



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Caracterización estructural y cinética de un inhibidor de proteasas bifuncional de la familia Kunitz-STI: posibles sitios de unión a proteasas serino y aspártico

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. Yasel Guerra Borrego

TUTOR PRINCIPAL:

Dr. Enrique Rudiño Piñera, Instituto de Biotecnología, UNAM

MIEMBROS DEL COMITÉ TUTOR:

Dra. Rosario Muñoz Clares, Facultad de Química, UNAM

Dra. Gloria Saab Rincón, Instituto de Biotecnología, UNAM

CIUDAD UNIVERSITARIA, CD. MX., ENERO DE 2018



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Este proyecto se llevó a cabo en el Laboratorio de Bioquímica Estructural perteneciente al Departamento de Medicina Molecular y Bioprocesos en el Instituto de Biotecnología de la Universidad Nacional Autónoma de México, bajo la dirección del Dr. Enrique Rudiño Piñera.

Agradecemos el apoyo técnico de la Biol. Sonia Patricia Rojas Trejo.

Agradecemos al Laboratorio Nacional de Estructuras de Macromoléculas (LANEM) del Instituto de Química de la UNAM por las facilidades prestadas.

Agradecemos el apoyo del Programa de Apoyo a los estudios de Posgrado (PAEP).

Agradecemos el apoyo del programa de becas nacionales y mixtas del CONACyT.

Este proyecto fue financiado por CONACyT ciencia básica proyecto No. 204639 y PAPIIT IN209114.

Agradecimientos

Al Dr. Enrique Rudiño Piñera, por su apoyo, dedicación, paciencia y su amistad, pero en especial por aceptarme en su laboratorio cuando todo parecía perdido.

A la Dra. Rosario Muñoz Clares y a la Dra. Gloria Saab Rincón, por sus oportunos consejos y sugerencias a lo largo de este proyecto.

A los miembros del jurado, el Dr. Lourival Possani Postay, el Dr. Alfredo Martínez Jiménez, la Dra. Nina Pastor Colón, la Dra. Isabel Gómez Gómez y la Dra. Adela Rodríguez Romero por las críticas constructivas, sugerencias y opiniones que indudablemente han contribuido a mejorar este trabajo.

A mi familia por el todo el cariño y el apoyo que me han brindado, aún con tanta distancia entre nosotros.

A mi esposa Lien, por su amor y su paciencia en todos estos años, y sobre todo en el tiempo que no hemos estado juntos físicamente. ¡Terminó al fin!

A todos los miembros del laboratorio 8, los actuales y los que ya no están, por los consejos, ayuda y apoyo durante estos años.

A la Biol. Sonia Rojas, por estar siempre dispuesta a ayudar en lo que sea necesario, y ser el primer abrazo en la mañana al llegar al laboratorio.

A todos los amigos que conocí durante mi estancia en México, gracias por hacer mi vida aquí una experiencia inolvidable.

Publicación derivada de este trabajo

- Guerra, Y., Valiente, P. A., Pons, T., Berry, C., & Rudiño-Piñera, E. (2016). Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: Could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop? *Journal of Structural Biology*. 195 (2): 259-271.

Índice

Abreviaturas y acrónimos	9
Abstract	10
Resumen	11
1. Introducción	12
1.1. Antecedentes	13
1.1.1 Generalidades de las proteasas. Características de las proteasas serino y aspártico.	13
1.1.2 Inhibidores de proteasas como mecanismo de regulación de la actividad enzimática. Mecanismos de inhibición.	16
1.1.3 Inhibidores de unión fuerte.....	19
1.1.4 Inhibidores de la familia Kunitz-STI	20
1.1.5 Inhibidores bifuncionales de la familia Kunitz-STI. Inhibidores de proteasas serino y aspártico.	24
2. Justificación	26
2.1 Hipótesis.....	26
2.2 Objetivo	26
2.3 Objetivos Particulares	26
3. Materiales y Métodos	27
3.1 Clonaje de las secuencias codificadoras de los inhibidores de proteasas.....	27
3.2 Obtención de los mutantes E3Ad_N19D y E3Ad_N19D_R151A, y sus variantes con una etiqueta de His6x.....	28
3.3 Expresión heteróloga de los inhibidores.....	29
3.4 Purificación de los inhibidores	30
3.5 Electroforesis de isoelectroenfoque.	31
3.6 Análisis por espectrometría de masas.	32
3.7 Cristalización de los inhibidores	32
3.8 Colecta y procesamiento de datos de la difracción de rayos X.	33
3.9 Análisis de secuencia y estructural de los inhibidores de proteasas.....	33

3.10 Obtención y purificación de la proteasa Plasmepsina II de <i>Plasmodium falciparum</i>	35
3.11 Ensayos de actividad inhibidora frente a proteasas aspárticas y serino.....	36
3.12 Simulaciones de Dinámica Molecular.....	37
3.12.1 Equilibración y simulaciones productivas.....	38
3.12.2 Análisis de las simulaciones de dinámica molecular.....	39
4. Resultados y discusión.....	39
4.1 Obtención, purificación y caracterización bioquímica de los inhibidores.....	39
4.2 Cristalización y determinación de la estructura tridimensional del inhibidor E3Ad y del mutante E3Ad_N19D.	45
4.3 Análisis de secuencia y estructural con respecto a otros miembros de la familia Kunitz-STI .51	
4.4 Predicción del sitio de interacción con proteasas de tipo serino.	55
4.5 Predicción del sitio de interacción con proteasas de tipo aspártico.	63
4.6 Caracterización cinética del mutante E3Ad_N19D_R151A_His.....	66
5. Conclusiones	71
6. Perspectivas	72
7. Bibliografía	73
Anexo 1	82
Anexo 2.....	89

Lista de Figuras

- Figura 1. Esquema de la nomenclatura de Berger & Schechter para los posiciones en el sustrato y sus subsitios de interacción en el centro activo de la proteasa..... 14
- Figura 2. Estructura general de los inhibidores Kunitz-STI..... 21
- Figura 3. Conformaciones canónicas y no canónicas del asa de interacción de los inhibidores de serino proteasas de la familia Kunitz-STI..... 23
- Figura 4. Perfil cromatográfico de la exclusión molecular en una columna Superdex 75 HR 10//30 para los inhibidores bifuncionales..... 40
- Figura 5. Electroforesis de SDS-PAGE de los inhibidores E3Ad, E3Ad_N19D, E3Ad_N19D_His6x y E3Ad_N19D_R151A_His6x. 41
- Figura 6. Perfiles de espectrometría de masas ESI-MS-MS del péptido correspondiente al extremo amino del inhibidor E3Ad y de sus precursores glicosilados 42
- Figura 7. Electroforesis de isoelectroenfoque de los inhibidores E3Ad y E3Ad_N19D 43
- Figura 8. Ajuste a la ecuación de Morrison de los resultados de los ensayos de inhibición de E3Ad_N19D frente a las proteasas Plm II y tripsina de páncreas bovino 44
- Figura 9. Cristales obtenidos para los inhibidores E3Ad y E3Ad_N19D 46
- Figura 10. Estructura tridimensional global del inhibidor E3Ad_N19D..... 49
- Figura 11. Sobreposición estructural de todos los modelos cristalográficos de los inhibidores E3Ad y E3Ad_N19D..... 50
- Figura 12. Alineamiento múltiple de secuencia de algunos inhibidores de proteasas de la familia Kunitz-STI con estructura tridimensional determinada..... 52
- Figura 13. Superposición de las estructuras de los inhibidores bifuncionales PDI y E3Ad..... 54
- Figura 14. Cambios conformacionales del asa L9 de los inhibidores PDI y E3Ad durante las simulaciones de dinámica molecular a diferentes valores de pH y temperatura..... 55
- Figura 15. Comparación estructural de las posiciones P3-P3' del asa L4 de inhibidores canónicos con las posiciones equivalentes en el inhibidor E3Ad 56
- Figura 16. Comparación estructural de las posibles posiciones P3-P3' en el asa L5 de los inhibidores E3Ad, PDI y PSPI con sus posiciones equivalentes en el asa no canónica de API-A..... 59
- Figura 17. Representación de las interacciones que estabilizan el asa L5 de los inhibidores API-A, E3Ad, PSPI y PDI..... 60
- Figura 18. Interfaz proteasa-inhibidor de los inhibidores canónicos y no canónicos de proteasas serino..... 61
- Figura 19. Árbol filogenético consenso (1000 réplicas) usando el método de máxima verosimilitud de inhibidores de la familia Kunitz-STI con estructura tridimensional..... 62
- Figura 20. Predicción de posibles residuos de una interfaz proteína-proteína en los inhibidores PDI y E3Ad, usando el servidor metaPPISP. 64
- Figura 21. Modelos obtenidos por anclaje molecular utilizando las coordenadas cristalográficas de la proteasa de tipo aspartático Plm II y del inhibidor E3Ad_N19D (5FNW) (A). 65

- Figura 22. Ajuste a la ecuación de Morrison de los resultados de los ensayos de inhibición frente a la proteasa Plm II de los inhibidores E3Ad_N19D_His₆ (A), E3Ad_N19D_R151A_His₆ (B), E3Ad_N19D_R151A (C).....67
- Figura 23. Modelos obtenidos por anclaje molecular utilizando las coordenadas cristalográficas de la proteasa de tipo aspártico Plm II y del inhibidor E3Ad_N19D (5FNW) (salmón) y el mutante E3Ad_N19D_R151A (verde).....68
- Figura 24. Esquema de la frecuencia de los aminoácidos en cada posición del asa L9. Se indica la numeración de cada posición para el inhibidor E3Ad.....69
- Figura A1. N-Terminal de la proteína.....84
- Figura A2. Análisis completo de iones precursores glicosilados85
- Figura A3 Fragmentación de uno de los péptidos glicosilados.....86
- Figura A4 Fragmentación del péptido glicosilado.....86

Listado de Tablas

- Tabla 1. Parámetros estadísticos de las estructuras cristalográficas de los inhibidores E3Ad y E3Ad_N19D.46
- Tabla 2. Valores de r.m.s.d. de C^α de las diferentes estructuras cristalográficas obtenidas para E3Ad y E3Ad_N19D.....51
- Tabla 3. Valores de r.m.s.d C^α (Å) del asa L4 (P70-I75) para las estructuras cristalográficas de los inhibidores E3Ad y el mutante E3Ad_N19D, con las posiciones P3-P3' de asas canónicas de algunos inhibidores Kunitz-STI.....56
- Tabla 4. Valores de r.m.s.d C^α (Å) asa L5 (E3Ad, T89-V94) para las estructuras cristalográficas de los inhibidores E3Ad y el mutante E3Ad_N19D con las posiciones P3-P3' de asas canónicas.....58

Abreviaturas y acrónimos

AOX1: Alcohol oxidasa 1

API-A: *Arrowhead protease inhibitor A* PDI: *Potato cathepsin D Inhibitor*

BASI: *Barley α -amylase/subtilisin inhibitor*

Bicina: *N, N-Bis(2-hydroxyethyl) glycine*

BMGY: *Buffered Glycerol-complex medium*

BMMY: *Buffered Methanol-complex medium*

EcTI: *Enterolobium cortisolum Trypsin inhibitor*

EDTA: *Ethylenediaminetetraacetic acid*

ESI-MS-MS: *Electrospray ionization tandem mass spectrometry*

ETI: *Erythrina caffra Trypsin inhibitor*

HEPES: *4-(2-Hydroxyethyl) piperazine-1-ethanesulfonic acid*

LB: *Luria-Bertani*

LSPI: *Latex serine protease inhibitor*

MES: *2-(N-Morpholino) ethanesulfonic acid*

MME: *DL-malic acid/MES/Tris base buffer*

MPD: *(+/-)-2-Methyl-2,4-pentanediol*

PCR: *Reacción en cadena de la polimerasa*

PDB: *Protein Data Bank*

PDI: *Potato Cathepsin D Inhibitor*

PEG: *Polietilén glicol*

PI3: *Ascaris sum pepsin inhibitor-3*

Plm II: *Plasmepsina II de Plasmodium falciparum*

PSPI: *Potato serine protease inhibitor*

RASI: *Rice α -amylase/subtilisin inhibitor*

r.m.s.d.: *root-mean-square deviation*

SDS-PAGE: *Sodium dodecyl sulphate-polyacrylamide gel electrophoresis*

SLAPI: *Solanum lycopersicum aspartic protease inhibitor*

SQAPI: *Squash aspartic acid proteinase inhibitor*

STI: *Soybean trypsin inhibitor*

TBE: *Tris/Boric acid/ EDTA buffer*

TKI: *Tamarind Kunitz Inhibitor*

VIH: *Virus de inmunodeficiencia humana*

WCI: *Winged bean chymotrypsin inhibitor*

Abstract

Bifunctional protease inhibitors belonging to the Kunitz-STI family are promising tools for targeting relevant proteases in biomedicine and biotechnology. Their ability to interact with proteases of different catalytic types is among their most interesting potentialities, together with recent reports of the use of β -trefoil fold as a scaffold for the design of new multifunctional proteins (Azarkan, Martinez-Rodriguez, Buts, Baeyens-Volant, & Garcia-Pino, 2011; Broom et al., 2012; Lee, Blaber, Dubey, & Blaber, 2011). In this work, we have obtained six crystallographic structures of a bifunctional protease inhibitor (E3Ad) from *Solanum tuberosum*, expressed in the yeast *Pichia pastoris*. These three-dimensional structures were determined for a glycosylated form as well as for a non-glycosylated mutant (N19D) of the inhibitor E3Ad, showing that deglycosylation does not affect the global tridimensional structure of the inhibitor. Besides, the different pH values of the crystallization conditions allowed us to analyze the effect of pH change upon the tridimensional structure of the inhibitor. The comparison of the crystallographic structures obtained for the E3Ad inhibitor with other Kunitz-STI family members showed very interesting differences from other bifunctional inhibitors within this family. First, the proposed aspartic protease-binding loop could adopt different conformations, even in inhibitors with high sequence identity percentage, such as E3Ad and PDI. Molecular dynamics simulations suggest that the loop L9 could adopt different conformations depending upon the pH and loop amino acid sequence. Second, we have proposed that the serine protease-binding loop for bifunctional protease inhibitors like E3Ad and PDI is the loop L5, which has a non-canonical conformation in contrast with most of the members of the Kunitz-STI family that have a canonical protease-binding loop. Phylogenetic analysis supports this hypothesis, which in the case of being confirmed, will mean that non-canonical serine-protease loops are not so uncommon in Kunitz-STI inhibitors as it was believed.

Kinetics evaluation of the mutant R151A showed a statistically significant difference in the K_i value compared to the wild type although it is very low (2.8 fold) to be considered a hot- or a warm-spot residue of the protease-inhibitor interface. The high amino acid sequence variability in loop L9 make it very difficult to identify essential interacting residues, which points out the necessity to obtain a tridimensional structure of the protease-inhibitor complex or a larger-scale mutagenesis study of loop L9.

Resumen

Los inhibidores de proteasas bifuncionales pertenecientes a la familia Kunitz-STI son una herramienta prometedora para controlar la actividad enzimática de proteasas de interés en la biomedicina y la biotecnología. Su versatilidad para inhibir diferentes familias de proteasas es uno de los motivos de sus potencialidades (Azarkan et al., 2011). Además, se ha demostrado que es posible usar el plegamiento que presentan estos inhibidores (β -trefoil), para diseñar nuevas proteínas con varios sitios de reconocimiento para sus proteínas dianas (Broom et al., 2012; Lee et al., 2011; Terada et al., 2017). En el presente trabajo se determinaron seis estructuras tridimensionales de un inhibidor bifuncional (E3Ad), aislado de *Solanum tuberosum* y expresado de forma heteróloga en la levadura *Pichia pastoris*. Se obtuvieron estructuras tridimensionales para el inhibidor en su forma glicosilada, así como para un mutante no glicosilado, demostrando que la ausencia de glicosilaciones no afecta la estructura tridimensional global del inhibidor. Además, los diferentes valores de pH de las condiciones de cristalización para las estructuras obtenidas, nos permitieron analizar el efecto del pH sobre la estructura del inhibidor. La comparación de las estructuras obtenidas en este trabajo con aquellas disponibles para otros miembros de la familia Kunitz-STI mostró elementos muy interesantes sobre estos inhibidores bifuncionales. Primero, la conformación del asa propuesta como la región involucrada en la interacción con las proteasas aspárticas puede adoptar diferentes conformaciones, aun cuando los inhibidores tengan una alta identidad de secuencia entre sí, como es el caso de E3Ad y PDI. Simulaciones de dinámica molecular sugieren que la conformación de esta asa puede incluso cambiar en función de la secuencia del asa y el pH. Segundo, los inhibidores bifuncionales E3Ad y PDI podrían interactuar con la tripsina a través de su asa L5, el cual presenta una conformación no canónica, diferente a la mayoría de los miembros de la familia Kunitz-STI. Análisis de filogenia apoyan esta propuesta, que de ser verificada significaría que el uso de conformaciones no canónicas por inhibidores de la familia Kunitz-STI para inhibir serino proteasas, no es algo tan inusual como se pensaba hasta ahora.

Los resultados de la evaluación cinética del mutante R151A mostraron cambios estadísticamente significativos en la K_i , pero muy pequeños (2.83 veces) como para ser considerado un residuo importante en la interfaz proteasa-inhibidor. La gran variabilidad de secuencia en esta asa entre los inhibidores caracterizados dificulta la identificación de residuos esenciales para esta interacción, lo que conduce a la necesidad de obtener un complejo proteasa-inhibidor o un estudio más amplio de mutagénesis en diferentes posiciones en el asa L9.

1. Introducción

Las proteasas son enzimas que están involucradas en etapas claves de múltiples procesos biológicos, lo que las convierte en una diana atractiva para el desarrollo de nuevas estrategias para controlar su actividad. Los inhibidores de proteasas han demostrado ser un mecanismo eficiente de control, con amplias aplicaciones en la biomedicina y la biotecnología. Entre las familias de inhibidores de naturaleza proteica descritas, la familia Kunitz-STI se caracteriza por su variedad de funciones de sus miembros, no solo como inhibidores de proteasas de diferentes clases mecanísticas sino por inhibir a otros tipos de enzimas. Algunos de estos inhibidores son capaces de interactuar con más de una proteasa diana. Esta capacidad de una molécula de interactuar con varias proteasas diana es lo que se denomina bifuncionalidad o multifuncionalidad, y es considerada una ventaja en términos de aplicación práctica, al ser capaz de regular la actividad de varias enzimas. En particular, dentro de la familia Kunitz-STI existe un grupo de inhibidores que son capaces de interactuar con proteasas serino y aspártico, y que han sido aislados fundamentalmente de *Solanum tuberosum* y *Solanum lycopersicum* (Cater et al., 2002; Keilová H. and Tomášek, 1976b; Mares, Meloun, Pavlik, Kostka, & Baudys, 1989). Si bien, estos inhibidores bifuncionales fueron descritos por primera vez desde hace alrededor de 40 años, hasta muy recientemente no se disponía de una estructura tridimensional de alguno de ellos (Keilová H. and Tomášek, 1976b). Esta ausencia de una estructura tridimensional limitaba las posibilidades de predecir los posibles sitios de interacción con su proteasa diana. Incluso, la región implicada en la inhibición de las proteasas serino, que ha sido ampliamente estudiada en otros miembros de la familia, presenta una mayor conservación en su conformación con respecto a su secuencia de aminoácidos. Un estudio previo usando herramientas computacionales permitió predecir el posible sitio de interacción con las proteasas aspártico, y lo ubicó en un asa con una gran hiper variabilidad dentro de la familia (Guerra, Valiente, Berry, & Pons, 2012). Anteriormente, se había detectado que la hidrólisis de un enlace peptídico en esa asa afectaba la constante de inhibición frente la proteasa aspártica sacaropepsina (Cater et al., 2002), abriendo la posibilidad de que puede ser parte de la interfaz proteasa-inhibidor.

Durante el desarrollo del presente proyecto otros autores informaron la determinación de la estructura tridimensional de un inhibidor de catepsina D y tripsina, con alta similitud de secuencia con el inhibidor estudiado por nosotros (Guo, Erskine, Coker, Wood, & Cooper, 2015). Sin embargo, como se demuestra durante la comparación de nuestras estructuras con la de estos autores, el asa predicha anteriormente como la región de interacción con las proteasas aspártico adopta conformaciones completamente diferentes en ambos inhibidores. Simulaciones de dinámica molecular realizadas con ambos inhibidores sugieren que el asa L9 puede adoptar diferentes conformaciones en función del pH del medio y de la secuencia del asa.

La obtención de estructuras del inhibidor bifuncional E3Ad de *S. tuberosum* en su forma glicosilada y no glicosilada, permitió evaluar el efecto de las glicosilaciones en la estructura global y local del inhibidor, concluyendo que su ausencia no afecta la estructura del inhibidor. Esto está en concordancia con la presencia de actividad inhibidora frente a proteasas serino y aspártico en ausencia de glicosilaciones.

Adicionalmente, en el presente trabajo se demostró que el inhibidor E3Ad es capaz de inhibir a la proteasa aspártico Plm II, lo que constituye el primer informe de inhibición frente a esta enzima y la convierte en un modelo para estudiar la inhibición de las proteasas aspártico por inhibidores bifuncionales de la familia Kunitz-STI. Sin embargo, los resultados cinéticos no nos permiten validar el modo de interacción modelado mediante acoplamiento molecular a partir de las estructuras tridimensionales de la proteasa y el inhibidor libre. Uno de los resultados no esperados derivados de este trabajo es el hecho que no encontramos un asa con una conformación canónica, similar a la descrita para la mayoría de los inhibidores de proteasas serino pertenecientes a esta familia. En su lugar, nuestro análisis estructural y filogenético sugiere que el inhibidor bifuncional E3Ad, así como otros casos aislados de *Solanum tuberosum*, podrían usar un asa con una conformación no canónica descrita previamente para un miembro de esta familia. En conjunto, todos estos resultados y análisis abren nuevas interrogantes interesantes, no solo acerca del modo de interacción de estos inhibidores bifuncionales con sus proteasas diana, sino además del proceso evolutivo para adquirir esta bifuncionalidad.

1.1. Antecedentes

1.1.1 Generalidades de las proteasas. Características de las proteasas serino y aspártico.

Las proteasas son enzimas hidrolíticas que catalizan la ruptura de un enlace peptídico. Estas enzimas son clasificadas como exoproteasas, en el caso de que uno de los aminoácidos que forma el enlace peptídico a ser escindido se encuentre en uno de los extremos de la cadena peptídica o a unos pocos residuos del extremo (aminopeptidasas o carboxipeptidasas). Por otro lado, si el enlace peptídico a hidrolizar se localiza en el interior de la cadena peptídica, entonces a la proteasa se le denomina endoproteasa. Otra clasificación se basa en la identidad de los aminoácidos involucrados en la catálisis enzimática, siendo los tipos reconocidos: Aspártico, Cisteínico, Glutámico, Metalo, Serino y Treonina (Barrett, Rawlings, Salvesen, & Fred Woessner, 2013). En el caso de las metaloproteasas, la clasificación se basa en la participación de iones metálicos en la catálisis. La subclasificación dentro de cada tipo catalítico se divide en clanes y familias. Los clanes agrupan a familias de proteínas homólogas, fundamentalmente con base en la similitud a nivel de estructura terciaria y/o en la organización de los aminoácidos catalíticos y su entorno. Las familias se organizan

basadas en la similitud de secuencia entre sus miembros (Rawlings, Waller, Barrett, & Bateman, 2014). Esta organización jerárquica es empleada por la base de datos MEROPS (<http://merops.sanger.ac.uk/>) tanto para las proteasas como para sus inhibidores de naturaleza proteica. La nomenclatura de los clanes utiliza letras acordes al grupo catalítico de las proteasas (ej. A para aspártico, S para serino y así sucesivamente).

De forma general el mecanismo catalítico de las proteasas caracterizadas puede ser una catálisis covalente, donde se forma un intermediario covalente con uno de los aminoácidos de la enzima, o una catálisis ácido-base general (Barrett et al., 2013). En el primer grupo se encuentran las proteasas de tipo serino, cisteína y treonino; mientras en el segundo encontramos a las proteasas aspárticas, glutámicas y metaloproteasas. También, existe una nomenclatura para las posiciones de los aminoácidos en el sustrato de estas enzimas y las regiones del centro activo de la proteasa que interactúa con cada una de ellas. Esta nomenclatura es conocida como de Berger & Schechter, y nombra las posiciones hacia el extremo amino del enlace hidrolizado como P1-Pn y P1'-Pn' a aquellas posiciones en la dirección del extremo carboxilo (Figura 1) (Schechter & Berger, 1967). Los sub-sitios del centro activo se nombran atendiendo a la posición del sustrato con la cual interactúan.

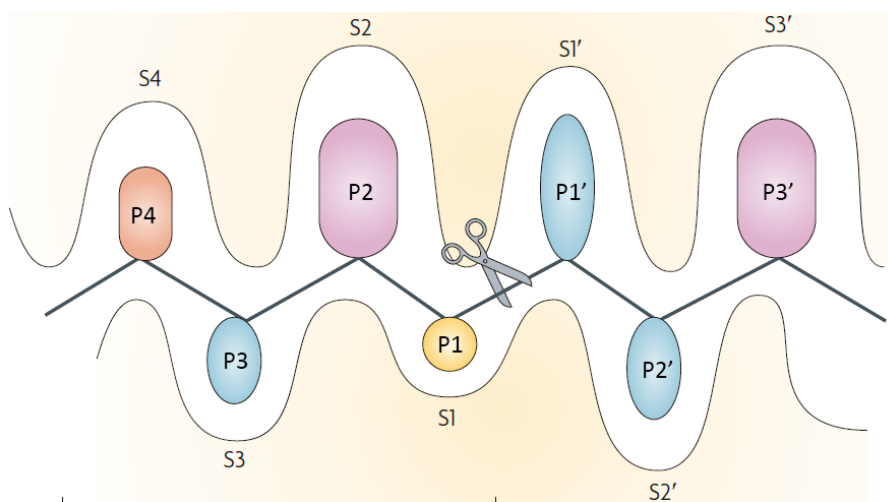


Figura 1. Esquema de la nomenclatura de Berger & Schechter para las posiciones en el sustrato (P) y sus subsitios de interacción en el centro activo de la proteasa (S). La figura fue modificada de Drag & Salvesen, 2010.

Las proteasas serino son posiblemente el grupo más estudiado de todas las proteasas. Entre las diferentes familias de este grupo la mejor caracterizada es la familia S1 del clan PA, subclan PA(S), cuyo miembro representativo es la quimiotripsina. Dentro de esta familia también se encuentra la tripsina, que ha sido ampliamente estudiada y usada como modelo de proteasa de tipo serino. Todos los miembros de la familia S1 tienen una tríada catalítica compuesta de una histidina, un ácido

aspártico y una serina, presentes en ese orden en la secuencia de aminoácidos. La totalidad de los miembros de esta familia son endopeptidasas, y presentan un plegamiento compuesto por dos dominios con la cavidad del centro activo ubicada entre ambos dominios. Cada dominio es tá compuesto por un barril β abierto y se encuentran orientados en un ángulo recto uno con respecto al otro. Como resultado del patrón cruzado de hebras β presente en los barriles β visto de lado, este plegamiento ha sido descrito como "llave griega". Además de la tríada catalítica, otra característica importante del centro activo de las proteasas serino es la presencia del sitio de unión oxianiónico. En el caso de la quimiotripsina, este sitio es formado por los grupos imino de los residuos G193 y S195 (parte de la tríada catalítica), y se encuentran involucrados en la estabilización del estado de transición a través de la formación de puentes de hidrógeno con un átomo de oxígeno con carga negativa.

El mecanismo catalítico de las proteasas serino tiene una primera etapa donde ocurre un ataque nucleofílico por parte de la Ser de la tríada catalítica (en la familia S1), sobre el grupo carbonilo del residuo en posición P1, formando un intermediario acil-enzima y un extremo amino en el residuo en posición P1'. En una segunda etapa ocurre un segundo ataque nucleofílico a través de una molécula de agua sobre el carbonilo del intermediario acil-enzima, lo que conduce a la hidrólisis del intermediario y a la consecuente liberación del extremo carboxilo y restauración del estado inicial de la serina catalítica. Los otros aminoácidos de la tríada catalítica participan en la estabilización de intermediarios y en la transferencia de protones durante la reacción. La especificidad de los sustratos está determinada de forma primaria por la identidad del residuo en la posición P1. En el caso de la quimiotripsina, existe una preferencia por aminoácidos como fenilalanina, triptófano, tirosina, leucina y metionina, mientras que en la tripsina la preferencia es por residuos básicos como arginina y lisina (Lu et al., 1997). Esta preferencia está determinada por los aminoácidos que componen la cavidad o subsitio S1. En el caso de la quimiotripsina al fondo de la cavidad se ubica una serina, mientras que en la tripsina encontramos un ácido aspártico (Lu et al., 1997).

Las proteasas aspártico son menos abundantes y están menos caracterizadas en comparación con las de tipo serino. Las familias A1 y A2 del clan AA se encuentran entre las más estudiadas. La proteasa representativa para la familia A1 es la pepsina, mientras que para la familia A2 es la retropepsina del VIH, la que ha sido ampliamente utilizada en el diseño de fármacos antivirales (Ghosh, Osswald, & Prato, 2016). Las proteasas de la familia A1 poseen un plegamiento bilobular, donde cada lóbulo está compuesto por un dominio de tipo barril β , denominados con base en la presencia del extremo amino o carboxilo de la proteína en ellos. Ambos dominios se encuentran conectados a través de una hoja β antiparalela compuesta por seis hebras β . Esta conexión es importante para la flexibilidad, evidenciada muchas veces como movimientos relativos de un dominio con respecto a otro. El centro activo de la proteína se encuentra en una cavidad entre los dominios,

con los dos aspárticos catalíticos en la base de dicha cavidad. Una de las características distintivas de las proteasas de la familia A1 es la presencia de un motivo catalítico compuesto por la secuencia D-T(S)-G-T(S) (Dunn, 2002). Una característica estructural importante de las proteasas aspártico de esta familia es la presencia de una horquilla que actúa como tapa ("flap") del centro activo y se encuentra involucrada en la orientación de forma correcta de los sustratos en el centro activo de la enzima (Dunn, 2002). Esta tapa posee una marcada flexibilidad que junto a su posición en la estructura le permiten regular la accesibilidad al centro activo de la proteasa. Las proteasas aspártico emplean un mecanismo catalítico de tipo ácido-base general, e involucra a los dos residuos de ácido aspártico y a una molécula de agua (Dunn, 2002), aunque en algunos casos se han propuesto variantes ante la ausencia de uno de los aspárticos catalíticos (Bjelic & Åqvist, 2004). Esta molécula de agua es activada a un ón hidróxido por uno de los ácidos aspárticos catalíticos y realiza un ataque nucleofílico sobre el grupo carbonilo del residuo en la posición P1. Como resultado de este ataque se forma un intermediario tetraédrico que es hidrolizado por transferencia de un protón de uno de los aspárticos catalíticos, acoplado a la ruptura del enlace peptídico y con la consecuente liberación de los extremos amino y carboxilo de las posiciones P1 y P1' respectivamente. De forma general las proteasas aspártico tienen preferencia por hidrolizar enlaces peptídicos donde las posiciones P1 y P1' estén ocupadas por aminoácidos hidrofóbicos, exceptuando las proteasas de hongos (Dunn, 2002). La especificidad respecto a las otras posiciones del sustrato varía en dependencia de la proteasa en cuestión.

1.1.2 Inhibidores de proteasas como mecanismo de regulación de la actividad enzimática. Mecanismos de inhibición.

Las proteasas son enzimas que se encuentran involucradas en múltiples procesos biológicos. Algunos ejemplos de estos procesos son la cascada de coagulación sanguínea, regulación del ciclo celular, remodelación tisular, apoptosis y respuesta inmune, entre otros (Turk, 2006; Turk, Turk, & Turk, 2012; Yike, 2011). Además, las proteasas han sido identificadas como componentes claves en el desarrollo de enfermedades, así como en el ciclo de vida de organismos de interés para la biomedicina y la biotecnología (López-Otín & Bond, 2008; McKerrow, Caffrey, Kelly, Loke, & Sajid, 2006; Yike, 2011). Esto ha propiciado que estas enzimas sean consideradas como un blanco terapéutico atractivo para el desarrollo de medicamentos o como herramientas biotecnológicas (Drag & Salvesen, 2010; Klemba & Goldberg, 2002).

Una de las formas de regular la actividad enzimática es a través de inhibidores, lo que ha demostrado ser un mecanismo útil para el control de la actividad de proteasas de interés (Abbenante & Fairlie, 2005; Rivard et al., 2006; Turk, 2006). Existen múltiples ejemplos de medicamentos y otros productos biotecnológicos que actúan como inhibidores de proteasas (Abbenante & Fairlie, 2005; Fear,

Komarnytsky, & Raskin, 2007; Turk, 2006). Los inhibidores usados específicamente en biomedicina son generalmente moléculas de baja masa molecular (< 500 Da) y de naturaleza no peptídica (Abbenante & Fairlie, 2005; Turk, 2006). Sin embargo, los inhibidores de naturaleza proteica presentan varias ventajas sobre los no peptídicos, entre las que se encuentran la especificidad y la afinidad por la proteasa blanco (Cohen et al., 2016). Existen ejemplos del uso de inhibidores de proteasas de naturaleza proteica en la terapia del cáncer (Srikanth & Chen, 2016). Además, los inhibidores de proteasas son considerados una estrategia promisoriosa para el control de varias enfermedades de interés agrícola (Castro et al., 2011; Gatehouse, 2011), así como en la regulación de la actividad proteolítica indeseada durante la expresión de proteínas recombinantes en diferentes tipos de hospederos (Gil et al., 2011; Pillay, Schlüter, Van Wyk, Kunert, & Vorster, 2013).

El uso de los inhibidores para el control de fitopatógenos e insectos herbívoros se basa en la inhibición de proteasas durante la etapa de infección y/o esenciales para el ciclo de vida del organismo y de enzimas digestivas en el caso de los insectos herbívoros (Lima et al., 2011; Mehrabadi, Franco, & Bandani, 2012; Schlüter et al., 2010). Entre los organismos cuyo crecimiento o desarrollo se afecta por la presencia y/o ingestión de inhibidores de proteasas encontramos la chinche del maíz (*Eurygaster integriceps*) (Saadati & Bandani, 2011), la polilla mediterránea de la harina (*Anagasta kuehniella*) (Rodríguez Macedo et al., 2010), el escarabajo de la papa (*Leptinotarsa decemlineata* Say) (Brunelle, Girard, Cloutier, & Michaud, 2005), *Clavibacter michiganense* subsp. *michiganense*, *Candida albicans* (Kim et al., 2005), *Rhizoctonia solani* (Huang et al., 2010), entre otros.

Sin embargo, uno de los problemas que ha sido detectado en este uso de los inhibidores de proteasas es el desarrollo de mecanismos de resistencia en los organismos diana. Dentro de estos mecanismos de resistencia se encuentran la expresión de proteasas que hidrolizan a los inhibidores o que son insensibles a ellos, o la sobreexpresión de las proteasas diana de los inhibidores (Ahn, Salzman, Braunagel, Koiwa, & Zhu-Salzman, 2004; Bayés et al., 2005; Kiggundu et al., 2010). Una de las alternativas propuestas para revertir o contrarrestar esta resistencia, es el uso de varios inhibidores activos frente a proteasas de diferentes clases mecanísticas (Amirhusin et al., 2007; Santamaria et al., 2012; Senthilkumar, Cheng, & Yeh, 2010). Una variante podría ser el uso de un solo inhibidor con capacidad de inhibir a proteasas de varias clases mecanísticas, en lugar de varios inhibidores.

Los inhibidores de proteasas de naturaleza proteica pueden usar diferentes mecanismos para interactuar con sus enzimas diana. Los inhibidores canónicos forman un complejo con la proteasa, similar al que se forma entre la enzima y su sustrato. Este mecanismo también es conocido como mecanismo estándar o de Laskowski (Laskowski & Kato, 1980). Este mecanismo de inhibición es utilizado por varias familias de inhibidores con plegamientos diferentes, pero que conservan una

conformación similar en el asa que interactúa con el centro activo de la enzima. Esta conformación canónica del asa de interacción con la proteasa, permite que la enzima interactúe con el inhibidor como si fuera un sustrato, formándose incluso el intermediario acil-enzima, pero el paso de hidrólisis es extremadamente lento y la orientación del extremo amino del residuo en posición P1' favorece la re-síntesis del enlace peptídico (Ardelt & Laskowski, 1985; Krowarsch, Cierpicki, Jelen, & Otlewski, 2003; Zakharova, Horvath, & Goldenberg, 2009).

Otro mecanismo empleado para inhibir a las proteasas es el denominado bloqueo estérico. En este caso el inhibidor impide el acceso del sustrato al centro activo. Dentro de los inhibidores que usan este mecanismo se encuentran las cistatinas (familia MEROPS I25), los inhibidores no canónicos de proteasas serino (familia MEROPS I14) y el inhibidor IA3 de proteasas aspártico (familia MEROPS I34) (Otlewski, Jelen, Zakrzewska, & Oleksy, 2005). Este último inhibidor es un caso muy particular, debido a que en ausencia de la enzima el inhibidor no adopta una estructura secundaria definida en solución, mientras que en presencia de la proteasa adopta una conformación de α -hélice que ocupa el centro activo de la sacaropepsina (M. Li et al., 2000). En el caso de los inhibidores no canónicos de proteasas serino la interacción entre la proteasa y el inhibidor involucra regiones externas (exositos) además del centro activo (Grütter et al., 1990).

Otro modo de interacción entre el inhibidor y la proteasa involucra la formación de un complejo enzima-producto. En este mecanismo de inhibición ocurre la hidrólisis del enlace peptídico entre los aminoácidos P1-P1', donde uno de los dos residuos es el extremo amino o carboxilo del inhibidor. El producto resultante no es liberado del centro activo de la enzima inhibiendo de esta forma su actividad catalítica. Entre los inhibidores de naturaleza proteica que usan este mecanismo de inhibición encontramos un inhibidor de carboxipeptidasa A obtenido de papa (*Solanum tuberosum*) (familia MEROPS I37) (Rees & Lipscomb, 1982) y un inhibidor de carboxipeptidasa obtenido de la sanguijuela (*Hirudo medicinalis*) (familia MEROPS I46) (Reverter et al., 2000). También, el inhibidor PI3 aislado de *Ascaris suum* (familia MEROPS I33) emplea este mecanismo de inhibición para con varias proteasas aspártico (Ng et al., 2000).

Todos los mecanismos antes descritos se clasifican como reversibles y, en la mayoría de los casos, son considerados como de unión fuerte. Sin embargo, existen inhibidores, como es el caso de las serpinas (familia MEROPS I4), que forman un intermediario acil-enzima que no puede ser hidrolizado por la proteasa para liberar los productos. Las serpinas inducen un cambio conformacional en la enzima que afecta la viabilidad del centro activo de la misma (Huntington, Read, & Carrell, 2000). La convergencia en términos estructurales y funcionales ha sido descrito en varios inhibidores de proteasas con plegamientos diferentes, lo que para algunos autores podría ser una evidencia de que existen un número limitado de modos de unión entre las proteasas y sus inhibidores (Otlewski et al., 2005).

1.1.3 Inhibidores de unión fuerte.

Los inhibidores de unión fuerte son aquellos cuyas constantes de inhibición por una enzima poseen valores menores a 10^{-7} - 10^{-8} mol L⁻¹. Esta característica cinética, derivada de su función biológica, garantiza una alta eficiencia como inhibidores. Estas moléculas generalmente son análogas del estado de transición y actúan como inhibidores competitivos de sus enzimas dianas. Consecuentemente, esto determina que sean eficientes a concentraciones en el intervalo de las concentraciones de la enzima que inhiben y, por tanto, los principios básicos de la cinética de Michaelis-Menten no se cumplan. De esta manera, los ensayos de actividad inhibitoria, expresados usualmente en porcentajes de inhibición o fracción de inhibición, no pueden reflejar adecuadamente la cantidad real de inhibidor presente en una muestra, ni su efectividad frente a la enzima ensayada. Por otra parte, la determinación de K_i no puede realizarse de acuerdo a los principios de la inhibición competitiva clásica y requiere de un tratamiento particular. Este parámetro cinético, en última instancia, es el único que posibilitará describir la eficiencia real de un inhibidor de unión fuerte frente a la enzima ensayada.

Las concentraciones efectivas de los inhibidores de unión fuerte están en el intervalo de las concentraciones de enzima utilizadas en los ensayos. Por lo tanto, al igual que se procede habitualmente con la enzima, se debe incluir una ecuación de conservación de masas para el inhibidor: $[I]_0 = [I] + [EI]$ (donde $[I]_0$: concentración inicial del inhibidor, $[I]$: concentración de inhibidor libre en el equilibrio, $[EI]$: concentración del complejo enzima-inhibidor en el equilibrio). Estas condiciones conducen a una ecuación cinética diferente a la ecuación clásica de Michaelis-Menten para la inhibición competitiva (Bieth, 1995). Además, el comportamiento gráfico de la curva de fracción de actividad enzimática residual ($a = v_i/v_0$) frente a la concentración de inhibidor inicial ($[I]_0$) depende de la relación entre la concentración inicial de enzima ($[E]_0$) y la constante de inhibición aparente (K_i^{ap}). A valores de $[E]_0/K_i^{ap}$ por encima de 10, la curva adopta un comportamiento bifásico, con una región lineal que significa que la unión es tan fuerte que prácticamente todo el inhibidor está en la forma combinada EI, es decir, el inhibidor es capaz de titular a la enzima. Este comportamiento se produce experimentalmente en presencia de un inhibidor de unión fuerte reversible o pseudo irreversible. Estas condiciones son las empleadas generalmente para la determinación de las concentraciones de enzima o inhibidor activo en la muestra (Bieth, 1995).

El procedimiento para determinar el valor de K_i requiere del ajuste de las condiciones experimentales para obtener una curva cóncava de $a(v_i/v_0)$, en función de concentraciones variables de inhibidor a una concentración constante de enzima y sustrato. Este comportamiento se produce en presencia de un inhibidor de unión fuerte reversible y se logra con condiciones en el ensayo de una relación de $[E]_0/K_i \leq 10$ y $[S] \approx K_M$ (Bieth, 1995; Copeland, 2000; Murphy, 2004). Este comportamiento se describe

por la ecuación general de la inhibición de unión fuerte planteada por Morrison (1969), que muestra claramente que tanto la K_i , como la $[E]_0$, son factores importantes en la interacción enzima-inhibidor. El valor de una curva cóncava de a en función de la $[I]_0$ reside en que demuestra la reversibilidad de la inhibición y la posibilidad de determinar el valor de K_i mediante esta ecuación. Sin embargo, el valor de K_i que se determina es aparente y su valor real solo se puede calcular, teniendo en cuenta el efecto del sustrato sobre la inhibición, el que varía de acuerdo al tipo de inhibición (Copeland, 2000).

1.1.4 Inhibidores de la familia Kunitz-STI

Una de las familias de inhibidores de proteasas de naturaleza proteica con mayor versatilidad respecto a su capacidad de inhibir diferentes tipos de proteasas es la familia Kunitz-STI (familia MEROPS I3). Entre las proteasas diana de los inhibidores de esta familia se encuentran proteasas de tipo serino como la tripsina (EC 3.4.21.4), la quimiotripsina (EC 3.4.21.1), la subtilisina (EC 3.4.21.62) y la elastasa pancreática (EC 3.4.21.36); de tipo aspartico como la catepsina D (EC 3.4.23.5), la sacaropepsina (EC 3.4.21.41), la memapsina-2 humana (β -secretasa) (3.4.23.46) y de tipo cisteino como la papaína (EC 3.4.22.2) y la catepsina K (EC 3.4.22.38) (Cater et al., 2002; De Meester, Brick, Lloyd, Blow, & Onesti, 1998; Fischer, Kuckenberg, Kastilan, Muth, & Gebhardt, 2015; Hansen et al., 2007; Krauchenco, Pando, Marangoni, & Polikarpov, 2003; Mares et al., 1989). Además, se ha informado de su capacidad de inhibir a las α -amilasas y la sucrosa invertasa (Glaczinski, Heibges, Salamini, & Gebhardt, 2002; Song et al., 1999). Esta versatilidad en cuanto a enzimas diana hace de estas proteínas una plataforma atractiva para el desarrollo de nuevas moléculas bi- y/o multifuncionales.

Los inhibidores Kunitz-STI presentan un plegamiento β -*trefoil* que está compuesto por 12 hebras β y en algunos casos pueden presentar de 1 a 3 hélices (Renko, Sabotič, & Turk, 2012). Del total de hebras β , seis interactúan de forma antiparalela para formar un barril β , mientras que las otras seis conforman 3 horquillas β antiparalelas que se ubican formando una especie de tapa del barril β (Figura 2A). El plegamiento β -*trefoil* presenta una pseudo-simetría o simetría interna de tercer orden, compuesta por tres subdominios (módulos o motivos estructurales) conformados por 4 hebras β antiparalelas, dos de ellas formando parte del barril β y las otras dos de la tapa del barril (Figura 2B). Este plegamiento presenta 10 asas con longitudes que pueden variar desde 2 hasta 16 aminoácidos. El plegamiento β -*trefoil* también es adoptado por otras familias de proteínas, algunas con actividad inhibidora de proteasas como las clitocipinas (familia MEROPS I48) y macrocipinas (Familia MEROPS I84), además de otras funciones como citocinas (Murzin, Lesk, & Chothia, 1992; Renko et al., 2010).

