee dos residuos de cisteína muy próximos entre sí, localizados en la segunda y la tercera hebra β , suficientemente cercanos como para formar un enlace disulfuro. La estructura global se describe como un plegado $\beta\alpha\beta\beta\alpha\beta$, excepto por pequeñas inserciones de elementos de estructura secundaria. Se observan dos lazos cortos (L1, residuos 64-72; L2, residuos 81-87) y una región desestructurada correspondiente a los residuos 66-69, que se superponen con el sitio de corte del dominio SEA de MUC1.

El dominio M16SEA está altamente cargado en forma positiva a pH neutro, y ya que una función propuesta para los dominios SEA es la unión de residuos de azúcares, se estudiaron sus propiedades de unión hacia distintos compuestos relacionados con azúcares, tales como ATP, sacarosa, mucina de estómago porcino y condroitín sulfato, con los cuales no se detectó interacción (Maeda, Inoue *et al.* 2004).

1.5. La familia de proteínas tirosina fosfatasa tipo receptor

Como se mencionó anteriormente, los dominios SEA han sido encontrados en proteínas de membrana pertenecientes a la familia de las tirosina fosfatasas. Muchos procesos de señalización intracelular son mediados por fosforilación de residuos de tirosina, a través de la acción coordinada y contrapuesta de proteínas tirosina quinasas (PTK) y de proteínas tirosina fosfatasas (PTP). Mientras que las PTK fosforilan residuos específicos de tirosina de sus moléculas blanco pertenecientes a rutas de señalización, las PTP los desfosforilan. Su acción conjunta regula la división, crecimiento, diferenciación y metabolismo celular.

La familia de las PTP puede clasificarse en cuatro grupos de acuerdo a las secuencias de sus dominios catalíticos, el grupo más grande es el de las PTP basadas en cisteína tipo-I,

que está caracterizado por la presencia del motivo C(X)₅R, que forma parte del sitio activo. Los miembros de este grupo pueden dividirse adicionalmente en dos categorías: las PTP clásicas, que desfosforilan específicamente residuos de tirosina, y las PTP de especificidad dual, que pueden desfosforilar también residuos de serina o de treonina (Patterson, Brummer *et al.* 2009).

Adicionalmente, el grupo de las PTP clásicas se puede subdividir en PTP citoplasmáticas (o no transmembrana, PTPNT) y PTP de membrana (o tipo receptor, PTPR), aunque esta no es una división absoluta, ya que el uso de promotores alternativos, o de *splicing* alternativo, puede llevar a que los marcos de lectura codificantes para PTPR se expresen de forma soluble sin la región transmembrana (Tiganis y Bennett 2007).

La mayoría de las PTPR suele tener dos tipos de dominios tirosina fosfatasa (PTP), uno más cercano a la membrana, que se denomina D1, y otro más alejado, llamado D2. Mientras que el dominio D1 es activo, el D2 es inactivo o con una actividad mucho menor, por lo que D1 suele considerarse el dominio catalítico y D2 como regulatorio (Tonks 2006).

Como es el caso de otros receptores, la porción extracelular de las PTPR está involucrada en la transducción de señales por unión a ligandos solubles, pero además se ha implicado en diversas interacciones célula-célula y célula-matriz extracelular (Burridge, Sastry *et al.* 2006; Sallee, Wittchen *et al.* 2006).

La función de las PTP puede regularse a nivel postraducciona mediante diversas modificaciones como proteólisis, fosforilación, sumoilación y oxidación (Stoker 2005). La activación proteolítica es un mecanismo muy frecuente entre las PTP (Aicher, Lerch *et al.* 1997). La oxidación reversible de la cisteína del sitio activo, suele ocurrir preferencialmente en D2 pero repercute en la actividad catalítica del dominio D1, aparentemente mediante cambios conformacionales (Tonks 2006). Otros mecanismos regulatorios importantes son la unión de ligandos, como proteínas de la matriz extracelular, y la dimerización, ya sea por la formación de homodímeros o de heterodímeros con otras PTP (Chin, Sachs *et al.* 2005; Tonks 2006).

1.5.1 Las PTPR ICA512 y fognina

Los receptores ICA512 (también conocido como IA-2, PTP35 y RPTPN) y fognina (también conocido como IA-2β, ICAAR, IAR y RPTPN2) son dos proteínas homólogas pertenecientes a la familia de las PTPR, que fueron identificados en células pancreáticas y se ha demostrado que son dos de los principales autoantígenos en diabetes mellitus tipo I (DMI) (Lan, Wasserfall *et al.* 1996; Solimena, Dirx *et al.* 1996; Bottazzo, Bosi *et al.* 2005). Se encuentran localizados en los gránulos de secreción de cerebro, hipófisis, islotes pancreáticos y en general, de células de origen neuroendocrino (Solimena, Dirx *et al.* 1996; Wasmeier y

Hutton 1996). Comparten una arquitectura común, con un péptido señal, un ectodominio, una región transmembrana y un único dominio tirosina fosfatasa (Torii 2009). Globalmente, fgrina posee un 42% de identidad de secuencia con ICA512, mientras que es de un 80% en el dominio tirosina fosfatasa, y sólo un 29% en el ectodominio. Tanto en ICA512 como en fgrina, el dominio PTPasa tiene mutaciones en el sitio activo que lo hacen inactivo hacia los sustratos fosforilados convencionales, aunque recientemente se ha demostrado que fgrina es capaz de desfosforilar fosfolípidos de inositol de la membrana plasmática (Caromile, Oganessian *et al.* 2010), y esto podría tener importantes implicancias biológicas en el mecanismo de exocitosis de insulina en respuesta a estímulos.

Aproximadamente el 50% de los diabéticos recientemente diagnosticados tienen autoanticuerpos contra fgrina, que son expresados algunos años antes del diagnóstico clínico (Notkins, Zhang *et al.* 1997). En el caso de ICA512, esta proporción se eleva a un 60-70%. Dado que ICA512 y fgrina se localizan junto con insulina en los gránulos de secreción (GS) de las células β -pancreáticas, y que en ratones *knock out* que carecen de uno o ambos receptores se observan alteraciones en la secreción de insulina inducida por glucosa (GSIS) y también de hormonas hipofisarias (Saeki, Zhu *et al.* 2002; Kubosaki, Gross *et al.* 2004; Kubosaki, Nakamura *et al.* 2005; Kubosaki, Nakamura *et al.* 2006), se ha propuesto que estos receptores son importantes reguladores de las rutas de secreción. En un trabajo reciente, se ha encontrado que tanto el silenciamiento como la sobreexpresión de fgrina, en líneas celulares de insulinoma y en cultivos de islotes aislados, no tiene efecto sobre la GSIS, aunque la razón de estas discrepancias aún no ha sido aclarada (Torii, Saito *et al.* 2009). Sin embargo, estos mismos estudios sugieren que ICA512 y fgrina son importantes reguladores de la proliferación de las células β -pancreáticas, posiblemente a través de su interacción directa con el receptor de insulina, el cual es activado de modo autocrino/paracrino por insulina, y lleva a la estabilización de la proteína IRS2, que es un regulador del crecimiento de las células β .

1.5.2. Dominio extracelular maduro de ICA512

La proteína ICA512 se sintetiza con 979 residuos de aminoácidos y un peso molecular de 106 kDa. Se ubica en los GS con su dominio intracelular expuesto al citoplasma celular y la región extracelular dentro del gránulo (Lan, Lu *et al.* 1994). Durante su maduración, ICA512 sufre diversas modificaciones postraduccionales, incluyendo el corte del péptido señal (residuos 1-34), glicosilaciones y el corte proteolítico entre los residuos 448 y 449 por una convertasa de tipo furina, lo que genera dos fragmentos (Hermel, Dirkx *et al.* 1999). Cuando la proteína madura alcanza la membrana plasmática durante la exocitosis, el dominio intracelular es escindido por una proteasa dependiente de calcio, y migra al núcleo donde

activa la expresión de genes involucrados en el proceso de secreción, incluido el de insulina (Ort, Voronov *et al.* 2001; Trajkovski, Mziaut *et al.* 2004; Mziaut, Trajkovski *et al.* 2006).

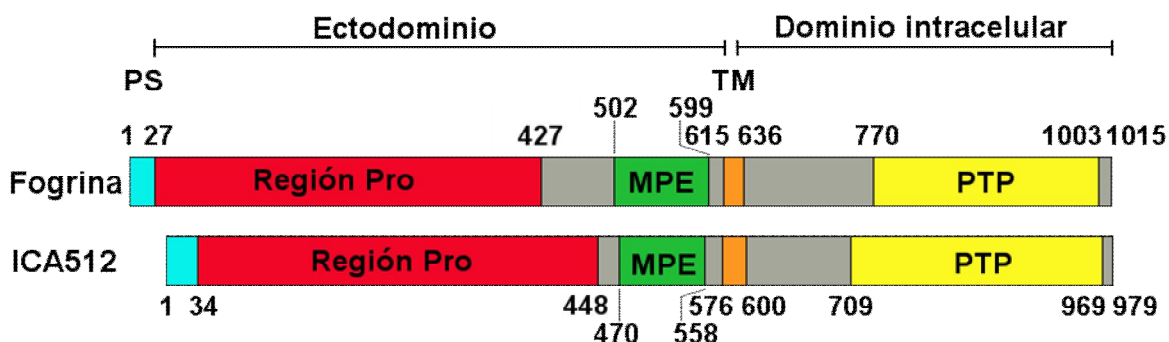


Figura 1.3. Arreglo de dominios de las proteínas ICA512 y fogrina. Se señalan distintas regiones, PS: péptido señal, TM: segmento transmembrana, PTP: dominio tirosina fosfatasa, MPE: dominio proximal a la membrana.

El estudio del segmento extracelular de la proteína ICA512 madura, llevó a nuestro grupo de investigación a determinar su estructura tridimensional (Primo, Klinke *et al.* 2008). Durante las pruebas de cristalización, se encontró que durante la incubación, el dominio recombinante y altamente purificado sufre un proceso de proteólisis espontánea mediado por radicales libres, que resulta en la remoción de unos veinte residuos de cada extremo, generando una variante acortada, MPE ICA512, a partir de la cual se obtuvieron los cristales que permitieron la resolución de la estructura tridimensional. De esta forma, la similitud estructural, su contexto dentro de la proteína completa, y su capacidad de sufrir proteólisis espontánea nos llevaron a relacionar este segmento de ICA512 con los dominios SEA, y esto también se confirmó mediante un análisis estadístico de las secuencias de los dominios SEA y de las PTP.

MPE ICA512 cristaliza en forma de dímeros o como un tetrámero, con dos posibles interfaces de dimerización (Primo, Klinke *et al.* 2008; Primo, Jakoncic *et al.* 2011). Los dos modos de interacción parecen ser de relevancia funcional, ya que los monómeros conservan elementos de simetría propios de los dímeros biológicos. Los modos de asociación de MPE ICA512 se discuten en la Sección 3.4.

Cada monómero en la estructura cristalográfica es compacto y bien estructurado. Los lazos son cortos y la segunda hélice α tiene el pliegue característico de los dominios SEA. La superficie está cargada negativamente a pH neutro y los extremos N y C se encuentran espacialmente cercanos.

1.5.3. Dominio extracelular maduro de fogrina

La proteína fgrina humana es sintetizada inicialmente como una glicoproteína de 130 kDa, que luego es procesada proteolíticamente eliminando primero el péptido señal (residuos 1-27) y luego la región Pro (residuos 28-427), este último corte ocurre en una secuencia dibásica que es un sitio de reconocimiento característico de proteasas denominadas prohormona convertasas, comunes en los tejidos neuroendocrinos, y también se han reportado otros cortes proteolíticos en cerebro y células β -pancreáticas de ratón (Tomiko Kawakami 2007). Una de estas isoformas involucra el corte entre los residuos 501 y 502 por una endopeptidasa no identificada, dejando unida a la membrana plasmática una región del ectodominio de fgrina que es equivalente a MPE ICA512, y que por lo tanto hemos denominado MPE fgrina.

Mientras que MPE ICA512 es un dímero que se encuentra en equilibrio con cantidades significativas del monómero en el rango de concentraciones micromolar, demostramos en esta tesis que MPE fgrina es un monómero en las mismas condiciones. Se ha propuesto que el proceso de homo y heterodimerización es una forma de regular la actividad de los receptores tirosina fosfatasa (Chin, Sachs *et al.* 2005), y se ha demostrado que tanto ICA512 como fgrina pueden formar dímeros con otras PTPR (Gross, Blanchetot *et al.* 2002), sin embargo, se desconoce el detalle molecular de esta interacción.

1.6. Autoproteólisis catalizada por estrés conformacional

El fenómeno de autoproteólisis, es decir, la hidrólisis de un enlace peptídico por una reacción intramolecular, se presenta en varias proteínas, como las inteínas y otros sistemas que sufren maduración por el corte auto-catalizado de un precursor proteico (Paulus 2000).

La hidrólisis de un enlace peptídico es una reacción termodinámicamente favorable, pero extremadamente lenta, con una vida media de aproximadamente 2100 años a pH neutro y temperatura ambiente (Radzicka y Wolfenden 1996). Para hacer esta reacción compatible con los tiempos fisiológicos, la proteólisis enzimática consigue aumentar la velocidad de reacción al disminuir la barrera energética que debe ser sobrepasada para llegar a los productos, mediante la estabilización del estado de transición, lo que suele explicarse mediante el modelo de pre-organización del sitio activo (Warshel, Sharma *et al.* 2006), como se ilustra esquemáticamente en la Figura 1.4A.

Un mecanismo catalítico diferente se ha propuesto para la autoproteólisis de los dominios SEA, y que probablemente pueda ser extendido a otros sistemas que sufren autoproteólisis (Macao, Johansson *et al.* 2006; Johansson, Macao *et al.* 2008; Sandberg, Johansson *et al.* 2008; Johansson, Wallin *et al.* 2009; Wallin, Hård *et al.* 2012). Este proceso se ha estudiado en mayor detalle en el dominio SEA de la mucina MUC1, y explica la

aceleración de la reacción por medio de la desestabilización conformacional de un precursor (Figura 1.4B).

La mucina MUC1 se encuentra unida a la membrana celular en forma de heterodímero, y el corte proteolítico se produce en el dominio SEA dentro de la secuencia consenso G↓SVVV (Parry, Silverman *et al.* 2001). Experimentos de incubación de la proteína recombinante, altamente purificada, mostraron que el procesamiento ocurre por autoproteólisis e inmediatamente luego de que la proteína es sintetizada (Levitin, Stern *et al.* 2005). Lo mismo ocurre cuando se reemplaza la serina en la secuencia consenso por otros residuos de carácter nucleofílico, como treonina y cisteína, pero no cuando se sustituye por el resto de los aminoácidos naturales. Este es el primer paso del mecanismo de autoproteólisis, el ataque nucleofílico sobre el carbono carbonílico del enlace peptídico precedente, lo que resulta en la formación de un éster, que es luego hidrolizado por un segundo nucleófilo para formar los nuevos extremos N y C. Este primer paso recibe el nombre de N→O (o N→S) *acyl shift*, y explica una parte del aumento de la velocidad de hidrólisis, sin embargo, la proteólisis de M1SEA se observa dentro de los pocos minutos posteriores al inicio del plegado del precursor, lo que sugiere una asociación entre el plegado y la proteólisis.

Las observaciones que llevaron a postular que el aumento de la velocidad de reacción proviene del estrés conformacional son: (i) al determinar la estructura tridimensional de M1SEA, no fue posible reconectar por modelado molecular los extremos N y C generados por autoproteólisis, sin introducir severas alteraciones en toda la estructura; (ii) se determinó que ningún otro grupo funcional de los residuos cercanos al sitio de corte participa en el proceso de proteólisis, sólo se requieren la serina catalítica y la integridad conformacional del *loop*; (iii) el plegado de la molécula al estado nativo, desde el precursor no escindido y desplegado químicamente, se produce simultáneamente con la autoproteólisis, es decir la autoproteólisis y el plegado son procesos acoplados; (iv) la velocidad de autoproteólisis disminuye cuando se desestabiliza la conformación del dominio, ya sea mediante mutagénesis o el empleo de aditivos químicos; y (v) la supresión de la tensión conformacional añadiendo flexibilidad por la inserción de glicinas en el sitio de corte inhibe la autoproteólisis.

Todos estos resultados convergen en un modelo donde las tensiones conformacionales que se generan cuando un precursor no escindido se pliega al estado nativo, resultan en la fractura de un enlace peptídico específico, lo que es catalizado por la acción de un residuo que actúa como nucleófilo, y también por la energía libre de plegado (Figura 1.4B). Medidas termodinámicas y cinéticas, y estudios computacionales, están de acuerdo en la magnitud de la energía almacenada como tensión conformacional. Estudios similares en otros dominios SEA, como el de distroglicano, el del receptor GPR116 y el de la mucina MUC3 también avalan esta hipótesis (Akhavan, Crivelli *et al.* 2008; Sandberg, Johansson *et al.* 2008; Peng, He *et al.* 2010).

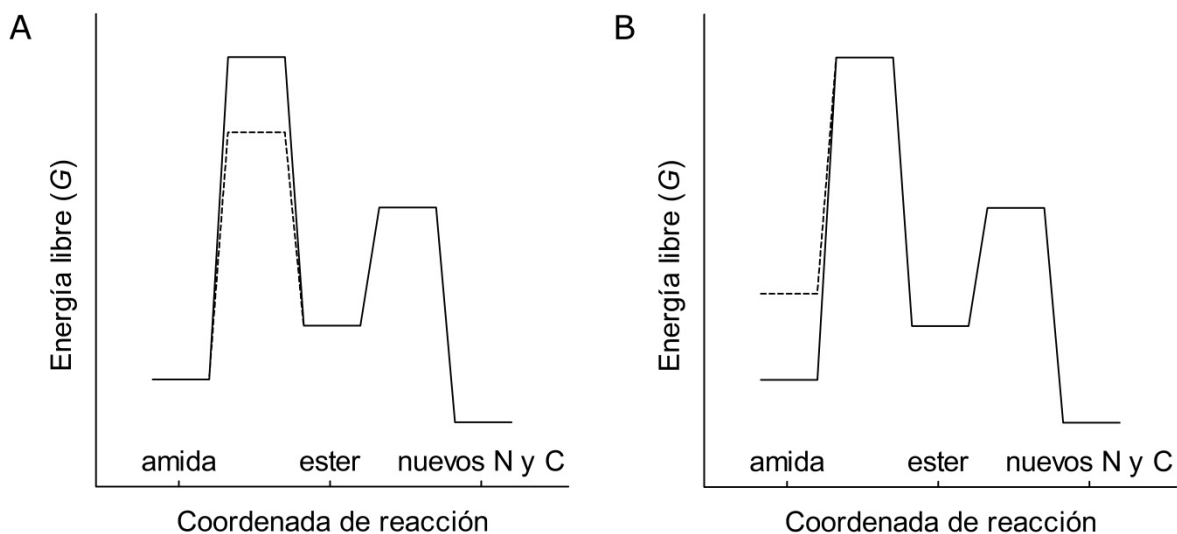


Figura 1.4. Modelos simplificados de catálisis de la hidrólisis del enlace peptídico. (A) Se ilustra un mecanismo de catálisis enzimática, por medio de la estabilización del estado de transición. (B) Mecanismo de autoproteólisis catalizada por estrés conformacional, donde se baja la barrera energética al aumentar la energía libre del reactivo (el enlace peptídico).

1.7. Objetivos

Los dominios SEA están presentes en la región extracelular de muchas proteínas de gran importancia en estados fisiológicos y patológicos, sin embargo, la función exacta de estos dominios se desconoce. Nuestro objetivo final es determinar si existen propiedades estructurales comunes a todos ellos, y qué relación existe entre estas propiedades y los aspectos de su función. Por ejemplo, se pretende establecer si los escasos residuos conservados en sus secuencias son importantes para alcanzar su estructura terciaria característica o son necesarios para su función biológica, y en este caso serían identificados residuos 'blanco' para el estudio de los detalles funcionales. Además, ya que esta familia de dominios presenta una alta similitud estructural, pero escasa identidad de secuencia, su estudio estructural comparativo es una contribución al estudio de aspectos elementales de la reacción de plegado de las proteínas.

Para este fin, los objetivos específicos que se abordan en este plan de tesis son:

I. Desarrollar métodos para la expresión heteróloga y purificación a escala de laboratorio de cuatro dominios SEA, que cubren un amplio rango de diversidad secuencial. Específicamente, estos dominios son: el perteneciente a la mucina humana MUC1 (M1SEA), el encontrado en el homólogo murino de la mucina MUC16 (M16SEA), y los dominios pertenecientes a los receptores ICA512 y fgrina (MPE ICA512 y MPE fgrina).

II. La realización de estudios biofísicos de los dominios enumerados en I, para determinar sus propiedades estructurales y relacionarlas con las similitudes y diferencias de sus secuencias y estructuras.

III. La determinación de la estructura tridimensional del dominio MPE fgrina.

IV. Llevar a cabo un estudio de la estabilidad termodinámica de los dominios mencionados, y relacionar sus mecanismos de desplegado con las particularidades de sus secuencias y estructuras tridimensionales.

V. Realizar un análisis bioinformático de la familia completa de dominios SEA, para integrar toda la información mencionada en los párrafos precedentes.

2. Propiedades estructurales de los dominios SEA

2.1. Introducción

El objetivo fundamental de la biología estructural es explicar la función de las macromoléculas biológicas a partir de su estructura. El mecanismo de funcionamiento de muchas familias proteicas se ha esclarecido, al menos en parte, mediante el estudio comparativo de sus miembros. La familia de los dominios SEA es muy numerosa y forma parte de proteínas muy diversas, asociadas con procesos fisiológicos y patológicos de gran importancia. Sin embargo, la función exacta de estos dominios se desconoce, aunque un gran número de estudios experimentales sugieren que están especializados en mediar interacciones entre células y con la matriz extracelular. Como se ha mencionado en la Introducción, se ha demostrado que algunos dominios SEA poseen la capacidad de formar dímeros y se cree que este tipo de interacción es necesaria en ciertos mecanismos de señalización. También se ha propuesto que la disociación de las dos cadenas, en los dominios que sufren autoproteólisis, conlleva un rol biológico semejante en la señalización celular (Wreschner, McGuckin *et al.* 2002).

En primer lugar, abordamos el estudio de los procesos de interacción intermolecular de los dominios SEA, buscando relacionar su estado de agregación con las características de su secuencia y estructura tridimensional. Para este fin, fue necesario optimizar las metodologías para expresar y purificar estos dominios. Se requerían protocolos robustos que garantizaran un muy alto nivel de pureza y rendimiento, ya que las proteínas estudiadas carecen de residuos de triptófano, el cromóforo intrínseco con mayor señal en la espectroscopía de absorción y emisión de luz.

En segundo lugar, determinamos las propiedades estructurales básicas de los dominios en solución, con el objeto de definir las características del estado nativo. En este punto, estudiamos su grado de compactación y rigidez, y también su estabilidad termodinámica.

En la Figura 2.1 se muestra la superposición de las estructuras de los cuatro dominios SEA estudiados en esta tesis, y en la Figura 2.2, se muestra un alineamiento de sus secuencias. De ambas figuras, resulta evidente que los dominios estudiados presentan alta similitud estructural en un contexto de gran diversidad secuencial.

2.2. Estructura primaria y definición de los dominios

Uno de los primeros pasos en el estudio estructural de un dominio aislado que proviene de una proteína más grande, consiste en definir los límites del dominio. Este problema no es trivial, ya que una elección incorrecta de los límites puede llevar a que se obtenga una proteína incapaz de plegarse eficientemente de manera estable. Una de las herramientas más útiles para definir las fronteras de un dominio, es el empleo de un alineamiento múltiple de secuencias diversas de los miembros de la familia. De esta forma, se define al dominio como la región más conservada dentro de cierto umbral de similitud secuencial. Usando este criterio, se habían definido exitosamente las secuencias que comprenden los dominios SEA de las mucinas MUC1 y Muc16, lo que permitió determinar sus estructuras tridimensionales mediante espectroscopía de RMN (Maeda, Inoue *et al.* 2004; Macao, Johansson *et al.* 2006). En este sentido, valiéndonos de la información estructural existente, en este trabajo únicamente fue necesario subclonar ambos dominios, eliminando las regiones adicionales presentes en las construcciones de origen, tales como secuencias de hexahistidinas empleadas en la purificación y otros artificios de clonado.

Por otra parte, en el caso de MPE ICA512, la identificación de la unidad de plegado autónomo estuvo dictada inicialmente por el resultado de predicciones bioinformáticas de dominios globulares a partir de su secuencia, y también por el conocimiento del procesamiento postraduccional de la proteína, ya que la secuencia de la región extracelular, que incluye este dominio, no guardaba similitud suficiente con ninguna otra proteína de estructura tridimensional conocida. Entonces, se eligió inicialmente para el estudio estructural a la región comprendida entre el sitio de corte de la enzima prohormona convertasa (448-449) y el comienzo del segmento transmembrana (que inicia en 576). Durante los ensayos de cristalización, se encontró que este segmento sufre un procesamiento proteolítico mediado por radicales libres que lleva a la remoción de unos 20 residuos de cada extremo, generando el fragmento MPE ICA512, que cristalizó y cuya estructura tridimensional se determinó por difracción de rayos x (Primo, Klinke *et al.* 2008). Los estudios en esta tesis se hicieron subclonando el dominio MPE ICA512, que comprende los residuos 470-558.

El dominio MPE fgrina se delimitó comparando la secuencia de fgrina con la de ICA512, y eligiendo como extremo N a la posición 502, que corresponde al sitio de corte proteolítico que genera la isoforma predominante en páncreas de ratón (Tomiko Kawakami 2007). El extremo C se fijó en una posición estructuralmente equivalente a la empleada en ICA512.

2.3. Expresión heteróloga de las proteínas recombinantes

La expresión en *E. coli* de los dominios aislados se ensayó inicialmente en condiciones estándar, en medio LB a 37 °C, y empleando IPTG para inducir la expresión por 3-4 h. Este procedimiento fue suficiente para un buen nivel de expresión de MPE ICA512, M1SEA y M16SEA, mientras que fue necesario ajustar las condiciones de expresión de MPE fogrina, variando la composición del medio de cultivo, la temperatura y el tiempo de inducción. Luego de una exploración sistemática de estas condiciones, se optó por expresar MPE fogrina por inducción en medio TB durante 20 h a 20 °C, lo que resultó en un rendimiento de proteína apropiado para los estudios biofísicos. Todas las proteínas se encontraron exclusivamente en la fracción soluble del lisado celular. Se empleó lactosa 1 % para la producción en mayor escala, ya que dio los mismos resultados que IPTG (no mostrado).

Todas las proteínas se purificaron mediante procedimientos cromatográficos convencionales, luego de lo cual su identidad se verificó por espectrometría de masa, los resultados de estas determinaciones se muestran en la Tabla 2.1.

Tabla 2.1. Propiedades de expresión de los dominios SEA.

Parámetro	MPE ICA512	MPE fogrina	M1SEA		M16SEA
			N-terminal	C-terminal	
PM calculado (Da)	9760.1	10813.1	6702.5	6107.7	13805.3
PM determinado (Da)	9759.5	10811.4	6702.3	6112	13805.4
Nivel de expresión (mg/l)	70	20	50		60

El peso molecular (PM) calculado se obtuvo a partir de la secuencia de aminoácidos, mientras que el PM determinado se obtuvo por análisis de masa por ESI o MALDI-TOF. El nivel de expresión está indicado como mg de proteína por litro de cultivo.

Un resultado notable se obtuvo con la proteína M16SEA. Al final de la purificación, esta proteína se observa como una sola banda en SDS-PAGE, y su peso molecular determinado mediante espectrometría de masas corresponde al esperado de acuerdo a su composición. Sin embargo, en su espectro de emisión de fluorescencia se observa, aparte de la emisión de tirosina, una banda a mayor longitud de onda, algo inesperado, en principio, para un dominio que no contiene triptófano (Figura 2.3).

Para investigar si la emisión anómala de M16SEA podía deberse a algún tipo de modificación química de la proteína o si era debida a la presencia de un contaminante, se realizó una separación adicional sobre una muestra purificada mediante el protocolo convencional, empleando cromatografía de HPLC de fase reversa. En este procedimiento, se observaron dos picos apenas resueltos (no mostrado), que se colectaron y analizaron por

espectrometría de masa. Estos análisis detectaron una especie con el peso molecular esperado para este dominio y una señal menor, que posee 490 Da adicionales. Luego de un tratamiento de proteólisis exhaustiva con tripsina y secuenciación parcial mediante MALDI TOF/TOF, no se encontraron diferencias significativas entre las dos especies. Paralelamente, se hizo un análisis por secuenciación N-terminal de una muestra de digestión parcial con tripsina, y sólo se obtuvieron las secuencias esperables para el corte en distintas posiciones a lo largo de M16SEA. Por todo esto, se razonó que la segunda especie podría tratarse de la misma proteína M16SEA, con algún tipo de modificación en su extremo C.

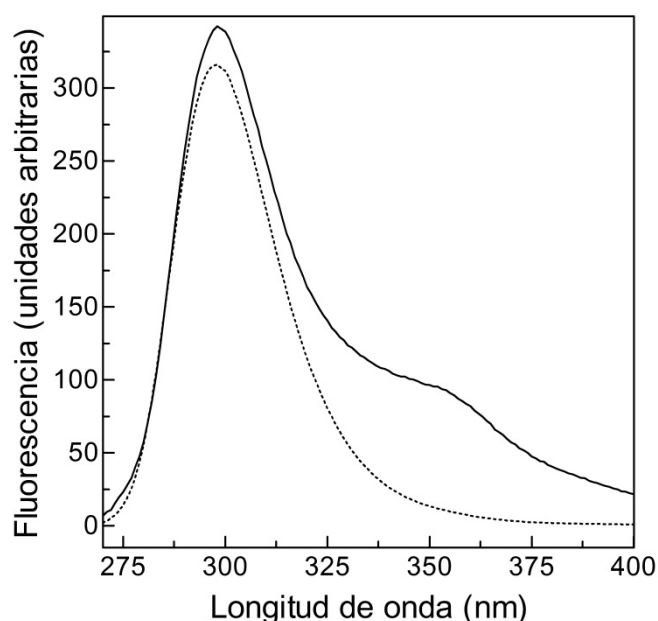


Figura 2.3. Espectros de fluorescencia de M16SEA. Se muestra en línea llena el espectro de emisión de M16SEA, luego de excitar a 276 nm. En línea de puntos, se muestra el espectro de la variante M16STOP, tomado en las mismas condiciones. El *buffer* empleado fue fosfato de sodio 25 mM pH 7.0 y la temperatura 20 °C.

Entonces se investigó cuál podría ser el efecto si la proteína se extendiera en su extremo C, como resultado de que el codón de *stop* no se reconociera como tal. Un cambio en la última base de la secuencia de este codón (TGG por TGA), conduce a la adición de la secuencia WGSGC hasta el siguiente codón de *stop*. Esta secuencia corresponde a un incremento de masa de 490.5 Da, lo que coincide con lo determinado por espectrometría de masa, y además introduce un residuo de triptófano, lo que explica el comportamiento de emisión de fluorescencia. Además, esto también añade un residuo de cisteína, el cual se pudo titular con el reactivo DTNB (no mostrado). Finalmente, esto también es consistente con

la dificultad encontrada en la separación de ambas especies mediante los procedimientos cromatográficos empleados para purificar este dominio.

La lectura incorrecta de un codón de *stop* TGA como codificante para triptófano (TGG) es un fenómeno que se ha reportado en otras oportunidades y recibe el nombre de "*codon readthrough*" (Lu, Rohde *et al.* 1995; MacBeath y Kast 1998). Este fenómeno se manifiesta, durante la expresión recombinante, por la aparición de una banda de inducción con un tamaño mayor que acompaña a la proteína de longitud correcta. En nuestro caso, la escasa diferencia de masa y la baja proporción de este sub-producto dificultaron su identificación, y sólo se hizo evidente a partir de la anomalía del espectro de fluorescencia. Notablemente, el empleo del mismo codón de *stop* en las construcciones de ICA512 y fognina no tuvo este efecto adverso, pero se reconoce que la tasa de error de lectura está influenciada por la secuencia de las bases que flanquean al codón de *stop*, y también por la identidad de los aminoácidos de la cadena polipeptídica naciente (Bjornsson, Mottagui-Tabar *et al.* 1996; Mottagui-Tabar y Isaksson 1997).

Como una prueba más directa de un mecanismo de "*codon readthrough*" en M16SEA, se clonó una variante donde el codón de *stop* se modificó de TGA ("*opal*") a TAA ("*ochre*"), formando parte del tetranucleótido TAAT que es la secuencia más favorecida como *stop* en *E. coli* (Poole, Brown *et al.* 1995). Cuando esta variante, llamada M16STOP, se estudió por espectroscopía de fluorescencia en las mismas condiciones (Figura 2.3), se observó la desaparición de la banda de emisión de ~350 nm, lo que confirma que la emisión anómala es debida a la incorporación errónea de triptófano en el extremo C de la proteína.

Para completar el estudio de la estructura primaria, se investigó la presencia de enlaces disulfuro en el dominio M16SEA, el único que contiene cisteínas. En la estructura tridimensional de este dominio, se observa que sus dos únicos residuos de cisteína se encuentran espacialmente cercanos, a una distancia de alrededor de 4 Å en cordones beta antiparalelos adyacentes (Figura 2.4). Esta cercanía y geometría llevó a considerar la posible formación de un puente disulfuro.

Cuando la proteína en el estado nativo se incubó con el reactivo DTNB, no se detectaron tioles libres, sin embargo, al inducir el desplegado por el agregado de cloruro de guanidinio, se lograron titular las dos cisteínas. Por lo tanto, en las condiciones de expresión recombinante, las dos cisteínas se encuentran reducidas y ocultas al solvente en el estado nativo.

Para probar si las uniones disulfuro podrían formarse durante el replegado *in vitro*, se desplegó la proteína con cloruro de guanidinio y luego se replegó por dilución en presencia de un par rédox (2-mercaptoetanol 1.15 mM, ácido 3,3'-ditiopropiónico 0.13 mM). Cuando esta muestra se ensayó con el reactivo DTNB, luego de pasarla por una columna de exclusión molecular (Sephadex G-25), la cantidad de tioles libres fue la misma que con la proteína sin

tratar, es decir, las cisteínas se encuentran igualmente reducidas y sólo accesibles al DTNB en el estado desplegado.

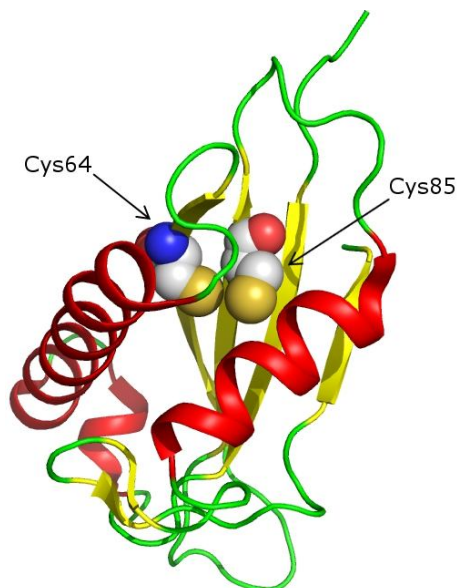


Figura 2.4. Residuos de cisteína en M16SEA. Los dos residuos de cisteína se muestran como esferas, con la numeración correspondiente al archivo PDB (1IVZ).

Un examen detallado de la estructura resuelta por RMN indicó que aunque la separación entre los átomos de azufre es demasiado grande para la formación de un puente disulfuro (4 vs 2.3 Å), una topología ideal de los cordones beta antiparalelos permitiría reducir la distancia, pero en este caso los cordones están distorsionados por el conjunto de interacciones entre átomos en el estado nativo. En particular, debido a la distorsión, no pueden alcanzarse en el estado nativo los rotámeros adecuados de las cisteínas que permitirían satisfacer los requerimientos de la formación de un puente disulfuro (Bhattacharyya, Pal *et al.* 2004). Por otra parte, el examen de la estructura también indicó que las cisteínas de referencia se hallan en un entorno altamente hidrofóbico y son muy poco accesibles al solvente, justificándose su baja reactividad frente al DTNB en condiciones nativas.

2.4. Estado de agregación

La cromatografía de exclusión molecular es una técnica estándar para la determinación de las propiedades hidrodinámicas de las proteínas. Una columna de exclusión molecular separa las proteínas de acuerdo a su coeficiente friccional, aunque esta propiedad suele expresarse de una manera más intuitiva definiendo el radio de Stokes (R_s), que es el radio de una esfera con el mismo coeficiente friccional que la proteína, lo que constituye una aproximación al volumen real que ocupa una macromolécula en solución. En estos estudios

se empleó una columna “Superose 12”, cuya utilidad para la determinación de las propiedades hidrodinámicas en el estado nativo y en estados no nativos de las proteínas está ampliamente documentada (Uversky 1993). La columna se calibró con un conjunto de proteínas globulares de tamaño molecular conocido. Los experimentos de exclusión molecular se realizaron a distintas concentraciones de proteína para estudiar si existe una dependencia del tiempo de retención con la concentración, lo que es indicativo de que el estado nativo se encuentra en equilibrio entre distintas especies oligoméricas. Para ello se inyectó cada proteína en un rango de concentraciones desde aproximadamente 5 μM hasta 200 μM . En la Figura 2.5 se muestran los perfiles cromatográficos y en la Tabla 2.2 se muestran los valores de radio de Stokes correspondientes.

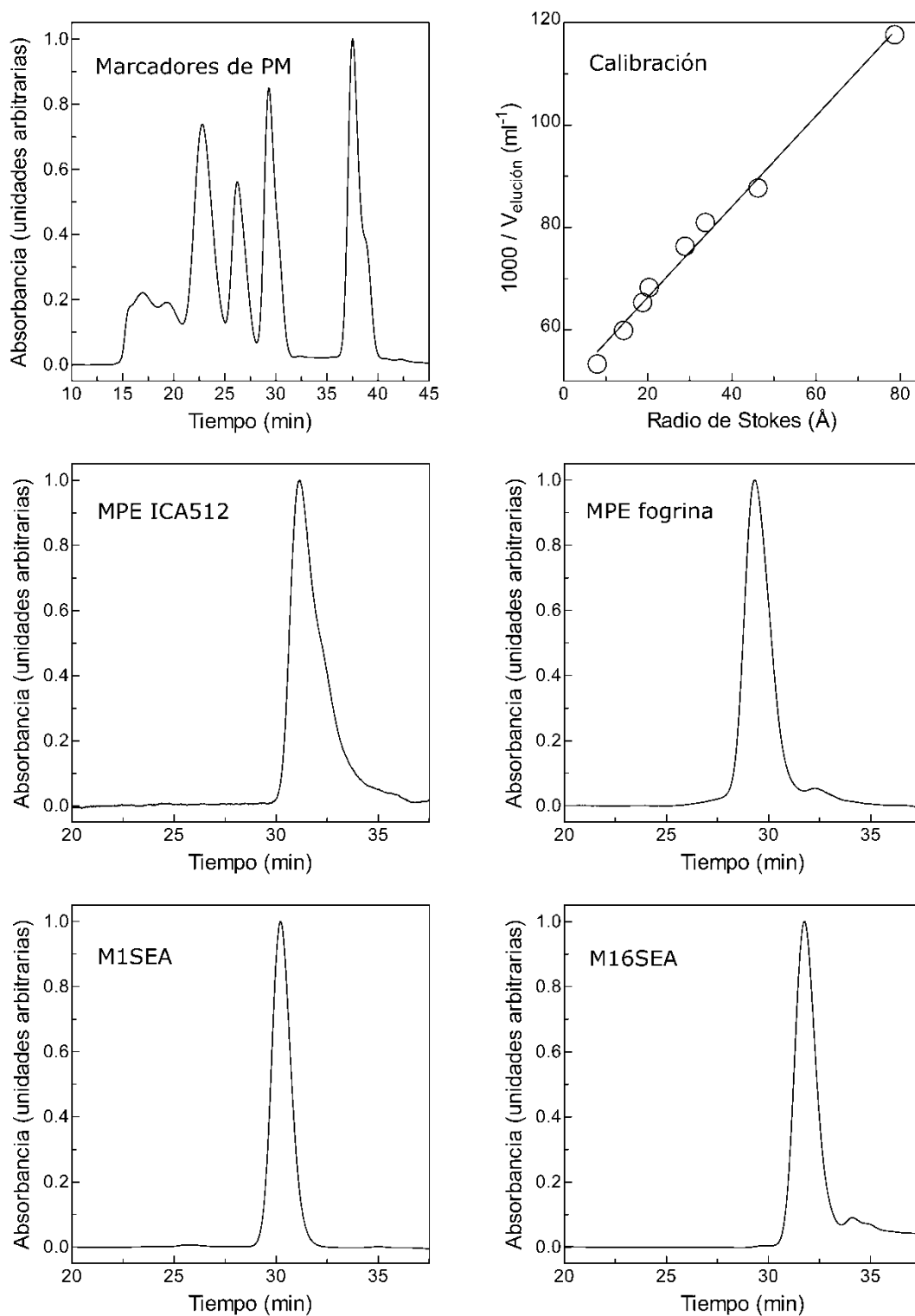


Figura 2.5. Comportamiento hidrodinámico seguido por SEC. Los perfiles cromatográficos se siguieron por absorbancia a 280 nm. Se muestra una corrida representativa de proteínas usadas como marcadores en la calibración, y la curva de calibración correspondiente. Los experimentos se realizaron en *buffer* fosfato de sodio 25 mM, NaCl 150 mM, pH 7.0.

En el caso de la proteína M16SEA, su perfil de elución es anómalo, caracterizado por un tiempo de retención mayor que el esperado para un monómero de su peso molecular. Esto posiblemente se deba a que la proteína posee algún grado de interacción con la matriz cromatográfica (compuesta de agarosa entrecruzada químicamente). Resultados similares se obtuvieron empleando una columna Superdex 75 (no mostrado), cuya matriz está compuesta de una mezcla de agarosa y dextrano.

Tabla 2.2. Comportamiento hidrodinámico seguido por SEC a pH 7.0.

Parámetro	MPE ICA512	MPE fogrina	M1SEA	M16SEA
Peso molecular (kDa)	9760.1	10813.1	12810.2	13805.3
Nº de residuos	90	99	112	121
R_s teórico ^a (Å)	17.5 (21.4)	18.0 (22.0)	18.7	19.1
R_s esperado ^b SEC (Å)	16.5 (21.3)	17.2 (22.2)	18.3 (23.6)	18.8
R_s experimental ^c (Å)	16.4 ± 0.2 (20.7 ± 0.2)	22.5 ± 0.3	20.6 ± 0.1	16.6 ± 0.3

^aCalculado a partir del número de residuos de acuerdo a la ecuación descrita por Dobson (Wilkins, Grimshaw *et al.* 1999). ^bCalculado a partir del peso molecular con la ecuación descrita por Uversky (Uversky 1993). ^cDeterminado a partir del tiempo de retención, por interpolación de la curva de calibración. Los números entre paréntesis representan los R_s teóricos o esperados para el dímero, según corresponda. En el caso de MPE ICA512, el R_s experimental depende de la concentración proteica y el extremo superior corresponde a altas concentraciones (predominio de la forma dimérica).

En el caso del dominio M1SEA, el perfil cromatográfico fue idéntico en todas las concentraciones ensayadas, con un tiempo de retención cercano al esperado para una proteína monomérica de su tamaño, con una forma globular algo expandida. Dado que este dominio está formado por dos cadenas no unidas covalentemente, podría esperarse que su disociación generara una dependencia en el tiempo de retención con la concentración de proteína inyectada (en caso de tratarse de un equilibrio de disociación rápido). Sin embargo, como se verá en la Sección 2.8.2, la conformación heterodimérica de este dominio es muy estable, con una constante de disociación de aproximadamente 50 fM, lo que explica que no se detecte disociación en el rango de concentraciones ensayado.

La proteína MPE fogrina exhibió también un perfil de elución invariable con la concentración, y coincidente con lo esperado para un dímero, lo que sugiere que forma un dímero de alta afinidad. Sin embargo, un examen usando otras técnicas (ver más abajo) demostró que el comportamiento de MPE fogrina en SEC es un artefacto y la proteína en solución es monomérica.

Respecto a la proteína MPE ICA512, hemos caracterizado extensivamente en nuestro laboratorio, la presencia de un equilibrio dímero-monómero en el estado nativo, que se

manifiesta en la dependencia del tiempo de retención con la concentración de proteína inyectada (Primo, Klinke *et al.* 2008). El análisis de los perfiles de elución según un modelo de equilibrio dímero-monómero, permitió estimar una constante de disociación de 0.8 μ M.

Un enfoque complementario para el estudio del estado de agregación, se obtuvo mediante experimentos de entrecruzamiento químico con glutaraldehído y DSS, visualizando los resultados por SDS-PAGE. En el caso del dominio MPE ICA512, los dos reactivos revelaron la presencia de dímeros. Al emplear DSS, se observaron solamente dímero y monómero (Figura 2.6), mientras que la reacción con glutaraldehído también dio cantidades menores de otros oligómeros de mayor tamaño (no mostrado). No se observó dependencia en la proporción de dímero respecto del tiempo de incubación con DSS (entre 5 y 15 min), y cuando se ensayaron distintas concentraciones de DSS, el único producto fue el dímero (empleando una concentración molar en exceso de entrecruzante de 2 a 20 veces la de proteína). Notablemente, la banda correspondiente al dímero persiste aún a altas concentraciones de GdmCl (Figura 2.6) o de urea (no mostrado).

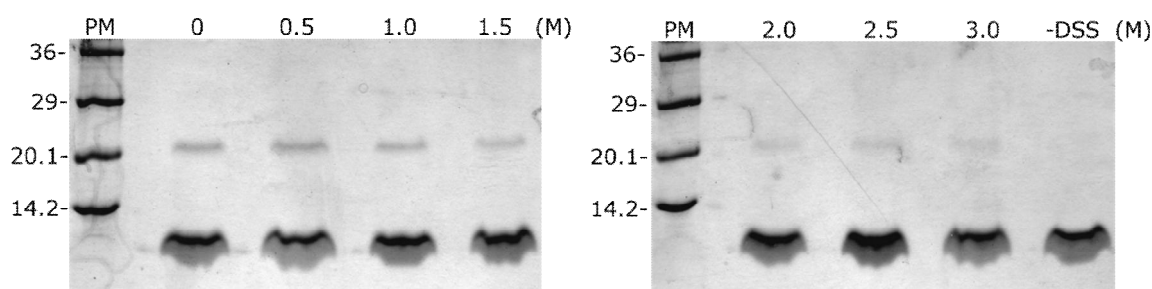


Figura 2.6. Entrecruzamiento con DSS de MPE ICA512. Las muestras se preincubaron a 25 °C con GdmCl en la concentración indicada, en *buffer* fosfato de sodio 25 mM pH 7.0. Luego de 3 h, se añadió DSS 300 μ M y se incubó a 27 °C por 15 min, frenando la reacción con Tris-Cl 1 M pH 6.8. Luego las muestras se dializaron y se precipitaron con TCA. A la muestra señalada como -DSS, no se le añadió DSS.

