

UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS



***PREDICCIÓN DE ESTRUCTURA Y SIMULACIÓN
MOLECULAR DE LA LIPOPROTEÍNA
LICANANTASA DE LA CEPA LICANANTAY:
ACIDITHIOBACILLUS THIOOXIDANS DSM 17318***

**Tesis presentada a la Universidad de Chile para optar al grado
de Magíster en Bioquímica área de especialización en
Bioquímica de Proteínas Recombinantes, y Memoria para
optar al Título de Bioquímico por:**

FERNANDO FABIÁN ABARCA FLORES

Directores de Tesis:

**Dr. Tomás Pérez-Acle
Dra. Pilar Parada**

Santiago, Chile

2012

UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS QUÍMICAS Y FARMACÉUTICAS
INFORME DE APROBACIÓN
TESIS DE MAGÍSTER Y TÍTULO DE BIOQUÍMICO

Se informa a la Dirección de la Escuela de Graduados de la Facultad de Ciencias Químicas y Farmacéuticas que la Tesis de Magíster y Memoria de Título presentada por el candidato

FERNANDO FABIÁN ABARCA FLORES

Ha sido aprobada por la Comisión de Evaluadora de Tesis como requisito para optar al grado de Magíster en Bioquímica, área de especialización Bioquímica de Proteínas Recombinantes, y Título de Bioquímico, en el examen público rendido el día _____.

Directores de Tesis:

Dr. Tomás Pérez-Acle

Dra. Pilar Parada

Comisión Evaluadora de Tesis:

Dr. Sergio Álvarez (Presidente)

Dr. Francisco Chávez

Dr. Davor Cotoras

Dedicada a mi familia:

María Teresa, Luis y Luis Pablo.

“Me lo contaron y lo olvidé.

Lo vi y lo entendí.

Lo hice y lo aprendí.”

Confucio

AGRADECIMIENTOS

Agradezco a mi director de tesis Dr. Tomás Pérez-Acle por permitirme incursionar en la biología computacional y brindarme su confianza, apoyo y guía durante esta tesis.

A la Dra. Pilar Parada, quien patrocinó esta tesis, por su dirección y apoyo en la realización de este trabajo.

Al Dr. Alejandro Maass, por dar origen a esta colaboración y apoyar económicamente el desarrollo de este proyecto.

A los profesores de mi comisión Dr. Sergio Álvarez, Dr. Francisco Chávez y Dr. Davor Cotoras por sus consejos y discusión que permitieron enriquecer este trabajo.

A mis compañeros y amigos de laboratorio (DLab), en particular a Sebastián G., Sebastián F., Felipe, Raúl, Héctor y Carlos, por la ayuda técnica, discusión y consejos dados, y por su buen ánimo que hicieron muy grata la estancia en el laboratorio.

A mis compañeros y amigos de la universidad, en especial a Bernardo, Sebastián, Pablo M., Cristian, Marianne, Pablo B., Damián, María, Camila y Catalina, por los buenos momentos compartidos, su amistad y apoyo.

A mis padres, María Teresa y Luis, y mi hermano, Luis Pablo, por su compañía, cariño y apoyo, que estuvieron siempre que lo necesité.

A la empresa BioSigma S.A. por facilitar la figura de entrecruzamiento químico, y a Patricio Martínez y Gloria Garín que fueron quienes obtuvieron estos resultados.

Al Laboratorio Nacional de Computación de Alto Rendimiento (National Laboratory for High Performance Computing, NLHPC), por proporcionar la infraestructura y tiempo de cómputo necesario para llevar a cabo este trabajo "Powered@NLHPC".

FINANCIAMIENTO

Esta Tesis fue realizada en el Laboratorio de Biología Computacional de la Fundación Ciencia & Vida, y contó con financiamiento de los proyectos FONDAP CRG 15090007 y ICM-ECONOMIA PROYECTO P09-022-F.

TABLA DE CONTENIDO

DEDICATORIA	II
AGRADECIMIENTOS.....	III
FINANCIAMIENTO	IV
TABLA DE CONTENIDO	V
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS.....	IX
ABREVIATURAS	X
RESUMEN.....	XII
SUMMARY	XIV
1 INTRODUCCIÓN	1
1.1 Biolixiviación	1
1.2 Licanantasa.....	3
1.3 Análisis basados en secuencias.....	4
1.4 Lipoproteína de Braun.....	6
1.5 Estructura de Lipoproteína de Braun y sus mutantes	7
1.6 Estructura de Licanantasa.....	9
2 HIPÓTESIS.....	12
3 OBJETIVO GENERAL	12
4 OBJETIVOS ESPECÍFICOS	12
5 METODOLOGÍA	13
5.1 Modelamiento <i>de novo</i> :.....	13
5.2 Construcción de Lpp-58:	13
5.3 Construcción y preparación de sistemas de simulación:	13
5.4 Simulación de Dinámica Molecular con Intercambio de Réplicas:	14

5.5	Análisis:	14
6	RESULTADOS Y DISCUSIÓN.....	17
6.1	Modelamiento <i>de novo</i> de la estructura trimérica de Licanantasa:	17
6.2	Simulación de Dinámica Molecular preparativa:	20
6.3	Simulación de Dinámica Molecular con Intercambio de Réplicas:	25
7	CONCLUSIONES	38
8	BIBLIOGRAFÍA	39
9	APÉNDICE.....	44
9.1	Dinámica Molecular.....	44
9.2	Campo de fuerza.....	45
9.3	Algoritmo de integración.....	46
9.4	Solvente implícito	47
9.5	Replica Exchange	48
9.6	Algoritmo de predicción <i>de novo</i> de Rosetta	49

ÍNDICE DE FIGURAS

Figura 1. Alineamiento de secuencia de aminoácidos de Licanantasa con lipoproteínas hipotéticas.....	4
Figura 2. Alineamiento de secuencia de aminoácidos de Licanantasa con la secuencia de la mutante Ala-14 de Lpp (PDB ID: 1JCD).....	5
Figura 3. Predicciones de dominios a partir de la secuencia de Licanantasa.....	6
Figura 4. Procesamiento de lipoproteínas.....	7
Figura 5. Representación de la estructura de Lpp-56.....	8
Figura 6. Representación tipo “helical wheel”.	9
Figura 7. Grupo principal de estructuras del monómero de Lic-80 predicho mediante modelamiento <i>de novo</i>	17
Figura 8. Estructuras triméricas de Lic-80, Lpp-58 y Ala-14.....	18
Figura 9. Entrecruzamiento químico de Licanantasa en <i>E. coli</i>	19
Figura 10. RMSD de carbonos alfa en las simulaciones preparativas.....	20
Figura 11. RMSF de carbonos alfa en las simulaciones preparativas.	21
Figura 12. Pérdida de puentes de hidrógeno inter-monómero en Lpp-58.	21
Figura 13. Puentes de hidrógeno y puentes salinos inter-monómeros en las simulaciones preparativas.	22
Figura 14. Contenido hélice alfa en las simulaciones preparativas.	23
Figura 15. Puentes de hidrógeno de las cadenas principales en las simulaciones preparativas.	23
Figura 16. Contactos hidrofóbicos en las simulaciones preparativas.	24
Figura 17. Porcentaje de contactos hidrofóbicos en las simulaciones de REMD.....	30
Figura 18. Porcentaje puentes de hidrógeno de cadena principal en las simulaciones de REMD.	31

Figura 19. Porcentaje de contenido de hélice alfa en las simulaciones de REMD.....	32
Figura 20. Porcentaje de puentes de hidrógeno inter-monómero en las simulaciones de REMD.	33
Figura 21. Porcentaje de puentes de puentes salinos en las simulaciones de REMD.	34
Figura 22. Probabilidad de intercambio en las simulaciones de REMD.....	35
Figura 23. Representación mediante WHAM del espacio conformacional muestreado.	36
Figura 24. Promedio de propiedades estructurales versus temperatura.	37
Figura A1. Método de integración Leap-Frog.....	46
Figura A2. Representación del protocolo de REMD para ocho replicas.	49
Figura A3. Algoritmo de plegamiento <i>de novo</i>	50

ÍNDICE DE TABLAS

Tabla I. Temperaturas de despliegamiento.....	29
---	----

ABREVIATURAS

Å: Angstrom (1×10^{-10} metros).

aa: Amino ácido.

A14: Proteína de 52 residuos derivada de Lpp-56 que contiene 14 alaninas en su núcleo hidrofóbico.

DSP: 3,3'-ditiobis(succinimidilpropionato).

DTSSP: 3,3'-ditiobis(sulfosuccinimidilpropionato).

fs: Femtosegundo (1×10^{-15} segundos).

GB/SA: Born Generalizado / Accesibilidad a Solvente (Generalized Born / Solvent Accessibility).

GROMACS: Máquina de Groningen para Simulaciones Químicas (Groningen Machine for Chemical Simulations).

kDa: Kilodalton (1×10^3 Daltons).

Lic-80: Proteína de 80 residuos correspondiente a la secuencia madura de Licanantasa.

Lpp: Lipoproteína de *Escherichia coli*.

Lpp-56: Proteína de 56 residuos derivada de la secuencia madura de Lpp sin la cisteína 1 ni la lisina 58.

Lpp-58: Proteína de 58 residuos correspondiente a la secuencia madura de Lpp.

µs: Microsegundo (1×10^{-6} segundos).

MD: Dinámica Molecular (Molecular Dynamics).

nm: Nanómetro (1×10^{-9} metros).

ns: Nanosegundo (1×10^{-9} segundos).

PDB: Banco de Datos de Proteínas (Protein Data Bank).

PDB ID: Identificador de Banco de Datos de Proteínas (Protein Data Bank Identifier).

ps: Picosegundo (1×10^{-12} segundos).

REMD: Dinámica Molecular con Intercambio de Réplicas (Replica-Exchange Molecular Dynamics).

RMSD: Desviación Cuadrática Media (Root-Mean Square Deviation).

RMSF: Fluctuación Cuadrática Media (Root-Mean Square Fluctuation).

SDS-PAGE: Electroforesis en Gel de Poliacrilamida con Dodecil Sulfato de Sodio (Sodium Dodecyl Sulfate – PolyAcrylamide Gel Electrophoresis).

T_m: Temperatura de desplegamiento (Melting Temperature).

VMD: Dinámica Molecular Visual (Visual Molecular Dynamics).

WHAM: Método de Análisis de Histogramas Ponderados (Weighted Histogram Analysis Method).

RESUMEN

Licanantasa es la proteína predominante en el secretoma de *Acidithiobacillus thiooxidans* cuando es cultivada con azufre elemental. Se ha descrito que Licanantasa es capaz de incrementar la recuperación de cobre desde calcopirita cuando se utiliza como aditivo para la biolixiviación. Sin embargo, el mecanismo por el cual se favorece la biolixiviación todavía no es entendido. Por este motivo es de nuestro interés lograr predecir la estructura de Licanantasa, lo que permitirá mejorar la comprensión sobre su función.

Mediante un análisis bioinformático se estudió la secuencia de aminoácidos de Licanantasa. Se encontró una gran similitud con la Lipoproteína de *Escherichia coli* (Lpp) que forma trímeros estables en solución, compartiendo motivos de secreción, procesamiento, estructura hélice alfa y distribución de residuos hidrofóbicos en heptadas que forman un núcleo hidrofóbico, organización clásica de hélices sobre-enrolladas. De esta manera se propuso que la estructura de Licanantasa más probable corresponde a una hélice sobre-enrollada trimérica que estaría estabilizada por interacciones no enlazantes inter-monómero, indiferentes a las condiciones de pH en la que ésta proteína ejerce su función.

En este trabajo se abordó el problema utilizando herramientas *in silico* para la predicción *de novo* de estructuras oligoméricas, simulaciones de Dinámica Molecular (MD) y exploración del espacio conformacional de proteínas mediante simulaciones de Dinámica Molecular con intercambio de réplicas (REMD).

La estructura propuesta se predijo utilizando el algoritmo “Fold and Dock” de Rosetta para la predicción de estructuras oligoméricas. El trímero presentó las características estructurales esperadas: una hélice sobre-enrollada, un dominio “alanine-zipper” y residuos en posición para formar interacciones electrostáticas inter-monómero.

Para estudiar la estabilidad estructural se realizaron simulaciones de MD y de REMD utilizando la estructura de Licanantasa, de Lpp y de la mutante con 14 alaninas de Lpp (Ala-14) que posee un dominio "alanine-zipper". Estas estructuras se simularon en sus estados de protonación a pH ácido y neutro. El proceso de desplegamiento fue similar para las tres proteínas y depende principalmente de las interacciones inter-monómeros, y de la conservación de la estructura secundaria. En las simulaciones, Licanantasa presentó el mayor número de contactos hidrofóbicos así como de puentes de hidrógeno que estabilizan la estructura secundaria. Esta mayor interfaz hidrofóbica se debe a su mayor extensión y a la presencia de residuos hidrofóbicos voluminosos en su extremo C-terminal. El mayor número de puentes de hidrógeno que estabilizan la estructura secundaria se debe a que posee una hélice alfa más extensa. De esta manera, Licanantasa presenta la mejor combinación de ambas características dentro de las proteínas estudiadas, requiriéndose una mayor cantidad de energía para producir el desplegamiento.

SUMMARY

Structure prediction and molecular simulation of lipoprotein Licanantase of strain Licanantay: *Acidithiobacillus thiooxidans* DSM 17318.

Licanantase was found as the major component of the secretome of *Acidithiobacillus thiooxidans* when grown in elemental sulfur. It has been described that Licanantase is able to improve copper recovery from bio-leaching when it is used as additive. However the mechanism by which the bio-leaching rate is enhanced is not yet understood. Because of this our interest is to predict the structure of Licanantase in order to improve the understanding of its role in the bio-leaching process.

A bioinformatics analysis was done using the amino acid sequence of Licanantase. Our findings show a great similarity with the Lipoprotein of *Escherichia coli* (Lpp) which can form stable trimers in solution, sharing secretion motif, processing and alpha helix structure, as well as the distribution of hydrophobic residues in heptads, forming a hydrophobic core, typical of coiled-coil structures. Thus it is proposed that the most probable structure for Licanantase is a trimeric coiled-coil, which would be stabilized by inter-monomer nonbonding interactions, insensitive to pH conditions in which this protein exerts its function.

In this work *in silico* tools were used for *de novo* prediction of oligomeric structures, Molecular Dynamics (MD) and for the exploration of conformational space by Replica Exchange Molecular Dynamics (REMD) simulations.

The proposed structure was predicted using the Rosetta Fold and Dock algorithm for oligomeric structures prediction. The trimer showed the expected structural features: a coiled-coil structure, an alanine-zipper domain and residues in position to form inter-monomer electrostatic interactions.

In order to study structural stability, MD and REMD simulations were performed for Licanantase, Lpp and 14-alanine Lpp mutant (Ala-14) which has an alanine-zipper

domain. These structures were simulated in its protonation states at acid and neutral pH. The unfolding process was similar for all proteins in study and it depends mainly of inter-monomer interactions and preservation of secondary structure. In simulations, Licanantase showed the greatest number of both hydrophobic contacts and hydrogen bonds that stabilize secondary structure. This greater hydrophobic interface is due to its greater extension and the presence of voluminous hydrophobic residues in its C-terminal end. The greater number of hydrogen bonds that stabilize secondary structure is due to a longer alpha helix structure. Thus, Licanantase shows the best combination of both features among studied proteins, requiring a greater amount of energy to produce unfolding.

1 INTRODUCCIÓN

1.1 Biolixiviación

La biolixiviación es un proceso de gran interés para la industria minera que consiste en la extracción de metales pesados desde minerales insolubles por medio de la oxidación biológica, como es el caso de la lixiviación bacteriana de cobre desde calcopirita (CuFeS_2). Su importancia radica en que es un proceso de bajo costo que permite ser aplicado a minerales de cobre de baja ley que no podrían ser extraídos por técnicas convencionales, lo que aumentaría las reservas de cobre económicamente explotables. Esta técnica también se ha empleado para otros metales como níquel, cobalto, zinc, que son extraídos desde sulfuros insolubles o, para el uranio, desde óxidos. También se utiliza el término de biooxidación, aunque preferentemente para la lixiviación de sulfuros metálicos presentes en el mineral que interfieren con la extracción de metales como oro y plata (Rohwerder et al., 2003).