De forma general, los inhibidores Kunitz-STI son proteínas de entre 170-200 aminoácidos, una masa molecular de alrededor de 20 kDa y con un número variable de puentes disulfuro (Oliva, Silva, Sallai, Brito, & Sampaio, 2010). Los puentes disulfuro más conservados se encuentran conectando las asas L2 y L5 y los extremos de las hebras $\beta 9$ y $\beta 10$. Existen algunos casos con una topología de puentes diferentes a los observados en el inhibidor STI y en la mayoría de los miembros de la familia (Oliva et al., 2010). Existen otros casos donde incluso en ausencia de puentes disulfuro se conservan la estructura tridimensional y la actividad inhibidora (Hansen et al., 2007; Zhou et al., 2015). Varios estudios han confirmado que las regiones responsables de la actividad inhibidora de proteasas se encuentran en las asas que conectan las hebras β , los que a su vez presentan una gran variabilidad de longitud y secuencia. Esta variabilidad ha sido sugerida como la responsable de la versatilidad de la actividad inhibidora de los miembros integrantes de la familia Kunitz-STI (Azarkan et al., 2011; Renko et al., 2012).

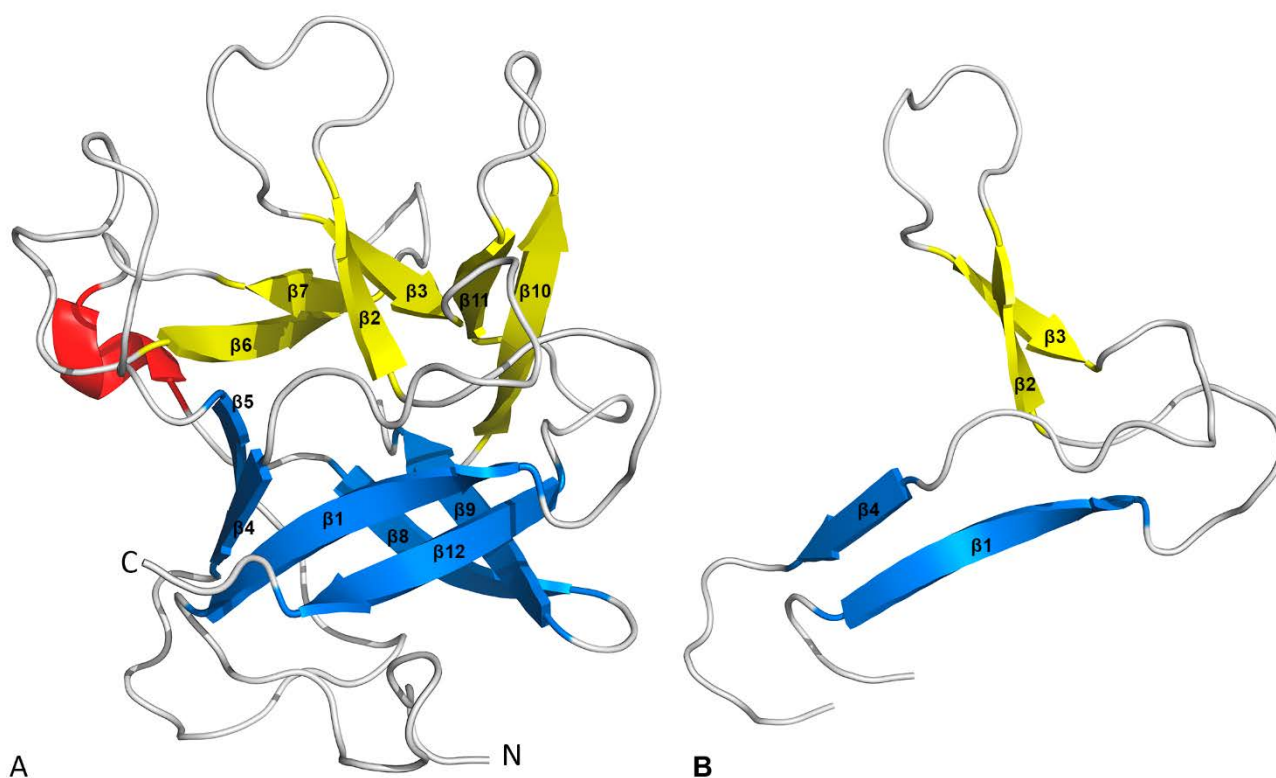


Figura 2. Estructura general de los inhibidores Kunitz-STI (A). En azul se representan las hebras β que componen el barril β , mientras en amarillo aquellas que se organizan en una especie de tapa de uno de los extremos del barril. Uno de los 3 subdominios o unidad estructural que compone el plegamiento β -trefoil (B).

La mayoría de las estructuras tridimensionales disponibles para inhibidores Kunitz-STI son de moléculas capaces de inhibir proteasas de tipo serino de las familia S1 y S8, como tripsina,

quimotripsina y subtilisina (De Meester et al., 1998; Krauchenco et al., 2003; Ravindranathan, Gallicchio, & Levy, 2005; Song & Suh, 1998), además de algunos casos de inhibición de proteasas de tipo cisteíno (Hansen et al., 2007) y de inhibidores bifuncionales que inhiben subtilisina y α -amilasas (Micheelsen et al., 2008; Vallée et al., 1998). También existen numerosos estudios estructurales sobre el mecanismo de inhibición de tripsina por esta familia de inhibidores (Bao et al., 2009; Khamrui, Majumder, Dasgupta, Dattagupta, & Sen, 2010; Patil, Chaudhary, Sharma, Tomar, & Kumar, 2012). En la mayoría de los casos, la interacción del inhibidor con la proteasa ocurre a través de un asa expuesto al solvente con una conformación descrita como “canónica”. La secuencia de las asas canónicas no se encuentra conservada, aunque existe preferencia por determinados aminoácidos en la posición P1, en función de la proteasa diana. En la mayoría de los inhibidores Kunitz-STI esta asa canónica se encuentra entre las hebras β 4 y β 5, de ahí que esta es la región que primero se inspecciona en cada nueva estructura de un inhibidor de esta familia. La estabilización de esta asa canónica se logra de varias maneras. En algunos inhibidores como el STI, existe una asparagina, N13 (la posición varía para otros inhibidores), cuya cadena lateral interactúa con la cadena principal de los aminoácidos en el asa canónica a través de puentes de hidrógeno (Dasgupta, Khamrui, Dattagupta, & Sen, 2006). En otros inhibidores descritos como canónicos, la estabilización del asa canónica ocurre a través de interacciones con una molécula de agua como es el caso del inhibidor de flamboyán (*Delonix regia*), DrTI (Krauchenco et al., 2003), aunque en realidad en este inhibidor el asa descrita como canónica se desvía de la conformación adoptada por otros miembros de la familia como el STI. En el inhibidor API-A, que posee dos sitios de unión a tripsina, la conformación del asa canónica entre las hebras β 9 y β 10 (L9) es estabilizada por la presencia de dos puentes disulfuro (Figura 3A) (Bao et al., 2009).

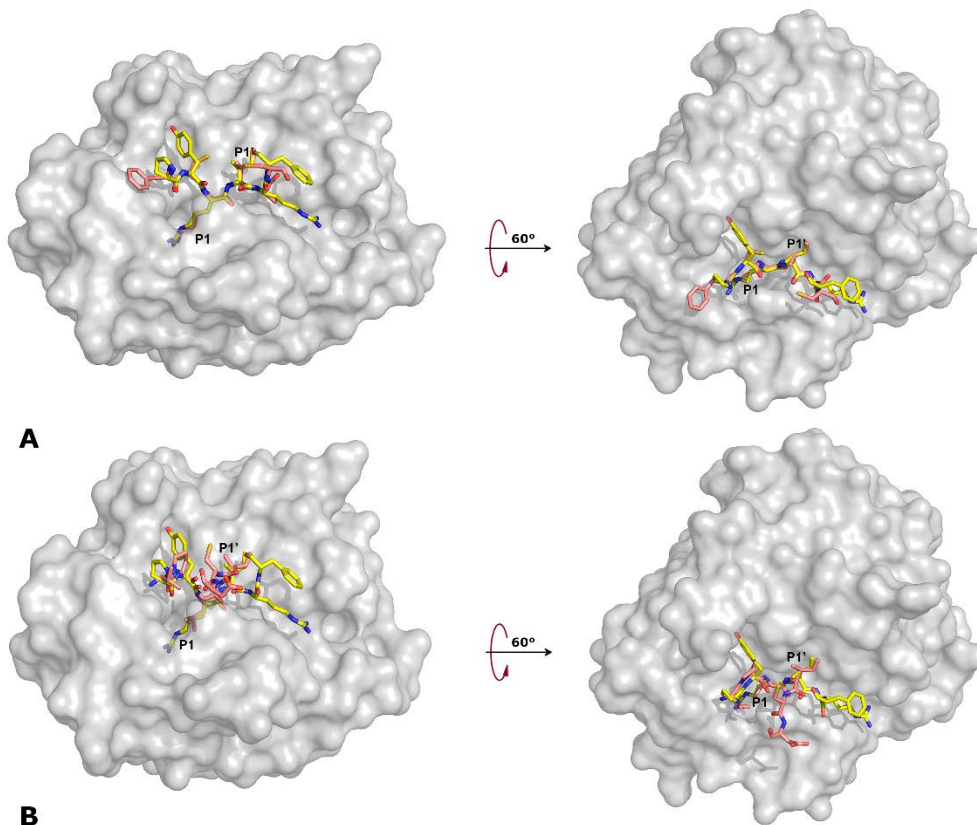


Figura 3. Conformaciones canónicas y no canónicas del asa de interacción de los inhibidores de serino proteasas de la familia Kunitz-STI. Comparación de la conformación canónica de las asas L9 del inhibidor API-A (3E8L, salmón) y L4 de STI (1AVW, amarillo) en el complejo tripsina-inhibidor (A). Comparación de la conformación no canónica del asa L5 del inhibidor API-A (3E8L, salmón) con la conformación canónica del asa L4 de STI (1AVW, amarillo) (B).

La interacción de los inhibidores canónicos con las proteasas serino también ha sido descrita como similar a la de un sustrato, por el hecho de que se forma rápidamente un intermediario covalente, igual a como sucede con los sustratos. Sin embargo, la hidrólisis de este intermediario resultante es muy lenta lo que promueve la reversibilidad del proceso y provoca que la enzima quede atrapada en un ciclo de hidrólisis/re-síntesis de este enlace peptídico (Michael Laskowski & Qasim, 2000; Radisky & Koshland, 2002; Zakharova et al., 2009). Por otro lado, existe al menos un caso de un inhibidor perteneciente a la familia Kunitz-STI que interacciona con la proteasa tripsina a través de un asa que adopta una conformación no canónica (Figura 3B) (Bao et al., 2009). Este sitio de interacción ha sido validado estructuralmente y por estudios de mutagénesis sitio-específica (Bao et al., 2009). Las diferencias en conformación entre ambos tipos de asas se encuentran fundamentalmente en las posiciones P2'-P3' (Figura 3).

1.1.5 Inhibidores bifuncionales de la familia Kunitz-STI. Inhibidores de proteasas serino y aspártico.

Dentro de la familia Kunitz-STI, existe un subgrupo de proteínas que han mostrado capacidad de inhibir tanto a proteasas de tipo serino como aspártico (Cater et al., 2002; Fischer et al., 2015; Heibges, Salamini, & Gebhardt, 2003; Mares et al., 1989). Estos inhibidores bifuncionales han sido aislados hasta el momento de papa (*S. tuberosum*) (Fischer et al., 2015; Heibges, Salamini, et al., 2003; Mares et al., 1989; Ritonja et al., 1990) y tomate (*Solanum lycopersicum*) (Cater et al., 2002). No obstante, existen proteínas anotadas como tales en otras especies relacionadas, aunque sin estudios funcionales que demuestren su actividad inhibidora frente a ambas familias de proteasas. Los primeros inhibidores bifuncionales de proteasas aspártico y serino inhibían a tripsina y catepsina D, de ahí que en muchos casos se les denomine como inhibidores de catepsina D (Keilová H. and Tomášek, 1976a, 1976b; Mares et al., 1989). Sin embargo, trabajos posteriores demostraron que también son capaces de inhibir a otras proteasas aspárticas de la familia A1, como por ejemplo sacaropepsina (Cater et al., 2002), memapsina-2 humana (Fischer et al., 2015) y Aspergillopepsin I (Heibges, Salamini, et al., 2003), aunque con diferentes niveles de potencia e especificidad.

La determinación de la secuencia de aminoácidos de estos inhibidores bifuncionales mostró una baja similitud con la de los inhibidores típicos de la familia (~25 % de identidad de secuencia con STI). En papa en específico, se han identificado un gran número de inhibidores Kunitz-STI, algunos de ellos descritos como inhibidores bifuncionales con base en su similitud de secuencia con los que están caracterizados bioquímicamente. En este organismo, los inhibidores Kunitz-STI se han clasificado en grupos que pueden variar en número, dependiendo de cada autor (Bauw et al., 2006; Heibges, Glaczinski, Ballvora, Salamini, & Gebhardt, 2003). De forma general los inhibidores de proteasas aspártico pertenecen al grupo A, aunque recientemente se informó que algunos inhibidores del grupo B son capaces de inhibir a la proteasa memapsin-2 (Fischer et al., 2015).

En contraste con los numerosos estudios sobre el mecanismo de inhibición de proteasas de tipo serino, no existe información estructural sobre el mecanismo de inhibición de estos inhibidores bifuncionales sobre las proteasas aspártico. De las seis familias de inhibidores de proteasas existentes en la base de datos MEROPS con miembros capaces de inhibir proteasas aspártico, solo tres de ellas poseen al menos un miembro con estructura tridimensional determinada (Headey et al., 2010; M. Li et al., 2000; Ng et al., 2000). Dos de estos casos son estructuras cristalográficas de complejos proteasa-inhibidor, una de ellas de un inhibidor aislado de *Saccharomyces cerevisiae* (IA₃) y la proteasa sacaropepsina (M. Li et al., 2000), y la otra es de un inhibidor aislado de *Ascaris sum* (PI3) que forma un complejo con la pepsina porcina (EC 3.4.23.1) (Ng et al., 2000). El tercer caso es una estructura en solución obtenida por resonancia magnética nuclear de un inhibidor aislado de *Cucurbita máxima* (SQAPI), perteneciente a la familia MEROPS I55 (Headey et al., 2010). En el caso

de SQAPI, estudios de mutagénesis sitio-dirigida, en combinación con modelos de anclaje molecular han permitido proponer un posible mecanismo de inhibición (Headey et al., 2010). Sin embargo, estos inhibidores poseen plegamientos diferentes y no tienen similitud en secuencia entre sí, ni con los miembros de otras familias de inhibidores de proteasas aspártico, lo que imposibilita extrapolar cualquier característica del mecanismo de inhibición a otros inhibidores.

Como antecedente directo del presente trabajo, nuestro grupo utilizó una combinación de herramientas computacionales para predecir la posible región de interacción de estos inhibidores bifuncionales con las proteasas de tipo aspártico. En ese trabajo se empleó como modelo un inhibidor aislado de tomate (SLAPI) y expresado de forma heteróloga en *Pichia pastoris* (Cater et al., 2002; Guerra et al., 2012). Los resultados derivados de este análisis sugieren que el asa L9 ($\beta 9$ - $\beta 10$) podría ser la región de interacción con la sacaropepsina, y por lo tanto con las proteasas aspártico. Adicionalmente, se propusieron aquellas posiciones en la secuencia del asa L9 que podrían ser las responsables de la diferencia en afinidad informada anteriormente entre el inhibidor SLAPI y un inhibidor de papa (Cater et al., 2002). Los modelos propuestos en ese trabajo no permitieron proponer un modo de interacción con la proteasa, debido a la posibilidad, al menos *in silico*, de interactuar a través de diferentes mecanismos de inhibición (Guerra et al., 2012). Previamente, se había sugerido la posibilidad de que el asa L9 pudiera estar implicada en la inhibición de las proteasas aspárticas, basados fundamentalmente en las diferencias en la secuencia de aminoácidos con STI (Ritonja et al., 1990). Recientemente se publicó una estructura tridimensional de un inhibidor bifuncional capaz de inhibir proteasas de tipo serino y aspártico y los autores también proponen entre los posibles sitios de interacción con las proteasas de tipo aspártico el asa L9, aunque los modelos resultantes del anclaje molecular proponen al asa L5 (Guo et al., 2015).

2. Justificación

Los inhibidores de proteasas de la familia Kunitz-STI presentan una gran versatilidad en cuanto a las proteasas que son capaces de inhibir. Esa versatilidad está relacionada con la variabilidad en longitud y secuencia de las asas involucradas en la interacción con las proteasas, así como con las características estructurales del plegamiento β -trefoil que presentan. La identificación de las regiones involucradas con las diferentes clases mecánicas de proteasas permitiría el diseño de nuevas moléculas bifuncionales y/o multifuncionales a través de la manipulación de las regiones responsables de interactuar con cada tipo de proteasas. Sin embargo, no existe información estructural ni funcional de las regiones involucradas en la inhibición de las proteasas aspárticas en los inhibidores bifuncionales de proteasas serino y aspártico. Trabajos anteriores proponen posibles regiones de interacción, basadas en análisis con herramientas computacionales. En el presente trabajo se determinaron y analizaron varias estructuras cristalográficas de un inhibidor bifuncional de la familia Kunitz-STI, las que también se compararon con una estructura descrita recientemente para un inhibidor con alta identidad en secuencia y con los otros miembros de la familia Kunitz-STI. Además, los resultados derivados de este trabajo permitieron ofrecer una nueva perspectiva acerca de los mecanismos de inhibición de los inhibidores bifuncionales.

2.1 Hipótesis

La determinación de la estructura tridimensional de un inhibidor bifuncional capaz de inhibir a proteasas serino y aspártico nos permitirá proponer los posibles sitios de interacción con estas proteasas.

2.2 Objetivo

Proponer los sitios de interacción con las proteasas serino y aspártico de un inhibidor bifuncional de la familia Kunitz-STI, basados en el análisis de su estructura tridimensional.

2.3 Objetivos Particulares

- Expresar el inhibidor bifuncional E3Ad de *Solanum tuberosum* en *Pichia pastoris*.
- Obtener un preparado homogéneo del inhibidor E3Ad.
- Caracterizar cinéticamente el inhibidor E3Ad.
- Determinar y analizar la estructura cristalográfica del inhibidor para proponer posibles sitios de interacción con las proteasas serino y aspártico.
- Obtener y caracterizar cinéticamente mutantes del posible sitio de unión.

3. Materiales y Métodos

Los reactivos que se emplearon en el presente trabajo fueron de Sigma-Aldrich (EUA), grado analítico, excepto en los casos que se indique lo contrario. Las cepas de *Escherichia coli* y *Pichia pastoris* provienen de Invitrogen™ Life Technologies (EUA).

3.1 Clonaje de las secuencias codificadoras de los inhibidores de proteasas.

Las secuencias codificantes de los inhibidores de proteasas se amplificaron a partir de ADN genómico de *Solanum tuberosum*, variedades Estima y Maris piper, en el laboratorio del Dr. Colin Berry de la Escuela de Biociencias de la Universidad de Cardiff en el Reino Unido. Las secuencias amplificadas fueron insertadas en el vector pGEMT (Promega, EUA) y secuenciadas en la unidad de secuenciación de la Universidad de Cardiff, Reino Unido. Entre las secuencias obtenidas, la correspondiente a la clona E3Ad, la cual no contiene intrones y coincide con las entradas del GenBank XP_006362668.1 y XM_006362606.2, fue seleccionada para su expresión. Con el propósito de introducir los sitios de corte para las enzimas de restricción *Xho* I y *Sac* II, la secuencia codificante E3Ad fue re-amplificada usando la enzima ADN polimerasa *PfuUltra* High-Fidelity (Agilent Technologies, EUA) y los siguientes cebadores (subrayados se indican los sitios de corte para las enzimas de restricción *Xho* I y *Sac* II):

Directo: 5' TCTCTCGAGAAGAGAGAATCTCCTCTACCTAAGCCGGTACTTG3'

Reverso: 5' ATAGTTCCGCGGTCAAACCTTCCTGGAATAAGACATCAAGAGGATTTTCG3'

Las condiciones empleadas en la reacción de amplificación fueron: un ciclo a 91 °C durante 3 minutos; 30 ciclos de 30 segundos a 91 °C, 30 segundos a 65 °C y 40 segundos a 72 °C y finalmente un ciclo de 5 minutos a 72 °C. Los productos de la reacción de amplificación fueron analizados mediante electroforesis en gel de agarosa (1.5 %) en TBE (Tris 89 mmol L⁻¹, ácido bórico 89 mmol L⁻¹, EDTA 2 mmol L⁻¹). El producto de PCR y el vector de expresión pPICZαC (Invitrogen™ Life Technologies, EUA) fueron incubados con las enzimas de restricción *Xho*I y *Sac*II (Promega, EUA), siguiendo las instrucciones del fabricante, y los productos resultantes se purificaron con el kit Wizard® SV Gel and PCR Clean-up System (Promega, EUA). Para la reacción de ligación se mezclaron diferentes proporciones molares de vector e inserto (1:3 y 1:1) en presencia de la enzima T4 ligasa (Promega, EUA) y se incubó toda la noche a 4 °C. Se utilizaron 2 µL de esta mezcla de reacción para transformar 50 µL de células quimiocompetentes de *E. coli* DH5α™ (Invitrogen™ Life Sciences, EUA). Se utilizó Zeocina™ (Invitrogen™ Life Sciences, EUA) a 25 µg mL⁻¹ en medio LB

(Tryptona 1 %, extracto de levadura 0.5 %, NaCl 1 %) bajo en sales para la selección de las células transformadas.

La presencia del inserto, así como su orientación, se verificaron mediante secuenciación de ADN con la compañía MacroGen Inc. (Seúl, Corea del Sur), usando los cebadores 5'AOX1 y 3' AOX1 (Invitrogen™ Life Technologies, EUA).

3.2 Obtención de los mutantes E3Ad_N19D y E3Ad_N19D_R151A, y sus variantes con una etiqueta de His6x.

Para la construcción del mutante E3Ad_N19D se utilizó el kit *QuickChange MultiSite-Directed Mutagenesis* (Agilent Technologies, EUA), usando como molde la construcción pPICZαC-E3Ad obtenida previamente y el cebador 5'GAACTCAATCCTGATTCGAGTTATCGCATTATTTCCATTGG 3' (base subrayada indica la mutación introducida). La presencia de la mutación deseada se comprobó mediante secuenciación de ADN en la unidad de secuenciación del Instituto de Biotecnología (US-IBt, UNAM), usando los cebadores 5' AOX1 y 3' AOX1 (Invitrogen™ Life Technologies, EUA). Para obtener la forma lineal del plásmido pPICZαC-E3Ad_N19D, se utilizó la enzima de restricción *Sac I* (Invitrogen™ Life Technologies, EUA), siguiendo las instrucciones del fabricante. La preparación y transformación de células quimiocompetentes de *Pichia pastoris* X33 se realizó mediante el kit *Pichia EasyComp™* (Invitrogen™ Life Technologies, EUA). La presencia del inserto en el genoma de *P. pastoris* se verificó mediante PCR de colonias, usando los cebadores 5'AOX1 y 3'AOX1 (Invitrogen™ Life Technologies, EUA).

El doble mutante E3Ad_N19D_R151A se obtuvo usando el plásmido pPICZαC-E3Ad_N19D como molde y los siguientes cebadores (subrayado se indican las mutaciones introducidas):

Directo: 5'CTTCTATTATTTGTCTAGCTTGCCCTGAGGATCAGTTTTG3'

Reverso: 5'CAA^AACTGATCCTCAGGGCAAGCTAGACAAATAATAGAAG3'

Las condiciones de la reacción de PCR empleando la enzima ADN polimerasa Phusion® (New England Biolabs, EUA) fueron: un ciclo de 30 segundos a 95 °C; 18 ciclos de 30 segundos a 95 °C, 60 segundos a 55 °C, 5 minutos a 68 °C y finalmente un ciclo de 5 minutos a 68 °C. Posteriormente, las reacciones fueron incubadas en hielo por 2 minutos. Seguidamente, se añadieron 10 U de la enzima de restricción *Dpn I* (Agilent Technologies, EUA) a la mezcla de reacción y se incubó por 90 minutos a 37 °C. Se utilizaron 2 μL de esta mezcla para transformar 50 μL de células quimiocompetentes de *E. coli* DH5α™. Se utilizó Zeocina™ (25 μg mL⁻¹) en medio LB bajo en sales para la selección de las células transformadas. La presencia de la mutación deseada se verificó mediante secuenciación de ADN en la Unidad de Síntesis y Secuenciación de ADN del Instituto de

Biología (US-IBt, UNAM), usando los cebadores 5' AOX1 y 3' AOX1 (Invitrogen™ Life Technologies, EUA). La transformación de *P. pastoris* X33 se realizó de forma similar a lo descrito anteriormente para E3Ad_N19D.

Para la obtención de una construcción de los inhibidores mutantes E3Ad_N19D y E3Ad_N19D_R151A con una etiqueta de histidinas (His6x) en su extremo carboxilo, se amplificaron sus secuencias a partir de los plásmidos pPICZαC-E3Ad_N19D y pPICZαC-E3Ad_N19D_R151A usando los siguientes cebadores (subrayados se indican los sitios de corte para las enzimas de restricción *Xho* I y *Sac* II):

Directo: 5' TCTCTCGAGAAGAGAGAATCTCCTCTACCTAAGCCGGTACTTG3'

Reverso:

5' GTTCCGCGGTCAATGATGATGATGATGATGAACTTCCTGGAATAAGACATCAAGAGGATTTTCG3'

Las condiciones de la reacción de PCR empleando la enzima ADN polimerasa P husion® (New England Biolabs, EUA) fueron: un ciclo de 30 segundos a 98 °C; 30 ciclos de 10 segundos a 98 °C, 30 segundos a 65 °C, 30 segundos a 72 °C y finalmente un ciclo de 5 minutos a 72 °C. Para el subclonaje en el vector de expresión pPICZαC se procedió de forma similar a lo descrito anteriormente para la secuencia de E3Ad en la sección 3.1.

3.3 Expresión heteróloga de los inhibidores

Se inocularon 40 mL de medio de cultivo BMGY (triptona 2 %, extracto de levadura 1 %, K_2HPO_4/KH_2PO_4 0.1 mol L⁻¹ pH 6.0, base nitrogenada para levaduras 1.34 %, biotina 40 μmol L⁻¹ y glicerol 1 %) complementado con ampicilina a una concentración final de 100 μg mL⁻¹ y se incubaron a 30 °C y 200 rpm de agitación hasta alcanzar una absorbancia a 600 nm entre 6-8. Se recuperaron las células por centrifugación a 5000 x g durante 10 minutos a temperatura ambiente. Las células se lavaron con 30 mL de medio de cultivo BMMY (triptona 2 %, extracto de levadura 1 %, K_2HPO_4/KH_2PO_4 0.1 mol L⁻¹ pH 6.0, base nitrogenada para levaduras 1.34 %, biotina 40 μmol L⁻¹ y metanol 1 %) y se recuperaron bajo las condiciones de centrifugación descritas anteriormente. Posteriormente, se re-suspendieron en un volumen final de medio de cultivo BMMY, complementado con ampicilina a una concentración final de 100 μg mL⁻¹, suficiente para obtener una absorbancia a 600 nm igual a 1. El cultivo se incubó a 28 °C y 150 rpm por 72 horas, con reposición del metanol (1 %) cada 24 horas. Al cabo de las 72 horas, se separaron las células del sobrenadante mediante centrifugación a 5000 x g durante 10 minutos a temperatura ambiente. El pH del sobrenadante se ajustó a 4.0 con ácido fórmico y se dializó contra una solución amortiguadora de formiato de sodio 10 mol L⁻¹ pH 4.0 a una temperatura de 4 °C. Se realizaron varios cambios de la solución

amortiguadora hasta alcanzar una relación de volúmenes 100:1. Se añadió cloruro de sodio a la solución dializada, hasta una concentración final de 150 mmol L⁻¹ y se eliminó cualquier material insoluble mediante centrifugación a 10000 x g y 4 °C, durante 30 minutos.

Para los inhibidores E3Ad_N19D_His6x y E3Ad_N19D_R151A_His6x, la expresión se realizó de la misma manera descrita para las variantes sin la etiqueta de histidinas. El sobrenadante de cultivo se diluyó al doble de su volumen y se ajustó a una concentración final de NaCl e imidazol igual a 0.2 mol L⁻¹ y 40 mmol L⁻¹, respectivamente. El pH de la solución resultante se ajustó en dependencia del inhibidor presente en cada caso, teniendo en cuenta el valor de su pI teórico y se mantuvo en agitación por 16 horas a una temperatura entre 18-20 °C. Para el inhibidor E3Ad_N19D_His6x el pH se ajustó a 7.2, mientras que para E3Ad_N19D_R151A_His6x fue de 8.0. Las soluciones fueron centrifugadas a 10000 x g, 18 °C, durante 30 minutos previo a su purificación mediante cromatografía de afinidad a Ni²⁺.

3.4 Purificación de los inhibidores

Para las construcciones con el extremo carboxilo nativo, la solución resultante de la diálisis y ajustada con cloruro de sodio (150 mmol L⁻¹) se aplicó a una columna de intercambio catiónico HiPrep™ SP Sepharosa™ HP 16/10 (GE Healthcare Life Sciences, EUA), previamente equilibrada con una solución amortiguadora de formiato de sodio 10 mmol L⁻¹, NaCl 150 mmol L⁻¹ pH 4.0, acoplada a un sistema cromatográfico Äktapure™ (GE Healthcare Life Sciences, EUA). La elución se realizó mediante un gradiente lineal de 150 a 490 mmol L⁻¹ de NaCl en formiato de sodio 10 mmol L⁻¹ pH 4.0 a un flujo de 60 cm h⁻¹. Las fracciones colectadas durante la elución se analizaron mediante una electroforesis SDS-PAGE (Laemmli, 1970), bajo condiciones desnaturizantes. Las electroforesis se realizaron con amortiguador Tris-glicina y a una corriente constante de 25 mA por cada gel. Las fracciones que contenían el inhibidor se concentraron usando un dispositivo de concentración por centrifugación Amicon Ultra-15 (Millipore, EUA), con límite de exclusión de 10 kDa. En el caso del mutante E3Ad_N19D, las fracciones del intercambio iónico fueron dializadas en una solución amortiguadora de formiato de sodio 10 mmol L⁻¹ pH 3.0, previo a su concentración.

La muestra concentrada se aplicó a una columna de exclusión molecular Superdex™ 75HR 10/30 (GE Healthcare Life Sciences, EUA), previamente equilibrada con la solución amortiguadora formiato de sodio 10 mmol L⁻¹, NaCl 150 mmol L⁻¹ pH 4.0 (pH 3.0 para E3Ad_N19D), acoplada a un sistema cromatográfico Äktapure™ 25 (GE Healthcare Life Sciences, EUA). La cromatografía se realizó a un flujo lineal de 15.3 cm h⁻¹ y las fracciones colectadas se evaluaron mediante SDS-PAGE (Laemmli, 1970), bajo las mismas condiciones descritas anteriormente. Las fracciones que contenían el inhibidor se diafiltraron con la solución amortiguadora formiato de sodio 10 mmol L⁻¹ pH 4.0 (pH 3.0

para E 3Ad_N19D), usando un dispositivo de concentración por centrifugación Amicon Ultra-4 (Millipore, EUA), con límite de exclusión de 10 kDa. Posteriormente, fueron concentradas hasta un valor de entre 10-15 mg mL⁻¹. Para el cálculo de la concentración de proteínas se utilizó un coeficiente de extinción molar a 280 nm igual a 16 305 L mol⁻¹ cm⁻¹, obtenido a través del programa ProtParam (<http://web.expasy.org/protparam/>), utilizando la secuencia de aminoácidos de los inhibidores.

Para los inhibidores que presentaron una etiqueta de His_{6x} en su extremo carboxilo, el sobrenadante de cultivo clarificado se aplicó a una columna de afinidad a Ni²⁺ HisTrap™ HP (5mL) (GE Healthcare Life Sciences, EUA), previamente equilibrada con un amortiguador NaH₂PO₄/Na₂HPO₄ 50 mmol L⁻¹, NaCl 0.2 mol L⁻¹, imidazol 40 mmol L⁻¹ pH 7.0-8.0 (en dependencia del inhibidor a purificar). La elución se realizó mediante un gradiente lineal de 40 a 500 mmol L⁻¹ de imidazol en un volumen de 100 mL a un flujo de 45 cm h⁻¹. Las fracciones colectadas se evaluaron mediante SDS-PAGE (Laemmli, 1970), bajo las mismas condiciones descritas anteriormente. Las fracciones que contenían al inhibidor se unieron y se diafiltraron con la solución amortiguadora NaAc 50 mmol L⁻¹, NaCl 100 mmol L⁻¹ pH 5.0, usando un dispositivo de concentración por centrifugación Amicon Ultra-4 (Millipore, EUA), con límite de exclusión de 10 kDa. Posteriormente, fueron concentradas hasta un valor de entre 5-10 mg mL⁻¹. La muestra concentrada se aplicó a una columna de exclusión molecular Superdex75HR 10/30, previamente equilibrada con la solución amortiguadora NaAc 50 mmol L⁻¹, NaCl 100 mmol L⁻¹ pH 5.0. La cromografía se realizó a un flujo lineal de 15.3 cm h⁻¹ y las fracciones colectadas se evaluaron mediante SDS-PAGE (Laemmli, 1970). Las fracciones con presencia del inhibidor se unieron y se concentraron hasta un valor de entre 10-15 mg mL⁻¹. Para el cálculo de la concentración de proteínas se utilizó un coeficiente de extinción molar a 280nm igual a 16 305 L mol⁻¹ cm⁻¹, obtenido a través del programa ProtParam.

3.5 Electroforesis de isoelectroenfoque.

Para la determinación del punto isoeléctrico de los inhibidores, las muestras purificadas se ajustaron a una concentración máxima de 1 mg mL⁻¹ y se aplicó 1 µL de muestra en un gel PhastGel IEF 3-9 (GE Healthcare Life Sciences, EUA). La electroforesis se realizó en un sistema Phastsystem™ (Pharmacia, Suecia), bajo las condiciones sugeridas por el fabricante. Para la detección de las bandas de proteína se utilizó PhastGel Blue Coomassie R (0.2 %) en una mezcla metanol (30 %) y ácido acético (10 %).

3.6 Análisis por espectrometría de masas.

Para el análisis de las muestras por espectrometría de masas ESI-MS-MS, el inhibidor E3Ad purificado se aplicó a un gel de SDS-PAGE al 15 % bajo las mismas condiciones descritas anteriormente. Las bandas se detectaron mediante tinción con azul Coomassie R-250 (2.5 %) en una solución de metanol (50 %) y ácido acético (10 %). Para la decoloración de los geles se empleó una mezcla de metanol (12.5 %) y ácido acético (10 %). Las dos bandas mayoritarias se recortaron y se enviaron para su análisis a la Unidad Universitaria de Proteómica del Instituto de Biotecnología de la UNAM. Las muestras se sometieron a secuenciación ESI-MS-MS, detección de glicosilaciones y caracterización de las mismas.

3.7 Cristalización de los inhibidores

Se evaluaron diferentes condiciones de cristalización del inhibidor E3Ad y del mutante E3Ad_N19D usando las matrices de condiciones de cristalización Wizard I-IV (Emerald Biosciences), PEG/Ion, Crystal Screen, Index, Crystal Screen Cryo (Hampton Research, EUA) y los métodos de microbatch y gota sedente (asentada). En el método de microbatch se mezclaron 1 μ L de la solución del inhibidor con 1 μ L de la solución de cristalización y se cubrieron con 12 μ L de aceite de parafina (Hampton Research, EUA). Se prepararon pruebas de cristalización a temperaturas de 4 y 18 °C. En el caso del método de gota apoyada se utilizó un robot de cristalización TTP Labtech's *Mosquito*®LCP (TTP Labtech, Reino Unido) para preparar gotas compuestas por 100 nL de la solución de proteína más 100 nL de la solución de cristalización. Se dispensaron 40 μ L de la solución de cristalización en el pocillo correspondiente a cada gota. Estas pruebas se realizaron a 18 °C.

El proceso de optimización de las condiciones de cristalización iniciales se realizó en cajas de 24 pozos, usando el método de la gota apoyada. Se evaluaron diferentes concentraciones de agente precipitante y de los otros componentes de la condición de cristalización, además del pH y la concentración de proteína. También, se utilizaron las soluciones del Additive Screen™ (Hampton Research, EUA).

Los cristales más promisorios se remojaron con una solución criopreservante (condición de cristalización con 30-40 % glicerol, etilenglicol o PEG 200-400), que se añadió lentamente (volúmenes de 1/10 de volumen de la gota) hasta reemplazar todo el contenido de la gota de cristalización. Posteriormente los cristales fueron montados a -173 °C en una crío-asa de nylon.

3.8 Colecta y procesamiento de datos de la difracción de rayos X.

La exposición de los cristales a una fuente de rayos X se realizó en la línea 19ID del *Advanced Photon Source* (APS) con un detector ADSC Q315r CCD y en la línea BL14-1 del *Stanford Synchrotron Radiation Lightsource* (SSRL) con un detector MARmosaic 325 CCD. Las imágenes de la difracción fueron indexadas e integradas con el programa XDS (Wolfgang Kabsch, 2010). El escalamiento, ordenamiento y empalme se realizó con el programa XSCALE (Wolfgang Kabsch, 2010). En el caso de los datos de las estructuras del inhibidor silvestre E3Ad (PDB ID: 5FZZ y 5G00), se realizó un escalado anisotrópico de los datos, usando el servidor *Diffraction Anisotropy Server* (<http://services.mbi.ucla.edu/anisoscale/>) (Strong et al., 2006). Para la conversión de los valores de intensidad de las reflexiones en estimados de los valores de los factores de estructura, se utilizó el programa XDSCONV. Las fases fueron resueltas mediante reemplazo molecular utilizando el programa PHASER (McCoy et al., 2007) y la estructura cristalográfica del inhibidor de papa PSPI (PDB ID: 3TC2) que presenta un 70.8 % de identidad de secuencia con E3Ad. Para los datos de difracción obtenidos en el SSRL, el reemplazo molecular se realizó usando como molde la estructura del mutante E3Ad_N19D (PDB ID: 5FNW), obtenida previamente con los datos del APS. Los modelos resultantes se afinaron mediante un ciclo bajo condiciones de cuerpos rígidos con el programa REFMAC5 (Murshudov et al., 2011). Posteriormente, se alternaron ciclos de reconstrucción manual y afinamiento restringido global usando los programas Coot (Emsley, Lohkamp, Scott, & Cowtan, 2010) y PHENIX (Adams et al., 2010), respectivamente. La validación final de las estructuras se realizó con el programa MolProbity (Chen et al., 2010).

3.9 Análisis de secuencia y estructural de los inhibidores de proteasas

La búsqueda de secuencias similares al inhibidor E3Ad en bases de datos no redundantes (nrNCBI) se realizó usando la herramienta de búsqueda DELTA-BLAST (Boratyn et al., 2012). Los alineamientos múltiples de secuencia y estructurales se realizaron con los servidores T-Coffee (<http://tcoffee.crg.cat/apps/tcoffee/index.html>) (Di Tommaso et al., 2011) y PD BeFold (<http://www.ebi.ac.uk/msdsrv/ssm/>) (Krissinel & Henrick, 2004), respectivamente. La información derivada de ambos alineamientos fue utilizada para la edición manual del alineamiento múltiple de secuencia. Para la estimación del modelo evolutivo que mejor describía nuestro grupo de inhibidores se utilizó el servidor ProtTest v2.4 (http://darwin.uvigo.es/software/prottest2_server.html) (Abascal, Zardoya, & Posada, 2005). Para la construcción del árbol filogenético se empleó el método de máxima verosimilitud y el modelo de mejor ajuste proporcionado por el servidor ProtTest usando el programa MEGA7 (Kumar, Stecher, & Tamura, 2016). Se obtuvo un árbol consenso a partir de 1000 réplicas de *bootstrap*.

Los alineamientos estructurales por pares globales y de regiones específicas de las estructuras se realizaron con los módulos de sobreposición del programa Coot, *Secondary Structure Matching* (SSM) y *Least-squares* (LSQ), respectivamente (Emsley et al., 2010).

Para la obtención del modelo del complejo E3Ad-Tripsina se alinearon las coordenadas de las posibles posiciones P3-P3' del inhibidor E3Ad_N19D (5FZU, T89-V94) con las posiciones equivalentes del asa no canónica del inhibidor API-A (3E8L, V85-D90). Se generó un archivo que contenía las coordenadas de la molécula de tripsina unida a la asa no canónica de API-A y las coordenadas del inhibidor E3Ad_N19D después del alineamiento con el asa no canónica de API-A. Los posibles impedimentos estéricos presentes en el modelo del complejo Tripsina-E3Ad_N19D se minimizaron con el servidor Chiron (<http://redshift.med.unc.edu/chiron/login.php>).

El servidor metaPPISP (<http://pipe.scs.fsu.edu/meta-ppisp.html>) se utilizó para la predicción de los residuos involucrados en una interfaz proteína-proteína, basados en las diferentes estructuras obtenidas para los inhibidores E3Ad, E3Ad_N19D y PDI (Qin & Zhou, 2007). Todas aquellas posiciones de la secuencia de aminoácidos con un valor de puntuación mayor a 0.34 fueron consideradas como una predicción positiva, en concordancia con lo sugerido por los desarrolladores del servidor.

El servidor de acoplamiento molecular ClusPro v2.0 (<https://cluspro.bu.edu/login.php>) se empleó para la obtención de posibles modelos del complejo Plasmepsina II (PlmII)-E3Ad_N19D y PlmII-PDI. Se utilizó la estructura 3D de la Plm II con código PDB 1LF4 y las diferentes estructuras obtenidas en este trabajo para E3Ad y E3Ad_N19D. En el caso del inhibidor PDI, se utilizaron las coordenadas del inhibidor a los 93.77 ns de una simulación de dinámica molecular a pH 4.4 y 37 °C. En todos los casos se evaluaron los modelos obtenidos con la función de energía para el coeficiente balanceado de ClusPro.

Para analizar la conservación de la secuencia del asa L9 en aquellos inhibidores con actividad frente a proteasas aspárticas se obtuvo un alineamiento múltiple de las secuencias de los inhibidores E3Ad, PDI (Keilová H. and Tomášek, 1976b), API9 (Cater et al., 2002; Ritonja et al., 1990), SLAPI (Cater et al., 2002), Cathepsin D inhibidor (Uniprot ID: Q8RXA3_SOLTU) (Heibges, Salamini, et al., 2003), clone PI2112 Kunitz-type inhibitor B (Uniprot ID: Q8S380_SOLTU), clone PI8234 Kunitz-type inhibitor B (Uniprot ID: A0A097H184_SOLTU), clone PI4587 Kunitz-type inhibitor B (Uniprot ID: A0A097H132_SOLTU) y PI5887 Kunitz-type inhibitor B (Uniprot ID: A0A097H174_SOLTU) (Fischer et al., 2015). Los cuatro primeros inhibidores pertenecen al grupo A, mientras que el resto son integrantes del grupo B.

3.10 Obtención y purificación de la proteasa Plasmepsina II de *Plasmodium falciparum*.

La obtención de la proteasa Plm II se realizó siguiendo un protocolo de expresión y purificación descrito anteriormente (Ramírez et al., 2009), con algunas modificaciones. La Plm II se expresó como un zimógeno y en forma de cuerpos de inclusión en *E. coli*.

Para ello, se inocularon cepas de *E. coli* BL21(DE3) pLysS, previamente transformadas con el plásmido pET3b-proPlmII (Hill et al., 1994) en 20 mL de medio MDG (Na_2HPO_4 25 mmol L⁻¹, KH_2PO_4 25 mmol L⁻¹, NH_4Cl 50 mmol L⁻¹, Na_2SO_4 5 mmol L⁻¹, MgSO_4 2 mmol L⁻¹, FeCl_3 100 $\mu\text{mol L}^{-1}$, glucosa 0.5 % y ácido aspártico 0.25 %), complementado con ampicilina (100 $\mu\text{g L}^{-1}$) y cloranfenicol (34 $\mu\text{g L}^{-1}$). El cultivo se incubó a 37 °C y 150 rpm por 18 horas. Se utilizaron 300 μL de este cultivo para inocular 300 mL de medio ZYM-5052 (triptona 1 %, extracto de levadura 0.5 %, Na_2HPO_4 25 mmol L⁻¹, KH_2PO_4 25 mmol L⁻¹, NH_4Cl 50 mmol L⁻¹, Na_2SO_4 5 mmol L⁻¹, MgSO_4 2 mmol L⁻¹, FeCl_3 100 $\mu\text{mol L}^{-1}$, glicerol 0.5 %, glucosa 0.05 % y lactosa 0.2 %), complementado con ampicilina (100 $\mu\text{g L}^{-1}$) y cloranfenicol (34 $\mu\text{g L}^{-1}$). El cultivo se incubó a 37 °C y 150 rpm por 18 horas. Las células se recuperaron mediante centrifugación a 5000 x g a temperatura ambiente por 10 minutos y se resuspendieron en 80 mL de una solución amortiguadora Tris-HCl 50 mmol L⁻¹, NaCl 150 mmol L⁻¹ pH 8.0. La ruptura celular se realizó mediante tres ciclos de congelación/descongelación, seguidos de 12 ciclos con 15 segundos de sonicación y 1 minuto de reposo. Se adicionó Triton x-100, a una concentración final de 0.5 % y se incubó a 37 °C durante 30 minutos. Se centrifugó a 10000 x g a 4 °C durante 30 minutos, y se recuperó el sedimento. Se utilizaron 250 mL de la solución amortiguadora Tris-HCl 50 mmol L⁻¹, NaCl 150 mmol L⁻¹ pH 8.0 para re-suspender el sedimento y se mantuvo en agitación a 4 °C durante toda la noche.