Con el dominio M16SEA se obtuvo una tenue banda de dímero cuando fue incubado con DSS (no mostrado), y en la incubación con glutaraldehído (Figura 2.7) se obtuvieron dímeros y especies de mayor tamaño, por lo que es probable que la reacción fuera inespecífica. En el caso de MPE fgrina, no se observa reacción con DSS y la reacción con glutaraldehído dio lugar a la aparición de especies de muy alto peso molecular, que no atraviesan el gel de corrida (no mostrado), con un patrón compatible con la colisión azarosa de monómeros.

La reacción de M1SEA con glutaraldehído o DSS genera un patrón de difícil interpretación, ya que hay que considerar que este dominio es ya un heterodímero en condiciones nativas (no mostrado).

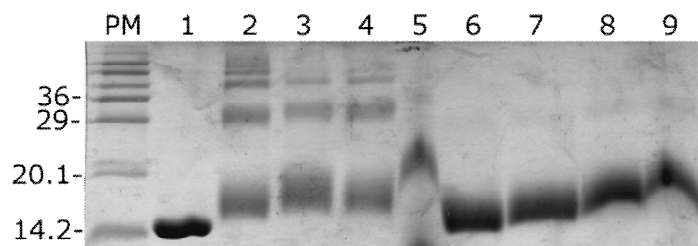


Figura 2.7. Entrecruzamiento con glutaraldehído de M16SEA. Las reacciones se hicieron en *buffer* fosfato de sodio 25 mM pH 7.0, a 27 °C. La proteína se incubó con glutaraldehído en distintas proporciones, que se expresan como múltiplos de la empleada en la calle 2 (1X). Calle 1: sin glutaraldehído; calles 2 a la 5: 1X, 2X, 4X, 8X, 16X, respectivamente. Las calles 6 a 9 corresponden a muestras que fueron preincubadas con SDS 8 % p/v y calentadas 5 min a 100 °C antes de agregar el glutaraldehído en las proporciones 2X, 4X, 8X y 16X, respectivamente.

Finalmente, el estado de agregación de los dominios se estudió por dispersión de luz estática. Las técnicas de dispersión de luz son metodologías espectroscópicas que permiten la determinación de los pesos moleculares de macromoléculas en solución. Una de estas metodologías es la dispersión de luz estática (SLS), que mide el promedio de la magnitud de la luz dispersada que, bajo ciertas condiciones, es proporcional al producto del peso molecular de la macromolécula por su concentración (Harding y Jumel 2001). En este trabajo empleamos la técnica *multi-angle laser light scattering*, acoplado a un fraccionamiento previo por exclusión molecular mediante una columna Superdex 200 (SEC-MALLS). A diferencia de la metodología convencional de exclusión molecular, la determinación del peso molecular por dispersión de luz no requiere del empleo de patrones y no se ve afectada por la forma de la macromolécula ni por artefactos de interacción inespecífica con la matriz de exclusión molecular.

Los resultados fueron esclarecedores en el caso de MPE fognina. A diferencia de lo estimado por SEC, el peso molecular de este dominio fue de 10.8 kDa calculado a partir de los datos de dispersión de la luz, lo que es consistente con un estado monomérico (Figura 2.8A). Además, el tiempo de retención en la matriz de exclusión usada en el experimento SEC-MALLS (Superdex 200), diferente de la empleada en el procedimiento de SEC convencional (Superose 12), fue compatible con un estado monomérico, lo que sugiere un artefacto en la interacción entre la matriz Superose 12 y MPE fognina. La naturaleza de la interacción de MPE fognina con la matriz de exclusión Superose 12 no se investigó adicionalmente, aunque parece estar relacionada con la carga superficial altamente negativa de la proteína.

En el caso de MPE ICA512, como parte de la caracterización del equilibrio dímero-monómero, en trabajos previos se han realizado en nuestro laboratorio mediciones de SEC-MALLS en las mismas condiciones que las descriptas aquí para fognina (no mostrado), que

muestran la presencia de masas moleculares en el rango del dímero (aproximadamente 20 kDa) hasta al monómero (10 kDa), dependiendo del tiempo de retención en la columna.

Como se mencionó en los párrafos precedentes, el estado de oligomerización de M16SEA no pudo obtenerse por mediciones convencionales de SEC, debido a su interacción inespecífica con la matriz, ni por entrecruzamiento químico, ya que la reacción fue inespecífica. Entonces se realizó la medición de masa molecular mediante SEC-MALLS, de la misma forma que con MPE fgrina. Aunque los resultados del experimento tienen una dispersión algo elevada (Figura 2.8B), son indicativos de un estado monomérico, con un peso molecular de 13.7 kDa.

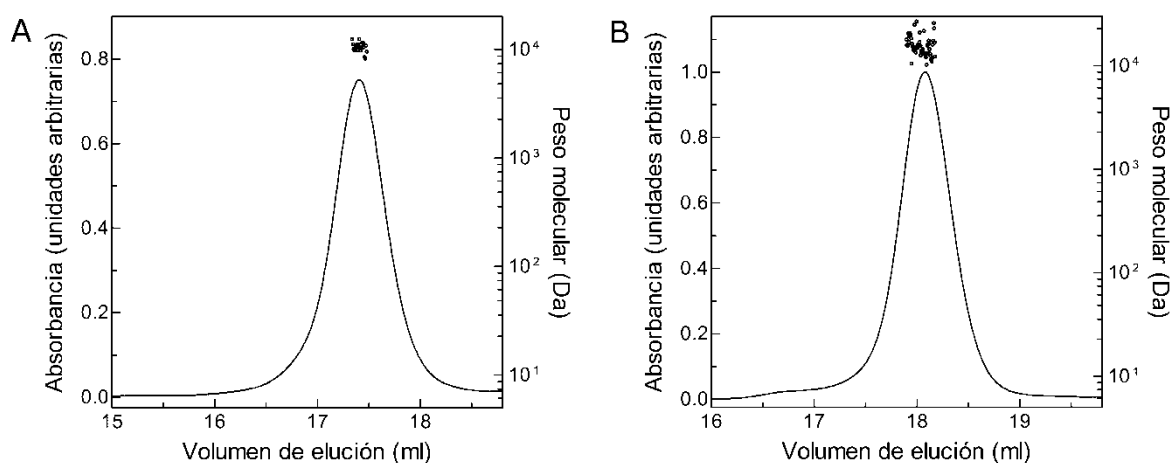


Figura 2.8. Peso molecular determinado por SEC-MALLS. Se muestra como línea continua el perfil cromatográfico seguido por absorbancia a 280 nm, y los puntos corresponden al peso molecular. **(A)** MPE fgrina. **(B)** M16SEA. Nótese que en los experimentos de SEC-MALLS se utilizó una columna cromatográfica empaquetada con la matriz Superdex 200, diferente a la empleada en los experimentos de SEC descritos en la Figura 2.5. Por lo tanto los perfiles cromatográficos no pueden ser comparados directamente. El *buffer* empleado fue el mismo que en los experimentos de SEC.

2.5. Flexibilidad conformacional explorada por proteólisis controlada

La flexibilidad conformacional de los dominios SEA se estudió por medio de experimentos de digestión controlada con tripsina. Para que una proteína pueda ser degradada por una proteasa debe existir, en la región de corte, una flexibilidad que le permita adaptarse al sitio activo de la enzima, lo que puede ser resultado de un desplegado local, en la cercanía del enlace peptídico blanco, o mediante un desplegado global, que involucre a toda la molécula. El método de digestión controlada con proteasas se ha utilizado para aislar

dominios, y también para identificar unidades autónomas de plegado dentro de los mismos dominios.

Cuando los dominios SEA se trataron con tripsina, se observaron distintos comportamientos (Figura 2.9). En el caso de MPE fognina, se observó que este dominio es marcadamente resistente a la proteólisis. La degradación ocurre lentamente y se manifiesta como una disminución en la intensidad de la banda de proteína de longitud completa, sin detección de fragmentos mayores. Se concluye de esta forma, que los fragmentos generados no son resistentes a la proteólisis ulterior, y que la proteína es susceptible a la hidrólisis por tripsina mediante un proceso de desplegamiento global.

El dominio MPE ICA512 exhibió una cinética de proteólisis aún más lenta que la de fognina (no mostrado). Aún después de 4 h de reacción no se observó degradación, sin embargo, con el agregado de urea en una concentración final de 2 M, se observa casi el 100 % de degradación en el mismo período. Eso es indicativo de una estructura esencialmente rígida, con potenciales sitios de corte que sólo se hacen accesibles luego de inducir el desplegamiento con urea.

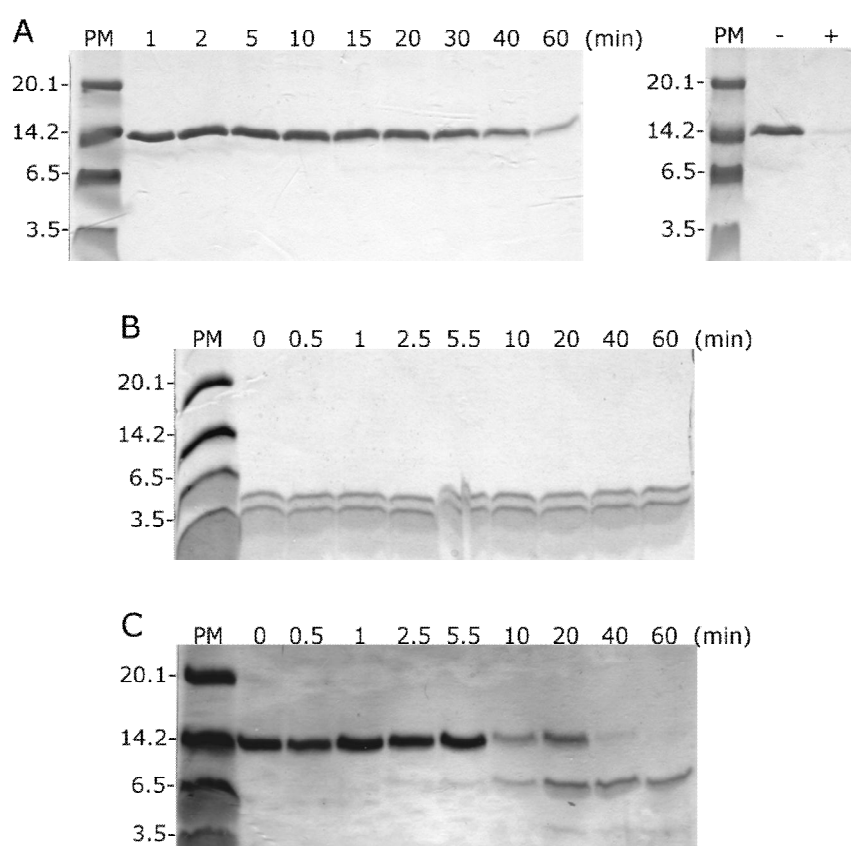


Figura 2.9. Proteólisis controlada con tripsina. (A) MPE fognina, la relación tripsina:proteína fue de 1:75, y también se muestran, en el panel de la derecha, muestras tomadas luego de 2 h de

reacción, con (+) y sin (-) tripsina. **(B)** M1SEA y **(C)**, M16SEA, ambas en una relación tripsina:proteína de 1:100 p/p.

El dominio M1SEA, el cual está constituido por dos cadenas, también se mostró resistente a la acción de tripsina (Figura 2.9B), no detectándose degradación en el tiempo ensayado. Esto es consistente con una estructura compacta y estable.

El comportamiento más interesante se observó con el dominio M16SEA (Figura 2.9C). Esta molécula es rápidamente degradada, pero los fragmentos generados son muy resistentes a la degradación posterior. El análisis del producto de digestión parcial por espectrometría de masa y secuenciación N-terminal, permitió asignar el sitio de corte entre las argininas 85 y 86, dentro de un *loop* altamente expuesto al solvente (Figura 2.10). Si se permite que la digestión continúe durante más tiempo, al final prácticamente todos los sitios de corte son blanco de la acción de tripsina, y los fragmentos generados pudieron identificarse por secuenciación parcial por espectrometría de masa MALDI TOF/TOF.

Segmento	Secuencia	Peso molecular (Da)	
	N-terminal	Esperado	Obtenido
S¹- R⁸⁵	SSSSQHFNL	9594.5	9595.1
R⁸⁶- G¹²¹	RVDRVAIYE	4228.8	4229.4

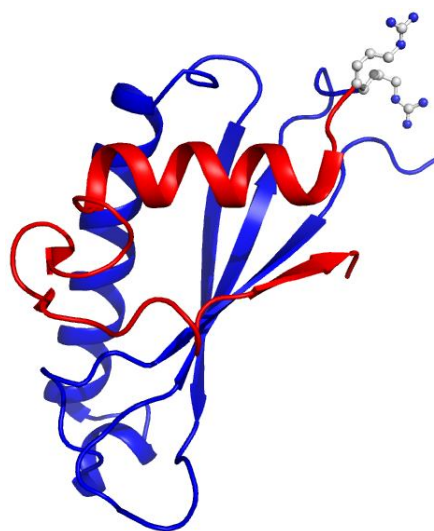


Figura 2.10. Fragmentos de proteólisis controlada del dominio M16SEA. Se indican los resultados del análisis por espectrometría de masas y secuenciación N-terminal, que permitieron asignar el enlace peptídico donde ocurre el primer evento de corte. A la derecha, se muestran sobre la estructura tridimensional los residuos de arginina 85 y 86 en representación de esferas, y en colores distintos los dos fragmentos generados.

2.6. Estructura secundaria

El dicroísmo circular (CD) es un fenómeno que resulta de la absorción diferencial de luz circularmente polarizada a la izquierda y derecha, por los cromóforos que son intrínsecamente quirales o que se hacen ópticamente activos al encontrarse en entornos asimétricos. Cuando las proteínas son estudiadas por CD con luz en la región del UV lejano (180-240 nm), el cromóforo que absorbe es el enlace peptídico, aunque también puede haber contribuciones de menor magnitud por parte de los residuos aromáticos. Las propiedades de absorción del enlace peptídico son modificadas al formar parte de arreglos de estructura regular, lo que resulta en espectros de CD característicos para los distintos elementos de estructura secundaria.

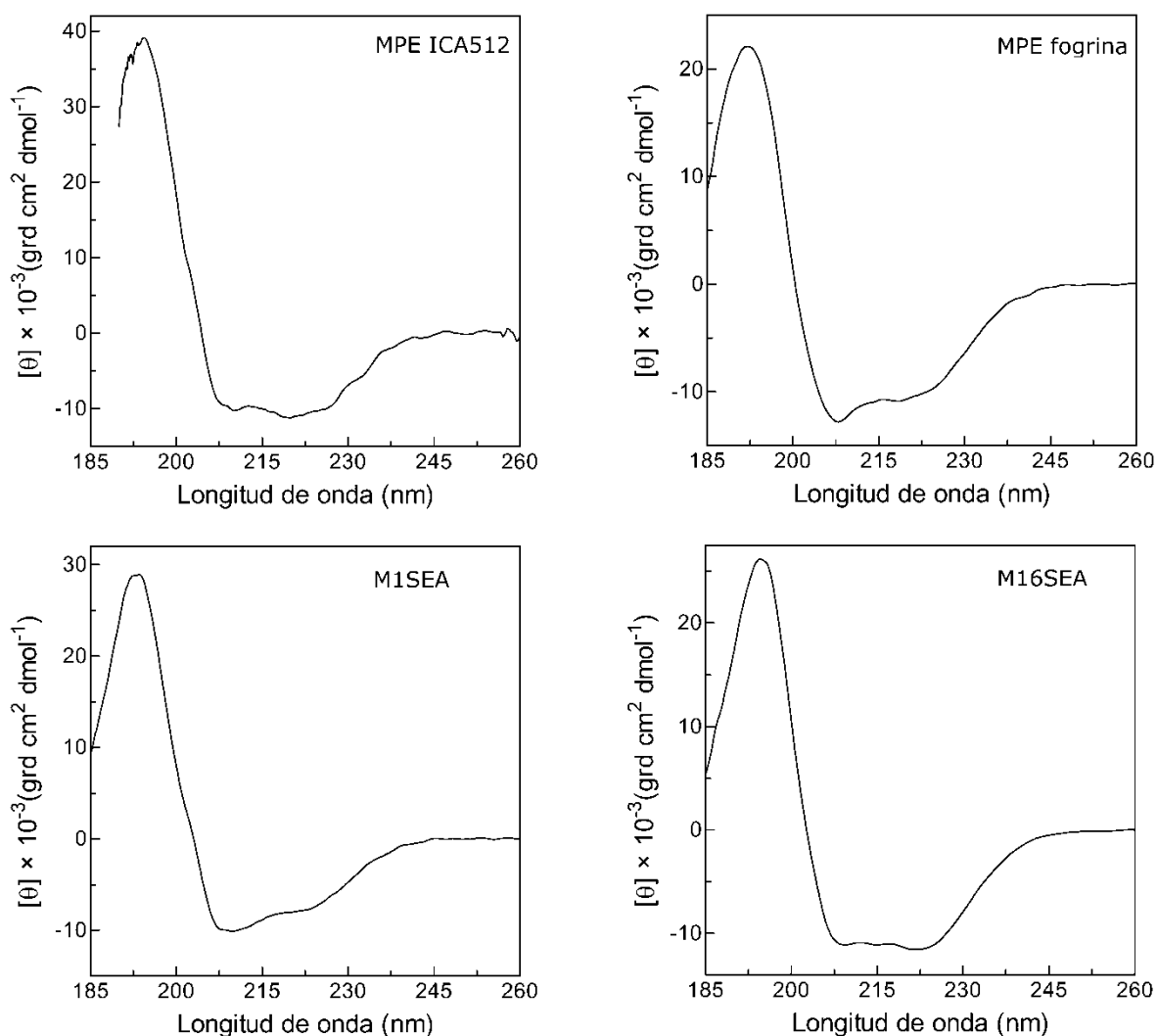


Figura 2.11. Espectros de CD en el UV lejano. Las medidas se hicieron a 20 °C en *buffer* fosfato de sodio 25 mM pH 7.0.

En la Figura 2.11 se muestran los espectros de CD en el UV lejano de los cuatro dominios SEA. En todos los espectros se observa una contribución mayoritaria de hélices α , observándose dos mínimos, a 208 y 220 nm, y un máximo, entre 190 y 195 nm. Se realizó una estimación del contenido de estructura secundaria a partir de los espectros de CD, empleando el conjunto de algoritmos de deconvolución de espectros reunido en la suite DICROPROT (Deleage y Geourjon 1993), y los resultados de este análisis están de acuerdo con la estructura tridimensional.

2.7. Estructura terciaria

En la región del UV cercano (250-300 nm), la absorción de los cromóforos está asociada con las transiciones electrónicas $\pi - \pi^*$. Dado que las cadenas laterales de los residuos aromáticos poseen planos de simetría, exhiben dicroísmo circular casi completamente debido a su entorno rígido asimétrico, como el que se encuentra en el interior de una proteína bien plegada.

Como se mencionó en la sección 2.1, ninguno de los dominios SEA estudiados aquí contiene residuos de triptófano, por lo tanto, las señales observadas en los espectros de CD en el UV cercano se atribuyen a las cadenas laterales de tirosina y fenilalanina (Tabla 2.3).

Tabla 2.3. Residuos aromáticos en los dominios SEA.

Residuo	MPE ICA512	MPE fogrina	M1SEA	M16SEA
Tirosina	2	1	5	4
Fenilalanina	2	2	12	9

Los espectros medidos para todos los dominios se muestran en la Figura 2.12. Una característica común a todos ellos es la presencia de estructura fina en longitudes de onda menores a 270 nm. Las características de estas bandas provienen de las transiciones entre los distintos niveles vibracionales del estado excitado de residuos de fenilalanina. Su destacada magnitud en los espectros mostrados, especialmente en MPE fogrina y M1SEA, es indicativa de un entorno apolar y relativamente rígido alrededor de estos residuos.

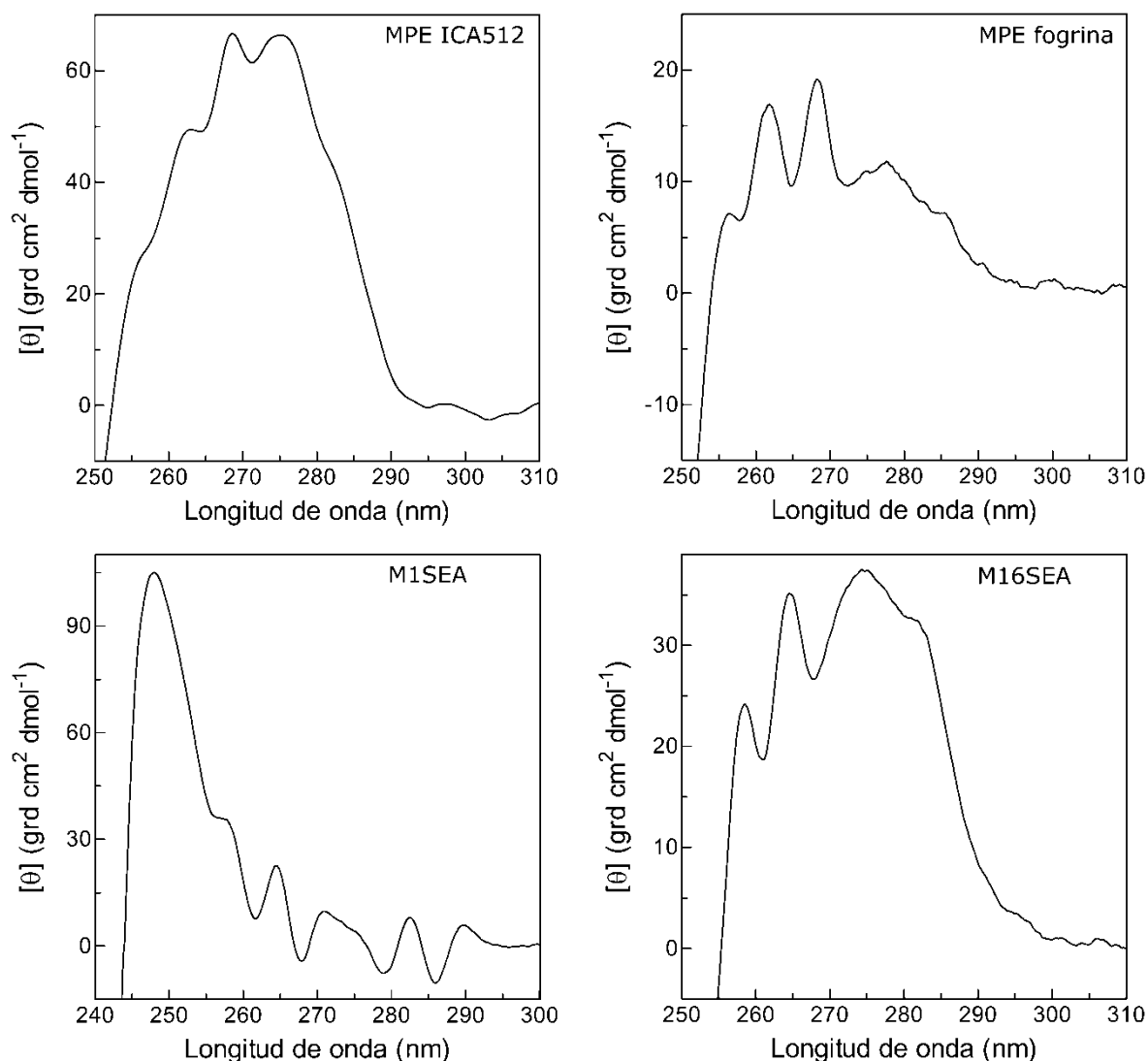


Figura 2.12. Espectros de CD en el UV cercano. Las medidas se hicieron a 20 °C en *buffer* fosfato de sodio 25 mM pH 7.0.

El espectro de la proteína M1SEA exhibe una muy prominente banda centrada en aproximadamente 250 nm. Esta característica inusual de su espectro de CD también ha sido notada previamente en una construcción que incluye el dominio M1SEA y 15 residuos más de su extremo N (lo que no añade ningún residuo aromático) (Johansson, Macao *et al.* 2008). La inspección de la estructura tridimensional de este dominio, permite distinguir un grupo de 4 residuos de fenilalanina y uno de tirosina (Figura 2.13), donde se observan geometrías favorables para la interacción π - π , con anillos aromáticos ubicados “de canto” hacia el plano de otros anillos (*T-shape*). El carácter agudo del pico a 250 nm sugiere un proceso de contribuciones negativas al espectro por parte de fenilalaninas y/o tirosinas que se 'restan' del espectro principal.

Por otra parte, en los espectros de MPE ICA512 y de M16SEA se aprecian claramente las bandas correspondientes a las transiciones 1L_b de las tirosinas que hacen una contribución positiva.

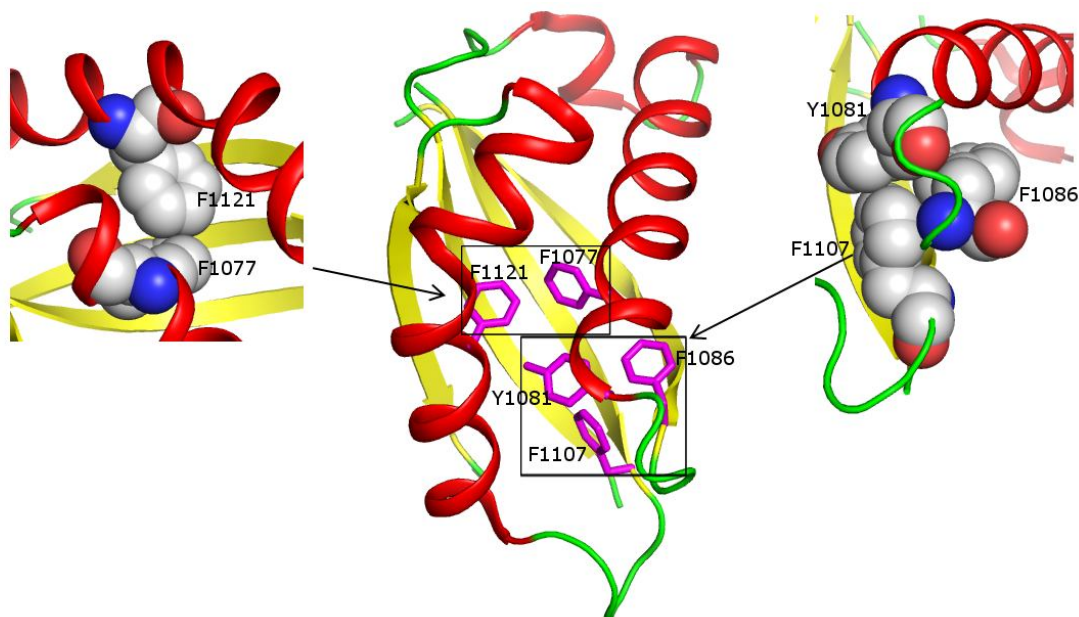


Figura 2.13. Agrupamiento de residuos aromáticos en M1SEA. En la figura central se muestra la ubicación de residuos aromáticos muy cercanos entre sí, y en las figuras laterales se muestra el detalle de las regiones recuadradas, rotadas para una mejor visualización. La numeración corresponde al archivo PDB (2ACM).

La estructura terciaria también se investigó mediante espectroscopía de emisión de fluorescencia, luego de excitar las cadenas laterales de tirosina (Figura 2.14). Los dominios MPE ICA512, M16SEA y MPE fogrina, presentan un máximo de emisión a 300 nm, cuya intensidad disminuye al desplegar la proteína con GdmCl 5-6 M, aunque la longitud de onda del máximo de emisión es invariable, como es en general el caso de proteínas que no contienen triptófano. Esto indica que los residuos de tirosina se encuentran ocultos al solvente y sólo resultan expuestos con el desplegado inducido por el desnaturizante. El dominio M1SEA tiene un espectro similar al de MPE ICA512 y de MPE fogrina, pero con una intensidad de emisión que no disminuye significativamente con el agregado de GdmCl.

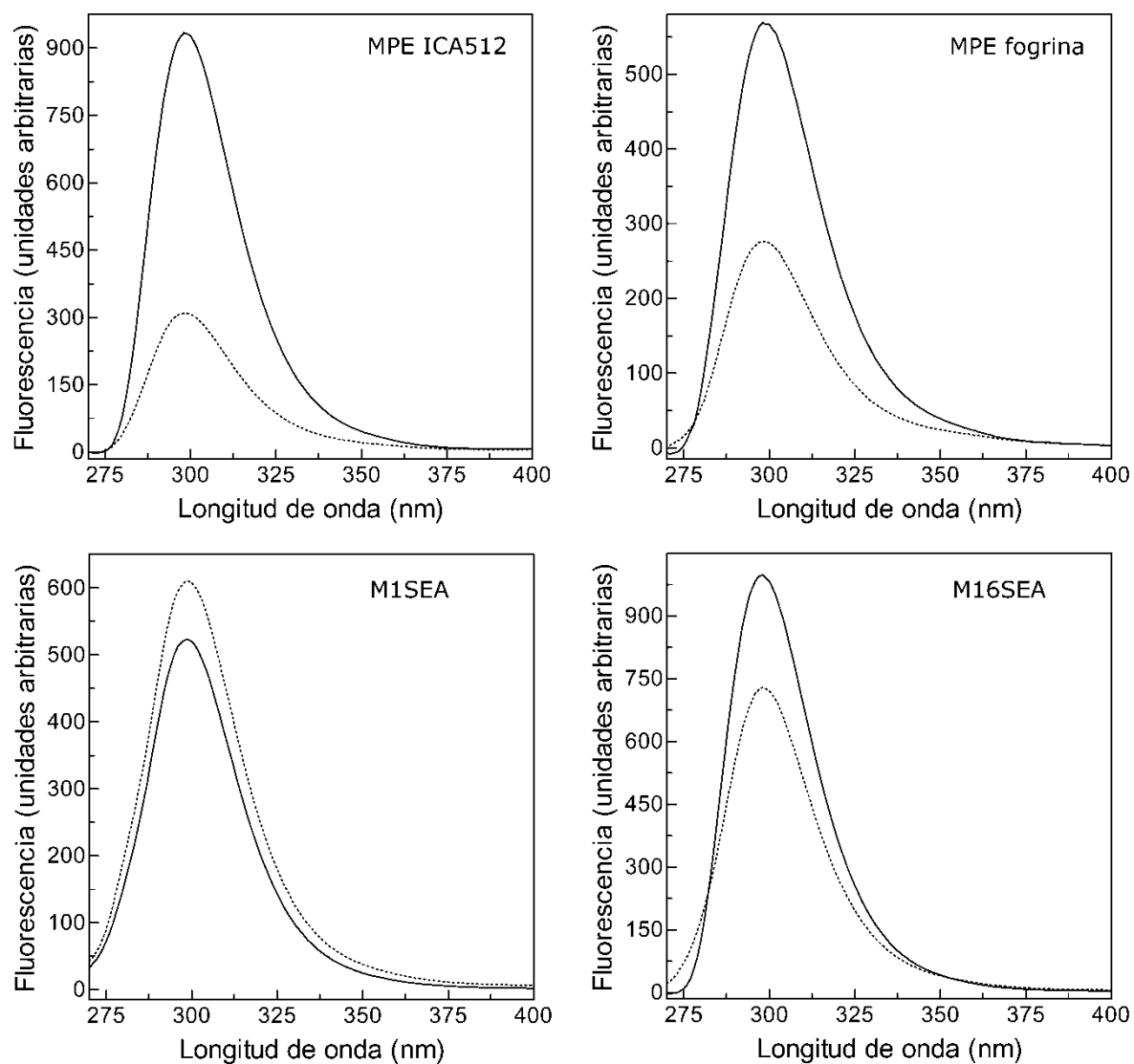


Figura 2.14. Espectros de fluorescencia. Las medidas se hicieron a 20 °C en fosfato de sodio 25 mM pH 7.0, con excitación a 276 nm. En línea punteada se muestran los espectros de la proteína desplegada con GdmCl 6 M. El espectro de M16SEA corresponde al de la variante M16STOP.

Para investigar la presencia de superficies hidrofóbicas expuestas al solvente, se realizaron espectros de emisión de fluorescencia en presencia de ANS (no mostrado). En ninguno de los dominios se detectó unión de esta sonda, lo cual es consistente con un empaquetamiento rígido, y sin superficie hidrofóbica expuesta al solvente.

2.8. Estabilidad conformacional de los dominios SEA

2.8.1. Desplegado por temperatura

La estabilidad de los dominios SEA frente al desplegado por temperatura, se estudió por medición de la señal dicroica a 220 nm, mientras se incrementaba linealmente la temperatura. De esta forma, se siguió el cambio en el contenido de estructura secundaria que acompaña al proceso de desplegado. Los resultados obtenidos se muestran en la Figura 2.15. En todos los casos, se observaron transiciones altamente cooperativas y con una elevada T_m , la temperatura en el punto medio de la transición. Esto indica que todos los dominios tienen una alta estabilidad estructural.

Las transiciones de los dominios M1SEA, M16SEA y MPE ICA512 son irreversibles, lo que impidió su análisis cuantitativo para obtener los parámetros termodinámicos del proceso de desplegado. En el caso de M16SEA, el desplegado ocurrió con una marcada precipitación de la muestra, mientras que para M1SEA y MPE ICA512, la intensidad de la señal nativa no se recuperó al retornarlas a la temperatura inicial, aunque tampoco se observaron signos de precipitación. El ajuste de las curvas de desplegado a un modelo de dos estados, permitió estimar una T_m de 74.6 °C para M16SEA y 76.4 °C para M1SEA, aunque es importante destacar que estos son parámetros aparentes, ya que pueden depender de la concentración de proteína y de la velocidad de calentamiento. Se ha reportado anteriormente que M1SEA, con una secuencia de hexahistidinas, posee una T_m de aproximadamente 75 °C a pH 6.4, siendo el proceso también irreversible (Macao, Johansson *et al.* 2006). No fue posible estimar la T_m de MPE ICA512, por la dificultad en establecer cuál es la señal del estado desplegado, ya que se observan dos fenómenos: el proceso de desplegado que disminuye la magnitud de la elipticidad, y una transición a alta temperatura con una ganancia de elipticidad negativa.

Una característica notable de la transición de desplegado de M1SEA, es que ocurre con un aumento en la intensidad negativa de la señal dicroica, mientras que lo usual es que disminuya, como se espera para una reducción del contenido de estructura secundaria con el desplegado. Una interpretación posible, es que el estado final no sea el “estado desplegado”, sino un intermediario parcialmente plegado, con estructura secundaria no nativa. Como se describe más adelante en los experimentos de desplegado químico, esta proteína es realmente muy estable, con lo cual se puede esperar estructura residual en condiciones altamente desnaturalizantes.

Dado que la transición de desplegado de MPE fognina es reversible, fue posible extraer los parámetros termodinámicos asociados al proceso (Tabla 2.4). El desplegado se efectuó en 3 condiciones de pH y a distintas concentraciones de proteína. El estudio de las

transiciones a distinto pH permite a menudo extraer el cambio en la capacidad calorífica que acompaña el proceso (ΔC_p), ya que existe poca información en una sola curva para estimarlo con certeza. El análisis de la dependencia de la concentración estuvo inicialmente motivado por el estudio de la posible dimerización de este dominio, ya que se espera en el caso de las proteínas que son dímeros (u oligómeros), que por la ley de acción de masas, exista una dependencia en la posición de las curvas con la concentración de proteína, lo que representa una forma adicional de estimar ΔC_p . En concordancia con su estado de agregación monomérico, y con la reversibilidad del proceso de desplegado, no se observó dependencia en la posición de las curvas con la concentración de proteína. Los datos se ajustaron empleando las ecuaciones correspondientes a un modelo desplegado de dos estados unimolecular. Los parámetros termodinámicos resultantes de este análisis se muestran en la Tabla 2.4.

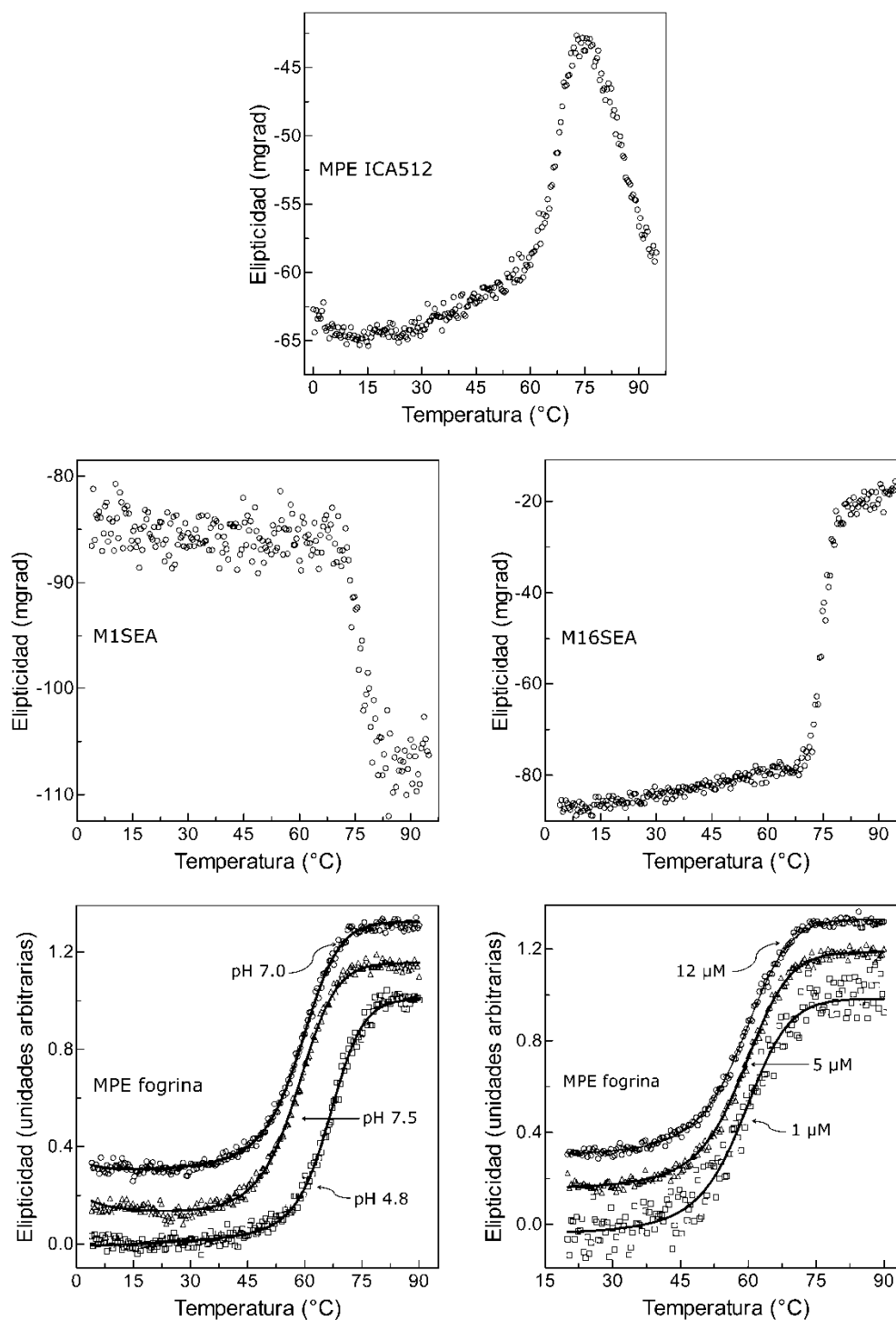


Figura 2.15. Desnaturalización por temperatura seguida mediante CD a 220 nm. A menos que se indique lo contrario, la incubación se hizo en *buffer* fosfato de sodio 25 mM, pH 7.0. En el panel inferior, los valores de elipticidad se normalizaron para llevarlos a una misma escala antes de hacer el ajuste, y se muestran desplazados arbitrariamente en el eje de ordenadas para mayor claridad. Las líneas representan el ajuste global de todos los datos de despliegue de MPE fogrina, a un modelo de dos estados unimolecular como se describe en el texto.

Tabla 2.4. Parámetros de desplegamiento por temperatura de MPE fognina^a.

pH	T_m^b (°C)	$\Delta H(T_m)^c$ (kcal mol⁻¹)	ΔC_p (kcal mol⁻¹ K⁻¹)	T_x^d (°C)	$\Delta G(T_x)^e$ (kcal mol⁻¹)
4.8	66.6 ± 0.1	50.9 ± 0.6	1.10 ± 0.03^f	23.4 ± 1.3	3.3 ± 1.5
7.0	59.1 ± 0.1	42.6 ± 0.4	1.10 ± 0.03	22.6 ± 1.1	2.4 ± 1.3
7.5	57.9 ± 0.1	41.3 ± 0.4	1.10 ± 0.03	22.5 ± 1.1	2.3 ± 1.2

^aLos parámetros se obtuvieron mediante regresión no lineal por cuadrados mínimos, aplicando las ecuaciones descritas en Materiales y Métodos. ^bTemperatura a la que $\Delta G = 0$ en la transición de desplegamiento. ^c ΔH a la temperatura T_m . ^dTemperatura en la que ΔG es máximo, $\Delta G = \Delta H$, y $\Delta S = 0$. ^e ΔG en T_x . ^f ΔC_p se ajustó globalmente a todos los datos.

Como se observa en la Tabla 2.4, la estabilidad de MPE fognina es dependiente del pH, con un ΔG de desplegamiento de mayor magnitud a pH 4.8, que es cercano al de los gránulos maduros de secreción de insulina. El cambio en la capacidad calorífica está de acuerdo con lo esperado para el desplegamiento completo de una proteína globular de este tamaño (1.15 kcal mol⁻¹ K⁻¹) (Myers, Pace *et al.* 1995), lo que indica que las transiciones reflejan el desplegamiento global de la proteína.

2.8.2. Desplegamiento inducido por agentes químicos

Las transiciones de desplegamiento químico de los cuatro dominios se siguieron por el cambio en la estructura secundaria, registrando los espectros de CD a distintas concentraciones de desnaturante. En el caso de MPE ICA512, se estudió también la pérdida de estructura terciaria por emisión de fluorescencia.

Para poder realizar un análisis cuantitativo de la termodinámica de plegado, es requisito que las transiciones sean reversibles y que se lleven a cabo en el equilibrio. Para asegurar que los tiempos de incubación de las proteínas en presencia de desnaturante permitan alcanzar el equilibrio químico, se prepararon réplicas, para cada proteína, a distintas concentraciones de desnaturante y se tomaron espectros a distintos tiempos. En base a experimentos preliminares, se tuvo especial precaución de incluir entre las concentraciones de desnaturante ensayadas a una que se encuentre cerca del punto medio de la transición, ya que comúnmente es esa la condición en la que más tiempo lleva alcanzar el equilibrio. La coincidencia de las medidas espectroscópicas registradas a distintos tiempos, se tomó como indicativa de que se había alcanzado el equilibrio. En la Figura 2.16, se muestra como ejemplo el desplegamiento de MPE ICA512 inducido por GdmCl, donde se observa

que los espectros tomados a distintos tiempos (se muestran solo dos para simplificar) coinciden dentro del error experimental. Los mismos resultados se obtienen con los otros tres dominios SEA. Además, experimentos cinéticos de desplegamiento, por mezclado manual de la proteína nativa con GdmCl, sugieren que el equilibrio se alcanza dentro de unos pocos minutos de iniciado el desplegamiento (no mostrado).

Para demostrar que el proceso es reversible, cada proteína se incubó en presencia de una alta concentración de desnaturalizante, de manera de asegurar su desplegamiento completo. Luego esta solución stock de proteína desplegada se diluyó con *buffer*, alcanzando distintas concentraciones de desnaturalizante y una concentración de proteína constante. Los espectros de CD, o las medidas de fluorescencia, de estas diluciones, se compararon con muestras de la misma concentración de proteína y desnaturalizante, pero preparadas a partir del stock de proteína nativa (sin desnaturalizante). La coincidencia de las señales espectroscópicas de estas dos muestras, preparadas por *desplegado* o por *replegado*, se consideró indicativa de que la transición es reversible. En la Figura 2.16 se ilustra este comportamiento con el dominio MPE ICA512, los cuatro dominios se comportan de manera semejante.

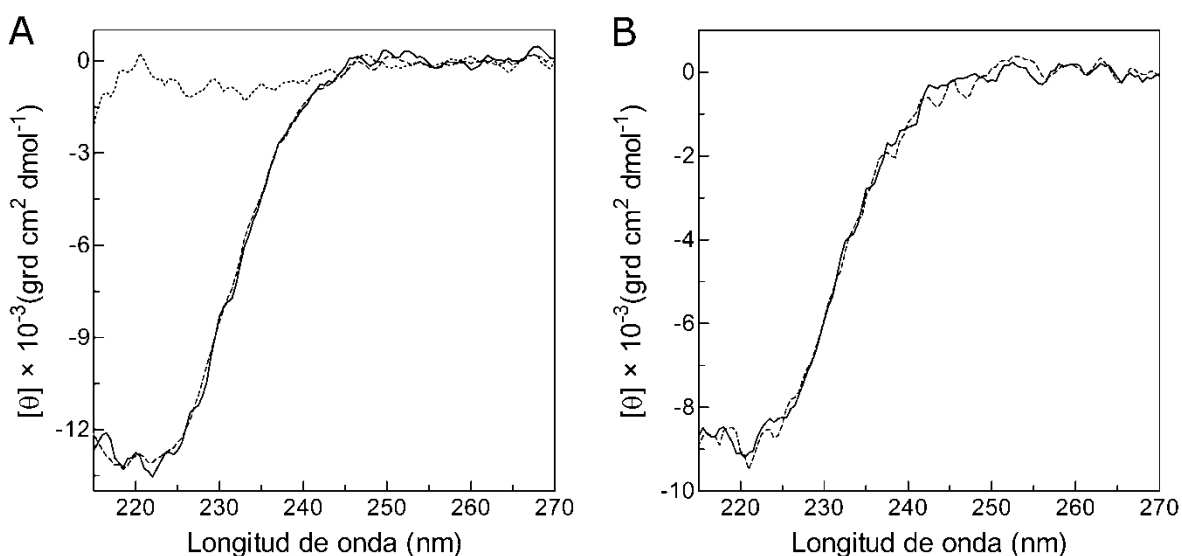


Figura 2.16. Reversibilidad y equilibrio del proceso de desplegamiento químico. (A) MPE ICA512 se desplegó por incubación con GdmCl 5.42 M por 6 horas, luego se diluyó a GdmCl 0.92 M y se tomó un espectro de CD luego de 4 horas de incubación (trazo grueso). Se compara con el espectro tomado luego de incubar la proteína directamente en presencia de GdmCl 0.92 M (trazo fino). En línea punteada se indica como referencia el espectro de la proteína desplegada en GdmCl 5.42 M. La concentración de proteína fue 15 μ M, y el *buffer* fosfato de sodio 25 mM, pH 7.0. (B) MPE ICA512 se incubó en GdmCl 2.03 M y se tomó un espectro de CD luego de 2.5 horas (trazo grueso) o 7 horas (trazo fino). La concentración de proteína fue 15 μ M, y el *buffer* acetato de sodio 25 mM, pH 4.8. La temperatura en todos los casos fue 20 °C.