En el proceso de biolixiviación es frecuente encontrar consorcios microbianos de especies extremófilas, pertenecientes a géneros como *Acidithiobacillus*, *Acidiphilium* y *Leptospirillum*, entre otros. Algunas de las características que poseen en común estos microorganismos son que crecen típicamente en ambientes ácidos (acidófilos); crecen autotróficamente mediante la fijación de CO_2 atmosférico; obtienen su energía por medio de hierro (II) o compuestos de azufre inorgánicos reducidos como dadores de electrones y utilizan generalmente oxígeno como aceptor de electrones; y poseen una destacable tolerancia a un amplio rango de iones metálicos (Rawlings, 2005).

Atención particular han recibido las especies *Acidithiobacillus ferrooxidans*, oxidante de hierro y azufre, junto con *Acidithiobacillus thiooxidans*, oxidante de azufre (Rawlings, 2005). Las bacterias biolixiviantes pueden actuar mediante dos mecanismos: de contacto y no contacto. El mecanismo de contacto se lleva a cabo en la interfaz generada entre las bacterias y la superficie mineral a la que se encuentran adheridas, mientras que el mecanismo sin contacto es llevado a cabo por las bacterias planctónicas que oxidan hierro (II) en solución. En el primer caso se propone que las

reacciones se llevan a cabo de manera más rápida y eficiente, gracias a la secreción de sustancias poliméricas extracelulares (SPE), que median la adherencia bacteriana a la superficie mineral y promueven el ataque oxidativo al concentrar hierro (III) por complejación (Gehrke et al., 1998; Sand y Gehrke, 2006). En el segundo caso, iones de hierro (III) y protones son los principales responsables de llevar a cabo las reacciones de lixiviación. El rol de los microorganismos es generar estos agentes oxidantes y crear el espacio adecuado para que las reacciones ocurran (Rawlings, 2005). Se han propuesto dos mecanismos para la disolución de sulfuros metálicos. El primero corresponde a la vía del tiosulfato para la oxidación de sulfuros metálicos insolubles en ácidos, como pirita, molibdenita y tungstenita (FeS_2 , MoS_2 y WS_2 , respectivamente). La solubilización es a través del ataque de iones de hierro (III), siendo el tiosulfato el principal producto intermediario y el sulfato el principal producto final. Microorganismos como *A. ferrooxidans* y *Leptospirillum ferrooxidans* son capaces de reoxidar los iones de hierro (II) producidos en esta vía a hierro (III). El segundo mecanismo corresponde a la vía de los polisulfuros, propuesto para sulfuros metálicos solubles en ácidos, como esfalerita, galena, arsenopirita, calcopirita y hauerita (ZnS , PbS , FeAsS , CuFeS_2 y MnS_2 , respectivamente). La solubilización es a través de un ataque combinado de iones de hierro (III) y protones, con azufre elemental como principal intermediario. El azufre elemental es relativamente estable, encontrándose como agregados libres y cristales, o formando una capa sobre sulfuros metálicos que modifica las propiedades electroquímicas de la superficie del mineral e inhibe del proceso de biolixiviación, sin embargo microorganismos como *A. thiooxidans* son capaces de oxidarlo a ácido sulfúrico, regenerando los protones y permitiendo que el proceso no sea inhibido (Rohwerder et al., 2003).

La oxidación de azufre elemental por bacterias acidófilas oxidantes de azufre es un proceso complejo que involucra principalmente la adhesión de células a partículas de azufre (mecanismo de contacto), transporte de azufre a través de la membrana externa y su oxidación en el espacio periplásmico (Zhang et al., 2008). Las SPE estarían encargadas de mediar el contacto para superar la hidrofobicidad de la superficie del azufre y asegurar la adsorción celular (Gehrke et al., 1998). Se ha observado que *A. thiooxidans* cultivada en azufre elemental presenta una vesiculación

(“blebbing”) de su membrana externa, compuesta por fosfolípidos y proteínas anfipáticas, que recubre las partículas de azufre elemental promoviendo su disolución a azufre coloidal (Knickerbocker et al., 2000).

Múltiples estudios de proteómica se han realizado bajo condiciones de crecimiento en azufre elemental para determinar cuáles son las principales proteínas que participan en su procesamiento. Se estudió la expresión de proteínas totales de *A. ferrooxidans* bajo condiciones de sulfuro metálico, tiosulfato, hierro (II) y de azufre elemental. Se observó que para esta última condición hubo aumento de una proteína de membrana externa de 44 kDa y una proteína hipotética de polisacárido capsular de 45 kDa, con roles propuestos para la adhesión al azufre elemental, además de observarse un aumento de proteínas posiblemente involucradas en el metabolismo del azufre (Ramírez et al., 2004). También se ha estudiado la fracción de proteínas extracelulares de *A. ferrooxidans* cultivada en azufre elemental y en sulfato ferroso, encontrándose proteínas como pilina, lipoproteínas, con posibles funciones en la adherencia celular, y otras proteínas hipotéticas, donde varias de estas últimas poseen motivos CXXC posiblemente involucradas en la activación del azufre mediante la unión de azufre a sus grupos tioles (Zhang et al., 2008).

1.2 Licanantasa

Recientemente, Bobadilla Fazzini et al. (2011) realizaron un estudio, en presencia de azufre elemental, del secretoma de *A. thiooxidans* y del metasecretoma de cultivos de *A. thiooxidans* con *A. ferrooxidans*. La proteína predominante en la fracción secretada por *A. thiooxidans* fue identificada como una lipoproteína a la que estos autores nombraron Licanantasa. Se realizaron ensayos de biolixiviación de calcopirita utilizando como aditivo un concentrado de proteínas secretadas, observándose un incremento significativo en la recuperación de cobre comparado con los controles, situación que se repitió al utilizar solamente una subfracción enriquecida en Licanantasa. En ninguno de los ensayos se observó un aumento en el crecimiento bacteriano. De esta manera se le atribuyó un rol principal a Licanantasa en el aumento de la biolixiviación de calcopirita, sin poder proponer un mecanismo claro de acción,

pero sugiriendo una posible participación como agente surfactante, ayudando a la superación de la barrera hidrofóbica de azufre elemental que se forma durante la biolixiviación de calcopirita. A partir de péptidos secuenciados de Licanantasa, los autores realizaron un análisis genético sobre una anotación parcial del genoma de *A. thiooxidans* DSM 17318 y encontraron un gen que codifica para una proteína que posee un 96% de identidad de secuencia de aminoácidos (aa) con una lipoproteína hipotética de *A. ferrooxidans* ATCC 53993 (Bobadilla Fazzini et al., 2011).

1.3 Análisis basados en secuencias

Para complementar esta información, procedimos a realizar una búsqueda con la secuencia de aa de Licanantasa mediante BLASTP (Altschul et al., 1997, 2005), encontrando coincidencias con alta identidad de secuencia correspondientes a lipoproteínas hipotéticas de *A. ferrooxidans* ATCC 53993 (96%), *Acidithiobacillus caldus* ATCC 51756 (95%) y *Acidithiobacillus ferrivorans* SS3 (93%). Estimamos el pI y el contenido de aa mediante PROTPARAM (Gasteiger et al., 2005), en todos los casos corresponden a proteínas con un pI básico debido principalmente a un alto contenido de lisina (~11%), además se puede observar la conservación de residuos de alanina que alcanzan un 25% del contenido total de aminoácidos (**Figura 1**).

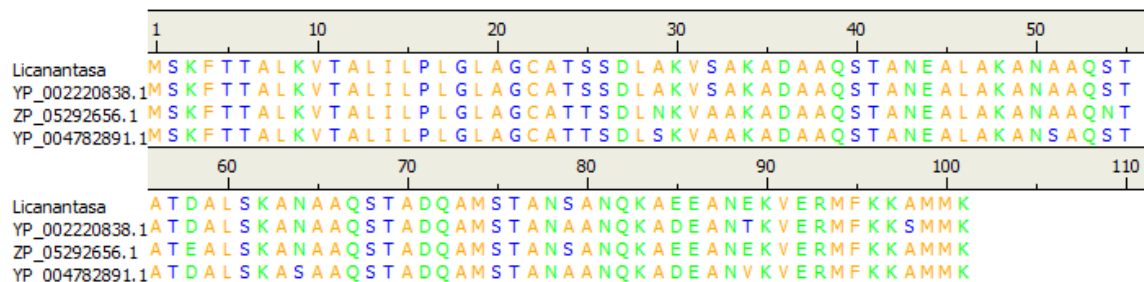


Figura 1. Alineamiento de secuencia de aminoácidos de Licanantasa con lipoproteínas hipotéticas.

Licanantasa de *A. thiooxidans* DSM 17318, YP_002220838.1 de *A. ferrooxidans* ATCC 53993 (96%), ZP_05292656.1 de *A. caldus* ATCC 51756 (95%) y YP_004782891.1 de *A. ferrivorans* SS3 (93%). Naranja: residuos hidrofóbicos; Azul: residuos neutros; Verde: residuos hidrofílicos.

Realizamos una búsqueda mediante FASTA (Pearson, 1990) contra la base de datos Protein Data Bank (PDB) encontrando coincidencias con la estructura 1JCD (**Figura 2**) y en menor medida con 1JCC y 1KFN, con identidades globales de

secuencia de 28.7 %, 24.2 % y 20.9 %, respectivamente. Las proteínas mencionadas corresponden a mutantes de Lipoproteína (Lpp) de *Escherichia coli* (PDB ID: 1EQ7) (Shu et al., 2000; Liu et al., 2002; Liu y Lu, 2002; Liu et al., 2003).

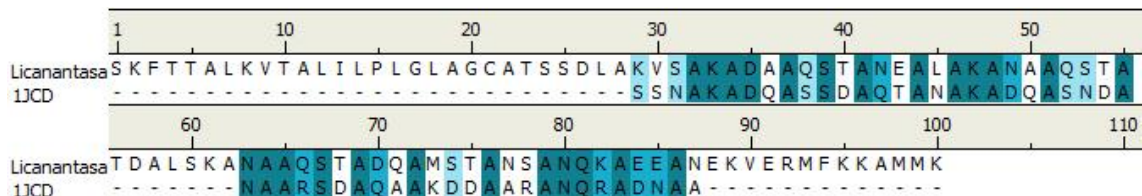


Figura 2. Alineamiento de secuencia de aminoácidos de Licanantasa con la secuencia de la mutante Ala-14 de Lpp (PDB ID: 1JCD).

Identidad de secuencia 28.7%, dada principalmente por las alaninas mutadas.

Analizamos la secuencia de Licanantasa mediante ScanProsite (de Castro et al., 2006) detectando un motivo conservado para el procesamiento de lipoproteínas, donde el primer segmento de 21 aa de Licanantasa correspondería a un péptido señal que sería cortado, dejando expuesta la cisteína 22 que puede ser modificada uniéndose covalentemente a una molécula de diacilglicerol mediante su grupo sulfhidrilo y a un palmitoílo mediante su grupo amino terminal, procesamiento clásico de lipoproteínas que coincide con el de Lpp (Tokuda, 2009). Obtuvimos predicciones congruentes con este procesamiento al utilizar las herramientas bioinformáticas LipoP y Signal Blast (Rahman et al., 2008; Frank y Sippl, 2008). Al analizar el perfil de hidrofobicidad mediante los servidores TMHMM, TMPRED y TOPPRED (Sonnhammer et al., 1998; Hofmann y Stoffel, 1993; von Heijne, 1992; Claros y von Heijne, 1994) no detectamos posibles dominios transmembrana, pero sí observamos una mayor hidropatía en el primer segmento de la proteína, coherente con su probable función de péptido señal. Realizamos la predicción de estructura secundaria mediante los servidores Jpred, NPS@, PCI-SS, Porter y SCRATCH (Cole et al., 2008; Combet et al., 2000; Green et al., 2009; Pollastri y McLysaght, 2005; Cheng et al., 2005), observando una estructura predominante tipo hélice alfa, constituida por dos segmentos helicoidales: el primero correspondería al péptido señal y el segundo se ubicaría aproximadamente desde la serina 26 hasta la metionina 99. Mediante el programa MultiCoil detectamos una alta probabilidad de que este segundo segmento hélice alfa forme hélices sobre-enrolladas triméricas (Wolf et al., 1997). Un resumen de estas predicciones se encuentra en la **Figura 3**.

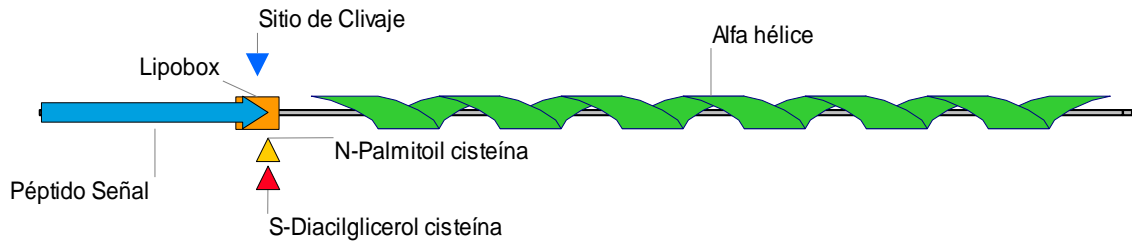


Figura 3. Predicciones de dominios a partir de la secuencia de Licanantasa.

Flecha celeste: Secuencia péptido señal residuos 1-22. Caja naranja: Motivo "lipobox" residuos 19-22. Triángulo azul: Sitio de clivaje residuos 21|22. Triángulo amarillo: Modificación N-palmitoil cisteína residuo 22. Triángulo rojo: Modificación S-diacilglicerol cisteína residuo 22. Hélice verde: Estructura secundaria tipo hélice alfa residuos 26-99.

1.4 Lipoproteína de Braun

La lipoproteína Lpp, también conocida como lipoproteína de Braun, es una de las proteínas más abundantes de *E. coli*. Su precursor de 78 aa posee un dominio péptido señal que le dirige a la membrana interna mediante el sistema Sec. Un motivo "lipobox" [L(A/S)(G/A)C] determina su corte en una proteína madura de 58 aa, dejando disponible una cisteína en el extremo N-terminal para su modificación y anclaje a la membrana. Luego, Lpp es translocada hacia la membrana externa por el sistema Lol (**Figura 4**). Su estructura secundaria es predominantemente hélice alfa y forma trímeros estables en solución, encontrándose en dos formas: libre o unida a péptidoglicano mediante su lisina final, en una proporción 2:1 (unida:libre). Lpp une la membrana externa al péptidoglicano teniendo un papel importante en la arquitectura y función de ésta. Se ha observado que mutantes de *E. coli* que carecen de esta lipoproteína forman vesículas de manera espontánea en su superficie y pierden enzimas periplásmicas que expulsan al medio de crecimiento (Hirota et al., 1977; Shu et al., 2000; Tokuda, 2009).

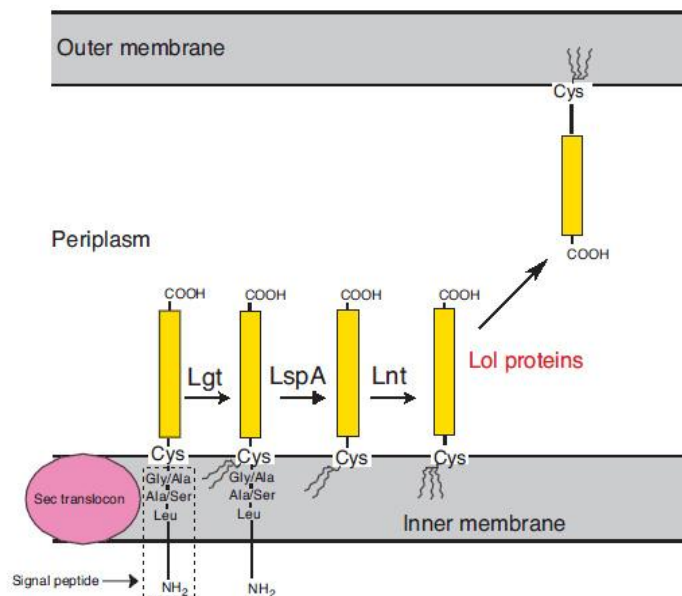


Figura 4. Procesamiento de lipoproteínas.

Las lipoproteínas son sintetizadas como precursores con un péptido señal N-terminal (caja punteada) que contiene un motivo llamado “lipobox”, [L(A/S)(G/A)C], que son luego traslocadas a través de la membrana interna por el sistema Sec. Las enzimas conservadas Lgt, LspA y Lnt se encargan de procesar los precursores de lipoproteínas a sus formas maduras. Luego, el sistema Lol transloca a la membrana externa las lipoproteínas que no posean un aspartato en posición 2, como es el caso de Lpp y de Licanantasa. Figura adaptada de Tokuda 2009.