La suspensión obtenida se centrifugó a 10000 x g a 4 °C durante 30 minutos, y se descartó el sobrenadante. Para la solubilización de los cuerpos de inclusión se utilizaron 50 mL de la solución amortiguadora urea 8 mol L⁻¹, Tris-HCl 100 mmol L⁻¹, β -mercaptoetanol 50 mmol L⁻¹, EDTA 1 mmol L⁻¹ glicina 1 mmol L⁻¹ pH 8.0 y se mantuvo en agitación a 4 °C durante dos horas. El material insoluble se eliminó mediante centrifugación a 15000 x g, a 7 °C durante 2 horas. El plegamiento de la proPlm II se realizó mediante dilución del sobrenadante de la centrifugación anterior, en 450 mL de amortiguador Tris-HCl 20 mmol L⁻¹ pH 8.0 en agitación constante a una temperatura de 18 °C. La solución se mantuvo en agitación durante toda la noche y posteriormente se dializó con 10 L de amortiguador Tris-HCl 20 mmol L⁻¹ pH 8.0.

Para la posterior purificación de la proPlmII, se aplicó la solución dializada en 26 mL de resina de intercambio aniónico Q-Sepharose™ HP (GE Healthcare Life Sciences, EUA), empacada en una columna Econo-Column® (2.5 x 10 cm, BioRad, EUA) acoplada a un sistema cromatográfico Äktaprime™ (GE Healthcare Life Sciences, EUA). La elución se realizó mediante un lavado con 100 mmol L⁻¹ de NaCl en el amortiguador Tris-HCl 20 mmol L⁻¹ pH 8.0, seguido por un gradiente lineal de

100 a 500 mmol L⁻¹ de NaCl en el referido amortiguador. Las fracciones colectadas durante la elución se analizaron mediante una electroforesis SDS-PAGE bajo condiciones reductoras (Laemmli, 1970). A las fracciones que contenían la proPlmII se les añadió NaAc 1 mol L⁻¹ pH 5.0 hasta alcanzar una concentración final de 50 mmol L⁻¹. Se verificó que el valor del pH (5.0) y se incubó con agitación a 37 °C durante una hora para inducir la autoactivación de la proPlmII. Posteriormente, se diluyó hasta alcanzar 4 veces su volumen inicial y se ajustó el pH a 8.0 antes de realizar una nueva cromatografía de intercambio aniónico, bajo las mismas condiciones descritas anteriormente. Las fracciones colectadas durante la elución se analizaron mediante una electroforesis SDS-PAGE bajo condiciones reductoras (Laemmli, 1970). Las fracciones que contenían a la Plm II se unieron y se concentraron usando un dispositivo de concentración por centrifugación Amicon Ultra-15 (Millipore, EUA), con límite de exclusión de 10 kDa hasta valores de concentración entre 5-10 mg mL⁻¹. Para el cálculo de la concentración de proteínas se utilizó un coeficiente de extinción molar a 280 nm igual a 42 080 L mol⁻¹ cm⁻¹, obtenido a través del programa ProtParam.

Se aplicaron 300 µL de la solución concentrada de Plm II a una columna Superdex™ 75HR 10/30, previamente equilibrada con la solución amortiguadora Tris-HCl 20 mmol L⁻¹, NaCl 150 mmol L⁻¹ pH 8.0, acoplada a un sistema cromatográfico Äktapure™25 (GE Healthcare Life Sciences, EUA). La cromatografía se realizó a un flujo lineal de 15.3 cm h⁻¹ y las fracciones colectadas se evaluaron mediante SDS-PAGE (Laemmli, 1970). Las fracciones que contenían a la Plm II se concentraron usando un Amicon Ultra-4 (Millipore, EUA), con límite de exclusión de 10 kDa hasta un valor de entre 10-15 mg mL⁻¹. Posteriormente, se añadió glicerol a una concentración final de 50 % y se almacenaron a -20 °C hasta su uso.

3.11 Ensayos de actividad inhibidora frente a proteasas aspárticas y serino

Los ensayos de actividad enzimática e inhibidora con la proteasa aspártico Plm II se realizaron usando el sustrato peptídico cromogénico KPFEFNphRL (Nph: paranitrofenilalanina, K_M: 50.8 µmol L⁻¹; Genscript, EUA) a una concentración final en el ensayo igual a 84.0-84.5 µmol L⁻¹ y usando el amortiguador NaAc 100 mmol L⁻¹ pH 4.4. Las reacciones fueron monitoreadas de forma continua a 310 nm en un espectrofotómetro Cary60 (Agilent Technologies, EUA) a una temperatura de 37 °C (Westling et al., 1999). La concentración de enzima activa se determinó mediante titulación con el inhibidor de unión fuerte Pepstatin A (Sigma Aldrich, EUA) y se ajustó en el ensayo enzimático a valores entre 1.0-2.5 x 10⁻⁷ mol L⁻¹.

Para los ensayos de actividad inhibidora con la proteasa de tipo serino tripsina de páncreas bovino, se empleó Nα-benzoyl-L-arginine p-nitroanilide (BAPNA; Sigma-Aldrich, EUA) como sustrato (K_M: 1 mmol L⁻¹), a una concentración final de 1 mmol L⁻¹ y el amortiguador Tris-HCl 20 mmol L⁻¹, NaCl 150

mmol L⁻¹, CaCl₂ 20 mmol L⁻¹ pH 8.0. Las reacciones fueron monitoreadas de forma continua a 405 nm en un espectrofotómetro Cary60 (Agilent Technologies, EUA) a una temperatura de 25 °C (Erlanger, Kokowsky, & Cohen, 1961). La concentración de enzima en el ensayo se ajustó a un valor de 6.6 x 10⁻⁸ mol L⁻¹.

Para ambos ensayos cinéticos, la enzima y el inhibidor se incubaron por 5 minutos previo a la adición del sustrato. La velocidad inicial de la reacción se determinó con el programa Cary WinUV Kinetics v5.0 (Agilent Technologies, EUA), a diferentes concentraciones del inhibidor. La fracción de actividad residual (v_i/v_0) se calculó a partir de la velocidad inicial de las reacciones en presencia de diferentes concentraciones de inhibidor (v_i) respecto a la velocidad de la reacción en ausencia de inhibidor (v_0).

Todos los ensayos se realizaron por triplicado. Los valores de la constante de inhibición aparente (K_i^{app}) se determinaron a través del ajuste por regresión no lineal de los datos a la ecuación de Morrison (Ecuación 1), usando el programa Statistica v7.0 (Statsoft, EUA). Los valores de la constante de inhibición se calcularon usando la Ecuación 2, considerando que los inhibidores descritos en este trabajo son de tipo competitivo, lo cual fue reportado anteriormente por Keilová H. and Tomášek, 1976, para otros inhibidores bifuncionales de este subgrupo. Los valores finales de las K_i representan la media de tres experimentos independientes.

$$\frac{v_i}{v_0} = 1 - \frac{([E]_0 + [I]_0 + K_i^{app}) - \sqrt{([E]_0 + [I]_0 + K_i^{app})^2 - 4[E]_0[I]_0}}{2[E]_0} \quad (\text{Ecuación 1})$$

$$K_i^{app} = K_i \left(1 + \frac{[S]_0}{K_M} \right) \quad (\text{Ecuación 2})$$

3.12 Simulaciones de Dinámica Molecular

Las simulaciones de dinámica molecular se realizaron en colaboración con el Dr. Pedro Alberto Valiente Flores, adscrito al Laboratorio de Bioinformática y Dinámica Molecular del Centro de Estudios de Proteínas (CEP) de la Universidad de la Habana, Cuba. Todos los sistemas moleculares se manipularon con las herramientas de GROMACS (Hess, Kutzner, van der Spoel, & Lindahl, 2008).

Los archivos de topologías y las estructuras tridimensionales de partida de los inhibidores PDI y E3Ad_N19D se crearon con la herramienta *pdb2gmx* (van der Spoel et al., 2010) y el conjunto de parámetros del campo de fuerzas AMBER99sb (Hornak et al., 2006). Ambos inhibidores se simularon a pH=4.4 y pH=8.0. El estado de protonación de los aminoácidos ionizables de ambos inhibidores a ambos valores de pH se asignó con el programa PROPKA (H. Li, Robertson, & Jensen, 2005), implementado en el servidor web PDB2PQR (<http://agave.wustl.edu/pdb2pqr/server.html>). Las configuraciones iniciales de los sistemas estudiados se generaron en una caja dodecahédrica

rómbica (xy cuadrada) bajo condiciones periódicas de contorno, con la herramienta *editconf* (van der Spoel et al., 2010). En estos sistemas se definió una distancia de 1 nm entre el soluto y las paredes de la caja. Las moléculas de solvente se añadieron a estos sistemas, empleando el modelo de agua TIP3P (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983), con la herramienta *genbox* (van der Spoel et al., 2010). Como resultado, se obtuvieron sistemas con un número total de átomos entre 24825 y 25000. Finalmente, la carga neta de estos sistemas se neutralizó, mediante la adición de 0.15 mol L^{-1} de NaCl, con la herramienta *genion* (van der Spoel et al., 2010).

Las ecuaciones de movimiento de los sistemas simulados se solucionaron por métodos numéricos, a partir del algoritmo de “salto de rana” (Hockney, 1970), del inglés “*leap frog*”, empleando un tiempo de integración de 2 fs. Para ello fue necesario limitar los movimientos de vibración de los enlaces covalentes en las moléculas simuladas, mediante los algoritmos LINCS (Hess, Bekker, Berendsen, & Fraaije, 1997) y SETTLE (Miyamoto & Kollman, 1992). Las velocidades iniciales de cada uno de los átomos de los sistemas estudiados se asignaron de forma aleatoria, a partir de la distribución de velocidades de Maxwell generadas a las temperaturas de 25 y 37 °C, con la herramienta *grompp* (van der Spoel et al., 2010). La temperatura y la presión de los sistemas estudiados se controló mediante el algoritmo de Berendsen (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984), a partir de un acoplamiento débil de la temperatura y la presión de los sistemas a un baño térmico de 310 K (constante de acoplamiento térmico, τ_T igual a 0.1 ps) y una presión de referencia de 100 kPa (constante de acoplamiento τ_P igual a 0.5 ps).

Durante las simulaciones de dinámica molecular, las interacciones electrostáticas de largo alcance y de van der Waals entre los átomos no unidos covalentemente, se calcularon mediante el método PME (del inglés “*Particle Mesh Ewald*”) (Darden, York, & Pedersen, 1993) y la expresión de Lennard-Jones, respectivamente. En el caso de las interacciones electrostáticas de largo alcance, se empleó como radio de corte para evaluar estas interacciones, una distancia interatómica mayor de 1.0 nm, mientras que para evaluar la expresión de Lennard-Jones se empleó como radio de corte una distancia menor de 1.1 nm. Estos dos tipos de interacciones se calcularon cada 5 tiempos de integración (10 fs), lo cual determina que la evaluación de las interacciones entre los átomos no unidos covalentemente ocurra después de la generación del nuevo “registro de vecinos” (del inglés “*neighbor-list*”).

3.12.1 Equilibración y simulaciones productivas

En la optimización de la estructura tridimensional de PDI y E3Ad_N19D se emplearon los métodos del descenso más marcado o empinado (del inglés “*steepest descent*”), y de gradientes conjugados (del inglés “*conjugate gradient*”) (Zimmermann, 1991). Para ambos métodos se utilizaron como

criterios de terminación de la optimización molecular: *i*) una variación en la energía potencial del sistema entre los pasos *i* e *i*+1 inferior a 10 kJ mol⁻¹ nm⁻¹ ó *ii*) un máximo de 10 000 iteraciones sin alcanzar este criterio de convergencia. Posteriormente, se realizaron simulaciones de MD de 500 ps para cada uno de estos sistemas. En estas simulaciones se aplicó una fuerza constante isotrópica de 4180 kJ mol⁻¹ nm⁻², sobre las posiciones de cada uno de los átomos pesados de PDI ó E3Ad_N19D, mediante la herramienta *genrestr* (van der Spoel et al., 2010). Estas simulaciones se realizaron con el objetivo de evitar deformaciones en los sistemas estudiados, como resultado del proceso de equilibración de las capas de solvente añadidas a los solutos. Seguidamente, se realizó una simulación de dinámica molecular de 200 ps en la cual se eliminó la restricción harmónica sobre las posiciones de los átomos de los inhibidores y de las cadenas laterales de aminoácidos de ambos inhibidores. Finalmente, se realizaron simulaciones de dinámica molecular de 100 ns, sin restricciones harmónicas sobre las posiciones de los átomos de ambas moléculas con la herramienta *mdrun* (van der Spoel et al., 2010).

3.12.2 Análisis de las simulaciones de dinámica molecular.

De manera general, antes de realizar los análisis relacionados con los resultados de las simulaciones de dinámica molecular, fue necesario realizar el ajuste de las coordenadas almacenadas en los archivos de trayectoria, de los sistemas simulados con respecto a sus estructuras iniciales, mediante la herramienta *trjconv* (van der Spoel et al., 2010). Este paso tiene como objetivo eliminar la influencia de los movimientos de rotación y traslación de las moléculas que componen los sistemas simulados. Los resultados de los análisis realizados se graficaron con el programa XMGRACE (<http://plasma-gate.weizmann.ac.il/Grace/>).

La desviación cuadrática media en distancia (r.m.s.d) del esqueleto polipeptídico de los inhibidores PDI y E3Ad_N19D durante el tiempo de simulación, en cada uno de los sistemas estudiados, con respecto a las coordenadas iniciales (estructuras cristalográficas) se calculó mediante la herramienta *g_rms* (van der Spoel et al., 2010).

4. Resultados y discusión

4.1 Obtención, purificación y caracterización bioquímica de los inhibidores.

La obtención de los inhibidores descritos en este trabajo se hizo en dos etapas. En la primera se obtuvo la forma nativa del inhibidor E3Ad sin modificaciones en sus extremos amino y carboxilo, además de su forma no glicosilada E3Ad_N19D. Las variantes con etiqueta de histidinas en el extremo carboxilo se obtuvieron una vez obtenidas las estructuras tridimensionales y como resultado

del análisis que sugiere que el extremo carboxilo no parece implicado en la interacción con las proteasas di ana. Todos los inhibidores fueron expresados en la levadura *P. pastoris* X33 y presentaban el factor α de secreción de *S. cerevisiae* fusionado a su extremo amino, lo que permitió su secreción al medio de cultivo, donde representaron alrededor del 30 % de las proteínas totales. En el caso de los inhibidores E3Ad y su variante no glicosilada E3Ad_N19D, la purificación hasta un porcentaje de pureza mayor al 90 % estimado por densitometría de geles de SDS-PAGE, se logró con una combinación de cromatografía de intercambio catiónico y exclusión molecular (Figuras 4 y 5A/B). En el caso de las variantes con etiqueta de histidinas (6x), la purificación se logró mediante una combinación de cromatografía de afinidad a Ni^{2+} y exclusión molecular con un grado de pureza superior al 95 % usando el mismo criterio que para los inhibidores E3Ad y E3Ad_N19D (Figuras 4 y 5C).

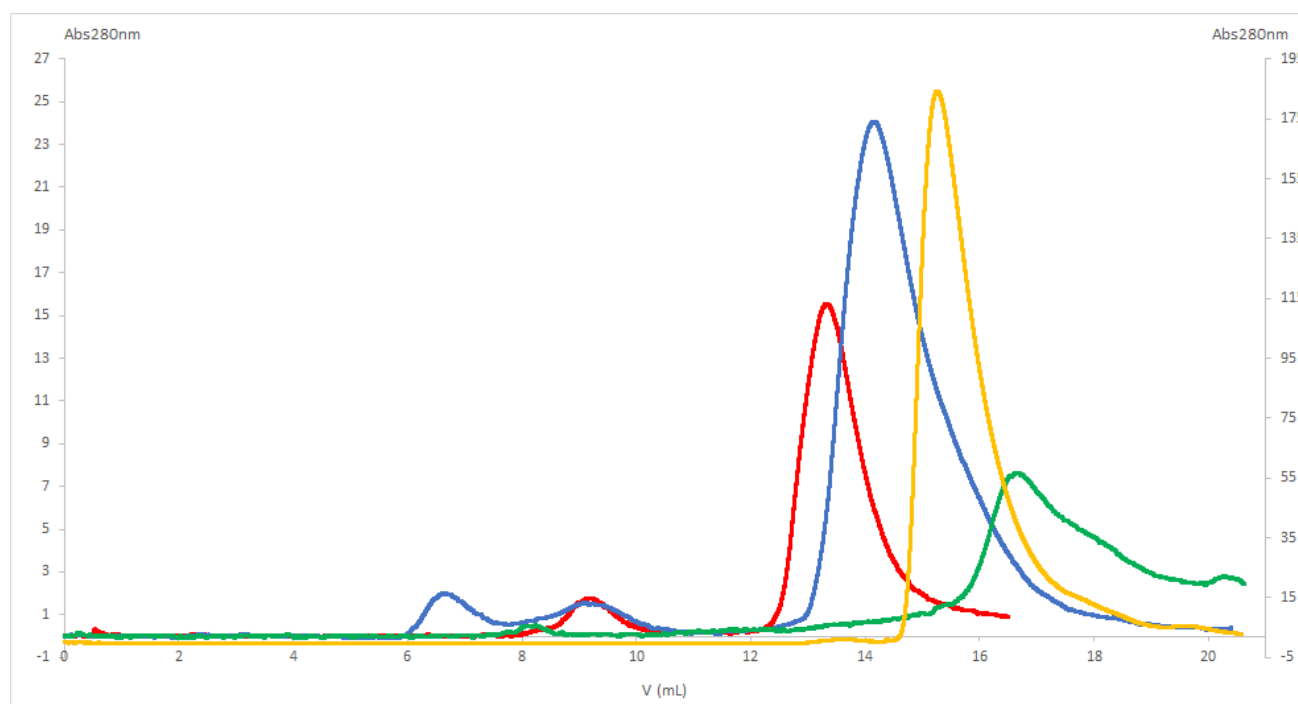


Figura 4. Perfil cromatográfico de la exclusión molecular en una columna Superdex 75 HR 10//30. para los inhibidores bifuncionales purificados. El eje secundario (lado derecho) se corresponde con la absorbancia a 280 nm durante la purificación del inhibidor E3Ad_N19D_R151A_His_{6x}. E3Ad (rojo), E3Ad_N19D (azul), E3Ad_N19D_His_{6x} (verde) y E3Ad_N19D_R151A_His_{6x} (amarillo).

En todos los casos los niveles de pureza alcanzados fueron adecuados para estudios estructurales y funcionales. El rendimiento en base a la cantidad de proteína total osciló entre 4-6 mg por litro de cultivo. Las diferencias en los volúmenes de elución entre las proteínas, sin y en presencia de la

etiqueta de histidinas, podría deberse a la composición y el pH de los buffers empleados o a cambios en el radio hidrodinámico de las proteínas como resultado de la presencia de la etiqueta de histidinas.

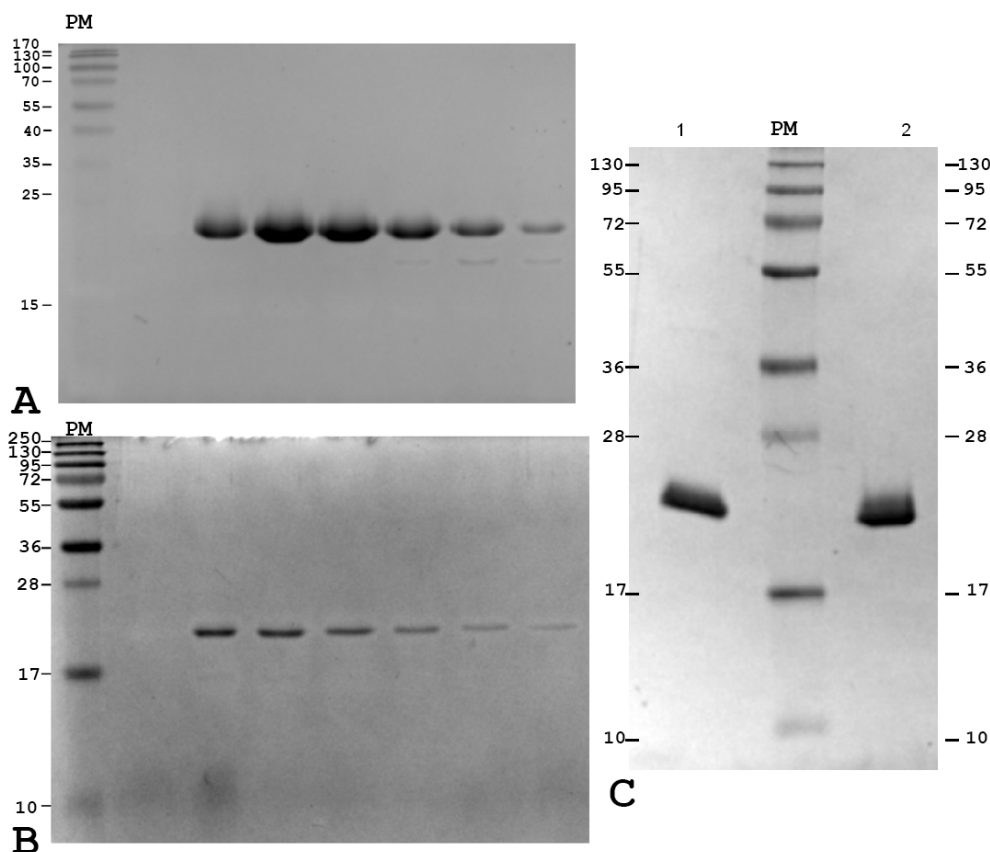


Figura 5. Electroforesis de SDS-PAGE de los inhibidores E3Ad (A), E3Ad_N19D (B), E3Ad_N19D_His6x (C, carril 1) y E3Ad_N19D_R151A_His6x (C, carril 2). PM: marcador de masa molecular (kDa).

En el caso del inhibidor silvestre E3Ad, el análisis por SDS-PAGE de las fracciones provenientes de la cromatografía de exclusión molecular, mostraron la presencia de una segunda banda minoritaria. La identidad de ambas bandas se determinó mediante espectrometría de masas ESI-MS-MS (Unidad de Proteómica del IBt) y los resultados mostraron que ambas corresponden al inhibidor E3Ad (95.7 % de cobertura de secuencia), pero que presentaron un número variable de carbohidratos unidos a la Asparagina 19 (Anexo 1). Adicionalmente, el análisis de espectrometría de masas nos permitió verificar la integridad del extremo amino (la eliminación del factor α luego de su secreción al medio de cultivo), así como la naturaleza de los carbohidratos unidos a la proteína, siendo estos, dos unidades de N-Acetilglucosamina unidas a un número variable de manosas (Figura 6). Este resultado mostró que nuestra muestra del inhibidor E3Ad, en realidad es una mezcla de al menos 6 glicofomas con diferencias de un residuo de manosa entre ellos, lo que afecta la heterogeneidad de la muestra.

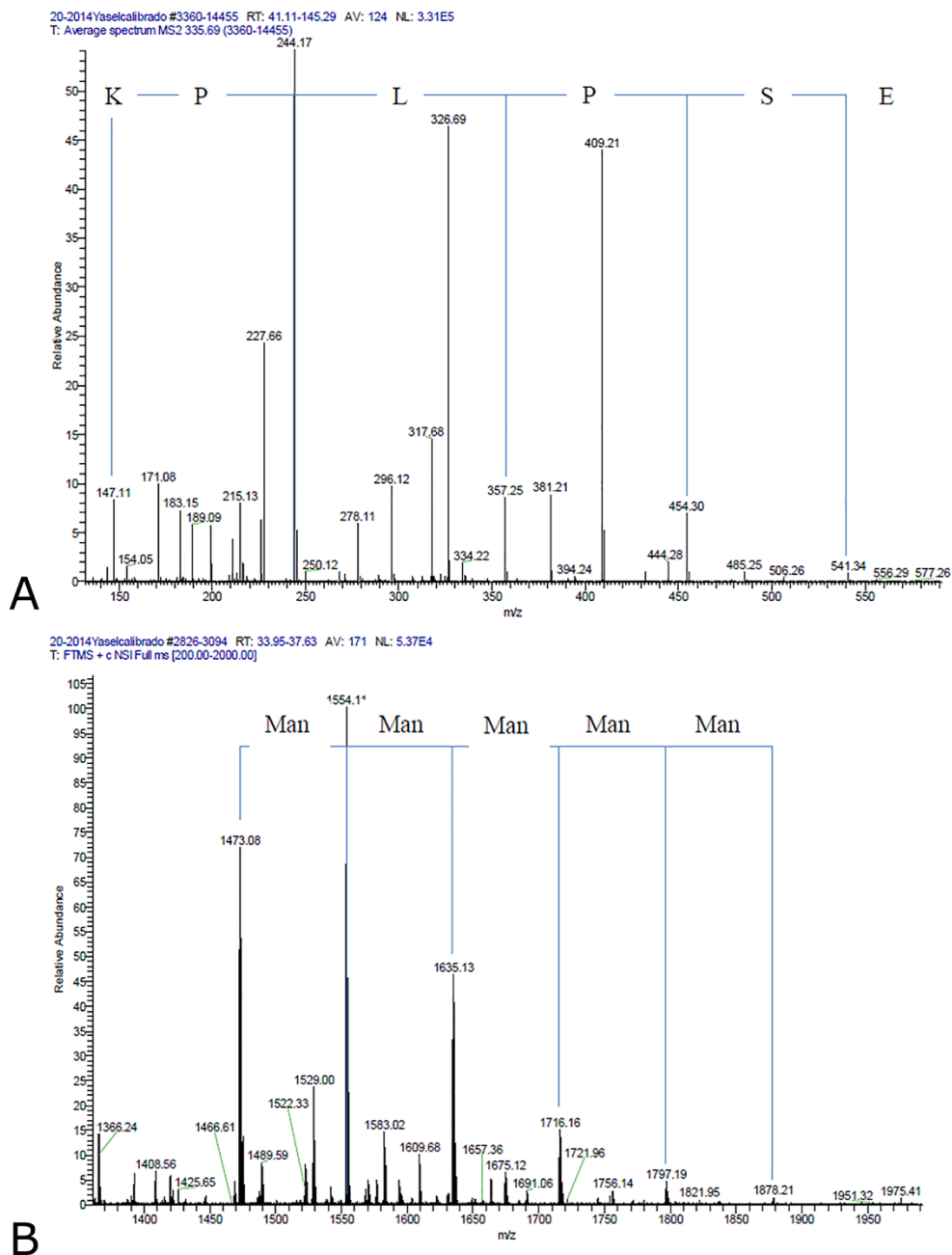


Figura 6. Perfiles de espectrometría de masas ESI-MS-MS del péptido correspondiente al extremo amino del inhibidor E3Ad (A) y de sus precursores glicosilados (B).

Es de destacar que esta heterogeneidad de la muestra de E3Ad producto de la presencia de múltiples glicofomas no fue detectada durante un análisis de la misma mediante dispersión dinámica de luz, a pesar de ser una técnica empleada ampliamente para analizar la dispersión de una solución. En nuestra opinión este resultado pudo deberse a la pequeña diferencia de masa entre las distintas formas glicosiladas de E3Ad, de solo 18 Da entre sí.

Los resultados de una electroforesis de isoelectroenfoque nos permitieron determinar los valores experimentales de pI de los inhibidores E3Ad y su mutante E3Ad_N19D (Figura 7), los cuales fueron un poco mayores a los calculados teóricamente con el programa ProtParam para E3Ad (8.26) y E3Ad_N19D (7.79).

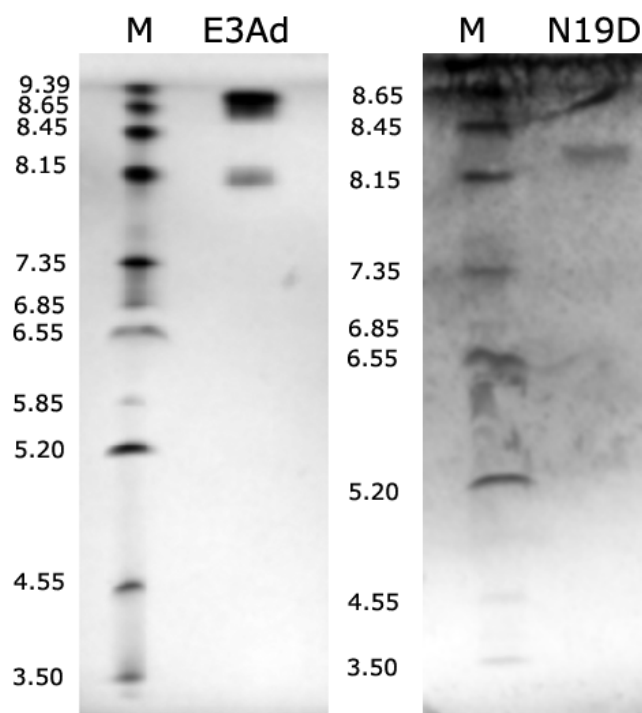


Figura 7. Electroforesis de isoelectroenfoque de los inhibidores E3Ad (A) y E3Ad_N19D (B).

Como parte de la caracterización del inhibidor E3Ad_N19D, se determinaron las constantes de inhibición (K_i) frente a las proteasas tripsina de páncreas bovino (Figura 8A) y Plm II (Figura 8B). En ambos casos se consideró que el inhibidor es de tipo competitivo, en base a b descrito anteriormente por Keilová H. y Tomášek (1976) para otro inhibidor bifuncional similar. Ensayos de inhibición preliminares mostraban que la capacidad inhibidora de E3Ad_N19D frente a Plm II era similar a la de E3Ad. Los valores de K_i obtenidos para el inhibidor E3Ad_N19D fueron de $43 \pm 2.9 \text{ nmol L}^{-1}$ para la tripsina y $10.5 \pm 2.5 \text{ nmol L}^{-1}$ para la Plm II, lo que nos permitió clasificarlo como un inhibidor de unión fuerte frente a ambas enzimas. Esta es la primera vez que se informa de la capacidad de estos inhibidores de inhibir Plm II, lo que amplía el número de proteasas del clan AA, familia A1, que son inhibidas por inhibidores de proteasas de tipo aspártico de la familia Kunitz-STI. Si bien, esta enzima no representa el blanco natural de estos inhibidores puede ser usada como una proteasa modelo para el estudio de las interacciones de estos inhibidores bifuncionales con las proteasas aspártico. Además, la disponibilidad de otras isoformas de esta enzima permitiría realizar estudios sobre la especificidad de la interacción proteasa-inhibidor.

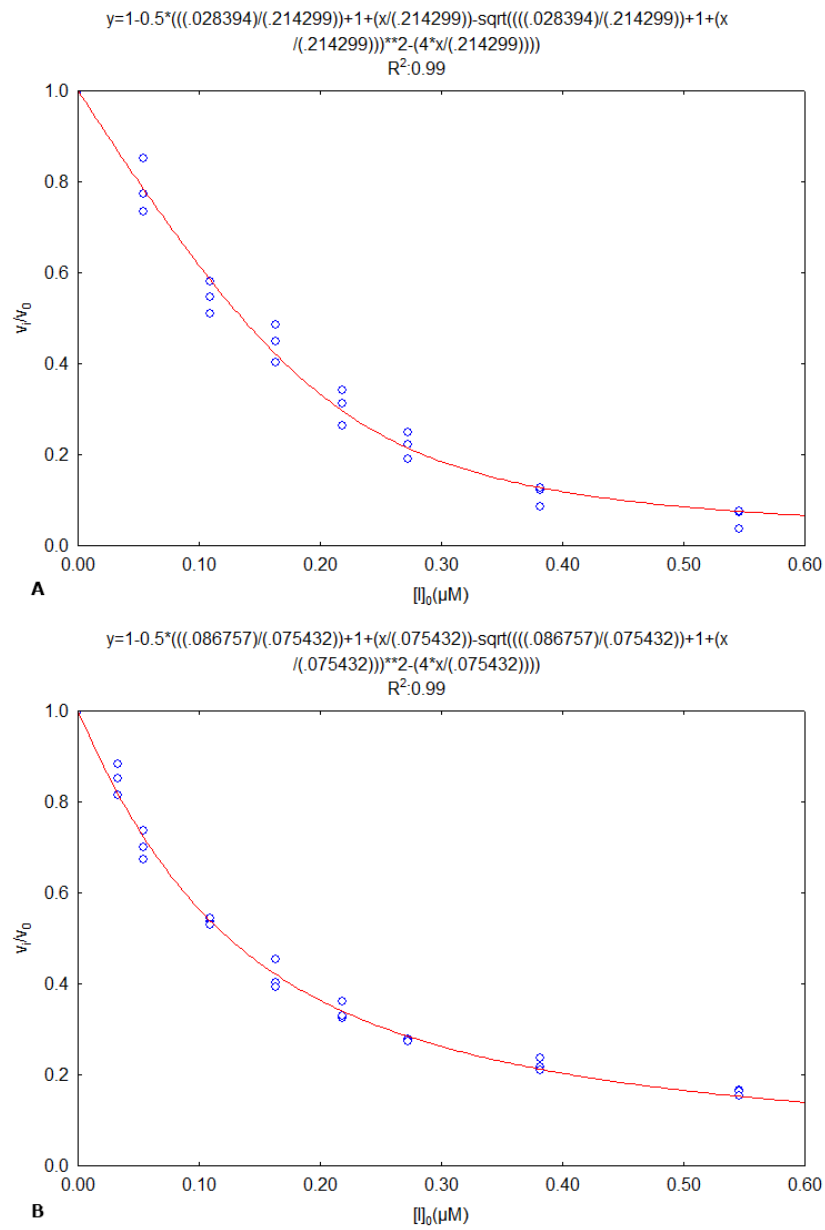


Figura 8. Ajuste a la ecuación de Morrison de los resultados de los ensayos de inhibición de E3Ad_N9D frente a las proteasas Plm II (A) y tripsina de páncreas bovino (B).

Trabajos anteriores con inhibidores de tripsina de la familia Kunitz-STI, obtenidos en *E. coli* demostraban su capacidad de inhibir a proteasas serino y cisteíno (Speranskaya et al., 2006), pero no existían evidencias de este comportamiento para las proteasas aspártico. Estos resultados muestran que la ausencia de las glicosilaciones no es esencial para la capacidad de inhibir proteasas serino, cisteíno y aspártico en los inhibidores de proteasas Kunitz-STI. En nuestra opinión las glicosilaciones podrían jugar un papel en el transporte de estas moléculas en la célula o de estabilización.

4.2 Cristalización y determinación de la estructura tridimensional del inhibidor E3Ad y del mutante E3Ad_N19D.

De un total de 530 condiciones iniciales de cristalización evaluadas a diferentes temperaturas y concentración de proteínas, se obtuvieron cristales en 42 de ellas. Las condiciones encontradas presentaron valores de pH en un rango desde 4 hasta 9, en la mayoría de los casos en presencia de polietilén glicol como agente precipitante, pero en ningún caso se obtuvieron patrones de difracción con resoluciones mayores a 3.2 Å. Durante el proceso de optimización posterior se modificaron las condiciones de cristalización y también se emplearon aditivos, lo que nos permitió aumentar el límite de resolución hasta alrededor de 2.5 Å. Sin embargo, los problemas relacionados con la posible presencia de varias redes cristalinas y/o altos valores de mosaicidad persistían en nuestros patrones de difracción de rayos X. Considerando la presencia de varias glicofomas del inhibidor E3Ad se eliminó el único sitio de glicosilación existente en su secuencia, que era la asparagina en la posición 19. Se prepararon pruebas de cristalización del mutante E3Ad_N19D usando las matrices comerciales, así como en aquellas condiciones donde se obtuvieron los mejores resultados con el inhibidor silvestre, en cuanto al límite de resolución. Se obtuvieron cristales del inhibidor E3Ad_N19D en forma de barras en varias de las condiciones exitosas identificadas previamente para el inhibidor silvestre (Figura 9). Los cristales aparecieron entre 2 a 15 días, en dependencia de la condición en cuestión.

Considerando la presencia de aditivos y la diferente composición de las condiciones de cristalización, decidimos verificar el pH final de las condiciones. Se mezclaron volúmenes iguales de la solución de cristalización y de la solución amortiguadora que contenía a las proteínas y se determinó el valor final del pH usando un potenciómetro. Las condiciones donde crecieron los cristales que se difractaron fueron: 1 (PEG 20 000 15 % (m/v) / HEPES 0.1 mol L⁻¹ pH 7.0 / Glicil-glicil-glicina 30 mmol L⁻¹, con un pH final igual a 3.0); 2 (MPD 10 % (v/v) / Bicina 0.1 mol L⁻¹ pH 9.0); 3 (PEG 8000 20 % (m/v) / HEPES 0.1 mol L⁻¹ pH 7.5 / Glicina 0.1 mol L⁻¹, con un pH final igual a 3.5) y 4 (PEG 8000 20 % (m/v) / HEPES 0.1 mol L⁻¹ pH 7.5 / Glicil-glicil-glicina 30 mmol L⁻¹, con un pH final igual a 7.2). En el caso del inhibidor E3Ad las condiciones fueron 5 (PEG 2000 MME 10 % (m/v) / Tris-HCl 0.1 mol L⁻¹ pH 8.5 / Trimetilamina N-óxido 0.2 mol L⁻¹ / Glicil-glicil-glicina 30 mmol L⁻¹, con un pH final igual a 7.4) y 6 (PEG 3350 20 % (m/v) / Tris-HCl 0.1 mol L⁻¹ pH 7.0 / CaCl₂ 0.2 mol L⁻¹) (Figura 9). Todos los cristales crecidos en estas condiciones fueron obtenidos a una temperatura de 18 °C, excepto en el caso de la condición 6 que se obtuvo a 4 °C. Es importante destacar que la forma de adición de la solución criopreservante fue decisiva para obtener patrones de difracción con la calidad requerida para determinar las estructuras cristalográficas, excepto en la condición 6, donde al ser por el método de batch no fue posible añadir lentamente el crioprotector.

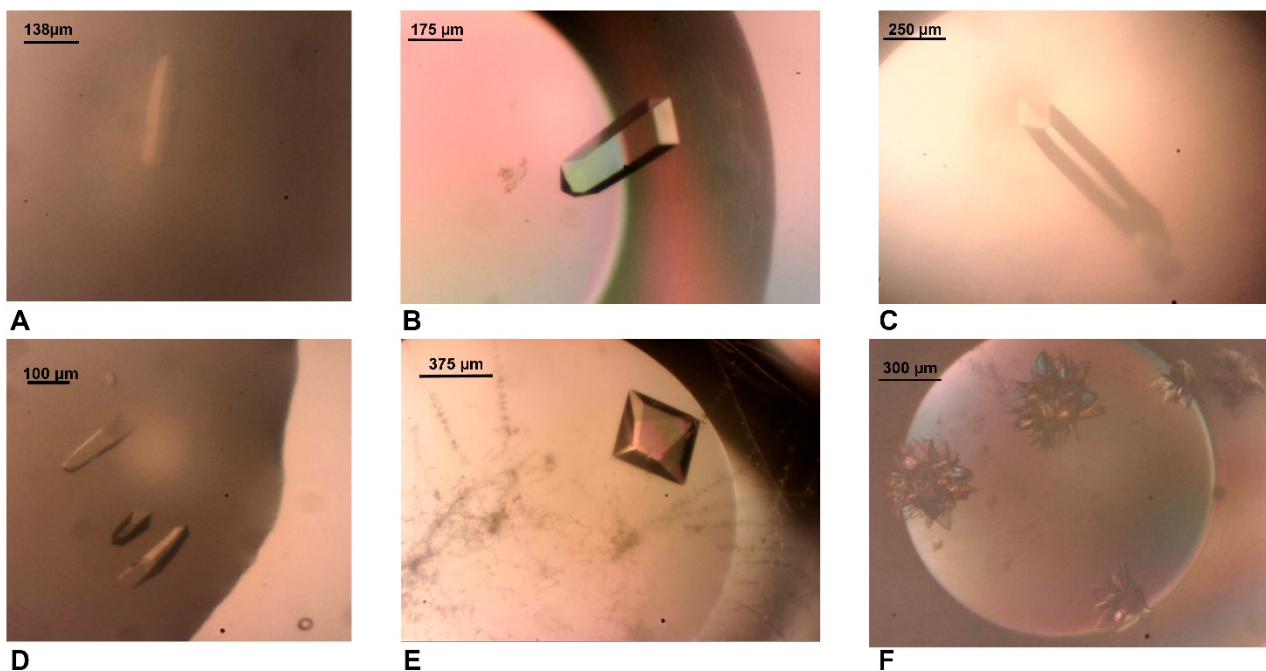


Figura 9. Cristales obtenidos para los inhibidores E3Ad (A-D, condiciones 1-4) y E3Ad_N19D (E (5) y F (6)).

Los resultados estadísticos de la colección de datos y de la determinación de las estructuras cristalógicas se pueden apreciar en la Tabla 1. De las 6 estructuras determinadas, 2 corresponden al inhibidor silvestre en los grupos espaciales ortorrómbicos $C222_1$ y $P22_12_1$ mientras que las otras 4 son del mutante no glicosilado, y son del grupo espacial tetragonal $P4_322$. El rango de resolución de las estructuras fluctúa entre 2.43 y 2.65 Å.

Tabla 1. Parámetros estadísticos de las estructuras cristalógicas de los inhibidores E3Ad y E3Ad_N19D. Los valores dentro del paréntesis se refieren a la franja de mayor resolución.

Proteína	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad	E3Ad
Condición (pH)	1 (pH 3.0)	2 (pH 9.0)	3 (pH 3.5)	4 (pH 7.2)	5 (pH 7.4)	6 (pH 7.0)
Fuente de rayos X	19ID (APS)	19ID (APS)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)
Longitud de onda (Å)	0.9792	0.9792	1.1807	1.1807	1.1807	1.1807
Grupo Espacial	$P4_322$	$P4_322$	$C222_1$	$P4_322$	$P4_322$	$P22_12_1$
Parámetros de la celda unitaria						
a (Å)	77.22	77.94	102.95	77.14	77.18	47.58
b (Å)	77.22	77.94	114.24	77.14	77.18	56.45
c (Å)	94.24	95.01	98.77	94.14	95.90	72.55
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0

Mosaicidad (°) ^a	0.14	0.25	0.21	0.13	0.38	0.55
Resolución (Å)	19.50-2.45 (2.50-2.45)	19.45-2.65 (2.75-2.65)	38.42-2.47 (2.57-2.47)	32.0-2.43 (2.53- 2.43)	29.5-2.5 (2.57- 2.5)	22.6-2.55 (2.62- 2.55)
R _{merge} (%) ^b	5.3 (63.5)	4.9 (57.1)	4.7 (63.9)	4.6 (62.8)	5.5 (72.8)	6.5 (24.1)
R _{meas} (%) ^c	5.4 (65.2)	5.1 (58.6)	5.0 (67.9)	4.7 (64.8)	5.7 (75.1)	7.0 (25.8)
CC _{1/2} (%) ^d	99.9 (96.2)	100 (95.4)	100 (95.1)	100 (94.3)	100 (94.2)	99.9 (99.2)
Integridad (%)	99.8 (100)	99.8 (100)	99.8 (100)	99.9 (100)	98.0 (74.1)	98.0 (75.8)
I/σ(I)	38.01 (5.6)	40.61 (6.0)	28.06 (3.2)	42.13 (5.2)	33.82 (4.9)	21.69 (8.0)
Multiplicidad	18.5 (19.3)	18.3 (19.2)	8.6 (8.7)	16.4 (17.0)	16.2 (16.6)	7.7 (8.1)
No. de reflexiones totales	20 3134 (12 053)	163 456 (17 575)	182 351 (20 413)	184 496 (21 020)	167 335 (9378)	51 340 (2876)
No. De reflexiones únicas	10 963 (625)	8 941 (913)	21 238 (2345)	11 223 (1240)	10 356 (564)	6 643 (354)
Factor B de Wilson (Å ²)	58.2	72.4	60.6	56.8	60.2	33.6
Contenido de solvente (%)	64.1	65.0	65.2	64.0	59.6	40.7
Factor R (%)	20.9	20.8	22.53	19.85	22.80	18.1
Factor R _{free} (%)	26.2	26.0	27.42	23.75	24.37	24.6
r.m.s.d. distancia enlaces (Å)	0.010	0.012	0.010	0.009	0.010	0.009
r.m.s.d. ángulos de enlace (°)	1.21	1.50	1.39	1.22	1.35	1.16
No. de reflexiones en el grupo de trabajo.	10 458	8 545	20 203	10 664	9 857	6 332
No. de reflexiones en el grupo de prueba.	503	391	1 035	559	499	311
Valor medio de l factor B (Å ²)	58.0	70.0	65.10	58.1	67.4	34.1
Gráfico de Ramachandran						
Permitidos (%)	94.02	90.76	92.35	95.63	92.90	96.20
Permitidos adicionales (%)	4.89	7.61	7.38	3.83	6.56	3.80
No permitidos (%)	1.09	1.63	0.27	0.55	0.55	0.0
Código PDB	5FNW	5FNX	5FZY	5FZU	5G00	5FZZ

^aCalculada con XDS (Wolfgang Kabsch, 2010).

