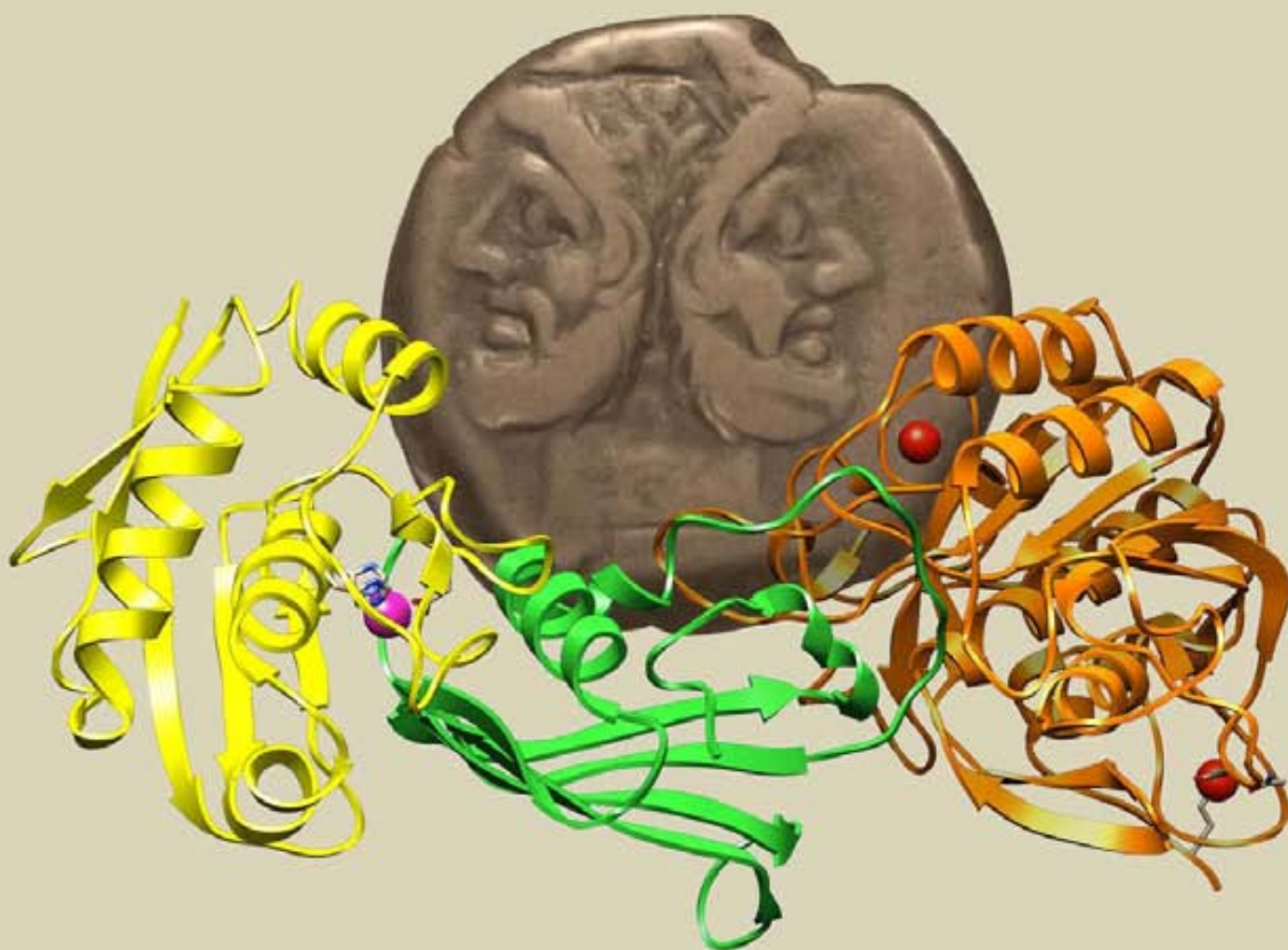


**E**STUDIO ESTRUCTURAL Y FUNCIONAL DE UN INHIBIDOR  
PROTEÍICO MONODOMINIO DE DOBLE FAZ,  
SERMETSTATINA, EN COMPLEJO CON DOS PEPTIDASAS  
DE DIFERENTE CLASE, SUBTILISINA Y ESNAPALISINA.



SERGIO TRILLO MUYO

2013



Portada: Fotografía de un as, una moneda de bronce utilizada durante la Segunda República Romana (siglo segundo antes de Cristo), la cual muestra la imagen del dios Jano (fotografía tomada en el Museu Nacional d'Art de Catalunya, Barcelona; [www.mnac.cat](http://www.mnac.cat)). Sobre la moneda se muestra el modelo de la estructura del complejo ternario subtilisina-sermetstatina-esnapalisina (en naranja, verde y amarillo, respectivamente) obtenido en la presente tesis doctoral.



UNIVERSITAT AUTÒNOMA DE BARCELONA

Institut de Biotecnologia i Biomedicina



CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Institut de Biologia Molecular de Barcelona

Departament de Biologia Estructural

Proteolysis Laboratory



**Estudio estructural y funcional de un inhibidor  
proteínico monodominio de doble faz,  
sermetstatina, en complejo con dos peptidasas  
de diferente clase, subtilisina y esnapalisina**

Sergio Trillo Muyo

Barcelona, Mayo 2013



**UNIVERSITAT AUTÒNOMA DE BARCELONA**

Institut de Biotecnologia i Biomedicina

**Programa de Doctorado en Estructura y Función de Proteínas 2008/2009**

Programa con mención de calidad otorgado por el Ministerio de Educación y Ciencia,  
enmarcado en las directrices establecidas por el Real Decreto 778/98

**Estudio estructural y funcional de un inhibidor proteínico  
monodominio de doble faz, sermetstatina, en complejo con dos  
peptidasas de diferente clase, subtilisina y esnapalisina**

Memoria presentada por Sergio Trillo Muyo  
para optar al título de doctor por la Universitat Autònoma de Barcelona

Sergio Trillo Muyo

Directores de Tesis:

Tutor:

Prof. F.Xavier Gomis Rùth

Dr. Joan López Arolas

Prof. Xavier Daura Ribera





**A mis padres**

**A mi tía**



## AGRADECIMIENTOS

La portada de esta memoria de tesis se ilustra con una imagen del dios romano Jano, el dios de las puertas, de las transiciones, de los comienzos y los finales. Precisamente, después de todos estos años, de tanto trabajo y de tantas horas, me encuentro aquí, en el final de una etapa y el inicio de otra. La representación de este dios, con dos caras mirando hacia ambos lados de su perfil, me recuerda que en los momentos de transición hay que saber mirar atrás y adelante, al pasado y al futuro. Así que sin apartar la vista del futuro, de los nuevos proyectos y oportunidades, es también hora de mirar atrás, de no dejar olvidadas las cosas que me han llevado hasta aquí. Mirando al pasado veo momentos, algunos buenos y otros no tanto, que de un modo u otro me han marcado en este periodo de crecimiento científico y personal que es la realización de una tesis. Pero, más importantes que esos momentos, son las personas con quienes los compartí, y este es el momento agradecerles el haber estado ahí.

Quiero empezar agradeciendo a Xavier Gomis la oportunidad que me dio de realizar esta tesis. Sin él, sin su tiempo, su apoyo, su guía, su motivación y sus ideas, este trabajo no hubiese sido posible.

También quiero agradecer especialmente a Joan L. Arolas haber codirigido esta tesis, por todo su trabajo, sus ideas y por todos los conocimientos que me ha transmitido.

Tengo que dar las gracias a todos los miembros del laboratorio Cri3 con los que he tenido la suerte de trabajar durante estos años y compartir el día a día. De todos ellos debo mencionar a Goretta Mallorquí, por guiar mis primeros pasos, por motivarme como lo hizo, porque sin ella esta etapa que ahora acaba no habría empezado. Gracias también a Tibisay Guevara, por haberme acompañado durante toda mi estancia en el laboratorio, por todo su trabajo y todo su apoyo. Además, debo dar las gracias a los miembros de los otros Cri con los que he coincidido, especialmente a Isabel Usón por haberme acogido temporalmente en su laboratorio.

Gracias también a Anna Mañosas, por su apoyo incondicional, por todas las transiciones que hemos vivido y viviremos juntos.

Finalmente, quiero dar las gracias a mi familia. A mis padres, por haber educado a dos hijos como lo han hecho, por todo su sacrificio. A mi hermano, por estar siempre ahí. A mi tía, que siempre tendré presente.



# ÍNDICE



---

<b>Índice</b> .....	<b>I</b>
Figuras .....	VI
Tablas .....	VIII
<b>Abreviaturas</b> .....	<b>IX</b>
<b>Nomenclatura</b> .....	<b>XV</b>
<b>Prefacio</b> .....	<b>XIX</b>
<b>1. Introducción</b> .....	<b>1</b>
1.1. Bacterias del suelo .....	3
1.1.1. Género <i>Streptomyces</i> .....	5
1.2. Peptidasas .....	8
1.2.1. Clasificación.....	9
1.2.1.1. Base de datos EC .....	10
1.2.1.2. Base de datos MEROPS .....	11
1.2.2. Metalopeptidasas .....	11
1.2.2.1. Metzinquinas.....	15
1.2.2.1.1. Esnapalisina.....	18
1.2.2.1.2. Regulación por zimogenicidad en metzinquinas.....	20
1.2.3. Serín peptidasas.....	26
1.2.3.1. Clan PA .....	29
1.2.3.2. Clan SB.....	31
1.3. Inhibidores proteínicos de peptidasas .....	33
1.3.1. Mecanismo estándar de inhibición.....	38
1.3.2. Inhibidores proteínicos de metalopeptidasas .....	40
1.3.3. Inhibidores proteínicos de serín peptidasas .....	43
1.3.3.1. Inhibidor de subtilisina de <i>Streptomyces</i> .....	45
1.3.3.2. Inhibidor de peptidasa neutra de <i>Streptomyces caespitosus</i> .....	49
<b>2. Objetivos</b> .....	<b>51</b>
<b>3. Materiales y métodos</b> .....	<b>55</b>
3.1. Producción y purificación de sermetstatina .....	57
3.2. Obtención de esnapalisina y proesnapalisina. Análisis del prodominio .....	58
3.2.1. Análisis bioinformático .....	59
3.2.2. Producción y purificación de proesnapalisina .....	59
3.2.3. Producción y purificación de esnapalisina.....	60

3.2.4. Estudios de activación <i>in-vitro</i> de proesnapalisina .....	61
3.2.5. Cristalización de proesnapalisina E164A.....	62
3.2.6. Estudios de actividad .....	62
3.3. Análisis de proteólisis y su inhibición por sermetstatina .....	63
3.4. Formación y purificación de complejos.....	64
3.4.1. Subtilisina-sermetstatina .....	64
3.4.2. Esnapalisina-sermetstatina .....	64
3.4.3. Subtilisina-sermetstatina-esnapalisina .....	65
3.5. Cristalización y colección de datos de difracción de rayos X .....	65
3.6. Resolución de estructuras y refinamiento .....	67
3.7. Experimentos de estabilidad de sermetstatina. Puentes disulfuro .....	68
3.8. Miscelánea .....	68
<b>4. Resultados y discusión .....</b>	<b>69</b>
4.1. Producción y purificación de sermetstatina.....	71
4.2. Obtención de esnapalisina y proesnapalisina. Análisis del prodominio .....	73
4.2.1. Análisis bioinformático.....	74
4.2.2. Ensayos de expresión y solubilidad de proesnapalisina y esnapalisina .....	75
4.2.3. Purificación de esnapalisina.....	76
4.2.4. Purificación y estudios de activación de proesnapalisina.....	77
4.2.5. Purificación y estudios de activación de proesnapalisina E164A.....	78
4.2.6. Cristalización de proesnapalisina E164A.....	79
4.2.7. Estudios de actividad .....	80
4.3. Análisis de proteólisis y su inhibición por sermetstatina .....	82
4.4. Formación y purificación de complejos.....	84
4.4.1. Subtilisina-sermetstatina .....	84
4.4.2. Esnapalisina-sermetstatina .....	85
4.4.3. Subtilisina-sermetstatina-esnapalisina .....	87
4.5. Cristalización y colección de datos de difracción de rayos X .....	90
4.6. Resolución de estructuras y refinamiento .....	94
4.7. Estructura de sermetstatina.....	96
4.8. Estructura del complejo subtilisina-sermetstatina .....	99
4.9. Estructura del complejo esnapalisina-sermetstatina.....	102
4.10. Modelo del complejo ternario subtilisina-sermetstatina-esnapalisina .....	105



---

4.11. Rediseño de la especificidad inhibitoria de sermetstatina y análisis de los sitios activos mediante mutagénesis.....	106
4.12. Experimentos de estabilidad de sermetstatina. Puentes disulfuro .....	107
<b>5. Discusión general .....</b>	<b>109</b>
<b>6. Conclusiones .....</b>	<b>117</b>
<b>7. Bibliografía .....</b>	<b>121</b>
<b>Apéndice I .....</b>	<b>145</b>
Mechanism of action of a Janus-faced single-domain protein inhibitor simultaneously targeting two peptidase classes	
<b>Apéndice II .....</b>	<b>169</b>
Ultratight cristal packing of a 10 kDa protein	
<b>Apéndice III .....</b>	<b>179</b>
Functional and structural insights into astacin metallopeptidases	
<b>Apéndice IV .....</b>	<b>197</b>
Structure of the catalytic domain of the <i>Tannerella forsythia</i> matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor	

## FIGURAS

Figura I.1. Colonia de <i>Streptomyces coelicolor</i> .....	5
Figura I.2. Representación esquemática del complejo enzima-substrato .....	9
Figura I.3. Clasificación de peptidasas.....	9
Figura I.4. Clasificación de metalopeptidasas .....	13
Figura I.5. Mecanismo de acción sugerido para metalopeptidasas.....	14
Figura I.6. Mecanismos de inhibición de la catálisis del enlace peptídico dependiente de zinc.....	14
Figura I.7. Diagrama topológico de algunas familias de metzincinas estructuralmente caracterizadas .....	17
Figura I.8. Estructuras tridimensionales de la familia de las metzincinas .....	18
Figura I.9. Estructura de esnapalisina .....	19
Figura I.10. Estructura de prometalopeptidasa de matriz 2 .....	23
Figura I.11. Estructura de proastacina .....	24
Figura I.12. Estructura de profragilisina .....	26
Figura I.13. Clasificación de serín peptidasas.....	27
Figura I.14. Mecanismo de acción sugerido para serín peptidasas con triada catalítica clásica Ser/His/Asp .....	28
Figura I.15. Clan PA de serín peptidasas .....	30
Figura I.16. Clan SB de serín peptidasas.....	32
Figura I.17. Tipos de inhibición de inhibidores proteínicos de peptidasas .....	35
Figura I.18. Inhibidor de doble faz en complejo con su peptidasa diana.....	37
Figura I.19. Mecanismo estándar de inhibición .....	38
Figura I.20. Inhibidores de metalopeptidasas en complejo con sus peptidasas diana .....	41
Figura I.21. Inhibidores canónicos de serín peptidasas en complejo con sus peptidasas diana .....	44
Figura I.22. Inhibidor de subtilisina de <i>Streptomyces</i> .....	46
Figura I.23. Esquema de las funciones biológicas del inhibidor de subtilisina de <i>Streptomyces</i> .....	49
Figura R.1. Gen sintético de sermetstatina .....	71
Figura R.2. Tricina-SDS-PAGE 15% de sermetstatina .....	72
Figura R.3. Cromatograma de exclusión molecular de sermetstatina .....	72

---

Figura R.4. Análisis bioinformático de esnapalisina .....	74
Figura R.5. Predicción de péptido señal en esnapalisina .....	75
Figura R.6. Tricina-SDS-PAGE 15% de la expresión de esnapalisina y proesnapalisina .....	76
Figura R.7. Activación de proesnapalisina .....	77
Figura R.8. Tricina-SDS-PAGE 15% de proesnapalisina E164A .....	79
Figura R.9. Cristalización de proesnapalisina E164A (segmento Q44-R227) .....	79
Figura R.10. Actividad proteolítica de proesnapalisina y esnapalisina .....	81
Figura R.11. Constantes de inhibición de sermetstatina frente a subtilisina y esnapalisina .....	83
Figura R.12. Complejo subtilisina-sermetstatina .....	84
Figura R.13. Complejo esnapalisina-sermetstatina(I) .....	86
Figura R.14. Complejo esnapalisina-sermetstatina (II) .....	87
Figura R.15. Complejo subtilisina-sermetstatina-esnapalisina (I) .....	88
Figura R.16. Complejo subtilisina-sermetstatina-esnapalisina (II) .....	89
Figura R.17. Cristales de sermetstatina nativa .....	90
Figura R.18. Cristales de complejo subtilisina-sermetstatina .....	91
Figura R.19. Cristales de complejo esnapalisina-sermetstatina .....	91
Figura R.20. Cristales de complejo subtilisina-sermetstatina-esnapalisina (I) .....	92
Figura R.21. Cristales de complejo subtilisina-sermetstatina-esnapalisina (II) .....	93
Figura R.22. Estructura de sermetstatina .....	96
Figura R.23. Familia MEROPS I16 y sermetstatina .....	98
Figura R.24. Estructura del complejo subtilisina-sermetstatina .....	100
Figura R.25. Flexibilidad en la interacción subtilisina-sermetstatina .....	101
Figura R.26. Estructura del complejo esnapalisina-sermetstatina .....	103
Figura R.27. Modelo de homología del complejo ternario subtilisina-sermetstatina-esnapalisina .....	105
Figura D.1. Esquema de las funciones biológicas propuestas para sermetstatina .....	115

## TABLAS

Tabla I.1. Substancias de interés terapéutico producidas por bacterias del género <i>Streptomyces</i> .....	6
Tabla I.2. Clasificación de peptidasas de acuerdo con las recomendaciones EC.....	10
Tabla I.3. Secuencias consenso de unión a zinc y giro-Met para cada una de las familias de metzinquinas caracterizadas estructuralmente hasta la fecha .....	15
Tabla I.4. Inhibidores monodominio de doble faz .....	36
Tabla R.1. Actividad inhibitoria de sermetstatina .....	82
Tabla R.2. Datos cristalográficos de las estructuras.....	95
Tabla R.3. Actividad inhibitoria de sermetstatina silvestre y sus variantes.....	106

# ABREVIATURAS



---

<b>A</b>	Adenosina
<b>Å</b>	Ångstrom
<b>Abz</b>	Aminobenzoil
<b>ADAM</b>	Desintegrina y metalopeptidasa ( <i>a disintegrin and a metalloprotease</i> )
<b>ADAMT</b>	Desintegrina y metalopeptidasa con motivos trombospondina ( <i>a disintegrin and a metalloprotease with thombospondin motifs</i> )
<b>ADN</b>	Ácido desoxirribonucleico
<b>API</b>	Inhibidor de peptidasas punta de flecha ( <i>arrowhead protease inhibitor</i> )
<b>APRin</b>	Inhibidor de proteinasas alcalinas ( <i>alkaline proteinase inhibitor</i> )
<b>ARN</b>	Ácido ribonucleico
<b><i>B. subtilis</i></b>	<i>Bacillus subtilis</i>
<b>Bim</b>	<i>Bimane</i>
<b>BPTI</b>	Inhibidor de tripsina pancreática bovina ( <i>bovine pancreatic trypsin inhibitor</i> )
<b>C</b>	Citosina
<b>C-terminal</b>	Carboxi-terminal
<b>Da</b>	Dalton
<b>DO</b>	Densidad óptica
<b>dnp</b>	2,4-dinitrofenilamino
<b>DS</b>	Desviación estándar
<b>DTT</b>	1,4-ditiotreitol
<b>EC</b>	<i>Enzyme commission</i>
<b><i>E. coli</i></b>	<i>Escherichia coli</i>
<b>EDTA</b>	Ácido etilendiaminetetraacético ( <i>ethylenediaminetetraacetic acid</i> )
<b>ESRF</b>	<i>European Synchrotron Radiation Facility</i>
<b>FRET</b>	<i>Fluorescence resonance energy transfer</i>
<b>G</b>	Guanina
<b>GABA</b>	Ácido gamma-aminobutírico
<b>GB1</b>	Proteína G
<b>GST</b>	<i>Glutathione S-transferase</i>
<b>IMPI</b>	Inhibidor de metalopeptidasas de insecto ( <i>insect metallopeptidase inhibitor</i> ) o inhibidor de metaloproteinasas inducible ( <i>inducible metalloproteinase inhibitor</i> )

---

<b>IPTG</b>	Isopropil- $\beta$ -D-tiogalactopiranosido
<b>K<sub>m</sub></b>	Constante de Michaelis-Menten
<b>K<sub>i</sub></b>	Constante de inhibición
<b>LB</b>	Luria Bertani
<b>LCI</b>	Inhibidor de carboxipeptidasas de sanguijuela ( <i>leech carboxypeptidase inhibitor</i> )
<b>MBP</b>	Proteína de unión a maltosa ( <i>maltose binding protein</i> )
<b>MMP</b>	Metalopeptidasa de matriz ( <i>matrix metalloproteinase</i> )
<b>MP</b>	Metalopeptidasa
<b>MT-MMP</b>	Metalopeptidasa de matriz anclada a membrana ( <i>membrane-type matrix metalloproteinase</i> )
<b>N-terminal</b>	Amino-terminal
<b>PAC</b>	Plataforma automatizada de cristalografía
<b>PAGE</b>	Electroforesis en gel de poliacrilamida ( <i>polyacrylamide gel electrophoresis</i> )
<b>PBL</b>	<i>Protein binding loop</i>
<b>PBS</b>	Tampón fosfato salino ( <i>phosphate buffered saline</i> )
<b>PCI</b>	Inhibidor de carboxipeptidasas de patata ( <i>potato carboxypeptidase inhibitor</i> )
<b>PCR</b>	Reacción en cadena de la polimerasa ( <i>polymerase chain reaction</i> )
<b>PDB</b>	Banco de datos de proteínas ( <i>Protein Data Bank</i> )
<b>PEG</b>	Polietilenglicol
<b>PMSF</b>	Fluoruro de fenilmetilsulfonilo ( <i>phenylmethylsulfonyl fluoride</i> )
<b>pNa</b>	Para-nitroanilida
<b>rpm</b>	Revoluciones por minuto
<b><i>S. caespitosus</i></b>	<i>Streptomyces caespitosus</i>
<b><i>S. coelicolor</i></b>	<i>Streptomyces coelicolor</i>
<b>ScNP</b>	Proteasa neutra de <i>Streptomyces caespitosus</i> ( <i>Streptomyces caespitosus neutral protease</i> )
<b>ScNPI</b>	Inhibidor de la proteasa neutra de <i>Streptomyces caespitosus</i> ( <i>Streptomyces caespitosus neutral protease inhibitor</i> )
<b>SDS</b>	Dodecilsulfato sódico ( <i>sodium dodecyl sulfate</i> )
<b>SGMP<sub>II</sub></b>	Metalopeptidasa II de <i>Streptomyces griseus</i> ( <i>Streptomyces griseus metalloendopeptidase II</i> )



---

<b>SmaPI</b>	Inhibidor de metalopeptidasas de <i>Serratia marcescens</i> ( <i>Serratia marcescens metallopeptidase inhibitor</i> )
<b>SMPI</b>	Inhibidor de metalopeptidasas de <i>Streptomyces</i> ( <i>Streptomyces metallopeptidase inhibitor</i> )
<b>SP</b>	Serín peptidasa
<b>SSI</b>	Inhibidor de subtilisina de <i>Streptomyces</i> ( <i>Streptomyces Subtilisin Inhibitor</i> )
<b>Suc</b>	Succinil
<b>T</b>	Timina
<b>TAP</b>	Péptido anticoagulante de garrapata ( <i>tick anticoagulant peptide</i> )
<b>TEV</b>	<i>Tobacco etch virus</i>
<b>TIMP</b>	Inhibidor de metalopeptidasas de tejidos ( <i>tissue inhibitor of metallopeptidases</i> )
<b>TRIS</b>	Tris(hidroximetil)aminometano
<b>trx</b>	Tiorredoxina
<b>V<sub>e</sub></b>	Volumen de elución
<b>V<sub>m</sub></b>	Velocidad máxima
<b>X</b>	Cualquiera de los 20 aminoácidos proteínogénicos

**Código de tres y una letra de los 20 aminoácidos fisiológicos proteinogénicos**

Alanina	<b>Ala</b>	<b>A</b>
Arginina	<b>Arg</b>	<b>R</b>
Asparagina	<b>Asn</b>	<b>N</b>
Ácido aspártico	<b>Asp</b>	<b>D</b>
Cisteína	<b>Cys</b>	<b>C</b>
Fenilalanina	<b>Phe</b>	<b>F</b>
Glicina	<b>Gly</b>	<b>G</b>
Ácido glutámico	<b>Glu</b>	<b>E</b>
Glutamina	<b>Gln</b>	<b>Q</b>
Histidina	<b>His</b>	<b>H</b>
Isoleucina	<b>Ile</b>	<b>I</b>
Leucina	<b>Leu</b>	<b>L</b>
Lisina	<b>Lys</b>	<b>K</b>
Metionina	<b>Met</b>	<b>M</b>
Prolina	<b>Pro</b>	<b>P</b>
Serina	<b>Ser</b>	<b>S</b>
Tirosina	<b>Tyr</b>	<b>Y</b>
Treonina	<b>Thr</b>	<b>T</b>
Triptófano	<b>Trp</b>	<b>W</b>
Valina	<b>Val</b>	<b>V</b>

# **NOMENCLATURA**



Los nombres de las proteínas y sustancias químicas que aparecen en el presente trabajo han sido traducidos o adaptados al castellano. A continuación se muestra su nombre en castellano (en negrita) seguido del original en inglés:

<b>Ácido clavulánico</b>	<b>Calicreína</b>	<b>Gametolisina</b>
Clavulanic acid	Kallikrein	Gametolysin
<b>Actinorodina</b>	<b>Caspasa</b>	<b>Geosmina</b>
Actinorhodin	Caspase	Geosmin
<b>Adamalisina</b>	<b>Catepsina</b>	<b>Glutamil peptidasa</b>
Adamalysin	Cathepsin	Glutamic peptidase
<b>Almelisina</b>	<b>Cisteín peptidasa</b>	<b>Gluzinquina</b>
Almelysin	Cysteine peptidase	Gluzincin
<b>Amfotericina B</b>	<b>Cloranfenicol</b>	<b>Griselinina</b>
Amphotericin B	Chloramphenicol	Griselysin
<b>Antistasina</b>	<b>Colerilisinas</b>	<b>Helicolisinas</b>
Antistasin	Cholerilysin	Helicolysin
<b>Archaemetzinquina</b>	<b>Coelilisinas</b>	<b>Hirudina</b>
Archaemetzincin	Coelilysin	Hirudin
<b>Ascomicolisinas</b>	<b>Daptomicina</b>	<b>Inverzinquina</b>
Ascomycolysin	Daptomycin	Inverzincin
<b>Asparagín peptidasa</b>	<b>Elastasa</b>	<b>Invermectina</b>
Asparagine peptidase	Elastase	Ivermectin
<b>Aspartil peptidasa</b>	<b>Equistatina</b>	<b>Leishmanolisina</b>
Aspartic peptidase	Equistatin	Leishmanolysin
<b>Aspergilopeptidasa</b>	<b>Esnapalisina</b>	<b>Lincomicina</b>
Aspergilopeptidase	Snalysin	Lincomycin
<b>Aspzinquina</b>	<b>Estreptomicina</b>	<b>Macroblobulina</b>
Aspzincin	Streptomycin	Macroglobulin
<b>Astacina</b>	<b>Fosfomicina</b>	<b>Neomicina</b>
Astacin	Fosfomycin	Neomycin
<b>Avermectina</b>	<b>Fragilisina</b>	<b>Nistatina</b>
Avermectin	Fragilysin	Nystatin
<b>Bleomicina</b>	<b>Gabexato mesilato</b>	<b>Quimotripsina</b>
Bleomycin	Gabexate mesilate	Chymotrypsin

**Ornitodorina**

Ornithodorin

**Orizina**

Oryzin

**Papaína**

Papain

**Papalisina**

Papalysin

**Plasmina**

Plasmin

**Plasminostreptina**

Plasminostreptin

**Puomicina**

Puromycin

**Rapamicina**

Rapamycin

**Sedolisina**

Sedolisin

**Serín peptidasa**

Serine peptidase

**Serpina**

Serpin

**Serralisina**

Serralysin

**Subtilisina**

Subtilisin

**Termolisina**

Thermolysin

**Tetraciclina**

Tetracycline

**Treonín peptidasa**

Threonine peptidase

**Tripsina**

Trypsin

**Trombina**

Thrombin

**Turingilisina**

Thuringilysin

**Kanamicina**

Kanamycin

**Metalopeptidasa**

Metallopeptidase

**Metzinquina**

Metzincin

**Migrastina**

Migrastatin

**Monastatina**

Monastatin

**Natamicina**

Natamycin

**Papalisina**

Pappalysin

**Sermetstatina**

Sermetstatin

**Serralisina**

Serralysin

**Subtilisina**

Subtilisin

**Vancomicina**

Vancomycin

**Vimelisina**

Vimelysin

**Voriconazol**

Voriconazole

**Zinquina**

Zincin

# **PREFACIO**





La presente tesis doctoral está centrada en el estudio estructural y funcional de un inhibidor proteínico monodominio de doble faz, sermetstatina, que es capaz de inhibir peptidasas de dos clases catalíticas diferentes (metalopeptidasas y serín peptidasas).

En primer lugar se desarrolla una **Introducción** dividida en tres subapartados. En el primero, con la finalidad de situar a sermetstatina en su contexto biológico, se describe brevemente el género bacteriano *Streptomyces* así como las complejas relaciones que se establecen entre las bacterias del suelo. En el segundo se describen las peptidasas, con especial atención a la estructura, función y regulación de la actividad proteolítica de metalopeptidasas y serín peptidasas. En el tercero se presentan los inhibidores proteínicos de peptidasas, de manera que juntamente con el subapartado anterior, introduce a sermetstatina en su contexto bioquímico, ayudando a la comprensión de los resultados de su estudio estructural y funcional. Posteriormente se describen los **Objetivos** planteados para esta tesis doctoral. En el apartado de **Materiales y métodos** se detallan los procedimientos experimentales seguidos para la obtención de sermetstatina, subtilisina y esnapalisina, la formación de los complejos, cristalización y resolución de sus estructuras, así como para la caracterización funcional del inhibidor y el estudio de la activación del zimógeno de esnapalisina. En el siguiente apartado, **Resultados y discusión**, se presentan los resultados obtenidos acompañados de una discusión de los mismos para cada uno de los subapartados en los que se divide. Seguidamente, en la **Discusión General** se discuten los resultados desde un punto de vista global, integrando todos los datos obtenidos de la realización de este trabajo. A continuación, se exponen las **Conclusiones** extraídas de la presente tesis, y finalmente se incluye toda la **Bibliografía** usada a lo largo del texto.

Este trabajo de tesis doctoral ha dado lugar a la publicación de un artículo científico en la revista *Chemical Science* (incluido en el **Apéndice 1**):

Trillo-Muyo, S., Martínez-Rodríguez S., Arolas J. L. and F. Xavier Gomis-Rüth (2013), 'Mechanism of action of a Janus-faced single-domain protein inhibitor simultaneously targeting two peptidase classes', *Chem Sci*, 4, 791-7.

Además, durante la realización de la presente tesis se ha participado en otros trabajos que han dado lugar a dos artículos científicos publicados (incluidos en los **Apéndices 2 y 3**, respectivamente) y uno aceptado pendiente de publicación (incluido en el **Apéndice 4**):

Trillo-Muyo, S., Domagalski, M. J., Chruszcz, M., Minor, W., Kuisiene, N., Arolas, J. L., Solà, M. and Gomis-Rüth, F. X. (2013), 'Ultratight crystal packing of a 10 kDa protein', *Acta Crystallogr D*, 69, 464-70.

Gomis-Rüth, F. X., Trillo-Muyo, S. and Stöcker, W. (2012), 'Functional and structural insights into astacin metallopeptidases', *Biol Chem*, 393, 1027-41.

Guevara, T., Ksiazek, M., Skottrup, P., Cerdà-Costa, N., Trillo-Muyo, S., de Diego, I., Riise, E., Potempab, J. and Gomis-Rüth, F. X. (2013), 'Structure of the catalytic domain of the *Tannerella forsythia* matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor', *Acta Crystallogr F* (in press).

# 1. INTRODUCCIÓN

Entonces el Dios (Jano), recostándose sobre el báculo que tenía en la diestra dijo:

- Los agujeros suelen estar en los principios. Apartáis los tímidos oídos a la primera voz, y el agorero consulta al ave que ve primero. Se abren los templos, y los oídos de los Dioses, y ninguna lengua pronuncia vanas súplicas; y las palabras surten efecto.
- Oh Jano, ¿por qué aunque quiera aplacar a otros Dioses primero te ofrezco a ti los inciensos y el vino?
- Para que puedas tener entrada por mí, que guardo las puertas a cualesquiera Dioses que quisiere.

Publio Ovidio Nasón (siglo I)

*Fastos*



## 1.1. BACTERIAS DEL SUELO

---

El suelo constituye un ambiente rico y lleno de biodiversidad, sobretodo en su capa más superficial. En él habitan desde insectos y anélidos hasta algas, hongos, protozoos, bacterias y virus. Se estima que los 15 cm superiores de un suelo fértil pueden contener más de 5 toneladas de hongos y bacterias por hectárea (Stainer et al. 1991). De hecho, es habitual considerar al suelo como un sistema vivo en sí mismo, dado que presenta las características de movimiento, respiración, generación de calor, digestión y evolución (Coyne 2000).

La microbiología del suelo es la rama de la edafología que se ocupa de estudiar los microorganismos que habitan en él, sus funciones y sus actividades. Las bacterias representan el grupo más abundante y más diverso de todos los organismos del suelo. La elevada concentración de materia orgánica presente en dicho ambiente favorece el desarrollo de microorganismos heterótrofos. Estos desarrollan un papel fundamental en los procesos de reciclado de los materiales biológicos, ya que descomponen gran parte de los restos de los seres vivos. Los elementos como el carbono, el nitrógeno, el oxígeno, el fósforo y el azufre provenientes de estos restos entran nuevamente en sus respectivos ciclos biogeoquímicos por acción de las bacterias del suelo, siendo estas por tanto indispensables para la vida en la tierra tal y como la conocemos (Stainer et al. 1991).

Las bacterias del suelo pueden clasificarse en dos grandes grupos: especies nativas o autóctonas y especies alóctonas. Las primeras son las propias del suelo. En él pueden sobrevivir, crecer y realizar sus funciones metabólicas, ocupando los nichos ambientales disponibles. Las especies alóctonas, por el contrario, son miembros temporales del hábitat, no ocupan nichos funcionales del ecosistema y normalmente llegan al suelo accidentalmente desde sus propios hábitats. Dentro de las bacterias autóctonas pueden diferenciarse aquellas, cuya actividad en el suelo es lenta y continua, en su mayoría bacilos Gram negativos y actinomicetos, de las bacterias zimogéneas u oportunistas. Estas últimas, típicamente pertenecientes a los géneros *Pseudomonas* y *Bacillus*, son generalmente incapaces de utilizar los compuestos húmicos, pero muestran una actividad intensa y un crecimiento rápido en presencia de substratos fácilmente utilizables como el manto vegetal, excrementos de animales y restos de animales muertos. Entre los géneros más frecuentes de bacterias del suelo se encuentran *Acinetobacter*, *Agrobacterium*, *Alcaligenes*, *Arthrobacter*, *Bacillus*, *Brevibacterium*, *Caulobacter*, *Cellulomonas*, *Clostridium*,

*Corynebacterium*, *Flavobacterium*, *Micrococcus*, *Mycobacterium*, *Nocardia*, *Pseudomonas*, *Staphylococcus*, *Streptococcus*, *Streptomyces* y *Xanthomonas* (Alexander 1977; Atlas and Bartha 2002). La interacción entre poblaciones microbianas es muy frecuente en los suelos. Estas pueden clasificarse como negativas (competencia y amensalismo), positivas (comensalismo, sinergismo y mutualismo) y positivas para una población y negativas para la otra (parasitismo y depredación).

La competencia representa una relación negativa dado que la supervivencia y el crecimiento de las poblaciones bacterianas implicadas se ven perjudicados. La competencia tiene lugar cuando las diferentes poblaciones utilizan un mismo recurso, ya sea el espacio o un nutriente limitante. Algunos de los recursos más comunes por los cuales pueden establecerse estas relaciones son las fuentes disponibles de carbono, nitrógeno, fósforo, hierro y oxígeno (Veldkamp et al. 1984).

El amensalismo, también denominado antagonismo, es otro tipo de relación negativa donde una población se ve perjudicada y la otra no experimenta ninguna alteración. En algunos casos, ciertos microorganismos producen sustancias tóxicas para sus competidores. Estas sustancias pueden afectar o no a la propia población que las produce, pero en todo caso ésta obtiene una ventaja competitiva natural. Para describir estos casos de inhibición competitiva se utilizan los términos antibiosis y alelopatía. El amensalismo puede conducir a una colonización preferente de un hábitat por parte de un microorganismo, donde una vez éste se ha establecido puede impedir que otras poblaciones sobrevivan en el mismo hábitat (Fredrickson and Stephanopoulos 1981).

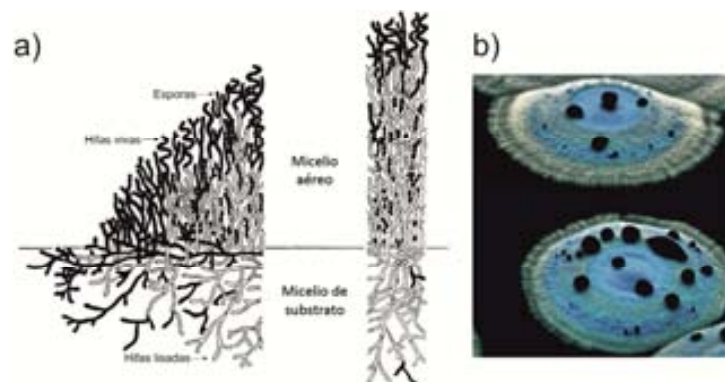
Los mecanismos por los cuales pueden establecerse relaciones de amensalismo son variados, e incluyen la alteración del hábitat mediante la producción de metabolitos primarios como el ácido láctico, ácidos grasos de bajo peso molecular, ácido sulfúrico y alcoholes, así como el consumo o producción de oxígeno (Atlas and Bartha 2002). Algunos microorganismos establecen este tipo de relaciones mediante la producción de antibióticos, sustancias químicas con un gran interés médico. Los antibióticos son metabolitos secundarios que, a bajas concentraciones, matan o inhiben el crecimiento de otros microorganismos. Presentan mecanismos de acción diversos, pudiendo actuar a nivel de pared celular, membrana celular, síntesis proteínica, síntesis de ADN y ARN o replicación del ADN. Por otro lado, se ha postulado la existencia de otros mecanismos implicados en el amensalismo, en los cuales intervienen péptidos o proteínas sintetizados por los microorganismos, como son las bacteriocinas, sideróforos, lisozimas, peptidasas e

inhibidores de peptidasas (Verschuere et al. 2000). Una de las primeras evidencias de la implicación de peptidasas en amensalismo fue hallada en ecosistemas marinos, donde se encontró que numerosas bacterias producían enzimas bacteriolíticas frente a *Vibrio parahaemolyticus* (Nair et al. 1985). En el mismo ecosistema se aisló la cepa *Alteromonas sp.* B-10-31, la cual producía un inhibidor de peptidasas alcalinas denominado monastatina con capacidad para inhibir la peptidasa de *Aeromonas hydrophila* y de *Vibrio anguillarum* (Imada et al. 1985).

### 1.1.1. GÉNERO *STREPTOMYCES*

*Streptomyces* es un género bacteriano perteneciente a la familia Streptomycetaceae dentro del orden de los Actinomycetales, un orden muy diverso de eubacterias Gram-positivas caracterizado por la capacidad para formar estructuras micelares. A su vez, el orden de los Actinomycetales se incluye en el filo Actinobacteria, compuesto por bacterias Gram positivas generalmente con alto contenido de guanina y citosina en su genoma (Stainer et al. 1991).

Las bacterias del género *Streptomyces* se encuentran predominantemente en suelos, principalmente en aquellos con una elevada cantidad de materia orgánica. Generalmente presentan un micelio de substrato incluido en el suelo así como un micelio aéreo que se prolonga fuera del mismo (Figura 1.1). En el extremo de las hifas aéreas se forman las esporas o conidios. En los micelos se alternan hifas vivas con lisadas, concentrándose estas últimas principalmente en el centro de la colonia. El diámetro de las hifas es constante, de aproximadamente 1  $\mu\text{m}$  de diámetro, debido a que el crecimiento ocurre por alargamiento sin incremento en el diámetro celular (Wildermuth 1970).



**Figura 1.1.** Colonia de *Streptomyces coelicolor*. **a)** Diagrama idealizado de la sección vertical del centro de una colonia en fase de esporulación. Se muestran las esporas (círculos negros contiguos), hifas vivas (negro) y lisadas (blancas). Adaptado de (Wildermuth 1970). **b)** Fotografía de una colonia produciendo y excretando el antibiótico pigmentado actinorodina (azul). Adaptado de la portada de la revista *Nature Chemical Biology* de octubre de 2005, volumen 1 (5).

Los actinomicetos fueron descubiertos en las últimas décadas del siglo XIX. En ese momento despertaron un interés minoritario hasta que en 1943 Selman Waksman (1888-1973) purificó el antibiótico estreptomicina a partir de bacterias del género *Streptomyces* (*S. griseus*), convirtiéndose en el primer tratamiento efectivo frente a la tuberculosis (Waksman et al. 1945; Zaffiri et al. 2012). En los años siguientes, nuevos antibióticos de *Streptomyces* fueron descubiertos y caracterizados, consolidando a este género bacteriano como uno de los productores de antibióticos más importantes en la naturaleza (Dyson 2011; Hopwood 2007; Zhou et al. 2011). Además, también se descubrió el ácido clavulánico, que utilizado en combinación con antibióticos  $\beta$ -lactámicos permite bloquear su inhibición por  $\beta$ -lactamasas (Reading and Cole 1977). Por otro lado, se caracterizaron otras sustancias derivadas del metabolismo secundario de *Streptomyces* con aplicaciones médicas como antiparasitarios, inmunosupresores y antineoplásicos (Tabla I.1).

**Tabla I.1.** Sustancias de interés terapéutico producidas por bacterias del género *Streptomyces*.

Producto	Organismo	Referencia
<b>Antibióticos antibacterianos</b>		
Cloranfenicol	<i>S. venezuelae</i> <sup>1</sup>	(Ehrlich et al. 1948)
Daptomicina	<i>S. roseosporus</i>	(Eliopoulos et al. 1986)
Estreptomicina	<i>S. griseus</i>	(Waksman et al. 1945)
Fosfomicina	<i>S. fradiae</i>	(Stapley et al. 1969)
Kanamicina	<i>S. kanamyceticus</i>	(Umezawa et al. 1957)
Lincomicina	<i>S. lincolnensis</i>	(Schaffer et al. 1963)
Neomicina	<i>S. fradiae</i>	(Waksman et al. 1949)
Puromicina	<i>S. alboniger</i>	(Trincao et al. 1954)
Tetraciclina	<i>S. rimosus</i> / <i>S. aureofaciens</i>	(Duggar 1948)
Vancomicina*	<i>S. orientalis</i> <sup>2</sup>	(McCormick et al. 1955)
<b>Antibióticos antifúngicos</b>		
Amfotericina B	<i>S. nodosus</i>	(Oura et al. 1955)
Natamicina	<i>S. natalensis</i>	(Struyk et al. 1957)
Nistatina	<i>S. noursei</i>	(Hazen and Brown 1951)
<b>Antiparasitarios</b>		
Ivermectina	<i>S. avermitilis</i>	(Chabala et al. 1980)
<b>Inmunosupresores</b>		
Rapamicina	<i>S. hygrosopicus</i>	(Vezina et al. 1975)
Tacrolimus	<i>S. tsukubaensis</i>	(Kino et al. 1987)
<b>Antineoplásicos</b>		
Bleomicina	<i>S. verticillus</i>	(Umezawa 1965)
Migrastatina	<i>S. platensis</i>	(Nakae et al. 2000)
<b>Otros</b>		
Ácido clavulánico	<i>S. clavuligerus</i>	(Reading and Cole 1977)

<sup>1</sup> S. es la abreviatura de *Streptomyces*. <sup>2</sup> Actualmente referida como *Amycolatopsis orientalis* (orden Actinomycetales). \* Tradicionalmente incluido como antibiótico producido por *Streptomyces* pese a no serlo.



No obstante, no todos los miembros del género *Streptomyces* producen sustancias beneficiosas para la salud humana. Algunos de ellos incluso pueden ser patogénicos en determinadas circunstancias y causar infecciones. La más común de todas ellas es el micetoma o pie de Madura, una enfermedad subcutánea crónica caracterizada por la formación de nódulos y trayectos fistulosos, causada, entre otros actinomicetos y algunos hongos filamentosos, por *S. somaliensis* y *S. sudanensis* (Welsh et al. 2007).

Las sustancias derivadas del complejo metabolismo secundario de este género bacteriano también pueden tener utilidades biotecnológicas. Unos de los compuestos antiparasitarios producidos por *Streptomyces*, las avermectinas, dieron lugar a una nueva familia de insecticidas (Pitterna et al. 2009). Estas estimulan la secreción del neurotransmisor inhibitorio ácido gamma-aminobutírico (GABA), de manera que bloquean la transmisión neuromuscular causando la muerte de los insectos. Además, algunas especies del género *Streptomyces*, junto con algunos hongos y otros géneros bacterianos pertenecientes en su mayoría a los filos Actinobacteria y Cyanobacteria, son responsables de la producción de geosmina, una sustancia volátil responsable del olor a tierra mojada (Gerber and Lechevalier 1965). Recientemente se ha descrito que insectos como *Drosophila melanogaster* reconocen específicamente geosmina a través de circuitos olfatorios exclusivos, de manera que es interpretada como una alerta ante la presencia de microorganismos dañinos (Stensmyr et al. 2012). Por tanto, esta sustancia sería candidata para su aplicación como repelente de moscas y para evitar la puesta de sus huevos de manera segura para la salud y respetuosa con el medio ambiente.

Por otro lado, las bacterias del género *Streptomyces* producen y segregan al medio extracelular un gran número de peptidasas e inhibidores de peptidasas. Estos intervienen en procesos de morfogénesis y de defensa frente a peptidasas extrínsecas (Taguchi 1995). La peptidasa esnapalisina y el inhibidor proteínico sermetstatina, ambos producidos por *S. caespitosus*, son objeto de estudios del presente trabajo.

La gran cantidad de sustancias de interés producidas por *Streptomyces* ha hecho de este género uno de los más extensamente estudiados. El genoma completo de *S. coelicolor* A3(2) fue publicado el año 2002, representando en ese momento el genoma bacteriano con mayor número de genes conocido hasta la fecha (Bentley et al. 2002). Posteriormente se completó la secuencia genómica de *S. avermitilis* (Ikeda et al. 2003). Cabe destacar que la cepa *S. coelicolor* A3(2) pertenece taxonómicamente a la especie *S. violaceoruber* y no a *S. coelicolor*. No obstante, por conveniencia es habitual referirse a *S. coelicolor* A3(2) como *S. coelicolor* (Chater et al. 2010).

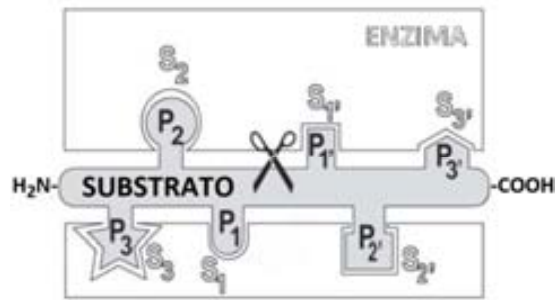
## 1.2. PEPTIDASAS

---

Las peptidasas, también denominadas proteasas, son enzimas que catalizan la hidrólisis del enlace peptídico en proteínas o sustratos peptídicos. La hidrólisis es un proceso generalmente favorable en términos termodinámicos, de manera que las peptidasas catalizan procesos en su mayoría irreversibles. Por tanto, es necesario que su actividad esté estrictamente regulada por la célula/organismo. Esta regulación puede ejercerse a diferentes niveles, incluyendo la regulación génica de la expresión, variaciones epigenéticas, expresión en forma de peptidasas latentes (zimógenos), control de la localización, coexpresión de inhibidores o necesidad de cofactores (Barrett et al. 1998; Pasic et al. 2012).

Son un grupo de enzimas extensamente distribuidas en todos los seres vivos, necesarias para el crecimiento y adaptación de los organismos al medio ambiente. Las peptidasas intervienen en multitud de procesos fisiológicos tales como la homeóstasis, digestión, remodelamiento tisular, ciclo celular, inmunidad, fertilización y apoptosis (Beynon and Bond 2001; Neurath and Walsh 1976; Neurath 1991). Además, en algunos microorganismos la secreción extracelular de peptidasas puede actuar en el establecimiento de relaciones de amensalismo o como factor de virulencia necesario para la infección (Bostanci and Belibasakis 2012; Nair et al. 1985). La actividad de las peptidasas puede resultar en un cambio destructivo (pérdida de función y digestión) o activador (activación de zimógenos). Estas pueden desarrollar su acción proteolítica de manera específica dependiendo de la secuencia aminoacídica del sustrato (proteólisis limitada) o degradándolos completamente (proteólisis de amplio espectro) (Barrett et al. 1998; Wheatley 1984).

El centro activo de las peptidasas tiene una doble función: participar en la unión enzima-sustrato y catalizar la hidrólisis del enlace peptídico. Por tanto, la eficiencia de ambas funciones determinará la actividad de la peptidasa frente a un sustrato. La nomenclatura general de las posiciones de los subsitios en el centro activo fue acuñada por Schechter y Berger (Schechter and Berger 1967) (en el presente trabajo se indicaran siempre en negrita). Se designó el sitio de corte como el enlace situado entre las posiciones **P1** y **P1'**. En dirección N-terminal las posiciones se numeran en orden creciente desde el sitio de corte (**P1+x**), así como en dirección C-terminal (**P1'+x**). Las posiciones de la enzima que encajan con las del sustrato se indican como **Sn** y **Sn'**, donde "n" es la posición (Figura 1.2). Las relaciones existentes entre las diferentes posiciones del sustrato y la peptidasa pueden ser más o menos específicas, condicionando la actividad proteolítica y su función fisiológica.



**Figura 1.2.** Representación esquemática del complejo enzima-substrato. Se indican los subsitios de la enzima (blanco) y del sustrato (negro). Las tijeras representan el sitio de corte. Adaptado de (Gomis-Ruth 2003).

### 1.2.1. CLASIFICACIÓN

Existen diferentes sistemas para la clasificación de las peptidasas dependiendo de las características consideradas. Los más extendidos hacen referencia a la localización del enlace peptídico a hidrolizar, pH de actividad óptima y mecanismo catalítico (Figura 1.3).

Según si desarrollan su acción proteolítica sobre enlaces internos de las cadenas polipeptídicas o a partir de sus extremos las peptidasas se clasifican en dos grandes grupos: endopeptidasas y exopeptidasas. A su vez, las exopeptidasas se dividen en aminopeptidasas y carboxipeptidasas dependiendo de si el aminoácido cortado corresponde al extremo amino-terminal (N-terminal) o carboxi-terminal (C-terminal). Por otro lado, según el pH de actividad óptima las peptidasas pueden ser clasificadas como peptidasas ácidas, peptidasas neutras y peptidasas básicas.

Finalmente, según el mecanismo catalítico que presentan y los residuos implicados en el mismo se clasifican principalmente en cuatro tipos: aspartil peptidasas, cisteín peptidasas, metalopeptidasas y serín peptidasas (Rawlings and Barrett 1993). Existen también otros tipos menos extendidos descritos posteriormente como las treonín peptidasas (Lowe et al. 1995), glutamín peptidasas (Fujinaga et al. 2004) y asparagín peptidasas (Reddy et al. 2004).



**Figura 1.3.** Clasificación de peptidasas. Clasificación según su sitio de acción, pH óptimo y mecanismo catalítico.

### 1.2.1.1. Base de datos EC

El sistema de clasificación EC (*Enzyme Commission*) utiliza la nomenclatura recomendada por el comité de nomenclatura de la unión internacional de bioquímica y biología molecular (*Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*; NC-IUBMB) con el asesoramiento de la Unión Internacional de Química Pura y Aplicada (*International Union of Pure and Applied Chemistry*; IUPAC).

Actualmente recomienda el uso del término “peptidasa” frente al sinónimo “proteasa”, aunque tradicionalmente, de acuerdo con *Enzyme Nomenclature* (International Union of Biochemistry 1984), reservaba el término “peptidasa” para referirse específicamente a exopeptidasas y “proteasa” para endopeptidasas. Esta base de datos puede ser consultada en la dirección electrónica <http://www.chem.qmul.ac.uk/iubmb/enzyme/EC3/4/>.

Según este sistema, las peptidasas se organizan en una clasificación de tres niveles: tipo catalítico, estructura molecular y peptidasas individuales. En el sistema EC las peptidasas se subdividen en 14 sub-subclases según su tipo catalítico (Tabla I.2). Una de las características más interesantes a nivel práctico de este sistema es que para cada peptidasa existe un único código EC asignado, de manera que puede ser utilizado como referencia inequívoca para la enzima. Además de este código, también se indica el nombre recomendado, así como la lista de otros nombres posibles presentes en la literatura científica.

**Tabla I.2.** Clasificación de peptidasas de acuerdo con las recomendaciones EC.

Sub-Subclase	Tipo de peptidasa
EC 3.4.11	Aminopeptidasas
EC 3.4.13	Dipeptidasas
EC 3.4.14	Dipeptidil-peptidasas y tripeptidil-peptidasas
EC 3.4.15	Peptidil-dipeptidasas
EC 3.4.16	Serín carboxipeptidasas
EC 3.4.17	Metalocarboxipeptidasas
EC 3.4.18	Cisteín carboxipeptidasas
EC 3.4.19	Omega peptidasas
EC 3.4.21	Serín endopeptidasas
EC 3.4.22	Cisteín endopeptidasas
EC 3.4.23	Aspartil endopeptidasas
EC 3.4.24	Metaloendopeptidasas
EC 3.4.25	Treonín endopeptidasas
EC 3.4.99	Endopeptidasas de mecanismo catalítico desconocido

### **1.2.1.2. Base de datos MEROPS**

El sistema de clasificación MEROPS se basa en la agrupación de las peptidasas en clanes y familias (Rawlings et al. 2012). Tradicionalmente recomienda el uso del término “peptidasa” frente a “proteasa”, poniendo énfasis en que las enzimas proteolíticas actúan más frecuentemente sobre péptidos que sobre proteínas. La base de datos puede ser consultada en la dirección electrónica <http://merops.sanger.ac.uk/>. Ofrece para cada peptidasa la información disponible sobre nomenclatura, clasificación, dominios, secuencia, distribución, actividad, substratos, inhibidores, estructura y referencias bibliográficas.

De igual forma que el sistema EC, la base de datos MEROPS se organiza según los niveles de tipo catalítico, estructura molecular y peptidasas individuales. Este sistema utiliza un sistema de clasificación jerárquico basado en la estructura primaria y terciaria de las peptidasas. En este, cada una de ellas es asignada a una familia de acuerdo con las similitudes estadísticamente significativas en sus secuencias aminoacídicas. Las similitudes en la estructura primaria reflejan un origen evolutivo compartido, permitiendo la agrupación de peptidasas relacionadas. Además, las familias se agrupan en clanes en base al plegamiento que presentan. Las similitudes en el plegamiento de las proteínas persisten evolutivamente más allá de las similitudes en la estructura primaria, de manera que el análisis de estas puede revelar relaciones distantes, dando lugar a la agrupación en clanes. Las peptidasas de mecanismo catalítico desconocido se agrupan en familias sin clan asignado (*unknown catalytic type*). Por otro lado, la base de datos MEROPS también ofrece información sobre inhibidores proteínicos. Estos se clasifican de igual manera que las peptidasas, agrupados en familias y clanes. Finalmente, se incluye un apartado con inhibidores de molécula pequeña.

### **1.2.2. METALLOPEPTIDASAS**

Las metalopeptidasas son peptidasas que requieren un metal para realizar su función catalítica. La mayoría de ellas requieren zinc, aunque existen metalopeptidasas que utilizan otros metales como cobalto o manganeso (Gomis-Ruth et al. 1994; Roderick and Matthews 1993). El átomo del metal catalítico se encuentra normalmente coordinado por las cadenas laterales de tres o cuatro aminoácidos y una molécula de agua esencial para la hidrólisis del enlace peptídico. Los ligandos más comunes en metalopeptidasas son por orden de ocurrencia histidina, ácido glutámico, ácido aspártico y lisina. En algunos casos también

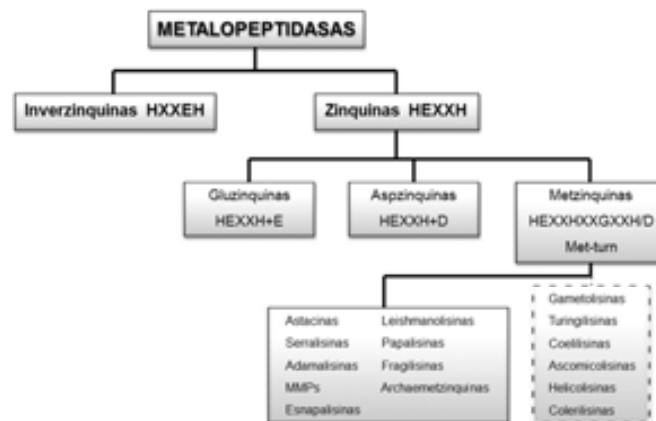
puede encontrarse una cisteína desapareada coordinando el metal. Este hecho es común en el caso de prodominios, donde una vez la peptidasa latente es activada una molécula de agua reemplaza a la cisteína en su interacción con el metal (Morgunova et al. 1999).

Estas enzimas participan en diversos procesos biológicos como reproducción, desarrollo embrionario, morfogénesis, remodelamiento tisular, procesamiento de hormonas peptídicas, liberación de citoquinas y factores de crecimiento, fusión celular, adhesión y migración celular, absorción intestinal de nutrientes, inmunidad, procesamiento de proteínas virales, infección, biosíntesis de la pared bacteriana y metabolismo de antibióticos (Barrett et al. 1998; Bartlett and Smith 2013; Kumar 2008; Wu and Chen 2011). El funcionamiento anormal de las metalopeptidasas implicadas en estos procesos, ya sea por exceso o por defecto de actividad, es responsable de numerosas patologías tales como nefritis, cáncer, enfermedad cardiovascular, infección, etc. (Nagase 2001). Por ello, estas peptidasas constituyen dianas terapéuticas para el diseño de fármacos destinados a combatir algunas de las enfermedades asociadas (Clutterbuck et al. 2009).

Tal como ocurre con las peptidasas en general, existen también diferentes métodos para la clasificación de las metalopeptidasas. Según la base de datos MEROPS se dividen en 67 familias con diversas subfamilias. Estas se agrupan por su relación evolutiva en 16 clanes. En los casos en que la estructura terciaria se encuentra disponible, esta se utiliza para establecer las relaciones entre las familias. Por el contrario, para metalopeptidasas de estructura desconocida se utilizan las similitudes en la estructura primaria del centro catalítico y alrededores. No obstante, la asignación de estas peptidasas no siempre es posible, de manera que seis familias (M73, M77, M79, M82, M87 y M88) no se encuentran asignadas a ningún clan a la espera de la resolución de su estructura terciaria. Los clanes se identifican con dos letras, la primera de las cuales en el caso de las metalopeptidasas es siempre una "M".

No obstante, otra clasificación (Figura I.4) propone la subdivisión de las metalopeptidasas en dos grandes subgrupos dependiendo de la secuencia aminoacídica de unión al zinc: zinquinas (motivo HEXXH) e inverzinquinas (HXEH) (Gomis-Ruth 2003). En ambos casos las histidinas actúan de ligando del zinc catalítico y el glutamato como ácido/base general en la reacción (Barret et al., 1998). Las zinquinas representan el grupo más extenso de metalopeptidasas. Estas se clasifican según el tercer ligando del zinc catalítico. Dependiendo de si el residuo implicado es glutamato o aspartato se denominan gluzinquinas (HEXXH+E) o aspzinquinas (HEXXH+D). Otro tipo de zinquinas, llamado

metzinquinas, está caracterizado por una secuencia consenso de unión a zinc más larga, HEXXHXXGXXH/D+M, con una glicina adicional altamente conservada, una histidina o aspartato como tercer ligando del zinc y una metionina invariable situada en la tercera posición de un giro de la cadena principal situado debajo del ion zinc catalítico (giro-Met) (Bode et al., 1993). A su vez, las metzinquinas se dividen en diferentes familias según el tercer residuo de unión al zinc y sus residuos circundantes, así como por los residuos presentes alrededor del giro-Met.



**Figura I.4.** Clasificación de metalopeptidasas. Clasificación en función de la secuencia consenso de unión a zinc. Detalle de las familias de metzinquinas, incluyendo familias de estructura conocida (línea continua) y desconocida (línea discontinua). MMPs son metalopeptidasas de matriz. Adaptado de (Gomis-Ruth 2003).

La determinación de las estructuras ternarias de diversas metalopeptidasas ha permitido postular el mecanismo de reacción de la hidrólisis del enlace peptídico catalizado por estas enzimas. A diferencia de serín y cisteín peptidasas, y tal como ocurre con aspartil peptidasas, las metalopeptidasas no forman intermediarios covalentes durante el proceso. Los elementos de la enzima implicados en el mecanismo son el ion metálico, la molécula de agua unida a él y la cadena lateral del ácido glutámico presente en el motivo HEXXH. Una de las propiedades más relevantes del zinc para su acción catalítica es que se trata de un ácido de Lewis fuerte. Durante la reacción, el ion zinc coordina una molécula de agua nucleofílica a la vez que coordina el oxígeno del grupo carboxilo del sustrato. Su número de coordinación flexible (cuatro-seis) unido a su capacidad para adoptar diferentes conformaciones geométricas (tetraédrico, trigonal piramidal, piramidal cuadrada y octaédrica) y de cambiar de ligando rápidamente, permite al zinc adaptarse a los reordenamientos estructurales que ocurren durante el proceso (Alberts et al. 1998). La hidrólisis está mediada por el ataque nucleofílico de la molécula de agua polarizada sobre el grupo carbonilo del enlace peptídico hidrolizable. Esta reacción está asistida por el grupo carboxilato del ácido glutámico. Este actúa como base captando el protón de la molécula de agua y posteriormente como ácido transfiriéndolo al nuevo extremo N-terminal (Gomis-Ruth 2008) (Figura I.5).

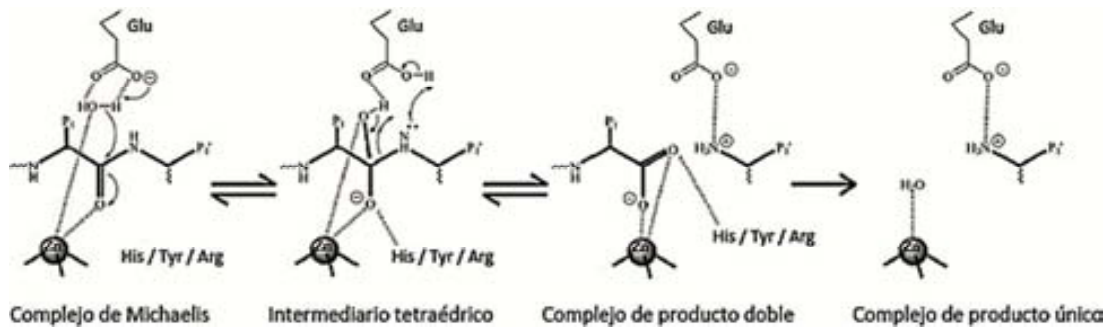


Figura I.5. Mecanismo de acción sugerido para metalopeptidasas. Adaptado de (Gomis-Ruth 2008).

Existen diversos mecanismos para inactivar la catálisis en metalopeptidasas. En apartados posteriores se tratará en detalle la inhibición a través de prodominios (en zimógenos; apartado 1.2.2.1.2) y la inhibición mediante inhibidores proteínicos (apartado 1.3.2). Otros de los mecanismos más destacados son el uso de inhibidores de molécula pequeña, la eliminación del zinc catalítico mediante quelantes y el exceso de zinc (Figura I.6).

Algunos inhibidores de molécula pequeña pueden substituir a la molécula de agua reactiva de la esfera de coordinación del zinc y unirse a la hendidura del centro activo, evitando que la peptidasa pueda actuar sobre su sustrato. Estos actúan mimetizando los estados intermediarios de la reacción de hidrólisis de los sustratos peptídicos bloqueando el mecanismo de acción. Algunos de estos inhibidores sintéticos contienen regiones quelantes como pueden ser grupos carboxilos, tioles, fósforo o ácido hidroxámico, unidas a otros grupos químicos, aminoácidos o no, que encajan en los subsitios de la enzima confiriéndoles especificidad (Woessner 1999).

Quelantes de metales como el ácido etilendiaminotetraacético (EDTA), ácido etilenglicol tetraacético (EGTA) y 1,10-fenantrolina presentan una elevada afinidad por el zinc, de manera que pueden entrar en el sitio activo y captar el zinc catalítico, dando lugar a apoenzimas inactivas (Barrett et al. 1998). Se trata de un proceso generalmente reversible, donde la eliminación del quelante y la adición del metal comportan la recuperación de la actividad proteolítica.

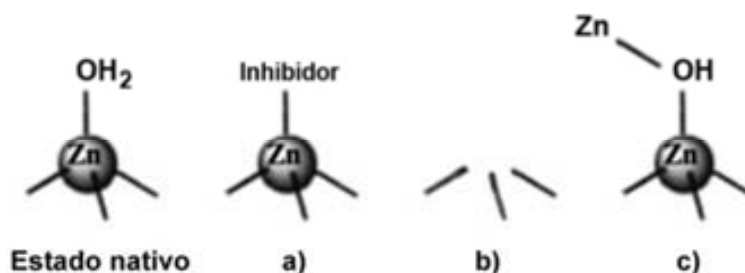


Figura 1.6. Mecanismos de inhibición de la catálisis del enlace peptídico dependiente de zinc. a) Unión de inhibidor específico al metal. b) Eliminación del zinc. c) Unión de un segundo átomo de zinc a la molécula de agua reactiva.



Además, la presencia de elevadas concentraciones de zinc, generalmente en el rango de milimolar, también suele comportar la inhibición de las metalopeptidasas. Esto es debido a la formación de monohidróxido de zinc que se une al ion de zinc catalítico del centro activo. Por tanto, se trata de un tipo de inhibición competitiva con el sustrato. Estudios estructurales demostraron que el zinc inhibitorio se une al centro activo formando un complejo de coordinación tetraédrico distorsionado, afectando al ácido glutámico implicado en el mecanismo de reacción. El segundo zinc mantiene al grupo hidróxido en una localización prácticamente idéntica a la que ocuparía la molécula de agua reactiva perturbando el posicionamiento del sustrato y los reajustes estereoquímicos requeridos durante la catálisis (Gomez-Ortiz et al. 1997).

### 1.2.2.1. Metzinquinas

Las metzinquinas se caracterizan por presentar la secuencia consenso HEXXHXXGXXH/D+M. Estas se dividen en diferentes familias según el tercer residuo de unión al zinc y sus residuos circundantes, así como por los residuos presentes alrededor del giro-Met (Tabla I.3).

**Tabla I.3.** Secuencias consenso de unión a zinc y giro-Met para cada una de las familias de metzinquinas caracterizadas estructuralmente hasta la fecha.

Familia	Secuencia consenso*
Astacinas	H-E-B-X-H-A/V-B-G-F-X-H-E/Q-X-X-R-X-D-R-D-X-O-[23-37]-Y-D-O-X-S-I/V-M-H-Y
ADAMs <sup>1</sup>	H-E-B-G-H-X-L/F-G-M/I-X-H-D-[8-31]-C/O-B-M-X-X
MMPs <sup>2</sup>	H-E-B-G-H-X-L-G-L-X-H-S-[4-6]-B-M-X-P
Leishmanolisinas	H-E-B-X-H-A-B-G-F-[58-96]-S-H-B-[6-7]-E-B-M-X-X
Serralisinas	H-E-I-G-H-A/T-L-G-L-X-H-P-G-D/E-Y-N-A-G-X-G-D-[1-7]-D-T/S-R-Q/G-F/Y-S-I/V-M-S-Y
Papalisinas	H-E-B-G-H-X-B-G-L-B-H-X-O-[38-63]-N-B-M-X-Y
Archametzinquinas	H-E-B-G-H-X-B-G-L-X-H-C-X-N-[1-2]-C-B-M-X-X
Fragilisinas	H-E-B-G-H-B-L-G-A-X-H-X-D-[]-D-B-M-O-X
Esnapalisinas	H-E-T-G-H-V-L-G-L-P-D-X-Y-S-G-P-C-S-X-L-M-S-G

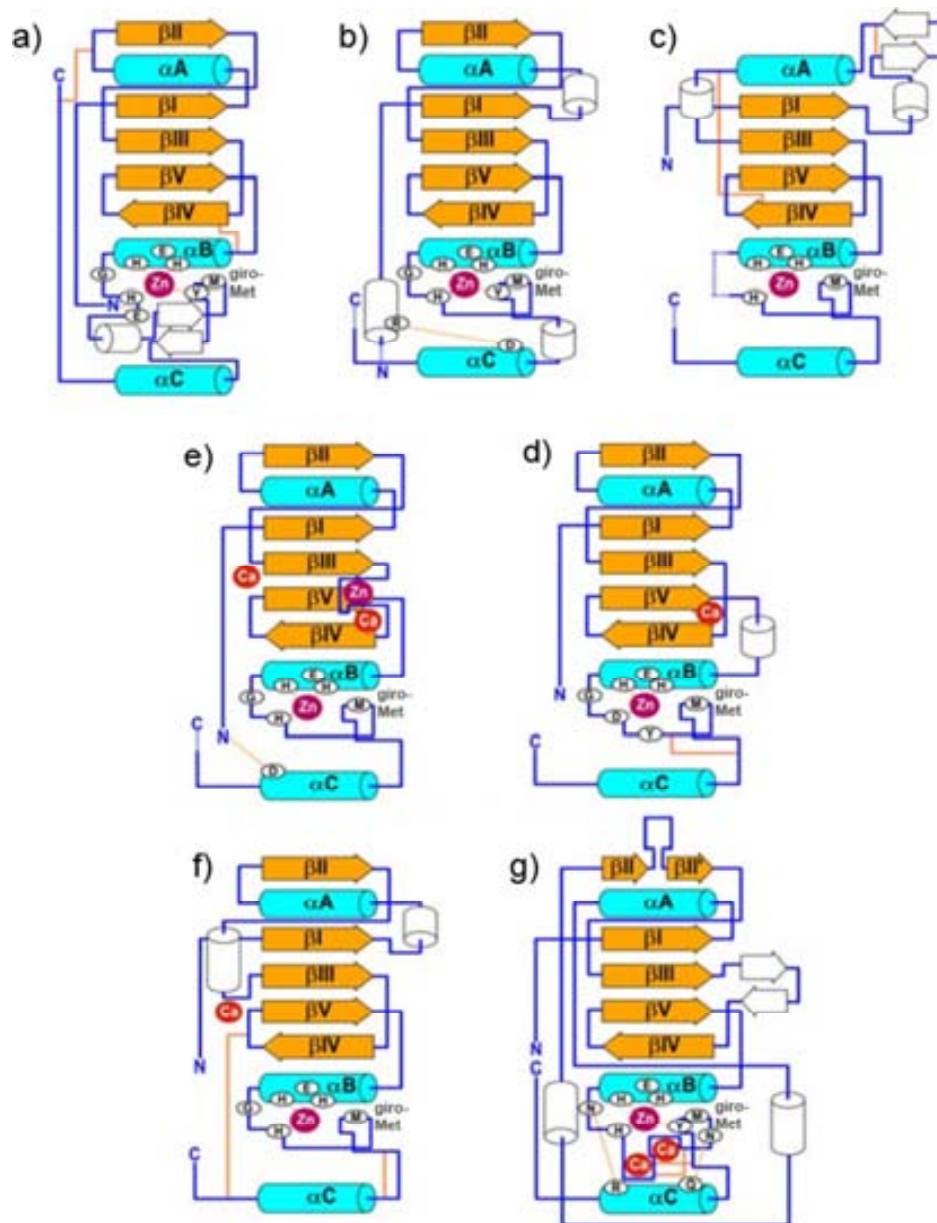
\*Se indican los aminoácidos de unión al ion zinc (rojo), el ácido glutámico, glicina y metionina conservadas en las metzinquinas (naranja) y los residuos más característicos de cada familia (verde). Código: B, residuo voluminoso apolar (y/o aromático), O, residuo aromático, X, cualquier aminoácido. <sup>1</sup> Desintegrinas y metalopeptidasas. <sup>2</sup> Metalopeptidasas de matriz.

Hasta la fecha se ha determinado la estructura tridimensional de diversas metzincinas dando lugar a su clasificación en nueve familias. Estas son astacinas (Bode et al. 1992), adamalinas (desintegrinas y metaloproteinasas; ADAMs) (Gomis-Ruth et al. 1993), matrixinas (metalopeptidasas de matriz celular; MMPs) (Li et al. 1995), leishmanolisinas (Stöcker et al., 1995), serralinas (Hamada et al. 1996), esnapalinas (Kurusu et al. 1997), papalinas (Tallant et al., 2006), archaemetzincinas (Waltersperger et al. 2010) y fragilinas (Goulas et al. 2011). No obstante, el estudio bioinformático de diferentes genomas ha permitido identificar hipotéticas metzincinas que podrían constituir nuevas familias. Algunas de estas nuevas familias propuestas son las gametolisinas, turingilinas, coelilinas, ascomicolinas, helicilinas y colerilinas (Gomis-Ruth 2003).

Los esqueletos de las cadenas principales de los miembros de las diferentes familias de metzincinas muestran un elevado grado de similitud estructural. Todos ellos están constituidos por un subdominio N-terminal y otro C-terminal separados por la hendidura del centro activo. Algunos de los elementos estructurales comunes son cuatro hebras  $\beta$  paralelas ( $\beta$ I,  $\beta$ II,  $\beta$ III y  $\beta$ V), una hebra  $\beta$  antiparalela ( $\beta$ IV) y tres hélices  $\alpha$  largas ( $\alpha$ A,  $\alpha$ B y  $\alpha$ C), las cuales se organizan secuencialmente de la misma manera:  $\beta$ I,  $\alpha$ A,  $\beta$ II,  $\beta$ III,  $\beta$ IV,  $\beta$ V,  $\alpha$ B,  $\alpha$ C (Figura I.7) (Gomis-Ruth 2003).

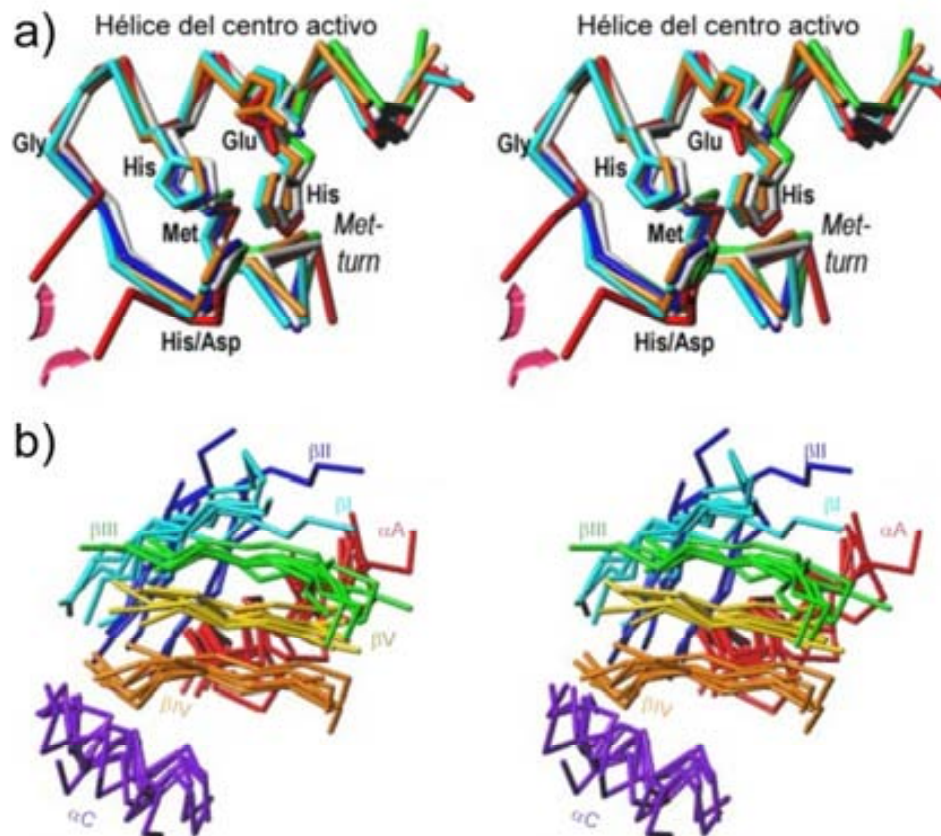
El subdominio N-terminal está compuesto por una hoja  $\beta$  formada por las hebras  $\beta$ I,  $\beta$ II,  $\beta$ III,  $\beta$ IV y  $\beta$ V (exceptuando las leishmanolisinas que tienen cuatro hebras) flanqueada por las hélices  $\alpha$ A y  $\alpha$ B, de manera que la hélice  $\alpha$ A se inserta entre las hebras  $\beta$ I y  $\beta$ II paralelas. La hebra  $\beta$ III, situada en el centro de la hoja  $\beta$ , es la hebra que presenta menor variabilidad entre las diferentes familias tanto en orientación como en longitud. La única hebra  $\beta$  antiparalela,  $\beta$ IV, forma el límite superior de la hendidura del centro activo. Esta interacciona con el sustrato de manera antiparalela, principalmente a través de su cadena principal. El lazo conector de las hebras  $\beta$ III y  $\beta$ IV interviene en la especificidad de las metzincinas, ya que condiciona la morfología y la carga de los subsitios **S1'** y **S2'**. La hélice  $\alpha$ B contiene el motivo HEXXH y conecta el subdominio N-terminal con el C-terminal.

El centro activo es similar en todas las familias de metzincinas. La superposición de las hélices  $\alpha$ B de las diferentes familias muestra un encaje prácticamente perfecto (Figura I.8). Las histidinas de unión al zinc se encuentran separadas por un único giro en la hélice permitiendo la coordinación del metal a través del grupo imidazol de sus cadenas laterales. A diferencia de otras metalopeptidasas, las metzincinas presentan un residuo de glicina conservado al final de esta hélice, permitiendo el giro brusco de la cadena polipeptídica.



**Figura 1.7.** Diagrama topológico de algunas familias de metzincinas estructuralmente caracterizadas. **a)** Astacinas. **b)** Serralisinas. **c)** Desintegrinas y metalopeptidasas (ADAMs). **d)** Metalopeptidasas de matriz (MMPs). **e)** Esnapalisinas. **f)** Leishmanolisinas. **g)** Papalisinas. Adaptado de (Gomis-Ruth 2009).

El subdominio C-terminal empieza a partir de la glicina conservada. Este presenta una gran variabilidad en tamaño y estructura entre las diferentes familias. No obstante contiene algunas regiones conservadas como una hélice larga C-terminal ( $\alpha C$ ), próxima al sitio de unión a zinc, y las regiones del tercer ligando del zinc catalítico y del giro-Met. Este último consiste en un giro 1,4- $\beta$  que contiene una metionina en posición 3. El giro-Met se sitúa por debajo de la base piramidal de los aminoácidos que coordinan el ion zinc separado del tercer ligando del zinc por un segmento de entre seis y 53 aminoácidos de estructura variable. El giro de la cadena principal por detrás del ion zinc crea un suelo hidrofóbico, pero en ningún caso existe contacto directo con el metal o los substratos (Gomis-Ruth 2009).



**Figura 1.8.** Estructuras tridimensionales de la familia de las metzincinas. Detalle en estereoscopia de ojos paralelos<sup>1</sup> de la superposición de los segmentos comunes de las cadenas principales de miembros representativos de siete familias de metzincinas. Adaptado de (Gomis-Ruth 2009). **a)** Hélice del centro activo ( $\alpha$ B) y giro-Met. Las familias representadas son astacinas (azul), desintegrinas y metalopeptidasas (ADAMs; cian), leishmanolisinas (rojo), metalopeptidasas de matriz (MMPs; verde), serralisinas (blanco), esnapalisinas (amarillo) y papalisinas (naranja). Las dos flechas indican la presencia de un dominio inserto en esa posición en leishmanolisinas. **b)** Hebras  $\beta$ I (cian),  $\beta$ II (azul),  $\beta$ III (verde),  $\beta$ IV (naranja) y  $\beta$ V (amarillo), y hélices  $\alpha$ A (rojo) y  $\alpha$ C (morado).

### 1.2.2.1.1. Esnapalisinas

La primera peptidasa identificada dentro de la familia de las esnapalisinas, y hasta el momento la mejor estudiada, es la esnapalisina de *Streptomyces caespitosus* (código UniProt P56406), también conocida como proteasa neutra extracelular pequeña (*extracellular small neutral protease*) o proteasa neutra de *S. caespitosus* (*Streptomyces caespitosus neutral protease*; ScNP). Es una zinc metalopeptidasa secretada al medio extracelular de *S. caespitosus* que cataliza específicamente la hidrólisis del enlace peptídico requiriendo un residuo aromático en posición **P1'** (Yokote 1969). Constituida por 132 residuos y con un peso molecular de 14,38 kDa se trata de una de las peptidasas más pequeñas descritas hasta la fecha.

<sup>1</sup> En la estereoscopia de ojos paralelos, la imagen izquierda es percibida por el ojo izquierdo y viceversa. Para apreciar el efecto 3D es necesario mirar "a través" del plano de proyección, como si se estuviera mirando al horizonte lejano, y tratar de enfocar la imagen.

Esta peptidasa, objeto de estudio del presente trabajo, contiene la secuencia *H*<sub>83</sub>*E*<sub>84</sub>*TGH*<sub>87</sub>*VLG*<sub>90</sub>*LPD*<sub>93</sub>*M*<sub>103</sub> (aminoácidos de esnapalisina representados en cursiva; números de residuos indicados como subíndice de acuerdo con la secuencia de la proteína madura; UniProt P56406) que excepto por *D*<sub>93</sub> se ajusta a la secuencia consenso característica de las metzincinas (HEXXHXXGXXH-M). El residuo del ácido aspártico actúa de la misma manera que la histidina presente en la posición equivalente de la mayoría de las metzincinas coordinando un átomo de zinc. Este presenta una coordinación tetraédrica con las cadenas laterales de *H*<sub>83</sub> (2,01 Å), *H*<sub>87</sub> (2,01 Å), *D*<sub>93</sub> (1,95 Å) y una molécula de agua (Kurusu et al. 2000).

La estructura de esnapalisina está constituida por cinco hebras  $\beta$ , cuatro hélices  $\alpha$ , un ion calcio y un ion zinc catalítico (Figura 1.9). Cuatro hebras  $\beta$  se disponen de manera paralela ( $\beta$ I,  $\beta$ II,  $\beta$ III y  $\beta$ V) y una de manera antiparalela ( $\beta$ IV) formando una hoja  $\beta$ . Su lado cóncavo se encuentra flanqueado por las hélices  $\alpha$ A y  $\alpha$ C, mientras el convexo está expuesto al solvente adoptando una estructura de sándwich abierto. Se pueden distinguir dos regiones, la región N-terminal (*T*<sub>1</sub>-*D*<sub>76</sub>), organizada en los elementos de estructura secundaria típicos de las metzincinas, que incluyen las cinco hebras  $\beta$  y las hélices  $\alpha$ A y  $\alpha$ B, y la región C-terminal (*G*<sub>90</sub>-*G*<sub>132</sub>) que presenta un plegamiento más irregular.