2.8.2.1. Estudios de desplegado en el equilibrio a pH 7.0

En todas las transiciones monitoreadas por CD, se eligió la elipticidad a 222 nm para seguir el desplegado. La transición correspondiente a MPE fgrina, inducida por GdmCl, se muestra en la Figura 2.17. El desplegado inicia a baja concentración de desnaturalizante, pero la transición es altamente cooperativa. Ya que la proteína es un monómero en solución, los datos se analizaron mediante un modelo de dos estados unimolecular. Este modelo fue suficiente para explicar satisfactoriamente los datos, lo que permitió derivar los parámetros termodinámicos listados en la Tabla 2.5. El valor de ΔG obtenido ($2.29 \pm 0.13 \text{ kcal mol}^{-1}$) indica que MPE fgrina es un dominio relativamente poco estable, cerca del límite inferior de estabilidad que suelen tener las proteínas monoméricas de este tamaño (Jackson 1998). Este resultado coincide, dentro del error experimental, con la estabilidad determinada a partir de los experimentos de desplegado térmico (Sección 2.8.1). El valor obtenido para el parámetro m , que está relacionado con el cambio en el área accesible al solvente que acompaña al desplegado (Myers, Pace *et al.* 1995), está de acuerdo con el desplegado total de una proteína del tamaño de MPE fgrina.

Como una forma adicional de estudiar el proceso, se usó SEC-FPLC para monitorear el comportamiento hidrodinámico a distintas concentraciones de GdmCl (Figuras 2.17A y 2.18). El desplegado seguido por SEC se superpone muy bien con el ajuste de los datos de CD, excepto a concentraciones de GdmCl menores a 0.5 M. Esto se debe a que, como se describe en la Sección 2.4, el comportamiento de MPE fgrina en SEC es anómalo en ausencia de desnaturalizante. La presencia de GdmCl parece normalizar este comportamiento, dando inicialmente un R_S cercano al esperado para un monómero, seguido de la transición principal correspondiente al desplegado.

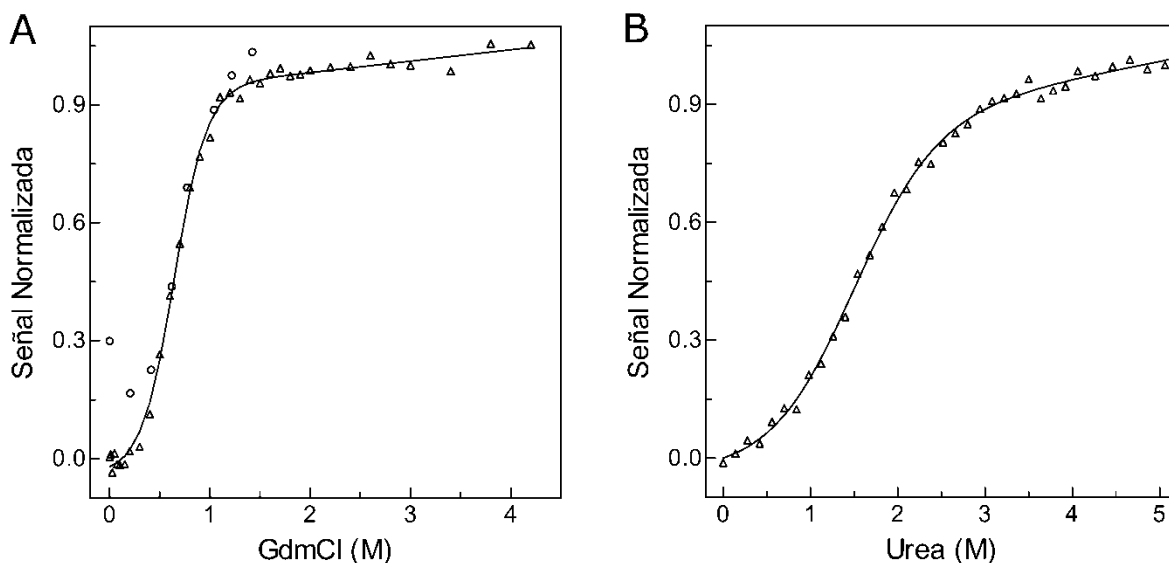


Figura 2.17. Desplegado químico de MPE fogrina en el equilibrio. La transición se siguió por CD a 222 nm (triángulos) o por SEC (círculos). La línea representa el ajuste al modelo de dos estados descrito en el texto. (A) Desplegado por GdmCl. (B) desplegado por urea. El *buffer* empleado fue fosfato de sodio 25 mM pH 7.0, y en los experimentos de SEC, al mismo *buffer* se le añadió cloruro de sodio 0.15 M. La temperatura en todos los casos fue 20 °C.

Finalmente, el mecanismo de desnaturalización química de MPE fogrina se estudió adicionalmente en presencia de urea (Figura 2.17B). De manera similar a lo observado con GdmCl, el proceso de desplegado con urea comienza a baja concentración de desnaturalizante, es monofásico y altamente cooperativo. El ajuste de un modelo unimolecular de dos estados explicó satisfactoriamente los datos, lo que permitió derivar que en ausencia de urea $\Delta G = 1.78 \pm 0.12 \text{ kcal mol}^{-1}$, y $m = 1.22 \pm 0.08 \text{ kcal mol}^{-1} \text{ M}^{-1}$. Por lo tanto, también el desplegado químico por urea es un proceso de dos estados, y revela que MPE fogrina es marginalmente estable.

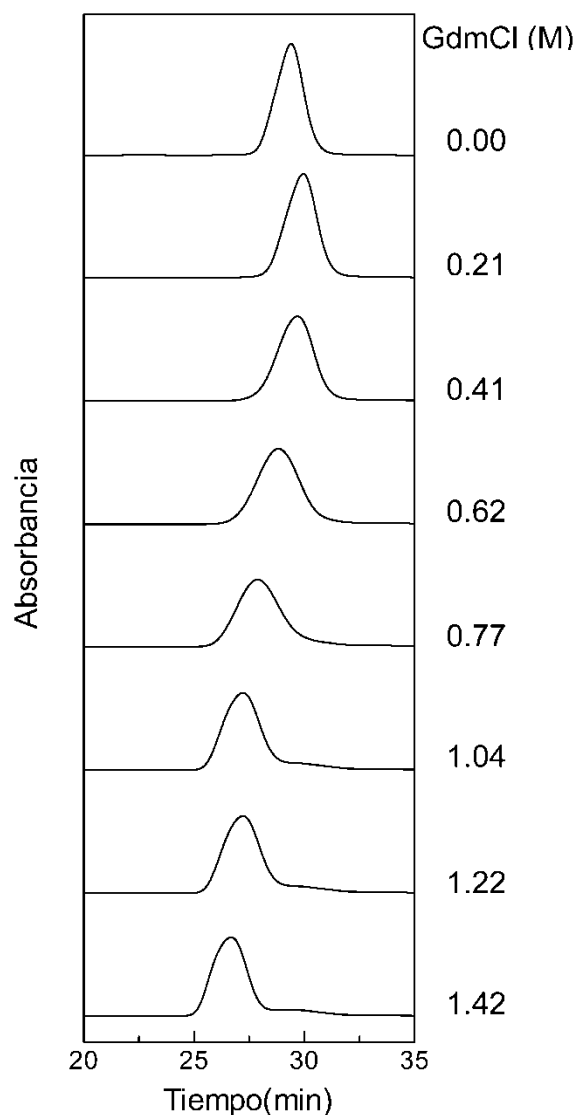


Figura 2.18. Desplegado químico de MPE fogrina en el equilibrio seguido por SEC. Se realizaron corridas de SEC en *buffer* fosfato de sodio 25 mM, cloruro de sodio 0.15 M, pH 7.0, con las concentraciones de GdmCl indicadas, registrando la absorbancia a 280 nm.

El proceso de desplegado de M16SEA también está caracterizado por una reducción sigmoidea de la elipticidad a 222 nm (Figura 2.19). Pero en este caso, la transición inicia a una concentración de desnaturalizante mucho más alta que en MPE fogrina, y es tan cooperativa como lo observado en esta última. Dado que se observa una sola transición, y que la proteína es monomérica, los datos se analizaron de acuerdo a un modelo de desplegado de dos estados unimolecular. Este modelo explica muy bien los datos obtenidos, y permitió derivar los parámetros termodinámicos indicados en la Tabla 2.5. El valor de ΔG confirma que este dominio es muy estable.

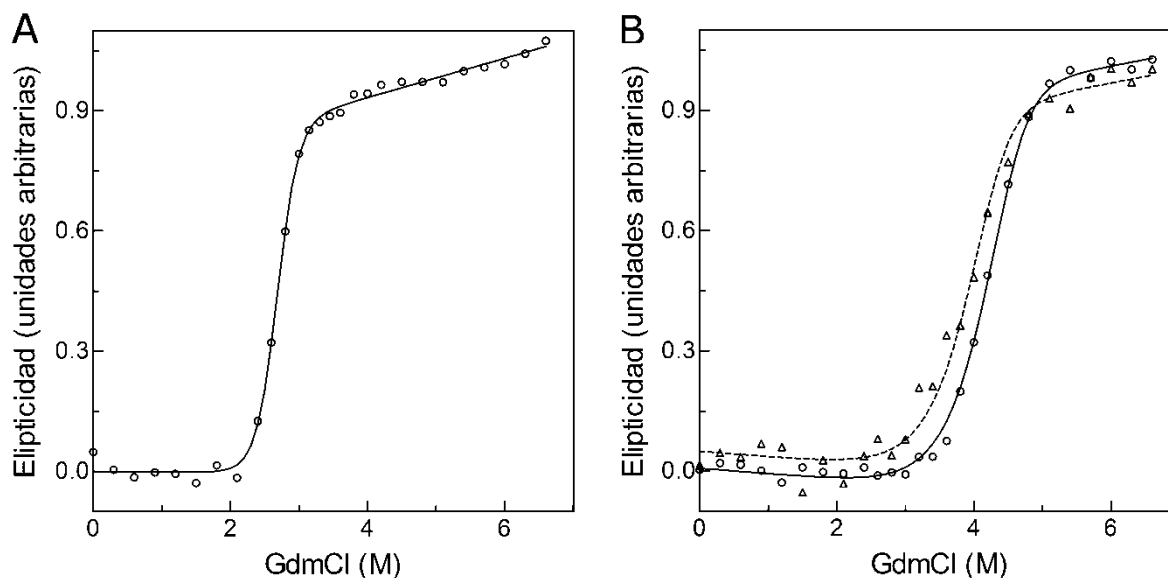


Figura 2.19. Desplegado en el equilibrio de M16SEA y M1SEA. Se prepararon tubos separados en las concentraciones de GdmCl indicadas y se tomaron espectros de CD en el UV lejano. **(A)** M16SEA. **(B)** M1SEA a dos concentraciones: 13 μM (círculos) y 4 μM (triángulos). Las líneas indican el ajuste a un modelo de dos estados, como se indica en el texto. En ambos casos, el *buffer* empleado fue fosfato de sodio 25 mM pH 7.0, y la temperatura fue de 20 $^{\circ}\text{C}$.

Las curvas de desplegamiento de MPE ICA512 se siguieron por CD y fluorescencia total de tirosinas. Las transiciones seguidas por uno u otro método se ven coincidentes dentro del error experimental. Dado que MPE ICA512 es un dímero en solución, el análisis de los datos debió contemplar el proceso de disociación además del desplegamiento. Existen en principio muchos posibles mecanismos, pero los tres más comúnmente observados son el de dos estados, donde el dímero nativo (N_2) se encuentra en equilibrio sólo con el monómero desplegado (U), y dos mecanismos de tres estados que involucran además a un intermediario dimérico (I_2) o uno monomérico (I). Además, hay que considerar la posibilidad de que el proceso de disociación no conlleve un cambio significativo de la señal espectroscópica o que el dímero se despliegue parcialmente sin llegar a disociarse en el rango de concentraciones de desnaturalizante ensayado. Estas dos situaciones se describen mediante mecanismos de desplegamiento unimolecular. Se puede obtener mayor información acerca de cuál mecanismo de desplegamiento está operando, realizando los experimentos a distintas concentraciones de proteína, ya que el cambio de molecularidad produce alteraciones características en la forma sigmoidal de las curvas, respecto a lo observado en las transiciones de dos estados unimoleculares (Rumfeldt, Galvagnion *et al.* 2008). Por esta razón, se realizaron los experimentos a distintas concentraciones de proteína, en el rango de 1-15 μM .

Tabla 2.5. Parámetros de estabilidad termodinámica de los dominios SEA^a.

Proteína	Modelo ^b	D_m (M)	m (kcal mol ⁻¹ M ⁻¹)	ΔG_0^c (kcal mol ⁻¹)
MPE fgrina	N ↔ U	0.64 ± 0.01	3.57 ± 0.19	2.29 ± 0.13
M16SEA	N ↔ U	2.68 ± 0.01	3.80 ± 0.26	10.19 ± 0.69
MPE ICA512	(N ₂ ↔) 2N' ↔ 2U	2.13 ± 0.02	2.68 ± 0.23	5.69 ± 0.49
M1SEA	N ₂ ↔ 2U	6.59 ± 0.15	2.71 ± 0.17	17.86 ± 1.18

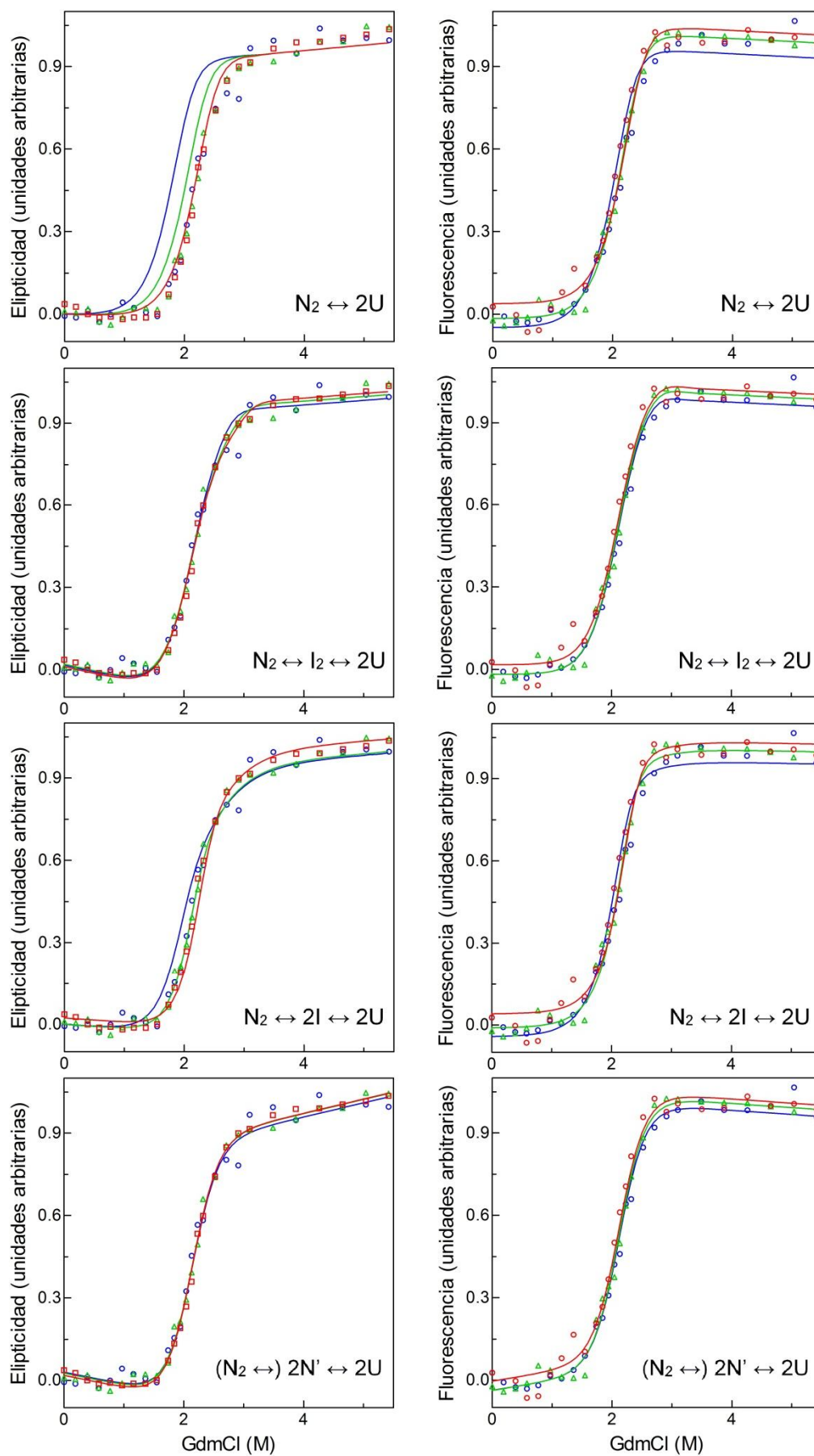
^aLos parámetros se obtuvieron mediante regresión no lineal por cuadrados mínimos, aplicando las ecuaciones descritas en Materiales y Métodos. ^bLos paréntesis en el primer equilibrio indican que esa etapa es silenciosa espectroscópicamente y no se ha tenido en cuenta en el análisis de los datos. ^c ΔG en ausencia de desnaturizante, $\Delta G_0 = mD_m$.

Los experimentos de desplegado de MPE ICA512, analizados mediante un ajuste global que incluye tanto a los datos de CD como de fluorescencia, de acuerdo a los tres modelos que incluyen disociación, se muestran en la Figura 2.20. A simple vista, se observa que las curvas de desplegado no dependen de la concentración de proteína, y esto se ve reflejado en el análisis de los datos con los modelos bimoleculares: el modelo de dos estados no ajusta satisfactoriamente los datos ya que predice una dependencia mucho mayor de la posición de las curvas con la concentración de proteína, mientras que los modelos de tres estados ajustan mejor pero arrojan valores de parámetros sin sentido físico, o que terminan reduciendo el mecanismo de tres estados en uno de dos estados.

En cambio, un modelo de dos estados unimolecular resulta en un muy buen ajuste a los datos de desplegado de MPE ICA512. En la Tabla 2.5 se indican los parámetros termodinámicos correspondientes. Esto podría significar que el desplegado ocurre sin la disociación del dímero, o bien que la disociación ocurre sin un cambio detectable en la señal espectroscópica. Ya que en ausencia de desnaturizante se observan cantidades significativas de tanto dímero como monómero (Sección 2.4), y que la constante de dimerización estimada por SEC corresponde a una afinidad relativamente baja, es poco probable que el dímero persista a altas concentraciones de GdmCl. Además, experimentos de desplegado con GdmCl seguidos por SEC, análogos a los realizados con MPE fgrina, muestran una reducción en el R_s que precede a la transición de desplegado. Por todo lo expuesto, el escenario más probable es que la disociación de las subunidades de MPE ICA512 sea espectroscópicamente silenciosa, y que preceda a la transición de desplegado, con lo cual los parámetros termodinámicos derivados corresponden a la estabilidad del monómero de MPE ICA512, y están dentro del rango comúnmente encontrado en proteínas globulares de este tamaño.

Dado que M1SEA es un único dominio formado por dos cadenas distintas, el análisis de los datos de desplegado involucró necesariamente un mecanismo que incluye desplegado y disociación. El mecanismo más sencillo, de dos estados, donde el desplegado y la disociación se producen simultáneamente, fue suficiente para explicar los datos (Figura 2.19B), lo que permitió derivar los parámetros termodinámicos indicados en la Tabla 2.5. El cambio de molecularidad que acompaña al desplegado se ve reflejado en la dependencia de la posición de las curvas con la concentración de proteína. Es oportuno recalcar dos aspectos importantes para la evaluación de los parámetros termodinámicos para proteínas oligoméricas, en comparación con los de proteínas monoméricas. En primer lugar, la concentración de desnaturizante en el punto medio de la transición, $[D]_{1/2}$, definida como la concentración a la que $F_u = 0.5$ (la fracción de proteína desplegada) constituye un parámetro poco útil para comparar la estabilidad de distintas proteínas, ya que depende de la concentración a la que se realizó el experimento, y por lo tanto, tampoco puede emplearse como un parámetro compartido en el ajuste global de curvas a distintas concentraciones de proteína. Por eso suele emplearse el D_m , la concentración de desnaturizante a la que $\Delta G = 0$ como parámetro para describir la estabilidad de la proteína. En proteínas monoméricas, $[D]_{1/2}$ y D_m coinciden, pero no es así en las proteínas oligoméricas, excepto en el estado estándar 1 M. En segundo lugar, la comparación de los valores de ΔG con los correspondientes a proteínas monoméricas no es directa. Por ejemplo, un dímero con la estabilidad de M1SEA, 17.86 kcal/mol, tiene la misma fracción de proteína desplegada (1.5×10^{-5}) que una proteína monomérica con un $\Delta G = 6.45$ kcal/mol a una concentración de dímero de 100 μM . Por otro lado, los datos de desplegado obtenidos en esta tesis son coincidentes con los ya reportados por otro grupo de investigación, empleando una construcción de M1SEA ligeramente diferente (que incluye una secuencia de hexahistidinas) y hechos a pH 7.5 (Sandberg, Johansson *et al.* 2008).

Figura 2.20. Desplegado en el equilibrio de MPE ICA512 a pH 7.0 (página siguiente). Se prepararon tubos separados con las concentraciones de GdmCl indicadas, y se tomaron espectros de CD en el UV lejano (panel izquierdo), o se midió la fluorescencia total (panel derecho). Las líneas representan el ajuste global de todos los datos a los modelos indicados. Las concentraciones de proteína en los experimentos de CD fueron: 1, 5 y 15 μM , representados en color azul, verde y rojo, respectivamente, y en los experimentos de fluorescencia fueron: 5, 10 y 15 μM , representados en azul, verde y rojo, respectivamente. La temperatura fue 20 °C.



2.8.2.2. Desplegado químico de MPE ICA512 a pH 4.8

El contenido de los gránulos de secreción sufre un cambio gradual de pH a lo largo del proceso de secreción, mientras que el pH del retículo endoplasmático es de 7.0 ± 0.2 , disminuye hasta cerca de 5.0 ± 0.4 en los GS maduros y es nuevamente cercano a 7.0 cuando es liberado en la membrana plasmática durante el proceso de exocitosis (Wu, Grabe *et al.* 2001; Paroutis, Touret *et al.* 2004). Es por ello importante comprender el efecto del pH sobre la estabilidad estructural de la región MPE de ICA512 y de fognrina. Para ampliar lo ya aprendido por desplegado térmico con MPE fognrina, estudiamos la estabilidad de MPE ICA512 por desplegado químico con GdmCl, ya que su desplegado térmico es irreversible.

Para este fin, se repitieron los experimentos de desplegado a distintas concentraciones de proteína pero a pH 4.8, siguiendo la transición por CD a 222 nm. Como puede observarse en la Figura 2.21, las transiciones son muy semejantes a lo observado a pH 7.0, por lo que los datos se analizaron de acuerdo a un modelo de dos estados unimolecular. Los parámetros resultantes de dicho análisis fueron $D_m = 2.03 \pm 0.03$ M, $m = 2.79 \pm 0.27$ kcal mol⁻¹ M⁻¹ y $\Delta G = 5.68 \pm 0.56$ kcal mol⁻¹, donde es evidente que su estabilidad termodinámica no se modifica significativamente al disminuir el pH.

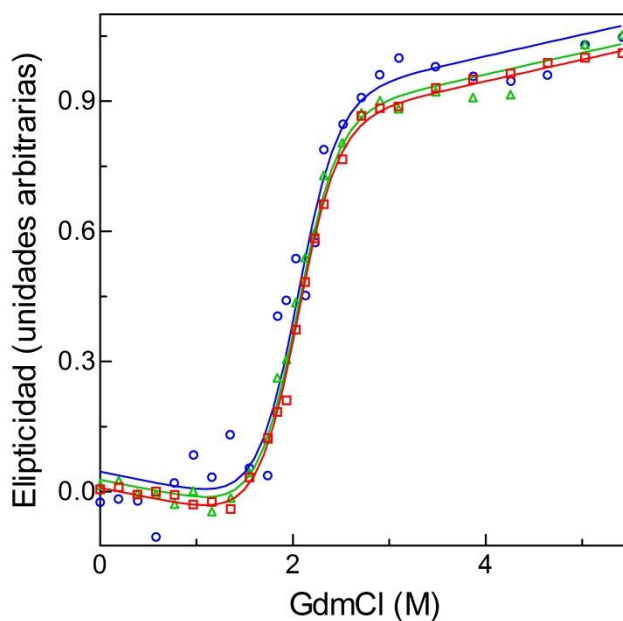


Figura 2.21. Desplegado en el equilibrio de MPE ICA512 a pH 4.8 (página anterior). Se prepararon tubos separados en las concentraciones de GdmCl indicadas, y se tomaron espectros de CD en el UV lejano. Las concentraciones de proteína fueron 1, 5 y 15 µM, representadas en azul, verde y rojo, respectivamente. El buffer empleado fue acetato de sodio 25 mM pH 4.8 y la temperatura 20 °C. Las líneas representan el ajuste global al modelo de dos estados descrito en el texto.

2.9. Proteólisis *in vitro* de MPE fogrina

En el laboratorio donde se llevó a cabo esta tesis se había demostrado previamente que el ectodominio maduro de ICA512 sufre un proceso de proteólisis *in vitro* mediado por especies reactivas de oxígeno (ROS). Proceso que resulta en la degradación de las regiones que flanquean al dominio MPE. Para investigar si fogrina podría experimentar un proceso de degradación semejante, se clonó y expresó una versión extendida de MPE fogrina, que comprende los residuos 486-614. En esta construcción, el extremo C es inmediatamente anterior al comienzo de la región transmembrana, mientras que el extremo N se escogió de manera de abarcar una región equivalente a la estudiada con ICA512 (Primo, Klinke *et al.* 2008). A esta construcción la denominamos IA-2b⁴⁸⁶⁻⁶¹⁴.

IA-2b⁴⁸⁶⁻⁶¹⁴ se expresó y purificó de la misma manera que MPE fogrina. Esta variante se comporta de manera semejante al dominio MPE por SEC y está bien plegada, aunque en comparación con MPE fogrina, posee una menor señal dicroica en el UV lejano y también en el UV cercano, lo que sugiere que los segmentos añadidos en ambos extremos no adquieren estructura secundaria rígida, e incluso perturban la estructura terciaria del dominio MPE. Notablemente, esta variante es también más estable que MPE fogrina (la caracterización estructural básica de esta variante se muestra en el APÉNDICE).

Cuando IA-2b⁴⁸⁶⁻⁶¹⁴, purificada con el método descripto para MPE fogrina, se incubó en *buffer* a 20 °C, se observó una marcada degradación en el término de unos pocos días (Figura 2.22). Esto también ocurrió a 4 °C durante su almacenamiento (no mostrado). Además, se observa que el proceso ocurre con la aparición de múltiples bandas de tamaño intermedio. Otra característica notable es que la reacción puede trascurrir hasta dar cerca de un 100 % de degradación.

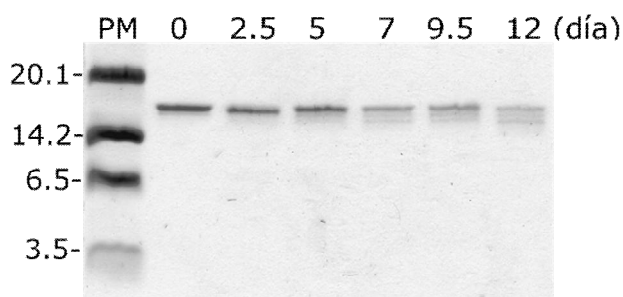


Figura 2.22. Proteólisis espontánea de fogrina. La proteína se incubó en buffer MES 50 mM pH 6.0, a 20 °C. Se tomaron alícuotas en los tiempos indicados y se corrieron por SDS-PAGE.

Para descartar la presencia de proteasas contaminantes, se ensayó en primer lugar la existencia de actividad proteolítica en las muestras, mediante un ensayo de hidrólisis de α -

caseína. En este experimento, se añade a las muestras una cierta cantidad de α -caseína, que es un sustrato reconocido por un gran número de proteasas, y luego de un cierto período de incubación, se precipitan todas las proteínas de la muestra, quedando en solución sólo los pequeños péptidos solubles resultantes de la hidrólisis de caseína, cuya concentración, medida por absorbancia, es una medida de la actividad proteolítica en la muestra. El resultado de dicho experimento fue negativo en los tiempos y condiciones de incubación en los cuales se observó degradación significativa de IA-2b⁴⁸⁶⁻⁶¹⁴ por SDS-PAGE. Como un control adicional, se sembró un exceso de IA-2b⁴⁸⁶⁻⁶¹⁴ en SDS-PAGE, lo que no mostró bandas adicionales contaminantes.

Para llevar a cabo una caracterización más detallada del proceso de degradación, se llevaron a cabo un número de experimentos adicionales de incubación de la proteína IA-2b⁴⁸⁶⁻⁶¹⁴ en distintas condiciones. En los experimentos mostrados en la Figura 2.23, la proteína sufrió una degradación significativa durante el proceso de purificación, por lo que se observan dos bandas en el tiempo cero. Por un lado, se estudió el efecto del PMSF, un inhibidor irreversible de serina proteasas. Como se observa en la Figura 2.23B, este aditivo no frena el proceso de degradación, aunque sí lo enlentece. Paralelamente, se determinó el pH óptimo de degradación incubando la proteína en distintos *buffers*: acetato de sodio pH 4.0 y 5.0, MES-Na pH 6.0 y 7.0, y Tris-Cl pH 8.0 y 9.0, todos ellos 50 mM (Figura 2.23E). Se observó que la degradación se encuentra favorecida a pH ácido, aunque sigue habiendo degradación significativa a pH 9.0. Esta es una diferencia notable respecto a lo observado con ICA512, donde la proteólisis se encuentra favorecida a pH neutro, y significativamente inhibida a pH ácido (Primo, Klinker *et al.* 2008).

Teniendo en cuenta que ICA512 y fognina pertenecen a una misma familia, es razonable investigar la posibilidad de que la proteólisis de fognina también estuviera mediada por ROS. Una forma de evidenciar la presencia de ROS es mediante el estudio del efecto de compuestos que son conocidos inhibidores de la generación de estos radicales, conocidos como atrapantes de ROS. Cuando el DMSO, un conocido atrapante de radicales, particularmente hidroxilos ($\text{HO}\cdot$), se añadió a la muestra de IA-2b⁴⁸⁶⁻⁶¹⁴, se observó una muy marcada desaceleración de la degradación (Figura 2.23D). Los radicales hidroxilos suelen generarse por la reacción de Fenton a partir de peróxidos o de radicales superóxido, mediante catálisis por Fe^{2+} y algunos otros metales de transición. Por esta razón, se probó el efecto de secuestrar los metales de la solución, mediante el agregado de los quelantes EDTA y DTPA. Es importante resaltar que la reactividad de los metales de transición, en cuanto a su capacidad de generar radicales libres, no se ve fuertemente afectada al estar complejados por quelantes, así que el efecto de los quelantes resulta principalmente de secuestrar estos metales de su sitio de unión en la macromolécula (Halliwell y Gutteridge 1986). Como los radicales libres son especies muy reactivas, con una vida media muy baja, tienen un radio de

difusión de unas pocas decenas de Å, así que deben estar ubicados en sitios específicos para que se pueda observar el clivaje. El agregado de EDTA prácticamente detuvo la degradación (Figura 2.23C), y el DTPA tuvo un efecto semejante (no mostrado).

En resumen, se ha encontrado que fognina experimenta un proceso de degradación *in vitro*, pero son necesarios experimentos adicionales para establecer la naturaleza exacta del mecanismo proteolítico, los cuales se están llevando a cabo actualmente. Los experimentos mostrados sugieren la participación de iones metálicos, y posiblemente de especies reactivas de oxígeno.

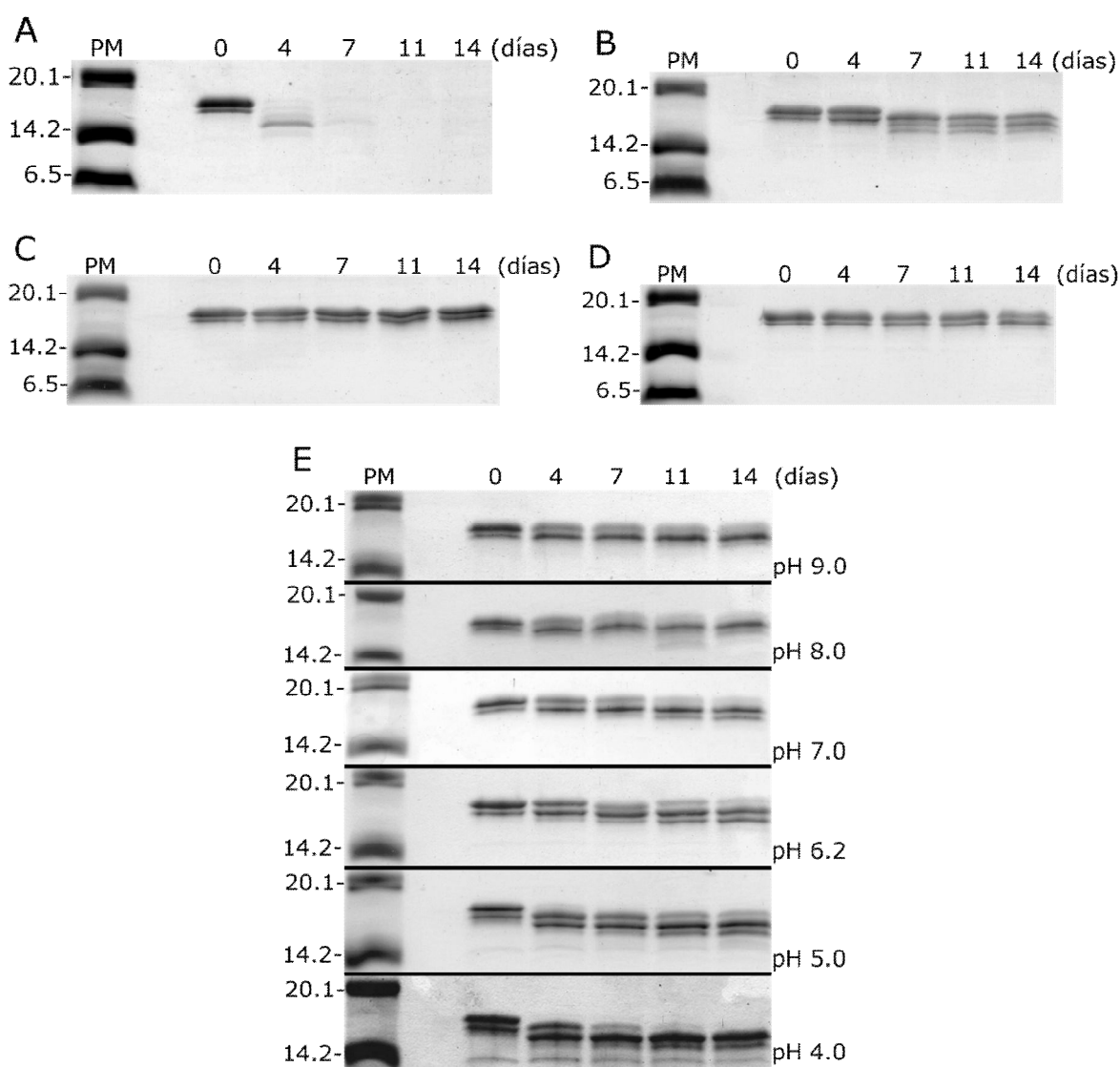


Figura 2.23. Proteólisis espontánea de fognina en distintas condiciones. La proteína se incubó en *buffer* fosfato de sodio 25 mM pH 7.0, a menos que se indique lo contrario, y la temperatura fue de 20 °C. Se tomaron alícuotas en los tiempos indicados y se corrieron por SDS-PAGE. Se probó el efecto del agregado de distintos agentes. (A) Sin agregado. (B) PMSF 1 mM. (C) EDTA 5 mM. (D) DMSO 10% v/v. (E) Incubación en distintos *buffers*, todos 50 mM: Tris·Cl pH 9.0 y pH 8.0, MES·Na pH 7.0 y pH 6.2, acetato de sodio pH 5.0 y pH 4.0.

3. Estructuras y secuencias de los dominios SEA

3.1. Introducción

Las proteínas que integran la familia de dominios SEA poseen una estructura tridimensional muy conservada y una identidad de secuencia baja. Cuando se compara el conjunto de secuencias de una familia de estas características, se suelen identificar patrones de residuos conservados que pueden formar parte de las interacciones que definen las características del plegado y estabilizan su conformación. Similarmente, también suelen estar muy conservados los residuos importantes para la función, ya sea que pertenezcan a un sitio activo, o que formen parte de las superficies de interacción proteína-proteína.

Para tratar de identificar las interacciones importantes para la estructura y función de los SEA, ampliamos la información estructural para esta familia con la determinación de la estructura cristalográfica del dominio MPE fogrina, y luego realizamos un análisis detallado de las secuencias y estructuras tridimensionales disponibles. Como se describió en los capítulos previos, ICA512 y fogrina son dos homólogos cercanos, que comparten muchas de sus propiedades funcionales y estructurales. Sin embargo, mientras que MPE ICA512 forma un dímero en solución, MPE fogrina es un monómero en las mismas condiciones. De esta forma, al disponer de la estructura tridimensional de ambos dominios, se puede intentar explicar cómo las diferencias en su estructura cuaternaria se encuentran determinadas a nivel de su secuencia. Finalmente, esto también permite derivar un número de hipótesis acerca de la relevancia *in vivo* de los procesos de oligomerización en estos dominios.

Hacia el final del capítulo, se analizan las estructuras tridimensionales y las secuencias de los dominios SEA, y se identifican residuos e interacciones importantes para la conservación del plegado, junto con otros que pueden ser importantes para su función.

3.2. Determinación de la estructura tridimensional de MPE fogrina por difracción de rayos x

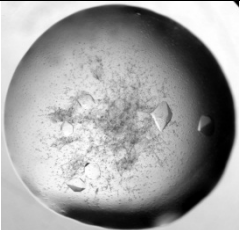
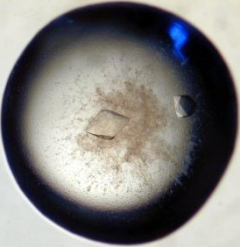
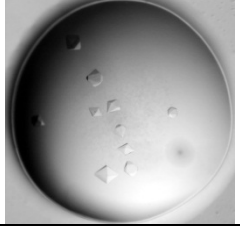
3.2.1. Cristalogénesis y resolución de la estructura

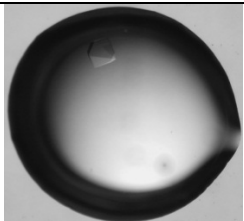
La cristalización MPE fogrina se llevó a cabo mediante la técnica de difusión de vapor en gota suspendida, explorando un amplio rango de condiciones de cristalización. La proteína se obtuvo siguiendo la misma metodología de purificación descrita para los estudios biofísicos del Capítulo 2, añadiendo únicamente el cambio del *buffer* fosfato por Tris 10 mM pH 7.4, para evitar la interferencia del ion fosfato con los aditivos presentes en las soluciones de

cristalización. Se obtuvieron cristales en diversas condiciones (Tabla 3.1), tanto a pH ácido como neutro, dentro de unos pocos días de incubación a 20 °C.

Para obtener los datos de difracción, en primer lugar los cristales debieron ser congelados y enviados al sincrotrón del Brookhaven National Laboratory (Nueva York, EE.UU.). Se ensayaron diversas condiciones de crioprotección, para lo cual se transfirieron los cristales individuales a soluciones de idéntica composición a la de cristalización, pero con el agregado de una sustancia criopreservante, para luego congelarlos en nitrógeno líquido. Los mejores datos de difracción se obtuvieron, a pH 7.4, en la condición 31 de la Tabla 3.1, sin el agregado de criopreservante, esto último se esperaba ya que la solución de cristalización contenía PEG, un polímero con propiedades crioprotectoras. A pH 4.6, se obtuvieron buenos datos de difracción para la condición 20, usando glicerol 25 % v/v como crioprotector.

Tabla 3.1. Condiciones de cristalización de MPE fogrina.

Condición	Composición	Tiempo de aparición del cristal	
30	Sulfato de amonio 0.2 M, PEG 8000 30 % p/v	< 1 día (luego de incubación ON)	
31	Sulfato de amonio 0.2 M, PEG 4000 30 % p/v	< 1 día (luego de incubación ON)	
47	Acetato de sodio 0.1 M pH 4.6, sulfato de amonio 2 M	2 - 3 días	
20	Acetato de sodio 0.1 M pH 4.6, sulfato de amonio 0.2 M, PEG 4000 25 % p/v	2 - 3 días	

32	Sulfato de amonio 2 M	5 semanas	
----	-----------------------	-----------	---

En las dos condiciones de pH estudiadas, MPE fognina cristalizó en el grupo espacial hexagonal $P6_122$ y con parámetros de celda similares (ver Tabla 3.2). A partir de esos parámetros, se analizó el contenido de la celda, lo que indicó que en ambas condiciones se encuentra sólo una molécula de proteína por unidad asimétrica.

Tabla 3.2. Estadísticas de resolución de las estructuras de MPE fognina.

pH de la muestra	7.4	4.6
RECOLECCIÓN DE DATOS ^a		
Longitud de onda (Å)	0.91840	0.91840
Grupo espacial	$P6_122$	$P6_122$
Parámetros de la celda unidad		
Dimensiones (Å)	$a=b=55.133$ $c=148.39$	$a=b=54.742$ $c=149.147$
Ángulos (°)	$\alpha=\beta=90$ $\gamma=120$	$\alpha=\beta=90$ $\gamma=120$
Límites de resolución (Å)	30.00-1.95 (1.98-1.95)	20.00-2.01 (2.03-2.01)
R_{merge}^b	0.06 (0.99)	0.06 (0.79)
$I/\sigma(I)$ promedio	104.6 (7.72)	41.2 (3.6)
Compleitud (%)	99.6 (100)	95.4 (61.9)
Redundancia	40.5 (41.4)	12.5 (9.7)
Coeficiente de Matthew (Å ³ /Da)	3.01	2.98
% de solvente	59.17	58.79
Moléculas/unidad asimétrica	1	1
REFINAMIENTO		
Límites de resolución (Å)	30.00-1.95	20.00-2.01
Número de reflexiones	10292	9206
$R_{\text{work}}/R_{\text{free}}^c$	0.221/ 0.243	0.203/0.238
Número de átomos proteicos	670	676
Número de moléculas de agua	40	34

Factores B promedio (Å ²)		
Átomos de la proteína	48.5	46.7
Moléculas de agua	45.6	48.7
Desviaciones cuadráticas medias		
Longitud de enlace (Å)	0.021	0.019
Ángulos de enlace (°)	1.563	2.000
Ramachandran		
Más favorecido (%)	100	100

^aLos valores entre paréntesis corresponden a la capa de mayor resolución.

^b $R_{\text{merge}} = \sum_h \sum_i (|I_h| - \langle I_h \rangle) / \sum_h \sum_i I_{h,i}$. ^c $R_{\text{work}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$, donde F_{obs} y F_{calc} son las amplitudes de los factores de estructura observados y calculados, respectivamente. R_{free} es igual que R_{work} , pero calculado con un 5 % de las reflexiones, elegidas al azar, que se omitieron del refinamiento.

Ya que la proteína cristalizada posee una alta similitud de secuencia con el dominio MPE ICA512 (43 % de identidad), se apuntó a resolver su estructura mediante remplazo molecular. Este procedimiento se aplicó exitosamente, empleando como modelo de búsqueda a las coordenadas de MPE ICA512 a pH 8.5 (PDB ID: 2QT7). El uso de las otras estructuras cristalográficas de MPE ICA512 (PDB ID: 3NO1 a pH 8.5 y 3NP5 a pH 4.5), no mejoró el contraste de las soluciones. Finalmente, ambas estructuras se refinaron y verificaron estereoquímicamente. Algunos parámetros estadísticos de la recolección de datos, del refinamiento y validación de las estructuras, se indican en la Tabla 3.2.

3.2.2. Estructura tridimensional de MPE fgrina

Las estructuras cristalográficas de MPE fgrina determinadas a pH 7.4 y a pH 4.6, y resueltas a 1.95 y 2.01 Å, respectivamente, resultaron esencialmente idénticas, con un RMSD de 0.29 Å para los átomos del *backbone*. En la Figura 3.1 se muestra la estructura obtenida a pH 7.4, y los parámetros cristalográficos y del refinamiento se consignan en la Tabla 3.2. En los dos cristales se observó una molécula de proteína por unidad asimétrica y su contenido de solvente es de ~60 %. En las dos estructuras, hay residuos para los cuales no se observa densidad electrónica interpretable. Estos son los primeros 8 residuos en ambas, el último residuo también en las dos, y el penúltimo a pH 7.4. Es muy probable que esto se deba a que son regiones muy móviles, ya que estos residuos son predichos como desestructurados por varios algoritmos de predicción de dominios globulares y de estructura secundaria a partir de la secuencia de aminoácidos. Además, la identidad de secuencia con ICA512, y con el resto de los dominios SEA, es muy baja para los primeros residuos de la región N-terminal. Ya que

las estructuras a los dos pH son muy similares, para simplificar la discusión, en los párrafos siguientes nos referiremos a la estructura de MPE fgrina como la obtenida a pH 7.4, a menos que se especifique lo contrario. Además, dado que las estructuras de los monómeros de MPE ICA512 son casi idénticas en una variedad de condiciones de pH ensayadas (Primo, Jakoncic *et al.* 2011), nos referiremos a la estructura de este dominio como la determinada a pH 8.5 (Primo, Klinke *et al.* 2008).

Como se esperaba, la estructura de MPE fgrina es muy similar a la de MPE ICA512 (Figura 3.1), con un plegado tipo ferredoxina, consistente de una lámina β de cuatro hebras antiparalelas, sobre las que se empaquetan dos hélices α , la segunda de ellas con un pliegue característico de estos dominios. Las estructuras se superponen con un RMSD de 1.36 Å para los átomos del *backbone*, con sólo sutiles diferencias en la orientación relativa de las hélices y en los *loops*, como se discute en la próxima sección.

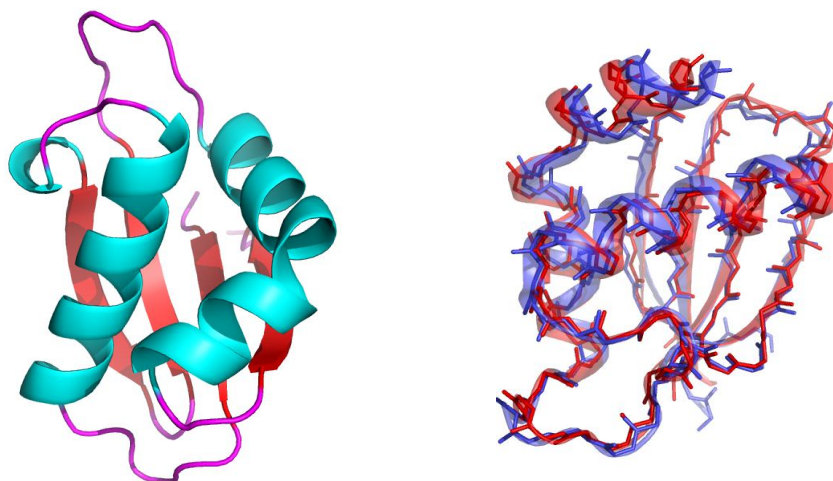


Figura 3.1. Estructura cristalográfica de MPE fgrina. Se muestra una ilustración de cintas, destacando los elementos de estructura secundaria. A la derecha, se superpusieron las estructuras de MPE fgrina (rojo) y de ICA512 (azul).