^b $R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, donde $I_i(hkl)$ and $\langle I(hkl) \rangle$ representa los valores de la intensidad de las reflexiones para m ediciones i individuales y $\langle I(hkl) \rangle$ los correspondientes valores de su promedio. La suma es respectivamente a todas las reflexiones únicas.

^c R_{meas} es una versión de R_{merge} independiente de la multiplicidad. $R_{meas} = \frac{\sum_{hkl} \{ [N(hkl)/[N(hkl)-1] \}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$.

^dCC_{1/2} coeficiente de correlación entre las dos mitades de los datos. (Karplus & Diederichs, 2012).

Todas las estructuras tridimensionales determinadas mostraron el plegamiento β -trefoil típico de los inhibidores de la familia Kunitz-STI, con 12 hebras β , que forman tres elementos de pseudo-simetría de tercer orden (Figura 10). El plegamiento β -trefoil está compuesto por 6 hebras formando un barril β (hebras 1, 4, 5, 8, 9 y 12) y las otras 6 formando una especie de cubierta del barril.

La densidad electrónica permitió modelar el 98 % de los aminoácidos, con la excepción de los dos primeros residuos (E1 y S2). Se asignaron 3 hélices de tipo 3_{10} en las estructuras obtenidas a pH 3.0 (PDB: 5FNW) y pH 3.5 (PDB: 5FZY) para el mutante E3Ad_N19D y a pH 7.0 (PDB: 5FZZ) y pH 7.4 (PDB: 5G00) para el inhibidor glicosilado E3Ad. Estas hélices se asignaron en las regiones comprendidas entre los residuos Arg29-Gly33 (3_{10} -A), Lys91-Tyr96 (3_{10} -B) y Cys152-Gln156 (3_{10} -C). En el caso de las estructuras del mutante E3Ad_N19D obtenidas a pH 7.2 (PDB: 5FZU) y pH 9.0 (PDB: 5FNX), solo se asignaron dos hélices 3_{10} en cada una, en los aminoácidos Arg29-Gly33 y Val90-Val94 para la estructura 5FZU, así como Val90-Val94 y Cys152-Gln156 en la estructura 5FNX. Esta diferencia pudiera estar relacionada con el algoritmo empleado por el programa de asignación de estructura secundaria, DSSP (W Kabsch & Sander, 1983) y los valores de resolución y de factor B de nuestras estructuras, debido a que estas regiones poseen conformaciones muy similares a sus equivalentes en las estructuras donde sí fueron asignadas hélices 3_{10} .

Hasta el momento estas son las primeras estructuras de un miembro de la familia Kunitz-STI con tres hélices en posiciones simétricamente equivalentes. Anteriormente, se habían obtenido estructuras de miembros con una hélice α o 3_{10} : STI (1AVW), API-A (3E8L), DrTI (1R8N), PDI (5DZU) o dos hélices 3_{10} , PSPI (3TC2). Algunas de estas hélices se ubican en posiciones equivalentes a las obtenidas para los inhibidores E3Ad y E3Ad_N19D. La densidad electrónica mostró claramente la presencia de tres puentes disulfuro en las posiciones C48-C93, C142-C158 y C148-C152 en las seis estructuras determinadas en el presente trabajo. La presencia de estos tres puentes disulfuro es una característica de todos los inhibidores de proteasas aspárticas, pertenecientes a la familia Kunitz-STI grupo A, aunque recientemente, se reportó que algunos miembros del grupo B con sólo dos puentes disulfuros son capaces de inhibir a la proteasa memapsin-2 (BACE) (Fischer et al., 2015).

En el caso de la estructura obtenida para el inhibidor silvestre con código PDB 5G00, fue posible modelar una molécula de N-Acetilglucosamina unida al residuo N19, pero la ausencia de densidad electrónica nos impidió modelar el resto de los monosacáridos presentes según los resultados obtenidos por espectrometría de masas. En el caso de la otra estructura obtenida para la variante silvestre del inhibidor E3Ad (5FZZ), no fue posible modelar ninguna molécula unida a la asparagina 19.

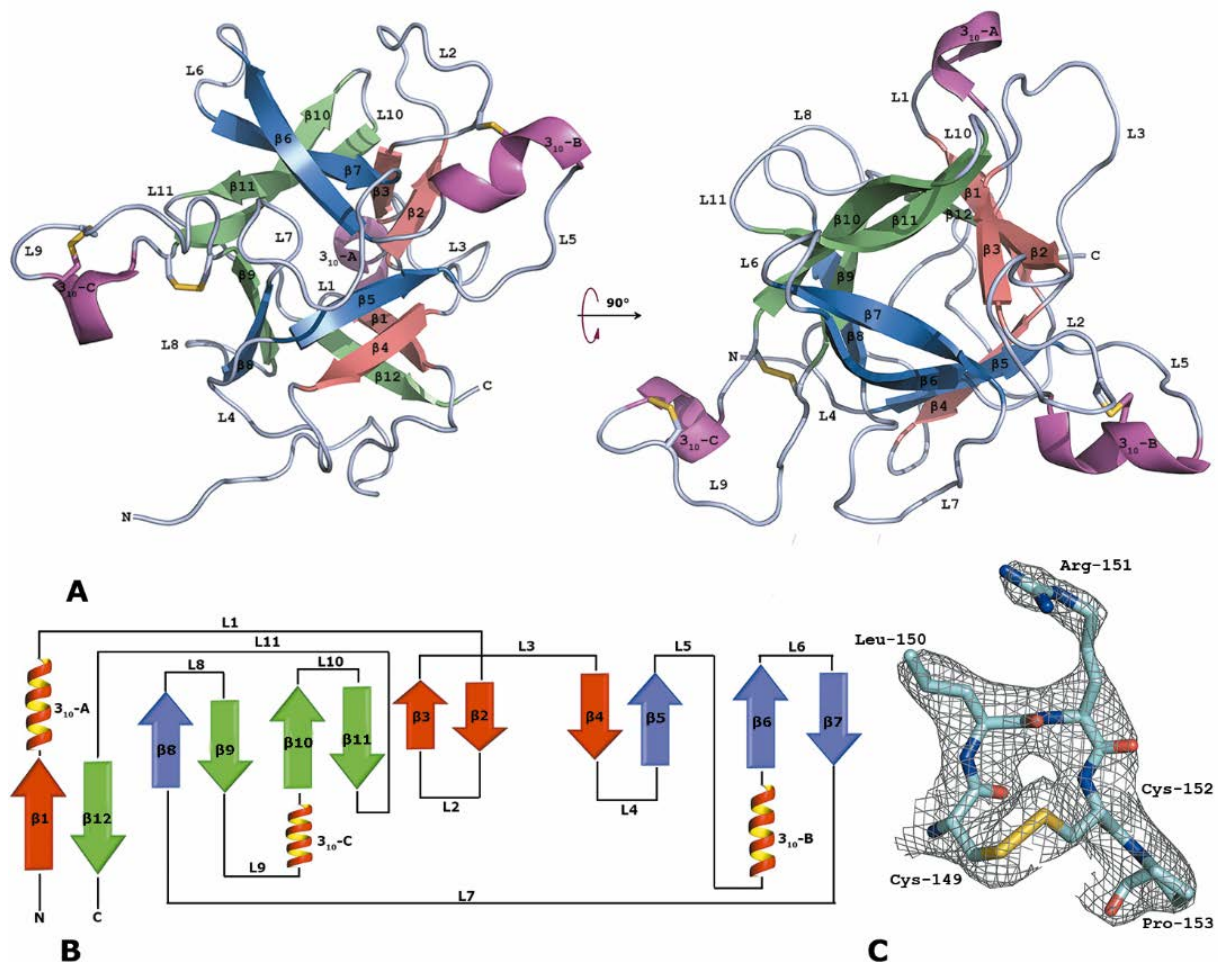


Figura 10. Estructura tridimensional global del inhibidor E3Ad_N19D (PDB ID: 5FNW). En colores similares se muestran los subdominios (módulos o motivos estructurales) relacionados por una pseudo-simetría de tercer orden (A). Topología del inhibidor bifuncional E3Ad (B). Mapa de densidad electrónica $2F_o-F_c$ alrededor de los aminoácidos C149-P153, con un contorno igual a 1σ (C). La figura fue preparada con PyMOL ("The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.," n.d.).

La comparación estructural de las diferentes estructuras obtenidas en el presente trabajo mostró que de forma general la estructura global es muy similar entre ellas, con las diferencias más marcadas en la región de las asas L1, L2 y L5 (Figura 11).

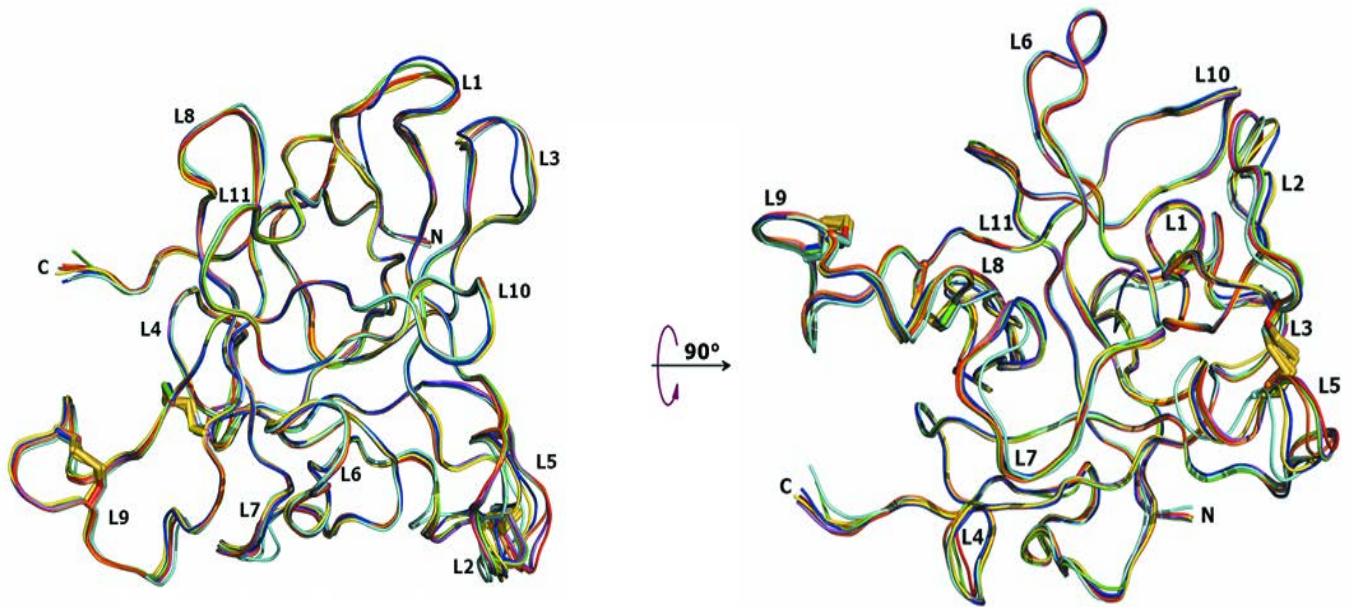


Figura 11. Sobreposición estructural de todos los modelos cristalográficos de los inhibidores E3Ad y E3Ad_N19D. 5FNW (rojo); 5FNX (verde); 5FZY, cadena A (azul); 5FZY, cadena B (amarillo); 5FZU (magenta); 5FZZ (cian) y 5G00 (naranja).

Con el propósito de analizar si las diferencias detectadas entre las diferentes estructuras obtenidas pueden deberse a los distintos valores de pH en las condiciones o a los diferentes arreglos cristalinos, se realizó una comparación estructural por pares (Tabla 2). Como se puede apreciar los resultados mostraron que las diferencias entre las estructuras a diferentes valores de pH y el mismo grupo espacial fueron menores entre sí (0.21-0.35 Å), en comparación con aquellas obtenidas en otros grupos espaciales (0.49-0.68 Å). Además, si consideramos el error estimado para las coordenadas de las estructuras en el grupo espacial $P4_322$ (0.27-0.43 Å), de forma general el r.m.s.d. se encuentra dentro del rango del error, lo que permite afirmar que las diferencias entre ellas no son significativas o al menos mínimas. Por el contrario, el r.m.s.d. entre estructuras de diferentes grupos espaciales superaron de forma general los valores de los errores estimados de coordenadas (0.31-0.59 Å). Esto nos permite sugerir que las diferencias conformacionales observadas en ciertas asas pudieran deberse a los diferentes contactos entre las moléculas como resultado de una distinta organización interna en los cristales.

Tabla 2. Valores de r.m.s.d. de C α de las diferentes estructuras cristalográficas obtenidas para E3Ad y E3Ad_N19D

Código	5FNW	5FZU	5G00	5FNX	5FZY: A	5FZY: B	5FZZ	Grupo Espacial	pH
	r.m.s.d. C α (Å)								
5FNW	-	-	-	-	-	-	-	P4 ₃ 22	3.0
5FZU	0.21	-	-	-	-	-	-	P4 ₃ 22	7.2
5G00	0.30	0.23	-	-	-	-	-	P4 ₃ 22	7.4
5FNX	0.32	0.32	0.35	-	-	-	-	P4 ₃ 22	9.0
5FZY: A	0.67	0.63	0.59	0.60	-	-	-	C222 ₁	3.5
5FZY: B	0.52	0.49	0.48	0.52	0.47	-	-	C222 ₁	3.5
5FZZ	0.68	0.68	0.64	0.52	0.58	0.57	-	P22 ₁ 2 ₁	7.0

4.3 Análisis de secuencia y estructural con respecto a otros miembros de la familia Kunitz-STI

La secuencia del inhibidor E3Ad coincide con la entrada de la base de datos UniProt M1AKE5_SOLTU, que a su vez es resultado de la traducción de las entradas de GenBank XP_006362668.1 y XM_006362606.2. La proteína con mayor identidad de secuencia depositada en UniProt y además caracterizada funcionalmente es el inhibidor API9_SOLTU (95.7 % de identidad), también aislado de papa (Ritonja et al., 1990). El inhibidor API9 es capaz de inhibir tripsina pancreática bovina, quimiotripsina bovina y elastasa de leucocitos humana; todas ellas proteasas de tipo serino de la familia S1. Además, se ha detectado actividad inhibidora frente a proteasas de tipo aspártico como Catepsina D humana y bovina y Sacaropepsina (Cater et al., 2002). El porcentaje de identidad de secuencia entre el inhibidor E3Ad con otros miembros de la familia caracterizados funcional y/o estructuralmente fue: 90.4 % con PDI (Guo et al., 2015), 76.1 % con SLAPI (Cater et al., 2002), 71.2 % con PSPI (Meulenbroek et al., 2012), 28.7 % con RASI (Ohtsubo & Richardson, 1992), 26.1 % con TKI (Patil et al., 2012), 25.5 % con EcTI (Zhou et al., 2013), 24.5 % con API-A (Bao et al., 2009), 23.6 % con BASI (Vallée et al., 1998) y 22.6 % con STI (Song & Suh, 1998) (Figura 12).

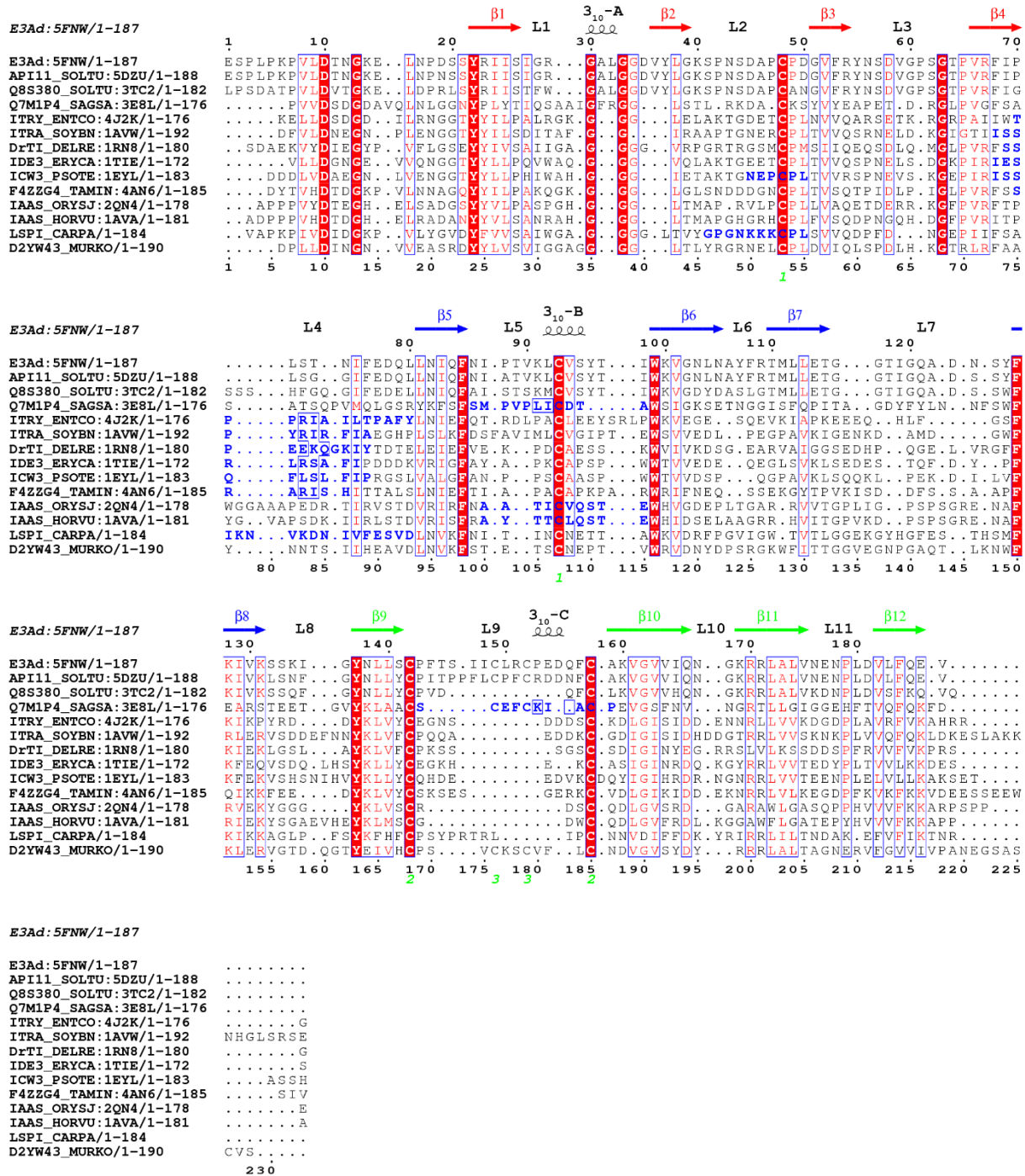


Figura 12. Alineamiento múltiple de secuencia de algunos inhibidores de proteasas de la familia Kunitz-STI con estructura tridimensional determinada. Los elementos de estructura secundaria que componen los motivos estructurales relacionados por una pseudo-simetría de tercer orden se encuentran representados en diferentes colores. La topología de los puentes disulfuro se indica con números en la parte inferior del alineamiento. Los identificadores de la base de datos UniProt se usaron para nombrar a cada secuencia: PDI (API11_SOLITU), PSI (Q8S380_SOLITU), PI-A (Q7M1P4_SAGSA), ECTI (ITRY_ENTCO), STI (ITRA_SOYBN), DRTI (DRTI_DELRE), RASI

(IAAS_ORYSJ), ET I (IDE3_ERYCA), BASI (IAAS_HORVU), WCI (ICW3_PSOTE), TKI (F4ZZG4_TAMIN), LSPI (LSPI_CARPA) y el inhibidor de tripsina de *Murraya koenigii* (D2YW43_MURKO). La figura fue elaborada con el servidor ESript 3.0 (<http://esript.ibcp.fr/ESript/cgi-bin/ESript.cgi>).

Como se puede apreciar en el alineamiento múltiple de secuencia, la familia Kunitz-STI es muy variable, incluso en los sitios de interacción con las proteasas diana. La conformación del asa canónica de interacción con las proteasas y el aminoácido en la posición P1 parecen ser los elementos más conservados en estos inhibidores. En la mayoría de los casos descritos, el asa canónica se ubica entre las hebras β 4- β 5 (L4), como es el caso del miembro representativo de la familia el inhibidor STI. Sin embargo, existen otros miembros donde este asa se encuentra en otras posiciones, como el inhibidor API-A (L5 y L9) (Bao et al., 2009), los inhibidores de subtilisina/ α -amilasa, BASI (L5) (Micheelsen et al., 2008) y el inhibidor LSPI (L2 y L4) (Azarkan et al., 2011). La comparación estructural global mostró que el plegamiento es conservado entre los diferentes miembros, con las principales diferencias localizadas en la conformación de las asas y la presencia/ausencia de hélices en las estructuras. Es de destacar el hecho de que, si bien el inhibidor E3Ad presentó una elevada identidad de secuencia con el inhibidor bifuncional PDI cuya estructura fue liberada recientemente, la superposición de las estructuras mostró una diferencia significativa en la conformación del asa L9 (Figura 13). Es de señalar que es precisamente en esta asa es donde se encuentran 10 de los 17 aminoácidos que son diferentes entre ambas proteínas (Figura 12).

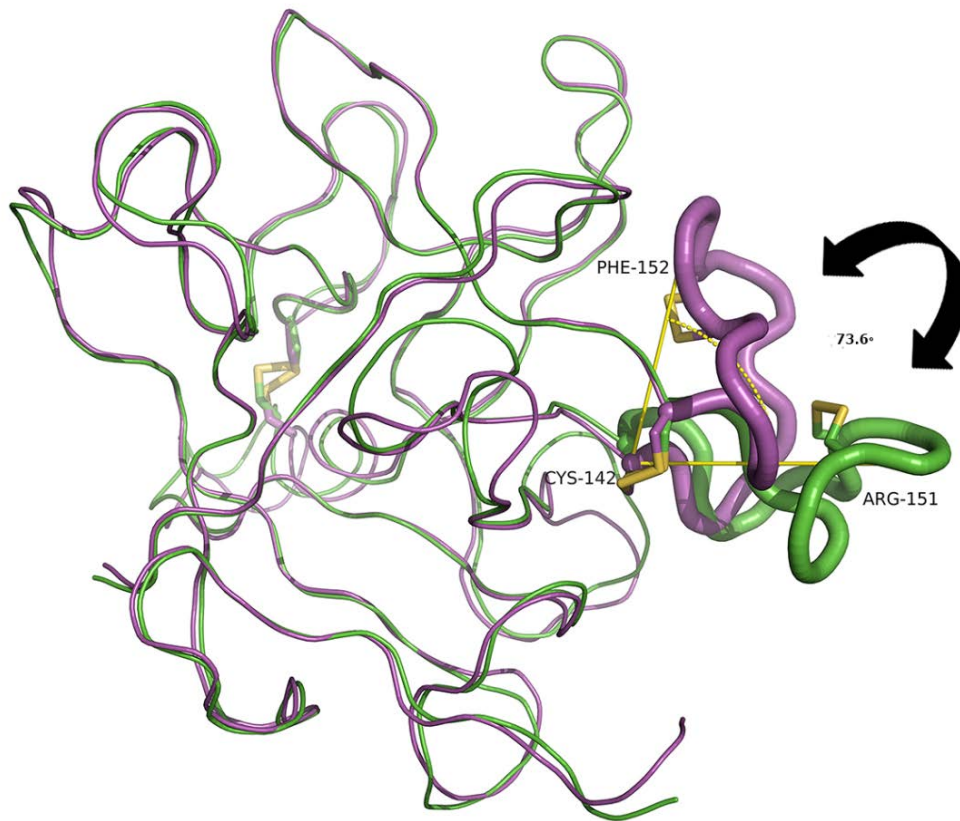


Figura 13. Superposición de las estructuras de los inhibidores bifuncionales PDI (5DZU, magenta) y E3Ad (5FNW, verde). La r egión del asa L9 se encuentra representada con un radio mayor y se muestra el ángulo entre los residuos equivalentes R151 (E3Ad) y F151 (PDI), usando la C142 como vértice.

Una de las interrogantes derivadas de esta diferencia entre las conformaciones en ambas asas es si es debida a los contactos que se establecen entre las moléculas como resultado de determinado orden interno en el cristal o si es una conformación que existe en solución. En nuestro caso decidimos usar simulaciones de dinámica molecular para intentar responderlo. Las simulaciones se realizaron a los valores de pH y temperatura a los cuales se realizaron los ensayos enzimáticos frente a la proteasa serino tripsina bovina y la de tipo aspártico Plasmepsina II. Los resultados de la simulación mostraron que en el caso del inhibidor E3Ad_N19D, la conformación del asa L9 no sufre un cambio significativo, libre de cualquier restricción impuesta por los contactos con otras moléculas dentro del cristal (Figura 14). Por el contrario, en el caso del inhibidor PDI, los resultados de la simulación a un pH 4.4 y una temperatura de 37 °C mostraron un cambio en la conformación del asa L9, en un momento cercano a los 20 ns (Figura 14). El cambio de conformación modelado consistió en un aumento de la exposición al solvente, o en una conformación extendida similar a la adoptada por la

misma asa en las estructuras del inhibidor E3Ad y su mutante no glicosilado. Este cambio de conformación no se manifestó cuando el valor de pH fue 8.0.

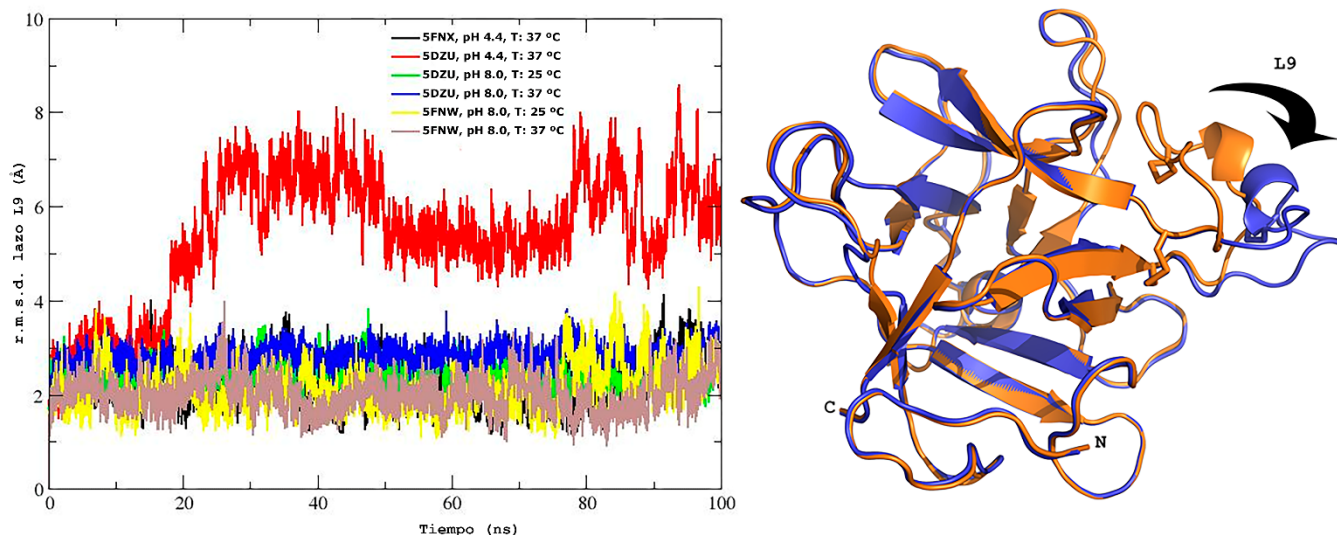


Figura 14. Cambios conformacionales del asa L9 de los inhibidores PDI y E3Ad, durante las simulaciones de dinámica molecular a diferentes valores de pH y temperatura (A). Conformación del asa L9 al inicio (naranja) y al final (azul) de la simulación de dinámica molecular a pH 4.4 y 37 °C (B).

Estamos conscientes de que flexibilidad del asa L9 en función del pH debe ser demostrada experimentalmente, pues no existe un precedente en inhibidores similares, al menos que sea de nuestro conocimiento. Sin embargo, en nuestra opinión esto abre nuevas interrogantes muy interesantes relacionadas con la posible capacidad de un asa de adoptar diferentes conformaciones en función del pH y su secuencia de aminoácidos.

4.4 Predicción del sitio de interacción con proteasas de tipo serino.

Los inhibidores de la familia Kunitz-STI se encuentran clasificados como inhibidores canónicos de las proteasas serino en la base de datos MEROPS y en varios trabajos de revisión. Sin embargo, existen miembros de esta familia que no poseen un asa canónica, o que interactúan con las proteasas serino a través de asas con otras conformaciones no canónicas (Bao et al., 2009).

Considerando que la mayoría de los inhibidores pertenecientes a la familia Kunitz-STI emplean un asa canónica en la posición L4, nuestra primera aproximación fue comparar la conformación de esta asa con las de inhibidores canónicos, específicamente las posiciones P3-P3'. Sin embargo, en el caso del inhibidor E3Ad esta asa es más corta en un residuo y el r.m.s.d. de los C α entre las

posiciones P70-I75 y las posiciones P3-P3' de tres inhibidores canónicos varía entre 1.77 Å y 2.40 Å (Tabla 3).

Tabla 3. Valores de r.m.s.d C^α (Å) del asa L4 (P70-I75) para las estructuras cristalográficas de los inhibidores E3Ad y E3Ad_N19D, con las posiciones P3-P3' del asa canónicas de algunos inhibidores Kunitz-STI.

	Grupo Espacial	P4 ₃ 22				C222 ₁		P22 ₁ 2 ₁
P3-P3'	Inhibidor/PDB ID	5FNX	5FNW	5FZU	5FZZ	5FZY: A	5FZY: B	5G00
P61-F66	STI (1AVW)	1.89	1.83	2.18	2.10	2.17	2.07	2.40
P62-I67	EcTI (4J2K)	1.84	1.79	2.10	2.00	2.07	1.99	2.10
R64-N69	TKI (4AN6)	1.81	1.77	2.09	1.93	2.05	1.93	2.08

Como se puede apreciar en la Figura 15, la conformación del asa canónica en los inhibidores STI, EcTi y TKI es muy similar, siendo al mismo tiempo diferente a la del asa L4 del inhibidor E3Ad_N19D (5FNW).

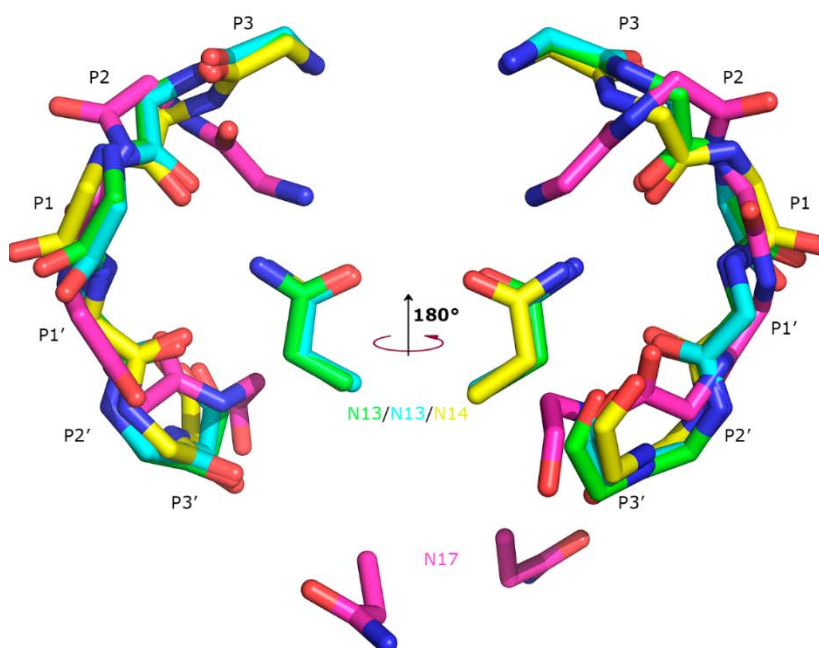


Figura 15. Comparación estructural de las posiciones P3-P3' del asa L4 de i nhibidores canónicos con las posiciones equivalentes en el inhibidor E3Ad. STI (1avw, azul celeste), EcTI (4J2K, verde), TKI (4AN6, amarillo) y E3Ad (5fnw, magenta).

Resulta interesante destacar el hecho de que en el inhibidor E3Ad (y en E3Ad_N19D), si bien existe una asparagina (N17) en una posición adyacente (-1) a la posición equivalente a la N14 que ayuda a estabilizar la conformación canónica en otros inhibidores Kunitz-STI, la conformación de la cadena lateral de la N17 en las estructuras obtenidas en el presente trabajo no se encuentra orientada en la dirección opuesta del asa L4 (Figura 15). En el caso de los inhibidores canónicos esta asparagina resulta esencial para mantener los productos de la hidrólisis a una distancia que permita que se forme nuevamente el enlace peptídico (Dasgupta et al., 2006). También, se ha demostrado que su presencia permite mantener la conformación canónica del asa (Dasgupta et al., 2006). En el caso del inhibidor E3Ad existe otra asparagina en la posición 19, la cual se encuentra unida covalentemente a una N-acetilglucosamina, como se demostró por espectrometría de masas y por difracción de rayos X. La ausencia de este aminoácido de soporte, unido a un asa L4 más corta en los inhibidores bifuncionales como E3Ad, podría explicar las diferencias en la conformación de esta asa con respecto a los inhibidores canónicos. Una alternativa es la presencia de una molécula de agua estableciendo puentes de hidrógeno con los residuos del asa, como ha sido descrito para el inhibidor DrTI (Krauchenco et al., 2003; Renko et al., 2012). El análisis de la densidad electrónica alrededor del asa L4 en algunas de nuestras estructuras, nos permitió modelar una molécula de agua cercana a los aminoácidos N74 (P2'), F76 (P4') y Q79. La posición de esta molécula de agua sugiere que podría estar interaccionando con el oxígeno del grupo carbonilo de la cadena principal de la N74, el grupo imino de la F76 y la cadena lateral de la Q79. Sin embargo, esto podría ser un mecanismo para estabilizar una conformación distinta a una canónica.

Una alternativa como sitio de interacción con las proteasas serino es el asa L9, la cual adopta una conformación canónica en el inhibidor API-A. Sin embargo, como se puede apreciar en la Figura 12, la longitud del L9 entre el inhibidor E3Ad y API-A es diferente en 8 residuos, lo que dificulta la comparación estructural entre ellos y resultando en valores de r.m.s.d. de alrededor de 2.5 Å entre las posiciones equivalentes en esta asa para todas las estructuras obtenidas para E3Ad y E3Ad_N19D.

Otra posibilidad es que el asa L5 que ha sido descrita como un asa de interacción con tripsina para el inhibidor API-A, adopte una conformación canónica. Sin embargo, el asa L5 de E3Ad y las asas canónicas muestran valores de r.m.s.d que son incluso mayores en comparación con los obtenidos para el L4 (Tabla 4). Estos resultados sugieren que las asas L4 y L5 del inhibidor E3Ad no presentan una conformación canónica.

Tabla 4. Valores de r.m.s.d C α (Å) asa L5 (E3Ad, T89-V94) para las estructuras cristalográficas de los inhibidores E3Ad y E3Ad_N19D con las posiciones P3-P3' de asas canónicas.

	Grupo Espacial	P4 ₃ 2 ₂				C222 ₁		P22 ₁ 2 ₁
P3-P3'	Inhibidor /PDB ID	5FNX	5FNW	5FZU	5FZZ	5FZY: A	5FZY: B	5G00
P61-F66	STI (1AVW)	2.46	2.44	2.48	2.42	2.43	2.55	2.76
P62-I67	EcTI (4J2K)	2.52	2.49	2.80	2.49	2.50	2.63	2.52
R64-N69	TKI (4AN6)	2.56	2.50	2.56	2.53	2.55	2.66	2.57

Por otro lado, cuando comparamos estructuralmente el asa L5 de E3Ad con el asa equivalente del inhibidor API-A, los valores de r.m.s.d. oscilaron entre 0.65 y 0.8 Å. Si solamente comparamos las posiciones P3-P2', los valores de r.m.s.d. disminuyeron hasta 0.13-0.42 Å. El inhibidor API-A es capaz de unir a dos moléculas de tripsina de forma simultánea, a través de un asa canónica (L9) y uno no canónico (L5). Cuando analizamos la secuencia del asa L5 nos percatamos de que en E3Ad existe un residuo de lisina (K91) en una posición equivalente con el residuo P1 de API-A (L87) (Figura 12 y 16). Los aminoácidos básicos (K,R), son frecuentes en las posiciones P1 de los inhibidores canónicos y se ha reportado que pueden ser responsables de hasta el 50 % del área de la interfaz proteasa-inhibidor (Krowarsch et al., 2003). Generalmente, aminoácidos como la leucina presentes en la posición P1 del asa no canónica L5 del inhibidor API-A, es más frecuente en sitios de interacción con proteasas como quimiotripsina; sin embargo, la densidad electrónica de este residuo es clara en la estructura del complejo terciario (Bao et al., 2009). Estudios de mutagénesis mostraron que la sustitución de esta leucina por arginina aumentó su afinidad por tripsina, mientras que disminuyó por quimiotripsina, (Xie, Luo, Xu, & Chi, 1997). En el caso del inhibidor PDI, se ha propuesto que el asa responsable de la interacción con tripsina podría ser el L4 o el L5, aunque los autores sugieren que el L5 también podría estar involucrado con la interacción con la proteasa aspártica catepsina D (Guo et al., 2015). La superposición de los C α correspondientes a las posibles posiciones P3-P3' de E3Ad (T89-V94), con las regiones equivalentes de los inhibidores PDI y PSPI, ambos aislados de papa, mostraron una gran similitud estructural entre ellos y con API-A (Figura 16).

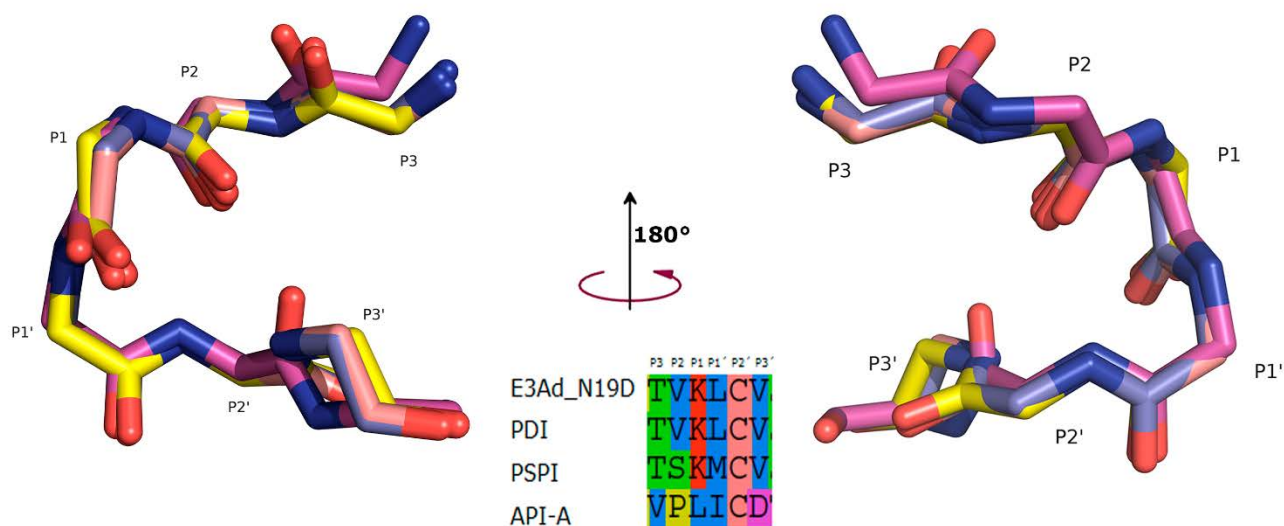


Figura 16. Comparación estructural de las posibles posiciones P3-P3' en el asa L5 de los inhibidores E3Ad (5FZU, amarillo), PDI (5DZU, salmón) y PSPI (3TC2, azul) con sus posiciones equivalentes en el asa no canónica de API-A (3E8L, magenta). Detalles de la secuencia de aminoácidos para las posiciones P3-P3' para cada inhibidor.

Una comparación de las interacciones que se establecen entre las posiciones P3-P3' entre estos inhibidores de papa y el API-A, mostró que algunas de ellas se conservan en los cuatro inhibidores, como por ejemplo la interacción entre el oxígeno del grupo imino del aminoácido en posición P2' con el grupo carbonilo de la posición P2 (Figura 17). En el caso del inhibidor PSPI, el grupo hidroxilo de la cadena lateral de la S94 (P2), establece un puente de hidrógeno adicional con el grupo imino en P2'. En los inhibidores PSPI, PDI y E3Ad, la presencia de una treonina en la posición P3, permite la formación de un puente de hidrogeno entre el hidroxilo de cadena lateral y el carbonilo de la cadena principal de P2. Para el inhibidor API-A la presencia de una valina en esta posición no permite establecer esta interacción. En el caso del inhibidor APIA la mayoría de los residuos en las posiciones P2'-P4' interactúan fundamentalmente con aminoácidos del asa L7 (D113 y F115), excepto el caso de la formación de un puente disulfuro con la C42, del asa L2. Este puente disulfuro entre la posición P2 y una cisteína del asa L2 se encuentra estrictamente conservado en los cuatro inhibidores y contribuye a la conformación del asa L5. En el caso de los 3 inhibidores de papa, se mantienen estas interacciones entre al menos un residuo de la región P4'-P5' y un aminoácido del asa L7. Para los inhibidores PDI y PSPI la interacción ocurre entre el grupo carbonilo de la posición P4' y una glicina, mientras que para el inhibidor E3Ad es entre el carbonilo de la posición P5' y el hidroxilo de la T115. En general, en los inhibidores PSPI, PDI y E3Ad hay un mayor número de interacciones entre los

residuos del mismo asa en comparación con API-A. De cualquier forma, las interacciones en todos ellos permiten conservar una conformación muy similar en la región P3-P3', a pesar de los cambios en la secuencia de aminoácidos.

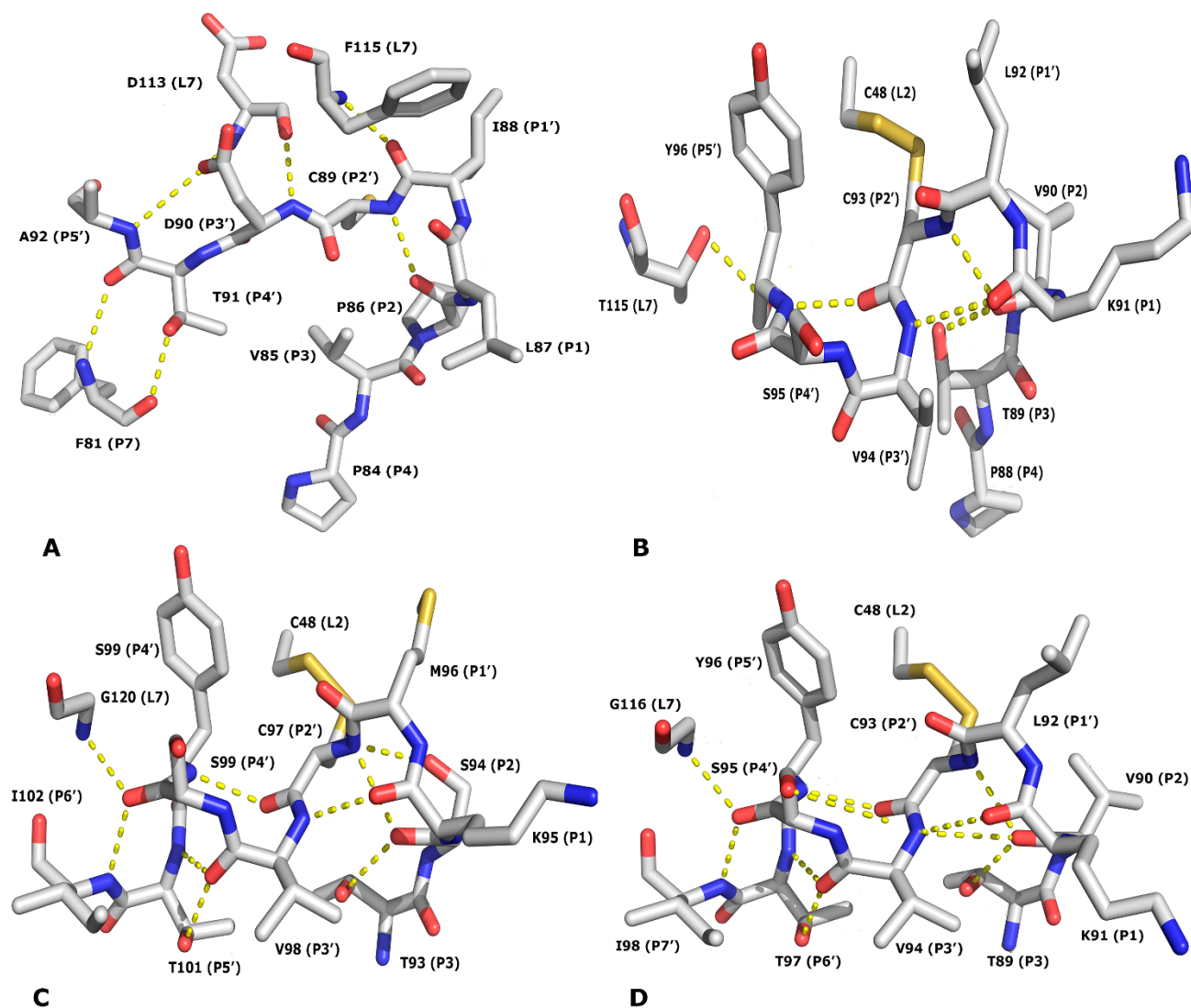


Figura 17. Representación de las interacciones que estabilizan el asa L5 de los inhibidores API-A (3E8L, A), E3Ad (5FZU, B), PSPI (3TC2, C), PDI (5DZU, D). Se hace énfasis en las posiciones P3-P3'.