**Figura 1.9.** Estructura de esnapalisina. Figura preparada con el programa PyMOL (Schrodinger 2010). Código *Protein Data Bank* (PDB) 1C7K (Kurusu et al. 2000). Se indican las hebras  $\beta$  ( $\beta$ I- $\beta$ V; magenta), hélices  $\alpha$  ( $\alpha$ A- $\alpha$ C; cian), ion calcio (rosa), ion zinc (gris), aminoácidos de unión a zinc (rojo), glicina y metionina conservada en metzincinas (naranja).

La hélice  $\alpha$ C, que se encuentra inmediatamente por detrás del zinc catalítico, conecta ambas regiones configurando el centro activo. Este se encuentra bordeado por la hebra  $\beta$ IV, la propia hélice  $\alpha$ C y un lazo adyacente al extremo C-terminal de esa hélice. La hélice  $\alpha$ C discurre paralela desde la hendidura hasta el centro de la peptidasa e incluye la secuencia HEXXH. Termina en *G*<sub>90</sub>, donde la cadena polipeptídica gira bruscamente

permitiendo a la cadena lateral de  $D_{93}$  interactuar con el ion zinc catalítico. Tras este residuo la cadena adopta una estructura de lazos, que consiste en una alternancia consecutiva de giros  $\beta$  tipo I y III, incluyendo  $H_{94}$ - $G_{97}$  (tipo I),  $P_{98}$ - $E_{101}$  (tipo III),  $E_{101}$ - $S_{104}$  (tipo I),  $G_{105}$ - $P_{108}$  (tipo III). La metionina incluida en el giro  $\beta$   $E_{101}$   $L_{102}$   $M_{103}$   $S_{104}$  corresponde al giro-Met propio de las metzinquinas. Las cadenas laterales de  $L_{102}$  y  $M_{103}$  proporcionan un entorno hidrofóbico cercano a  $H_{83}$  afectando a la acidez del zinc catalítico y promoviendo la disociación del protón de  $Zn^{2+}$ - $OH_2$ . Tras  $S_{104}$ , la cadena polipeptídica adopta un plegamiento irregular ( $G_{105}$ - $N_{114}$ ), estabilizado en gran medida por la formación de un puente disulfuro entre  $C_{99}$  y  $C_{112}$  (Kurusu et al. 1997).

Respecto a su actividad proteolítica, esnapalisina presenta una actividad óptima a pHs de entre 6,4 y 7,0. Esta se incrementa exponencialmente con la adición de NaCl, siendo esta 3,7 veces mayor a una concentración de 3,6 M que a 0 M. Por otro lado, la actividad disminuye con la adición de LiCl, KCl, NaBr, LiBr, KBr o  $NaClO_4$ . La adición de alcoholes también comporta una reducción en la actividad de esnapalisina, presentando mayor inhibición frente a alcoholes de cadena larga y alcoholes terciarios (Inouye et al. 2007).

Hasta el momento, todos los estudios publicados sobre esnapalisina se han realizado sobre proteína purificada de medio de cultivo de *S. caespitosus*. Al proceder de esta manera se parte de la peptidasa activa en su forma madura descrita anteriormente. No obstante, esnapalisina se sintetiza en forma de zimógeno en el interior de la bacteria y es tras su secreción que se produce la activación. Se desconoce la secuencia del prodominio de esnapalisina de *S. caespitosus* debido a que no se ha podido aislar el zimógeno y el genoma de este organismo no ha sido secuenciado. Sin embargo, existen homólogos de esnapalisina en organismos muy próximos en términos evolutivos, como es el caso de *S. coelicolor*, cuyo genoma sí ha sido secuenciado (Bentley et al. 2002). La existencia de esnapalisina en *S. coelicolor* ha sido inferida por homología, de manera que pese a ser un organismo ampliamente estudiado, no existen evidencias reportadas a nivel proteínico.

#### **1.2.2.1.2. Regulación por zimogenicidad en metzinquinas**

La actividad proteolítica de la mayoría de metalopeptidasas está estrictamente regulada, de manera que únicamente tiene lugar donde y cuando es requerida (Barrett et al. 1998; Holzer and Heinrich 1980). Este control puede ejercerse mediante la expresión de las peptidasas como zimógenos, la mayoría de los cuales contienen prosegmentos inhibidores N-terminales.

Estos prosegmentos, también denominados prodominios, bloquean el acceso de los sustratos al centro activo de la proteína y son eliminados durante la maduración de las enzimas (Bryan 2002; Khan and James 1998). Por otro lado, los prosegmentos suelen plegarse de manera independiente actuando como guía en el proceso de plegamiento del resto de la proteína a modo de chaperonas intramoleculares (Bryan 2002). La importancia del mantenimiento del estado de latencia durante el tránsito intracelular hacia su localización definitiva hace que el estudio de los mecanismos moleculares mediante los cuales las metalopeptidasas mantienen este estado, así como el estudio de su activación, sean imprescindible para el entendimiento de su modo de acción. Además, su estudio facilita el diseño de inhibidores que puedan mimetizar la interacción del prodominio con el centro activo de la peptidasa, para así modular su actividad como parte de aproximaciones terapéuticas o con aplicaciones biotecnológicas.

Como se ha comentado anteriormente, hasta la fecha se han resuelto las estructuras de miembros de nueve subfamilias dentro de la familia de las metzincinas. De todas ellas, únicamente se ha obtenido la estructura del zimógeno para MMPs, astacinas y fragilisinias.

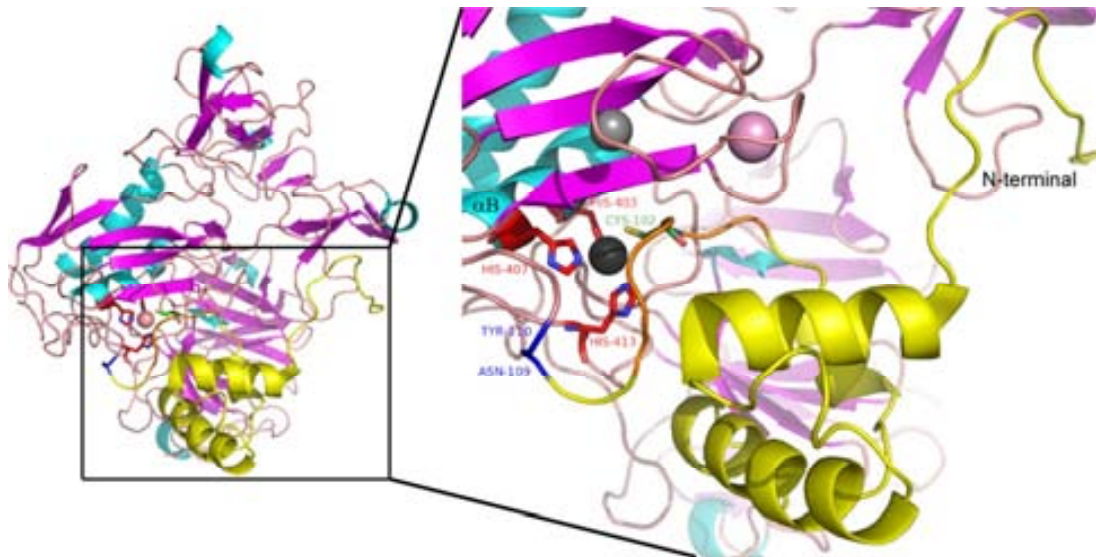
La familia de las MMPs ha sido ampliamente estudiada desde que su primer miembro fuese identificado hace ya 50 años en células de anfibio (Gross and Lapiere 1962). Desde entonces numerosas estructuras de miembros de esta familia han sido resueltas y publicadas. No obstante, únicamente cuatro de estas estructuras, todas ellas pertenecientes a proteínas humanas, corresponden a zimógenos (colagenasa-1 [proMMP-1] (Jozic et al. 2005), progelatinasa A [proMMP-2] (Morgunova et al. 1999), proestromelisin-1 [proMMP-3] (Becker et al. 1995) y progelatinasa B [proMMP-9] (Elkins et al. 2002)). Estas revelan que la inhibición se produce por un bloqueo del acceso al centro activo vía prodominios globulares y que el dominio catalítico se encuentra ya preformado en el zimógeno. Por otro lado, se ha demostrado que contrariamente a otros zimógenos previamente estudiados, los prodominios de las MMPs no son imprescindibles para el correcto replegamiento *in-vitro* del dominio catalítico (Wetmore and Hardman 1996).

El prodominio de MMPs presenta un residuo de cisteína dentro de un motivo conservado (PRCGVPD), cuya cadena lateral substituye a la molécula de agua catalítica en la esfera de coordinación del zinc. La eliminación del prodominio permite el paso de la forma inactiva a la activa mediante el mecanismo conocido como “*cysteine-switch*” o “*velcro*”. En otras familias de metzincinas, cuyas estructuras zimogénicas no ha sido resuelta aún, como es el caso de adamalisinias y leishmanolisinas, se ha postulado la existencia de mecanismos de

latencia similares. Adamalisin y lesihmanolisinas presentan motivos conservados (PKMCGV y HRCIHD, respectivamente) que contienen un residuo de cisteína que podría estar implicado en este mecanismo (Gomis-Ruth 2003). En el caso de MMP-2, el prodominio, formado por 79 residuos, está estabilizado por un puente disulfuro que no está presente en otras MMPs (Figura I.10). Presenta forma ovoide e interacciona con el centro activo catalítico que lo rodea. El extremo N-terminal del prodominio, que comienza con el residuo 30 (los residuos 1-29 corresponden al péptido señal), se halla en la parte superior del tercer dominio de unión a fibronectina tipo II. Discurre en dos segmentos de conformación extendida hasta el dominio globular mantenido por un extenso núcleo hidrofóbico constituido por las cadenas laterales de dos hélices  $\alpha$ . La primera empieza en la posición 46 y discurre de manera antiparalela hacia la hendidura del centro activo. Un conector da lugar a la segunda hélice que se orienta con su eje perpendicular a la primera. Un segundo conector enlaza con la tercera hélice, que también discurre perpendicular a la que le precede. El final de la última hélice, constituido por una prolina (P100), se sitúa en el extremo de la región de interacción con el sustrato. Esta prolina introduce un giro brusco en la cadena polipeptídica induciendo su penetración en la hendidura del centro activo en configuración extendida, de manera que discurre a través de la zona de anclaje del sustrato en dirección opuesta a la de sustratos e inhibidores peptidomiméticos. La cisteína del *cysteine-switch* (C102), localizada dos residuos después de la mencionada prolina, interacciona con el zinc catalítico mediante su átomo S $\gamma$ . Este átomo substituye a la molécula de solvente catalítica y juntamente con tres átomos N $\epsilon$ 2 de histidina (H403, H407, H413) da lugar a una coordinación tetraédrica casi perfecta del metal. Dos residuos después, una asparagina (N104) se sitúa ocupando parcialmente el bolsillo **S1** y uniendo los átomos de la cadena principal de A194. Finalmente, tras una prolina, un residuo de ácido aspártico (D106) estabiliza la conformación mediante un doble puente de hidrógeno con una arginina localizada cinco residuos por delante en la cadena polipeptídica (Morgunova et al. 1999).

La activación de proMMP-2 se consigue mediante el procesado proteolítico del enlace N109-Y110 situado en la región cebo (*bait region*) de la superficie del prodominio. Han sido descritos otros sitios de corte secundarios, de manera que tras un primer corte en la región cebo se producirían otros cortes posteriores, dando lugar a un proceso de activación secuencial. Una vez producidos los cortes, los siete primeros residuos de la enzima madura experimentan un desplazamiento de aproximadamente 20 Å, situando al nuevo extremo N-terminal a una distancia de enlace del primer residuo de aspartato conservado del dominio catalítico (D436) situado al principio de la hélice C-terminal  $\alpha$ C.





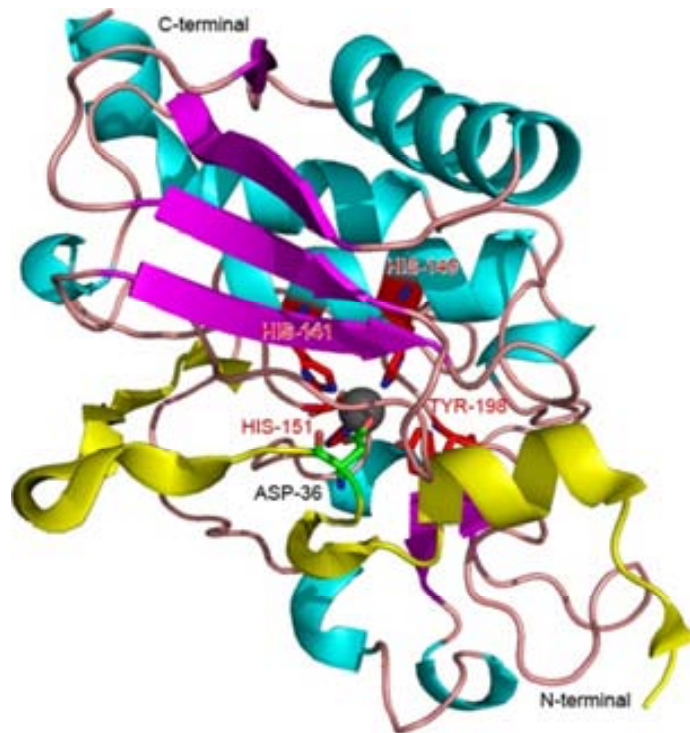
**Figura I.10.** Estructura de prometalopeptidasa de matriz 2 (proMMP-2). Figura preparada con el programa PyMOL (Schrodinger 2010). Código *Protein Data Bank* (PDB) 1CK7 (Morgunova et al. 1999). Se indica el ion zinc catalítico (gris oscuro), ion zinc estructural (gris claro), iones calcio (rosa), aminoácidos de unión a zinc (rojo), prodominio (amarillo), motivo conservado del prodominio (naranja), región cebo (azul) y cisteína implicada en el *cysteine-switch* (verde).

Algunas MMPs, en particular las MMPs ancladas a membrana (MT-MMPs), presentan hasta once residuos antes del inicio del dominio catalítico. Esta región finaliza en una secuencia consenso (RX[K/R]R) que permite su reconocimiento y activación intracelular por parte de las serín peptidasas asociadas al aparato de Golgi de la familia subtilisina tipo Kex-2/furina (Yana and Weiss 2000). Por tanto, estas enzimas serían secretadas en su forma activa. Otra forma de activar MMPs es mediante agentes caotrópicos o reactivos con sulfhidrido, los cuales permiten la salida de la cisteína bloqueante y el acceso del zinc al solvente, favoreciendo la unión del metal a la molécula de agua catalítica y la consiguiente activación de la proteína (Murphy et al. 1999; Nagase 1997).

Por otro lado, las estructuras de proastacina (Guevara et al. 2010) y profragilisina (Goulas et al. 2011) muestran un mecanismo diferente para el mantenimiento del zimógeno basado en la coordinación del ion zinc catalítico por parte de un residuo de aspartato (*aspartate-switch*), en lugar de la cisteína presente en MMPs, adamalinas y leshmanolisinas.

Proastacina, aislada del cangrejo *Astacus astacus*, presenta el prodominio más corto de los zimógenos de metalopeptidasas descritos hasta la fecha. Consta de únicamente 34 residuos, y de igual forma que en MMPs, se ha demostrado que no es imprescindible para el repliegamiento de la enzima *in-vitro* (Reyda et al. 1999). El prodominio discurre a través de la superficie del dominio catalítico estabilizado por múltiples interacciones intramoleculares, bloqueando el acceso a la hendidura del centro activo (Figura I.11). Su

interacción con el dominio catalítico cubre  $1.580 \text{ \AA}^2$  de la interfase de la proteína y está basada en 75 contactos próximos ( $< 4 \text{ \AA}$ ), de los cuales 20 son puentes de hidrógeno, dos son interacciones organometálicas y el resto son contactos hidrofóbicos entre 27 pares de residuos. El extremo N-terminal (S16) se ancla al dominio catalítico mediante un puente de hidrógeno mediado por solvente con I205, residuo presente en el segmento que conecta el giro-Met con la hélice  $\alpha$  C-terminal. El prodominio se extiende a lo largo de la superficie del dominio catalítico hacia la hendidura del centro activo (Guevara et al. 2010).



**Figura I.11.** Estructura de proastacina. Figura preparada con el programa PyMOL (Schrodinger 2010). Código *Protein Data Bank* (PDB) 3LQ0 (Guevara et al. 2010). Se indica el ion zinc catalítico (gris), aminoácidos de unión a zinc (rojo), prodominio (amarillo) y el aspartato implicado en el *aspartate-switch* (verde).

Desde E21 hasta R40 la cadena discurre a través de la hendidura en sentido inverso a la orientación de un sustrato adoptando una conformación de hélice  $\alpha$  desde el inicio hasta Y27 (hélice  $\alpha_1$ ). L26 se dispone ocupando parcialmente el bolsillo  $S'_1$ . La cadena lateral de D36 coordina de manera bidentada al ion zinc catalítico, que juntamente con los átomos  $N\epsilon_2$  de H141, H145 y H151, y el  $O\eta$  de Y198 crea una esfera de coordinación octaédrica muy inusual para el zinc (Guevara et al. 2010).

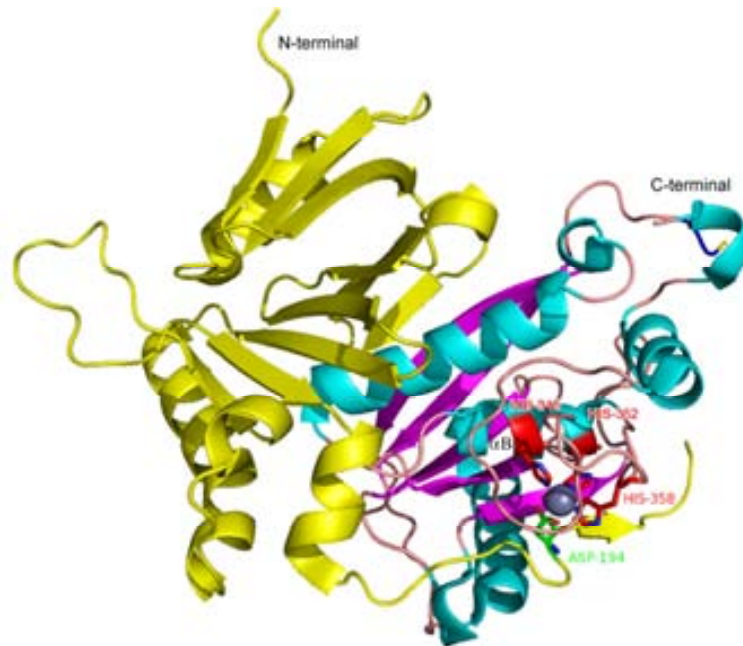
La conformación en hélice  $\alpha$  podría contribuir a la prevención de la autólisis, dado que los sustratos suelen interactuar en conformación extendida con el centro activo. Por otro lado, la orientación inversa del prodominio en la hendidura del centro activo, igual que ocurre en proMMP y profragilisina, también sería un mecanismo de protección contra la autólisis.

Estudios bioquímicos han demostrado que la activación de proastacina se produce por un mecanismo de dos pasos (Yiallourous et al. 2002). El primer paso es un proceso heterolítico en el cual tripsina realiza una serie de cortes en la superficie expuesta del prodominio, facilitados por la elevada flexibilidad del fragmento entre A41 y Q44. Estos cortes producen la exposición de R47-V48, último lugar de corte de tripsina, induciendo cambios conformacionales importantes en el segmento comprendido entre D178 y G187. Este segmento, análogo al presente en múltiples serín peptidasas, es conocido como dominio de activación. Su estructura desordenada en zimógenos bloquea parcialmente la unión del sustrato con el centro activo. Una vez se produce el corte entre R47 y V48 el dominio adquiere una estructura rígida permitiendo la interacción y dando lugar a formas premaduras catalíticamente activas. Esta reordenación da lugar al segundo paso de activación, un proceso autolítico durante el cual se produce el corte entre G49 y A50. El corte genera un reordenamiento de los seis primeros residuos de la forma madura (A50-D56) clave para el mantenimiento de la integridad estructural de la peptidasas (Guevara et al. 2010).

En profragilisina, el prodominio, compuesto por aproximadamente 170 residuos, muestra una gran complementabilidad con el dominio catalítico, cubriendo  $1.996 \text{ \AA}^2$  en la interfase de la proteína. En la interacción intervienen 63 residuos con un total de 98 contactos próximos. A diferencia de en MMPs, el prodominio no interacciona por el centro del dominio catalítico, sino que lo hace por la parte lateral derecha del mismo (Figura I.12). Previene el acceso a la hendidura del centro catalítico mediante su segmento C-terminal, el cual discurre en conformación extendida a lo largo de este en orientación opuesta al sustrato.

El segmento más importante para el mantenimiento de la latencia incluye la hélice  $\alpha_3$ , la hebra  $\beta_{11}$  y el conector presente entre ambas. Este atraviesa frontalmente el centro catalítico de derecha a izquierda, estableciendo una interacción de hebras  $\beta$  paralelas en la región menor de la hendidura catalítica entre la hebra  $\beta_{11}$  del prodominio y la  $\beta_{15}$  del dominio catalítico. Una protuberancia prominente procedente de  $\beta_{11}$  en la región comprendida entre Y191 y D194 da origen a un giro  $\beta$  de tipo 1,4, que produce la penetración de sus cadenas laterales en la hendidura del centro activo. La cadena lateral de Y191 ocupa el bolsillo **S1** del centro activo, mientras que D194 coordina el átomo de zinc catalítico en substitución de la molécula de solvente presente en la forma madura. Los dos oxígenos terminales del grupo carboxilato de D194 crean una unión bidentada con el zinc, que juntamente con los átomos  $N\epsilon_2$  de H348, H352 y H358, dan lugar a una esfera de coordinación tetraédrica +1. Además, el átomo  $O\delta_2$  de D194 se sitúa orientado hacia el átomo  $O\epsilon_1$  del ácido/base general (E349) (Goulas et al. 2011).

La maduración de esta peptidasa *in-vivo* ocurre mediante un proceso de activación heterolítico. Uno de los principales candidatos para llevar a cabo esta activación es tripsina, dado que profragilisina es producida por *Bacteroides fragilis*, una bacteria patógena del aparato digestivo, y tripsina es una peptidasa presente en el duodeno de forma abundante. Además, el sitio de corte final para la activación de profragilisina, R211-A212, encaja con la especificidad de tripsina y ensayos de proteólisis limitada con ella han demostrado la obtención de la forma madura de fragilisina (Goulas et al. 2011).



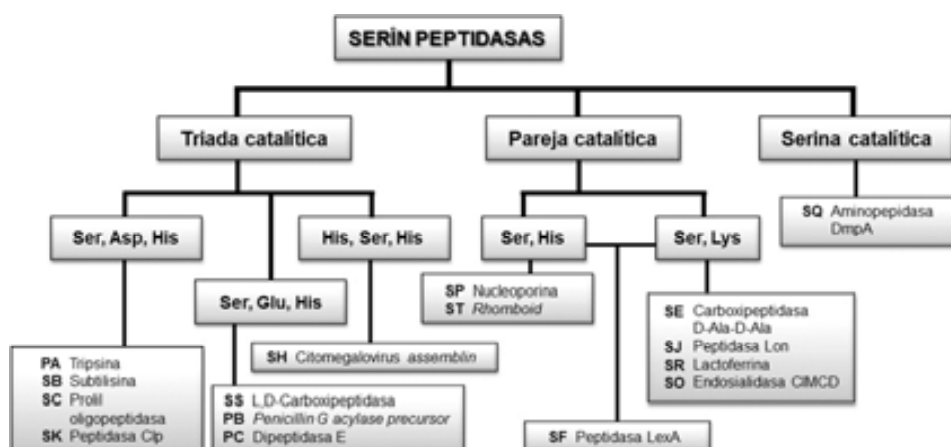
**Figura I.12.** Estructura de profragilisina. Figura preparada con el programa PyMOL (Schrodinger 2010). Código *Protein Data Bank* (PDB) 3P24 (Goulas et al. 2011). Se indica el ion zinc catalítico (gris), aminoácidos de unión a zinc (rojo), prodominio (amarillo), el sitio de corte final (azul) y el aspartato implicado en el *aspartate-switch* (verde).

### 1.2.3. SERÍN PEPTIDASAS

Las serín peptidasas son enzimas que poseen en su centro activo un aminoácido de serina nucleófilo esencial para su actividad hidrolítica. Este actúa atacando al grupo carbonilo del enlace peptídico a escindir del substrato para formar un intermediario covalente acil-enzima. La nucleofilia de la serina catalítica depende típicamente de la triada catalítica, constituida por un ácido aspártico, una histidina y la propia serina catalítica (Hedstrom 2002). No obstante, algunas peptidasas presentan diferentes residuos en la triada, pudiéndose producir la substitución del ácido aspártico por ácido glutámico o por una segunda histidina. Además, otras serín peptidasas emplean únicamente una pareja catalítica, donde una serina y una histidina o lisina son suficientes para el desarrollo de la actividad catalítica.

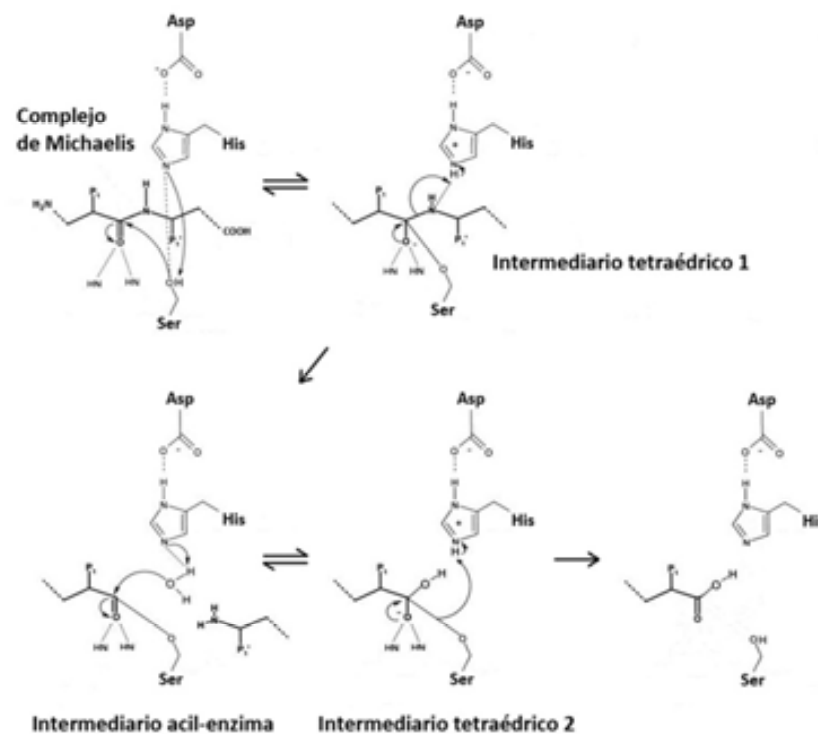
Estas enzimas están involucradas en multitud de procesos fisiológicos actuando como peptidasas reguladoras mediante la activación de precursores proteínicos, además de como enzimas digestivas. Algunos de los procesos en los que intervienen son la activación del tripsinógeno, fibrinólisis y cascada de coagulación de la sangre, diferenciación celular y desarrollo (Perona and Craik 1995), apoptosis, sustitución de proteínas intracelulares (Page and Di Cera 2008), exportación extracelular de proteínas (Dalbey et al. 1997), respuesta inmunitaria y reproducción (Hedstrom 2002). De igual forma que en otros tipos de peptidasas, su función anormal es responsable de numerosas patologías, teniendo especial relevancia los procesos relacionados con la coagulación sanguínea (Meeks and Abshire 2008; Soff 2012).

La base de datos MEROPS clasifica las serín peptidasas según las similitudes en su secuencia y estructura. Tradicionalmente, se habían clasificado en dos grandes grupos: el clan quimotripsina (PA), caracterizado por presentar un plegamiento en doble barril  $\beta$ , y el clan subtilisina (SB), con una estructura en hebras  $\beta$  paralelas flanqueadas por hélices  $\alpha$  (Polgar 2005). No obstante, a pesar de las diferencias estructurales, el mecanismo de acción de ambos clanes basado en la triada catalítica clásica es esencialmente el mismo, por lo que se consideran el resultado de una evolución convergente. Posteriormente se identificaron nuevos clanes con diferencias en la triada, con sus consiguientes modificaciones en el mecanismo de acción. Actualmente existen 16 clanes de serín peptidasas en base a su estructura y mecanismo catalítico, designados con un código de dos letras, la primera de las cuales es una "P" o una "S" (Figura I.13). Por otro lado, la base de datos EC las clasifica en dos grandes grupos según el lugar de corte en la cadena polipeptídica: serín carboxipeptidasas (EC 3.4.16) y serín endopeptidasas (EC 3.4.21).



**Figura I.13.** Clasificación de serín peptidasas. Clasificación según los residuos implicados en su mecanismo catalítico. Se indican los clanes (negrita) según la base de datos MEROPS y un ejemplo de serín peptidasa para cada uno de ellos (Page and Di Cera 2008).

El mecanismo de acción de las serín peptidasas se basa en la formación de un complejo de Michaelis entre enzima y sustrato y el posterior ataque nucleofílico del grupo hidroxilo de la serina catalítica sobre el átomo de carbono del carbonilo del enlace peptídico a escindir del sustrato (Figura I.14). Este ataque, en el que el grupo imidazol de la histidina actúa como base general (en algunos clanes es el grupo amino de la cadena lateral de una lisina), comporta la creación de un intermediario tetraédrico y un ion imidazolio (reacción de adición). El intermediario tetraédrico se rompe por catálisis ácida en un intermediario covalente acil-enzima, una base imidazol y un producto amino (reacción de eliminación). Durante el paso de acilación, el ion imidazolio transfiere el protón captado del grupo hidroxilo de la serina al grupo amino del grupo saliente. Posteriormente, el grupo acil-enzima es desacetilado mediante una segunda reacción de adición-eliminación con los mismos pasos catalíticos que la primera. Esta se produce por el ataque nucleofílico por parte de una molécula de agua situada en el espacio dejado por el grupo saliente. La segunda adición da lugar a la formación de un segundo intermedio tetraédrico y nuevamente a un ion imidazolio, para finalmente regenerar la enzima por eliminación, una base imidazol y un producto ácido. La cadena lateral del residuo de ácido aspártico está implicada en la estabilización del par iónico generado entre el ion imidazolio y el intermediario tetraédrico cargado, así como en la correcta orientación del tautómero de histidina respecto a la serina catalítica.



**Figura I.14.** Mecanismo de acción sugerido para serín peptidasas con triada catalítica clásica Ser/His/Asp. Adaptado de (Ekici et al. 2008).

Estudios de resonancia magnética nuclear (Sakiyama and Kawata 1983) y difracción de neutrones (Kossiakoff and Spencer 1980) han confirmado que es el residuo de histidina y no el de ácido aspártico el que se protona durante la reacción, tal y como se deducía de las estructuras previas obtenidas por difracción de rayos X (Birktoft and Blow 1972; Fehlhammer et al. 1977). Este último residuo está substituido por ácido glutámico o histidina en algunos clanes de serín peptidasas, de manera que estos estabilizarían a la histidina que actúa como ácido/base general. En otros clanes no se requiere la acción de este residuo (Polgar 2005).

Por otro lado, estudios cristalográficos y funcionales han demostrado la existencia de un sitio de unión que contribuye a la estabilización de los intermedios de reacción sin afectar a la unión entre la peptidasa y el sustrato. Tras las reacciones de adición, el oxígeno del grupo carbonilo del enlace hidrolizable genera un oxoanión centrado en un carbono tetrahédrico. Este se estabiliza en el sitio de unión a oxoanión mediante la interacción con dos grupos amino pertenecientes a la cadena principal de la enzima (Polgar 2005).

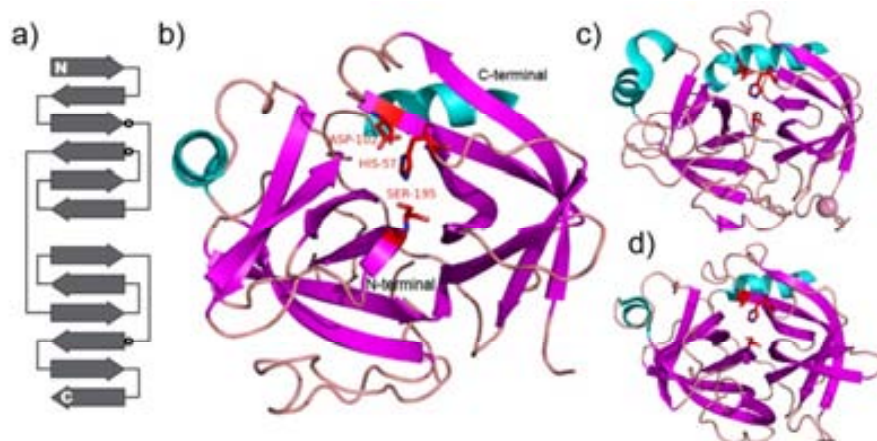
Las serín peptidasas son inhibidas por diferentes grupos de inhibidores específicos. Estos incluyen inhibidores sintéticos de molécula pequeña e inhibidores proteínicos naturales. Estos últimos serán descritos en detalle en el apartado 1.3.3. Respecto a los inhibidores de molécula pequeña, algunos de los más representativos son el fluoruro de fenilmetilsulfóxido (PMSF) y el clorhidrato del fluoruro de 4-(2-aminoetil)-fenil sulfonilo (AEBSF; Pefabloc). Ambos reaccionan con el grupo hidroxilo de la serina del centro activo, de manera que tras la reacción de eliminación se genera una forma anhidra de la peptidasa en un proceso irreversible (Gold and Fahrney 1964). Existen también inhibidores generales con aplicaciones terapéuticas como el gabexato mesilato, así como inhibidores específicos para trombina (argatroban, napsagatran, inogatran, CVS-1123, LY178550 y melagatran), factor Xa (DX-9065a, YM-60828, ZK-807191 y FX-2212), elastasa (MR889, L-658.758, ONO-5046, WIN63759, DMP-777 y CE-1037), tripsina (BABIM), convertasas del complemento (sepimostato mesilato y nafamostato mesilato) y peptidasas C-NS3 (RD4-6204, RD4-6205) (Leung et al. 2000).

### **1.2.3.1. Clan PA**

El clan PA es el clan de peptidasas más grande y el mejor estudiado. Contiene 21 familias (según la base de datos MEROPS) que muestran un origen común con quimotripsina, una de las primeras serín peptidasas caracterizadas estructuralmente (Matthews et al. 1967).

Para 11 de ellas existe al menos un miembro representativo cuya estructura ha sido resuelta. Cabe destacar que el clan PA no es estrictamente un clan de serín peptidasas, ya que 7 de las 21 familias pertenecen al grupo de las cisteín peptidasas. No obstante, todos los miembros del clan presentan estructuras similares, son endopeptidasas y la mayoría de ellos contienen una triada catalítica. La diferencia entre las serín peptidasas y las cisteín peptidasas del clan PA radica en el aminoácido responsable del ataque nucleofílico, de manera que se dividen en dos subclanes: PA(S) y PA(C). La triada catalítica del primero está constituida por histidina, ácido aspártico y serina, mientras que la del segundo está formada por histidina, ácido aspártico y cisteína. En ambos casos el orden de los residuos catalíticos en la secuencia es siempre el mismo (H, D, S/C). Todos los miembros del subclan PA(C) pertenecen a virus de ARN, por lo que se postula que estos adquirieron los genes de estas peptidasas a partir de sus huéspedes por transferencia génica horizontal. En la mayoría de virus persisten varias familias de serín peptidasas, pero en algunos de ellos se habría producido el reemplazamiento de la serina catalítica por cisteína con retención de la actividad y sin comportar grandes diferencias en su mecanismo de acción (Rawlings et al. 2012).

Quimotripsina (familia S1) es una enzima pancreática digestiva presente únicamente en mamíferos, aunque existen peptidasas tipo quimotripsina en otros organismos. Es sintetizada en forma de zimógeno, quimotripsinógeno, y activada por tripsina y por la propia quimotripsina (Appel 1986). La forma activa resultante está compuesta por tres cadenas polipeptídicas unidas mediante puentes disulfuro. Muestra una estructura compuesta por dos dominios en barril  $\beta$  probablemente formados por duplicación génica (Figura I.15).



**Figura I.15.** Clan PA de serín peptidasas. **a)** Topología de las hebras  $\beta$  característica del clan PA. Los círculos indican la posición de los residuos catalíticos. Estructura de **b)** quimotripsinógeno (código *Protein Data Bank* [PDB] 1CHG (Freer et al. 1970)), **c)** tripsinógeno (código PDB 1TGB (Fehlhammer et al. 1977)) y **d)** elastasa (código PDB 3EST (Meyer et al. 1988)). Se representan las hebras  $\beta$  (magenta), las hélices  $\alpha$  (cian) y la triada catalítica (rojo). Figuras **b**, **c** y **d** preparadas con el programa PyMOL (Schrodinger 2010).



Entre los dos dominios de quimotripsina se encuentra el centro activo. El barril  $\beta$  N-terminal contiene los residuos catalíticos histidina y ácido aspártico, mientras que la serina catalítica se encuentra en el C-terminal. A pesar de que se describe como una proteína todo  $\beta$ , en el extremo C-terminal de cada uno de los dominios presenta una hélice  $\alpha$  que estabiliza la interacción entre ambos.

Los miembros del clan PA, pese a tener una estructura muy similar, pueden presentar diferente especificidad de sustrato. Quimotripsina cataliza la hidrólisis del enlace peptídico en sustratos con tirosina, fenilalanina, leucina o triptófano en posición **P1** en este orden de preferencia. Los grupos aromáticos y/o hidrofóbicos de estos aminoácidos encajan en el bolsillo hidrofóbico de la posición **S1** de la enzima (Appel 1986). En cambio, en tripsina la posición **S1** es ácida, de manera que esta enzima tiene una elevada especificidad hacia sustratos con lisina o arginina en posición **P1** (Walter et al. 1982).

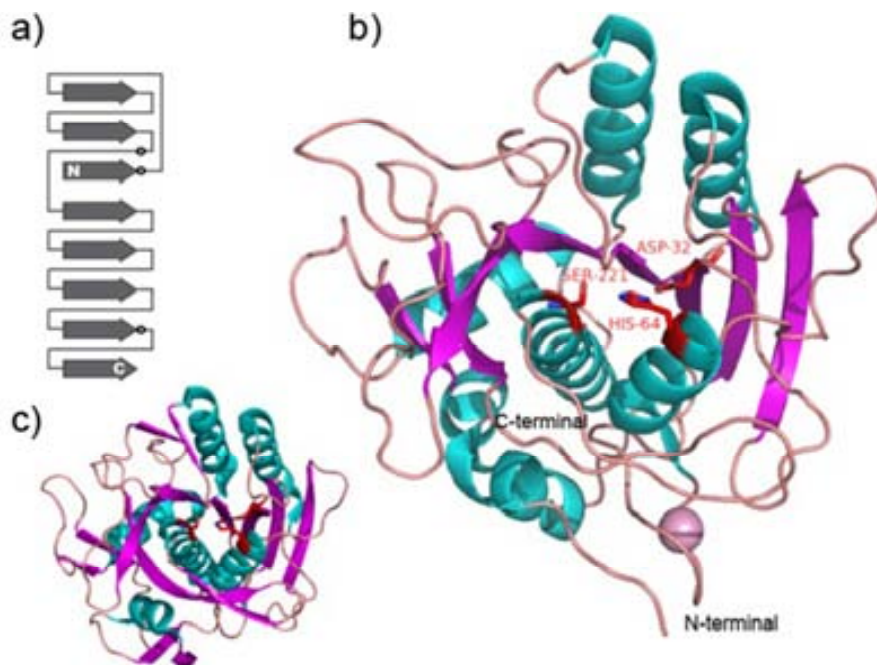
### **1.2.3.2. Clan SB**

Las endopeptidasas y exopeptidasas del clan SB se encuentran ampliamente distribuidas en bacterias y plantas, aunque también están presentes en otros organismos. Contiene únicamente dos familias, S8 y S53, las cuales presentan mecanismos catalíticos diferentes pero un plegamiento común. Se trata de proteínas  $\alpha/\beta$  constituidas por siete-ocho hebras  $\beta$  asociadas en una hoja  $\beta$  paralela en orden  $\beta 2$ - $\beta 3$ - $\beta 1$ - $\beta 4$ - $\beta 5$ - $\beta 6$ - $\beta 7$ - $\beta 8$  y flanqueadas en ambas caras por diversas hélices  $\alpha$ . Debido a las notables diferencias estructurales del clan SB respecto al clan PA, cabe suponer que ambos han evolucionado de manera independiente (Hartley 1970).

El arquetipo de este clan es subtilisina (familia S8), descubierta originalmente en el medio extracelular de la bacteria Gram positiva *Bacillus subtilis* (Wright et al. 1969) y objeto de estudio del presente trabajo. De la misma manera que quimotripsina, subtilisina fue una de las primeras proteínas cuya estructura fue determinada cristalográficamente (Alden et al. 1971). Presenta una triada catalítica en orden ácido aspártico, histidina y serina. No obstante, pese al diferente orden de la triada respecto al existente en el clan PA y a presentar estructuras tridimensionales completamente diferentes, se observa una organización geométrica idéntica de los residuos catalíticos (Figura I.16).

Respecto a su actividad catalítica, se trata de una peptidasa no específica. Está implicada en el procesamiento de sustratos peptídicos para su posterior asimilación por parte de las bacterias que la secretan (Wells et al. 1983). Además, su expresión ha sido asociada a procesos de esporulación, de manera que podría asistir la autólisis de la bacteria para la liberación de las esporas (Balaban et al. 2002; Wells et al. 1983). Su expresión también ha sido asociada a la competencia entre poblaciones microbianas (Taguchi 1995). Por otro lado, peptidasas tipo subtilisina han sido descritas como factores de virulencia en microorganismos patógenos (Monod et al. 2002).

Se han descrito diversas aplicaciones biotecnológicas para subtilisina. Esta peptidasa se utiliza frecuentemente en el procesamiento de alimentos (Wong 1995) así como en detergentes para la ropa, lavavajillas, limpiadores industriales y productos de cuidado personal (Polaina and MacCabe 2007; van Ee and Misset 1997). Debido a su elevada estabilidad, su baja especificidad y su capacidad para degradar completamente sus sustratos, subtilisina es útil para la eliminación de manchas y contaminantes proteínicos. Existen diversas patentes sobre sus usos así como sobre mutaciones para el incremento de su actividad proteolítica y estabilidad en las condiciones de utilización (Aaslyng et al. 2000; Bryan et al. 1993; Graycar et al. 1994; Hastrup et al. 1989; Zurowski and Stabinsky 1988).



**Figura I.16.** Clan SB de serín peptidasas. **a)** Topología de las hebras  $\beta$  característica del clan SB. Los círculos negros indican la posición de los residuos catalíticos. Estructura de **b)** subtilisina (código *Protein Data Bank* [PDB] 1SBC (Neidhart and Petsko 1988)) y **c)** proteinasa K (código PDB 2PRK (Betz et al. 1988)). Se representan las hebras  $\beta$  (magenta), las hélices  $\alpha$  (cian) y la triada catalítica (rojo). Figuras **b** y **c** preparadas con el programa PyMOL (Schrodinger 2010).

Por su parte, la familia S53 contiene una tétrada catalítica en orden ácido glutámico (base general), ácido aspártico, ácido aspártico (sitio de unión al oxoanión) y serina. Únicamente la serina de las familias S53 y S8, la histidina de la familia S8 y el ácido glutámico de la S53 se encuentran en posiciones equivalentes. En las dos familias las serinas catalíticas están situadas en un motivo G-T-S-Xa-Xa-Xb-P, donde Xa es un aminoácido alifático (A, G, I, L o V) y Xb un aminoácido pequeño (A, G, S o T). Los miembros más representativos de esta familia son las sedolisinas, unas peptidasas de amplio espectro con preferencia para sustratos con aminoácidos hidrofóbicos en posición **P1** y **P2**. A diferencia de subtilisina, cuya actividad óptima se encuentra a pH 9 (Peng et al. 2003), estas suelen presentar una actividad óptima a pH ácido, en gran parte debido a la presencia del residuo catalítico de ácido aspártico adicional (Wlodawer et al. 2003).

### **1.3. INHIBIDORES PROTEÍNICOS DE PEPTIDASAS**

---

Debido al carácter esencialmente irreversible de la hidrólisis del enlace peptídico, las peptidasas están sometidas a una estricta regulación con la finalidad de prevenir una actividad dañina para el organismo en el espacio y en el tiempo (Barrett et al. 1998). Uno de los mecanismos de regulación consiste en la colocalización de las enzimas con inhibidores proteínicos, la mayoría de los cuales actúan sobre una o varias peptidasas de una clase determinada a través de una única región inhibitoria bloqueando la hendidura del centro activo de la enzima (Bode and Huber 1991, 1992).

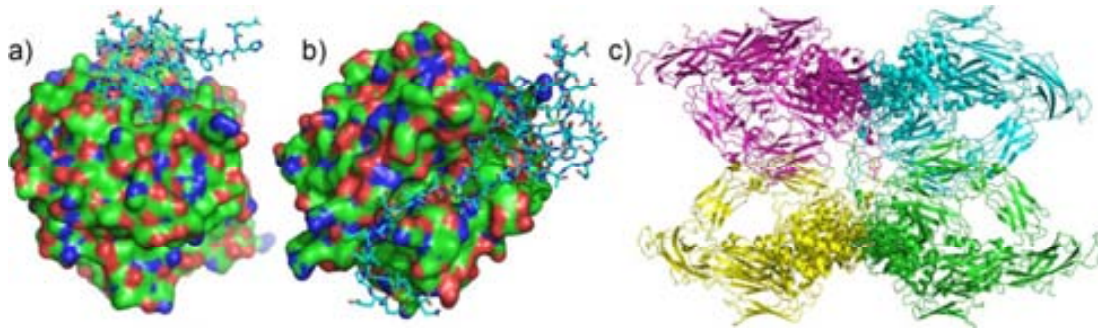
Los inhibidores endógenos actúan sobre peptidasas del mismo organismo que produce el inhibidor, mientras que los exógenos lo hacen sobre peptidasas de otros organismos. Algunos microorganismos patogénicos pueden sintetizar inhibidores de molécula pequeña no proteínicos que actúan sobre las peptidasas del huésped (Fallon et al. 2011; Sin et al. 1997). En cambio, en organismos superiores los inhibidores endógenos son siempre proteínicos (Bode and Huber 1992).

El número de inhibidores proteínicos de peptidasas aislados e identificados hasta el momento es muy elevado y existen numerosas estructuras cristalográficas disponibles, tanto de inhibidores aislados como en complejo con sus peptidasas diana. La base de datos MEROPS propone su clasificación en 74 familias y 38 clanes de una forma análoga a la clasificación de las peptidasas. Las familias de inhibidores, identificadas por una "I" seguida de dos números, están formadas por una serie de inhibidores de peptidasas homólogos. Su

homología se determina en base a su secuencia aminoacídica, ya sea de la proteína completa o únicamente de su región inhibitoria. Por tanto, una familia puede contener un único inhibidor si no se conocen otros homólogos, de la misma manera que un inhibidor que contenga diferentes dominios puede ser asignado a diferentes familias. Estas se agrupan en clanes, si bien existen numerosas familias que aún no han podido ser asignadas a ninguno. Actualmente existen 38 clanes, cada uno de los cuales contiene todos los inhibidores surgidos a partir de un único origen evolutivo. Por tanto, cada clan agrupa inhibidores de una o más familias que a través de las similitudes en sus estructuras terciarias muestran una relación evolutiva. Tienen un identificador formado por dos letras, la primera de las cuales puede ser una "I" o una "J".

Por otro lado, y de forma complementaria, los inhibidores proteínicos también pueden clasificarse según la clase de peptidasa sobre la que actúan. De esta manera encontramos inhibidores de serín peptidasas, metalopeptidasas, cisteín peptidasas y aspartil peptidasas. Además, los inhibidores proteínicos de peptidasas pueden agruparse según el mecanismo por el cual desarrollan su acción (Figura 1.17). Pueden diferenciarse dos grandes grupos según si interaccionan directamente con el centro catalítico de la peptidasa o con regiones adyacentes (Bode and Huber 1992). En los primeros, la unión puede producirse mimetizando la interacción del producto o del sustrato, de manera que el inhibidor utiliza los sitios de reconocimiento de la enzima para su especificidad. En caso de uniones tipo producto las interacciones entre la peptidasa y el inhibidor son suficientemente fuertes para evitar una disociación rápida. En la unión tipo sustrato las interacciones intramoleculares e intermoleculares de la región inhibitoria con el núcleo del inhibidor (frecuentemente a través de puentes disulfuro) y con el sitio de unión de la enzima producen su estabilización mutua, y son tan estrechas que la hidrólisis ocurre muy lentamente (mecanismo estándar de inhibición; apartado 1.3.1). Por otro lado, los inhibidores que no interaccionan directamente con el centro activo pueden hacerlo a través del sitio de unión a sustrato u otras zonas. En los casos en que existen regiones específicas en la superficie de la peptidasa involucradas en la interacción la inhibición puede llegar a ser muy selectiva.

Otro tipo de inhibidores según su mecanismo de acción son los tipo trampa molecular (*molecular trap*), como es el caso de  $\alpha_2$ -macroglobulina (Marrero et al. 2012). Estos pueden inhibir cualquier tipo de endopeptidasa sin importar su mecanismo catalítico. Poseen una región cebo, la cual puede ser hidrolizada por múltiples peptidasas provocando un cambio estructural en el inhibidor de manera que la enzima queda atrapada en su interior.



**Figura I.17.** Tipos de inhibición de inhibidores proteínicos de peptidasas. Figuras preparadas con el programa PyMOL (Schrodinger 2010). **a)** Interacción directa con el centro activo. Estructura del complejo carboxipeptidasa A (representación de superficie) con el inhibidor de carboxipeptidasas de patata (representación en *sticks*), código *Protein Data Bank* (PDB) 4CPA (Rees and Lipscomb 1982). **b)** Interacción con las regiones adyacentes al centro activo. Estructura del complejo trombina (representación de superficie) con hirudina (representación en *sticks*), código PDB 1HRT (Vitali et al. 1992). En **a** y **b** se representan los átomos de carbono en cian (inhibidores) y verde (enzimas), nitrógeno en azul, oxígeno en rojo y azufre en amarillo, **c)** Trampa molecular. Estructura del complejo homotetramérico de  $\alpha_2$ -macroglobulina, código PDB 4ACQ (Marrero et al. 2012). Cada subunidad se representa en un color diferente.

Los inhibidores proteínicos también pueden clasificarse a partir de características estructurales y funcionales específicas. En base a estas podemos definir los inhibidores de doble faz o tipo *Janus-faced* (Tabla I.4). Se trata de un tipo especial de inhibidores proteínicos, poco frecuentes entre la plétora de los inhibidores descritos, caracterizado por presentar dos regiones inhibitorias en un solo dominio de proteína mimetizando la arquitectura de la cabeza del dios romano Jano<sup>2</sup>. Estos inhibidores son capaces de inhibir dos moléculas de enzima simultáneamente (Bao et al. 2009).

Los inhibidores de doble faz muestran en su mayoría un origen común con inhibidores con un único sitio reactivo, de manera que habrían evolucionado a partir de estos adquiriendo un nuevo sitio reactivo. Debido a su proceso evolutivo es frecuente que sus sitios reactivos actúen sobre diferentes tipos de peptidasas diana, aunque también pueden actuar sobre peptidasas del mismo tipo (Figura I.18). Es razonable asumir que la obtención de un segundo sitio reactivo ocurre para la inhibición de una peptidasa presente en su mismo ambiente natural y que por tanto colocalize con la diana inicial. En consecuencia, es frecuente que este segundo sitio reactivo actúe sobre peptidasas funcionalmente relacionadas con las del primero, la mayoría de veces pertenecientes a la misma clase. Un ejemplo de este tipo de inhibidores son los inhibidores de doble faz de la familia Bowman-Birk (familia MEROPS I12), muchos de los cuales actúan sobre tripsina y quimotripsina,

<sup>2</sup> Jano (en latín Janus, Ianus) es, en la mitología romana, un dios representado con dos caras mirando hacia ambos lados de su perfil. Jano era el dios de las puertas, los comienzos y los finales, al que se le atribuye, entre otras cosas, la invención del dinero, las leyes y la agricultura.

enzimas con funciones digestivas en el procesado de nutrientes proteínicos. Actualmente existen unas pocas estructuras cristalográficas de inhibidores de doble faz, pero hasta la fecha no se ha determinado la estructura de ninguno de ellos en complejo con peptidasas diferentes a través de sus dos sitios reactivos.

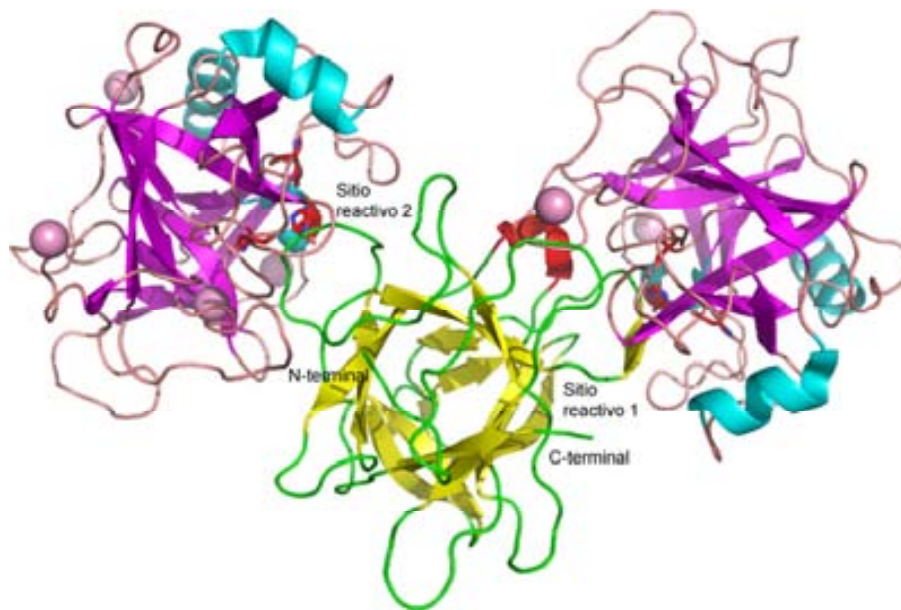
**Tabla 1.4.** Inhibidores monodominio de doble faz.

Familia (MEROPS)	Nombre	Organismo	Código UniProt	Clase de peptidasa	Peptidasa diana	Código PDB	Referencias
Kunitz (13)	ApKTI	<i>Adenanthera pavonina</i>	P09941	SP <sup>1</sup> /CP <sup>2</sup>	T <sup>3</sup> /P <sup>4a</sup>	-	(Migliolo et al. 2010)
	PKPI	<i>Solanum tuberosum</i>	Q66LL2	SP	T/Q <sup>5</sup>	3TC2	(Meulenbroek et al. 2012; Valueva et al. 2012)
	API	<i>Sagittaria sagittifolia</i>	Q7M1P4	SP	T/Q/C <sup>6</sup>	3E8L <sup>b</sup>	(Bao et al. 2009; Xie et al. 1997)
	PPI	<i>Carica papaya</i>	P80691	SP	T	3S8K, 3S8J	(Azarkan et al. 2011)
	WCI	<i>Psophocarpus tetragonolobus</i>	P10822	SP	Q	2WBC	(Dattagupta et al. 1999)
Bowman-Birk (112)	MSTI	<i>Medicago scutellata</i>	P80321	SP	T	2ILN <sup>b</sup>	(Capaldi et al. 2007)
	sBBI	<i>Glycine max</i>	P01055	SP	T/Q	1PI2, 1D6R <sup>b</sup>	(Chen et al. 1992; Koepke et al. 2000)
	LCTI	<i>Lens culinaris</i>	Q8W4Y8	SP	T/Q	2AIH	(Ragg et al. 2006)
	BTCl	<i>Vigna unguiculata</i>	P17734	SP	T/Q	2G81 <sup>c</sup> , 2OT6	(Barbosa et al. 2007; Rao and Suresh 2007)
	HGI-III	<i>Dolichos biflorus</i>	Q9S9E3	SP	T/Q	-	(Sreerama and Gowda 1997)
	CLTI	<i>Canavalia lineata</i>	Q7M1Q0	SP	T/Q	-	(Terada et al. 1994)
	A-II	<i>Arachis hypogaea</i>	P01066	SP	T/Q	No depositada	(Nomura and Suzuki 1993; Norioka and Ikenaka 1983)
	TcTI	<i>Amburana acreana</i>	P83284	SP	T/Q/PI <sup>7</sup>	-	(Tanaka et al. 1997)
	I-2B	<i>Triticum aestivum</i>	P09863	SP	T	-	(Odani et al. 1986)
	II-4	<i>Triticum aestivum</i>	P09864	SP	T	-	(Odani et al. 1986)
	LBI	<i>Phaseolus lunatus</i>	P01056	SP	T/Q	1H34	(Debreczeni et al. 2003; Krahn and Stevens 1971)
Clitocipina (148)	Clitocipina	<i>Clitocybe nebularis</i>	Q3Y9I4	CP	P/Ct <sup>9</sup>	3H6R, 3H6S <sup>d</sup>	(Renko et al. 2010)
		<i>Macrolepiota procera</i>	B9V973	CP/SP	P/Cu/T <sup>3</sup>	3H6Q	(Renko et al. 2010)
Macrocipina (185)	Macrocipina	<i>Macrolepiota procera</i>	B9V973	CP/SP	P/Cu/T <sup>3</sup>	3H6Q	(Renko et al. 2010)
Otros	Inhibidor de tripsina	<i>Trichosanthes</i>	P01069	SP	T	-	(Tan et al. 1984)
		<i>Kirilowii</i>					

<sup>1</sup> Serín peptidasas. <sup>2</sup> Cisteín peptidasas. <sup>3</sup> Tripsina. <sup>4</sup> Papaína. <sup>5</sup> Quimotripsina. <sup>6</sup> Calicreína. <sup>7</sup> Plasmina. <sup>8</sup> Elastasa. <sup>9</sup> Catepsina L, V, S y K. <sup>a</sup> Inhibición ligera <sup>b</sup> Complejo con dos moléculas de tripsina. <sup>c</sup> Complejo con tripsina. <sup>d</sup> Complejo con catepsina.

En un sentido más amplio, también pueden considerarse como inhibidores de doble faz los que presentan más de un sitio reactivo con independencia de su estructura terciaria. Esto incluiría a inhibidores multidominio, formados frecuentemente por procesos de duplicación génica puntuales o múltiples, de manera que presentan repeticiones del dominio original, y por tanto de su sitio reactivo, a lo largo de su estructura. Algunos de los dominios de estos

inhibidores de doble faz multidominio pueden haber evolucionado de manera independiente adquiriendo nuevas funciones como la capacidad de inhibir diferentes peptidasas incluso de clases distintas. Uno de estos inhibidores es el recientemente descrito SmCI, un inhibidor bifuncional de metalocarboxipeptidasas y serín peptidasas con tres dominios tipo Kunitz aislado del anélido *Sabellastarte magnifica* (Alonso-del-Rivero et al. 2012). Otro ejemplo de este tipo de inhibidores es esquistatina, un inhibidor también con tres dominios, en este caso tipo tiroglobulina tipo 1, aislado de la anémona *Actinia equina*. Este inhibe a las cisteín peptidasas papaina y catepsina B y L, y a la aspartil peptidasa catepsina D (Lenarcic et al. 1997; Lenarcic and Turk 1999).



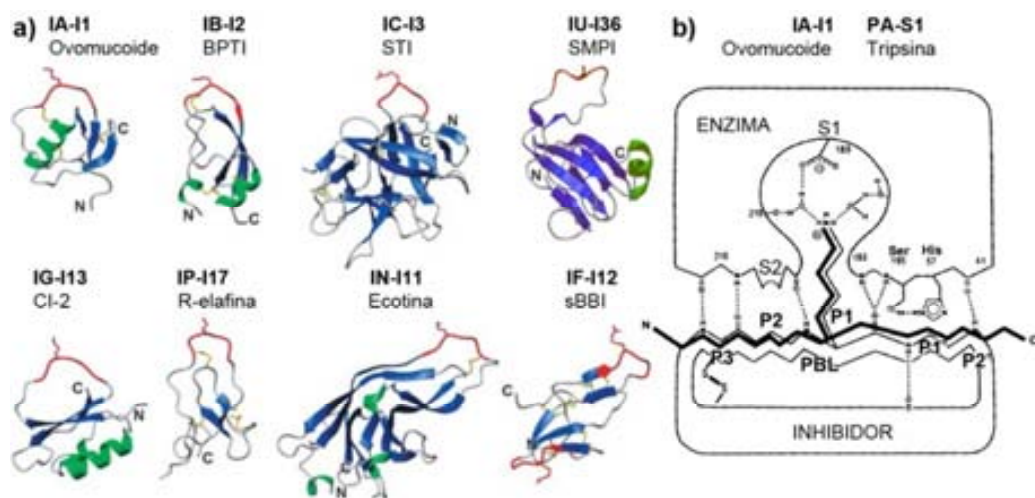
**Figura I.18.** Inhibidor de doble faz en complejo con su peptidasa diana. Estructura del complejo entre dos moléculas de tripsina (hebras  $\beta$  en magenta, hélices  $\alpha$  en cian, lazos en rosa, residuos catalíticos en rojo e iones calcio en rosa) y el inhibidor de peptidasas punta de flecha (*arrowhead protease inhibitor* [API-A]; hebras  $\beta$  en amarillo, hélices  $\alpha$  en rojo y lazos en verde). Código *Protein Data Bank* (PDB) 3E8L (Bao et al. 2009). Figura preparada con el programa PyMOL (Schrodinger 2010).

Finalmente, los inhibidores proteínicos de peptidasas, de igual forma que los inhibidores de molécula pequeña, pueden clasificarse según si su acción es reversible o irreversible. La mayoría de los inhibidores proteínicos son reversibles, aunque existen excepciones como  $\alpha_2$ -macroglobulina (Salvesen and Barrett 1980) o serpinas. Estas últimas son un grupo de inhibidores principalmente de serín peptidasas que actúan formando un complejo acil-enzima covalente irreversible (Matheson et al. 1991). La inhibición reversible competitiva, la más frecuente en los inhibidores proteínicos, puede describirse cuantitativamente en términos de unión del inhibidor con la enzima y su efecto en las constantes cinéticas de la enzima. La constante de inhibición ( $K_i$ ) se define como el producto de las concentraciones

de enzima e inhibidor libres divididas por la del complejo enzima-inhibidor ( $K_i = [E][I]/[EI]$ ), de manera que cuanto más pequeña sea la  $K_i$  mayor será la eficiencia inhibitoria. Por otro lado, la constante de Michaelis-Menten ( $K_m$ ) se define como la concentración de sustrato a la cual la velocidad de la reacción es la mitad de la velocidad máxima. Este tipo de inhibidores produce el incremento de la  $K_m$  aparente ( $K_m^{ap}$ ) sin modificar la velocidad máxima de la reacción de acuerdo con la relación  $K_m^{ap} = K_m(1+[Sustrato]/K_i)$ .

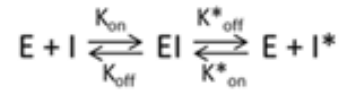
### 1.3.1. MECANISMO ESTÁNDAR DE INHIBICIÓN

Una parte importante de los inhibidores proteínicos descritos interaccionan con sus peptidasas diana siguiendo el mecanismo estándar de inhibición (Bode and Huber 1991, 1992; Laskowski Jr. and Kato 1980). Este se basa en la interacción de un lazo reactivo superficial del inhibidor (*reactive-site loop*) en una conformación canónica extendida con el centro activo de la peptidasa mimetizando la unión de la enzima con su sustrato (Figura I.19). De esta manera se forma un complejo similar al de Michaelis.





El lazo reactivo del inhibidor contiene un enlace peptídico hidrolizable que puede ser cortado, generalmente de manera muy lenta, por la peptidasa (Laskowski and Qasim 2000). Este tipo de interacción sigue el siguiente esquema donde “E” es la enzima libre, “I” el inhibidor libre nativo, “I\*” el inhibidor cortado y “EI” el complejo enzima inhibidor:



Por tanto se trata de una reacción reversible, donde el complejo puede disociarse liberando inhibidor en forma cortada o nativa. Típicamente, la constante de hidrólisis ( $K_h = [I^*]/[I]$ ) en condiciones fisiológicas tiende a la unidad, pero puede incrementarse notablemente mediante variaciones en el pH del medio de reacción (Ardelt and Laskowski 1983). No obstante, existen diferencias significativas en la cinética del mecanismo estándar de inhibición entre las diferentes peptidasas. En casos como el del inhibidor pancreático bovino de tripsina (BPTI), la vida media del inhibidor intacto en presencia de la enzima diana es de años (Ascenzi et al. 2003), mientras que en del inhibidor de metaloproteinasas de insecto (IMPI) el producto mayoritario de la disociación del inhibidor en complejo con la enzima es inhibidor cortado (Arolas et al. 2011). La reversibilidad del proceso queda patente en experimentos en los cuales se ha demostrado que la incubación del inhibidor cortado con la enzima da lugar a inhibidor nuevamente unido (Arolas et al. 2011; Zakharova et al. 2009).

Inicialmente, el mecanismo estándar de reacción fue descrito para inhibidores de serín peptidasas y ha sido demostrado o sugerido para 18 familias (Krowarsch et al. 2003). En los casos de peptidasas tipo tripsina, el segmento N-terminal del enlace peptídico hidrolizable se une a la enzima a modo de lámina  $\beta$  antiparalela a través de sus cadenas principales mediante puentes de hidrógeno en las posiciones **P3** y **P1**, a la vez que el C-terminal interacciona a través de **P2'**. En peptidasas tipo subtilisina la interacción se da en forma de lámina  $\beta$  antiparalela de tres hebras. El sitio reactivo se sitúa próximo a los residuos catalíticos de la peptidasa. Las interacciones de la cadena lateral del residuo situado en **P1** con el subsitio **S1** de la enzima son determinantes para la especificidad del inhibidor frente a la enzima, siendo estas especialmente energéticas en peptidasas tipo quimotripsina (Bode and Huber 1992).

Posteriormente, este mecanismo de inhibición fue atribuido también a inhibidores de metalopeptidasas, como son el inhibidor proteínico de metalopeptidasas de *Streptomyces* (SMPI) (Seeram et al. 1997) e IMPI (Arolas et al. 2011). En ambos casos se ha demostrado la

reversibilidad del proceso de hidrólisis del enlace peptídico reactivo, proceso que ocurre de una manera mucho más rápida que para la mayoría de inhibidores de serín peptidasas. De los dos inhibidores, únicamente se dispone de datos estructurales de la interacción con su peptidasa diana para IMPI. De igual manera que los inhibidores de serín peptidasas, las posiciones **P3** y **P1** del lazo reactivo de este inhibidor establecen una interacción en lámina  $\beta$  antiparalela con la enzima, con la diferencia de que el contacto se realiza a través del límite superior de la hendidura del centro activo en lugar de con su segmento inferior. Además, en este caso la especificidad de la inhibición viene conferida por el subsitio **S1'** y no por el **S1** (Arolas et al. 2011).

### 1.3.2. INHIBIDORES PROTEÍNICOS DE METALOPEPTIDASAS

Los inhibidores proteínicos de metalopeptidasas pueden dividirse en dos grandes grupos según si actúan sobre metalocarboxipeptidasas o metaloendopeptidasas. Algunos de los más destacados del primer grupo son el inhibidor de carboxipeptidasas de patata (PCI) y el inhibidor de carboxipeptidasas de sanguijuela (LCI). Por otro lado, los inhibidores proteínicos de metaloendopeptidasas más representativos y ampliamente estudiados son SMPI, IMPI, los inhibidores de metalopeptidasas de tejidos (TIMP) 1-4 y los inhibidores de serralisinas como el inhibidor de *Pseudomonas aeruginosa* (APRin), el inhibidor de *Serratia marcescens* (SmaPI) y el inhibidor de *Erwinia chrysanthemi* (Arolas et al. 2011; Hege et al. 2001; Murphy 2011).

Los inhibidores de metalocarboxipeptidasas PCI y LCI (familia MEROPS I37 e I46, respectivamente), así como otros relacionados, actúan interaccionando con la peptidasa mimetizando la interacción de un producto sin comportar cambios conformacionales significativos en la estructura del inhibidor. Forman un complejo “enzima-producto intermediario del mecanismo catalítico” de forma tan estrecha que su disociación es muy lenta. Interaccionan a través de su extremo C-terminal, pero también a través de otras regiones de su estructura, lo que da estabilidad al complejo y evita su disociación. La cola C-terminal se inserta en la hendidura del centro activo (Figura I.20) de manera que la cadena lateral del aminoácido situado en **P1** tiene un papel clave en la estabilización de la interacción (Rees and Lipscomb 1982; Reverter et al. 2000).

En cambio, los inhibidores proteínicos de metaloendopeptidasas presentan mecanismos de inhibición completamente diferentes. Se dispone de datos estructurales en complejo con sus peptidasas diana para algunos de ellos: IMPI, TIMP-1, TIMP-2, TIMP-3, APRin e inhibidor

de *Erwinia chrysanthemi* (códigos PDB 3SSB (Arolas et al. 2011), 1UEA (Gomis-Ruth et al. 1997), 2E2D (Maskos et al. 2007), 3CKI (Wisniewska et al. 2008), 1JIW (Hege et al. 2001) y 1SMP (Baumann et al. 1995), respectivamente). El análisis de estas estructuras revela tres modos distintos de inhibición representados por los complejos termolisina-IMPI, MMP-TIMP y serralisina-inhibidor (Figura I.20).



**Figura I.20.** Inhibidores de metalopeptidasas en complejo con sus peptidasas diana. Figuras preparadas con el programa PyMOL (Schrodinger 2010). Las peptidasas se representan con las hebras  $\beta$  en magenta, hélices  $\alpha$  en cian, lazos en rosa, residuos de unión al zinc catalítico en rojo, iones calcio en rosa, ion zinc estructural en gris claro e iones zinc catalíticos en gris oscuro. Los inhibidores se representan con las hebras  $\beta$  en amarillo, hélices  $\alpha$  en rojo y lazos en verde. Estructura del **a)** complejo entre carboxipeptidasa A e inhibidor de carboxipeptidasas de sanguijuela (LCI; código *Protein Data Bank* [PDB] 1DTD (Reverter et al. 2000)), **b)** complejo entre prometalopeptidasa de matriz 3 (MMP-3) e inhibidor de metalopeptidasas de tejidos 1 (TIMP-1; código PDB 1UEA (Gomis-Ruth et al. 1997)), **c)** complejo entre termolisina e inhibidor de metaloproteinasas de insecto (IMPI; código PDB 3SSB (Arolas et al. 2011)) y **d)** complejo entre serralisina e inhibidor de *Pseudomonas aeruginosa* (APRin; código PDB 1JIW (Hege et al. 2001)).

Los inhibidores SMPI e IMPI (familia MEROPS I36 e I08, respectivamente) son inhibidores principalmente de termolisinas (familia M04) que siguen el mecanismo estándar de inhibición tal y como se ha comentado en el apartado anterior.

Los inhibidores TIMP-1, 2, 3 y 4 (familia MEROPS I35), por su parte, son inhibidores de MMPs, peptidasas implicadas en la degradación de la matriz extracelular durante los procesos de desarrollo embrionario, morfogénesis y remodelación de tejidos (Murphy 2011). Los inhibidores TIMP regulan la actividad proteolítica de estas enzimas de manera precisa, de forma que alteraciones en el balance de inhibición pueden resultar en graves enfermedades como artritis, arterosclerosis, crecimiento tumoral y metástasis (Gomis-Ruth et al. 1997). También se ha observado inhibición frente a ADAMs y ADAMTs (ADAMs con motivos trombospondina), peptidasas igualmente implicadas en procesos de remodelación tisular (Brew and Nagase 2010). Los TIMPs interaccionan con sus peptidasas diana mediante uniones estrechas no covalentes con una especificidad moderada, por lo que cada TIMP puede inhibir diferentes peptidasas. Su región de contacto con la enzima está compuesta por el segmento N-terminal juntamente con cinco lazos inhibitorios, de manera que la superficie de interacción entre la peptidasa y el inhibidor es mayor que en los inhibidores de metalopeptidasas anteriormente descritos. El residuo situado en la posición N-terminal coordina el ion zinc catalítico, mientras que las interacciones con el bolsillo de especificidad de la enzima ocurren a través de la posición **P1'**. El estudio de las interacciones entre TIMPs y MMPs abre nuevas posibilidades para el diseño de variantes de TIMPs así como de nuevos inhibidores sintéticos con aplicaciones terapéuticas (Gomis-Ruth et al. 1997; Nagase and Brew 2002).

Finalmente, APRin, SmaPI y el inhibidor de *E. chrysanthemi* (familia MEROPS I38) son inhibidores de la familia de las serralisinas. Sus  $k_s$  son 4 pM, 0,7 mM y 1-10 mM, respectivamente, por lo que presentan cinéticas de inhibición muy diferentes (Hege et al. 2001). No obstante, presentan un mecanismo de inhibición común a través de las interacciones de los cinco primeros residuos del extremo N-terminal con el centro activo de la peptidasa estabilizado por múltiples interacciones secundarias en otras regiones. El residuo N-terminal coordina el ion zinc catalítico de forma análoga a los TIMPs (Baumann et al. 1995).

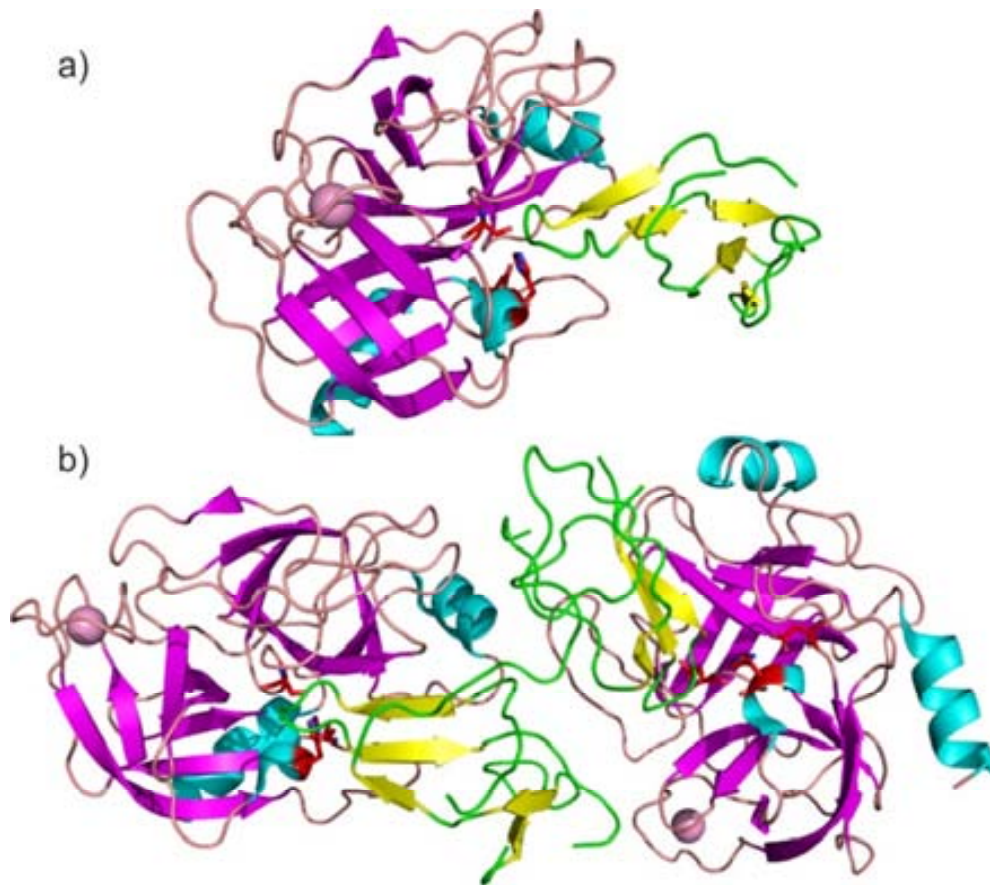
### 1.3.3. INHIBIDORES PROTEÍNICOS DE SERÍN PEPTIDASAS

Los inhibidores proteínicos de serín peptidasas se agrupan dependiendo de su mecanismo de acción en inhibidores canónicos, inhibidores no canónicos y serpinas (Otlewski et al. 2005).

Los inhibidores canónicos representan el mayor grupo de inhibidores proteínicos de peptidasas. Interaccionan con sus peptidasas dianas mediante el mecanismo estándar de inhibición (Figura I.21). Un gran número de ellos ha sido descrito y existen numerosas estructuras cristalográficas, tanto de sus formas libres como en complejo con sus peptidasas diana. Los inhibidores canónicos se encuentran ampliamente distribuidos en prácticamente todos los grupos de organismos. Se acumulan frecuentemente en grandes cantidades en semillas de plantas, huevos de aves y fluidos biológicos, tienen entre 14 y 200 aminoácidos y se agrupan en diferentes familias. Inicialmente fueron clasificadas en 8 familias de acuerdo con la topología de sus puentes disulfuro, la localización del sitio reactivo y la homología de sus secuencias (Laskowski Jr. and Kato 1980). Actualmente se reconocen 18 familias, la estructura terciaria de las cuales puede ser completamente diferente. A menudo presentan una estructura mixta  $\alpha/\beta$  o todo  $\beta$ , aunque también pueden ser todo  $\alpha$  o proteínas irregulares ricas en puentes disulfuro. Este tipo de inhibidores representa un claro ejemplo de convergencia evolutiva, ya que la conformación del lazo canónico inhibitorio puede encontrarse en diversas proteínas no relacionadas evolutivamente. Esta conformación habría aparecido en varias ocasiones a lo largo de la evolución (Krowarsch et al. 2003).

La mayoría de inhibidores canónicos son proteínas monodominio. Su dominio inhibitorio contiene un único sitio reactivo a excepción de los inhibidores de doble faz, cuya práctica totalidad son inhibidores canónicos de serín peptidasas. En algunas familias como antistasina, Bowman-Birk, BPTI o Kazal, el dominio puede encontrarse repetido de dos a 15 veces formando un inhibidor multidominio capaz de unir diferentes peptidasas a través de sus sitios reactivos de manera independiente (Figura I.21).

Los inhibidores no canónicos interaccionan con su peptidasa diana a través de una extensa región de su superficie fuera del centro activo, contribuyendo así a la intimidad, velocidad y especificidad del reconocimiento. Además, este tipo de inhibidores interacciona a través de su segmento N-terminal uniéndose al centro activo de la peptidasa formando una lámina  $\beta$  corta. Son inhibidores mucho menos frecuentes que los canónicos o las serpinas y únicamente han sido encontrados en organismos hematófagos. Estos los utilizan para inhibir trombina o factor  $X_a$ , peptidasas implicadas en la coagulación (Krowarsch et al. 2003).



**Figura I.21.** Inhibidores canónicos de serín peptidasas en complejo con sus peptidasas diana. Figuras preparadas con el programa PyMOL (Schrodinger 2010). Estructura del **a)** complejo entre tripsina e inhibidor pancreático bovino de tripsina (BPTI; código *Protein Data Bank* [PDB] 2G81 (Barbosa et al. 2007)) y **b)** complejo entre dos moléculas de tripsina y una del inhibidor de tripsina tipo Bowman-Birk bidominio de semillas de cebada (código PDB 1TX6 (Park et al. 2004)). Las moléculas de tripsina se representan con las hebras  $\beta$  en magenta, hélices  $\alpha$  en cian, lazos en rosa, residuos catalíticos en rojo e iones calcio en rosa. Los inhibidores se representan con las hebras  $\beta$  en amarillo y lazos en verde.

Un ejemplo típico de la interacción de un inhibidor no canónico con su peptidasa diana es el complejo entre el inhibidor hirudina de sanguijuela (familia MEROPS I14) y trombina (Figura I.17b). Otros miembros representativos de este tipo de inhibidores son el péptido anticoagulante de garrapata (TAP) y ornitorina, también de garrapata (familia MEROPS I52 e I12, respectivamente) (Otlewski et al. 2005). Ambos presentan el plegamiento típico de los inhibidores canónicos tipo BPTI/Kunitz pero muestran una conformación distorsionada del lazo reactivo.

Las serpinas (familia MEROPS I4), acrónimo de *serine protease inhibitor*, son inhibidores de serín peptidasas de las familia MEROPS S1 (quimotripsinas) y S8 (subtilisinas), y de cisteín peptidasas C1 (papainas) y C14 (caspasas), que actúan mediante la formación de un complejo intermedio covalente tetraédrico acil-enzima irreversible (Matheson et al. 1991). Este mecanismo, conocido como trampa para ratones (*mouse-trap*), implica importantes

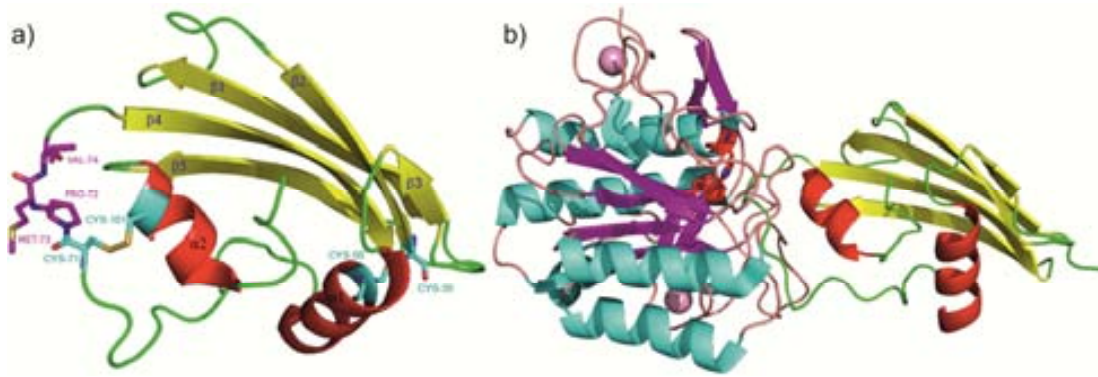
cambios conformacionales en el inhibidor y la alteración permanente del centro activo de la peptidasa. Las primeras serpinas identificadas fueron  $\alpha_1$ -antitripsina y antitrombina, proteínas plasmáticas humanas implicadas en el control de la inflamación y de la coagulación sanguínea, respectivamente (Carrell and Owen 1985). Deficiencias en estos inhibidores causan diferentes enfermedades como trombosis, en el caso de antitrombina (Kubier and O'Brien 2012), y enfisema, en el de  $\alpha_1$ -antitripsina (Brode et al. 2012), por lo que han sido estudiados exhaustivamente. Hasta la fecha se han identificado cientos de miembros de esta familia distribuidos en animales superiores, nematodos, insectos, plantas y virus (Irving et al. 2000), así como en procariontes y hongos (Irving et al. 2002).

### **1.3.3.1. Inhibidor de subtilisina de *Streptomyces***

El inhibidor de subtilisina de *Streptomyces* (SSI; familia MEROPS I16) es un inhibidor canónico de serín peptidasas de tipo subtilisina. Fue aislado y descrito hace más de 40 años en el medio extracelular de *Streptomyces albogriseolus* S-3253 (Muraio 1972) y desde entonces ha sido ampliamente estudiado. En su momento fue el primer inhibidor descrito de peptidasas alcalinas producido en microorganismos. SSI, con una masa molecular de 11,5 kDa, adopta una estructura dimérica compuesta por dos subunidades idénticas de 133 residuos (Figura 1.22). El enlace reactivo para subtilisina ha sido identificado como M73-V74. Presenta una  $k_i = 5$  pM frente a subtilisina BPN' utilizando succinil-Ala-Ala-Pro-X-p-nitroanilida como sustrato (Kojima et al. 1991). Por otro lado, se ha descrito que SSI también inhibe ligeramente a tripsina y quimotripsina (Muraio 1972; Takeuchi et al. 1992).