1.5 Estructura de Lipoproteína de Braun y sus mutantes

La estructura 1EQ7 descrita por Shu et al. (2000) fue obtenida a partir de una proteína recombinante a la que llamaron Lpp-56, que corresponde a Lpp madura sin su cisteína 1 ni su lisina 58. Lpp-56 posee una estructura de tipo hélice sobre-enrollada, constituida por tres monómeros orientados de manera paralela (**Figura 5**). Una característica destacable de este tipo de estructuras es la organización de cada monómero en heptadas, secuencias de 7 aa (*a b c d e f g*) que corresponden a ~2 vueltas de hélice y que definen la orientación de los residuos dentro de la hélice alfa. Es frecuente encontrar residuos hidrofóbicos en las posiciones *a* y *d* orientados hacia el núcleo de la proteína, como es el caso de 1EQ7 (**Figura 6.a**), en la que además se observa un pequeño dominio “alanine-zipper” formado por 3 alaninas (Ala-3). El mismo grupo de investigación evaluó la estructura y estabilidad de mutantes en las que se reemplazaron los residuos hidrofóbicos voluminosos de las

posiciones *a* y *d* por un número creciente de alaninas, llamándolas Ala-5 (PDB ID: 1KFM), Ala-7 (PDB ID: 1KFN), Ala-10 (PDB ID: 1JCC) y Ala-14 (PDB ID: 1JCD), conteniendo esta última sólo alaninas en su núcleo hidrofóbico. A medida que se aumenta el contenido de alaninas las estructuras pierden estabilidad (Ala-5 y Ala-7), lo que se traduce en una disminución considerable en su temperatura de desplegamiento (T_m) hasta que finalmente ya no forman estructuras estables en solución (Ala-10 y Ala-14) y sólo son capaces de constituir trímeros durante la cristalización gracias a la estabilización por interacciones inter-trímeros. La pérdida de estabilidad puede ser explicada por la disminución de interacciones tipo Van der Waals en la interfaz hidrofóbica inter-monómero (Shu et al., 2000; Liu et al., 2002; Liu y Lu, 2002; Liu et al., 2003).

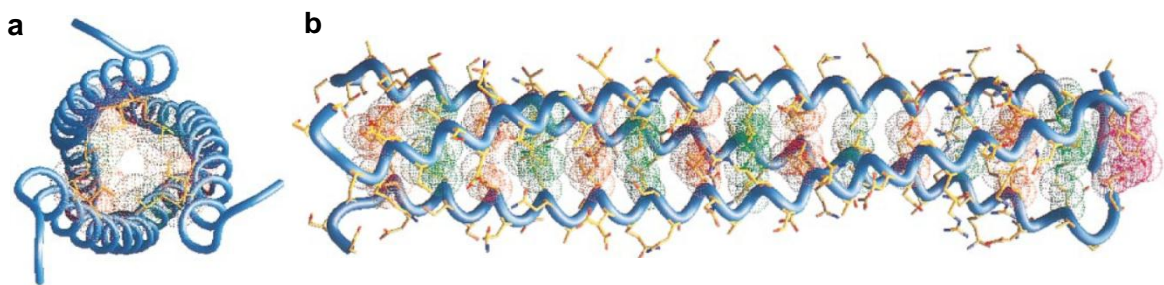


Figura 5. Representación de la estructura de Lpp-56.

(a) Vista axial del trímero Lpp-56, con el extremo amino terminal apuntando hacia el lector. (b) Vista lateral del trímero Lpp-56, con el extremo amino terminal hacia la izquierda. Superficies de van der Waals de los residuos en posiciones *a* y *d* en rojo y verde, respectivamente. Figura adaptada de Shu et al., 2000.

Los puentes de hidrógeno y los puentes salinos también participan en la mantención de la estructura de Lpp-56. Las posiciones *e* y *g* de cada hélice suelen ser ocupadas por residuos polares o cargados de forma complementaria de manera que pueden constituir un sistema de interacciones inter-hélice que estabilizan el trímero (Shu et al., 2000).

Se han realizado estudios de plegamiento por dicroísmo circular, calorimetría y agentes desnaturantes revelando que Lpp-56 posee una alta estabilidad y un proceso muy lento de plegamiento/desplegamiento (Dragan et al., 2004; Bjelić et al., 2006). También se ha estudiado el efecto de la disminución de pH en la estabilidad de Lpp-56 encontrando una disminución de la T_m y del calor necesario para desplegar la

proteína, fenómeno atribuido a la protonación de residuos ácidos que ocasiona una pérdida de su capacidad de formar puentes salinos con residuos básicos. La participación de las interacciones electrostáticas fue corroborada *in silico* mediante simulaciones de dinámica molecular de Lpp-56, observándose una alta prevalencia de este tipo de interacciones a lo largo de la simulación (Bjelić et al., 2008).

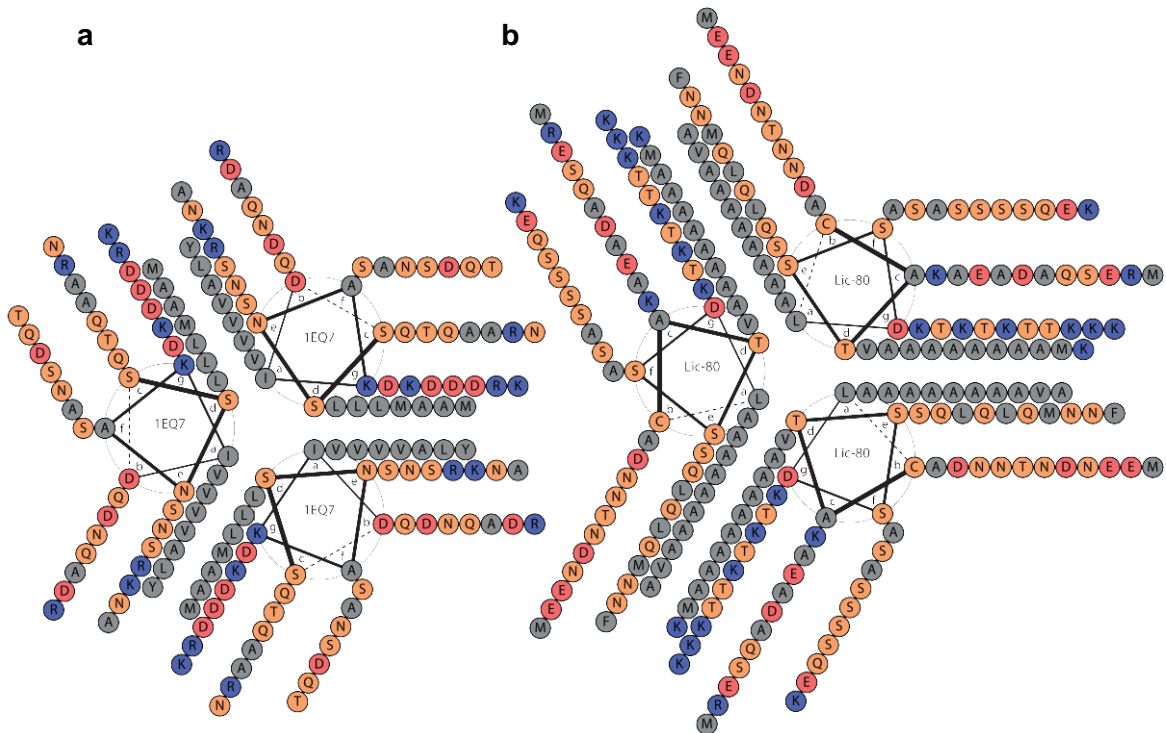


Figura 6. Representación tipo “helical wheel”.

(a) Representación de Lpp residuos 22-77 presentes en el PDB 1EQ7. (b) Representación de Licanantasa (Lic-80) residuos 22-101. Gris: residuos hidrofóbicos; Naranja: residuos polares; Rojo: residuos ácidos; Azul: residuos básicos.

1.6 Estructura de Licanantasa

Al hacer un estudio de los residuos 22-101 de Licanantasa (Lic-80) observamos la presencia de 11 heptadas consecutivas, donde se alinean 21 residuos hidrofóbicos en las posiciones *a* y *d*, siendo 17 de ellos alaninas (**Figura 6.b**). Esta disposición de residuos nos permite proponer una estructura de tipo “alanine-zipper” similar a la descrita para la mutante Ala-14. Sin embargo, la disminución en la estabilidad de esta

mutante, ocasionada por la sustitución de residuos hidrofóbicos más voluminosos por alaninas que son de pequeño tamaño, se podría ver compensada en el caso de Lic-80 por su mayor extensión y la presencia de residuos hidrofóbicos más voluminosos que corresponden a cuatro metioninas y una fenilalanina por monómero, ubicadas al extremo C-terminal (**Figura 6.b**). De esta manera, Lic-80 poseería una mayor superficie de contacto inter-monómero y, por lo tanto, también las interacciones de tipo Van der Waals. Este tipo de residuos más voluminosos sí están presentes en Lpp-58, que posee dos metioninas y una tirosina por monómero (**Figura 6.a**). Por otra parte, se ha postulado que el efecto de Licanantasa en la biolixiviación se debe a que actuaría como surfactante (Bobadilla Fazzini et al., 2011). En este caso, la presencia de los residuos hidrofóbicos mencionados anteriormente podría estar mejorando su capacidad de solubilizar el azufre elemental.

Lic-80 posee residuos básicos en la posición *g* que podrían establecer interacciones con los residuos ácidos ubicados en la posición *b* de la hélice contigua (**Figura 6.b**). También se ha descrito que aa polares y cargados cumplen un rol importante en el proceso de oligomerización, favoreciendo el alineamiento y la direccionalidad adecuada de los monómeros (Bjelić et al., 2006).

Se ha observado que proteínas periplásmicas de *A. ferrooxidans* son predominantemente básicas en comparación con organismos neutrofílicos como *E. coli*, característica que los autores sugieren como una adaptación al medio ácido del periplasma, siendo la mayoría de estas proteínas tolerantes o estables en ácido (Chi et al., 2007). De esta manera, es posible que la diferencia de pI de Lpp con Licanantasa dé cuenta de una adaptación específica al ambiente de cada organismo y que sea capaz de proveerle a Licanantasa una mayor estabilidad a pH ácido, a diferencia de Lpp-56 que pierde estabilidad cuando se encuentra a dicho pH.

El alto contenido de hélice alfa que se predice para Licanantasa, la presencia de residuos hidrofóbicos en heptadas, en conjunto con la evidencia estructural de 1EQ7 y sus mutantes, nos permite proponer que la estructura más probable de Licanantasa en solución es una hélice sobre-enrollada trimérica.

La presente tesis aborda este problema utilizando herramientas *in silico* para la predicción de estructuras oligoméricas mediante el software Rosetta “Fold and Dock” (Das et al., 2009), simulaciones de Dinámica Molecular usando el paquete de simulación GROMACS (van der Spoel et al., 2005) y exploración del espacio conformacional de plegamiento/desplegamiento de proteínas mediante la técnica de simulación “Replica Exchange” (Sugita y Okamoto, 1999).

2 HIPÓTESIS

"La lipoproteína Licanantasa posee una estructura trimérica tipo hélice sobre-enrollada cuya arquitectura macromolecular estaría estabilizada principalmente por interacciones no enlazantes inter-monómero, indiferentes a las condiciones de pH en la que ésta proteína ejerce su función."

3 OBJETIVO GENERAL

Predecir la estructura trimérica de Licanantasa mediante modelamiento *de novo*, y determinar mediante simulaciones de dinámica molecular cuáles de sus propiedades estructurales son importantes para la mantención de su arquitectura macromolecular.

4 OBJETIVOS ESPECÍFICOS

1. Determinar si la estructura macromolecular predicha para Licanantasa mediante el protocolo de modelamiento *de novo*, corresponde a una hélice sobre-enrollada trimérica.
2. Estudiar la estabilidad de las interacciones no enlazantes que serían responsables de mantener la arquitectura macromolecular de Licanantasa, en condiciones de pH neutro y en pH ácido, mediante el protocolo de Replica Exchange.

5 METODOLOGÍA

5.1 Modelamiento *de novo*:

Se utilizó el protocolo “Fold and Dock” del software Rosetta para estructuras triméricas con simetría cíclica (3C) (Das et al., 2009). Se generó la biblioteca de fragmentos para la secuencia de Lic-80 mediante el servidor Robetta (Kim et al., 2004). Se generaron 100.000 modelos de los cuales se seleccionaron los 400 con mejor puntaje de predicción, utilizando como criterio el menor valor de energía. Se procedió a realizar un agrupamiento estructural mediante la aplicación de “clustering” del programa Rosetta, utilizando como criterio la desviación cuadrática media (RMSD) de sus carbonos alfa (Shortle et al., 1998). El modelo final elegido corresponde a la estructura central del grupo con mayor número de miembros. La estructura final se estudió mediante el programa VMD (Humphrey et al., 1996).

5.2 Construcción de Lpp-58:

Mediante el programa MODELLER (Eswar et al., 2007) se agregaron los residuos faltantes al cristal de Lpp-56 (PDB ID: 1EQ7) correspondientes a la cisteína N-terminal y la lisina C-terminal.

5.3 Construcción y preparación de sistemas de simulación:

Con el fin de poder estimar la influencia del pH en la estabilidad de Lic-80, Lpp-58 y Ala-14 se procedió a generar dos nuevos modelos estructurales para cada proteína mediante el servidor web PDB2PQR (Dolinsky et al., 2004), de forma que el estado de protonación de sus residuos correspondan a los encontrados en un ambiente ácido (pH 1,6) o neutro (pH 7,4). Las simulaciones de Dinámica Molecular (MD) se realizaron utilizando el campo de fuerza CHARMM implementado en GROMACS (Bjellmar et al., 2010). Se utilizó el modelo OBC para la aproximación de solvente implícito Generalized Born / Solvent Accessibility (GB/SA) (Onufriev, Bashford y Case, 2004; Larsson y Lindahl, 2010). Cada una de las estructuras se minimizó utilizando el

algoritmo Steepest Descent con una tolerancia de $100 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. A continuación se simularon durante 500 ps utilizando una restricción armónica para los carbonos alfa con una fuerza constante de $10 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Este tipo de protocolo es ampliamente utilizado en MD ya que permite minimizar la energía inicial de los sistemas y llevarlos a un estado cercano al equilibrio de manera suave para evitar artefactos generados por estructuras iniciales inestables. Finalmente, cada estructura fue simulada durante 100 ns sin restricciones energéticas. El paso de integración utilizado fue 2 fs. Se utilizó un integrador dinámico estocástico con un coeficiente de fricción de 0,01 ps. Los componentes energéticos de Coulomb y Lennard Jones fueron calculados utilizando una distancia de corte de 1,3 nm. La temperatura se mantuvo controlada a 303.15 K usando un algoritmo de reescalamiento de velocidades con un término estocástico. Los enlaces covalentes fueron constreñidos usando el algoritmo LINCS (Hess, 2008). Las trayectorias fueron almacenadas con una separación temporal de 1 ps entre cada cuadro.

5.4 Simulación de Dinámica Molecular con Intercambio de Réplicas:

A partir de los sistemas finales obtenidos en el paso anterior se realizó simulación molecular utilizando el protocolo de intercambio de réplicas (REMD) de GROMACS (Seibert et al., 2005). Se simularon 64 réplicas por cada sistema para un rango de temperaturas entre 303,15 K y 809,57 K. La distribución de temperaturas se calculó utilizando el servidor web “Temperature generator for REMD-simulations” (Patriksson y van der Spoel, 2008) de manera de obtener una probabilidad de intercambio de $\sim 0,5$. Se simuló durante 100 ns cada réplica alcanzando un tiempo combinado total de $6,4 \mu\text{s}$ para cada sistema. El algoritmo intentó intercambiar las réplicas cada 1 ps. El resto de los parámetros de simulación fueron los mismos utilizados en la etapa anterior.

5.5 Análisis:

Para el estudio de las simulaciones se utilizaron herramientas de análisis disponibles en VMD y GROMACS. El cálculo de la RMSD de carbonos alfa se realizó

comparando cada cuadro de la trayectoria contra el cuadro inicial. La fluctuación cuadrática media (RMSF) de carbonos alfa se calculó para cada residuo, promediando los resultados de los tres monómeros en cada proteína.