La hoja β posee un patrón regular de puentes de hidrógeno, excepto por dos β -*bulges*, uno al inicio de β 2 (residuos 546 y 547) y otro al final de β 4 (residuos 590-591). Hay 3 β -*turns*, el primero involucra a los residuos 518-521 y conecta la hebra β 1 con la hélice α 1, pertenece a la categoría VIb, e incluye a la prolina 520 que adopta una configuración *cis*. El segundo β -*turn* involucra los residuos 550-553, se clasifica como tipo II' y conecta β 2 y β 3, donde también interviene un residuo de prolina (P552) que se encuentra en configuración *trans*. El último de los β -*turns* comprende a los residuos 561-564, es de tipo I y conecta a α 1 con β 2. Otro segmento de estructura regular es una pequeña hélice 3_{10} de 3 residuos (una vuelta de hélice), que involucra a los residuos 541-543.

3.3. Comparación de las estructuras y secuencias de MPE ICA512 y MPE fognina

Como se mencionó en la sección precedente, los monómeros de MPE fognina y de MPE ICA512 poseen estructuras tridimensionales muy similares. Los elementos de estructura secundaria se superponen muy bien, y comparten una gran similitud de secuencia (Figura 3.2). Una diferencia notable se encuentra en la hélice $\alpha 1$, pues mientras que existe un marcado predominio de residuos hidrofóbicos en ICA512, en fognina hay más residuos hidrofílicos. Aun así, las interacciones en esta región están conservadas, pues las mismas se forman con la parte no polar de los residuos hidrofílicos de MPE fognina. Las secuencias de las hebras $\beta 1$ y $\beta 4$ son las más conservadas, mientras que en las hélices la identidad es más baja (Figura 3.3).

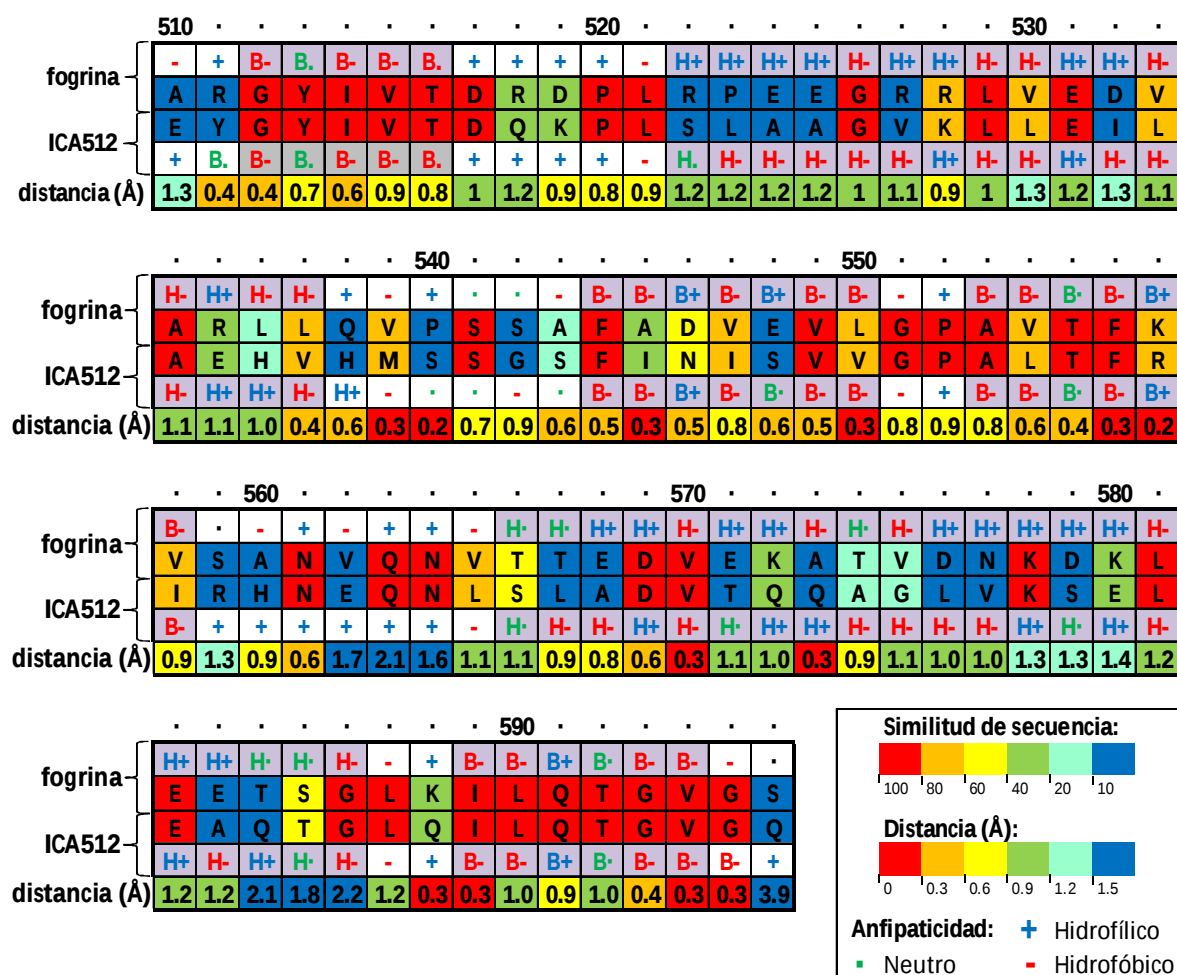


Figura 3.2. Alineamiento estructural de MPE ICA512 y MPE fognina. Se superpusieron las estructuras de ambos dominios con el programa SSM, y se extrajo el alineamiento de secuencias correspondiente. Los residuos se colorearon de acuerdo a su grado de similitud secuencial (definiendo 100% de similitud como los residuos idénticos y 0% a las sustituciones altamente no

conservativas). Se indican los elementos de estructura secundaria, α hélice (H) y cordón β (B). La numeración corresponde a la secuencia de fgrina.

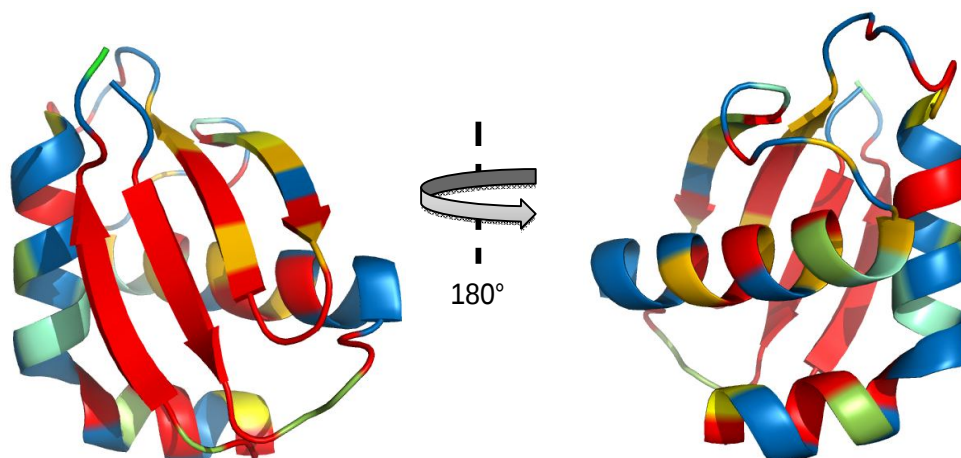


Figura 3.3. Alineamiento estructural de MPE ICA512 y MPE fgrina. Representación de cintas de la estructura de MPE fgrina con el mismo esquema de colores que en la Figura 3.2, ilustrando la disposición de los residuos conservados en los dos dominios.

Para investigar la flexibilidad estructural de los dominios, sus estructuras se superpusieron y se comparó la distribución de los parámetros de desplazamiento atómicos (ADP, también conocidos como factores de temperatura o *B-factors*). Los ADP representan la incertidumbre en la posición de cada átomo dentro del cristal en torno a su posición promedio. Dicha incertidumbre puede provenir del movimiento térmico de cada átomo, y es por esta razón que los ADP resultan importantes indicadores de la dinámica y flexibilidad conformacional de las proteínas. Hay que considerar que los ADP medidos en distintas estructuras pueden estar afectados por diferencias en la recolección y procesamiento de los datos cristalográficos, es por eso que restringimos el análisis a su distribución dentro de cada estructura y no comparamos la magnitud de los ADP entre estructuras distintas.

El análisis de los ADP muestra que la lámina β es la región más rígida, mientras que la más dinámica corresponde a la segunda mitad de $\alpha 2$, justo luego del pliegue (Figura 3.4). Esta región móvil establece pocos contactos con el resto del dominio, y no participa en las interfaces de dimerización de MPE ICA512. Los ADP son mayores en los dos lazos más largos, en especial el que vincula $\beta 3$ con $\alpha 2$, el cual en ICA512 posee tanta movilidad como la segunda mitad de $\alpha 2$. En cambio, en MPE fgrina, los ADP de este lazo son menores que los correspondientes al último segmento de $\alpha 2$.

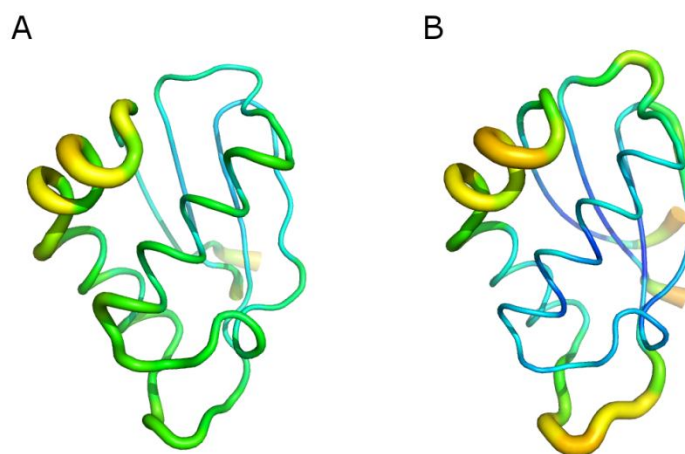


Figura 3.4. Distribución de ADP en MPE ICA512 y fogrina. En esta representación, el diámetro del tubo es proporcional al valor de ADP y el color muestra un gradiente desde los valores mayores de ADP (naranja) a los menores (azul). (A) MPE fogrina. (B) MPE ICA512.

En las estructuras cristalográficas de MPE ICA512, se encontraron dos tipos de interfaces de dimerización (Primo, Klinke *et al.* 2008; Primo, Jakoncic *et al.* 2011). La primera ocurre por la interacción de las hebras $\beta 4$ de cada monómero (y en menor medida también de $\alpha 2$), y la segunda involucra a las hebras $\beta 2$ (y en menor medida $\alpha 1$). Además, el estudio en solución de estas interfaces mediante mutagénesis sitio-dirigida del dominio no glicosilado, demostró que sólo el dímero $\beta 2$ - $\beta 2$ se encuentra significativamente poblado en solución (Primo, Jakoncic *et al.* 2011). Sin embargo, a pesar de la alta similitud estructural y secuencial que comparten, MPE fogrina es un monómero en solución (ver Sección 2.4), y en concordancia se observa sólo una molécula de proteína por unidad asimétrica en el cristal. Esto lleva a preguntarse de inmediato qué características de la estructura de estos dominios determinan la diferencia en su estado de agregación.

Para intentar responder a esta pregunta, se empleó modelado molecular para ubicar los monómeros de la estructura cristalográfica de MPE fogrina en la misma disposición que en los dímeros de MPE ICA512 (Figura 3.5). Existen 3 versiones del dímero $\beta 4$ - $\beta 4$ en ICA512, observados en cristales obtenidos a distinto pH y con diferentes celdas unitarias, por lo que se construyeron los posibles dímeros de MPE fogrina para todos ellos.

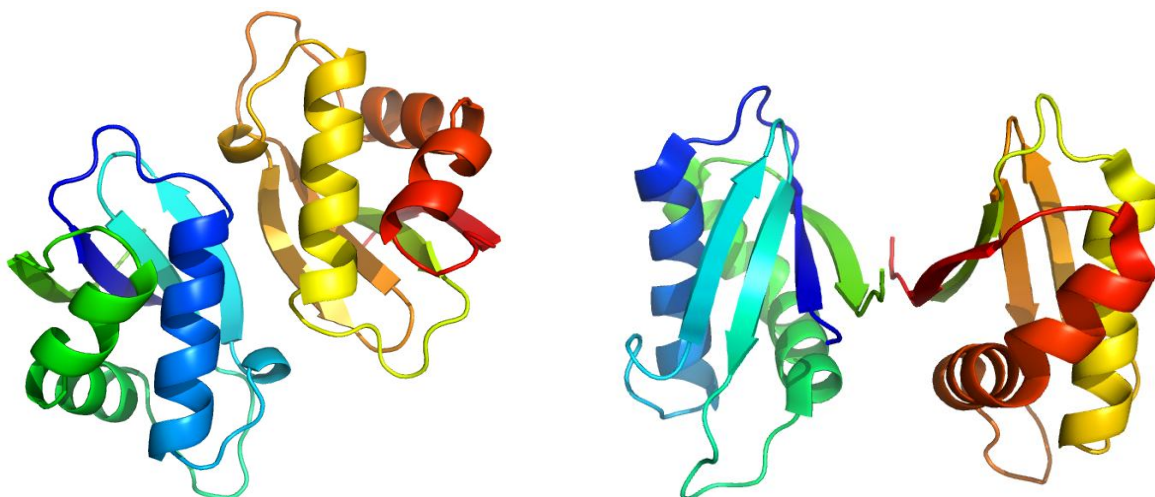


Figura 3.5. Posibles dímeros de MPE fognina generados por modelado molecular. La estructura cristalográfica de MPE fognina se rotó como un cuerpo rígido para imitar las interfaces observadas en la estructura de MPE ICA512 (PDB ID: 2QT7 y 3NP5). A la izquierda, dímero $\beta 2$ - $\beta 2$, y a la derecha, dímero $\beta 4$ - $\beta 4$.

En la evaluación de estos modelos diméricos, es importante considerar que los residuos pertenecientes a la interfaz $\beta 4$ - $\beta 4$ son lo más conservados entre MPE ICA512 y MPE fognina (ver Figura 3.3). El análisis del modelo construido a partir de la interfaz $\beta 4$ - $\beta 4$ de MPE ICA512 a pH 8.5, muestra que este dímero es muy improbable en MPE fognina, ya que los cambios en su secuencia resultan en choques estéricos y repulsiones de carga, que no pueden acomodarse en este arreglo cuaternario sin efectuar cambios mayores en la estructura terciaria. Dos cambios en la secuencia, A568E⁴ y Q572K, son los más importantes para explicar este efecto, ya que implican el remplazo por residuos más voluminosos y con carga. Por otro lado, estos remplazos pueden ajustarse satisfactoriamente a la interfaz $\beta 4$ - $\beta 4$ de ICA512 observada a pH 4.5, que es el modelo que se muestra en la Figura 3.5.

Los residuos involucrados en la interfaz $\beta 2$ - $\beta 2$ están menos conservados, sin embargo, estos cambios pueden ajustarse apropiadamente al modelo de MPE fognina. Hay cuatro cambios en la secuencia particularmente importantes. El primero es el remplazo N546D, lo que impide que esta posición pueda ser N-glicosilada en MPE fognina, y además introduce una repulsión de cargas que resultaría en la desestabilización del dímero. El segundo cambio es la sustitución S548E, que también introduce repulsión de cargas, y además se ha encontrado que el residuo de serina en la posición 548 de ICA512 juega un papel importante en la formación del dímero $\beta 2$ - $\beta 2$ (Primo, Jakoncic *et al.* 2011). El tercero y cuarto cambios son las sustituciones A524E and V527R, que también contribuyen a la repulsión de cargas.

⁴ Se usará la numeración de fognina en toda esta sección, la primera letra indica el residuo correspondiente a MPE ICA512, y la segunda el de MPE fognina.

En resumen, es posible modelar el dímero $\beta 4$ - $\beta 4$ de MPE fognina en un arreglo cuaternario que tiene sentido estereoquímico, sin embargo, se ha encontrado en ICA512 que este dímero es poco estable, y requiere de una fuerza de estabilización adicional para que pueda formarse. En el arreglo cristalino, los responsables de la estabilización de esta interfaz pueden ser las fuerzas de empaquetamiento cristalino, mientras que *in vivo*, puede ser resultado de la interacción de otras regiones del receptor, como los dominios intracelular y/o transmembrana. El dímero $\beta 2$ - $\beta 2$ se encuentra significativamente poblado en solución en ICA512 y puede ser perturbado por mutaciones sitio-dirigidas en su interfaz. Sin embargo, un residuo de asparagina perteneciente a esta interfaz se encuentra glicosilado *in vivo* (Hermel, Dirkx *et al.* 1999), y esta modificación introduce restricciones estéricas que hacen improbable la formación de este dímero *in vivo*. En el caso de MPE fognina, esta interfaz de dimerización se ve impedida, no por glicosilación, sino por la acumulación de un número de cambios en su secuencia que desestabilizan el dímero por repulsión de cargas.

3.4. Patrones de residuos conservados en las secuencias y estructuras de la región MPE de las RPTP

Con el objetivo de identificar las interacciones conservadas que pudieran estar involucradas en la estabilización de la estructura tridimensional, o en la función de la región MPE de las RPTP, se realizó un análisis en profundidad de las secuencias y estructuras disponibles de estos dominios homólogos.

En primer lugar, se emplearon las secuencias de las regiones MPE de ICA512 y de fognina para buscar las secuencias de homólogos en la base de datos de proteínas del NCBI, mediante el procedimiento iterativo de búsqueda PSI-BLAST. Con estas secuencias, se construyó un alineamiento múltiple, en el cual se penalizó fuertemente la aparición de inserciones o eliminaciones dentro de los segmentos que corresponden a los elementos de estructura secundaria observados en las estructuras cristalográficas. Para este fin, se eliminaron del alineamiento inicial aquellas secuencias con extensas eliminaciones en los bloques de estructura secundaria, provenientes principalmente de las isoformas de fognina que carecen del exón correspondiente a la región 519-547 de MPE fognina humana (que comprende $\alpha 1$ y gran parte de $\beta 2$). Además, para asegurar la generación de un alineamiento secuencialmente diverso, con el menor sesgo posible de las secuencias sobrerrepresentadas, se excluyeron las secuencias redundantes, eliminando aquellas por encima de un valor de corte de 90 % de identidad de secuencia. De esta forma, se llegó a un alineamiento final de 36 secuencias, donde las inserciones o eliminaciones se encuentran entre los elementos de estructura secundaria, a excepción de un grupo de inserciones cortas (de uno a tres residuos) que caen dentro de $\alpha 2$, justo en el pliegue característico de esta hélice. En la Figura 3.6 se

secuencia), y se obtuvo la entropía posicional. Esta medida está relacionada con la entropía de la información (o entropía de Shannon), y tiene la ventaja de que las frecuencias de ocurrencia de los aminoácidos se traducen en un valor acotado entre 1 y 20. Un valor de entropía igual a n , es equivalente a la situación donde n residuos aparecen con idéntica frecuencia ($1/n$) en la posición en cuestión. De esta forma, los valores más bajos de entropía se corresponden con las posiciones más conservadas. Una lista de las 25 posiciones más conservadas, de acuerdo a su valor de entropía posicional, se muestra en la Tabla 3.3.

Para intentar explicar el patrón de conservación encontrado, se analizó la ubicación de los residuos conservados en las estructuras tridimensionales de MPE ICA512 y MPE fogrina. Es un hecho reconocido que las posiciones del *core* hidrofóbico están entre las más conservadas. Para identificar los residuos que participan del *core*, se analizó en las dos estructuras el área accesible al solvente (ASA) en el estado nativo, empleando el monómero de cada estructura cristalográfica. Ya que los residuos más voluminosos pueden hacer mayores contribuciones al área total de la proteína, los valores de ASA para cada posición deben normalizarse, comparándolos con los valores estimados para el estado desplegado. En lugar de emplear el procedimiento habitual de asignar un valor de ASA genérico para cada residuo, derivado del valor computado para el aminoácido en un tripéptido Gly-X-Gly (u otro compuesto modelo), se obtuvo una estimación más realista mediante la asignación de los valores de ASA dependientes del contexto de secuencia, que se obtienen mediante simulaciones de mecánica molecular del estado desplegado (Estrada, Bernado *et al.* 2009). De todas formas, el empleo de uno u otro método de normalización no afectó las conclusiones. Las posiciones del *core*, corresponden a aquellas que poseen una baja accesibilidad al solvente (menor al 15 %) y un carácter hidrofóbico conservado. La hidrofobicidad de cada posición corresponde a la frecuencia de cada residuo (teniendo en cuenta el peso estadístico asociado a la secuencia) multiplicada por el valor de hidrofobicidad derivado de la energía de transferencia de los aminoácidos de agua a octanol (Fauchère y Pliska 1983). De esta forma, se identificaron las posiciones del *core* hidrofóbico listadas en la Tabla 3.4, y representadas en la estructura cristalográfica de MPE fogrina en la Figura 3.7.

Tabla 3.3. Residuos conservados en el alineamiento de las RPTP.

Orden	Posición	Consenso	Entropía Posicional	Hidrofobicidad	ASA %	SS^a	RMSD (Å)^b
1	33	G	1.00	0.00	0.06	H	1.04
2	79	F	1.00	1.80	1.88	B	0.26
3	114	G	1.42	-0.08	103.66		2.22
4	65	V	1.43	1.12	36.56	B	0.47
5	125	G	1.51	-0.04	20.59	B	0.44
6	127	G	1.69	-0.01	40.10	B	0.33
7	105	K	1.72	-1.00	58.19	H	1.27
8	84	N	1.76	-0.23	28.34		0.62
9	77	V	2.01	1.20	0.00	B	0.6
10	81	V	2.08	1.23	24.98	B	0.94
11	6	I	2.15	1.42	0.22	B	0.63
12	78	T	2.17	0.25	11.37	B	0.35
13	109	L	2.22	1.63	0.79	H	1.21
14	60	F	2.27	1.74	4.65	B	0.51
15	117	I	2.30	1.46	7.30	B	0.34
16	40	V	2.38	1.42	0.68	H	1.12
17	97	V	2.57	1.12	0.81	H	0.3
18	5	Y	2.57	1.35	32.34	B	0.65
19	45	L	2.59	1.52	17.58	H	0.41
20	101	A	2.65	0.83	0.79	H	0.92
21	98	A	2.71	0.57	17.82	H	1.09
22	4	G	2.80	0.36	2.19	B	0.35
23	80	K	2.82	-0.90	51.93	B	0.24
24	42	A	2.96	0.13	2.23	H	1.05
25	87	N	3.06	-0.39	112.21		1.55

^aEstructura secundaria, H = hélice α , B = cordón β . ^bRMSD entre las coordenadas de MPE fogrina y MPE ICA512.

La mayoría de las posiciones del core hidrofóbico se ubican dentro de los segmentos de estructura secundaria, y además se encuentran muy conservadas, ya que de las 16 posiciones identificadas, 10 se encuentran en regiones de estructura regular, con valores entropía posicional entre 1 y 3. Aunque la posición 8 se encuentra ocupada por un residuo de carácter polar (treonina) en la mayoría de las secuencias, se incluye como parte del *core* ya que el examen de la estructura tridimensional muestra que su grupo metilo se encuentra oculto en el núcleo hidrofóbico de la proteína, mientras que su hidroxilo se encuentra expuesto al solvente. Además, esta posición se encuentra sustituida solo por residuos de carácter hidrofóbico. En este sentido, hay otro residuo de treonina que se encuentra entre las posiciones más conservadas del alineamiento (en la posición 78), pero ya que se encuentra sustituido por serina en algunas de las secuencias, no se considera como parte del *core*.

Tabla 3.4. Residuos del core hidrofóbico de la región MPE de las RPTP.

Orden^a	Posición	Consenso	Entropía Posicional	Hidrofobicidad	ASA %	SS^b
2	79	F	1.00	1.80	1.88	B
9	77	V	2.01	1.20	0.00	B
11	6	I	2.15	1.42	0.22	B
13	109	L	2.22	1.63	0.79	H
14	60	F	2.27	1.74	4.65	B
15	117	I	2.30	1.46	7.30	B
16	40	V	2.38	1.42	0.68	H
17	97	V	2.57	1.12	0.81	H
20	101	A	2.65	0.83	0.79	H
24	42	A	2.96	0.13	2.23	H
28	18	L	3.25	1.59	2.57	
30	47	L	3.30	1.44	12.71	
35	36	L	3.61	1.54	0.01	H
38	8	T	3.93	1.01	2.26	B
44	37	L	4.47	1.30	1.41	H
55	115	L	5.36	1.47	4.70	

^aLas posiciones del alineamiento se ordenaron de acuerdo a su entropía posicional. ^bEstructura secundaria, H = hélice α , B = cordón β .

Se identificaron otras tres posiciones hidrofóbicas conservadas, con entropías posicionales entre 2.1 y 2.7, pero que no se encuentran suficientemente ocultas al solvente para considerarlas parte del core. Son las posiciones 45 y 81, ocupadas siempre por residuos hidrofóbicos, y la 98, que también está ocupada siempre por residuos hidrofóbicos, excepto por tres secuencias donde hay treonina o ácido glutámico. Estas dos excepciones se presentan en ICA512 y en fogrina, por lo que puede analizarse su efecto en las estructuras cristalográficas. En ICA512 se encuentra una treonina, que se orienta de manera análoga a lo observado en la posición 8: el grupo metilo se oculta en el core hidrofóbico, mientras que el hidroxilo se expone al solvente. En MPE fogrina ocurre algo semejante, la cadena carbonada del ácido glutámico establece contactos con el core hidrofóbico, mientras que el grupo carboxilo resulta expuesto. De esta forma, estas tres posiciones pueden clasificarse como integrantes del core “periférico” del dominio.

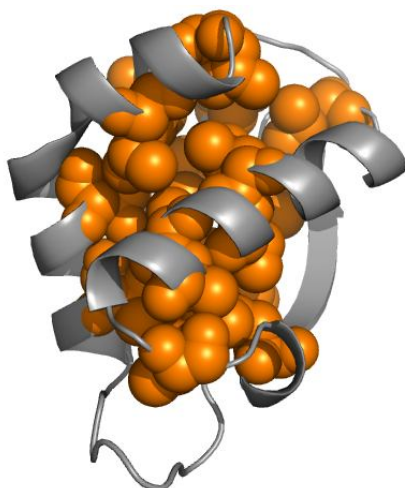


Figura 3.7. Residuos del core hidrofóbico de la región MPE de las RPTP. La estructura cristalográfica de MPE fgrina se muestra en un modelo de cintas, con los residuos del core hidrofóbico mostrados como esferas.

Entre las posiciones más conservadas del alineamiento, se destacan varios lugares ocupados por residuos de glicina, que pueden ser importantes desde un punto de vista estructural. Uno de estos residuos, en la posición 125, se ha estudiado en nuestro laboratorio, mediante una mutación puntual por ácido aspártico en ICA512. Esta variante es capaz de plegarse al estado nativo, sin embargo, su estructura se encuentra marcadamente desestabilizada, con una reducción en la T_m de aproximadamente 15 °C (resultados no publicados). En la posición 4, se observa un predominio de glicinas u otros residuos pequeños. El examen de las estructuras cristalográficas, sugiere que esta posición se encuentra sujeta a una restricción de tipo estérico, ya que está en contacto con una fenilalanina altamente conservada en la posición 79.

Otras dos posiciones conservadas han sido analizadas en nuestro laboratorio por mutagénesis sitio dirigida del dominio MPE ICA512. La posición 5 se encuentra ocupada por residuos hidrofóbicos, predominantemente aromáticos, localizados en una región expuesta al solvente. El remplazado de tirosina por alanina resulta en una variante inestable, con rápida degradación *in vivo* durante la expresión recombinante (resultados no publicados). La posición 64 (con entropía posicional de 4.52) se encuentra ocupada siempre por residuos polares. El remplazo de serina por alanina no afecta la estructura terciaria, pero sí desestabiliza fuertemente al dímero $\beta 2$ - $\beta 2$ (Primo, Jakoncic *et al.* 2011). Como se discutió en la Sección 3.2.3, fgrina posee un residuo de ácido glutámico en esa posición, y aproximadamente la mitad de las secuencias en el alineamiento tienen residuos cargados, por lo que es posible que esta posición sea importante para modular la estabilidad de la estructura cuaternaria.

Por último, la posición 87 está ocupada mayoritariamente por residuos de asparagina, y forma parte del motivo conservado N-X-T/S, característico de los sitios de N-glicosilación.

3.5. Comparación de las estructuras 3D de los SEA

Las estructuras tridimensionales de los dominios SEA tienen en común la disposición espacial de los elementos de estructura secundaria $\beta\alpha\beta\alpha\beta$, característicos del plegado tipo ferredoxina. Como se mencionó en el Capítulo 1, a la fecha se han reportado las estructuras de los dominios SEA de 5 proteínas: las mucinas MUC1 (Macao, Johansson *et al.* 2006) y Muc16 (Maeda, Inoue *et al.* 2004) se determinaron ambas por RMN, la proteasa transmembrana AsP (PDB ID: 2E7V) y las RPTP ICA512 (Primo, Klinke *et al.* 2008; Primo, Jakoncic *et al.* 2011) y fognrina (esta tesis), se determinaron por difracción de rayos x. La estructura del dominio de AsP es muy similar a la de M1SEA, excepto que la región correspondiente a la segunda mitad de $\alpha 2$ y la hebra $\beta 4$ adoptan una estructura irregular en esta última (Figura 3.8). Esto no se relaciona con la formación de una estructura dimérica, ya que hay sólo una molécula de proteína por unidad asimétrica, y la aplicación de operaciones de simetría no reproduce ningún dímero de relevancia. Además, la similitud de secuencia con el resto de los SEA, y las predicciones de estructura secundaria a partir de la secuencia, están de acuerdo con que esta región debiera integrarse en la lámina β de manera análoga a lo observado con M1SEA. Ya que no hay otra información sobre este dominio aparte de la estructura tridimensional, es difícil establecer el significado de esta irregularidad, si es de relevancia biológica o solo un artefacto de la estructura cristalográfica, por esa razón omitiremos a esta estructura del análisis general de los dominios SEA.

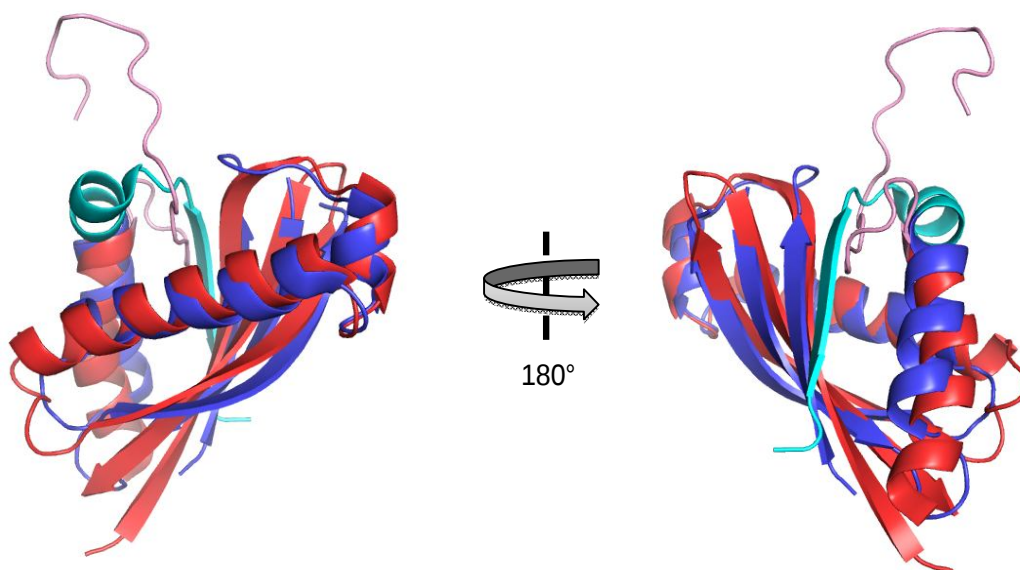


Figura 3.8. Superposición de los dominios SEA de MUC1 y AsP. La estructura tridimensional de AsP (rojo) se superpone a la de M1SEA (azul). Se destaca en magenta al segmento de AsP que no superpone con $\beta 4$ de Muc1 (cian).

Las estructuras de los otros cuatro SEA se muestran en la Figura 3.9. Los dominios MPE ICA512 y MPE fgrina tienen una forma más esférica y compacta, con un radio de giro de 11.94 Å en ambos casos. En cambio las mucinas tienen una forma menos esférica, M1SEA y M16SEA tienen radios de radio de giro de 13.16 y 14.47 Å, respectivamente.

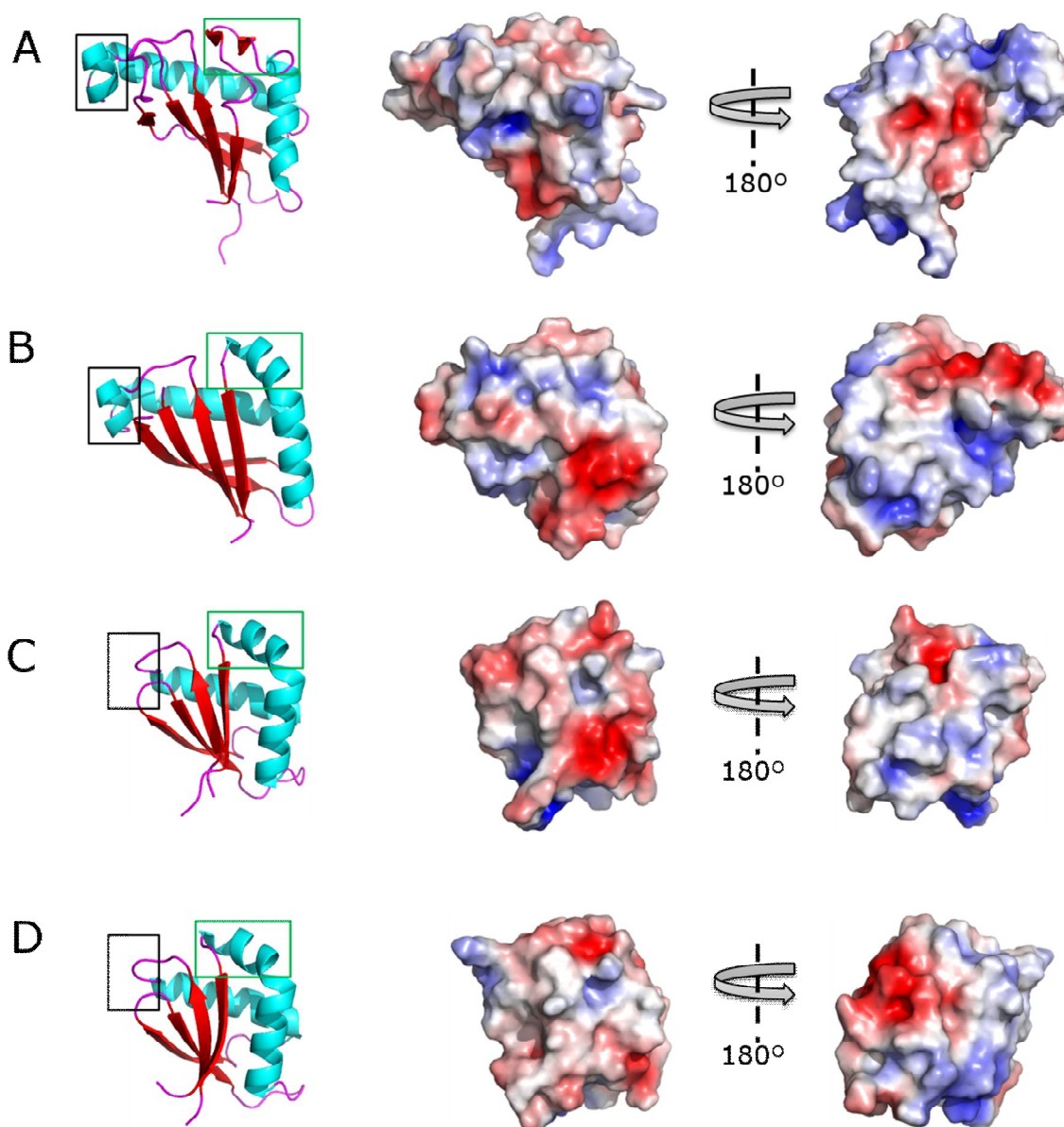


Figura 3.9. Estructuras tridimensionales de los dominios SEA. Se muestran en la misma orientación los dominios: (A) M16SEA. (B) M1SEA. (C) MPE fgrina y (D) MPE ICA512. A la derecha se muestra a cada dominio con la superficie accesible al solvente coloreada de acuerdo al

potencial electrostático, en el panel del medio la orientación es la misma que a la izquierda. Los recuadros señalan las diferencias estructurales entre los SEA que se discuten en el texto.

Aunque presentan una elevada similitud global, las estructuras de las mucinas tienen algunas diferencias respecto de las RPTP (Figura 3.9). En las mucinas, el lazo que conecta $\beta 1$ y $\alpha 1$, posee un pequeño segmento de estructura hélice 3_{10} a la que siguen los residuos que adoptan una conformación llamada “*TY turn*” (pues forman parte del motivo conservado T/S-X-X-Y/F), ambas características están ausentes en las RPTP (recuadro negro en la Figura 3.9). Además, la hélice $\alpha 1$ se encuentra extendida una vuelta más respecto a las RPTP, que sobresale de la lámina β . Esta protrusión de $\alpha 1$ impediría la posible formación de un dímero semejante al $\beta 2$ - $\beta 2$ de MPE ICA512.

M16SEA difiere de las otras estructuras en que no presenta el pliegue característico en $\alpha 2$, en cambio, esta región adopta una estructura irregular con un pequeño segmento de lámina β de dos cordones (Figura 3.9, recuadro verde). Esta región tiene muy pocas interacciones con el resto del dominio.

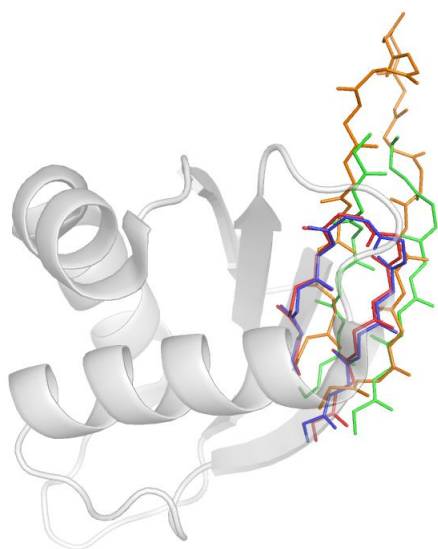


Figura 3.10. Diferencias en el lazo $\beta 3$ - $\beta 4$. Se muestran en detalle los átomos del *backbone* correspondientes a la región $\beta 3$ - $\beta 4$, luego de superponer las estructuras completas de MPE fognina (rojo), MPE ICA512 (azul), M1SEA (verde) y M16SEA (naranja). Como referencia, se muestra la estructura de MPE fognina en un modelo de cintas.

Otra importante diferencia se localiza en la región comprendida entre $\beta 3$ y $\beta 4$, donde se encuentra el sitio de corte autoproteolítico de M1SEA (Figura 3.10). En MPE ICA512 y MPE fognina, la conexión ocurre por medio de un lazo corto, mientras que en M16SEA se presenta un lazo extenso y de gran movilidad, como se ha determinado a partir de experimentos de RMN de relajación $T1/T2$ (Maeda, Inoue *et al.* 2004). Se ha postulado que la presencia de ese lazo flexible en M16SEA podría aliviar las tensiones conformacionales que llevan a la autoproteólisis en M1SEA. Como se demostró en el Capítulo 2, M16SEA es lo

suficientemente estable para catalizar la hidrólisis del enlace peptídico mediante un mecanismo de estrés conformacional, por lo tanto, es de gran interés determinar si este dominio puede también sufrir autoproteólisis si se acorta este lazo para disminuir su flexibilidad. Para este fin, actualmente se están caracterizando en nuestro laboratorio una serie de variantes del dominio M16SEA con eliminaciones en este *loop*.

Se ha propuesto que la función de los dominios SEA es la de establecer interacciones con otras proteínas y componentes de la matriz extracelular. En la Figura 3.9 se muestran para cada dominio los potenciales electrostáticos sobre la superficie accesible al solvente. M1SEA, MPE fogrina y MPE ICA512 muestran “parches” de potencial electrostático negativo (rojo), mientras que M16SEA presenta más áreas de potencial positivo (azul). Estos parches pueden servir como sitios para la interacción con receptores electrostáticamente complementarios y, en el caso de M16SEA, para la interacción con los residuos ácidos de carbohidratos que forman parte de los proteoglicanos de la matriz extracelular.

Adicionalmente, se exploró la exposición de los residuos hidrofóbicos en la superficie de los dominios. Cuando se representa la superficie expuesta al solvente de cada dominio (Figura 3.11), con los residuos coloreados de acuerdo a una escala de hidrofobicidad (Fauchère y Pliska 1983), se observa que tanto M16SEA como MPE fogrina no presentan grandes áreas de residuos hidrofóbicos expuestos al solvente. En cambio, M1SEA y MPE ICA512 exponen superficie hidrofóbica que pueden formar parte de sitios de interacción proteína-proteína. En el caso de MPE ICA512, dicha región hidrofóbica se encuentra oculta al solvente con la formación del dímero $\beta 4$ - $\beta 4$.

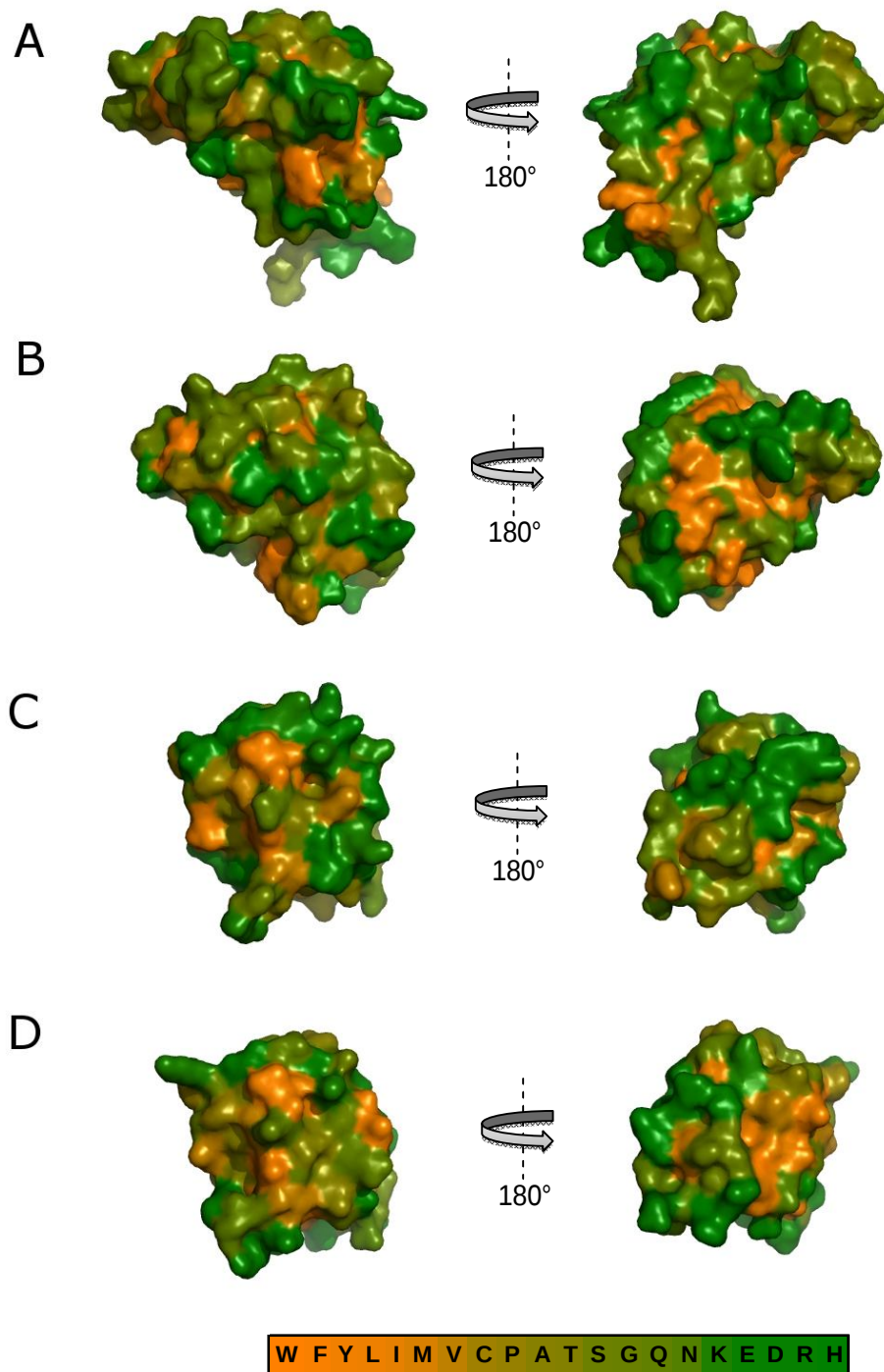
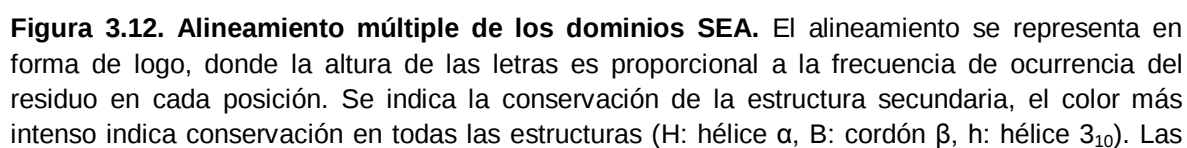


Figura 3.11. Superficie hidrofóbica expuesta. Los residuos se colorearon de acuerdo a la escala de hidrofobicidad de Pliska (Fauchère y Pliska 1983), donde los residuos hidrofóbicos se muestran en una escala de color desde el naranja (hidrofóbico) al verde (hidrofílico). La orientación de los dominios es la misma que en la Figura 3.9. (A) M16SEA. (B) M1SEA. (C) MPE fogrina. (D) MPE ICA512.