Contrariamente a lo que cabría esperar para un inhibidor de serín peptidasas, SSI presenta actividad inhibitoria significativa frente a la metaloendopeptidasa II de *Streptomyces griseus* (SGMP II, también conocida como griselisina, familia MEROPS M4) con una  $K_i = 41$  nM utilizando Bim-SCH<sub>2</sub>CO-Phe-Trp-Leu-OH como sustrato. Otros miembros de la familia I16, como son API-2c' y plasminostreptina, también presentan actividad contra griselisina ( $K_i = 1,3$  nM y 74 pM, respectivamente). La actividad de termolisina de *Bacillus thermoproteolyticus*, también miembro de la familia M4, ha sido analizada en presencia de SSI confirmando la especificidad de este frente a griselisina (Kajiwara et al. 1991). Estudios posteriores con mutantes de SSI demostraron que la inhibición de subtilisina y griselisina se producía a través del mismo lazo reactivo (Kumazaki et al. 1993). El enlace peptídico reactivo para griselisina está situado entre P72 y M73, una posición anterior al enlace implicado en la inhibición de subtilisina.





**Figura 1.22.** Inhibidor de subtilisina de *Streptomyces*. Figuras preparadas con el programa PyMOL (Schrodinger 2010).

**a)** Estructura del inhibidor de subtilisina de *Streptomyces* (SSI). Código *Protein Data Bank* (PDB) 3SSI (artículo no publicado). Se representa con las hebras  $\beta$  en amarillo, hélices  $\alpha$  en rojo, lazos en verde, residuos reactivos en magenta y cisteínas en cian. **b)** Estructura del complejo entre subtilisina (hebras  $\beta$  en magenta, hélices  $\alpha$  en cian, lazos en rosa, residuos catalíticos en rojo e iones calcio en rosa) y SSI (hebras  $\beta$  en amarillo, hélices  $\alpha$  en rojo y lazos en verde). Código PDB 2SIC (Takeuchi et al. 1991a).

La estructura de SSI fue resuelta en 1979 a una resolución de 2,6 Å (código PDB 1SSI) (Mitsui et al. 1979). Posteriormente ha sido remplazada en dos ocasiones (códigos PDB 2SSI (Satow et al. 1980) y 3SSI [artículo no publicado, 1996]) incrementando la resolución hasta 2,3 Å y el número de átomos identificados.

Cada subunidad del inhibidor está compuesta por cinco hebras  $\beta$  antiparalelas, asociadas en una hoja  $\beta$ , y dos hélices  $\alpha$  cortas. La dimerización se produce a través de la interacción de las caras convexas de las hojas  $\beta$ . Dos puentes disulfuro, C35-C50 y C71-C101, aportan estabilidad y rigidez al inhibidor. Concretamente, el puente C71-C101 estabiliza el lazo reactivo compuesto por los residuos comprendidos entre E67 y Y75 (Mitsui et al. 1979). Este incluye los enlaces reactivos para griselisina (P72-M73) y subtilisina (M73-V74).

La resolución de la estructura del complejo entre subtilisina BPN' y SSI demostró la formación de un complejo 2:2 entre ambas proteínas y confirmó la implicación del lazo anteriormente descrito en la inhibición (Hirono et al. 1984).

La mutación de M73, situada en la posición **P1**, comporta importantes variaciones en la afinidad entre SSI y diferentes serín peptidasas. El cambio de metionina por ácido aspártico, ácido glutámico, valina, isoleucina, glicina o prolina comporta una gran disminución de la afinidad de SSI con subtilisina (variación en la  $K_i$  = 5 pM a 5,8 nM, 0,31 nM, 72 pM, 0,39 nM, 0,11 nM y 24 nM, respectivamente) (Kojima et al. 1991). Por otro lado, la mutación de metionina por lisina o arginina comporta un incremento de su afinidad por tripsina, mientras que la mutación por tirosina y triptófano incrementa su afinidad por quimotripsina (Kojima et al. 1990). Mediante la mutación de M73 a K y de M70 a G



(posición **P<sub>4</sub>**) se consigue un complejo estable 2:2 entre SSI y tripsina cuya estructura pudo ser resuelta por cristalografía de rayos X (código PDB 2TLD (Takeuchi et al. 1992)).

Hasta el momento se han descrito diferentes aplicaciones biotecnológicas y terapéuticas para SSI, dando lugar a diversas publicaciones científicas y patentes (Fox and Wang 2006; Ganz et al. 2004; Saunders 1999; Saunders et al. 1999). Uno de los usos más destacados de SSI es el de estabilizar y conservar la subtilisina presente en detergentes para la ropa, lavavajillas, limpiadores industriales y productos de cuidado personal. Para ello se utiliza el mutante de SSI A62K+L63I+M73P+D83C+S98E (Ganz et al. 2004). Las mutaciones A62K, M73P y S98E contribuyen a disminuir la afinidad frente a subtilisina y así permitir la disociación del complejo al diluir el detergente. La mutación D83C permite la formación de un enlace disulfuro entre las subunidades del dímero de SSI incrementando substancialmente su termoestabilidad. Esta mutación, junto con L63I, incrementa la estabilidad del inhibidor en presencia de surfactantes. Este mutante múltiple de SSI puede ser añadido a la composición de diferentes detergentes que contengan subtilisina de manera económica y segura (Saunders et al. 1999) o puede ser coexpresado como proteína de fusión junto con la propia peptidasa y añadido conjuntamente al detergente (Saunders 1999).

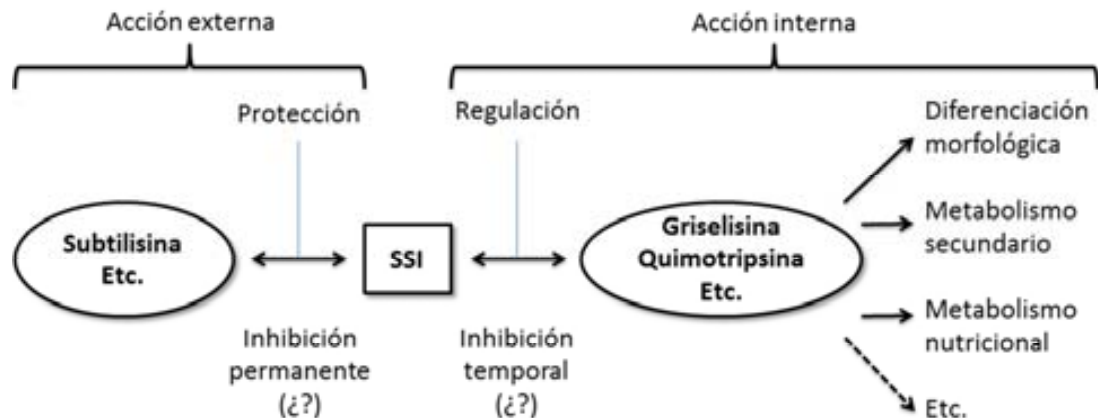
Desde un punto de vista terapéutico, SSI podría ser utilizado para el tratamiento y prevención de la aspergilosis pulmonar invasiva. Se trata de una enfermedad que presenta una elevada tasa de mortalidad en aplicación de anfotericina B (29%), tradicionalmente el tratamiento más eficaz (Herbrecht et al. 2002). Actualmente, la anfotericina B está siendo reemplazada como fármaco de elección por voriconazol (Walsh et al. 2008), aunque existe cierta controversia (Agarwal and Singh 2006). Voriconazol presenta una mejor respuesta terapéutica y menos efectos secundarios en comparación a anfotericina B, reduciendo la mortalidad al 13% (Herbrecht et al. 2002). La infección por *Aspergillus* es oportunista, de manera que afecta principalmente a pacientes inmunodeprimidos. El incremento en el uso de fármacos inmunosupresores, quimioterapia y radioterapia ha supuesto un aumento en la incidencia de enfermedades fúngicas. En pacientes que reciben alguno de estos tratamientos, así como en enfermos de SIDA, la aspergilosis representa la segunda enfermedad fúngica más frecuente por detrás de la candidiasis. El uso de SSI para el tratamiento y prevención de la aspergilosis pulmonar invasiva, vía gotas nasales, aerosoles nasales o aerosoles orales se encuentra en fase de desarrollo. Se ha demostrado un descenso de la mortalidad en animales de experimentación (Kolattukudy 1995) y cuenta con la ventaja de tratarse de un medicamento no tóxico en contraposición a anfotericina B,

fármaco marcadamente nefrotóxico (Pound et al. 2011), y a voriconazol (Epaulard et al. 2010). SSI inhibe a la peptidasas alcalina I de *Aspergillus*, también conocida como aspergilopeptidasa B u orizina ( $K_i = 1 \text{ nM}$  [*Aspergillus fumigatus*]), una peptidasas de tipo subtilisina (familia S08 según MEROPS) implicada en la asimilación de substratos proteínicos por parte del microorganismo. En las especies de *Aspergillus* patogénicas, orizina juega un papel clave en la invasión tisular, de manera que su inhibición podría ayudar a la prevención de la aspergilosis pulmonar invasiva (Guo and Ma 2008; Kolattukudy 1995; Markaryan et al. 1996).

En lo referente al papel biológico de SSI, se ha postulado su implicación en el control del desarrollo micelial en base a su regulación por el regulón AdpA en respuesta a factor A (2-isocaproilol-3R-hidroximetil- $\gamma$ -butirolactona) en *S. griseus*. Este compuesto actúa como hormona microbiana o molécula de señalización que desencadena el metabolismo secundario y la diferenciación celular. Dentro del mismo regulón se encontrarían los genes codificantes para peptidasas implicadas en la formación del micelio aéreo como la metalopeptidasa griselisina y las serín peptidasas tipo tripsina SprT y SprU y tipo quimotripsina SprA, SprB y SprD (Hirano et al. 2006). SSI regularía la actividad de todas ellas, pero su función sigue siendo controvertida ya que diferentes experimentos han aportado resultados contradictorios. Se ha observado que la adición del inhibidor al medio de cultivo retrasa la formación del micelio aéreo en *S. coelicolor* (Hirano et al. 2006) y que la mutación del gen en *S. albogriseolus* afecta a su capacidad para esporular (Taguchi et al. 1995). Por el contrario, se ha demostrado que la mutación del gen en *S. coelicolor* no comporta cambios en la diferenciación morfológica, formación de micelio o esporulación (Kato et al. 2005). Por otro lado, esta función no explica la elevada afinidad de SSI frente a subtilisina. La presencia de peptidasas tipo subtilisina ha sido demostrada en *S. albogriseolus* (Suzuki et al. 1997) e inferida por homología en otras familias. Por el momento en *Streptomyces* no ha sido descrita su implicación en estos procesos, aunque se sabe que en otros géneros bacterianos participan en la esporulación (Balaban et al. 2002). Si esto fuese cierto para *Streptomyces* reforzaría un rol de SSI en la regulación del desarrollo micelial (Taguchi et al. 1995).

Por otro lado, SSI podría estar implicado en la protección frente a peptidasas exógenas. Las bacterias del suelo establecen relaciones complejas de competencia y amensalismo, en las que están implicadas peptidasas e inhibidores de peptidasas (ver apartado 1.1). Las bacterias del género *Bacillus*, muchas de las cuales comparten hábitat con *Streptomyces*, segregan al medio elevadas cantidades de subtilisina, por lo que la expresión de SSI podría

ser un mecanismo defensivo. No obstante, las dos hipótesis sobre su función biológica no tienen por qué ser excluyentes. De este modo se ha postulado una teoría por la cual SSI actuaría como regulador de peptidasas endógenas y protector contra peptidasas exógenas (Taguchi 1995) (Figura 1.23).



**Figura 1.23.** Esquema de las funciones biológicas del inhibidor de subtilisina de *Streptomyces*. Mecanismo propuesto por (Taguchi 1995).

### 1.3.3.2. Inhibidor de la proteasa neutra de *Streptomyces caespitosus*

El inhibidor de la proteasa neutra de *Streptomyces caespitosus* (*Streptomyces caespitosus neutral proteinase inhibitor*; ScNPI), a partir de ahora denominado sermetstatina, es objeto de estudio del presente trabajo. Es un inhibidor monodominio de 113 aminoácidos de secreción extracelular perteneciente a la familia MEROPS I16, purificado por primera vez de medio de cultivo de *Streptomyces sp. I-355* (Hiraga et al. 2000). Estudios previos (Hiraga et al. 2000) han determinado la actividad inhibitoria de sermetstatina frente a diferentes serín peptidasas y frente a la metalopeptidasa tipo metzinquina esnapalisina. Se demostró también la ausencia de actividad frente a la cisteín peptidasa catepsina B, la aspartil peptidasa pepsina, la metalopeptidasa de familia desconocida almelisina (Shibata et al. 1997) y las metalopeptidasas vimelisina, elastasa de *Pseudomonas aureginosa* y termolisina, todas ellas gluzinquinas pertenecientes a la familia MEROPS M04 (Barrett et al. 1998; Rawlings et al. 2012).



## 2. OBJETIVOS

Jano, teniendo el báculo en la diestra, y la llave en la siniestra, me habló estas palabras con su boca de delante:

- Oh poeta estudioso de los tiempos, dejando el miedo, aprende lo que pides y percibe en tu mente mis voces.

Publio Ovidio Nasón (siglo I)

*Fastos*



---

El objetivo general del presente trabajo es estudiar la actividad inhibitoria y estructura tridimensional de sermetstatina. Los objetivos concretos son los siguientes:

1. Sobreexpresar heterológamente sermetstatina en *Escherichia coli* y purificarla a homogeneidad en cantidades suficientes para realizar estudios funcionales y estructurales.
2. Investigar el estado oligomérico de sermetstatina en solución.
3. Analizar la actividad inhibitoria de sermetstatina frente a diferentes metalopeptidasas de la familia de las metzincinas y diferentes serín peptidasas.
4. Determinar la estructura tridimensional de sermetstatina y realizar un análisis comparativo frente a la estructura del inhibidor de subtilisina de *Streptomyces SSI*.
5. Investigar la estequiometria del complejo subtilisina-sermetstatina en solución.
6. Determinar la estructura tridimensional del complejo subtilisina-sermetstatina y analizar las zonas de interacción entre ambas proteínas, así como comparar sus estructuras en el complejo con sus respectivas estructuras en forma libre y con la del complejo subtilisina-SSI.
7. Sobreexpresar heterológamente esnapalisina en *E. coli* y purificarla a homogeneidad, en cantidades suficientes para realizar estudios funcionales y estructurales.
8. Determinar la importancia del prodominio de esnapalisina en el plegamiento de la proteína y analizar el proceso de activación de la peptidasa a partir del zimógeno.
9. Investigar la estequiometria del complejo esnapalisina-sermetstatina en solución.
10. Determinar la estructura tridimensional del complejo esnapalisina-sermetstatina y analizar las zonas de interacción entre ambas proteínas, así como comparar sus estructuras en el complejo con sus respectivas estructuras en forma libre.
11. Investigar la estequiometria del complejo subtilisina-sermetstatina-esnapalisina en solución para confirmar la inhibición simultánea de ambas peptidasas por parte del inhibidor.
12. Determinar la estructura tridimensional del complejo subtilisina-sermetstatina-esnapalisina y compararla con la de cada uno de los complejos binarios.
13. Analizar el papel de los residuos determinantes en la interacción entre subtilisina y sermetstatina y entre esnapalisina y sermetstatina mediante mutagénesis.
14. Evaluar la importancia de los puentes disulfuro en el mantenimiento de la estructura de sermetstatina así como en la inhibición de subtilisina y esnapalisina mediante mutagénesis.





# 3. MATERIALES Y MÉTODOS

- ¿Por qué el año comienza por fríos, el cual mejor comenzaría por la Primavera?
- Yo entregué los primeros tiempos para hacer las cosas, porque todo el año no se pasase sin arte del agüero, cada uno demuestra sus artes obrando.

Publio Ovidio Nasón (siglo I)

*Fastos*



### 3.1. PRODUCCIÓN Y PURIFICACIÓN DE SERMETSTATINA

---

Un gen sintético (GenScript) codificante para sermetstatina de *S. caespitosus* (código UniProt Q9FDS0), también conocido como inhibidor de la peptidasa neutra de *S. caespitosus* (*Streptomyces caespitosus neutral proteinase inhibitor*; ScNPI) (Hiraga et al. 2000), sin los primeros 28 residuos correspondientes al péptido señal, fue clonado en un vector pET-32a (Novagen) modificado utilizando los sitios de restricción *Bgl*III y *Hind*III y verificado por secuenciación de ADN. Con este vector se obtiene una fusión N-terminal de sermetstatina con tiorredoxina y una cola de seis histidinas separadas por un conector y un sitio de reconocimiento para la peptidasas de *tobacco etch virus* (TEV). Para la determinación del péptido señal se utilizó el servidor SignalP 4.0 (Petersen et al. 2011). Sermetstatina fue producida por sobreexpresión heteróloga en células *E. coli* Origami2 (DE3) (Novagen). Estas fueron crecidas a 37°C en medio Luria-Bertani (LB) suplementado con 100 µg/ml de ampicilina y 10 µg/ml de tetraciclina. La expresión de tiorredoxina-His(6x)-sermetstatina fue inducida a una densidad óptica a 550 nm de longitud de onda (OD<sub>550</sub>) de 0,6 con isopropil-β-D-tiogalactopiranosido (IPTG) a una concentración final de 0,25 mM durante 12 h a 18°C. El cultivo fue centrifugado a 7.000xg durante 30 min a 4°C. El precipitado fue lavado dos veces con tampón A (50 mM Tris-HCl, 500 mM NaCl, pH 8,0) y resuspendido en el mismo tampón con 20 mM de imidazol suplementado con comprimidos de cóctel de inhibidor de proteasas libre de EDTA (Roche Diagnostics) y DNasa I (Roche Diagnostics). Las células fueron lisadas a 4°C utilizando un disruptor celular (Constant Systems) a una presión de 1,35 kbar. Posteriormente, los restos celulares fueron eliminados por centrifugación a 50.000xg durante 1 h a 4°C.

El sobrenadante fue filtrado a través de unos dispositivos con un tamaño de poro de 0,22 µm (Millipore) e incubado con resina de ácido níquel-nitrilotriacético (Ni-NTA; Invitrogen) previamente equilibrada con tampón A con 20 mM imidazol. La proteína fue eluida con tampón A con 350 mM imidazol y dializada durante 12 h a temperatura ambiente contra tampón B (50 mM Tris-HCl, 150 mM NaCl, 0,5 mM glutatión oxidado [GSSG], 3 mM glutatión reducido [GSH], pH 8,0) en presencia de peptidasa His(6x)-TEV en una relación enzima:substrato de 1:50 (p/p). La muestra digerida fue pasada repetidas veces a través de resina Ni-NTA previamente equilibrada con tampón A con la finalidad de eliminar His(6x)-TEV, tiorredoxina-His(6x) y trazas de tiorredoxina-His(6x)-sermetstatina no cortada presentes en la misma. Posteriormente, la muestra fue purificada por cromatografía de exclusión molecular en una columna HiLoad 16/60 Superdex 75 (GE Healthcare)

previamente equilibrada con tampón C (20 mM Tris-HCl, 150 mM NaCl, pH 7,4). El análisis de la estructura cuaternaria del inhibidor se realizó previa calibración de la misma columna con unos kits de calibración de filtración en gel de bajo y alto peso molecular (GE Healthcare), aplicando la muestra repetidas veces en la columna equilibrada en tampón 20 mM Tris-HCl pH 7,4 y concentraciones crecientes de NaCl (50 mM - 1 M).

La identidad y la pureza de sermetstatina fue determinada mediante espectrometría de masas y electroforesis en gel de tricina y poliacrilamida al 15% con dodecilsulfato sódico (Tricina-SDS-PAGE) teñido con Coomassie blue. Finalmente, la muestra fue concentrada por ultrafiltración con dispositivos Vivaspin 15 y Vivaspin 500 (Sartorius Stedim Biotech) con un tamaño de poro de 5 kDa. La concentración de proteína fue determinada por absorbancia a  $\lambda = 280$  utilizando un espectrofotómetro (NanoDrop) y un coeficiente de absorción molar teórico  $E_{0,1\%} = 0,86$ .

La variante de sermetstatina con selenometionina fue obtenida de la misma manera que la forma nativa, a excepción de que las células fueron crecidas en medio mínimo conteniendo selenometionina en lugar de metionina.

La serie de mutantes puntuales y dobles H<sup>3</sup>E, H<sup>3</sup>R, Y<sup>33</sup>P+T<sup>34</sup>G, M<sup>71</sup>K, Y<sup>72</sup>V+F<sup>73</sup>Y, C<sup>31</sup>S+C<sup>46</sup>S y C<sup>69</sup>S+C<sup>99</sup>S (números de residuos de sermetstatina indicados como superíndices de acuerdo a la secuencia de la proteína madura; código uniprot Q9FDS0) fueron generados usando el kit *QuickChange Site-Directed Mutagenesis* (Stratagene). Los mutantes fueron verificados por secuenciación de ADN y producidos y purificados de igual forma que la forma silvestre.

## **3.2. OBTENCIÓN DE ESNAPALISINA Y PROESNAPALISINA. ANÁLISIS DEL PRODOMINIO**

---

*S. coelicolor* fue cultivado en diferentes condiciones de temperatura (25°C, 37°C y 42°C) y composición del medio de cultivo (medio LB y otros medios de cultivo descritos para la producción de peptidasas en bacterias del género *Streptomyces* (Maldonado-Salvador 1984)) durante 24-96 h. Los cultivos fueron centrifugados a 7.000xg durante 30 min a 4°C. Posteriormente, los sobrenadantes fueron filtrados a través de unos dispositivos con un tamaño de poro de 0,22  $\mu$ m, incubados con sulfato de amonio al 80% de saturación durante 30 min a 0°C y centrifugados a 50.000xg durante 1 h a 4°C. El precipitado fue resuspendido en tampón C. La actividad proteolítica de estas muestras fue estudiada frente a sustratos conjugados fluorescentes BODIPY FL caseína y DQ gelatina (65  $\mu$ g/ml,

Invitrogen) en presencia y ausencia de EDTA (1 mM) y sermetstatina (10  $\mu$ M). Se utilizaron microplacas (Nunc) y un fluorímetro de microplacas (FLx800, BioTek) a  $\lambda_{ex} = 485$  nm /  $\lambda_{em} = 528$  nm.

Las cantidades de esnapalisina expresadas no fueron suficientes para abordar estudios posteriores de estructura y función, por lo se optó por la expresión heteróloga de esnapalisina en *E. coli* a partir de genes sintéticos.

### 3.2.1. ANÁLISIS BIOINFORMÁTICO

La secuencia de esnapalisina de *S. caespitosus* (código UniProt P56406) fue comparada con la de *S. coelicolor* (código UniProt P0A3Z7) mediante ClustalW (Thompson et al. 1994).

Para la determinación del péptido señal de proesnapalisina de *S. coelicolor* se utilizó el servidor SignalP 4.0 (<http://www.cbs.dtu.dk/services/SignalP-4.0>) (Petersen et al. 2011).

### 3.2.2. PRODUCCIÓN Y PURIFICACIÓN DE PROESNAPALISINA

Un gen sintético (GenScript) codificante para proesnapalisina de *S. coelicolor* (código UniProt P0A3Z7), sin los primeros 29 residuos correspondientes al péptido señal y la alanina N-terminal remplazada por metionina, fue clonado en una batería de vectores modificados a partir de pET-28a y pET-30a (Novagen) utilizando los sitios de restricción *Nco*I y *Xho*I. Se obtuvieron ocho vectores con fusiones en el extremo N-terminal de proesnapalisina con sitio de reconocimiento para la peptidasa TEV (His[6x]-GB1 His[6x]-Ztag, His[6x]-trx, His[6x]-MBP, His[6x]-GST, His[6x]-Nus, His[6x] y señal de exportación periplasmática) y uno con fusión C-terminal (His[6x]). El mutante proesnapalisina *E164A* (aminoácidos de proesnapalisina representados en cursiva; números de residuos indicados de acuerdo a la secuencia de la proteína completa; código uniprot P0A3Z7) fue obtenido usando el kit *QuickChange Site-Directed Mutagenesis* y verificado por secuenciación de ADN. Se realizaron pruebas de expresión con todos los vectores en diferentes cepas de *E. coli* DE3 (Origami2, pLysS, GroES y BL21) a diferentes tiempos y temperaturas de inducción (20°C - 8 h y 37°C - 3 h). Su solubilidad fue estudiada mediante Tricina-SDS-PAGE 10% teñido con Coomassie blue tras 10 ciclos de congelación descongelación de las células en nitrógeno líquido y una centrifugación a 15.000xg durante 20 min a 4°C. Para el vector con señal de exportación periplasmática se estudió la presencia de la proteína en el medio de cultivo.

Finalmente, ambas proteínas (silvestre y mutante) fueron sobreexpresadas con una cola de seis histidinas N-terminal en células de *E. coli* BL21 (DE3) crecidas a 37°C en medio LB suplementado con kanamicina hasta una concentración de 30 µg/ml. Las células fueron inducidas a una OD<sub>550</sub> de 0,8 con IPTG a una concentración de 1 mM y se dejaron crecer durante 5 h a 37°C. Tras una centrifugación a 7.000xg durante 30 min a 4°C el precipitado fue resuspendido en tampón D (PBS 1x, 1% Triton X-100, pH 7,4) conteniendo comprimidos de cóctel de inhibidor de proteasas libre de EDTA y DNAsa I. Las células fueron lisadas a 4°C utilizando un disruptor celular a una presión de 1,9 kbar e incubadas durante 30 min a 37°C.

Los cuerpos de inclusión resultantes fueron recuperados por centrifugación a 7.000xg durante 30 min a 4°C y lavados dos veces con tampón D. Seguidamente se resuspendieron en tampón E (50 mM Tris-HCl, 500 mM NaCl, 8 M urea, 25 M DTT, pH 8,0) y se incubaron durante 5 h a temperatura ambiente en agitación vigorosa. La muestra fue centrifugada a 50.000xg durante 1 h a 20°C. El sobrenadante fue dializado durante una noche a temperatura ambiente contra tampón F (50 mM Tris-HCl, 500 mM NaCl, 0,5 mM GSSG, 1 mM GSH, 10 mM CaCl<sub>2</sub>, 1 mM ZnSO<sub>4</sub>, 0,4 M L-arginina, pH 8,0), y posteriormente una noche a 4°C contra tampón C (20 mM Tris-HCl, 150 mM NaCl, pH 7,4). La cola de histidinas de proesnapalisina mutante E164A se cortó mediante una incubación con peptidasa TEV a una relación enzima:substrato de 1:50 (p/p) en tampón C con 0,5 mM GSSG y 3 mM GSH a temperatura ambiente durante una noche. Las proteínas fueron concentradas por ultracentrifugación y posteriormente purificadas por cromatografía de exclusión molecular en una columna HiLoad 16/60 Superdex 75 previamente equilibrada con tampón C. La identidad y la pureza de las proteínas fue determinada por degradación de Edman y Tricina-SDS-PAGE 15% teñido con Coomassie blue. Los pasos de ultrafiltración fueron realizados con dispositivos de filtrado Vivaspin 15 y Vivaspin 500 con un tamaño de poro de 5 kDa. La concentración de proteína fue determinada por absorbancia a  $\lambda = 280$  nm utilizando un espectrofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 1,38$  para ambas variantes.

### 3.2.3. PRODUCCIÓN Y PURIFICACIÓN DE ESNAPALISINA

Un gen sintético (GenScript) codificante para la forma madura de esnapalisina de *S. caespitosus* (código UniProt P56406), en el cual la treonina N-terminal fue reemplazada por metionina debido a la estrategia de clonación, fue clonado en la misma batería de vectores utilizada para proesnapalisina de *S. coelicolor* (ver apartado 3.2.2) utilizando los sitios de

restricción *Nco*I y *Xho*I. Se realizaron los ensayos de expresión y solubilidad descritos para proesnnapalisina. Finalmente la peptidasa fue sobreexpresada con una cola de seis histidinas N-terminal en células de *E. coli* BL21 (DE3) crecidas a 37°C en medio LB suplementado con kanamicina hasta una concentración de 30 µg/ml. Las células fueron inducidas a una OD<sub>550</sub> de 0,8 con IPTG a una concentración de 1 mM y se dejaron crecer durante 5 h a 37°C. Tras una centrifugación a 7.000xg durante 30 min a 4°C el precipitado fue resuspendido en tampón D conteniendo comprimidos de cóctel de inhibidor de peptidasas libre de EDTA y DNasa I. Las células fueron lisadas a 4°C utilizando un disruptor celular a una presión de 1,9 kbar e incubadas durante 30 min a 37°C.

Los cuerpos de inclusión resultantes fueron recuperados por centrifugación a 7.000xg durante 30 min a 4°C y lavados dos veces con tampón D. Se resuspendieron en tampón G (50 mM Tris-HCl, 500 mM NaCl, 8 M urea, pH 8,0) y se incubaron durante 5 h a temperatura ambiente en agitación vigorosa. La muestra fue centrifugada a 50.000xg durante 1 h a 20°C, y el sobrenadante fue incubado con resina Ni-NTA previamente equilibrada con tampón E. La proteína de fusión fue eluida con tampón H (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 350 mM NaCl, 8 M urea, pH 4,0). Finalmente, la muestra fue dializada durante una noche contra tampón F a temperatura ambiente y posteriormente una noche a 4°C contra tampón C. La proteína fue concentrada por ultracentrifugación y purificada por cromatografía de exclusión molecular en una columna Superdex 75 10/300 (GE Healthcare) previamente equilibrada con tampón C. La identidad y la pureza de la misma fue determinada por degradación de Edman y Tricina-SDS-PAGE 15% teñidos con Coomassie blue. Los pasos de ultrafiltración fueron realizados con dispositivos de filtrado Vivaspin 15 y Vivaspin 500 con un tamaño de poro de 5 kDa. La concentración de proteína fue determinada por absorbancia a  $\lambda = 280$  nm utilizando un espectrofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 1,7$ .

### 3.2.4. ESTUDIOS DE ACTIVACIÓN *IN-VITRO* DE PROESNAPALISINA

Se realizaron diferentes incubaciones en tampón C tanto de proesnnapalisina silvestre como del mutante *E164A* a diferentes tiempos (1 h - 1 mes), temperaturas (4°C, 20°C y 37°C) y concentraciones (0,5-5 mg/ml) en ausencia y presencia de EDTA (1-5 mM) y de esnapalisina madura de *S. caespitosus* (1-10 µM [0,015-0,15 mg/ml]). Los cortes producidos fueron monitorizados por espectrometría de masas, Tricina-SDS-PAGE 15% y degradación de Edman de bandas transferidas a una membrana Immun-Blot PVDF (BioRad).

### 3.2.5. CRISTALIZACIÓN DE PROESNAPALISINA E164A

Los ensayos de cristalización de proesnapalisina mutante E164A de *S. coelicolor* se llevaron a cabo en gota sentada mediante el método de difusión de vapor. Las soluciones de reservorio fueron preparadas por un robot Tecan. Gotas de 200 nL, equívolumétricas en reservorio y solución de proteína, fueron dispensadas en placas MRC de 96x2 pocillos (Innovadyne) por un robot Cartesian nanodrop (Genomic Solutions) en la Plataforma Automatizada de Cristalografía (PAC) del Institut de Biologia Molecular de Barcelona / Institut de Recerca Biomèdica (IBMB/IRB). Las placas de cristalización fueron almacenadas en dos incubadores Bruker Crystal Farm a una temperatura controlada y constante de 4°C y 20°C. Las condiciones positivas fueron escaladas a gotas de 2 µl (1 µl proteína + 1 µl reservorio) en placas de cristalización de 24 pocillos (Hampton Research).

Los cristales obtenidos fueron crioprottegidos mediante sucesivos pases a través de soluciones de reservorio con cantidades crecientes de glicerol (25-30% [v/v]). Posteriormente fueron difractados a 100K (Oxford Cryosystems 700 series cryostrem) en el *European Synchrotron Radiation Facility* (ESRF, Grenoble, Francia) dentro del “*Block Allocation Group* (BAG) Barcelona”. Para la comprobación de la integridad de la proenzima algunos de los cristales fueron lavados con solución de reservorio y analizados por Tricina-SDS-PAGE 15% con tinción de plata.

### 3.2.6. ESTUDIOS DE ACTIVIDAD

La actividad proteolítica de esnapalisina madura de *S. caespitosus*, de proesnapalisina silvestre de *S. coelicolor* y de su mutante E164A, así como de los fragmentos resultantes de la activación de los zimógenos, fue estudiada a una concentración de enzima de 0,3-1 µM en tampón C a 37°C frente a sustratos conjugados fluorescentes BODIPY FL caseína y DQ gelatina (65 µg/ml). El experimento se llevó a cabo en microplacas utilizando un fluorímetro de microplacas a  $\lambda_{ex} = 485 \text{ nm}$  /  $\lambda_{em} = 528 \text{ nm}$ . Para los ensayos de inhibición se utilizó EDTA (1 mM) y 1,10-fenantrolina (0,1 mM).



### 3.3. ANÁLISIS DE PROTEÓLISIS Y SU INHIBICIÓN POR SERMETSTATINA

---

Ulilisina de *Methanosarcina acetivorans*, aeruginolisina de *Pseudomonas aeruginosa*, y fragilisina-3 de *Bacteroides fragilis* fueron producidas y purificadas siguiendo los protocolos previamente publicados para cada una de ellas (Baumann et al. 1993; Goulas et al. 2011; Hege et al. 2001; Tallant et al. 2006). El vector codificante para aeruginolisina fue cedido por Ulrich Baumann (Universidad de Colonia, Alemania). Para la obtención de esnapalisina de *S. coelicolor* se siguió el protocolo descrito en el apartado 3.2.3. Astacina y las MMPs 1, 3, y 13 fueron provistas por Walter Stöcker (Universidad de Mainz, Alemania) y Hideaki Nagase (Imperial College London, Reino Unido), respectivamente. La actividad proteolítica de ADAM-17 en presencia de sermetstatina fue analizada por Jordi Malapeira y Joaquín Arribas (Instituto de Oncología de la Vall d'Hebron, España) utilizando vasorina como sustrato en un experimento basado en células previamente descrito (Malapeira et al. 2011). Subtilisina Carlsberg de *Bacillus licheniformis*, tripsina pancreática bovina, quimotripsina pancreática bovina, elastasa pancreática porcina y proteinasa K de *Engyodontium album* fueron obtenidos de Sigma.

La actividad proteolítica de todas estas enzimas, a excepción de ADAM-17, fue medida por triplicado a una concentración de 1  $\mu\text{M}$  en tampón C a 37°C utilizando los conjugados fluorescentes BODIPY FL caseína y DQ gelatina (65  $\mu\text{g}/\text{ml}$ ). El experimento se llevó a cabo en microplacas utilizando un fluorímetro de microplacas a  $\lambda_{\text{ex}} = 485 \text{ nm}$  /  $\lambda_{\text{em}} = 528 \text{ nm}$ . Cantidades crecientes de sermetstatina nativa purificada (1-100  $\mu\text{M}$ ) así como de los diferentes mutantes fueron añadidas al experimento para determinar su efecto inhibitor.

Por otro lado, se determinaron las  $K_i$  de sermetstatina silvestre contra subtilisina y esnapalisina mediante sendos diagramas de Lineweaver-Burk (inverso de la velocidad frente al inverso de la concentración de sustrato) y Dixon (inverso de la velocidad frente a la concentración de inhibidor). Para el cálculo de la  $K_i$  frente a subtilisina se utilizó el sustrato succinil-Ala-Ala-Pro-Phe-*p*-nitroanilida (Suc-AAPF-*p*NA; Sigma) en un lector de absorbancias de microplacas (ELx808, BioTek) a  $\lambda = 410 \text{ nm}$ . Las reacciones se llevaron a cabo en tampón D (20 mM Tris-HCl, 2 mM dimetilsulfóxido [DMSO], pH 7,5) con diferentes concentraciones de inhibidor (0 / 3,87 / 5,16 / 12,90 / 18,10 [nM]) y sustrato (0,2 / 0,3 / 0,4 / 0,5 / 0,75 / 1 / 1,5 [mM]). La  $K_i$  de sermetstatina frente a esnapalisina fue determinada utilizando el sustrato aminobenzoil-Val-Lys-Phe-Tyr-Asp-Ile-Lys(2,4-dinitrofenilamino)

(Abz-VKFYDIK[dnp]; cedido por Jean-Louis Reymond [Universidad de Berna, Suiza]) en un fluorímetro de microplacas (Infinite 200, Tecan) a  $\lambda_{\text{ex}} = 260 \text{ nm}$  /  $\lambda_{\text{em}} = 420 \text{ nm}$ . Las reacciones se realizaron en tampón C con concentraciones variables de inhibidor (0 / 2,73 / 3,64 / 5,45 / 6,36 / 7,27 [nM]) y sustrato (5,45 / 7,27 / 9,09 / 10,91 / 12,73 / 14,54 / 16,36 / 18,18 [nM]).

## **3.4. FORMACIÓN Y PURIFICACIÓN DE COMPLEJOS**

---

### **3.4.1. SUBTILISINA-SERMETSTATINA**

El liofilizado de subtilisina Carlsberg comercial fue reconstituido en tampón C y purificado por cromatografía de exclusión molecular en una columna HiLoad 16/60 Superdex 75 previamente equilibrada con el mismo tampón. Su concentración fue determinada midiendo la absorbancia a  $\lambda = 280 \text{ nM}$  utilizando un espectofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 0,86$ .

El complejo subtilisina-sermetstatina fue formado por incubación de cantidades equimolares de inhibidor nativo y peptidasa en tampón C y posteriormente purificado por exclusión molecular utilizando la columna HiLoad 16/60 Superdex 75. La pureza del mismo fue determinada mediante Tricina-SDS-PAGE 15% teñido con Coomassie blue, previa precipitación de las muestras con ácido tricloroacético (TCA) al 25% (p/v) y posterior lavado con acetona a  $-20^{\circ}\text{C}$  (este paso previo es necesario a fin de evitar la autodegradación de subtilisina durante el proceso normal de preparación de las muestras para SDS-PAGE, de manera que se realizó en todas las electroforesis que contenían esta enzima, se indique específicamente o no). Posteriormente, el complejo fue concentrado por ultracentrifugación hasta 5,5 mg/ml utilizando dispositivos Vivaspin 15 y Vivaspin 500 de 30 kDa de tamaño de poro. Su concentración fue determinada midiendo la absorbancia a  $\lambda = 280 \text{ nM}$  utilizando un espectofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 0,89$ .

### **3.4.2. ESNAPALISINA-SERMETSTATINA**

Se obtuvieron cuerpos de inclusión de esnapalisina tal como se describe en el apartado 3.2.3. Esnapalisina desplegada fue diluida en proporción 1:9 en tampón I (50 mM Tris-HCl, 500 mM NaCl, 0,55 mM GSSG, 1,1 mM GSH, 11 mM  $\text{CaCl}_2$ , 1,1 mM  $\text{ZnSO}_4$ , 0,44 M L-arginina)

junto con una concentración equimolar de sermetstatina nativa. La muestra fue incubada durante 4 h a 4°C y dializada durante una noche contra tampón C. El complejo esnapalisina-sermetstatina fue concentrado por ultracentrifugación y posteriormente purificado mediante repetidas cromatografías de exclusión molecular en una columna HiLoad 16/60 Superdex 75 previamente equilibrada con tampón C. La pureza del complejo y la identidad de sus componentes fueron determinados mediante Tricina-SDS-PAGE 15% teñido con Coomassie blue, permitiendo también una valoración semicuantitativa de las concentraciones de proteína y en consecuencia de la relación estequiométrica entre enzima e inhibidor. La muestra fue dializada durante 12 h a 4°C contra tampón J (20 mM Tris-HCl, 1 M NaCl, pH 7,4) y concentrada con dispositivos Vivaspin 15 y Vivaspin 500 de 10 kDa de tamaño de poro hasta 4,5 mg/ml. Su concentración fue determinada midiendo la absorbancia a  $\lambda = 280$  nm utilizando un espectrofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 1,32$ .

### **3.4.3. SUBTILISINA-SERMETSTATINA-ESNAPALISINA**

El complejo ternario entre subtilisina, sermetstatina y esnapalisina fue obtenido añadiendo una concentración equimolar de subtilisina sobre el complejo esnapalisina-sermetstatina. Posteriormente fue purificado mediante cromatografía de exclusión molecular en una columna Superdex 200 10/300 (GE Healthcare) previamente equilibrada con tampón C. La muestra fue concentrada con dispositivos Vivaspin 15 y Vivaspin 500 de 10 kDa de tamaño de poro hasta 7,5 mg/ml. Su concentración fue determinada midiendo la absorbancia a  $\lambda = 280$  nm utilizando un espectrofotómetro y un coeficiente de absorción molar teórico  $E_{0,1\%} = 1,10$ .

## **3.5. CRISTALIZACIÓN Y COLECCIÓN DE DATOS DE DIFRACCIÓN DE RAYOS X**

---

Los experimentos de cristalización de sermetstatina, de los complejos binarios subtilisina-sermetstatina y esnapalisina-sermetstatina, y del complejo ternario subtilisina-sermetstatina-esnapalisina se llevaron a cabo en gota sentada por el método de difusión de vapor. Las soluciones de reservorio fueron preparadas con un robot Tecan. Gotas de 200 nl, en proporción equivolúmica de solución de proteína y reservorio, fueron dispensadas por un robot nanodrop Cartesian en la PAC. Las placas de cristalización fueron almacenadas

en dos incubadores Bruker Crystal Farm a una temperatura controlada y constante de 4°C y 20°C, respectivamente. Las condiciones positivas para sermetstatina nativa y su derivado con selenometionina, complejo subtilisina-sermetstatina y complejo subtilisina-sermetstatina-esnapalisina fueron escaladas a gotas de 2 µl (1 µl proteína + 1 µl reservorio) en placas de cristalización de 24 pocillos.

Los mejores cristales de sermetstatina nativa y de su derivado se obtuvieron a 20°C con solución de proteína a 10 mg/ml en tampón C y solución de reservorio 100 mM HOC(COONa)(CH<sub>2</sub>COONa)<sub>2</sub>·2H<sub>2</sub>O (citrato de sodio dihidratado), 200 mM CH<sub>3</sub>COONH<sub>4</sub> (acetato de amonio), 10% p/v PEG 3.350, pH 5,6. Los mejores cristales del complejo subtilisina-sermetstatina aparecieron a 4°C con solución de proteína a 5,5 mg/ml en tampón C y solución de reservorio 100 mM (CH<sub>3</sub>)<sub>2</sub>AsOO<sup>-</sup> (cacodilato), 200 mM Zn(CH<sub>3</sub>COO)<sub>2</sub>·2H<sub>2</sub>O (acetato de zinc dihidratado), 10% (v/v) CH<sub>3</sub>-CHOH-CH<sub>3</sub> (isopropanol), pH 6,5. Los mejores cristales del complejo esnapalisina-sermetstatina se obtuvieron a 20°C con solución de proteína a 4,5 mg/ml en tampón J y solución de reservorio 100 mM HEPES, 200 mM MgCl<sub>2</sub>, 15,0% (v/v) CH<sub>3</sub>-CH<sub>2</sub>OH (etanol), pH 7,5. Finalmente, los mejores cristales del complejo subtilisina-sermetstatina-esnapalisina se obtuvieron a 20°C con una solución de proteína a 7,5 mg/ml en tampón C con soluciones de reservorio 100 mM NaCH<sub>3</sub>COO (acetato de sodio), 10 mM CoCl<sub>2</sub>, 1,2 M HOCH<sub>2</sub>(CH<sub>2</sub>)<sub>4</sub>CH<sub>2</sub>OH (1,6-Hexanediol), pH 4,6 mediante la técnica de microsembrado con Seed Bead (Hampton) y con reservorio 100 mM Na/K fosfato, 41,0% v/v HOCH<sub>2</sub>CH(CH<sub>3</sub>)CH<sub>2</sub>OH (MPD), pH 6,2 y nucleantes de Naomi (Molecular Dimensions) (Saridakis and Chayen 2009).

Los cristales obtenidos fueron crioprottegidos mediante sucesivos pases a través de soluciones de reservorio con cantidades crecientes de glicerol (25-30% [v/v]). Posteriormente fueron difractados a 100K en el ESRF dentro del "BAG Barcelona". Para la recolección de los datos de difracción se utilizaron los detectores MarCCD (línea de luz de sincrotrón ID23-2, sermetstatina nativa y complejo esnapalisina-sermetstatina) y ADSC Q315R CCD (línea ID29, derivado de selenometionina de sermetstatina; ID23-1, complejo subtilisina-sermetstatina). Los datos fueron integrados, escalados, fusionados y reducidos con los programas XDS (Kabsch 2010), XSCALE, y SCALA (Evans 2006) del paquete CCP4 (CCP4 1994).

### 3.6. RESOLUCIÓN DE ESTRUCTURAS Y REFINAMIENTO

---

Se trató de resolver la estructura de sermetstatina mediante reemplazamiento molecular con la función de Patterson (Huber 1965) utilizando las coordenadas de SSI (código PDB 3SSI (Mitsui et al. 1977)) de *S. albogriseolus* como modelo de búsqueda. Finalmente, la estructura fue resuelta por difracción anómala a longitud de onda simple (*Single wavelength Anomalous Diffraction; SAD*) utilizando el derivado con selenometionina de sermetstatina y el programa SHELXD (Sheldrick 2010). Los datos de difracción recogidos a la longitud de onda del pico de absorción del selenio, previamente determinada mediante un escáner de fluorescencia XANES, fueron procesados mediante los programas XDS (Kabsch 2010), XSCALE, y SCALA (Evans 2006). Estos datos permitieron encontrar las tres posiciones de selenio presentes en la unidad asimétrica del cristal con el programa SHELXD (Sheldrick 2010). La obtención de las fases se realizó mediante el programa SHELXE (Sheldrick 2010). Un modelo inicial fue construido en la densidad electrónica resultante con el programa COOT (Emsley et al. 2010) y refinado contra los datos de difracción del derivado de selenometionina con el programa REFMAC5 (Murshudov et al. 2011). Este modelo fue utilizado para determinar la estructura nativa por reemplazamiento molecular con el programa PHASER (McCoy et al. 2007). Posteriormente se completó el modelo manualmente con los programas COOT y TURBO-Frodo (Carranza et al. 1999) alternado con un refinamiento cristalográfico (incluyendo refinamiento *Translation Libration Screw-motion; TLS*) con los programas REFMAC5 y BUSTER/TNT (Blanc et al. 2004).

Las estructuras de los complejo subtilisina-sermetstatina y esnapalisina-sermetstatina fueron resueltas por reemplazamiento molecular con el programa PHASER utilizando como modelo en ambos casos las coordenadas de sermetstatina, además de las coordenadas de subtilisina Carlsberg de *B. licheniformis* (código PDB 1SBC (Neidhart and Petsko 1988)) y esnapalisina de *S. caespitosus* (código PDB 1KUH (Kurusu et al. 1997)), respectivamente. Se procedió a la construcción y refinamiento del modelo de la manera descrita anteriormente para sermetstatina, utilizando restricciones de simetría no cristalográficas automáticas.

Un modelo de homología del complejo ternario entre subtilisina, sermetstatina y esnapalisina fue obtenido con TURBO-Frodo tomando como base los complejos subtilisina-sermetstatina y esnapalisina-sermetstatina después de superponer las regiones del inhibidor de ambos complejos.

### 3.7. EXPERIMENTOS DE ESTABILIDAD DE SERMETSTATINA. PUENTES DISULFURO

---

Sermetstatina silvestre, así como los mutantes C<sup>31</sup>S+C<sup>46</sup>S y C<sup>69</sup>S+C<sup>99</sup>S, fueron incubados por separado con cantidades equimolares de subtilisina y esnapalisina a temperatura ambiente durante 12 h a 0,5 mg/ml en tampón C. Las mezclas de reacción fueron analizadas en busca de cortes en sermetstatina en un espectrómetro de masas Bruker Autoflex III. Cada muestra fue mezclada en una relación 1:1 (v/v) con una solución matriz de ácido sinapínico (10 mg/ml, Sigma) diluido en acetonitrilo 30% y ácido trifluoroacético 0,1% y posteriormente fueron aplicadas en la placa utilizando el método *dried-droplet*. El espectro fue adquirido en modo geométrico lineal con > 1.000 disparos de laser utilizando una mezcla de proteínas en el rango de 5 kDa a 17,5 kDa (Protein Calibration Standard I, Bruker) como calibrador.

### 3.8. MISCELÁNEA

---

Las figuras de las estructuras proteínicas fueron preparadas con los programas TURBO-Frodo, CHIMERA (Pettersen et al. 2004) y PyMOL (Schrodinger 2010). Las interacciones entre las superficies (tomadas como la mitad de la superficie comprendida en la interfase de los complejos) y los contactos próximos (< 4 Å) fueron determinados con CNS (Brünger et al. 1998). Las similitudes entre estructuras fueron analizadas con DALI (Holm et al. 2006). La validación de los modelos fue realizada con MOLPROBITY (Chen. et al. 2010) y la rutina WHATCHECK de WHATIF (Vriend 1990). Para los alineamientos de secuencia y cálculos filogenéticos se utilizó MULTALIN (Corpet 1988).

# 4. RESULTADOS Y DISCUSIÓN

- Pero, oh Jano biforme, ¿qué Dios diré que eres? Porque Grecia ninguna deidad tiene igual a ti. Di asimismo la causa por qué solo tú de los Dioses ves delante y a la espalda.

Publio Ovidio Nasón (siglo I)

*Fastos*





## 4.1. PRODUCCIÓN Y PURIFICACIÓN DE SERMETSTATINA

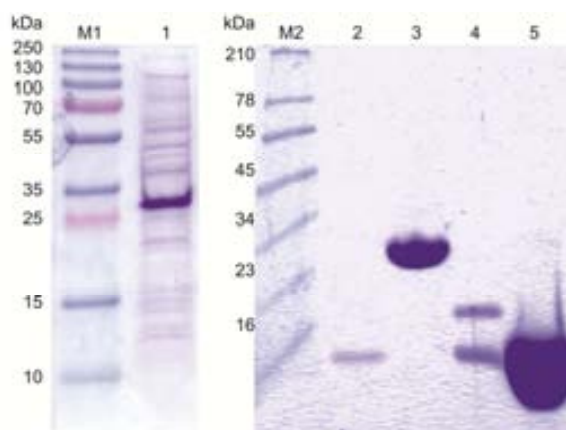
El análisis de péptido señal mediante el servidor SignalP 4.0 (Petersen et al. 2011) reveló la presencia de un péptido señal en sermetstatina correspondiente a los residuos M<sup>1</sup>-A<sup>28</sup>, confirmando resultados anteriores (Hiraga et al. 2000). La clonación en un vector pET-32a modificado de un gen sintético codificante para sermetstatina sin los residuos del péptido señal (Figura R.1) permitió su expresión en *E. coli*. El diseño de dicho gen sintético con codones optimizados para su expresión en este organismo permitió optimizar su rendimiento, evitando el problema del sesgo en el uso de codones (Ikemura 1981).



**Figura R.1.** Gen sintético de sermetstatina. **a)** Secuencia nucleotídica del gen sintético de sermetstatina y traducción proteínica (ExPASy Translate tool; <http://web.expasy.org/translate>). **b)** Análisis de uso de codones del gen sintético de sermetstatina en *Escherichia coli*. Las barras del gráfico representan el porcentaje de adaptabilidad relativa (Sharp and Li 1987) de cada uno de los codones (Graphical codon usage analyzer; <http://gcu.schoedl.de/index.html>).

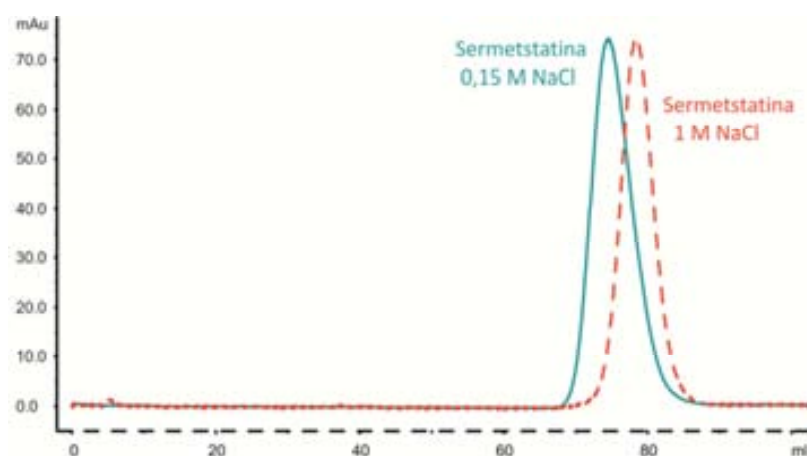
La expresión heteróloga de la proteína de fusión tiorredoxina-His(6x)-sermetstatina en células *E. coli* Origami2 (DE3) resultó en la obtención 13 mg de proteína soluble purificada por litro de cultivo (Figura R.2). La utilización de la proteína de fusión tiorredoxina y de las células Origami2 favorece la creación de un ambiente reductor durante la expresión intracelular, lo que permite la formación de los puentes disulfuro, y por tanto, la obtención de la proteína en forma soluble. En este caso se hizo posible la formación de los dos puentes disulfuro presentes en sermetstatina, C<sup>31</sup>-C<sup>46</sup> y C<sup>69</sup>-C<sup>99</sup>.

Las seis histidinas añadidas al extremo N-terminal de sermetstatina permitieron la purificación de la fusión por cromatografía de afinidad a Ni<sup>2+</sup> en un proceso eficiente y notablemente específico. El corte con peptidasa TEV resultó también muy eficiente obteniéndose rendimientos cercanos al 100% en las condiciones utilizadas. El corte se realizó en presencia de concentraciones adecuadas de agentes redox que permitieron el mantenimiento de los puentes disulfuro nativos de sermetstatina. La purificación por cromatografía de exclusión molecular en una columna HiLoad 16/60 Superdex 75 permitió la purificación final de sermetstatina G<sup>-1</sup>-F<sup>113</sup> hasta un nivel superior al 95%. Finalmente, el inhibidor pudo ser concentrado hasta los 10 mg/ml.



**Figura R.2.** Tricina-SDS-PAGE 15% de sermetstatina. **M1.** Marcador *PageRuler Plus Prestained Ladder* (Fermentas). **M2.** Marcador *SeeBlue Pre-Stained Standard* (Invitrogen). **1.** Fracción soluble de la expresión de tiorredoxina-His(6x)-sermetstatina (28 kDa) en células *Escherichia coli* Origami2 (DE3). **3.** Purificación de tiorredoxina-His(6x)-sermetstatina por cromatografía de afinidad a  $\text{Ni}^{2+}$ . **4.** Digestión con peptidasa TEV. La banda superior corresponde a tiorredoxina (16 kDa) y la inferior a sermetstatina (12 kDa). **2 y 5.** Sermetstatina tras la cromatografía de exclusión molecular, aplicada en diferentes cantidades. El carril 2 permite apreciar la homogeneidad de la banda de proteína, mientras que en el 5 se observa la pureza de la muestra.

Por otro lado, la calibración previa de la columna de exclusión molecular permitió determinar que el inhibidor adopta una estructura cuaternaria dimérica en solución. La interacción entre las subunidades del dímero se estabiliza a bajas concentraciones de NaCl, de manera que entre 50 mM y 500 mM únicamente se observa la forma dimérica por cromatografía de exclusión molecular. A concentraciones superiores a 500 mM se empieza a observar un equilibrio entre la forma monomérica y dimérica, hasta llegar a 1 M, donde únicamente se observa la forma monomérica (Figura R.3).



**Figura R.3.** Cromatograma de exclusión molecular de sermetstatina. Superposición de los cromatogramas de sermetstatina en tampón C (0,15 M NaCl; volumen de elución ( $V_e$ ) = 73,0 ml [20 kDa]; línea cian continua) y J (1 M NaCl;  $V_e$  = 79,2 ml [12 kDa]; línea roja discontinua), donde se observa la variación en su estructura cuaternaria en función de la fuerza iónica del medio. Columna Superdex 75 16/60 calibrada con aldolasa ( $V_e$  = 53,9 ml [158 kDa]), conalbúmina ( $V_e$  = 61,5 ml [75 kDa]), ovalbúmina ( $V_e$  = 67,2 ml [43 kDa]) y citocromo C ( $V_e$  = 80,2 ml [12 kDa]). Absorbancia medida a  $\lambda$  = 280 nm.

## 4.2. OBTENCIÓN DE ESNA PALISINA Y PROESNA PALISINA. ANÁLISIS DEL PRODOMINIO

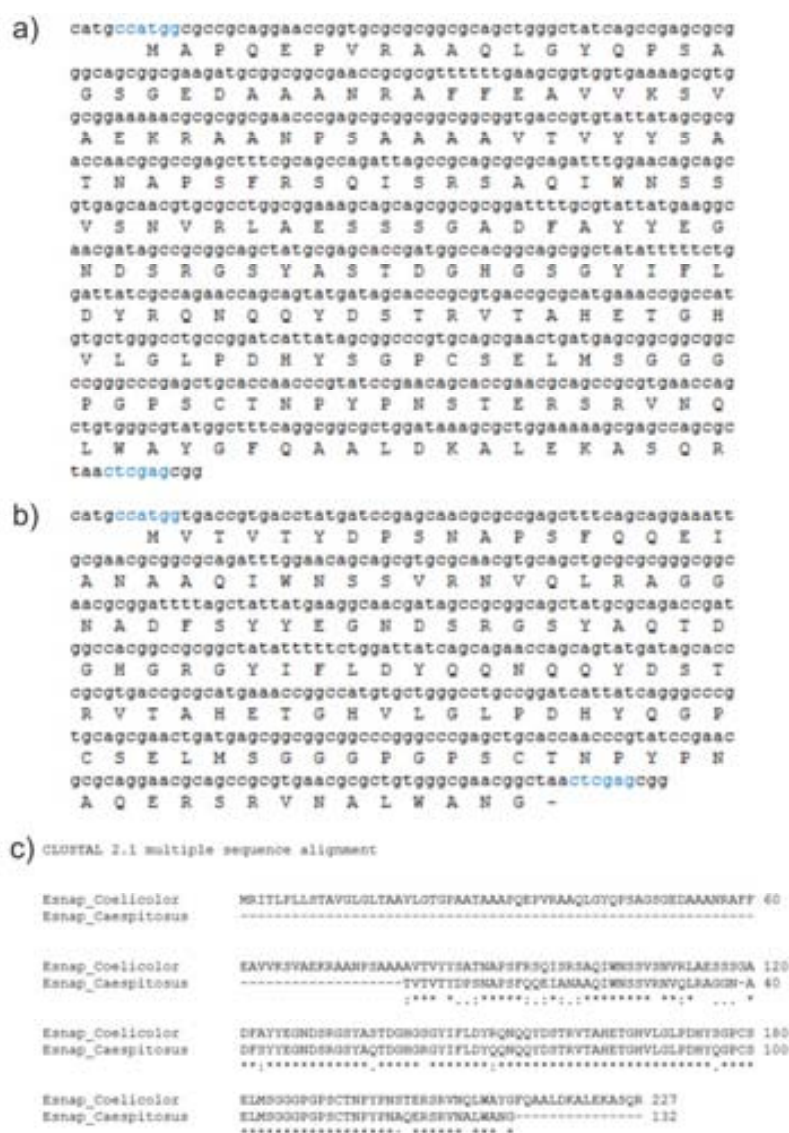
---

La mayor parte de los estudios previos publicados sobre esnapalisina se realizaron partiendo del liofilizado del medio extracelular de *S. caespitosus* (*dark brown lyophilized powder of the crude filtrate of S. caespitosus*) comercializado por la empresa Kyowa Hakko Kogyo (Harada et al. 1991; Harada et al. 1995; Inouye et al. 2007). Esta empresa, actualmente denominada Kyowa Hakko Kirin, ya no distribuye este producto, por lo que se tuvieron que buscar alternativas para la obtención de esnapalisina. En un primer momento se intentó purificar esnapalisina a partir del cultivo de *S. coelicolor*. Las condiciones para su expresión en *S. caespitosus* fueron previamente descritas en una publicación en japonés sin posibilidad de acceso de forma electrónica (Yokote 1969), por lo que fue necesaria la realización de experimentos en distintas condiciones de temperatura y diferentes medios de cultivo. Estudios de actividad con sustrato BODIPY FL caseína y DQ gelatina determinaron indirectamente la presencia de esnapalisina en varias de las condiciones estudiadas. El análisis del medio extracelular del cultivo de *S. coelicolor* crecido en LB a 37°C tras 72 h de crecimiento determinó que el 86% de la actividad proteolítica era inhibida por EDTA y el 47% por sermetstatina. El importante descenso de la actividad al añadir sermetstatina solo sería explicable por la presencia de esnapalisina en la muestra. Los resultados de los estudios de inhibición de sermetstatina (apartado 4.3) y la inhibición por EDTA, que muestra que la mayor parte de la actividad proteolítica sobre el sustrato utilizado es causada por metalopeptidasas, sugieren la presencia de esnapalisina. Estos resultados sugieren la existencia de esnapalisina en *S. coelicolor*, donde había sido inferida por homología, pero indican que su expresión en *S. caespitosus* es probablemente más abundante, ya que las cantidades expresadas fueron muy pequeñas. Estas cantidades no fueron suficientes para abordar estudios posteriores de estructura y función.

Así pues, se optó por la expresión heteróloga de esnapalisina en *E. coli* a partir de genes sintéticos optimizados para la expresión en este organismo. Esta opción presentaba la ventaja de permitir el estudio del prodominio de esta peptidasa, ya que no se ha descrito la presencia del zimógeno intacto en el medio de cultivo de *Streptomyces*. Además hacía posible la obtención de mutantes funcionales útiles para este estudio y otros futuros. Por tanto, en lugar de tratar de cultivar *S. caespitosus* y purificar esnapalisina del medio de cultivo, tal como se hizo para *S. coelicolor*, se procedió a desarrollar los protocolos de expresión heteróloga en *E. coli* de esnapalisina de *S. caespitosus* y proesnapalisina de *S. coelicolor*.

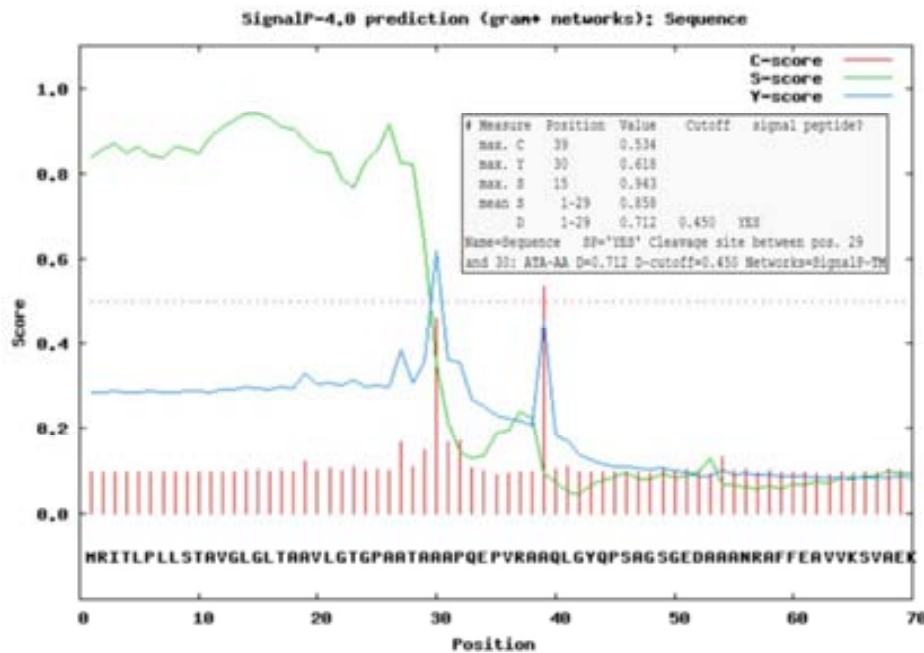
### 4.2.1. ANÁLISIS BIOINFORMÁTICO

El análisis bioinformático de las secuencias aminoacídicas de esnapalisina de *S. caespitosus* y *S. coelicolor* reveló un 79,4% de identidad entre ambas proteínas (Figura R.4). Pese a no haberse descrito evidencias a nivel proteínico de la existencia de esnapalisina en *S. coelicolor* en la literatura científica, el alto grado de homología hace plausible su existencia y permite su utilización para el estudio del prodominio y activación de esta familia de peptidasas. Además, los estudios de expresión de esnapalisina en *S. coelicolor* descritos en el apartado anterior indicaron su presencia en el medio extracelular reforzando esta hipótesis.



**Figura R.4.** Análisis bioinformático de esnapalisina. Secuencia nucleotídica del gen sintético y traducción proteínica de **a)** proesnapalisina de *Streptomyces coelicolor* y **b)** esnapalisina de *S. caespitosus*. En azul se indican los sitios de restricción utilizados en el clonaje (*NcoI* y *XhoI*). **c)** Alineamiento de las secuencias proteínicas de esnapalisina de *S. caespitosus* y proesnapalisina de *S. coelicolor* mediante Clustal 2.1. (\*) Posición totalmente conservada. (:) Posición conservada entre residuos muy similares. (.) Posición conservada entre residuos ligeramente similares.

La predicción de péptido señal en la secuencia de esnapalisina de *S. coelicolor* mediante el servidor SignalP 4.0 (Petersen et al. 2011) reveló la existencia de uno correspondiente a los residuos M1-A29 (Figura R.5).

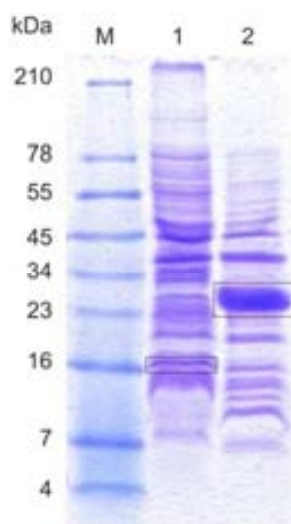


**Figura R.5.** Predicción de péptido señal en esnapalisina. El resultado gráfico ofrecido por el servidor SignalP 4.0 (Petersen et al. 2011) incluye tres puntuaciones diferentes: *C-score*, *S-score* y *Y-score*, más dos puntuaciones adicionales ofrecidas únicamente en forma numérica: *S-mean* y *D-score*. El análisis de todas ellas permite la predicción de un péptido señal entre los aminoácidos 1 y 29.

La existencia de péptido señal es consistente con el hecho de que se trata de una proteína de secreción extracelular. Para su expresión en *E. coli* se procedió al diseño del gen de proesnapalisina sin la secuencia del péptido señal.

#### 4.2.2. ENSAYOS DE EXPRESIÓN Y SOLUBILIDAD DE PROESNAPALISINA Y ESNAPALISINA

El análisis de expresión de proesnapalisina silvestre y proesnapalisina nativa E164A para los diferentes vectores utilizados mostró elevados niveles de expresión para todos ellos a excepción de los que presentaban una señal de exportación periplasmática y una His(6x) C-terminal. En contraste, el rendimiento de expresión de esnapalisina madura resultó notablemente inferior (~20 veces) utilizando los mismos vectores. Por otro lado, se determinó que las células *E. coli* BL21 (DE3) mostraban niveles de expresión superiores al resto, siendo óptima la expresión a 37°C durante 5 horas tras la inducción con 1 mM IPTG a una OD<sub>550</sub> de 0,8 (Figura R.6).



**Figura R.6.** Tricina-SDS-PAGE 15% de la expresión de esnapalisina y proesnapalisina. Expresión en células *Escherichia coli* BL21 (DE3). Las proteínas de interés se muestran recuadradas. Su identidad fue confirmada mediante análisis de huella peptídica (*peptide mass fingerprint*) **M.** Marcador *SeeBlue Pre-Stained Standard* (Invitrogen). **1.** His(6x)-esnapalisina (16,8 kDa). **2.** His(6x)-proesnapalisina (23,5 kDa).

El estudio de solubilidad mostró que para todos los vectores de expresión en todas las cepas de *E. coli* DE3 la proteína expresada era totalmente insoluble. Variaciones en la temperatura de expresión, medio de cultivo, concentraciones de inducción con IPTG y tampones de lisis se mostraron ineficaces a la hora de incrementar la solubilidad. En consecuencia se procedió al desarrollo de protocolos de purificación de cuerpos de inclusión y replegamiento proteínico. Para ello se partió de los vectores His(6x) N-terminal, utilizando las condiciones óptimas de expresión determinadas anteriormente.

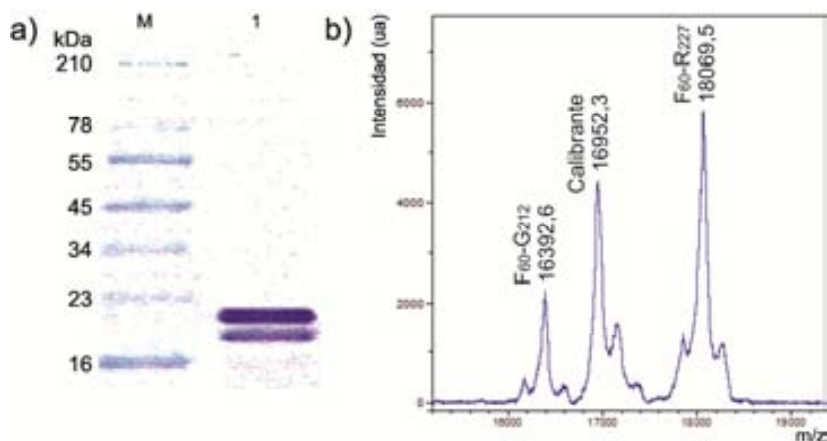
#### 4.2.3. PURIFICACIÓN DE ESNAPALISINA

El protocolo de expresión, purificación y replegamiento de la forma madura de esnapalisina permitió la obtención de la misma en cantidades muy bajas (~50 µg por litro de cultivo), pero suficientes para la realización de ensayos de actividad proteolítica y de activación de proesnapalisina. Las causas del escaso rendimiento del proceso son los relativamente bajos niveles de expresión y principalmente las pérdidas durante el proceso de replegamiento, siendo estas superiores al 98%.

Durante el proceso de replegamiento se produce un corte autolítico en el aminoácido anterior al sitio de corte de la peptidasa TEV, entre los residuos  $F_{-4}$  y  $Q_{-3}$ , en una región expuesta y compatible con la afinidad de esnapalisina. De esta manera se elimina la cola de histidinas obteniendo esnapalisina con un segmento N-terminal  $Q_{-3}+G_{-2}+P_{-1}$ .

#### 4.2.4. PURIFICACIÓN Y ESTUDIOS DE ACTIVACIÓN DE PROESNAPALISINA

Tras el aislamiento de los cuerpos de inclusión, replegamiento y purificación de proesnnapalisina silvestre se obtuvo un rendimiento de 2 mg a una concentración de 5 mg/ml partiendo de un litro de cultivo, gracias a los elevados niveles de expresión y a un rendimiento en el replegamiento superior al 55%. El estudio del proceso de activación indicó que proesnnapalisina silvestre sufre un proceso rápido de activación autolítica *in-vitro* a la concentración resultante del proceso de purificación, no siendo posible aislar el zimógeno intacto. Tras este proceso se observan dos bandas mayoritarias en Tricina-SDS-PAGE 15% (Figura R.7) correspondientes al corte *F59-F60* (*F60-R227*; 18.062,5 Da) y *F59-F60 G212-F213* (*F60-G212*; 16.404,6 Da). La especificidad de estos cortes corresponde a la descrita en estudios previos para esnapalisina (Inouye et al. 2007). Por tanto, en el proceso de activación se produciría un primer corte autolítico con gran eficiencia entre *F59-F60* eliminando el extremo N-terminal y posteriormente otro menos eficiente entre *G212-F213* que eliminaría 15 residuos del extremo C-terminal. Este mecanismo de autoactivación recuerda al descrito previamente para la metalopeptidasa ulilisina de *Methanosarcina acetivorans* (Tallant et al. 2006).



**Figura R.7.** Activación de proesnnapalisina. **a)** Tricina-SDS-PAGE 15%. **M.** Marcador *SeeBlue Pre-Stained Standard* (Invitrogen). **1.** Muestra de proesnnapalisina silvestre de *Streptomyces coelicolor* tras el proceso de replegamiento y purificación. **b)** Espectro de masas de proesnnapalisina purificada realizado en un MALDI-TOF Mass Spectrometer Autoflex Speed (Bruker). Calibración interna con citocromo c ( $m/z$  12.361 Da; no se muestra en la figura) y mioglobina ( $m/z$  16.952 Da).

La incubación posterior de la proteína a 37°C durante 3-4 días mostró que se producen diversos cortes en la parte restante de 49 residuos del prodominio para finalmente obtener la forma madura de 133 residuos, A80-G212. La lentitud con la que se producen dichos cortes, la no existencia de residuos aromáticos en el fragmento y el hecho de que la



activación no se inhiba con EDTA (1-5 mM) y que la adición de esnapalisina madura de *S. caespitosus* (1-10  $\mu$ M) no la favorezca, hace suponer que estos cortes son producidos por peptidasas contaminantes provenientes de *E. coli* presentes en la muestra. Por tanto, es de esperar que *in-vivo* otro tipo de peptidasas, probablemente serán peptidasas, asistan a esnapalisina en su proceso de activación, tal y como ha sido descrito para otras muchas peptidasas (Appel 1986; Goulas et al. 2011; Yiallourous et al. 2002).

#### **4.2.5. PURIFICACIÓN Y ESTUDIOS DE ACTIVACIÓN DE PROESNAPALISINA E164A**

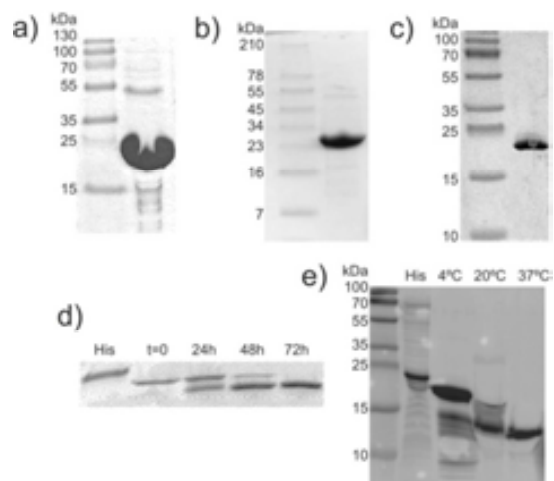
Tras el aislamiento de los cuerpos de inclusión, replegamiento y purificación de proesnapalisina E164A se obtuvo un rendimiento de 5 mg a una concentración de 5 mg/ml partiendo de un litro de cultivo, gracias a los elevados niveles de expresión y a un rendimiento en el replegamiento superior al 55% (Figura R.8). Su purificación y posterior digestión con peptidasa TEV resultó en la obtención de una única banda en Tricina-SDS-PAGE 15% correspondiente a la proteína completa G(-2)-R227 (21.260,0 Da). Tras la incubación durante 24 horas a 37°C se observó un corte entre Y43 y Q44 (19.756,3 Da). Pasados tres días la totalidad de la proteína apareció cortada en esta posición.

La forma resultante presentó una gran estabilidad, llegando a poder ser concentrada hasta un máximo de 20 mg/ml en contraposición a los 5 mg/ml de la forma completa, de manera que la gran estabilidad del producto resultó en un corte energéticamente muy favorable. La mutación del ácido/base general del mecanismo catalítico en ocasiones no inhibe completamente la actividad de las metalopeptidasas. Este hecho se ha reportado en peptidasas como profragilisina (Goulas et al. 2011) o prouilisina (información no publicada) donde la mutación ralentiza su maduración pero no evita que esta se produzca en determinadas condiciones. Por tanto, la actividad remanente del mutante E164A sería capaz de catalizar la hidrólisis del enlace entre Y43 y Q44. La existencia de un residuo aromático en posición P1' y la inhibición del corte en presencia de EDTA apoyan la hipótesis del corte autolítico.

Por otro lado, la actividad proteolítica del mutante E164A es insuficiente para producir los cortes F59-F60 y G212-F213, de manera que la forma obtenida se mantiene intacta durante días de incubación a 37°C. Además, los fragmentos Q44-F59 y F213-R227 aportarían a la proteína resistencia contra la acción de peptidasas contaminantes, incrementando su estabilidad.



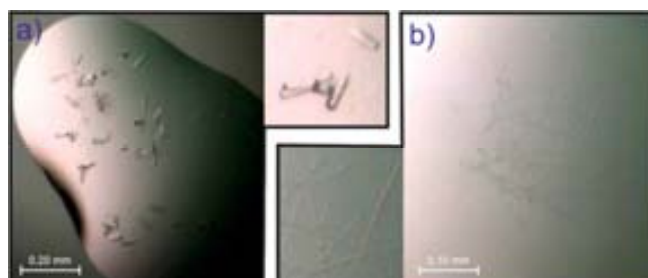
No obstante, tras su incubación durante 4 semanas a 1 mg/ml y 37°C se observó una banda mayoritaria correspondiente a la forma madura, en un proceso muy lento en comparación con la proenzima silvestre (Figura R.8). Este proceso sería catalizado por trazas de peptidasas contaminantes provenientes de *E. coli*.



**Figura R.8.** Tricina-SDS-PAGE 15% de proenkefalina E164A. Marcadores *PageRuler Plus Prestained Ladder* (Fermentas; a, c y e) y *SeeBlue Pre-Stained Standard* (Invitrogen; b). En d) y e) se utiliza proenkefalina mutante con cola de histidinas (His) como referencia. **a)** Purificación de cuerpos de inclusión de proenkefalina mutante expresada en *Escherichia coli* BL21 (DE3) a 37°C durante 5 h. **b)** Proenkefalina mutante tras el proceso de repliegamiento. **c)** Proenkefalina mutante digerida con peptidasa TEV y purificada por cromatografía de exclusión molecular. **d)** Evolución de proenkefalina mutante G(-2)-R227 incubada a 37°C. **e)** Evolución de proenkefalina mutante Q44-R227 incubada durante un mes a diferentes temperaturas.

#### 4.2.6. CRISTALIZACIÓN DE PROENKEFALISINA E164A

Los ensayos de cristalización de proenkefalina E164A de *S. coelicolor* se llevaron a cabo con el fragmento Q44-R227 tal y como se detalla en el apartado 3.2.5. Los mejores cristales aparecieron después de tres días a 4°C a una concentración de 6 mg/ml y 0,5 M fosfato de amonio, 0,2 M citrato de trisodio, pH 5,6 como reservorio (Figura R.9b).



**Figura R.9.** Cristalización de proenkefalina E164A (segmento Q44-R227). Gotas de 1 µl (proteína) + 1 µl (reservorio). **a)** Proenkefalina mutante a 20 mg/ml en 20mM HEPES, 30% PEG 8.000, pH 7,5 como reservorio a 20°C. Da lugar a cristales de la forma madura. **b)** Proenkefalina mutante a 6 mg/ml en 0,5 M fosfato de amonio, 0,2 M citrato de trisodio, pH 5,6. 4°C.

La integridad del fragmento de proesnapalisina mutante en estos cristales fue confirmada por Tricina-SDS-PAGE y tinción con plata. No obstante, estos no difractaron por debajo de 6 Å y estaban desordenados, no siendo posible la resolución de la estructura. Pasadas 4 semanas aparecieron cristales en numerosas condiciones en placas incubadas a 20°C, pero se observó que estos correspondían a la forma madura de esnapalisina (Figura R.9a).

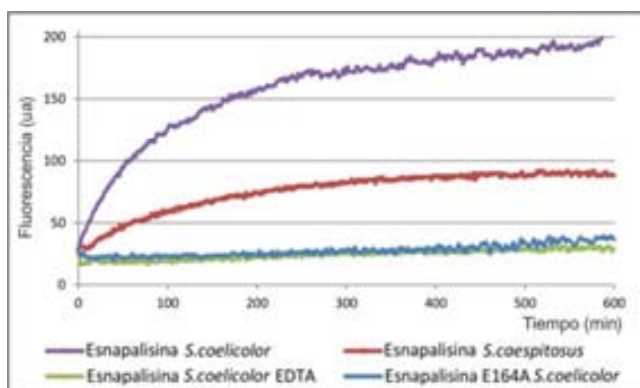
Los intentos de optimización mediante variaciones en la concentración de proteína, relación proteína/reservorio en la gota y cribaje con aditivos (Additive Screen [Hampton Research]) no aportaron resultados positivos. En general, se obtuvieron cristales de proesnapalisina durante los primeros días que no difractaban o lo hacían pobremente, para posteriormente dar lugar a la aparición de cristales de la forma madura. La adición de inhibidores de peptidasas a las gotas y la incubación de las placas de cristalización a 4°C consiguieron frenar la activación, pero no fue posible la obtención de cristales de proesnapalisina que difractaran dentro de los límites adecuados para la resolución de su estructura.

La maduración del fragmento Q44-R227 de proesnapalisina mutante en las condiciones de cristalización puede deberse, tal y como se ha comentado anteriormente, a la presencia de trazas de peptidasas contaminantes provenientes de *E. coli*. Las elevadas concentraciones de proteína utilizadas para su cristalización podrían potenciar la actividad de dichas peptidasas produciendo la maduración del fragmento Q44-R227 más rápidamente que en los estudios realizados previamente, donde se determinó que era necesaria la incubación a 37°C durante 4 semanas para obtener la forma madura. Para la cristalización se tuvieron que utilizar concentraciones seis veces superiores, de manera que posiblemente se aceleró el proceso de maduración aún a 20°C. La inestabilidad del zimógeno mutante pudo ser la causa de la obtención de cristales pequeños y de baja calidad que no permitieron la resolución de su estructura.

#### **4.2.7. ESTUDIOS DE ACTIVIDAD**

La actividad proteolítica de esnapalisina madura de *S. caespitosus* fue estudiada frente substratos de caseína y gelatina fluorescentes obteniendo claros perfiles de actividad (Figura R.10). Se observó que la actividad proteolítica se inhibía un 99% con EDTA (1 mM) y 1,10-fenantrolina (0,1 mM).

Por tanto, se puede concluir que el proceso de replegamiento y purificación, pese a resultar ineficiente desde el punto de vista del rendimiento, permite obtener esnapalisina correctamente plegada y activa.



**Figura R.10.** Actividad proteolítica de proesnapalisina y esnapalisina. Experimentos realizados frente a caseína fluorescente BODIPY FL (65 µg/ml) a 37°C. **Verde y Lila)** Fragmentos resultantes de la activación de proesnapalisina de *Streptomyces coelicolor* a una concentración final de 1 µM con y sin 1 mM EDTA, respectivamente. **Azul)** Proesnapalisina E164A Q44-R227 a una concentración final de 1 µM. **Rojo)** Esnapalisina de *S. caespitosus* a una concentración final de 0,3 µM.

Las formas proteínicas purificadas a partir del replegamiento de proesnapalisina silvestre de *S. coelicolor* también mostraron actividad contra gelatina y caseína susceptible de inhibición con EDTA y 1,10-fenantrolina (Figura R.10). No se pudo discriminar si la actividad observada provenía de los fragmentos F60-R227 o F60-G212, o bien de trazas de la forma madura presentes en la muestra. No obstante, asumiendo que esnapalisina de *S. caespitosus* y de *S. coelicolor* presenten niveles de actividad similares, la relación entre concentración de peptidasa y actividad indica que ambos fragmentos pueden tener actividad.

Por otro lado, no se observó actividad proteolítica significativa en las condiciones ensayadas para proesnapalisina mutante E164A en ninguna de sus formas, incluyendo la forma madura. La baja concentración de proteína a la que se realizó el estudio en comparación a la utilizada en ensayos de estabilidad o de cristalización, así como el tiempo relativamente corto analizado, hacen que no se observen trazas de actividad significativa ni para proesnapalisina mutante E164A ni para proesnapalisina nativa con EDTA. Por tanto, los resultados serían compatibles con la hipótesis de la activación exógena por parte de peptidasas contaminantes y de la actividad del mutante, ya que se trataría de actividades residuales no detectables en este tipo de experimentos.

### 4.3. ANÁLISIS DE PROTEÓLISIS Y SU INHIBICIÓN POR SERMETSTATINA

Los análisis de inhibición frente a varias metalopeptidasas del clan de las metzincinas por parte de sermetstatina confirmaron la especificidad del inhibidor por esnapalisina. En los experimentos de inhibición de astacina, ulilisina, aeruginolisina, fragilisina, ADAM-17, MMP-1, MMP-3 y MMP-13 se observó como la adición de sermetstatina en exceso, en una relación molar enzima:inhibidor de hasta 1:100, no alteraba la actividad proteolítica de las peptidasas. En contraposición, la inhibición de esnapalisina en presencia de cantidades equimolares de sermetstatina era total (Tabla R.1).