Basados en la similitud estructural del asa L5 de E3Ad con el asa L4 del inhibidor API-A, obtuvimos un modelo de un complejo tripsina bovina-E3Ad_N19D. El análisis de los contactos del asa L5 con la tripsina muestra que la mayoría de los contactos en la interfaz son similares a los de API-A (Figura 18 A y C), lo cual era de esperarse debido a la similitud entre ellas. Una diferencia en los contactos es la ausencia de la interacción entre el Nε2 de la Q189 y el oxígeno del carbonilo de la C93I del inhibidor E3Ad_N19D. Es precisamente la orientación del grupo carbonilo de la cisteína en posición

P2 una diferencia entre el inhibidor API-A y los inhibidores de papa E3Ad, PDI y PSPI (Figura 16). En el modelo del complejo tripsina-E3Ad_N19D, la Q 189 se encuentra interactuando con el carbonilo de la K91 (Figura 18C).

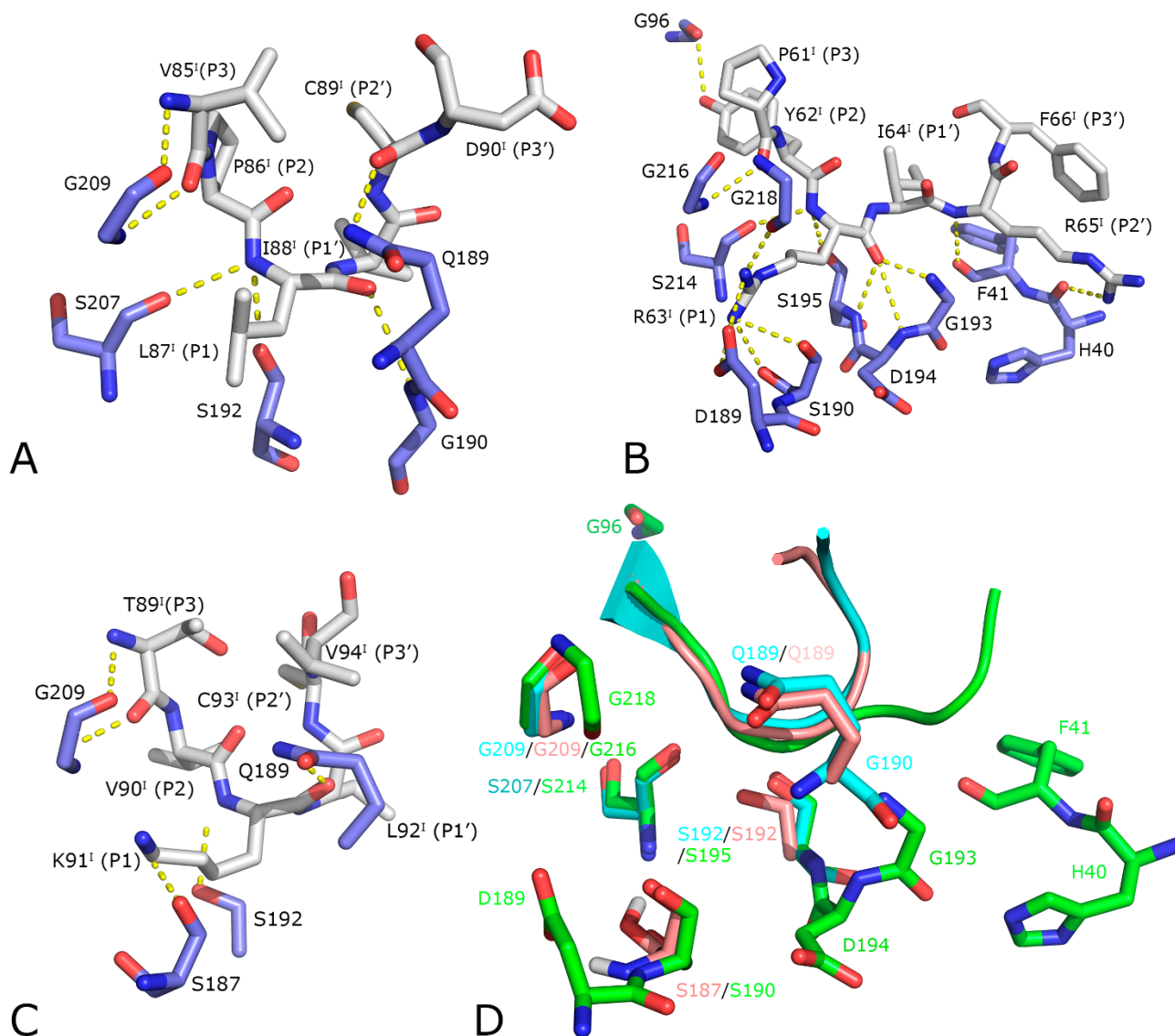


Figura 18. Interfaz proteasa-inhibidor de los inhibidores canónicos y no canónicos de proteasas serino. Complejos tripsina bovina-API-A (3E8L) (A) tripsina porcina-STI (1AVW) (B), modelo por comparación del complejo Tripsina-E3Ad_N19D (C). Superposición de las interfaces anteriores: Tripsina porcina-STI (verde), Tripsina bovina-API-A (celeste), modelo por comparación de complejo Tripsina-E3Ad_N19D (salmón). Las etiquetas de los aminoácidos correspondientes al inhibidor presentan la letra I como superíndice.

En comparación con los inhibidores canónicos, los inhibidores no canónicos establecen un menor número de interacciones proteasa-inhibidor, siendo en su mayoría interacciones entre la cadena

principal del asa del inhibidor y el centro activo de la proteasa (Figura 18A, B). Otra diferencia es que la conformación del asa no canónica en las posiciones P2'-Pn parece alejarse del centro activo de la proteasa (Figura 3 y 18D), mientras que en los inhibidores canónicos se mantiene sobre la superficie de la proteasa.

El análisis de un alineamiento múltiple de las secuencias presentadas en la Figura 12 con el servidor ProtTest propuso que el modelo evolutivo que mejor describía nuestros datos es el de Whelan & Goldman (WAG) (Whelan & Goldman, 2001), con una distribución gamma (+G, 4 categorías, parámetro=3.24) (Yang, 1994) con sitios invariables (+I, 8.34 % de los sitios) (Reeves, 1992). La topología del árbol filogenético consenso construido usando el método de máxima verosimilitud sugiere que los inhibidores de papa E3Ad, PDI y PSPI se agrupan en un mismo subgrupo (*phyla*) independiente respecto a los inhibidores con asas canónicas como STI, EcTI y TKI (Figura 19). Esto apunta una ruta evolutiva distinta para ambos grupos de inhibidores canónicos y no canónicos. Una reconstrucción anterior usando un modelo evolutivo más simple (Jones-Taylor-Thorton) agrupó al inhibidor API-A junto a los inhibidores de papa E3Ad, PDI y PSPI (Guerra, Valiente, Pons, Berry, & Rudiño-Piñera, 2016).

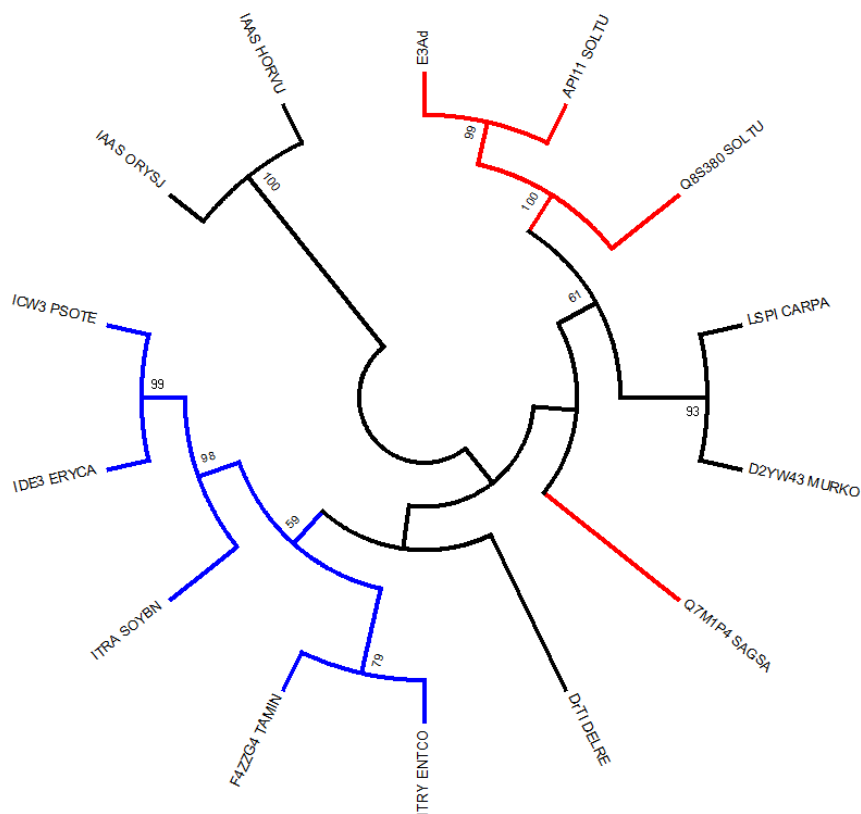


Figura 19. Árbol filogenético consenso (1000 réplicas) usando el método de máxima verosimilitud de inhibidores de la familia Kunitz-STI con estructura tridimensional. En colores se indican los inhibidores

canónicos (azul) y no canónicos (rojo). Para la reconstrucción se empleó el método de máxima verosimilitud y el modelo de evolución WAG +G +I.

Resultados similares fueron obtenidos de forma independiente para otro inhibidor bifuncional de la familia Kunitz-STI (Azarkan et al., 2011). En resumen, la comparación estructural del inhibidor E3Ad con otros miembros de la familia Kunitz-STI sugiere que este podría usar su asa L5 para inhibir a las proteasas serino como la tripsina, usando una conformación no canónica similar al inhibidor API-A.

Es importante destacar el hecho de que si bien se dispone de la estructura de un complejo entre el inhibidor PDI y la tripsina (PDB ID: 5H9Z), la cual fue depositada a finales del año 2015, no existe una publicación asociada y sus autores no autorizaron la liberación de las coordenadas al cumplirse un año de su depósito. Consideramos que esta estructura permitirá verificar o refutar nuestra propuesta acerca del mecanismo de interacción de estos inhibidores con las proteasas serino.

4.5 Predicción del sitio de interacción con proteasas de tipo aspártico.

La predicción de la región que participa en la interacción con las proteasas de tipo aspártico es un poco más complicada con respecto a las proteasas serino, debido a que no existe información estructural de esta interacción para inhibidores de la familia Kunitz-STI. Los inhibidores proteicos de proteasas aspártico con estructura tridimensional determinada hasta el momento poseen plegamientos distintos entre sí y con los inhibidores de la familia Kunitz-STI. La obtención reciente de la estructura tridimensional del inhibidor bifuncional PDI permitió sugerir a los autores posibles sitios de interacción con las proteasas aspártico a las asas L5 y L9 (Guo et al., 2015). Sin embargo, la conformación adoptada por el asa L9 en la estructura de PDI, no favorece una interacción con el sitio activo de una proteasa de tipo aspártico, como muestran la solución de acoplamiento molecular (“docking”) obtenida por ellos (Guo et al., 2015).

Con anterioridad, nuestro grupo había propuesto al asa L9 como el responsable de la interacción entre un inhibidor bifuncional aislado de tomate, basados en una combinación de herramientas computacionales (Guerra et al., 2012). Teniendo en cuenta la disponibilidad de estructuras tridimensionales experimentales para PDI y E3Ad, empleamos el servidor metaPPISP para predecir posibles regiones de interacción proteína-proteína para ambos inhibidores. En la Figura 20 se puede observar que, independientemente de las diferencias conformacionales, el asa L9 en ambos inhibidores es la que con mayor frecuencia alcanza una puntuación superior a 0.34. Ciertamente, este metaservidor no nos predice las regiones que interaccionarán específicamente con otra proteína en particular, pero nos dan una aproximación de aquellas regiones con mayores probabilidades de estar involucradas en una interacción proteína-proteína. Es de destacar el hecho de que el asa L5

no aparece predicho como una posible región de interacción proteína-proteína. Estos resultados son similares a los obtenidos por nuestro grupo al emplear diferentes modelos por comparación de un inhibidor bifuncional de tomate (Guerra et al., 2012), indicando que la predicción del asa L9 es independiente de los cambios en su conformación.

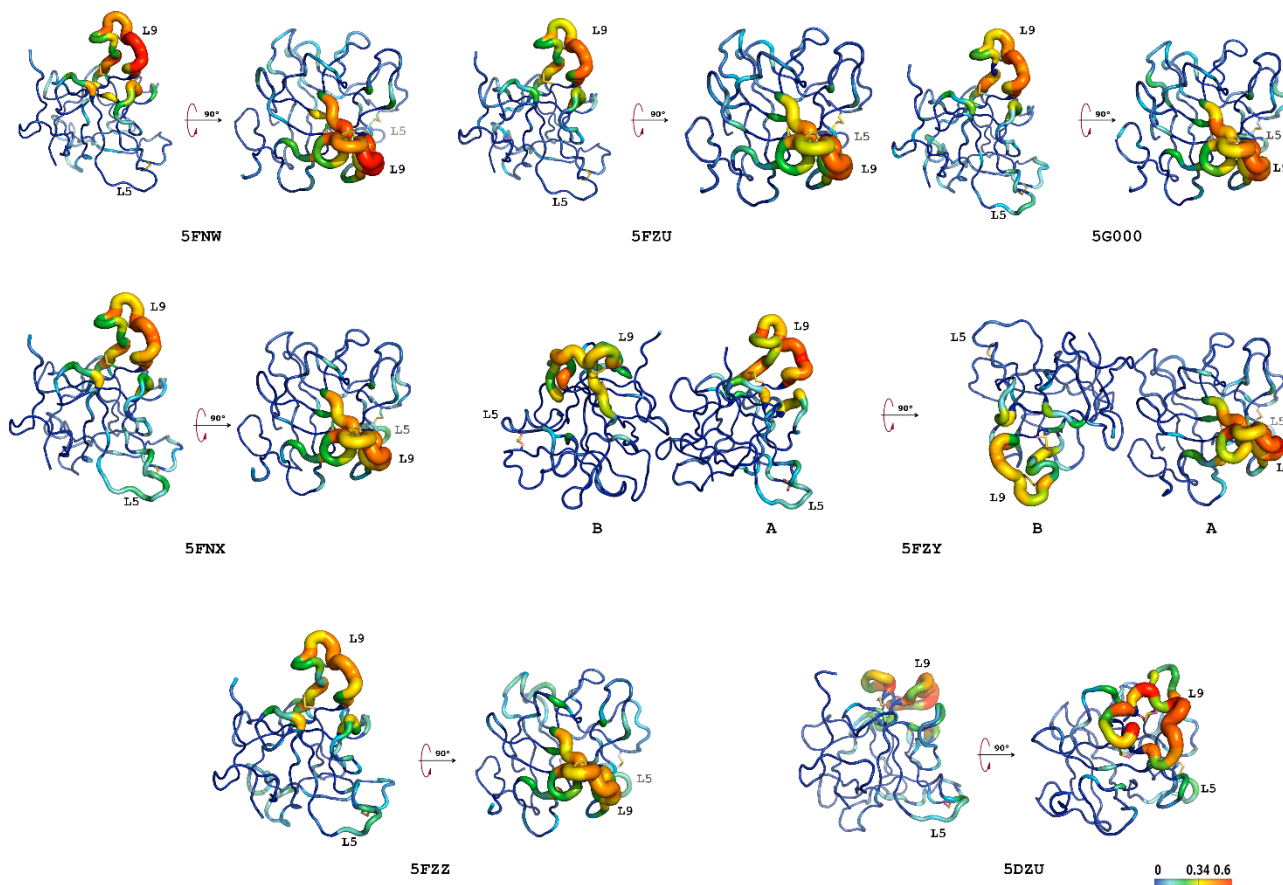


Figura 20. Predicción de posibles residuos de una interfaz proteína-proteína en los inhibidores PDI y E3Ad, usando el servidor metaPPISP. El servidor usa un límite de corte igual a 0.34.

Adicionalmente, realizamos un análisis de acoplamiento molecular de cuerpos rígidos empleando el servidor ClusPro (Kozakov et al., 2013), considerando que este es el servidor con mejores resultados en las competencias CAPRI hasta la fecha (Vajda et al., 2017), pero también para comparar nuestros resultados con los obtenidos para el inhibidor PDI (Guo et al., 2015). Los resultados mostraron que las soluciones con mejor puntuación ubicaron al asa L9 en el centro activo de la proteasa Plm II, para la mayoría de las estructuras del inhibidor. Este análisis mostró que la R151 interacciona con los aspárticos catalíticos (D34 y D214) de la enzima (Figura 21A). Un mecanismo de inhibición similar fue obtenido para el inhibidor PDI usado el mismo servidor, pero la interacción en ese modelo era

entre la K91 y los aspárticos catalíticos de la catepsina D (Guo et al., 2015). La interacción de un aminoácido básico (R, K) con los residuos catalíticos de las proteasas de tipo aspártico ya ha sido descrita en el mecanismo de inhibición de los zimógenos progastricina (PDB: 1HTR) (Moore, Sielecki, Chernai, Tarasova, & James, 1995) y pepsinógeno (PDB: 2PSG). Sin embargo, nuestro análisis sugiere que el asa L5, y específicamente la K91, podrían estar involucradas en la interacción con la tripsina y no con las proteasas aspárticas. Podría ser posible que una misma región esté involucrada en la interacción con proteasas de diferentes clases mecanísticas, como se ha propuesto para un inhibidor de papaína aislado de *Prosopis juliflora* (Franco et al., 2002). Sin embargo, en nuestra opinión tener los sitios de interacción en regiones distintas sería más conveniente evolutivamente que el que se encuentren sobrepuestos.

Teniendo en cuenta los resultados de las simulaciones de dinámica molecular y la flexibilidad del asa L9, nos preguntamos si la solución del anclaje molecular obtenida para el inhibidor PDI podría haber sido diferente si el asa L9 tuviera la conformación modelada a pH 4.4. Para respondernos esta pregunta tomamos las coordenadas del inhibidor a los 93.77 ns de la simulación y lo usamos como la estructura de PDI para un anclaje molecular frente a PlmII. Como se puede apreciar en la Figura 21B, la solución con mejor puntuación ubica al asa L9 en el centro activo de la enzima, aunque no se observa interacción directa con los residuos catalíticos. Este modelo sugiere que a pH 4.4, el inhibidor podría interactuar con las proteasas aspártico a través de su asa L9 en lugar del asa L5.

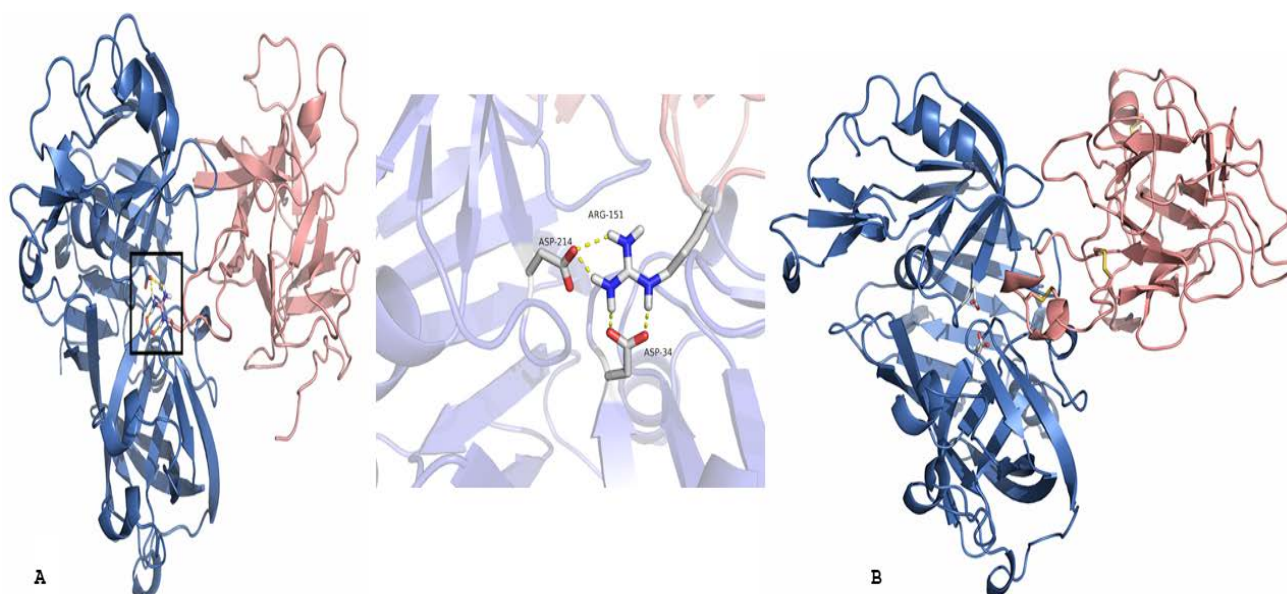


Figura 21. Modelos obtenidos por anclaje molecular utilizando las coordenadas cristalográficas de la proteasa de tipo aspártico Plm II y del inhibidor E3Ad_N19D (5FNW) (A). Para el inhibidor PDI se utilizaron las coordenadas a los 93.77 ns de una simulación de dinámica molecular a pH 4.4 y 37 °C (B).

4.6 Caracterización cinética del mutante E3Ad_N19D_R151A_His

Con el propósito de validar nuestro modelo de acoplamiento molecular mutamos la arginina en la posición 151 por alanina en la variante no glicosilada del inhibidor. Además, agregamos una etiqueta de histidinas en el extremo carboxilo de los inhibidores para facilitar la purificación de los mismos. Curiosamente los resultados de la caracterización cinética mostraron que la sustitución de la R151 por alanina no produce una disminución de la afinidad del inhibidor por la Plasmepsina II, sino más bien un pequeño incremento en la afinidad (Figura 22). Los valores de K_i fueron estadísticamente diferentes pero muy cercanos entre sí, 4.45 ± 0.06 nM y 1.57 ± 0.22 nM para E3Ad_N19D_His y E3Ad_N19D_R151A_His, respectivamente. Estos valores de K_i descartan el hecho de que la arginina 151 sea un residuo determinante en la interacción entre el inhibidor E3Ad_N19D_His y la proteasa aspártica Plm II. Un análisis más detallado del comportamiento del ajuste de los datos cinéticos a la ecuación de Morrison para los inhibidores de unión fuerte, mostró un cambio en la concavidad de la curva para el mutante R151A respecto al silvestre, lo que se puede justificar por un cambio en la relación $[E]_0/K_i^{app}$. En el caso de E3Ad_N19D_His, el ajuste mostró una relación igual a 7.8, mientras que para el mutante E3Ad_N19D_R151A_His este valor se incrementó hasta 38.4. Este cambio de comportamiento es típico de los inhibidores de unión fuerte y en nuestra opinión es una evidencia adicional de que hay un aumento en la afinidad de la interacción entre el mutante E3Ad_N19D_R151A_His y la Plm II. Los resultados de la determinación de las constantes de inhibición de las variantes con etiqueta de histidina también sugieren que la adición de esta tiene un efecto, aunque mínimo, sobre la afinidad del inhibidor por la Plm II. Sin embargo, no tenemos una explicación estructural para este efecto hasta el momento.

Lamentablemente, debido a las limitaciones de sensibilidad del sustrato, no fue posible usar una menor concentración de enzima en los ensayos para alcanzar una $[E]_0/K_i^{app}$ similar en ambos ensayos. Un análisis preliminar de la K_i del inhibidor E3Ad_N19D_R151A, comprobó la tendencia a un aumento en la afinidad del inhibidor con el cambio R151A (3.8 veces) (Figura 22C) y además un cambio en la concavidad de la curva como resultado de un aumento en la relación $[E]_0/K_i^{app}$.

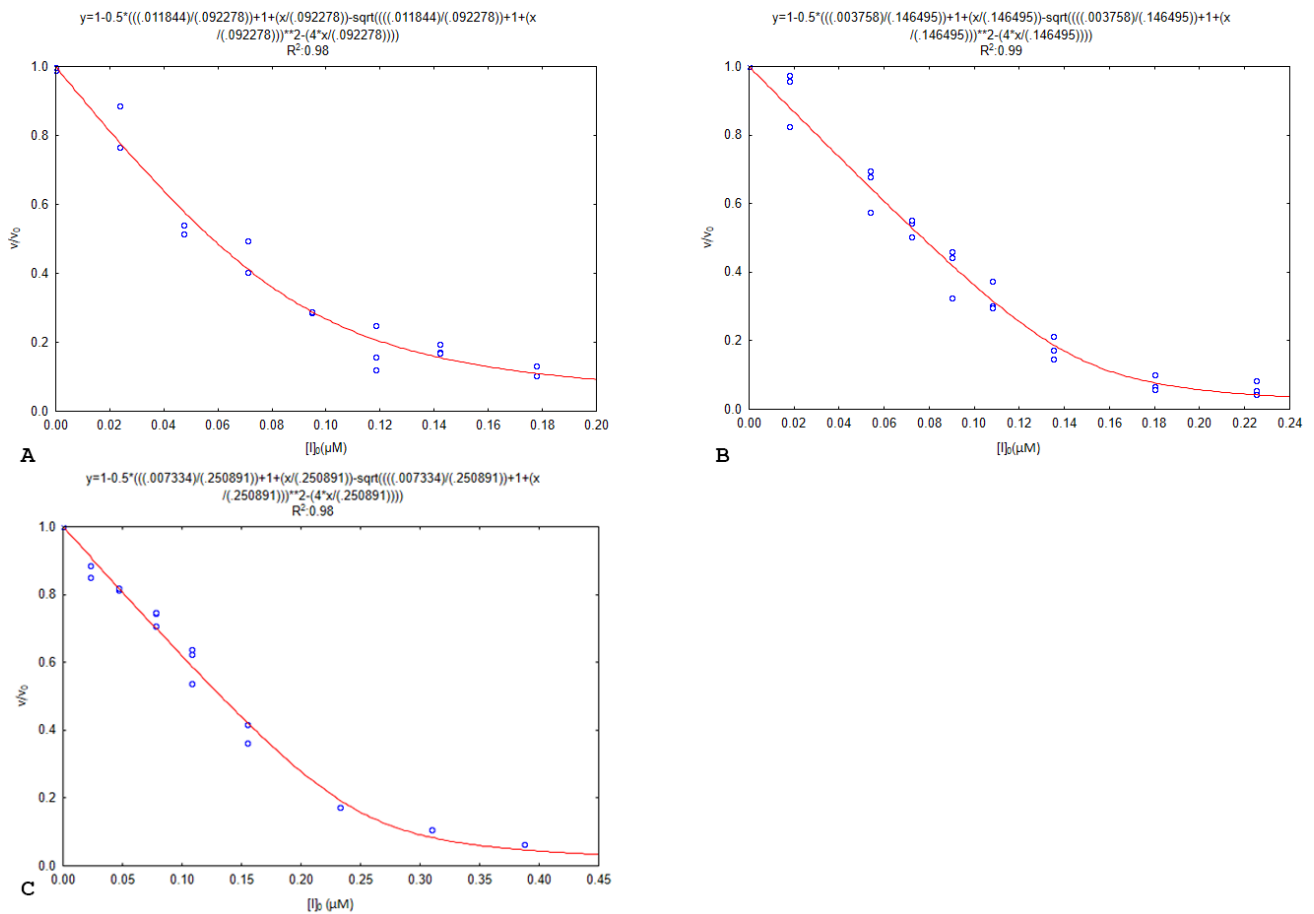


Figura 22. Ajuste a la ecuación de Morrison de los resultados de los ensayos de inhibición frente a la proteasa P Im I I de los inhibidores E3Ad_N19D_His₆ (A), E3Ad_N19D_R151A_His₆ (B), E3Ad_N19D_R151A (C).

Una explicación alternativa a este comportamiento cinético es que ocurra algún tipo de cambio conformacional en el modo de interacción como resultado de este cambio en la secuencia del inhibidor. Se obtuvo un posible complejo entre la PIm II y el mutante E3Ad_N19D_R151A (generado a partir de la estructura 5FNW) usando el servidor ClusPro, y se observó que el asa L9 estuvo ubicada en el centro activo de la enzima. Sin embargo, el modelo muestra cambios respecto al obtenido para E3Ad_N19D, específicamente un cambio en la conformación del asa L9 (F144-C149) y una mayor área de contacto entre ambas proteínas (2293 vs 2081 Å²), según el análisis realizado con el servidor ProFace (<http://resources.boseinst.ernet.in/resources/bioinfo/interface/>) (Saha, Bahadur, Pal, Mandal, & Chakrabarti, 2006). El modelo del complejo con el mutante R151A muestra que la cadena lateral de la A151 se ubica cerca del anillo aromático de la Y77 (Figura 23).

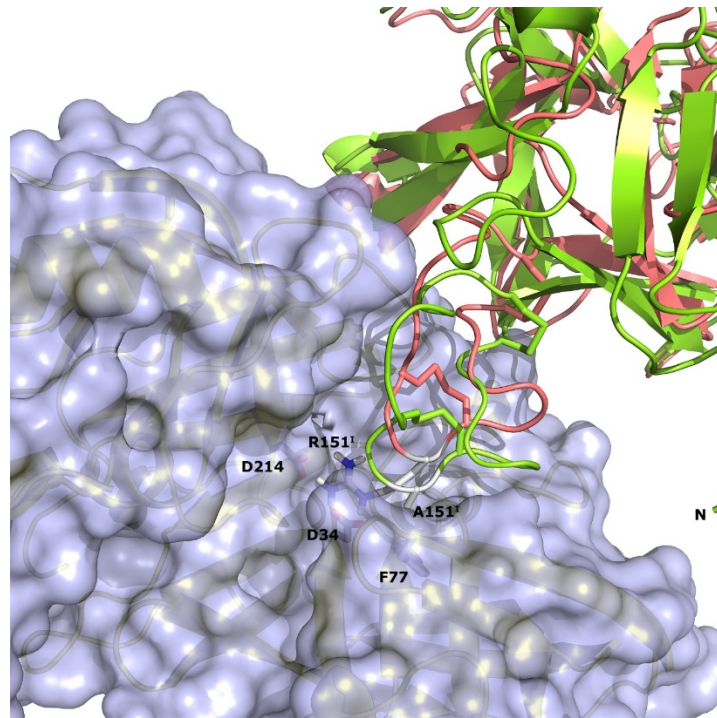


Figura 23. Modelos obtenidos por acoplamiento molecular utilizando las coordenadas cristalográficas de la proteasa de tipo aspártico P1m II y del inhibidor E3Ad_N19D (5FNW) (salmón) y el mutante E3Ad_N19D_R151A (verde). Para el mutante E3Ad_N19D_R151A se utilizó un modelo obtenido a partir la estructura 5FNW de E3Ad_N19D. Las etiquetas de los aminoácidos correspondientes al inhibidor presentan la letra I como superíndice.

El análisis de secuencia del asa L9 en aquellos inhibidores donde se ha de mostrado experimentalmente la actividad inhibidora de proteasas aspártico de la familia A1, mostró que la posición 152 (posición 151 en E3d) no está conservada (Figura 24). De hecho, solo los residuos C142, P143, T145, D155 y C158 se encuentran estrictamente conservados en el asa L9. Si el análisis se realiza separándolos por los grupos a los que pertenecen, las posiciones P147, C149 y C152 se conservaron para los inhibidores del grupo A. Esta variabilidad de secuencia podría ser la explicación a la diferencia en potencia entre el inhibidor SLAPI y API9 frente a las proteasas sacaropepsina y catepsina D humana. Conocer el modo de interacción de estos inhibidores requerirá de estudios estructurales adicionales, lo más probable la determinación de la estructura de un complejo proteasa-inhibidor.

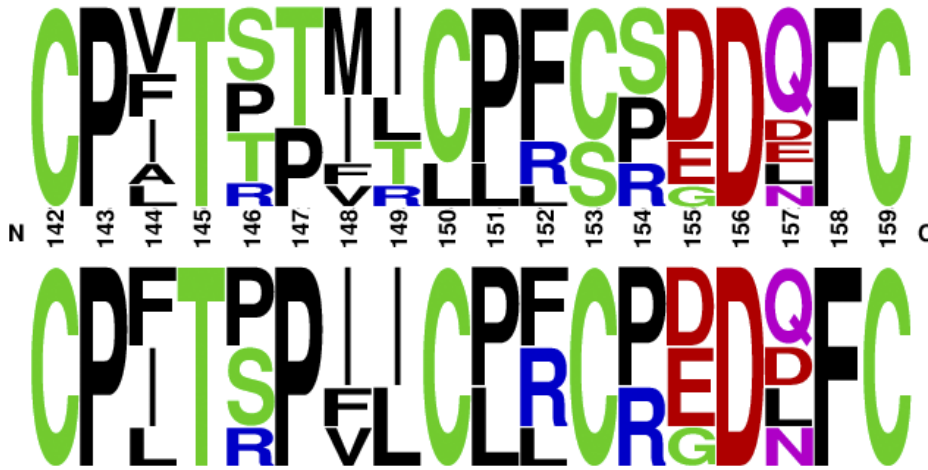


Figura 24. Esquema de la frecuencia de los aminoácidos en cada posición del asa L9. El logo superior representa la frecuencia para todos los inhibidores capaces de inhibir proteasas aspártico, mientras que el inferior es para aquellos que pertenecen al grupo A. La figura se elaboró con el servidor WebLogo (<http://weblogo.berkeley.edu/logo.cgi>) (Crooks, Hon, Chandonia, & Brenner, 2004).

Un aspecto interesante derivado de este trabajo es la posibilidad de que el uso de un asa no canónica para inhibir proteasas serino por miembros de la familia Kunitz sea algo relativamente común. Además, que el asa de interacción con las proteasas aspárticas se encuentre en una posición equivalente con un asa canónica existente en otro inhibidor de la misma familia para inhibir a las proteasas serino, lo que abre nuevas interrogantes acerca de su posible relación evolutiva. Otros autores han sugerido la posibilidad de que los inhibidores de proteasas aspártico que pertenecen a familias cuyas proteasas diana primarias son de otra clase mecanística, pueden haber sido “reclutados” para interactuar con nuevas proteasas a partir de la modificación de los sitios de interacción con las proteasas primarias de la familia (Christeller, 2005). Una de las ventajas evolutivas propuestas para plegamientos que se han originado como resultado de procesos de duplicación y fusión génica es la adquisición o el desarrollo de nuevas funciones (Eisenbeis & Höcker, 2010). El razonamiento detrás de ello es que como homotrímico, la región funcional es idéntica en las tres cadenas por lo que un cambio en una afecta a todas. Sin embargo, posterior al proceso de duplicación y fusión, cada región funcional puede cambiar de forma independiente. En el caso particular de un inhibidor como API-A, uno de los sitios de interacción con proteasas serino puede cambiar su especificidad para inhibir a una nueva clase mecanística, sin perder la capacidad de inhibir a las proteasas serino. Si algo similar es lo que ocurrió con los inhibidores E3Ad y PDI es algo que queda por demostrar.

Una posible ventaja práctica de que los sitios de interacción con las proteasas serino y aspártico se ubiquen en subdominios distintos es que posibilitaría el diseño de nuevos inhibidores totalmente

simétricos con un número variable de sitios de interacción con proteasas serino y/o aspártico. Esto se ha logrado con otras proteínas con el plegamiento β -trefoil (Broom et al., 2012; Lee et al., 2011; Terada et al., 2017), por lo que sería una poderosa herramienta en la obtención de nuevos inhibidores multifuncionales.

5. Conclusiones

- El análisis estructural y cinético de las variantes glicosilada y no glicosilada del inhibidor E3Ad mostró que las glicosilaciones no son elemento esencial en la conservación de la estructura-función de este inhibidor. Estas glicosilaciones podrían estar involucradas en la estabilización y/o en mecanismos de señalización en el ambiente natural de estos inhibidores.
- Los análisis estructurales y de filogenia sugieren que los inhibidores bifuncionales, como E3Ad y PDI, podrían usar un mecanismo de inhibición para las proteasas serino diferente al mecanismo estándar empleado por la mayoría de los miembros de la familia Kunitz-STI. Se propone que la interacción de estos inhibidores bifuncionales con la tripsina podría ocurrir a través del asa L5, la cual presenta una conformación no canónica similar a la descrita para el inhibidor API-A.
- Los resultados obtenidos no apoyan el modo de interacción donde la R151 interactúa con ambos sitios catalíticos, abriendo la posibilidad de la existencia de cierto nivel de tolerancia en la conservación de la secuencia de la interface proteasa-inhibidor.
- Las simulaciones de dinámica molecular sugieren que el asa L9 de los inhibidores bifuncionales de proteasas serino y aspártico es capaz de adoptar diferentes conformaciones en función del valor del pH y que esta adaptabilidad conformacional pudiera ser dependiente de la secuencia del asa. Esta capacidad podría ser útil en el diseño de proteínas que adopten diferentes conformaciones a conveniencia de su aplicación.

6. Perspectivas

- Obtener la estructura de un complejo PlmII-E3Ad_N19D/R15A_His₆ mediante cristalografía de proteínas. La obtención de ambos complejos permitiría establecer los detalles estructurales de la interfaz proteasa-inhibidor, profundizar acerca de la tolerancia en la conservación de la interfaz y la posible existencia de cambios conformacionales como resultado de la mutación R151A.
- Realizar estudios de SAXS del complejo PlmII-E3Ad_N19D/R151A y Tripsina-E3Ad_N19D/R151A, como una vía de baja resolución, pero más rápida de obtener información estructural de las regiones de interacción del inhibidor con ambas proteasas.
- Explorar la posibilidad de obtener y caracterizar estructuralmente un complejo ternario PlmII-E3Ad_N19D_Tripsina.
- Realizar estudios termodinámicos de la interacción proteasa-inhibidor

7. Bibliografía

- Abascal, F., Zardoya, R., & Posada, D. (2005). ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics*, 21(9), 2104–2105. <http://doi.org/10.1093/bioinformatics/bti263>
- Abbenante, G., & Fairlie, D. P. (2005). Protease inhibitors in the clinic. *Medicinal Chemistry (Sharjah (United Arab Emirates))*, 1(1), 71–104. <http://doi.org/10.2174/1573406053402569>
- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., ... Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, 66(Pt 2), 213–221. <http://doi.org/10.1107/S0907444909052925>
- Ahn, J. E., Salzman, R. A., Braunagel, S. C., Koiwa, H., & Zhu-Salzman, K. (2004). Functional roles of specificbruchid protease isoforms in adaptation to a soybean protease inhibitor. *Insect Molecular Biology*, 13(6), 649–657. <http://doi.org/10.1111/j.0962-1075.2004.00523.x>
- Amirhusin, B., Shade, R. E., Koiwa, H., Hasegawa, P. M., Bressan, R. A., Murdock, L. L., & Zhu-Salzman, K. (2007). Protease inhibitors from several classes work synergistically against *Callosobruchus maculatus*. *Journal of Insect Physiology*, 53(7), 734–740. <http://doi.org/10.1016/j.jinsphys.2007.03.008>
- Ardelt, W., & Laskowski, M. (1985). Turkey ovomucoid third domain inhibits eight different serine proteinases of varied specificity on the same ...Leu18-Glu19 ... reactive site. *Biochemistry*, 24(20), 5313–5320.
- Azarkan, M., Martinez-Rodriguez, S., Buts, L., Baeyens-Volant, D., & Garcia-Pino, A. (2011). The plasticity of the β -trefoil fold constitutes an evolutionary platform for protease inhibition. *Journal of Biological Chemistry*, 286(51), 43726–43734. <http://doi.org/10.1074/jbc.M111.291310>
- Bao, R., Zhou, C. Z., Jiang, C., Lin, S. X., Chi, C. W., & Chen, Y. (2009). The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsins reveals a novel reactive site conformation. *Journal of Biological Chemistry*, 284(39), 26676–26684. <http://doi.org/10.1074/jbc.M109.022095>
- Barrett, A. J., Rawlings, N. D., Salvesen, G., & Fred Woessner, J. (2013). Introduction BT - Handbook of Proteolytic Enzymes (pp. li–liv). Academic Press. <http://doi.org/http://dx.doi.org/10.1016/B978-0-12-382219-2.00838-3>
- Bauw, G., Nielsen, H. V., Emmersen, J., Nielsen, K. L., Jørgensen, M., & Welinder, K. G. (2006). Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. *FEBS Journal*, 273(15), 3569–3584. <http://doi.org/10.1111/j.1742-4658.2006.05364.x>
- Bayés, A., Comellas-Bigler, M., Rodríguez de la Vega, M., Maškos, K., Bode, W., Aviles, F. X., ... Vendrell, J. (2005). Structural basis of the resistance of an insect carboxypeptidase to plant protease inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16602–16607. <http://doi.org/10.1073/pnas.0505489102>
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8), 3684–3690. <http://doi.org/10.1063/1.448118>
- Bieth, J. G. (1995). Theoretical and practical aspects of proteinase inhibition kinetics. *Methods in Enzymology*, 248, 59–84. [http://doi.org/10.1016/0076-6879\(95\)48007-2](http://doi.org/10.1016/0076-6879(95)48007-2)
- Bjelic, S., & Åqvist, J. (2004). Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site. *Biochemistry*, 43(46), 14521–14528. <http://doi.org/10.1021/bi048252q>
- Boratyn, G. M., Schäffer, A. A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology Direct*, 7, 12. <http://doi.org/10.1186/1745-6150-7-12>
- Broom, A., Doxey, A. C., Lobsanov, Y. D., Berthin, L. G., Rose, D. R., Howell, P. L., ... Meiering, E. M. (2012). Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. *Structure (London, England : 1993)*, 20(1), 161–71. <http://doi.org/10.1016/j.str.2011.10.021>

- Brunelle, F., Girard, C., Cloutier, C., & Michaud, D. (2005). A hybrid, broad-spectrum inhibitor of Colorado potato beetle aspartate and cysteine dipeptidases. *Archives of Insect Biochemistry and Physiology*, *60*(1), 20–31. <http://doi.org/10.1002/arch.20078>
- Castro, H. C., Abreu, P. a., Geraldo, R. B., Martins, R. C. a, Dos Santos, R., Loureiro, N. I. V, ... Rodrigues, C. R. (2011). Looking at the proteases from a simple perspective. *Journal of Molecular Recognition*, *24*(2), 165–181. <http://doi.org/10.1002/jmr.1091>
- Cater, S. a., Lees, W. E., Hill, J., Brzin, J., Kay, J., & Phylip, L. H. (2002). Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, *1596*(1), 76–82. [http://doi.org/10.1016/S0167-4838\(02\)00206-6](http://doi.org/10.1016/S0167-4838(02)00206-6)
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., ... Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D*, *66*(1), 12–21. <http://doi.org/10.1107/S0907444909042073>
- Christeller, J. T. (2005). Evolutionary mechanisms acting on proteinase inhibitor variability. *FEBS Journal*, *272*(22), 5710–5722. <http://doi.org/10.1111/j.1742-4658.2005.04975.x>
- Cohen, I., Kayode, O., Hockla, A., Sankaran, B., Radisky, D. C., Radisky, E. S., & Papo, N. (2016). Combinatorial protein engineering of proteolytically resistant mesotrypsin inhibitors as candidates for cancer therapy. *The Biochemical Journal*, *473*(10), 1329–41. <http://doi.org/10.1042/BJ20151410>
- Copeland, R. a. (2000). Tight binding inhibitors. In *Enzymes: A practical Introduction to Structure, Mechanism and Data Analysis* (Vol. 7, pp. 305–317). <http://doi.org/10.1002/0471220639.ch9>
- Crooks, G., Hon, G., Chandonia, J., & Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Res*, *14*, 1188–1190. <http://doi.org/10.1101/gr.849004.1>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, *98*(12).
- Dasgupta, J., Khamrui, S., Dattagupta, J. K., & Sen, U. (2006). Spacer Asn determines the fate of Kunitz (STI) inhibitors, as revealed by structural and biochemical studies on WCI mutants. *Biochemistry*, *45*(22), 6783–6792. <http://doi.org/10.1021/bi060374q>
- De Meester, P., Brick, P., Lloyd, L. F., Blow, D. M., & Onesti, S. (1998). Structure of the Kunitz-type soybean trypsin inhibitor (STI): Implication for the interactions between members of the STI family and tissue-plasminogen activator. *Acta Crystallographica Section D: Biological Crystallography*, *54*(4), 589–597. <http://doi.org/10.1107/S0907444997015849>
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J.-M., ... Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, *39*(suppl 2), W13–W17. <http://doi.org/10.1093/nar/gkr245>
- Drag, M., & Salvesen, G. S. (2010). Emerging principles in protease-based drug discovery. *Nature Reviews. Drug Discovery*, *9*(9), 690–701. <http://doi.org/10.1038/nrd3053>
- Dunn, B. M. (2002). Structure and mechanism of the Pepsin-like family of aspartic peptidases. *Chemical Reviews*, *102*(12), 4431–4458. <http://doi.org/10.1021/cr010167q>
- Eisenbeis, S., & Höcker, B. (2010). Evolutionary mechanism as a template for protein engineering. *Journal of Peptide Science*, *16*(10), 538–544. <http://doi.org/10.1002/psc.1233>
- Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*(4), 486–501. <http://doi.org/10.1107/S0907444910007493>
- Erlanger, B. F., Kokowsky, N., & Cohen, W. (1961). The preparation and properties of two new chromogenic substrates of trypsin. *Archives of Biochemistry and Biophysics*, *95*, 271–278.
- Fear, G., Komarnytsky, S., & Raskin, I. (2007). Protease inhibitors and their peptidomimetic derivatives as potential drugs. *Pharmacology and Therapeutics*, *113*(2), 354–368. <http://doi.org/10.1016/j.pharmthera.2006.09.001>