3.6. Residuos conservados en la familia de dominios SEA

En la Sección 3.4 se analizaron los residuos conservados entre los homólogos cercanos de las RPTP. Para determinar cuáles de estos son también preservados en todos los SEA, se construyó un alineamiento de toda la familia. La construcción de este alineamiento, donde también se ha favorecido la ubicación de las inserciones y eliminaciones entre los elementos de estructura secundaria, ya ha sido descrita en la literatura (Primo, Klinke *et al.* 2008), y puede consultarse en el APÉNDICE. El alineamiento de todos los SEA se compara con el de las RPTP en la Figura 3.12. Para analizar los patrones de conservación, se ponderó a cada secuencia mediante un peso estadístico de la misma forma que con las RPTP, y luego se calculó la entropía posicional y la hidrofobicidad para cada posición. Como antes, se calculó la accesibilidad al solvente de cada estructura y esto se usó, junto con la hidrofobicidad, para definir las posiciones del *core* hidrofóbico.

En la Tabla 3.5 se listan las 25 posiciones más conservadas de toda la familia SEA. De las 16 posiciones asignadas al *core* hidrofóbico en las RPTP, nueve también pertenecen al *core* conservado de todos los SEA (Tabla 3.6), y su ubicación en las estructuras tridimensionales está conservada (posiciones 77, 8, 6, 37, 117, 60, 97, 115 y 101, Figura 3.13). Otras dos posiciones también forman parte del *core* en todas las estructuras, pero su localización exacta difiere en la estructura tridimensional. Dado que M16SEA no posee el segundo segmento de $\alpha 2$ (luego del pliegue en las otras estructuras), la posición 109 se ubica en una región estructuralmente diferente. También la posición 18, a pesar de ser la más conservada en el alineamiento, se encuentra en un contexto estructural diferente, ya que en las mucinas se ubica en la región de hélice 3_{10} que está ausente de las RPTP. Finalmente, otros residuos que forman parte del *core* de las RPTP adoptan roles diferentes en el resto de los SEA, estos son las posiciones 36, 40, 42, 47 y 79, que no forman parte del *core* conservado ya que se encuentran sustituidos por residuos polares en un número de secuencias.



posiciones recuadradas deben interpretarse con cautela ya que corresponden a regiones con gaps en más del 20% de las secuencias.

Tabla 3.5. Residuos conservados en el alineamiento de los SEA.

Orden	Posición	Consenso	Entropía Posicional	Hidrofobicidad	ASA %	SS^a
1	18	Y	4.11	1.68	16.23	
2	65	F	4.12	1.57	24.71	B
3	62	V	4.23	1.02	57.31	
4	77	V	4.77	0.83	0.31	B
5	30	Y	5.02	1.37	50.96	H
6	8	I	5.07	1.23	1.72	
7	6	F	5.08	1.48	0.51	B
8	37	I	5.49	1.45	1.50	H
9	75	V	5.55	1.06	21.57	
10	45	F	5.57	1.58	10.85	H
11	33	L	6.43	1.17	6.49	H
12	4	G	6.72	0.85	1.43	B
13	81	V	7.04	1.17	12.49	B
14	117	I	7.07	1.07	8.77	
15	109	L	7.17	0.91	0.77	
16	126	V	7.30	0.86	25.45	B
17	83	F	7.36	1.24	71.78	
18	123	S	7.73	0.00	70.98	
19	10	N	7.73	-0.24	53.78	
20	19	S	7.90	0.09	64.18	
21	66	R	8.11	-0.39	70.12	B
22	67	P	8.34	-0.18	86.62	
23	34	A	8.70	0.02	39.17	H
24	60	V	8.72	0.88	13.24	B
25	79	F	9.06	1.22	2.44	B

^aEstructura secundaria, H = hélice α , B = cordón β .

Otras cuatro posiciones conservadas en todos los SEA, que son también hidrofóbicas y de baja accesibilidad al solvente, no fueron asignadas al *core* de las RPTP. Dos de ellas, 45 y 81, habían sido clasificadas como del “*core* periférico”, ya que presentaban valores de ASA entre 18 y 25 %. Sin embargo, su carácter hidrofóbico está conservado y están completamente ocultas al solvente en M1SEA y M16SEA, por lo que resulta apropiado asignarlas al *core*. Las otras dos posiciones, 4 y 33, también estaban conservadas y ocultas en las RPTP, pero al estar ocupadas por residuos pequeños, especialmente de glicina, no fue posible inferir su rol estructural. Notablemente, se observa que la posición 4 está ocupada por

residuos hidrofóbicos dentro del *core*, siempre que en la posición 79, que se encuentra espacialmente cercana, no ocurran residuos aromáticos como se observa frecuentemente en las secuencias de las RPTP.

Tabla 3.6. Residuos del core de los SEA.

Orden^a	Posición	Consenso	Entropía Posicional	Hidrofobicidad	ASA %	SS^b
4	77	V	4.77	0.83	0.31	B
6	8	I	5.07	1.23	1.72	
7	6	F	5.08	1.48	0.51	B
8	37	I	5.49	1.45	1.50	H
10	45	F	5.57	1.58	10.85	H
11	33	L	6.43	1.17	6.49	H
12	4	G	6.72	0.85	1.43	B
13	81	V	7.04	1.17	12.49	B
14	117	I	7.07	1.07	8.77	
15	109	L	7.17	0.91	0.77	
24	60	V	8.72	0.88	13.24	B
26	97	V	9.22	0.83	1.44	H
31	115	L	9.58	1.06	12.71	
33	101	L	9.70	0.87	0.87	H

^aLas posiciones del alineamiento se ordenaron de acuerdo a su entropía posicional. ^bEstructura secundaria, H = hélice α , B = cordón β .

En resumen, han podido identificarse residuos de importancia estructural, que forman parte del *core* hidrofóbico estabilizando la conformación nativa. Aun así, muchos de los residuos conservados en los SEA no presentan patrones evidentes que justifiquen su conservación. Por ejemplo las posiciones 65 y 62 son la segunda y tercera más conservadas, y son hidrofóbicas pero expuestas al solvente, por lo que podrían estar involucradas en interacciones intermoleculares. En definitiva, puede que muchos de estos residuos conservados sean importantes para determinar la ruta de plegado, y también para la desestabilización de conformaciones alternativas. Además, es posible que algunos de ellos sean importantes desde un punto de vista funcional. Es por ello que este “inventario” de residuos conservados constituye una herramienta valiosa para la elaboración de estrategias experimentales tendientes a esclarecer nuevos aspectos de la estructura y función de los SEA.

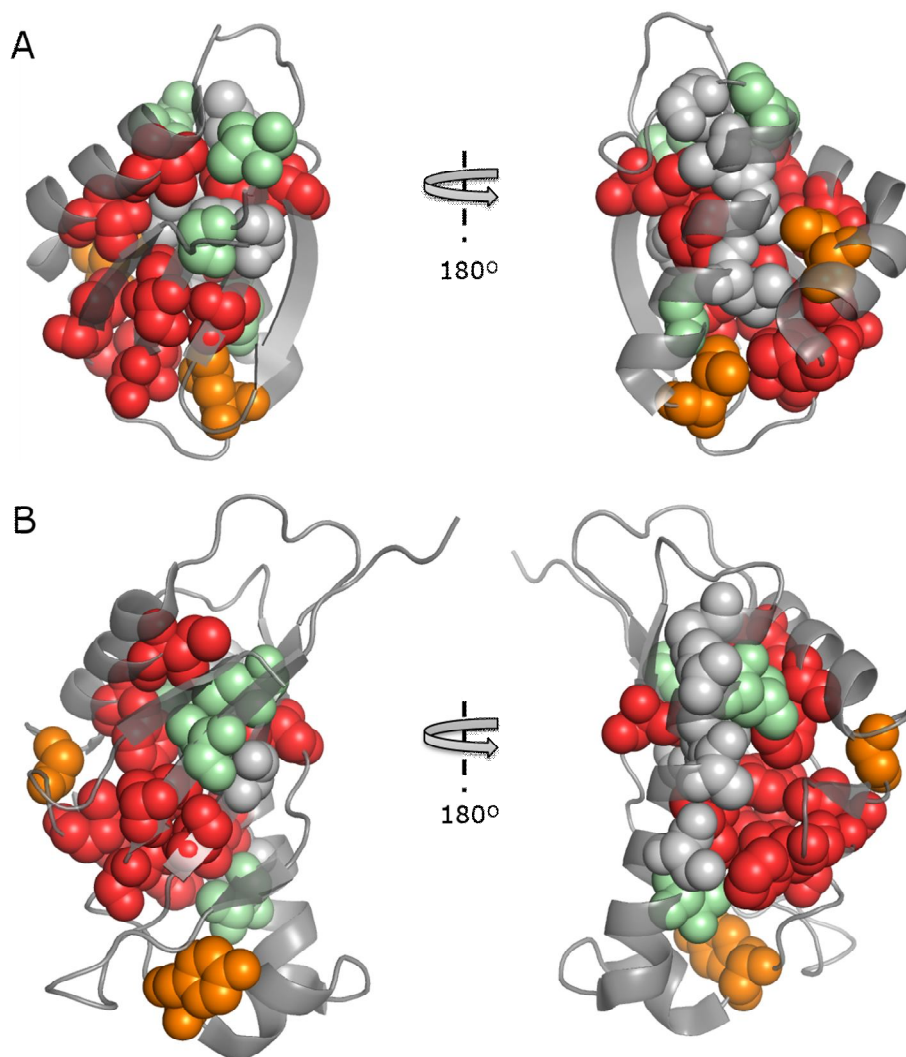


Figura 3.13. Conservación de los residuos del core. (A) MPE fogrina. (B) M16SEA. En rojo, se muestran los residuos que pertenecen al core de las RPTP, y también al de todos los SEA. En gris, los residuos que pertenecen al core de las RPTP, pero no se pueden generalizar a todos los SEA. En naranja, residuos del core de los SEA, que se ubican en contextos estructurales distintos en las diferentes estructuras, y en verde, las posiciones que se pueden asignar al core de los SEA, luego de analizar el patrón de conservación y las estructuras de todos los SEA.

4. Discusión general y conclusiones

Los dominios SEA fueron reconocidos como una familia de proteínas a mediados de los '90, a partir del estudio de sus secuencias (Bork y Patthy 1995). Se encontró que eran conspicuos, y que formaban parte de la región extracelular de numerosas proteínas, muchas de ellas de membrana, y a menudo interactuando con la matriz extracelular, o incluso formando parte de ella. Nada se sabía acerca de sus funciones específicas. Permanecieron en un misterio durante mucho tiempo, hasta que se obtuvo la primera estructura tridimensional, perteneciente a la mucina Muc16 (Maeda, Inoue *et al.* 2004). A esto le siguió la determinación por RMN del dominio de la mucina MUC1 (Macao, Johansson *et al.* 2006), y la estructura cristalográfica del SEA de la proteasa AsP. Más recientemente, se resolvió la estructura cristalográfica del dominio MPE ICA512 (Primo, Klinke *et al.* 2008), en el laboratorio donde se llevó a cabo esta tesis.

Dado que los dominios SEA son tan diversos desde un punto de vista secuencial, nuestro conocimiento de los mismos, basado en tan pocos ejemplos caracterizados, podría considerarse al menos como endeble. Existía entonces la necesidad de, por un lado, extender el estudio estructural a otros miembros de la familia, y por otro, de realizar estudios comparativos sistemáticos entre los miembros de la misma. La estrategia adoptada para el desarrollo de esta tesis consistió en la determinación de la estructura tridimensional del dominio SEA de fognina y la caracterización en paralelo, desde un punto de vista bioquímico y biofísico, de esta última junto con otros miembros de la familia pertenecientes a las proteínas MUC1, Muc16 e ICA512. Esta estrategia permitió ofrecer el primer cuerpo consistente de datos e información experimental comparativa de una muestra significativa de las proteínas SEA.

4.1. Características generales de los dominios SEA

Los estudios en esta tesis sobre cuatro dominios SEA, confirman que la unidad de plegado autónomo está constituida por el empaquetamiento de dos hélices α sobre una lámina β de cuatro cordones antiparalelos, en el orden $\beta\alpha\beta\alpha\beta$. No se requiere de la formación de contactos con otras regiones contiguas de la proteína para la estabilización de la estructura nativa. Esto se evidencia en primer lugar, en sus propiedades de expresión heteróloga: los dominios aislados pudieron obtenerse con altos rendimientos y correctamente plegados en todos los casos. La delimitación de un módulo en base a la homología de secuencia no siempre conduce a una proteína estable, por ejemplo, en el caso de los dominios KH, las características del dominio se definieron en base a la conservación de secuencia (Gibson, Thompson *et al.* 1993). Los intentos de expresar el sexto módulo KH de la

proteína *vigilin* fracasaron al obtenerse variantes muy inestables que eran rápidamente degradadas (Castiglione Morelli, Stier *et al.* 1995). En estudios posteriores, se encontró que una región adyacente en el extremo C, con alta propensión a formar α hélice, era esencial para la estabilización del dominio ya que contribuía a la formación del *core* hidrofóbico. De esta forma, fue necesario cambiar la definición original del módulo, para incluir a todos los elementos que contribuyen a la unidad de plegado autónomo (Musco, Stier *et al.* 1996).

Un aspecto adicional en el estudio de los dominios aislados, consiste en definir los límites precisos del dominio para una proteína en particular, ya que se ha encontrado que esto puede tener un fuerte impacto sobre sus propiedades estructurales. Por ejemplo, a partir del estudio estructural del dominio Tnf3 de tenascina, se encontró que cuando éste es acortado en tan solo dos residuos del extremo C, se remueven interacciones de largo alcance que tienen un marcado efecto sobre su estabilidad y dinámica conformacional (Hamill, Meekhof *et al.* 1998). Durante el desarrollo de esta tesis, la definición precisa de los límites del dominio representó un desafío para la región extracelular de fognina, ya que su estructura tridimensional se desconocía, y porque se han reportado múltiples isoformas de fognina, que dan lugar a distintos extremos N (Tomiko Kawakami 2007). Empleando la información del procesamiento proteolítico y la similitud de secuencia con ICA512, se diseñó una construcción que abarcara al dominio SEA completo, a la que denominamos MPE fognina. Los estudios biofísicos realizados en esta tesis, demostraron que MPE fognina posee las características de un dominio bien plegado, y constituyen la primera evidencia experimental de que fognina posee una región de plegado autónomo en su ectodominio.

Se distinguen dos características destacadas en las secuencias de los SEA. Una de ellas es la secuencia de corte proteolítico, típicamente GSVVV, y la otra es la presencia de un par conservado de residuos de cisteína. Los dos tipos de motivos están presentes en un gran número de secuencias de la familia, pero se ha observado que no se presentan simultáneamente en una misma secuencia. Esto llevó a proponer que mientras algunos dominios SEA evolucionaron para escindirse y formar un heterodímero, otros lo hicieron para ser estables con la formación de un enlace disulfuro (Macao, Johansson *et al.* 2006). En esta tesis se demostró que las cisteínas de M16SEA no forman un enlace disulfuro, y esto tiene una explicación estructural, ya que a pesar de que las cisteínas se encuentran muy próximas entre sí, existen distorsiones en la cadena principal que impiden que se adopten los rotámeros necesarios para la formación del enlace. En un estudio previo tampoco se había detectado la formación de un disulfuro en M16SEA (Maeda, Inoue *et al.* 2004), en esta tesis demostramos que el dominio es muy estable, aún en ausencia del entrecruzamiento por un disulfuro.

4.2. Autoproteólisis de los dominios SEA

La autoproteólisis de los SEA es considerada la característica más notable de estos dominios, ya que puede ser vista como un mecanismo por el cual un mismo gen codifica tanto al receptor como a su ligando específico (Wreschner, McGuckin *et al.* 2002). Los resultados de esta tesis sugieren que la autoproteólisis es un fenómeno general o al menos presente en muchos miembros de la familia. Existen variaciones en el mecanismo de autoproteólisis, ya que en los SEA que poseen la secuencia consenso de corte, el mecanismo está mediado por estrés conformacional, mientras que en la región MPE de las RPTP, es resultado de la susceptibilidad a la degradación por especies reactivas de oxígeno en la cercanía de los extremos del dominio. Hay un número importante de dominios SEA que carecen de la secuencia consenso de corte y para los cuales no se ha reportado que sufran procesamiento proteolítico alguno. Este es el caso del dominio M16SEA y en general de todos aquellos que poseen el par de cisteínas conservado. Futuros estudios deberían demostrar si estos dominios han sufrido modificaciones que impiden el procesamiento proteolítico, es decir, que se han especializado en ser monoméricos, o si por el contrario, son los dominios que sufren autoproteólisis los que se han especializado para formar heterodímeros.

Estudios previos del laboratorio demostraron que el ectodominio maduro de ICA512 sufre un proceso de degradación *in vitro* que resulta en la generación del dominio MPE (Primo, Klinke *et al.* 2008). Los experimentos realizados en esta tesis, con una versión extendida de MPE fgrina, mostraron que la misma también es degradada *in vitro*. Actualmente se están realizando experimentos adicionales, que permitirán establecer si el mecanismo proteolítico es el mismo en ambos dominios, lo que es sugerido por los resultados mostrados en la Sección 2.9, donde se observó que la degradación depende de la presencia de metales y es inhibida por DMSO, un conocido atrapante de especies reactivas de oxígeno. ¿Es este procesamiento proteolítico de relevancia *in vivo*, en condiciones normales o patológicas? En este sentido es importante destacar que el extremo N de MPE fgrina se eligió de manera tal que coincidiera con la isoforma predominante en páncreas (Tomiko Kawakami 2007). La búsqueda bioinformática de proteasas que pudieran generar ese patrón de corte, no arrojó resultados positivos entre las enzimas propias del tejido neuroendocrino, a diferencia de otras isoformas de fgrina y de ICA512 cuyos patrones de corte coinciden con los generados por las enzimas prohormona convertasas que son propias de este tipo de tejidos. Futuros experimentos deberán ocuparse de establecer si esta isoforma de fgrina es resultado del mismo procesamiento observado *in vitro*. Por otro lado, actualmente se reconoce que el estrés oxidativo juega un rol importante en la patología diabética (Kajimoto y Kaneto 2004), por lo que resulta de gran interés establecer si este proceso de degradación puede darse *in vivo* y cuáles son sus consecuencias.

4.3. Propiedades estructurales de los dominios SEA

Un mecanismo habitual en las rutas de transducción de señales a través de la membrana plasmática, es la oligomerización de los receptores de la superficie celular en respuesta a la unión de un agonista, con su consecuente activación. Por ejemplo, este es el mecanismo empleado por muchos receptores tirosina quinasa y receptores acoplados a proteína-G (GPCR) (Hebert y Bouvier 1998). En algunos casos, la unión del agonista induce la dimerización de los receptores, o la estabiliza, mientras que en otros, la dimerización es un prerequisite para la unión del agonista (Rios, Jordan *et al.* 2001). Un gran número de dominios SEA forman parte de la región extracelular de receptores de membrana, como es el caso del receptor GPR116 y de las RPTP ICA512 y fgrina. Lo mismo sucede con la mayoría de las mucinas unidas a la membrana, ya que además de su función clásica en la protección de la superficie epitelial, juegan un rol importante en rutas de señalización a través de su región intracitoplasmática. Por esa razón, comprender el potencial de oligomerización de estos dominios es importante para explicar la función de las proteínas en las que están presentes. Los resultados de esta tesis, muestran que el estado de agregación de los SEA es particular de cada dominio, lo que es comúnmente observado en otras familias proteicas.

Se ha propuesto que los procesos de homo y heterodimerización son importantes mecanismos regulatorios de la función de las RPTP (Chin, Sachs *et al.* 2005). La formación de heterodímeros puede explicar que la coexpresión de ICA512 o de fgrina, junto con RPTP α , redujo la actividad de esta última (Gross, Blanchetot *et al.* 2002). En los estudios estructurales de MPE ICA512, se demostró que este dominio forma dímeros a través de dos tipos de interfaces de dimerización. El hallazgo de que MPE fgrina, un homólogo cercano de ICA512 que comparte muchas de sus propiedades funcionales, es un monómero en las mismas condiciones, resultó en principio sorprendente. La resolución de la estructura tridimensional de MPE fgrina, permitió racionalizar esta diferencia, ya que algunos cambios en su secuencia afectan a las posibles interfaces de dimerización, lo que resultaría en la desestabilización de las interfaces $\beta 2$ - $\beta 2$ y $\beta 4$ - $\beta 4$ encontradas en ICA512 a pH 8.5. Esto es resultado de que los cambios en la secuencia introducen repulsiones de carga e incompatibilidades estéricas, que no pueden ajustarse satisfactoriamente a la estructura dimerica, sin alterar sustancialmente a la estructura terciaria. Estas sustituciones son, sin embargo, compatibles con el dímero $\beta 4$ - $\beta 4$ de ICA512 encontrado a pH ácido, pero el dímero $\beta 4$ - $\beta 4$ se encuentra poco poblado en solución. Aun así, este dímero podría formarse si resultara estabilizado por la interacción de otras partes del receptor, por lo que se requieren experimentos adicionales para demostrar si el mismo se forma *in vivo*.

Como se describe en la Sección 2.4, M16SEA y M1SEA no oligomerizan, y eso se puede explicar porque algunas diferencias estructurales con MPE ICA512 resultan en incompatibilidades geométricas con las formas diméricas de esta última.

Los resultados de esta tesis son consistentes con la noción de que la estructura cuaternaria no contribuye significativamente a la estabilización del estado nativo de MPE ICA512. Los experimentos de desplegamiento químico en el equilibrio, mostraron transiciones monofásicas que no dependen de la concentración de proteína. Adicionalmente, durante el desplegamiento de un mutante puntual de este dominio (S508A), que es esencialmente monomérico, se observan transiciones idénticas a las de la proteína salvaje (resultados no publicados). Esto se puede explicar asumiendo que la disociación es silenciosa a las sondas espectroscópicas empleadas, algo que también ha sido observado con otros dímeros, como la creatinina quinasa (Fan, Zhou *et al.* 1998) y la proteína Ure2 (Perrett, Freeman *et al.* 1999). Una evidencia adicional de que la disociación no altera la estructura del monómero, proviene del hecho de que MPE fognina es un monómero, mientras que posee una estructura terciaria casi idéntica a la de MPE ICA512. Por lo tanto, se puede concluir que las especies dimérica y monomérica de MPE ICA512 coexisten en solución con estructuras tridimensionales muy similares.

Aunque las fuerzas que estabilizan la conformación proteica están ampliamente caracterizadas, de qué forma éstas se traducen en la conformación nativa, es un interrogante que aún no posee una respuesta satisfactoria. Las diferencias de estabilidad entre proteínas estructuralmente relacionadas, parecen ser resultado de la acumulación de un número de sutiles cambios en su estructura, tales como el número de interacciones electrostáticas, diferencias en la extensión de estructura no regular, o la estabilización local de hélices α (Szilagyi y Zavodszky 2000; Kumar, Tsai *et al.* 2001). En este trabajo se encontró que tres de los dominios SEA siguen un mecanismo de desplegamiento en el equilibrio de dos estados. Este mecanismo también puede explicar la estabilidad “relevante” de MPE ICA512, ya que la dimerización no contribuye a la estabilización de la estructura terciaria. Se observaron diferencias importantes en la estabilidad de estos dominios. El orden de estabilidad conformacional sigue la misma tendencia que el contenido de residuos aromáticos de las secuencias, un importante factor que contribuye a la estabilidad de las proteínas: M1SEA es la más estable ($\Delta G = 17.9$ kcal/mol, 15 % de residuos aromáticos), seguida de M16SEA ($\Delta G = 10.2$ kcal/mol, 11 %), luego MPE ICA512 (5.7 kcal/mol y 4.4 %), y finalmente MPE fognina (2.3 kcal/mol y 3%),

Una razón adicional para estudiar la estabilidad de los SEA, es que en algunos miembros de la familia el proceso de plegamiento proteico está acoplado a la autoproteólisis. Como se describe en la Sección 1.6, experimentos cinéticos y estudios teóricos permitieron estimar que la energía libre de plegamiento contribuye en aproximadamente 7 kcal/mol a la

escisión del enlace peptídico en forma de estrés conformacional. M16SEA es lo suficientemente estable para superar esta barrera, por lo que es razonable preguntarse por qué esta molécula no experimenta autoproteólisis. El examen de la estructura tridimensional de este dominio, muestra que la región equivalente al sitio de proteólisis de M1SEA presenta un lazo extenso y poco estructurado formado por residuos polares. Ya que se ha demostrado que la inserción de residuos de glicina en el sitio de proteólisis de M1SEA inhibe la autoproteólisis, por el aumento de la flexibilidad y consecuente reducción del estrés conformacional, una posible explicación es que de manera análoga, el lazo desestructurado de M16SEA impide la autoproteólisis al reducir el estrés conformacional. Para corroborar esta hipótesis, hemos diseñado en nuestro laboratorio un conjunto de mutantes de M16SEA con acortamientos en este lazo y mutaciones en los cordones $\beta 2$ y $\beta 3$, que los hacen más parecidos a los de M1SEA. En experimentos preliminares con estos mutantes, no se detectó proteólisis *in vivo* en los extractos de *E. coli*, y se observó que algunas de estas variantes se encuentran marcadamente desestabilizadas. Una caracterización en mayor profundidad de las propiedades estructurales de estos mutantes está actualmente en curso.

En el marco de esta tesis doctoral, se comparó un conjunto muy diverso de secuencias y estructuras tridimensionales de los dominios SEA, lográndose identificar un conjunto de interacciones conservadas que constituyen el *core* hidrofóbico del dominio. Este resultado está de acuerdo con estudios similares en otras familias proteicas, como las globinas, inmunoglobulinas y dominios SH3 (Bashford, Chothia *et al.* 1987; Chothia, Gelfand *et al.* 1998; Larson y Davidson 2000), que muestran que las posiciones hidrofóbicas ocultas en la estructura tridimensional están altamente conservadas. Es importante destacar que en nuestro análisis de los SEA se encontraron otras posiciones conservadas que no están involucradas en la formación del *core*, ni tienen un rol estructural evidente. Por esta razón, el análisis realizado en esta tesis constituye un punto de partida para estudios más detallados tendientes a establecer cuáles de estas posiciones están involucradas en la ruta de plegado, y si algunas de ellas podrían estar asociadas a aspectos funcionales comunes de todos los miembros de la familia SEA.

4.4. Conclusiones

Las conclusiones generales de estos estudios son:

1. Los dominios SEA adquieren una estructura tridimensional muy similar, a pesar de tener una similitud de secuencia muy baja.

2. El dominio SEA está formado por un arreglo $\beta\alpha\beta\beta\alpha\beta$, y estos son los únicos elementos necesarios para la estabilización de la estructura.
3. La estabilidad de los dominios SEA no depende de la formación de enlaces disulfuro.
4. El estado de agregación no es una propiedad común a todos los SEA. Cada miembro de la familia necesita ser estudiado para definir su estado de agregación. El ensamblado de los SEA en un arreglo cuaternario necesita ser estabilizado por una fuerza externa, como la inserción en la membrana plasmática, o la interacción entre otros módulos de la misma proteína, o bien por la unión de otra proteína para formar un hetero-oligómero.
5. La estabilidad conformacional es muy variable entre los distintos dominios SEA. Su estabilidad no es un reflejo de aspectos funcionales comunes a toda la familia, sino más bien de las particularidades fisiológicas de cada miembro. Se conoce que el tiempo de vida de una proteína está determinado por la relación entre sus velocidades de síntesis y de degradación, lo que está ajustado a los requerimientos de la célula. La estabilidad conformacional tiene una gran influencia sobre esta velocidad de recambio. Creemos que la mayor estabilidad de los SEA de las mucinas está relacionada con su rol estructural en la protección de los epitelios, lo exige una alta estabilidad para hacer frente al estrés mecánico y, como es el caso de otras proteínas estructurales, está asociado a una baja tasa de recambio. Por el contrario, los procesos de secreción celular en respuesta a estímulos requieren de una regulación rápida y precisa, para lo cual es necesario una estabilidad conformacional más baja, como es el caso de los SEA de las RPTP.
6. La estructura de MPE fognina es robusta a los cambios de pH que ocurren durante el ciclo de vida de los gránulos de secreción. La estructura terciaria es idéntica y monomérica a pH 7.4 y 4.6. Por otro lado, el estudio cristalográfico de MPE ICA512 también indica que su estructura terciaria es muy similar a pH 4.5 u 8.5, aunque se forman distintas interfaces de dimerización dependiendo de las condiciones de cristalización. Los dominios MPE fognina y MPE ICA512 experimentan una acidificación a pH $\sim$ 4.6 con la maduración del gránulo, y luego llegan hasta pH $\sim$ 7.4 con la exposición a la membrana celular, pero estos cambios de pH no tienen efecto sobre su estructura terciaria.
7. El procesamiento autoproteolítico es una característica general, o al menos altamente extendida entre los miembros de la familia SEA, aunque el mecanismo exacto por el que ocurre este fenómeno es variable.
8. El plegado característico de los dominios SEA puede alcanzarse por medio de secuencias muy diferentes. Aparte de las interacciones conservadas que definen el *core* hidrofóbico, no se encontraron otros elementos comunes que definan las características del plegado. Esto conduce a una gran variabilidad en las propiedades de la superficie, tales como la distribución de cargas y de parches hidrofóbicos. De esta forma, parece que la ubicuidad de los SEA es resultado de su capacidad de ser un andamiaje robusto, que puede

soportar grandes variaciones de secuencia para exhibir un amplio rango de propiedades en su superficie. Es decir, el dominio SEA constituye un marco adecuado para una gran diversidad de funciones.

4.5. Proyecciones de esta tesis y futuros experimentos

Uno de los objetivos generales de la línea de investigación en la que se encuadra esta tesis, es la obtención de un modelo de los receptores completos ICA512 y fognina en la membrana celular. Como se mencionó en la Sección 4.3, es muy probable que estos receptores formen dímeros en la superficie celular, aun cuando la región MPE no sea determinante para la estabilidad del dímero. En la construcción de los modelos mencionados, resulta entonces esencial considerar la geometría de la interacción con la membrana, pues ambos extremos C del dímero debieran penetrar en la membrana, y esto se traduce en importantes restricciones geométricas. Desafortunadamente, no hemos podido obtener información estructural sobre el segmento que conecta a la región MPE con la región transmembrana, ya que tanto ICA512 como fognina experimentan procesos de degradación *in vitro*. Por esta razón, nuestro conocimiento actual de la topología del receptor en la membrana, es resultado de estudios de modelado molecular, que muestran que el arreglo más parsimonioso se consigue mediante la interacción $\beta 4$ - $\beta 4$ (Figura 4.1). Habiendo observado que la degradación de ICA512 y de fognina resulta inhibida al secuestrar los metales del medio y por el agregado de atrapantes de ROS, actualmente se está investigando si es posible cristalizar a estos dominios extendidos en el extremo C, en condiciones que inhiben la degradación.

Por otro lado, el estudio de proteínas con la misma estructura y secuencias diferentes, ha hecho importantes contribuciones a nuestro conocimiento actual del plegado proteico. El descubrimiento en los años '90 de que la topología nativa es el principal determinante de la ruta de plegado proteico (Plaxco, Simons *et al.* 1998), llevó a postular que las proteínas con la misma estructura deberían plegarse siguiendo una misma ruta (Baker 2000). Estudios posteriores, demostraron que esta simplificación es excesiva, ya que muchas proteínas, aún con alta similitud secuencial, pueden alcanzar el estado nativo siguiendo rutas diferentes (Zarrine-Afsar, Larson *et al.* 2005). Actualmente, se considera que existen dos tipos de escenarios. Por un lado, algunas topologías poseen fuertes restricciones para alcanzar el estado nativo, lo que lleva a que sigan la misma ruta de plegado aun aquellos miembros cuyas secuencias son muy diferentes. Otras topologías no poseen restricciones tan fuertes, por lo que incluso proteínas con gran identidad de secuencia pueden plegarse siguiendo rutas

muy diferentes. La caracterización de las rutas de plegado requiere de análisis cinéticos sobre un gran número de mutantes puntuales, por lo que el número de familias analizadas en este aspecto es relativamente pequeño hasta la fecha. Por esta razón, aun no ha podido establecerse cual de los dos tipos de comportamiento es el más general. Se espera que pueda alcanzarse una respuesta a este interrogante conforme se aumente el número de familias analizadas. Los dominios SEA constituyen buenos candidatos para este tipo de estudio, ya que en ellos puede construirse una estructura muy similar a partir de secuencias muy diferentes.

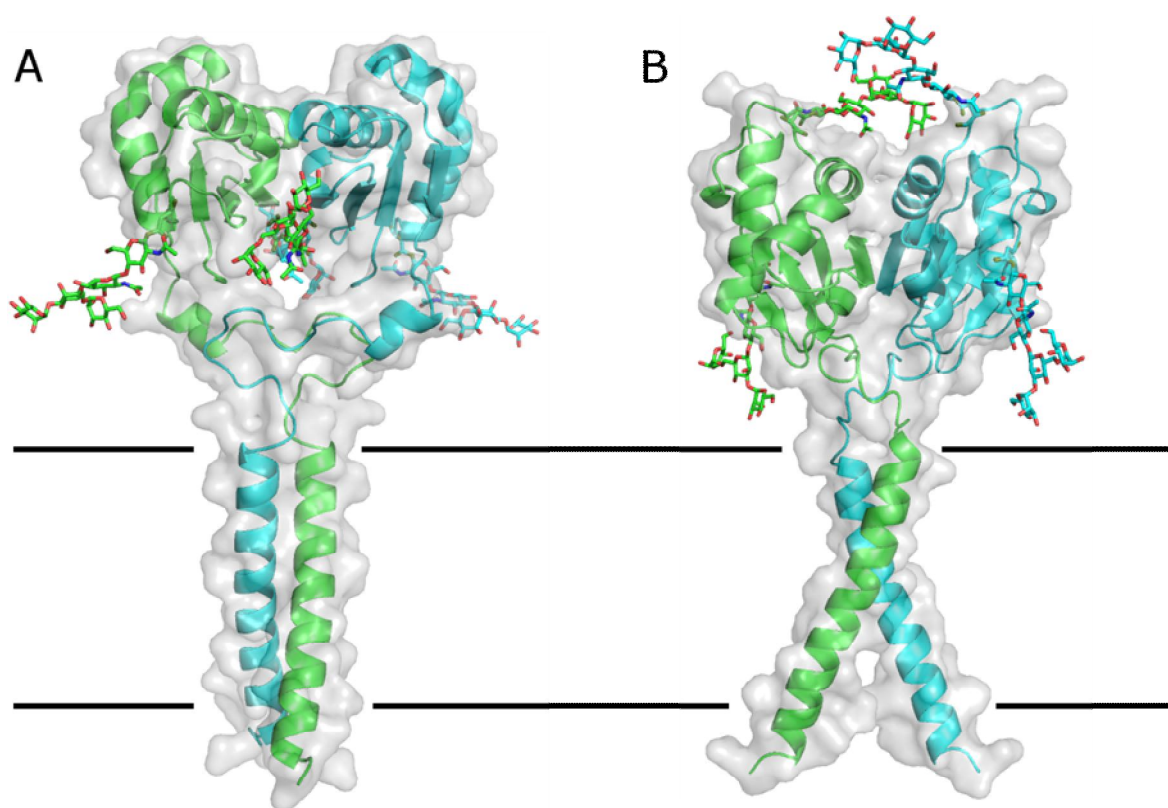


Figura 4.1. Modelos de la proteína completa ICA512 en la membrana. (A) Dímero $\beta 2$ - $\beta 2$. (B) Dímero $\beta 4$ - $\beta 4$. Las líneas horizontales representan la bicapa lipídica de la membrana celular. Se muestra una representación de varillas de los residuos de azúcares. Nótese que los azúcares en el modelo $\beta 2$ - $\beta 2$ se localizan entre medio de las dos subunidades.

5. Materiales y métodos

5.1. Generalidades

5.1.1. Electroforesis de proteínas

La técnica de electroforesis en geles de poliacrilamida con dodecil sulfato de sodio (SDS-PAGE) se realizó siguiendo el método de Schgger y von Jagow (Schgger y von Jagow 1987), con una concentraci3n de mon3meros (%T) de 16.5 % y un grado de entrecruzamiento (%C) de 6 %. Las muestras se prepararon adicionndoles 1 volumen de *buffer* muestra (Tris·Cl 100 mM, glicerol 24 %, azul de bromofenol 0.01 %, SDS 8.0 %, pH 6.8, al que se agreg3 2-mercaptoetanol, 8 % v/v, justo antes de usar) y se calentaron a 100  C por 5 minutos. Luego de 1 minuto de centrifugaci3n a 14000 rpm, se sembraron 5-20  L de cada muestra en el gel. Las corridas electrofor3ticas se iniciaron con un voltaje de 30 V, que se mantuvo mientras las muestras se encontraban en el *stacking* gel, y luego se increment3 a 100-150 V, finalizndolas cuando el frente de corrida migr3 al extremo an3dico del gel. Las prote3nas se visualizaron incubando el gel con una soluci3n de Coomassie Brilliant Blue 0.3 % p/v, en una mezcla de metanol, cido ac3tico y agua (3:1:6), seguida de lavados sucesivos con una soluci3n de etanol, cido ac3tico y agua (3:1:6). Como patr3n de referencia para estimar los pesos moleculares y las masas de las bandas, se us3 una mezcla de prote3nas de pesos moleculares conocidos sembradas en las mismas condiciones.

5.1.2. Espectrometr3a de masas y secuenciaci3n N-terminal

Las determinaciones de masa se hicieron a trav3s del servicio del LANAIS-PRO (UBA, CONICET), o de la Unidad de Bioqu3mica y Prote3mica Anal3ticas del Institut Pasteur de Montevideo. En el LANAIS-PRO, se emplearon un espectr3metro de masa LCQ DUO ESI ion trap (Thermo Finnigan, San Jos3, CA) y un espectr3metro 4800 Plus MALDI-TOF/TOF Analyzer (AB Sciex, Framingham, MA, USA), y en el Institut Pasteur se us3 un equipo 4800 MALDI-TOF/TOF (Applied Biosystems, Framingham, MA, USA).

La secuencia N-terminal se obtuvo por degradaci3n de Edman, empleando un secuenciador Applied Biosystems modelo 477A, con un analizador de feniltiohidanto3na modelo 120. Este anlisis se hizo tambi3n en el LANAIS-PRO.

5.1.3. Determinación de la concentración de proteínas

La concentración de proteínas se determinó por absorbancia UV a 280 nm en condiciones nativas, aplicando la ley de Lambert-Beer. El coeficiente de extinción molar a 280 nm del estado desplegado, ϵ_{280} (GdmCl 6 M), se calculó a partir de la composición de aminoácidos, aplicando la aproximación propuesta por Gill & von Hippel (Gill y von Hippel 1989). El valor correspondiente al estado nativo se obtuvo haciendo diluciones idénticas de un stock concentrado de proteína nativa, en fosfato de sodio 50 mM pH 7.0, y en el mismo *buffer* con una concentración final de GdmCl 6 M. Luego de incubar a temperatura ambiente por al menos 4 horas, se midieron los espectros de absorción de estas diluciones entre 240 y 340 nm. El valor de ϵ_{280} para la proteína nativa fue calculado como:

$$\epsilon_{280}(\text{Nativo}) = A_{280}(\text{Nativo}) / [A_{280}(\text{GdmCl 6 M}) / \epsilon_{280}(\text{GdmCl 6 M})]$$

donde $\epsilon_{280}(\text{GdmCl 6 M}) = 1280 \times (\text{N}^\circ \text{ de residuos de tirosina})$

Como se describe en el Capítulo 2, la única proteína con residuos de cisteína, M16SEA, no posee enlaces disulfuro, y por esa razón, sólo los residuos de tirosina se tuvieron en cuenta para el cálculo de $\epsilon_{280}(\text{GdmCl 6 M})$.

Los espectros se midieron con un espectrofotómetro Jasco UV-550 (Jasco Corporation, Tokio, Japón) mediante un barrido lento (40 nm/min), luego de centrifugar la muestra extensivamente a 4 °C, para eliminar cualquier material precipitado que pudiera estar presente. Los espectros se corrigieron restando el blanco correspondiente, y sustrayendo además la contribución de la luz dispersada elásticamente (*light scattering*), que se estimó por regresión no lineal de los datos entre 320 y 340 nm, fuera de la región de absorción de las proteínas. Para este fin, se empleó una ecuación que describe la dependencia de la luz dispersada en función de la longitud de onda, aplicada a partículas del tamaño esperado de los agregados proteicos:

$$S = a + b \lambda^{-c}$$

donde a , b y c son los parámetros de ajuste. Se espera que la constante c se aproxime a 4.0 para soluciones de moléculas cuyos radios son mucho menores que la longitud de onda empleada.

Los coeficientes de extinción nativos a 280 nm, determinados con esta metodología, fueron $1300 \text{ M}^{-1}\text{cm}^{-1}$ para MPE fgrina, $3377 \text{ M}^{-1}\text{cm}^{-1}$ para MPE ICA512, $5207 \text{ M}^{-1}\text{cm}^{-1}$ para M16SEA y $7785 \text{ M}^{-1}\text{cm}^{-1}$ para M1SEA, donde la molaridad se refiere a los moles de monómero para MPE fgrina, MPE ICA512 y M16SEA, y a moles de heterodímero escindido para M1SEA.

5.1.4. Concentración y almacenamiento de proteínas

Para concentrar proteínas en el estado nativo, se emplearon dispositivos que consisten en dos compartimientos separados por una membrana de corte de 10 kDa. Se emplearon distintas versiones del mismo dispositivo, según el volumen a concentrar, se usaron CENTRIPREP® YM10 (para volúmenes de hasta 15 ml), CENTRICON YM10 (2 ml), AMICON 4 (4 ml), todos ellos fabricados por Millipore (Bedford, MA), y Vivaspin500 (0.5 ml, GE Healthcare Bio-Sciences AB, Uppsala, Suecia). En todos los casos se siguieron las instrucciones del fabricante, empleando una centrífuga refrigerada a 4 °C.

Los experimentos descriptos en esta tesis se llevaron a cabo con preparaciones frescas de proteína (dentro de 5 días posteriores a su purificación), conservando dichas preparaciones en la heladera a 4 °C.

5.2. Ingeniería genética

5.2.1. Generalidades

Los oligonucleótidos iniciadores (*primers*) empleados en las reacciones de PCR se diseñaron usando el programa Amplify (Bill Engels® 1992, Universidad de Wisconsin, Madison, WI, EE.UU.) para simular los posibles productos de amplificación, y el programa RNAdraw de Ole Matzura (Departamento de Biofísica, Karolinska Institute, Solna, Suecia), para evitar la formación de estructura secundaria que podría disminuir el rendimiento de la PCR. Los oligonucleótidos fueron sintetizados por Operon Biotechnologies GmbH (Cologne, NRW, Alemania). Las reacciones se llevaron a cabo en un termostatzador programable Minicycler PTC-150 de MJ Research Inc. (Waltham, MA, EE.UU.), involucrando 30 ciclos de desnaturalización, apareamiento y elongación.

Las reacciones de PCR emplearon la enzima DNA polimerasa Platinum *Pfx* (Invitrogen, Carlsbad, CA, EE.UU.). Las endonucleasas de restricción se obtuvieron de Promega Corporation (Madison, WI, EE.UU.), y la enzima ligasa del fago T4 se adquirió también a Invitrogen. Las reacciones se llevaron a cabo siguiendo las instrucciones del fabricante. La relación molar aproximada en las reacciones de ligación fue de 5:1 (inserto:vector).

Los procedimientos de manipulación de ácidos nucleicos no descriptos en esta sección se llevaron a cabo siguiendo técnicas estándares de biología molecular (Sambrook, Fritsh *et al.* 1989). La separación de ácidos nucleicos se hizo por electroforesis en geles de agarosa 1 % en *buffer* TBE (Tris-Borato 0.045 M, EDTA 1 mM), conteniendo bromuro de etidio 0.5

µg/mL. La visualización y cuantificación de ácidos nucleicos se hizo por iluminación con luz UV.

Para purificar ADN de las mezclas de reacción, como de PCR y digestión con endonucleasas, la porción de agarosa conteniendo la banda de interés se cortó con un bisturí y se extrajo el ADN mediante el kit Silica Bead DNA Gel Extraction Kit (Fermentas, Vilnius, Lithuania).

La identidad de las construcciones genéticas fue verificada por secuenciación de ADN (Macrogen, Seúl, Corea).

5.2.2. Clonado del dominio M1SEA

El plásmido conteniendo la región codificante para el dominio SEA de la mucina humana MUC1 (Uniprot ID: P15941) fue gentilmente cedido por el Dr. Torleif Härd, de la Universidad de Göteborg, Suecia. El mismo codifica los residuos 1024-1152 de MUC1, precedido de la secuencia MGSSHHHHH. Se diseñaron oligonucleótidos iniciadores que permitieron amplificar la región 1041-1152 correspondiente al dominio SEA, y que por lo tanto hemos denominado M1SEA, conteniendo además las secuencias de reconocimiento para las enzimas de restricción *NdeI* y *BamHI* (*primers* SEA-F y SEA-R en la Tabla 5.1).

El producto de PCR se ligó al vector pGemT Easy (Promega), siguiendo las instrucciones del fabricante. La mezcla de ligación se transformó en bacterias JM109 electrocompetentes, preparadas según el método descrito en la Sección 5.2.5. Las características de todas las cepas de *E. coli* empleadas se muestran en la Tabla 5.2. Se seleccionaron las colonias portadoras de la construcción buscada mediante el sistema de α -complementación (parte del sistema pGemT) y posterior mapeo por restricción. A continuación, la región codificante para el dominio SEA se escindió del vector pGemT por digestión con las enzimas *NdeI* y *BamHI*, y se ligó al vector pET9b (Novagen, EMD Chemicals Inc., San Diego, CA, EE.UU.) previamente digerido con las mismas enzimas. Con la reacción de ligación, se transformaron células electrocompetentes JM109, a partir de las cuales se recuperó la construcción buscada.

Tabla 5.1. Oligonucleótidos iniciadores empleados en las reacciones de PCR.

Primer	Secuencia (escrita en el sentido 5'→3')
F-pET9(971)	GGAAGGAGCTGACTGGGTTGAAG
R-pET9(4299)	AATTGCAGGTTTCATTTGATGCT

F-469	<u>CATATGGAATATGGCTACATCGT</u>
R-557	<u>GGATCCTCACTCCCTCTGTCCCACT</u>
SEA-F	<u>CATATGTCTTTCTTTTTCCTGTCTTTTCACA</u>
SEA-R	<u>CGGATCCTTAAGACTGGGCAGAGAAAGG</u>
M16SEAF	<u>CATATGAGTTCCAGTTCTCAG</u>
M16SEAR	<u>GGATCCTCACCCCGAATCCACA</u>
M16SEASTOP	CAT <u>GGATCCTTATTACCCCGAATCAACAAAG</u>
IA2b-502F	<u>CATATGGAGGTCCAGCCTTCTGAGGA</u>
IA2b-599R	<u>GGATCCTCATTTGCTTTTCGACCCGA</u>
IA2b-486F	<u>CATATGAGTCTTCCGGCGGGTGCTC</u>
IA2b-614R	<u>GGATCCTCACTTGGTAGAGTCTTCT</u>
IA2b2F	GGATGTCGCCCGTCTCCTGCAGGTGCCGTCTAGTGCGTTCGCTGACGTGGAGGTTCTCG
IA2b2R	GGCGACATCCTCAACAAGACGACGTCTTCTCTGGACGCAGCGGATCTCTGTCTGTC

Tabla 5.2. Cepas de *E. coli* empleadas para el clonado y expresión de proteínas.