Los puentes de hidrógenos se calcularon utilizando un punto de corte de 3,5 Å para la distancia entre los átomos dador y aceptor, y un punto de corte de 30° para el ángulo. Los puentes salinos se calcularon utilizando una distancia de corte de 4,0 Å entre los átomos dador y aceptor. Los contactos hidrofóbicos se calcularon utilizando una distancia de corte de 7,0 Å entre los centros de masa de las cadenas laterales de los residuos hidrofóbicos. Para los puentes salinos y contactos hidrofóbicos se contabilizaron sólo las interacciones inter-monómeros. Para los puentes de hidrógeno se calcularon por separado las interacciones inter-monómeros y las interacciones que se forman en las cadenas principales. Para las simulaciones preparativas se informaron los valores totales de puentes de hidrógeno, puentes salinos y contactos hidrofóbicos. Para las simulaciones de REMD se informaron los porcentajes respecto al valor máximo alcanzado para cada proteína a lo largo de todas las temperaturas.

El contenido de hélice alfa se calculó como el porcentaje de residuos que fueron identificados como hélice alfa mediante el programa STRIDE implementado en VMD (Frishman y Argos, 1995).

Las temperaturas de desplegamiento (T_m) corresponden al valor mínimo de la derivada de la propiedad observada. Para los puentes de hidrógeno inter-monómeros y los puentes salinos se calcularon los puntos medios como una aproximación a la temperatura de desplegamiento, ya que los datos no presentaron un comportamiento sigmoidal.

Para graficar el proceso de desplegamiento se calculó la probabilidad de encontrar cada estructura utilizando como ejes de coordenadas los contactos hidrofóbicos y contenido de hélice alfa. Este cálculo se realizó para cada temperatura y la información se combinó mediante el método de análisis de histogramas ponderados

(WHAM), de manera de obtener un solo gráfico que represente la probabilidad de encontrar una estructura a 303,15 K (Sindhikara, 2011).

Los cálculos se realizaron sobre datos extraídos cada 10 ps de simulación, a lo largo de la trayectoria completa, en el caso de las simulaciones preparativas, o de los últimos 50 ns, en el caso de las simulaciones de REMD. Los gráficos se construyeron con el programa de estadística R (R Development Core Team, 2011).

6 RESULTADOS Y DISCUSIÓN

6.1 Modelamiento *de novo* de la estructura trimérica de Licanantasa:

Se realizó un agrupamiento estructural utilizando una RMSD de 3,0 Å a los 400 modelos de menor energía obteniéndose un grupo principal de 71 estructuras. El modelo de Lic-80 seleccionado corresponde al centro de este grupo (**Figura 7**). La variabilidad con respecto a los otros miembros del grupo se encuentra principalmente hacia los extremos amino y carboxilo. El modelo predicho presentó la estructura tipo hélice sobre-enrollada esperada, orientando los residuos hidrofóbicos hacia el interior, formando un dominio “alanine-zipper” (**Figura 8.a**) similar al formado por Ala-14 (**Figura 8.e**). Se observa en la estructura la presencia de residuos polares y cargados en posición adecuada para la formación de interacciones electrostáticas inter-monómero (**Figura 8.b**), fenómeno que se evaluó durante las simulaciones. También se puede observar el mayor tamaño de Lic-80 (**Figura 8.b**) con respecto a Lpp-58 y Ala-14 (**Figura 8.d y 8.f**).

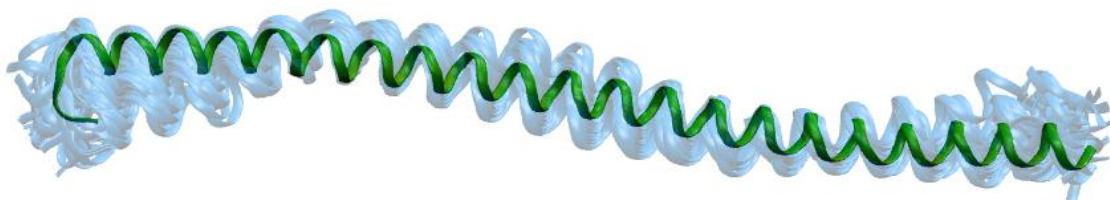


Figura 7. Grupo principal de estructuras del monómero de Lic-80 predicho mediante modelamiento *de novo*.

En verde se muestra el monómero de la estructura central del grupo. En celeste se muestran los monómeros de las 70 estructuras restantes del grupo principal. Extremos amino y carboxilo orientados hacia la izquierda y derecha, respectivamente.

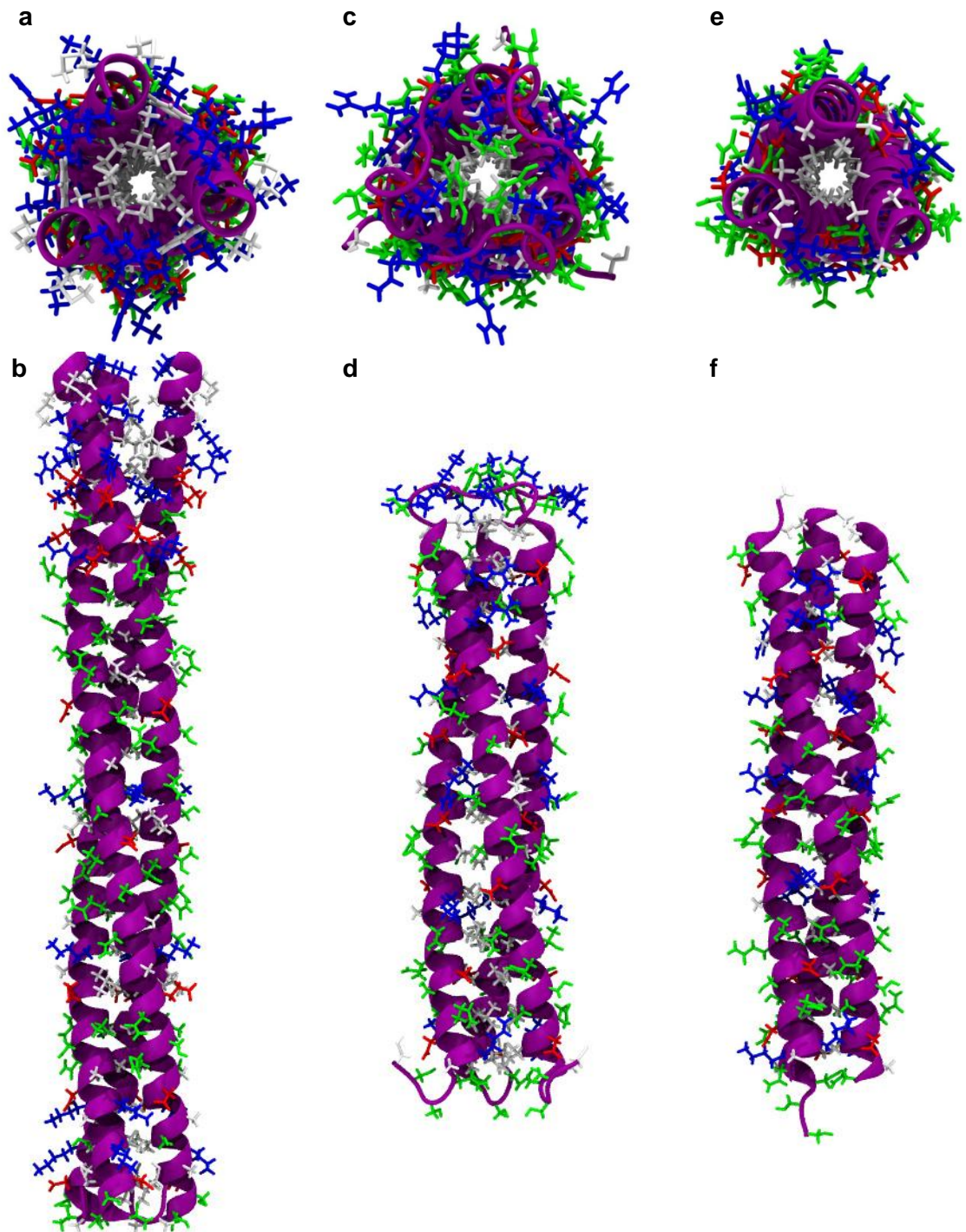


Figura 8. Estructuras triméricas de Lic-80, Lpp-58 y Ala-14.

Vista superior (a), (c) y (e) orientan sus extremos carboxilo hacia el lector. Vista lateral (b), (d) y (f) orientan sus extremos carboxilo hacia la parte superior de la página.

Lic-80 (a) y (b) se obtuvo a partir de modelamiento *de novo*. Lpp-58 (c) y (d) se obtuvo a partir del cristal PDB ID: 1EQ7 adicionándole los residuos faltantes. Ala-14 (e) y (f) corresponde al cristal PDB ID: 1JCD. Verde: residuos polares; Blanco: residuos hidrofóbicos; Azul: residuos básicos; Rojo: residuos ácidos.

El grado de oligomerización propuesto en este trabajo fue corroborado por la empresa BioSigma S.A. mediante experimentos llevados a cabo en Licanantasa expresada en forma heteróloga en *E. coli*. Para esto, realizaron ensayos de entrecruzamiento químico y observaron la presencia de trímeros de Licanantasa cuando se trataron las células con agentes entrecruzantes (**Figura 9**). Resulta importante destacar que los patrones de bandas que obtuvieron se asemejan a los previamente descritos para Lpp cuando se trata con agentes entrecruzantes (Choi et al., 1986). El protocolo utilizado de Rosetta nos permite entonces explorar las distintas conformaciones que puede poseer el trímero de Licanantasa, y dentro de las estructuras triméricas de menor energía, la que posee mayor probabilidad es la que forma hélices sobre-enrolladas.

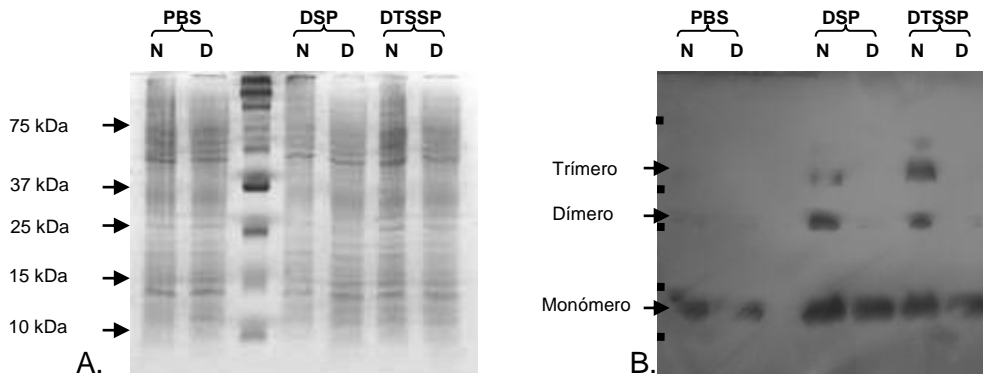


Figura 9. Entrecruzamiento químico de Licanantasa en *E. coli*.

Las células fueron tratadas con dos tipos de agentes entrecruzantes: DSP, reactivo hidrofóbico que puede atravesar membranas e ingresar a la célula y DTSSP, reactivo hidrofílico incapaz de cruzar membranas celulares. Las células fueron analizadas en gel SDS-PAGE 15% (A) y Western Blot His-probe (B). En la imagen se indica el entrecruzante utilizado para cada muestra y su estado de carga; N: muestra nativa; D: muestra desnaturalizada mediante β -mercaptoetanol. (Figura facilitada por BioSigma S.A., obtenida por Patricio Martínez y Gloria Garín.)

6.2 Simulación de Dinámica Molecular preparativa:

Se cambió el grado de protonación de las estructuras de Lic-80, Lpp-58 y Ala-14 a condiciones de pH 1,6 y pH 7,4, obteniéndose 2 estructuras para cada proteína. Con el fin de evitar artefactos durante la simulación, primero se minimizaron las estructuras y luego se relajaron durante 500 ps con una restricción armónica en sus carbonos alfa. A continuación las estructuras fueron simuladas durante 100 ns a 303,15 K. La RMSD de carbonos alfa se estabilizó rápidamente a valores alrededor de 2,0 Å para Lic-80 y Ala-14, y alrededor de 3 Å para Lpp-58, indicando que las estructuras son estables en las condiciones de simulación (**Figura 10**). Se observaron valores mayores de RMSD para Lpp-58 cuando fue simulada a pH ácido, mientras que los valores de RMSD para Lic-80 y Ala-14 no presentaron cambios debido al estado de protonación.

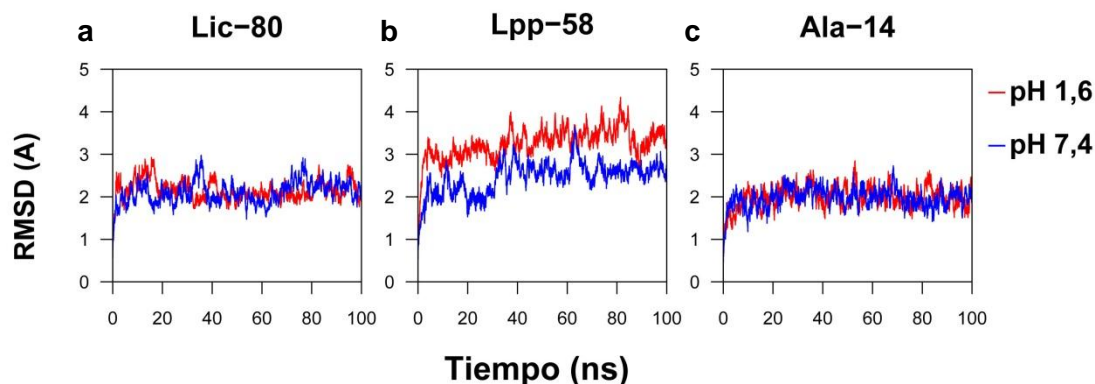


Figura 10. RMSD de carbonos alfa en las simulaciones preparativas. Lic-80 (a), Lpp-58 (b) y Ala-14 (c) en pH ácido (línea roja) y pH neutro (línea azul).

La RMSF indica que la mayor variabilidad estructural se encuentra en los extremos de los monómeros (**Figura 11**), característica ya observada en el modelamiento *de novo*. En el caso de Lpp-58 la RMSF se ve aumentada alrededor del residuo 31 a pH ácido. En esta zona a pH neutro la arginina 32 forma puentes de hidrogeno con la asparagina 29 de su mismo monómero y con el aspartato 27 del monómero contiguo (**Figura 12**). A pH ácido el aspartato 27 se encuentra protonado y pierde su interacción con la arginina 32, lo que brinda mayor flexibilidad al trímero, permitiendo que residuos voluminosos como valina 28, metionina 31 y valina 35 se muevan y acomoden al

interior del trímero. Esta mayor flexibilidad explicaría por qué la RMSD aumentaba en la simulación de Lpp-58 a pH ácido. Este fenómeno no se observa en Ala-14 ni en Lic-80, probablemente debido a que sus núcleos hidrofóbicos están constituidos principalmente por alaninas, las cuales son capaces de empaquetarse ordenadamente, lo que evitaría que se produzcan acomodaciones de residuos hidrofóbicos.

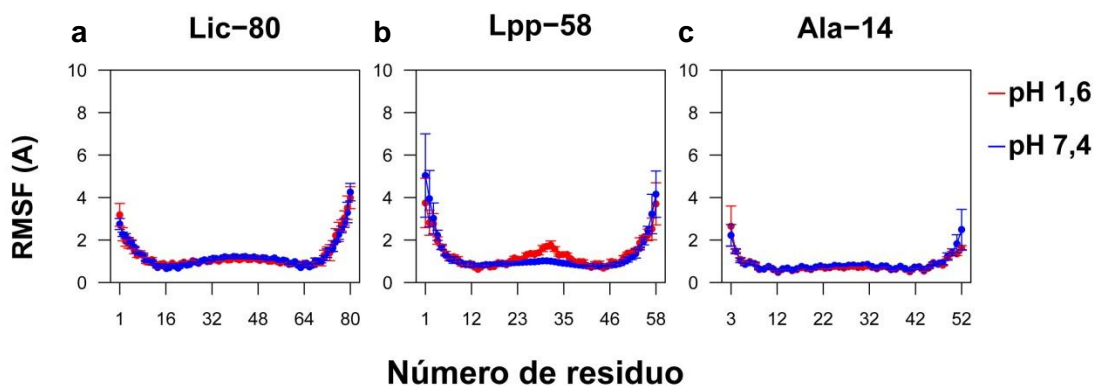


Figura 11. RMSF de carbonos alfa en las simulaciones preparativas. Lic-80 (a), Lpp-58 (b) y Ala-14 (c) en pH ácido (línea roja) y pH neutro (línea azul).