Tabla R.1. Actividad inhibitoria de sermetstatina.

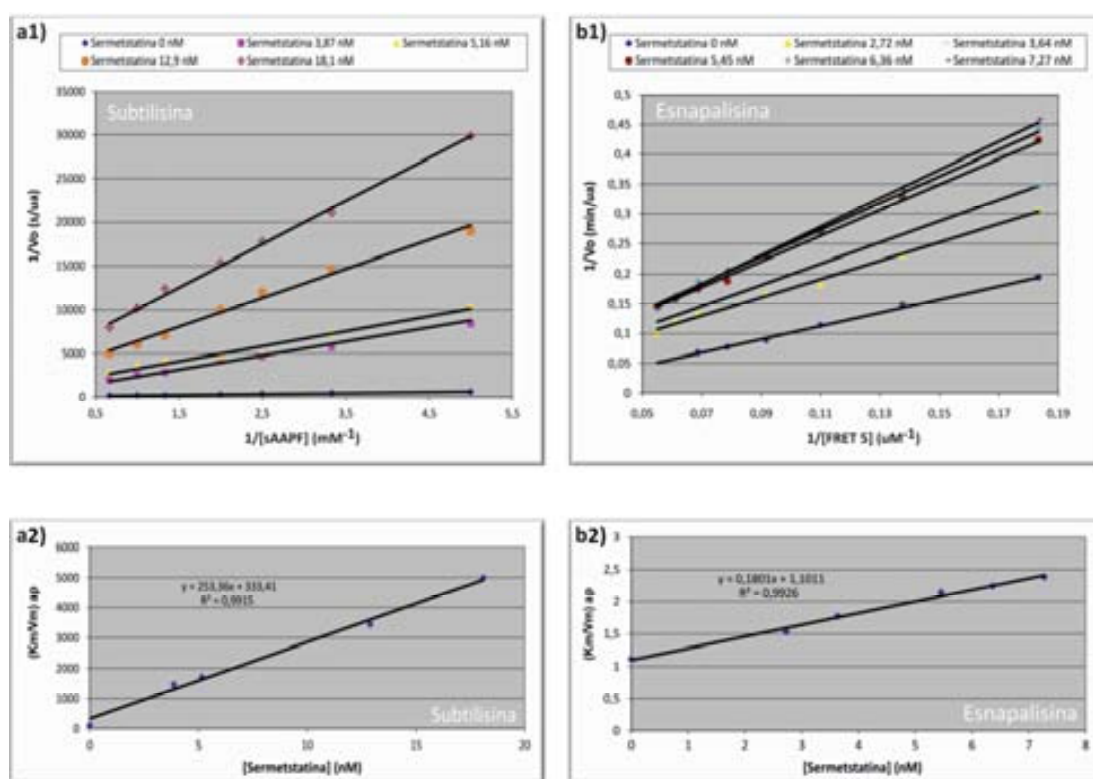
Peptidasa	Relación molar (Inh./Enz.)	Actividad Inhibitoria* (%)
<b><i>Metalopeptidasas</i></b>		
Esnapalisina	1	99
	10	100
Astacina	100	0
Ulilisina	100	0
Aeruginolisina	100	0
Fragilisina	100	0
ADAM <sup>a</sup> -17	-	NI <sup>c</sup>
MMP <sup>b</sup> -1,-3,-13	100	0
<b><i>Serín peptidasas</i></b>		
Subtilisina	1	95
Tripsina	10	21
Quimotripsina	10	69
Elastasa	10	68
Proteinasa K	1	82
	10	100

\* Valores correspondientes a tres mediciones independientes con una desviación estándar de  $\pm 5$ . <sup>a</sup> Desintegrina y metalopeptidasa. <sup>b</sup> Metalopeptidasa de matriz. <sup>c</sup> No se observa inhibición a una concentración de sermetstatina 4  $\mu$ M en un experimento basado en células.

Por otro lado, los experimentos de inhibición sobre serín peptidasas mostraron que sermetstatina es un potente inhibidor, además de subtilisina BPN' de *B. amyloliquefaciens* (Hiraga et al. 2000), de subtilisina Carlsberg de *B. licheniformis* y proteinasa K de *E. album* (Tabla R.1), todas ellas muy próximas en términos evolutivos, funcionales y estructurales.

Se observó también una notable inhibición de la actividad de quimotripsina y elastasa en condiciones de exceso moderado de sermetstatina. Por el contrario, únicamente se apreció una ligera inhibición en el caso de tripsina. Por tanto, en comparación a la zona reactiva contra metalopeptidasas, el sitio reactivo para serín peptidasas presenta una menor selectividad. Sermetstatina, respecto a esta cara inhibitoria, es un potente inhibidor de amplio espectro de serín peptidasas tipo subtilisina y, parcialmente, tipo quimotripsina.

Los experimentos para el cálculo de las constantes de inhibición de sermetstatina frente a subtilisina y esnapalisina revelaron unas  $K_i = 1,3 \pm 0,3$  nM y  $6,1 \pm 0,5$  nM, respectivamente (Figura R.11), demostrando la elevada afinidad del inhibidor para ambas peptidasas.



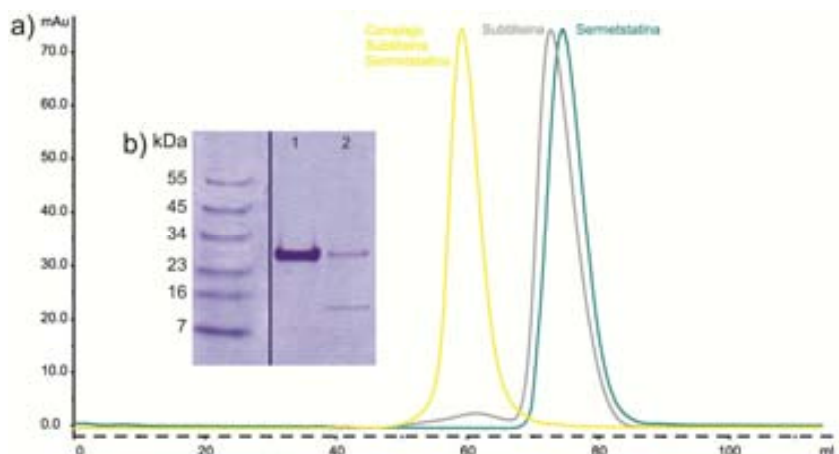
**Figura R.11.** Constantes de inhibición de sermetstatina frente a subtilisina y esnapalisina. **a)** Enzima: Subtilisina. Substrato: Succinil-Ala-Ala-Pro-Phe-*p*-nitroanilida (sAAPF). **b)** Enzima: Esnapalisina. Substrato: Aminobenzoil-Val-Lys-Phe-Tyr-Asp-Ile-Lys(2,4-dinitrofenilamino) (FRET 5). **1)** Diagrama de Lineweaver-Burk para las diferentes concentraciones de sermetstatina analizadas. Las pendientes de las rectas corresponden a la constante de Michaelis-Menten dividida por la velocidad máxima ( $K_m/V_m$ ) para cada concentración. **2)** Representación de  $K_m/V_m$  aparente frente a la concentración de inhibidor. La intersección de la recta con el eje de las abscisas corresponde al opuesto de la constante de inhibición ( $-K_i$ ).

## 4.4. FORMACIÓN Y PURIFICACIÓN DE COMPLEJOS

### 4.4.1. SUBTILISINA-SERMETSTATINA

Aunque subtilisina Carlsberg comercial presenta un elevado nivel de pureza, se realizó una purificación por cromatografía de exclusión molecular para eliminar los productos de degradación de bajo peso molecular así como las sales presentes en el liofilizado. De esta manera se obtuvo subtilisina pura reconstituida en un tampón de composición conocida (ver apartado 3.4.1). Tras la cromatografía, la concentración de la peptidasa pudo ser determinada con exactitud por absorbancia a  $\lambda = 280$  nm, condición necesaria para la formación del complejo con sermetstatina en la relación estequiométrica adecuada. La presencia de productos de degradación y de sales de composición desconocida en la muestra interferirían de manera muy significativa en la determinación de la concentración.

La incubación de subtilisina Carlsberg con sermetstatina permitió la obtención de un complejo estable en las condiciones analizadas. Su posterior purificación por cromatografía de exclusión molecular en una columna HiLoad 16/60 Superdex 75, previamente calibrada, permitió determinar que el complejo adopta una estructura cuaternaria en forma de heterotetrámero 2+2, en la cual un dímero de inhibidor interacciona con dos moléculas de subtilisina (Figura R.12).



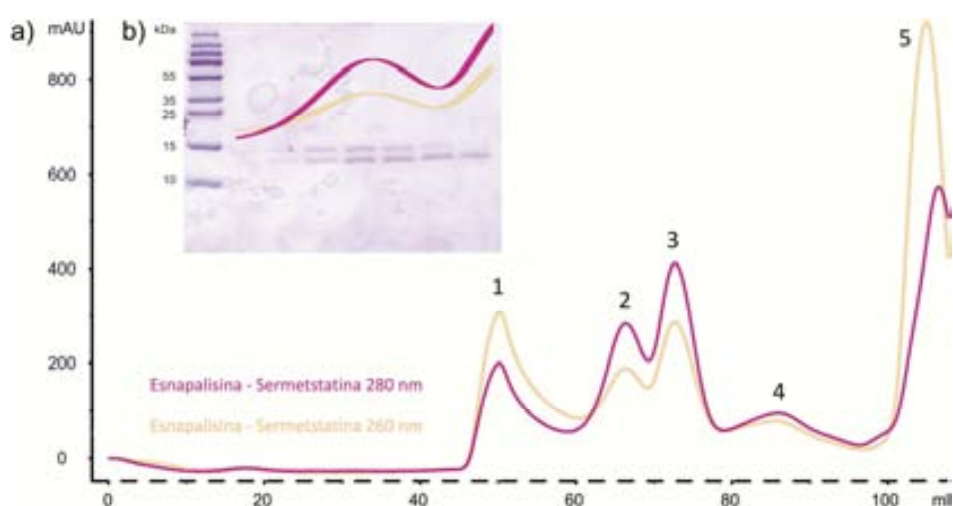
**Figura R.12.** Complejo subtilisina-sermetstatina **a)** Superposición de los cromatogramas de exclusión molecular de sermetstatina pura (volumen de elución ( $V_e$ ) = 73,0 ml [20 kDa]; línea verde), subtilisina pura ( $V_e$  = 71,2 ml [28 kDa]; línea gris) y complejo subtilisina-sermetstatina ( $V_e$  = 59,9 ml [85 kDa]; línea amarilla), donde se aprecia el desplazamiento de los picos cromatográficos confirmando la formación del complejo. Columna Superdex 75 16/60 calibrada con aldolasa ( $V_e$  = 53,9 ml [158 kDa]), conalbúmina ( $V_e$  = 61,5 ml [75 kDa]), ovalbúmina ( $V_e$  = 67,2 ml [43 kDa]) y citocromo C ( $V_e$  = 80,2 ml [12 kDa]). Absorbancia medida a  $\lambda = 280$  nm. **b)** Tricina-SDS-PAGE 15%. Marcador *SeeBlue Pre-Stained Standard* (Invitrogen). La muestra 1 corresponde al pico de subtilisina (27 kDa) y la 2 al pico de complejo subtilisina-sermetstatina ((27 kDa+12 kDa)x2).

#### 4.4.2. ESNAPALISINA-SERMETSTATINA

Los resultados del apartado 4.2.3 han demostrado que el proceso de obtención de esnapalisina de *S. caespitosus* en *E. coli* es muy poco eficiente debido al bajo rendimiento de expresión y sobretodo de replegamiento, obteniéndose cantidades de esnapalisina insuficientes para abordar estudios estructurales. Por otro lado, la activación de proesnapalisina de *S. coelicolor* para obtener peptidasa activa resultó ser un proceso poco eficiente y reproducible (apartado 4.2.4). Para resolver este problema se procedió a optimizar el proceso de replegamiento de esnapalisina de *S. caespitosus*. Para ello, dado que el replegamiento de proesnapalisina es un proceso eficiente debido al efecto chaperona del prodominio, se intentó mimetizar este efecto sobre la forma madura de esnapalisina mediante la adición de sermetstatina en el tampón de replegamiento. La interacción del inhibidor con esnapalisina estabiliza la enzima y guía su replegamiento, tal y como se describió anteriormente para el replegamiento de tripsina en presencia de un inhibidor sintético (Nohara et al. 2004) y de un mutante de SSI con afinidad modificada por tripsina (Nohara et al. 1999). De esta manera se consiguió un incremento notable del rendimiento del proceso de replegamiento, pasando del ~2% sin inhibidor al ~30% con inhibidor. Durante el proceso se produce el mismo corte entre los residuos F(-4) y Q(-3) observado en el replegamiento sin inhibidor (apartado 4.2.3). En las condiciones de replegamiento, debido a la elevada fuerza iónica del medio y a la presencia de agentes caotrópicos (0,8 M urea proveniente del tampón de solubilización de los cuerpos de inclusión), la interacción entre inhibidor y enzima no es suficientemente estable como para inhibir totalmente a esnapalisina, pudiendo ocurrir el corte autolítico. Este corte hace innecesaria la digestión con peptidasa TEV y evita pasos de purificación adicionales, evitando sus pérdidas asociadas y por tanto incrementando el rendimiento del proceso.

Tras el replegamiento, al aplicar la muestra en una columna HiLoad 16/60 Superdex 75 previamente calibrada se obtuvieron un total de cinco picos, uno de ellos correspondiente al heterotetrámero 2+2 de sermetstatina y esnapalisina (Figura R.13). El análisis de la relación entre la absorbancia a  $\lambda = 280$  nm y  $\lambda = 260$  nm permitió intuir que los picos 1 y 5 podrían corresponder a contaminantes no proteínicos provenientes de los cuerpos de inclusión. Al analizar muestras de todos los picos por Tricina-SDS-PAGE se confirmó que los picos 1 y 5 no estaban compuestos por proteínas, y se determinó que el pico 2 correspondía al complejo esnapalisina-sermetstatina, el 3 a sermetstatina y el 4 a contaminantes proteínicos. La aparición de un pico de sermetstatina libre es debido a las

pérdidas de esnapalisina producidas durante el proceso de replegamiento. Su presencia en dicho paso de purificación resultó inevitable debido a que la disminución de la concentración inicial de inhibidor en el paso de replegamiento conlleva un descenso del rendimiento del proceso. En cambio, la adición de un exceso inicial de sermetstatina no resulta en un incremento evidente del rendimiento y además dificulta la posterior purificación del complejo, siendo óptimo partir de cantidades equimolares. El solapamiento de los picos de sermetstatina y complejo esnapalisina-sermetstatina dificultó notablemente la purificación, ya que únicamente en el primer tercio del pico 2 se pudo obtener el complejo binario en una relación estequiométrica adecuada (Figura R.13). Las fracciones del primer tercio fueron reservadas, mientras que las de los dos tercios siguientes, más las del primer tercio del pico 3 fueron concentradas y purificadas de nuevo mediante cromatografía de exclusión molecular. La repetición del proceso en varias ocasiones permitió la purificación del complejo binario minimizando las pérdidas de esnapalisina.

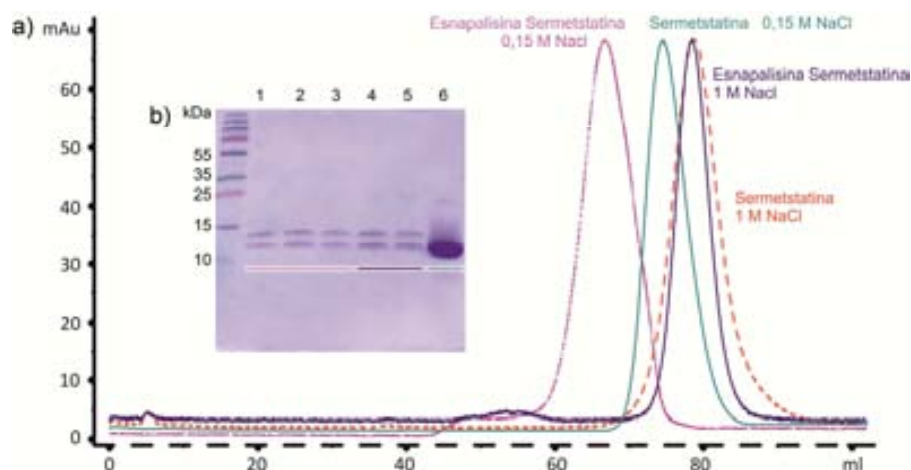


**Figura R.13.** Complejo esnapalisina-sermetstatina (I) **a)** Cromatograma de exclusión molecular del complejo esnapalisina-sermetstatina tras el proceso de replegamiento (línea rosa [ $\lambda = 280$  nm] y naranja pálido [ $\lambda = 260$  nm]). Se observa el solapamiento entre el pico del complejo binario (2; volumen de elución ( $V_e$ ) = 66,6 ml [45 kDa]) y el de sermetstatina libre (3;  $V_e = 73,0$  ml [20 kDa]). Columna Superdex 75 16/60 calibrada con aldolasa ( $V_e = 53,9$  ml [158 kDa]), conalbúmina ( $V_e = 61,5$  ml [75 kDa]), ovalbúmina ( $V_e = 67,2$  ml [43 kDa]) y citocromo C ( $V_e = 80,2$  ml [12 kDa]). **b)** Tricina-SDS-PAGE 15%. Marcador *PageRuler Plus Prestained Ladder* (Fermentas). Muestras correspondientes a las fracciones del pico 2 y del inicio del 3 (esnapalisina = 15 kDa, sermetstatina = 12 kDa).

Al concentrar el complejo esnapalisina-sermetstatina en el tampón de la cromatografía de exclusión molecular (tampón C; 20 mM Tris-HCl, 150 mM NaCl, pH 7,4) se alcanzó una concentración máxima de 1,5 mg/ml. Estudios posteriores de cristalización determinaron que se trataba de una concentración insuficiente para la obtención de cristales. Para incrementar la concentración del complejo se procedió a la diálisis de la muestra frente a tampón J (20 mM Tris-HCl, 1 M NaCl, pH 7,4). En estas condiciones, debido al incremento



de la fuerza iónica, el homodímero de sermetstatina se disocia pasando a adoptar una estructura monomérica (Figura R.14). A su vez, la interacción entre sermetstatina y esnapalisina se ve igualmente afectada, produciéndose la disociación del complejo. En condiciones fisiológicas la interacción entre una enzima y un inhibidor normalmente estabiliza a la enzima, siendo esta la base de algunas aplicaciones biotecnológicas de los inhibidores (Saunders et al. 1999). No obstante, esnapalisina tiene la propiedad de ser particularmente estable a elevadas concentraciones de sal, presentando su máxima actividad a 3,6 M NaCl (Inouye et al. 2007). Por tanto, el tampón J, pese a producir la disociación del complejo, incrementa notablemente la estabilidad de la enzima. Esto permitió la concentración de la muestra hasta los 4,5 mg/ml, condición que resultó necesaria para la obtención de cristales del complejo.

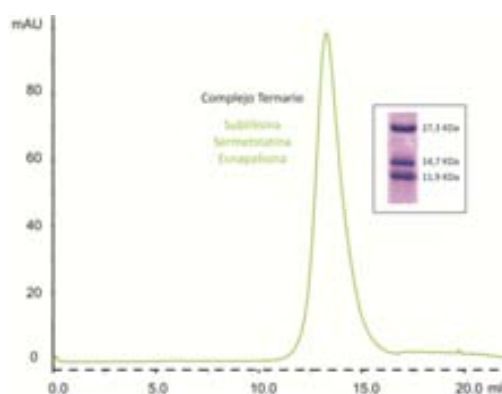


**Figura R.14.** Complejo esnapalisina-sermetstatina (II) **a)** Superposición de los cromatogramas de exclusión molecular de sermetstatina pura en tampón C (0,15 M NaCl; volumen de elución ( $V_e$ ) = 73,0 ml [20 kDa]; línea verde) y J (1 M NaCl;  $V_e$  = 79,2 ml [12 kDa]; línea roja discontinua) y complejo esnapalisina-sermetstatina en tampón C ( $V_e$  = 66,6 ml [45 kDa]; línea rosa discontinua) y J ( $V_e$  = 79,1 ml [13 kDa]; línea azul oscuro continua). La disociación del complejo causada por la elevada fuerza iónica del medio queda patente por la desaparición del pico del complejo en 1 M NaCl y la aparición de un único pico en el cual colocan el monómero de sermetstatina y esnapalisina. Columna Superdex 75 16/60 calibrada con aldolasa ( $V_e$  = 53,9 ml [158 kDa]), conalbúmina ( $V_e$  = 61,5 ml [75 kDa]), ovalbúmina ( $V_e$  = 67,2 ml [43 kDa]) y citocromo C ( $V_e$  = 80,2 ml [12 kDa]). Absorbancia medida a  $\lambda$  = 280 nm. **b)** Tricina-SDS-PAGE 15%. Marcador PageRuler Plus Prestained Ladder (Fermentas). Las muestras 1-3 corresponden al pico esnapalisina-sermetstatina ([15 kDa+12 kDa] $\times$ 2) en 0,15 M NaCl, las muestras 4 y 5 al de esnapalisina-sermetstatina en 1 M NaCl y la muestra 6 al de sermetstatina ([12 kDa] $\times$ 2) en 0,15 M NaCl.

#### 4.4.3. SUBTILISINA-SERMETSTATINA-ESNAPALISINA

La incubación de subtilisina Carlsberg con el complejo esnapalisina-sermetstatina permitió la obtención de un complejo ternario estable en las condiciones analizadas. Su posterior purificación por cromatografía de exclusión molecular en una columna Superdex 200

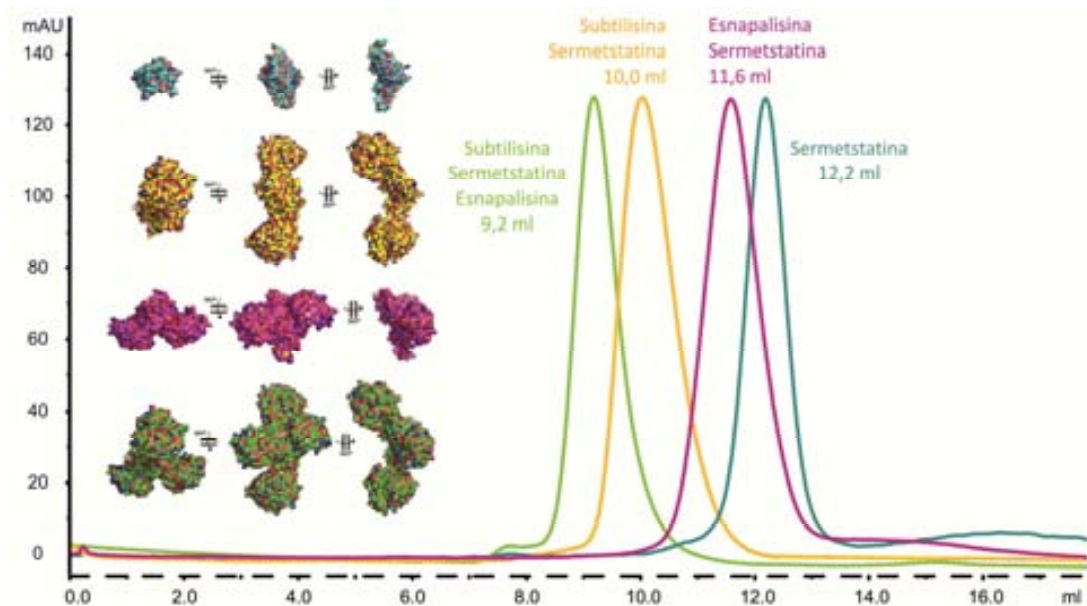
10/300, previamente calibrada, permitió determinar que el complejo adopta una estructura cuaternaria en forma de heterohexámero 2+2+2, en la cual un dímero de inhibidor interacciona con dos moléculas de subtilisina y dos de esnapalisina (Figura R.15). El análisis mediante Tricina-SDS-PAGE determinó la colocalización de las tres proteínas en las fracciones correspondientes al pico cromatográfico. El posterior proceso de concentración de estas fracciones permitió la obtención del complejo ternario a una concentración de 7,5 mg/ml.



**Figura R.15.** Complejo subtilisina-sermetstatina-esnapalisina (I). **a)** Cromatograma de exclusión molecular del complejo ternario (volumen de elución ( $V_e$ ) = 13,3 ml [102 kDa]). Columna Superdex 200 10/300 calibrada con aldolasa ( $V_e$  = 12,5 ml [158 kDa]), conalbúmina ( $V_e$  = 13,9 ml [75 kDa]), ovalbúmina ( $V_e$  = 14,8 ml [43 kDa]) y citocromo C ( $V_e$  = 17,8 ml [12 kDa]). Absorbancia medida a  $\lambda$  = 280 nm. **b)** Tricina-SDS-PAGE 15%. Muestra correspondiente al pico del complejo subtilisina-sermetstatina-esnapalisina. Se confirma la colocalización de las tres proteínas en el pico cromatográfico ((27 kDa+12 kDa+15 kDa)x2) y por tanto la formación del complejo ternario.

Para demostrar la formación del complejo ternario se comparó su migración en la cromatografía de exclusión molecular con la de los complejos subtilisina-sermetstatina y esnapalisina-sermetstatina y sermetstatina pura (Figura R.16). La adición de subtilisina (27 kDa) al dímero de sermetstatina (24 kDa) y al complejo esnapalisina-sermetstatina (54 kDa) supone un incremento de peso teórico de 54 kDa, resultando en la formación del complejo subtilisina-sermetstatina (78 kDa) y subtilisina-sermetstatina-esnapalisina (108 kDa), respectivamente. En cambio, la adición de esnapalisina (15 kDa) sobre sermetstatina para la formación del complejo únicamente supone un aumento teórico de 30 kDa, pasando de los 24 kDa a los 54 kDa. Por tanto, se hace evidente que la adición de subtilisina comportaría un desplazamiento mayor de los picos cromatográficos que el producido por la adición de esnapalisina. No obstante, las diferencias observadas experimentalmente en la migración de los diferentes complejos no pudieron explicarse únicamente por las diferencias de masa. Los datos de la calibración previa de la columna eran coherentes con el peso teórico del monómero de sermetstatina (tampón J), pero mostraron un ~15% de error respecto al dímero del inhibidor (tampón C) y al complejo esnapalisina-sermetstatina, obteniéndose

valores de pesos aparentes menores a los esperados. Por otro lado, se obtuvo una sobreestimación del ~10% en el peso del complejo subtilisina-sermetstatina. Esta diferencia es debida a que la cromatografía de exclusión molecular no separa las proteínas únicamente por su masa, sino que también está condicionada por la forma de las mismas. El volumen de elución de las diferentes proteínas depende por tanto del radio hidrodinámico ( $R_H$ ) o radio efectivo (Dubin and Principi 1989). Este se define como el radio de una esfera rígida hipotética cuya velocidad de migración es la misma que la media de las velocidades de migración de las diferentes orientaciones espaciales de la proteína. En la formación de complejos, la naturaleza de la interacción entre las proteínas puede condicionar las variaciones del radio hidrodinámico. Por tanto, la cromatografía de exclusión molecular no permite el cálculo exacto de las masas moleculares de algunos complejos proteínicos. No obstante, es una técnica útil para la determinación de la formación de dichos complejos mediante la evaluación de la aparición, desaparición y desplazamiento de los picos cromatográficos al estudiar conjuntamente y por separado las proteínas que los componen. Además, el análisis de estos picos mediante Tricina-SDS-PAGE permite determinar la colocación de las diferentes proteínas así como la apreciación aproximada de la relación estequiométrica entre ellas.

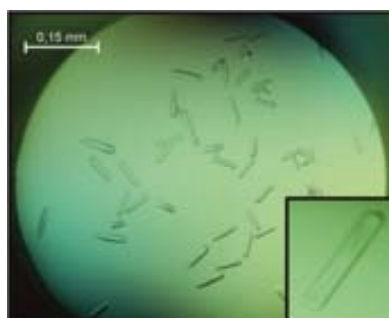


**Figura R16.** Complejo subtilisina-sermetstatina-esnapalisin (II). Cromatograma de exclusión molecular en columna Superdex 75 10/300. Absorbancia medida a  $\lambda = 280$  nm. Superposición de los cromatogramas de sermetstatina (línea cyan), complejo esnapalisin-sermetstatina (línea rosa), complejo subtilisina-sermetstatina (línea amarilla) y complejo subtilisina-sermetstatina-esnapalisin (línea verde). Se muestran las imágenes de las superficies moleculares de sermetstatina y de los diferentes complejos (mismos colores que el cromatograma) a la misma escala. Se presentan las moléculas centrales rotadas  $90^\circ$  horizontalmente (izquierda) y verticalmente (derecha) para la apreciación de los radios hidrodinámicos.

Los resultados de las cromatografías de exclusión molecular de los diferentes complejos sugirieron que la interacción de sermetstatina con subtilisina resulta en un incremento mayor del radio hidrodinámico que la interacción con esnapalisina y que este no se debe únicamente a la diferencia de tamaño de ambas proteínas. Esto sugiere que la interacción entre el inhibidor y las dos peptidasas ocurre a través de regiones diferentes. Los resultados estructurales (apartados 4.8 y 4.9) confirman este hecho, mostrando como subtilisina se une a los extremos del inhibidor incrementando notablemente el radio hidrodinámico, y esnapalisina a la zona central, generando un incremento más discreto.

## 4.5. CRISTALIZACIÓN Y COLECCIÓN DE DATOS DE DIFRACCIÓN DE RAYOS X

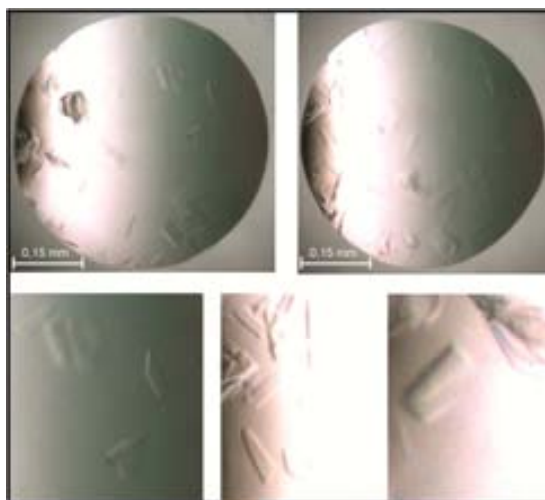
Los mejores cristales de sermetstatina nativa y de su derivado con selenometioninas en gotas de 200 nl aparecieron a 20°C con solución de proteína a 10 mg/ml en tampón C y solución de reservorio 100 mM citrato de sodio dihidratado, 200 mM acetato de amonio, 30% p/v PEG 4.000, pH 5,6. El proceso de escalado y optimización permitió la obtención de cristales de mayor tamaño y calidad en condición de reservorio 100 mM citrato de sodio dihidratado, 200 mM acetato de amonio, 10% p/v PEG 3.350, pH 5,6 (Figura R.17). Estos fueron crioprottegidos y difractados (apartado 3.5). Tanto los cristales de sermetstatina nativa como los de su derivado con selenometionina resultaron ser trigonales, con una molécula por unidad asimétrica (coeficiente de Matthews (Matthews 1968)  $[V_M] = 3,2 \text{ \AA}^3/\text{Da}$ ; contenido de solvente = 62%).



**Figura R.17.** Cristales de sermetstatina nativa. Ampliación de uno de ellos en la esquina inferior derecha, donde se aprecian los detalles de su morfología. Se trata de barras acabadas en punta con pequeñas imperfecciones en su superficie.

Los mejores cristales del complejo subtilisina-sermetstatina en gotas de 200 nl fueron obtenidos a 4°C con solución de proteína a 5,5 mg/ml en tampón C y solución de reservorio 100 mM cacodilato, 200 mM acetato de zinc dihidratado, 10% (v/v) isopropanol, pH 6,5. El

proceso de escalado y optimización permitió la obtención de cristales de mayor tamaño en las mismas condiciones de temperatura y composición de solución de proteína y de solución reservorio (Figura R.18). Los cristales fueron crioprottegidos y difractados (apartado 3.5), resultando ser monoclinicos. Se determinó la presencia de dos complejos por unidad asimétrica ( $V_M = 3,6 \text{ \AA}^3/\text{Da}$ ; contenido de solvente = 65%).



**Figura R.18.** Cristales de complejo subtilisina-sermetstatina. Ampliación de algunos de ellos en la parte inferior. Se observan placas delgadas trapezoidales, muchas de ellas agrupadas.

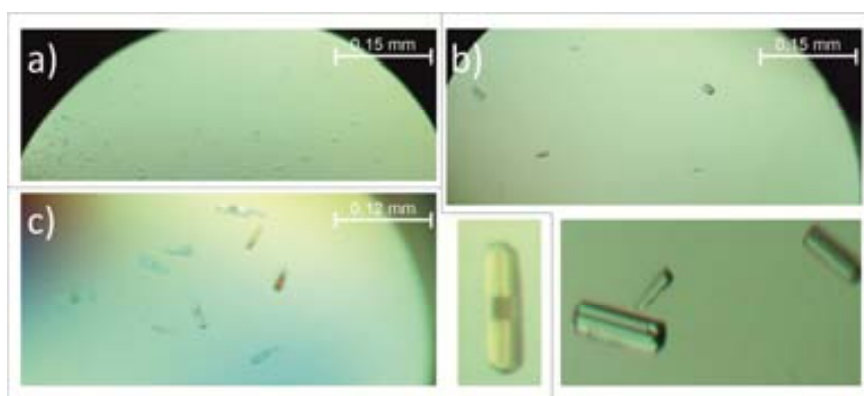
Los mejores cristales del complejo esnapalisina-sermetstatina se obtuvieron a 20°C con solución de proteína a 4,5 mg/ml en tampón J y solución de reservorio 100 mM HEPES, 200 mM  $\text{MgCl}_2$ , 15.0% (v/v) etanol, pH 7,5 (Figura R.19). El proceso de escalado no permitió la obtención de cristales, por lo que los cristales de las gotas de 200 nl fueron crioprottegidos y difractados (apartado 3.5). La utilización de una línea de luz de sincrotrón de microfoco resultó clave para la difracción ya que los cristales obtenidos eran muy pequeños. Estos resultaron ser trigonales y contenían seis complejos por unidad asimétrica ( $V_M = 2,9 \text{ \AA}^3/\text{Da}$ ; contenido de solvente = 58%). Se utilizó una colección helicoidal a lo largo del cristal aprovechando su forma de barra alargada, minimizando así el daño por radiación.



**Figura R.19.** Cristales de complejo esnapalisina-sermetstatina. Presentan forma de barra alargada.

Finalmente, la cristalización del complejo ternario entre subtilisina, sermetstatina y esnapalisina resultó en la obtención de cristales en numerosas condiciones. Estos aparecieron en gotas de 200 nl con una solución de proteína a 7,5 mg/ml en tampón C con soluciones de reservorio R1 (100 mM acetato de sodio, 10 mM cloruro de cobalto (II), 1 M 1,6-Hexanediol, pH 4,6), R2 (100 mM Na/K fosfato, 35,0% v/v MPD, pH 6,2), R3 (100 mM imidazol, 200 mM cloruro de magnesio, 35,0% v/v MPD, pH 8,0) y R4 (100 mM acetato de sodio, 200 mM acetato de zinc dihidratado, 10,0% p/v PEG 3.000, pH 4,5) a 20°C; y con R5 (100 mM HEPES, 200 mM citrato de sodio dihidratado, 30,0% v/v MPD, pH 7,5) y R6 (1 M amonio dihidrógenofosfato, 100 mM citrato de sodio dihidratado, pH 5,6) a 4°C.

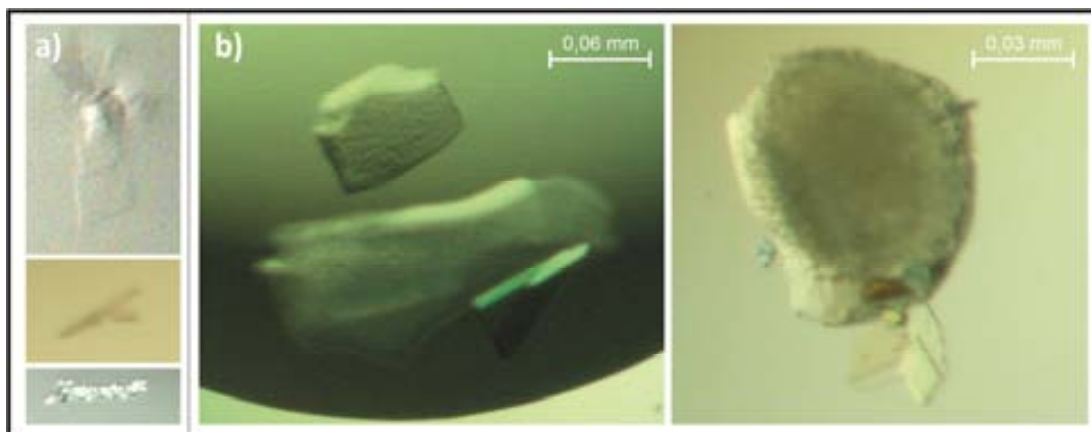
Los cristales obtenidos fueron demasiado pequeños para hacer viable su difracción, por lo que se procedió a su escalado, obteniéndose resultados positivos para las condiciones R1 y R2 modificadas (R1: 100 mM acetato de sodio, 10 mM cloruro de cobalto (II), 1,2 M 1,6-Hexanediol, pH 4,6. R2: 100 mM Na/K fosfato, 41,0% v/v MPD, pH 6,2). En el primer caso, los cristales obtenidos fueron igualmente pequeños, observándose una escasa nucleación en las gotas, por lo que se realizaron dos pasos sucesivos de microsebrado para su mejora. Se procedió aislando los cristales y preparando la solución de sembrado con Seed Beads (Hampton Research) de acuerdo con las instrucciones del fabricante. Tras el primer paso se obtuvieron cristales de mayor tamaño, los cuales fueron utilizados para la preparación de una nueva solución de sembrado (Figura R.20). La utilización de sucesivos pasos de sembrado para la mejora de los cristales, descrita en estudios anteriores (Stura and Wilson 1991), permitió finalmente la obtención de cristales de un tamaño adecuado para su crioprotección y difracción.



**Figura R.20.** Cristales de complejo subtilisina-sermetstatina-esnapalisina (I). Se observa el incremento del tamaño de los mismos durante las dos etapas de sembrado. **a)** Fase de escalado. **b)** Primer paso de sembrado. Detalle de unos cristales en la parte inferior. Adoptan forma de barra ancha acabada en picos redondeados y con irregularidades en la superficie, mostrando la apariencia de un conjunto barras alargadas fusionadas. **c)** Segundo paso de sembrado. Detalle de un cristal en la parte derecha. Presenta la misma morfología que los del paso anterior con la aparición de una muesca cuadrada en el centro.

La técnica de sembrado permite el crecimiento de cristales en la zona metaestable del diagrama de cristalización (Asherie 2004) debido a la adición de núcleos a la gota estabilizada en esta zona. En esta fase no es posible la formación de nuevos núcleos pero sí el crecimiento de cristales a partir de ellos, por lo que es una técnica muy eficiente en casos de escasa nucleación como el presente.

Por otro lado, el escalado de la condición R2 permitió también la obtención de cristales. En estas condiciones la nucleación resultó ser nula, obteniéndose los cristales a partir de impurezas presentes en las gotas como motas de polvo o fibras. Para la optimización de los cristales las técnicas de microsembrado y macrosembrado no resultaron eficaces, por lo que se procedió a utilizar los denominados nucleantes de Naomi (Saridakis and Chayen 2009). Estos nucleantes están compuestos por  $\text{CaO-P}_2\text{O}_5\text{-SiO}_2$ , son altamente porosos y presentan en su superficie cavidades en las cuales las proteínas pueden ser atrapadas, fomentando la nucleación y la formación de cristales (Chayen et al. 2006). Su utilización resultó en la obtención de cristales susceptibles de ser crioprottegidos y difractados (Figura R.21).



**Figura R.21.** Cristales de complejo subtilisina-sermetstatina-esnapalisina (II). **a)** Cristales obtenidos en la fase de escalado. En las dos primeras imágenes se observa como aparecen a partir de una mota de polvo, mientras que en la tercera lo hacen a partir de una fibra textil. **b)** Fase de optimización. Cristales adheridos a la superficie de los nucleantes de Naomi. En la primera fotografía se observa una placa triangular ancha, mientras que en la segunda hay diversas placas trapezoides muy finas.

No obstante, tanto los cristales obtenidos en la condición de reservorio R1 con microsembrado como en R2 con nucleantes de Naomi no difractaron por debajo de  $8 \text{ \AA}$  y estaban desordenados, por lo que la estructura del complejo ternario no pudo ser resuelta, optándose por la construcción de un modelo de homología (ver apartado 4.10).

## 4.6.- RESOLUCIÓN DE ESTRUCTURAS Y REFINAMIENTO

---

No se consiguió resolver la estructura de sermetstatina mediante reemplazamiento molecular utilizando las coordenadas de SSI como modelo de búsqueda pese a presentar una identidad de secuencia del 43%.

Los motivos por los cuales se cree que no funcionó se discuten en el apartado 4.7. Por otro lado, no se pudo utilizar la estructura de plasminostreptina de *S. antifibrinolyticus* (Kamiya et al. 1984) como modelo debido a que a pesar de que su estructura ha sido reportada, las coordenadas de la misma no se encuentran disponibles. Así pues, la estructura fue resuelta por difracción anómala a una sola longitud de onda (SAD) usando un derivado de sermetstatina con selenometioninas. Los datos de difracción recogidos en la longitud de onda del pico de absorción de selenio permitieron la identificación de los tres sitios de selenio del monómero de proteína presente en la unidad asimétrica. El subsiguiente faseado confirmó que el grupo espacial P3<sub>1</sub>21 era el correcto. Tras la construcción de un modelo inicial y posterior refinamiento, el modelo parcial fue utilizado para la determinación de la estructura nativa por reemplazamiento molecular. Los valores de *Z-score* (para su definición, ver (Evans and McCoy 2008; Larsen and Marx 2010)) obtenidos para las funciones de rotación y translación fueron 20,4 y 46,0, respectivamente. El modelo final de sermetstatina fue refinado con datos hasta una resolución de 1,9 Å. Este comprende los residuos G<sup>-1</sup>+S<sup>1</sup>-F<sup>113</sup> (cadena A), además de 126 moléculas de agua, una molécula de glicerol y tres iones acetato (Tabla R.2). Dos moléculas de sermetstatina aparecían íntimamente relacionadas por un eje cristalográfico, por lo que la estructura cristalográfica confirmó la estructura dimérica observada en solución (ver apartado 4.1).

La estructura del complejo subtilisina-sermetstatina fue resuelta por reemplazamiento molecular utilizando las coordenadas de subtilisina y sermetstatina en forma libre. Dos resultados significativos fueron hallados para ambas moléculas, los cuales mostraron unos *Z-scores* para las funciones de rotación y translación de 20,8/22,8 y 18,7/46,2 (subtilisina) y 6,0/36,2 y 6,8/44,5 (sermetstatina). El ordenamiento cuaternario encontrado en la unidad asimétrica cristalina muestra un heterotetrámero 2+2 (cadenas A-D). El modelo final, refinado con datos hasta una resolución de 2,25 Å, contiene los residuos A<sub>1</sub>-Q<sub>275</sub> (números de residuos de subtilisina indicados como subíndice de acuerdo con la secuencia de la proteína madura; UniProt P00780) de las moléculas de subtilisina (cadenas A y C) y sus respectivos iones calcio estructurales (Ca<sub>999</sub> and Ca<sub>998</sub>), G<sup>-1</sup>+S<sup>1</sup>-F<sup>113</sup> de la cadena B de sermetstatina y G<sup>4</sup>-F<sup>113</sup> de la cadena D; además de otros ligandos y solventes (Tabla R.2).



Tabla R.2. Datos cristalográficos de las estructuras.

Set de datos	Sermetstatina		Complejo Subtilisina Sermetstatina	Complejo Esnapalisina Sermetstatina
	Nativo	Selenometionina (pico de absorción) <sup>a</sup>	Nativo	Nativo
Grupo espacial	P3 <sub>1</sub> 21	P3 <sub>1</sub> 21	C2	P2 <sub>1</sub> 2 <sub>1</sub>
Constantes de celda (a, b, c, en Å; β en ° si ≠90)	71,04, 71,04, 52,40	70,81, 70,81, 52,25	183,97, 83,62, 77,62; 110,78	116,54, 121,81, 130,67
Longitud de onda (Å)	0,8726	0,9791	0,9763	0,8726
N.º de medidas / reflexiones únicas	134.041 / 12.234	218.877 / 10.514	312.908 / 52.346	211.0198 / 51.602
Rango de resolución (Å)	39,9 – 1,90	26,4 – 2,00	48,6 – 2,25	89,1 – 2,70
(outermost shell) <sup>b</sup>	(2,00 – 1,90)	(2,11 – 2,00)	(2,37 – 2,25)	(2,77 – 2,70)
Completitud	99,3 (98,9)	99,8 (100,0)	99,9 (99,7)	99,4 (95,8)
[Completitud anom.] (%)		99,7 (100,0)		
R <sub>merge</sub> <sup>c</sup>	0,038 (0,596)	0,074 (0,757)	0,097 (0,751)	0,091 (0,704)
R <sub>r.i.m.</sub> (= R <sub>meas</sub> ) <sup>c</sup>	0,040 (0,625)	0,077 (0,793)	0,106 (0,842)	0,105 (0,845)
[R <sub>p.i.m.</sub> ] <sup>c</sup>	0,012 (0,185)	0,023 (0,235)	0,043 (0,368)	
Intensidad media (<I<I> / σ<I>)>	40,3 (4,9)	35,4 (8,0)	13,2 (2,0)	14,7 (1,9)
B-Factor (Wilson) (Å <sup>2</sup> ) / Multiplicidad media	32,7 / 11,0 (11,1)	29,5 / 20,8 (21,8)	41,9 / 6,0 (4,9)	51,9 / 4,1 (3,2)
Resolution range used for refinement (Å)	∞ – 1,90		∞ – 2,25	∞ – 2,70
N.º de reflexiones usadas (set test)	11.844 (479)		51.487 (797)	50.699 (782)
Factor R cristalográfico (Factor R libre) <sup>d</sup>	0,187 (0,214)		0,177 (0,217)	0,195 (0,242)
N.º átomos de proteína / moléc. de solvente	834 / 126		5.483 / 305	11.043 / 275
Ligandos neutro	1 glicerol		1 diglicerol, 3 2-propanol, 2 glicerol	2 glicerol
Ligandos iónicos	3 CH <sub>3</sub> COO <sup>-</sup>		4 Ca <sup>2+</sup> , 2 Zn <sup>2+</sup> , 2 K <sup>+</sup> , 1 Cl <sup>-</sup> , 1 (CH <sub>3</sub> ) <sub>2</sub> AsOO <sup>-</sup> , 1 CH <sub>3</sub> COO <sup>-</sup> , 1 PO <sub>4</sub> <sup>3-</sup>	6 Zn <sup>2+</sup>
Rmsd				
Enlaces (Å) / Ángulos (°)	0,010 / 1,14		0,010 / 1,12	0,010 / 1,06
Factores B medios de átomos de prot. (Å <sup>2</sup> )	40,4		51,1	58,2
Análisis conformacional de ángulo de la cadena principal <sup>e</sup>				
Residuos en regiones favorecidas / atípicos / total	109 / 1 / 112		744 / 1 / 764	1.406 / 2 / 1.436

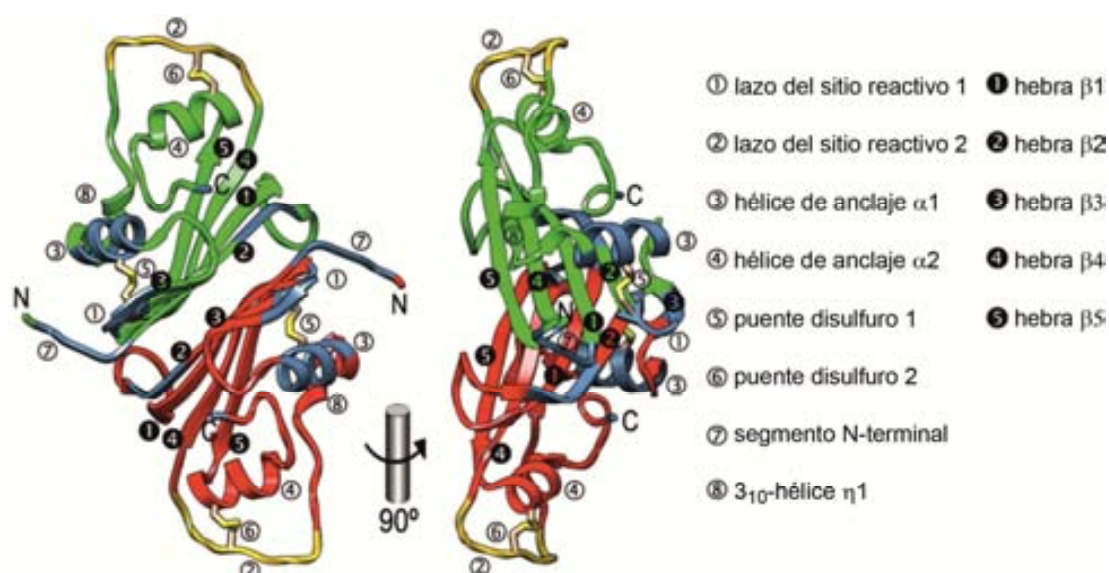
<sup>a</sup> Las parejas de Friedel fueron tratadas como reflexiones separadas. <sup>b</sup> Los valores entre paréntesis se refieren a la resolución del *outermost shell*. <sup>c</sup>  $R_{r.i.m.} = \sum_{hkl} (n_{hkl} / [n_{hkl} - 1]^{1/2}) \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$  y  $R_{p.i.m.} = \sum_{hkl} (1 / [n_{hkl} - 1]^{1/2}) \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , donde  $I_i(hkl)$  es la medida de intensidad  $i$ -th y  $n_{hkl}$  el número de observaciones de reflexiones  $hkl$ , incluyendo las reflexiones relacionadas por simetría, y  $\langle I(hkl) \rangle$  su intensidad media.  $R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ . <sup>d</sup> Factor R cristalográfico =  $\sum_{hkl} |F_{obs} - k| F_{calc} / \sum_{hkl} |F_{obs}|$ , con  $F_{obs}$  y  $F_{calc}$  como los factores estructurales de amplitud observados y calculados. Factor R libre, calculado de la misma manera para el set test de datos de reflexiones no usados durante el refinamiento. <sup>e</sup> De acuerdo con MOLPROBITY (Chen. et al. 2010).

La estructura del complejo esnapalisina-sermetstatina fue igualmente resuelta por reemplazamiento molecular usando las coordenadas de esnapalisina y sermetstatina en forma libre. La presencia de 12 moléculas en la unidad asimétrica obligó a la realización de las búsquedas de Patteson en dos pasos. En el primero, cuatro copias de cada molécula fueron encontradas independientemente, permitiendo la identificación de dos complejos los cuales mostraban la misma disposición relativa entre peptidasa e inhibidor. En una segunda búsqueda se hallaron seis soluciones reordenadas como tres heterotetrámeros 2+2 (cadenas A-D, E-H, I-L), utilizando uno de los complejos anteriores como modelo, con

unos *Z-score* para las respectivas funciones de rotación y translación de 6,7/15,3, 8,1/26,2, 7,3/29,8, 8,8/35,9, 5,5/37,8 y 6,6/35,7. El modelo final fue refinado con datos hasta una resolución de 2,7 Å. Este contiene los residuos de esnapalisina  $G_{-2}+P_{-1}+M_1-G_{132}$  en la cadena A, C, e I; los residuos  $P_{-1}+M_1-G_{132}$  de la cadena G y K; y los residuos  $M_1-G_{132}$  de la cadena E. También contiene los residuos de sermetstatina  $S^1-F^{113}$  (cadena B y L),  $A^2-F^{113}$  (cadena F),  $S^1-A^{62}+L^{67}-F^{113}$  (cadena D),  $S^1-L^{60}+L^{67}-F^{113}$  (cadena H) y  $S^1-P^{65}+V^{68}-F^{113}$  (cadena J), además de ligandos y solventes (275 moléculas de agua, dos moléculas de glicerol y seis iones de zinc; Tabla R.2). Debido a las diferencias en el entorno de empaquetamiento de las distintas moléculas presentes en el cristal, las cadenas E y K presentan una menor definición en la densidad de electrónica en comparación a las otras cadenas. Esto se traduce en un valor significativamente mayor en los parámetros promedio de desplazamiento térmico para estas cadenas (76,0 y 90,3 Å<sup>2</sup>, respectivamente) respecto al resto de moléculas (44,8-64,5 Å<sup>2</sup>).

#### 4.7. ESTRUCTURA DE SERMETSTATINA

Sermetstatina es una proteína  $\alpha/\beta$ -sándwich alargada de unas dimensiones máximas de 45x30x25 Å. Consta de cinco hebras  $\beta$  antiparalelas conectadas entre sí ( $\beta_3-\beta_2-\beta_1-\beta_4-\beta_5$ ; conectividad -1, -1, +3, +1) que dan lugar a una hoja  $\beta$  girada  $\sim 30^\circ$ , en cuya cara cóncava se acomodan dos hélices  $\alpha$  ( $\alpha_1$  y  $\alpha_2$ ) y una hélice  $3_{10}$  ( $\eta_1$ ) (Figuras R.22 y R.23).



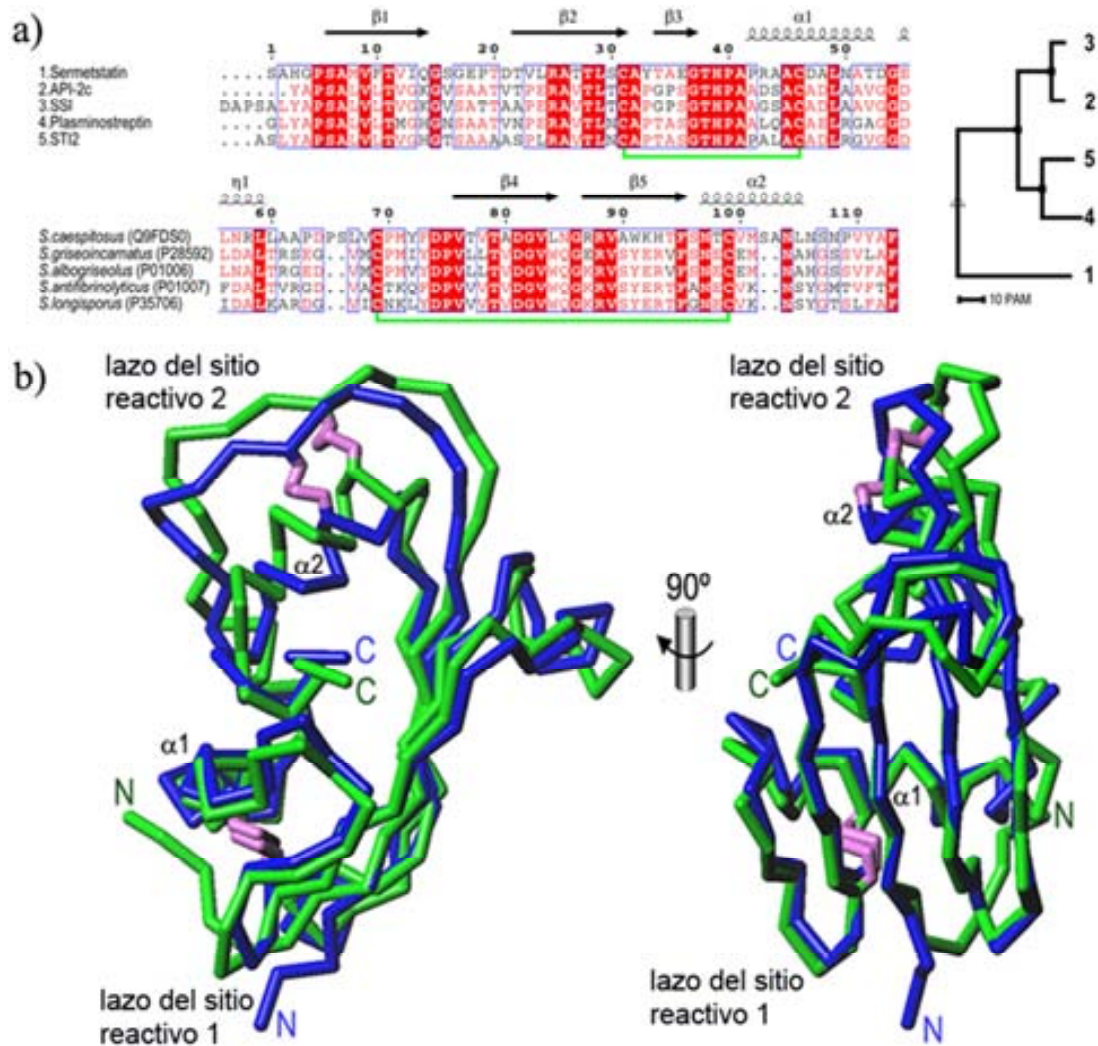
**Figura R.22.** Estructura de sermetstatina. Dímero de sermetstatina (en verde y rojo) en la orientación de referencia elegida (izquierda) y después de una rotación vertical de  $90^\circ$  (derecha). Los segmentos característicos están indicados por números. Las regiones implicadas en la unión con la metalopeptidasa (①) y las serín peptidasas (②) están representadas en azul y amarillo, respectivamente.

La estructura general está cohesionada por un gran núcleo hidrofóbico constituido por las cadenas laterales de los residuos de la cara cóncava de las hebras  $\beta$ , de las dos hélices  $\alpha$  y de la cola C-terminal. Además, la estructura está mantenida por dos puentes disulfuro. El primero ( $C^{31}-C^{46}$ ) une la “hélice de anclaje” ( $\alpha 1$ ) con la segunda hebra  $\beta$  ( $\beta 2$ ) y el subsiguiente lazo conector de las hebras  $\beta 2$  y  $\beta 3$  ( $L\beta 2\beta 3$ ), el cual incluye el “sitio reactivo 1” ( $C^{31}-A^{35}$ ). El segundo puente disulfuro ( $C^{69}-C^{99}$ ) conecta la “hélice de anclaje” ( $\alpha 2$ ) con  $L\alpha 2\beta 4$ , lazo en el que se encuentra el “sitio reactivo 2” ( $P^{65}-F^{73}$ ).

Los dos sitios reactivos sobresalen de la superficie de la molécula y son completamente accesibles al solvente. El carboxilato C-terminal está anclado a la cadena lateral de  $R^{25}$  de  $\beta 2$ . En contraste, el extremo N-terminal sobresale ligeramente de la superficie molecular y únicamente adquiere rigidez a partir de  $P^5$ , residuo que precede al primero de la hebra  $\beta 1$ .

Dos moléculas de sermetstatina se asocian íntimamente a través de las caras convexas de las hojas  $\beta$  resultando en un dímero relacionado cristalográficamente, en concordancia con el estado oligomérico determinado en solución. Los planos de las hojas  $\beta$  están rotados  $\sim 35^\circ$  el uno del otro, presentando una superficie de interacción de  $\sim 800 \text{ \AA}^2$ , lo que representa el 25% de la superficie total del monómero. La interacción es mayoritariamente hidrofóbica, formada por 31 contactos próximos, entre los cuales se incluyen las interacciones hidrofóbicas simétricas entre 10 residuos provenientes de cada una de las hojas  $\beta$  de cada monómero y únicamente dos puentes salinos entre  $R^{88}$  de una cadena y  $D^{81}$  de la otra.

La búsqueda por similitud de estructura de sermetstatina contra el PDB únicamente identificó al miembro de la familia MEROPS I16 SSI, el cual adopta también una estructura dimerica y se encuentra evolutivamente relacionado. No obstante, se hallaron variaciones estructurales significativas, particularmente alrededor del sitio reactivo 2, la estructura de las hélices de anclaje y los puentes disulfuro, y principalmente en el segmento N-terminal (Figura R.23). En las diferentes estructuras de SSI depositadas (códigos PDB 3SSI (Mitsui et al. 1977), 2SIC (Takeuchi et al. 1991a) y 2TLD (Strickler et al. 1992)), el segmento N-terminal es más largo y presenta una secuencia aminoacídica distinta, y, a diferencia de sermetstatina, está desordenado hasta las posiciones cuatro-siete. Estas variaciones son consistentes con las diferencias funcionales entre ambos inhibidores y pueden ser la causa de que sea inviable la resolución de la estructura de sermetstatina por reemplazamiento molecular utilizando SSI como modelo de búsqueda, motivo por el cual se tuvo que usar un derivado con selenometioninas.



**Figura R.23.** Familia MEROPS I16 y sermetstatina. **a)** (Izquierda) Alineamiento de secuencia de miembros representativos de la familia de inhibidores de peptidasas I16 según la base de datos MEROPS. Api-2c' (inhibidor de proteasas alcalinas 2c'), SSI (inhibidor de subtilisina de *Streptomyces*), STI2 (inhibidor de tripsina STI2). Los dos puentes disulfuro presentes en estas proteínas están indicados con líneas verdes. Los nombres de los organismos (*S.* abreviatura de *Streptomyces*) y el código de acceso UniProt se presentan a principio del segundo bloque. Los elementos de estructura secundaria indicados corresponden a la estructura de sermetstatina. (Derecha) Análisis filogenético de los cinco inhibidores (números correspondientes a los del bloque superior del alineamiento), en el cual se refleja la larga distancia evolutiva entre sermetstatina y las otras cuatro secuencias. MPA = mutaciones puntuales aceptadas. **b)** (Izquierda) Superposición del trazado de los C $\alpha$  de sermetstatina (verde) y SSI (azul; código *Protein Data Bank* [PDB] 2SIC (Takeuchi et al. 1991a)) tomados de sus respectivos complejos con subtilisina en la vista frontal de referencia y (derecha) una vista lateral. Se observan 96 átomos de C $\alpha$  con una desviación inferior a los 3 Å y una desviación cuadrática media de 1,48 Å. Se indican los lazos de los sitios reactivos 1 y 2, los extremos N-terminal y C-terminal, y las hélices de anclaje  $\alpha$ 1 y  $\alpha$ 2. Los puentes disulfuro se representan en rosa.

La estructura de sermetstatina ha sido depositada en el PDB y está disponible en la dirección electrónica <http://www.rcsb.org>. El código asignado para esta estructura es 4HWX.

## 4.8. ESTRUCTURA DEL COMPLEJO SUBTILISINA-SERMETSTATINA

---

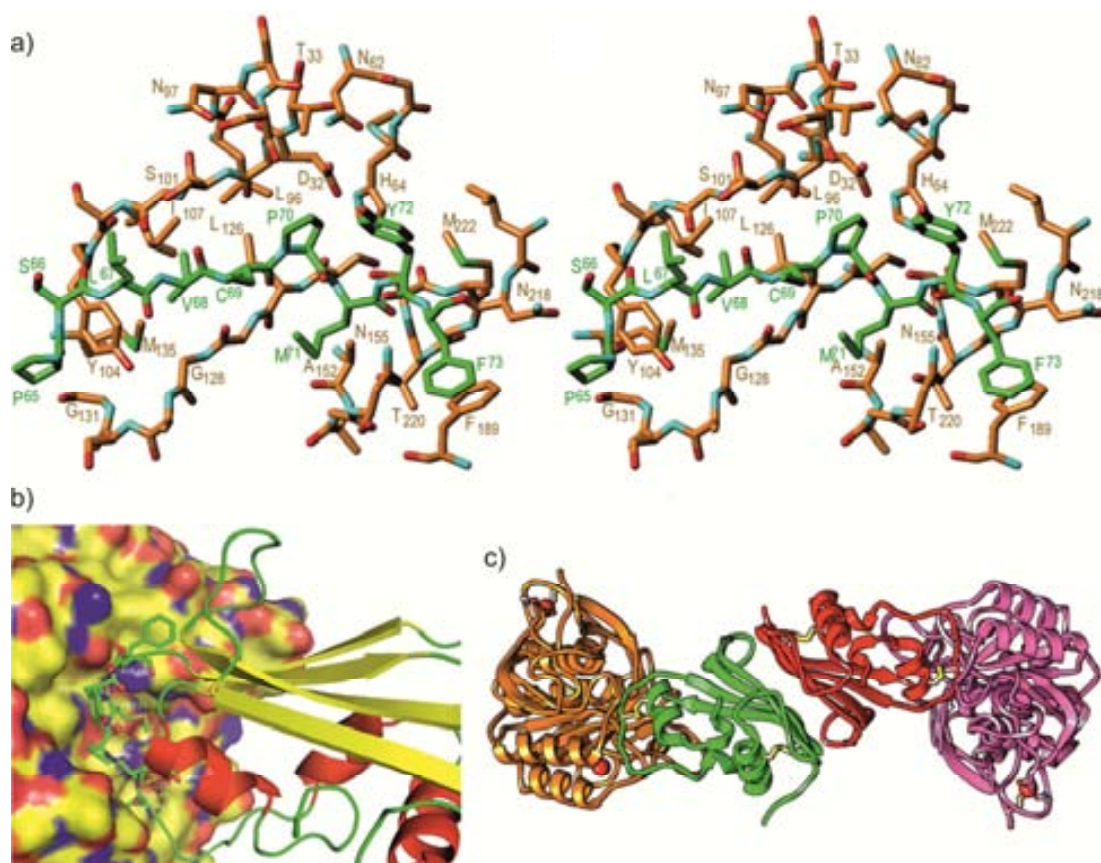
La estructura del complejo entre subtilisina Carlsberg y sermetstatina muestra un heterotetrámero 2+2, tal como anticiparon los estudios en solución, con un dímero de inhibidor intacto en el centro (de la misma manera que en la estructura del inhibidor en solitario) y una molécula de subtilisina unida a cada monómero de sermetstatina (Figura R.24).

Únicamente dos segmentos del inhibidor adoptan una conformación diferente respecto al inhibidor no unido. Estos son el extremo flexible N-terminal hasta A<sup>7</sup> y el segmento precedente al lazo y al propio lazo del sitio reactivo 2 hasta C<sup>69</sup>. La segunda variación es debida al volteo del enlace peptídico entre L<sup>60</sup> y A<sup>61</sup>. La única interacción entre peptidasa e inhibidor se realiza precisamente a través del sitio reactivo 2 del inhibidor (P<sup>65</sup>-F<sup>73</sup>), posicionado adecuadamente por su hélice de anclaje, y de la hendidura del centro activo de subtilisina.

La rigidez del sitio reactivo 2 se debe a la prolina presente en posición **P2** y al puente disulfuro C<sup>69</sup>-C<sup>99</sup>. El lazo, que incluye el enlace reactivo M<sup>71</sup>-Y<sup>72</sup>, está insertado a modo de cuña mimetizando un substrato en conformación extendida y canónica en la hendidura del centro activo de la enzima, siguiendo por tanto el mecanismo estándar de inhibición (apartado 1.3.1). Este mecanismo ha sido únicamente demostrado estructuralmente para inhibidores de serín peptidasas en general y específicamente para el inhibidor de insecto de metalopeptidasas (IMPI) (Arolas et al. 2011).

En la zona *no prima* de la hendidura del centro activo de subtilisina, la cadena principal del inhibidor se ancla a través de las interacciones de las hebras  $\beta$  al segmento G<sub>100</sub>-Y<sub>104</sub>, que crean el límite superior de la hendidura, a través de tres puentes de hidrógeno. Por debajo de la hendidura, el segmento de subtilisina S<sub>125</sub>-G<sub>128</sub> liga al inhibidor a través de tres puentes de hidrogeno adicionales entre cadenas principales. V<sup>68</sup> y M<sup>71</sup> ocupan, respectivamente, los dos subsitios de la hendidura del centro activo importantes para la especificidad en subtilisina, **S4** y **S1** (Takeuchi et al. 1991b). La serina catalítica, S<sub>221</sub>, forma un enlace de hidrógeno a través de su átomo O $\gamma$  con el N de M<sup>71</sup>. Por debajo del subsitio **P1** la cadena principal del inhibidor sufre un giro brusco que probablemente previene el corte del enlace M<sup>71</sup>-Y<sup>72</sup>, resultando en la formación de enlaces de hidrógeno entre la cadena

principal de los dos residuos subsiguientes y N<sub>318</sub>-T<sub>220</sub> de forma antiparalela. Contrariamente a como ocurre generalmente en metalopeptidasas, la posición **P1'** del sustrato no suele ser importante para la especificidad en serín peptidasas (Gomis-Ruth 2003; Page and Di Cera 2008). Esta se encuentra ocupada por Y<sup>72</sup>, cuya cadena lateral está en su mayoría expuesta al solvente. En cambio, la cadena lateral de F<sup>73</sup>, presente en la posición **P3'**, se orienta hacia la enzima formando una interacción hidrofóbica entre residuos aromáticos con F<sub>189</sub>.



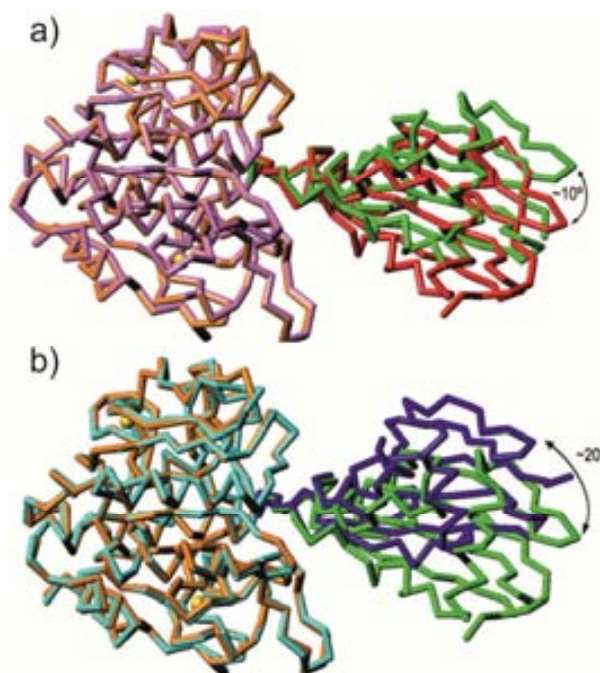
**Figura R.24.** Estructura del complejo subtilisina-sermetstatina. **a)** Detalle en estereoscopia de ojos cruzados<sup>3</sup> de la interacción entre el lazo del sitio reactivo 2 (átomos de carbono y etiquetas en verde) y subtilisina Carlsberg (átomos de carbono y etiquetas en naranja) mostrada en orientación estándar y en la orientación de referencia del inhibidor (ver figura R.22). La triada catalítica de subtilisina es S<sub>221</sub>-H<sub>64</sub>-D<sub>32</sub>. **b)** Detalle de la interacción de sermetstatina (lazos en verde, hojas  $\beta$  en amarillo, hélices  $\alpha$  en rojo y etiquetas en blanco; los aminoácidos del centro activo se representan con los átomos de carbono en verde, oxígeno en rojo y nitrógeno en azul) en la hendidura del centro activo de subtilisina (superficie representada con los átomos de carbono en amarillo, oxígeno en rojo y nitrógeno en azul). **c)** Estructura del complejo heterotetramérico entre subtilisina (naranja y rosa) y sermetstatina (verde y rojo) en orientación estándar del inhibidor subsecuentemente rotada 90° hacia la izquierda.

<sup>3</sup> En la estereoscopia de ojos cruzados, la imagen izquierda es percibida por el ojo derecho y viceversa. Para lograr que cada ojo vea la imagen que le corresponde es necesario cruzar la vista por delante de la imagen y tratar de enfocarla, permitiendo apreciar el efecto 3D en la imagen virtual que aparece en medio de ambas.



A pesar de la elevada potencia inhibitoria la superficie de interacción entre enzima e inhibidor es relativamente pequeña. Este hecho conlleva que tras la superposición de las moléculas de subtilisina de los dos complejos del heterotetrámero (Figura R.25a) las moléculas de sermetstatina muestren una rotación relativa de  $\sim 10^\circ$  alrededor de un eje horizontal a través de la hendidura del sitio activo. La pequeña superficie de interacción entre sermetstatina y subtilisina, junto con la ausencia de exositos empleados para la unión y que pudieran conferir especificidad, explican por qué diversas serin peptidasas pueden ser inhibidas por este inhibidor.

Finalmente, la superposición del complejo subtilisina-sermetstatina con el reportado para subtilisina BPN' y SSI (código PDB 2SIC (Takeuchi et al. 1991a)) revela la existencia de una unión muy similar. La superposición de las respectivas partes de la enzima en complejo (Figura R.25b) revela una reordenación relativa entre las moléculas de sermetstatina y SSI similar a la observada en la unidad asimétrica de la estructura del complejo subtilisina-sermetstatina, siendo en este caso de  $\sim 20^\circ$ .



**Figura R.25.** Flexibilidad en la interacción subtilisina-sermetstatina. **a)** Superposición de los dos complejos del complejo heterotetramérico (cadenas B/A en verde/naranja y D/C en rojo/rosa) después de la alineación óptima de las moléculas de la peptidasa. Los dos átomos de calcio de la cadena A se muestran como esferas amarillas para referencia. La vista corresponde a la orientación de la figura R.24 tras sendas rotaciones de  $90^\circ$  verticalmente y en el plano. **b)** Superposición del complejo entre subtilisina Carlsberg y sermetstatina (cadenas B/A en verde/naranja) sobre el complejo entre subtilisina BPN' y el inhibidor de subtilisina de Streptomyces (SSI; código *Protein Data Bank* [PDB] 2SIC (Takeuchi et al. 1991a); cian/azul) después de la alineación óptima de las moléculas de la peptidasa.

La estructura del complejo ha sido depositada en el PDB y está disponible en la dirección electrónica <http://www.rcsb.org>. El código asignado para esta estructura es 4HX2.

## 4.9. ESTRUCTURA DEL COMPLEJO ESNAPALISINA-SERMETSTATINA

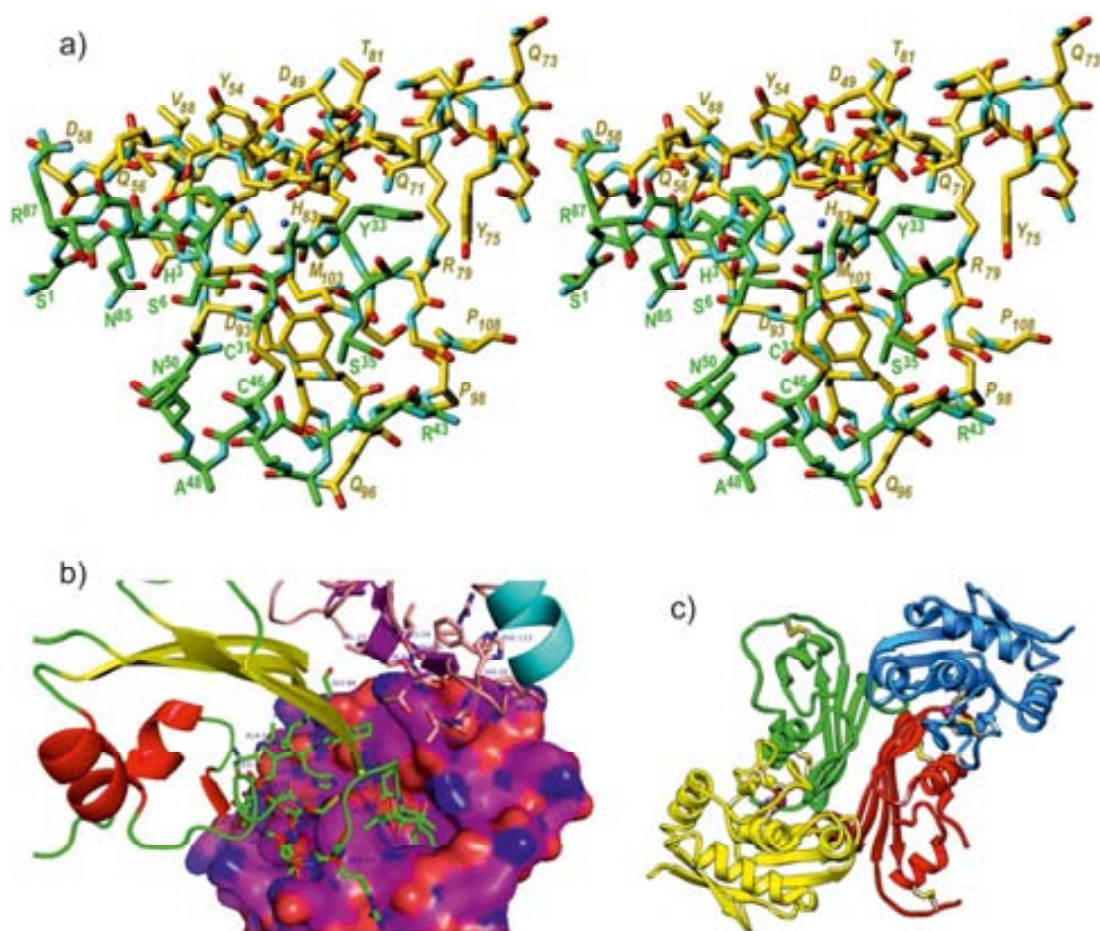
---

La estructura del complejo entre esnapalisina y sermetstatina muestra una estructura cuaternaria heterotetramérica 2+2, de nuevo de acuerdo con los resultados obtenidos en solución. En la unidad asimétrica cristalina se encontraron un total de 12 cadenas peptídicas (A-L) formando tres heterotetrámeros estructuralmente equivalentes (la siguiente discusión se centrará en el tetrámero ABCD). La estructura del núcleo del dímero de sermetstatina central muestra un inhibidor intacto con prácticamente las mismas desviaciones menores con respecto a la estructura del inhibidor no unido que en la estructura del complejo con subtilisina. No obstante, a diferencia de las dos anteriores, el sitio reactivo 2 es flexible y está parcialmente desordenado en algunas de las cadenas peptídicas en la estructura del complejo esnapalisina-sermetstatina. Por su lado, esnapalisina muestra una conformación virtualmente idéntica a la de la estructura no unida (código PDB 1KUH (Kurisu et al. 1997)) a lo largo de toda su estructura, a excepción del lazo  $A_{36}-A_{40}$ , el cual se encuentra en posición distal respecto a la hendidura del centro activo y de la zona de interacción con sermetstatina.

En el complejo, la enzima está unida a sermetstatina por una región localizada en la superficie opuesta al sitio de unión de subtilisina. Esta unión, en la cual participan diversos segmentos del inhibidor, presenta una gran superficie de contacto. Este hecho resulta en una escasa flexibilidad en la zona de interacción entre ambas proteínas, como se pudo comprobar por la superposición de los distintos complejos encontrados en la unidad asimétrica.

La interacción entre esnapalisina (cadena A) y sermetstatina (cadena B) abarca una superficie de  $\sim 870 \text{ \AA}^2$ , lo que representa el  $\sim 13\%$  de la superficie de enzima e inhibidor. Esta incluye 43 contactos próximos, entre ellos 11 puentes de hidrógeno e interacciones hidrofóbicas entre ocho residuos de esnapalisina y siete de sermetstatina. Los segmentos del inhibidor implicados en la interacción son la cola N-terminal ( $S^1-P^5$ ), el sitio reactivo 1 ( $C^{31}-A^{35}$ ), la hélice  $\alpha 1$  ( $R^{43}-N^{50}$ ) y la región  $L\beta 4\beta 5$  ( $N^{85}-G^{86}$ ). Por otro lado, también están implicados los segmentos  $R_{51}-D_{58}$ ,  $Q_{71}-Y_{75}$ ,  $R_{79}-H_{87}$ ,  $D_{93}-G_{97}$  y  $G_{105}-P_{108}$  de la enzima (Figura R.26). Esto se traduce en que la hendidura del centro activo de esnapalisina no es la única diana para su inhibición por parte de sermetstatina, sino que existen una serie de exositos en la superficie de la enzima. La existencia de los mismos explica la gran especificidad de sermetstatina a través de su segunda cara inhibitoria.





**Figura R.26.** Estructura del complejo esnapalísina-sermetstatina **a)** Detalle en estereoscopia de ojos cruzados de la interacción entre sermetstatina (cadena B; átomos de carbono y etiquetas en verde) y esnapalísina (cadena A; átomos de carbono y etiquetas en ocre). La peptidasa se muestra en orientación estándar con su ion de zinc catalítico representado como una pequeña esfera magenta. Las dos moléculas de solvente encontradas en la hendidura del centro catalítico se muestran como dos esferas azules: La primera es la molécula de solvente catalítico unida al ion zinc (derecha) y la segunda forma un puente entre el átomo  $H^3Ne2$  de sermetstatina y la molécula de solvente anterior (izquierda). La interacción del segmento de la protuberancia  $R_{51}-D_{58}$  de esnapalísina con las regiones de la segunda molécula del inhibidor del complejo heterotetramérico 2+2 (cadena D) no se muestra para facilitar la visualización. **b)** Detalle de la interacción de la cadena B de sermetstatina (lazos en verde, hojas  $\beta$  en amarillo, hélices  $\alpha$  en rojo y etiquetas en azul); los aminoácidos implicados en la interacción se representan con los átomos de carbono en verde, oxígeno en rojo y nitrógeno en azul) y la cadena D (lazos en naranja, hojas  $\beta$  en lila, hélices  $\alpha$  en cian y etiquetas en azul); los aminoácidos implicados en la interacción se representan con los átomos de carbono en verde, oxígeno en rojo y nitrógeno en azul) con esnapalísina (superficie representada con los átomos de carbono en lila, oxígeno en rojo y nitrógeno en azul). **c)** Estructura del complejo heterotetramérico entre esnapalísina (amarillo y azul) y sermetstatina (verde y rojo) en orientación estándar del inhibidor.

Este modo de inhibición representa un mecanismo novedoso para inhibidores de metalopeptidasas. El segmento N-terminal se encuentra insertado en el lado *no prima* de la hendidura del centro activo a modo de sustrato en conformación canónica, estableciendo interacciones tipo hoja  $\beta$  con las hebras  $\beta$  localizadas encima de la hendidura ( $A_{55}-D_{58}$ ). La cadena lateral de  $H^3$  apunta hacia la enzima interaccionando mediante un puente de solvente con la molécula de agua catalítica unida al ion de zinc.

La segunda región principal de interacción para la inhibición entre esnapalisina y sermetstatina es la que compromete al sitio reactivo 1, el cual penetra en el subsitio **S1'** de la hendidura mediante Y<sup>33</sup>, residuo que encaja con la especificidad de sustrato de la metalopeptidasa (Harada et al. 1995; Kurisu et al. 1997). La cadena lateral aromática interacciona con R<sub>79</sub>, aminoácido que crea el fondo del bolsillo hidrofóbico **S1'** de la enzima a través de la parte alifática de su cadena lateral. Debido a la intrusión del lazo reactivo 1, Y<sub>95</sub>, residuo hallado en una posición indistinguible a la observada en la estructura no unida y supuestamente implicado en la estabilización del intermediario de reacción tetraédrico durante la catálisis (Gomis-Ruth 2009), contacta con el carbonilo de la cadena principal del lazo en A<sup>32</sup> a través de su cadena lateral.

Además, la hélice  $\alpha$ 1 está implicada en una interacción secundaria entre esnapalisina y sermetstatina, contactando con el segmento inmediatamente anterior al residuo del tercer sitio de unión al ion de zinc catalítico de la enzima (D<sub>93</sub>-G<sub>97</sub>) y con L $\beta$ 4 $\beta$ 5, lazo que interacciona con la cadena del límite superior del sitio *no prima* de la hendidura catalítica. Concretamente D<sup>47</sup> de  $\alpha$ 1 interacciona a través de su cadena lateral con los átomos de nitrógeno de los grupos amida de la cadena principal de Y<sub>95</sub> y Q<sub>96</sub>. Esta interacción causa la reorientación de la cadena lateral de Q<sub>96</sub>, que se redirige hacia el solvente cuando se forma el complejo. Finalmente, la protuberancia frontal del segmento situado por encima del sitio *prima* de la hendidura del centro activo, R<sub>51</sub>-D<sub>58</sub>, interacciona con tres pequeños segmentos de la segunda molécula de sermetstatina (cadena D) del heterotetrámero en V<sup>23</sup>-T<sup>28</sup>, G<sup>37</sup>-P<sup>40</sup> y F<sup>122</sup> mediante cuatro puentes de hidrógeno.