Cepa	Genotipo
JM109	<i>endA1, recA1, gyrA96, thi, hsdR17</i> (r_k^- , m_k^+), <i>relA1, supE44</i> , $\Delta(lac-proAB)$, [<i>F'</i> , <i>traD36, proAB, laqI</i> ^q $\Delta M15$]
BL21(DE3)pLysS	<i>F</i> ⁻ , <i>ompT, hsdS_B</i> (r_B^- , m_B^-), <i>dcm, gal</i> , λ (DE3), pLysS, Cm ^r
BL21(DE3)RIL	<i>F</i> ⁻ , <i>ompT, hsdS_B</i> (r_B^- , m_B^-), <i>dcm, gal</i> , λ (DE3), Tet ^r , <i>endA</i> , Hte [<i>argU, ileY, leuW</i> Cam ^r]

5.2.2. Clonado del dominio M16SEA

El material genético que contiene la región codificante para el dominio SEA del homólogo murino de la mucina humana MUC16 (Uniprot ID: Q9D1H) fue cedido por el Dr. Shigeyuki Yokoyama de la Universidad de Tokio, Japón. Esta construcción codifica el dominio

SEA (residuos 67-185), junto con otros residuos derivados del vector, una secuencia de seis histidinas en el extremo N y una secuencia de reconocimiento para la proteasa TEV. Se diseñaron *primers* que permitieran eliminar todas las secuencias ajenas al dominio SEA, aunque se preservaron de la construcción original dos residuos del extremo C, S120 y G121, que aunque están ausentes en la proteína silvestre, forman interacciones por puente de hidrógeno en la hoja β , como se observa en la estructura tridimensional de este dominio. Las secuencias de los *primers* empleados, M16SEAF y M16SEAR, se muestran en la Tabla 5.1. El segmento correspondiente al dominio SEA, al cual denominamos M16SEA, se insertó en el plásmido de expresión pET9b siguiendo la misma metodología descrita para el dominio M1SEA.

La variante M16STOP se preparó a partir del vector de expresión codificante para M16SEA, empleando los *primers* M16STOP y F-pET9(971). El producto de PCR se cortó con las enzimas *Sall* y *Bam*HI, y se ligó al vector pET9b cortado con las mismas enzimas.

5.2.3. Clonado del dominio MPE ICA512

El plásmido de expresión con el segmento codificante para el dominio MPE ICA512, residuos 470-558 de la proteína ICA512 humana (Uniprot ID: Q16849), se construyó previamente en nuestro laboratorio como está descrito (Primo, Jakoncic *et al.* 2011). Para ello se empleó el mismo plásmido de expresión y la metodología ya descrita para M1SEA, usando los *primers* F-469 y R-557 listados en la Tabla 5.1.

5.2.4. Clonado del dominio MPE fgrina

El ADN codificante para la isoforma 2 de la proteína fgrina humana (UniProt ID: Q92932-2), fue gentilmente cedido por el Dr. John C. Hutton del Barbara Davis Center for Childhood Diabetes, Universidad de Colorado, EE.UU. El dominio MPE fgrina comprende los residuos 502-599 (de acuerdo a la numeración de la isoforma 1). La isoforma 2 carece de un exón que codifica los residuos 519-547, por esta razón, la estrategia de clonado se hizo en dos etapas. En primer lugar se clonó la región 502-599 de la isoforma 2, y en segundo lugar, se añadieron los residuos faltantes de la isoforma 1. La primera etapa se hizo con la misma metodología ya descrita para los otros dominios SEA, empleando el vector de expresión pET9b, y los *primers* IA2b-502F y IA2b-599R. En la segunda etapa, los residuos faltantes se introdujeron usando un par *primers* solapantes (IA2b2F, IA2b2R) y otro par de *primers* externos, F-pET9(971) y R-pET9(4299), que poseen la misma secuencia que la región que flanquea al sitio de clonado múltiple del vector pET9b. Los productos de cada PCR se

purificaron como se describió en la Sección 5.2.1, y se mezclaron en una nueva PCR siguiendo el método denominado “*overlap extension PCR*” (Higuchi, Krummel *et al.* 1988). Luego de purificar del gel la banda de amplificación correspondiente, ésta se cortó con las enzimas *Sall* y *EcoRI*, ligándose a continuación al vector pET9b.

Para clonar la variante IA-2b⁴⁸⁶⁻⁶¹⁴, se siguió la misma estrategia en dos etapas descripta para la región 502-599, usando en la primera etapa los *primers* IA2b-486F, e IA2b-614R.

5.2.5. Preparación y transformación de bacterias competentes

Durante los pasos de clonado y propagación de plásmidos, se empleó la cepa de *E. coli* JM109 (Promega). Para la sobreexpresión de las proteínas de interés, se usaron las cepas de *E. coli* BL21(DE3)pLysS (para MPE ICA512) y BL21(DE3)RIL (en el resto de las construcciones), adquiridas respectivamente a Novagen y a Stratagene (La Jolla, CA, EE.UU.). Las características de estas cepas se indican en la Tabla 5.2.

Las bacterias competentes se transformaron con el método de electroporación, para ello se empleó un equipo *E. coli* Pulser de Bio-Rad Laboratories Inc. (Hercules, CA, EE.UU.). Las células se prepararon siguiendo las instrucciones del equipo, y se guardaron en alícuotas de 80 µl a -86 °C hasta su uso. En el procedimiento de transformación, se adicionaron 1 a 2.5 µl de mezcla de ligación, o 10 ng de plásmido superenrollado, y se emplearon celdas de 0.2 cm. Las células transformadas se sembraron en placas de Petri con medio LB agar, conteniendo los antibióticos correspondientes, y se incubaron ON a 37 °C.

Para la conservación a largo plazo de bacterias transformadas, se emplearon colonias aisladas para inocular cultivos en medio LB con los antibióticos correspondientes y crecidos hasta saturación. A una alícuota de cada cultivo se le agregó glicerol en una concentración final de 18 % v/v, y luego se almacenó a -86 °C o en nitrógeno líquido.

5.3. Expresión y purificación de proteínas

5.3.1. Dominio M16SEA

Se inocularon 10 ml de medio LB, suplementado con 50 µg/ml de kanamicina y 34 µg/ml de cloranfenicol, con una ansada del stock de bacterias transformadas correspondiente, y se crecieron a 37 °C toda la noche. Al día siguiente, se emplearon los 10 ml del cultivo anterior para inocular 1000 ml de medio LB con los mismos antibióticos, y se creció a 37 °C hasta $DO_{600\text{ nm}} = 1$. La expresión de proteínas se indujo con lactosa monohidrato 1 % p/v, incubando

durante 4 horas a una temperatura de 37 °C, y con agitación a 200 rpm. Al cabo de este tiempo, las bacterias se colectaron por centrifugación (7000 x g, 15 minutos, 4 °C). Se descartó el sobrenadante y el *pellet* se resuspendió en 100 ml de *buffer* de lisis (Tris·Cl 25 mM pH 7.2) y se repitió la centrifugación. A continuación, se resuspendió el *pellet* en 50 ml del mismo *buffer*, se repitió la centrifugación y finalmente se almacenó el *pellet* a -20 °C hasta su uso.

El *pellet* bacteriano se resuspendió en 20 ml de *buffer* de lisis y se agregó PMSF en una concentración final de 1 mM, a partir de un stock 100 mM preparado en isopropanol. A continuación, se lisaron las células usando una *French Press* (Thermo, Rochester, NY, USA), a una presión de 1000 psi. Luego se separó la fracción soluble por centrifugación (dos ciclos, a 17000 y 20000 x g, por 15 y 20 minutos, respectivamente, a 4 °C). El sobrenadante se filtró a través de un filtro de jeringa de 0.22 µm y se vertió sobre 4 ml de resina de intercambio aniónico Q-Sepharose Fast Flow (GE Healthcare, Uppsala, Suecia), equilibrada previamente en *buffer* de lisis. Se llevó el volumen final a 50 ml con *buffer* de lisis y se incubó durante 30 minutos en el refrigerador, a 4 °C y con agitación eventual. Finalmente se transfirió la suspensión a una columna de polipropileno de 30 ml, con la cual se retuvo la matriz y se dejó pasar la fracción no retenida. Dado que la proteína de interés posee un pI de 9.10, no se retiene en la matriz en estas condiciones, pero sí lo hacen el ADN y otras proteínas y componentes celulares. A continuación, se lavó la matriz con 10 ml de *buffer* de lisis, y esta fracción se combinó con la fracción no retenida. La preparación con las fracciones no retenida y de lavado combinadas, se dializó a 4 °C contra 1 litro de *buffer* fosfato de sodio 25 mM a pH 7.0, en presencia de 1,4-dithio-D-threitol (DTT) 1 mM, usando una membrana Spectra/Por 6000-8000 Da (Spectrum Medical Industries Inc., Houston, TX, EE.UU.). Luego 3 horas, la muestra de diálisis se cambió a otro recipiente con 1 litro del mismo *buffer*, y se continuó el proceso por 3 horas más.

El paso siguiente de purificación consistió de un intercambio catiónico mediante una columna MonoS HR 5/5 (GE Healthcare, Uppsala, Suecia), conectada a un cromatógrafo líquido de alta presión Jasco (Jasco Corporation, Japón). Para este fin, se filtró la muestra dializada con un filtro de jeringa de 0.22 µm y se inyectó en la columna, previamente equilibrada con *buffer* fosfato de sodio 25 mM pH 7.0. Finalmente, la proteína se separó de la columna con un gradiente lineal de NaCl desde 0 a 1 M en el mismo *buffer*. Las fracciones con M16SEA se identificaron por SDS-PAGE y se guardaron en el refrigerador a 4 °C hasta el día siguiente, agregando previamente DTT hasta una concentración de 5 mM y también EDTA 2.5 mM, para proteger los tioles.

Mediante el protocolo descripto, se obtuvo la proteína con elevado nivel de pureza, que se observa como una sola banda en SDS-PAGE. De todas formas, rutinariamente se hizo un paso de purificación adicional, concentrando la proteína y sembrándola en una columna de

exclusión molecular Superose 12 HR 10/30 (GE Healthcare, Uppsala, Suecia), equilibrada en *buffer* fosfato de sodio 25 mM pH 7.0, con lo cual también se eliminan el DTT y el EDTA, lo que deja la muestra en las condiciones apropiadas para los estudios biofísicos.

5.3.2. Dominio M1SEA

El protocolo de expresión, lisis y separación de la fracción soluble del lisado celular fue idéntico al ya descrito para el dominio M16SEA. Esta fracción se pasó por un filtro de jeringa de 0.22 μ m y se sometió a precipitación con sulfato de amonio, añadiendo lentamente sulfato de amonio sólido con agitación, en un baño de hielo, hasta que la concentración alcanzó un 57 % de saturación. Se mantuvo con agitación en hielo durante 30 min, y se separó la fracción soluble por centrifugación a 10000 x g, por 15 minutos a 0 °C. La fracción insoluble, conteniendo a M1SEA, se resuspendió en 5 ml de *buffer* de lisis.

El paso siguiente de purificación consistió en una cromatografía de intercambio aniónico mediante una columna MonoQ HR 5/5 (GE Healthcare, Uppsala, Suecia), conectada al cromatógrafo líquido de alta presión Jasco. Para este fin, se eliminó previamente el sulfato de amonio pasando la muestra a través de una columna de exclusión molecular G-25 Medium (PD10 column, GE Healthcare, Uppsala, Suecia), equilibrada en *buffer* de lisis. Las fracciones recogidas se inyectaron en la columna de intercambio aniónico, previamente equilibrada con *buffer* de lisis, y las proteínas se separaron de la columna con un gradiente lineal de NaCl, desde 0 a 1 M en el mismo *buffer*. Las fracciones conteniendo a M1SEA se identificaron por SDS-PAGE, y se guardaron en el refrigerador a 4 °C hasta el día siguiente.

El último paso de purificación consistió de una cromatografía de exclusión molecular, que se hizo de la misma forma descrita en la Sección 5.3.1 para el dominio M16SEA.

5.3.3. Dominio MPE ICA512

El dominio MPE ICA512 se obtuvo creciendo las bacterias transformadas en medio LB con kanamicina, y todos los aspectos del cultivo, inducción y lisis bacteriana se llevaron a cabo como se detalló para el dominio M16SEA.

El primer paso de purificación consistió en una cromatografía de intercambio aniónico con una resina Q-Sepharose Fast Flow. La fracción soluble de lisis se pasó a través de un filtro de jeringa de 0.22 μ m y se vertió sobre 4 ml de resina, equilibrada previamente con *buffer* de lisis. Se llevó el volumen final a 50 ml con *buffer* de lisis y se incubó durante 30 minutos en el refrigerador, a 4 °C y con agitación eventual. Finalmente se transfirió la suspensión a una columna de polipropileno de 30 ml, mediante la cual se retuvo la matriz. Se

hicieron lavados de 20 ml con *buffer* de lisis, y luego con el mismo *buffer* suplementado con 0.1 M de NaCl y posteriormente con 0.15 M de NaCl. La proteína MPE ICA512 se separa de la matriz en estas dos etapas de lavado, las fracciones correspondientes se combinaron y se dializaron de la misma forma descripta para el dominio M16SEA, pero usando un *buffer* Tris·Cl 25 mM pH 8.5.

La muestra dializada se pasó a través de un filtro de jeringa de 0.22 μm , y se inyectó en una columna intercambio aniónico MonoQ HR 5/5, conectada al cromatógrafo líquido de alta presión Jasco. La columna se equilibró previamente con el *buffer* Tris·Cl 25 mM pH 8.5. Las proteínas se separaron de la matriz mediante un gradiente lineal de NaCl desde 0 a 1 M en el mismo *buffer*. Las fracciones conteniendo a MPE ICA512 se identificaron por SDS-PAGE y se guardaron en el refrigerador a 4 °C hasta el día siguiente.

El paso final de purificación consistió en una cromatografía de intercambio catiónico mediante una columna MonoS HR 5/5 conectada al cromatógrafo mencionado, para lo cual fue necesario previamente hacer un cambio del *buffer* de las fracciones de proteína, pasándolas a través de una columna de exclusión molecular Sephadex G-25 Medium, equilibrada con *buffer* acetato de sodio 25 mM pH 4.8. Las fracciones recogidas se inyectaron en la columna de intercambio catiónico, equilibrada con el mismo *buffer*, y las proteínas se separaron de la columna por un gradiente lineal de NaCl desde 0 a 1 M en el mismo *buffer*. Las fracciones conteniendo a MPE ICA512 se identificaron por SDS-PAGE.

5.3.4. Dominio MPE fogrina

En el caso de MPE fogrina, el protocolo de crecimiento y lisis de bacterias es similar al ya descripto para la purificación del dominio SEA de MUC16, excepto que en este caso se empleó el medio de cultivo Terrific Broth y la inducción se hizo por 20 horas a 20 °C. La fracción soluble de lisis se pasó por un filtro de 0.22 μm , y se sometió a precipitación con sulfato de amonio. Para ello se agregó lentamente el reactivo sólido, con agitación en un baño de hielo, hasta que la concentración alcanzó un 60 % de saturación. Se mantuvo con agitación en hielo durante 30 minutos, y se separó la fracción soluble por centrifugación a 10000 x g (15 minutos a 0 °C). Se repitió la centrifugación, y la fracción soluble se llevó a 80 % de saturación, mediante el agregado de la cantidad necesaria de sulfato de amonio sólido, de la misma forma que en el paso anterior. Luego de 1 hora de incubación, se repitieron los pasos de centrifugación, y se reservó la fracción insoluble.

La fracción insoluble de precipitación con sulfato de amonio se resuspendió en 5 ml de *buffer* de lisis. Los pasos siguientes de separaciones cromatográficas se desarrollaron de la misma forma y en la misma secuencia descripta para M1SEA.

5.4. Reactividad química de tioles

La determinación de tioles libres se llevó a cabo empleando el reactivo 5,5'-ditiobis-(2-nitrobenzónico) (DTNB). Las reacciones se efectuaron en *buffer* fosfato de sodio 25 mM, pH 7.0. A las muestras se les agregó DTNB de un *stock* preparado en el mismo *buffer*, en un exceso de 3 veces la cantidad de tioles esperada para una muestra totalmente reducida, y se determinó la absorbancia a 412 nm hasta lectura constante (5-10 min). Se hizo un blanco en las mismas condiciones, pero sin proteína, que se restó a la lectura anterior. Luego se calculó la concentración de tioles reactivos usando un coeficiente de extinción $\epsilon_{412\text{ nm}} = 14150\text{ M}^{-1} \times \text{cm}^{-1}$ (Riddles, Blakeley *et al.* 1983). Se siguió el mismo procedimiento para mediciones en el estado desplegado, pero usando un $\epsilon_{412\text{ nm}} = 13700\text{ M}^{-1} \times \text{cm}^{-1}$.

5.5. Modificación química con agentes entrecruzantes

Las reacciones de entrecruzamiento se hicieron con los reactivos glutaraldehído y disuccinimidil suberato (DSS). Las reacciones se llevaron a cabo en *buffer* fosfato de sodio 25 mM pH 7.0, con el agregado de las cantidades indicadas de glutaraldehído (50 % en H₂O, Fluka) o de un *stock* de DSS 20 µg/ml (preparado en dimetilsulfóxido) o 10 mg/ml, según se indique. La reacción tuvo lugar durante 15 minutos a 27 °C y se frenó con el agregado de *buffer* muestra de SDS-PAGE. Los resultados se analizaron por SDS-PAGE.

Cuando las reacciones se llevaron a cabo a bajas concentraciones de proteína, o en presencia de GdmCl, las muestras se precipitaron con TCA antes de sembrarlas en el gel. En estos experimentos, la reacción de entrecruzamiento se detuvo agregando Tris·Cl pH 7.5, en una concentración final de 50 mM. Cuando las muestras contenían GdmCl, la precipitación se llevó a cabo por el agregado de deoxicolato de sodio, en una concentración final de 2 % p/v y TCA 15 % p/v, e incubando a temperatura ambiente por 15 minutos. Luego las muestras se centrifugaron por 10 minutos y se lavaron los *pellets* con 100 µl de acetona a -20 °C. Luego de resuspender completamente, se incubaron toda la noche a -20 °C, o por 30 minutos a -80 °C. Se repitió la centrifugación, y los *pellets* se dejaron secar por 5 minutos, resuspendiéndolos finalmente en *buffer* muestra de SDS-PAGE. En el caso de las reacciones a baja concentración de proteína, pero sin GdmCl, se siguió el mismo procedimiento de precipitación, pero sin el agregado de deoxicolato de sodio.

5.6. Digestión controlada con tripsina

La digestión con tripsina se hizo en *buffer* fosfato de sodio 25 mM pH 7.0 y a 27 °C, utilizando tripsina (SIGMA) en relación 1:100 en masa (enzima:sustrato). Se tomaron alícuotas a distintos intervalos de tiempo y se mezclaron con 1 volumen de *buffer* muestra de SDS-PAGE. Luego se calentaron a 100 °C por 5 minutos y se congelaron a -20 °C hasta analizarlas por SDS-PAGE.

5.7. Comportamiento hidrodinámico

5.7.1. Exclusión molecular (SEC)

El radio hidrodinámico y estado de agregación se estudió por cromatografía de exclusión molecular usando una columna Superose 12 HR 10/30 (GE Healthcare, Uppsala, Suecia), acoplada a un cromatógrafo líquido de alta presión Jasco (Jasco Corporation, Japón). La columna se calibró utilizando una mezcla de 5 patrones de proteínas globulares de propiedades hidrodinámicas conocidas (Gel Filtration Estándar, adquirido a Bio-Rad Laboratories Inc., consistente de tiroglobulina (670 kDa), IgG bovina (158 kDa), ovoalbúmina de pollo (43 kDa), mioglobina equina (17 kDa) y vitamina B12 (1.3 kDa)). En algunos experimentos también se usó en la calibración albúmina sérica bovina (BSA, 67 kDa, adquirida a SIGMA), RNasa A (13.7 kDa, CALBIOCHEM) y aprotinina de pulmón bovino (6.5 kDa). La columna se equilibró previamente con al menos dos volúmenes de columna del *buffer* indicado en cada experimento, y el flujo empleado fue de 0.5 ml/min. En cada corrida se sembraron 100 µl de muestra.

El radio de Stokes (R_S) teórico de las proteínas se calculó a partir de su peso molecular, usando la ecuación descripta por Uversky (Uversky 1993):

$$\log R_S = 0.369 \times \log(PM) - 0.254$$

Con esta ecuación, se obtiene habitualmente el R_S de proteínas patrones (Uversky 1993). Para determinar el R_S experimental de la proteína estudiada, se construyó una curva de calibración relacionando el R_S de las proteínas patrón con su volumen de elución ($1000/V_{\text{elución}}$ en función de R_S), y los puntos se ajustaron a una recta, que se usó para interpolar el R_S correspondiente al volumen de elución de la proteína en estudio.

El R_S también se estimó de manera teórica a partir del número de residuos de aminoácidos (N_u), empleando la ecuación descripta por Dobson (Wilkins, Grimshaw *et al.* 1999):

$$R_S = (4.75 \pm 1.11) N_u^{0.29}$$

5.7.2. Dispersión de luz estática (SLS)

Para la determinación de masa por SEC-MALLS se usó un equipo miniDawn (Wyatt Technology) acoplado a una columna Superdex 200 10/300 GL (GE Healthcare) y un detector de absorbancia Jasco UV-2075 Plus. El *buffer* de elución fue fosfato de sodio 25 mM, NaCl 0.15 M, pH 7.0. Los experimentos se realizaron a 20 °C, inyectando 100 µl de muestra, con un flujo de 0.3 o 0.4 ml/min. Los datos se analizaron con el programa Astra 6.0 (Wyatt Technology).

5.8. Dicroísmo circular

Se estudió el contenido de estructura secundaria y terciaria por espectroscopía de dicroísmo circular en el UV lejano (180 y 240 nm) y en el UV cercano (240-340 nm), respectivamente (Kelly, Jess *et al.* 2005). Se utilizó un equipo JASCO J-810 (Jasco Corp., Tokio, Japón) calibrado de acuerdo a las instrucciones de los fabricantes con ácido (+)-10-canforsulfónico. Las mediciones se realizaron bajo un flujo de nitrógeno gaseoso de 8 l/h, a una temperatura de 20 °C, controlada con un sistema tipo Peltier (JASCO).

5.8.1. CD en el UV lejano

Se midieron espectros entre 180 y 340 nm, utilizando una celda de 0.1 cm de paso óptico. La concentración de proteína fue 5-12 µM. La sensibilidad fue de 100 miligrados. Se usó una velocidad de barrido de 50 nm/min, un tiempo de respuesta de 1 s y un ancho de banda de 1 nm. Se realizó un promedio de diez espectros para cada muestra. El promedio fue corregido por absorción del *buffer* y posteriormente por la línea de base utilizando como cero el promedio de lecturas entre 310 y 340 nm. Finalmente, los datos fueron suavizados mediante un polinomio de Savitzky-Golay de cuarto grado, con una ventana de diez puntos (Press, Teukolsky *et al.* 1992). Los espectros se convirtieron a elipticidad molar media por residuo usando la relación:

$$[\theta] = \theta / (10 \times c \times n \times d)$$

donde $[\theta]$ es la elipticidad molar (en grados $\times$ cm² $\times$ dmol⁻¹), θ es la elipticidad en miligrados, n el número de residuos de la proteína, c la concentración molar de proteína y d la longitud de la celda en centímetros.

5.8.2. CD en el UV cercano

Los espectros se registraron entre 240 y 340 nm, y se obtuvieron a una velocidad de barrido de 20 nm/min con una sensibilidad de 5 miligrados. La concentración de proteína fue de 50-96 μ M y se usó una celda de 1 cm de paso óptico. Los datos se procesaron de igual manera que lo descripto para el UV lejano, expresando el resultado en forma de elipticidad molar $[\theta]$:

$$[\theta] = \theta / (10 \times c \times d)$$

siendo θ la elipticidad en miligrados, c la concentración molar de proteína y d la longitud de la celda en centímetros.

5.9. Espectroscopía de fluorescencia

Los espectros de emisión de fluorescencia se registraron usando un fluorómetro ISS K2 multifrecuencia (ISS, Champaign, Illinois, USA) equipado con un compartimiento de celda conectado a un baño termostático fijado a 20 °C. En los espectros de emisión de tirosinas, la longitud de onda de excitación fue 276 nm (con un ancho de banda de 8 nm) y la intensidad de la emisión de fluorescencia fue medida entre 250 y 450 nm. La concentración de proteína fue entre 30 y 50 μ M, y una absorbancia no mayor a 0.1 para evitar el efecto filtro interno.

5.9.1. Fluorescencia total de tirosinas

La emisión total de tirosinas se midió interponiendo un filtro Schott WG320 en el canal izquierdo del instrumento, siendo la excitación a 276 nm, con un ancho de banda de 8 nm. El valor registrado fue un promedio de las mediciones realizadas cada 1 segundo durante 30 segundos.

5.9.2. Unión de ANS

Se midió la unión de la sonda fluorescente 1-anilinaftaleno- 8-sulfonato (ANS) a proteínas mediante la emisión de fluorescencia a 480 nm (Semisotnov, Rodionova *et al.* 1991). La concentración de proteína fue la misma empleada en los espectros de emisión de tirosinas. Se agregaron de manera creciente alícuotas de una solución *stock* de ANS 530 μ M

a la celda conteniendo la solución de proteína o *buffer*. La solución *stock* de ANS se preparó en agua y se cuantificó por absorbancia usando un $\epsilon_{350\text{ nm}} = 4950\text{ cm}^{-1}\text{ M}^{-1}$ (Daniel y Weber 1966). Los espectros de emisión se registraron entre 250 y 450 nm, luego de la excitación a 350 nm. Tanto para la excitación como para la emisión se utilizaron rendijas de 8 nm.

5.10. Desnaturalización por temperatura

Las transiciones de desplegado inducidas por temperatura se estudiaron por dicroísmo circular, midiendo la elipticidad a 220 nm mientras se incrementaba linealmente la temperatura entre 4 y 95 °C a una velocidad de 2 °C/min. Salvo que se indique lo contrario, la concentración de proteína fue 5 μM . Se usó un espectropolarímetro JASCO J-810, con una celda de 1 cm de paso óptico y la sensibilidad fue de 100 miligrados. Los datos se analizaron como se indica en la Sección 5.12.

5.11. Desnaturalización química

Los experimentos de desplegado isotérmico se llevaron a cabo a 20 °C, registrando el espectro de CD en el UV lejano, o de fluorescencia intrínseca de tirosinas, usando los equipos descritos en las secciones precedentes. Para las mediciones de CD, se usaron celdas de 0.1 cm, salvo en los experimentos en los que la concentración de proteína fue 1 μM , donde se usaron alternativamente celdas de 0.1 o de 0.5 cm, siendo las últimas preferidas por presentar una mayor relación señal/ruido. Se tomaron dos espectros por muestra, entre 210 y 320 nm, con los mismos ajustes instrumentales descritos anteriormente, excepto que la velocidad de escaneo fue de 100 nm/min. Los espectros promediados y corregidos por el blanco, fueron adicionalmente suavizados para mejorar la relación señal/ruido, usando un filtro digital no paramétrico implementado en el programa PeakFit versión 4 (SeaSolve Software Inc). Las transiciones se monitorearon usando la elipticidad a 222 nm. Las mediciones de fluorescencia total de tirosinas se llevaron a cabo siguiendo el mismo procedimiento descrito anteriormente.

Se prepararon soluciones frescas de GdmCl o de urea en el *buffer* correspondiente, se las llevó a pH y se calculó su concentración exacta midiendo su índice de refracción como está descrito (Pace y Scholtz 1997), usando un refractómetro digital Palm Abbe (Misco, Cleveland, OH, EE.UU.). Se prepararon tubos en paralelo con el mismo volumen final, conteniendo distintas proporciones de *buffer* y desnaturalizante, y se les añadió la misma cantidad de un *stock* concentrado de proteína, o de *buffer* en el caso de los blancos. Se

incubaron las mezclas durante más de 4 horas a temperatura ambiente y se tomaron las medidas espectroscópicas. Los datos se analizaron como se indica en la sección siguiente.

5.12. Análisis de datos de equilibrio de desplegado

5.12.1. Generalidades

Las transiciones de desplegado por desnaturalizante químico se analizaron de acuerdo al modelo de extrapolación lineal, el cual establece que el cambio en la energía libre de Gibbs estándar asociado al desplegado, $\Delta G = -RT \ln(K_u)$, varía linealmente con la concentración de desnaturalizante (D) (Pace y Scholtz 1997):

$$\Delta G = \Delta G_0 - mD \quad (1)$$

donde ΔG_0 es la energía libre de desplegado en ausencia de desnaturalizante y m es una constante de proporcionalidad. Se puede describir la ecuación 1 de una forma más conveniente para el ajuste, definiendo D_m , la concentración de desnaturalizante a la que $\Delta G = 0$, entonces:

$$\Delta G = m(D_m - D) \quad (2)$$

Las señales espectroscópicas observadas (Y_0) se modelaron como la suma de las señales intrínsecas de cada estado presente en el equilibrio (Y_i):

$$Y_0 = \sum F_i Y_i \quad (3)$$

donde el subíndice i , identifica a cada estado y F_i es su fracción molar. Además, se asume que Y_i varía linealmente con la concentración de desnaturalizante (líneas de base pre- y post-transición):

$$Y_i = Y_{i,0} + m_i D \quad (4)$$

donde $Y_{i,0}$ es la señal intrínseca del estado i en ausencia de desnaturalizante, y m_i es la dependencia de la señal del estado i con la concentración de desnaturalizante.

Para realizar el ajuste sobre una misma escala, a partir de datos registrados con distintos métodos espectroscópicos y varias concentraciones de proteína, cada conjunto de datos se normalizó aplicando la ecuación:

$$Y_0 = (S_0 - S_u)/(S_u - S_n) \quad (5)$$

donde S_0 es la señal espectroscópica observada, sin normalizar, mientras que S_n y S_u son las señales sin normalizar del estado nativo y desplegado, que se estimaron como el promedio de los primeros y de los últimos datos, respectivamente.

En el caso de las transiciones de desplegado térmico, se describió la dependencia de la energía libre de Gibbs con la temperatura mediante la ecuación de Gibbs-Helmholtz:

$$\Delta G(T) = \Delta H(T_m) + \Delta C_p(T - T_m) - T \left(\frac{\Delta H(T_m)}{T_m} + \Delta C_p \ln \left(\frac{T}{T_m} \right) \right) \quad (6)$$

en la cual T_m es una temperatura de referencia, donde $\Delta G = 0$. Las fracciones de las distintas especies se obtuvieron de la ecuación 3, asumiendo una variación lineal de las señales espectroscópicas con la temperatura, de manera análoga a la ecuación 4. En los experimentos de desplegado a distintos pH, los parámetros termodinámicos ΔH y T_m variaron conjuntamente de acuerdo a:

$$\Delta H = \Delta H^0 + \Delta C_p(T_m - T_m^0) \quad (7)$$

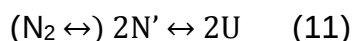
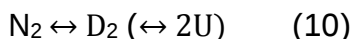
donde ΔH^0 y T_m^0 son los valores correspondientes a un pH de referencia (Privalov 1979).

Las ecuaciones se ajustaron a los datos empleando el complemento Solver de Microsoft Excel 2010, y los errores de los parámetros se calcularon mediante el método Monte Carlo como está descrito (Motulsky y Christopoulos 2003), empleando también Microsoft Excel 2010.

5.12.2. Modelo de desplegado unimolecular de dos estados

Este modelo se aplica a situaciones donde se observa una única transición, y sin cambios en el estado de asociación de la proteína. Este puede ser el caso de una proteína monomérica (ecuación 8), o también cuando la proteína es un dímero (u oligómero) y la transición no involucra disociación (ecuación 9), o bien cuando la disociación es indetectable

para la sonda espectroscópica (ecuaciones 10 y 11, los paréntesis encierran a la transición silenciosa, que no se tiene en cuenta al realizar el ajuste):



La constante de equilibrio para el modelo representado por la ecuación 8 se define como $K_u = [U]/[N]$, y para los otros modelos se define de manera semejante. Con lo cual, la fracción del estado desplegado se calcula como:

$$F_u = K_u / (1 + K_u) \quad (12)$$

$$Y_0 = Y_n(1 - F_u) + Y_u F_u \quad (13)$$

En el caso del desplegado por agentes químicos, las expresiones 2, 12 y 13 se combinan para dar la ecuación de ajuste:

$$Y_0 = \frac{(Y_{n,0} + m_n D) + (Y_{u,0} + m_u D) \exp\{m(D - D_m)/RT\}}{1 + \exp\{m(D - D_m)/RT\}} \quad (14)$$

5.12.3. Modelo de desplegado unimolecular de tres estados

Este modelo se emplea cuando existen dos transiciones, y asume la presencia de un intermediario (I). El esquema de reacción correspondiente es:



Las constantes de equilibrio para la primera y segunda etapa son, respectivamente, $K_1 = [I]/[N]$ y $K_2 = [U]/[I]$, y la fracciones de cada estado se calculan como:

$$F_n = 1 / (1 + K_1(1 + K_2)) \quad (16)$$

$$F_i = K_1 / (1 + K_1(1 + K_2)) \quad (17)$$

$$F_u = K_1 K_2 / (1 + K_1(1 + K_2)) \quad (18)$$

La ecuación de ajuste se obtiene combinando las ecuaciones 2-4 y 16-18.

5.12.4. Modelo de desplegado dimérico de dos estados

En este modelo, desplegado y disociación son procesos concertados, por lo que el estado nativo dimérico se encuentra en equilibrio con el monómero desplegado según:



Se define la constante de equilibrio como $K_u = [U]^2/[N_2]$, y la concentración total de proteína es, en términos del monómero, $P_t = 2[N_2] + [U]$, por lo que las fracciones de dímero nativo y monómero desplegado son, respectivamente, $F_n = 2[N_2]/P_t$ y $F_u = [U]/P_t$. Entonces, la constante de equilibrio se puede expresar como:

$$K_u = \frac{2P_t F_u^2}{1 - F_u} \quad (20)$$

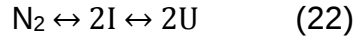
A partir de esta expresión se puede obtener F_u en función de K_u y P_t :

$$F_u = \frac{\sqrt{K_u^2 + 8K_u P_t} - K_u}{4P_t} \quad (21)$$

Finalmente, se obtiene la ecuación de ajuste reemplazando la expresión para F_u en la ecuación 3, asumiendo nuevamente que Y_n e Y_u varían linealmente con la temperatura o la concentración de desnaturalizante.

5.12.5. Modelos de desplegado diméricos de tres estados

Son posibles dos modelos de tres estados, donde el dímero nativo (N_2) se encuentra en equilibrio con el monómero desplegado (U) y un intermediario monomérico (I) o dimérico (I_2):



Cuando el desplegado ocurre a través de un intermediario monomérico, se definen las constantes para la primera y segunda transición como $K_1 = [I]^2/[N_2]$ y $K_2 = [U]/[I]$, respectivamente. La concentración total de proteína es $P_t = 2[N_2] + [I] + [U]$, y las respectivas fracciones de dímero nativo, intermediario y monómero desplegado son respectivamente, $F_n = 2[N_2]/P_t$, $F_i = [I]/P_t$ y $F_u = [U]/P_t$, con $F_n + F_i + F_u = 1$. Por lo tanto, $F_n = 2P_t F_i^2 / K_1$ y $F_u = K_2 F_i$, combinando estas expresiones puede resolverse F_i en función de K_1 , K_2 y P_t :

$$F_i = \frac{-K_1(1+K_2) + \sqrt{K_1^2(1+K_2)^2 + 8P_t K_1}}{4P_t} \quad (24)$$

Finalmente se obtiene la ecuación de ajuste desarrollando la ecuación 3, y sustituyendo la expresión anterior:

$$Y_o = Y_n F_n + Y_i F_i + Y_u F_u \quad (25)$$

$$Y_o = Y_n \left(\frac{2P_t F_i^2}{K_1} \right) + Y_i F_i + Y_u K_2 F_i \quad (26)$$

Para un modelo de tres estados con intermediario dimérico, se definen las constantes de equilibrio para la primera y segunda transición como $K_1 = [I_2]/[N_2]$ y $K_2 = [U]^2/[I_2]$, respectivamente. La concentración total de proteína es entonces $P_t = 2[N_2] + 2[I_2] + [U]$, y las respectivas fracciones de dímero nativo, intermediario y monómero desplegado son, respectivamente, $F_n = 2[N_2]/P_t$, $F_i = 2[I_2]/P_t$ y $F_u = [U]/P_t$, donde $F_n + F_i + F_u = 1$. Por lo tanto $F_i = F_n K_1 = 2P_t F_u^2 / K_2$, lo que permite resolver F_u en función de K_1 , K_2 y P_t :

$$F_u = \frac{-K_1 K_2 + \sqrt{(K_1 K_2)^2 + 8(1+K_1)(K_1 K_2)P_t}}{4P_t(1+K_1)} \quad (27)$$

La ecuación de ajuste se obtiene reemplazando estas expresiones en la ecuación 25:

$$Y_o = Y_n \left(\frac{2P_t F_u^2}{K_1 K_2} \right) + Y_i \left(\frac{2P_t F_u^2}{K_2} \right) + Y_u F_u \quad (28)$$

5.13. Detección de proteasas contaminantes

Para ensayar la presencia de proteasas contaminantes se empleó el método de digestión de α -caseína (Laskowski 1955). Se preparó una solución de caseína aproximadamente 10 mg/ml en el *buffer* de reacción y se pasó a través un filtro de 0.1 μ m de poro. Se mezclaron 100 μ l de la solución de caseína con 100 μ l de proteína para dar una concentración final de 75 μ M. El *buffer* de reacción fue fosfato de sodio 25 mM pH 7.0. Se preparó un blanco en las mismas condiciones, pero omitiendo la proteína. Las muestras se incubaron a 20 °C durante 9 días, tomando alícuotas a distintos tiempos para medir el curso de la reacción. Para este fin, las muestras se precipitaron con el agregado de 1 volumen de ácido tricloroacético 1.8 M, se incubaron durante 30 minutos a 0 °C, y luego se centrifugaron a 12000 rpm durante 10 minutos. Finalmente, se midieron las absorbancias a 280 nm de los sobrenadantes conteniendo los péptidos solubles.

5.14. Determinación de la estructura tridimensional de proteínas por difracción de rayos x

5.14.1. Cristalización de proteínas

Una vez purificadas las proteínas como se describió en las secciones correspondientes, se llevaron a la condición de cristalización, remplazando el *buffer* fosfato para evitar su posible interferencia con los experimentos de cristalogénesis. Para ello se concentraron las muestras como se describió en la Sección 5.1.4 y se diluyeron repetidas veces con el *buffer* de cristalización, Tris-Cl 10 mM pH 7.40, hasta que la concentración de fosfato fuera aproximadamente 300 μ M. Luego se continuó concentrando hasta alcanzar una concentración final de 10 mg/ml.

Para explorar las condiciones de cristalización, se utilizaron los kits comerciales Crystal Screen 1 y 2 (Hampton Research, Aliso Viejo, California, EE.UU.) y JBScreen Classic 9 (Jena Bioscience, Jena, Alemania), que abarcan una variedad de condiciones, incluyendo distintos tipos de *buffer* (pH), sales y precipitantes (sales, polímeros y aditivos orgánicos volátiles y no volátiles). El método utilizado fue el de difusión de vapor en gota suspendida, en placas de 24 pocillos. Para este fin, se agregó a cada pocillo 1 ml de solución de cristalización, y se depositaron sobre un cubreobjetos siliconado 2 μ l de esta solución, mezclados con un volumen igual de la solución de proteína. Luego se colocó el cubreobjetos invertido sobre el reservorio, sellando sus bordes con vaselina sólida. Las placas se almacenaron a 20 °C y se observaron regularmente con una lupa binocular.

Para el transporte y recolección de datos en el sincrotrón, los cristales obtenidos fueron criopreservados, ensayando distintas condiciones de crioprotección. En algunas condiciones de cristalización, fue posible omitir la transferencia a soluciones crioprotectoras, ya que la misma condición ya contenía agentes criopreservantes, igualmente, para todas las condiciones se prepararon cristales con criopreservante. Para este fin, se montaron los cristales mediante un lazo de 0.3-0.4 μ m (Hampton Research), se transfirieron a la solución con criopreservante por unos pocos segundos y luego se congelaron en nitrógeno líquido. Las soluciones de criopreservación estaban compuestas del mismo *buffer* de cristalización, suplementado con PEG400 a una concentración de 10 ó 25 % p/v, o con glicerol 25 % v/v o, cuando la solución de cristalización ya lo incluía como precipitante, sulfato de amonio en una concentración de 3.4 M. El cristal que mejor difractó a pH 4.6 contenía glicerol 25 % como criopreservante, mientras que el cristal a pH 7.4 no fue suplementado con crioprotector.

5.14.2. Recolección de datos y procesamiento

Se recogieron los datos de difracción en la estación X6A del National Synchrotron Light Source (Brookhaven National Laboratory, NY, EE.UU.), usando un detector ADSC Q270 (Area Detector Systems Corp., Poway, CA, EE.UU.), y una temperatura de 100 K. Los datos fueron indexados, integrados, escalados y reducidos usando el paquete de programas HKL2000 (Otwinowski y Minor 1997). Un 5 % de las reflexiones, seleccionadas al azar, fueron asignadas para los cálculos del R_{free} .

5.14.3. Resolución de la estructura, construcción del modelo y refinamiento

Las estructuras de MPE fognina fueron resueltas por el método de remplazo molecular, usando las coordenadas de MPE ICA512 (PDB ID: 2QT7) como modelo de búsqueda. La solución se obtuvo usando Molrep (Vagin y Teplyakov 2010) como parte del programa MrBUMP (Keegan y Winn 2008). Los modelos fueron corregidos manualmente usando Coot (Emsley y Cowtan 2004), y luego sujetos a refinamiento (*restrained refinement*) con REFMAC5 (Murshudov, Vagin *et al.* 1997). Tanto MrBUMP como REFMAC5 forman parte de la suite CCP4 (Cowtan, Emsley *et al.* 2011).

Los modelos iniciales fueron adicionalmente refinados usando REFMAC5 en modo “TLS and restrained refinement”. En el caso de la estructura a pH 7.4, toda la proteína fue incluida en un mismo grupo TLS, mientras que en la estructura a pH 4.6 se usaron 14 grupos TLS, que se definieron usando el servidor TLSMD (Painter y Merritt 2006). De esta forma se llevaron a cabo rondas de refinamiento con REFMAC5 seguidas de ajustes manuales del modelo con Coot, hasta llegar a convergencia.

La calidad estereoquímica de los modelos finales fue verificada con el servidor Molprobability (Chen, Arendall *et al.* 2010). Las coordenadas y factores de estructura fueron depositados en el Banco de Datos de Proteínas (PDB, *Protein Data Bank*) con los códigos 4HTI y 4HTJ.

5.15. Métodos bioinformáticos

En la construcción del alineamiento múltiple de secuencias de las RPTP, en primer lugar se realizó una búsqueda de homólogos en la base de datos no redundante del NCBI, mediante el algoritmo PSI-BLAST (Altschul, Madden *et al.* 1997), usando los parámetros *default*. Se emplearon como secuencias de búsqueda a MPE ICA512 y a MPE fognina, generándose dos alineamientos múltiples con la ayuda del programa BioEdit 7.13 (Ibis

Biosciences, Carlsbad, CA, EE.UU.). Se eliminaron del alineamiento las secuencias con grandes deleciones en la región de los elementos de estructura secundaria. El alineamiento estuvo guiado por la matriz de sustitución BLOSUM62, pero se hicieron algunas pocas correcciones manuales para favorecer la ubicación de las inserciones y eliminaciones en la región entre los elementos de estructura secundaria.

Finalmente, los dos alineamientos se unificaron, eliminando las secuencias redundantes, definidas como aquellas con una identidad mayor al 90 %, empleando el algoritmo CD-HIT (Huang, Niu *et al.* 2010).

Para analizar los patrones de conservación, se asignaron los pesos estadísticos a cada secuencia (Henikoff y Henikoff 1994), para lo cual se calcula primero el peso asociado a cada uno de los residuos en cada posición del alineamiento (w_{kp}) de acuerdo a la expresión:

$$w_{kp} = 1/r \times s$$

donde w_{kp} se refiere al peso del residuo k en la posición p , r es el número de residuos distintos en esa posición y s es el número de ocurrencias del aminoácido en cuestión. Finalmente se suman los pesos de los residuos correspondientes para cada secuencia, obteniéndose el peso de la secuencia en cuestión, el cual luego se normaliza al dividirlo por la suma de los pesos de todas las secuencias. Este peso normalizado de cada secuencia es el que luego se utiliza para ponderar la frecuencia de cada posición en el alineamiento (Larson y Davidson 2000).

Para calcular la entropía posicional, primero se obtuvo la entropía de la información de Shannon, empleando las frecuencias ponderadas de cada posición:

$$H(i) = - \sum_{r=A}^Y p_i(r) \ln(p_i(r))$$

donde $H(i)$ es la entropía para la posición i , para los residuos $r = (A, C, D, ..., W, Y)$, que ocurren con una frecuencia $p_i(r)$. La entropía posicional se obtiene mediante:

$$N(i) = e^{H(i)}$$

Todos los cálculos se realizaron con planillas de Microsoft Excel 2010.

Los alineamientos se representan en forma de *sequence-logo* en las Figuras 3.6 y 3.12, para lo cual se empleó el servidor Web-Logo (Crooks, Hon *et al.* 2004), que emplea como dato de entrada a un alineamiento múltiple. Los resultados no son corregidos por el peso estadístico asociado a cada secuencia, el propósito de estas representaciones es ilustrativo y es necesario referirse a los alineamientos completos mostrados en el APÉNDICE, o a las Tablas 3.3-3.6 para una medida más exacta de la conservación.

Para calcular el área accesible al solvente en el estado nativo y desplegado, se usó el servidor Protsa (Estrada, Bernado *et al.* 2009).

Las representaciones moleculares se hicieron con el programa Pymol 1.4.1 (Schrodinger LCC), y en la Figura 2.1 se usó el programa Discovery Studio 3.1 Visualizer (Accelrys Software Inc). Para calcular el potencial electrostático, se empleó el complemento APBS 2.1 de Pymol, la temperatura fue de 20 °C y el resto de los parámetros se fijaron en los valores *default*.