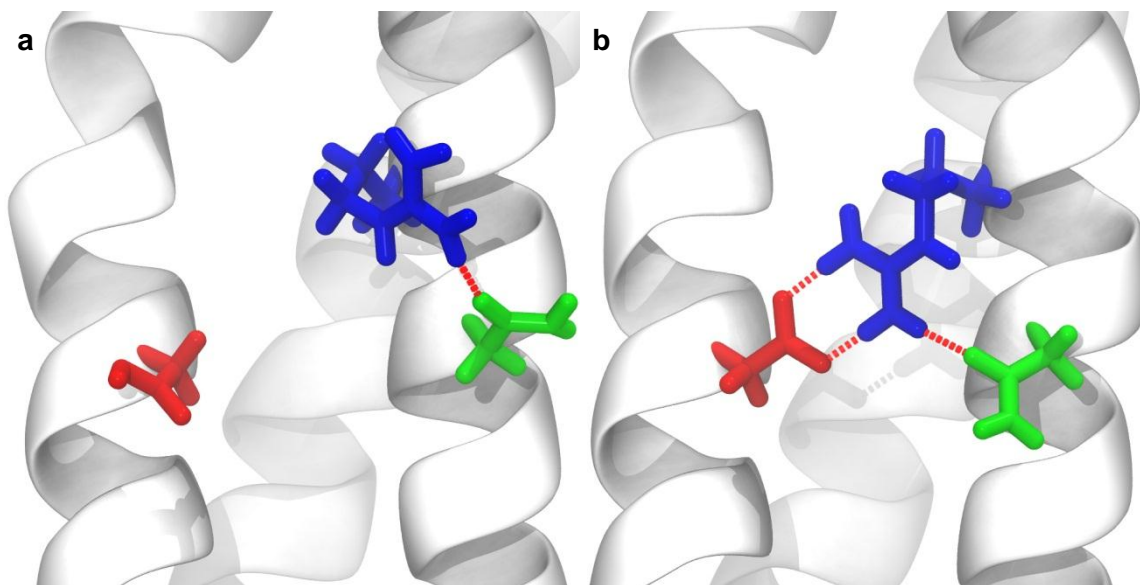


Figura 12. Pérdida de puentes de hidrógeno inter-monómero en Lpp-58. Estructura de Lpp-58 a pH ácido (a) y neutro (b). A pH ácido la arginina 32 (azul) no es capaz de interactuar con el aspartato 27 (rojo) ya que se encuentra protonado y solo interactúa con la asparagina 29 (verde) de su mismo monómero. A pH neutro la arginina 32 interactúa tanto con la asparagina 29, como con el aspartato 27 del monómero contiguo.

Como se esperaba, a pH neutro se observó un mayor número de puentes de hidrógenos inter-monómeros y puentes salinos que a pH ácido (**Figura 13**). Lic-80 a pH neutro presentó un menor número de este tipo de interacciones en comparación con Lpp-58 y Ala-14. Así, la pérdida de interacciones electrostáticas inter-monómeros que sufre Licanantasa a pH ácido es menor que la sufrida por Lpp-58 y Ala-14, por lo que el impacto del pH ácido sobre la estabilidad de Licanantasa también debería ser menor.

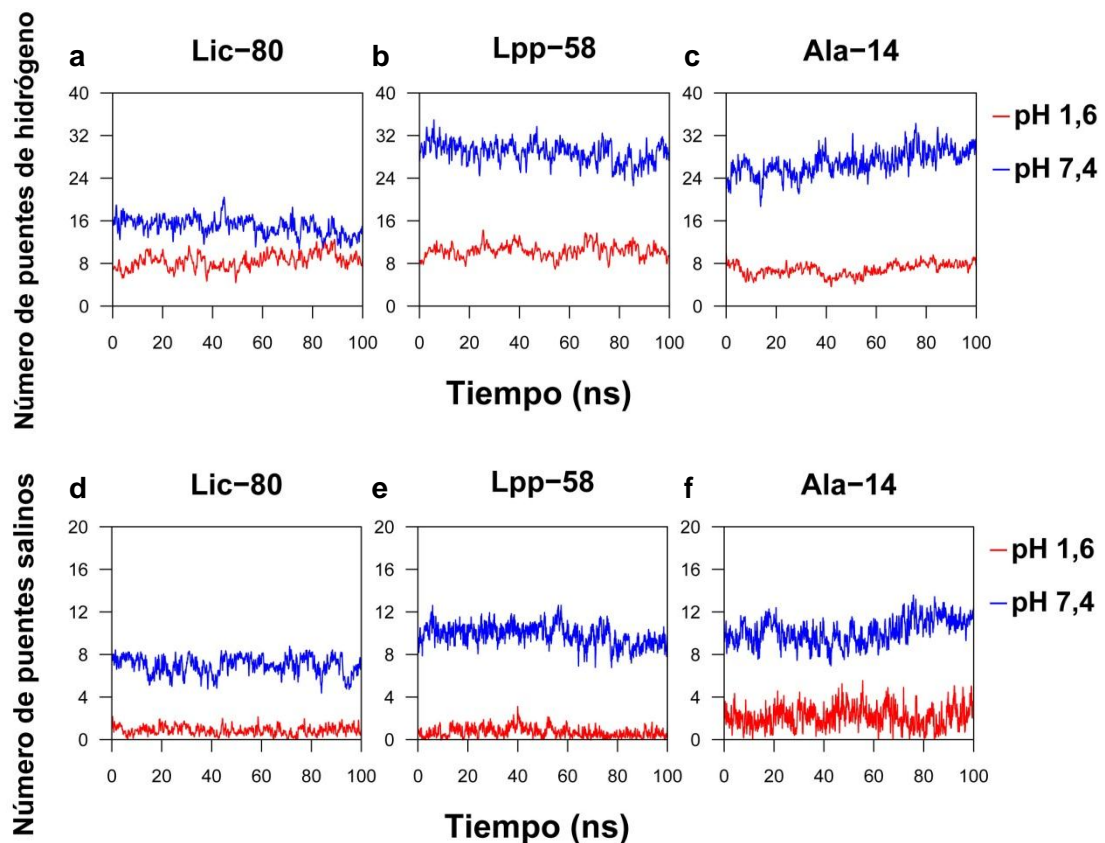


Figura 13. Puentes de hidrógeno y puentes salinos inter-monómeros en las simulaciones preparativas.

Lic-80 (a y d), Lpp-58 (b y e) y Ala-14 (c y f) en pH ácido (línea roja) y pH neutro (línea azul).

El contenido hélice alfa aumenta levemente para Lpp-58 en ambas condiciones hasta valores de ~90%, mientras que Lic-80 y Ala-14 se mantienen sobre un 90% de hélice alfa durante toda la simulación (**Figura 14**). Licanantasa presenta el mayor

número de puentes de hidrógeno que participan en la estabilización de la hélice alfa, esto se debe a que Licanantasa es una proteína más extensa que Lpp-58 y Ala-14 (Figura 15).

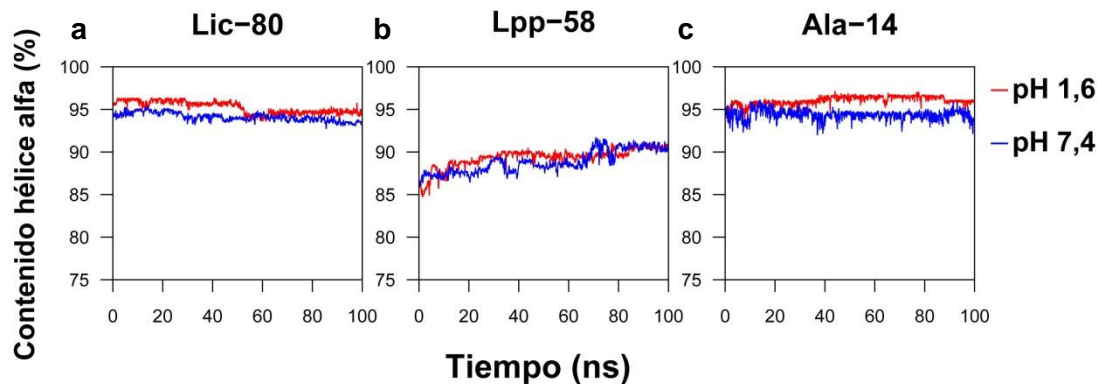


Figura 14. Contenido hélice alfa en las simulaciones preparativas. Lic-80 (a), Lpp-58 (b) y Ala-14 (c) en pH ácido (línea roja) y pH neutro (línea azul).

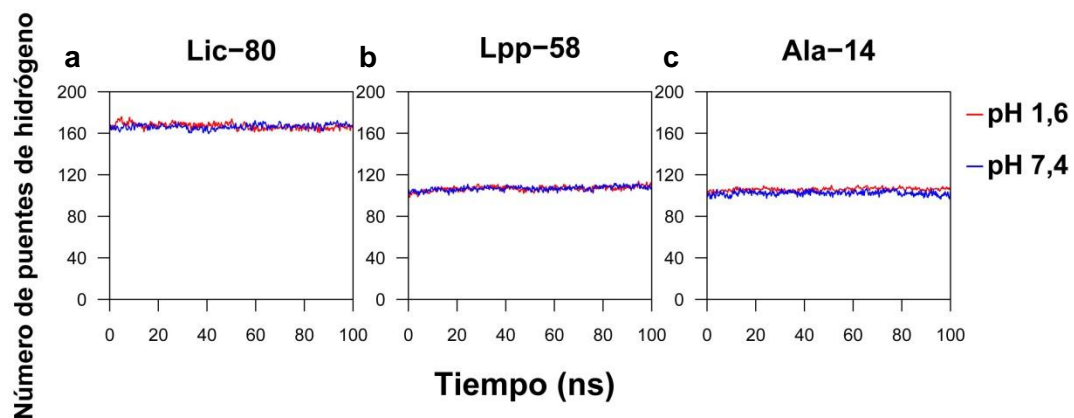


Figura 15. Puentes de hidrógeno de las cadenas principales en las simulaciones preparativas. Lic-80 (a), Lpp-58 (b) y Ala-14 (c) en pH ácido (línea roja) y pH neutro (línea azul).

El número de contactos hidrofóbicos se mantiene estable para las tres proteínas. Lic-80 posee un número considerablemente mayor de interacciones hidrofóbicas, en comparación con Lpp-58 y Ala-14, debido a que su mayor longitud da lugar a una interfaz hidrofóbica más extensa (Figura 16).

Durante esta simulación preparativa Ala-14 se mantuvo plegada y el efecto del estado de protonación en la estabilidad de Lpp-58 fue menor, lo que muestra la

dificultad de sacar las estructuras del pozo de energía potencial en que se encuentran y resalta la necesidad de utilizar métodos que mejoren la exploración del espacio conformacional como “Replica Exchange”.

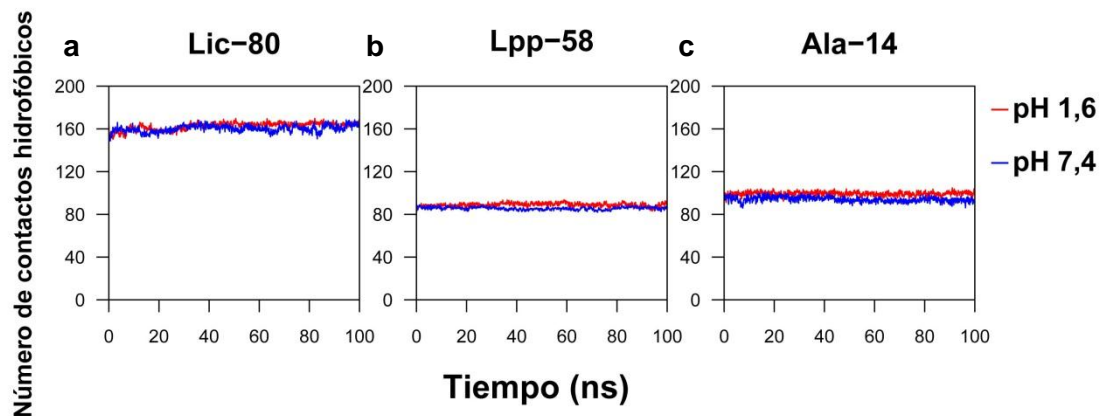


Figura 16. Contactos hidrofóbicos en las simulaciones preparativas. Lic-80 (a), Lpp-58 (b) y Ala-14 (c) en pH ácido (línea roja) y pH neutro (línea azul).

6.3 Simulación de Dinámica Molecular con Intercambio de Réplicas:

Con la estructura final de cada simulación anterior se procedió a realizar una simulación de REMD, utilizando 64 réplicas distribuidas en un rango de temperaturas entre 303,15 K y 809,57 K. Se definió el número de contactos hidrofóbicos inter-monómeros como criterio para estudiar la estabilidad del trímero (estructura cuaternaria), mientras que el contenido hélice alfa se utilizó para evaluar la estabilidad en general ya que es equivalente a las mediciones que se realizan experimentalmente por dicroísmo circular.

Para todos los sistemas simulados las estructuras parten plegadas y se comienzan a desplegar rápidamente durante los primeros 10 ns en las temperaturas más altas. Luego el desplegamiento sigue avanzando a temperaturas menores pero de manera más lenta (**Figuras 17, 18, 19, 20 y 21**). Las tres proteínas poseen una alta estabilidad en las condiciones de simulación, necesitando temperaturas muy elevadas para que se produzca el desplegamiento (**Tabla I**). Este fenómeno da cuenta de la dificultad para sacar las estructuras de los pozos energéticos en que se encontraban. La T_m de Lpp-56 a pH 7,4 está documentada alrededor de los 338 K cuando se utiliza una tasa de calentamiento de 1,0 K/min y de ~334 K a 0,125 K/min, existiendo una dependencia con la velocidad de calentamiento al necesitarse una tasa de calentamiento más lenta para obtener una T_m más cercana al proceso de desplegamiento en equilibrio (Dragan et al., 2004). En la simulación a pH 7,4 Lpp-58 presentó una T_m de 539,71 K, al menos 200 K más que el valor reportado experimentalmente (**Tabla I**). Este fenómeno se puede explicar en parte a la aproximación de solvente implícito utilizada, que se ha descrito que produce una sobreabundancia de estructuras secundarias tipo hélice alfa (Roe et al., 2007). Por estos motivos, es importante utilizar métodos que mejoren la exploración del espacio conformacional como REMD, ya que si se hubiese buscado observar el desplegamiento a las temperaturas descritas experimentalmente con una simulación tradicional, se hubiesen necesitado tiempos de simulación prácticamente inalcanzables. De esta manera, al estudiar proteínas con procesos de desplegamiento muy lentos siempre es recomendable utilizar los tiempos de simulación más extensos que la infraestructura disponible permita.

El protocolo de REMD se sustenta en mantener probabilidades de intercambio relativamente uniformes de manera que cada réplica pueda recorrer libremente las diferentes temperaturas. Sin embargo, en la distribución de temperaturas utilizadas, a pesar de tener una buena probabilidad de alrededor de 0,5, se generaron “cuellos de botella” en los que ocurrieron escasos eventos de intercambio (**Figura 22**). La probabilidad depende de la diferencia de energía potencial entre las dos réplicas a intercambiar, mientras menor sea esta diferencia mayor es la probabilidad de intercambio. Como estos “cuellos de botella” aparecen a temperaturas cercanas a las T_m , su ocurrencia se explicaría debido a la gran diferencia de energía potencial entre las estructuras plegadas y desplegadas. De esta manera, se podría suponer que el desplegamiento observado es de carácter irreversible, ya que una vez que se despliega una estructura su energía potencial aumenta y deja de intercambiarse con las réplicas de menor temperatura que le podría permitir replegarse.