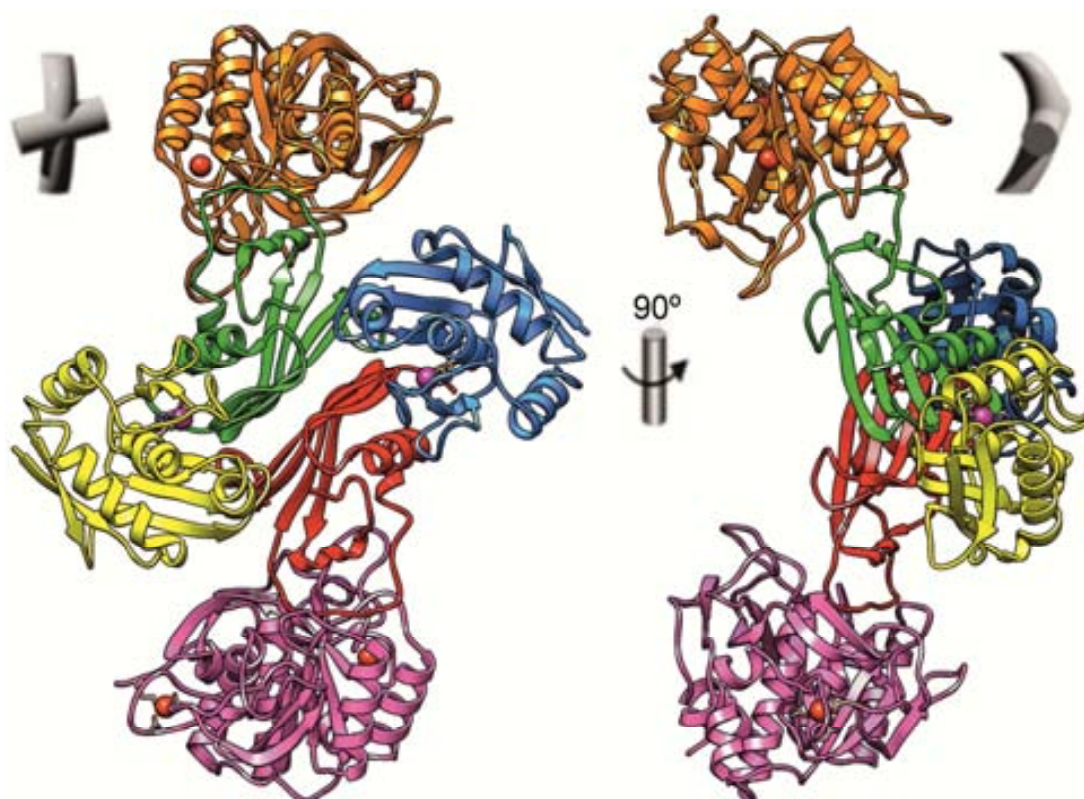
Globalmente, se puede considerar que este tipo de interacción entre inhibidor y enzima es una reminiscencia distante del modo inhibitorio de los TIMPs en su unión con MMPs (Gomis-Ruth et al. 1997) así como del modo inhibitorio del inhibidor de serralisina sobre la metalopeptidasa serralisina (Baumann et al. 1995). En estos casos, la cola N-terminal de los inhibidores también interacciona con la peptidasa en conformación extendida, pero a diferencia de en el caso de sermetstatina, lo hace en el lado *prima* de la hendidura del centro activo (apartado 1.3.2).

La estructura del complejo esnapalisina-sermetstatina ha sido depositada en el PDB y está disponible en la dirección electrónica <http://www.rcsb.org>. El código asignado para esta estructura es 4HX3.

#### 4.10. MODELO DEL COMPLEJO TERNARIO SUBTILISINA-SERMETSTATINA-ESNAPALISINA

Toda la información estructural descrita anteriormente permitió la construcción del modelo ternario entre subtilisina, sermetstatina y esnapalisina (Figura R.27). Para ello, únicamente el segmento N-terminal L $\beta$ 2 $\beta$ 3 y el segmento  $\beta$ 1-L $\beta$ 1 $\beta$ 2- $\beta$ 2 del complejo subtilisina-sermetstatina tuvieron que ser ligeramente remodelado para encajar con la conformación encontrada en el complejo esnapalisina-sermetstatina.

Este modelo muestra una gran partícula inhibitoria de unas dimensiones máximas de 120x85x60 Å, la cual es notablemente asimétrica y recuerda a una cruz deformada con una barra larga arqueada unos  $\sim 55^\circ$  y una barra corta horizontal, las cuales interseccionan en un ángulo de  $\sim 75^\circ$ . El modelo muestra una conformación heterohexamérica 2+2+2, tal y como se determinó en solución.



**Figura R.27.** Modelo de homología del complejo ternario subtilisina-sermetstatina-esnapalisina. Complejo heterohexamérico entre subtilisina (naranja y rosa), sermetstatina (verde y rojo) y esnapalisina (amarillo y azul) basado en las estructuras de los complejos binarios subtilisina-sermetstatina y esnapalisina-sermetstatina. Figura en orientación estándar del inhibidor (izquierda) y estándar rotada  $90^\circ$  verticalmente (derecha).

No se observan impedimentos estéricos entre las peptidasas unidas a los sitios distales del inhibidor, de manera que las enzimas se encuentran separadas por al menos ~15 Å de distancia. Por tanto, se puede concluir que sermetstatina es capaz de interaccionar simultáneamente con subtilisina y esnapalisina, tal y como se observa en solución (Figuras R.15 y R.16).

#### 4.11. REDISEÑO DE LA ESPECIFICIDAD INHIBITORIA Y ANÁLISIS DE LOS SITIOS ACTIVOS MEDIANTE MUTAGÉNESIS

Los estudios de inhibición de sermetstatina silvestre revelan una escasa actividad del inhibidor frente tripsina, posiblemente debido a la presencia de una metionina en posición P1 (M<sup>71</sup>). Con el objetivo de diseñar un inhibidor mejor para esta enzima se procedió a la producción de un mutante M<sup>71</sup>K de manera que encajara con la especificidad de tripsina. La expresión y purificación de este mutante resultó en su obtención con un rendimiento similar al de la forma silvestre. Esta mutación no mostró ninguna influencia sobre la inhibición de las metalopeptidasas analizadas. Sin embargo, se dobló la actividad inhibitoria frente a tripsina a la vez que se disminuyó la inhibición de serín peptidasas específicas con S1 hidrofóbico del tipo subtilisina (subtilisina Carlsberg y proteinasa K) y tipo quimotripsina (quimotripsina y elastasa) (Tabla R.3). Por tanto, la mutación M<sup>71</sup>K es capaz de alterar notablemente la especificidad de sermetstatina.

Tabla R.3. Actividad inhibitoria de sermetstatina silvestre y sus variantes.

Peptidasa	Relación molar (Inh./Enz.)	Silvestre	Y <sup>33</sup> P+T <sup>34</sup> G	H <sup>3</sup> R	H <sup>3</sup> E	M <sup>71</sup> K	Y <sup>72</sup> V+F <sup>73</sup> Y
		Actividad Inhibitoria* (%)	Actividad inhibitoria (%)	Actividad inhibitoria (%)	Actividad inhibitoria (%)	Actividad inhibitoria (%)	Actividad inhibitoria (%)
<b>Metalopeptidasas</b>							
Esnapalisina	1	99	1	96	95	90	96
	10	100	52	100	100	100	100
Astacina	100	0	0	0	0	0	0
Ulilisina	100	0	0	0	0	0	0
Aeruginolisina	100	0	0	0	0	0	0
Fragilisina	100	0	0	0	0	0	0
MMP <sup>3</sup> -1,-3,-13	100	0	0	0	0	0	0
<b>Serín peptidasas</b>							
Subtilisina	1	95	95	-	-	24	93
Tripsina	10	21	29	-	-	53	31
Quimotripsina	10	69	47	-	-	29	56
Elastasa	10	68	67	-	-	0	62
Proteinasa K	1	82	75	-	-	5	77
	10	100	100	-	-	36	100

\* Valores correspondientes a tres mediciones independientes con una desviación estándar de ±5. <sup>a</sup> Metalopeptidasa de matriz.

Por otro lado, se procedió a reemplazar las posiciones **P1'** y **P2'** del inhibidor por los presentes en SSI ( $Y^{72}V+F^{73}Y$ ; mutante obtenido con rendimientos similares a la forma silvestre) para determinar la importancia de estos sitios, no siendo posible observar ninguna diferencia inhibitoria significativa con respecto a sermetstatina silvestre (Tabla R.3). Los resultados son coherentes de acuerdo con la disposición que adoptan estos residuos en la estructura del complejo subtilisina-sermetstatina, ya que la cadena lateral de  $V^{72}$  quedaría expuesta al solvente y  $Y^{73}$  podría interactuar de una manera similar con  $F_{189}$  que  $F^{73}$ .

Para el estudio de los aminoácidos implicados en la inhibición de esnapalisina se obtuvieron los mutantes de sermetstatina  $Y^{33}P+T^{34}G$ ,  $H^3R$  y  $H^3E$ , todos ellos con un rendimiento similar al de la forma silvestre. Para evaluar la interacción a través del segmento N-terminal, concretamente en la interacción de  $H^3$  con la molécula de solvente unida al ion de zinc catalítico, se construyeron dos mutantes puntuales:  $H^3R$  y  $H^3E$ . Estos presentaron una actividad indistinguible respecto a la forma silvestre (Tabla R.3). Este resultado se explica por la capacidad de sermetstatina, observada a partir de su estructura, para acomodar los residuos cargados mediante la simple eliminación de la molécula de solvente. Para determinar la importancia del lazo reactivo 1 se construyó el doble mutante  $Y^{33}P+T^{34}G$  mimetizando la secuencia de SSI en esta región. El mutante mostró actividad únicamente residual frente a esnapalisina, mientras que su comportamiento contra serin peptidasas se mantuvo inalterado. Este hecho confirma el papel clave de esta segunda cara inhibitoria para la inhibición de la metalopeptidasa. Este sitio reactivo está ausente en SSI y muy probablemente en los otros miembros de la familia MEROPS I16 más alejados evolutivamente de sermetstatina, lo que explica que estos últimos no inhiban metalopeptidasas.

#### **4.12. EXPERIMENTOS DE ESTABILIDAD DE SERMETSTATINA. PUENTES DISULFURO**

---

La presencia de puentes disulfuro es un rasgo común en los inhibidores de peptidasas que siguen el mecanismo estándar de inhibición, por lo que se estudió la importancia de los dos puentes presentes en sermetstatina. Los puentes disulfuro del inhibidor unen los lazos de los dos sitios reactivos con las hélices de anclaje subyacentes, por lo que se procedió a su eliminación mediante la mutación de las cisteínas implicadas en su formación por serinas, obteniendo los dobles mutantes  $C^{31}S+C^{46}S$  y  $C^{69}S+C^{99}S$ .

La incubación por separado del inhibidor silvestre con subtilisina y esnapalisina mostró que sermetstatina no era cortada por ninguna de estas peptidasas, tal y como se confirmó en las estructuras resueltas de los respectivos complejos. En cambio, el mutante C<sup>69</sup>S+C<sup>99</sup>S, que fue obtenido con un rendimiento similar a la forma silvestre, fue completamente degradado por subtilisina aunque no por esnapalisina. El resultado sugiere que el inhibidor mutante mantiene una integridad estructural suficiente para la inhibición de la metalopeptidasa pero no para prevenir el colapso del sitio reactivo para la serín peptidasas (lazo 2). En contraste, el mutante C<sup>31</sup>S+C<sup>46</sup>S, obtenido con un rendimiento muy inferior al de la forma nativa (~100 veces menor), fue rápidamente degradado por ambas peptidasas, indicando que este puente disulfuro es esencial para la conformación estructural general del inhibidor.

Por tanto, los puentes disulfuro de sermetstatina son necesarios para su doble actividad inhibitoria ya que reducen la flexibilidad conformacional de la cadena principal alrededor de los sitios reactivos, contribuyendo al mantenimiento de la estructura general en una conformación inhibitoria competente.

## 5. DISCUSIÓN GENERAL

- Entonces el sagrado Jano, admirable en su duplicada forma, de repente ofreció los dos rostros a mis ojos. Me pasmé, y sentí que los cabellos se me habían espeluznado, y luego al punto sentí el corazón helado con el frío.

Publio Ovidio Nasón (siglo I)

*Fastos*





El nombre propuesto para el inhibidor de la peptidasa neutra de *Streptomyces caespitosus*, **sermetstatina**, hace referencia a que se trata de un inhibidor bifuncional de **serín** peptidasas y **metalopeptidasas**. Es un inhibidor monodominio de doble faz genuino con dos sitios reactivos distales que pueden unir diferentes peptidasas de manera simultánea e independiente. A través de una de sus caras (sitio reactivo 2) puede inhibir diversas serín peptidasas mediante el mecanismo estándar de inhibición, presentando una gran afinidad hacia peptidasas del tipo subtilisina. En cambio, su otra cara (sitio reactivo 1) presenta un mecanismo de inhibición novedoso en metalopeptidasas que le confiere una especificidad muy elevada hacia una única peptidasa, esnapalisina. Ambos sitios reactivos están estabilizados por sendos puentes disulfuro imprescindibles para su acción inhibitoria.

Estudios previos (Hiraga et al. 2000) habían demostrado la actividad inhibitoria de sermetstatina frente a serín peptidasas y esnapalisina, determinando también la ausencia de actividad frente a la cisteín peptidasa catepsina B, la aspartil peptidasa pepsina, la metalopeptidasa de familia desconocida almelisina (Shibata et al. 1997) y las metalopeptidasas gluzinquinas pertenecientes a la familia MEROPS M04 vimelisina, elastasa de *Pseudomonas aureginosa* y termolisina. La elección de estas metalopeptidasas para ensayos de inhibición responde probablemente al hecho de que el inhibidor de subtilisina SSI de *S. albogriseolus* presenta actividad inhibitoria frente a griselina (familia MEROPS M04). Este último inhibidor pertenece a la misma familia MEROPS que sermetstatina (I16) y por tanto está relacionado evolutivamente. No obstante, en SSI la inhibición de la metalopeptidasa y las serín peptidasas ocurre aparentemente a través del mismo sitio reactivo (situado en una posición equivalente al sitio reactivo 2 de sermetstatina) (Kumazaki et al. 1993) al contrario de lo que ocurre en sermetstatina donde existen dos sitios reactivos específicos e independientes. Por tanto, los resultados anteriores son útiles para descartar que el sitio reactivo 2 comparta con SSI la capacidad de inhibir metalopeptidasas, pero aportan una información limitada sobre la especificidad del sitio reactivo 1. En consecuencia, los experimentos de inhibición de metalopeptidasas en el presente trabajo se han centrado en metzinquinas, metalopeptidasas relacionadas con esnapalisina desde el punto de vista estructural y mecanístico.

Los estudios funcionales realizados han permitido determinar la afinidad de sermetstatina frente a diferentes peptidasas revelando información sobre la especificidad de cada uno de los sitios reactivos. El análisis de los datos estructurales obtenidos para los complejos subtilisina-sermetstatina y esnapalisina-sermetstatina ofrece una explicación coherente de los resultados de inhibición observados, aportando los detalles de las interacciones entre las enzimas y el inhibidor.

En el primer caso se observa como la interacción con subtilisina se realiza únicamente a través de un lazo de sermetstatina (lazo del sitio reactivo 2) que se inserta en el centro activo de la peptidasa. La escasa superficie de interacción y el hecho de que el lazo adopte una estructura parecida a un sustrato hacen que este pueda interactuar con diversas peptidasas. Por tanto, es de esperar que pueda inhibir diferentes serín peptidasas como son tripsina, quimotripsina y elastasa. Las diferencias en la afinidad del inhibidor frente a las diferentes peptidasas dependen del encaje de los residuos del lazo con los subsitios de las enzimas responsables de su especificidad respectiva. La estructura del complejo con subtilisina muestra como la posición **P1** del inhibidor interactúa con el subsitio **S1** del sustrato, siendo esta interacción determinante para la estabilización del complejo. La mutación de la metionina presente en **P1** por lisina y su consiguiente capacidad inhibitoria frente a tripsina, demostró la importancia de esta posición y que es posible alterar la especificidad de sermetstatina frente diferentes serín peptidasas modificando la secuencia aminoacídica del lazo reactivo.

Por el contrario, en la interacción sermetstatina-esnapalisina existe una gran superficie de contacto. En esta participan diversos segmentos del inhibidor. Por tanto, se trata de una interacción muy específica que hace improbable la unión de otro tipo de peptidasas a pesar de presentar estructuras similares. Este hecho es coherente con los datos funcionales que descartaron la inhibición de otros miembros del clan de las metzincinas. La mutación del sitio reactivo 1 de sermetstatina comporta una gran reducción de la actividad inhibitoria frente a esnapalisina, pero no la suprime totalmente, lo que indica la importancia del conjunto de las interacciones en el mecanismo de inhibición.

En el presente trabajo no se han realizado experimentos para el estudio de la función *in-vivo* de sermetstatina. No obstante, los datos funcionales y estructurales obtenidos, así como la información disponible sobre otros inhibidores relacionados permiten el planteamiento de diferentes hipótesis. Se desconoce cuál es su función biológica, en gran parte debido a que la fisiología de esnapalisina y subtilisina no está del todo establecida. En la cepa C5 de *Streptomyces sp.* se determinó que el gen de esnapalisina (*snpA*) se encontraba en el loci *snp* junto al gen *snpR*. Este codifica para un regulador de la transcripción perteneciente a la familia LysR que activa la transcripción de *snpA*, acción que ocurre principalmente durante las fases de crecimiento estacionario (DeSanti and Strohl 2003). La expresión del inhibidor supondría un modo de regulación adicional para la actividad de esta peptidasa.

Por otro lado, sermetstatina también podría estar regulando la actividad de peptidasas tipo subtilisina propias o actuando con función protectora frente a subtilisinas exógenas. La presencia de dos sitios reactivos independientes en sermetstatina podría suponer una ventaja evolutiva frente a otros inhibidores de la familia MEROPS I16, ya que permite la regulación de la actividad de la metalopeptidasa y de las serín peptidasas sin interferencias y de manera precisa, a la vez que optimiza el uso de los recursos al combinar las diferentes funciones en una única proteína.

Las bacterias del género *Streptomyces* son bacterias autóctonas del suelo que presentan un crecimiento lento y continuo. En cambio, las del género *Bacillus*, caracterizadas por secretar al medio elevadas cantidades de subtilisina, son bacterias generalmente oportunistas de crecimiento rápido que aparecen en momentos puntuales (Alexander 1977; Atlas and Bartha 2002). Por tanto, es de esperar que los miembros del género *Streptomyces* presenten mecanismos defensivos frente a este tipo de microorganismo. Este hecho queda patente en la multitud de antibióticos aislados de este género bacteriano (Tabla I.1). El género *Streptomyces* es muy variado, de manera que existen maneras muy diversas de afrontar esta situación de competencia según la especie.

En SSI una misma región inhibe griselisina y diversas serín peptidasas tipo tripsina y quimotripsinas, todas ellas implicadas en el desarrollo micelar, aunque es probable que su acción reguladora principal se ejerza sobre griselisina ( $K_i = 41 \text{ nM}$ ) (Kajiwara et al. 1991; Taguchi 1995). Al no haberse encontrado peptidasas tipo subtilisina implicadas en este proceso de desarrollo, es factible que la elevada afinidad de SSI frente a subtilisina se deba a un mecanismo de defensa contra subtilisinas exógenas, probablemente procedentes de bacterias oportunistas, ya que con una  $K_i = 5 \text{ pM}$  (Kojima et al. 1991) desplazaría a cualquier otra peptidasa en su unión con SSI. Este desplazamiento podría ser un mecanismo de defensa transitorio en caso de emergencia para evitar los efectos deletéreos de subtilisina en *S. albogriseolus* y/o un mecanismo de señalización donde la liberación de griselisina del complejo con el inhibidor promovería el desarrollo micelar y la esporulación, proceso que se da habitualmente frente a condiciones adversas y escasez de nutrientes (Daza et al. 1989).

La producción de sermetstatina, asumiendo la hipótesis de las subtilisinas exógenas, tendría sentido en un ambiente de competencia estable, donde la aparición de bacterias productoras de subtilisina no suponga la eliminación de la población de *Streptomyces* por falta de nutrientes ni se produzca la eliminación inmediata de la otra población mediante

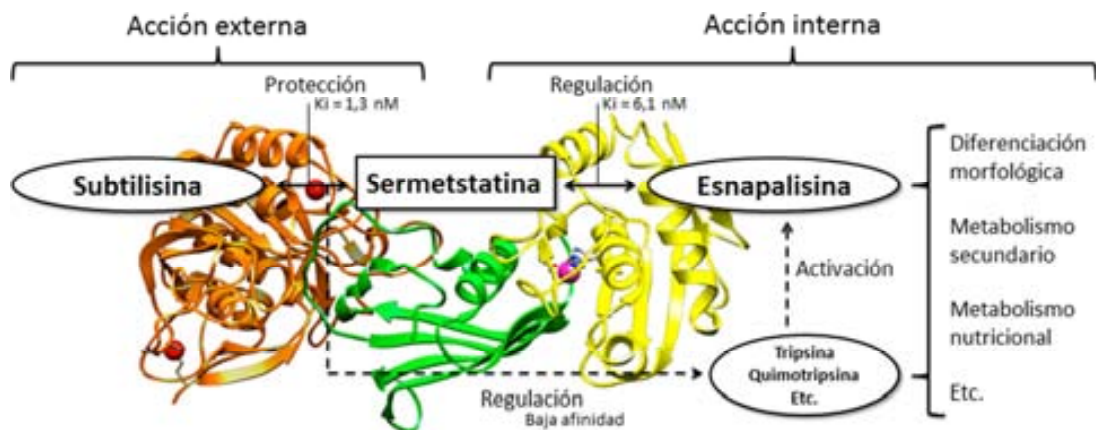
antibióticos. Por tanto, se podría tratar de competencia con bacterias autóctonas productoras de subtilisina o probablemente con bacterias oportunistas del género *Bacillus* en suelos ricos en materia orgánica donde su aparición fuese recurrente. El inhibidor podría desarrollar su función reguladora sobre esnapalisina de manera independiente sin importar la presencia de subtilisina en el medio, de manera que ejercería simultáneamente las funciones de regulación y protección. Sermetstatina habría evolucionado a partir de inhibidores de serín peptidasas de la familia MEROPS I16 con un único sitio reactivo adoptando nuevas funciones biológicas en un proceso de adaptación al medio. El estudio de estas funciones puede explicar algunas de las diferencias observadas entre sermetstatina y SSI, como es el caso de las diferentes afinidades observadas frente a subtilisina. En el caso de sermetstatina la  $K_i$  de unión a subtilisina ( $k_i = 1,3 \text{ nM}$ ) es mucho más elevada que en SSI ( $k_i = 5 \text{ pM}$ ) y similar a la de esnapalisina ( $k_i = 6,1 \text{ nM}$ ), probablemente debido a que en sermetstatina no existe competencia entre la serín peptidasa y la metalopeptidasa para su unión al inhibidor. En cambio, esta competencia podría ser clave para la función de SSI.

Tal y como se ha comentado, uno de los mayores escollos para la determinación de la función biológica de sermetstatina es el desconocimiento de la función de esnapalisina. Uno de los aspectos que ha quedado patente en la realización de este trabajo es la estricta regulación a la que esta enzima está sometida. Esta se realiza mediante una compleja regulación génica de su expresión, la expresión de la enzima en forma de zimógeno y la expresión de un inhibidor específico con una elevada afinidad. El principal objetivo del trabajo ha sido caracterizar este inhibidor, pero también se ha estudiado la regulación del zimógeno y su mecanismo de activación a nivel molecular.

De igual manera que el estudio de los inhibidores proteínicos, el estudio de los prodominios ofrece información sobre la regulación de las peptidasas y el conocimiento obtenido puede ser utilizado para el diseño de inhibidores de interés terapéutico y/o biotecnológico. Además, aporta información sobre los procesos de plegamiento de las proteínas, a la vez que puede ayudar a esclarecer aspectos de la función biológica de las peptidasas mediante el estudio de su activación.

Pese a no haberse conseguido la resolución de la estructura de proesnapalisina, los datos obtenidos permiten un mayor entendimiento de las funciones del prodominio en esta enzima. En primer lugar, se ha determinado su importancia en el plegamiento de esnapalisina gracias a su acción como chaperona. Esta observación fue clave para el

desarrollo del protocolo de replegamiento de esnapalisina madura en presencia de sermetstatina que concluyó en la resolución de la estructura del complejo binario. *In-vivo*, esta acción chaperona del inhibidor es improbable, pero representa un ejemplo de como la información extraída del estudio de los prodominios puede extrapolarse a otros ámbitos. Por otro lado, los estudios de activación determinaron que proesnapalisina requiere de la acción de otras peptidasas para alcanzar su forma madura. Debido al complejo ambiente en el que esnapalisina desarrolla su acción proteolítica no es de extrañar que colocalice con diversas enzimas tipo tripsina y quimotripsina producidas por el propio organismo, o incluso con subtilisinas endógenas o exógenas, que puedan ayudar en su activación. En todo caso, es probable que las enzimas activadoras sean serín peptidasas y que por tanto también puedan ser inhibidas por sermetstatina, lo que añadiría complejidad al modelo de las funciones biológicas propuesto para el inhibidor (Figura D.1).



**Figura D.1.** Esquema de las funciones biológicas propuestas para sermetstatina. Este se basa en los resultados funcionales y estructurales obtenidos y en su homología con el inhibidor de subtilisina de *Streptomyces* (SSI). Se muestra superpuesto el modelo del complejo ternario entre subtilisina (naranja), sermetstatina (verde) y esnapalisina (amarillo).

A lo largo de la introducción se ha puesto de manifiesto las diversas funciones fisiológicas en las que intervienen las peptidasas y como en ocasiones son causa de patología. El estudio de inhibidores proteínicos es siempre interesante, ya que aporta un valioso conocimiento sobre como la actividad de las peptidasas puede ser regulada. Los datos estructurales de las interacciones entre estos inhibidores y sus enzimas diana pueden ser utilizados en el diseño de nuevos inhibidores proteínicos optimizados o inhibidores sintéticos con aplicaciones terapéuticas y/o biotecnológicas. Además, entender cómo se producen estas uniones puede permitir la extrapolación de los mecanismos de inhibición sobre otras peptidasas de interés de estructura conocida.

Las estructuras de los complejos resueltas en el presente trabajo pasan a engrosar el conocimiento sobre las interacciones inhibidor-peptidasa, teniendo una especial relevancia la del complejo esnapalisina-sermetstatina, ya que representa un mecanismo novedoso de inhibición. Hasta el momento se habían descrito principalmente tres tipos de interacciones entre metaloendopeptidasas y sus inhibidores: IMPI-termolisina, TIMPs-MMPs y serralisina-inhibidor. Por tanto, la unión esnapalisina-sermetstatina representa un nuevo tipo de interacción para estas peptidasas y sus inhibidores, el tercero descrito para metzincinas.

Además, se ha obtenido por primera vez las estructuras de un inhibidor de doble faz monodominio en complejo con peptidasas diferentes a través de sus dos sitios reactivos. A diferencia de la mayoría de inhibidores de doble faz, incluidos los no genuinos, sermetstatina inhibe a dos peptidasas de diferente clase, por lo que se ha caracterizado un tipo de inhibidor realmente infrecuente en la naturaleza nunca antes descrito con este grado de detalle. La información obtenida ayuda a entender la complejidad de este tipo de inhibidores.

Por otro lado, cabe destacar que sermetstatina presenta una identidad de secuencia del 43% con SSI, un inhibidor con aplicaciones biotecnológicas asociadas a la estabilización y conservación de la subtilisina presente en detergentes para la ropa, lavavajillas, limpiadores industriales y productos de cuidado personal, y que ha sido postulado como tratamiento alternativo frente a la aspergilosis pulmonar invasiva. La elevada similitud de sermetstatina con este inhibidor hace de él un candidato para las mismas aplicaciones biotecnológicas y terapéuticas, pudiendo representar sus singularidades una ventaja notable en algunas de ellas. Ambos comparten una estructura similar estabilizada por dos puentes disulfuros y su capacidad de inhibir subtilisina, pero a diferencia de sermetstatina, SSI lo hace de manera tan estable que se requieren mutaciones en su sitio reactivo para permitir la disociación del complejo en sus aplicaciones biotecnológicas.

En resumen, en el presente trabajo se ha determinado que sermetstatina es un inhibidor de doble faz genuino dimérico monodominio que ha evolucionado a partir de inhibidores de serín peptidasas de la familia MEROPS I16 con un único sitio reactivo que siguen el mecanismo estándar de inhibición. Dicha evolución ha dado lugar a una proteína bifuncional capaz de inhibir simultáneamente diferentes serín peptidasas y una metalopeptidasa específica a través de dos sitios reactivos distales compatibles.

## 6. CONCLUSIONES

- Toda puerta tiene de aquí y de allí duplicadas frentes, de las cuales una mira al pueblo, la otra a las casas. Y a la manera que vuestro portero, sentado junto al umbral del primer cuarto, mira las entradas y las salidas: Así yo, portero de la casa celestial, miro juntamente las partes del oriente y las de occidente. Y a mí me es lícito, por no perder tiempo con la vuelta de la cerviz, mirar a dos partes sin mover el cuerpo.

Publio Ovidio Nasón (siglo I)

*Fastos*





---

De la realización del presente trabajo se pueden extraer las siguientes conclusiones:

1. Se ha sobreexpresado heterológamente el inhibidor sermetstatina en forma soluble en células de *E. coli*.
2. Se ha observado mediante cromatografía de exclusión molecular que sermetstatina adopta una estructura homodimérica estable en solución.
3. Se ha caracterizado la actividad inhibitoria de sermetstatina encontrándose que esta inhibe serín peptidasas de tipo subtilisina en el rango nanomolar. También inhibe en menor medida quimotripsina y de forma muy discreta tripsina. Sermetstatina inhibe de manera muy específica la metalopeptidasa de la familia de las metzincinas esnapalisina, con una constante de inhibición nanomolar.
4. Se ha determinado la estructura tridimensional de sermetstatina mediante el método de difracción anómala a una sola longitud de onda a partir de un derivado con selenometionina, obteniéndose una conformación en  $\alpha/\beta$ -sándwich alargado constituido por cinco hebras  $\beta$  antiparalelas, en cuya cara cóncava se acomodan dos hélices  $\alpha$  y una hélice  $3_{10}$ . Presenta dos sitios reactivos distales que sobresalen de la superficie de la molécula y son completamente accesibles al solvente.
5. Se ha observado mediante cromatografía de exclusión molecular que sermetstatina y subtilisina forman un complejo heterotetramérico estable en solución de estequiometría 2:2.
6. Se ha determinado la estructura tridimensional del complejo subtilisina-sermetstatina mediante reemplazamiento molecular utilizando sus estructuras aisladas como modelo. Sermetstatina interacciona con el centro activo de subtilisina a través del lazo del sitio reactivo 2 (P<sup>65</sup>-F<sup>73</sup>) siguiendo el mecanismo estándar de inhibición.
7. Se ha sobreexpresado heterológamente esnapalisina y proesnapalisina en forma de cuerpos de inclusión en *E. coli* y se ha desarrollado un protocolo de replegamiento para obtener ambas proteínas en forma soluble.
8. Se ha determinado que la activación *in vitro* de proesnapalisina se produce mediante tres cortes autolíticos de manera secuencial. La forma madura se obtiene mediante diversos cortes en la parte restante de 49 residuos del prodominio N-terminal, los cuales son probablemente asistidos *in vivo* por otras peptidasas.
9. Se ha observado que el prodominio de esnapalisina tiene un efecto chaperona en el plegamiento de la peptidasa. Sermetstatina mimetiza este efecto en el replegamiento de la forma madura de esnapalisina.

10. Se ha observado mediante cromatografía de exclusión molecular que sermetstatina y esnapalisina forman un complejo heterotetramérico estable en solución de estequiometría 2:2.
11. Se ha determinado la estructura tridimensional del complejo esnapalisina-sermetstatina mediante reemplazamiento molecular utilizando las estructuras aisladas como modelos. Sermetstatina interacciona con el centro activo y con regiones de la superficie de esnapalisina a través de la cola N-terminal, el lazo del sitio reactivo 1 (C<sup>31</sup>-A<sup>35</sup>), la hélice  $\alpha$ 1 y la región L $\beta$ 4 $\beta$ 5, así como por otros segmentos accesorios de la otra subunidad del dímero de sermetstatina, presentando un mecanismo de inhibición novedoso.
12. Se ha observado mediante cromatografía de exclusión molecular que subtilisina, sermetstatina y esnapalisina forman un complejo heterohexamérico 2:2:2 estable en solución.
13. Se ha obtenido un modelo de homología del complejo ternario entre subtilisina, sermetstatina y esnapalisina basado en las estructuras cristalográficas de los complejos binarios subtilisina-sermetstatina y esnapalisina-sermetstatina. Se ha confirmado que sermetstatina es capaz de inhibir simultáneamente a subtilisina y esnapalisina al no observarse impedimentos estéricos entre las peptidasas unidas a los sitios distales del inhibidor.
14. Se ha caracterizado mediante mutagénesis el papel fundamental de los residuos M<sup>71</sup> (sitio reactivo 2) y Y<sup>33</sup> (sitio reactivo 1) de sermetstatina en la inhibición de serín peptidasas y esnapalisina, respectivamente. La substitución de los residuos de cada uno de los sitios reactivos por separado no afecta a la actividad del otro sitio reactivo.
15. Se ha caracterizado mediante mutagénesis la importancia de los puentes disulfuro en el mantenimiento de la estructura de sermetstatina así como de su actividad inhibitoria. El puente disulfuro C<sup>69</sup>-C<sup>99</sup> mantiene la integridad estructural del sitio reactivo 2 frente a serín peptidasas, mientras que el puente C<sup>31</sup>-C<sup>46</sup> contribuye al mantenimiento de la estabilidad conformacional general del inhibidor.

## 7. BIBLIOGRAFÍA

Reconocerás las cosas sagradas, sacadas de los antiguos anales, y por qué razón cada día se haya señalado.

Publio Ovidio Nasón (siglo I)

*Fastos*



**A**

- Aaslyng, D., et al. (2000), 'Modified subtilisins and detergent compositions containing same', Novozymes and Unilever PLC, U. S. Patent 6632646, October 2.
- Agarwal, R. and Singh, N. (2006), 'Amphotericin B is still the drug of choice for invasive aspergillosis', *Am J Respir Crit Care Med*, 174, 102-3.
- Alberts, I. L., Nadassy, K. and Wodak, S. J. (1998), 'Analysis of zinc binding sites in protein crystal structures', *Protein Sci*, 7, 1700-16.
- Alden, R. A., et al. (1971), 'Atomic coordinates for subtilisin BPN' (or Novo)', *Biochem Biophys Res Commun*, 45, 337-44.
- Alexander, M. (1977), 'Introduction to soil microbiology' (2 edn.; New York: Wiley).
- Alonso-del-Rivero, M., et al. (2012), 'Tri-domain bifunctional inhibitor of metallocarboxypeptidases A and serine proteases isolated from marine annelid *Sabellastarte magnifica*', *J Biol Chem*, 287, 15427-38.
- Appel, W. (1986), 'Chymotrypsin: molecular and catalytic properties', *Clin Biochem*, 19, 317-22.
- Ardelt, W. and Laskowski, M., Jr. (1983), 'Thermodynamics and kinetics of the hydrolysis and resynthesis of the reactive site peptide bond in turkey ovomucoid third domain by aspergillopeptidase B', *Acta Biochim Pol*, 30, 115-26.
- Arolas, J. L., et al. (2011), 'Structural evidence for standard-mechanism inhibition in metallopeptidases from a complex poised to resynthesize a peptide bond', *Angew Chem Int Ed Engl*, 50, 10357-60.
- Ascenzi, P., et al. (2003), 'The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein', *Curr Protein Pept Sci*, 4, 231-51.
- Asherie, N. (2004), 'Protein crystallization and phase diagrams', *Methods*, 34, 266-72.
- Atlas, R. M. and Bartha, R. (2002), 'Ecología microbiana y microbiología ambiental' (4 edn.; Madrid: Addison Wesley).
- Azarkan, M., et al. (2011), 'The plasticity of the  $\beta$ -trefoil fold constitutes an evolutionary platform for protease inhibition', *J Biol Chem*, 286, 43726-34.

**B**

- Balaban, N. P., et al. (2002), 'Proteinases from *Bacillus intermedius* secreted in the late stages of sporulation', *Med Sci Monit*, 8, BR168-71.
- Bao, R., et al. (2009), 'The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation', *J Biol Chem*, 284, 26676-84.
- Barbosa, J. A., et al. (2007), 'Crystal structure of the Bowman-Birk Inhibitor from *Vigna unguiculata* seeds in complex with  $\beta$ -trypsin at 1.55 Å resolution and its structural properties in association with proteinases', *Biophys J*, 92, 1638-50.
- Barrett, A. J., Rawlings, N. D. and Woessner, J. F. (1998), 'Handbook of Proteolytic Enzymes' (London: Academic Press).
- Bartlett, J. D. and Smith, C. E. (2013), 'Modulation of cell-cell junctional complexes by matrix metalloproteinases', *J Dent Res*, 92, 10-7.
- Baumann, U., et al. (1993), 'Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: a two-domain protein with a calcium binding parallel  $\beta$  roll motif', *EMBO J*, 12, 3357-64.
- Baumann, U., et al. (1995), 'Crystal structure of a complex between *Serratia marcescens* metallo-protease and an inhibitor from *Erwinia chrysanthemi*', *J Mol Biol*, 248, 653-61.
- Becker, J. W., et al. (1995), 'Stromelysin-1: three-dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme', *Protein Sci*, 4, 1966-76.
- Bentley, S. D., et al. (2002), 'Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)', *Nature*, 417, 141-7.
- Betzl, C., Pal, G. P. and Saenger, W. (1988), 'Synchrotron X-ray data collection and restrained least-squares refinement of the crystal structure of proteinase K at 1.5 Å resolution', *Acta Crystallogr B*, 44, 163-72.
- Beynon, R. R. J. and Bond, J. S. (2001), 'Proteolytic Enzymes: A Practical Approach' (2 edn.; New York: Oxford University Press).
- Birktoft, J. J. and Blow, D. M. (1972), 'Structure of crystalline  $\alpha$ -chymotrypsin. The atomic structure of tosyl- $\alpha$ -chymotrypsin at 2 Å resolution', *J Mol Biol*, 68, 187-240.
- Blanc, E., et al. (2004), 'Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT', *Acta Crystallogr D*, 60, 2210-21.

- Bode, W. and Huber, R. (1991), 'Proteinase-protein inhibitor interaction', *Biomed Biochim Acta*, 50, 437-46.
- Bode, W. and Huber, R. (1992), 'Natural protein proteinase inhibitors and their interaction with proteinases', *Eur J Biochem*, 204, 433-51.
- Bode, W., et al. (1985), 'The crystal and molecular structure of the third domain of silver pheasant ovomucoid (OMSVP3)', *Eur J Biochem*, 147, 387-95.
- Bode, W., et al. (1992), 'Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases', *Nature*, 358, 164-7.
- Bostanci, N. and Belibasakis, G. N. (2012), '*Porphyromonas gingivalis*: an invasive and evasive opportunistic oral pathogen', *FEMS Microbiol Lett*, 333, 1-9.
- Brew, K. and Nagase, H. (2010), 'The tissue inhibitors of metalloproteinases (TIMPs): an ancient family with structural and functional diversity', *Biochim Biophys Acta*, 1803, 55-71.
- Brode, S. K., Ling, S. C. and Chapman, K. R. (2012), ' $\alpha$ -1 antitrypsin deficiency: a commonly overlooked cause of lung disease', *CMAJ*, 184, 1365-71.
- Brünger, A. T., et al. (1998), 'Crystallography & NMR System: a new software suite for macromolecular structure determination', *Acta Crystallogr D*, 54, 905-21.
- Bryan, P. N. (2002), 'Prodomains and protein folding catalysis', *Chem Rev*, 102, 4805-16.
- Bryan, P. N., Alexander, P. A. and Strausberg, S. L. (1993), 'Calcium free subtilisin mutants', University of Maryland, U. S. Patent 5470733, June 1.

## C

- Capaldi, S., et al. (2007), 'Crystal structure of the anticarcinogenic Bowman-Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin', *J Struct Biol*, 158, 71-9.
- Carranza, C., et al. (1999), 'Turbo-Frodo', *AFMB Activity Report 1996-1999* (Marseille: CNRS-UPR 9039), 89-90.
- Carrell, R. W. and Owen, M. C. (1985), 'Plakalbumin,  $\alpha$ 1-antitrypsin, antithrombin and the mechanism of inflammatory thrombosis', *Nature*, 317, 730-2.
- CCP4 (1994), 'The CCP4 suite: programs for protein crystallography', *Acta Crystallogr D*, 50, 760-63.

- Clutterbuck, A. L., et al. (2009), 'Targeting matrix metalloproteinases in inflammatory conditions', *Curr Drug Targets*, 10, 1245-54.
- Corpet, F. (1988), 'Multiple sequence alignment with hierarchical clustering', *Nucl Acids Res*, 16, 10881-90.
- Coyne, M. (2000), 'Microbiología del suelo: un enfoque exploratorio' (Madrid: Cengage Learning).
- Chabala, J. C., et al. (1980), 'Ivermectin, a new broad-spectrum antiparasitic agent', *J Med Chem*, 23, 1134-6.
- Chater, K. F., et al. (2010), 'The complex extracellular biology of *Streptomyces*', *FEMS Microbiol Rev*, 34, 171-98.
- Chayen, N. E., Saridakis, E. and Sear, R. P. (2006), 'Experiment and theory for heterogeneous nucleation of protein crystals in a porous medium', *Proc Natl Acad Sci U S A*, 103, 597-601.
- Chen, P., et al. (1992), 'Reactive sites of an anticarcinogenic Bowman-Birk proteinase inhibitor are similar to other trypsin inhibitors', *J Biol Chem*, 267, 1990-4.
- Chen., V. B., et al. (2010), 'MolProbity: all-atom structure validation for macromolecular crystallography', *Acta Crystallogr D*, 66, 12-21.

## D

- Dalbey, R. E., et al. (1997), 'The chemistry and enzymology of the type I signal peptidases', *Protein Sci*, 6, 1129-38.
- Dattagupta, J. K., et al. (1999), 'Refined crystal structure (2.3 Å) of a double-headed winged bean  $\alpha$ -chymotrypsin inhibitor and location of its second reactive site', *Proteins*, 35, 321-31.
- Daza, A., et al. (1989), 'Sporulation of several species of *Streptomyces* in submerged cultures after nutritional downshift', *J Gen Microbiol*, 135, 2483-91.
- Debreczeni, J. E., et al. (2003), 'In-house phase determination of the lima bean trypsin inhibitor: a low-resolution sulfur-SAD case', *Acta Crystallogr D*, 59, 393-5.
- DeSanti, C. L. and Strohl, W. R. (2003), 'Characterization of the *Streptomyces* sp. strain C5 *snp* locus and development of *snp*-derived expression vectors', *Appl Environ Microbiol*, 69, 1647-54.



Dubin, P. L. and Principi, J. M. (1989), 'Optimization of size-exclusion separation of proteins on a Superose column', *J Chromatogr*, 479, 159-64.

Duggar, B. M. (1948), 'Aureomycin; a product of the continuing search for new antibiotics', *Ann N Y Acad Sci*, 51, 177-81.

Dyson, P. (2011), '*Streptomyces: Molecular Biology and Biotechnology*' (Norfolk: Caister Academic Press).

## E

Eggers, C. T., et al. (2001), 'The role of ecotin dimerization in protease inhibition', *J Mol Biol*, 308, 975-91.

Ehrlich, J., Gottlieb, D. and et al. (1948), '*Streptomyces venezuelae*, N. sp., the source of chloromycetin', *J Bacteriol*, 56, 467-77.

Ekici, O. D., Paetzel, M. and Dalbey, R. E. (2008), 'Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration', *Protein Sci*, 17 (12), 2023-37.

Eliopoulos, G. M., et al. (1986), 'In vitro and in vivo activity of LY 146032, a new cyclic lipopeptide antibiotic', *Antimicrob Agents Chemother*, 30, 532-5.

Elkins, P. A., et al. (2002), 'Structure of the C-terminally truncated human ProMMP9, a gelatin-binding matrix metalloproteinase', *Acta Crystallogr D*, 58, 1182-92.

Emsley, P., et al. (2010), 'Features and development of Coot', *Acta Crystallogr D*, 66, 486-501.

Epaulard, O., et al. (2010), 'Multiple aggressive squamous cell carcinomas associated with prolonged voriconazole therapy in four immunocompromised patients', *Clin Microbiol Infect*, 16, 1362-4.

Evans, P. (2006), 'Scaling and assessment of data quality', *Acta Crystallogr D*, 62, 72-82.

Evans, P. and McCoy, A. (2008), 'An introduction to molecular replacement', *Acta Crystallogr D*, 64, 1-10.

## F

Fallon, J. P., Reeves, E. P. and Kavanagh, K. (2011), 'The *Aspergillus fumigatus* toxin fumagillin suppresses the immune response of *Galleria mellonella* larvae by inhibiting the action of haemocytes', *Microbiology*, 157, 1481-8.

- Fehlhammer, H., Bode, W. and Huber, R. (1977), 'Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin', *J Mol Biol*, 111, 415-38.
- Fox, B. P. and Wang, H. (2006), 'Expression of *Streptomyces* subtilisin inhibitor (SSI) proteins in *Bacillus* and *Streptomyces sp.*', Genencor International, U. S. Patent Application 20080090745, October 13.
- Francart, C., et al. (1997), 'Solution structure of R-elafin, a specific inhibitor of elastase', *J Mol Biol*, 268, 666-77.
- Fredrickson, A. G. and Stephanopoulos, G. (1981), 'Microbial competition', *Science*, 213, 972-9.
- Freer, S. T., et al. (1970), 'Chymotrypsinogen: 2.5 Å crystal structure, comparison with α-chymotrypsin, and implications for zymogen activation', *Biochemistry*, 9, 1997-2009.
- Fujinaga, M., et al. (2004), 'The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*', *Proc Natl Acad Sci U S A*, 101, 3364-9.

## G

- Ganz, P. J., et al. (2004), 'Stabilized variant of *Streptomyces* subtilisin inhibitor and its use in stabilizing subtilisin BPN', *Protein Eng Des Sel*, 17, 333-9.
- Gerber, N. N. and Lechevalier, H. A. (1965), 'Geosmin, an earthy-smelling substance isolated from Actinomycetes', *Appl Microbiol*, 13, 935-8.
- Gold, A. M. and Fahrney, D. (1964), 'Sulfonyl fluorides as inhibitors of esterases. Formation and reactions of phenylmethanesulfonyl α-chymotrypsin', *Biochemistry*, 3, 783-91.
- Gomez-Ortiz, M., et al. (1997), 'Inhibition of carboxypeptidase A by excess zinc: analysis of the structural determinants by X-ray crystallography', *FEBS Lett*, 400, 336-40.
- Gomis-Ruth, F. X. (2003), 'Structural aspects of the metzincin clan of metalloendopeptidases', *Mol Biotechnol*, 24, 157-202.
- Gomis-Ruth, F. X. (2008), 'Structure and mechanism of metallocarboxypeptidases', *Crit Rev Biochem Mol Biol*, 43, 319-45.
- Gomis-Ruth, F. X. (2009), 'Catalytic domain architecture of metzincin metalloproteases', *J Biol Chem*, 284, 15353-7.

- Gomis-Ruth, F. X., Kress, L. F. and Bode, W. (1993), 'First structure of a snake venom metalloproteinase: a prototype for matrix metalloproteinases/collagenases', *EMBO J*, 12, 4151-7.
- Gomis-Ruth, F. X., et al. (1994), 'Crystal structures, spectroscopic features, and catalytic properties of cobalt(II), copper(II), nickel(II), and mercury(II) derivatives of the zinc endopeptidase astacin. A correlation of structure and proteolytic activity', *J Biol Chem*, 269, 17111-7.
- Gomis-Ruth, F. X., et al. (1997), 'Mechanism of inhibition of the human matrix metalloproteinase stromelysin-1 by TIMP-1', *Nature*, 389, 77-81.
- Goulas, T., Arolas, J. L. and Gomis-Ruth, F. X. (2011), 'Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog', *Proc Natl Acad Sci U S A*, 108, 1856-61.
- Graycar, T. P., Bott, R. R. and Wilson, L. J. (1994), 'Subtilisin variants', Genencor International, European Patent 0723590, October 13.
- Gross, J. and Lapiere, C. M. (1962), 'Collagenolytic activity in amphibian tissues: a tissue culture assay', *Proc Natl Acad Sci U S A*, 48, 1014-22.
- Guevara, T., et al. (2010), 'Proenzyme structure and activation of astacin metalloproteinase', *J Biol Chem*, 285, 13958-65.
- Guo, J. P. and Ma, Y. (2008), 'High-level expression, purification and characterization of recombinant *Aspergillus oryzae* alkaline protease in *Pichia pastoris*', *Protein Expr Purif*, 58, 301-8.

## H

- Hamada, K., et al. (1996), 'Crystal structure of *Serratia* protease, a zinc-dependent proteinase from *Serratia sp.* E-15, containing a  $\beta$ -sheet coil motif at 2.0 Å resolution', *J Biochem*, 119, 844-51.
- Harada, S., et al. (1991), 'Crystallization and main-chain structure of neutral protease from *Streptomyces caespitosus*', *J Biochem*, 110, 46-9.
- Harada, S., et al. (1995), 'Complete amino acid sequence of a zinc metalloendoprotease from *Streptomyces caespitosus*', *Eur J Biochem*, 233, 683-6.
- Hartley, B. S. (1970), 'Homologies in serine proteinases', *Philos Trans R Soc Lond B Biol Sci*, 257, 77-87.

- Hastrup, S., et al. (1989), 'Mutated subtilisin genes', Novo Nordisk A/s, European Patent 0396608, January 6.
- Hazen, E. L. and Brown, R. (1951), 'Fungicidin, an antibiotic produced by a soil Actinomycete', *Proc Soc Exp Biol Med*, 76, 93-7.
- Hedstrom, L. (2002), 'Serine protease mechanism and specificity', *Chem Rev*, 102, 4501-24.
- Hege, T., et al. (2001), 'Crystal structure of a complex between *Pseudomonas aeruginosa* alkaline protease and its cognate inhibitor: inhibition by a zinc-NH<sub>2</sub> coordinative bond', *J Biol Chem*, 276, 35087-92.
- Herbrecht, R., et al. (2002), 'Voriconazole versus amphotericin B for primary therapy of invasive aspergillosis', *N Engl J Med*, 347, 408-15.
- Hiraga, K., Suzuki, T. and Oda, K. (2000), 'A novel double-headed proteinaceous inhibitor for metalloproteinase and serine proteinase', *J Biol Chem*, 275, 25173-9.
- Hirano, S., et al. (2006), 'Control of the *Streptomyces* subtilisin inhibitor gene by AdpA in the A-factor regulatory cascade in *Streptomyces griseus*', *J Bacteriol*, 188, 6207-16.
- Hirono, S., et al. (1984), 'Crystal structure at 2.6 Å resolution of the complex of subtilisin BPN' with *Streptomyces* subtilisin inhibitor', *J Mol Biol*, 178, 389-414.
- Holm, L., et al. (2006), 'Using Dali for structural comparison of proteins', *Curr Protoc Bioinformatics*, Chapter 5, Unit 5 5.
- Holzer, H. and Heinrich, P. C. (1980), 'Control of proteolysis', *Annu Rev Biochem*, 49, 63-91.
- Hopwood, D.A. (2007), '*Streptomyces* in nature and medicine: The antibiotic makers' (New York: Oxford University Press).
- Huber, R. (1965), 'Die automatisierte Faltmolekülmethode', *Acta Crystallogr A*, 19, 353-56.
- I
- Ikeda, H., et al. (2003), 'Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*', *Nat Biotechnol*, 21, 526-31.
- Ikemura, T. (1981), 'Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system', *J Mol Biol*, 151, 389-409.
- Imada, C., Maeda, M, and Taga, N (1985), 'Purification and characterization of the protease inhibitor "monastatin" from a marine *Alteromonas* sp. with reference to inhibition

of the protease produced by a bacterium pathogenic to fish', *Can J Microbiol*, 31, 1089–94.

Inouye, K., Shimada, T. and Yasukawa, K. (2007), 'Effects of neutral salts and alcohols on the activity of *Streptomyces caespitosus* neutral protease', *J Biochem*, 142, 317-24.

International Union of Biochemistry (1984), 'Enzyme nomenclature 1984: recommendations of the Nomenclature Committee of the International Union of Biochemistry on the nomenclature and classification of enzyme-catalysed reactions' (New York: Academic Press).

Irving, J. A., et al. (2000), 'Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function', *Genome Res*, 10, 1845-64.

Irving, J. A., et al. (2002), 'Serpins in prokaryotes', *Mol Biol Evol*, 19, 1881-90.

## J

Jozic, D., et al. (2005), 'X-ray structure of human proMMP-1: new insights into procollagenase activation and collagen binding', *J Biol Chem*, 280, 9578-85.

## K

Kabsch, W. (2010), 'XDS.', *Acta Crystallogr D*, 66, 125-32.

Kajiwara, K., et al. (1991), 'Interactions of *Streptomyces* serine-protease inhibitors with *Streptomyces griseus* metalloendopeptidase II', *J Biochem*, 110, 350-4.

Kamiya, N., Matsushima, M. and Sugino, H. (1984), 'The three-dimensional structure of plasminostreptin, a bacterial protein protease inhibitor, at 2.8 Å resolution', *Bull Chem Soc Jpn*, 57, 2075-81.

Kato, J. Y., et al. (2005), 'The *Streptomyces* subtilisin inhibitor (SSI) gene in *Streptomyces coelicolor* A3(2)', *Biosci Biotechnol Biochem*, 69, 1624-9.

Khan, A. R. and James, M. N. (1998), 'Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes', *Protein Sci*, 7, 815-36.

Kino, T., et al. (1987), 'FK-506, a novel immunosuppressant isolated from a *Streptomyces*. Fermentation, isolation, and physico-chemical and biological characteristics', *J Antibiot*, 40, 1249-55.

- Koepke, J., et al. (2000), 'Crystal structure of cancer chemopreventive Bowman-Birk inhibitor in ternary complex with bovine trypsin at 2.3 Å resolution. Structural basis of Janus-faced serine protease inhibitor specificity', *J Mol Biol*, 298, 477-91.
- Kojima, S., et al. (1990), 'Alteration of the specificity of the *Streptomyces* subtilisin inhibitor by gene engineering', *Biotechnology*, 8, 449-52.
- Kojima, S., et al. (1991), 'Inhibition of subtilisin BPN' by reaction site P1 mutants of *Streptomyces* subtilisin inhibitor', *J Biochem*, 109, 377-82.
- Kolattukudy, P. E., Markaryan, A. N., Copelan, E. A., Ceselski, S. K. (1995), 'Prevention of invasive pulmonary aspergillosis with serine proteinase inhibitors', The Ohio State Research Foundation, U. S. Patent 5739283, June 7.
- Kossiakoff, A. A. and Spencer, S. A. (1980), 'Neutron diffraction identifies His 57 as the catalytic base in trypsin', *Nature*, 288, 414-6.
- Krahn, J. and Stevens, F. C. (1971), 'Stoichiometry of the interaction of lima bean protease inhibitor with trypsin and/or chymotrypsin', *FEBS Lett*, 13, 339-41.
- Krowarsch, D., et al. (2003), 'Canonical protein inhibitors of serine proteases', *Cell Mol Life Sci*, 60, 2427-44.
- Kubier, A. and O'Brien, M. (2012), 'Endogenous anticoagulants', *Top Companion Anim Med*, 27, 81-7.
- Kumar, J. K. (2008), 'Lysostaphin: an antistaphylococcal agent', *Appl Microbiol Biotechnol*, 80, 555-61.
- Kumazaki, T., et al. (1993), 'Interaction of *Streptomyces* subtilisin inhibitor (SSI) with *Streptomyces griseus* metallo-endopeptidase II (SGMP II)', *J Biochem*, 114, 570-5.
- Kurusu, G., Kai, Y. and Harada, S. (2000), 'Structure of the zinc-binding site in the crystal structure of a zinc endoprotease from *Streptomyces caespitosus* at 1 Å resolution', *J Inorg Biochem*, 82, 225-8.
- Kurusu, G., et al. (1997), 'Structure of the zinc endoprotease from *Streptomyces caespitosus*', *J Biochem*, 121, 304-8.

## L

- Larsen, R. J. and Marx, M. L. (2010), 'An introduction to mathematical statistics and its applications' (Boston: Prentice Hall PTR).

- Laskowski Jr., M. and Kato, I. (1980), 'Protein inhibitors of proteinases', *Annu Rev Biochem*, 49, 593-626.
- Laskowski, M. and Qasim, M. A. (2000), 'What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme substrate complexes?', *Biochim Biophys Acta*, 1477, 324-37.
- Lenarcic, B. and Turk, V. (1999), 'Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D', *J Biol Chem*, 274, 563-6.
- Lenarcic, B., et al. (1997), 'Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain', *J Biol Chem*, 272, 13899-903.
- Leung, D., Abbenante, G. and Fairlie, D. P. (2000), 'Protease inhibitors: current status and future prospects', *J Med Chem*, 43, 305-41.
- Li, J., et al. (1995), 'Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed  $\beta$ -propeller', *Structure*, 3, 541-9.
- Lowe, J., et al. (1995), 'Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution', *Science*, 268, 533-9.

## M

- Malapeira, J., et al. (2011), 'ADAM17 (TACE) regulates TGF $\beta$  signaling through the cleavage of vasorin', *Oncogene*, 30, 1912-22.
- Maldonado-Salvador, F. A. (1984), 'Proteasas de *Streptomyces fradiae*: obtención por fermentación y aplicaciones industriales', Tesis Doctoral, (Madrid: Universidad Complutense de Madrid).
- Markaryan, A., Beall, C. J. and Kolattukudy, P. E. (1996), 'Inhibition of *Aspergillus* serine proteinase by *Streptomyces subtilisin* inhibitor and high-level expression of this inhibitor in *Pichia pastoris*', *Biochem Biophys Res Commun*, 220, 372-6.
- Marrero, A., et al. (2012), 'The crystal structure of human  $\alpha_2$ -macroglobulin reveals a unique molecular cage', *Angew Chem Int Ed Engl*, 51, 3340-4.
- Maskos, K., et al. (2007), 'Flexibility and variability of TIMP binding: X-ray structure of the complex between collagenase-3/MMP-13 and TIMP-2', *J Mol Biol*, 366, 1222-31.

- Matheson, N. R., van Halbeek, H. and Travis, J. (1991), 'Evidence for a tetrahedral intermediate complex during serpin-proteinase interactions', *J Biol Chem*, 266, 13489-91.
- Matthews, B. W. (1968), 'Solvent content of protein crystals', *J Mol Biol*, 33, 491-7.
- Matthews, B. W., et al. (1967), 'Three-dimensional structure of tosyl- $\alpha$ -chymotrypsin', *Nature*, 214, 652-6.
- McCormick, M. H., et al. (1955), 'Vancomycin, a new antibiotic. Chemical and biologic properties', *Antibiot Annu*, 3, 606-11.
- McCoy, A. J., et al. (2007), 'Phaser crystallographic software', *J Appl Crystallogr*, 40, 658-74.
- McPhalen, C. A. and James, M. N. (1987), 'Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds', *Biochemistry*, 26, 261-9.
- Meeks, S. L. and Abshire, T. C. (2008), 'Abnormalities of prothrombin: a review of the pathophysiology, diagnosis, and treatment', *Haemophilia*, 14, 1159-63.
- Meulenbroek, E. M., et al. (2012), 'Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato', *Acta Crystallogr D*, 68, 794-9.
- Meyer, E., et al. (1988), 'Structure of native porcine pancreatic elastase at 1.65 Å resolutions', *Acta Crystallogr B*, 44, 26-38.
- Migliolo, L., et al. (2010), 'Structural and mechanistic insights into a novel non-competitive Kunitz trypsin inhibitor from *Adenanthera pavonina* L. seeds with double activity toward serine- and cysteine-proteinases', *J Mol Graphics Modell*, 29, 148-56.
- Mitsui, Y., et al. (1977), 'Crystal structure of a protein proteinase inhibitor, *Streptomyces subtilisin* inhibitor, at 2.3 Å resolution', *J Biochem*, 82, 295-98.
- Mitsui, Y., et al. (1979), 'Crystal structure of a bacterial protein proteinase inhibitor (*Streptomyces subtilisin* inhibitor) at 2.6 Å resolution', *J Mol Biol*, 131, 697-724.
- Monod, M., et al. (2002), 'Secreted proteases from pathogenic fungi', *Int J Med Microbiol*, 292, 405-19.
- Morgunova, E., et al. (1999), 'Structure of human pro-matrix metalloproteinase-2: activation mechanism revealed', *Science*, 284, 1667-70.
- Murao, S., Sato, S. (1972), 'S-SI, a new alkaline protease inhibitor from *Streptomyces albobrisesolus* S-3253', *Agr Biol Chem*, 36, 4.
- Murphy, G. (2011), 'Tissue inhibitors of metalloproteinases', *Genome Biol*, 12, 233.



---

Murphy, G., et al. (1999), 'Mechanisms for pro matrix metalloproteinase activation', *APMIS*, 107, 38-44.

Murshudov, G.N., et al. (2011), 'REFMAC5 for the refinement of macromolecular crystal structures', *Acta Crystallogr D*, 67, 355-67.

## N

Nagase, H. (1997), 'Activation mechanisms of matrix metalloproteinases', *Biol Chem*, 378, 151-60.

Nagase, H. (2001), 'Metalloproteases', *Curr Protoc Protein Sci*, Chapter 21, Unit 21 4.

Nagase, H. and Brew, K. (2002), 'Engineering of tissue inhibitor of metalloproteinases mutants as potential therapeutics', *Arthritis Res*, 4 Suppl 3, S51-61.

Nair, S., Tsukamoto, K. and Shimidu, U. (1985), 'Distribution of bacteriolytic bacteria in the coastal marine environments of Japan', *Bull Japan Soc Sci Fish*, 51, 1469-73.

Nakae, K., et al. (2000), 'Migrastatin, a novel 14-membered lactone from *Streptomyces* sp. MK929-43F1', *J Antibiot*, 53, 1228-30.

Neidhart, D. J. and Petsko, G. A. (1988), 'The refined crystal structure of subtilisin Carlsberg at 2.5 Å resolution', *Protein Eng*, 2, 271-76.

Neurath, H. (1991), 'Proteolytic processing and regulation', *Enzyme*, 45, 239-43.

Neurath, H. and Walsh, K. A. (1976), 'Role of proteolytic enzymes in biological regulation (a review)', *Proc Natl Acad Sci U S A*, 73, 3825-32.

Nohara, D., et al. (2004), 'High performance in protease refolding by application of a continuous system using immobilized inhibitor', *J Biosci Bioeng*, 98, 482-6.

Nohara, D., et al. (1999), 'High performance in refolding of *Streptomyces griseus* trypsin by the aid of a mutant of *Streptomyces* subtilisin inhibitor designed as trypsin inhibitor', *J Biochem*, 125, 343-7.

Nomura, K. and Suzuki, N. (1993), 'Stereo-specific inhibition of sea urchin envelysin (hatching enzyme) by a synthetic autoinhibitor peptide with a cysteine-switch consensus sequence', *FEBS Lett*, 321, 84-8.

Norioka, S. and Ikenaka, T. (1983), 'Amino acid sequences of trypsin-chymotrypsin inhibitors (A-I, A-II, B-I, and B-II) from peanut (*Arachis hypogaea*): a discussion on the molecular evolution of legume Bowman-Birk type inhibitors', *J Biochem*, 94, 589-99.

**O**

- Odani, S., Koide, T. and Ono, T. (1986), 'Wheat germ trypsin inhibitors. Isolation and structural characterization of single-headed and double-headed inhibitors of the Bowman-Birk type', *J Biochem*, 100, 975-83.
- Ohno, A., et al. (1998), 'NMR structure of the *Streptomyces* metalloproteinase inhibitor, SMPI, isolated from *Streptomyces nigrescens* TK-23: another example of an ancestral  $\beta$   $\gamma$ -crystallin precursor structure', *J Mol Biol*, 282, 421-33.
- Otlewski, J., et al. (2005), 'The many faces of protease-protein inhibitor interaction', *EMBO J*, 24, 1303-10.
- Oura, M., Sternberg, T. H. and Wright, E. T. (1955), 'A new antifungal antibiotic, amphotericin B', *Antibiot Annu*, 3, 566-73.

**P**

- Page, M. J. and Di Cera, E. (2008), 'Serine peptidases: classification, structure and function', *Cell Mol Life Sci*, 65, 1220-36.
- Park, E. Y., et al. (2004), 'Crystal structure of the Bowman-Birk inhibitor from barley seeds in ternary complex with porcine trypsin', *J Mol Biol*, 343, 173-86.
- Pasic, M. D., et al. (2012), 'Epigenetic regulation of kallikrein-related peptidases: there is a whole new world out there', *Biol Chem*, 393, 319-30.
- Peng, Y., et al. (2003), 'Purification and characterization of a fibrinolytic enzyme produced by *Bacillus amyloliquefaciens* DC-4 screened from douchi, a traditional Chinese soybean food', *Comp Biochem Physiol B*, 134, 45-52.
- Perona, J. J. and Craik, C. S. (1995), 'Structural basis of substrate specificity in the serine proteases', *Protein Sci*, 4, 337-60.
- Petersen, T. N., et al. (2011), 'SignalP 4.0: discriminating signal peptides from transmembrane regions', *Nat Methods*, 8, 785-6.
- Pettersen, E.F., et al. (2004), 'UCSF Chimera--a visualization system for exploratory research and analysis', *J Comput Chem*, 25, 1605-12.
- Pitterna, T., et al. (2009), 'New ventures in the chemistry of avermectins', *Bioorg Med Chem*, 17, 4085-95.
- Polaina, J. and MacCabe, A. P. (2007), 'Industrial Enzymes: Structure, Function and Applications' (Dordrecht: Springer).

- Polgar, L. (2005), 'The catalytic triad of serine peptidases', *Cell Mol Life Sci*, 62, 2161-72.
- Pound, M. W., et al. (2011), 'Overview of treatment options for invasive fungal infections', *Med Mycol*, 49, 561-80.

## R

- Ragg, E. M., et al. (2006), 'Inhibitory properties and solution structure of a potent Bowman-Birk protease inhibitor from lentil (*Lens culinaris*, L) seeds', *Eur J Biochem*, 273, 4024-39.
- Rao, K. N. and Suresh, C. G. (2007), 'Bowman-Birk protease inhibitor from the seeds of *Vigna unguiculata* forms a highly stable dimeric structure', *Biochim Biophys Acta*, 1774, 1264-73.
- Rawlings, N. D. and Barrett, A. J. (1993), 'Evolutionary families of peptidases', *Biochem J*, 290, 205-18.
- Rawlings, N. D., Barrett, A. J. and Bateman, A. (2012), 'MEROPS: the database of proteolytic enzymes, their substrates and inhibitors', *Nucleic Acids Res*, 40, D343-50.
- Reading, C. and Cole, M. (1977), 'Clavulanic acid: a  $\beta$ -lactamase-inhibiting  $\beta$ -lactam from *Streptomyces clavuligerus*', *Antimicrob Agents Chemother*, 11, 852-7.
- Reddy, A., Schneemann, A. and Johnson, J. E. (2004), 'Nodavirus endopeptidase', in A. J. Barrett, N. D. Rawlings and J. F. Woessner (eds.), *Handbook of Proteolytic Enzymes* (2 edn.; London: Elsevier), 197-201.
- Rees, D. C. and Lipscomb, W. N. (1982), 'Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution', *J Mol Biol*, 160, 475-98.
- Renko, M., et al. (2010), 'Versatile loops in mycocypins inhibit three protease families', *J Biol Chem*, 285, 308-16.
- Reverter, D., et al. (2000), 'Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2', *Nat Struct Biol*, 7, 322-8.
- Reyda, S., et al. (1999), 'cDNA cloning, bacterial expression, in vitro renaturation and affinity purification of the zinc endopeptidase astacin', *Biochem J*, 344, 851-7.
- Roderick, S. L. and Matthews, B. W. (1993), 'Structure of the cobalt-dependent methionine aminopeptidase from *Escherichia coli*: a new type of proteolytic enzyme', *Biochemistry*, 32, 3907-12.

**S**

- Sakiyama, F. and Kawata, Y. (1983), 'NMR titration studies of histidine 57 and the [methylene-<sup>13</sup>C]PMS group in the phenylmethanesulfonyl (PMS) derivative of *Streptomyces erythraeus* trypsin', *J Biochem*, 94, 1661-9.
- Salvesen, G. S. and Barrett, A. J. (1980), 'Covalent binding of proteinases in their reaction with  $\alpha_2$ -macroglobulin', *Biochem J*, 187, 695-701.
- Saridakis, E. and Chayen, N. E. (2009), 'Towards a 'universal' nucleant for protein crystallization', *Trends Biotechnol*, 27, 99-106.
- Satow, Y., Watanabe, Y. and Mitsui, Y. (1980), 'Solvent accessibility and microenvironment in a bacterial protein proteinase inhibitor SSI (*Streptomyces* subtilisin inhibitor)', *J Biochem*, 88, 1739-55.
- Saunders, C. W. (1999), 'Proteases fused with variants of *Streptomyces* subtilisin Inhibitor', Procter & Gamble Company, WIPO Patent 2000001831, July 7.
- Saunders, C. W., et al. (1999), 'Stabilized variants of *Streptomyces* subtilisin Inhibitor', Procter & Gamble Company, WIPO Patent 2000001826, July 7.
- Schaffer, L., et al. (1963), 'Lincomycin--a New Antibiotic', *Clin Pediatr*, 2, 642-5.
- Schechter, I. and Berger, A. (1967), 'On the size of the active site in proteases. Papain', *Biochem Biophys Res Commun*, 27, 157-62.
- Schrodinger, LLC (2010), 'The PyMOL Molecular Graphics System, Version 1.3r1'.
- Seeram, S. S., Hiraga, K. and Oda, K. (1997), 'Resynthesis of reactive site peptide bond and temporary inhibition of *Streptomyces* metalloproteinase inhibitor', *J Biochem*, 122, 788-94.
- Sharp, P. M. and Li, W. H. (1987), 'The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications', *Nucleic Acids Res*, 15, 1281-95.
- Sheldrick, G. M. (2010), 'Experimental phasing with SHELXC/D/E: combining chain tracing with density modification.', *Acta Crystallogr D*, 66, 479-85.
- Shibata, M., et al. (1997), 'A novel metalloproteinase, almelysin, from a marine bacterium, *Alteromonas* sp. No. 3696: purification and characterization', *Biosci Biotechnol Biochem*, 61, 710-5.

- Sin, N., et al. (1997), 'The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2', *Proc Natl Acad Sci U S A*, 94, 6099-103.
- Soff, G. A. (2012), 'A new generation of oral direct anticoagulants', *Arterioscler Thromb Vasc Biol*, 32, 569-74.
- Song, H. K. and Suh, S. W. (1998), 'Kunitz-type soybean trypsin inhibitor revisited: refined structure of its complex with porcine trypsin reveals an insight into the interaction between a homologous inhibitor from *Erythrina caffra* and tissue-type plasminogen activator', *J Mol Biol*, 275, 347-63.
- Sreerama, Y. N. and Gowda, L. R. (1997), 'Antigenic determinants and reactive sites of a trypsin/chymotrypsin double-headed inhibitor from horse gram (*Dolichos biflorus*)', *Biochim Biophys Acta*, 1343, 235-42.
- Stainer, R. Y., et al. (1991), 'Microbiología' (2 edn.; Barcelona: Editorial Reverté).
- Stapley, E. O., et al. (1969), 'Phosphonomycin. Discovery and in vitro biological characterization', *Antimicrob Agents Chemother*, 9, 284-90.
- Stensmyr, M. C., et al. (2012), 'A conserved dedicated olfactory circuit for detecting harmful microbes in *Drosophila*', *Cell*, 151, 1345-57.
- Strickler, J. E., et al. (1992), 'Two novel *Streptomyces* protein protease inhibitors. Purification, activity, cloning, and expression', *J Biol Chem*, 267, 3236-41.
- Struyk, A. P., et al. (1957), 'Pimaricin, a new antifungal antibiotic', *Antibiot Annu*, 5, 878-85.
- Stura, E. A. and Wilson, I. A. (1991), 'Applications of the streak seeding technique in protein crystallization', *J Cryst Growth*, 110, 13.
- Suzuki, M., et al. (1997), 'A novel member of the subtilisin-like protease family from *Streptomyces albogriseolus*', *J Bacteriol*, 179, 430-8.

## T

- Taguchi, S. (1995), 'Communication between protease and protease inhibitor in the *Streptomyces* world', *Actinomycetologica*, 9, 12.
- Taguchi, S., et al. (1995), 'Molecular characterization of a gene encoding extracellular serine protease isolated from a subtilisin inhibitor-deficient mutant of *Streptomyces albogriseolus* S-3253', *Appl Environ Microbiol*, 61, 180-6.
- Takeuchi, Y., et al. (1991a), 'Refined crystal structure of the complex of subtilisin BPN' and *Streptomyces* subtilisin inhibitor at 1.8 Å resolution', *J Mol Biol*, 221, 309-25.

- Takeuchi, Y., et al. (1992), 'Crystal structure of an engineered subtilisin inhibitor complexed with bovine trypsin', *Proc Natl Acad Sci U S A*, 89, 4407-11.
- Takeuchi, Y., et al. (1991b), 'Molecular recognition at the active site of subtilisin BPN': crystallographic studies using genetically engineered proteinaceous inhibitor SSI (*Streptomyces subtilisin inhibitor*)', *Protein Eng*, 4, 501-08.
- Tallant, C., et al. (2006), 'Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases', *J Biol Chem*, 281, 17920-8.
- Tan, F. L., et al. (1984), 'Purification, characterization and sequence determination of a double-headed trypsin inhibitor peptide from *Trichosanthes kirilowii* (a Chinese medical herb)', *H-S Z Physiol Chem*, 365, 1211-7.
- Tanaka, A. S., et al. (1997), 'Purification and primary structure determination of a Bowman-Birk trypsin inhibitor from *Torresea cearensis* seeds', *J Biol Chem*, 378, 273-81.
- Terada, S., Fujimura, S. and Kimoto, E. (1994), 'Amino acid sequences of double-headed proteinase inhibitors from the seeds of *Canavalia lineata*', *Biosci, Biotechnol, Biochem*, 58, 376-9.
- Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994), 'CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Res*, 22, 4673-80.
- Trincao, C., et al. (1954), 'Effects of a new antibiotic (puromycin) on sleeping disease (*T. gambiense*)', *Med Contemp*, 72, 391-2.

## U

- Umezawa, H. (1965), 'Bleomycin and other antitumor antibiotics of high molecular weight', *Antimicrob Agents Chemother*, 5, 1079-85.
- Umezawa, H., et al. (1957), 'Production and isolation of a new antibiotic: kanamycin', *J Antibiot*, 10, 181-8.

## V

- Valueva, T. A., et al. (2012), 'Structure and properties of the potato chymotrypsin inhibitor', *Plant Physiol Bioch*, 52, 83-90.
- van Ee, J.H. and Misset, O. (1997), 'Enzymes in Detergency' (New York: Marcel Dekker).

- Veldkamp, H., et al. (1984), 'Competition among bacteria: An overview', in A. J. Klug and C.A. Reddy (eds.), *Current Perspectives in Microbial Ecology* (Washington, D.C.: American Society for Microbiology).
- Verschuere, L., et al. (2000), 'Probiotic bacteria as biological control agents in aquaculture', *Microbiol Mol Biol Rev*, 64, 655-71.
- Vezina, C., Kudelski, A. and Sehgal, S. N. (1975), 'Rapamycin (AY-22,989), a new antifungal antibiotic. Taxonomy of the producing Streptomyces and isolation of the active principle', *J Antibiot*, 28, 721-6.
- Vitali, J., et al. (1992), 'The structure of a complex of bovine  $\alpha$ -thrombin and recombinant hirudin at 2.8 Å resolution', *J Biol Chem*, 267, 17670-8.
- Vriend, G. (1990), 'WHAT IF: a molecular modelling and drug design program', *J Mol Graph*, 8, 52-56.

## W

- Waksman, S. A., Reilly, H. C. and Schatz, A. (1945), 'Strain specificity and production of antibiotic substances: Strain resistance of bacteria to antibiotic substances, especially to streptomycin', *Proc Natl Acad Sci U S A*, 31, 157-64.
- Waksman, S. A., Lechevalier, H. A. and Harris, D. A. (1949), 'Neomycin-Production and Antibiotic Properties', *J Clin Invest*, 28, 934-9.
- Walsh, T. J., et al. (2008), 'Treatment of aspergillosis: clinical practice guidelines of the Infectious Diseases Society of America', *Clin Infect Dis*, 46, 327-60.
- Walter, J., et al. (1982), 'On the disordered activation domain in trypsinogen. Chemical labelling and low-temperature crystallography', *Acta Crystallogr B*, 38, 11.
- Waltersperger, S., et al. (2010), 'Crystal structure of archaemetzincin AmzA from *Methanopyrus kandleri* at 1.5 Å resolution', *Proteins*, 78, 2720-3.
- Welsh, O., Vera-Cabrera, L. and Salinas-Carmona, M. C. (2007), 'Mycetoma', *Clin Dermatol*, 25, 195-202.
- Wells, J. A., et al. (1983), 'Cloning, sequencing, and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*', *Nucleic Acids Res*, 11, 7911-25.
- Wetmore, D. R. and Hardman, K. D. (1996), 'Roles of the propeptide and metal ions in the folding and stability of the catalytic domain of stromelysin (matrix metalloproteinase 3)', *Biochemistry*, 35, 6549-58.

- Wheatley, D. N. (1984), 'Intracellular protein degradation: basis of a self-regulating mechanism for the proteolysis of endogenous proteins', *J Theor Biol*, 107, 127-49.
- Wildermuth, H. (1970), 'Development and organization of the aerial mycelium in *Streptomyces coelicolor*', *J Gen Microbiol*, 60, 43-50.
- Wilson, K. A. and Laskowski, M., Sr. (1975), 'The partial amino acid sequence of trypsin inhibitor II from garden bean, *Phaseolus vulgaris*, with location of the trypsin and elastase-reactive sites', *J Biol Chem*, 250, 4261-7.
- Wisniewska, M., et al. (2008), 'Structural determinants of the ADAM inhibition by TIMP-3: crystal structure of the TACE-N-TIMP-3 complex', *J Mol Biol*, 381, 1307-19.
- Wlodawer, A., et al. (1984), 'Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II', *J Mol Biol*, 180, 301-29.
- Wlodawer, A., et al. (2003), 'Structural and enzymatic properties of the sedolisin family of serine-carboxyl peptidases', *Acta Biochim Pol*, 50, 81-102.
- Woessner, J. F., Jr. (1999), 'Matrix metalloproteinase inhibition. From the Jurassic to the third millennium', *Ann N Y Acad Sci*, 878, 388-403.
- Wong, D.W.S. (1995), 'Food Enzymes: Structure and Mechanism' (New York: Springer).
- Wright, C. S., Alden, R. A. and Kraut, J. (1969), 'Structure of subtilisin BPN' at 2.5 Å resolution', *Nature*, 221, 235-42.
- Wu, J. W. and Chen, X. L. (2011), 'Extracellular metalloproteases from bacteria', *Appl Microbiol Biotechnol*, 92, 253-62.

## X

- Xie, Z. W., et al. (1997), 'Two reactive site locations and structure-function study of the arrowhead proteinase inhibitors, A and B, using mutagenesis', *Biochemistry*, 36, 5846-52.

## Y

- Yana, I. and Weiss, S. J. (2000), 'Regulation of membrane type-1 matrix metalloproteinase activation by proprotein convertases', *Mol Biol Cell*, 11, 2387-401.
- Yiallourous, I., et al. (2002), 'Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing', *J Mol Biol*, 324, 237-46.



Yokote, Y., Kawasaki, K., Nakajima, J. and Noguchi, Y. (1969), 'Studies on enzymes produced by *Streptomyces caespitosus*. Production conditions and some properties of neutral proteases.', *Nippon Nôgeikagaku Kaishi*, 43, 7.

## Z

Zaffiri, L., Gardner, J. and Toledo-Pereyra, L. H. (2012), 'History of antibiotics. From salvarsan to cephalosporins', *J Invest Surg*, 25, 67-77.

Zakharova, E., Horvath, M. P. and Goldenberg, D. P. (2009), 'Structure of a serine protease poised to resynthesize a peptide bond', *Proc Natl Acad Sci U S A*, 106, 11034-9.

Zhou, Z., et al. (2011), 'The -omics era- toward a systems-level understanding of *Streptomyces*', *Curr Genomics*, 12, 404-16.

Zurowski, M. M. and Stabinsky, Y. L. M. (1988), 'Subtilisin analogs', Amgen Inc., European Patent 0309565, March 28.



# APÉNDICE I

Mechanism of action of a Janus-faced single-domain protein inhibitor simultaneously targeting two peptidase classes



## Mechanism of action of a Janus-faced single-domain protein inhibitor simultaneously targeting two peptidase classes†

Cite this: *Chem. Sci.*, 2013, **4**, 791

Sergio Trillo-Muyo,<sup>a</sup> Sergio Martínez-Rodríguez,<sup>b</sup> Joan L. Arolas<sup>\*a</sup> and F. Xavier Gomis-Rüth<sup>\*a</sup>

Protein inhibitors provide a physiological mechanism for the regulation of proteolytic enzymes. While most single-domain inhibitors have one reactive site with which they target peptidases of a specific catalytic class, selected specimens inhibit two peptidase molecules simultaneously, thus giving rise to ternary complexes. To study such inhibition, we analyzed the function of one of these proteins, *sermetstatin*, which strongly binds as a dimer to serine proteinases (SPs) and a metallopeptidase (MP). In addition, we determined the structures of the isolated inhibitor dimer and its heterotetrameric complexes with the SP subtilisin and the MP snapalysin, which reveal that inhibition occurs through two independent distal reactive sites. These structures and the derived model for the heterohexameric complex provide for the first time a detailed view of the molecular mechanism of simultaneous inhibition of proteinases belonging to two distinct mechanistic classes by a single-domain protein.

Received 9th October 2012

Accepted 13th November 2012

DOI: 10.1039/c2sc21712k

[www.rsc.org/chemicalscience](http://www.rsc.org/chemicalscience)

### Introduction

Due to the essentially irreversible nature of peptide-bond hydrolysis, proteolytic enzymes must be tightly regulated to prevent spatially and temporally deleterious activity. One such regulatory mechanism is provided by co-localizing protein inhibitors which mostly target one or several peptidases of a certain catalytic class through a single inhibitory region by blocking the active-site cleft of the enzyme.<sup>1,2</sup> A special case is provided by inhibitors with two inhibitory sites which mimic the head architecture of the Roman god Janus and are capable of inhibiting two peptidase molecules simultaneously.<sup>3</sup> Such Janus-faced inhibitors (see ESI Table S1†) are infrequent among the plethora of inhibitors described, and none of them has yet been structurally characterized in a ternary complex with two different targets.

To address this question, we focused on the bacterial genus *Streptomyces*, which occurs in soil and aquatic habitats and is the source of many medicinal, pharmaceutical, and industrial products of natural origin such as antibiotics.<sup>4</sup> These bacteria secrete a number of peptidases and cognate protein inhibitors

thought to participate in morphological development and defense against extrinsic proteinases. Among these inhibitors are members of family I16 (according to MEROPS inhibitor database, <http://merops.sanger.ac.uk> (ref. 5)), which only occur within the bacterial order *Actinomycetales*. These inhibitors generally target serine proteinases (SPs) of the chymotrypsin- and subtilisin-type; some also bind tightly to *Streptomyces griseus* griselysin but apparently not the structural and functional relative thermolysin from *Bacillus thermoproteolyticus*, which is in MEROPS metallopeptidase (MP) family M4.<sup>5</sup> This inhibition is exerted by the same region employed for SPs,<sup>6</sup> so that, although they are bi-functional, dual inhibition cannot occur simultaneously. Previously studied I16 inhibitors include *Streptomyces* subtilisin inhibitor (SSI) from *Streptomyces albogriseolus*, plasminostreptin from *Streptomyces antifibrinolyticus*, alkaline protease inhibitor API-2c' from *Streptomyces pseudogriseolus*, and trypsin inhibitor STI2 from *Streptomyces longisporus*<sup>5,7–11</sup> (see Fig. 1a). Among them is also sermetstatin—also known as *Streptomyces caespitosus* neutral proteinase inhibitor<sup>8</sup>—which is a 113-residue secreted single-domain molecule and the main object of study of the present work.