Bibliografía

- Aicher, B., M. M. Lerch, *et al.* (1997). "Cellular redistribution of protein tyrosine phosphatases LAR and PTPsigma by inducible proteolytic processing." *J Cell Biol* **138**(3): 681-696.
- Akhavan, A., S. N. Crivelli, *et al.* (2008). "SEA domain proteolysis determines the functional composition of dystroglycan." *FASEB J* **22**(2): 612-621.
- Altamura, S., F. D'Alessio, *et al.* (2010). "A novel TMPRSS6 mutation that prevents protease auto-activation causes IRIDA." *Biochem J* **431**(3): 363-371.
- Altschul, S. F., T. L. Madden, *et al.* (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res* **25**(17): 3389-3402.
- Bafna, S., S. Kaur, *et al.* (2010). "Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells." *Oncogene* **29**(20): 2893-2904.
- Baker, D. (2000). "A surprising simplicity to protein folding." *Nature* **405**(6782): 39-42.
- Baldwin, R. L. y G. D. Rose (1999). "Is protein folding hierarchic? I. Local structure and peptide folding." *Trends Biochem Sci* **24**(1): 26-33.
- Bashford, D., C. Chothia, *et al.* (1987). "Determinants of a protein fold. Unique features of the globin amino acid sequences." *J Mol Biol* **196**(1): 199-216.
- Bast, R. C., Jr., M. Feeney, *et al.* (1981). "Reactivity of a monoclonal antibody with human ovarian carcinoma." *J Clin Invest* **68**(5): 1331-1337.
- Battaglia, C., U. Mayer, *et al.* (1992). "Basement-membrane heparan sulfate proteoglycan binds to laminin by its heparan sulfate chains and to nidogen by sites in the protein core." *Eur J Biochem* **208**(2): 359-366.
- Bhattacharyya, R., D. Pal, *et al.* (2004). "Disulfide bonds, their stereospecific environment and conservation in protein structures." *Protein Eng Des Sel* **17**(11): 795-808.
- Bjornsson, A., S. Mottagui-Tabar, *et al.* (1996). "Structure of the C-terminal end of the nascent peptide influences translation termination." *EMBO J* **15**(7): 1696-1704.
- Bork, P. y L. Patthy (1995). "The SEA module: a new extracellular domain associated with O-glycosylation." *Protein Sci* **4**(7): 1421-1425.
- Bottazzo, G. F., E. Bosi, *et al.* (2005). "IA-2 antibody prevalence and risk assessment of early insulin requirement in subjects presenting with type 2 diabetes (UKPDS 71)." *Diabetologia* **48**(4): 703-708.
- Bouanene, H. y A. Miled (2010). "Conflicting views on the molecular structure of the cancer antigen CA125/MUC16." *Dis Markers* **28**(6): 385-394.
- Bugge, T. H., K. List, *et al.* (2007). "Matriptase-dependent cell surface proteolysis in epithelial development and pathogenesis." *Front Biosci* **12**: 5060-5070.
- Burridge, K., S. K. Sastry, *et al.* (2006). "Regulation of cell adhesion by protein-tyrosine phosphatases. I. Cell-matrix adhesion." *J Biol Chem* **281**(23): 15593-15596.
- Caromile, L. A., A. Oganessian, *et al.* (2010). "The Neurosecretory Vesicle Protein Phogrin Functions as a Phosphatidylinositol Phosphatase to Regulate Insulin Secretion." *Journal of Biological Chemistry* **285**(14): 10487-10496.
- Castiglione Morelli, M. A., G. Stier, *et al.* (1995). "The KH module has an alpha beta fold." *FEBS Lett* **358**(2): 193-198.
- Cowtan, K., P. Emsley, *et al.* (2011). "From crystal to structure with CCP4." *Acta Crystallogr D Biol Crystallogr* **67**(Pt 4): 233-234.
- Crooks, G. E., G. Hon, *et al.* (2004). "WebLogo: a sequence logo generator." *Genome Res* **14**(6): 1188-1190.

- Chauhan, S. C., A. P. Singh, *et al.* (2006). "Aberrant expression of MUC4 in ovarian carcinoma: diagnostic significance alone and in combination with MUC1 and MUC16 (CA125)." *Mod Pathol* **19**(10): 1386-1394.
- Chen, V. B., W. B. Arendall, 3rd, *et al.* (2010). "MolProbity: all-atom structure validation for macromolecular crystallography." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 12-21.
- Chin, C. N., J. N. Sachs, *et al.* (2005). "Transmembrane homodimerization of receptor-like protein tyrosine phosphatases." *FEBS Lett.* **579**(17): 3855-3858.
- Cho, E. G., M. G. Kim, *et al.* (2001). "N-terminal processing is essential for release of epithin, a mouse type II membrane serine protease." *J Biol Chem* **276**(48): 44581-44589.
- Chothia, C., I. Gelfand, *et al.* (1998). "Structural determinants in the sequences of immunoglobulin variable domain." *J Mol Biol* **278**(2): 457-479.
- Daniel, E. y G. Weber (1966). "Cooperative effects in binding by bovine serum albumin. I. The binding of 1-anilino-8-naphthalenesulfonate. Fluorimetric titrations." *Biochemistry* **5**(6): 1893-1900.
- Deleage, G. y C. Geourjon (1993). "An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum." *Comput Appl Biosci* **9**(2): 197-199.
- Du, X., E. She, *et al.* (2008). "The serine protease TMPRSS6 is required to sense iron deficiency." *Science* **320**(5879): 1088-1092.
- Duraisamy, S., S. Ramasamy, *et al.* (2006). "Distinct evolution of the human carcinoma-associated transmembrane mucins, MUC1, MUC4 AND MUC16." *Gene* **373**: 28-34.
- Emsley, P. y K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.
- Estrada, J., P. Bernado, *et al.* (2009). "ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble." *BMC Bioinformatics* **10**: 104.
- Fan, Y. X., J. M. Zhou, *et al.* (1998). "Unfolding and refolding of dimeric creatine kinase equilibrium and kinetic studies." *Protein Sci.* **7**(12): 2631-2641.
- Fauchère, J. L. y V. Pliska (1983). "Hydrophobic parameters π of amino acid side chains from the partitioning of N-acetyl-amino acid amines." *Eur J Med Chem* **18**: 369-375.
- Finberg, K. E., M. M. Heeney, *et al.* (2008). "Mutations in TMPRSS6 cause iron-refractory iron deficiency anemia (IRIDA)." *Nat Genet* **40**(5): 569-571.
- Fitzkee, N. C., P. J. Fleming, *et al.* (2005). "Are proteins made from a limited parts list?" *Trends Biochem Sci* **30**(2): 73-80.
- Fowler, J., L. Vinall, *et al.* (2001). "Polymorphism of the human muc genes." *Front Biosci* **6**: D1207-1215.
- Gendler, S. J. (2001). "MUC1, the renaissance molecule." *J Mammary Gland Biol Neoplasia* **6**(3): 339-353.
- Gibson, T. J., J. D. Thompson, *et al.* (1993). "The KH domain occurs in a diverse set of RNA-binding proteins that include the antiterminator NusA and is probably involved in binding to nucleic acid." *FEBS Lett* **324**(3): 361-366.
- Gill, S. C. y P. H. von Hippel (1989). "Calculation of protein extinction coefficients from amino acid sequence data." *Analytical Biochemistry* **182**(2): 319-326.
- Gross, S., C. Blanchetot, *et al.* (2002). "Multimerization of the protein-tyrosine phosphatase (PTP)-like insulin-dependent diabetes mellitus autoantigens IA-2 and IA-2beta with receptor PTPs (RPTPs). Inhibition of RPTPalpha enzymatic activity." *J. Biol. Chem.* **277**(50): 48139-48145.

- Halliwell, B. y J. M. Gutteridge (1986). "Oxygen free radicals and iron in relation to biology and medicine: some problems and concepts." *Arch Biochem Biophys* **246**(2): 501-514.
- Hamill, S. J., A. E. Meekhof, *et al.* (1998). "The effect of boundary selection on the stability and folding of the third fibronectin type III domain from human tenascin." *Biochemistry* **37**(22): 8071-8079.
- Harding, S. E. y K. Jumel (2001). "Light scattering." *Curr Protoc Protein Sci* **Chapter 7**: Unit 7 8.
- Hartwell, L. H., J. J. Hopfield, *et al.* (1999). "From molecular to modular cell biology." *Nature* **402**(6761 Suppl): C47-52.
- Hatrup, C. L. y S. J. Gendler (2008). "Structure and function of the cell surface (tethered) mucins." *Annu Rev Physiol* **70**: 431-457.
- Hebert, T. E. y M. Bouvier (1998). "Structural and functional aspects of G protein-coupled receptor oligomerization." *Biochem Cell Biol* **76**(1): 1-11.
- Henikoff, S. y J. G. Henikoff (1994). "Position-based sequence weights." *J Mol Biol* **243**(4): 574-578.
- Hermel, J. M., R. Dirx, Jr., *et al.* (1999). "Post-translational modifications of ICA512, a receptor tyrosine phosphatase-like protein of secretory granules." *Eur. J. Neurosci.* **11**(8): 2609-2620.
- Higuchi, R., B. Krummel, *et al.* (1988). "A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions." *Nucleic Acids Res* **16**(15): 7351-7367.
- Huang, X., F. Poy, *et al.* (2000). "Structure of a WW domain containing fragment of dystrophin in complex with beta-dystroglycan." *Nat Struct Biol* **7**(8): 634-638.
- Huang, Y., B. Niu, *et al.* (2010). "CD-HIT Suite: a web server for clustering and comparing biological sequences." *Bioinformatics* **26**(5): 680-682.
- Iozzo, R. V., I. R. Cohen, *et al.* (1994). "The biology of perlecan: the multifaceted heparan sulphate proteoglycan of basement membranes and pericellular matrices." *Biochem J* **302** (Pt 3): 625-639.
- Jackson, S. E. (1998). "How do small single-domain proteins fold?" *Fold Des* **3**(4): R81-91.
- Johansson, D. G., B. Macao, *et al.* (2008). "SEA domain autoproteolysis accelerated by conformational strain: mechanistic aspects." *J Mol Biol* **377**(4): 1130-1143.
- Johansson, D. G., G. Wallin, *et al.* (2009). "Protein autoproteolysis: conformational strain linked to the rate of peptide cleavage by the pH dependence of the N --> O acyl shift reaction." *J Am Chem Soc* **131**(27): 9475-9477.
- Kajimoto, Y. y H. Kaneto (2004). "Role of oxidative stress in pancreatic beta-cell dysfunction." *Ann N Y Acad Sci* **1011**: 168-176.
- Keegan, R. M. y M. D. Winn (2008). "MrBUMP: an automated pipeline for molecular replacement." *Acta Crystallogr D Biol Crystallogr* **64**(Pt 1): 119-124.
- Kelly, S. M., T. J. Jess, *et al.* (2005). "How to study proteins by circular dichroism." *Biochim Biophys Acta* **1751**(2): 119-139.
- Krissinel, E. y K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2256-2268.
- Kubosaki, A., S. Gross, *et al.* (2004). "Targeted Disruption of the IA-2 β Gene Causes Glucose Intolerance and Impairs Insulin Secretion but Does Not Prevent the Development of Diabetes in NOD Mice." *Diabetes* **53**(7): 1684-1691.

- Kubosaki, A., S. Nakamura, *et al.* (2006). "Disruption of the transmembrane dense core vesicle proteins IA-2 and IA-2beta causes female infertility." *Endocrinology* **147**(2): 811-815.
- Kubosaki, A., S. Nakamura, *et al.* (2005). "Dense core vesicle proteins IA-2 and IA-2beta: metabolic alterations in double knockout mice." *Diabetes* **54 Suppl 2**: S46-51.
- Kumar, S., C. J. Tsai, *et al.* (2001). "Thermodynamic differences among homologous thermophilic and mesophilic proteins." *Biochemistry* **40**(47): 14152-14165.
- Lan, M. S., J. Lu, *et al.* (1994). "Molecular cloning and identification of a receptor-type protein tyrosine phosphatase, IA-2, from human insulinoma." *DNA Cell Biol* **13**(5): 505-514.
- Lan, M. S., C. Wasserfall, *et al.* (1996). "IA-2, a transmembrane protein of the protein tyrosine phosphatase family, is a major autoantigen in insulin-dependent diabetes mellitus." *Proceedings of the National Academy of Sciences* **93**(13): 6367-6370.
- Larson, S. M. y A. R. Davidson (2000). "The identification of conserved interactions within the SH3 domain by alignment of sequences and structures." *Protein Sci* **9**(11): 2170-2180.
- Laskowski, M. (1955). "Trypsinogen and trypsin." *Methods in Enzymology* **2**: 26-36.
- Lattman, E. E. y G. D. Rose (1993). "Protein folding--what's the question?" *Proc Natl Acad Sci U S A* **90**(2): 439-441.
- Levitin, F., O. Stern, *et al.* (2005). "The MUC1 SEA module is a self-cleaving domain." *J Biol Chem* **280**(39): 33374-33386.
- Light, A. y H. Janska (1989). "Enterokinase (enteropeptidase): comparative aspects." *Trends Biochem Sci* **14**(3): 110-112.
- Ligtenberg, M. J., L. Kruijsaar, *et al.* (1992). "Cell-associated episialin is a complex containing two proteins derived from a common precursor." *J Biol Chem* **267**(9): 6171-6177.
- Lu, K. V., M. F. Rohde, *et al.* (1995). "Mistranslation of a TGA termination codon as tryptophan in recombinant platelet-derived growth factor expressed in Escherichia coli." *Biochem J* **309 (Pt 2)**: 411-417.
- Macao, B., D. G. Johansson, *et al.* (2006). "Autoproteolysis coupled to protein folding in the SEA domain of the membrane-bound MUC1 mucin." *Nat Struct Mol Biol* **13**(1): 71-76.
- MacBeath, G. y P. Kast (1998). "UGA read-through artifacts--when popular gene expression systems need a pATCH." *Biotechniques* **24**(5): 789-794.
- Maeda, T., M. Inoue, *et al.* (2004). "Solution structure of the SEA domain from the murine homologue of ovarian cancer antigen CA125 (MUC16)." *J Biol Chem* **279**(13): 13174-13182.
- Maggino, T. y A. Gadducci (2000). "Serum markers as prognostic factors in epithelial ovarian cancer: an overview." *Eur J Gynaecol Oncol* **21**(1): 64-69.
- Melis, M. A., M. Cau, *et al.* (2008). "A mutation in the TMPRSS6 gene, encoding a transmembrane serine protease that suppresses hepcidin production, in familial iron deficiency anemia refractory to oral iron." *Haematologica* **93**(10): 1473-1479.
- Mottagui-Tabar, S. y L. A. Isaksson (1997). "Only the last amino acids in the nascent peptide influence translation termination in Escherichia coli genes." *FEBS Lett* **414**(1): 165-170.
- Motulsky, H. y A. Christopoulos (2003). *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*. San Diego, GraphPad Software Inc.
- Murshudov, G. N., A. A. Vagin, *et al.* (1997). "Refinement of macromolecular structures by the maximum-likelihood method." *Acta Crystallogr D Biol Crystallogr* **53**(Pt 3): 240-255.
- Musco, G., G. Stier, *et al.* (1996). "Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome." *Cell* **85**(2): 237-245.

- Mushegian, A. R., D. E. Bassett, *et al.* (1997). "Positionally cloned human disease genes: Patterns of evolutionary conservation and functional motifs." *Proceedings of the National Academy of Sciences* **94**(11): 5831-5836.
- Myers, J. K., C. N. Pace, *et al.* (1995). "Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding." *Protein Sci.* **4**(10): 2138-2148.
- Mziaut, H., M. Trajkovski, *et al.* (2006). "Synergy of glucose and growth hormone signalling in islet cells through ICA512 and STAT5." *Nat Cell Biol* **8**(5): 435-445.
- Nastuk, M. A. y J. R. Fallon (1993). "Agrin and the molecular choreography of synapse formation." *Trends Neurosci* **16**(2): 72-76.
- Notkins, A. L., B. Zhang, *et al.* (1997). "Comparison of IA-2 with IA-2beta and with six other members of the protein tyrosine phosphatase family: recognition of antigenic determinants by IDDM sera." *J Autoimmun* **10**(3): 245-250.
- O'Brien, T. J., J. B. Beard, *et al.* (2001). "The CA 125 gene: an extracellular superstructure dominated by repeat sequences." *Tumour Biol* **22**(6): 348-366.
- Oberst, M. D., C. A. Williams, *et al.* (2003). "The activation of matriptase requires its noncatalytic domains, serine protease domain, and its cognate inhibitor." *J Biol Chem* **278**(29): 26773-26779.
- Ort, T., S. Voronov, *et al.* (2001). "Dephosphorylation of beta2-syntrophin and Ca²⁺/mu-calpain-mediated cleavage of ICA512 upon stimulation of insulin secretion." *EMBO J* **20**(15): 4013-4023.
- Otwinowski, Z. y W. Minor (1997). [20] Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* Charles W. Carter, Jr., Academic Press. **Volume 276**: 307-326.
- Pace, C. N. (1986). "Determination and analysis of urea and guanidine hydrochloride denaturation curves." *Methods Enzymol* **131**: 266-280.
- Pace, C. N. y M. Scholtz (1997). Measuring the Conformational Stability of a Protein. *Protein Structure: A Practical Approach*. T. Creighton. New York, Oxford University Press: 299-320.
- Painter, J. y E. A. Merritt (2006). "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion." *Acta Crystallogr D Biol Crystallogr* **62**(Pt 4): 439-450.
- Paroutis, P., N. Touret, *et al.* (2004). "The pH of the secretory pathway: measurement, determinants, and regulation." *Physiology (Bethesda)* **19**: 207-215.
- Parry, S., H. S. Silverman, *et al.* (2001). "Identification of MUC1 proteolytic cleavage sites in vivo." *Biochem Biophys Res Commun* **283**(3): 715-720.
- Patterson, K. I., T. Brummer, *et al.* (2009). "Dual-specificity phosphatases: critical regulators with diverse cellular targets." *Biochem J* **418**(3): 475-489.
- Patthy, L. y K. Nikolics (1993). "Functions of agrin and agrin-related proteins." *Trends Neurosci* **16**(2): 76-81.
- Paulus, H. (2000). "Protein splicing and related forms of protein autoprocessing." *Annu Rev Biochem* **69**: 447-496.
- Peng, Z., Y. He, *et al.* (2010). "Autoproteolysis of the SEA module of rMuc3 C-terminal domain modulates its functional composition." *Arch Biochem Biophys* **503**(2): 238-247.
- Perrett, S., S. J. Freeman, *et al.* (1999). "Equilibrium folding properties of the yeast prion protein determinant Ure2." *J. Mol. Biol.* **290**(1): 331-345.
- Plaxco, K. W., K. T. Simons, *et al.* (1998). "Contact order, transition state placement and the refolding rates of single domain proteins." *J Mol Biol* **277**(4): 985-994.

- Poole, E. S., C. M. Brown, *et al.* (1995). "The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*." *EMBO J* **14**(1): 151-158.
- Press, W. H., S. A. Teukolsky, *et al.* (1992). *Numerical Recipes in C. The art of scientific computing*. New York, Cambridge University Press.
- Primo, M. E., J. Jakoncic, *et al.* (2011). "Protein-protein interactions in crystals of the human receptor-type protein tyrosine phosphatase ICA512 ectodomain." *PLoS One* **6**(9): e24191.
- Primo, M. E., S. Klinke, *et al.* (2008). "Structure of the Mature Ectodomain of the Human Receptor-type Protein-tyrosine Phosphatase IA-2." *Journal of Biological Chemistry* **283**(8): 4674-4681.
- Privalov, P. L. (1979). "Stability of proteins: small globular proteins." *Adv Protein Chem* **33**: 167-241.
- Radzicka, A. y R. Wolfenden (1996). "Rates of Uncatalyzed Peptide Bond Hydrolysis in Neutral Solution and the Transition State Affinities of Proteases." *J Am Chem Soc* **118**(26): 6105-6109.
- Riddles, P. W., R. L. Blakeley, *et al.* (1983). "Reassessment of Ellman's reagent." *Methods Enzymol* **91**: 49-60.
- Rios, C. D., B. A. Jordan, *et al.* (2001). "G-protein-coupled receptor dimerization: modulation of receptor function." *Pharmacol Ther* **92**(2-3): 71-87.
- Rumfeldt, J. A., C. Galvagnion, *et al.* (2008). "Conformational stability and folding mechanisms of dimeric proteins." *Prog Biophys Mol Biol* **98**(1): 61-84.
- Rustin, G. J., M. Marples, *et al.* (2001). "Use of CA-125 to define progression of ovarian cancer in patients with persistently elevated levels." *J Clin Oncol* **19**(20): 4054-4057.
- Saeki, K., M. Zhu, *et al.* (2002). "Targeted disruption of the protein tyrosine phosphatase-like molecule IA-2 results in alterations in glucose tolerance tests and insulin secretion." *Diabetes* **51**(6): 1842-1850.
- Sallee, J. L., E. S. Wittchen, *et al.* (2006). "Regulation of cell adhesion by protein-tyrosine phosphatases: II. Cell-cell adhesion." *J Biol Chem* **281**(24): 16189-16192.
- Sambrook, J., E. F. Fritsh, *et al.* (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press.
- Sandberg, A., D. G. Johansson, *et al.* (2008). "SEA domain autoproteolysis accelerated by conformational strain: energetic aspects." *J Mol Biol* **377**(4): 1117-1129.
- Sasaki, T., W. Gohring, *et al.* (1995). "Binding of mouse and human fibulin-2 to extracellular matrix ligands." *J Mol Biol* **254**(5): 892-899.
- Schägger, H. y G. von Jagow (1987). "Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa." *Analytical Biochemistry* **166**(2): 368-379.
- Semisotnov, G. V., N. A. Rodionova, *et al.* (1991). "Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe." *Biopolymers* **31**(1): 119-128.
- Solimena, M., R. Dirkx, Jr., *et al.* (1996). "ICA 512, an autoantigen of type I diabetes, is an intrinsic membrane protein of neurosecretory granules." *EMBO J* **15**(9): 2102-2114.
- Sonnhammer, E. L., S. R. Eddy, *et al.* (1998). "Pfam: multiple sequence alignments and HMM-profiles of protein domains." *Nucleic Acids Res* **26**(1): 320-322.
- Stoker, A. W. (2005). "Protein tyrosine phosphatases and signalling." *J Endocrinol* **185**(1): 19-33.

- Szilagyi, A. y P. Zavodszky (2000). "Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey." *Structure* **8**(5): 493-504.
- Teichmann, S. A., C. Chothia, *et al.* (1999). "Advances in structural genomics." *Curr Opin Struct Biol* **9**(3): 390-399.
- Tiganis, T. y A. M. Bennett (2007). "Protein tyrosine phosphatase function: the substrate perspective." *Biochem J* **402**(1): 1-15.
- Tomiko Kawakami , K. S., Natsumi Takeyama , Guoying Wu , Akikazu Sakudo , Yoshitsugu Matsumoto , Toshiharu Hayashi , Takashi Onodera (2007). "Detection of proteolytic cleavages of diabetes-associated protein IA-2 β in the pancreas and the brain using novel anti-IA-2 β monoclonal antibodies." *International Journal of Molecular Medicine* **20**(12): 177-185.
- Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." *Nat Rev Mol Cell Biol* **7**(11): 833-846.
- Torii, S. (2009). "Expression and function of IA-2 family proteins, unique neuroendocrine-specific protein-tyrosine phosphatases." *Endocr. J.* **56**(5): 639-648.
- Torii, S., N. Saito, *et al.* (2009). "Gene Silencing of Phogrin Unveils Its Essential Role in Glucose-Responsive Pancreatic β -Cell Growth." *Diabetes* **58**(3): 682-692.
- Trajkovski, M., H. Mziaut, *et al.* (2004). "Nuclear translocation of an ICA512 cytosolic fragment couples granule exocytosis and insulin expression in β -cells." *J Cell Biol* **167**(6): 1063-1074.
- Tsen, G., W. Halfter, *et al.* (1995). "Agrin is a heparan sulfate proteoglycan." *J Biol Chem* **270**(7): 3392-3399.
- Uversky, V. N. (1993). "Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule." *Biochemistry* **32**(48): 13288-13298.
- Vagin, A. y A. Teplyakov (2010). "Molecular replacement with MOLREP." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 1): 22-25.
- Velasco, G., S. Cal, *et al.* (2002). "Matriptase-2, a membrane-bound mosaic serine proteinase predominantly expressed in human liver and showing degrading activity against extracellular matrix proteins." *J Biol Chem* **277**(40): 37637-37646.
- Wallin, G., T. Hård, *et al.* (2012). "Folding-Reaction Coupling in a Self-Cleaving Protein." *Journal of Chemical Theory and Computation*.
- Warshel, A., P. K. Sharma, *et al.* (2006). "Electrostatic basis for enzyme catalysis." *Chem Rev* **106**(8): 3210-3235.
- Wasmeier, C. y J. C. Hutton (1996). "Molecular Cloning of Phogrin, a Protein-tyrosine Phosphatase Homologue Localized to Insulin Secretory Granule Membranes." *Journal of Biological Chemistry* **271**(30): 18161-18170.
- Weiland, F., K. Martin, *et al.* (2012). "Deciphering the Molecular Nature of Ovarian Cancer Biomarker CA125." *Int J Mol Sci* **13**(8): 10568-10582.
- Wilkins, D. K., S. B. Grimshaw, *et al.* (1999). "Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques." *Biochemistry* **38**(50): 16424-16431.
- Winder, S. J. (2001). "The complexities of dystroglycan." *Trends Biochem Sci* **26**(2): 118-124.
- Wreschner, D. H., M. A. McGuckin, *et al.* (2002). "Generation of ligand-receptor alliances by "SEA" module-mediated cleavage of membrane-associated mucin proteins." *Protein Sci* **11**(3): 698-706.

- Wu, M. M., M. Grabe, *et al.* (2001). "Mechanisms of pH regulation in the regulated secretory pathway." *J Biol Chem* **276**(35): 33027-33035.
- Zarrine-Afsar, A., S. M. Larson, *et al.* (2005). "The family feud: do proteins with similar structures fold via the same pathway?" *Curr Opin Struct Biol* **15**(1): 42-49.

Apéndices

Apéndice 1: Caracterización estructural de IA-2b⁴⁸⁶⁻⁶¹⁴

Secuencias:

IA-2b⁴⁸⁶⁻⁶¹⁴:

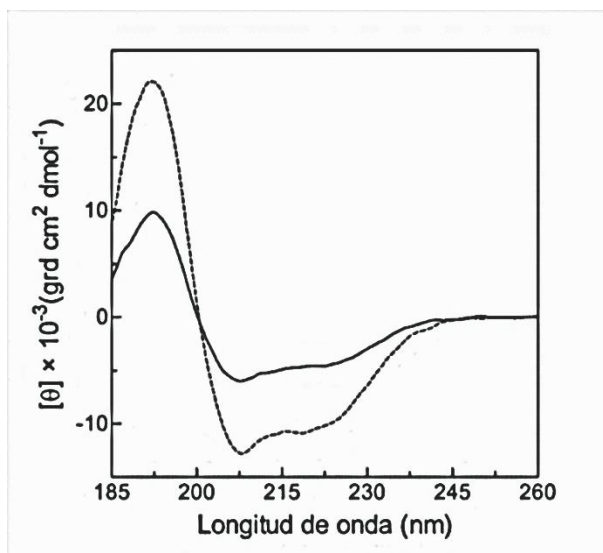
SLPAGAEALSDGLQLEVQPSEEEARGY | ... | KILQTGVGSKSKLKFLPPQAEQEDSTK

MPE fogrina:

MEVQPSEEEARGY | ... | KILQTGVGSKSK

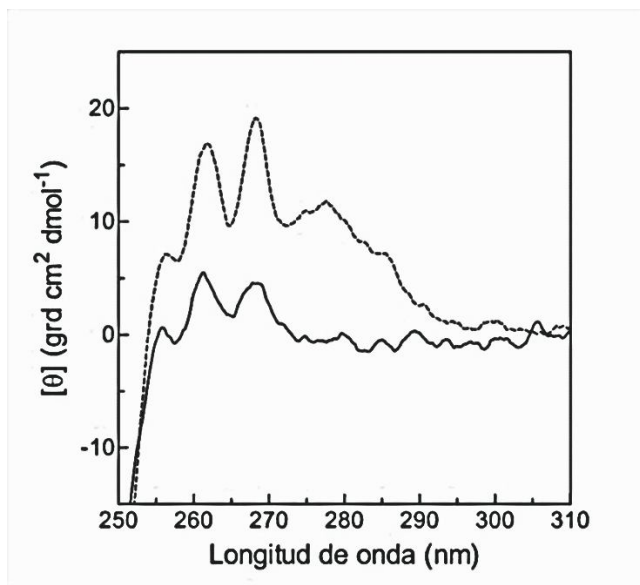
En la realización de todos estos experimentos se asumió un coeficiente de extinción molar de IA-2b⁴⁸⁶⁻⁶¹⁴ idéntico al determinado para MPE fogrina, lo que es razonable dado que poseen el mismo contenido de residuos aromáticos.

Espectro de CD en el UV lejano:



En línea continua se muestra el espectro de IA-2b⁴⁸⁶⁻⁶¹⁴ y en línea de trazos el correspondiente a MPE fogrina. El *buffer* fue fosfato de sodio 25 mM pH 7.0 y la temperatura 20 °C. Se observa que el contenido de estructura secundaria es menor que en MPE fogrina (tanto al expresar la elipticidad por mol de residuos, como se muestra en la figura, como al hacerlo por mol de proteína, no mostrado).

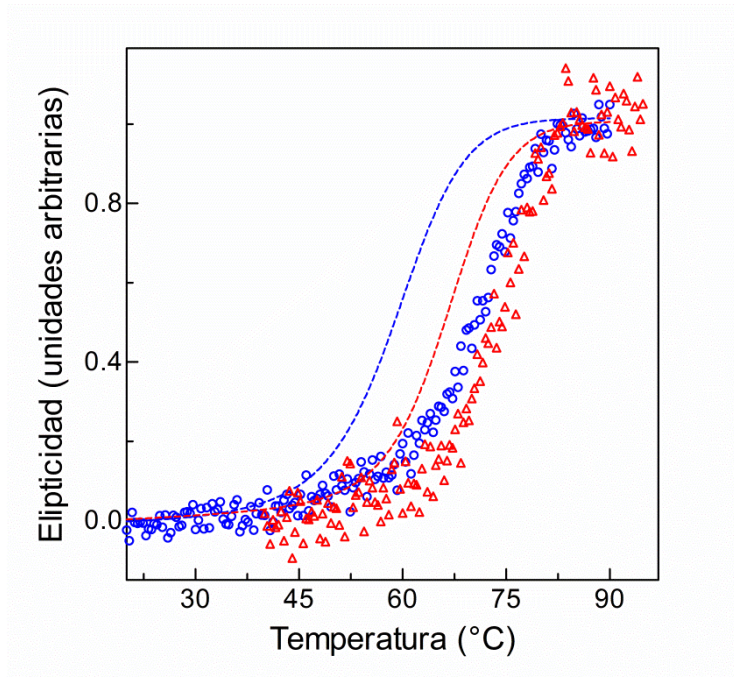
Espectro de CD en el UV cercano:



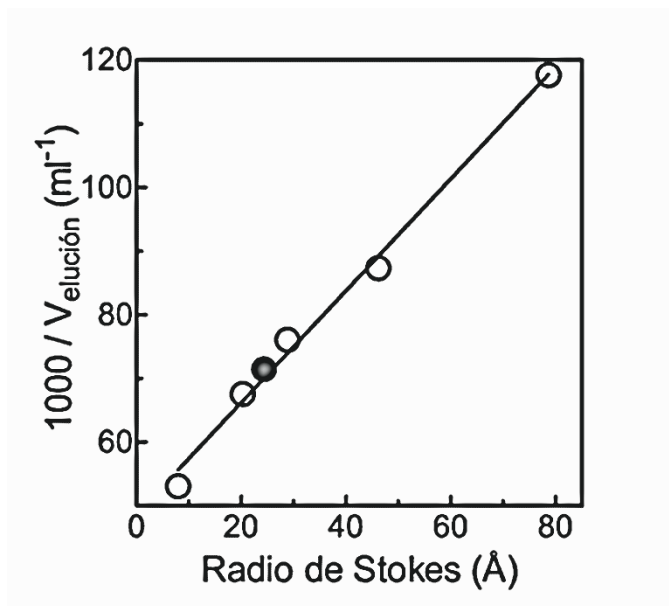
En línea continua se muestra el espectro de IA-2b⁴⁸⁶⁻⁶¹⁴ y en línea de trazos el correspondiente a MPE fogrina. El *buffer* fue fosfato de sodio 25 mM pH 7.0 y la temperatura 20 °C. La señal de estructura terciaria es menor tanto si se expresa la señal por mol de residuos (mostrada en la figura) como por mol de proteína (no mostrado). Se observa la desaparición de la señal de la única tirosina de la molécula, que está ubicada en el extremo N.

Estabilidad frente al desplegado por temperatura:

Se monitoreó la elipticidad a 220 nm mientras se calentaba a velocidad constante. Se estudió el efecto a pH 7.0 y 4.5 (puntos azules y rojos, respectivamente) y para comparar se muestran las transiciones de MPE fogrina a pH 7.0 y 4.8 (trazos azules y rojos, respectivamente). IA-2b⁴⁸⁶⁻⁶¹⁴ es más estable en ambas condiciones (los datos no se ajustaron al modelo de desplegado porque la reversibilidad a pH 7.0 fue de 91 % y a pH 4,5 fue de 86 % y además la muestra ya se encontraba parcialmente degradada por el mecanismo descrito en la Sección 2.9).



Estado de agregación:



Se hicieron corridas de exclusión molecular como se indica en Materiales y métodos, en un rango de concentraciones de proteína inyectada desde 3.5 a 220 μM . El cambio en el volumen de elución no depende significativamente de la concentración de proteína inyectada. En el gráfico se muestra la curva de calibración usando las proteínas estándar, y se indica como un punto negro al volumen de elución observado, con un R_s correspondiente al esperado para una proteína dimerica del PM de IA-2b⁴⁸⁶⁻⁶¹⁴. Esto indica que IA-2b⁴⁸⁶⁻⁶¹⁴ se comporta de la misma forma que MPE fogrina en SEC, pero como se explica en la Sección

2.4, esta metodología no es apropiada para determinar el estado de agregación de fognina por problemas de interacción con la matriz de exclusión molecular.

Apéndice 2: Alineamiento de secuencias de la región MPE de las RPTP

El alineamiento completo se muestra a continuación, dividido en tres segmentos:

SS MPE ICA512	-	-	B	B	B	B	B	B	C	C	C	C	C	C	-	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	C	
	1	2	3	4	5	6	7	8	9	10	11	12	18	19	20	30	31	32	33	34	35	36	37	38	39	40	42	43	44	45	46	47	
[MPE ICA512	E	E	Y	G	Y	I	V	T	D	Q	K	P	L	S	-	L	A	A	G	V	K	L	L	E	I	L	A	E	H	V	H	M	
gi 363736259	D	E	Y	G	Y	I	V	T	D	Q	K	P	L	G	-	L	A	A	G	V	R	L	L	E	I	L	A	K	R	V	H	L	
gi 327260586	E	E	Y	G	Y	I	V	T	D	Q	K	P	L	G	-	L	A	A	G	I	R	L	L	E	I	L	A	E	H	V	H	L	
gi 344298205	E	E	Y	G	Y	I	V	T	E	N	S	P	L	S	-	L	E	K	G	T	K	L	I	E	D	I	A	H	L	L	K	V	
gi 334348781	E	E	Y	G	Y	I	M	T	E	E	D	P	L	S	-	V	E	K	G	K	K	L	I	K	E	V	A	Q	L	L	K	L	
gi 327274430	E	E	Y	G	Y	I	V	T	E	K	D	P	L	S	-	E	E	K	G	L	E	L	I	R	E	V	A	P	V	L	L	K	L
gi 345324216	E	E	Y	G	Y	I	V	T	E	K	D	P	L	S	-	V	E	K	G	K	K	L	E	N	V	A	P	P	L	L	K	L	
gi 189230214	E	E	Y	G	Y	I	V	T	N	Q	K	P	L	S	-	L	L	S	G	A	Q	L	L	E	T	L	A	Q	R	V	H	L	
gi 194578855	E	E	Y	G	Y	I	V	T	N	Q	S	H	L	S	-	L	Y	D	G	V	K	L	L	E	T	L	A	E	R	I	H	L	
gi 348515725	E	E	Y	G	Y	I	V	T	N	Q	S	P	L	S	-	L	Y	D	G	V	K	M	L	E	L	L	A	D	K	I	H	L	
gi 224044668	E	E	Y	G	Y	I	V	T	V	K	D	P	L	S	-	V	E	K	G	L	E	L	I	K	E	V	A	D	L	L	K	L	
gi 126540835	G	D	F	G	Y	I	I	T	E	-	D	S	L	S	-	T	D	Q	G	L	N	L	M	E	I	L	A	E	R	V	R	L	
gi 348520004	D	V	F	G	Y	V	I	T	D	T	D	G	L	Q	-	T	D	E	G	L	H	L	M	E	I	L	A	R	M	A	K	I	
gi 348503674	R	H	F	G	Y	I	I	T	G	T	D	S	L	T	-	T	D	Q	G	T	D	L	M	E	H	L	T	Q	R	L	S	L	
gi 348567400	E	E	P	G	Y	I	V	T	G	S	D	L	L	S	-	P	E	K	G	K	Q	L	M	S	D	V	A	H	L	L	Q	V	
gi 194666632	-	E	Y	G	Y	I	L	M	E	N	D	S	L	S	-	L	D	E	G	T	Q	A	L	E	G	V	A	R	L	L	E	V	
gi 317420011	R	H	F	G	Y	I	I	T	G	S	D	S	L	T	-	A	D	Q	G	S	D	L	M	E	R	L	T	Q	R	V	N	L	
[MPE fognina	E	A	R	G	Y	I	V	T	D	R	D	P	L	R	-	P	E	E	G	R	R	L	V	E	D	V	A	R	L	L	Q	V	
gi 194210195	-	-	S	G	Y	I	V	T	E	T	D	P	L	S	-	L	E	K	G	K	E	L	L	A	G	V	A	R	L	L	E	V	
gi 354489421	E	Q	Q	G	Y	I	L	T	G	N	D	P	L	S	-	P	E	K	G	K	Q	L	M	D	E	V	A	H	I	L	G	V	
gi 13928818	E	Q	Q	G	Y	I	L	T	G	N	N	P	L	S	-	P	E	K	G	K	Q	L	M	D	E	V	A	H	L	L	R	V	
gi 351715014	E	E	L	G	Y	I	V	T	D	T	D	P	L	S	-	P	E	K	G	R	Q	L	M	E	D	V	A	R	L	L	Q	V	
gi 345781298	E	E	L	G	Y	I	V	T	G	N	D	P	L	S	-	P	G	D	G	K	E	L	L	V	S	V	A	R	L	L	E	V	
gi 328719556	T	S	Y	V	C	I	R	L	K	E	-	K	I	S	N	P	D	R	G	E	K	L	L	K	T	L	A	S	I	I	N	V	
gi 118789314	T	E	Y	A	H	V	I	L	K	N	-	P	I	D	N	W	K	D	G	S	R	I	V	A	A	L	A	E	L	L	N	M	
gi 189241159	T	D	Y	V	Y	V	E	F	E	N	-	D	V	Y	K	W	A	E	G	K	Q	L	I	G	N	I	S	D	L	L	G	L	
gi 170029348	T	E	Y	A	H	V	I	L	R	N	-	P	I	D	N	W	K	D	G	S	R	I	V	T	A	L	A	E	L	L	N	M	
gi 350398975	L	D	H	V	Y	V	E	F	Q	K	-	P	F	H	R	W	S	E	G	E	H	V	V	E	E	V	E	K	L	L	G	L	
gi 242021897	E	D	Y	A	Y	V	I	L	K	N	-	K	V	N	T	W	Q	E	G	S	R	I	V	R	K	V	E	E	L	L	D	L	
gi 321477883	T	N	Y	A	F	I	T	I	K	G	-	E	F	L	N	W	R	Q	G	T	E	F	V	K	A	L	C	D	V	L	K	I	
gi 260788939	R	N	H	A	F	L	V	F	D	R	-	K	V	T	-	Y	D	D	G	T	K	V	R	D	A	L	A	D	I	I	G	V	
gi 307175617	L	D	Y	V	Y	I	I	F	E	Q	-	E	F	H	T	W	L	E	G	G	H	V	I	A	K	V	E	E	L	L	G	L	
gi 307209208	L	D	Y	V	Y	I	E	F	E	Q	-	E	F	H	K	W	L	E	G	E	H	V	V	A	K	V	G	E	L	L	D	L	
gi 332019898	L	D	Y	V	Y	I	E	F	E	Q	-	E	F	H	T	W	L	E	G	E	H	V	V	N	K	V	E	E	L	L	G	L	
gi 195437839	T	E	Y	A	H	V	F	V	K	N	-	P	I	D	S	W	Y	D	G	Q	R	I	M	N	E	L	A	Q	I	L	H	M	
gi 357624369	P	S	Y	A	F	V	T	F	H	N	R	F	L	T	D	W	E	K	G	I	S	F	I	T	R	L	E	E	M	L	G	L	

SS MPE ICA512		C	C	C	C	B	B	B	B	B	B	B	C	C	B	B	B	B	B	B	B	C	C	-	C	C	C	C	C	H	H	H	H		
		48	49	50	51	60	61	62	63	64	65	66	67	75	76	77	78	79	80	81	82	83	84	84 _a	85	86	87	92	93	94	95	96	97		
MPE ICA512	gi 363736259	S	S	G	S	F	I	N	I	S	V	V	G	P	A	L	T	F	R	I	R	H	N	-	E	Q	N	L	S	L	A	D	V		
	gi 327260586	S	T	A	S	F	I	N	I	S	V	V	G	P	A	L	T	F	R	I	R	Q	N	-	H	Q	N	L	S	L	G	D	V		
	gi 344298205	P	V	S	T	F	A	D	I	S	V	L	G	P	A	V	T	F	R	V	T	S	N	-	I	Q	N	V	T	T	A	A	V		
	gi 334348781	P	P	T	A	I	V	D	V	N	V	L	G	P	A	V	T	F	K	V	S	S	S	-	F	Q	N	M	T	T	A	E	V		
	gi 327274430	Q	M	T	A	F	A	D	I	N	I	L	G	A	A	V	T	F	K	V	H	S	N	-	A	L	N	V	T	T	A	D	V		
	gi 345324216	Q	T	S	I	F	A	D	I	N	V	L	G	P	A	V	T	F	K	V	S	S	N	-	D	Q	N	I	T	T	A	D	V		
	gi 189230214	S	T	A	S	F	I	N	I	S	V	V	G	S	A	L	T	F	R	I	R	Q	N	-	N	Q	N	L	S	L	T	D	V		
	gi 194578855	S	T	S	S	F	I	N	I	S	V	V	G	S	A	L	T	F	R	I	R	Q	N	-	A	L	N	L	S	A	E	D	V		
	gi 348515725	T	T	S	S	F	I	N	I	S	V	V	G	P	A	L	T	F	R	I	R	Q	N	-	E	H	N	M	T	A	A	E	V		
	gi 224044668	Q	M	S	V	F	D	D	V	N	V	L	G	P	A	V	T	F	R	V	H	A	N	-	L	Q	N	I	S	T	A	D	V		
	gi 126540835	H	V	T	D	F	L	Q	L	S	V	L	G	P	A	V	T	F	R	L	R	P	N	-	A	R	N	I	S	T	A	Q	V		
	gi 348520004	Q	M	S	D	F	A	E	L	S	V	V	G	P	A	V	T	F	K	V	S	P	N	-	A	Q	N	V	S	T	A	D	V		
	gi 348503674	H	A	A	D	L	T	Q	L	S	V	L	G	P	A	L	T	F	R	V	G	P	N	-	P	K	N	V	T	T	A	D	V		
	gi 348567400	P	P	S	I	F	E	D	I	E	V	V	G	P	A	V	T	F	K	V	G	A	N	-	A	R	N	M	T	T	A	D	V		
	gi 194666632	P	A	G	V	F	T	D	V	E	V	E	G	P	A	V	T	F	R	V	R	A	N	-	A	Q	N	V	T	T	A	A	A	V	
	gi 317420011	H	A	T	D	L	T	Q	L	S	V	L	G	P	A	L	T	F	R	I	R	G	P	N	-	P	K	N	V	T	T	A	D	L	
MPE fogrina	gi 194210195	P	S	S	A	F	A	D	V	E	V	L	G	P	A	V	T	F	K	V	S	A	N	-	V	Q	N	V	T	T	E	D	V	V	
	gi 354489421	P	M	S	V	F	V	D	I	D	V	V	L	G	P	A	V	T	F	K	V	S	A	N	-	V	L	N	V	T	T	A	E	A	V
	gi 13928818	P	S	G	F	F	A	D	V	R	V	L	G	P	A	V	T	F	K	V	S	A	N	-	I	Q	N	M	T	T	A	D	V	V	
	gi 13928818	P	S	S	F	F	A	D	V	K	V	L	G	P	A	V	I	F	K	V	S	A	N	-	I	Q	N	M	T	T	A	D	V	V	
	gi 351715014	P	P	N	I	F	E	D	I	D	V	L	G	P	A	V	T	F	K	V	G	A	N	-	V	R	N	V	T	T	A	D	V	V	
	gi 345781298	P	M	S	V	F	A	D	I	T	V	L	G	P	A	V	T	F	R	V	S	A	S	-	I	W	N	L	T	T	A	D	V	V	
	gi 328719556	E	F	K	S	M	V	N	A	R	F	E	E	S	E	I	T	F	Q	I	L	P	S	-	-	A	G	I	N	V	T	E	L	V	
	gi 118789314	Q	G	-	Y	F	T	H	P	R	V	D	R	H	E	V	S	F	R	V	E	A	N	-	P	D	R	K	T	A	A	D	V	V	
	gi 189241159	E	R	S	A	F	S	H	G	R	V	D	D	H	N	E	V	S	F	K	V	E	P	N	-	N	K	G	W	N	S	T	Q	V	
	gi 170029348	Q	G	-	Y	F	T	H	P	R	V	D	R	H	E	V	S	F	R	V	E	S	N	-	P	D	K	K	T	A	A	D	V	V	
	gi 350398975	S	S	G	S	L	K	D	I	R	V	G	R	A	E	V	T	F	K	V	V	K	N	-	N	K	D	Y	S	A	T	D	V	V	
	gi 242021897	P	Q	N	T	F	T	N	I	R	V	D	Q	N	S	V	K	F	K	I	D	S	N	-	R	L	N	L	N	A	S	E	T	V	
	gi 321477883	H	H	N	N	L	Y	D	V	S	V	L	K	N	K	V	S	F	R	I	Q	T	Q	-	D	K	G	L	N	P	S	L	I	V	
	gi 260788939	P	T	E	I	F	S	D	I	S	V	L	D	N	H	V	S	F	L	V	R	P	N	-	D	K	G	L	N	A	S	T	V	V	
	gi 307175617	P	S	G	T	L	T	D	I	R	V	G	R	A	E	V	T	F	K	V	N	Q	N	-	D	K	N	Y	N	A	T	D	V	V	
	gi 307209208	P	P	G	T	L	K	D	I	R	V	G	R	A	E	V	T	F	K	V	V	P	N	-	F	K	N	Y	N	A	T	D	V	V	
	gi 332019898	P	S	G	T	L	K	D	I	R	V	G	R	A	E	V	T	F	K	V	I	Q	N	-	D	K	N	Y	N	A	T	E	V	V	
	gi 195437839	Q	G	-	Y	F	S	Y	L	T	V	Q	P	H	E	V	S	F	R	V	D	S	N	N	-	P	E	R	K	T	A	G	D	V	
	gi 357624369	E	K	N	T	F	T	N	P	R	V	D	P	S	E	V	T	F	K	V	E	K	N	-	S	K	G	Y	D	A	A	D	V	V	