Para reconstruir el espacio conformacional total explorado por las réplicas de cada sistema se procedió a utilizar WHAM ocupando como ejes de coordenadas el número de contactos hidrofóbicos y el contenido hélice alfa (**Figura 23**). El espacio conformacional de las tres proteínas en estudio fue similar, observándose dos pozos con mayor probabilidad (color verde oscuro). El primero, en la esquina superior derecha de cada panel, corresponde al pozo principal donde se encuentran las estructuras plegadas y recibe principalmente la contribución de las simulaciones a bajas temperaturas. El segundo pozo más pequeño, ubicado en la esquina inferior izquierda de cada panel, corresponde a las estructuras desplegadas, recibe principalmente la contribución de las simulaciones a temperaturas más altas. En este gráfico queda en evidencia la dificultad para sacar las proteínas desde el primer pozo hasta el segundo. Además, en la transición entre ambos pozos no se identificaron terceros pozos de probabilidad, es decir, no se observan configuraciones intermediarias estables. El camino de transición de Licanantasa resultó el menos probable (color rojo), es decir que se necesita una mayor cantidad de energía para llevar a Licanantasa desde el primer pozo hasta el segundo. Las diferencias ocasionadas por el pH ácido en el espacio conformacional fueron mínimas, observando un pequeño ensanchamiento del pozo principal y un aumento del área del

espacio conformacional explorado, fenómeno que se explica por la mayor flexibilidad que adquieren las proteínas al perder sus puentes salinos y puentes de hidrógeno inter-monómero.

En el proceso de desplegamiento observado, primero se pierden los puentes de hidrógeno inter-monómero y los puentes salinos, luego los contactos hidrofóbicos, los puentes de hidrógeno de las cadenas principales y finalmente la estructura secundaria (**Figura 24, Tabla I**).

Ala-14 presentó la menor estabilidad de las tres proteínas en ambas condiciones de pH. No obstante, esta estabilidad fue mucho mayor a la esperada. Se ha reportado que la mutante Ala-14 no forma trímeros y sólo presenta ~20% de hélice alfa a 273,15 K (Liu et al., 2002), por lo que se esperaba que su T_m fuera cerca de 65 K menor que la T_m de Lpp-58.

Se ha descrito experimentalmente que Lpp-56 disminuye su T_m desde 338 K a pH 7,4 hasta 316 K a pH 3,0 (Dragan et al., 2004). Sin embargo, esta disminución de la T_m no se observó en ninguna de las simulaciones realizadas, por el contrario, se observó que la T_m aumentó en condiciones de protonación a pH ácido. Esta anomalía podría explicarse por la aproximación de solvente implícito utilizada en las simulaciones. Aun cuando el modelo GB/SA para solvente implícito contempla el efecto de apantallamiento de cargas y de la tensión superficial, no considera el impacto de los puentes de hidrógenos que se forman entre el agua y la proteína, ni el efecto de los iones. Por esta razón no es posible afirmar ni descartar *a priori* que una solución ácida produzca una disminución de la T_m de Licanantasa. Sin embargo, aunque así ocurriese, la diferencia de estabilidad entre Licanantasa a pH ácido y a pH neutro debería ser menor que en el caso de Lpp, ya que la estructura de Licanantasa depende menos de puentes de hidrógeno y puentes salinos sensibles al cambio de protonación que la estructura de Lpp. Estos antecedentes sumados a que Lic-80 presenta la mayor estabilidad *in silico* de las proteínas en estudio, nos permiten proponer que la estabilidad de Licanantasa no se vería comprometida de forma

importante al encontrarse en un ambiente ácido como el de los medios de biolixiviación.

La estabilidad de la hélice sobre-enrollada en estas simulaciones se sustenta por una parte en conservar las interacciones entre los monómeros, siendo los contactos hidrofóbicos los más importantes, y por otra parte en la mantención de la estructura secundaria por medio de los puentes de hidrógeno de la cadena principal. Licanantasa al ser más extensa posee una mayor interfaz hidrofóbica que estabiliza el trímero y un mayor número de puentes de hidrógeno que estabiliza la hélice alfa. Por lo tanto, se requiere una mayor cantidad de energía para producir el desplegamiento.

La baja identidad de secuencia entre Licanantasa y Lpp da cuenta del fenómeno que la estructura es más conservada que la secuencia (Chothia y Lesk, 1986). En este caso, lo que se mantiene es la arquitectura de hélice sobre-enrollada más que aa específicos. Las diferencias encontradas darían cuenta de adaptaciones específicas al ambiente y al rol biológico de cada proteína.

Tabla I. Temperaturas de desplegamiento

	Tm					
Proteína	Contactos hidrofóbicos		Puentes de hidrógeno intra-monómero cadena principal		Hélice alfa	
	pH 1,6	pH 7,4	pH 1,6	pH 7,4	pH 1,6	pH 7,4
Ala-14	523,11 K	499,18 K	523,11 K	507,03 K	539,71 K	515,01 K
Lpp-58	531,35 K	507,03 K	547,95 K	523,11 K	556,59 K	539,71 K
Lic-80	556,59 K	556,59 K	565,34 K	556,59 K	574,24 K	556,59 K
	Puntos medios					
Proteína	Puentes de hidrógeno inter-monómero cadena lateral		Puentes salinos inter-monómero			
	pH 7,4		pH 7,4			
Ala-14	483,83 K		483,83 K			
Lpp-58	468,95 K		491,44 K			
Lic-80	407,48 K		483,83 K			

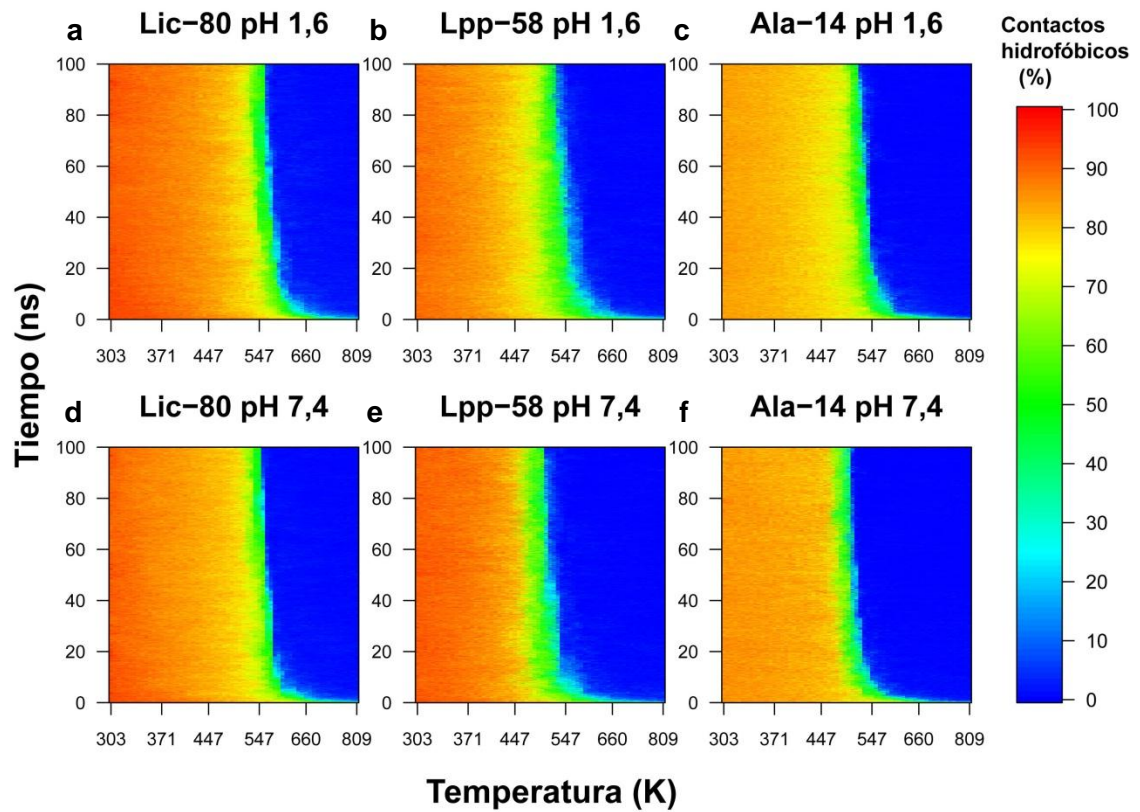


Figura 17. Porcentaje de contactos hidrofóbicos en las simulaciones de REMD. Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

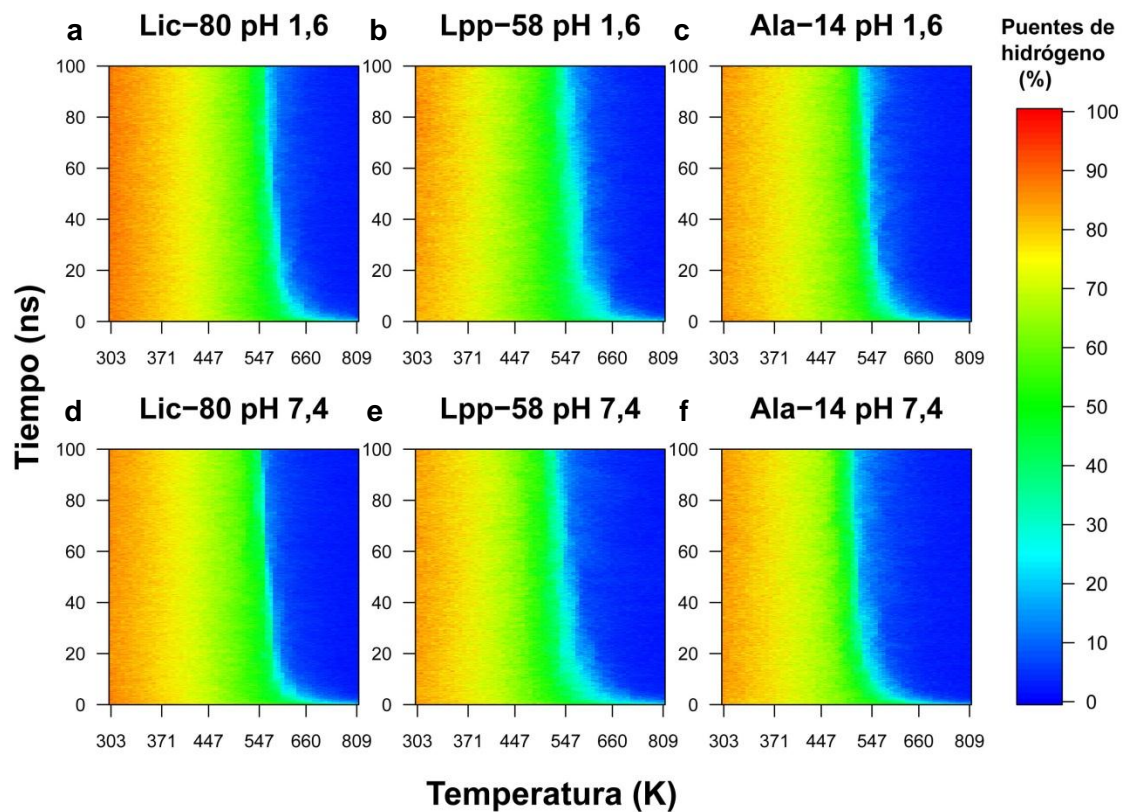


Figura 18. Porcentaje puentes de hidrógeno de cadena principal en las simulaciones de REMD. Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

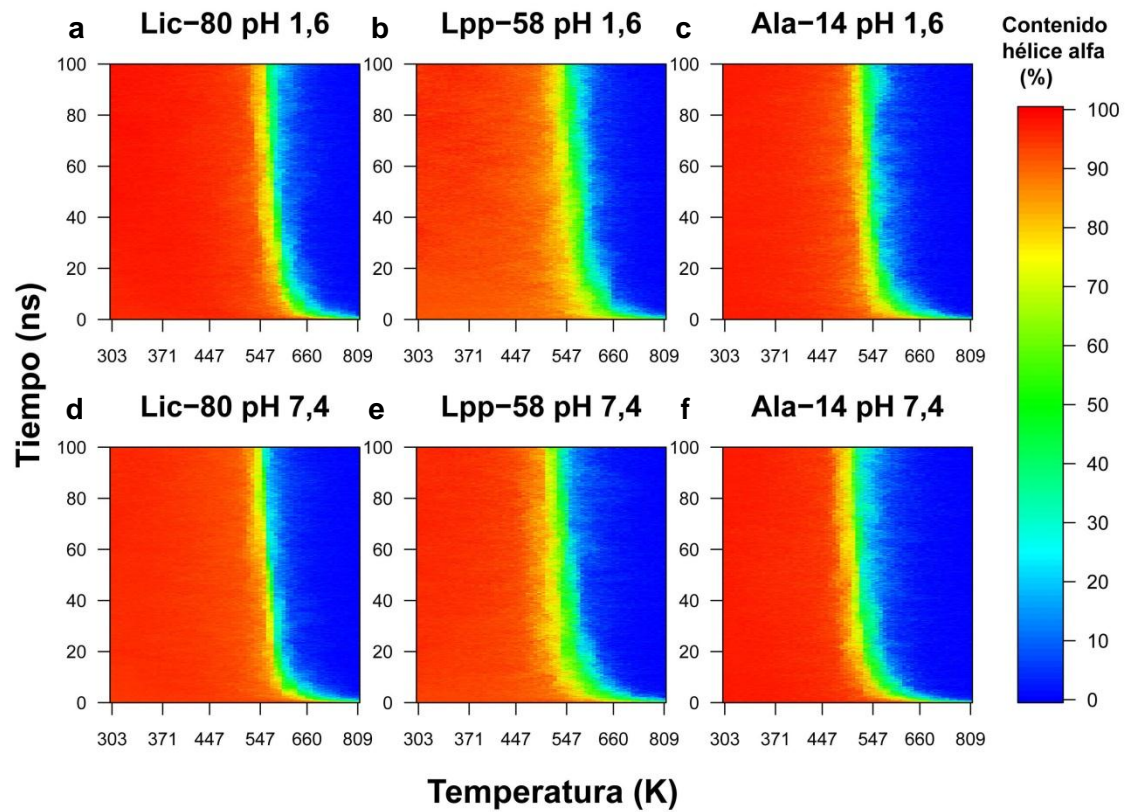


Figura 19. Porcentaje de contenido de hélice alfa en las simulaciones de REMD. Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

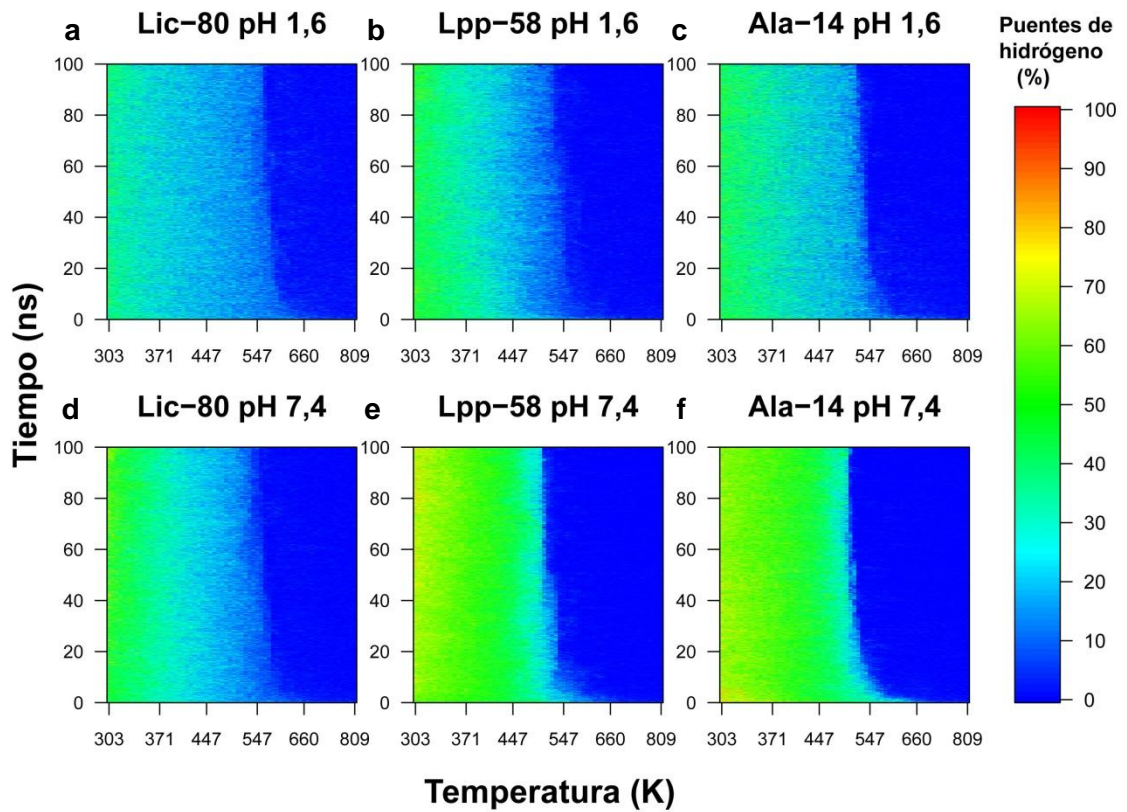


Figura 20. Porcentaje de puentes de hidrógeno inter-monómero en las simulaciones de REMD. Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