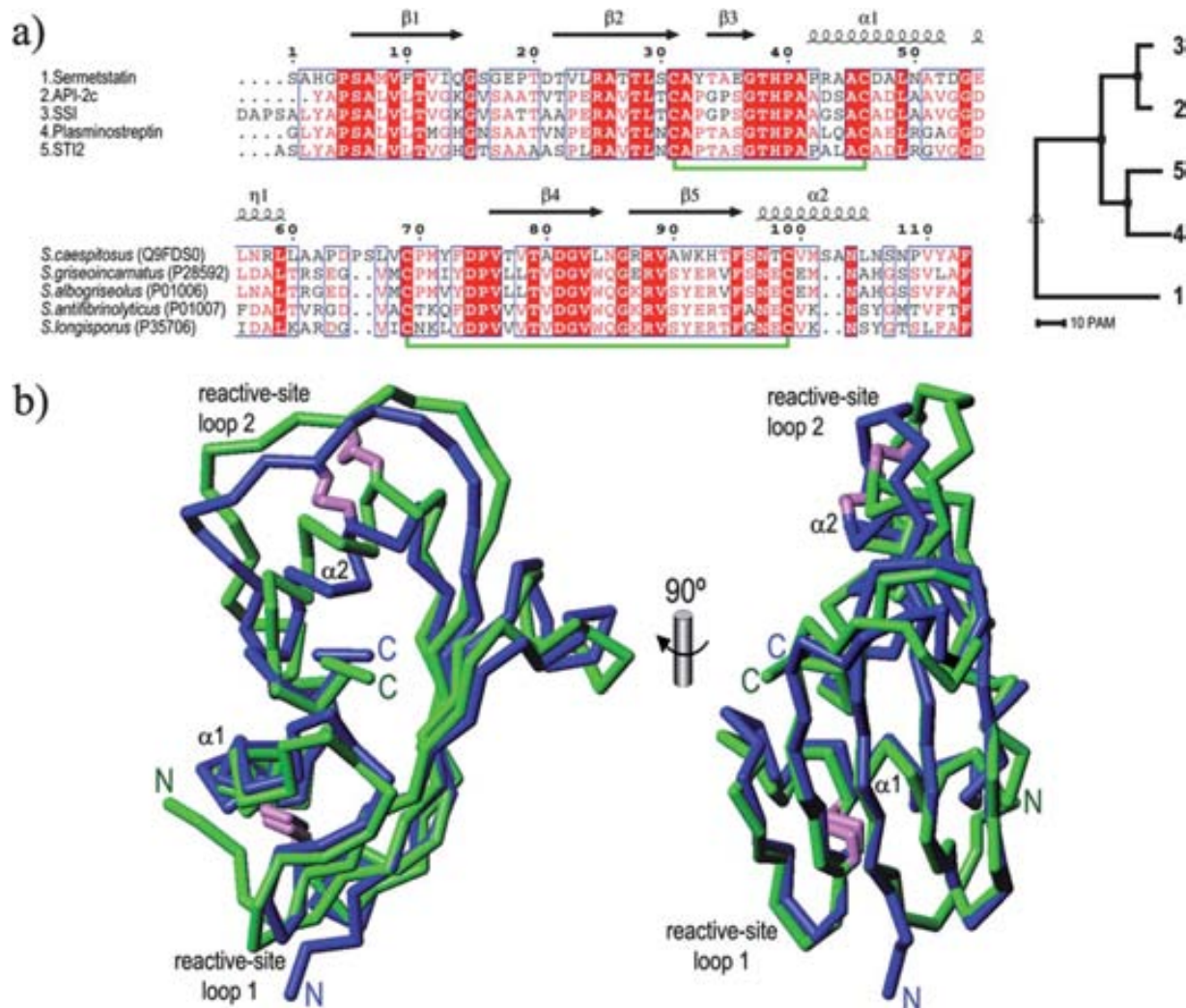
### Inhibitory profile of sermetstatin

We produced sermetstatin by recombinant overexpression in *Escherichia coli* (see *Experimental procedures* in the ESI†) and found it to be a homodimer in solution, thus confirming earlier results.<sup>8</sup> We assessed its inhibitory function, which had been previously ruled out for the cysteine proteinase cathepsin B and the aspartic proteinase pepsin,<sup>8</sup> against a broad panel of model

<sup>a</sup>Proteolysis Lab, Molecular Biology Institute of Barcelona, Spanish Research Council CSIC, Barcelona Science Park; c/Baldiri Reixac, 15-21, 08028 Barcelona, Catalonia, Spain. E-mail: [xgrcri@ibmb.csic.es](mailto:xgrcri@ibmb.csic.es); [jlacri@ibmb.csic.es](mailto:jlacri@ibmb.csic.es); Fax: +34 934034979; Tel: +34 934020186

<sup>b</sup>Department of Physical Chemistry, Biochemistry, and Inorganic Chemistry, University of Almería, 04210 Almería, Spain

† Electronic supplementary information (ESI) available: Experimental procedures, acknowledgements, supplemental references, three supplemental tables and one supplemental figure. See DOI: 10.1039/c2sc21712k



**Fig. 1** (a) (left) Sequence alignment of representative members of the I16 family of peptidase inhibitors according to MEROPS database (<http://merops.sanger.ac.uk>). API-2c' stands for alkaline protease inhibitor-2c', SSI for *Streptomyces subtilisin* inhibitor, and STI2 for trypsin inhibitor STI2. The two disulfide bonds found in these proteins are depicted as green lines. The corresponding organism names ("S." stands for *Streptomyces*) and UniProt database access codes are provided in the second block. The regular secondary-structure elements correspond to the structure of sermetstatin. (right) Phylogenetic analysis of the five inhibitors (numbering corresponds to that of the upper block of the alignment) reflecting the large evolutionary distance of sermetstatin from the other four sequences. PAM = point accepted mutation. (b) (Left) Superposition of the  $\alpha$ -traces of sermetstatin (green) and SSI (blue; PDB 2SIC) taken from their respective complexes with subtilisin in the front reference view and (right) a lateral view. 96 common  $\alpha$  atoms deviating less than 3 Å showed an rmsd of 1.48 Å. The respective termini are labeled as are the reactive-site loops 1 and 2, as well as scaffold-helices  $\alpha$ 1 and  $\alpha$ 2. Disulfide bonds are in pink.

peptidases (ESI Table S2†). To this end, we recombinantly produced and purified the MP snapalysin from *S. caespitosus*, which is a member of the metzincin clan of MPs—unrelated to griselysin and other M4 peptidases.<sup>12–14</sup> We found it is potently inhibited by sermetstatin with an apparent inhibition constant  $K_i$  of 6 nM, which is consistent with reported results.<sup>8</sup> These inhibition studies were complemented by autolytic activation studies with recombinant wild-type prosnapalysin from *Streptomyces coelicolor* (the sequence of the propeptide of the *S. caespitosus* zymogen is not known). We found the zymogen underwent autolytic activation *in vitro* at the concentration resulting from the purification strategy so fast that the intact zymogen could actually not be isolated. This cleavage also removed the N-terminal fusion protein employed for purification. Two initial cleavages were observed at bonds F59–F60 and

G212–F213 (*S. coelicolor* prosnapalysin numbering in regular characters; see UniProt sequence database access code P0A3Z7), which removed not only the N-terminal part but also a 15-residue C-terminal fragment, as previously described for the MP ulilysin from *Methanosarcina acetivorans*.<sup>15</sup> Subsequently, several cleavages occurred in the remaining part of the 49-residue propeptide to finally yield the mature 133-residue form spanning residues A80–G212. Consistent with an autolytic mechanism *in vitro*, mutant E164A, which affects the glutamate that plays the role of a general base–acid in the catalytic process in MPs,<sup>12,16,17</sup> displays only very slow activation over a period of weeks. In nature, however, prosnapalysin activation is likely to occur under assistance of other proteinases.

Further to snapalysin, sermetstatin did not inhibit any of the other MPs assayed, which included thermolysin and metzincins

of the astacin, pappalysin, serralyisin, fragilyisin, ADAM/ada-malysin and matrix metalloproteinase families. Taken together with previous results showing no inhibition of the M4-family MPs *Pseudomonas aeruginosa* elastase and vimelysin, as well as the unassigned MP almelysin,<sup>8</sup> we conclude that sermetstatin is most likely a specific inhibitor of snapalysins through one of its Janus faces. In addition, sermetstatin potently inhibited subtilisin Carlsberg from *Bacillus licheniformis* (with an apparent  $K_i$  of 1.3 nM) and proteinase K (ESI Table S2<sup>†</sup>), as well as subtilisin BPN' from *Bacillus amyloliquefaciens*,<sup>8</sup> all of them peptidases of biotechnological relevance. However, it only moderately inhibited chymotrypsin and elastase, and only weakly targeted trypsin (ESI Table S2<sup>†</sup>). These results confirm that sermetstatin is a potent, though rather broad-spectrum inhibitor of subtilisin-type—and, partially, chymotrypsin-type—SPs through its other Janus face.

### Structure of sermetstatin

We crystallized and solved the structure of sermetstatin by single-wavelength anomalous diffraction with the help of a selenomethione derivative (see ESI and Table S3<sup>†</sup>). The protein is an elongated  $\alpha/\beta$ -sandwich of maximal dimensions  $45 \times 30 \times 25$  Å. It consists of a five-stranded antiparallel  $\beta$ -sheet ( $\beta3$ - $\beta2$ - $\beta1$ - $\beta4$ - $\beta5$ ; connectivity  $-1, -1, +3, +1$ ) twisted by  $\sim 30^\circ$ , whose concave face accommodates two  $\alpha$ -helices ( $\alpha1$  and  $\alpha2$ ) and a  $3_{10}$ -helix ( $\eta1$ ; see Fig. 1a and 2a). The overall structure is tethered by a large hydrophobic core provided by side chains of the  $\beta$ -sheet on its concave face, the two  $\alpha$ -helices, and the C-terminal tail. The structure is further maintained by two disulfide bonds. The first ( $C^{31}$ - $C^{46}$ ; sermetstatin mature-protein residues numbered in superscript; see Fig. 1a and subtract 28 from UniProt Q9FDS0 numbers) links the first “scaffold helix”  $\alpha1$  with  $\beta2$  and the subsequent loop connecting strands  $\beta2$  with  $\beta3$  (L $\beta2\beta3$ ), which includes the “reactive-site loop 1'” ( $C^{31}$ - $A^{35}$ ). The second disulfide bond ( $C^{69}$ - $C^{99}$ ) connects the second “scaffold helix”  $\alpha2$  with L $\alpha2\beta4$ , which contains the “reactive-site loop 2'” ( $P^{65}$ - $F^{73}$ ). Both reactive-site loops protrude from the molecular surface and are fully solvent accessible (Fig. 2a). The C-terminal carboxylate is anchored to the side chain of R<sup>25</sup> within  $\beta2$ ; by contrast, the N-terminus loosely protrudes from the molecular surface and becomes only rigid at P<sup>5</sup>, which leads to the first residue of strand  $\beta1$ .

Two sermetstatin molecules intimately associate through the convex faces of their  $\beta$ -sheets to yield a crystallographically related dimer (Fig. 2a), in agreement with the oligomeric state found in solution. The sheet planes are rotated by  $\sim 35^\circ$  away from each other and the interaction surface spans  $\sim 800$  Å<sup>2</sup>, *viz.* 25% of the total surface of a monomer. The interaction is mainly hydrophobic and made by 31 close contacts, which include symmetric hydrophobic interactions between ten residues provided by each sheet of either monomer, and only two salt bridges between R<sup>88</sup> of one protein chain and D<sup>81</sup> of the other.

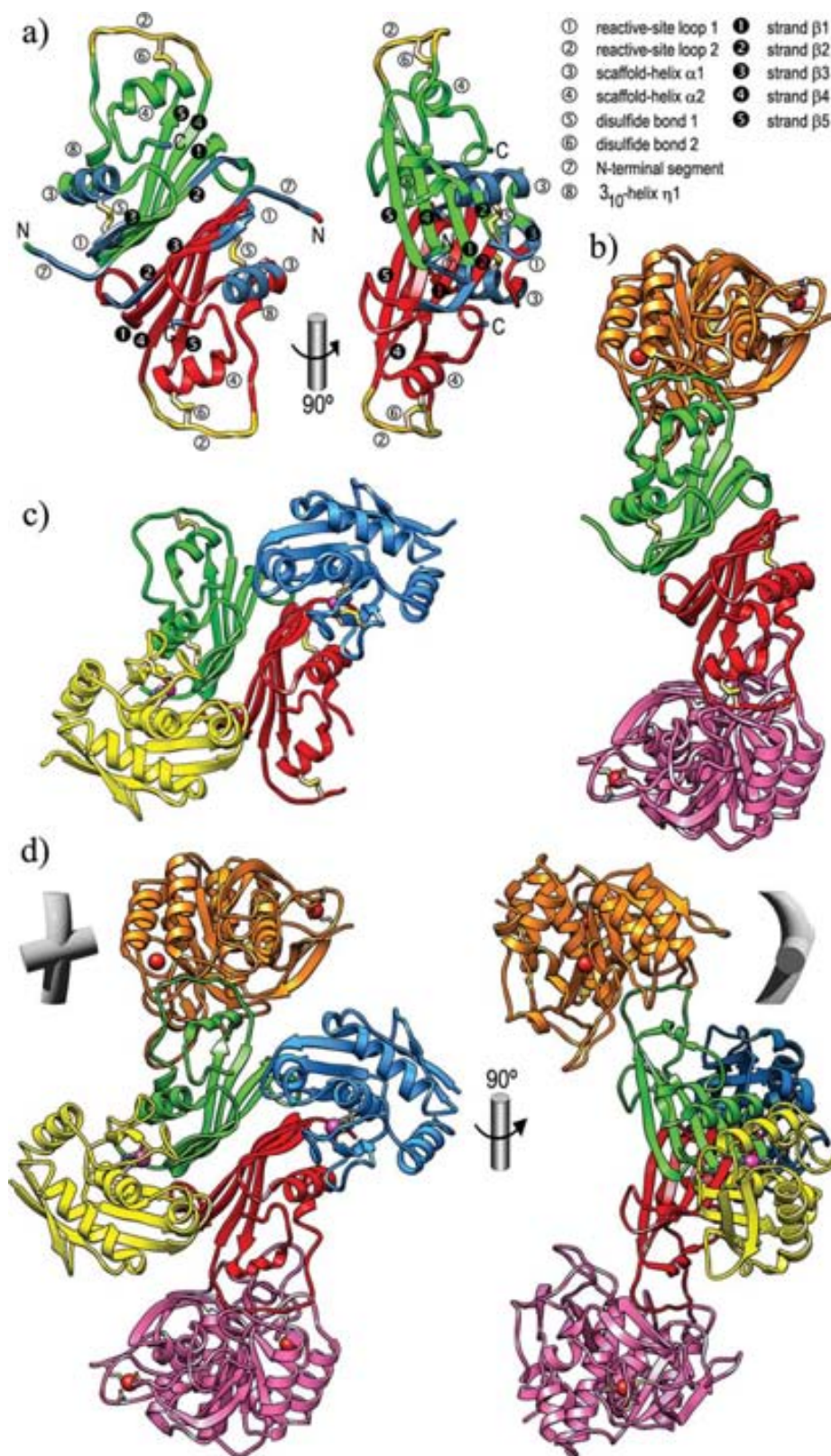
Structure similarity searches with sermetstatin against the Protein Data Bank (PDB) only identified the I16-family relative SSI, which is also a dimer, as being closely related. However, significant variations, in particular around reactive-site loop 2,

the scaffold helices and the linking disulfide bonds, and, most importantly, the N-terminal segment—which is longer, differs in sequence (Fig. 1a), and is disordered until positions 4–7 in the different wild-type SSI structures reported (PDB access codes 3SSI,<sup>18</sup> 2SIC,<sup>19</sup> and 2TLD<sup>20</sup>)—account for functional differences (see below). Consistently, the sermetstatin structure could not be solved by Patterson-search techniques using SSI as a search model (see ESI<sup>†</sup> and Fig. 1b). In addition to SSI, the isolated structure of plasminostreptin has been reported<sup>21</sup> but the coordinates are not available.

### Structure of the subtilisin–sermetstatin complex

The crystal structure of the complex between subtilisin Carlsberg and sermetstatin was determined by Patterson search and reveals a 2 + 2 heterotetramer as anticipated by studies in solution, with a central non-cleaved inhibitor dimer—as present in the unbound structure—and one enzyme moiety bound to each inhibitor monomer (Fig. 2b and ESI Table S3<sup>†</sup>). Only two segments of sermetstatin adopt a different conformation from the unbound structure: the flexible N-terminal segment until A<sup>7</sup> and the segment preceding and including reactive-site loop 2 until C<sup>69</sup>. The latter variation is due to the flip of the peptide bond L<sup>60</sup>-A<sup>61</sup>. The only enzyme–inhibitor interaction is precisely made by reactive-site loop 2, adequately positioned by its scaffold helix, and the active-site cleft of the enzyme (Fig. 3a). The loop, which includes the reactive-site bond M<sup>71</sup>-Y<sup>72</sup>, is inserted wedge-like mimicking a substrate in extended, “canonical” conformation in the active-site cleft of the enzyme following the “standard mechanism” of inhibition, terms which were coined by Laskowski, Bode, Huber, and others.<sup>1,2,22</sup> This mechanism has only been structurally proven to date for SP inhibitors in general and for the specific insect metalloproteinase inhibitor of thermolysin-family MPs.<sup>23</sup> Rigidity to reactive-site loop 2 is conferred by the proline at position P2 (active-site cleft and substrate sub-site nomenclature according to ref. 17 and 24 in bold) and by disulfide bond C<sup>69</sup>-C<sup>99</sup>. On the non-primed side of the subtilisin cleft, the inhibitor main chain is anchored *via* a  $\beta$ -ribbon interaction to segment G<sub>100</sub>-Y<sub>104</sub> (subtilisin mature-protein residue numbering in subscript; subtract 104 from UniProt P00780 numbers), which creates an upper rim of the cleft, through three hydrogen bonds. Below the cleft, subtilisin segment S<sub>125</sub>-G<sub>128</sub> further tethers the inhibitor through three more inter-main-chain hydrogen bonds. V<sup>68</sup> and M<sup>71</sup>, respectively, occupy the two sub-sites of the active-site cleft important for specificity in subtilisin, S4 and S1.<sup>25</sup> The catalytic serine, S<sub>221</sub>, hydrogen bonds through its O $\gamma$  atom the inhibitor at M<sup>71</sup> N and, downstream of P1, the polypeptide undergoes a kink that probably prevents cleavage at M<sup>71</sup>-Y<sup>72</sup> and results in the main chain at the two subsequent residues being hydrogen-bonded in an antiparallel manner by N<sub>318</sub>-T<sub>220</sub>. Contrary to MPs, substrate position P1' is not important for specificity in SPs in general, and Y<sup>72</sup> points toward bulk solvent; F<sup>73</sup> in P3', by contrast, approaches the enzyme and performs a stacking interaction with F<sub>189</sub> (Fig. 3a). The small interacting surface between enzyme and inhibitor—despite high inhibitory potency (see above)—means that upon superposition of the proteinase moieties of the two complexes within the





**Fig. 2** (a) (Left) Structure of the unbound sermetstatin dimer (in red and green) in the reference orientation chosen and (Right) after a vertical rotation of  $90^\circ$ . Characteristic segments are pinpointed by encircled numbers. Regions engaged in MP/SP binding are in blue/yellow, respectively. (b) Structure of the heterotetrameric subtilisin–sermetstatin and (c) snapalysin–sermetstatin complexes. Proteinase moieties are in orange/pink and blue/yellow, respectively. (d) Model of the heterohexameric inhibitor complex derived from (b) and (c) in two orientations.



heterotetramer, the inhibitor moieties display a relative rotation of  $\sim 10^\circ$  around an axis horizontally traversing the active-site cleft (Fig. 3b). In addition, it also explains why several SPs are targeted by the inhibitor, as no exosites are employed for binding that could confer specificity. Finally, superposition of the complex with the one reported between SSI and subtilisin BPN' (PDB 2SIC; (ref. 19)) reveals a very similar binding. As observed in the asymmetric unit of the subtilisin–sermetstatin structure, a relative reorientation between the sermetstatin and SSI inhibitor moieties—here of  $\sim 20^\circ$ —is found upon superposition of the respective enzyme parts (Fig. 3c).

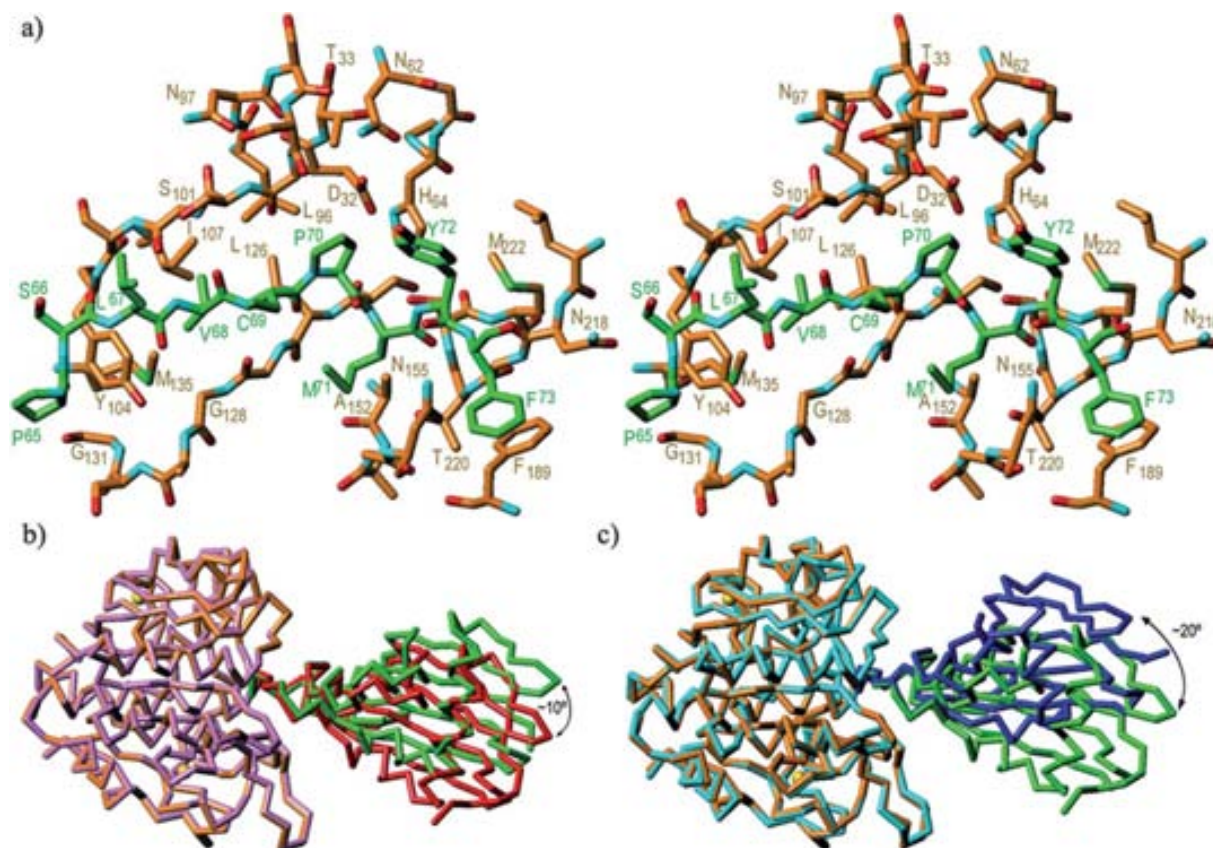
### Redesign of inhibitory specificity

Inhibitory studies with wild-type sermetstatin revealed only weak activity against trypsin (ESI Table S2 and (ref. 8)†), possibly owing to the presence of a methionine in P1 (M<sup>71</sup>). In order to design a better inhibitor of the latter SP, we produced a mutant, M<sup>71</sup>K, to match the specificity of trypsin. While this mutant did not have any influence on the MPs tested, it showed a twofold increase in trypsin inhibition and a substantial decrease in the inhibition of hydrophobic S1-specific SPs of subtilisin type (subtilisin Carlsberg and proteinase K) and chymotrypsin type

(chymotrypsin and elastase), *i.e.* we were able to alter the specificity of sermetstatin. In addition, we also replaced positions P1' and P2' of sermetstatin with the residues found in SSI (mutant Y<sup>72</sup>V + F<sup>73</sup>Y; see Fig. 1a) to assess the importance of these sites but could not observe any significant inhibitory difference with the wild-type (ESI Table S2†).

### Structure of the snapalysin–sermetstatin complex

The crystal structure of the complex between *S. caespitosus* snapalysin and sermetstatin was determined by multi-step Patterson search and also shows a heterotetrameric 2 + 2 quaternary structure (Fig. 2c and ESI Table S3†), again in accordance with results in solution. Three such structurally equivalent heterotetramers totalling 12 protein chains (A–L) were found in the crystal asymmetric unit, and the discussion hereafter is centered on tetramer ABCD (see ESI†). The structure of the central dimeric sermetstatin core shows an intact inhibitor and almost exactly the same minor structural deviations from the unbound structure as seen in the subtilisin complex. However, unlike the unbound structure and the latter complex, reactive-site loop 2 is flexible and partially disordered in some of the polypeptide chains. Snapalysin, in turn, is in a



**Fig. 3** (a) Detail, in cross-eye stereo, of the interaction between sermetstatin reactive-site loop 2 (green carbon atoms and residue labels) and subtilisin Carlsberg (orange carbon atoms and residue labels) shown in standard orientation and in the reference front view of the inhibitor (see Fig. 2a). The catalytic triad is S221–H64–D32. (b) Superposition of the two complexes within the heterotetrameric complex (chains A/B in orange/green and C/D in pink/red) after optimal fit of the proteinase moieties. The two calcium cations of chain A are shown as yellow spheres for reference. The view corresponds to (a) after a vertical and an in-plane rotation of  $90^\circ$  each. (c) Superposition of the sermetstatin–subtilisin Carlsberg complex (chains A/B in orange/green) onto the complex between SSI and subtilisin BPN' (PDB 2SIC) in blue/cyan after optimal fit of the proteinase moieties. View as in (b).

virtually identical conformation to the unbound structure (PDB 1KUH; (ref. 26)) throughout its entire structure except for loop  $A_{36}$ – $A_{40}$  (italicized *S. caespitosus* snapalysin residues and subscript numbers according to the mature MP sequence; see UniProt P56406), which is distal from the active-site cleft and thus the zone of interaction with sermetstatin. Within the complex, the proteinase is bound by a region on the opposite surface to the subtilisin binding site (Fig. 2c) through a large surface contributed to by several protein segments. This results in little flexibility at the inter-protein junction as seen upon superposition of the distinct complexes. The interaction between snapalysin (chain A) and sermetstatin (chain B) buries an interface of  $\sim 870 \text{ \AA}^2$  ( $\sim 13\%$  of the total surface of both enzyme and inhibitor). It includes 43 close contacts, among them 11 hydrogen bonds, and hydrophobic interactions between eight peptidase and seven inhibitor residues. Segments involved include the N-terminal tail ( $S^1$ – $P^5$ ), reactive-site loop 1, helix  $\alpha 1$  (at  $R^{43}$ – $N^{50}$ ), and  $L\beta 4\beta 5$  (at  $N^{85}$ – $G^{86}$ ) of the inhibitor; and  $R_{51}$ – $D_{58}$ ,  $Q_{71}$ – $Y_{75}$ ,  $R_{79}$ – $H_{87}$ ,  $D_{93}$ – $G_{97}$ , and  $G_{105}$ – $P_{108}$  of the proteinase moiety. This implies that not only the active-site cleft is targeted during inhibition but also exosites on the protease surface, and this, in turn, explains the specificity of this second Janus face for snapalysin.

The present inhibition modus follows a novel mechanism for MP inhibitors: the N-terminal segment is inserted in a substrate-like manner in canonical conformation into the non-primed side of the active-site cleft, thus establishing a  $\beta$ -ribbon-type interaction with the  $\beta$ -strand above the active-site cleft at  $A_{55}$ – $D_{58}$  (see Fig. 4). The side chain of  $H^3$  is directed toward the peptidase moiety and approaches the zinc-bound catalytic solvent through a solvent bridge. To assess this interaction, we constructed two point mutants,  $H^3R$  and  $H^3E$ , which showed inhibitory activity indistinguishable from that of the wild type (data not shown), which is logical as the structure should be able to accommodate both changes simply through removal of the bridging solvent. The second main interacting region for inhibition between snapalysin and sermetstatin is provided by reactive-site loop 1, which penetrates cleft sub-site  $S1'$  with  $Y^{33}$ , thus matching the substrate specificity of the  $MP^{26,27}$  (Fig. 4). The aromatic side chain approaches  $R_{79}$ , which creates the bottom of the hydrophobic  $S1'$  pocket of the enzyme through the aliphatic part of its side chain. Due to the intrusion of reactive-site loop 1,  $Y_{95}$ , which is in a position indistinguishable from that seen in the unbound structure<sup>26</sup> and is generally supposed to stabilize the tetrahedral reaction intermediate during catalysis,<sup>12</sup> contacts the main-chain carbonyl of the loop at  $A^{32}$  through its side chain. To assess the importance of this loop *in vitro*, we constructed a double point mutant,  $Y^{33}P + T^{34}G$ , which mimics the sequence of SSI at this site (Fig. 1a), and found only residual inhibitory activity against snapalysin, while the behavior against SPs was unaltered (ESI Table S2†). This supports an authentic key role in inhibition for this second Janus-face site, which is absent in SSI and, most likely, other I16 inhibitors studied, which are evolutionarily more distant from sermetstatin than from each other (Fig. 1a). In addition, secondary snapalysin–sermetstatin interaction sites are provided by sermetstatin helix  $\alpha 1$ , which interacts with the segment immediately downstream of the third zinc-binding

residue of snapalysin ( $D_{93}$ – $G_{97}$ ), and  $L\beta 4\beta 5$ , which contacts the upper-rim strand of the peptidase on its non-primed side (Fig. 4). In particular,  $D^{47}$  of  $\alpha 1$  approaches the main-chain amide nitrogen atoms of both  $Y_{95}$  and  $Q_{96}$  through its side chain. This causes the side chain of the latter residue to be pulled out towards the bulk solvent in the complex. Finally, the frontal bulge segment above the primed-side of the active-site cleft of snapalysin,  $R_{51}$ – $D_{58}$ , interacts with three short segments of the second sermetstatin molecule (chain D) within the heterotetramer at  $V^{23}$ – $T^{28}$ ,  $G^{37}$ – $P^{40}$ , and  $F^{113}$  through four hydrogen bonds. Overall, this type of enzyme–inhibitor interaction is only distantly reminiscent of the inhibitory mode of tissue inhibitors of metalloproteinases on their target matrix metalloproteinases<sup>28</sup> and of serralysin inhibitors on their cognate serralysin MPs.<sup>29</sup> In these cases, N-terminal tails likewise bind in extended conformation but are found in the primed side of the cleft.

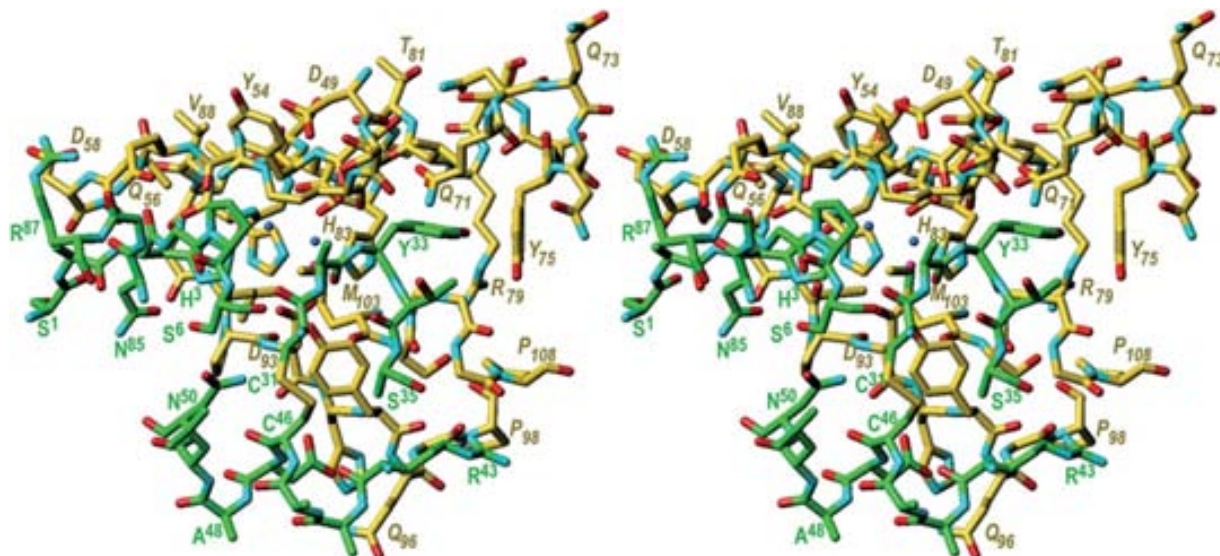
### Disulfide bonds in sermetstatin

Disulfide bonds are a common feature of proteinase inhibitors that follow the standard mechanism, and we set out to assess the importance of the two that link the reactive-site loops to their subjacent scaffold helices by mutating the cysteine residues to serine. Wild-type sermetstatin, as well as mutants  $C^{31}S + C^{46}S$  and  $C^{69}S + C^{99}S$ , were incubated with either subtilisin or snapalysin. While the wild-type protein was not cleaved by either peptidase, mutant  $C^{69}S + C^{99}S$ , which was obtained with similar yields to the wild-type protein (see ESI†), was completely degraded by subtilisin but not snapalysin. This suggested that the mutant inhibitor still had sufficient structural integrity to maintain its MP site functional but not to prevent the SP site from collapsing. By contrast, mutant  $C^{31}S + C^{46}S$ , which was obtained with a much lower yield than the wild-type, was readily degraded by both peptidases, strongly suggesting that this disulfide is essential for the overall structural conformation of the inhibitor. Accordingly, the disulfide bonds of sermetstatin are required for its double inhibitory activity because they reduce the conformational flexibility of the main chain around the reactive site and contribute to maintaining the overall scaffold in a competent conformation.

### Model for the ternary complex

All the structural information described has allowed us to construct a composite model of the ternary complex between subtilisin, sermetstatin, and snapalysin (Fig. 2d). This model shows a large inhibitory particle of maximal dimensions  $120 \times 85 \times 65 \text{ \AA}$ , which is rather asymmetric and resembles a distorted cross with a large vertical bar arched by  $\sim 55^\circ$  and a short horizontal bar, which intersect at an angle of  $\sim 75^\circ$ . The model confirms the heterohexameric  $2 + 2 + 2$  arrangement found in solution (see ESI and Fig. S1†); no steric hindrance is observed upon peptidase binding to the distal inhibitory sites, and the proteinase surfaces are at least  $\sim 15 \text{ \AA}$  apart.

In summary, sermetstatin is a genuine Janus-faced dimeric single-domain inhibitor which has evolved from single-site standard-mechanism SP inhibitors of family I16 to give a



**Fig. 4** Detail in cross-eye stereo of the interaction of sermetstatin (chain B; green carbon atoms and residue labels) and snapalysin (chain A; yellow carbons and ochre residue labels). The peptidase is shown in standard orientation and its active-site zinc ion is depicted as a small magenta sphere. The two solvent molecules found in the active-site cleft are shown as blue spheres: one is the catalytic solvent molecule attached to the zinc ion (right) and the second one bridges  $H^2 N\epsilon 2$  of sermetstatin and the former solvent molecule (left). Selected residues of either polypeptide chain are labeled for reference. The interaction of the snapalysin bulge segment ( $R_{51}$ – $D_{58}$ ) with regions of the second inhibitor molecule within the heterotetrameric 2 + 2 complex (chain D) is not shown for clarity.

protein capable of simultaneous inhibition of SPs in general and a specific MP through distinct but compatible sites.

## Notes and references

- W. Bode and R. Huber, *Biomed. Biochim. Acta*, 1991, **50**, 437–446.
- W. Bode and R. Huber, *Eur. J. Biochem.*, 1992, **204**, 433–451.
- R. Bao, C. Z. Zhou, C. Jiang, S. X. Lin, C. W. Chi and Y. Chen, *J. Biol. Chem.*, 2009, **284**, 26676–26684.
- Z. Zhou, J. Gu, Y. L. Du, Y. Q. Li and Y. Wang, *Curr. Genomics*, 2011, **12**, 404–416.
- N. D. Rawlings, A. J. Barrett and A. Bateman, *Nucleic Acids Res.*, 2012, **40**, D343–D350.
- T. Kumazaki, K. Kajiwara, S. Kojima, K. Miura and S. Ishii, *J. Biochem.*, 1993, **114**, 570–575.
- M. Uyeda, S. Yoshida, K. Suzuki and M. Shibata, *Agric. Biol. Chem.*, 1976, **40**, 1237–1238.
- K. Hiraga, T. Suzuki and K. Oda, *J. Biol. Chem.*, 2000, **275**, 25173–25179.
- S. Murao and S. Sato, *Agric. Biol. Chem.*, 1972, **36**, 160–163.
- J. E. Strickler, T. R. Berka, J. Gorniak, J. Fornwald, R. Keys, J. J. Rowland, M. Rosenberg and D. P. Taylor, *J. Biol. Chem.*, 1992, **267**, 3236–3241.
- A. Kakinuma, H. Sugino, N. Moriya and M. Isono, *J. Biol. Chem.*, 1978, **253**, 1529–1537.
- F. X. Gomis-Rüth, *J. Biol. Chem.*, 2009, **284**, 15353–15357.
- W. Bode, F. X. Gomis-Rüth and W. Stöcker, *FEBS Lett.*, 1993, **331**, 134–140.
- W. Stöcker, F. Grams, U. Baumann, P. Reinemer, F. X. Gomis-Rüth, D. B. McKay and W. Bode, *Protein Sci.*, 1995, **4**, 823–840.
- C. Tallant, R. García-Castellanos, J. Seco, U. Baumann and F. X. Gomis-Rüth, *J. Biol. Chem.*, 2006, **281**, 17920–17928.
- B. W. Matthews, *Acc. Chem. Res.*, 1988, **21**, 333–340.
- F. X. Gomis-Rüth, T. O. Botelho and W. Bode, *Biochim. Biophys. Acta, Proteins Proteomics*, 2012, **1824**, 157–163.
- Y. Mitsui, Y. Satow, T. Sakamaki and Y. Iitaka, *J. Biochem.*, 1977, **82**, 295–298.
- Y. Takeuchi, Y. Satow, K. T. Nakamura and Y. Mitsui, *J. Mol. Biol.*, 1991, **221**, 309–325.
- Y. Takeuchi, T. Nonaka, K. T. Nakamura, S. Kojima, K. Miura and Y. Mitsui, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 4407–4411.
- N. Kamiya, M. Matsushima and H. Sugino, *Bull. Chem. Soc. Jpn.*, 1984, **57**, 2075–2081.
- M. Laskowski Jr and I. Kato, *Annu. Rev. Biochem.*, 1980, **49**, 593–626.
- J. L. Arolas, T. O. Botelho, A. Vilcinskas and F. X. Gomis-Rüth, *Angew. Chem., Int. Ed.*, 2011, **50**, 10357–10360.
- I. Schechter and A. Berger, *Biochem. Biophys. Res. Commun.*, 1967, **27**, 157–162.
- Y. Takeuchi, S. Noguchi, Y. Satow, S. Kojima, I. Kumagai, K. Miura, K. T. Nakamura and Y. Mitsui, *Protein Eng., Des. Sel.*, 1991, **4**, 501–508.
- G. Kurisu, T. Kinoshita, A. Sugimoto, A. Nagara, Y. Kai, N. Kasai and S. Harada, *J. Biochem.*, 1997, **121**, 304–308.
- S. Harada, T. Kinoshita, N. Kasai, S. Tsunasawa and F. Sakiyama, *Eur. J. Biochem.*, 1995, **233**, 683–686.
- F. X. Gomis-Rüth, K. Maskos, M. Betz, A. Bergner, R. Huber, K. Suzuki, N. Yoshida, H. Nagase, K. Brew, G. P. Bourenkov, H. Bartunik and W. Bode, *Nature*, 1997, **389**, 77–81.
- U. Baumann, M. Bauer, S. Letoffe, P. Delepeleire and C. Wandersman, *J. Mol. Biol.*, 1995, **248**, 653–661.



## ELECTRONIC SUPPLEMENTARY INFORMATION

### Table of Contents

<b>1. Experimental procedures</b>	<b>1</b>
<i>Production and purification of sermetstatin</i>	<b>1</b>
<i>Production and purification of snapalysin</i>	<b>2</b>
<i>In vitro activation studies of prosnapalysin</i>	<b>3</b>
<i>Proteolytic and inhibition assays</i>	<b>4</b>
<i>Cleavage of sermetstatin mutants</i>	<b>4</b>
<i>Complex formation and purification</i>	<b>5</b>
<i>Crystallization and X-ray diffraction data collection</i>	<b>5</b>
<i>Structure solution and refinement</i>	<b>6</b>
<i>Miscellaneous</i>	<b>8</b>
<b>2. Acknowledgments</b>	<b>8</b>
<b>3. Supplemental References</b>	<b>9</b>
<b>4. Supplemental Tables</b>	<b>11</b>
<i>Supplemental Table S1</i>	<b>11</b>
<i>Supplemental Table S2</i>	<b>12</b>
<i>Supplemental Table S3</i>	<b>13</b>
<b>5. Supplemental Figures</b>	<b>14</b>
<i>Supplemental Figure S1</i>	<b>14</b>

### 1. Experimental procedures

***Production and purification of sermetstatin*** – A synthetic gene coding for sermetstatin from *Streptomyces caespitosus* (UniProt database code Q9FDS0), also known as *Streptomyces caespitosus* neutral proteinase inhibitor<sup>1</sup>, was purchased (GenScript) and cloned into a modified pET-32a vector between the *Bgl*II and *Hind*III restriction sites. This vector attaches an N-terminal thioredoxin-His<sub>6</sub> fusion construct followed by a tobacco-etch-virus (TEV) protease recognition site. Sermetstatin was produced by heterologous overexpression in *Escherichia coli* Origami2 (DE3) cells (Novagen). These were grown at 37°C in Luria-Bertani (LB) medium containing 100µg/ml ampicillin and 10µg/ml tetracycline, induced at an OD<sub>550</sub> of 0.6 with isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.25mM, and subsequently incubated overnight at 18°C. Cultures were centrifuged at 7,000xg for 30min at 4°C. Pellets were washed twice with buffer A (50mM Tris-HCl, 500mM NaCl, pH8.0) and resuspended in the same buffer further containing 20mM imidazole and supplemented with EDTA-free protease inhibitor cocktail tablets (Roche Diagnostics) and DNase I (Roche Diagnostics). Cells were lysed at 4°C using a cell disrupter (Constant Systems) at a pressure of 1.35Kbar, and the cell debris was removed by centrifugation at 50,000xg for 1h at 4°C. The

supernatant was filtered (0.22µm pore size; Millipore) and incubated with nickel-nitrilotriacetic acid (Ni-NTA) resin (Invitrogen) previously equilibrated with buffer A, 20mM imidazole. The protein was eluted using buffer A, 350mM imidazole. Subsequently, the sample was dialyzed overnight at room temperature against buffer B (50mM Tris-HCl, 150mM NaCl, 0.5mM oxidized glutathione [GSSG], 3mM reduced glutathione [GSH], pH8.0) in the presence of His<sub>6</sub>-tagged TEV protease at an enzyme:substrate ratio of 1:50 (w/w). TEV cleavage leaves an extra glycine residue at the N-terminus of the protein (termed G<sup>-1</sup>; superscripted sermetstatin amino acid numbering hereafter corresponds to the mature protein without the 28-residue signal peptide [see UniProt Q9FDS0]). The digested sample was passed several times through Ni-NTA resin previously equilibrated with buffer A to remove His<sub>6</sub>-containing molecules. The flow-through was collected, concentrated by ultrafiltration, and further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column (GE Healthcare) previously equilibrated with buffer C (20mM Tris-HCl, 150mM NaCl, pH7.4). Previous column calibration revealed that the protein eluted as a dimer, and its identity and purity were assessed by mass spectrometry and 15% Tricine-SDS-PAGE stained with Coomassie blue. Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices with a 5-KDa cut-off (Sartorius Stedim Biotech). Protein concentration was determined by measuring the absorbance at 280 nm using a spectrophotometer (NanoDrop) and a calculated absorption coefficient  $E_{0.1\%} = 0.86$ .

The present production system based on a fusion with thioredoxin and expression in Origami2 cells, which is known to assist multi-disulfide proteins in folding correctly during intracellular biosynthesis, and further TEV-mediated removal of the fusion construct in selected concentrations of redox agents, yielded approx. 10mg of purified natively-folded protein per liter of cell culture. The selenomethionine variant of the protein was obtained in the same way as the wild-type except that cells were grown in minimal medium containing selenomethionine (Sigma) instead of methionine. A series of sermetstatin single- and double-point mutants, namely, H<sup>3</sup>E, H<sup>3</sup>R, Y<sup>33</sup>P+T<sup>34</sup>G, M<sup>71</sup>K, Y<sup>72</sup>V+F<sup>73</sup>Y, C<sup>31</sup>S+C<sup>46</sup>S, and C<sup>69</sup>S+C<sup>99</sup>S, was generated using the QuickChange Site-Directed Mutagenesis kit (Stratagene). All constructs were verified by DNA sequencing and mutant variants were produced and purified as for the wild-type protein, with comparable yields except for C<sup>31</sup>S+C<sup>46</sup>S, which was obtained with an approx. 100-fold lower yield.

***Production and purification of snapalysin*** – A synthetic gene coding for mature snapalysin from *S. caespitosus*, in which the N-terminal threonine had been replaced by methionine due to the cloning strategy, was purchased (GenScript) and cloned into a modified pET-28a vector between the *Nco*I and *Xho*I restriction sites. This vector attaches an N-terminal His<sub>6</sub> fusion tag, followed by a TEV protease recognition site. The protein was overproduced as inclusion bodies in *E. coli* BL21 (DE3) cells, which were grown at 37°C in LB medium supplemented with 30µg/ml kanamycin, induced at an OD<sub>550</sub> of 0.8 with 1mM IPTG, and grown for a further 5h at 37°C. After centrifugation at 7,000xg for 30min at 4°C, the pellet was resuspended in buffer D (PBS, 1% Triton X-100, pH7.4) containing EDTA-free protease inhibitor cocktail tablets and DNase I. Cells were lysed at 4°C using a cell

disrupter at a pressure of 1.9Kbar and incubated for 30min at 37°C. The inclusion bodies were recovered by centrifugation at 7,000xg for 30min at 4°C and washed twice with buffer D, resuspended in buffer A, 8M urea, and incubated for 5h at room temperature under vigorous shaking. The sample was centrifuged at 50,000xg for 1h at 20°C, the supernatant was incubated with Ni-NTA resin previously equilibrated with buffer A, 8M urea, and the fusion protein was eluted using buffer E (50mM sodium phosphate, 250mM NaCl, 8M urea, pH4.0). The sample was dialyzed overnight at room temperature against buffer A plus 0.5mM GSSG, 1mM GSH, 10mM CaCl<sub>2</sub>, 1mM ZnSO<sub>4</sub>, and 0.4M L-arginine, and subsequently dialyzed overnight at 4°C against buffer C. The N-terminal extension of the snapalysin construct was removed by autolytic cleavage as determined by Edman degradation, yielding an N-terminal segment *Q<sub>-3</sub>+G<sub>-2</sub>+P<sub>-1</sub>+M<sub>1</sub>* (italicized snapalysin amino acids subscripted numbering correspond to UniProt P56406 *plus* the additional N-terminal tag residues and the *T<sub>1</sub>M* mutation; although this is a secreted protein, the signal-peptide sequence is unknown and the database entry corresponds to the authentic protein purified from culture supernatants of *S. caespitosus*). The protein was concentrated by ultrafiltration, and further purified by size-exclusion chromatography on a Superdex 75 10/300 column previously equilibrated with buffer C. Protein identity and purity were assessed by Edman degradation and 15% Tricine-SDS-PAGE stained with Coomassie blue. Ultrafiltration steps were performed with Vivaspin 15 and Vivaspin 500 filter devices with 5-KDa cut-off. Protein concentration was determined by measuring the absorbance at 280nm using a NanoDrop spectrophotometer and a calculated absorption coefficient  $E_{0.1\%} = 1.7$ . This procedure yielded functional snapalysin at very low concentration (approx. 50µg per liter of cell culture), which was used for activity assays (see below).

**In vitro activation studies of prosnapalysin** – To study the transition between latent and mature snapalysins, a synthetic gene encompassing the chemical sequence of prosnapalysin from *Streptomyces coelicor*—the sequence of the pro-peptide of the zymogen from *S. caespitosus* is not known—lacking the 29-residue signal peptide and with the N-terminal alanine replaced with methionine (A30M; numbering in regular characters according to UniProt P0A3Z7) due to the cloning strategy, was purchased (GenScript) and cloned into the aforementioned modified pET-28a vector using the *Nco*I and *Xho*I restriction sites. Prosnapalysin active-site mutant E164A was obtained using the QuickChange Site-Directed Mutagenesis kit and verified by DNA sequencing. Both proteins were overproduced in *E. coli* BL21 (DE3) cells, which were grown at 37°C in LB medium supplemented with kanamycin to a final concentration of 30µg/ml. Thereafter, cells were induced at an OD<sub>550</sub> of 0.8 with IPTG to a final concentration of 1mM and grown for a further 5h at 37°C. After centrifugation at 7,000xg for 30min at 4°C, the pellet was resuspended in buffer D containing EDTA-free protease inhibitor cocktail tablets and DNase I. Cells were lysed at 4°C using a cell disrupter at a pressure of 1.9 Kbar and incubated for 30min at 37°C. The resulting inclusion bodies were recovered by centrifugation at 7,000xg for 30min at 4°C and washed twice with buffer D, resuspended in buffer A, 8M urea, and incubated for 5h at room temperature under vigorous shaking. The sample was

centrifuged at 50,000xg for 1h at 20°C, the supernatant was incubated with Ni-NTA resin previously equilibrated with buffer A, 8M urea, and the fusion protein was eluted using buffer E. The sample was then first dialyzed overnight at room temperature against buffer A plus 0.5mM GSSG, 1mM GSH, 10mM CaCl<sub>2</sub>, 1mM ZnSO<sub>4</sub>, and 0.4M L-arginine, and then dialyzed overnight at 4°C against buffer C. The fusion tag of prosnapalysin mutant E164A was cleaved overnight at room temperature in buffer C, 0.5mM GSSG, 3mM GSH by incubation with TEV proteinase (enzyme:substrate ratio of 1:50 [w/w]). Proteins were concentrated by ultrafiltration, and further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer C. The final yield for the wild-type and the mutant was approx. 2mg and 5mg, respectively, per liter of cell culture. Protein identity and purity were assessed by Edman degradation and 15% Tricine-SDS-PAGE stained with Coomassie blue. Cleavage of wild-type and mutant prosnapalysin was monitored by mass spectrometry, SDS-PAGE and Edman degradation of electro-blotted bands onto an Immun-Blot PVDF membrane (BioRad). Ultrafiltration steps were performed with VivaSpin 15 and VivaSpin 500 filter devices of 5-KDa cut-off. Protein concentration was determined by measuring the absorbance at 280 nm using a NanoDrop and a calculated absorption coefficient  $E_{0.1\%} = 1.38$  for both variants.

**Proteolytic and inhibition assays** – Ulilysin from *Methanosarcina acetivorans*, aeruginolysin from *Pseudomonas aeruginosa*, and fragilysin-3 from *Bacteroides fragilis* were produced and purified as previously described<sup>2-5</sup>. The vector coding for aeruginolysin was a kind gift from Ulrich Baumann (University of Cologne, Germany). Astacin and matrix metalloproteinases-1, -3, and -13 were kindly provided by Walter Stöcker (Johannes Gutenberg-University of Mainz, Germany) and Hideaki Nagase (Imperial College London, UK), respectively. Proteolytic activity of ADAM-17 was kindly measured by Jordi Malapeira and Joaquin Arribas (Vall d'Hebron Institute of Oncology, Spain) using the substrate vasorin in a cell-based assay as previously published<sup>6</sup>. Subtilisin Carlsberg from *Bacillus licheniformis*, thermolysin from *Bacillus thermoproteolyticus*, bovine pancreatic trypsin, bovine pancreatic chymotrypsin, porcine pancreatic elastase, and fungal proteinase K from *Engyodontium album* were purchased from Sigma. Evident proteolytic activity of all these proteinases and recombinant snapalysin from *S. caespitosus* was measured at 1μM enzyme concentration in buffer C at 37°C with the fluorescein conjugates BODIPY FL casein and DQ gelatin (65μg/ml, Invitrogen) as substrates by using microplates (Nunc) and a microplate fluorimeter (FLx800, BioTek) at  $\lambda_{ex} = 485\text{nm}$  /  $\lambda_{em} = 528\text{nm}$ . Increasing amounts of purified wild-type and mutant variants of sermetstatin were added to the assays to determine inhibition (see Suppl. Table S2). In addition, inhibition of wild-type sermetstatin against subtilisin and snapalysin was tested by using the substrates succinyl-A-A-P-F-p-nitroanilide (Sigma) and aminobenzoyl-V-K-F-Y-D-I-K(2,4-dinitrophenylamino), respectively,—kindly provided by Jean-Louis Reymond (University of Berne, Switzerland)—in buffer C, and the associated apparent inhibition constants ( $K_i$ ) were derived from a Dixon plot.

**Cleavage of sermetstatin mutants** – Wild-type sermetstatin, as well as mutants C<sup>31</sup>S+C<sup>46</sup>S and C<sup>69</sup>S+C<sup>99</sup>S (at 0.5mg/ml in 20mM Tris-HCl, 50mM NaCl, pH7.4), were incubated with equimolar

amounts of either subtilisin or snapalysin overnight at room temperature. The reactions were then assessed for cleavage in a Bruker Autoflex III mass spectrometer. Each sample was mixed at a 1:1 ratio (v/v) with a matrix solution of sinapinic acid (10mg/ml, Sigma) dissolved in 30% acetonitrile and 0.1% trifluoroacetic acid, and subsequently spotted onto the plate using the dried-droplet method. Mass spectra were acquired in linear mode geometry with >1,000 laser shots and using a protein mixture ranging from 5 KDa to 17.5 KDa (Protein Calibration Standard I, Bruker) as a calibrator.

**Complex formation and purification** – Commercial subtilisin Carlsberg from *B. licheniformis* was purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer C. The subtilisin:sermetstatin complex was formed by incubation of equimolar amounts of purified wild-type inhibitor and protease, and it eluted as a 2+2 heterotetramer in calibrated size-exclusion chromatography. The complex was concentrated by ultrafiltration up to 5.5mg/ml using Vivaspin 15 and Vivaspin 500 filter devices of 30-KDa cut-off. Protein concentration was determined by measuring the absorbance at 280nm using a NanoDrop spectrophotometer and a calculated absorption coefficient  $E_{0.1\%} = 0.89$ . To produce sufficient amounts of snapalysin for structural studies, inclusion bodies of snapalysin were obtained and purified as described above. Next, unfolded snapalysin dissolved in buffer E was diluted 1:9 against buffer A, 0.55mM GSSG, 1.1mM GSH, 11mM CaCl<sub>2</sub>, 1.1mM ZnSO<sub>4</sub>, 0.44M L-arginine and an equimolar amount of wild-type sermetstatin. The sample was incubated for 4h at 4°C and dialyzed overnight at 4°C against buffer C. The snapalysin:sermetstatin complex was concentrated by ultrafiltration, and further purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer C and calibrated, which revealed that the complex was a 2+2 heterotetramer in solution. The sample was dialyzed overnight at 4°C against buffer F (20mM Tris-HCl, 1M NaCl, pH7.4) and concentrated by ultrafiltration to 4.5mg/ml using Vivaspin 15 and Vivaspin 500 filter devices of 10-KDa cut-off. The protein concentration was determined by measuring the absorbance at 280nm using a NanoDrop spectrophotometer and a calculated absorption coefficient  $E_{0.1\%} = 1.32$ . Finally, the ternary complex between subtilisin, sermetstatin, and snapalysin was obtained by adding an equimolar amount of subtilisin to the snapalysin:sermetstatin complex. A further size-exclusion chromatography step in a calibrated Superdex 200 10/300 column previously equilibrated with buffer C revealed that the quaternary arrangement in solution was a 2+2+2 heterohexamer (see Suppl. Fig. S4). The complex was concentrated by ultrafiltration to 7.5mg/ml with Vivaspin 15 and Vivaspin 500 filter devices of 10-KDa cut-off. Protein concentration was determined by measuring the absorbance at 280nm using a NanoDrop spectrophotometer and a calculated absorption coefficient  $E_{0.1\%} = 1.1$ .

**Crystallization and X-ray diffraction data collection** – Crystallization assays of sermetstatin, its two binary complexes with snapalysin and subtilisin, and the ternary complex with both peptidases were carried out by the sitting-drop vapor diffusion method. Reservoir solutions were prepared by a Tecan robot, and 200-nl crystallization drops, equivolumetric in protein solution and reservoir solution, were dispensed on 96x2-well MRC plates (Innovadyne) by a Cartesian nanodrop robot



(Genomic Solutions) at the IBMB/IRB joint High-Throughput Automated Crystallography Platform. Crystallization plates were stored in Bruker steady-temperature crystal farms at 4°C and 20°C. Successful hits were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research) whenever possible.

The best crystals of wild-type and selenomethionine-derivatized sermetstatin appeared at 20°C with protein solution (10mg/ml in buffer C) and 100mM sodium citrate dihydrate, 200mM ammonium acetate, 10% (w/v) PEG3,350, pH5.6 as reservoir solution in microliter drops. The best crystals of the subtilisin:sermetstatin complex were obtained at 4°C with protein solution (5.5mg/ml in buffer C) and 100mM cacodylate, 200mM zinc acetate dihydrate, 10% (v/v) 2-propanol, pH6.5 as reservoir solution from microliter drops. Suitable crystals of the snapalysin:sermetstatin complex were obtained from 200-nanoliter drops at 20°C with protein solution (4.5mg/ml in buffer F) and 100mM HEPES, 200mM magnesium chloride, 15.0% (v/v) ethanol, pH7.5 as reservoir solution. No suitable crystals of the ternary complex could be obtained despite extensive trials. Crystals were cryo-protected by successive passages through reservoir solutions containing increasing amounts of glycerol (up to 25-30% [v/v]). Complete diffraction datasets were collected at 100K from liquid-N<sub>2</sub> flash-cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on MarCCD (beam line ID23-2, native sermetstatin, snapalysin:sermetstatin complex) and ADSC Q315R CCD (ID29, sermetstatin selenomethionine derivative; and ID23-1, subtilisin:sermetstatin complex) detectors at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group “BAG Barcelona”. Crystals of both native and selenomethionine-derivatized sermetstatin were trigonal, with one molecule per asymmetric unit ( $V_M = 3.2\text{\AA}^3/\text{Da}$ ; 62% solvent content). Crystals of subtilisin:sermetstatin were monoclinic, with two complexes per asymmetric unit ( $V_M = 3.6\text{\AA}^3/\text{Da}$ ; 65% solvent content). Crystals of snapalysin:sermetstatin were orthorhombic with six complexes per asymmetric unit ( $V_M = 2.9\text{\AA}^3/\text{Da}$ ; 58% solvent content). Diffraction data were integrated, scaled, merged, and reduced with programs XDS<sup>7</sup>, XSCALE, and SCALA<sup>8</sup> within the CCP4 suite<sup>9</sup> (see Suppl. Table S3).

**Structure solution and refinement** – All attempts to solve the structure of unbound sermetstatin by Patterson search<sup>10</sup> using the coordinates of the related *Streptomyces* subtilisin inhibitor (SSI; Protein Data Bank [PDB] access code 3SSI;<sup>11</sup>) from *Streptomyces albogriseolus* as a searching model failed despite a sequence identity of 43%. Therefore, the structure was solved by single-wavelength anomalous diffraction using a selenomethionine derivative and program SHELXD<sup>12</sup>. Diffraction data of a crystal collected at the selenium absorption-peak wavelength, as inferred from a XANES fluorescence scan, enabled the program to identify all three selenium sites of the monomer present in the asymmetric unit. Subsequent phasing with SHELXE<sup>12</sup> resolved the twofold ambiguity intrinsic to a SAD experiment due to the difference in the values of the pseudo-free correlation coefficient of the two possible hands, which confirmed P3<sub>1</sub>21 as the correct space group. An initial model was built with program COOT<sup>13</sup> and refined against the selenomethionine dataset with program REFMAC5<sup>14</sup>. The

resulting partial model was used to determine the native structure by Patterson-search methods with program PHASER<sup>15</sup>. One unambiguous hit was found, which yielded Z-scores for the rotation and translation functions of 20.4 and 46.0, respectively. Subsequently, manual model building with COOT and TURBO-Frodo<sup>16</sup> alternated with crystallographic refinement (including TLS refinement) with programs REFMAC5 and BUSTER/TNT<sup>17</sup> until completion of the model (Suppl. Table S3). The final model of sermetstatin comprised residues G<sup>-1</sup>+S<sup>1</sup>-F<sup>113</sup> (chain A) *plus* ligands and solvent molecules (see Suppl. Table S3). The dimeric quaternary arrangement observed in solution (see above) was likewise found in the crystal formed by a crystallographic twofold axis.

The structure of the subtilisin:sermetstatin complex was solved by Patterson-search methods with program PHASER using the co-ordinates of subtilisin Carlsberg from *B. licheniformis* (Protein Data Bank [PDB] access code 1SBC;<sup>18</sup>) and sermetstatin as search models. Two unambiguous hits were found for each of the two molecules, which showed Z-scores for the rotation/translation functions of 20.8/22.8 and 18.7/46.2 (subtilisin) and 6.0/36.2 and 6.8/44.5 (sermetstatin). The quaternary arrangement in the crystal asymmetric unit is a 2+2 heterotetramer (chains A-D). Model building and refinement—with automatic non-crystallographic symmetry restraints—proceeded as before (Suppl. Table S3). The final model contained residues A<sub>1</sub>-Q<sub>275</sub> (subscripted subtilisin residue numbers according to the mature enzyme sequence; see UniProt P00780) and two calcium ions (Ca<sub>999</sub> and Ca<sub>998</sub>) for the two subtilisin moieties (chains A and C), G<sup>-1</sup>+S<sup>1</sup>-F<sup>113</sup> for sermetstatin chain B, and G<sup>4</sup>-F<sup>113</sup> for sermetstatin chain D *plus* ligands and solvent molecules (see Suppl. Table S3).

The structure of the snapalysin:sermetstatin complex was likewise solved by Patterson search using the coordinates of *S. caespitosus* snapalysin (PDB 1KUH;<sup>19</sup>) and sermetstatin. Due to the presence of 12 molecules in the asymmetric unit in total, searches had to be split into two steps: first, four copies of each protein were searched for independently, enabling us to identify two complexes showing the same relative arrangement between proteinase and inhibitor. In a second search, one such complex was used as a search model, which yielded six solutions arranged as three 2+2 heterotetramers (chains A-D, E-H, I-L), with Z-scores for the respective rotation/translation functions of 6.7/15.3, 8.1/26.2, 7.3/29.8, 8.8/35.9, 5.5/37.8, and 6.6/35.7. Model building and refinement—with automatic non-crystallographic symmetry restraints—proceeded as before (Suppl. Table S3). The final model contained snapalysin residues G<sub>2</sub>+P<sub>1</sub>+M<sub>1</sub>-G<sub>132</sub> in chains A, C, and I; residues P<sub>1</sub>+M<sub>1</sub>-G<sub>132</sub> in chains G and K; and residues M<sub>1</sub>-G<sub>132</sub> in chain E. In addition, sermetstatin chains spanned residues S<sup>1</sup>-F<sup>113</sup> (chains B and L), A<sup>2</sup>-F<sup>113</sup> (chain F), S<sup>1</sup>-A<sup>62</sup>+L<sup>67</sup>-F<sup>113</sup> (chain D), S<sup>1</sup>-L<sup>60</sup>+L<sup>67</sup>-F<sup>113</sup> (chain H), and S<sup>1</sup>-P<sup>65</sup>+V<sup>68</sup>-F<sup>113</sup> (chain J). Further ligands and solvent molecules completed the model (see Suppl. Table S3). Due to the distinct packing environments of the molecules within the crystal, chains E and K were overall less well-defined by electron density than the other ten chains; this is indicated by the significantly higher average thermal displacement parameters for their protein parts (76.0 and 90.3 Å<sup>2</sup>, respectively) than the remaining molecules (44.8-64.5 Å<sup>2</sup>).

A composite model of the ternary 2+2+2 complex between subtilisin, sermetstatin, and snapalysin was obtained with TURBO-Frodo starting from the subtilisin:sermetstatin complex after superposition of the dimeric inhibitor moieties of the binary complexes. Only the N-terminal segment, L $\beta$ 2 $\beta$ 3, and segment  $\beta$ 1-L $\beta$ 1 $\beta$ 2- $\beta$ 2 had to be slightly remodeled to match the conformation found in the snapalysin:sermetstatin complex. These segments were subjected to geometric refinement with TURBO-Frodo.

**Miscellaneous** – Figures were prepared with programs TURBO-Frodo and CHIMERA<sup>20</sup>. Interaction surfaces (taken as half of the surface area buried at a complex interface) and close contacts (<4Å) were determined with CNS<sup>21</sup>. Structure similarities were investigated with DALI<sup>22</sup>. Model validation was performed with MOLPROBITY<sup>23</sup> and the WHATCHECK routine of WHATIF<sup>24</sup>. Sequence alignments and phylogenetic calculations were performed with MULTALIN<sup>25</sup>. The final coordinates of sermetstatin, the subtilisin:sermetstatin complex, and the snapalysin:sermetstatin complex have been deposited with the PDB at [www.pdb.org](http://www.pdb.org) (access codes 4HWX, 4HX2, and 4HX3). The model of the ternary complex is available from the authors upon request.

## 2. Acknowledgments

We thank Ulrich Baumann, Walter Stöcker, Hideaki Nagase, Jean-Louis Reymond, Jordi Malapeira, and Joaquin Arribas for providing reagents or performing assays. We are also grateful to Tibisay Guevara and the Automated Crystallography Platform at IBMB/IRB for assistance during crystallization experiments. This study was supported in part by grants from European, Spanish, and Catalan agencies (FP7-HEALTH-F3-2009-223101 “AntiPathoGN”; FP7-HEALTH-2010-261460 “Gums&Joints”; FP7-PEOPLE-2011-ITN-290246 “RAPID”; BIO2009-10334; BFU2012-32862; CSD2006-00015; a JAE postdoctoral contract from CSIC; an FPU Ph.D. fellowship from the Spanish Ministry for Science and Technology; a grant from the Andalusian Regional Government; Fundació “La Marató de TV3” grant 2009-100732; and 2009SGR1036). We acknowledge the help provided by ESRF synchrotron local contacts. Funding for data collection was provided in part by ESRF.

### 3. Supplemental References

1. K. Hiraga, T. Suzuki and K. Oda, *J. Biol. Chem.*, 2000, **275**, 25173-25179.
2. C. Tallant, R. García-Castellanos, J. Seco, U. Baumann and F. X. Gomis-Rüth, *J. Biol. Chem.*, 2006, **281**, 17920-17928.
3. T. Goulas, J. L. Arolas and F. X. Gomis-Rüth, *Proc. Natl. Acad. Sci. USA*, 2010, **108**, 1856-1861.
4. U. Baumann, S. Wu, K. M. Flaherty and D. B. McKay, *EMBO J.*, 1993, **12**, 3357-3364.
5. T. Hege, R. E. Feltzer, R. D. Gray and U. Baumann, *J. Biol. Chem.*, 2001, **276**, 35087-35092.
6. J. Malapeira, C. Esselens, J. J. Bech-Serra, F. Canals and J. Arribas, *Oncogene*, 2011, **30**, 1912-1922.
7. W. Kabsch, *Acta Crystallogr. sect. D*, 2010, **66**, 125-132.
8. P. Evans, *Acta Crystallogr. sect. D*, 2006, **62**, 72-82.
9. CCP4, *Acta Crystallogr. sect. D*, 1994, **50**, 760-763.
10. R. Huber, *Acta Crystallogr. sect. A*, 1965, **19**, 353-356.
11. Y. Mitsui, Y. Satow, T. Sakamaki and Y. Iitaka, *J. Biochem.*, 1977, **82**, 295-298.
12. G. M. Sheldrick, *Acta Crystallogr. sect. D*, 2010, **66**, 479-485.
13. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr. sect. D*, 2010, **66**, 486-501.
14. G. N. Murshudov, P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr. sect. D*, 2011, **67**, 355-367.
15. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J. Appl. Crystallogr.*, 2007, **40**, 658-674.
16. C. Carranza, A.-G. Inisan, E. Mouthuy-Knoops, C. Cambillau and A. Roussel, in *AFMB Activity Report 1996-1999*, CNRS-UPR 9039, Marseille, 1999, pp. 89-90.
17. E. Blanc, P. Roversi, C. Vornrhein, C. Flensburg, S. M. Lea and G. Bricogne, *Acta Crystallogr. sect. D*, 2004, **60**, 2210-2221.
18. D. J. Neidhart and G. A. Petsko, *Protein Eng.*, 1988, **2**, 271-276.
19. G. Kurisu, T. Kinoshita, A. Sugimoto, A. Nagara, Y. Kai, N. Kasai and S. Harada, *J. Biochem.*, 1997, **121**, 304-308.
20. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.
21. A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J.-S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren, *Acta Crystallogr. sect. D*, 1998, **54**, 905-921.
22. L. Holm, S. Kaariainen, C. Wilton and D. Plewczynski, *Curr. Protoc. Bioinformatics*, 2006, **Chapter 5**, Unit 5 5.
23. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. Bryan Arendall, 3rd, J. Snoeyink, J. S. Richardson and D. C. Richardson, *Nucl. Acids Res.*, 2007, **35 (Web Server issue)**, W375-W383.
24. G. Vriend, *J. Mol. Graph.*, 1990, **8**, 52-56.
25. F. Corpet, *Nucl. Acids Res.*, 1988, **16**, 10881-10890.
26. L. Migliolo, A. S. de Oliveira, E. A. Santos, O. L. Franco and M. P. de Sales, *J. Mol. Graph. Model.*, 2010, **29**, 148-156.
27. E. M. Meulenbroek, E. A. Thomassen, L. Pouvreau, J. P. Abrahams, H. Gruppen and N. S. Pannu, *Acta Crystallogr. sect. D*, 2012, **68**, 794-799.
28. T. A. Valueva, I. A. Parfenov, T. A. Revina, E. V. Morozkina and S. V. Benevolensky, *Plant. Physiol. Biochem.*, 2012, **52**, 83-90.
29. R. Bao, C. Z. Zhou, C. Jiang, S. X. Lin, C. W. Chi and Y. Chen, *J. Biol. Chem.*, 2009, **284**, 26676-26684.
30. Z. W. Xie, M. J. Luo, W. F. Xu and C. W. Chi, *Biochemistry*, 1997, **36**, 5846-5852.
31. M. Azarkan, S. Martinez-Rodriguez, L. Buts, D. Baeyens-Volant and A. Garcia-Pino, *J. Biol. Chem.*, 2011, **286**, 43726-43734.

32. J. K. Dattagupta, A. Podder, C. Chakrabarti, U. Sen, D. Mukhopadhyay, S. K. Dutta and M. Singh, *Proteins*, 1999, **35**, 321-331.
33. S. Capaldi, M. Perduca, B. Faggion, M. E. Carrizo, A. Tava, L. Ragona and H. L. Monaco, *J. Struct. Biol.*, 2007, **158**, 71-79.
34. P. Chen, J. Rose, R. Love, C. H. Wei and B. C. Wang, *J. Biol. Chem.*, 1992, **267**, 1990-1994.
35. J. Koepke, U. Ermler, E. Warkentin, G. Wenzl and P. Flecker, *J. Mol. Biol.*, 2000, **298**, 477-491.
36. E. M. Ragg, V. Galbusera, A. Scarafoni, A. Negri, G. Tedeschi, A. Consonni, F. Sessa and M. Duranti, *Febs J.*, 2006, **273**, 4024-4039.
37. J. A. Barbosa, L. P. Silva, R. C. Teles, G. F. Esteves, R. B. Azevedo, M. M. Ventura and S. M. de Freitas, *Biophys. J.*, 2007, **92**, 1638-1650.
38. K. N. Rao and C. G. Suresh, *Biochim. Biophys. Acta*, 2007, **1774**, 1264-1273.
39. Y. N. Sreerama and L. R. Gowda, *Biochim. Biophys. Acta*, 1997, **1343**, 235-242.
40. S. Terada, S. Fujimura and E. Kimoto, *Biosci. Biotechnol. Biochem.*, 1994, **58**, 376-379.
41. S. Norioka and T. Ikenaka, *J. Biochem.*, 1983, **94**, 589-599.
42. K. Nomura and N. Suzuki, *FEBS Lett.*, 1993, **321**, 84-88.
43. A. S. Tanaka, M. U. Sampaio, S. Marangoni, B. de Oliveira, J. C. Novello, M. L. Oliva, E. Fink and C. A. Sampaio, *Biol. Chem.*, 1997, **378**, 273-281.
44. S. Odani, T. Koide and T. Ono, *J. Biochem.*, 1986, **100**, 975-983.
45. J. E. Debreczeni, G. Bunkoczi, B. Girmann and G. M. Sheldrick, *Acta Crystallogr. sect. D*, 2003, **59**, 393-395.
46. J. Krahn and F. C. Stevens, *FEBS Lett.*, 1971, **13**, 339-341.
47. K. A. Wilson and M. Laskowski, Sr., *J. Biol. Chem.*, 1975, **250**, 4261-4267.
48. M. Renko, J. Sabotic, M. Mihelic, J. Brzin, J. Kos and D. Turk, *J. Biol. Chem.*, 2010, **285**, 308-316.
49. F. L. Tan, G. D. Zhang, J. F. Mu, N. Q. Lin and C. W. Chi, *Hoppe Seylers Z Physiol. Chem.*, 1984, **365**, 1211-1217.
50. N. Mallorquí-Fernández, S. P. Manandhar, G. Mallorquí-Fernández, I. Usón, K. Wawrzonek, T. Kantyka, M. Solà, I. B. Thogersen, J. J. Enghild, J. Potempa and F. X. Gomis-Rüth, *J. Biol. Chem.*, 2008, **283**, 2871-2882.
51. V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. sect. D*, 2010, **66**, 12-21.

4. Supplemental Tables

Suppl. Table S1. Janus-faced single-domain inhibitors with at least two reactive sites.								
Family (MEROPS)	Inhibitor name	Source	UniProt code	Class of targets	Targets	PDB code	Ref.	
Kunitz (I3)	ApKTI	<i>Adenanthera pavonina</i>	P09941	SP <sup>a</sup> (and CP <sup>a</sup> )	Trypsin and papain (very weakly)	-	26	
	PKPI	<i>Solanum tuberosum</i>	Q66LL2	SP	Trypsin and chymotrypsin	3TC2	27, 28	
	API	<i>Sagittaria sagittifolia</i>	Q7M1P4	SP	Trypsin, chymotrypsin and kallikrein	3E8L <sup>b</sup>	29, 30	
	PPI	<i>Carica papaya</i>	P80691	SP	Trypsin	3S8K, 3S8J	31	
	WCI	<i>Psophocarpus tetragonolobus</i>	P10822	SP	Chymotrypsin	2WBC	32	
Bowman-Birk (I12)	MSTI	<i>Medicago scutellata</i>	P80321	SP	Trypsin	2ILN <sup>b</sup>	33	
	sBBI	<i>Glycine max</i>	P01055	SP	Trypsin and chymotrypsin	1PI2, 1D6R <sup>b</sup>	34, 35	
	LCTI	<i>Lens culinaris</i>	Q8W4Y8	SP	Trypsin and chymotrypsin	2AIH	36	
	BTCI	<i>Vigna unguiculata</i>	P17734	SP	Trypsin and chymotrypsin	2G81 <sup>c</sup> , 2OT6	37, 38	
	HGI-III	<i>Dolichos biflorus</i>	Q9S9E3	SP	Trypsin and chymotrypsin	-	39	
	CLTI	<i>Canavalia lineata</i>	Q7M1Q0	SP	Trypsin and chymotrypsin	-	40	
	A-II	<i>Arachis hypogaea</i>	P01066	SP	Trypsin and chymotrypsin	Not deposited	41, 42	
	TcTI	<i>Amburana acreana</i>	P83284	SP	Trypsin, plasmin and chymotrypsin	-	43	
	I-2B	<i>Triticum aestivum</i>	P09863	SP	Trypsin	-	44	
	II-4	<i>Triticum aestivum</i>	P09864	SP	Trypsin	-	44	
	LBI	<i>Phaseolus lunatus</i>	P01056	SP	Trypsin and chymotrypsin	1H34	45, 46	
	GBI	<i>Phaseolus vulgaris</i>	P01060	SP	Trypsin, chymotrypsin and elastase	-	47	
	Clitocypin (I48)	Clitocypin	<i>Clitocybe nebularis</i>	Q3Y9I4	CP	Papain and cathepsins L, V, S and K	3H6R, 3H6S <sup>d</sup>	48
	Macrocyprin (I85)	Macrocyprin 1	<i>Macrolepiota procera</i>	B9V973	CP (and SP)	Papain and cathepsins L, V, S and K (Macrocyprin 4 inhibits trypsin weakly)	3H6Q	48
Others	Trypsin inhibitor	<i>Trichosanthes kirilowii</i>	P01069	SP	Trypsin	-	49	

<sup>a</sup> SP and CP stand for serine and cysteine proteases, respectively. <sup>b</sup> Complex with two molecules of trypsin. <sup>c</sup> Complex with trypsin. <sup>d</sup> Complex with cathepsin V.

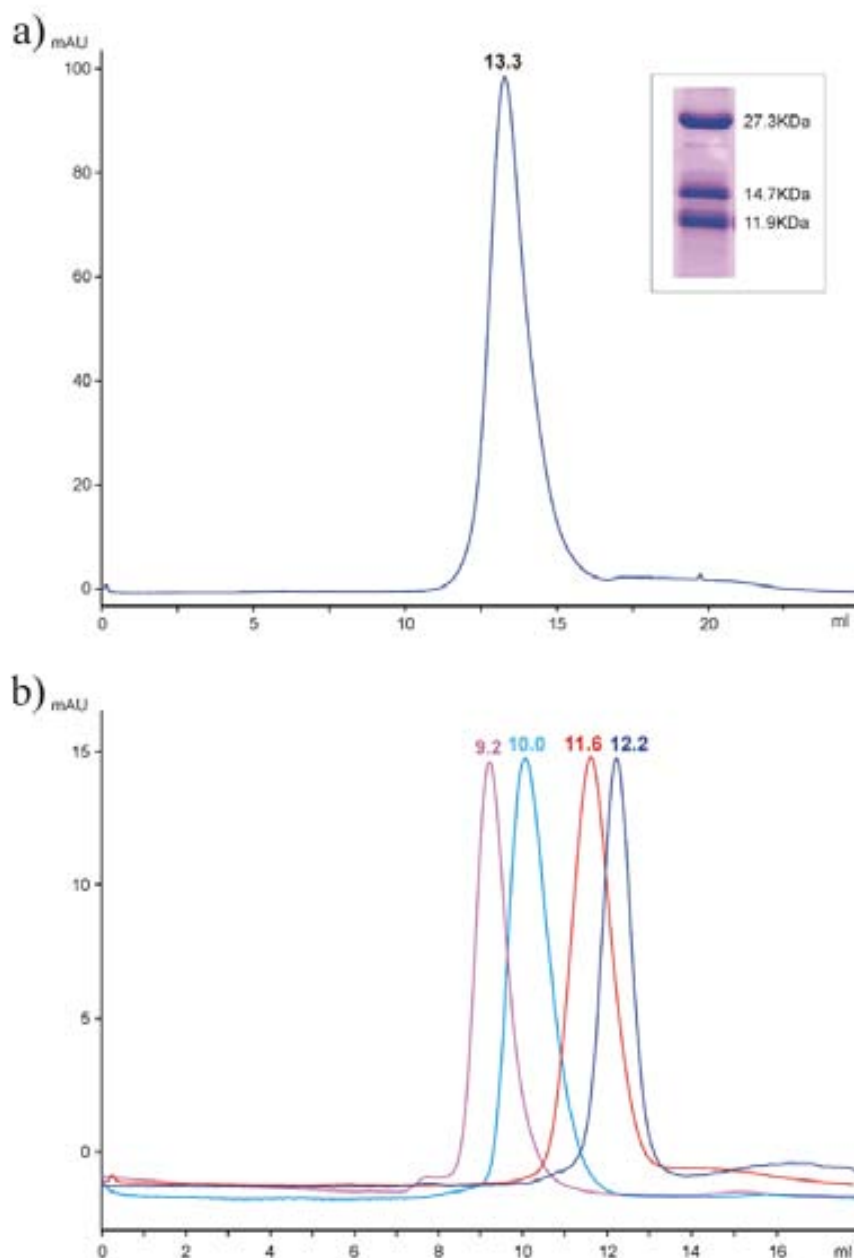
<b>Suppl. Table S2. Inhibitory activity of sermetstatin variants.</b>					
		Wild-type	Y <sup>33</sup> P+T <sup>34</sup> G	M <sup>71</sup> K	Y <sup>72</sup> V+F <sup>73</sup> Y
Protease	Molar ratio (Inh./Enz.)	Inhibitory activity (%)	Inhibitory activity (%)	Inhibitory activity (%)	Inhibitory activity (%)
<b><i>Metallopeptidases</i></b>					
Snapalysin	1	99	1	90	96
	10	100	52	100	100
Astacin	100	0	0	0	0
Ulilysin	100	0	0	0	0
Aeruginolysin	100	0	0	0	0
Thermolysin	100	0	0	0	0
Fragilysin	100	0	0	0	0
ADAM-17 <sup>a</sup>	-	NI <sup>b</sup>	-	-	-
MMP-1,-3,-13 <sup>a</sup>	100	0	0	0	0
<b><i>Serine proteinases</i></b>					
Subtilisin	1	95	95	24	93
Trypsin	10	21	29	53	31
Chymotrypsin	10	69	47	29	56
Elastase	10	68	67	0	62
Proteinase K	1	82	75	5	77
	10	100	100	36	100
Values as mean of three independent measurements, SD within ±5. <sup>a</sup> ADAM stands for a disintegrin and metalloprotease, MMP for matrix metalloproteinase. <sup>b</sup> NI, no inhibition at 4 μM inhibitor concentration in a cell-based assay.					