- Fischer, M., Kuckenberger, M., Kastilan, R., Muth, J., & Gebhardt, C. (2015). Novel in vitro inhibitory functions of potato tuber proteinaceous inhibitors. *Molecular Genetics and Genomics*, 290(1), 387–398. <http://doi.org/10.1007/s00438-014-0906-5>
- Franco, O. L., De Grossi Sá, M. F., Sales, M. P., Mello, L. V., Oliveira, A. S., & Rigden, D. J. (2002). Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from *Prosopis juliflora*. *Proteins: Structure, Function and Genetics*, 49(3), 335–341. <http://doi.org/10.1002/prot.10228>
- Gatehouse, J. a. (2011). Prospects for using proteinase inhibitors to protect transgenic plants against attack by herbivorous insects. *Current Protein & Peptide Science*, 12(5), 409–416. <http://doi.org/10.2174/138920311796391142>
- Ghosh, A. K., Osswald, H. L., & Prato, G. (2016). Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of Medicinal Chemistry*, 59(11), 5172–5208. <http://doi.org/10.1021/acs.jmedchem.5b01697>
- Gil, D. F., García-Fernández, R., Alonso-del-Rivero, M., Lamazares, E., Pérez, M., Varas, L., ... Mansur, M. (2011). Recombinant expression of ShPI-1A, a non-specific BPTI-Kunitz-type inhibitor, and its protection effect on proteolytic degradation of recombinant human miniproinsulin expressed in *Pichia pastoris*. *FEMS Yeast Research*, 11(7), 575–586. <http://doi.org/10.1111/j.1567-1364.2011.00749.x>
- Glaczinski, H., Heibges, A., Salamini, F., & Gebhardt, C. (2002). Members of the Kunitz-type protease inhibitor gene family of potato inhibit soluble tuber invertase in vitro. *Potato Research*, 45(2–4), 163–176. <http://doi.org/10.1007/BF02736112>
- Grütter, M. G., Priestle, J. P., Rahuel, J., Grossenbacher, H., Bode, W., Hofsteenge, J., & Stone, S. R. (1990). Crystal structure of the thrombin-hirudin complex: a novel mode of serine protease inhibition. *The EMBO Journal*, 9(8), 2361–2365. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC552259/>
- Guerra, Y., Valiente, P. a., Berry, C., & Pons, T. (2012). Predicting functional residues of the *Solanum lycopersicum* aspartic protease inhibitor (SLAPI) by combining sequence and structural analysis with molecular docking. *Journal of Molecular Modeling*, 18(6), 2673–2687. <http://doi.org/10.1007/s00894-011-1290-2>
- Guerra, Y., Valiente, P. A., Pons, T., Berry, C., & Rudiño-Piñera, E. (2016). Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: Could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop? *Journal of Structural Biology*, 195(2), 259–271. <http://doi.org/10.1016/j.jsb.2016.06.014>
- Guo, J., Erskine, P. T., Coker, A. R., Wood, S. P., & Cooper, J. B. (2015). Structure of a Kunitz-type potato cathepsin D inhibitor. *Journal of Structural Biology*, 192(3), 554–560. <http://doi.org/http://dx.doi.org/10.1016/j.jsb.2015.10.020>
- Hansen, D., Macedo-Ribeiro, S., Verissimo, P., Yoo Im, S., Sampaio, M. U., & Oliva, M. L. V. (2007). Crystal structure of a novel cysteineless plant Kunitz-type protease inhibitor. *Biochemical and Biophysical Research Communications*, 360(4), 735–740. <http://doi.org/10.1016/j.bbrc.2007.06.144>
- Headey, S. J., MacAskill, U. K., Wright, M. a., Claridge, J. K., Edwards, P. J. B., Farley, P. C., ... Pascal, S. M. (2010). Solution structure of the squash aspartic acid proteinase inhibitor (SQAPI) and mutational analysis of pepsin inhibition. *Journal of Biological Chemistry*, 285(35), 27019–27025. <http://doi.org/10.1074/jbc.M110.137018>
- Heibges, a, Glaczinski, H., Ballvora, a, Salamini, F., & Gebhardt, C. (2003). Structural diversity and organization of three gene families for Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Molecular Genetics and Genomics: MGG*, 269(4), 526–34. <http://doi.org/10.1007/s00438-003-0860-0>
- Heibges, a, Salamini, F., & Gebhardt, C. (2003). Functional comparison of homologous members of three groups of Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Molecular Genetics and Genomics: MGG*, 269(4), 535–41. <http://doi.org/10.1007/s00438-003-0861-z>
- Hess, B., Bekker, H., Berendsen, H. J. C., & Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12), 1463–1472. [http://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](http://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)

- Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447. <http://doi.org/10.1021/ct700301q>
- Hill, J., Tyas, L., Phylip, L. H., Kay, J., Dunn, B. M., & Berry, C. (1994). High level expression and characterisation of Plasmepsin I, a naspatic protease from *Plasmodium falciparum*. *FEBS Letters*, *352*(2), 155–158. [http://doi.org/10.1016/0014-5793\(94\)00940-6](http://doi.org/10.1016/0014-5793(94)00940-6)
- Hockney, R. W. (1970). The potential calculation and some applications. In *Methods in Computational Physics* (pp. 136–211).
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, *65*(3), 712–725. <http://doi.org/10.1002/prot.21123>
- Huang, H., Qi, S. D., Qi, F., Wu, C. A., Yang, G. D., & Zheng, C. C. (2010). NtKT11, a Kunitz trypsin inhibitor with antifungal activity from *Nicotiana tabacum*, plays an important role in tobacco's defense response. *FEBS Journal*, *277*(19), 4076–4088. <http://doi.org/10.1111/j.1742-4658.2010.07803.x>
- Huntington, J. A., Read, R. J., & Carrell, R. W. (2000). Structure of a serpin-protease complex shows inhibition by deformation. *Nature*, *407*(6806), 923–926. Retrieved from <http://dx.doi.org/10.1038/35038119>
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, *79*(2), 926–935. <http://doi.org/10.1063/1.445869>
- Kabsch, W. (2010). Xds. *Acta Crystallographica Section D: Biological Crystallography*, *66*(2), 125–132. <http://doi.org/10.1107/S0907444909047337>
- Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. <http://doi.org/10.1002/bip.360221211>
- Karplus, P. a., & Diederichs, K. (2012). Linking Crystallographic Model and Data Quality. *Science*, *336*(6084), 1030–1033. <http://doi.org/10.1126/science.1218231>
- Keilová H. and Tomášek, V. (1976a). Further characteristics of cathepsin D inhibitor from potatoes. *Collect. Czech. Chem. Commun.*, *41*, 2440–2447.
- Keilová H. and Tomášek, V. (1976b). Isolation and properties of Cathepsin D inhibitor from potatoes. *Collect Czech Chem Commun*, *41*, 489–497.
- Khamrui, S., Majumder, S., Dasgupta, J., Dattagupta, J. K., & Sen, U. (2010). Identification of a novel set of scaffolding residues that are instrumental for the inhibitory property of Kunitz (STI) inhibitors. *Protein Science*, *19*(3), 593–602. <http://doi.org/10.1002/pro.338>
- Kiggundu, A., Muchwezi, J., van der Vyver, C., Viljoen, A., Vorster, J., Schlüter, U., ... Michaud, D. (2010). Deleterious effects of plant cystatins against the banana weevil *Cosmopolites sordidus*. *Archives of Insect Biochemistry and Physiology*, *73*(2), 87–105. <http://doi.org/10.1002/arch.20342>
- Kim, J. Y., Park, S. C., Kim, M. H., Lim, H. T., Park, Y., & Hahn, K. S. (2005). Antimicrobial activity studies on a trypsin-chymotrypsin protease inhibitor obtained from potato. *Biochemical and Biophysical Research Communications*, *330*(3), 921–927. <http://doi.org/10.1016/j.bbrc.2005.03.057>
- Klemba, M., & Goldberg, D. E. (2002). Biological roles of proteases in parasitic protozoa. *Annual Review of Biochemistry*, *71*, 275–305. <http://doi.org/10.1146/annurev.biochem.71.090501.145453>
- Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., Hall, D. R., & Vajda, S. (2013). How good is automated protein docking? *Proteins: Structure, Function and Bioinformatics*, *81*(12), 2159–2166. <http://doi.org/10.1002/prot.24403>
- Krauchenco, S., Pando, S. C., Marangoni, S., & Polikarpov, I. (2003). Crystal structure of the Kunitz (STI)-type inhibitor from

- Delonix regia seeds. *Biochemical and Biophysical Research Communications*, 312(4), 1303–1308. <http://doi.org/10.1016/j.bbrc.2003.11.062>
- Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica Section D*, 60(12), 2256–2268. Retrieved from <http://dx.doi.org/10.1107/S0907444904026460>
- Krowarsch, D., Cierpicki, T., Jelen, F., & Otlewski, J. (2003). Canonical protein inhibitors of serine proteases. *Cellular and Molecular Life Sciences*, 60(11), 2427–2444. <http://doi.org/10.1007/s00018-003-3120-x>
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), msw054. <http://doi.org/10.1093/molbev/msw054>
- Laemmli, U. K. (1970). Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature*, 227(5259), 680–685. Retrieved from <http://dx.doi.org/10.1038/227680a0>
- Laskowski, M., & Kato, I. (1980). Protein Inhibitors of Proteinases. *Annual Review of Biochemistry*, 49(1), 593–626. <http://doi.org/10.1146/annurev.bi.49.070180.003113>
- Laskowski, M., & Qasim, M. a. (2000). What can the structures of enzyme-inhibitor complexes tell us about the structures of enzyme-substrate complexes? *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, 1477(1–2), 324–337. [http://doi.org/10.1016/S0167-4838\(99\)00284-8](http://doi.org/10.1016/S0167-4838(99)00284-8)
- Lee, J., Blaber, S. I., Dubey, V. K., & Blaber, M. (2011). A polypeptide “building block” for the β -trefoil fold identified by “top-down symmetric deconstruction”. *Journal of Molecular Biology*, 407(5), 744–63. <http://doi.org/10.1016/j.jmb.2011.02.002>
- Li, H., Robertson, A. D., & Jensen, J. H. (2005). Very fast empirical prediction and rationalization of protein pKa values. *Proteins*, 61(4), 704–721. <http://doi.org/10.1002/prot.20660>
- Li, M., Philip, L. H., Lees, W. E., Winther, J. R., Dunn, B. M., Wlodawer, P., ... Gustchina, A. (2000). The aspartic proteinase from *Saccharomyces cerevisiae* folds into its own inhibitor into a helix. *Nature Structural Biology*, 7(2), 113–117. <http://doi.org/10.1038/72378>
- Lima, T. B., Silva, O. N., Migliolo, L., Souza-Filho, C. R., Gonçalves, E. G., Vasconcelos, I. M., ... Franco, O. L. (2011). A Kunitz proteinase inhibitor from corms of *Xanthosoma blandum* with bactericidal activity. *Journal of Natural Products*, 74(5), 969–975. <http://doi.org/10.1021/np200312r>
- López-Otín, C., & Bond, J. S. (2008). Proteases: Multifunctional enzymes in life and disease. *Journal of Biological Chemistry*, 283(45), 30433–30437. <http://doi.org/10.1074/jbc.R800035200>
- Lu, W., Apostol, I., Qasim, M. A., Warne, N., Wynn, R., Zhang, W. L., ... Laskowski Jr, M. (1997). Binding of amino acid side-chains to S1 cavities of serine proteinases. *Journal of Molecular Biology*, 266(2), 441–461. <http://doi.org/http://dx.doi.org/10.1006/jmbi.1996.0781>
- Mares, M., Meloun, B., Pavlik, M., Kostka, V., & Baudys, M. (1989). Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Letters*, 251(1–2), 94–98. [http://doi.org/10.1016/0014-5793\(89\)81435-8](http://doi.org/10.1016/0014-5793(89)81435-8)
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *Journal of Applied Crystallography*, 40(4), 658–674. <http://doi.org/10.1107/S0021889807021206>
- McKerrow, J. H., Caffrey, C., Kelly, B., Loke, P., & Sajid, M. (2006). Proteases in parasitic diseases. *Annual Review of Pathology*, 1, 497–536. <http://doi.org/10.1146/annurev.pathol.1.110304.100151>
- Mehrabadi, M., Franco, O. L., & Bandani, A. R. (2012). Plant Proteinaceous α -Amylase and Proteinase Inhibitors and Their Use in Insect Pest Control.
- Meulenbroek, E. M., Thomassen, E. a. J., Pouvreau, L., Abrahams, J. P., Gruppen, H., & Pannu, N. S. (2012). Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato. *Acta*

- Crystallographica Section D: Biological Crystallography*, 68(7), 794–799. <http://doi.org/10.1107/S090744491201222X>
- Micheelsen, P. O., Vévodová, J., De Maria, L., Østergaard, P. R., Friis, E. P., Wilson, K., & Skjøt, M. (2008). Structural and Mutational Analyses of the Interaction between the Barley α -Amylase/Subtilisin Inhibitor and the Subtilisin Savinase Reveal a Novel Mode of Inhibition. *Journal of Molecular Biology*, 380(4), 681–690. <http://doi.org/10.1016/j.jmb.2008.05.034>
- Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8), 952–962. <http://doi.org/10.1002/jcc.540130805>
- Moore, S. A., Sielecki, A. R., Chernai, M. M., Tarasova, N. I., & James, M. N. G. (1995). Crystal and Molecular Structures of Human Progastricsin at 1.62 Å Resolution. *Journal of Molecular Biology*, 247(3), 466–485. <http://doi.org/http://dx.doi.org/10.1006/jmbi.1994.0154>
- Morrison, J. F. (1969). Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochimica et Biophysica Acta (BBA) - Enzymology*, 185(2), 269–286. [http://doi.org/http://dx.doi.org/10.1016/0005-2744\(69\)90420-3](http://doi.org/http://dx.doi.org/10.1016/0005-2744(69)90420-3)
- Murphy, D. J. (2004). Determination of accurate K_i values for tight-binding enzyme inhibitors: An in silico study of experimental error and assay design. *Analytical Biochemistry*, 327(1), 61–67. <http://doi.org/10.1016/j.ab.2003.12.018>
- Murshudov, G. N., Skubák, P., Lebedev, A. a., Pannu, N. S., Steiner, R. a., Nicholls, R. a., ... Vagin, A. a. (2011). REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallographica Section D: Biological Crystallography*, 67(4), 355–367. <http://doi.org/10.1107/S0907444911001314>
- Murzin, A. G., Lesk, A. M., & Chothia, C. (1992). β -Trefold fold. *Journal of Molecular Biology*, 223(2), 531–543. [http://doi.org/http://dx.doi.org/10.1016/0022-2836\(92\)90668-A](http://doi.org/http://dx.doi.org/10.1016/0022-2836(92)90668-A)
- Ng, K. K., Petersen, J. F., Cherney, M. M., Garen, C., Zalatoris, J. J., Rao-Naik, C., ... James, M. N. (2000). Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3. *Nature Structural Biology*, 7(8), 653–657. <http://doi.org/10.1038/77950>
- Ohtsubo, K., & Richardson, M. (1992). The amino acid sequence of a 20 kDa bifunctional subtilisin/ α -amylase inhibitor from bran [correction of brain] of rice (*Oryza sativa* L.) seeds. *FEBS Letters*, 309(1), 68–72.
- Oliva, M. L. V., Silva, M. C. C., Sallai, R. C., Brito, M. V., & Sampaio, M. U. (2010). A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie*, 92(11), 1667–1673. <http://doi.org/10.1016/j.biochi.2010.03.021>
- Otlewski, J., Jelen, F., Zakrzewska, M., & Oleksy, A. (2005). The many faces of protease-protein inhibitor interaction. *The EMBO Journal*, 24(7), 1303–1310. <http://doi.org/10.1038/sj.emboj.7600611>
- Patil, D. N., Chaudhary, A., Sharma, A. K., Tomar, S., & Kumar, P. (2012). Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. *FEBS Journal*, 279(24), 4547–4564. <http://doi.org/10.1111/febs.12042>
- Pillay, P., Schlüter, U., Van Wyk, S., Kurnert, K. J., & Vorster, B. J. (2013). Proteolysis of recombinant proteins in bioengineered plant cells. *Bioengineered*, 5(1), 1–6. <http://doi.org/10.4161/bioe.25158>
- Qin, S., & Zhou, H.-X. (2007). meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics*, 23(24), 3386–7. <http://doi.org/10.1093/bioinformatics/btm434>
- Radisky, E. S., & Koshland, D. E. (2002). A clogged gutter mechanism for protease inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16), 10316–10321. <http://doi.org/10.1073/pnas.112332899>
- Ramírez, A. R., Guerra, Y., Otero, A., García, B., Berry, C., Mendiola, J., ... Chávez, M. de los A. (2009). Generation of an affinity matrix useful in the purification of natural inhibitors of plasmepepsin II, an antimalarial-drug target. *Biotechnology and Applied Biochemistry*, 52(2), 149–157. <http://doi.org/10.1042/BA20080015>
- Ravindranathan, K. P., Gallicchio, E., & Levy, R. M. (2005). Conformational equilibria and free energy profiles for the allosteric transition of the ribose-binding protein. *Journal of Molecular Biology*, 353(1), 196–210.

<http://doi.org/10.1016/j.jmb.2005.08.009>

- Rawlings, N. D., Waller, M., Barrett, A. J., & Bateman, A. (2014). MEROPS: The database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, *42*(D1), 503–509. <http://doi.org/10.1093/nar/gkt953>
- Rees, D. C., & Lipscomb, W. N. (1982). Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *Journal of Molecular Biology*, *160*(3), 475–498.
- Reeves, J. H. (1992). Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. *Journal of Molecular Evolution*, *35*(1), 17–31. <http://doi.org/10.1007/BF00160257>
- Renko, M., Sabotič, J., Mihelič, M., Brzin, J., Kos, J., & Turk, D. (2010). Versatile loops in mycocypins inhibit three protease families. *Journal of Biological Chemistry*, *285*(1), 308–316. <http://doi.org/10.1074/jbc.M109.043331>
- Renko, M., Sabotič, J., & Turk, D. (2012). β -Trefol inhibitors - From the work of Kunitz onward. *Biological Chemistry*, *393*(10), 1043–1054. <http://doi.org/10.1515/hsz-2012-0159>
- Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., ... Aviles, F. X. (2000). Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nature Structural Biology*, *7*(4), 322–328. <http://doi.org/10.1038/74092>
- Ritonja, a, Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., ... Turk, V. (1990). The amino acid sequence of a novel inhibitor of cathepsin D from potato. *FEBS Letters*, *267*(1), 13–15.
- Rivard, D., Anguenot, R., Brunelle, F., Le, V. Q., Vézina, L. P., Trépanier, S., & Michaud, D. (2006). An in-built proteinase inhibitor system for the protection of recombinant proteins recovered from transgenic plants. *Plant Biotechnology Journal*, *4*(3), 359–368. <http://doi.org/10.1111/j.1467-7652.2006.00187.x>
- Rodrigues Macedo, M. L., Durigan, R. A., da Silva, D. S., Marangoni, S., Freire, M. G. M., & Parra, J. R. P. (2010). Adenanthera pavonina trypsin inhibitor retard growth of Anagasta kuehniella (Lepidoptera: Pyralidae). *Archives of Insect Biochemistry and Physiology*, *73*(4), 213–231. <http://doi.org/10.1002/arch.20352>
- Saadati, F., & Bandani, A. R. (2011). Effects of Serine Protease Inhibitors on Growth and Development and Digestive Serine Proteinases of the Sunn Pest, Eurygaster integriceps. *Journal of Insect Science (Online)*, *11*(72), 72. <http://doi.org/10.1673/031.011.7201>
- Saha, R. P., Bahadur, R. P., Pal, A., Mandal, S., & Chakrabarti, P. (2006). ProFace: a server for the analysis of the physicochemical features of protein-protein interfaces. *BMC Structural Biology*, *6*, 11. <http://doi.org/10.1186/1472-6807-6-11>
- Santamaria, M. E., Cambra, I., Martinez, M., Pozancos, C., González-Melendi, P., Grbic, V., ... Diaz, I. (2012). Gene pyramiding of peptidase inhibitors enhances plant resistance to the spider mite tetranychus urticae. *PLoS ONE*, *7*(8), 2–11. <http://doi.org/10.1371/journal.pone.0043011>
- Schechter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, *27*(2), 157–162. [http://doi.org/http://dx.doi.org/10.1016/S0006-291X\(67\)80055-X](http://doi.org/http://dx.doi.org/10.1016/S0006-291X(67)80055-X)
- Schlüter, U., Benchabane, M., Munger, A., Kiggundu, A., Vorster, J., Goulet, M. C., ... Michaud, D. (2010). Recombinant protease inhibitors for herbivore pest control: A multitrophic perspective. *Journal of Experimental Botany*, *61*(15), 4169–4183. <http://doi.org/10.1093/jxb/erq166>
- Senthilkumar, R., Cheng, C. P., & Yeh, K. W. (2010). Genetically pyramiding protease-inhibitor genes for dual broad-spectrum resistance against insect and phytopathogens in transgenic tobacco. *Plant Biotechnology Journal*, *8*(1), 65–75. <http://doi.org/10.1111/j.1467-7652.2009.00466.x>
- Song, H. K., Kim, Y. S., Yang, J. K., Moon, J., Lee, J. Y., & Suh, S. W. (1999). Crystal structure of a 16 kDa double-headed Bowman-Birk trypsin inhibitor from barley seeds at 1.9 Å resolution. *Journal of Molecular Biology*, *293*(5), 1133–1144. <http://doi.org/10.1006/jmbi.1999.3239>
- Song, H. K., & 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. *Journal of Molecular Biology*, 275(2), 347–363. <http://doi.org/10.1006/jmbi.1997.1469>
- Speranskaya, a S., Krinitsina, a a, Revina, T. a, Gerasimova, N. G., Keruchen'ko, Y. S., Shevelev, a B., & Valueva, T. a. (2006). Heterologous expression, purification, and properties of a potato protein inhibitor of serine proteinases. *Biochemistry. Biokhimiia*, 71(11), 1176–1182. <http://doi.org/10.1134/S0006297906110022>
- Srikanth, S., & Chen, Z. (2016). Plant Protease Inhibitors in Therapeutics-Focus on Cancer Therapy. *Frontiers in Pharmacology*, 7(December), 470. <http://doi.org/10.3389/fphar.2016.00470>
- Strong, M., Sawaya, M. R., Wang, S., Phillips, M., Cascio, D., & Eisenberg, D. (2006). Toward the structural genomics of complexes: Crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, 103(21), 8060–8065. <http://doi.org/10.1073/pnas.0602606103>
- Terada, D., Voet, A. R. D., Noguchi, H., Kamata, K., Ohki, M., Addy, C., ... Zhang, K. Y. J. (2017). Computational design of a symmetrical β -trefoil lectin with cancer cell binding activity. *Scientific Reports*, 7(1), 1–13. <http://doi.org/10.1038/s41598-017-06332-7>
- The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC. (n.d.).
- Turk, B. (2006). Targeting proteases: successes, failures and future prospects. *Nature Reviews. Drug Discovery*, 5(9), 785–799. <http://doi.org/10.1038/nrd2092>
- Turk, B., Turk, D., & Turk, V. (2012). Protease signalling: the cutting edge. *The EMBO Journal*, 31(7), 1630–1643. <http://doi.org/10.1038/emboj.2012.42>
- Vajda, S., Yueh, C., Beglov, D., Bohnuud, T., Mottarella, S. E., Xia, B., ... Kozakov, D. (2017). New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics*, 85(3), 435–444. <http://doi.org/10.1002/prot.25219>
- Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K. W., Svensson, B., & Haser, R. (1998). Barley α -amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 Å resolution. *Structure*, 6(5), 649–659. [http://doi.org/http://dx.doi.org/10.1016/S0969-2126\(98\)00066-5](http://doi.org/http://dx.doi.org/10.1016/S0969-2126(98)00066-5)
- van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., ... Berendsen, H. J. C. (2010). Gromacs User Manual version 4.5.
- Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B., & Dunn, B. M. (1999). Active site specificity of plasmepsin II. *Protein Science: A Publication of the Protein Society*, 8(10), 2001–2009. <http://doi.org/10.1110/ps.8.10.2001>
- Whelan, S., & Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution*, 18(5), 691–699. Retrieved from <http://dx.doi.org/10.1093/oxfordjournals.molbev.a003851>
- Xie, Z.-W., Luo, M.-J., Xu, W.-F., & Chi, C.-W. (1997). Two Reactive Site Locations and Structure–Function Study of the Arrowhead Proteinase Inhibitors, A and B, Using Mutagenesis. *Biochemistry*, 36(19), 5846–5852. <http://doi.org/10.1021/bi962993c>
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *Journal of Molecular Evolution*, 39(3), 306–314. <http://doi.org/10.1007/BF00160154>
- Yike, I. (2011). Fungal Proteases and Their Pathophysiological Effects. *Mycopathologia*, 171(5), 299–323. <http://doi.org/10.1007/s11046-010-9386-2>
- Zakharova, E., Horvath, M. P., & Goldenberg, D. P. (2009). Structure of a serine protease poised to resynthesize a peptide bond. *Proceedings of the National Academy of Sciences of the United States of America*, 106(27), 11034–11039. <http://doi.org/10.1073/pnas.0902463106>
- Zhou, D., Hansen, D., Shabalin, I. G., Gustchina, A., Vieira, D. F., de Brito, M. V., ... Wlodawer, A. (2015). Structure of BbKI, a disulfide-free plasma kallikrein inhibitor. *Acta Crystallographica Section F*, 71(8), 1055–1062. <http://doi.org/10.1107/S2053230X15011127>
- Zhou, D., Lobo, Y. a., Batista, I. F. C., Marques-Porto, R., Gustchina, A., Oliva, M. L. V., & Wlodawer, A. (2013). Crystal

Structures of a Plant Trypsin Inhibitor from *Enterolobium contortisiliquum* (EcTI) and of Its Complex with Bovine Trypsin. *PLoS ONE*, 8(4). <http://doi.org/10.1371/journal.pone.0062252>

Zimmermann, K. (1991). ORAL: All purpose molecular mechanics simulator and energy minimizer. *Journal of Computational Chemistry*, 12(3), 310–319. <http://doi.org/10.1002/jcc.540120305>

Anexo 1

Informe del análisis del inhibidor E3Ad mediante espectrometría de masas ESI-MS-MS.

No. de folio: 20-2014

Laboratorio Universitario de Proteómica

IBT/UNAM

Servicios de identificación, caracterización estructural y cuantificación de proteínas a través de la Espectrometría de Masas de alta resolución.

Usuario: Yasel Guerra Borrego

Teléfono: 27610

E. mail: yaselg@ibt.unam.mx

Jefe de Grupo: Dr. Enrique Rudiño P.

Institución: IBT

Número de Muestras: 02

Organismo: *Pichia pastoris*

Tipo de servicio: Identificación de proteína + ID-Gly + ID-Nterminal

Fecha de entrega de la muestra al LUP: 03/04/2014

RESULTADOS

La secuenciación “de novo” de los espectros de MS/MS de las dos muestras (M1 y M2) demostró que las dos bandas analizadas contienen la misma proteína. En los dos análisis, M1 y M2, el porcentaje de cobertura de ambas secuencias fue de 95.7%. Observe las dos secuencias, los amino ácidos en verde fueron todos secuenciados “de novo”

Para determinación del N-terminal de la proteína fueron analizados los espectros de masas de los péptidos precursores que correspondían a los valores teóricos de los péptidos trípticos obtenidos teóricamente. El último péptido tríptico en contrario fue el ESPLPK. Ningún fragmento correspondiente al péptido EEGVSLEKR fue encontrado. Por lo tanto, el péptido ESPLPK es el N-terminal de la proteína analizada (Figura A1).

Para la determinación de las posibles glicosilaciones, los péptidos trópicos que no dieron match con los teóricos esperados fueron evaluados de forma individual. De esta forma determinamos la presencia de por lo menos 6 isoformas glicosiladas presentes en ambas muestras como se puede observar en la Figura A2. A través de la fragmentación del precursor de carga 2⁺ a *m/z* 1554.11 (Masa molecular monoisotópica: 3106.22) se pudo identificar el péptido que contiene la glicosilación y secuenciarlo totalmente (Figura A3).

El análisis del mismo espectro de fragmentación con *m/z* 2⁺ a 1554.1 se pudo determinar que el mismo sitio del péptido está glicosilado (en esta isoforma, las otras isoformas poseen una glucosa de diferencia ellas) por 10 glucosas y dos acetilhexoaminas. Sin embargo, no es posible determinar la disposición estructural de los doce monosacáridos. La suma de la masa molecular del péptido 1079.51 + 10(162.052) hexosas + 2(203.079) N-acetilhexosaminas es igual a 3106.188. Una diferencia (0.032 Da) bastante aceptable de la masa experimental determinada para el péptido glicosilado. En la Figura A4 están marcadas las diferencias de masa con carga 2⁺ de 81.03 (162.052=Hexose) y 101.54 (203.079=N-acetilhexosamina) respectivamente.

Found Modifications:

- C Carbamidomethyl (C)
- O Oxidation (M)

M1

C

EEGVSLEKRE SPLPKPVLDT NGKELNPNS YRIISIGRGA LGGDVYLGKS PNSDAPCPDG VFRYNSDVGP

C O

SGTPVRFIPL STNIFEDQLL NIQFNIPTVK LCVSYTIWKV GNLNAYFRM LLETGGTIGQ ADNSYFKIVK

C C C C

SSKIGYNLLS CPFTSIICLR CPEDQFCAKV GVVIQNGKRR LALVNEPLD VLFQEV

M2

C

EEGVSLEKRE **SPLPKPVLDT** NGKELNPNS YRIISIGRGA LGGDVYLGKS PNSDAPCPDG VFRYNSDVGP

C O

SGTPVRFIPL STNIFEDQLL NIQFNIP**TVK** **LCVSYTIWKV** GNLNAYFR**TM** LLETGGTIGQ ADNSYFK**IVK**

C C C C

SSK **IGYNLLS** CPFTSI**ICLR** CPEDQF**CAKV** GVVIQNGK**RR** **LALVNENPLD** VLFQ**EV**

C:\Xcalibur\data\20-2014Yaselcalibrado

4/30/2014 1:42:35 PM

20-2014Yaselcalibrado #3360-14455 RT: 41.11-145.29 AV: 124 NL: 3.31E5
T: Average spectrum MS2 335.69 (3360-14455)

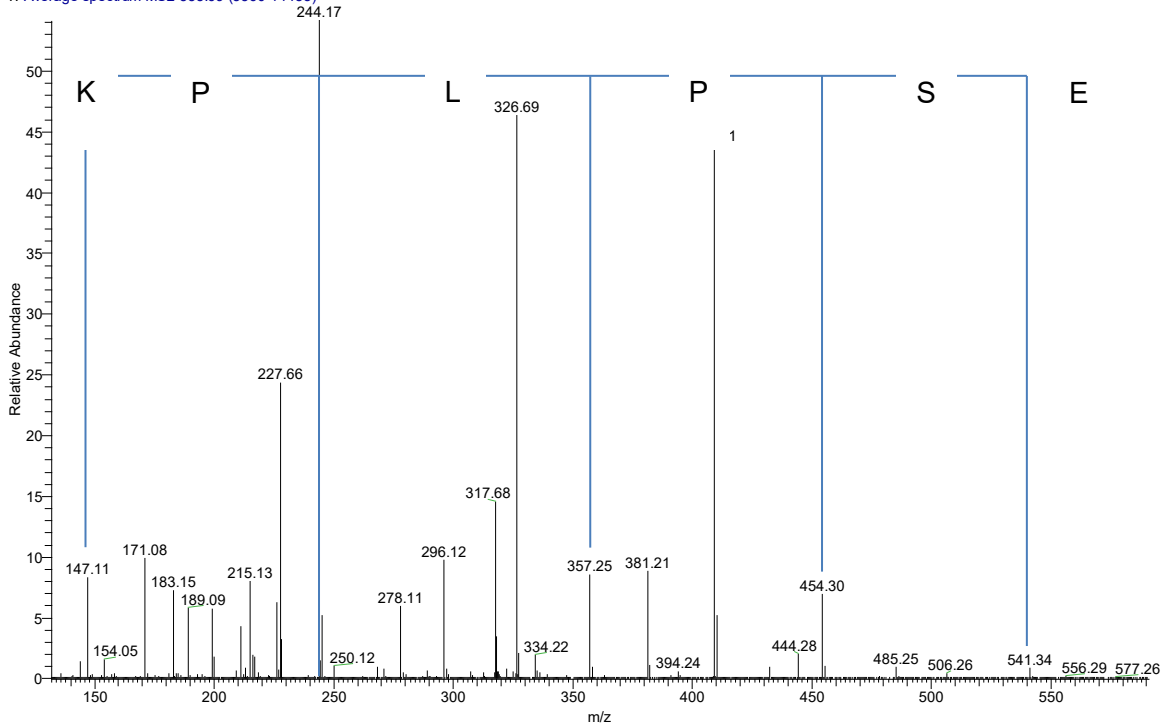


Figura A1. N-Terminal de la proteína. Espectro de MS/MS del precursor a m/z $^{2+}$ 335.69 con secuencia ESPLPK.

20-2014Yaselcalibrado #2826-3094 RT: 33.95-37.63 AV: 171 NL: 5.37E4
T: FTMS + c NSI Full ms [200.00-2000.00]

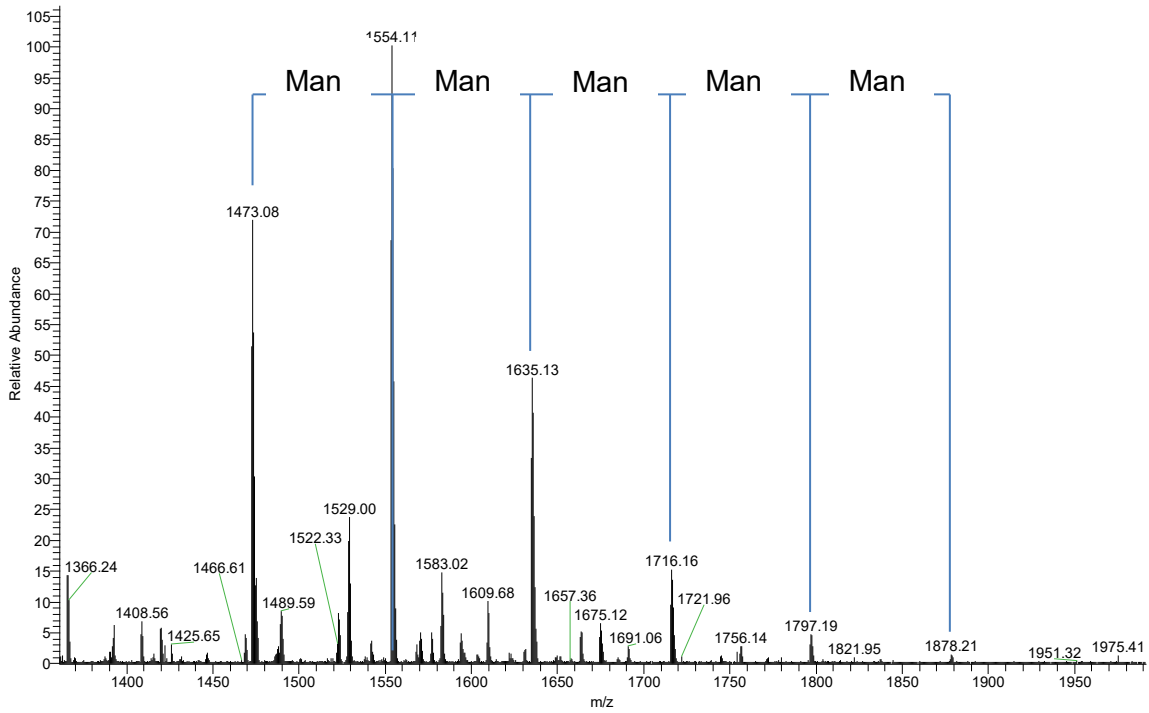


Figura A2. Análisis completo de iones precursores glicosilados. Se observa 6 precursores con diferencia de apenas una glucosa.

20-2014YaselM1solo1554 #1-812 RT: 0.01-8.14 AV: 812 NL: 8.18E4
 T: Average spectrum MS2 1554.11 (1-812)

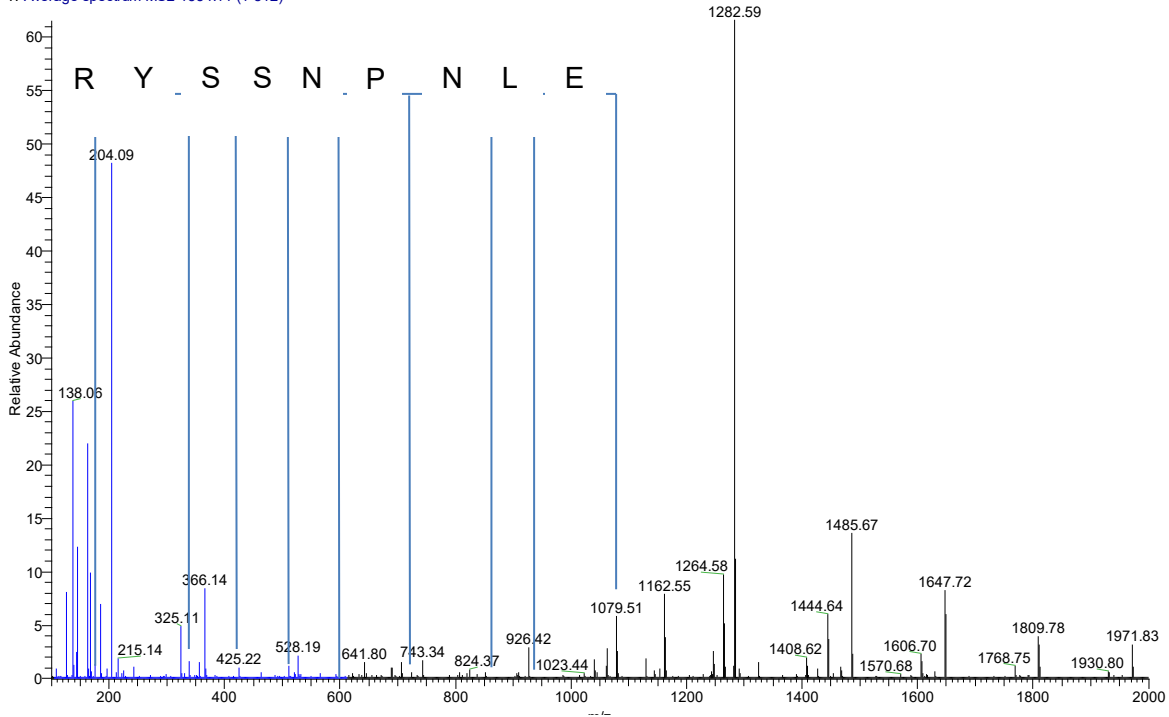


Figura A3. Espectro de fragmentación de uno de los péptidos glicosilados. Se observa la secuencia del péptido que contiene la glicosilación y fragmentos de azúcares con picos más intensos a la derecha del espectro.

E:\r
 ruc
 T:

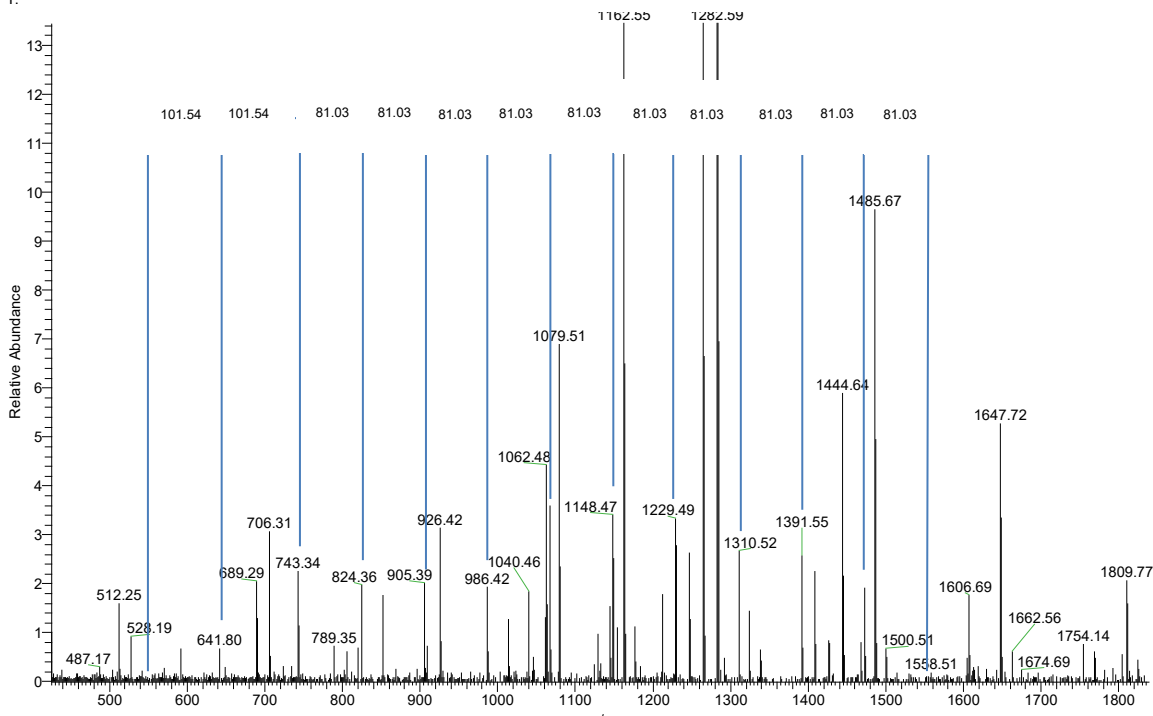


Figura A4. Fragmentación del péptido glicosilado. Se observa claramente las diferencias de masas con carga 2+ de hexosas y acetilhexosaminas.

METODOLOGÍA

Las muestras fueron previamente reducidas con ditioneitol (DTT), alquiladas con iodoacetamida (obtenida con Sigma-Aldrich) digerida "in gel" con tripsina (Promega Sequencing Grade Modified Trypsin). Los péptidos resultantes fueron aplicados en un sistema LC-MS bomba de nanoflujo EASY-nLC II (Thermo-Fisher Co. San Jose, CA) acoplado a un espectrómetro de masas LTQ-Orbitrap Velos (Thermo-Fisher Co., San Jose, CA) con sistema de ionización tipo nano-electrospray (ESI). La calibración del espectrómetro fue realizada con una solución Calmix (N-butilamina, cafeína, Met-Arg-Phe-Ala (MRFA) y Ultramark 1621 se utilizan para calibrar Thermo Scientific LTQ Velos de trampa de iones y Orbitrap espectrómetros de masas en el modo de ionización positiva ESI. La N-butilamina se incluye para extender a menor rango de calibración de masas) esto permite determinaciones con exactitudes menores a 5 ppm (partes/millón). En el sistema de cromatografía de líquidos se utilizó un sistema gradiente de 10-80 % de solvente B (agua/acetonitrilo con 0.1 % de ácido fórmico) en 120 minutos utilizando una columna capilar hecha en casa (ID 0.75 μm y 10cm largo RP-C18). El flujo del sistema LC fue de 300 nanolitros/minuto. Para la fragmentación de los péptidos se utilizó los métodos de CID (Collision-Induced Dissociation) y HCD (High energy Collision Dissociation) donde solamente los iones con carga 2⁺, 3⁺ y 4⁺ fueron seleccionados para los eventos de fragmentación. Fueron desconsiderados los iones con cargas 1⁺, superiores a 5⁺ y de cargas indefinidas. Todos los espectros fueron adquiridos en modo de detección positivo. La ejecución y captura de los datos de fragmentación fueron realizados de forma dependiente del escaneo total de iones según las cargas pre-determinadas con un ancho de aislamiento de 3.0 (m/z), energía de colisión normalizada de 35 unidades arbitrarias, activación Q de 0.250, tiempo de activación de 10 milisegundos y tiempo máximo de inyección de 10 milisegundos por micro-escaneo. Durante la captura automática de los datos fue utilizado la exclusión dinámica de iones: (i) lista de exclusión de 400 iones, (ii) tiempo de pre-exclusión de 30 segundos y (iii) tiempo de exclusión de 300 segundos. Los datos espectrométricos fueron sometidos a la búsqueda contra la secuencia proporcionada por el usuario a través del programa Proteome Discoverer 1.4 para confirmar la identidad de la proteína. Todos los espectros producidos por los dos modos de fragmentación utilizados (CID y HCD) fueron exhaustivamente analizados de forma manual.