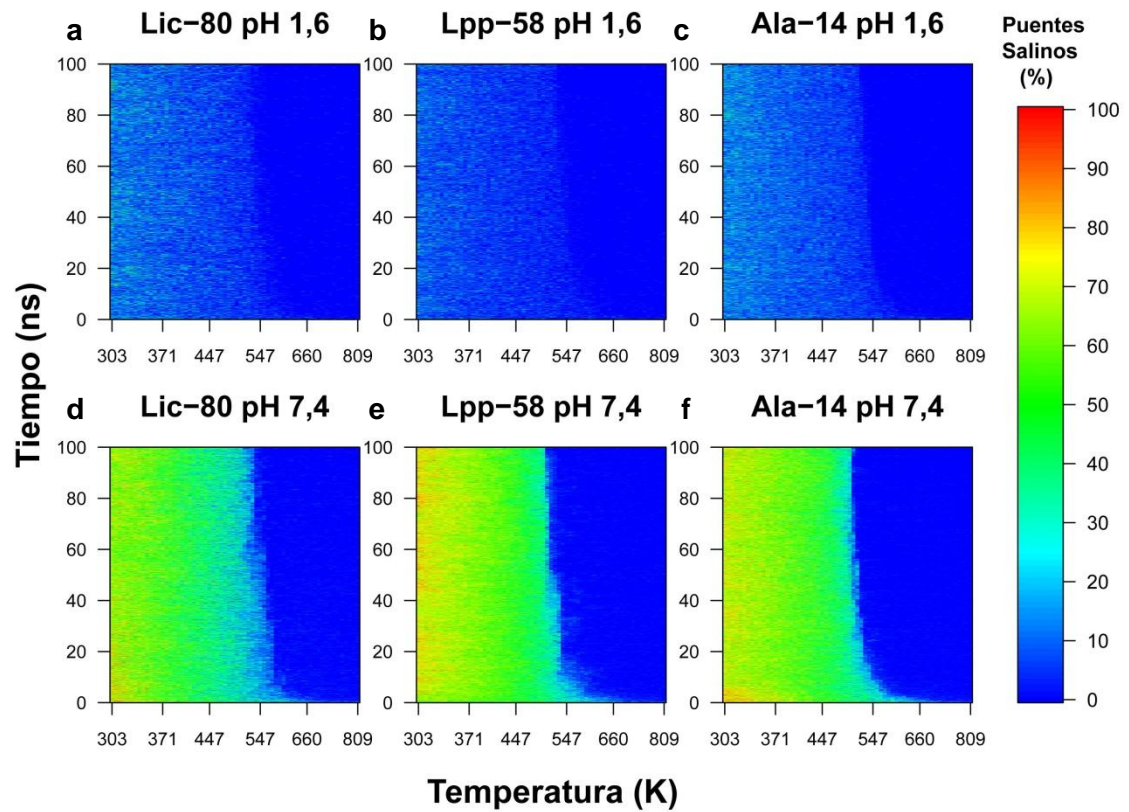


Figura 21. Porcentaje de puentes de puentes salinos en las simulaciones de REMD. Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

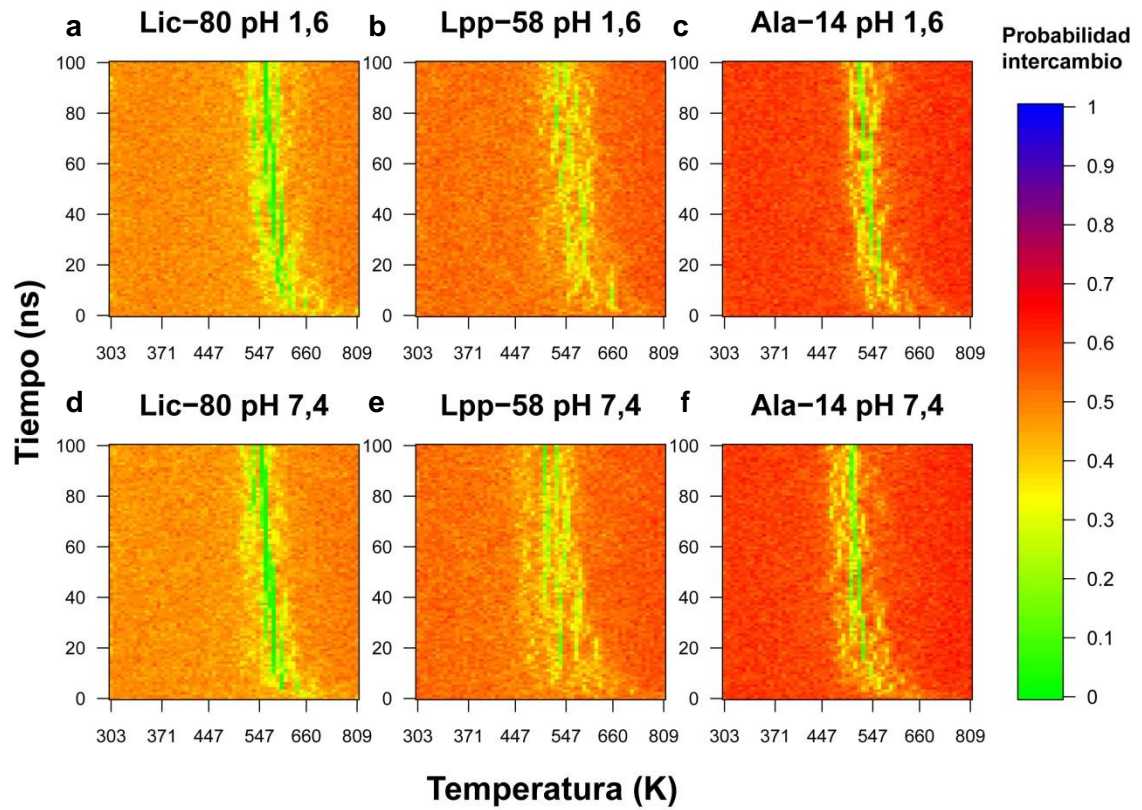


Figura 22. Probabilidad de intercambio en las simulaciones de REMD.
 Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

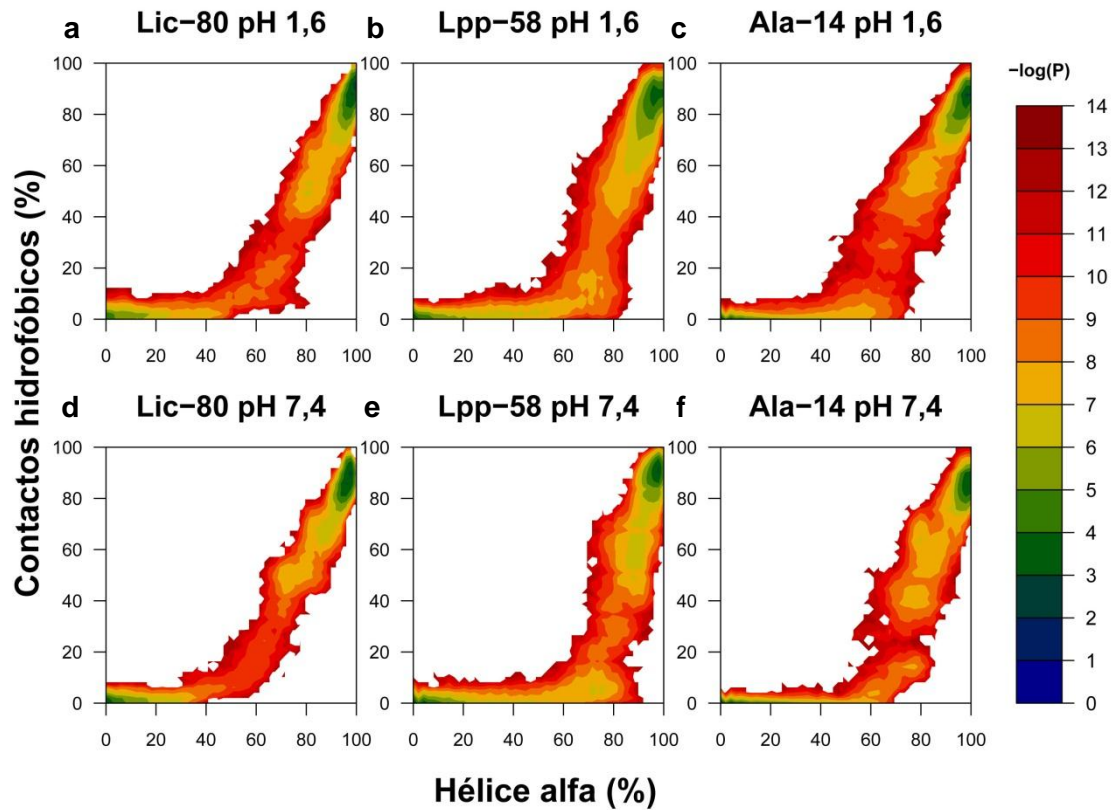


Figura 23. Representación mediante WHAM del espacio conformacional muestreado. Probabilidad de encontrar una estructura con un determinado porcentaje de hélice alfa y de contactos hidrofóbicos a 303.15 K en las simulaciones de REMD para Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro.

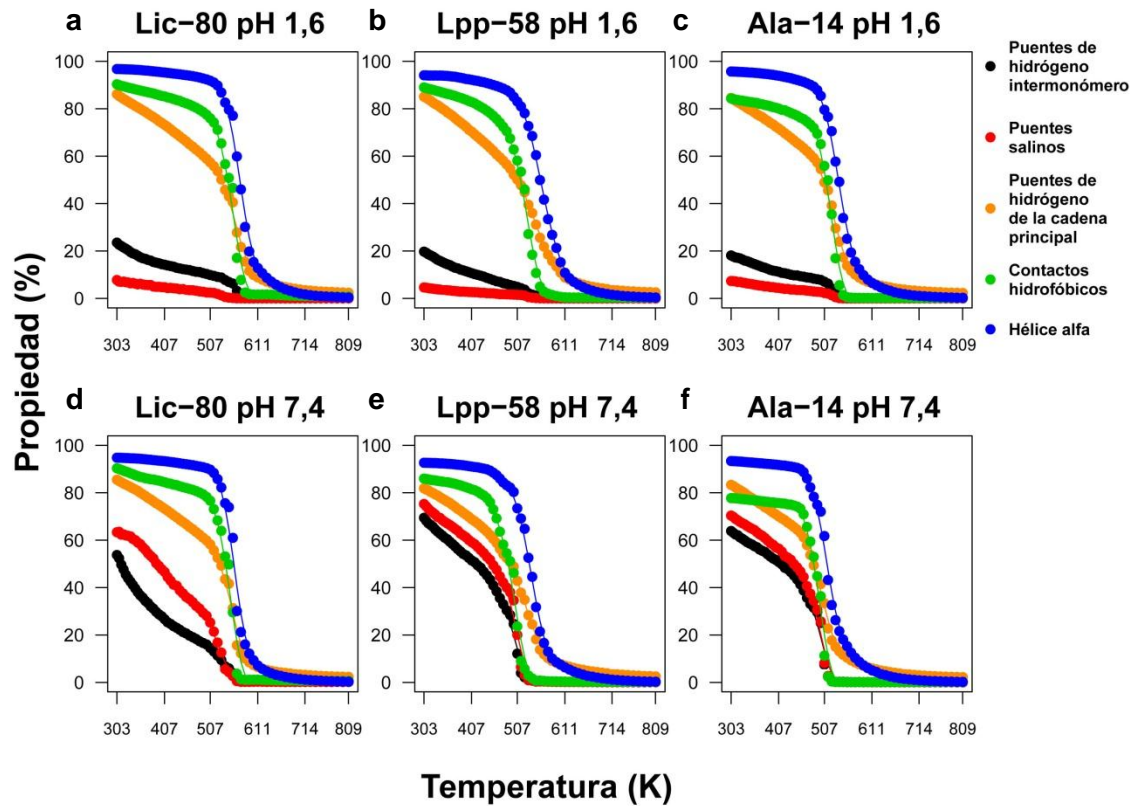


Figura 24. Promedio de propiedades estructurales versus temperatura.

Puentes de hidrógeno inter-monómero, puentes salinos, contactos hidrofóbicos, y contenido de hélice alfa en las simulaciones de REMD para Lic-80, Lpp-58 y Ala-14 en pH ácido y pH neutro. Se utilizaron los datos de los últimos 50 ns de cada simulación.

7 CONCLUSIONES

1. La estructura obtenida mediante modelamiento *de novo* para el trímero de Licanantasa, presentó las características estructurales propuestas: una hélice sobre-enrollada, un dominio “alanine-zipper” y residuos en posiciones apropiadas para formar interacciones no enlazantes inter-monómero.
2. La pérdida de interacciones electrostáticas inter monómeros que presentó Licanantasa a pH ácido fue menor que la presentada por Lpp y Ala-14. De esta forma, Licanantasa podría soportar mejor ambientes ácidos, como lo son los medios en que ocurre el proceso de biolixiviación. En consecuencia, la estabilidad de su arquitectura macromolecular sería indiferente a la pérdida de puentes de hidrógeno y puentes salinos producto del cambio de protonación.
3. La metodología utilizada no fue capaz de dar cuenta de la disminución de la T_m descrita experimentalmente para Lpp cuando se encuentra a pH ácido. Esta situación podría explicarse por el uso de una aproximación de solvente implícito en el protocolo de Replica Exchange.
4. Dentro de las interacciones no enlazantes inter-monómeros analizadas, los contactos hidrofóbicos y los puentes de hidrógeno de cadena principal son los principales encargados de la mantención de la arquitectura macromolecular de Licanantasa. Comparando con los controles utilizados, la estabilidad aumentaría a medida que: mayor sea el número de puentes de hidrógeno de cadena principal, mayor sea el número de contactos hidrofóbicos y mayor sea el tamaño de los residuos hidrofóbicos involucrados. De esta manera, Licanantasa presenta la mejor combinación de estas características dentro de las proteínas estudiadas, siendo la más estable dado que se requiere de una mayor cantidad de energía para producir su desplegamiento.

8 BIBLIOGRAFÍA

- Altschul, S. F., Madden, T. L., Schäffer, a a, Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389–3402.
- Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. a, and Yu, Y.-K. (2005). Protein database searches using compositionally adjusted substitution matrices. *The FEBS journal* 272, 5101–5109.
- Bjelić, S., Karshikoff, A., and Jelesarov, I. (2006). Stability and folding/unfolding kinetics of the homotrimeric coiled coil Lpp-56. *Biochemistry* 45, 8931–8939.
- Bjelić, S., Wieninger, S., Jelesarov, I., and Karshikoff, A. (2008). Electrostatic contribution to the thermodynamic and kinetic stability of the homotrimeric coiled coil Lpp-56: A computational study. *Proteins: Structure, Function, and Bioinformatics* 70, 810–822.
- Bjelkmar, P., Larsson, P., Cuendet, M. A., Hess, B., and Lindahl, E. (2010). Implementation of the CHARMM force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *Journal of Chemical Theory and Computation* 6, 459–466.
- Bobadilla Fazzini, R. a, Levican, G., and Parada, P. (2011). Acidithiobacillus thiooxidans secretome containing a newly described lipoprotein Licanantase enhances chalcopyrite bioleaching rate. *Applied Microbiology and Biotechnology* 89, 771–780.
- de Castro, E., Sigrist, C. J. a, Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., and Hulo, N. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* 34, W362–5.
- Cheng, J., Randall, a Z., Sweredoski, M. J., and Baldi, P. (2005). SCRATCH: a protein structure and structural feature prediction server. *Nucleic Acids Research* 33, W72–6.
- Chi, A., Valenzuela, L., Beard, S., Mackey, A. J., Shabanowitz, J., Hunt, D. F., and Jerez, C. a (2007). Periplasmic proteins of the extremophile Acidithiobacillus ferrooxidans: a high throughput proteomics analysis. *Molecular & Cellular Proteomics : MCP* 6, 2239–2251.
- Choi, D. S., Yamada, H., Mizuno, T., and Mizushima, S. (1986). Trimeric structure and localization of the major lipoprotein in the cell surface of Escherichia coli. *The Journal of Biological Chemistry* 261, 8953–8957.