<b>Suppl. Table S3. Crystallographic data.</b>				
	Sermetstatin		Subtilisin:sermetstatin	Snalypsin:sermetstatin
Dataset	Native	Selenomethionine (absorption peak) <sup>a</sup>	Native	Native
Space group	P <sub>3</sub> ,21	P <sub>3</sub> ,21	C2	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell constants (a, b, c, in Å; β in ° if ≠90)	71.04, 71.04, 52.40	70.81, 70.81, 52.25	183.97, 83.62, 77.62; 110.78	116.54, 121.81, 130.67
Wavelength (Å)	0.8726	0.9791	0.9763	0.8726
No. of measurements / unique reflections	134,041 / 12,234	218,877 / 10,514	312,908 / 52,346	211,0198 / 51,602
Resolution range (Å) (outermost shell) <sup>b</sup>	39.9 – 1.90 (2.00 – 1.90)	26.4 – 2.00 (2.11 – 2.00)	48.6 – 2.25 (2.37– 2.25)	89.1 – 2.70 (2.77– 2.70)
Completeness [ / Anom. completeness] (%)	99.3 (98.9)	99.8 (100.0) / 99.7 (100.0)	99.9 (99.7)	99.4 (95.8)
R <sub>merge</sub> <sup>c</sup>	0.038 (0.596)	0.074 (0.757)	0.097 (0.751)	0.091 (0.704)
R <sub>r.i.m.</sub> (= R <sub>meas</sub> ) <sup>c</sup> [ / R <sub>p.i.m.</sub> ] <sup>c</sup>	0.040 (0.625) / 0.012 (0.185)	0.077 (0.793) / 0.023 (0.235)	0.106 (0.842) / 0.043 (0.368)	0.105 (0.845)
Average intensity (<[I] / σ(<[I]>)]>	40.3 (4.9)	35.4 (8.0)	13.2 (2.0)	14.7 (1.9)
B-Factor (Wilson) (Å <sup>2</sup> ) / Aver. multiplicity	32.7 / 11.0 (11.1)	29.5 / 20.8 (21.8)	41.9 / 6.0 (4.9)	51.9 / 4.1 (3.2)
Resolution range used for refinement (Å)	∞ – 1.90		∞ – 2.25	∞ – 2.70
No. of reflections used (test set)	11,844 (479)		51,487 (797)	50,699 (782)
Crystallographic R <sub>factor</sub> (free R <sub>factor</sub> ) <sup>c</sup>	0.187 (0.214)		0.177 (0.218)	0.195 (0.242)
No. of protein atoms / solvent molecules / neutral ligands / ionic ligands	834 / 126 / 1 glycerol / 3 CH <sub>3</sub> COO <sup>-</sup>		5,487 / 359 / 1 diglycerol, 3 2-propanol, 2 glycerol / 4 Ca <sup>2+</sup> , 2 Zn <sup>2+</sup> , 2 K <sup>+</sup> , 1 Cl <sup>-</sup> , 1 (CH <sub>3</sub> ) <sub>2</sub> AsOO <sup>-</sup> , 1 CH <sub>3</sub> COO <sup>-</sup> , 1 PO <sub>4</sub> <sup>3-</sup>	11,043 / 275 / 2 glycerol / 6 Zn <sup>2+</sup>
R <sub>msd</sub> from target values				
bonds (Å) / angles (°)	0.010 / 1.14		0.010 / 1.12	0.010 / 1.06
Average B-factors for protein atoms (Å <sup>2</sup> )	40.4		51.1	58.2
Main-chain conformational angle analysis <sup>d</sup>				
Residues in favored regions / outliers / all residues	109 / 1 / 112		746 / 1 / 764	1,406 / 2 / 1,436
<sup>a</sup> Friedel-mates were treated as separate reflections. <sup>b</sup> Values in parentheses refer to the outermost resolution shell. <sup>c</sup> For definitions, see Table 1 in <sup>50</sup> . <sup>d</sup> According to MOLPROBITY <sup>51</sup> .				



### 5. Supplemental Figures



**Suppl. Figure S1 – Size-exclusion chromatography of the ternary complex.** (a) The complex between subtilisin, sermetstatin and snapalysin was loaded onto a Superdex 200 10/300 column. The complex eluted after 13.3ml, which corresponds to 102KDa and reveals the formation of a 2+2+2 heterohexameric complex in solution (theoretical mass 107.8KDa). The column was calibrated with the following protein standards: aldolase (158KDa; elution after 12.5ml), conalbumin (75KDa; 13.9ml), ovalbumin (43KDa; 14.8ml), and cytochrome C (12.4KDa; 17.8ml). The presence of the three molecules (subtilisin 27KDa, snapalysin 15KDa, and sermetstatin 12KDa) was further confirmed by SDS-PAGE (see inset) after trichloroacetic acid-precipitation of the collected sample. (b) The isolated inhibitor and the binary and ternary complexes were loaded onto a Superdex 75 10/300 column for comparison. The sermetstatin eluted after 12.2ml (1+1), the snapalysin:sermetstatin complex after 11.6ml (2+2), the subtilisin:sermetstatin complex after 10.0ml (2+2), and the ternary complex after 9.2ml (2+2+2).



# APÉNDICE II

Ultratight crystal packing of a 10 kDa protein



## Ultratight crystal packing of a 10 kDa protein

Sergio Trillo-Muyo,<sup>a</sup>  
Andrius Jasilionis,<sup>b</sup> Marcin J.  
Domagalski,<sup>c</sup> Maksymilian  
Chruszcz,<sup>d</sup> Wlodek Minor,<sup>c</sup>  
Nomeda Kuisiene,<sup>b</sup> Joan L.  
Arolas,<sup>a</sup> Maria Solà<sup>a</sup> and  
F. Xavier Gomis-Rüth<sup>a\*</sup>

<sup>a</sup>Proteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona, Spanish Research Council CSIC, Barcelona Science Park, c/Baldiri Reixac 15-21, 08028 Barcelona, Spain, <sup>b</sup>Department of Microbiology and Biotechnology, Vilnius University, M. K. Čiurlionio 21/27, 03101 Vilnius, Lithuania, <sup>c</sup>Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908-0736, USA, and <sup>d</sup>Department of Chemistry and Biochemistry, University of South Carolina, 631 Sumter Street, Columbia, SC 29208, USA

Correspondence e-mail: xgrcri@ibmb.csic.es

While small organic molecules generally crystallize forming tightly packed lattices with little solvent content, proteins form air-sensitive high-solvent-content crystals. Here, the crystallization and full structure analysis of a novel recombinant 10 kDa protein corresponding to the C-terminal domain of a putative U32 peptidase are reported. The orthorhombic crystal contained only 24.5% solvent and is therefore among the most tightly packed protein lattices ever reported.

Received 7 November 2012

Accepted 7 December 2012

**PDB References:** C-terminal domain of putative U32 peptidase, native, 4he6; selenomethionine variant, 4he5

## 1. Introduction

Unlike inorganic and organic crystals, protein crystals are very fragile (McPherson, 1999). This is a result of far fewer and much weaker interactions contributing to the crystalline architecture in proportion to the molecular mass of the molecule composing the crystal (Drenth & Haas, 1992). This also affects the kinetic parameters of crystal growth: whereas small-molecule crystals crystallize within minutes or a few days, proteins generally require a much longer time, sometimes weeks or months. The majority of these properties result from a single feature: the solvent content (McPherson, 1999). While metal and atomic crystals are completely anhydrous, inorganic ionic crystals may be anhydrous, like rock salt, or may contain a few water molecules per molecular unit. The same holds for organic small-molecule crystals, for which the solvent content is generally between 23 and 35% (Kitaigorodskii, 1973). In contrast, protein crystals have a high solvent content per volume. For most protein crystals, this varies from 40 to 60%; very few protein crystals have been reported with solvent contents of <30% or >90% (McPherson, 1999; Matthews, 1968). In addition, solvent molecules are generally ordered in all parts of organic and inorganic crystals; in contrast, macromolecular crystals have large channels of only partially ordered solvent. As a result, protein crystals cannot be dehydrated and prolonged exposure to air results in destruction of the crystalline order (Bernal & Crowfoot, 1934).

Here, we report the crystal structure determination and crystal-packing analysis of a 10 kDa protein, the C-terminal domain of the putative U32 peptidase from *Geobacillus thermoleovorans*, and analyze the findings in the general context of the Protein Data Bank.

## 2. Materials and methods

### 2.1. Protein production and purification

A gene coding for a hypothetical U32-type peptidase from *G. thermoleovorans* (previously *G. lituanicus*; Dinsdale *et al.*,

**Table 1**

Crystallographic data.

Values in parentheses are for the outermost resolution shell.

Data set	Native	Selenomethionine (absorption peak) <sup>†</sup>
Space group	$P2_12_12_1$	$P3_221$
Unit-cell parameters (Å)	$a = 28.40,$ $b = 36.30,$ $c = 65.67$	$a = b = 45.55,$ $c = 60.90$
Wavelength (Å)	0.9724	0.9791
No. of measurements	200643	131989
No. of unique reflections	27941	26349
Resolution range (Å)	32.8–1.10 (1.16–1.10)	39.5–1.15 (1.21–1.15)
Completeness (%)	98.8 (96.0)	99.7 (98.8)
Anomalous completeness (%)		97.2 (91.0)
$R_{\text{merge}}^{\ddagger}$	0.116 (0.159)	0.036 (0.348)
$R_{\text{r.i.m.}} (= R_{\text{meas}})^{\ddagger}$	0.125 (0.172)	0.045 (0.459)
$R_{\text{p.i.m.}}^{\ddagger}$	0.046 (0.066)	0.026 (0.296)
Average intensity $\{\{I\}/\sigma(I)\}$	13.7 (8.6)	19.0 (3.3)
$B$ factor (Wilson) (Å <sup>2</sup> )	6.6	11.4
Average multiplicity	7.2 (6.3)	5.0 (3.7)
Resolution range used for refinement (Å)	$\infty$ –1.10	$\infty$ –1.15
No. of reflections used	27158	25557
No. of reflections used for test set	734	792
Crystallographic $R$ factor <sup>‡</sup>	0.123	0.141
Free $R$ factor <sup>‡</sup>	0.163	0.170
No. of protein atoms	760	692
No. of solvent molecules	108	80
No. of ligands	1 acetate, 1 cation	1 sulfate
R.m.s.d. from target values, bonds (Å)	0.015	0.014
R.m.s.d. from target values, angles (°)	2.33	2.44
Average $B$ factor for protein atoms (Å <sup>2</sup> )	10.4	18.3
Main-chain conformational angle analysis <sup>§</sup>		
Residues in favoured regions	85	78
Outliers	0	0
All residues	87	80

<sup>†</sup> Friedel mates were treated as separate reflections. <sup>‡</sup> For definitions, see Table 1 in Mallorquí-Fernández *et al.* (2008). <sup>§</sup> According to *MolProbity* (Chen *et al.*, 2010).

2011; 422 residues; 47 928 Da; UniProt code G5DCB7; L2V mutation for cloning strategy) was cloned into a modified pET-28a vector using *NcoI* and *SalI* restriction sites and was verified by DNA sequencing. The protein was produced by heterologous overexpression in *Escherichia coli* BL21 (DE3) cells, which were grown at 310 K in Luria–Bertani medium supplemented with kanamycin to a final concentration of 30  $\mu\text{g ml}^{-1}$ . Cell cultures were induced at an OD<sub>550</sub> of 0.8 with isopropyl  $\beta$ -D-1-thiogalactopyranoside to a final concentration of 1 mM and growth was continued for 5 h at 310 K. The selenomethionine variant was obtained in the same way except that the cells were grown in minimal medium containing selenomethionine (Sigma) instead of methionine. After centrifugation at 7000g for 30 min at 277 K, the pellet was washed twice with buffer *A* (50 mM Tris–HCl, 250 mM NaCl, 20 mM  $\beta$ -mercaptoethanol pH 8.0) and resuspended in the same buffer containing EDTA-free protease-inhibitor cocktail tablets (Roche Diagnostics) and DNase I (Roche Diagnostics). The cells were lysed at 277 K using a cell disrupter (Constant Systems) at a pressure of 135 MPa, the cell debris was removed by centrifugation at 50 000g for 1 h at 277 K and the supernatant was filtered (0.22  $\mu\text{m}$  pore size;

Millipore). The protein was found to serendipitously bind nickel–nitrilotriacetic acid (Ni–NTA) resin despite the absence of a polyhistidine tag. As such, the sample was incubated with Ni–NTA resin (Invitrogen) previously equilibrated with buffer *A* and eluted using the same buffer plus 150 mM imidazole. The protein was subsequently purified by size-exclusion chromatography on a HiLoad 26/60 Superdex 75 column previously equilibrated with buffer *B* [20 mM Tris–HCl, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), pH 8.0]. The protein identity and purity were assessed by Edman degradation, peptide mass fingerprinting and 15% Tricine–SDS–PAGE stained with Coomassie Blue. Fractions containing the  $\sim$ 48 kDa protein were pooled, concentrated to 30 mg ml<sup>−1</sup> by ultrafiltration using Vivaspin 15 filter devices with a 10 kDa cutoff (Sartorius Stedim Biotech) and incubated for 48 h at 310 K to test the stability of the protein over time. After incubation, the sample showed strong precipitation and was therefore centrifuged at 16 000g for 10 min at 277 K; the supernatant was subsequently filtered. SDS–PAGE analysis of this supernatant showed the presence of a major band at  $\sim$ 10.4 kDa, which we attribute to reproducible heterologous cleavage by a contaminating peptidase. This protein species was purified by size-exclusion chromatography on a HiLoad 16/60 Superdex 75 column previously equilibrated with buffer *C* (20 mM Tris–HCl, 50 mM NaCl, 1 mM TCEP pH 7.5) and eluted as a monomer. Edman-degradation and mass-spectrometric analyses (10 392 Da; experimental molecular mass with internal calibration) revealed that this species corresponded to the C-terminal region of the U32-type peptidase spanning the segment Ser334–Asn422 (hereafter referred to as GT-U32-CTD; 89 residues; calculated molecular mass 10 395 Da). The protein was further concentrated to 75 mg ml<sup>−1</sup> using Vivaspin 15 and 500 filter devices with a 5 kDa cutoff. The protein concentration was determined by measuring the absorbance at 280 nm using a spectrophotometer (NanoDrop) and a calculated absorption coefficient  $\epsilon_{0.1\%}$  of 0.82.

## 2.2. Crystallization and X-ray diffraction data collection

Crystallization assays for GT-U32-CTD were carried out by the sitting-drop vapour-diffusion method using 96  $\times$  2-well MRC plates (Innovadyne) and a Cartesian nanodrop robot (Genomic Solutions) at the IBMB/IRB crystallization service. Crystallization plates were stored in Bruker steady-temperature crystal farms at 277 and 293 K. Successful hits were scaled up to the microlitre range in 24-well Cryschem crystallization plates (Hampton Research). The best crystals of native GT-U32-CTD appeared at 293 K using protein solution at 75 mg ml<sup>−1</sup> in buffer *C* containing 1 mM phenylmethylsulfonyl fluoride and reservoir solution consisting of 100 mM sodium acetate, 200 mM ammonium acetate, 30% (w/v) polyethylene glycol 4000 pH 4.6 in microlitre plates. The best crystals of the selenomethionine-derivatized protein were obtained from microlitre drops at 293 K using protein solution at 75 mg ml<sup>−1</sup> in buffer *C* and reservoir solution consisting of 100 mM sodium citrate, 2 M ammonium sulfate pH 5.5.

Crystals were cryoprotected by successive passages through reservoir solution containing increasing amounts of glycerol [up to 20–25% (v/v)]. Complete diffraction data sets were collected at 100 K from liquid-nitrogen flash-cryocooled crystals (Oxford Cryosystems 700 series cryostream) using a PILATUS 6M pixel detector (Dectris) on beamline ID29 of the European Synchrotron Radiation Facility (ESRF) Grenoble, France within the Block Allocation Group 'BAG Barcelona' (native protein) and using an ADSC Q315R CCD detector on beamline PROXIMA 1 of synchrotron SOLEIL, Paris, France (selenomethionine-derivatized protein). The crystals of native and selenomethionine-derivatized protein were orthorhombic (maximal resolution 1.10 Å) and trigonal (maximal resolution 1.15 Å), respectively, with one molecule per asymmetric unit. Mass-spectrometric analysis of carefully washed crystals revealed the presence of full-length GT-U32-CTD (approximately 60%) and three shorter forms that lacked two, three and five N-terminal residues, respectively (approximately 40% in total). Diffraction data were integrated, scaled, merged and reduced using the programs *XDS* (Kabsch, 2010) and *SCALA* (Evans, 2006) within the *CCP4* suite (Winn *et al.*, 2011; Table 1).

### 2.3. Structure determination and refinement

The structure was determined by single-wavelength anomalous diffraction using the selenomethionine derivative and the program *SHELXD* (Sheldrick, 2010). Diffraction data were collected from a crystal at the selenium absorption-peak wavelength, as inferred from a previous XANES fluorescence scan, and enabled the program to identify the two selenium sites of the monomer present in the asymmetric unit. Subsequent phasing with *SHELXE* (Sheldrick, 2010), including implementation of the 'free-lunch' algorithm and auto-tracing, resolved the twofold ambiguity intrinsic to a SAD experiment owing to the difference in the values of the pseudo-free correlation coefficients of the two possible hands (79.2% versus 49.8%), which confirmed  $P3_221$  as the correct space group. An initial model was built into the experimental electron density with the program *TURBO-FRODO* (Carranza *et al.*, 1999) on a Silicon Graphics Octane2 workstation and was refined with *PHENIX* (Adams *et al.*, 2002), *BUSTER* (Blanc *et al.*, 2004) and *SHELXL* (Sheldrick, 2008). This model was employed to solve the native structure (true space group  $P2_12_12_1$ ) by molecular replacement (Huber, 1965) using *Phaser* (McCoy, 2007), which rendered one unambiguous solution with *Z*-scores of 9.7 and 13.8 for the rotation and translation functions, respectively. Model building and refinement, which included TLS refinement and refinement of anisotropic thermal displacement parameters, proceeded as described above. The final native model comprised all of the residues of the protein (Ser334–Asn422) plus one oxygen-pentacoordinated cation, an acetate ion and 108 solvent molecules. Table 1 provides a summary of data collection and final model refinement.

### 2.4. Solvent-content analysis

A solvent-content distribution for all X-ray crystal structures deposited in the Protein Data Bank (PDB) before September 2012 was calculated for a nonredundant set of unique crystal forms. To obtain this data set, structures that share equivalent asymmetric units were clustered together. The crystal structures in each cluster possessed the same polypeptide composition, were determined in the same space group and showed differences of up to 5% in unit-cell volume. The working pipeline was composed of the following steps. Firstly, to determine the polypeptide composition, each protein structure was labelled with a formula describing the number of polypeptides and their sequence-cluster identifiers. The pre-calculated sequence clusters had been generated by *BLAST* (Altschul *et al.*, 1990) using a 90% identity threshold (obtained from <ftp://resources.rcsb.org/sequence/clusters/>). At this stage, the data set contained clusters of all possible protein assemblies. Sequences shorter than 20 amino acids, split entries, virus capsids, synthetic constructs, complexes with nucleic acids, polysaccharides and D-polypeptides were excluded from the analysis. Structures missing 20% or more of the residues of the associated sequence were likewise rejected. Next, the aforementioned clusters were subdivided based on the crystal space group and unit-cell volume. The latter were calculated based on the unit-cell dimensions defined in the mmCIF files using the *MATTHEWS\_COEF* program from the *CCP4* suite (Matthews, 1968; Kantardjieff & Rupp, 2003). The molecular weight of the polypeptides in the asymmetric unit was calculated from the sequences reported by the authors using the *ProtParam* methodology (Gasteiger *et al.*, 2005). Representatives of each final cluster were chosen based on a simple quality factor defined as  $1/\text{resolution} - R$  value. For the resulting data set, the solvent content of each protein crystal structure was determined using *MATTHEWS\_COEF* (Kantardjieff & Rupp, 2003; Matthews, 1968). Subsequently, manual curation of the final data set enabled the removal of special cases in which the structures were the result of ensemble refinement (PDB entries 1cwq and 1gtv), desiccated crystals and those obtained by phase transition (PDB entries 1c0c, 1xek, 1xej and 1v7t), crystals with 'order-disorder' (Pletnev *et al.*, 2009; PDB entry 3h1r) and other nonstandard protocols for structure determination (PDB entries 3gi0 and 2xge). Cases for which the solvent content was artificially low owing to long missing N- and C-terminal fragments were not considered either (PDB entries 2xnq, 2duy, 2axo, 3bqh, 2f9l, 3m4s, 3nzl, 2xjx, 4eti, 2ds8 and 1vfq). The final resulting data set consisted of 35 656 X-ray structures.

### 2.5. Miscellaneous

Figures were prepared with *TURBO-FRODO* and *UCSF Chimera* (Pettersen *et al.*, 2004). The total interaction surface (taken as half of the surface area buried at a complex interface; probe radius 1.4 Å) and close contacts (defined as atoms separated by less than 4 Å) were determined using *CNS* (Brünger *et al.*, 1998). Pairwise interaction surfaces were calculated with the *PISA* server (<http://www.ebi.ac.uk/msd-srv/>

prot\_int/pistart.html; Krissinel & Henrick, 2005). Surface shape complementarity was calculated with the program *SC* (Lawrence & Colman, 1993) within *CCP4* with a probe radius of 1.4 Å. Structure similarities were investigated with *DALI* (Holm & Rosenström, 2010). Model validation was performed with *MolProbity* (Chen *et al.*, 2010) and the *WHAT\_CHECK* routine of *WHAT IF* (Vriend, 1990). The final coordinates of native and selenomethione-derivatized GT-U32-CTD have been deposited in the PDB (<http://www.pdb.org>; accession codes 4he5 and 4he6).

### 3. Results and discussion

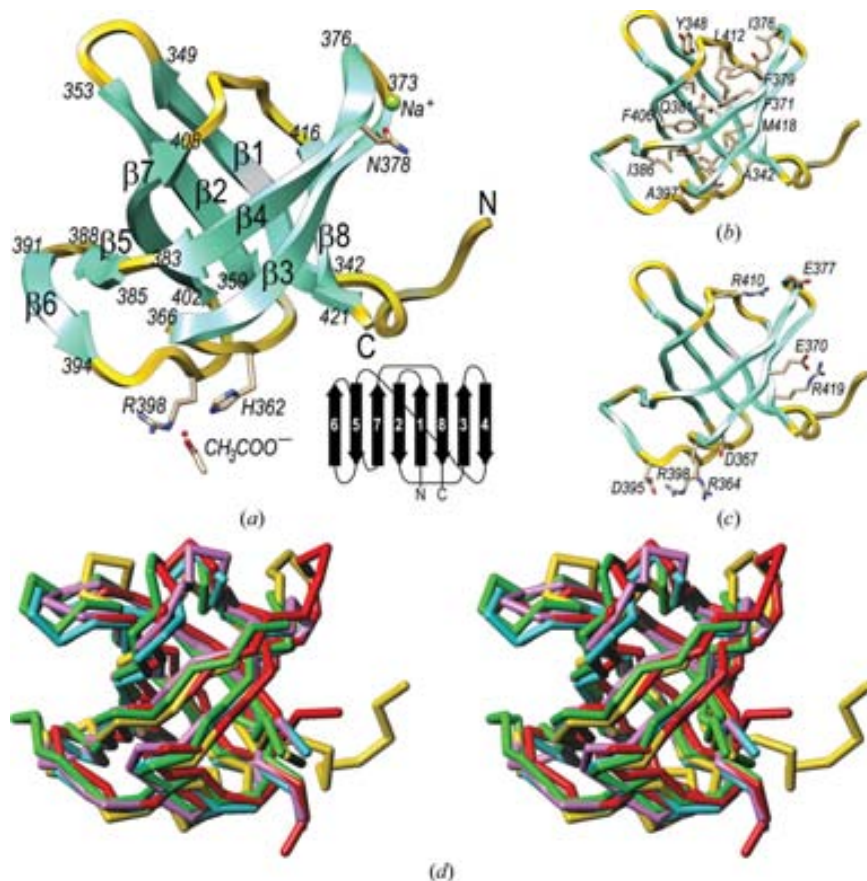
#### 3.1. Crystallization and structure determination

We studied the crystal structure of the C-terminal domain of a putative peptidase from *G. thermoleovorans* belonging to MEROPS (<http://merops.sanger.ac.uk>; Rawlings *et al.*, 2012) family U32 (GT-U32-CTD), which was obtained by spontaneous proteolytic fragmentation of the recombinant full-length protein. Native GT-U32-CTD crystallized in space

group  $P2_12_12_1$  using 100 mM sodium acetate, 200 mM ammonium acetate, 30% (*w/v*) polyethylene glycol 4000 pH 4.6 as the reservoir solution and the crystals diffracted to 1.10 Å resolution (see §2 and Table 1). The structure was determined by single-wavelength anomalous diffraction of a selenomethione-derivatized variant of the protein, which crystallized in space group  $P3_221$  using 100 mM sodium citrate, 2 M ammonium acetate pH 5.5 as the reservoir solution; the crystals diffracted to 1.15 Å resolution. The resulting molecular model was employed to solve the structure of the orthorhombic crystal form, which revealed all of the residues of the only chain present in the asymmetric unit plus one (hypothetical) sodium cation, an acetate ion and 108 solvent molecules. Table 1 provides a summary of the crystallographic data-collection and final model-refinement parameters.

#### 3.2. Architecture of GT-U32-CTD

The structure shows a compact distorted open  $\beta$ -barrel made up of eight  $\beta$ -strands ( $\beta 1$ – $\beta 8$ ), into which the first eight residues (Ser334–Phe341) and the last single residue (Asn422) are inserted (Fig. 1*a*). The first six residues are present with a refined occupancy of 60%, which is in accordance with mass-spectrometric analysis of carefully washed crystals (see §2). The barrel is actually arranged as a strongly twisted, curled and arched antiparallel  $\beta$ -sheet ( $\beta 6$ – $\beta 5$ – $\beta 7$ – $\beta 2$ – $\beta 1$ – $\beta 8$ – $\beta 3$ – $\beta 4$ ; connectivity  $-1, +3, +1, -6, -1, +2, +3$ ); the flanking strands,  $\beta 4$  and  $\beta 6$ , do not interact with each other through their main chains. In addition,  $\beta 4$  and  $\beta 7$  are close to giving rise to a perfect barrel with  $\beta 1$ – $\beta 3$  and  $\beta 8$  (Fig. 1*a*), but they do not contact each other through their main chains either. Instead, the gap is closed by the side chains of Phe379, Gln381, Lys409 and Arg410 (Fig. 1*b*). The sheet wraps around a central hydrophobic core formed by the side chains of residues Ala342, Val345, Tyr348, Ala355, Val357, Ala359, Phe363, Val369, Phe371, Ile376, Phe379, Ile383, Gln381, Ala397, Val404, Phe406, Val408, Leu412, Asn416 and Met418. In contrast, the surface is mainly hydrophilic and shows seven lysines, five arginines, five aspartates and ten glutamates, which contribute to four salt bridges (Asp367–Arg364, Glu370–Arg419, Glu377–Arg410 and Asp395–Arg398; Fig. 1*c*). Furthermore, a potential sodium cation was tentatively assigned on the surface based on very short binding distances to five liganding O atoms (Asn378 O <sup>$\delta 1$</sup> , 2.35 Å; Gly373 O, 2.38 Å; Ile376 O, 2.28 Å; a solvent molecule, 2.32 Å; a symmetry-related Asp395 O <sup>$\delta 2$</sup>  atom, 2.38 Å) arranged in a trigonal bipyramidal coordination sphere. Finally, an



**Figure 1**

(*a*) Ribbon-type plot of GT-U32-CTD with the eight  $\beta$ -strands labelled and marked with their flanking residues. Residues participating in ion binding are also shown and labelled. Inset (lower right): topology scheme of the protein. (*b*) Cartoon showing the side chains participating in the central hydrophobic core. (*c*) As (*b*) but showing the charged residues of the protein engaged in salt bridges. (*d*)  $C^\alpha$ -type plot in stereo showing the superposed structures of GT-U32-CTD (yellow), LepA (green), eRF3 (lilac), EF-1a (cyan) and EF-Tu (red).



acetate ion was likewise found bound to the protein surface anchored to His362 N<sup>e2</sup> (2.90 Å), Arg398 N<sup>η2</sup> (2.85 Å) and a solvent molecule (2.64 Å) through one of its carboxylate O atoms. The other O atom is bound by Arg398 N<sup>ε</sup> (2.89 Å), a solvent molecule (2.74 Å) and a symmetry-related Gln358 N<sup>ε2</sup> atom (3.20 Å).

### 3.3. Structural relatives

Structure-similarity searches with GT-U32-CTD identified the LepA protein (PDB entry 3cb4; Evans *et al.*, 2008), release factor eRF3 (PDB entry 3e20; Cheng *et al.*, 2009), elongation factor 1a (EF-1a; PDB entry 1skq; Vitagliano *et al.*, 2004) and elongation factor Tu (EF-Tu; PDB entry 2c77; Parmeggiani *et al.*, 2006) as the most similar structures, with Z-scores of 10.9–11.0, r.m.s.d. values of 1.7–2.0 Å and aligned stretches of 78–80 residues. Superposition of these structures onto GT-U32-CTD (Fig. 1*d*) revealed equivalent architectures, topologies and connectivities despite negligible sequence identity (8–19%). These structural relatives are all found in proteins that further comprise an additional N-terminal guanine-nucleotide binding domain and they participate in ribosomal protein translation elongation or termination (Andersen *et al.*, 2003). For the elongation factors, it has been shown that the GT-U32-CTD-like domains interact with aminoacyl-loaded tRNAs and with antibiotics that target the protein-synthesis machinery (Andersen *et al.*, 2003; Parmeggiani *et al.*, 2006). Accordingly, a similar function to a protein engaged in binding is conceivable for the C-terminal domain of the putative U32-type peptidase of *G. thermoleovorans*.

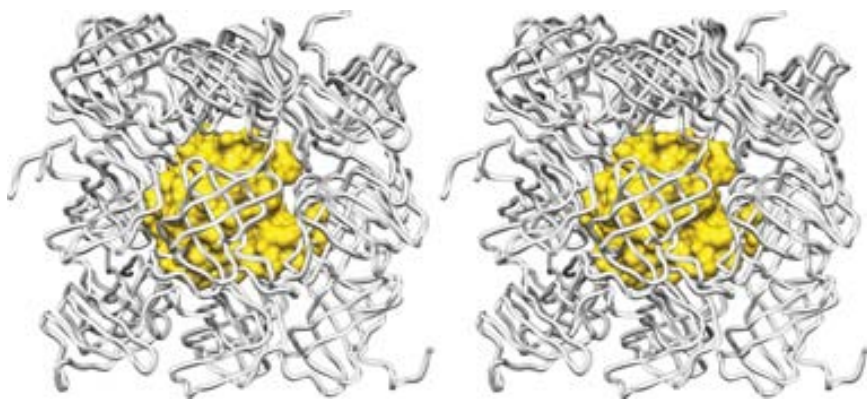
### 3.4. Crystal-packing analysis

Notwithstanding the interest of the abovementioned results, the most striking feature of the present study is that the native protein crystallized in an orthorhombic crystal form with an extremely low solvent content of 24.5% (Matthews parameter  $V_M$  of 1.63 Å<sup>3</sup> Da<sup>-1</sup>; Matthews, 1968), which may at least partially account for its strong diffraction power; crystals with less solvent tend to diffract better (Kantardjieff & Rupp,

2003). In contrast, the selenomethionine-derivatized protein, which only differs from the former by the replacement of two S atoms by selenium within the 10 kDa molecule, crystallized in a trigonal (*i.e.* higher symmetry) space group with 30.0% solvent content ( $V_M = 1.75$  Å<sup>3</sup> Da<sup>-1</sup>). In addition, none of the abovementioned structural relatives packed with less than 39% solvent. Detailed inspection of the native crystal lattice (Fig. 2) confirmed very tight binding of the protein molecules within the crystal, with almost no solvent channels and only ordered solvent molecules ‘gluing’ the protein molecules together. Each protein molecule interacts directly with its surrounding crystallographic neighbours through a total of 136 close contacts and a total contact surface of 3480 Å<sup>2</sup>, which is in good agreement with the value obtained by adding the binary interaction surfaces with each neighbour when calculated by another approach (3590 Å<sup>2</sup>; see §2). The total contact surface represents 65% of the total surface of a monomer, which is far larger than generally reported for crystal contact surfaces (15–49%; Carugo & Argos, 1997). In comparison, tight protein–protein complexes, which exist not only in the crystalline state but also in solution, such as proteinase–inhibitor and antibody–antigen complexes, interact through surfaces of 600–1000 Å<sup>2</sup>, which correspond to 5–20% of their total surfaces (Janin & Chothia, 1990). The shape complementarity between the surface of one GT-U32-CTD protomer and its surrounding crystal relatives is 0.69. This falls within the ranges of values for proteinase–inhibitor complexes (0.70–0.76) and antibody–antigen complexes (0.64–0.68) (Lawrence & Colman, 1993), which is noteworthy taking into account the overall size of the total contact surface. Taken together, all of these features explain the very tightly packed crystal lattice.

### 3.5. Low-solvent crystal structures in the PDB

To assess the uniqueness of the present packing in the general context of the PDB, we investigated the cluster of tightly packed protein crystal structures that have been deposited. Detailed studies of the solvent-content distribution have been published based on 116 (Matthews, 1968), 15 641 (Kantardjieff & Rupp, 2003) and 9081 (Chruszcz *et al.*, 2008) macromolecular crystal structures. These studies concluded that ~0.5% of nonredundant protein structures contain 25% solvent or less (Kantardjieff & Rupp, 2003), which is consistent with recent estimates that less than 1% of proteins pack with less than 25–26% solvent (C. Weichenberger & B. Rupp, personal communication). As the annotation of PDB entries with respect to solvent content is inaccurate (Andersson & Hovmöller, 2000), we decided to perform a comprehensive search for low-solvent-content crystal structures (see §2). We calculated the solvent contents from the molecular masses of the annotated protein sequences of 35 656 nonredundant protein



**Figure 2**

Cartoon in stereo depicting the packing of one GT-U32-CTD molecule (shown with its Connolly surface in yellow) surrounded by its crystallographic symmetry equivalents (shown as white ropes).

**Table 2**

Crystal structures with &lt;25% solvent content.

PDB code	Solvent content (%)	Space group	Cell volume (Å <sup>3</sup> )	Theoretical sequence length (amino acids)	No. of residues actually observed	Quality factor†
3jvb	21.32	I23	1074906	243	204	0.2968
2uwr	21.77	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	57793	79	79	0.5393
1zp5	22.53	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	114865	163	163	0.3366
2f2v	22.64	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	45796	62	62	0.3653
2oh7	22.72	I23	1084536	248	247	0.2862
2guv	24.12	P12 <sub>1</sub> 1	109779	280	280	0.5156
1p9g	24.29	P12 <sub>1</sub> 1	13536	41	40	1.1221
3gw1	24.48	P12 <sub>1</sub> 1	66477	176	168	0.1520
2erl	24.50	C121	28766	40	40	0.8710
3zr8	24.86	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	48935	65	65	0.9844

† Defined as 1/resolution – R value.

crystal forms from the PDB determined using X-ray crystallography (Berman *et al.*, 2000), excluding virus capsids, protein–DNA and protein–RNA complexes, synthetic constructs and peptides (defined as comprising 20 residues or less). Artificially reduced values such as those obtained by crystal desiccation or treatment with organic solvents were also omitted. These calculations revealed only ten non-redundant structures (~0.028%) with a solvent content equal to or less than 26% (see Table 2). These are human CD59 (79 residues; PDB entry 2uwr; Leath *et al.*, 2007), human matrix metalloproteinase 8 (163 residues; PDB entry 1zp5; Campestre *et al.*, 2006), the SH3 domain of chicken  $\alpha$ -spectrin (62 residues; PDB entry 2f2v; Casares *et al.*, 2007), two structures of cypoviral polyhedrines from *Wiseana signata* NPV (243 residues; PDB entry 3jvb; Coulibaly *et al.*, 2007, 2009) and *Bombyx mori* (248 residues; PDB entry 2oh7; Coulibaly *et al.*, 2007), a five-stranded phenylalanine zipper (56 residues; PDB entry 2guv; Liu *et al.*, 2006), the CLP-protease adaptor protein (176 residues; PDB entry 3gw1; Román-Hernández *et al.*, 2009), a small antifungal protein from *Ecommia ulmoides* (41 residues; PDB entry 1p9g; Xiang *et al.*, 2004) and the mating pheromone Er-1 from *Euplotes raikovi* (40 residues; PDB entry 2erl; Anderson *et al.*, 1996), with five and three disulfide bridges, respectively, and effector protein AVR3A11 from *Phytophthora capsici* (65 residues; PDB entry 3zr8; Boutemy *et al.*, 2011). Interestingly, four of the ten abovementioned structures crystallized in space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, which is also the presently reported space group (see Table 1), and the other three in P2<sub>1</sub>. This correlates well with the observation by Chruszcz and coworkers that these two space groups, together with P1, show the lowest mean solvent content in general (see Table 3 in Chruszcz *et al.*, 2008).

#### 4. Conclusion

Accordingly, the present orthorhombic crystal form of GT-U32-CTD represents one of the most tightly packed protein structures reported to date; it shows a solvent content that is more similar to those attributable to crystals of small organic compounds than to a 10 kDa protein.

We are particularly indebted to Bernhard Rupp for his interest, assistance and helpful discussion. We also thank Tibusay Guevara and the IBMB/IRB crystallization service for assistance during crystallization experiments. This study was supported in part by grants from the American, European, Lithuanian, Spanish and Catalan agencies (FP7-HEALTH-F3-2009-223101 'AntiPathoGN', FP7-HEALTH-2010-261460 'Gums & Joints', FP7-PEOPLE-2011-ITN-290246 'RAPID', 31V-151 'COSMETIZYM', BIO2009-10334, BFU2009-07134/BMC, BFU2012-32862, BFU2012-33516, CSD2006-00015, a JAE postdoctoral contract from CSIC, an FPU PhD fellowship from the Spanish Ministry for Science and Technology and Fundació 'La Marató de TV3' grants 2009-100732 and 2009SGR1036). We acknowledge the help provided by the ESRF synchrotron local contacts and, in particular, the excellent assistance and outstanding dedication of Andrew Thompson of SOLEIL beamline PROXIMA 1. Funding for travelling and data collection was provided in part by the ESRF and SOLEIL.

#### References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* **215**, 403–410.
- Andersen, G. R., Nissen, P. & Nyborg, J. (2003). *Trends Biochem. Sci.* **28**, 434–441.
- Anderson, D. H., Weiss, M. S. & Eisenberg, D. (1996). *Acta Cryst.* **D52**, 469–480.
- Andersson, K. M. & Hövmöller, S. (2000). *Acta Cryst.* **D56**, 789–790.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
- Bernal, J. D. & Crowfoot, D. (1934). *Nature (London)*, **133**, 794–795.
- Blanc, E., Roversi, P., Vornrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 2210–2221.
- Boutemy, L. S., King, S. R., Win, J., Hughes, R. K., Clarke, T. A., Blumenschein, T. M., Kamoun, S. & Banfield, M. J. (2011). *J. Biol. Chem.* **286**, 35834–35842.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Campestre, C., Agamennone, M., Tortorella, P., Preziuso, S., Biasone, A., Gavuzzo, E., Pochetti, G., Mazza, F., Hiller, O., Tschesche, H., Consalvi, V. & Gallina, C. (2006). *Bioorg. Med. Chem. Lett.* **16**, 20–24.
- Carranza, C., Inisan, A.-G., Mouthuy-Knoops, E., Cambillau, C. & Roussel, A. (1999). *AFMB Activity Report 1996–1999*, pp. 89–90. Marseille: CNRS-UPR 9039.
- Carugo, O. & Argos, P. (1997). *Protein Sci.* **6**, 2261–2263.
- Casares, S., López-Mayorga, O., Vega, M. C., Cámara-Artigas, A. & Conejero-Lara, F. (2007). *Proteins*, **67**, 531–547.
- Cheng, Z., Saito, K., Pisarev, A. V., Wada, M., Pisareva, V. P., Pestova, T. V., Gajda, M., Round, A., Kong, C., Lim, M., Nakamura, Y., Svergun, D. I., Ito, K. & Song, H. (2009). *Genes Dev.* **23**, 1106–1118.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Chruszcz, M., Potrzebowski, W., Zimmerman, M. D., Grabowski, M., Zheng, H., Lasota, P. & Minor, W. (2008). *Protein Sci.* **17**, 623–632.

- Coulibaly, F., Chiu, E., Gutmann, S., Rajendran, C., Haebel, P. W., Ikeda, K., Mori, H., Ward, V. K., Schulze-Briese, C. & Metcalf, P. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 22205–22210.
- Coulibaly, F., Chiu, E., Ikeda, K., Gutmann, S., Haebel, P. W., Schulze-Briese, C., Mori, H. & Metcalf, P. (2007). *Nature (London)*, **446**, 97–101.
- Dinsdale, A. E., Halket, G., Coorevits, A., Van Landschoot, A., Busse, H. J., De Vos, P. & Logan, N. A. (2011). *Int. J. Syst. Evol. Microbiol.* **61**, 1802–1810.
- Drenth, J. & Haas, C. (1992). *J. Cryst. Growth*, **122**, 107–109.
- Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.
- Evans, R. N., Blaha, G., Bailey, S. & Steitz, T. A. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 4673–4678.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa: Humana Press.
- Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
- Huber, R. (1965). *Acta Cryst.* **19**, 353–356.
- Janin, J. & Chothia, C. (1990). *J. Biol. Chem.* **265**, 16027–16030.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- Kitaigorodskii, A. I. (1973). *Molecular Crystals and Molecules*. London: Academic Press.
- Krissinel, E. & Henrick, K. (2005). *CompLife 2005*, edited by M. R. Berthold, R. Glen, K. Diederichs, O. Kohlbacher & I. Fischer, pp. 163–174. Berlin, Heidelberg: Springer.
- Lawrence, M. C. & Colman, P. M. (1993). *J. Mol. Biol.* **234**, 946–950.
- Leath, K. J., Johnson, S., Roversi, P., Hughes, T. R., Smith, R. A. G., Mackenzie, L., Morgan, B. P. & Lea, S. M. (2007). *Acta Cryst.* **F63**, 648–652.
- Liu, J., Zheng, Q., Deng, Y., Kallenbach, N. R. & Lu, M. (2006). *J. Mol. Biol.* **361**, 168–179.
- Mallorquí-Fernández, N., Manandhar, S. P., Mallorquí-Fernández, G., Usón, I., Wawrzonek, K., Kantyka, T., Solà, M., Thøgersen, I. B., Engchild, J. J., Potempa, J. & Gomis-Rüth, F. X. (2008). *J. Biol. Chem.* **283**, 2871–2882.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J. (2007). *Acta Cryst.* **D63**, 32–41.
- McPherson, A. (1999). *Crystallization of Biological Macromolecules*. New York: Cold Spring Harbor Laboratory Press.
- Parmeggiani, A., Krab, I. M., Okamura, S., Nielsen, R. C., Nyborg, J. & Nissen, P. (2006). *Biochemistry*, **45**, 6846–6857.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Pletnev, S., Morozova, K. S., Verkhusha, V. V. & Dauter, Z. (2009). *Acta Cryst.* **D65**, 906–912.
- Rawlings, N. D., Barrett, A. J. & Bateman, A. (2012). *Nucleic Acids Res.* **40**, D343–D350.
- Román-Hernández, G., Grant, R. A., Sauer, R. T. & Baker, T. A. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 8888–8893.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M. (2010). *Acta Cryst.* **D66**, 479–485.
- Vitagliano, L., Ruggiero, A., Masullo, M., Cantiello, P., Arcari, P. & Zagari, A. (2004). *Biochemistry*, **43**, 6630–6636.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Xiang, Y., Huang, R.-H., Liu, X.-Z., Zhang, Y. & Wang, D.-C. (2004). *J. Struct. Biol.* **148**, 86–97.



# APÉNDICE III

Functional and structural insights into astacin  
metallopeptidases



## Review

# Functional and structural insights into astacin metallopeptidases

F. Xavier Gomis-Rüth<sup>1,\*</sup>, Sergio Trillo-Muyo<sup>1</sup>  
and Walter Stöcker<sup>2,\*</sup>

<sup>1</sup>Proteolysis Lab, Molecular Biology Institute of Barcelona, CSIC, Barcelona Science Park, Helix Building, c/Baldiri Reixac, 15-21, E-08028 Barcelona, Spain

<sup>2</sup>Institute of Zoology, Cell and Matrix Biology, Johannes Gutenberg University, Johannes-von-Müller-Weg 6, D-55128 Mainz, Germany

\*Corresponding authors

e-mail: fxgr@ibmb.csic.es; stoecker@uni-mainz.de

## Abstract

The astacins are a family of multi-domain metallopeptidases with manifold functions in metabolism. They are either secreted or membrane-anchored and are regulated by being synthesized as inactive zymogens and also by co-localizing protein inhibitors. The distinct family members consist of N-terminal signal peptides and pro-segments, zinc-dependent catalytic domains, further downstream extracellular domains, transmembrane anchors, and cytosolic domains. The catalytic domains of four astacins and the zymogen of one of these have been structurally characterized and shown to comprise compact ~200-residue zinc-dependent moieties divided into an N-terminal and a C-terminal sub-domain by an active-site cleft. Astacins include an extended zinc-binding motif (HEXXHXXGXXH) which includes three metal ligands and groups them into the metzincin clan of metallopeptidases. In mature, unbound astacins, a conserved tyrosine acts as an additional zinc ligand, which is swung out upon substrate or inhibitor binding in a ‘tyrosine switch’ motion. Other characteristic structural elements of astacin catalytic domains are three large  $\alpha$ -helices and a five-stranded  $\beta$ -sheet, as well as two or three disulfide bonds. The N-terminal pro-segments are variable in length and rather unstructured. They inhibit the catalytic zinc following an ‘aspartate-switch’ mechanism mediated by an aspartate embedded in a conserved motif (FXGD). Removal of the pro-segment uncovers a deep and extended active-site cleft, which in general shows preference for aspartate residues in the specificity pocket ( $S_1'$ ). Furthermore, astacins undergo major rearrangement upon activation within an ‘activation domain,’ and show a slight hinge movement when binding substrates or inhibitors. In this review, we discuss the overall architecture of astacin catalytic domains and their involvement in function and zymogenic activation.

**Keywords:** bone morphogenetic protein; catalytic domain; meprin; metzincin; tolloid; zinc metallopeptidase.

## Introduction: a short historical background

The first report on the digestive protease astacin from the European freshwater crayfish, *Astacus astacus* L. – then termed ‘crayfish small-molecule protease’ or ‘*Astacus* protease’ – dates back to the late 1960s (Sonneborn et al., 1969). Protein sequencing by Zwilling and co-workers in the 1980s did not reveal homology to any other protein (Titani et al., 1987). Shortly after, the enzyme was identified as a zinc metallopeptidase (Stöcker et al., 1988), and other family members emerged. The first of these was bone morphogenetic protein 1 (BMP1), a protease co-purified with TGF $\beta$ -like growth factors termed bone morphogenetic proteins due to their capacity to induce ectopic bone formation in mice (Wozney et al., 1988). Later, prompted by the discovery of vertebrate meprins, the term ‘astacins’ was coined in 1991 to refer to a family of extracellular zinc endopeptidases encompassing them all (Dumermuth et al., 1991; Stöcker et al., 1991b). In the following years, cloning and sequence analysis of a variety of other astacins was reported, and most of them showed a multi-domain structure. They were shown to be involved in developmental processes, tissue differentiation, and embryonic hatching, as exemplified by UVS.2 from claw frog (Sato and Sargent, 1990), tolloid from fruit fly (Shimell et al., 1991), the low (LCE) and high (HCE) choriolytic enzymes from medaka fish (Yasumasu et al., 1992), and SPAN and blastula protein BP10 from sea urchin (Lepage et al., 1992; Reynolds et al., 1992).

## Physiological background

In the human and mouse genomes, there are six genes encoding astacin proteases, namely, *bmp1*, *tll1*, *tll2*, *mep1a*, *mep1b*, and *astl* (see <http://degradome.uniovi.es/met.html>). The first three code for the tolloid subgroup, which includes protein BMP1 and its major splice variant, mammalian tolloid. These two are also known as procollagen C-proteases and are important for extracellular matrix assembly (Kessler et al., 1996; Li et al., 1996). Closely related to them are mammalian tolloid-like proteins 1 and 2 (TLL1 and TLL2), whose genes are differentially expressed when compared with that of BMP1. Knock-out mice for *bmp1*, *tll1*, and *tll2* have severe



defects in connective tissue assembly and heart and skeletal muscle development (for reviews, see Ge and Greenspan, 2006; Hopkins et al., 2007). These enzymes cleave precursors of fibrillar procollagens for proper matrix assembly. They also process other matrix proteins including proteoglycans, laminins, and anchoring fibrils. In addition, tolloids also cleave growth factors and their antagonists, which are crucial for dorso-ventral patterning during gastrulation in the embryo (Shimell et al., 1991; Holley et al., 1996; Ge and Greenspan, 2006).

Genes *mep1a* and *mep1b* encode the multi-domain proteins meprin  $\alpha$  and meprin  $\beta$ , respectively. These are translated as membrane-bound proteins containing C-terminal MAM domains (meprin, A5 protein, and receptor protein tyrosine phosphatase  $\mu$ ) (Beckmann and Bork, 1993), TRAF domains (tumor necrosis factor receptor-assoiated factor) (Rothe et al., 1994; Zapata et al., 2001), and EGF-like, transmembrane, and cytosolic domains (Figure 1A). The  $\alpha$  subunit is post-translationally cleaved within a unique I (inserted) domain (Figure 1A) during the passage through ER and Golgi, and therefore is found as high-molecular-weight soluble multimers. By contrast, meprin  $\beta$  homodimers and  $\alpha/\beta$  heterodimers remain cell-surface-bound unless shed proteolytically (Hahn et al., 2003). Meprins are involved in tissue differentiation and pericellular signaling. In this context, a variety of meprin substrates have been reported *in vitro*, including biologically active peptides such as gastrin and cholecystokinin, substance P, cytokines, and chemokines (reviewed by Sterchi et al., 2008). Of special interest is the fact that meprins cleave components of the extracellular matrix, in particular the basal lamina but also adhesion proteins at the cell-cell interface (Sterchi et al., 2008; Ambort et al., 2010; Vazeille et al., 2011). Recent proteomics approaches have identified previously known and new physiologically relevant *in vivo* substrates such as vascular endothelial growth factor (Schütte et al., 2010), amyloid precursor protein (Jefferson et al., 2011), procollagens I and III (Kronenberg et al., 2010), interleukin-1 $\beta$  (Herzog et al., 2005), interleukin 18 (Banerjee and Bond, 2008), prokallikrein 7 (Ohler et al., 2010), and fibroblast growth factor 19 (Becker-Pauly et al., 2011).

The third subgroup of astacins in vertebrates comprises the so-called hatching enzymes, represented by just one member in mammals termed ovastacin by Carlos López-Otín and colleagues (Quesada et al., 2004). The protein is encoded by the gene *astl* and expressed in the oocyte and in the developing embryo. A recent report (Sachdev et al., 2012) suggested a role in sperm-egg interaction reminiscent of that of non-proteolytic members of the ADAM (a disintegrin and metalloproteinase) family of metalloproteases (Wolfsberg et al., 1993; Takeda, 2009; Takeda et al., 2012), which, like astacins, belong to the metzincin clan (Bode et al., 1993; Stöcker et al., 1995; Gomis-Rüth, 2003, 2009) (see chapter 'Overall structure of mature astacin catalytic domains' below).

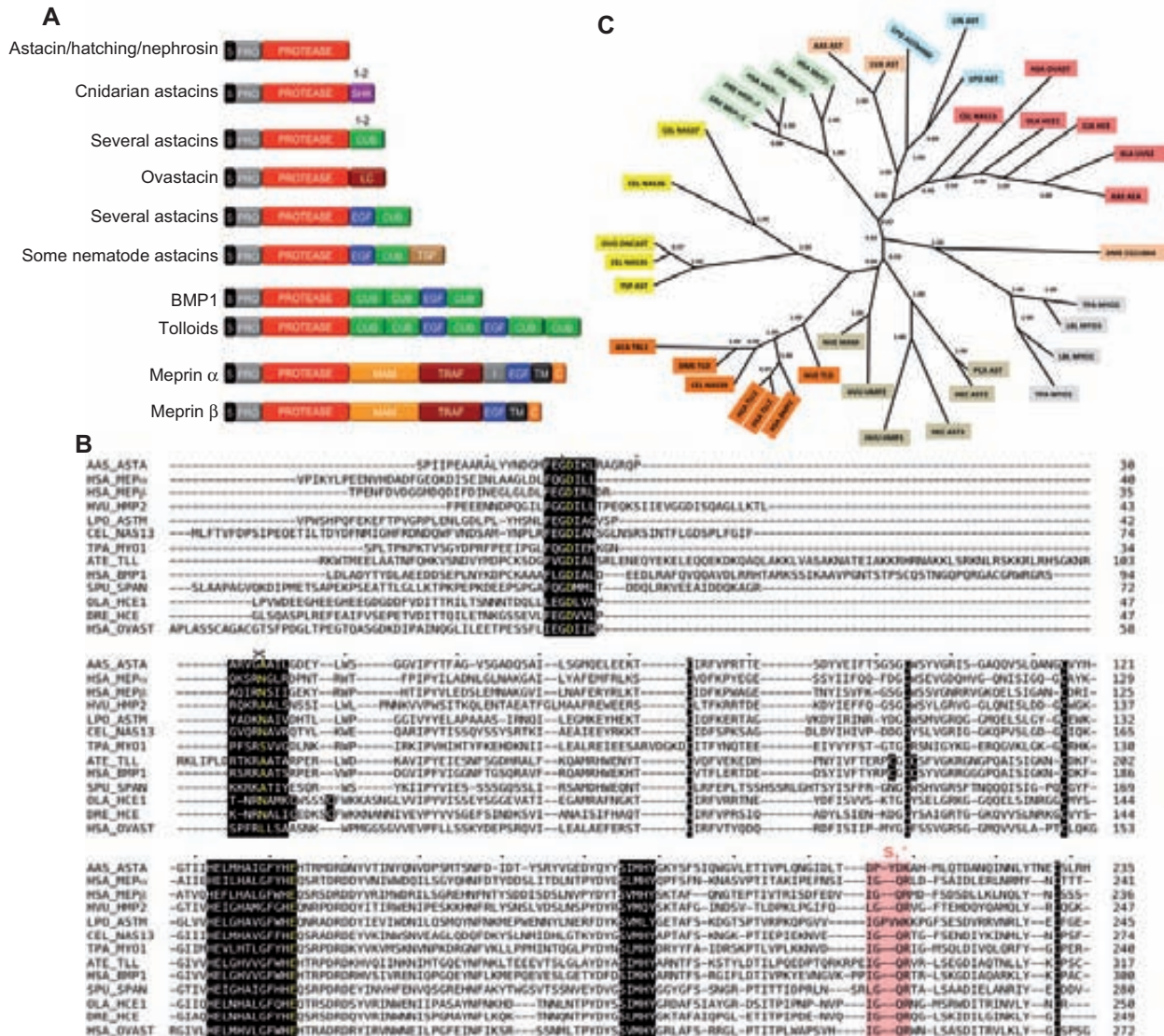
Surprisingly, the genomes of lower vertebrates and invertebrates generally contain more astacin genes than mammalian genomes according to the MEROPS database (<http://merops.sanger.ac.uk>): 7–18 in amphibians and fish, at least 4 in

cnidarians, 13–25 in insects, and up to 40 in nematodes such as *Caenorhabditis elegans* (Möhrle et al., 2003). The large number in the latter example can at least partially be attributed to the parasitic lifestyle of nematodes, which requires an array of proteases to break down host connective tissue. Characterized examples are enzymes from *Trichinella spiralis* (Lun et al., 2003) and *Onchocerca volvulus* (Borchert et al., 2007). Also abundant in lower vertebrates and invertebrates are hatching enzymes, which degrade embryonic envelopes during the free water developmental stage of crustaceans, fish, frogs, and birds. The peptidases of this heterogeneous group often contain C-terminal cysteine-rich and CUB domains (complement C1r/C1s, Uegf, and BMP1) (Bork and Beckmann, 1993), yet some consist of just a single catalytic domain. Examples include crayfish embryonic astacin (Geier and Zwillig, 1998), fish alveolin (Shibata et al., 2000), fish proteins LCE and HCE (Yasumasu et al., 1996), nephrosin from carp head kidney (Hung et al., 1997), protein UVS.2 from frog (Fan and Katagiri, 2001), and protein CAM1 from bird (Elaroussi and DeLuca, 1994).

### Modular organization of astacins and evolutionary aspects

The minimal structure of an astacin protease is a catalytic domain of approximately 200 amino acid residues as found in bacteria. In eukaryotes, this minimal structure is extended by at least an N-terminal pro-peptide which confers latency, so that most astacins, including the prototypical crayfish enzyme (Yiallourous et al., 2002), are secreted as inactive zymogens (Figure 1B). Additional downstream domains include EGF-like modules and/or one or more copies of CUB modules (Bond and Beynon, 1995; Stöcker and Bode, 1995). These domains may be involved in calcium-binding and protein-protein or enzyme-substrate interactions. It has been demonstrated that the C-terminal CUB- and EGF-like domains of procollagen C-peptidase are important for selectivity in substrate recognition (Sieron et al., 2000; Garrigue-Antar et al., 2004; Hintze et al., 2006; Wermter et al., 2007). By virtue of their similar, but not identical, domain composition, sea urchin astacins like SPAN and BP10 are related to tolloids (Lepage et al., 1992; Reynolds et al., 1992). They likewise contain a unique serine/threonine-rich region, which could be the target of O-glycosylation. Interestingly, some *C. elegans* astacins contain thrombospondin type 1 repeats, which are also found in another family of metzincins, the ADAMTS peptidases (Apte, 2009). Further C-terminal domains described include the ShK toxin domain of some coelenterate astacins. Such domains – also called six-cysteine (SXC) domains – were originally identified in metridin, a toxin from sea anemone, and several hypothetical *C. elegans* proteins. Other notable domains are the aforementioned TRAF and MAM domains (see section 'Physiological background'), found in meprins (see above) but also in HMP2 from hydra (Yan et al., 2000a,b) and LAST-MAM from the horseshoe crab (Becker-Pauly et al., 2009). Furthermore, regions of generally low compositional complexity and similarity to other protein modules





**Figure 1** Architecture and evolution of astacins.

(A) Scheme depicting the distinct domains observed in astacins. In addition to N-terminal signal peptides and pro-peptides, most astacins contain further domains C-terminally attached to the catalytic protease domain. These are termed EGF (epidermal growth factor-like; UniProt accession number PF00008); CUB (named after their occurrence in complement component C1r/1s, embryonic sea urchin Uegf, and BMP1; PF00431); TSP (thrombospondin type 1 repeats; PF00090); ShK (ShK toxin domain; PF01549); EB (associated with Kunitz domains and found in several *C. elegans* proteins; PF01683); MAM (meprin, A5 receptor protein, tyrosine phosphatase  $\mu$ ; PF00629); TRAF (found in intracellular signaling proteins; PF00917); bacterial TT domains (PF02957); LC (low complexity domains); C (cytosolic domains); I (inserted domain); and TM (transmembrane anchor). (B) Sequence alignment of the pro-domains and catalytic domains of selected astacins. Over black background: the Met-turn, the zinc binding motif, cysteines, the aspartate-switch residue in the pro-peptide, and the activation site (scissors); in pink: residues chiefly shaping the  $S_1'$  sub-site. (C) Family tree based on the catalytic domains of astacins. The numbers indicate the relative probability of branching. UniProt database accession numbers: AAS AEA (O44072); AAS AST (P07584); ACA TBL1 (P91972); ATE\_TLL (Q75UQ6); CEL\_NAS35 (P98060); CEL\_NAS36 (Q18206); CEL\_NAS37 (Q93243); CEL\_NAS39 (Q20176); CJA CAM1 (P42662); DME TLD (P25723); DME\_AST (CG11864); DME\_CG11864 (Q9VJN9); DRE\_AST (=ZHE1) (Q75NR9); DRE MEP $\alpha$ 1 (Q5RHM1); DRE MEP $\alpha$ 2 (Q5RHM2); DRE MEP $\beta$  (Q08CC4); HEC AST2 (Q2MCX8); HEC AST3 (Q2MCX7); HSA BMP1 (P13497); HSA MEP $\alpha$  (Q16819); HSA MEP $\beta$  (Q16820); HSA OVAST (Q6HA08); HSA TLL1 (O43897); HSA TLL2 (Q9Y6L7); HVU HMP1 (Q25174); HVU HMP2 (Q9XZG0); LBL MYOI (Q8IU47); LBL MYOIII (Q8IU44); LIN AST (A0FKN6); LPO LAST (B4F319); LPO LAST\_MAM (B4F320); NVE MAM (A7SJ13); NVE TLD (Q27W05); OLA HCE1 (P31580); OVO ONCHAST (Q2YFS7); PCA PMP1 (O62558); LVA AST (Q20AS7); SPU\_SPAN (P98068); TPA MYOI (Q8IU46); TPA MYOII (Q8IU45); TSP MP (Q8T5Z5); and XLA UVS2 (P42664).

have been discovered and termed LC domains. Such regions have been observed in *C. elegans* astacins and sea urchin astacins SPAN and BP10 (Lepage et al., 1992; Reynolds et

al., 1992). Moreover, mouse and human ovastacin contain a distinct C-terminal domain of approximately 150 residues with little similarity to other reported proteins (Figure 1A);

this domain might be heavily *O*-glycosylated (Quesada et al., 2004). Several other domains can be inferred from the more than 1000 astacin entries in the MEROPS database but they are not dealt with here as they have not been characterized at the protein level. Examples are the EB module found in some *C. elegans* proteins and the TT domain of bacterial astacins, which has been named after the viral ORF2 of the TT virus. The interested reader is referred to <http://merops.sanger.ac.uk> (Rawlings et al., 2010).

Alignment of representative pro- and catalytic domains of astacin peptidases reveals characteristic structural features which are associated with conserved functions (Figure 1B). There is the typical ‘aspartate-switch’ region in the pro-peptide (see chapter ‘Zymogen structure and activation mechanism’), the zinc-binding consensus sequence, the ‘Met-turn’ (both discussed in the following section), and the unique  $S_1'$  sub-site, responsible for cleavage specificity (see chapter ‘Active-site cleft, substrate specificity, and zinc-binding site’). A phylogenetic analysis based merely on the catalytic domains, and thus omitting pro-peptide regions and C-terminal domains, shows the interrelationships of astacin proteases (Figure 1C). There are several clusters of astacins. Some, like the tolloids, are present throughout the animal kingdom. Others, such as meprins, which have only been observed in vertebrates so far, are confined to distinct taxonomic groups. By contrast, hatching enzymes have diverged into many paralog lineages, especially in amphibians and fish. Other specialized subgroups of astacins seem to exist in nematodes, cnidarians, insects, and molluscs.

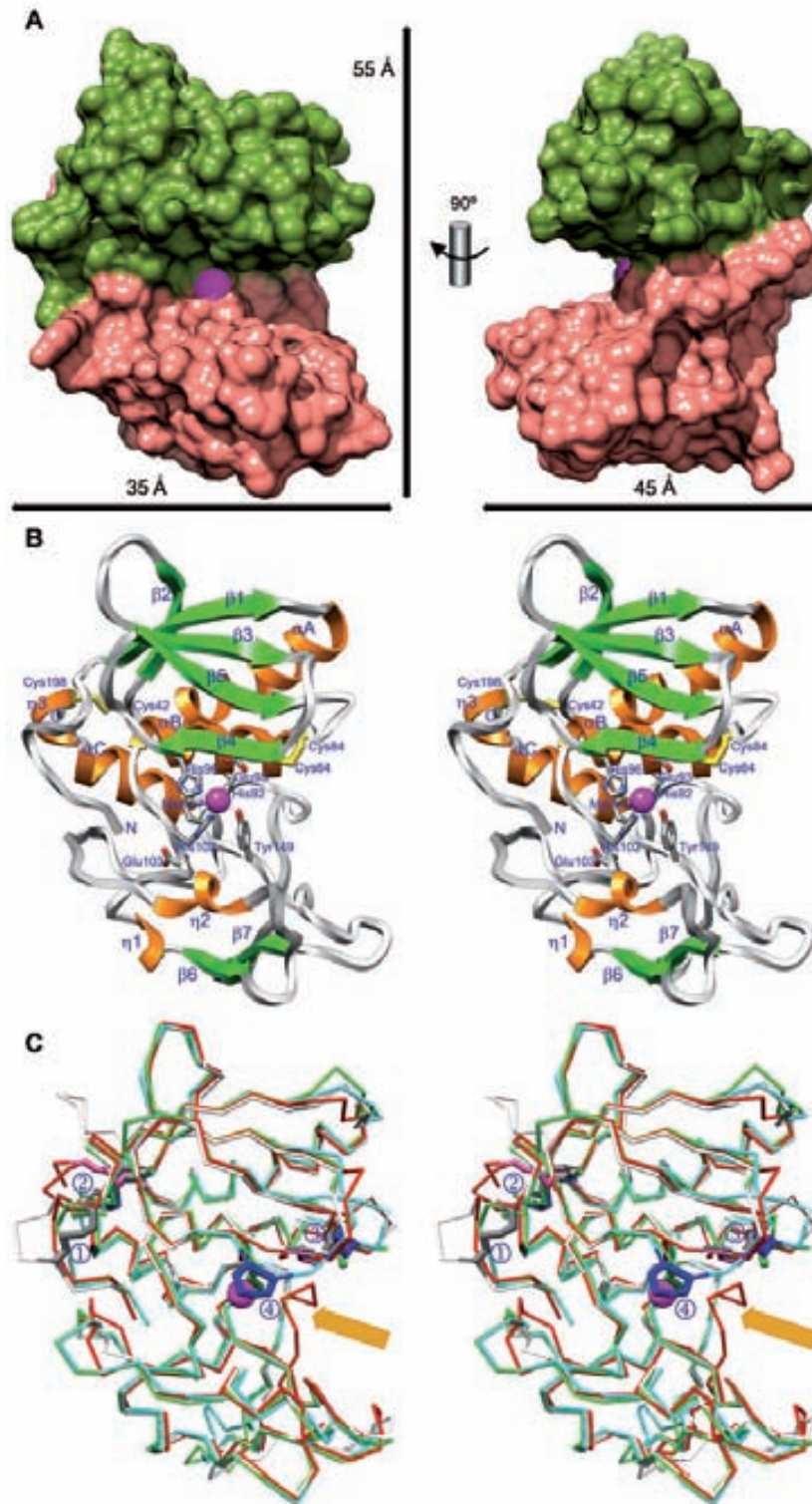
## Overall structure of mature astacin catalytic domains

The first structure solved of a family member was that of crayfish astacin (Bode et al., 1992; Gomis-Rüth et al., 1993; Stöcker et al., 1993). It was the first metalloendopeptidase prototype to be structurally analyzed after thermolysin from *Bacillus thermoproteolyticus* (Matthews et al., 1972) and two closely related thermolysin-family members (Paupitit et al., 1988; Thayer et al., 1991). In contrast to other metallopeptidase (MP) groups such as matrix metalloproteinases (Tallant et al., 2010b), funnelin and cowrin metallocarboxypeptidases (Gomis-Rüth, 2008), the aforementioned thermolysins, and adamalysins/ADAMs (Takeda et al., 2012), few structures of astacins have been reported. To date, only human BMP1, human TLL1, and hatching enzyme 1 from the zebrafish *Danio rerio* (ZHE1) have been described in addition to the crayfish enzyme (Mac Sweeney et al., 2008; Okada et al., 2010). Furthermore, the structure of the astacin zymogen has also been published recently (Guevara et al., 2010). Overall, astacin catalytic domains (CDs) show a compact ellipsoidal shape, reminiscent of a kidney or a Pac-Man, with maximal dimensions of approximately 55×45×35 Å (Figure 2A). A deep and narrow active-site cleft divides the CDs into two sub-domains of approximately 100 residues when viewed in standard orientation (Gomis-Rüth et al., 2012), an upper N-terminal (NTS) and a lower C-terminal sub-domain (CTS;

Figure 2A–C). Superposition of BMP1, ZHE1, and TLL1 onto astacin reveals very similar chain traces, and this results in low overall *rmsd* values between these structures and astacin: 1.3 Å for BMP1 (for 174–180 equivalent C $\alpha$ -atoms deviating <3 Å), 1.1 Å for TLL1 (175 equivalent C $\alpha$ -atoms), and 1.0 Å for ZHE1 (178 equivalent C $\alpha$ -atoms; Figure 2C). Accordingly, the main structural features, as well as residue numbering, will hereafter refer to astacin (see Bode et al., 1992; Gomis-Rüth et al., 1993) unless otherwise stated.

The NTS consists of a strongly twisted five-stranded  $\beta$ -sheet ( $\beta_1$ – $\beta_5$ ; connectivity  $-1\times, +2\times, +2, -1$ ), whose strands parallel the active-site cleft except for its lowermost strand ( $\beta_4$ ) which creates the ‘upper-rim’ of the active-site cleft (see also the section ‘Active-site cleft, substrate specificity, and zinc-binding site’) and runs antiparallel (Figure 2B). The sheet is flanked on its top convex side by a long characteristic loop, which connects strands  $\beta_2$  and  $\beta_3$  (L $\beta_2\beta_3$ ), and by two helices on its bottom concave side, the ‘backing helix’ ( $\alpha_A$ ) and the ‘active-site helix’ ( $\alpha_B$ ), which run nearly parallel to the strands of the sheet. Helix  $\alpha_B$  includes the first three residues of the long zinc-binding consensus sequence H<sup>92</sup>E<sup>93</sup>X<sup>94</sup>H<sup>95</sup>X<sup>96</sup>G<sup>97</sup>X<sup>98</sup>H<sup>99</sup><sup>102</sup> (amino-acid one-letter code; X stands for any residue), which is characteristic of astacins but also metzincins in general (Bode et al., 1993; Stöcker et al., 1993, 1995; Gomis-Rüth, 2003, 2009). G<sup>99</sup> within the consensus sequence is the endpoint of both helix  $\alpha_2$  and NTS. This glycine allows for a sharp turn in the trajectory of the polypeptide in order to enter the CTS, and the values of its main-chain angles in the different structures ( $\Phi=122^\circ$ – $140^\circ$ ;  $\Psi=10^\circ$ – $23^\circ$ ) indicate that any other amino acid would be in a high-energy conformation (Davis et al., 2007). The CTS contains the third zinc-binding residue H<sup>102</sup>, which is followed by the ‘family-specific’ residue of astacins (E<sup>103</sup>) (Stöcker et al., 1993) (see also section ‘Active-site cleft, substrate specificity, and zinc-binding site’). The rest of the CTS is characterized by few regular secondary structure elements, and only three short  $3_{10}$ -helices ( $\eta_1$ – $\eta_3$ ) and two short strands ( $\beta_6$  and  $\beta_7$ ) are found in addition to the major ‘C-terminal helix’ ( $\alpha_C$ ; Figure 2B). Special mention should be given to a tight 1,4-turn situated below the catalytic zinc-site, the Met-turn, which is characterized by a strictly conserved methionine (M<sup>147</sup>), both in sequence and side-chain conformation, within astacins and also all other metzincins structurally analyzed to date (Gomis-Rüth, 2009; Goulas et al., 2010; Waltersperger et al., 2010). It has been proposed that the Met-turn acts as a plug that inserts laterally into a core structure created by the protein segment engaged in zinc binding, thus contributing to the structural integrity that is indispensable for function, but there is still debate on its significance in metzincins (Pieper et al., 1997; Boldt et al., 2001; Hege and Baumann, 2001; Butler et al., 2004; Walasek and Honek, 2005; Pérez et al., 2007; Oberholzer et al., 2009; Tallant et al., 2010a). Finally, a tyrosine two positions downstream of the methionine (Y<sup>149</sup>) is also engaged in zinc binding and catalysis (see also the chapter ‘Active-site cleft, substrate specificity, and zinc-binding site’).





**Figure 2** Overall structure of astacin catalytic domains.

(A) Mature astacin catalytic domain displayed with its Connolly surface in standard orientation according to Gomis-Rüth et al. (2012) (left) and after a vertical rotation of  $90^\circ$  (right). The NTS is shown in green, the CTS in salmon, and the zinc ion in magenta. (B) Richardson plot in stereo of astacin in standard orientation depicting its regular secondary structure elements as green arrows ( $\beta$ -strands  $\beta 1$ – $\beta 7$ ) and orange ribbons ( $\alpha$ -helices  $\alpha A$ – $\alpha C$  and  $3_{10}$ -helices  $\eta 1$ – $\eta 3$ ). The two disulfide bonds are shown as yellow sticks and labeled, as are the catalytic zinc in magenta and its protein ligands, the termini, the Met-turn methionine, and the family-specific glutamate ( $E^{103}$ ). (C) Superposition in stereo of the  $C\alpha$ -traces of astacin (red; pink disulfide bonds), BMP1 (cyan; dark blue disulfide bonds), TLL1 (green; dark green disulfide bonds), and ZHE1 (white; gray disulfide bonds). The four possible disulfide site locations are indicated by ①–④. The 170-loop is marked with an orange arrow.

## Disulfide bonds

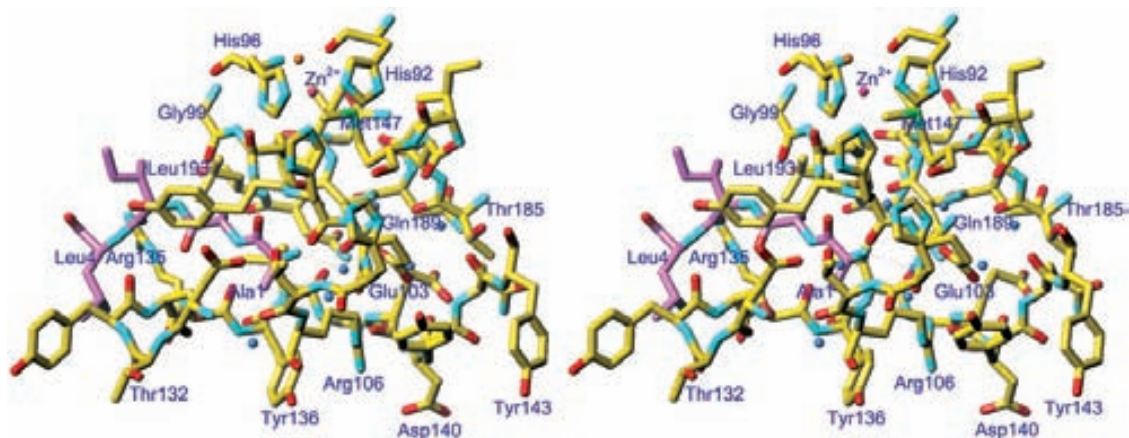
The CD of astacins is cross-linked by two or three disulfide bonds. Roughly common to all structures are the two found between C<sup>42</sup> and C<sup>198</sup> and between C<sup>64</sup> and C<sup>84</sup> in the crayfish prototype (Figure 1B and 2C; ② and ③). The former links the C-terminal segment of the CD to the NTS at the loop connecting helix  $\alpha$ A with strand  $\beta$ 2; the latter links L $\beta$ 3 $\beta$ 4 and the beginning of  $\beta$ 4 to L $\beta$ 5 $\alpha$ B and thus contributes to shaping the active-site cleft at its primed site and to substrate binding (see also chapter ‘Active-site cleft, substrate specificity, and zinc-binding site’). Sequence alignment with structurally non-characterized astacin family members (see Figure 1B and Stöcker et al., 1993) indicates that these two bonds are likely to be conserved among all astacins. In addition, selected members may show additional SS-bridges. ZHE1, as a member of the hatching subgroup, shows a unique cross-link between two cysteine residues in the N-terminal segment of the CD [C<sup>5</sup>–C<sup>10</sup> according to ZHE1 numbering; see Protein Data Bank (PDB) access code 3LQB; Okada et al., 2010; Figure 2C; ①]. This may be required to fix the very N-terminus in a competent position (see next section). BMP1 and TLL1, in turn, show a slight displacement of the disulfide bond at position ③ in Figure 2C and a further unique bond between two consecutive residues of the upper-rim strand  $\beta$ 4 enabled by a *cis*-peptide bond between them (C<sup>64</sup>–C<sup>65</sup> according to BMP1 numbering; PDB 3EDH; Figure 2C; ④). This segment, termed ‘cysteine-rich loop,’ is unique for the tollid subfamily (Figure 1C) and has implications for substrate binding (see section ‘Active-site cleft, substrate specificity, and zinc-binding site’).

## A buried N-terminus in mature astacins

In general, proteins that undergo proteolytic maturation and possess additional flanking domains have their chain termini

located on the molecular surface. While this is the case for the C-terminus of the CDs of mature astacins, the N-terminus is buried within the molecule for its first three or four residues (Figure 2B and C). In astacin, the first three residues A<sup>1</sup>–A<sup>2</sup>–I<sup>3</sup> are inserted like a plug in an internal cavity framed by segments D<sup>131</sup>–Q<sup>142</sup>, F<sup>100</sup>–M<sup>107</sup>, and T<sup>185</sup>–Q<sup>189</sup>. Residues from these segments, together with eight solvent molecules and the aforementioned N-terminal residues, are involved in an intricate, completely buried hydrogen-bonding network that is key for structural integrity of the enzyme (see Figure 3 of Bode et al., 1992) and incompatible with N-terminally elongated polypeptide chains. Most importantly, the  $\alpha$ -amino group of A<sup>1</sup> establishes a solvent-mediated salt bridge with E<sup>103</sup>, which is the ‘family-specific’ residue found immediately after the third zinc-binding residue (H<sup>102</sup>; Figure 3). The solvent molecule further hydrogen-bonds the side chain of Q<sup>189</sup>. These interactions contribute to structure and stability in astacin CDs. Mutants of pro-astacin, in which E<sup>103</sup> had been replaced with glutamine and alanine, displayed unaltered catalytic efficiency but much lower thermal stability (Yiallourou et al., 2002). This structural rather than functional importance is supported by a superposable position and conformation of E<sup>103</sup> in the zymogen and the mature structures (see also section ‘Zymogen structure and activation mechanism’).

By contrast, in both BMP1 and TLL1 structures, the first residue is an N-acetylated alanine. This means that the bridging solvent molecule is replaced by the carbonyl oxygen of the acetyl group, which is within hydrogen-binding distance of both Q<sup>189</sup> and E<sup>103</sup> (same numbering as in astacin; Mac Sweeney et al., 2008). In ZHE1, in turn, the structure of the mature enzyme starts at the position equivalent to A<sup>2</sup> of astacin, likewise with an alanine, so that its  $\alpha$ -amino group is too far away from the E<sup>103</sup>-equivalent to establish an interaction, and the empty space is occupied by six solvent molecules (Okada et al., 2010). However, the latter glutamate is maintained in a very similar side-chain conformation to that in astacin, TLL1, and BMP1, although in this case, an interaction takes places



**Figure 3** The mature N-terminus.

Stereographic picture centered on the buried N-terminus of the structure of astacin as a stick model colored according to atom types. For the N-terminal tail, carbons are shown in pink, otherwise in yellow. The zinc ion is depicted as a magenta sphere and labeled, and the eight solvent molecules of the internal cavity are shown as blue spheres. Selected residues are labeled for reference.



with an arginine at position 189, which replaces Q<sup>189</sup> in the latter three proteins. This scenario also explains why ZHE1 possesses a unique disulfide bond that bridges the cysteine residues at positions 5 and 10 (see chapter ‘Disulfide bonds’ above), as this maintains the plugging function of the N-terminal segment despite the lack of the pivotal salt bridge with the family-specific residue. Inspection of the mature N-termini of representative astacin family members (Figure 1B and Figure 7 in Stöcker et al., 1993) shows that most structures should be compatible with the termini of either astacin/BMP1/TLL1 or ZHE1. N-terminal residues are almost exclusively alanine or asparagine, which is also compatible with the aforementioned structural features of a direct or water-mediated interaction with the family-specific residue. Overall, the presence of a buried N-terminus is unique to astacins within MPs and does not allow for variation in the length of the N-terminal segment which is essential for the maintenance of this specific structural feature. The latter finding is reminiscent of the internal salt bridge observed between the mature N-terminus of trypsin-like serine proteinases and an aspartate immediately adjacent to the catalytic serine residue (Fehlhammer et al., 1977; Huber and Bode, 1986).

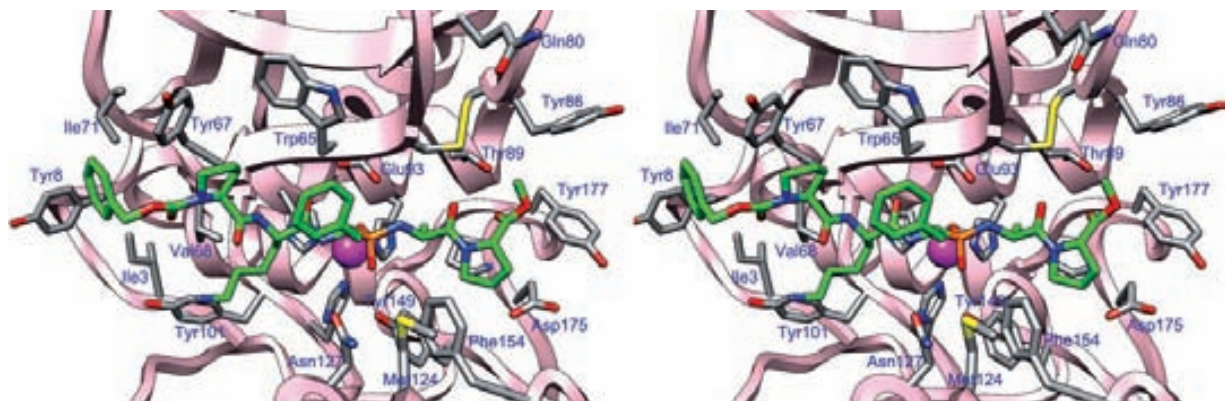
### Active-site cleft, substrate specificity, and zinc-binding site

In MPs, substrate hydrolysis proceeds via a high-energy reaction intermediate centered on a tetrahedral carbon bound to two *gem*-diolate oxygens, the scissile-bond nitrogen, and the preceding C $\alpha$  atom; the tetrahedral carbon itself is derived from the original scissile-bond carbonyl group by the nucleophilic attack of a solvent molecule polarized by a general base/acid glutamate (Matthews, 1988; Bayés et al., 2007; Gomis-Rüth, 2008). A complex of astacin with a reaction-intermediate analogue of peptide Phe-Pro-Lys-Phe- $\Psi$ -Ala-Pro, in which the tetrahedral carbon and the downstream nitrogen flanking the scissile bond are replaced by a phosphinic group (Figure 4), allows us to delimit the deep and extended active-site cleft,

which horizontally traverses the catalytic domains of astacins. It explains why only elongated substrates – ideally longer than seven residues – are efficiently cleaved (Stöcker et al., 1990). In addition, comparison of this and other complexed forms of astacins with unbound enzymes further shows that astacins may undergo a slight overall hinge motion upon substrate, inhibitor, or ligand binding, which brings the CTS and the NTS closer by approximately 1 Å (Grams et al., 1996), in a fashion similar to that described for thermolysins (Holland et al., 1992).

In astacin and ZHE11, the upper-rim strand  $\beta$ 4 frames the top of the cleft on its primed side, together with L $\beta$ 5 $\alpha$ B and the disulfide bond at site  $\textcircled{3}$  in Figure 2C. By contrast, in BMP1 and TLL1, this disulfide bond is slightly displaced and a further, unique SS-bridge is found between two consecutive cysteines within a cysteine-rich loop that replaces the upper-rim strand of astacin and ZHE1 (see section ‘Disulfide bonds’ above). This gives rise to an eight-membered, largely hydrophobic ring above the S<sub>1</sub> pocket (Figure 2C), which prevents substrate binding to the cleft and causes the upper rim to no longer be a  $\beta$ -strand. This cysteine-rich loop is disordered in the unbound structures, and it has been proposed to act as a mobile flap that clamps substrates into a competent position for a Michaelis complex (Mac Sweeney et al., 2008). On its non-primed side, the cleft is delimited in astacins by the end of strand  $\beta$ 4 and the subsequent L $\beta$ 4 $\beta$ 5, as well as the N-terminal segment at E<sup>7</sup>-Y<sup>8</sup>. At its bottom, the cleft is constrained on its non-primed side by I<sup>4</sup>-G<sup>5</sup> and the loop after  $\alpha$ B, and, on its primed side, by the Met-turn and the subsequent segment up to W<sup>158</sup>, and, importantly, the ‘170 loop’ mainly at T<sup>174</sup>-D<sup>178</sup>.