CUENTA

Identificación de 02 proteínas por LC-MS/MS + determinación del N-terminal + caracterización de glicosilación de dos proteínas:

Costo total :6,500.00 M.N.

Iva N/A

Cuernavaca a 05 de Mayo del 2014.

Atentamente,

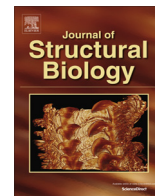
Dr. Cesar V. F. Batista

Jefe del L.U.P.

*****Nota: Conservar el número de folio para cualquier aclaración.*****

Anexo 2

Producción científica



Structures of a bi-functional Kunitz-type STI family inhibitor of serine and aspartic proteases: Could the aspartic protease inhibition have evolved from a canonical serine protease-binding loop?



Yasel Guerra^{a,*}, Pedro A. Valiente^b, Tirso Pons^c, Colin Berry^d, Enrique Rudiño-Piñera^{a,*}

^aDepartamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos CP 62210, Mexico

^bLaboratorio de Biología Computacional y Diseño de Proteínas, Centro de Estudios de Proteínas (CEP), Facultad de Biología, Universidad de La Habana, Cuba

^cStructural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), C/Melchor Fernández Almagro 3, Madrid 28029, Spain

^dCardiff School of Biosciences, Cardiff University, Cardiff CF10 3AT, Wales, UK

ARTICLE INFO

Article history:

Received 9 April 2016

Received in revised form 14 June 2016

Accepted 18 June 2016

Available online 18 June 2016

Keywords:

Aspartic protease inhibitors
Kunitz-type STI family inhibitors
Bi-functional inhibitors
Plant protease inhibitors
 β -Trefoil fold

ABSTRACT

Bi-functional inhibitors from the Kunitz-type soybean trypsin inhibitor (STI) family are glycosylated proteins able to inhibit serine and aspartic proteases. Here we report six crystal structures of the wild-type and a non-glycosylated mutant of the bifunctional inhibitor E3Ad obtained at different pH values and space groups. The crystal structures show that E3Ad adopts the typical β -trefoil fold of the STI family exhibiting some conformational changes due to pH variations and crystal packing. Despite the high sequence identity with a recently reported potato cathepsin D inhibitor (PDI), three-dimensional structures obtained in this work show a significant conformational change in the protease-binding loop proposed for aspartic protease inhibition. The E3Ad binding loop for serine protease inhibition is also proposed, based on structural similarity with a novel non-canonical conformation described for the double-headed inhibitor API-A from the Kunitz-type STI family. In addition, structural and sequence analyses suggest that bifunctional inhibitors of serine and aspartic proteases from the Kunitz-type STI family are more similar to double-headed inhibitor API-A than other inhibitors with a canonical protease-binding loop.

© 2016 Published by Elsevier Inc.

1. Introduction

Protease inhibitors are an effective way to regulate the activity of target proteases, with great interest in those involved in biomedical and biotechnological processes. Among the different protease inhibitor families, proteinaceous serine protease inhibitors are probably the most studied and widespread in nature (Otlewski et al., 2001; Tyndall et al., 2005).

Inhibitors of the Kunitz-type soybean trypsin inhibitor (STI) family are among the most versatile protease inhibitors reported, being able to interact with proteases belonging to different mech-

anistic classes like serine, aspartic and cysteine proteases, as well as other enzymes: α -amylase and sucrose invertase (Azarkan et al., 2011; Cater et al., 2002; Fischer et al., 2015; Franco et al., 2002; Glaczinski et al., 2002; Keilová and Tomášek, 1976a). These inhibitors have a β -trefoil fold, composed of 12 β -strands arranged in threefold pseudo-symmetry units or subdomains with the presence of helices in some cases (Broom et al., 2012; McLachlan, 1979; Murzin et al., 1992; Renko et al., 2012). The pseudo-symmetry units are composed of four β -strands, two of them forming a six-stranded β -barrel with the other subdomains, and the other two of each unit forming a β -hairpin acting like the cap for the β -barrel. The structural regions reported to be involved in protease-inhibitor interactions are loops connecting the β -strand elements (i.e. loop L4 between β -strands 4 and 5 in STI), which can be highly variable in sequence and length (Renko et al., 2012). The most accepted model for the β -trefoil fold evolution proposes that the fold arose from a duplication and fusion of the monomeric unit from an ancestral trimeric protein (Broom et al., 2012; Ponting and Russell, 2000).

Abbreviations: BASI, barley alpha-amylase/subtilisin inhibitor; EcTI, *Enterolobium contortisiliquum* trypsin inhibitor; PSPI, potato serine protease inhibitor; API-A, arrowhead proteinase inhibitor A; PDI, potato cathepsin D inhibitor; STI, soybean trypsin inhibitor; SLAPI, *Solanum lycopersicum* aspartic protease inhibitor; TKI, Tamarind Kunitz-type inhibitor.

* Corresponding authors.

E-mail addresses: yaselg@ibt.unam.mx (Y. Guerra), rudiño@ibt.unam.mx (E. Rudiño-Piñera).

There are several structural and functional studies defining the inhibition mechanism for serine proteases (Laskowski and Kato, 1980; Radisky and Koshland, 2002; Song and Suh, 1998; Zakharova et al., 2009). Among the proteinaceous protease inhibitors, almost all Kunitz-type STI inhibitors have a “canonical” conformation in the protease-binding loop that interacts with the active site of the serine proteases forming a substrate-like “Michaelis” complex and thereby implicated directly in their biological activity (Song and Suh, 1998; Zhou et al., 2013). In the inhibition mechanism, the enzyme is able to hydrolyze the peptidic bond between P1-P1' positions (Schechter & Berger Pn-Pn' nomenclature) (Schechter and Berger, 1967) but the product retains a similar conformation with respect to the intact inhibitor, trapping the enzyme in a hydrolysis/re-synthesis cycle (Zakharova et al., 2009). The amino acid sequence of these “canonical” protease-binding loops can be variable, although there are certain preferences for basic amino acids (Lys and Arg) in the P1 position for trypsin (E.C. 3.4.21.4) and the amino acids Tyr, Trp, Phe, Leu and Met in the case of chymotrypsin (E.C. 3.4.21.1) (Tyndall et al., 2005). Despite the variability among amino acid sequences, the conformation of the protease-binding loops are kept unchanged (Laskowski and Qasim, 2000). Generally, positions P3-P3' are key determinants of the reactive loop conformation, with the P1 position comprising up to 50% of the enzyme-inhibitor interface contact area (Krowarsch et al., 2003). However, there is at least one case reported where a double-headed inhibitor has a binding loop with a non-canonical conformation, forming an enzyme-inhibitor complex with trypsin (Bao et al., 2009).

In the case of Kunitz-type STI inhibitors from legume seeds, a sub-classification has been proposed based on the numbers of cysteines in the sequence (Oliva et al., 2010). Specifically in potato, a different classification has been proposed to divide the proteinaceous inhibitors into five groups with the best characterized proteins belonging to groups A, B and C (Bauw et al., 2006; Heibges et al., 2003a). It has been shown that there is a great sequence variability in the Kunitz-type STI inhibitors in potato, which has been linked to an adaptive mechanism in response to pathogens or even the acquisition of new functions to enhance their adaptability (Heibges et al., 2003b). Recently, a cDNA library was obtained from several potato cultivars and some proteins were heterologously expressed and tested against different classes of proteases. This work showed that inhibition of aspartic proteases is also possible in potato protease inhibitors belonging to group B (Fischer et al., 2015).

Despite the wealth of data regarding the inhibition process for serine proteases, very little is known concerning the structural determinants for the inhibition of aspartic proteases by inhibitors from the Kunitz-type STI family. Although first studies on bi-functional inhibitors of the Kunitz-type STI family against aspartic proteases were performed using human cathepsin D (EC 3.4.23.5) (Keilová and Tomášek, 1976b), to date there is evidence that Kunitz-type STI family members are also able to inhibit other aspartic proteases of the clan AA, family A1, such as saccharopepsin (EC 3.4.21.41) (Cater et al., 2002), memapsin 1 (EC 3.4.23.45) (Fischer et al., 2015), and Plasmepsin II (EC 3.4.23.39) (this work).

Until recently, structural studies of proteinaceous inhibitors of aspartic proteases were limited with only four three-dimensional coordinates available in the PDB from four different families in the MEROPS database (Rawlings et al., 2014): inhibitor SQAPI, family I25 (Headey et al., 2010, PDB code: 2KXG); inhibitor PI3, family I33 (Ng et al., 2000, PDB codes: 1F32 and 1F34), inhibitor IA3, family I34 (Li et al., 2000, PDB code 1DPJ) and potato cathepsin D inhibitor (PDI), family I3 (Guo et al., 2015, PDB code: 5DZU). Only for the inhibitors PI3 and IA3 are there three-dimensional coordinates of the inhibitors in complex with target aspartic proteases (Pepsin

and saccharopepsin; PDB codes: 1F34 and 1DPJ, respectively). However, there are no similarities in the amino acid sequences, three-dimensional structures or inhibition binding modes between these distinct families, making extrapolations to other proteinaceous protease inhibitors impossible.

Previous computational analyses based on comparative protein modeling and docking studies, predicted loop L9 in the *Solanum lycopersicum* (tomato) aspartic protease inhibitor (SLAPI), as a key region involved in the inhibitor interaction with saccharopepsin's active site (Guerra et al., 2012). A recent paper that describes the crystallographic structure of PDI also suggests the same structural region – the loop from Cys142 to Ala160 in PDI – as a binding loop for cathepsin D (Guo et al., 2015). Based on docking studies, Guo and colleagues, propose a second possible binding loop (loop L5) for cathepsin D running from Phe85 to Ile98.

In the present work we report six crystal structures, obtained at different pH values (3.0, 3.5, 7.0, 7.2, 7.4 and 9.0) and space groups (C222₁, P22₁2₁ and P4₃22), of a potato bi-functional inhibitor of the Kunitz-type STI family that is able to inhibit the serine protease bovine trypsin and the aspartic protease Plasmepsin II. Despite, the close sequence identity with the potato cathepsin D inhibitor recently reported by Guo and colleagues, our crystal structures show significant differences in loop conformations, especially in loop L9. The analysis of the structural similarities with other inhibitors of the same family, as well as docking-modeling studies suggests loop L5 and L9 as potential binding loops for serine and aspartic proteases, respectively.

2. Methods

2.1. DNA isolation and cloning

Genomic DNA from *Solanum tuberosum* (potato) Estima variety was isolated using the plant CTAB method (ABgene). Primers designed based on the API-8 inhibitor sequence (Strukelj et al., 1992) were used to amplify the inhibitor sequences and clone them in the pGEMT vector (Promega) and positive clones were sequenced. Among the sequences obtained, clone E3Ad, which does not contain introns and matched with GenBank entries XP_006362668.1 and XM_006362606.2, was selected for protein expression and re-amplified to introduce restriction sites for *XhoI* and *Sac II*, for subcloning in the vector pPICZ α C, in frame with the α -factor secretion signal (Table S1). The mutant E3Ad_N19D was generated to delete the only glycosylation site on the E3Ad inhibitor using the construct pPICZ α C-E3Ad as template (Table S1) with the QuickChange Multi-Site-Directed Mutagenesis Kit (Agilent). The clones were sequenced to verify the N19D mutation. The plasmid construct was linearized with the restriction enzyme *Sac I* and transformed into the yeast *Pichia pastoris* (GS115 strain) using the Pichia EasyComp Kit (Life Sciences). Positive clones were checked by colony PCR with 5'AOX1 and 3'AOX1 primers.

2.2. Expression and purification of inhibitors

A single colony expressing the inhibitor E3Ad or E3Ad-N19D was inoculated and pre-cultured in 120 mL of BMGY medium (2% tryptone, 1% yeast extract, 0.1 M potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 40 μ M biotin, 1% glycerol) in baffled flasks at 30 °C and 200 rpm for 24 h. Cells were harvested by centrifugation at 3000g for 10 min at RT and washed with BMMY (2% tryptone, 1% yeast extract, 0.1 M potassium phosphate, pH 6.0, 1.34% yeast nitrogen base, 40 μ M biotin, 1% methanol) medium and harvested as previously described. Expression of the inhibitors was achieved by re-suspension in BMMY medium to a final

OD_{600nm} of 1.0. Recombinant *P. pastoris* cells were incubated with continuous shaking at 28 °C and 200 rpm for 72 h. Methanol (100%) was added every 24 h to a final concentration of 1% to maintain induction. Cells were harvested by centrifugation at 3000g for 10 min at RT and discarded. The supernatant was used for protein purification.

Culture supernatant was extensively dialyzed against 10 mM sodium formate pH 4.0 buffer at 4 °C. The dialysate was adjusted to 150 mM NaCl and applied to a HiPrep™ SP-Sepharose HP column (GE Healthcare Life Sciences). Inhibitors were eluted using a linear gradient of 150–490 mM of NaCl in 10 mM sodium formate pH 4.0 buffer. Fractions were analyzed by SDS-PAGE, pooled and concentrated using 10 kDa cut-off centrifugal filter units (Millipore). Concentrated samples of the inhibitors were applied to a Superdex 75 10/30 HR column (GE Healthcare Life Sciences) equilibrated with 10 mM sodium formate, 150 mM NaCl pH 4.0, coupled to an Äkta-pure system (GE Healthcare Life Sciences). Fractions were analyzed by SDS-PAGE and those containing the inhibitor were pooled.

2.3. Crystallization

Purified inhibitors were concentrated up to 10 mg/mL using centrifugal filter units with a 10 kDa cut-off (Millipore).

Crystallization conditions were screened by the microbatch method using Wizard Classics (I–IV) (Emerald Biosciences) kits at 4 °C and 18 °C. The sitting-drop vapour-diffusion method was used for crystal optimization, with drops consisting of 1 µL of protein and 1 µL of reservoir solution against 0.5 mL of reservoir solution. Promising conditions were further optimized with the use of Hampton Research Additive Screen™. Final pH values were measured using a pH meter.

Diffraction quality crystals for mutant E3Ad_N19D were obtained from 15% (w/v) PEG 20,000, 0.1 M HEPES pH 7.0, 30 mM glycyl-glycyl-glycine (Condition 1, final pH 3.0), 10% (v/v) MPD, 0.1 M Bicine pH 9.0 (Condition 2, final pH 9.0), 20% (w/v) PEG 8000, 0.1 M HEPES pH 7.5, 100 mM glycine (Condition 3, final pH 3.5) and 20% (w/v) PEG 8000, 0.1 M HEPES pH 7.5, 30 mM glycyl-glycyl-glycine (Condition 4, final pH 7.2). Crystals for E3Ad were grown in 10% (w/v) PEG 2000 MME, 0.1 M Tris-HCl pH 8.5, 0.2 M trimethylamine N-oxide, 30 mM glycyl-glycyl-glycine (Condition 5, final pH 7.4) and 20% (w/v) PEG 3350, 0.1 M Tris-HCl pH 7.0, 0.2 M CaCl₂ (Condition 6, final pH 7.0). Single crystals were soaked in reservoir solution containing a cryoprotectant agent (30% (v/v) glycerol or 35% (v/v) PEG 200) and mounted in loops, before flash-cooling in liquid nitrogen at 100 K.

2.4. Data collection and processing

Data were collected at beamline 19ID of the Advanced Photon Source (APS) on an ADSC Q315r CCD detector and at beamline BL14-1 of the Stanford Synchrotron Radiation Lightsource (SSRL) on a MARmosaic 325 CCD detector. Indexing, integration and scaling was performed with XDS and XSCALE (Kabsch, 2010). The reflections of the crystals obtained in conditions 5 and 6 were anisotropic scaled using the web server Diffraction Anisotropy Server (<http://services.mbi.ucla.edu/anisotrope/>) (Strong et al., 2006). The Matthews coefficient was calculated using MATTPROB program (Kantardjieff and Rupp, 2003; Weichenberger and Rupp, 2014). Molecular replacement was performed with PHASER (McCoy et al., 2007) from the CCP4 suite (Winn et al., 2011) using the crystallographic structure of the inhibitor PSPI (PDB code 3TC2, %ID 70.8) as the initial search model for data obtained in the APS. For diffraction data obtained in the SSRL, the structure obtained for inhibitor E3Ad_N19D at 2.45 Å (PDB code 5FNW) was used as template for molecular replacement. In all cases, a cycle of rigid-body refinement was performed with REFMAC5 (Murshudov et al.,

2011). Model refinement was performed with alternate cycles of Phenix (Adams et al., 2010) and manual rebuilding with Coot (Emsley et al., 2010). Model validation was performed with MolProbity (Chen et al., 2010).

2.5. Enzymatic assays

The aspartic protease Plasmepsin II was expressed and purified as described previously (Ramírez et al., 2009). A continuous enzymatic assay was performed using the chromogenic peptide Lys-Pro-Phe-Glu-Phe-Nph-Arg-Leu (Nph: *para*-nitrophenylalanine) as substrate (84 µM final concentration, K_M: 50.8 µM) and 100 mM NaAc pH 4.4 as assay buffer. The reactions were followed at 310 nm in a Cary 60 spectrophotometer (Agilent, USA) at 37 °C. The enzyme concentration in the assay was 180 nM, which was determined by active-site titration with Pepstatin A.

Serine protease inhibition assay was performed with bovine trypsin (EC 3.4.21.4) from Sigma. A continuous enzymatic assay was performed using N α -benzoyl-L-arginine *p*-nitroanilide (BAPNA) as substrate (1 mM final concentration, K_M: 1 mM) and 20 mM Tris-HCl, 150 mM NaCl, 20 mM CaCl₂ pH 8.0 as assay buffer. The reactions were followed at 405 nm in a Cary 60 spectrophotometer (Agilent, USA) at 25 °C. The enzyme concentration in the assay was 66 nM.

For both assays, the enzyme and inhibitor were incubated for 5 min prior the addition of the substrate. The initial velocity was determined at different concentrations of inhibitor E3Ad_N19D using the Cary WinUV Kinetics v5.0 software. Fractional residual activity (v_i/v_0) was calculated from initial reaction velocities for different inhibitor concentrations with respect to an assay in the absence of inhibitor. All assays were performed in triplicate. K_i^{app} values were obtained by non-linear regression fitting of the data to the Morrison equation (Eq. (1)) (Bieth, 1995; Morrison, 1969) with the program Statistica v7.0 (Statsoft). K_i values were calculated using Eq. (2), considering a competitive binding mode, as previously reported for other related bifunctional inhibitors (Keilová and Tomášek, 1976b).

$$\frac{v_i}{v_0} = 1 - \frac{([E]_0 + [I]_0 + K_i^{\text{app}}) - \sqrt{([E]_0 + [I]_0 + K_i^{\text{app}})^2 - 4[E]_0[I]_0}}{2[E]_0} \quad (1)$$

To calculate the stoichiometry of the Trypsin-inhibitor interaction, the enzyme concentration in the assay was increased to be under titration conditions ($[E]_0/K_i^{\text{app}} = 15$). Under these conditions, the curve v_i/v_0 vs $[I]_0/[E]_0$ shows a biphasic behaviour with the x axis intercept being the point where all the enzyme is bound to the inhibitor present in the assay.

$$K_i^{\text{app}} = K_i \left(1 + \frac{[S]_0}{K_M} \right) \quad (2)$$

2.6. Sequence and structural comparison

Multiple sequence and structural alignments were performed using T-Coffee (<http://tcoffee.crg.cat/apps/tcoffee/index.html>) (Di Tommaso et al., 2011) and PDBFold (<http://www.ebi.ac.uk/msd-srv/ssm/>) (Krissinel and Henrick, 2004) web servers, respectively. The final multiple sequence alignment (MSA) was manually parsed using the results obtained from both alignments. A phylogenetic consensus tree was calculated from the curated MSA obtained previously using the Maximum Likelihood method based on the JTT matrix-model and 1000 bootstrap replicates, using the program MEGA6 (Jones et al., 1992; Tamura et al., 2013). Pairwise whole protein and specific loop structural superpositions were performed with the Secondary Structure Matching (SSM) and Least-squares

(LSQ) superposition modules respectively in the program Coot (Emsley et al., 2010).

The root mean square fluctuation (RMSF) among the heavy atoms of all the X-ray structures determined for E3Ad and E3Ad_N19D inhibitors were calculated using the *g_rmsf* tool of the GROMACS software package (version 4.6.5) (Páll et al., 2015). For this analysis the structure of the inhibitor E3Ad_N19D at pH 3.0 (PDB code: 5FNW) was used as the reference.

The metaPPISP web server (<http://pipe.scs.fsu.edu/meta-ppisp.html>) (Qin and Zhou, 2007) was used to predict potential protein-protein interaction sites based on the three-dimensional structures obtained for the inhibitors E3Ad_N19D and PDI. All the residues with a final score above 0.34 were considered as a positive prediction, as suggested by the authors. Rigid body docking was performed with the ClusPro 2.0 web server (<https://clus-pro.bu.edu/login.php>) (Comeau et al., 2004a) using a crystal structure of Plm II in its free form (PDB code: 1LF4) and the structure of the inhibitor E3Ad_N19D at pH 3.0 (PDB code: 5FNW).

2.7. Molecular Dynamics

Energy Minimizations (EM) and Molecular Dynamics (MD) simulations were carried out using the GROMACS software package (version 4.6.5) (Páll et al., 2015) with the AMBER99sb force field (Hornak et al., 2006), and the TIP3P water model (Price and Brooks, 2004). The simulation systems consist of each inhibitor (PDI or E3Ad_N19D) solvated in a dodecahedron box with 7414 water molecules. The protonation states of ionizable residues were assigned at pH 4.4 or pH 8.0 with the program PDB2PQR (http://nbc-222.ucsd.edu/pdb2pqr_1.8/), which uses PROPKA for the prediction of pK_a values (Dolinsky et al., 2007). To achieve the electroneutrality in the system, Na^+ and Cl^- ions were added at an appropriate ratio to reach a NaCl concentration of 0.15 M. At each step, the electrostatic interactions were calculated with the particle-mesh Ewald method (Darden et al., 1993). Van der Waals interactions were described by a Lennard-Jones potential with a cut-off of 1.1 nm that was switched to zero between 1.0 and 1.1 nm. Dispersion corrections for energy and pressure were applied. The SETTLE algorithm (Miyamoto and Kollman, 1992) was used to constrain bonds and angles of water molecules and LINCS (Hess et al., 1997) was used for all other bonds, allowing a time step of 2 fs. The equilibrium simulations were calculated with the leapfrog dynamic integrator (Hockney, 1970) at 310 K and a constant pressure of 1 atm, using the weak coupling to the velocity rescaling thermostat (Bussi et al., 2007) and Parrinello-Rahman barostat (Nosé and Klein, 1983; Parrinello and Rahman, 1981), respectively. Snapshots were saved at 10 ps intervals.

3. Results and discussion

3.1. Overall structure of the bi-functional inhibitors E3Ad and mutant E3Ad_N19D

Inhibitors E3Ad and E3Ad_N19D crystallized in several conditions, with most of them belonging to tetragonal space group $P4_32_1$, with one molecule in the asymmetric unit, and two crystals belonging to orthorhombic space groups $C222_1$ and $P22_12_1$ with two and one molecules in the asymmetric unit, respectively. Even though space group $P22_12_1$ is non-standard, systematic absences and the final refinement of the model supported the space group assignment. Structures were obtained at different pH values and resolutions from 2.43 Å to 2.65 Å (Table 1). Coordinates for more than 98% of the residues were clearly defined by the $2F_o - F_c$ electron density map, with the exception of residues Glu1 and Ser2 in all structures.

High solvent content in most of the crystals may be responsible for the high B-values calculated for most of the structures (Table 1). For the structures reported in this work, a sequential addition of the cryoprotectant to the crystallization drop was crucial in order to get well defined spots with a mosaicity below 1 degree. When the crystals were directly soaked in the cryoprotectant solution both resolution and mosaicity were always negatively affected, which seems to be a result of the high solvent content in most of our crystals.

All structures determined show the β -trefoil fold typical of Kunitz-type STI family inhibitors, with 12 antiparallel β -strands arranged in three pseudo-symmetry elements composed of four β -strands each (Fig. 1). The β -trefoil fold is composed of six β -strands forming a β -barrel (strands β_1 , β_4 , β_5 , β_8 , β_9 and β_{12}) and the other six β -strands arranged as a lid for the barrel. Two β -strands from the barrel and two from the lid form each pseudo-symmetry unit or subdomain. Three short 3_{10} helices have been assigned in the E3Ad_N19D structures at pH 3.0 (PDB code: 5FNW) and pH 3.5 (PDB code: 5FZY), as well as E3Ad structures at pH 7.0 (PDB code: 5FZZ) and pH 7.4 (PDB code: 5G00) comprising residues Arg29-Gly33 (3_{10} -A), Lys91-Tyr96 (3_{10} -B) and Cys152-Gln156 (3_{10} -C). However, in the E3Ad_N19D structures at pH 7.2 (PDB code: 5FZU) and pH 9.0 (PDB code: 5FNX) just two 3_{10} helices have been assigned, running from residues Arg29-Gly33 and Val90-Val94 (PDB code: 5FZU) or Val90-Val94 and Cys152-Gln156 (PDB code: 5FNX). The $2F_o - F_c$ electron density maps of the E3Ad and E3Ad_N19D inhibitor structures clearly showed the presence of three disulfide bridges between positions Cys48-Cys93, Cys142-Cys158 and Cys148-Cys152. These disulfide bridges are present in all Kunitz-type STI family inhibitors able to inhibit aspartic proteases characterized to date, except those classified as part of the group B (Cater et al., 2002; Fischer et al., 2015; Keilová and Tomášek, 1976a; Mares et al., 1989).

In the coordinates of the inhibitor 3EAd (PDB code: 5G00), an N-Acetyl-glucosamine (NAG) molecule was modeled linked to the Asn19. Mass spectrometry analysis of the purified inhibitor E3Ad showed that several glycoforms are present with two molecules of NAG linked to the Asn19 and a variable number of mannose units (results not shown). However, electron density was just visible for one molecule of NAG.

The three-dimensional structures of the inhibitor E3Ad and the mutant E3Ad_N19D at different pH values and space groups are very similar, with RMSD values from 0.21 Å to 0.68 Å over 185 $C\alpha$ atoms (Fig. 2A). However, a more detailed analysis of RMSD differences among all the structures suggests that crystal packing is more important than pH for such differences (Supplementary Table S2). The main differences in the structures are located in loops L1 (28–32), L2 (42–46), and L5 (89–92) (Fig. 2A and B). The presence of three 3_{10} helices present in four crystal structures reported in this work (PDB codes: 5FNW, 5FZY, 5FZZ, 5G00) are located in a pseudo-symmetry equivalent position connected to one β -strand from the β -barrel and one β -strand from the lid (Figs. 1 and 3). To our knowledge, this is the first report of a protein of the Kunitz-type STI family with three helices, with most of the other family members having one α -helix, like STI (PDB code: 1AVW), API-A (PDB code: 3E8L), DrTI (PDB code: 1R8N) or one (PDI (PDB code: 5DZU)) or two (PSPI (PDB code: 3TC2)) 3_{10} helices. In this last structure, the two 3_{10} helices are in equivalent positions to our structures.

3.2. Sequence and structural analysis

The amino acid sequence of inhibitor E3Ad matches with a computationally predicted protein from *S. tuberosum* (potato) (UniProt identifier: M1AKE5_SOLTU). The closest functionally characterized Kunitz-type STI inhibitor deposited in the UniProt database is

Table 1

X-ray statistics for the E3Ad and E3Ad_N19D structures. Values in parenthesis are for the highest resolution shell.

Protein	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad_N19D	E3Ad	E3Ad
Condition (pH)	1 (pH 3.0)	2 (pH 9.0)	3 (pH 3.5)	4 (pH 7.2)	5 (pH 7.4)	6 (pH 7.0)
Beamline	19ID (APS)	19ID (APS)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)	BL14-1 (SSRL)
Wavelength (Å)	0.9792	0.9792	1.1807	1.1807	1.1807	1.1807
Space group	P4 ₃ 22	P4 ₃ 22	C222 ₁	P4 ₃ 22	P4 ₃ 22	P22 ₁ 2 ₁
Unit-cell parameters						
a (Å)	77.22	77.94	102.95	77.14	77.18	47.58
b (Å)	77.22	77.94	114.24	77.14	77.18	56.45
c (Å)	94.24	95.01	98.77	94.14	95.90	72.55
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Mosaic spread (°) ^a	0.14	0.25	0.21	0.13	0.38	0.55
Resolution (Å)	19.50–2.45 (2.50–2.45)	19.45–2.65 (2.75–2.65)	38.42–2.47 (2.57–2.47)	32.0–2.43 (2.53–2.43)	29.5–2.5 (2.57–2.5)	22.6–2.55 (2.62–2.55)
R _{merge} (%) ^b	5.3 (63.5)	4.9 (57.1)	4.7 (63.9)	4.6 (62.8)	5.5 (72.8)	6.5 (24.1)
R _{meas} (%) ^c	5.4 (65.2)	5.1 (58.6)	5.0 (67.9)	4.7 (64.8)	5.7 (75.1)	7.0 (25.8)
CC _{1/2} (%) ^d	99.9 (96.2)	100 (95.4)	100 (95.1)	100 (94.3)	100 (94.2)	99.9 (99.2)
Completeness (%)	99.8 (100)	99.8 (100)	99.8 (100)	99.9 (100)	98.0 (74.1)	98.0 (75.8)
I/σ(I)	38.01 (5.58)	40.61 (5.98)	28.06 (3.24)	42.13 (5.21)	33.82 (4.89)	21.69 (8.08)
Multiplicity	18.5 (19.28)	18.3 (19.24)	8.59 (8.7)	16.4 (16.95)	16.2 (16.63)	7.73 (8.12)
No. of observed reflections	203,134 (12,053)	163,456 (17,575)	182,351 (20,413)	184,496 (21,020)	167,335 (9378)	51,340 (2876)
No. of unique reflections	10,963 (625)	8941 (913)	21,238 (2345)	11,223 (1240)	10,356 (564)	6643 (354)
Wilson plot B factor (Å ²)	58.2	72.4	60.57	56.77	60.16	33.56
Solvent content (%)	64.1	65.0	65.24	63.99	59.62	40.67
R-factor (%)	20.9	20.8	22.53	19.85	22.80	18.1
Free R-factor (%)	26.2	26.0	27.42	23.75	24.37	24.6
RMSD bond lengths (Å)	0.010	0.012	0.010	0.009	0.010	0.009
RMSD bond angles (°)	1.21	1.50	1.39	1.22	1.35	1.16
No. of reflections in working set	10,458	8545	20,203	10,664	9857	6332
No. of reflections in test set	503	391	1035	559	499	311
Mean protein B-factor (Å ²)	58.0	70.0	65.10	58.10	67.40	34.10
Ramachandran plot ^e						
Most favoured (%)	94.02	90.76	92.35	95.63	92.90	96.20
Additional allowed (%)	4.89	7.61	7.38	3.83	6.56	3.80
Outliers (%)	1.09	1.63	0.27	0.55	0.55	0.0
PDB code	5FNW	5FNX	5FZY	5FZU	5G00	5FZZ

^a Calculated with XDS.^b $R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$, where $I_i(hkl)$ and $\langle I(hkl) \rangle$ represent the diffraction-intensity values of the individual measurements and the corresponding mean values. The summation is over all unique measurements.^c Rmeas is a redundancy-independent version of Rmerge: $R_{meas} = \frac{\sum_{hkl} \{N(hkl)/[N(hkl) - 1]\}^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|}{\sum_{hkl} \sum_i I_i(hkl)}$.^d CC_{1/2} values are the half-set correlation coefficients (Karplus and Diederichs, 2012).^e Categories defined by Molprobit.

API9_SOLTU (95.7% sequence identity) (Ritonja et al., 1990), which inhibit porcine pancreatic trypsin (E.C. 3.4.21.4), bovine chymotrypsin (E.C. 3.4.21.1), human leukocyte elastase (EC 3.4.21.37), human and bovine cathepsin D (EC 3.4.23.5) and Saccharopepsin (EC 3.4.21.41) (Cater et al., 2002; Pouvreau et al., 2001). The sequence identity of E3Ad with other inhibitors of the Kunitz-type STI family characterized so far are: PDI (90.4%) (Keilová and Tomášek, 1976a), SLAPI (76.1%) (Cater et al., 2002), PSPI (71.2%) (Meulenbroek et al., 2012), *Murraya koenigii* miraculin-like protein (27.7%) (Gahloth et al., 2010), TKI (26.1%) (Patil et al., 2012) EcTI (25.5%) (Zhou et al., 2013), API-A (24.5%) (Bao et al., 2009), BASI (23.6%) (Vallée et al., 1998) and STI (22.6%) (Song and Suh, 1998). Details of amino acid substitutions in key regions for Kunitz-type STI inhibitors are shown in Fig. 3.

The three-dimensional structure of a potato cathepsin D inhibitor PDI that is closely-related to E3Ad (90.4% sequence identity) was reported very recently (Guo et al., 2015). The structural superposition of E3Ad_N19D structure (PDB code: 5FZU) with PDI, showed that they are very similar in the protein core but differ in the conformation of some loops, like L5 (Ans86-Ile98) L7 (Gly116-Tyr126) L8 (Ser132-Gly136) and L9 (Cys142-Cys158). The E3Ad_N19D structure used for the superposition was selected considering similarity with the pH of the crystallization condition for PDI. One of most interesting features of this structural comparison is the significant conformational difference of the loop L9 (S141–A159) in the two structures (Fig. 4). Most of the sequence

differences between the inhibitors are located in this loop (10 out 17 residues) and also crystallographic packing is different in both cases, which could explain this difference in conformations. Interestingly, this loop was predicted in a previous study, using a combination of computational tools, as the region involved in aspartic protease inhibition (Guerra et al., 2012). In addition, in that work it was also predicted that residues I144, V148, L49, P151, F152 and R154 (tomato inhibitor numbering) could be involved in the difference in potency against the proteases cathepsin D and saccharopepsin reported previously for tomato and potato inhibitors (Cater et al., 2002), and all these residues are localized in loop L9.

A molecular dynamics simulation study was performed to explore whether these conformations could be a result of the crystal packing or could be possible in solution. This analysis did not reveal large conformational changes at loop L9 of the E3Ad_N19D inhibitor but the results showed that loop L9 in PDI structure (5DZU) changes its conformation after 20 ns at pH 4.4 and 310 K (Fig. S2). This conformational rearrangement was not observed in simulations of PDI at pH 8. This is quite interesting because it suggests that loop L9 in PDI could adopt a conformation more similar to the equivalent loop in the E3Ad inhibitor reported in this work.

The analysis performed with the PISA server based on the crystal arrangement, suggested that the E3Ad_N19D inhibitor could exist as a dimer in solution, however, size exclusion chromatogra-

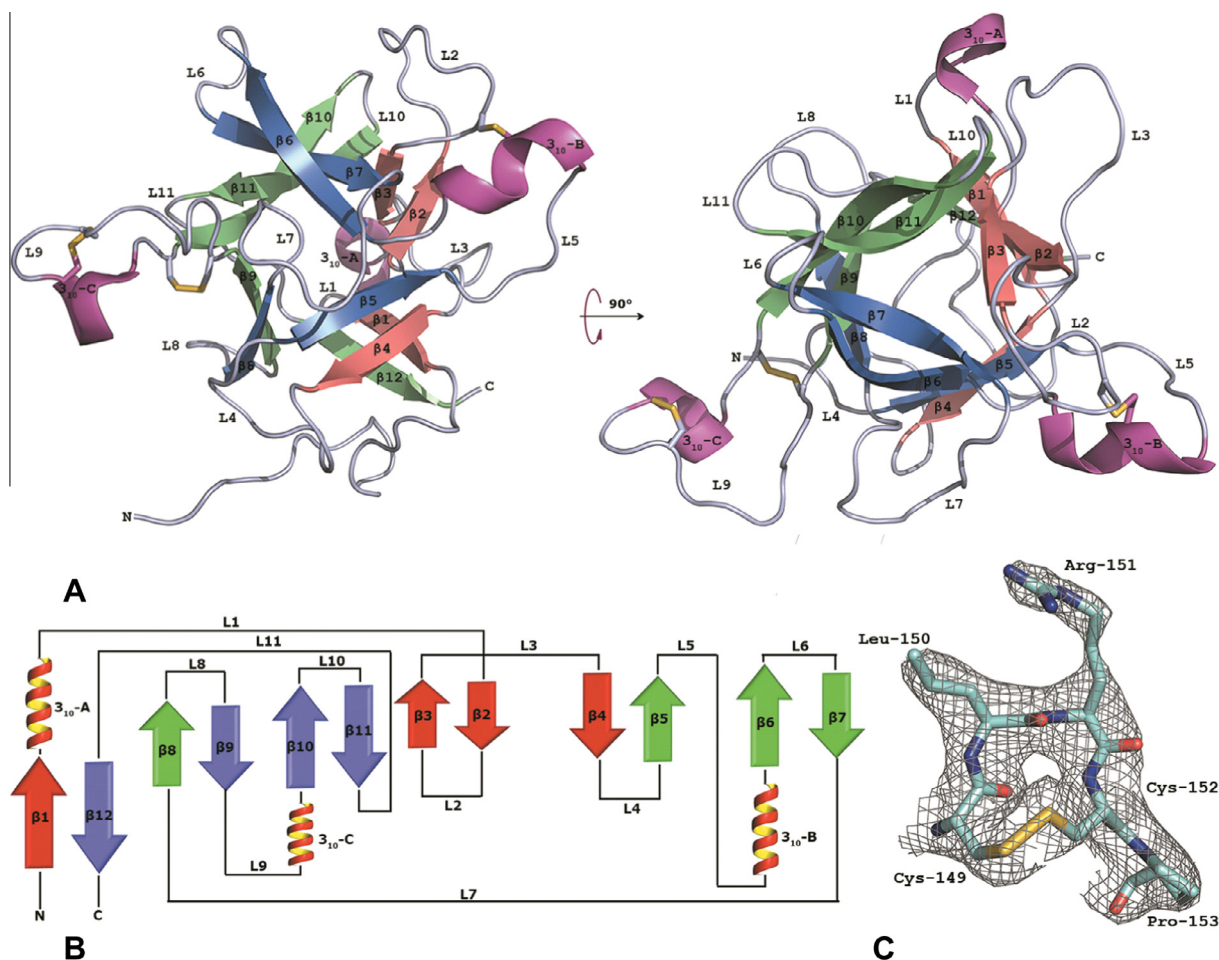


Fig. 1. Overall three-dimensional structure of inhibitor E3Ad_N19D at pH 3.0. β -sheets are colored according the pseudo threefold symmetry elements (A). Topology of the inhibitor E3Ad_N19D (B) Final ($2F_o - F_c$) electron density map around residues Cys149-Pro153 (Loop L9) contoured at 1.0σ (C). The figure was created with PyMOL Molecular Graphics System, Schrödinger, LLC (<http://www.pymol.org>).

phy profiles show that, at least at pH 4.0, E3Ad_N19D behaves as a monomer (Fig. S3). A similar prediction was made for inhibitor PSPI (PDB code: 3TC2), which has been proved to be a monomer in solution (Meulenbroek et al., 2012).

3.3. Enzyme inhibition

Inhibitor E3Ad_N19D showed a typical tight-binding behavior against the aspartic protease Plm II as well as the serine protease bovine trypsin with K_i values of 10.5 and 43 nM respectively (Fig. 5). These results prove that glycosylation is not essential for aspartic protease activity. Previous work has demonstrated that glycosylation was not essential for the inhibition of the serine protease trypsin by Kunitz-type STI family inhibitors (Bao et al., 2009; Speranskaya et al., 2006). The structural superposition of coordinates of the inhibitor E3Ad and E3Ad_N19D showed that the presence of the sugar moiety does not make any significant change to the overall three-dimensional structure of the protein, which correlates well with the inhibitory activity assay results.

Based on structural analysis and the possibility that E3Ad_N19D could be able to interact with two molecules of trypsin in a similar way to that described for the inhibitor PSPI (PDB code: 3TC2), the stoichiometry of the trypsin-E3Ad_N19D binding was calculated using an enzyme titration assay, and an $[I]_0/[E]_0$ value of 1.18 was obtained (Fig. S1). This result proved that the inhibitor E3Ad_N19D interacts with just one molecule of trypsin.

3.4. Protease-binding loop for serine proteases

The Kunitz-type STI inhibitor protease-binding loops involved in serine protease inhibition are highly variable in sequence and length (Fig. 3) with almost all cases adopting a canonical conformation at positions P3-P3'. For most of the Kunitz-type STI family inhibitors structurally characterized so far, the protease-binding loop for serine proteases is located between strands $\beta 4$ and $\beta 5$ (Loop L4), however there are some inhibitors where it lies between strands $\beta 2$ - $\beta 3$ (Loop L2), $\beta 5$ - $\beta 6$ (Loop L5) or $\beta 9$ - $\beta 10$ (Loop L9) (Fig. 3) (Azarkan et al., 2011; Bao et al., 2009). In addition, there is just one double-headed inhibitor (API-A) where one of its protease-binding loops (loop L5) adopts a novel, non-canonical conformation, and even has an unusual Leu P1 site, that was confirmed by site-directed mutagenesis experiments (Bao et al., 2009).

The loop L4 in the E3Ad_N19D inhibitor is shorter and does not contain any equivalent amino acids for positions P1-P1' in comparison with other members with a canonical conformation in this region (Fig. 3). Besides, in all three-dimensional structures of the E3Ad/E3Ad_N19D inhibitors determined in this work, the conformation of the loop L4 is such that the end of the loop is oriented toward the N-terminus (or N-terminal part) of the protein and not protruding to the solvent, an orientation that might make it difficult to allow any interaction with the active site of a protease. Structural comparison of the E3Ad_N19D loop L5 with identified protease-binding loops for serine proteases present in other

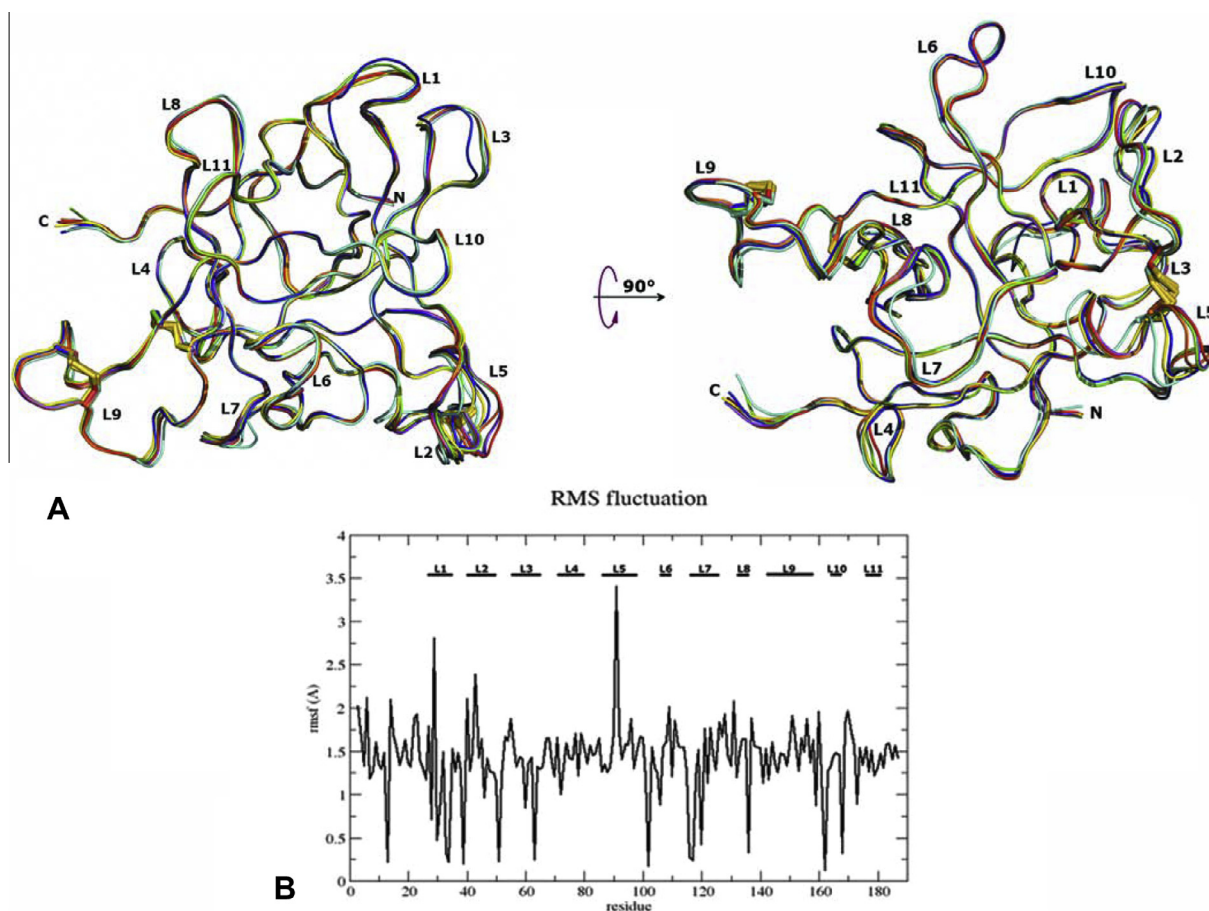


Fig. 2. Structural superposition of all structures of inhibitor E3Ad and mutant E3A_N19D. 5FNW (red), 5FNX (green), 5FZY chain A (blue), 5FZY chain B (yellow), 5FZU (magenta), 5FZZ (cyan) and 5G00 (orange) (A). The root mean square fluctuation (RMSF) among the heavy atoms of all the X-ray structures determined using E3Ad_N19D structure with PDB code 5FNW as the reference (B).