SS MPE ICA512		H	H	H	H	H	H	-	-	-	H	H	H	H	H	H	H	C	C	C	B	B	B	B	B	B	B	-	-			
		98	99	100	101	102	103	104	104 _a	104 _b	104 _c	105	107	108	109	110	111	112	113	114	115	116	117	118	123	124	125	126	127	128	129	
MPE ICA512	gi 363736259	T	Q	Q	A	G	L	V	-	-	-	K	S	E	L	E	A	Q	T	G	L	Q	I	L	Q	T	G	V	G	Q	R	
	gi 327260586	A	N	R	A	E	Q	M	-	-	-	K	A	E	L	E	A	E	L	G	L	K	I	V	Q	T	G	V	G	Q	R	
	gi 344298205	A	K	A	A	V	D	N	-	-	-	K	D	K	L	E	A	T	S	R	V	K	I	L	R	A	G	I	G	S	K	
	gi 334348781	A	K	A	T	A	D	N	-	-	-	K	E	K	L	E	K	A	S	G	L	K	I	L	R	T	G	I	G	S	K	
	gi 327274430	A	K	A	A	A	D	N	-	-	-	K	D	M	L	E	K	T	T	G	L	K	I	V	Q	I	G	M	G	E	K	
	gi 345324216	A	K	A	A	V	D	N	-	-	-	K	D	K	L	E	K	L	S	G	L	K	I	L	Q	T	G	I	G	A	K	
	gi 189230214	A	D	R	A	V	A	L	-	-	-	K	S	D	L	E	T	E	T	G	V	K	I	I	Q	T	G	V	G	Q	R	
	gi 194578855	A	N	K	A	V	A	E	-	-	-	K	N	F	L	E	G	E	T	G	L	K	I	V	Q	T	G	V	G	E	R	
	gi 348515725	A	A	K	A	V	S	E	-	-	-	K	N	F	L	E	S	E	T	G	L	K	I	V	Q	T	G	V	G	E	R	
	gi 224044668	A	K	E	A	A	I	N	-	-	-	K	E	K	L	E	K	T	T	G	L	K	I	L	Q	T	G	V	G	E	K	
	gi 126540835	A	R	V	A	V	E	Q	-	-	-	K	G	Q	S	L	E	K	E	S	G	L	R	I	L	E	A	G	L	S	D	R
	gi 348520004	A	N	V	A	V	Q	K	-	-	-	K	V	S	L	E	K	E	A	K	V	S	I	L	L	E	A	G	I	A	D	G
	gi 348503674	V	Q	V	A	V	Q	Q	-	-	-	K	Q	Q	S	L	E	K	E	T	G	L	K	I	V	E	A	G	V	S	D	-
	gi 348567400	V	K	A	T	A	D	H	-	-	-	R	N	Q	L	E	E	A	S	G	L	R	I	L	Q	S	G	T	G	P	E	
	gi 194666632	A	K	A	A	A	D	N	-	-	-	R	D	K	L	Q	T	M	S	G	L	R	V	L	Q	A	G	V	G	S	R	
	gi 317420011	V	Q	V	A	V	Q	Q	-	-	-	K	Q	Q	L	E	K	E	T	G	M	K	I	V	E	A	G	V	S	D	R	
MPE fogrina	gi 194210195	E	K	A	T	V	D	N	-	-	-	K	D	K	L	E	E	T	S	G	L	K	I	L	Q	T	G	V	G	S	K	K
	gi 354489421	V	E	A	A	V	E	N	-	-	-	K	D	N	L	E	K	T	S	G	L	K	I	L	Q	T	G	V	G	S	K	K
	gi 13928818	I	K	A	T	V	D	N	-	-	-	K	D	Q	L	E	K	T	S	G	L	T	I	L	Q	S	G	I	P	K	K	
	gi 351715014	T	K	A	A	V	D	N	-	-	-	K	D	E	L	E	K	A	T	G	L	T	I	L	Q	S	G	I	R	P	K	
	gi 345781298	A	R	A	T	A	D	H	-	-	-	K	N	Q	L	A	D	T	S	E	L	K	I	L	Q	S	G	T	G	P	G	
	gi 328719556	V	Q	A	T	V	N	N	-	-	-	K	E	K	L	E	A	A	S	G	L	K	I	L	H	V	G	S	G	L	K	
	gi 118789314	V	K	K	I	E	T	M	-	-	-	K	N	K	F	A	K	K	T	G	Y	T	I	D	F	V	S	I	G	D	K	
	gi 189241159	A	K	Q	I	N	D	S	R	-	-	F	K	N	T	L	S	R	R	L	G	V	I	V	I	Q	A	G	V	G	D	K
	gi 170029348	A	E	Q	I	G	K	I	-	-	-	K	D	T	L	R	E	V	T	G	V	I	V	I	T	A	A	G	I	G	D	K
	gi 350398975	A	K	Q	I	N	D	S	R	-	-	V	R	N	N	L	S	R	R	F	G	V	V	V	D	S	A	G	V	G	D	K
	gi 242021897	V	N	N	I	D	D	I	-	-	-	R	G	K	L	K	N	S	L	G	V	E	V	I	R	A	G	I	G	D	K	
	gi 321477883	I	K	K	I	D	S	M	-	-	-	K	D	S	I	Q	R	A	T	G	I	E	I	V	S	T	K	V	G	N	Q	
	gi 260788939	A	S	K	A	D	E	M	-	-	-	R	L	K	I	K	E	V	T	G	F	E	L	D	T	T	G	S	G	E	Q	N
	gi 307175617	A	H	L	A	E	N	N	-	-	-	K	E	A	L	Q	K	A	V	G	M	E	L	V	G	T	G	I	G	S	N	
	gi 307209208	A	N	N	I	D	T	I	-	-	-	R	G	K	L	Q	D	A	L	G	I	R	V	I	R	A	G	I	G	N	K	
	gi 332019898	A	E	N	I	G	N	I	-	-	-	R	G	K	L	K	E	T	L	G	V	K	V	I	R	A	G	I	G	D	K	
	gi 195437839	V	N	N	I	D	N	I	-	-	-	R	G	K	L	Q	N	A	L	G	I	E	V	I	R	A	G	I	G	D	K	
	gi 357624369	A	R	I	I	N	E	N	R	G	V	K	N	N	I	Q	R	R	V	G	F	Y	V	L	H	A	G	V	G	D	K	
		A	K	Q	I	D	V	I	-	-	-	K	E	K	V	R	K	D	T	G	A	Q	I	Q	S	A	G	V	G	D	K	

SS MPE ICA512
SS M1SEA
SS M16SEA

B B B B B B
B B B B B B
B B B B B B

H H H H H H H
H H H H H H H
H H

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
>pdb 1IVZ A_12-132:	N	L	N	F	T	I	T	N	L	P	-	-	-	-	-	Y	S	Q	D	I	A	Q	P	S	T
>sp Q16849 PTPRN_HUMAN:	Y	G	Y	I	V	T	D	Q	K	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>gi 74005774.565-658:	Y	G	Y	I	V	T	D	Q	K	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>gi 126337782.514-607:	Y	G	Y	I	V	T	D	Q	K	P	-	-	-	-	-	L	G	-	-	-	-	-	-	-	-
>sp Q8WXI7 MUC16_HUMAN_21966-22:	Y	L	N	F	T	I	T	N	L	P	-	-	-	-	-	Y	S	Q	D	K	A	Q	P	G	T
>pdb 2ACN MUC1_HUMAN_1041-1148:	F	L	S	F	H	I	S	N	L	Q	-	-	-	-	-	F	N	S	S	L	E	D	P	S	T
>sp Q29435 MUC1_HYLLA_261-368:	F	L	S	F	H	I	S	N	L	Q	-	-	-	-	-	F	N	S	S	L	E	D	P	S	T
>gi 2289161.384-495:	F	L	S	F	Y	I	T	N	L	Q	-	-	-	-	-	F	N	S	S	L	E	D	P	S	T
>gi 41386778.363-473:	F	L	S	F	R	I	T	N	L	Q	-	-	-	-	-	F	N	S	S	L	E	D	P	S	T
>gi 125825580.335-428:	Y	G	Y	I	V	T	N	Q	S	H	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>gi 47226294.315-408:	Y	G	Y	I	F	T	N	K	S	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>sp Q02496 MUC1_MOUSE_415-523:	F	L	F	F	Y	I	Q	N	H	P	-	-	-	-	-	F	N	S	S	L	E	D	P	S	S
>sp Q60528 MUC1_MESAU_465-570:	L	L	S	F	H	I	W	N	H	Q	-	-	-	-	-	F	N	S	S	L	E	D	P	S	S
>gi 118085571.510-603:	Y	G	Y	I	V	T	V	K	D	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>sp Q02695 PTPR2_MACNE_506-599:	R	G	Y	I	V	T	D	R	D	P	-	-	-	-	-	L	R	-	-	-	-	-	-	-	-
>sp Q92932 PTPR2_HUMAN_508-601:	R	G	Y	I	V	T	D	R	D	P	-	-	-	-	-	L	R	-	-	-	-	-	-	-	-
>sp P80560 PTPR2_MOUSE_494-587:	Q	G	Y	I	L	T	G	N	N	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>sp Q63475 PTPR2_RAT_497-590:	Q	G	Y	I	L	T	G	N	N	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>gi 73979061.518-609:	L	G	Y	I	V	T	G	N	D	P	-	-	-	-	-	L	S	-	-	-	-	-	-	-	-
>sp Q8WXI7 MUC16_HUMAN_18941-19:	T	L	N	F	T	I	T	N	L	Q	-	-	-	-	-	Y	E	E	D	M	H	R	P	G	S
>sp Q8WXI7 MUC16_HUMAN_20649-20:	T	L	N	F	T	I	T	N	L	Q	-	-	-	-	-	Y	E	E	D	M	R	H	P	G	S
>sp Q8WXI7 MUC16_HUMAN_12076-12:	T	L	N	F	T	I	T	N	L	Q	-	-	-	-	-	Y	E	E	D	M	R	H	P	G	S
>sp Q8WXI7 MUC16_HUMAN_21714-21:	T	L	N	F	T	I	S	N	L	Q	-	-	-	-	-	Y	S	P	D	M	G	K	-	G	S
>sp Q8WXI7 MUC16_HUMAN_21569-21:	T	L	N	F	T	I	N	N	L	R	-	-	-	-	-	Y	M	A	D	M	G	Q	P	G	S
>gi 126339723.35-158:	I	L	N	F	T	I	T	N	M	F	-	-	-	-	-	Y	T	T	K	M	A	Q	R	D	S
>sp Q96MS5 IMP61_BOVIN_565-695:	F	F	S	L	R	V	A	N	V	P	-	-	-	-	-	F	S	T	D	L	F	N	K	S	S
>sp Q8RIW8 IMP61_MOUSE_577-688:	F	F	S	L	R	V	A	N	M	P	-	-	-	-	-	F	S	Y	D	L	F	N	K	S	S
>gi 109658622.573-684:	F	F	S	L	R	V	A	N	M	A	-	-	-	-	-	F	S	N	D	L	F	N	K	S	S
>sp Q8JIR8 IMP61_CHICK_737-848:	F	F	S	L	R	V	T	N	M	H	-	-	-	-	-	F	S	D	D	L	F	N	K	S	S
>gi 20806127.893-1009:	F	F	S	L	R	V	T	N	M	L	-	-	-	-	-	F	S	E	D	L	F	N	K	N	S
>sp Q5T601 GP110_HUMAN_145-257:	I	W	G	T	F	K	I	N	E	R	-	-	-	-	-	F	T	N	D	L	L	N	S	S	S
>gi 57157125.143-255:	V	W	G	T	F	E	I	D	E	N	-	-	-	-	-	F	P	E	D	L	W	N	S	S	S
>sp Q9WVT0 GP116_RAT_158-271:	T	L	K	I	K	V	R	L	N	I	G	-	-	-	-	Y	Q	E	K	D	L	E	N	T	S
>gi 2583092.125-241:	G	V	S	V	T	V	T	S	Q	E	-	-	-	-	-	Y	S	E	K	L	Q	D	R	K	S
>gi 91982772.4179-4294:	E	L	T	V	T	V	T	S	V	K	-	-	-	-	-	F	T	E	E	L	K	N	H	S	S
>gi 121934070.276-394:	G	M	T	V	K	V	T	Y	R	N	-	-	-	-	-	F	T	E	K	M	N	D	A	S	S
>sp Q02505 MUC3A_HUMAN_2233-235:	E	V	G	M	E	V	S	V	D	Q	Q	-	-	-	-	F	S	P	D	L	N	D	N	T	S
>sp Q50ID0 UROL1_HUMAN_391-513:	E	V	T	I	K	I	V	N	H	N	-	-	-	-	-	L	T	E	K	L	L	N	R	S	S
>sp Q50ID3 UROL1_MOUSE_384-506:	Q	V	T	I	R	I	M	D	R	N	-	-	-	-	-	L	T	E	Q	I	L	D	C	S	S
>sp Q50ID3 UROL1_MOUSE_789-901:	A	G	N	V	R	I	T	S	M	Q	-	-	-	-	-	Y	S	E	S	F	L	N	T	S	S
>sp Q50ID0 UROL1_HUMAN_783-897:	I	G	K	V	R	I	K	N	V	R	-	-	-	-	-	Y	S	E	S	F	R	N	A	S	S
>gi 17531707.1336-1449:	G	L	R	V	R	I	L	N	R	E	-	-	-	-	-	F	N	E	A	L	N	D	P	K	T
>gi 118777774.886-1006:	G	A	S	I	K	L	T	N	L	A	-	-	-	-	-	A	G	F	Q	V	P	D	S	G	T
>gi 24658804.928-1047:	A	V	A	V	K	L	M	N	L	G	-	-	-	-	-	E	S	Y	E	A	P	K	P	G	T
>sp Q8VHJ4 TM11D_RAT_48-160:	H	S	N	F	H	I	L	N	V	D	-	-	-	-	-	Y	T	E	A	L	N	S	P	A	T
>sp Q8VHK8 TM11D_MOUSE_48-160:	H	S	S	F	Q	I	L	N	V	E	-	-	-	-	-	Y	T	E	A	L	N	S	P	A	T
>sp Q060235 TM11D_HUMAN_42-164:	R	S	S	F	Q	L	L	N	V	E	-	-	-	-	-	Y	N	S	Q	L	N	S	P	A	T
>gi 51571899.55-175:	H	V	S	F	K	V	N	N	I	D	-	-	-	-	-	Y	D	S	K	F	A	K	P	Y	S
>sp Q5QSK2 DESC4_RAT_45-165:	Q	A	S	F	W	I	P	S	I	K	-	-	-	-	-	Y	S	S	D	L	S	E	Q	S	S
>sp Q8BZ10 DESC4_MOUSE_45-165:	Q	V	S	F	W	I	P	S	I	N	-	-	-	-	-	Y	S	S	D	L	S	K	E	Q	S
>gi 45861650.41-161:	H	G	S	F	K	I	L	D	P	Q	-	-	-	-	-	I	N	N	N	F	G	Q	S	N	T
>sp Q6ZWK6 Q6ZWK6_HUMAN_52-173:	L	A	S	F	K	V	T	N	I	K	-	-	-	-	-	Y	K	E	N	Y	G	I	R	S	S
>sp Q6IE14 Q6IE14_RAT_43-163:	Q	G	S	F	K	V	L	N	I	P	-	-	-	-	-	Y	D	R	N	Y	E	R	E	T	S
>sp Q5S248 TM11E_MOUSE_62-183:	Y	S	T	L	S	F	T	S	D	K	-	-	-	-	-	L	Y	S	E	F	G	R	E	A	S
>sp Q9UL52 TM11E_HUMAN_44-163:	Y	S	T	L	S	F	T	S	D	K	-	-	-	-	-	L	Y	A	E	F	G	R	E	A	S
>sp P98074 ENTK_PIG_56-171:	R	G	T	M	K	I	T	S	G	V	T	-	-	-	-	Y	N	P	N	L	Q	D	K	L	S
>sp P98072 ENTK_BOVIN_56-171:	R	G	T	L	K	I	I	S	G	A	T	-	-	-	-	Y	N	P	H	L	Q	D	K	L	S
>sp P98073 ENTK_HUMAN_56-171:	R	A	T	F	K	I	T	S	G	V	T	-	-	-	-	Y	N	P	N	L	Q	D	K	L	S
>sp P97435 ENTK_MOUSE_56-171:	R	G	T	F	K	I	T	S	G	V	T	-	-	-	-	Y	N	P	N	L	Q	D	K	L	S
>gi 118780624.566-687:	A	G	T	I	R	I	V	D	G	V	K	-	-	-	-	W	V	P	E	L	L	D	H	N	T
>gi 24654487.1166-1287:	A	A	T	I	K	I	I	D	G	V	A	-	-	-	-	W	K	P	E	L	L	D	H	N	T
>sp Q5BIK4 Q5BIK4_DROME_224-347:	R	G	R	F	M	V	L	N	D	K	-	-	-	-	-	W	S	M	G	L	A	N	Q	N	S
>sp Q7QF48 Q7QF48_ANOGA_52-177:	A	R	L	R	V	I	E	G	D	R	-	-	-	-	-	W	T	P	E	L	A	D	Q	N	T
>sp Q07929 SP63_STRPU_73-205:	A	G	S	E	S	V	T	Q	V	G	S	N	V	L	-	Y	S	A	D	L	A	D	T	D	S
>sp Q17R60 IMP61_HUMAN_227-349:	E	L	S	V	S	L	V	N	Q	K	-	-	-	-	-	F	K	A	E	L	A	D	S	Q	S
>sp Q96MS5 IMP61_BOVIN_225-365:	E	L	S	I	S	L	A	N	Q	K	-	-	-	-	-	F	K	S	E	L	D	N	S	Q	S
>sp Q8RIW8 IMP61_MOUSE_230-352:	E	F	S	I	S	L	P	N	H	R	-	-	-	-	-	F	K	A	E	L	T	N	S	G	S
>sp Q8JIR8 IMP61_CHICK_226-348:	E	F	S	V	T	L	T	D	Q	E	-	-	-	-	-	Y	T	A	E	L	S	D	P	N	S
>gi 20806127.228-347:	E	F	S	I	Q	L	L	G	K	Q	-	-	-	-	-	Y	S	E	E	L	R	D	P	S	S
>sp Q00468 AGRIN_HUMAN_1132-125:	Q	G	V	L	E	L	E	G	V	E	G	Q	E	L	F	Y	T	P	E	M	A	D	P	K	S
>sp P25304 AGRIN_RAT_1025-1145:	Q	G	V	L	E	L	E	G	V	E	G	Q	E	L	F	Y	T	P	E	M	A	D	P	K	S
>sp P31696 AGRIN_CHICK_1146-126:	Q	G	V	L	I	L	E	E	V	E	G	Q	E	L	F	Y	T	P	E	M	A	D	P	K	S
>sp Q90404 AGRIN_DISOM_413-533:	Q	G	V	L	I	V	E	E	V	E	G	Q	E	L	F	Y	T	P	E	M	D	D	P	K	S
>sp P98160 PGBM_HUMAN_75-188:	R	A	L	V	N	F	T	R	S	I	E	-	-	-	-	Y	S	P	Q	L	E	D	A	G	S
>gi 94536813.76-189:	R	A	L	V	N	F	T	R	S	I	D	-	-	-	-	F	S	P	R	L	E	D	P	N	S
>gi 52346064.69-191:	T	G	Y	L	R	I	A	N	T	Q	-	-	-	-	-	F	V	E	A	Y	E	N	S	T	T
>gi 975																									

[illegible]

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes

SS MPE ICA512
SS M1SEA
SS M16SEA

B B B B B B B B B B B B
B B B B B B B B B B B B
B B B B B B B B B B B B

	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77
>pdb 1TVZ A_12-132:	K	S	-	-	Y	F	S	D	C	Q	V	L	A	F	R	S	V	S	N	N	N	N	H	T	G	V
>sp Q16849 PTPRN_HUMAN:	-	-	-	-	-	-	-	-	F	I	N	I	S	V	V	G	-	-	-	-	-	-	-	P	A	L
>gi 74005774.565-658:	-	-	-	-	-	-	-	-	F	I	N	I	S	V	V	G	-	-	-	-	-	-	-	P	A	L
>gi 126337782.514-607:	-	-	-	-	-	-	-	-	F	I	N	I	S	V	V	G	-	-	-	-	-	-	-	S	A	L
>sp Q8WXI7 MUC16_HUMAN_21966-22:	K	S	-	-	Y	F	S	D	C	Q	V	S	T	F	R	S	V	P	-	N	R	H	H	T	G	V
>pdb 2ACM MUC1_HUMAN_1041-1148:	-	-	-	-	F	L	G	L	S	N	I	K	F	R	P	G	S	-	-	-	-	-	-	V	V	V
>sp Q29435 MUC1_HYLLA_261-368:	-	-	-	-	F	L	G	V	S	N	I	K	F	R	P	G	S	-	-	-	-	-	-	V	V	V
>gi 2289161.384-495:	-	-	-	-	F	L	G	L	S	E	I	K	F	R	P	G	S	-	-	-	-	-	-	V	V	V
>gi 41386778.363-473:	-	-	-	-	F	L	G	L	S	E	I	K	F	R	P	G	S	-	-	-	-	-	-	V	V	V
>gi 125825580.335-428:	-	-	-	-	-	-	-	-	F	I	N	I	S	V	V	G	-	-	-	-	-	-	-	S	A	L
>gi 47226294.315-408:	-	-	-	-	-	-	-	-	F	I	N	I	S	V	V	G	-	-	-	-	-	-	-	P	A	L
>sp Q02496 MUC1_MOUSE_415-523:	G	-	-	-	-	-	-	-	I	S	S	I	K	F	R	S	G	S	-	-	-	-	-	V	V	V
>sp Q60528 MUC1_MESAU_465-570:	G	-	-	-	-	-	-	-	I	S	T	I	E	F	R	S	G	S	-	-	-	-	-	V	V	V
>gi 118085571.510-603:	-	-	-	-	-	-	-	-	F	D	D	V	N	M	L	G	-	-	-	-	-	-	-	P	A	V
>sp Q02695 PTPR2_MACNE_506-599:	-	-	-	-	-	-	-	-	F	A	D	V	E	V	L	G	-	-	-	-	-	-	-	P	A	V
>sp Q92932 PTPR2_HUMAN_508-601:	-	-	-	-	-	-	-	-	F	A	D	V	E	V	L	G	-	-	-	-	-	-	-	P	A	V
>sp P80560 PTPR2_MOUSE_494-587:	-	-	-	-	-	-	-	-	F	A	D	V	E	V	L	G	-	-	-	-	-	-	-	P	A	V
>sp Q63475 PTPR2_RAT_497-590:	-	-	-	-	-	-	-	-	F	A	D	V	E	V	L	G	-	-	-	-	-	-	-	P	A	V
>gi 73979061.518-609:	-	-	-	-	-	-	-	-	F	A	D	V	E	V	L	G	-	-	-	-	-	-	-	P	A	V
>sp Q8WXI7 MUC16_HUMAN_18941-19:	G	P	L	-	Y	S	G	C	R	L	T	S	L	R	P	E	K	-	D	G	A	A	T	G	M	-
>sp Q8WXI7 MUC16_HUMAN_20649-20:	G	P	L	-	Y	S	G	C	R	L	T	S	L	R	P	E	K	-	D	K	A	A	T	R	V	-
>sp Q8WXI7 MUC16_HUMAN_12076-12:	E	Y	L	-	Y	S	G	C	R	L	A	S	L	R	P	E	K	-	D	S	S	A	M	A	V	-
>sp Q8WXI7 MUC16_HUMAN_21714-21:	G	P	F	-	Y	L	G	C	Q	L	I	S	L	R	P	E	K	-	D	G	A	A	T	G	V	-
>sp Q8WXI7 MUC16_HUMAN_21569-21:	G	A	R	-	Y	T	G	C	R	V	I	A	L	R	S	V	K	-	N	G	A	E	T	R	V	-
>gi 126339723.35-158:	G	S	Q	-	Y	C	G	C	N	V	T	L	L	R	S	M	K	-	N	G	A	A	T	G	V	-
>sp Q9GMS5 IMP61_BOVIN_565-695:	-	-	-	-	F	K	Q	L	E	I	L	N	F	R	N	G	S	-	-	-	-	-	-	V	I	V
>sp Q8R1W8 IMP61_MOUSE_577-688:	-	-	-	-	F	K	Q	L	E	I	L	N	F	R	N	G	S	-	-	-	-	-	-	V	I	V
>gi 109658622.573-684:	-	-	-	-	F	K	Q	L	E	I	L	N	F	R	N	G	S	-	-	-	-	-	-	V	I	V
>sp Q8JIR8 IMP61_CHICK_737-848:	-	-	-	-	F	K	Q	L	E	I	L	N	F	R	N	G	S	-	-	-	-	-	-	V	I	V
>gi 20806127.893-1009:	-	-	-	-	F	Q	N	L	E	I	L	N	F	R	N	G	S	-	-	-	-	-	-	I	V	A
>sp Q5T601 GP110_HUMAN_145-257:	-	-	-	-	F	E	S	V	Q	V	T	Q	F	R	N	G	S	-	-	-	-	-	-	I	V	A
>gi 57157125.143-255:	-	-	-	-	F	E	S	V	R	V	T	Q	F	R	K	G	S	-	-	-	-	-	-	I	V	V
>sp Q9WV70 GP116_RAT_158-271:	-	-	-	-	F	R	S	V	T	V	T	Q	F	T	K	G	S	-	-	-	-	-	-	V	V	V
>gi 2583092.125-241:	-	-	-	-	Y	E	G	V	I	K	N	L	S	K	G	S	-	-	-	-	-	-	-	I	V	V
>gi 91982772.4179-4294:	-	-	-	-	Y	V	G	V	N	I	T	K	L	R	L	G	S	-	-	-	-	-	-	V	V	V
>gi 121934070.276-394:	P	Q	-	-	Y	R	G	V	N	I	R	R	L	L	N	G	S	-	-	-	-	-	-	I	V	V
>sp Q02505 MUC3A_HUMAN_2233-235:	-	F	T	-	F	K	G	V	E	I	L	S	L	R	N	G	S	-	-	-	-	-	-	I	V	V
>sp Q5DID0 UROL1_HUMAN_391-513:	L	Y	R	S	G	K	L	R	M	Q	I	V	S	L	Q	A	G	S	-	-	-	-	-	V	V	V
>sp Q5DID3 UROL1_MOUSE_384-506:	L	Y	R	Q	G	R	L	R	M	Q	I	V	S	L	Q	A	G	S	-	-	-	-	-	L	V	V
>sp Q5DID3 UROL1_MOUSE_789-901:	H	M	D	A	G	R	I	R	V	D	I	N	I	T	N	G	S	-	-	-	-	-	-	I	V	V
>sp Q5DID0 UROL1_HUMAN_783-897:	H	M	D	A	G	G	V	R	M	E	V	V	S	V	T	N	G	S	-	-	-	-	-	I	V	V
>gi 17531707.1336-1449:	-	-	-	-	M	R	V	S	K	I	V	G	Y	E	K	G	S	-	-	-	-	-	-	V	I	A
>gi 118777774.886-1006:	-	-	-	-	F	Y	R	T	D	V	Q	E	F	E	K	Y	V	-	D	A	D	G	S	V	V	V
>gi 24658804.928-1047:	-	-	-	-	Y	Y	K	T	E	I	L	G	F	E	K	E	G	-	-	-	-	-	-	S	T	M
>sp Q8VHJ4 TM11D_RAT_48-160:	S	E	-	-	F	I	R	T	H	V	V	K	L	R	K	E	G	-	-	-	-	-	-	S	G	V
>sp Q8VHK8 TM11D_MOUSE_48-160:	S	E	-	-	F	I	R	T	H	V	V	K	L	R	K	E	G	-	-	-	-	-	-	S	G	V
>sp Q60235 TM11D_HUMAN_42-164:	N	Q	-	-	F	I	R	A	H	V	A	K	L	R	Q	D	G	-	-	-	-	-	-	S	G	V
>gi 51571899.55-175:	K	Q	-	-	Y	V	K	A	H	T	V	Q	V	S	K	A	K	-	-	-	-	-	-	G	K	V
>sp Q5QK2 DESC4_RAT_45-165:	H	H	-	-	Y	V	K	S	Q	V	V	N	F	R	P	S	N	-	-	-	-	-	-	D	G	V
>sp Q8BZ10 DESC4_MOUSE_45-165:	H	H	-	-	Y	V	K	S	Q	V	V	N	F	R	P	S	N	-	-	-	-	-	-	D	G	V
>gi 45861650.41-161:	K	N	-	-	Y	I	K	N	Q	V	V	R	L	T	P	E	E	-	-	-	-	-	-	D	G	V
>sp Q6ZWK6 Q6ZWK6_HUMAN_52-173:	G	R	-	-	Y	I	K	S	H	V	R	L	S	P	D	E	-	-	-	-	-	-	-	Q	G	V
>sp Q6IE14 Q6IE14_RAT_43-163:	K	Q	-	-	Y	I	N	S	Q	I	I	T	L	V	P	E	N	-	-	-	-	-	-	N	S	V
>sp Q5S248 TM11E_MOUSE_62-183:	G	Q	-	-	L	V	K	A	H	I	I	K	F	S	K	E	D	-	-	-	-	-	-	D	G	V
>sp Q9UL52 TM11E_HUMAN_44-163:	E	E	-	-	F	V	K	S	Q	V	I	K	F	S	Q	Q	K	-	-	-	-	-	-	H	G	V
>sp P98074 ENTK_PIG_56-171:	N	E	-	-	Y	K	N	S	R	V	L	Q	F	E	N	G	S	-	-	-	-	-	-	-	V	I
>sp P98072 ENTK_BOVIN_56-171:	N	E	-	-	Y	K	N	S	R	V	L	Q	F	E	N	G	S	-	-	-	-	-	-	-	I	I
>sp P98073 ENTK_HUMAN_56-171:	N	E	-	-	Y	K	N	S	R	V	L	Q	F	E	N	G	S	-	-	-	-	-	-	-	I	I
>sp P97435 ENTK_MOUSE_56-171:	N	E	-	-	Y	E	K	S	K	V	F	F	E	N	G	S	-	-	-	-	-	-	-	-	V	I
>gi 118780624.566-687:	S	K	W	-	Y	K	K	V	R	I	D	S	F	N	K	G	S	-	-	-	-	-	-	-	V	L
>gi 24654487.1166-1287:	N	K	W	-	Y	K	K	V	R	I	D	S	F	N	K	G	S	-	-	-	-	-	-	-	V	L
>sp Q5BIK4 Q5BIK4_DROME_224-347:	E	A	-	-	Y	E	G	S	E	I	L	A	L	D	G	S	E	-	-	-	-	-	-	D	N	N
>sp Q7QF48 Q7QF48_ANOGA_52-177:	E	P	-	-	F	E	G	S	E	I	L	A	L	D	G	N	E	-	-	-	-	-	-	E	P	G
>sp Q07929 SP63_STRPU_73-205:	D	I	-	-	Y	L	G	S	E	V	W	G	V	P	E	W	L	-	-	-	-	-	-	-	Y	R
>sp Q17R60 IMP61_HUMAN_227-349:	-	-	-	-	F	K	K	I	H	V	L	G	F	R	P	K	K	-	E	K	D	G	S	S	S	T
>sp Q9GMS5 IMP61_BOVIN_225-365:	-	-	-	-	F	K	E	I	H	V	S	G	F	R	P	K	K	-	E	R	D	G	T	S	S	T
>sp Q8R1W8 IMP61_MOUSE_230-352:	-	-	-	-	F	G	E	I	R	V	L	G	F	R	P	K	K	-	E	E	D	G	T	S	S	T
>sp Q8JIR8 IMP61_CHICK_226-348:	-	-	-	-	F	K	E	I	H	V	L	G	F	R	P	K	K	-	E	K	D	G	T	S	S	T
>gi 20806127.228-347:	-	-	-	-	Y	K	G	I	H	V	L	D	F	R	S	P	K	-	E	N	G	S	G	I	D	V
>sp Q080468 AGRIN_HUMAN_1132-125:	K	D	-	-	F	R	S	V	R	L	R	D	L	G	P	G	K	-	-	-	-	-	-	S	V	R
>sp P25304 AGRIN_RAT_1025-1145:	K	D	-	-	F	W	S	V	R	L	R	E	L	G	P	G	K	-	-	-	-	-	-	L	V	R
>sp P31696 AGRIN_CHICK_1146-126:	N	D	-	-	F	K	S	I	R	V	R	D	L	G	Q	S	S	-	-	-	-	-	-	A	V	R
>sp Q90404 AGRIN_DISOM_413-533:	K	D	-	-	F	K	S	V	R	V	H	G	L	G	P	S	D	-	-	-	-	-	-	P	V	R
>sp P98160 PGBM_HUMAN_75-188:																										

SS MPE ICA512
SS M1SEA
SS M16SEA

B B B B B
B B B B B B
B B B B B B

H H H H H H H H H
H H H H H H H H H
H H H H H H H H H

	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102		
>pdb 1IVZ A_12-132:	D	S	L	C	N	F	S	P	L	A	R	R	-	-	V	D	R	V	A	I	Y	E	E	F	L		
>sp Q16849 PTPRN_HUMAN:	T	F	R	I	R	H	N	E	Q	N	-	-	-	-	L	S	L	A	D	V	T	Q	Q	A	G		
>gi 74005774.565-658:	T	F	R	I	R	H	N	E	Q	N	-	-	-	-	L	S	L	A	D	V	T	Q	Q	A	G		
>gi 126337782.514-607:	T	F	R	I	R	N	N	E	Q	N	-	-	-	-	M	S	L	A	D	V	T	Q	Q	A	G		
>sp Q8WXI7 MUC16_HUMAN_21966-22:	D	S	L	C	N	F	S	P	L	A	R	R	-	-	V	D	R	V	A	I	Y	E	E	F	L		
>pdb 2ACN MUC1_HUMAN_1041-1148:	Q	L	T	L	A	F	R	E	G	T	-	-	-	-	I	N	V	H	D	V	E	T	Q	F	N		
>sp Q29435 MUC1_HYLLA_261-368:	Q	S	T	L	A	F	R	E	G	T	-	-	-	-	T	N	V	H	D	V	E	T	Q	F	N		
>gi 2289161.384-495:	D	L	I	L	A	F	Q	E	G	S	-	-	-	-	T	N	E	V	T	V	Q	S	Q	F	I		
>gi 41386778.363-473:	E	L	T	L	A	F	R	E	G	T	-	-	-	-	-	T	A	E	W	V	K	A	Q	F	S		
>gi 125825580.335-428:	T	F	R	I	R	Q	N	E	H	N	-	-	-	-	L	S	A	E	D	V	A	N	K	A	V		
>gi 47226294.315-408:	T	F	R	I	R	Q	N	E	H	N	-	-	-	-	M	T	A	A	E	V	A	A	K	A	V		
>sp Q02496 MUC1_MOUSE_415-523:	E	S	T	V	V	F	R	E	G	T	-	-	-	-	F	S	A	S	D	V	K	S	Q	L	L		
>sp Q60528 MUC1_MESAU_465-570:	D	S	T	V	I	F	R	E	G	A	-	-	-	-	V	N	A	S	E	V	K	S	Q	L	L		
>gi 118085571.510-603:	T	F	R	V	H	S	N	L	Q	N	-	-	-	-	I	S	T	A	D	V	A	K	A	A	V		
>sp 002695 PTPR2_MACNE_506-599:	T	F	K	V	G	A	N	V	Q	N	-	-	-	-	V	T	T	A	D	V	E	K	A	T	V		
>sp Q92932 PTPR2_HUMAN_508-601:	T	F	K	V	S	A	N	V	Q	N	-	-	-	-	V	T	T	E	D	V	E	K	A	T	V		
>sp P80560 PTPR2_MOUSE_494-587:	T	F	K	V	S	A	N	I	Q	N	-	-	-	-	M	T	T	A	D	V	I	K	A	A	V		
>sp Q63475 PTPR2_RAT_497-590:	I	F	K	V	S	A	N	I	Q	N	-	-	-	-	M	T	T	A	D	V	T	K	A	A	V		
>gi 73979061.518-609:	T	F	R	V	S	A	S	I	W	N	-	-	-	-	L	T	R	A	D	V	V	Q	A	T	V		
>sp Q8WXI7 MUC16_HUMAN_18941-19:	D	A	V	C	L	Y	H	S	P	N	P	K	R	P	G	L	D	R	E	Q	L	Y	C	E	L	S	
>sp Q8WXI7 MUC16_HUMAN_20649-20:	D	A	I	C	T	H	H	P	D	P	Q	S	P	G	L	L	N	R	E	Q	L	Y	W	E	L	S	
>sp Q8WXI7 MUC16_HUMAN_12076-12:	D	A	I	C	T	H	R	P	D	P	E	D	L	G	L	L	D	R	E	R	L	Y	W	E	L	S	
>sp Q8WXI7 MUC16_HUMAN_21714-21:	D	T	T	C	T	Y	H	P	D	P	V	G	P	G	L	D	I	Q	Q	L	Y	W	E	L	S		
>sp Q8WXI7 MUC16_HUMAN_21569-21:	D	L	L	C	T	Y	L	Q	P	L	S	G	P	G	L	P	I	K	Q	V	F	H	E	L	S		
>gi 126339723.35-158:	E	V	I	C	F	Y	K	E	V	F	S	V	P	V	L	D	K	K	K	I	Y	Q	E	L	S		
>sp Q96MS5 IMP61_BOVIN_565-695:	N	S	K	V	R	F	A	K	S	V	-	-	-	-	P	Y	N	L	T	K	A	V	R	G	V	L	E
>sp Q8R1W8 IMP61_MOUSE_577-688:	N	S	K	V	R	F	A	K	A	V	-	-	-	-	P	Y	N	L	T	Q	A	V	R	G	V	L	E
>gi 109658622.573-684:	N	S	K	M	K	F	A	K	S	V	-	-	-	-	P	Y	N	L	T	K	A	V	H	G	V	L	E
>sp Q8JIR8 IMP61_CHICK_737-848:	N	S	K	M	K	F	A	R	T	V	-	-	-	-	P	Y	N	I	T	E	A	V	H	C	V	L	E
>gi 20806127.893-1009:	N	S	R	V	K	F	A	E	S	V	-	-	-	-	P	P	N	V	N	N	A	I	Y	M	I	L	E
>sp Q5T601 GP110_HUMAN_145-257:	G	Y	E	V	V	G	S	S	S	A	S	-	-	-	E	L	L	S	A	I	E	H	V	A	E	E	
>gi 57157125.143-255:	G	Y	E	V	T	G	S	T	S	P	-	-	-	-	P	E	L	L	F	A	I	E	Q	E	A	E	
>sp Q9WVT0 GP116_RAT_158-271:	D	Y	I	V	E	V	A	S	A	P	-	-	-	-	-	L	P	G	S	I	H	K	A	N	E		
>gi 2583092.125-241:	D	Y	D	V	I	L	K	A	K	Y	T	-	-	-	P	G	-	F	E	N	T	L	D	T	V	V	K
>gi 91982772.4179-4294:	E	H	D	V	I	L	R	T	K	Y	T	-	-	-	P	-	E	Y	K	T	V	L	D	N	A	T	E
>gi 121934070.276-394:	K	N	D	V	I	L	E	A	D	Y	T	-	-	-	L	E	Y	E	E	L	F	E	N	L	A	E	E
>sp Q02505 MUC3A_HUMAN_2233-235:	D	Y	L	V	L	E	M	P	F	S	-	-	-	-	-	-	P	Q	L	E	S	E	Y	E	E	E	
>sp Q50ID0 UROL1_HUMAN_391-513:	R	L	K	L	T	V	Q	D	P	G	-	-	-	-	-	-	F	P	M	G	I	S	T	L	A	E	
>sp Q50ID3 UROL1_MOUSE_384-506:	T	L	R	L	T	L	Q	D	P	D	-	-	-	-	-	-	F	S	V	G	V	H	T	L	T	E	
>sp Q50ID3 UROL1_MOUSE_789-901:	E	F	N	L	L	M	T	A	D	L	-	-	-	-	-	-	-	D	V	R	E	V	S	A	A	E	
>sp Q50ID0 UROL1_HUMAN_783-897:	E	F	H	L	I	I	A	D	V	-	-	-	-	-	-	-	-	D	V	Q	E	V	S	A	A	E	
>gi 17531707.1336-1449:	E	F	E	V	V	A	M	E	D	T	-	-	-	-	-	-	A	K	P	M	E	L	K	S	M	V	
>gi 118777774.886-1006:	K	M	N	I	L	V	D	K	E	M	K	N	G	-	-	-	T	K	N	L	L	A	D	D	P	A	
>gi 24658804.928-1047:	K	V	Q	A	M	F	D	K	A	D	-	-	-	-	-	V	E	K	K	H	A	L	E	T	N	E	V
>sp Q8VHJ4 TM11D_RAT_48-160:	D	V	V	M	K	F	R	S	S	K	R	-	-	-	-	-	N	N	K	K	A	I	K	T	R	I	Q
>sp Q8VHK8 TM11D_MOUSE_48-160:	D	V	V	M	K	F	R	S	S	K	R	-	-	-	-	-	N	N	R	K	V	M	K	T	R	I	Q
>sp Q60235 TM11D_HUMAN_42-164:	D	V	V	M	K	F	Q	F	T	R	N	-	-	-	-	-	N	N	G	A	S	M	K	S	R	I	E
>gi 51571899.55-175:	H	G	I	V	L	K	F	K	S	C	Y	K	-	-	-	-	N	N	V	E	K	F	W	D	S	V	E
>sp Q5QSK2 DESC4_RAT_45-165:	D	G	L	I	K	F	Q	I	P	R	K	-	-	-	-	-	N	A	D	T	L	R	S	E	A	D	D
>sp Q8BZ10 DESC4_MOUSE_45-165:	D	V	L	I	K	F	Q	I	P	R	K	-	-	-	-	-	N	A	G	T	L	K	R	Q	A	D	D
>gi 45861650.41-161:	D	V	I	M	V	F	Q	F	P	S	T	-	-	-	-	-	E	Q	R	A	V	R	E	K	K	I	Q
>sp Q6ZWK6 Q6ZWK6_HUMAN_52-173:	L	I	V	L	I	F	R	Y	P	S	T	-	-	-	-	-	D	S	A	E	Q	I	K	K	K	I	E
>sp Q6IE14 Q6IE14_RAT_43-163:	H	I	W	L	V	F	K	D	P	W	S	-	-	-	-	-	N	K	E	N	L	R	R	R	I	E	E
>sp Q5S248 TM11E_MOUSE_62-183:	H	M	L	L	I	F	R	I	R	S	T	-	-	-	-	-	E	D	P	E	T	V	H	K	I	I	E
>sp Q9UL52 TM11E_HUMAN_44-163:	H	M	L	L	I	C	R	F	H	S	T	-	-	-	-	-	E	D	P	E	T	V	D	K	I	V	Q
>sp P98074 ENTK_PIG_56-171:	I	F	D	L	L	F	A	Q	W	V	S	-	-	-	-	-	D	E	N	I	K	E	E	L	I	Q	Q
>sp P98072 ENTK_BOVIN_56-171:	I	F	D	L	L	F	A	Q	W	V	S	-	-	-	-	-	D	K	N	V	K	E	E	L	I	Q	Q
>sp P98073 ENTK_HUMAN_56-171:	V	F	D	L	F	A	Q	W	V	S	-	-	-	-	-	-	Q	N	V	K	E	E	L	I	Q	Q	
>sp P97435 ENTK_MOUSE_56-171:	L	F	D	L	F	A	Q	W	V	S	-	-	-	-	-	-	D	K	N	V	K	E	E	L	I	Q	Q
>gi 118780624.566-687:	D	Y	F	V	E	L	T	D	L	T	R	-	-	-	-	-	-	D	V	N	T	L	E	I	R	R	
>gi 24654487.1166-1287:	D	Y	F	V	E	L	A	N	I	T	-	-	-	-	-	-	-	E	D	V	D	T	L	E	I	R	
>sp Q5BIK4 Q5BIK4_DROME_224-347:	H	F	N	M	I	F	D	P	Y	A	G	-	-	-	-	-	L	V	S	S	G	D	L	L	A	A	
>sp Q7QF48 Q7QF48_ANOGA_52-177:	H	F	A	L	Y	F	D	P	Y	A	G	-	-	-	-	-	-	E	L	V	S	A	A	D	L	A	
>sp Q07929 SP63_STRPU_73-205:	R	L	H	V	L	F	A	T	E	D	A	G	-	-	-	Q	P	V	L	V	N	S	T	D	A	T	E
>sp Q17R60 IMP61_HUMAN_227-349:	E	M	Q	L	T	A	I	F	K	R	-	-	-	-	-	-	H	S	A	E	A	K	S	P	A	S	
>sp Q96MS5 IMP61_BOVIN_225-365:	E	M	Q	L	T	A	I	F	K	K	G	-	-	-	-	-	-	K	A	E	A	K	S	P	A	S	
>sp Q8R1W8 IMP61_MOUSE_230-352:	E	I	Q	L	M	A	I	F	K	R	-	-	-	-	-	-	D	H	A	E	A	K	S	P	D	S	
>sp Q8JIR8 IMP61_CHICK_226-348:	I	A	R	Y	M	V	N	F	E	R	-	-	-	-	-	-	D	G	S	E	I	K	S	T	A	D	
>gi 20806127.228-347:	H	Y	A	V	T	F	N	G	E	A	-	-	-	-	-	-	-	I	S	N	T	T	W	D	L	I	
>sp Q00468 AGRIN_HUMAN_1132-125:	I	V	D	V	H	F	D	P	T	T	-	-	-	-	-	-	A	F	R	A	P	D	V	A	R	A	
>sp P25304 AGRIN_RAT_1025-1145:	I	V	D	V	H	F	D	P	T	T	-	-	-	-	-	-	A	F	R	A	S	D	V	G	Q	A	
>sp P31696 AGRIN_HUMAN_1146-126:	I	V	E	S	H	F	D	P	A	T	S	-	-	-	-	-	-	Y	T	A	A	D	V	Q	R	A	
>sp Q90404 AGRIN_DISOM_413-533:	I	V	E	V	H	F	D	P	R	T	S	-	-	-	-	-	-	Y	N	S	H	D	V	Q	R	A	
>sp P98160 PGBM_HUMAN_75-188:	E	L	D	V	G	S	E	G	N	A	D	G	-	-													

Repositorio Institucional Digital de Acceso Abierto, Universidad Nacional de Quilmes