A substrate binds to astacins in an extended conformation and is anchored to the cleft in an antiparallel manner by the upper-rim strand  $\beta$ 4 through inter-main-chain interactions, on both the primed and non-primed sides (Figure 4). The reaction-intermediate complex further reveals that cleft sub-site S<sub>4</sub> is framed in astacin by I<sup>71</sup>, Y<sup>8</sup>, I<sup>3</sup>, V<sup>68</sup>, and Y<sup>67</sup>. In contrast, substrate residues in P<sub>3</sub> and P<sub>1</sub> protrude from the cleft toward the bulk solvent and lie up against the side chains of the upper-rim-strand residues W<sup>65</sup> and Y<sup>67</sup>. Sub-site S<sub>2</sub> is shaped by H<sup>96</sup>, H<sup>102</sup>, V<sup>68</sup>, and Y<sup>101</sup>. On the opposite side of the cleft, S<sub>2</sub>' is created by



**Figure 4** Substrate binding and enzymatic reaction.

Detail of the active-site cleft of astacin in stereo (pink ribbons with stick models of selected side chains colored according to atom types; carbons in gray) with the bound reaction-intermediate analog BOC-Pro-Lys-Phe $\Psi$ (PO<sub>2</sub>)-CH<sub>2</sub>-C(CH<sub>3</sub>)CO-Pro-OCH<sub>3</sub> (PDB 1QJI; Grams et al., 1996) likewise colored according to atom types (carbons in green, phosphorous in orange).

Y<sup>149</sup>, D<sup>175</sup>-P<sup>176</sup>, and S<sup>153</sup>-F<sup>154</sup>; and additional downstream sub-sites are likely to be conditioned by D<sup>175</sup>, Y<sup>177</sup>, and W<sup>158</sup>. As is usual for MPs, the most relevant sub-site for specificity is S<sub>1</sub>' , which is delimited in astacin by H<sup>92</sup> and the first helical turn of helix  $\alpha$ B, in particular through the side chain of T<sup>89</sup>. The most important structural element for primed sub-sites in astacins is the 170-loop, which shows disparate chain traces in ZHE1, TLL1, and BMP1, on the one hand, and astacin, on the other (see Figures 2C, 4 and Figure 7 in Okada et al., 2010). This leads to a conserved arginine (R<sup>182</sup> in ZHE1 and R<sup>176</sup> in BMP1/TLL1) shaping the bottom of a deep S<sub>1</sub>' pocket in ZHE1, TLL1, and BMP1, while the rather shallow pocket in astacin is shaped by P<sup>176</sup>. This explains why most family members, including ZHE1, BMP1, TLL1, meprins  $\alpha$  and  $\beta$ , and horseshoe-crab LAST-MAM, prefer aspartate residues in P<sub>1</sub>' while astacin prefers small aliphatic residues (Stöcker et al., 1990; Okada et al., 2010; Becker-Pauly et al., 2011). Accordingly, astacins, with the notable exception of their founding prototype and a minority of other family members, are the only eukaryote proteases to cleave N-terminal to an aspartate in proteins (Becker-Pauly et al., 2011). This preference is seen even more strongly in enzymes like meprin  $\beta$  and tolloids (e.g., BMP1), in which the S<sub>2</sub>' sub-site (see above) also contributes to the binding of acidic substrate side chains with basic residues.

The catalytic zinc ion resides at the bottom of the active-site cleft (Figure 2A and B). The active-site helix  $\alpha$ B provides two histidine ligands of the metal (H<sup>92</sup> and H<sup>96</sup>), separated by a single helical turn that allows a concerted approach to the metal together with the third zinc-binding histidine, H<sup>102</sup>. In unbound astacin, the N $\epsilon$ 2 atoms of these residues coordinate the metal together with a catalytic solvent molecule, also anchored to the general base/acid E<sup>93</sup>, and Y<sup>149</sup> O $\eta$  (Gomis-Rüth et al., 1993). The overall metal coordination is trigonal-bipyramidal, with H<sup>96</sup> N $\epsilon$ 2 and Y<sup>149</sup> O $\eta$  at apical, somewhat larger distances (2.2–2.5 Å), and the remaining three ligands on a plane with the metal,  $\sim$ 2 Å apart (Figure 12 in Gomis-Rüth et al., 1993). In ligand- or inhibitor-bound astacins, the catalytic solvent is replaced by one or two oxygens, and the Y<sup>149</sup> ligand is pulled away from its metal-binding position. In the complex with the reaction-intermediate analogue, the O $\eta$  atom stabilizes one of the *gem*-diolate oxygen atoms of the tetrahedral carbon (Figure 4). This motion has been termed a 'tyrosine switch' and is also found in serralsins and – most likely – pappalysin family members within the metzincins (Gomis-Rüth, 2003, 2009; Baumann, 2004; Tallant et al., 2006). Finally, studies on apo-astacin and metal-replaced enzyme revealed that the metal site is preformed and that it can accommodate the ligand requirements of distinct divalent metals through minor rearrangement and recruitment of a variable number of solvent molecules. Overall, this gives rise to bipyramidal-trigonal, tetrahedral, or octahedral coordination spheres (Gomis-Rüth et al., 1994).

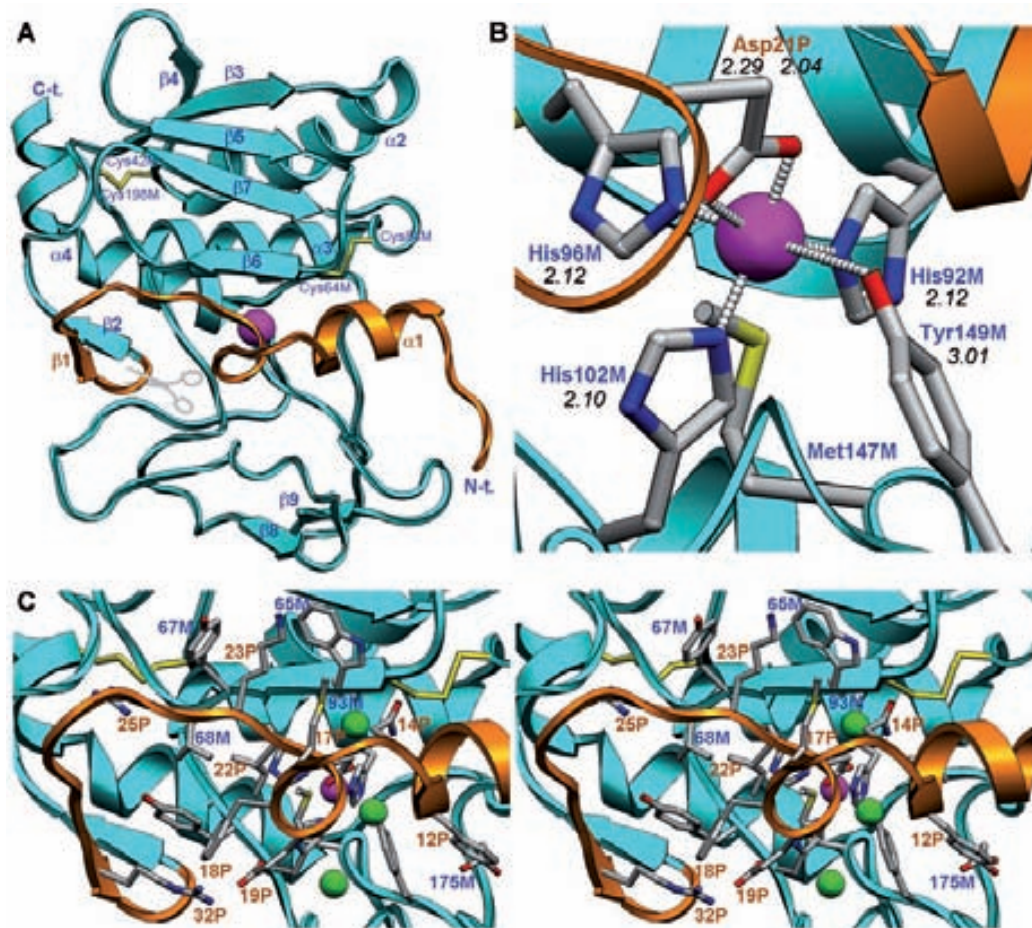
### Zymogen structure and activation mechanism

Virtually all eukaryotic astacins are synthesized as inactive precursors, which prevents the occurrence of temporally and

spatially inappropriate proteolytic activity. As an example, pro-astacin is only transiently found within the ducts between the hepatopancreas – where it is synthesized – and the stomach, where it is activated in order to participate in collagenolysis and gelatinolysis during digestion (Yiallourous et al., 2002). Comparison of the pro-segments of distinct family members (see Figure 1B and Figure 1 in Guevara et al., 2010) revealed that these vary greatly in length (from 34 to 393 residues) and that only a short consensus sequence F<sup>18P</sup>XGD<sup>21P</sup> (residues of the astacin pro-segment carry the suffix 'P') is revealed by sequence alignments.

The only structure of an astacin-family zymogen reported to date is that of the crayfish enzyme (Guevara et al., 2010; PDB 3LQ0; Figure 5A). With merely 34 residues, pro-astacin possesses the shortest pro-segment structurally analyzed for an MP and, in contrast with other peptidases, it is not required for proper folding in the crayfish enzyme, which could be purified from *Escherichia coli* inclusion bodies and correctly folded as a recombinant mature protease, devoid of the pro-peptide (Reyda et al., 1999; Yiallourous et al., 2000, 2002). Of course, this might be different in pro-astacins with longer pro-domains, which in some cases like *Drosophila* tolloid-like are much longer than the catalytic protease domain itself. In pro-astacin, the pro-segment is elongated and structured by means of several intramolecular interactions, and it runs across the front surface of the mature enzyme moiety in the opposite orientation to that of a substrate (Figure 5A–C). This contributes to the prevention of self-cleavage, as found in cysteine-protease and matrix metalloproteinase zymogens (Khan and James, 1998). The N-terminus of the pro-segment at S<sup>1P</sup> is anchored to the mature part through a hydrophilic interaction with the main chain of I<sup>156</sup>, within the segment connecting the Met-turn and the C-terminal helix  $\alpha$ C. The polypeptide runs along the molecular surface toward the active-site cleft and adopts a helical conformation from E<sup>6P</sup> to Y<sup>12P</sup> (helix  $\alpha$ 1 in Figure 5A). This segment nestles in the primed side of the cleft approximately until sub-site S<sub>1</sub>' , which is partially occupied by the side chain of L<sup>11P</sup>, and this contributes, together with a flexible segment within the CTS of the mature moiety of the zymogen (the 'activation-domain'; D<sup>129</sup>-G<sup>138</sup>; see below), to a larger separation between the two sub-domains than in the unbound mature enzyme. At N<sup>14P</sup>, the chain projects toward bulk solvent and enters a wide loop that ends at D<sup>21P</sup> (Figure 5A). This loop comprises two successive 1,4-turns, which together with hydrophobic interactions of M<sup>17P</sup> with both K<sup>23P</sup> and W<sup>65</sup> give rise to a small globular nucleus, which places D<sup>21P</sup> in contact with the catalytic zinc ion (Figure 5B and C). The loop structure is further stabilized by electrostatic interactions of E<sup>19P</sup> with the zinc-binding residue H<sup>102</sup>, T<sup>105</sup>, and, most relevantly, the penultimate residue of the pro-segment (R<sup>32P</sup>; Figure 5C). D<sup>21P</sup> approaches the catalytic metal from the top in a bidentate manner and its O $\delta$ 2 atom substitutes for the zinc-bound solvent molecule in the unbound mature enzyme. Overall, the zinc ion shows a distorted octahedral coordination sphere (Figure 5B), which is unusual for zinc (McCall et al., 2000) and is reminiscent of the structure of the catalytically inert nickel-substituted





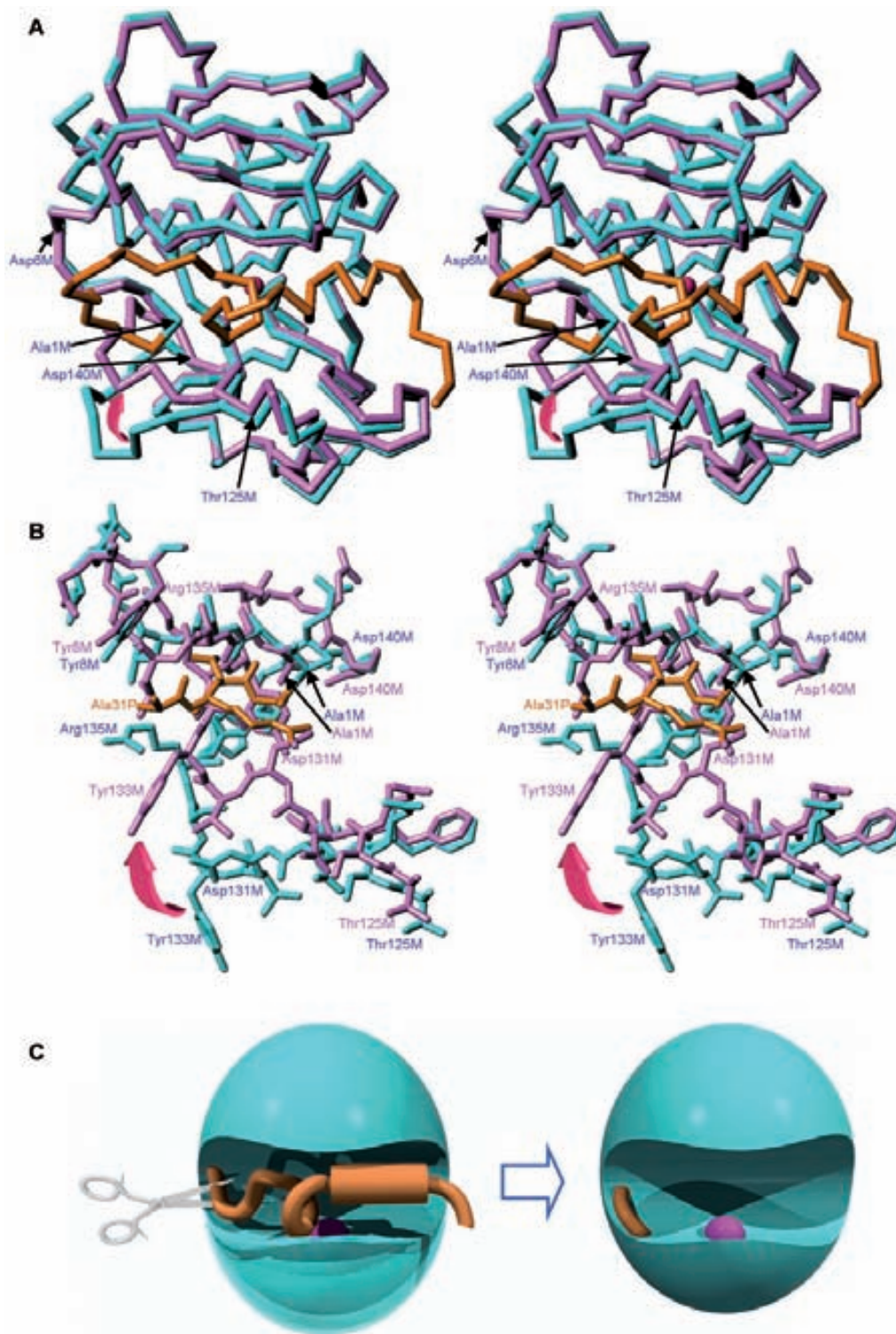
**Figure 5** Structural determinants of zymogenicity.

(A) Richardson-type plot of pro-astacin in standard orientation (Gomis-Rüth et al., 2012) depicting the pro-segment in orange and the mature enzyme moiety in cyan. Repetitive secondary structure elements are shown as ribbons (helices  $\alpha 1$ – $\alpha 4$ ) and arrows (strands  $\beta 1$ – $\beta 9$ ), the catalytic zinc ion as a magenta sphere, and the two disulfide bonds as yellow sticks. The latter are labeled, as are the termini of the molecule. The final activation cleavage site ( $G^{34P}$ - $A^{1M}$ ; mature enzyme residues within the zymogen structure carry suffix 'M,' pro-segment residues carry suffix 'P') is shown by the scissors. For clarity, only one conformation has been displayed for segment  $D^{129M}$ - $P^{135M}$ . (B) Close-up view of (A) depicting the catalytic zinc ion with its six ligands, which are labeled. The respective bonding distances (in Å) are shown below each residue label. The Met-turn methionine is also shown and labeled. (C) Close-up view of (A) in stereo to illustrate the major interactions between the pro-segment and the mature enzyme moiety. Participating residues are labeled (mature enzyme residues in blue, pro-segment residues in brown), except those already labeled in (B). Relevant solvent molecules are displayed as green spheres (reproduced from Guevara et al., 2010 © The American Society for Biochemistry and Molecular Biology).

astacin (Gomis-Rüth et al., 1994). By contrast, the side chain of tyrosine-switch residue  $Y^{149}$  is closer to the competent conformation of the unbound mature structure than to the structure bound to the reaction-intermediate analog, although somewhat further from the metal ion (see Figure 5B and Figure 12 in Gomis-Rüth et al., 1993). The four residues after  $D^{21P}$  run in extended conformation along the cleft between sub-sites  $S_2$  and  $S_5$  and bind the upper-rim strand (strand  $\beta 6$  in the pro-astacin structure; see Figure 5A and C). The polypeptide chain reaches the molecular surface after  $R^{25P}$ , whose side chain is anchored to the mature enzyme backbone, and thereafter, a flexible segment leads to a short  $\beta$ -hairpin structure made up by  $\beta 1$  ( $A^{31P}$ - $V^{33P}$ ) and  $\beta 2$  ( $A^2$ - $L^4$ ) (see Figure 5A and C). The main maturation cleavage point ( $G^{34P}$ - $A^1$ ), which is buried within the molecular structure

of the zymogen like the N-terminus of the mature form (see section 'A buried N-terminus in mature astacins'), is found at the tip of this hairpin.

Activation of pro-astacin entails removal of the inhibiting pro-segment through successive cleavage events, which eventually replace the zinc-binding aspartate with the catalytic solvent molecule following an aspartate-switch mechanism and render the mature N-terminus at  $A^1$  (Yiallourous et al., 2002; Guevara et al., 2010). In the first activation step, successive exogenous cleavages render an intermediate species starting at  $V^{33}$  (according to Yiallourous et al., 2002), which disrupts the aforementioned salt bridge ( $R^{32P}$ - $E^{19P}$ ) (see above), and triggers removal of most of the pro-segment. This causes major rearrangement of the subjacent 'activation domain,' which adopts the rigid and competent conformation



**Figure 6** Structural transitions upon activation.

(A) Superimposition in stereo of the C $\alpha$ -traces of pro-astacin (pro-segment in orange, catalytic moiety in cyan) and mature astacin (purple) in standard orientation (Gomis-Rüth et al., 2012). The catalytic zinc ion of pro-astacin is shown as a magenta sphere for reference. Segments involved in activation (magenta arrow) are indicated. Only one conformation is shown for segment D<sup>129M</sup>-P<sup>135M</sup> (mature enzyme residues within the zymogen structure carry suffix 'M'; pro-segment residues carry suffix 'P'). (B) Close-up view of (A) in stereo showing only the first residues of the mature moiety and the activation domain of both the pro-enzyme (cyan sticks; blue labels) and the mature enzyme (purple sticks and labels), as well as the four last pro-enzyme residues (orange sticks and labels). (C) Schematic representation illustrating the transition between the zymogen, with a flexible activation domain in the lower sub-domain of the molecule, and the rigid mature enzyme (reproduced from Guevara et al., 2010 © The American Society for Biochemistry and Molecular Biology).



found in the mature enzyme (Figure 6A–C). This is analogous to serine proteinases, in which the activation domain likewise becomes rigid and functional upon activation (Bode and Huber, 1978; Khan and James, 1998). In addition, the initial multiple exogenous cleavages are reminiscent of matrix metalloproteinase activation during which trimming cuts by other proteinases in a so-called ‘bait region’ are observed (Nagase, 1997). This contrasts with funnelins and trypsin-like serine proteinases, in which the first cut during activation generates the mature N-terminus (Avilés et al., 1993; Khan and James, 1998). Subsequently, the pre-mature astacin variants, which are catalytically active (Yiallourous et al., 2002), carry out further autolytic cleavages, eventually giving rise to the competent N-terminus at A<sup>1</sup> (see chapter ‘A buried N-terminus in mature astacins’). In matrix metalloproteinases, similar trimming yields the competent N-terminus needed to form a salt bridge with a conserved aspartate (Reinemer et al., 1994; Nagase, 1997). Finally, comparison of zymogenic and mature astacin further reveals that upon cleavage at G<sup>34P</sup>-A<sup>1</sup>, the main chain must undergo a 180° rotation around the Ψ main-chain angle of the new N-terminal residue to allow segment A<sup>1</sup>-D<sup>6</sup> to insert into the molecular moiety and fulfill the plugging function described in the chapter ‘A buried N-terminus in mature astacins’ above.

## Protein inhibitors

Astacins are not inhibited by tissue inhibitors of metalloproteinases (TIMPs). A natural inhibitor of astacin is the general protein scavenger α<sub>2</sub>-macroglobulin (Stöcker et al., 1991a; Meier et al., 1994; Zhang et al., 2006; Marrero et al., 2012). However, α<sub>2</sub>-macroglobulin will only entrap endopeptidases of limited size. Larger oligomeric astacins, such as the meprins, are not inhibited by this regulator of vascular and interstitial proteolysis. Interestingly, there is a fish astacin from carp head kidney (a hematopoietic organ) which circulates in the blood stream in complex with a specific protein inhibitor. Originally, this enzyme was termed ‘nephrosin’ (Hung et al., 1997) and its inhibitor ‘nephrosin inhibitor’ (Tsai et al., 2004). The nephrosin inhibitor is a homolog of fetuin, a large plasma protein with many functions. Fish fetuin, like its mammalian counterpart fetuin A, contains cystatin-like domains and is related to cystatin C-like inhibitors of cysteine cathepsins. More recently, it has been demonstrated that the plasma proteins fetuin A and cystatin C act as physiological inhibitors of human astacin proteases such as meprins and also block crayfish astacin (Hedrich et al., 2010).

## Conclusions

Since the first report on astacins back in 1991 (Dumermuth et al., 1991), the family has expanded to several hundreds of animal and bacterial – though not plant or archeal – sequences (Sterchi et al., 2008). Six are found in humans and up to 40 in *C. elegans* (Möhrlen et al., 2003), and putative uncharacterized proteins from *Fusarium oxysporum* (UniProt entries

F9FJL4, F9FJL6, and F9FF60), *Phaeosphaeria nodorum* (Q0U0C2 and Q0UTK3), *Pyrenophora teres* (E3RUK5), *Pyrenophora tritici-repentis* (B2WI05), and *Unicarpus reesii* (C4JMI3) may represent the first fungal members. Through their degradatory potential, astacins are involved in embryonic development, tissue differentiation, and extracellular matrix assembly, and they are thus therapeutic targets (Mac Sweeney et al., 2008; Sterchi et al., 2008; Okada et al., 2010). Physiologically, they are regulated by zymogenicity and co-localizing protein inhibitors; however, to be able to act upon them from a therapeutic perspective, it will be necessary to design highly specific small-molecule inhibitors that target the catalytic moieties. This is where detailed knowledge of their three-dimensional structure, tentatively presented in this review, may prove helpful.

## Acknowledgments

This study is dedicated to Robert Huber and Wolfram Bode on the occasion of their 75th and 70th birthdays, respectively. Financial support was provided by grants from European, Spanish, German, and Catalan agencies [FP7-HEALTH-F3-2009-223101 ‘AntiPathoGN,’ FP7-HEALTH-2010-261460 ‘Gums & Joints,’ FP7-PEOPLE-2011-290246 ‘RAPID,’ BIO2009-10334, CSD2006-00015, 2009SGR1036, Fundació ‘La Marató de TV3’ 100372, DFG Graduate School 1043 Immunotherapy; Natural Sciences and Medical Research Center (NMFZ) Mainz]. S.T.-M. is the recipient of an FPI fellowship from the former Spanish Ministry for Science and Technology.

## References

- Ambort, D., Brellier, F., Becker-Pauly, C., Stöcker, W., Andrejevic-Blant, S., Chiquet, M., and Sterchi, E.E. (2010). Specific processing of tenascin-C by the metalloprotease meprin β neutralizes its inhibition of cell spreading. *Matrix Biol.* 29, 31–42.
- Apte, S.S. (2009). A disintegrin-like and metalloprotease (reprolysin-type) with thrombospondin type 1 motif (ADAMTS) superfamily: functions and mechanisms. *J. Biol. Chem.* 284, 31493–31497.
- Avilés, F.X., Vendrell, J., Guasch, A., Coll, M., and Huber, R. (1993). Advances in metallo-procarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. *Eur. J. Biochem.* 211, 381–389.
- Banerjee, S. and Bond, J. (2008). Prointerleukin-18 is activated by meprin β *in vitro* and *in vivo* in intestinal inflammation. *J. Biol. Chem.* 283, 31371–31377.
- Baumann, U. (2004). 152. Serralyisin and related enzymes. In: A.J. Barrett, N.D. Rawlings, and J.F. Woessner Jr., eds. *Handbook of Proteolytic Enzymes*, Vol. 1 (London: Elsevier Academic Press), pp. 579–581.
- Bayés, A., Fernández, D., Solà, M., Marrero, A., García-Piqué, S., Avilés, F.X., Vendrell, J., and Gomis-Rüth, F.X. (2007). Caught after the act: a human A-type metalloprotease in a product complex with a cleaved hexapeptide. *Biochemistry* 46, 6921–6930.
- Becker-Pauly, C., Bruns, B.C., Damm, O., Schütte, A., Hammouti, K., Burmester, T., and Stöcker, W. (2009). News from an ancient world: two novel astacin metalloproteases from the horseshoe crab. *J. Mol. Biol.* 385, 236–248.

- Becker-Pauly, C., Barré, O., Schilling, O., Auf dem Keller, U., Ohler, A., Broder, C., Schutte, A., Kappelhoff, R., Stöcker, W., and Overall, C.M. (2011). Proteomic analyses reveal an acidic prime side specificity for the astacin metalloprotease family reflected by physiological substrates. *Mol. Cell. Proteomics* *10*, M111.009233.
- Beckmann, G. and Bork, P. (1993). An adhesive domain detected in functionally diverse receptors. *Trends Biochem. Sci.* *18*, 40–41.
- Bode, W. and Huber, R. (1978). Crystal structure analysis and refinement of two variants of trigonal trypsinogen: trigonal trypsin and PEG (polyethylene glycol) trypsinogen and their comparison with orthorhombic trypsin and trigonal trypsinogen. *FEBS Lett.* *90*, 265–269.
- Bode, W., Gomis-Rüth, F.X., Huber, R., Zwilling, R., and Stöcker, W. (1992). Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* *358*, 164–167.
- Bode, W., Gomis-Rüth, F.X., and Stöcker, W. (1993). Astacins, seralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins.' *FEBS Lett.* *331*, 134–140.
- Boldt, H.B., Overgaard, M.T., Laursen, L.S., Weyer, K., Sottrup-Jensen, L., and Oxvig, C. (2001). Mutational analysis of the proteolytic domain of pregnancy-associated plasma protein-A (PAPP-A): classification as a metzincin. *Biochem. J.* *358*, 359–367.
- Bond, J.S. and Beynon, R.J. (1995). The astacin family of metalloendopeptidases. *Protein Sci.* *4*, 1247–1261.
- Borchert, N., Becker-Pauly, C., Wagner, A., Fischer, P., Stöcker, W., and Brattig, N.W. (2007). Identification and characterization of onchoastacin, an astacin-like metalloproteinase from the filaria *Onchocerca volvulus*. *Microbes Infect.* *9*, 498–506.
- Bork, P. and Beckmann, G. (1993). The CUB domain. A widespread module in developmentally regulated proteins. *J. Mol. Biol.* *231*, 539–545.
- Butler, G.S., Tam, E.M., and Overall, C.M. (2004). The canonical methionine 392 of matrix metalloproteinase 2 (gelatinase A) is not required for catalytic efficiency or structural integrity: probing the role of the methionine-turn in the metzincin metalloprotease superfamily. *J. Biol. Chem.* *279*, 15615–15620.
- Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Bryan Arendall, W. 3rd, Snoeyink, J., Richardson, J.S., et al. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* *35* (Web Server issue), W375–W383.
- Dumermuth, E., Sterchi, E.E., Jiang, W.P., Wolz, R.L., Bond, J.S., Flannery, A.V., and Beynon, R.J. (1991). The astacin family of metalloendopeptidases. *J. Biol. Chem.* *266*, 21381–21385.
- Elaroussi, M.A. and DeLuca, H.F. (1994). A new member to the astacin family of metalloendopeptidases: a novel 1,25-dihydroxyvitamin D-3-stimulated mRNA from chorioallantoic membrane of quail. *Biochim. Biophys. Acta* *1217*, 1–8.
- Fan, T.J. and Katagiri, C. (2001). Properties of the hatching enzyme from *Xenopus laevis*. *Eur. J. Biochem.* *268*, 4892–4898.
- Fehlhammer, H., Bode, W., and Huber, R. (1977). Crystal structure of bovine trypsinogen at 1.8 Å resolution. II. Crystallographic refinement, refined crystal structure and comparison with bovine trypsin. *J. Mol. Biol.* *111*, 415–438.
- Garrigue-Antar, L., François, V., and Kadler, K.E. (2004). Deletion of epidermal growth factor-like domains converts mammalian tolloid into a chordinase and effective procollagen C-proteinase. *J. Biol. Chem.* *279*, 49835–49841.
- Ge, G. and Greenspan, D.S. (2006). Developmental roles of the BMP1/TLD metalloproteinases. *Birth Defects Res., Part C* *78*, 47–68.
- Geier, G. and Zwilling, R. (1998). Cloning and characterization of a cDNA coding for *Astacus* embryonic astacin, a member of the astacin family of metalloproteases from the crayfish *Astacus astacus*. *Eur. J. Biochem.* *253*, 796–803.
- Gomis-Rüth, F.X. (2003). Structural aspects of the metzincin clan of metalloendopeptidases. *Mol. Biotechnol.* *24*, 157–202.
- Gomis-Rüth, F.X. (2008). Structure and mechanism of metallocarboxypeptidases. *Crit. Rev. Biochem. Mol. Biol.* *43*, 319–345.
- Gomis-Rüth, F.X. (2009). Catalytic domain architecture of metzincin metalloproteases. *J. Biol. Chem.* *284*, 15353–15357.
- Gomis-Rüth, F.X., Stöcker, W., Huber, R., Zwilling, R., and Bode, W. (1993). Refined 1.8 Å X-ray crystal structure of astacin, a zinc-endopeptidase from the crayfish *Astacus astacus* L. Structure determination, refinement, molecular structure and comparison with thermolysin. *J. Mol. Biol.* *229*, 945–968.
- Gomis-Rüth, F.X., Nar, H., Grams, F., Yallouros, I., Küsthardt, U., Zwilling, R., Bode, W., and Stöcker, W. (1994). Crystal structures, spectroscopic features and catalytic properties of cobalt(II)-, copper(II)-, nickel(II)-, and mercury(II)-derivatives of the zinc-endopeptidase astacin. A correlation of structure and proteolytic activity. *J. Biol. Chem.* *269*, 17111–17117.
- Gomis-Rüth, F.X., Botelho, T.O., and Bode, W. (2012). A standard orientation for metallopeptidases. *Biochim. Biophys. Acta* *1824*, 157–163.
- Goulas, T., Arolas, J.L., and Gomis-Rüth, F.X. (2010). Structure, function and latency regulation of a bacterial enterotoxin potentially derived from a mammalian adamalysin/ADAM xenolog. *Proc. Natl. Acad. Sci. USA* *108*, 1856–1861.
- Grams, F., Dive, V., Yiotakis, A., Yallouros, I., Vassiliou, S., Zwilling, R., Bode, W., and Stöcker, W. (1996). Structure of astacin with a transition-state analogue inhibitor. *Nat. Struct. Biol.* *3*, 671–675.
- Guevara, T., Yallouros, I., Kappelhoff, R., Bissdorf, S., Stöcker, W., and Gomis-Rüth, F.X. (2010). Proenzyme structure and activation of astacin metallopeptidase. *J. Biol. Chem.* *285*, 13958–13965.
- Hahn, D., Pischitzis, A., Roesmann, S., Hansen, M.K., Leuenberger, B., Luginbuehl, U., and Sterchi, E.E. (2003). Phorbol 12-myristate 13-acetate-induced ectodomain shedding and phosphorylation of the human meprin β metalloprotease. *J. Biol. Chem.* *278*, 42829–42839.
- Hedrich, J., Lottaz, D., Meyer, K., Yallouros, I., Jahnen-Dechent, W., Stöcker, W., and Becker-Pauly, C. (2010). Fetuin-A and cystatin C are endogenous inhibitors of human meprin metalloproteases. *Biochemistry* *49*, 8599–8607.
- Hege, T. and Baumann, U. (2001). The conserved methionine residue of the metzincins: a site-directed mutagenesis study. *J. Mol. Biol.* *314*, 181–186.
- Herzog, C., Kaushal, G.P., and Haun, R.S. (2005). Generation of biologically active interleukin-1β by meprin B. *Cytokine* *31*, 394–403.
- Hintze, V., Höwel, M., Wermter, C., Grosse Berkhoff, E., Becker-Pauly, C., Beermann, B., Yallouros, I., and Stöcker, W. (2006). The interaction of recombinant subdomains of the procollagen C-proteinase with procollagen I provides a quantitative explanation for functional differences between the two splice variants, mammalian tolloid and bone morphogenetic protein 1. *Biochemistry* *45*, 6741–6748.
- Holland, D.R., Tronrud, D.E., Pley, H.W., Flaherty, K.M., Stark, W., Jansonius, J.N., McKay, D.B., and Matthews, B.W. (1992). Structural comparison suggests that thermolysin and related neutral proteases undergo hinge-bending motion during catalysis. *Biochemistry* *31*, 11310–11316.
- Holley, S., Neul, J., Attisano, L., Wrana, J., Sasai, Y., O'Connor, M., De Robertis, E., and Ferguson, E. (1996). The *Xenopus*

- dorsalizing factor noggin ventralizes *Drosophila* embryos by preventing DPP from activating its receptor. *Cell* 86, 607–617.
- Hopkins, D.R., Keles, S., and Greenspan, D.S. (2007). The bone morphogenetic protein 1/Tolloid-like metalloproteinases. *Matrix Biol.* 26, 508–523.
- Huber, R. and Bode, W. (1986). Structural basis for the activation and action of trypsin. *Acc. Chem. Res.* 11, 114–122.
- Hung, C.H., Huang, H.R., Huang, C.J., Huang, F.L., and Chang, G.D. (1997). Purification and cloning of carp nephrosin, a secreted zinc endopeptidase of the astacin family. *J. Biol. Chem.* 272, 13772–13778.
- Jefferson, T., Čaušević, M., Auf dem Keller, U., Schilling, O., Isbert, S., Geyer, R., Maier, W., Tschickardt, S., Jumpertz, T., Weggen, S., et al. (2011). The metalloprotease meprin  $\beta$  generates non-toxic N-terminal amyloid precursor protein fragments *in vivo*. *J. Biol. Chem.* 286, 27741–27750.
- Kessler, E., Takahara, K., Biniaminov, L., Brusel, M., and Greenspan, D.S. (1996). Bone morphogenetic protein-1: the type I procollagen C-proteinase. *Science* 271, 360–362.
- Khan, A.R. and James, M.N. (1998). Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* 7, 815–836.
- Kronenberg, D., Bruns, B.C., Moali, C., Vadon-Le Goff, S., Sterchi, E.E., Traupe, H., Böhm, M., Hulmes, D.J., Stöcker, W., and Becker-Pauly, C. (2010). Processing of procollagen III by meprins: new players in extracellular matrix assembly? *J. Invest. Dermatol.* 130, 2727–2735.
- Lepage, T., Ghiglione, C., and Gache, C. (1992). Spatial and temporal expression pattern during sea urchin embryogenesis of a gene coding for a protease homologous to the human protein BMP-1 and to the product of the *Drosophila* dorsal-ventral patterning gene tolloid. *Development* 114, 147–163.
- Li, S.W., Sieron, A.L., Fertala, A., Hojima, Y., Arnold, W.V., and Prockop, D.J. (1996). The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenetic protein-1. *Proc. Natl. Acad. Sci. USA* 93, 5127–5130.
- Lun, H.M., Mak, C.H., and Ko, R.C. (2003). Characterization and cloning of metallo-proteinase in the excretory/secretory products of the infective-stage larva of *Trichinella spiralis*. *Parasitol. Res.* 90, 27–37.
- Mac Sweeney, A., Gil-Parrado, S., Vinzenz, D., Bernardi, A., Hein, A., Bodendorf, U., Erbel, P., Logel, C., and Gerhartz, B. (2008). Structural basis for the substrate specificity of bone morphogenetic protein 1/tolloid-like metalloproteases. *J. Mol. Biol.* 384, 228–239.
- Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G.R., Navaza, J., Sottrup-Jensen, L., and Gomis-Rüth, F.X. (2012). The crystal structure of human  $\alpha_2$ -macroglobulin reveals a unique molecular cage. *Angew. Chem. Int. Ed.* 51, 3340–3344.
- Matthews, B.W. (1988). Structural basis of the action of thermolysin and related zinc peptidases. *Acc. Chem. Res.* 21, 333–340.
- Matthews, B.W., Jansonius, J.N., Colman, P.M., Schoenborn, B.P., and Dupourque, D. (1972). Three-dimensional structure of thermolysin. *Nature* 238, 37–41.
- McCall, K.A., Huang, C., and Fierke, C.A. (2000). Function and mechanism of zinc metalloenzymes. *J. Nutr.* 130, 1437S–1446S.
- Meier, U.C., Boetzel, J., Kellermann, J., Mann, K., Billich, A., Stöcker, W., and Schramm, H.J. (1994). The cleavage of the bait region of  $\alpha_2$ -macroglobulin by human immunodeficiency virus proteinases and by astacin. *Ann. NY Acad. Sci.* 737, 431–433.
- Möhrlen, F., Hutter, H., and Zwilling, R. (2003). The astacin protein family in *Caenorhabditis elegans*. *Eur. J. Biochem.* 270, 4909–4920.
- Nagase, H. (1997). Activation mechanisms of matrix metalloproteinases. *Biol. Chem.* 378, 151–160.
- Oberholzer, A.E., Bumann, M., Hege, T., Russo, S., and Baumann, U. (2009). Metzincin's canonical methionine is responsible for the structural integrity of the zinc-binding site. *Biol. Chem.* 390, 875–881.
- Ohler, A., Debela, M., Wagner, S., Magdolen, V., and Becker-Pauly, C. (2010). Analyzing the protease web in skin: meprin metalloproteases are activated specifically by KLK4, 5 and 8 *vice versa* leading to processing of proKLK7 thereby triggering its activation. *Biol. Chem.* 391, 455–460.
- Okada, A., Sano, K., Nagata, K., Yasumasu, S., Ohtsuka, J., Yamamura, A., Kubota, K., Iuchi, I., and Tanokura, M. (2010). Crystal structure of zebrafish hatching enzyme 1 from the zebrafish *Danio rerio*. *J. Mol. Biol.* 402, 865–878.
- Pauptit, R.A., Karlsson, R., Picot, D., Jenkins, J.A., Niklaus-Reimer, A.S., and Jansonius, J.N. (1988). Crystal structure of neutral protease from *Bacillus cereus* refined at 3.0 Å resolution and comparison with the homologous but more thermostable enzyme thermolysin. *J. Mol. Biol.* 199, 525–537.
- Pérez, L., Kerrigan, J.E., Li, X., and Fan, H. (2007). Substitution of methionine 435 with leucine, isoleucine, and serine in tumor necrosis factor  $\alpha$  converting enzyme inactivates ectodomain shedding activity. *Biochem. Cell Biol.* 85, 141–149.
- Pieper, M., Betz, M., Budisa, N., Gomis-Rüth, F.X., Bode, W., and Tschesche, H. (1997). Expression, purification, characterization, and X-ray analysis of selenomethionine 215 variant of leukocyte collagenase. *J. Protein Chem.* 16, 637–650.
- Quesada, V., Sánchez, L.M., Alvarez, J., and López-Otín, C. (2004). Identification and characterization of human and mouse ovas-tacin: a novel metalloproteinase similar to hatching enzymes from arthropods, birds, amphibians, and fish. *J. Biol. Chem.* 279, 26627–26634.
- Rawlings, N.D., Barrett, A.J., and Bateman, A. (2010). MEROPS: the peptidase database. *Nucleic Acids Res.* 38, D227–D233.
- Reinemer, P., Grams, F., Huber, R., Kleine, T., Schriener, S., Piper, M., Tschesche, H., and Bode, W. (1994). Structural implications for the role of the N-terminus in the 'superactivation' of collagenases. A crystallographic study. *FEBS Lett.* 338, 227–233.
- Reyda, S., Jacob, E., Zwilling, R., and Stöcker, W. (1999). cDNA cloning, bacterial expression, *in vitro* renaturation and affinity purification of the zinc endopeptidase astacin. *Biochem. J.* 344, 851–857.
- Reynolds, S.D., Angerer, L.M., Palis, J., Nasir, A., and Angerer, R.C. (1992). Early mRNAs, spatially restricted along the animal-vegetal axis of sea urchin embryos, include one encoding a protein related to tolloid and BMP-1. *Development* 114, 769–786.
- Rothe, M., Wong, S.C., Henzel, W.J., and Goeddel, D.V. (1994). A novel family of putative signal transducers associated with the cytoplasmic domain of the 75 kDa tumor necrosis factor receptor. *Cell* 78, 681–692.
- Sachdev, M., Mandal, A., Mulders, S., Digilio, L.C., Panneerdoss, S., Suryavathi, V., Pires, E., Klotz, K.L., Hermens, L., Herrero, M.B., et al. (2012). Oocyte specific oolemmal SAS1B involved in sperm binding through intra-acrosomal SLLP1 during fertilization. *Dev. Biol.* 363, 40–51.
- Sato, S.M. and Sargent, T.D. (1990). Molecular approach to dorsoanterior development in *Xenopus laevis*. *Dev. Biol.* 137, 135–141.
- Schütte, A., Hedrich, J., Stöcker, W., and Becker-Pauly, C. (2010). Let it flow: Morpholino knockdown in zebrafish embryos reveals



- a pro-angiogenic effect of the metalloprotease meprin  $\alpha 2$ . *PLoS ONE* 5, e8835.
- Shibata, Y., Iwamatsu, T., Oba, Y., Kobayashi, D., Tanaka, M., Nagahama, Y., Suzuki, N., and Yoshikuni, M. (2000). Identification and cDNA cloning of alveolin, an extracellular metalloproteinase, which induces chorion hardening of medaka (*Oryzias latipes*) eggs upon fertilization. *J. Biol. Chem.* 275, 8349–8354.
- Shimell, M.J., Ferguson, E.L., Childs, S.R., and O'Connor, M.B. (1991). The *Drosophila* dorsal-ventral patterning gene tolloid is related to human bone morphogenetic protein 1. *Cell* 67, 469–481.
- Sieron, A.L., Tretiakova, A., Jameson, B.A., Segall, M.L., Lund-Katz, S., Khan, M.T., Li, S.W., and Stöcker, W. (2000). Structure and function of procollagen C-proteinase (mTolloid) domains determined by protease digestion, circular dichroism, binding to procollagen type I, and computer modeling. *Biochemistry* 39, 3231–3239.
- Sonneborn, H.H., Zwilling, R., and Pfeleiderer, G. (1969). [Evolution of endopeptidases. X. Cleavage specificity of low molecular weight protease from *Astacus leptodactylus* Esch]. *Hoppe-Seyler's Z. Physiol. Chem.* 350, 1097–1102.
- Sterchi, E.E., Stöcker, W., and Bond, J.S. (2008). Mepriins, membrane-bound and secreted astacin metalloproteinases. *Mol. Aspects Med.* 29, 309–328.
- Stöcker, W. and Bode, W. (1995). Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Curr. Opin. Struct. Biol.* 5, 383–390.
- Stöcker, W., Wolz, R.L., Zwilling, R., Strydom, D.J., and Auld, D. (1988). *Astacus* protease, a zinc metalloenzyme. *Biochemistry* 27, 5026–5032.
- Stöcker, W., Ng, M., and Auld, D.S. (1990). Fluorescent oligopeptide substrates for kinetic characterization of the specificity of *Astacus* protease. *Biochemistry* 29, 10418–10425.
- Stöcker, W., Breit, S., Sottrup-Jensen, L., and Zwilling, R. (1991a).  $\alpha_2$ -Macroglobulin from hemolymph of the freshwater crayfish *Astacus astacus*. *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.* 98, 501–509.
- Stöcker, W., Sauer, B., and Zwilling, R. (1991b). Kinetics of nitroanilide cleavage by astacin. *Biol. Chem. Hoppe-Seyler* 372, 385–392.
- Stöcker, W., Gomis-Rüth, F.X., Bode, W., and Zwilling, R. (1993). Implications of the three-dimensional structure of astacin for the structure and function of the astacin-family of zinc-endopeptidases. *Eur. J. Biochem.* 214, 215–231.
- Stöcker, W., Grams, F., Baumann, U., Reinemer, P., Gomis-Rüth, F.X., McKay, D.B., and Bode, W. (1995). The metzincins – topological and sequential relations between the astacins, adamalysins, serralyins, and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci.* 4, 823–840.
- Takeda, S. (2009). Three-dimensional domain architecture of the ADAM family proteinases. *Semin. Cell Dev. Biol.* 20, 146–152.
- Takeda, S., Takeya, H., and Iwanaga, S. (2012). Snake venom metalloproteinases: structure, function and relevance to the mammalian ADAM/ADAMTS family proteins. *Biochim. Biophys. Acta* 1824, 164–176.
- Tallant, C., García-Castellanos, R., Seco, J., Baumann, U., and Gomis-Rüth, F.X. (2006). Molecular analysis of ulilysin, the structural prototype of a new family of metzincin metalloproteases. *J. Biol. Chem.* 281, 17920–17928.
- Tallant, C., García-Castellanos, R., Baumann, U., and Gomis-Rüth, F.X. (2010a). On the relevance of the Met-turn methionine in metzincins. *J. Biol. Chem.* 285, 13951–13957.
- Tallant, C., Marrero, A., and Gomis-Rüth, F.X. (2010b). Matrix metalloproteinases: fold and function of their catalytic domains. *Biochim. Biophys. Acta, Mol. Cell Res.* 1803, 20–28.
- Thayer, M.M., Flaherty, K.M., and McKay, D.B. (1991). Three-dimensional structure of the elastase of *Pseudomonas aeruginosa* at 1.5-Å resolution. *J. Biol. Chem.* 266, 2864–2871.
- Titani, K., Torff, H.J., Hormel, S., Kumar, S., Walsh, K.A., Rödl, J., Neurath, H., and Zwilling, R. (1987). Amino acid sequence of a unique protease from the crayfish *Astacus fluviatilis*. *Biochemistry* 26, 222–226.
- Tsai, P.L., Chen, C.H., Huang, C.J., Chou, C.M., and Chang, G.D. (2004). Purification and cloning of an endogenous protein inhibitor of carp nephrosin, an astacin metalloproteinase. *J. Biol. Chem.* 279, 11146–11155.
- Vazeille, E., Bringer, M.A., Gardarin, A., Chambon, C., Becker-Pauly, C., Pender, S.L., Jakob, C., Müller, S., Lottaz, D., and Darfeuille-Michaud, A. (2011). Role of mepriins to protect ileal mucosa of Crohn's disease patients from colonization by adherent-invasive *E. coli*. *PLoS ONE* 6, e21199.
- Walasek, P. and Honek, J.F. (2005). Nonnatural amino acid incorporation into the methionine 214 position of the metzincin *Pseudomonas aeruginosa* alkaline protease. *BMC Biochem.* 6, 21.
- Waltersperger, S., Widmer, C., Wang, M., and Baumann, U. (2010). Crystal structure of archaemetzincin AmzA from *Methanopyrus kandleri* at 1.5 Å resolution. *Proteins* 78, 2720–2723.
- Wermter, C., Höwel, M., Hintze, V., Bombosch, B., Aufenvenne, K., Yiallourous, I., and Stöcker, W. (2007). The protease domain of procollagen C-proteinase (BMP1) lacks substrate selectivity, which is conferred by non-proteolytic domains. *Biol. Chem.* 388, 513–521.
- Wolfsberg, T., Bazan, J., Blobel, C., Myles, D., Primakoff, P., and White, J. (1993). The precursor region of a protein active in sperm-egg fusion contains a metalloprotease and a disintegrin domain: structural, functional, and evolutionary implications. *Proc. Natl. Acad. Sci. USA* 90, 10783–10787.
- Wozney, J.M., Rosen, V., Celeste, A.J., Mitscock, L.M., Whitters, M.J., Kriz, R.W., Hewick, R.M., and Wang, E.A. (1988). Novel regulators of bone formation: molecular clones and activities. *Science* 242, 1528–1534.
- Yan, L., Fei, K., Zhang, J., Dexter, S., and Sarras, M.P. Jr. (2000a). Identification and characterization of hydra metalloproteinase 2 (HMP2): a meprin-like astacin metalloproteinase that functions in foot morphogenesis. *Development* 127, 129–141.
- Yan, L., Leontovich, A., Fei, K., and Sarras, M.P. Jr. (2000b). Hydra metalloproteinase 1: a secreted astacin metalloproteinase whose apical axis expression is differentially regulated during head regeneration. *Dev. Biol.* 219, 115–128.
- Yasumasu, S., Yamada, K., Akasaka, K., Mitsunaga, K., Iuchi, I., Shimada, H., and Yamagami, K. (1992). Isolation of cDNAs for LCE and HCE, two constituent proteases of the hatching enzyme of *Oryzias latipes*, and concurrent expression of their mRNAs during development. *Dev. Biol.* 153, 250–258.
- Yasumasu, S., Shimada, H., Inohaya, K., Yamazaki, K., Iuchi, I., Yasumasu, I., and Yamagami, K. (1996). Different exon-intron organizations of the genes for two astacin-like proteases, high choriolytic enzyme (choriolytin H) and low choriolytic enzyme (choriolytin L), the constituents of the fish hatching enzyme. *Eur. J. Biochem.* 237, 752–758.
- Yiallourous, I., Grosse-Berkhoff, E., and Stöcker, W. (2000). The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. *FEBS Lett.* 484, 224–228.
- Yiallourous, I., Kappelhoff, R., Schilling, O., Wegmann, F., Helms, M.W., Auge, A., Brachtendorf, G., Berkhoff, E.G., Beermann,

B., Hinz, H.J., et al. (2002). Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoproteolytic cleavage and importance of precise amino-terminal processing. *J. Mol. Biol.* 324, 237–246.

Zapata, J.M., Pawlowski, K., Haas, E., Ware, C.F., Godzik, A., and Reed, J.C. (2001). A diverse family of proteins containing tumor necrosis factor receptor-associated factor domains. *J. Biol. Chem.* 276, 24242–24252.

Zhang, Y., Ge, G., and Greenspan, D.S. (2006). Inhibition of bone morphogenetic protein 1 by native and altered forms of  $\alpha_2$ -macroglobulin. *J. Biol. Chem.* 281, 39096–39104.

Received March 9, 2012; accepted April 18, 2012



F. Xavier Gomis-Rüth performed his PhD on protein crystallography under supervision of Wolfram Bode in the department of Robert Huber at the Max Planck Institute of Biochemistry (MPIB) in Martinsried (Germany) in 1992. Post-doctoral stays followed at the Autonomous University of Barcelona, the MPIB, and the Molecular Biology Institute

of Barcelona (IBMB) of the Spanish Research Council (CSIC), where he became a tenured research scientist in 2000. Since 2008, he has been a Research Professor at CSIC. He is the PI of the Proteolysis Lab at the Department of Structural Biology of the IBMB and author of over 100 publications in the field.



Sergio Trillo-Muyo studied human biology at the Pompeu Fabra University of Barcelona, Spain, and obtained his BSc in 2006. In 2008, he obtained an MSc in Biochemistry and Molecular Biology at the Autonomous University of Barcelona in Bellaterra, Spain. He is currently a PhD student in the Proteolysis Lab at the Department of Structural

Biology of the Molecular Biology Institute of Barcelona, which is part of the Spanish Research Council (CSIC).



Walter Stöcker studied chemistry and biology at the University of Munich, Germany. After receiving his doctoral degree in Munich in 1984, he was a scientific assistant at the Institute of Zoology, University of Heidelberg, Germany, and did research as a visiting scientist at the Harvard Medical School, Boston, MA, USA, and the Max Planck Institute

of Biochemistry, Martinsried, Germany. He received his habilitation in biology from the University of Heidelberg in 1992, and became Associate Professor of Zoology at the University of Münster, Germany in 1996. Since 2003, he has been a full professor and chair at the University of Mainz, Germany. His research focuses on the structure and function of astacin-like zinc peptidases and their interactions in the proteolytic web.



# APÉNDICE IV

Structure of the catalytic domain of the *Tannerella forsythia* matrix metallopeptidase karilysin in complex with a tetrapeptidic inhibitor





Tibisay Guevara,<sup>a</sup> Mirosław Ksiazek,<sup>b</sup> Peter Skottrup,<sup>c,‡</sup> Núria Cerdà-Costa,<sup>a</sup> Sergio Trillo-Muyo,<sup>a</sup> Iñaki de Diego,<sup>a</sup> Erik Riise,<sup>c</sup> Jan Potempa<sup>b,d</sup> and F. Xavier Gomis-Rüth<sup>a\*</sup>

<sup>a</sup>Proteolysis Lab, Molecular Biology Institute of Barcelona, Spanish Research Council CSIC, Barcelona Science Park, c/Baldri Reixac 15-21, 08028 Barcelona, Catalonia, Spain,

<sup>b</sup>Department of Microbiology, Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Ul. Gronostajowa 7, 30-387 Kraków, Poland, <sup>c</sup>Biomolecular Interaction Group, Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark, and <sup>d</sup>Oral Health and Systemic Diseases Research Group, University of Louisville School of Dentistry, Louisville, Kentucky, USA

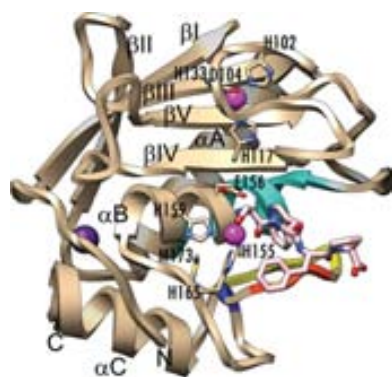
‡ Present address: DAKO Denmark, an Agilent Technologies Company, Produktionsvej 42, DK-2600 Glostrup, Denmark.

Correspondence e-mail: xgrcri@ibmb.csic.es

Received 9 January 2013

Accepted 18 March 2013

**PDB Reference:** Kly18–SWFP complex, 4in9



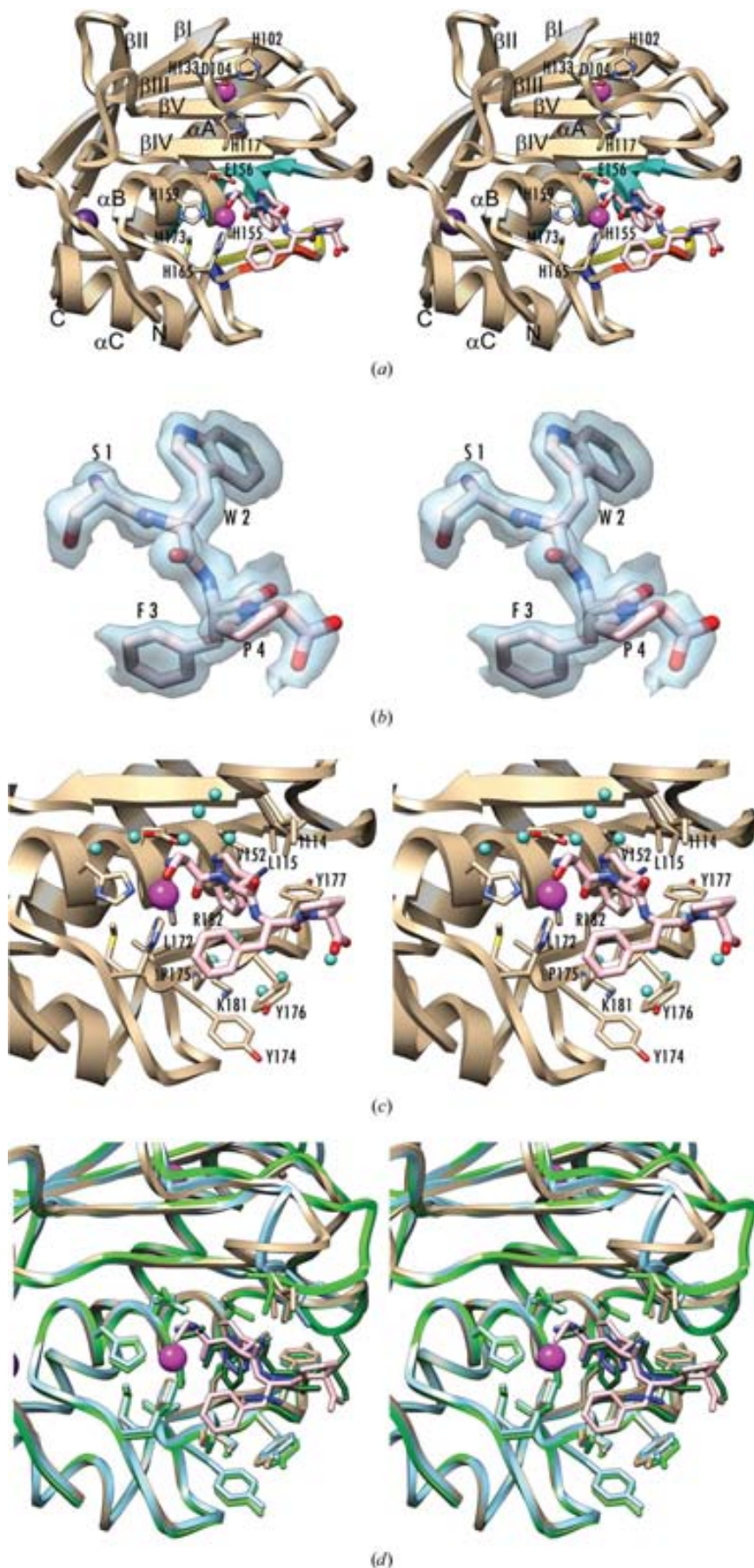
© 2013 International Union of Crystallography  
All rights reserved

## Structure of the catalytic domain of the *Tannerella forsythia* matrix metalloproteinase karilysin in complex with a tetrapeptidic inhibitor

Karilysin is the only metalloproteinase identified as a virulence factor in the odontopathogen *Tannerella forsythia* owing to its deleterious effect on the host immune response during bacterial infection. The very close structural and sequence-based similarity of its catalytic domain (Kly18) to matrix metalloproteinases suggests that karilysin was acquired by horizontal gene transfer from an animal host. Previous studies by phage display identified peptides with the consensus sequence XWFPXXXGGG (single-letter amino-acid codes; X represents any residue) as karilysin inhibitors with low-micromolar binding affinities. Subsequent refinement revealed that inhibition comparable to that of longer peptides could be achieved using the tetrapeptide SWFP. To analyze its binding, the high-resolution crystal structure of the complex between Kly18 and SWFP was determined and it was found that the peptide binds to the primed side of the active-site cleft in a substrate-like manner. The catalytic zinc ion is clamped by the  $\alpha$ -amino group and the carbonyl O atom of the serine, thus distantly mimicking the general manner of binding of hydroxamate inhibitors to metalloproteinases and contributing, together with three zinc-binding histidines from the protein scaffold, to an octahedral-minus-one metal-coordination sphere. The tryptophan side chain penetrates the deep partially water-filled specificity pocket of Kly18. Together with previous serendipitous product complexes of Kly18, the present results provide the structural determinants of inhibition of karilysin and open the field for the design of novel inhibitory strategies aimed at the treatment of human periodontal disease based on a peptidic hit molecule.

### 1. Introduction

Bacterial infection is responsible for periodontitis, an inflammatory gum disease that affects 10–15% of adults worldwide (Petersen & Ogawa, 2005). In the most severe forms of the disease, which may result in tooth loss, three bacterial species are found which form the ‘red complex’: *Porphyromonas gingivalis*, *Treponema denticola* and *Tannerella forsythia* (Socransky *et al.*, 1998; Rôças *et al.*, 2001). Infection entails the secretion of virulence factors that contribute to disease progression, and among such factors are proteolytic enzymes. In *T. forsythia*, proteolytic virulence factors have been found to include the cysteine proteinase PrtH, the trypsin-like serine proteinase BspA and, most recently, the metalloproteinase (MP) karilysin, which is responsible for the degradation of bactericidal peptides and the inactivation of host complement (Sharma, 2010; Karim *et al.*, 2010; Cerdà-Costa *et al.*, 2011; Koziel *et al.*, 2010; Jusko *et al.*, 2012). Karilysin and *Bacillus anthracis* MmpZ (Pomerantsev *et al.*, 2011) are the only members of the matrix metalloproteinase (MMP) family within the metzincin clan of MPs (Gomis-Rüth, 2003, 2009) characterized from bacteria, which strongly suggests that these proteins arose from eukaryotic-to-prokaryotic horizontal gene transfer, resulting in xenology (Cerdà-Costa *et al.*, 2011). Recently, the three-dimensional structure of the catalytic MMP-like domain of karilysin, Kly18, was analyzed with and without magnesium (Cerdà-Costa *et al.*, 2011), as alkaline-earth metals had been reported to be required for activity (Karim *et al.*, 2010). In addition, initial inhibitory studies of karilysin identified several phage display-selected peptides with apparent inhibition constants ( $K_i$ ) in the micromolar range, among



**Figure 1**  
 The complex between Kly18 and the SWFP tetrapeptide. (a) Ribbon-type plot in cross-eyed stereo of Kly18 in the standard MP orientation (Gomis-Rüth *et al.*, 2012) showing the regular secondary-structure elements ( $\alpha$ -helices as ribbons labelled  $\alpha A$ – $\alpha C$  and  $\beta$ -strands as arrows labelled  $\beta I$ – $\beta V$ ). The segment delimiting the  $S_1'$  pocket is shown with its constituent parts in blue (Met-turn), orange ( $S_1'$  wall-forming segment) and yellow (specificity loop). Residues participating in the binding of the two zinc ions (magenta spheres) are shown with their side chains as tan C atoms and are labelled. The tetrapeptide is shown as a stick model with pink C atoms. The tentative potassium cation is displayed as a purple sphere. (b) Initial OMIT ( $2F_{obs} - F_{calc}$ )-type electron density in cross-eyed stereo centred at  $0.8\sigma$  above the threshold and centred on the final refined model of the inhibitory tetrapeptide (Ser1-Trp2-Phe3-Pro4). (c) Close-up view of (a) in cross-eyed stereo centred on the active-site cleft. Protein residues participating in delimiting the specificity pocket are shown as sticks with tan C atoms and are labelled if not labelled in (a). Solvent molecules are shown as cyan spheres. The side chain of Leu115 is in a double conformation. (d) Superposition in cross-eyed stereo of Kly18 in the present complex (protein in tan, inhibitor in pink) and in the serendipitous product complexes of its previously described magnesium-depleted (PDB entry 2xs3; protein in pale green, peptide in dark green) and magnesium-bound (PDB entry 2xs4; protein in pale blue, peptide in dark blue) forms (Cerdà-Costa *et al.*, 2011). The view is similar to that in (c). The cations correspond to those of the SWFP complex.



**Table 1**

Data-collection and processing statistics.

Values in parentheses are for the outermost resolution shell.

Synchrotron	ESRF
Beamline	ID14-4
Detector	ADSC Q315R
Data-collection date	31 July 2012
No. of frames	180
Exposure time (s)	1
Rotation increment (°)	1
Refined crystal-to-detector distance (mm)	239.4
Crystal mosaicity (°)	0.424
Space group	<i>I</i> 4
Unit-cell parameters (Å)	<i>a</i> = 86.05, <i>c</i> = 49.44
Wavelength (Å)	0.9393
No. of measurements/unique reflections	183548/26344
Resolution range (Å)	42.9–1.55 (1.59–1.55)
Completeness (%)	99.9 (98.7)
$R_{\text{merge}}^{\dagger}$	0.081 (0.749)
$R_{\text{r.i.m.}} (= R_{\text{meas}})^{\ddagger}$	0.088 (0.854)
$\langle I/\sigma(I) \rangle$	18.4 (2.1)
Wilson <i>B</i> factor (Å <sup>2</sup> )	21.2
Average multiplicity	7.0 (4.3)

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ ;  $R_{\text{r.i.m.}} = \sum_{hkl} \{N(hkl) / [N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  is the *i*th intensity measurement and  $n_{hkl}$  is the redundancy of reflection *hkl*, including symmetry-related reflections, and  $\langle I(hkl) \rangle$  is its average intensity.  $R_{\text{r.i.m.}}$ , the redundancy-independent merging *R* factor (also known as  $R_{\text{meas}}$ ), is an improved multiplicity-weighted indicator of the quality of the data. For details, see Weiss (2001) and Evans (2006).

which was the tetrapeptide SWFP ( $K_i = 10.7 \mu\text{M}$ ; Skottrup *et al.*, 2012).

In order to understand the structural determinants of this inhibition, which would allow us to identify hit compounds for the development of drugs against Kly18, we determined the high-resolution X-ray crystal structure of the Kly18–SWFP complex. The results pave the way for the design of novel therapeutic strategies to palliate the effects of periodontal disease.

## 2. Materials and methods

### 2.1. Crystallization

Kly18 was prepared as reported by Karim *et al.* (2010). The complex between Kly18 and the SWFP peptide (Skottrup *et al.*, 2011) was obtained by mixing them in a 1:2 (enzyme:inhibitor) molar ratio (protein at 8.0 mg ml<sup>-1</sup> and peptide at 5.3 mg ml<sup>-1</sup>, both in buffer consisting of 5 mM Tris–HCl, 5 mM calcium chloride) and incubating the mixture at room temperature for 1 h. Crystallization assays were performed by the sitting-drop vapour-diffusion method. Reservoir solutions were prepared using a Tecan robot and 100 nl crystallization drops were dispensed onto 96 × 2-well MRC plates (Innovadyne) by a Phoenix nanodrop robot (Art Robbins) at the High-Throughput Crystallography Platform (PAC) at Barcelona Science Park. Plates were stored in Bruker steady-temperature crystal farms at 277 and 293 K. Successful conditions were scaled up to the microlitre range in 24-well Cryschem crystallization dishes (Hampton Research). The best crystals were obtained at 293 K from 1:1 μl drops of complex solution (8 mg ml<sup>-1</sup> in 5 mM Tris–HCl pH 8, 5 mM calcium chloride, 0.02% sodium azide) with 0.4 M sodium/potassium tartrate as the reservoir solution. Crystals were cryoprotected by immersion into cryosolution [0.32 M sodium/potassium tartrate, 25% (v/v) glycerol].

### 2.2. Diffraction data collection

A complete diffraction data set was collected at 100 K from a liquid-nitrogen flash-cryocooled crystal (Oxford Cryosystems 700 Series Cryostream) using an ADSC Q315R CCD detector on

**Table 2**

Structure-refinement and validation statistics.

Resolution range used for refinement (Å)	∞–1.55
No. of reflections in working set/test set	24983/1335
Crystallographic <i>R</i> factor/free <i>R</i> factor <sup>†</sup>	0.145/0.171
No. of protein atoms <sup>‡</sup>	1311
No. of peptide atoms	39
No. of solvent molecules	208
No. of ligands	4 (CH <sub>2</sub> OH) <sub>2</sub> CHOH
No. of ions	2 Zn <sup>2+</sup> , 1 K <sup>+</sup> , 1 Na <sup>+</sup>
R.m.s.d. from target values	
Bond lengths (Å)	0.010
Bond angles (°)	0.92
Overall average <i>B</i> factor (Å <sup>2</sup> )	17.6
Model validation <sup>§</sup>	
Main-chain conformational angle analysis	
Residues in favoured regions	156 (97.5%)
Outliers	0
All residues	160
Poor rotamers (%)	0.72
Residues with bad bonds	0
Residues with bad angles	0
<i>C</i> <sup>β</sup> deviations > 0.25 Å	0

<sup>†</sup> Crystallographic *R* factor =  $\sum_{hkl} | |F_{\text{obs}}| - k|F_{\text{calc}}| | / \sum_{hkl} |F_{\text{obs}}|$ , where *k* is a scaling factor and  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure-factor amplitudes, respectively. This factor is calculated for the working-set reflections; the free *R* factor is calculated in the same way for a test set of reflections (>500) not used during refinement. <sup>‡</sup> Including atoms with double occupancy. <sup>§</sup> According to *MolProbity* (Chen *et al.*, 2010).

beamline ID14-4 of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) within the Block Allocation Group ‘BAG Barcelona’. This crystal was body-centred tetragonal, with one complex per asymmetric unit, and was isomorphous to the previously reported magnesium-liganded form of Kly18 (Cerdà-Costa *et al.*, 2011). Diffraction data to 1.55 Å resolution were integrated, scaled, merged and reduced with the programs *XDS* and *XSCALE* (Kabsch, 2010; see Table 1).

### 2.3. Structure determination

The structure of the Kly18–SWFP complex (Fig. 1*a*) was solved by Fourier synthesis using the coordinates of the protein part only of the magnesium-liganded structure (PDB entry 2xs4; Cerdà-Costa *et al.*, 2011). Initial crystallographic refinement with the program *BUSTER/TNT* (Blanc *et al.*, 2004), which included TLS refinement, yielded an electron-density map that readily revealed that the inhibitor was anchored to the active cleft of the enzyme (Fig. 1*b*). Subsequent cycles of model building with *Coot* (Emsley *et al.*, 2010) were alternated with crystallographic refinement with *BUSTER/TNT* until completion of the model. The final model contained residues Tyr35–Asn53 and Leu58–Phe200 (protein residue numbering from UniProt sequence database access code D0EM77), four cations (two zinc, one potassium and one sodium, which replaces a solvent molecule found in the previous magnesium-bound structure), four glycerol molecules and 208 solvent molecules (see Table 2). Only the loop connecting strand βI to helix αA (*LβIαA*; for the nomenclature and the extent of the regular secondary-structure elements, see Figs. 1*a* and 2*c* in Cerdà-Costa *et al.*, 2011) was partially disordered and was not traced for Ser54–His57.

### 2.4. Miscellaneous

Figures were prepared with *CHIMERA* (Pettersen *et al.*, 2004). Model validation was performed with *MolProbity* (Chen *et al.*, 2010) and the *WHATCHECK* routine of *WHAT IF* (Vriend, 1990). The final coordinates of the complex between *T. forsythia* Kly18 and the SWFP tetrapeptide have been deposited in the PDB (<http://www.pdb.org>) as entry 4in9. The apparent inhibition constant ( $K_i$ ) of

the peptide AFTS against Kly18 was determined as described by Skottrup *et al.* (2012).

### 3. Results and discussion

#### 3.1. Structure of Kly18

The successful crystallization of Kly18–SWFP complex previously formed in solution only occurred under conditions containing sodium/potassium tartrate; these conditions were thus unrelated to those employed to obtain the magnesium-unbound and magnesium-bound forms (Cerdà-Costa *et al.*, 2011). The latter crystallized using 45% 2-methyl-2,4-pentanediol, 0.2 M ammonium acetate, 0.1 M Tris–HCl pH 8.5 and 14% sodium polyacrylate 5100, 20 mM magnesium chloride, 0.1 M HEPES pH 7.5, respectively, as reservoir solutions. The present crystals, however, were isomorphous to the magnesium-bound form and belonged to the tetragonal space group *I4* (compare Table 1 with Table 2 of Cerdà-Costa *et al.*, 2011).

The structure of Kly18 conforms to the overall fold of MMPs (Tallant *et al.*, 2010) as described previously (Cerdà-Costa *et al.*, 2011). Briefly, it has a compact globular shape (Fig. 1*a*) traversed by a shallow active-site cleft with a catalytic zinc ion at its bottom at half cleft-width. The former separates an upper N-terminal subdomain (NTS) from a lower C-terminal subdomain (CTS). The NTS consists of a twisted five-stranded  $\beta$ -sheet (strands  $\beta$ I– $\beta$ V), two  $\alpha$ -helices (the ‘backing helix’  $\alpha$ A and the ‘active-site helix’  $\alpha$ B) and an ‘S-loop’ (L $\beta$ III $\beta$ IV), which encompasses the binding site for a second structural zinc ion. The CTS contains the ‘Met-turn’, centred on Met173, which creates a hydrophobic pillow for the catalytic metal-binding site and is essential for the integrity of the site in MMPs and other members of the metzincin clan of MPs (Tallant *et al.*, 2010). In addition, the ‘C-terminal helix’  $\alpha$ C and a second structural cation-binding site, which is most likely to contain a potassium ion owing to geometric considerations (Harding, 2002) and its abundance in the crystallization conditions, are found in the CTS. The possible potassium is liganded to a distorted hexagonal coordination by six O atoms (2.7–3.0 Å apart) provided by the main-chain carbonyls of Ser75, Ser78 and Leu80 and by three solvent molecules.

As in MMPs in general, the catalytic zinc ion is bound by three histidines (here His155, His159 and His165) contained in a long consensus sequence characteristic to metzincins, HEXXHXXGXXH, which also contains the general base/acid glutamate, Glu156, required for catalysis (Tallant *et al.*, 2010). The top of the active-site cleft is shaped on its nonprimed side (for cleft-site nomenclature, see Gomis-Rüth *et al.*, 2012; Schechter & Berger, 1967) by the ‘upper-rim strand’  $\beta$ IV of the NTS  $\beta$ -sheet and on its primed side by the final stretch of the S-loop, termed the ‘bulge-edge segment’. The bottom of the active-site cleft on its primed side is in turn delimited by the segment bridging the Met turn and helix  $\alpha$ C, which includes the ‘S<sub>1</sub> wall-forming segment’ (Pro175–Tyr177) at the front and the ‘specificity loop’ (Gly179–Gln183) at the back (Fig. 1*a*). Together with the first turn of the active-site helix  $\alpha$ B, the latter three structural elements contribute to shaping the size and the chemical nature of the S<sub>1</sub> (or specificity) pocket in Kly18 and also in MMPs in general (Tallant *et al.*, 2010).

#### 3.2. Inhibitor binding

The structure revealed the unambiguous presence of the tetrapeptide (*Ser1*–*Trp2*–*Phe3*–*Pro4*; peptide residues are shown in italics) in the initial (OMIT) electron-density maps (Fig. 1*b*). It is firmly anchored in a substrate-like manner to the primed side of the cleft, with the N-terminal  $\alpha$ -amino group of *Ser1* bound to the catalytic zinc

(at a distance of 2.2 Å), the O<sup>ε2</sup> atom of Glu156 (2.9 Å) and a solvent molecule (2.9 Å). The latter is itself further anchored to Glu156 O<sup>ε1</sup> (3.0 Å) and a second solvent molecule (2.7 Å; Fig. 1*c*). These two solvent molecules possibly mimic authentic substrate-atom positions on the nonprimed side of the cleft in Michaelis complexes. The overall arrangement of the inhibitor N-terminus strongly suggests that *Ser1* N is in its uncharged NH<sub>2</sub> protonation state, while Glu156 is possibly present as an anionic carboxylate. In contrast to true peptidic substrates in Michaelis complexes, none of the peptide main-chain atoms contact the protein main chain directly, with the only exception being *Phe3* N, which forms a hydrogen bond to Pro175 O of the S<sub>1</sub> wall-forming segment (at a distance of 2.8 Å).

Considering the peptide side chains, *Ser1* points away from the protein moiety into bulk solvent and is in van der Waals contact with the zinc-binding His165 (*Ser1* O<sup>γ</sup>...His165 C<sup>ε1</sup>, 3.1 Å). The bulky aromatic side chain of *Trp2* penetrates the S<sub>1</sub> pocket of the enzyme, which is lined by the side chains of Leu73, Ile114–Leu115, Val152, His155, His165, Leu172–Tyr177 and Lys181–Arg182 (Fig. 1*c*). However, the presence of six solvent molecules at the bottom of the pocket indicates that the indole side chain far from fills it completely. In addition, Lys181 from the specificity loop closes the bottom of the pocket and is kept in place by a double hydrogen bond to the main-chain O atoms of the segment preceding the Met-turn (Lys181 N<sup>ε</sup>...Ser169 O and Lys181 N<sup>ε</sup>...Ala171 O; both 2.9 Å). The positive charge of the lysine  $\epsilon$ -amino group is not compensated by any protein residue. Downstream of *Trp2*, *Phe3* again points away from the cleft, but its side chain resides on a hydrophobic base which gives rise to the cleft subsite S<sub>2</sub> and is created by the side chains of the zinc-binding His165 and the S<sub>1</sub> wall residues Pro175 and Tyr177. After *Phe3*, the inhibitor peptide chain leaves the molecular surface of Kly18, so that *Pro4*, which is not so well defined by electron density (Fig. 1*b*), projects into the bulk solvent.

#### 3.3. Comparison with previous Kly18 structures

Overall, the structure of SWFP-inhibited Kly18 is very similar to the previously reported structures of its magnesium-unbound (PDB entry 2xs3) and magnesium-bound (PDB entry 2xs4) variants (Fig. 1*d*; Cerdà-Costa *et al.*, 2011). In these structures, which were determined to resolutions of 2.40 and 1.70 Å, respectively, we had serendipitously trapped short peptides, which were probably C-terminal cleavage products resulting from the purification procedure, of four and three residues, respectively, on the primed side of the cleft, which were tentatively interpreted as AFT(S). Interestingly, we could not detect any inhibitory activity of the tetrapeptide AFTS towards Kly18 even at 0.125 mM. Superposition of the three structures reveals that they coincide overall except for the disorder found in L $\beta$ I $\alpha$ A in the present complex and in the bulge-edge segment spanning Asp109–Ile114 (Fig. 1*d*). In the magnesium-unbound form this segment is completely folded out towards downstream primed positions of the cleft when compared with the other two structures, resulting in a maximal displacement of  $\sim$ 12 Å for Thr112 C $\alpha$  between the magnesium-unbound structure and the present complex structure. This suggests that the bulge-edge segment is flexible and is capable of adopting several conformations without affecting substrate, product or inhibitor binding, at least with respect to short peptides covering positions up to P<sub>4</sub>.

In addition, a magnesium ion is found in the magnesium-bound form of Kly18 coordinated by an octahedral-minus-one ligand sphere, with the apical position that points towards bulk solvent being unoccupied, close to the position of the structural potassium of the SWFP complex and with similar ligands (except for Leu80 O) but

much shorter binding distances (2.1–2.5 Å). Overall, the polypeptide chain traces (Ser75–Ser81) coincide not only in the two ion-bound structures but also in the ion-depleted structures, suggesting that there is no evident structural explanation for the requirement of calcium for activity or ion binding in this region in general (Karim *et al.*, 2010; Cerdà-Costa *et al.*, 2011).

The overall structural similarity of the Kly18 complexes extends to the binding mode of the peptides to the active-site cleft (Fig. 1*d*), in particular for their side chains, thus supporting the structural conclusions derived from the present inhibitory-peptide complex structure. In particular, the preference of Kly18 for bulky hydrophobic residues in the P<sub>1</sub>' position of substrates, as determined by applying the CLiPS methodology (Karim *et al.*, 2010), is met by Trp2 in the present structure and by a phenylalanine in the other two structures.

#### 4. Conclusion

We conclude that the inhibitory tetrapeptide SWFP binds to Kly18 in a very similar manner to the products left behind in the primed side of the active-site cleft after cleavage of substrates. Both the presence of solvent molecules in the specificity pocket below the indole side chain of Trp2 and the unbalanced positive charge of the pocket-bottom residue Lys181 strongly suggest that a chemically modified side chain that is bulky and hydrophobic at the peptide-proximal end but that is provided with a negatively charged group ~8.6 Å from the position of the Trp2 C<sup>α</sup> atom could yield a much stronger and specific inhibitor of Kly18 than SWFP. In addition, the lack of contact of Pro4 with the protein moiety suggests that substrate-subsite positions beyond P<sub>3</sub>' are probably not relevant for the design of specific active-site inhibitors of Kly18. Further experiments will be required to verify these points and could pave the way to the design of small-molecule inhibitors that may assist in the development of novel drugs to combat *Tannerella* infection.

We are grateful to the Automated Crystallography Platform at IBMB/IRB/PCB for assistance during crystallization experiments. We acknowledge the help provided by local contacts at the ESRF synchrotron. Funding for data collection was provided in part by the ESRF. This study was supported in part by grants from European, American, Polish, Spanish, Danish and Catalan agencies (2012/04/A/NZ1/00051, 2011/03/N/NZ1/00586, 2137/7.PR-EU/2011/2, DE09761, FP7-HEALTH-F3-2009-223101 'AntiPathoGN', FP7-HEALTH-2010-261460 'Gums&Joints', FP7-PEOPLE-2011-ITN-290246 'RAPID', BIO2009-10334, BFU2012-32862, CSD2006-00015, Lundbeck Foundation grant R54-A5291 and Fundació 'La Marató de TV3' grants 2009-100732 and 2009SGR1036). The Faculty of Biochemistry,

Biophysics and Biotechnology of the Jagiellonian University in Kraków (Poland) is a beneficiary of structural funds from the European Union (grant No POIG.02.01.00-12-064/08 'Molecular Biotechnology for Health'). The funders had no role in the study design, data collection and analysis, decision to publish or preparation of this manuscript.

#### References

- Blanc, E., Roversi, P., Vonnrhein, C., Flensburg, C., Lea, S. M. & Bricogne, G. (2004). *Acta Cryst. D60*, 2210–2221.
- Cerdà-Costa, N., Guevara, T., Karim, A. Y., Ksiazek, M., Nguyen, K. A., Arolas, J. L., Potempa, J. & Gomis-Rüth, F. X. (2011). *Mol. Microbiol.* **79**, 119–132.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst. D66*, 12–21.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst. D66*, 486–501.
- Evans, P. (2006). *Acta Cryst. D62*, 72–82.
- Gomis-Rüth, F. X. (2003). *Mol. Biotechnol.* **24**, 157–202.
- Gomis-Rüth, F. X. (2009). *J. Biol. Chem.* **284**, 15353–15357.
- Gomis-Rüth, F. X., Botelho, T. O. & Bode, W. (2012). *Biochim. Biophys. Acta*, **1824**, 157–163.
- Harding, M. M. (2002). *Acta Cryst. D58*, 872–874.
- Jusko, M., Potempa, J., Karim, A. Y., Ksiazek, M., Riesbeck, K., Garred, P., Eick, S. & Blom, A. M. (2012). *J. Immunol.* **188**, 2338–2349.
- Kabsch, W. (2010). *Acta Cryst. D66*, 133–144.
- Karim, A. Y., Kulczycka, M., Kantyka, T., Dubin, G., Jabaiah, A., Daugherty, P. S., Thogersen, I. B., Enghild, J. J., Nguyen, K. A. & Potempa, J. (2010). *Biol. Chem.* **391**, 105–117.
- Koziel, J., Karim, A. Y., Przybyszewska, K., Ksiazek, M., Rapala-Kozik, M., Nguyen, K.-A. & Potempa, J. (2010). *J. Innate Immun.* **2**, 288–293.
- Petersen, P. E. & Ogawa, H. (2005). *J. Periodontol.* **76**, 2187–2193.
- Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.
- Pomerantsev, A. P., Pomerantseva, O. M., Moayeri, M., Fattah, R., Tallant, C. & Leppla, S. H. (2011). *Protein Expr. Purif.* **80**, 80–90.
- Rôças, I. N., Siqueira, J. F. Jr, Santos, K. R. N. & Coelho, A. M. A. (2001). *Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod.* **91**, 468–471.
- Schechter, I. & Berger, A. (1967). *Biochem. Biophys. Res. Commun.* **27**, 157–162.
- Sharma, A. (2010). *Periodontol.* **54**, 106–116.
- Skottrup, P. D., Leonard, P., Kaczmarek, J. Z., Veillard, F., Enghild, J. J., O'Kennedy, R., Sroka, A., Clausen, R. P., Potempa, J. & Riise, E. (2011). *Anal. Biochem.* **415**, 158–167.
- Skottrup, P. D., Sorensen, G., Ksiazek, M., Potempa, J. & Riise, E. (2012). *PLoS One*, **7**, e48537.
- Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C. & Kent, R. L. Jr (1998). *J. Clin. Periodontol.* **25**, 134–144.
- Tallant, C., Marrero, A. & Gomis-Rüth, F. X. (2010). *Biochim. Biophys. Acta*, **1803**, 20–28.
- Vriend, G. (1990). *J. Mol. Graph.* **8**, 52–56.
- Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.