Kunitz-type STI family inhibitors, showed that the closest structural similarity is with the non-canonical loop L5 of API-A. The RMSD values ranged from 0.71 Å to 0.94 Å for residues Val85-Asp90 (P3-P3') for the API-A inhibitor and residues Thr89-Val94 of the E3Ad and E3Ad_N19D inhibitors (Fig. 6A, C). These RMSD values become even lower if we only consider positions P3-P2' (0.13Å to 0.42 Å). In contrast, when the structural superposition was made with the canonical binding loops (mainly L4), the RMSD values were much higher: 2.75 Å (PDB code: 4J2K), 2.79 Å (PDB code: 4AN6) and 2.72 Å (PDB code: 1AVW). Interestingly, the loop L5 of the inhibitors PDI and PSPI (both also isolated from potato), not only have a similar conformation (Fig. 6A), but also have a 3_{10} helix in equivalent residues to API-A and E3Ad_N19D. Phylogenetic analysis showed that inhibitors E3Ad, PDI and PSPI are closer to API-A than to the canonical Kunitz-type STI family inhibitors such as, for example STI (Fig. 6B). These results are similar to those obtained by Azarkan et al., 2011, where they showed that API-A is evolutionarily closer to API8_SOLTU (an aspartic protease inhibitor from the Kunitz-type STI family) than to the canonical inhibitors of the Kunitz-type STI family. It is worth noting that proposed P1 residues for inhibitors PSPI and PDI are also located in loop L5, and, in both cases, they have a Lys in this position as in the E3Ad inhibitor (Fig. 6C) (Guo et al., 2015; Meulenbroek et al., 2012). Lys is a preferred amino acid for the P1 position in protease-binding loop for serine proteases (Krowarsch et al., 2003; Tyndall et al., 2005). On the other hand, despite the fact that amino acid Leu92 present in E3Ad and PDI is not a common P1' residue, the similar amino acid Ile is present in that position in inhibitors with canonical confor-

mations like STI (PDB code: 1AVW), EcTI (PDB code: 4J2K), TKI (PDB code: 4AN4) (Patil et al., 2012) and in both protease-binding loops of the double-headed inhibitor API-A (PDB code: 3TC2) (Fig. 4).

Considering the phylogenetic relationship between E3Ad and API-A, a structural superposition of loop L9 of both inhibitors was performed taking into account that API-A's loop L9 is a canonical protease-binding loop for serine proteases that has been validated by site-directed mutagenesis (Bao et al., 2009). However, there are several insertions in the loop L9 of E3Ad_N19D with respect to API-A and the RMSD values between both loops, considering Phe143-Cys148 (P3-P3') of API-A and Ile148-Pro153 of E3Ad, is higher than 2.0 Å. This difference indicates that loop L9 in E3Ad inhibitor does not adopt a canonical conformation.

In the case of inhibitor PSPI, Meulenbroek et al., 2012 proposed Phe75 as a P1 residue in a secondary protease-binding loop, however, there is no equivalent amino acid in inhibitor E3Ad_N19D (Fig. 3), which could explain why PSPI is able to interact with two serine proteases simultaneously while E3Ad_N19D interacts with just one (Fig. S1). Considering all this evidence, we propose that inhibitor E3Ad (and E3Ad_N19D) could interact with trypsin through a non-canonical protease-binding loop (loop L5), with Lys91 and Leu92 as the P1-P1' residues.

3.5. Protease-binding loop for aspartic proteases

The case of proteinaceous inhibitors of aspartic proteases is quite different from serine proteases. To date, there are only two

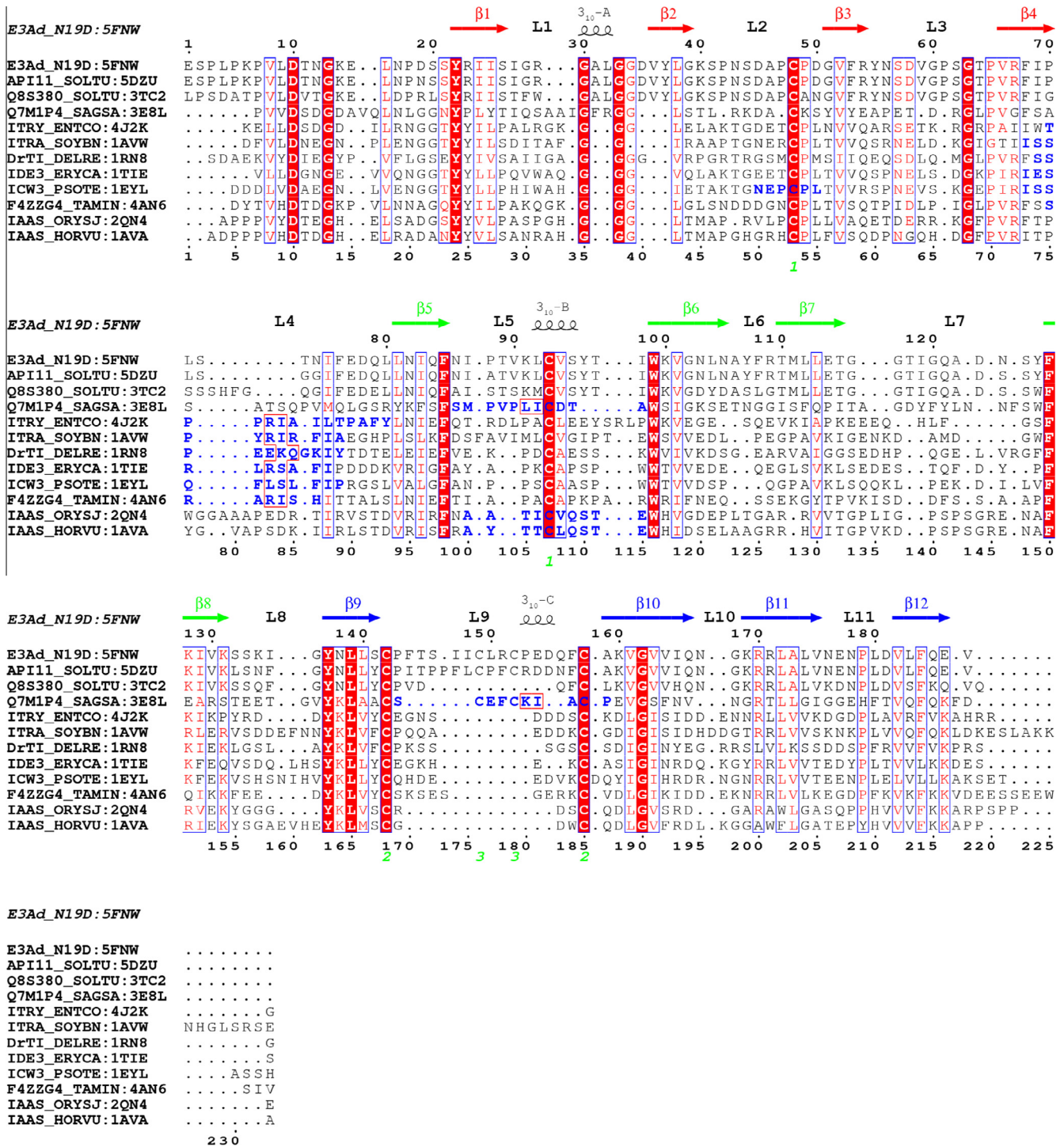


Fig. 3. Multiple sequence alignment of inhibitor E3Ad_N19D with other members of the Kunitz-type STI family for which three-dimensional structures are available, with the protease-binding loops identified. β -sheets are coloured according to the threefold pseudo-symmetry sub-domain to which they belong. Blue and bold: protease-binding loops for serine proteases, red β : P1-P1' residues, green numbers at the bottom show the disulfide bridge topology. The following UniProt identifiers were used for inhibitors: PDI (API11_SOLTU), PSP1 (Q8S380_SOLTU), API-A (Q7M1P4_SAGSA), EcTI (ITRY_ENTCO), STI (ITRA_SOYBN), DrTI (DRTI_DELRE), α -amylase/subtilisin inhibitor from rice (IAAS_ORYSJ), ETI (IDE3_ERYCA), BASI (IAAS_HORVU), double-headed winged bean α -chymotrypsin inhibitor (ICW3_PSOTE) and Tamarind Kunitz inhibitor (F4ZZG4_TAMIN). The figure was made with the ESPript 3.0 webserver (<http://esprict.ibcp.fr/ESPript/cgi-bin/ESPript.cgi>).

three-dimensional structures of protease-inhibitor complexes, with no amino acid sequence or structural similarity between the two inhibitors, or with other inhibitor families (Li et al., 2000; Ng et al., 2000). Previously, combining different computational tools, loop L9 of a tomato aspartic protease inhibitor of the Kunitz-type STI family was predicted as the region responsible for the inhibition of aspartic proteases (Guerra et al., 2012). More recently,

Guo et al., 2015 also suggested that equivalent loop L9 in PDI inhibitor could be implicated in the inhibition of cathepsin D, although their favoured docking models place residue Lys91 (loop L5) interacting with catalytic aspartic acids of cathepsin D.

The analysis of three-dimensional structures of inhibitor E3Ad_N19D and PDI with the metaPPISP web server predicted that in both cases loop L9 could be part of a Protein-Protein interface

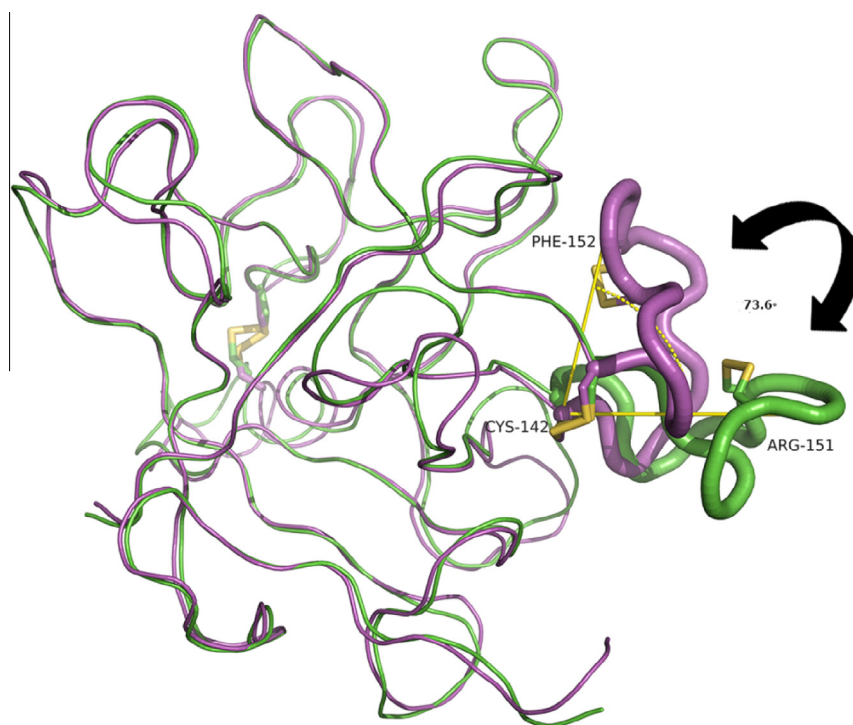


Fig. 4. Structural superposition of E3Ad_N19D (5FZU, green) and PDI (5DZU chain A, magenta). Loop L9 is represented with a higher ribbon radius and the angle between equivalent residues Arg151 (5FZU) and Phe152 (5DZU) is shown.

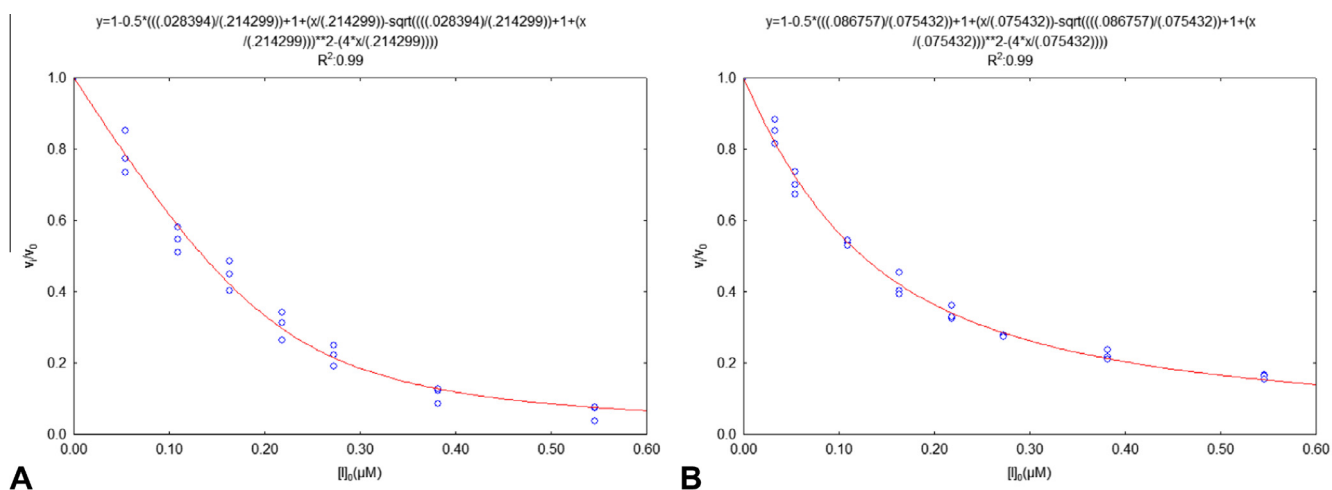


Fig. 5. Inhibition curves fitted to Morrison's equation for inhibitor E3Ad_N19D against the proteases Plm II (A) and bovine trypsin (B).

(Fig. S4). In addition, inhibitor E3Ad_N19D (PDB code: 5FNW) was docked to a three-dimensional apo structure of Plm II (PDB code: 1LF4) using the ClusPro 2.0 web server (Comeau et al., 2004b). The solutions of the balanced and electrostatic coefficients were analyzed, and the top solutions for both coefficients placed the residue Arg151 (loop L9) at interaction distance from both catalytic aspartic acids (Asp34 and Asp214) (Fig. 7A). The roles of basic residues in the inhibition of aspartic proteases have been described for self-inhibition of the aspartic proteases progastricin (Moore et al., 1995), pepsinogen (James and Sielecki, 1986), cathepsin D (Lee et al., 1998) and also for the complex saccharopepsin-inhibitor IA3 (Li et al., 2000). Loop L9 has been previously proposed as the region involved in aspartic protease inhibition (Guerra et al., 2012). Also, it is important to note that Cater et al., 2002 reported that bond hydrolysis between residues Arg154–Gly155 (loop L9) of

the tomato aspartic protease inhibitor SLAPI affects aspartic protease inhibition, doubling the K_i value against saccharopepsin. It is also interesting that loop L9 is in an equivalent sequence position to a canonical protease-binding loop used by the double-headed inhibitor API-A to interact with trypsin (Fig. 3). Considering the present structural and computational analyses, results from Guo et al., 2015 may be explained by a conformational change of loop L9.

Despite the fact that the docking results obtained by Guo et al., 2015 with the inhibitor PDI suggested a different protease-binding loop (loop L5) as their favoured interaction site for aspartic proteases, the results discussed in the present work reinforce our previous hypothesis about the protease-binding loop for aspartic protease inhibition (Guerra et al., 2012). A plausible explanation is that the specific conformation observed in the PDI structure

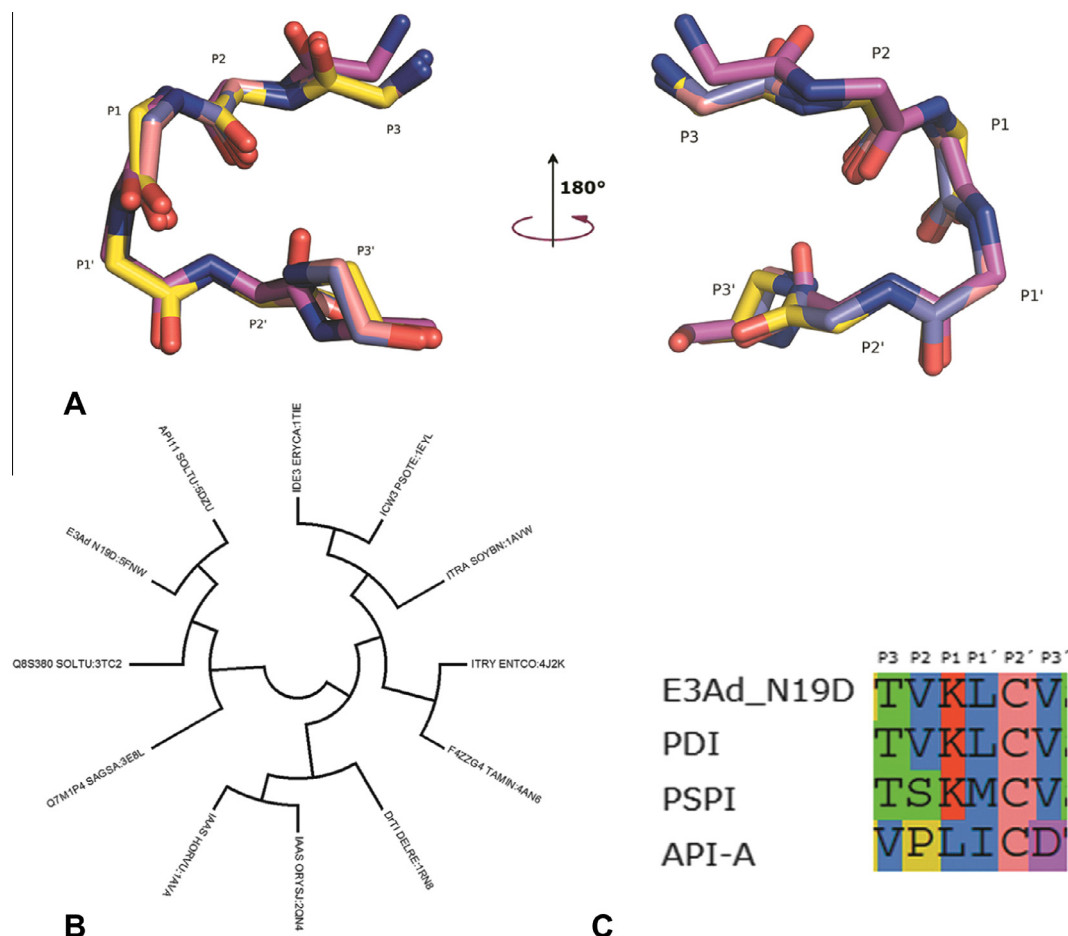


Fig. 6. Structural superposition of loop L5 of inhibitor E3Ad_N19D (5FZU, yellow) with non-canonical protease-binding loop of API-A (3E8L, magenta), loop L5 from PDI (5DZU, salmon) and PSPI (3TC2, blue) (A). Consensus tree (1000 replicates) using the maximum likelihood method calculated for the multiple sequence alignment of Fig. 3. (B). Sequences of the proposed P3-P3' positions in loop L5 for inhibitors E3Ad_N19D, PDI, PSPI and API-A (C).

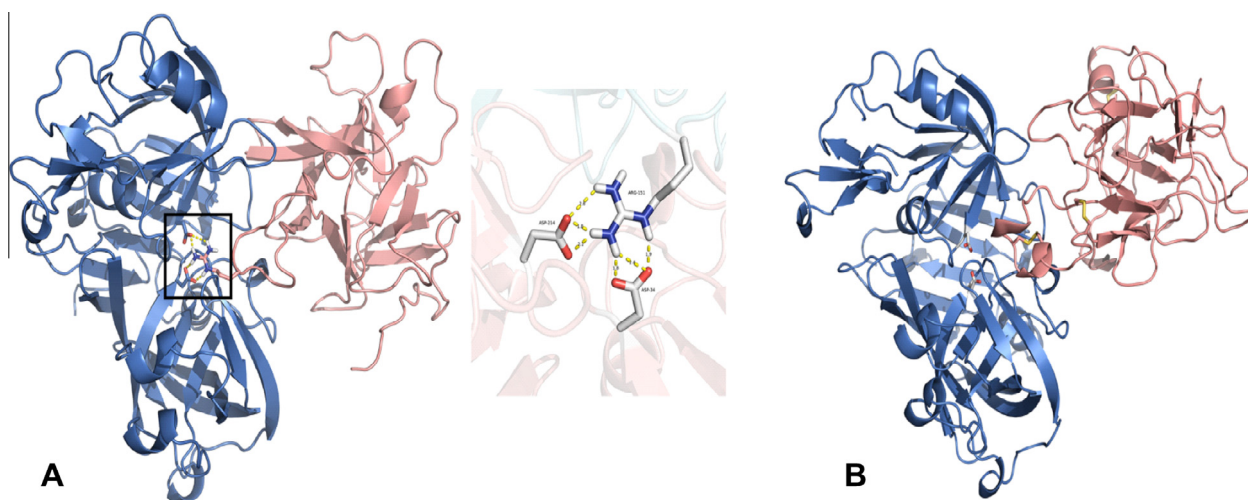


Fig. 7. Docking models of the complex between inhibitors E3Ad_N19D and PDI with Plm II. Interactions of the Arg151 of the inhibitor E3Ad_N19D (PDB code: 5FNW) with Plm II's catalytic aspartic acids (A). Modeled interaction of coordinates of PDI (PDB code: 5DZU), after 93.77 ns of molecular dynamic simulation, with Plm II (B). Plm II and inhibitors (E3Ad_N19D and PDI) are shown in blue and salmon, respectively.

for the loop L9 led to docking solutions where Lys91 is interacting with the active site of cathepsin D. To explore this option we performed a docking study with the coordinates of a molecular dynamics simulation (93.77 ns) with the structure 5DZU at pH 4.4 that showed a change in the conformation of the loop L9 and

the same Plm II apo structure used in docking with E3Ad_N19D. The results showed that in top solutions PDI's loop L9 is placed in the active site of Plm II, although there is no direct interaction with catalytic aspartic acids as shown for E3AD_N19D (Fig. 7B). Even though this approach is theoretical, it is remarkable that

the simulation suggested that if PDI's loop L9 could adopt an extended conformation it could be possible for it to interact with the active site of an aspartic protease, just as has been modeled for the E3Ad_N19D inhibitor.

The presence of a pseudo threefold symmetry, as a result of double gene duplication and fusion events, has been suggested as a selective advantage to develop new functions (Eisenbeis and Höcker, 2010). Bi-functional inhibitors of the I85 family which shares the β -trefoil fold, although having low sequence similarity with Kunitz-type STI family inhibitors, use loops located in different pseudo-symmetry elements to interact with serine and cysteine proteases (Renko et al., 2010). Maybe this fact could be related to a hypothesis proposing that ancestral reactive sites could evolve independently allowing the acquisition of a new functionality or sub-specialization (Christeller, 2005). Considering the present structural and computational analysis of this work, this hypothesis could explain what happened with loop L9 in bifunctional inhibitors of aspartic and serine proteases of the Kunitz-type STI family. Of course this will need experimental studies to confirm or refute the relationship between aspartic and canonical serine protease-binding loops.

4. Conclusions

The crystallographic structures reported here show that inhibitor E3Ad maintains basically the same conformation at different pH values and crystal packings, with only minor changes in some of its loops L1, L2 and L5. Moreover, the structural comparison with the closely-related inhibitor PDI suggests that loop L9 can adopt different conformations depending on pH, loop sequence and also crystal packing. Based on sequence and structural analysis we propose that protease-binding loops for serine and aspartic protease inhibition in E3Ad could be the loops L5 and L9, respectively. The proposed binding loops for serine and aspartic proteases for the bifunctional inhibitor E3Ad are in equivalent loops with two binding loops for serine proteases in the double-headed inhibitor API-A. In addition, phylogenetic analysis suggests that inhibitor E3Ad (and also PDI and PSPI) could have a closer evolutionary relationship with double-headed inhibitor API-A than with members of the Kunitz-type STI-family with canonical protease-binding loops. Whether the inhibition of aspartic proteases in the particular case of the Kunitz-type STI inhibitor family could have evolved from a loop used for serine protease inhibition or not, is an intriguing question that a three-dimensional structure of a complex between an aspartic protease and a bifunctional inhibitor like E3Ad could help to answer.

Acknowledgements

ER-P acknowledges financial support from CONACYT project No. 204639 and PAPIIT IN209114. YG is supported by a PhD fellowship from CONACYT and by Grant F/4927-1 (YG) from International Foundation for Science (Sweden). X-ray diffraction data were collected at beamline 19-ID at the Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source (APS). Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. The authors thank the Argonne staff for their support during data collection, particularly Dr. Norma Duke. This research was also performed at BL14-1 at the SSRL under the agreement between SSRL-SMB and NSLSII-LSBR. SMB operates under NIH-NIGMS grant P41GM103393 and a DOE-BER; and SSRL operates under DOE-BES contract No. DE-AC02-76SF00515. LSBR program is supported by DOE BER contract DE-SC0012704 and by NIH-NIGMS grant P41GM111244. NSLSII is

operated under DOE BES contract # DE-AC02-98CH10886. The authors thank the SSRL and NSLS staff for their support during data collection, particularly Dr. Vivian Stojanoff.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jsb.2016.06.014>.

References

- Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 213–221. <http://dx.doi.org/10.1107/S0907444909052925>.
- Azarkan, M., Martinez-Rodriguez, S., Buts, L., Baeyens-Volant, D., Garcia-Pino, A., 2011. The plasticity of the β -trefoil fold constitutes an evolutionary platform for protease inhibition. *J. Biol. Chem.* 286, 43726–43734. <http://dx.doi.org/10.1074/jbc.M111.291310>.
- Bao, R., Zhou, C.Z., Jiang, C., Lin, S.X., Chi, C.W., Chen, Y., 2009. The ternary structure of the double-headed arrowhead protease inhibitor API-A complexed with two trypsin reveals a novel reactive site conformation. *J. Biol. Chem.* 284, 26676–26684. <http://dx.doi.org/10.1074/jbc.M109.022095>.
- Bauw, G., Nielsen, H.V., Emmersen, J., Nielsen, K.L., Jørgensen, M., Welinder, K.G., 2006. Patatins, Kunitz protease inhibitors and other major proteins in tuber of potato cv. Kuras. *FEBS J.* 273, 3569–3584. <http://dx.doi.org/10.1111/j.1742-4658.2006.05364.x>.
- Bieth, J.G., 1995. Theoretical and practical aspects of proteinase inhibition kinetics. *Methods Enzymol.* 248, 59–84. [http://dx.doi.org/10.1016/0076-6879\(95\)48007-2](http://dx.doi.org/10.1016/0076-6879(95)48007-2).
- Broom, A., Doxey, A.C., Lobsanov, Y.D., Berthin, L.G., Rose, D.R., Howell, P.L., McConkey, B.J., Meiering, E.M., 2012. Modular evolution and the origins of symmetry: reconstruction of a three-fold symmetric globular protein. *Structure* 20, 161–171. <http://dx.doi.org/10.1016/j.str.2011.10.021>.
- Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126. <http://dx.doi.org/10.1063/1.2408420>.
- Cater, S.A., Lees, W.E., Hill, J., Brzin, J., Kay, J., Phylip, L.H., 2002. Aspartic proteinase inhibitors from tomato and potato are more potent against yeast proteinase A than cathepsin D. *Biochim. Biophys. Acta – Protein Struct. Mol. Enzymol.* 1596, 76–82. [http://dx.doi.org/10.1016/S0167-4838\(02\)00206-6](http://dx.doi.org/10.1016/S0167-4838(02)00206-6).
- 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. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21. <http://dx.doi.org/10.1107/S0907444909042073>.
- Christeller, J.T., 2005. Evolutionary mechanisms acting on proteinase inhibitor variability. *FEBS J.* 272, 5710–5722. <http://dx.doi.org/10.1111/j.1742-4658.2005.04975.x>.
- Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004a. ClusPro: a fully automated algorithm for protein-protein docking. *Nucleic Acids Res.* 32, 96–99. <http://dx.doi.org/10.1093/nar/gkh354>.
- Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004b. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 20, 45–50. <http://dx.doi.org/10.1093/bioinformatics/btg371>.
- Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobítz, M., Montanyola, A., Chang, J.-M., Taly, J.-F., Notredame, C., 2011. T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* 39, W13–W17. <http://dx.doi.org/10.1093/nar/gkr245>.
- Dolinsky, T.J., Czodrowski, P., Li, H., Nielsen, J.E., Jensen, J.H., Klebe, G., Baker, N.A., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* 35, W522–W525. <http://dx.doi.org/10.1093/nar/gkm276>.
- Eisenbeis, S., Höcker, B., 2010. Evolutionary mechanism as a template for protein engineering. *J. Pept. Sci.* 16, 538–544. <http://dx.doi.org/10.1002/psc.1233>.
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501. <http://dx.doi.org/10.1107/S0907444910007493>.
- Fischer, M., Kuckenberg, M., Kastilan, R., Muth, J., Gebhardt, C., 2015. Novel in vitro inhibitory functions of potato tuber proteinaceous inhibitors. *Mol. Genet. Genomics* 290, 387–398. <http://dx.doi.org/10.1007/s00438-014-0906-5>.
- Franco, O.L., De Grossi Sá, M.F., Sales, M.P., Mello, L.V., Oliveira, A.S., Rigden, D.J., 2002. Overlapping binding sites for trypsin and papain on a Kunitz-type proteinase inhibitor from *Prosopis juliflora*. *Proteins Struct. Funct. Genet.* 49, 335–341. <http://dx.doi.org/10.1002/prot.10228>.
- Gahltho, D., Selvakumar, P., Shee, C., Kumar, P., Sharma, A.K., 2010. Cloning, sequence analysis and crystal structure determination of a miraculin-like protein from *Murraya koenigii*. *Arch. Biochem. Biophys.* 494, 15–22. <http://dx.doi.org/10.1016/j.abb.2009.11.008>.

- Glaczinski, H., Heibges, A., Salamini, F., Gebhardt, C., 2002. Members of the Kunitz-type protease inhibitor gene family of potato inhibit soluble tuber invertase in vitro. *Potato Res.* 45, 163–176. <http://dx.doi.org/10.1007/BF02736112>.
- Guerra, Y., Valiente, P.A., Berry, C., Pons, T., 2012. Predicting functional residues of the *Solanum lycopersicum* aspartic protease inhibitor (SLAPI) by combining sequence and structural analysis with molecular docking. *J. Mol. Model.* 18, 2673–2687. <http://dx.doi.org/10.1007/s00894-011-1290-2>.
- Guo, J., Erskine, P.T., Coker, A.R., Wood, S.P., Cooper, J.B., 2015. Structure of a Kunitz-type potato cathepsin D inhibitor. *J. Struct. Biol.* 192, 554–560. <http://dx.doi.org/10.1016/j.jbsb.2015.10.020>.
- Headey, S.J., MacAskill, U.K., Wright, M.A., Claridge, J.K., Edwards, P.J.B., Farley, P.C., Christeller, J.T., Laing, W.A., Pascal, S.M., 2010. Solution structure of the squash aspartic acid proteinase inhibitor (SQAPI) and mutational analysis of pepsin inhibition. *J. Biol. Chem.* 285, 27019–27025. <http://dx.doi.org/10.1074/jbc.M110.137018>.
- Heibges, A., Glaczinski, H., Ballvora, A., Salamini, F., Gebhardt, C., 2003a. Structural diversity and organization of three gene families for Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Mol. Genet. Genomics* 269, 526–534. <http://dx.doi.org/10.1007/s00438-003-0860-0>.
- Heibges, A., Salamini, F., Gebhardt, C., 2003b. Functional comparison of homologous members of three groups of Kunitz-type enzyme inhibitors from potato tubers (*Solanum tuberosum* L.). *Mol. Genet. Genomics* 269, 535–541. <http://dx.doi.org/10.1007/s00438-003-0861-z>.
- Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINC3: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. [http://dx.doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](http://dx.doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).
- Hockney, R.W., 1970. The potential calculation and some applications. In: *Methods in Computational Physics*, pp. 136–211.
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., Simmerling, C., 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* 65, 712–725. <http://dx.doi.org/10.1002/prot.21123>.
- James, M.N.G., Sielecki, A.R., 1986. Molecular structure of an aspartic proteinase zymogen, porcine pepsinogen, at 1.8 Å resolution. *Nature* 319, 33–38.
- Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282. <http://dx.doi.org/10.1093/bioinformatics/8.3.275>.
- Kabsch, W., 2010. Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66, 125–132. <http://dx.doi.org/10.1107/S0907444909047337>.
- Kantardjiev, K.A., Rupp, B., 2003. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci.* 12, 1865–1871. <http://dx.doi.org/10.1110/ps.0350503>.
- Karplus, P.A., Diederichs, K., 2012. Linking crystallographic model and data quality. *Science* 336, 1030–1033. <http://dx.doi.org/10.1126/science.1218231> (80-).
- Keilová, H., Tomášek, V., 1976a. Isolation and properties of Cathepsin D inhibitor from potatoes. *Collect. Czech. Chem. Commun.* 41, 489–497.
- Keilová, H., Tomášek, V., 1976b. Further characteristics of cathepsin D inhibitor from potatoes. *Collect. Czech. Chem. Commun.* 41, 2440–2447.
- Krissinel, E., Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* 60, 2256–2268.
- Krowarsch, D., Cierpicki, T., Jelen, F., Otlewski, J., 2003. Canonical protein inhibitors of serine proteases. *Cell. Mol. Life Sci.* 60, 2427–2444. <http://dx.doi.org/10.1007/s00181-003-3120-x>.
- Laskowski, M., Kato, I., 1980. Protein inhibitors of proteinases. *Annu. Rev. Biochem.* 49, 593–626. <http://dx.doi.org/10.1146/annurev.bi.49.070180.003113>.
- Laskowski, M., Qasim, M.A., 2000. What can the structures of enzyme–inhibitor complexes tell us about the structures of enzyme substrate complexes? *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* 1477, 324–337. [http://dx.doi.org/10.1016/S0167-4838\(99\)00284-8](http://dx.doi.org/10.1016/S0167-4838(99)00284-8).
- Lee, A.Y., Gulnik, S.V., Erickson, J.W., 1998. Conformational switching in an aspartic proteinase. *Nat. Struct. Biol.* 5, 866–871. <http://dx.doi.org/10.1038/2306>.
- Li, M., Phylip, L.H., Lees, W.E., Winther, J.R., Dunn, B.M., Wlodawer, A., Kay, J., Gustchina, A., 2000. The aspartic proteinase from *Saccharomyces cerevisiae* folds its own inhibitor into a helix. *Nat. Struct. Biol.* 7, 113–117. <http://dx.doi.org/10.1038/72378>.
- Mares, M., Meloun, B., Pavlik, M., Kostka, V., Baudys, M., 1989. Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Lett.* 251, 94–98. [http://dx.doi.org/10.1016/0014-5793\(89\)81435-8](http://dx.doi.org/10.1016/0014-5793(89)81435-8).
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674. <http://dx.doi.org/10.1107/S0021889807021206>.
- McLachlan, A.D., 1979. Three-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J. Mol. Biol.* 133, 557–563. [http://dx.doi.org/10.1016/0022-2836\(79\)90408-X](http://dx.doi.org/10.1016/0022-2836(79)90408-X).
- Meulenbroek, E.M., Thomassen, E.A.J., Pouvreau, L., Abrahams, J.P., Gruppen, H., Pannu, N.S., 2012. Structure of a post-translationally processed heterodimeric double-headed Kunitz-type serine protease inhibitor from potato. *Acta Crystallogr. D Biol. Crystallogr.* 68, 794–799. <http://dx.doi.org/10.1107/S090744491201222X>.
- Miyamoto, S., Kollman, P.A., 1992. Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* 13, 952–962. <http://dx.doi.org/10.1002/jcc.540130805>.
- Moore, S.A., Sielecki, A.R., Chernai, M.M., Tarasova, N.I., James, M.N.G., 1995. Crystal and molecular structures of human progastricin at 1.62 Å resolution. *J. Mol. Biol.* 247, 466–485. <http://dx.doi.org/10.1006/jmbi.1994.0154>.
- Morrison, J.F., 1969. Kinetics of the reversible inhibition of enzyme-catalysed reactions by tight-binding inhibitors. *Biochim. Biophys. Acta – Enzymol.* 185, 269–286. [http://dx.doi.org/10.1016/0005-2744\(69\)90420-3](http://dx.doi.org/10.1016/0005-2744(69)90420-3).
- Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., Vagin, A.A., 2011. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* 67, 355–367. <http://dx.doi.org/10.1107/S0907444911001314>.
- Murzin, A.G., Lesk, A.M., Chothia, C., 1992. β -Trefoil fold. *J. Mol. Biol.* 223, 531–543. [http://dx.doi.org/10.1016/0022-2836\(92\)90668-A](http://dx.doi.org/10.1016/0022-2836(92)90668-A).
- Ng, K.K., Petersen, J.F., Cherney, M.M., Garen, C., Zalatoris, J.J., Rao-Naik, C., Dunn, B.M., Martzen, M.R., Peanasky, R.J., James, M.N., 2000. Structural basis for the inhibition of porcine pepsin by Ascaris pepsin inhibitor-3. *Nat. Struct. Biol.* 7, 653–657. <http://dx.doi.org/10.1038/77950>.
- Nosé, S., Klein, M.L., 1983. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* 50, 1055–1076. <http://dx.doi.org/10.1080/00268978300102851>.
- Oliveira, M.L.V., Silva, M.C.C., Sallai, R.C., Brito, M.V., Sampaio, M.U., 2010. A novel subclassification for Kunitz proteinase inhibitors from leguminous seeds. *Biochimie* 92, 1667–1673. <http://dx.doi.org/10.1016/j.biochi.2010.03.021>.
- Otlewski, J., Jaskólski, M., Buczek, O., Cierpicki, T., 2001. Structure–function relationship of serine protease–protein inhibitor interaction. *Acta Biochim. Pol.* 48, 419–428.
- Páll, S., Abraham, M.J., Carsten, K., Hess, B., Lindahl, E., 2015. Tackling Exascale Software Challenges in Molecular Simulations with GROMACS, Vol. 8759, pp. 3–27. <http://dx.doi.org/10.1007/978-3-319-15976-8>.
- Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* 52.
- Patil, D.N., Chaudhary, A., Sharma, A.K., Tomar, S., Kumar, P., 2012. Structural basis for dual inhibitory role of tamarind Kunitz inhibitor (TKI) against factor Xa and trypsin. *FEBS J.* 279, 4547–4564. <http://dx.doi.org/10.1111/febs.12042>.
- Ponting, C.P., Russell, R.B., 2000. Identification of distant homologues of fibroblast growth factors suggests a common ancestor for all β -trefoil proteins 1. *J. Mol. Biol.* 302, 1041–1047. <http://dx.doi.org/10.1006/jmbi.2000.4087>.
- Pouvreau, L., Gruppen, H., Piersma, S.R., van den Broek, L.A., van Koningsveld, G.A., Voragen, A.G., 2001. Relative abundance and inhibitory distribution of protease inhibitors in potato juice from cv. Elkana. *J. Agric. Food Chem.* 49, 2864–2874. <http://dx.doi.org/10.1021/jf010126v>.
- Price, D.J., Brooks, C.L., 2004. A modified TIP3P water potential for simulation with Ewald summation. *J. Chem. Phys.* 121.
- Qin, S., Zhou, H.-X., 2007. Meta-PPISP: a meta web server for protein–protein interaction site prediction. *Bioinformatics* 23, 3386–3387. <http://dx.doi.org/10.1093/bioinformatics/btm434>.
- Radisky, E.S., Koshland, D.E., 2002. A clogged gutter mechanism for protease inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 99, 10316–10321. <http://dx.doi.org/10.1073/pnas.112332899>.
- Ramírez, A.R., Guerra, Y., Otero, A., García, B., Berry, C., Mendiola, J., Hernández-Zanui, A., Chávez, M., de los, A., 2009. Generation of an affinity matrix useful in the purification of natural inhibitors of plasmepsin II, an antimalarial-drug target. *Biotechnol. Appl. Biochem.* 52, 149–157. <http://dx.doi.org/10.1042/BA20080015>.
- Rawlings, N.D., Waller, M., Barrett, A.J., Bateman, A., 2014. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* 42, 503–509. <http://dx.doi.org/10.1093/nar/gkt953>.
- Renko, M., Sabotič, J., Mihelič, M., Brzin, J., Kos, J., Turk, D., 2010. Versatile loops in mycopycybins inhibit three protease families. *J. Biol. Chem.* 285, 308–316. <http://dx.doi.org/10.1074/jbc.M109.043331>.
- Renko, M., Sabotič, J., Turk, D., 2012. β -Trefoil inhibitors – from the work of Kunitz onward. *Biol. Chem.* 393, 1043–1054. <http://dx.doi.org/10.1515/hsz-2012-0159>.
- Ritonja, A., Krizaj, I., Mesko, P., Kopitar, M., Lucovnik, P., Strukelj, B., Pungercar, J., Buttje, D.J., Barrett, A.J., Turk, V., 1990. The amino acid sequence of a novel inhibitor of cathepsin D from potato. *FEBS Lett.* 267, 13–15.
- Schechter, I., Berger, A., 1967. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* 27, 157–162. [http://dx.doi.org/10.1016/S0006-291X\(67\)80055-X](http://dx.doi.org/10.1016/S0006-291X(67)80055-X).
- Song, H.K., 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, 347–363. <http://dx.doi.org/10.1006/jmbi.1997.1469>.
- Speranskaya, A.S., Krinitsina, A.A., Revina, T.A., Gerasimova, N.G., Keruchen'ko, Y.S., Shevelev, A.B., Valueva, T.A., 2006. Heterologous expression, purification, and properties of a potato protein inhibitor of serine proteinases. *Biochemistry (Moscow)* 71, 1176–1182. <http://dx.doi.org/10.1134/S0006297906110022>.
- Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., Eisenberg, D., 2006. Toward the structural genomics of complexes: crystal structure of a PE/PPe protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8060–8065. <http://dx.doi.org/10.1073/pnas.0602606103>.
- Strukelj, B., Pungercar, J., Mesko, P., Barlic-Maganja, D., Gubensek, F., Kregar, I., Turk, V., 1992. Characterization of aspartic proteinase inhibitors from potato at the gene, cDNA and protein levels. *Biol. Chem. Hoppe Seyler* 373, 477–482.
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. <http://dx.doi.org/10.1093/molbev/mst197>.

- Tyndall, J.D.A., Nall, T., Fairlie, D.P., 2005. Proteases universally recognize beta strands in their active sites. *Chem. Rev.* 105, 973–999. <http://dx.doi.org/10.1021/cr040669e>.
- Vallée, F., Kadziola, A., Bourne, Y., Juy, M., Rodenburg, K.W., Svensson, B., Haser, R., 1998. Barley α -amylase bound to its endogenous protein inhibitor BASI: crystal structure of the complex at 1.9 Å resolution. *Structure* 6, 649–659. [http://dx.doi.org/10.1016/S0969-2126\(98\)00066-5](http://dx.doi.org/10.1016/S0969-2126(98)00066-5).
- Weichenberger, C.X., Rupp, B., 2014. Ten years of probabilistic estimates of biocrystal solvent content: new insights via nonparametric kernel density estimate. *Acta Crystallogr. D* 70, 1579–1588.
- Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., McNicholas, S.J., Murshudov, G. N., Pannu, N.S., Potterton, E.A., Powell, H.R., Read, R.J., Vagin, A., Wilson, K.S., 2011. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67, 235–242. <http://dx.doi.org/10.1107/S0907444910045749>.
- Zakharova, E., Horvath, M.P., Goldenberg, D.P., 2009. Structure of a serine protease poised to resynthesize a peptide bond. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11034–11039. <http://dx.doi.org/10.1073/pnas.0902463106>.
- Zhou, D., Lobo, Y.A., Batista, I.F.C., Marques-Porto, R., Gustchina, A., Oliva, M.L.V., Wlodawer, A., 2013. Crystal structures of a plant trypsin inhibitor from *Enterolobium contortisiliquum* (EcTI) and of its complex with bovine trypsin. *PLoS ONE* 8. <http://dx.doi.org/10.1371/journal.pone.0062252>.