- Chothia, C., and Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The European Molecular Biology Organization Journal* 5, 823–826.
- Claros, M. G., and von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Bioinformatics* 10, 685–686.
- Cole, C., Barber, J. D., and Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* 36, W197–201.
- Combet, C., Blanchet, C., Geourjon, C., and Deléage, G. (2000). NPS@: network protein sequence analysis. *Trends in Biochemical Sciences* 25, 147–150.
- Das, R., André, I., Shen, Y., Wu, Y., Lemak, A., Bansal, S., Arrowsmith, C. H., Szyperski, T., and Baker, D. (2009). Simultaneous prediction of protein folding and docking at high resolution. *Proceedings of the National Academy of Sciences of the United States of America* 106, 18978–18983.
- Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* 32, W665–7.
- Dragan, A. I., Potekhin, S. a, Sivolob, A., Lu, M., and Privalov, P. L. (2004). Kinetics and thermodynamics of the unfolding and refolding of the three-stranded alpha-helical coiled coil, Lpp-56. *Biochemistry* 43, 14891–14900.
- Eswar, N., Webb, B., Marti-Renom, M. a, Madhusudhan, M. S., Eramian, D., Shen, M.-Y., Pieper, U., and Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.] Chapter 2, Unit 2.9.*
- Frank, K., and Sippl, M. J. (2008). High-performance signal peptide prediction based on sequence alignment techniques. *Bioinformatics (Oxford, England)* 24, 2172–2176.
- Frishman, D., and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics* 23, 566–579.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, J. M. Walker, ed. (Humana Press), pp. 571–607.
- Gehrke, T., Telegdi, J., Thierry, D., and Sand, W. (1998). Importance of Extracellular Polymeric Substances from *Thiobacillus ferrooxidans* for Bioleaching. *Applied and Environmental Microbiology* 64, 2743–2747.

- Green, J. R., Korenberg, M. J., and Aboul-magd, M. O. (2009). PCI-SS : MISO dynamic nonlinear protein secondary structure prediction. *BMC Bioinformatics* 12, 1–12.
- von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *Journal of Molecular Biology* 225, 487–494.
- Hess, B. (2008). P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *Journal of Chemical Theory and Computation* 4, 116–122.
- Hirota, Y., Suzuki, H., Nishimura, Y., and Yasuda, S. (1977). On the Process of Cellular Division in *Escherichia coli*: A Mutant of *E. coli* Lacking a Murein-Lipoprotein. *Proceedings of the National Academy of Sciences* 74, 1417–1420.
- Hofmann, K., and Stoffel, W. (1993). TMbase - A database of membrane spanning proteins segments. *Biol Chem HoppeSeyler* 374, 166.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of Molecular Graphics* 14, 33–38.
- Kim, D. E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research* 32, W526–31.
- Knickerbocker, C., Nordstrom, D. K., and Southam, G. (2000). The role of “blebbing” in overcoming the hydrophobic barrier during biooxidation of elemental sulfur by *Thiobacillus thiooxidans*. *Chemical Geology* 169, 425–433.
- Larsson, P., and Lindahl, E. (2010). A high-performance parallel-generalized born implementation enabled by tabulated interaction rescaling. *Journal of Computational Chemistry* 31, 2593–2600.
- Liu, J., Cao, W., and Lu, M. (2002). Core side-chain packing and backbone conformation in Lpp-56 coiled-coil mutants. *Journal of Molecular Biology* 318, 877–888.
- Liu, J., Dai, J., and Lu, M. (2003). Zinc-mediated helix capping in a triple-helical protein. *Biochemistry* 42, 5657–5664.
- Liu, J., and Lu, M. (2002). An alanine-zipper structure determined by long range intermolecular interactions. *The Journal of Biological Chemistry* 277, 48708–48713.
- Onufriev, A., Bashford, D., and Case, D. a (2004). Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 55, 383–394.

- Patriksson, A., and van der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Physical chemistry chemical physics : PCCP* 10, 2073–2077.
- Pearson, W. R. (1990). Rapid and sensitive sequence comparison with FASTP and FASTA. *Methods in Enzymology* 183, 63–98.
- Pollastri, G., and McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics (Oxford, England)* 21, 1719–1720.
- Rahman, O., Cummings, S. P., Harrington, D. J., and Sutcliffe, I. C. (2008). Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria. *World Journal of Microbiology and Biotechnology* 24, 2377–2382.
- Ramírez, P., Guiliiani, N., Valenzuela, L., Beard, S., and Jerez, C. A. (2004). Differential Protein Expression during Growth of *Acidithiobacillus ferrooxidans* on Ferrous Iron, Sulfur Compounds, or Metal Sulfides. *Applied and Environmental Microbiology* 70, 4491–4498.
- Rawlings, D. E. (2005). Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microbial Cell Factories* 4, 13.
- Roe, D. R., Okur, A., Wickstrom, L., Hornak, V., and Simmerling, C. (2007). Secondary structure bias in generalized Born solvent models: comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *The Journal of Physical Chemistry. B* 111, 1846–1857.
- Rohwerder, T., Gehrke, T., Kinzler, K., and Sand, W. (2003). Bioleaching review part A: progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation. *Applied Microbiology and Biotechnology* 63, 239–248.
- Sand, W., and Gehrke, T. (2006). Extracellular polymeric substances mediate bioleaching/biocorrosion via interfacial processes involving iron(III) ions and acidophilic bacteria. *Research in Microbiology* 157, 49–56.
- Seibert, M. M., Patriksson, A., Hess, B., and van der Spoel, D. (2005). Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *Journal of Molecular Biology* 354, 173–183.
- Shortle, D., Simons, K. T., and Baker, D. (1998). Clustering of low-energy conformations near the native structures of small proteins. *Proceedings of the National Academy of Sciences of the United States of America* 95, 11158–11162.

- Shu, W., Liu, J., Ji, H., and Lu, M. (2000). Core structure of the outer membrane lipoprotein from *Escherichia coli* at 1.9 Å resolution. *Journal of Molecular Biology* 299, 1101–1112.
- Sindhikara, D. J. (2011). Modular reweighting software for statistical mechanical analysis of biased equilibrium data. *Computer Physics Communications* 182, 2227–2231.
- Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* 6, 175–182.
- van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: fast, flexible, and free. *Journal of Computational Chemistry* 26, 1701–1718.
- Sugita, Y., and Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* 314, 141–151.
- Tokuda, H. (2009). Biogenesis of Outer Membranes in Gram-Negative Bacteria. *Bioscience, Biotechnology, and Biochemistry* 73, 465–473.
- Wolf, E., Kim, P. S., and Berger, B. (1997). MultiCoil: a program for predicting two- and three-stranded coiled coils. *Protein science : a publication of the Protein Society* 6, 1179–1189.
- Zhang, C., Zhang, R., Xia, J., Zhang, Q., and Nie, Z. (2008). Sulfur activation-related extracellular proteins of *Acidithiobacillus ferrooxidans*. *Transactions of Nonferrous Metals Society of China* 18, 1398–1402.

9 APÉNDICE

9.1 Dinámica Molecular¹

Se entiende por Dinámica Molecular (MD) a la técnica de simulación computacional donde la evolución temporal de un grupo de átomos que interactúan se obtiene a partir de las ecuaciones de movimiento de Newton:

$$m_i \frac{\partial^2 r_i}{\partial t^2} = F_i$$

Para cada átomo i en un sistema constituido por N átomos, donde m_i es la masa del átomo, r_i es su posición, t es el tiempo y F_i es la fuerza que actúa sobre él debido a las interacciones que presenta con los otros átomos. Las fuerzas son obtenidas a partir de las derivadas negativas de una función potencial $V(r_1, r_2, \dots, r_N)$:

$$F_i = -\frac{\partial V}{\partial r_i}$$

Las ecuaciones son resueltas simultáneamente en pequeños pasos de tiempo, del orden de femtosegundos, para garantizar la conservación de energía total del sistema. A medida que el sistema avanza las coordenadas de los átomos se van escribiendo a un archivo de salida a intervalos regulares de tiempo. Estas coordenadas en función del tiempo representan una trayectoria del sistema. MD al ser una técnica basada en la mecánica estadística permite relacionar el comportamiento microscópico al macroscópico. El valor de una propiedad macroscópica se puede obtener a partir del promedio de un conjunto de configuraciones microscópicas. Una trayectoria obtenida mediante MD provee este conjunto de configuraciones. Por lo tanto, la medición de una propiedad en una simulación corresponde simplemente a la media aritmética de sus varios valores instantáneos durante la simulación de MD. En el límite de simulaciones muy largas, uno podría esperar que el espacio de fase sea completamente muestreado y que, en ese límite, el proceso de promediar produjera las propiedades termodinámicas. En la práctica, las simulaciones siempre son de duración finita, y uno debe tener precaución al estimar cuando el muestreo ha sido suficiente o no.

¹ Adaptado de Ercolessi (1997)

9.2 Campo de fuerza²

La energía potencial del sistema es representada mediante un campo de fuerza, que se constituye por dos componentes que describen la interacción entre las partículas. El primero corresponde al conjunto de ecuaciones usadas para generar las energías potenciales y sus derivadas, las fuerzas. El segundo componente corresponde a los parámetros usados en este conjunto de ecuaciones. La función de energía potencial se compone de las siguientes contribuciones:

$$V_{total} = V_{enlace} + V_{ángulo} + V_{diedro} + V_{vdW} + V_{Coulomb}$$

Los tres primeros términos describen las interacciones de tipo enlazante. La energía potencial de enlace describe el movimiento vibracional armónico entre dos átomos unidos covalentemente:

$$V_{enlace} = \sum_{enlace\ i} k_i^{enlace} (r_i - r_{0i})^2$$

La energía potencial angular describe el movimiento vibracional del ángulo que se forma entre tres átomos unidos de forma covalente:

$$V_{ángulo} = \sum_{ángulo\ i} k_i^{ángulo} (\theta_i - \theta_{0i})^2$$

La energía potencial de ángulo torsional describe el ángulo diedro formado entre cuatro átomos enlazados consecutivamente:

$$V_{diedro} = \sum_{diedro\ i} \begin{cases} k_i^{diedro} [1 + \cos(n_i \psi_i - \phi_i)] & n_i \neq 0 \\ k_i^{diedro} (\psi_i - \phi_i)^2 & n_i = 0 \end{cases}$$

Los dos últimos términos de la función de energía potencial describen las interacciones entre pares de átomos no unidos. El potencial Lennard-Jones 6-12 describe las fuerzas de van der Waals:

$$V_{vdW} = \sum_i \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]$$

² Adaptado de Phillips et al. (2005)

El potencial de Coulomb describe las interacciones electrostáticas:

$$V_{Coulomb} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

9.3 Algoritmo de integración³

El motor de MD es el algoritmo de integración, que es el encargado de resolver numéricamente las ecuaciones de movimiento de Newton y de esta manera generar las trayectorias. Existen varias aproximaciones de integración numérica, entre ellas las más conocidas son los algoritmos de Verlet, Leap-Frog y Velocity-Verlet. En GROMACS el algoritmo de integración por defecto es Leap-Frog que usa las posiciones r al tiempo t y las velocidades v al tiempo $t - \frac{1}{2}\Delta t$; y se actualizan las posiciones y velocidades usando las fuerzas $F(t)$ determinadas por las posiciones al tiempo t :

$$v\left(t + \frac{1}{2}\Delta t\right) = v\left(t - \frac{1}{2}\Delta t\right) + \frac{\Delta t}{m}F(t)$$

$$r(t + \Delta t) = r(t) + \Delta t v\left(t + \frac{1}{2}\Delta t\right)$$

El funcionamiento de este algoritmo se puede observar en la **Figura A1**.

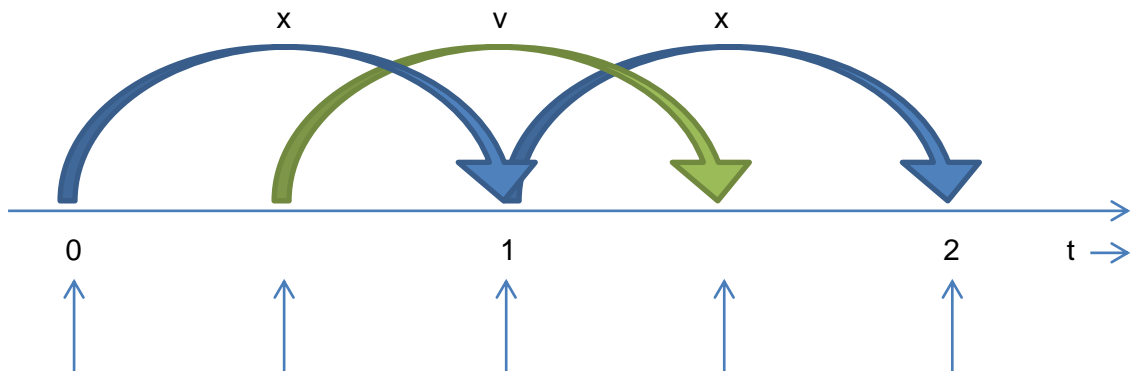


Figura A1. Método de integración Leap-Frog.

El algoritmo es llamado de esta forma dado que r y v saltan como ranas sobre las espaldas del otro. Figura adaptada del manual de usuario de GROMACS versión 4.5.4.

³ Adaptado del manual de usuario de GROMACS 4.5.4

9.4 Solvente implícito⁴

Los modelos de solvente implícito proveen una forma eficiente de representar los efectos electrostáticos de las moléculas de solvente, a la vez que permiten ahorrar una gran parte del cómputo asociado a la descripción precisa del agua del medio en las simulaciones de MD. Los modelos de solvente implícito ofrecen varias ventajas sobre la solvatación explícita, incluyendo la eliminación de la necesidad de equilibrar el agua alrededor del soluto y la ausencia de viscosidad que permite a la proteína explorar más rápidamente el espacio conformacional.

Para estimar la energía libre de solvatación total de una molécula ΔG_{solv} , uno supone típicamente que puede ser descompuesta en sus partes electrostática y no electrostática:

$$\Delta G_{solv} = \Delta G_{el} + \Delta G_{surf}$$

Donde ΔG_{surf} es la energía libre de solvatación de una molécula a la cual se le removieron todas las cargas, y ΔG_{el} es la energía libre de primero remover todas las cargas en el vacío y luego agregarlas en la presencia de un ambiente con solvente. Esta descripción es la base del esquema ampliamente usado PB/SA. En esta aproximación, ΔG_{el} es computado por la metodología numérica Poisson-Boltzmann (PB), y ΔG_{surf} se considera proporcional al total del área de superficie accesible a solvente (SA) de la molécula, con una constante de proporcionalidad derivada de energía de solvatación experimentales de pequeñas moléculas no polares.

La aproximación analítica Generalized Born (GB) es una forma de calcular la parte electrostática de la energía libre de solvatación ΔG_{el} . Esta metodología se ha vuelto popular debido a su relativa simplicidad y eficiencia computacional, comparada con la solución numérica estándar de la ecuación PB. El modelo GB representa cada átomo en la molécula como una esfera de radio ρ_i con una carga q_i en su centro, el interior de cada átomo se supone lleno de un material con una constante dieléctrica igual a 1.

⁴ Adaptado de Onufriev, Bashford y Case (2004)

La molécula es rodeada por un solvente con un valor dieléctrico alto ϵ_w , 80 para el agua a 300 K. El modelo GB aproxima ΔG_{el} mediante una formula analítica:

$$\Delta G_{el} \approx \Delta G_{GB} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f^{GB}(r_{ij}, R_i, R_j)} \left(1 - \frac{e^{-K f_{ij}^{GB}}}{\epsilon_w} \right)$$

Donde r_{ij} es la distancia entre los átomos i y j , R_i son los llamados radios de Born efectivos de los átomos i y j , y f^{GB} es una cierta función de suavizado de sus argumentos. Una elección común para f^{GB} es:

$$f^{GB} = \sqrt{r_{ij}^2 + R_i R_j \exp\left(-\frac{r_{ij}^2}{4R_i R_j}\right)}$$

El radio efectivo de Born de un átomo refleja su grado de entierro al interior de la molécula. Para un ion aislado R_i es igual su radio de vdW ρ_i , mientras para un ion profundamente enterrado $R_i \gg \rho_i$. El radio efectivo depende de la conformación de la molécula y debe ser recalculado cada vez que la conformación cambie. Los modelos más conocidos para calcular el radio de Born son Still, HCT y OBC, siendo este último el que produce mejores resultados.

9.5 Replica Exchange⁵

La Dinámica molecular con intercambio de réplicas (REMD) es un método que puede ser utilizado para mejorar el muestreo de cualquier tipo de simulación, especialmente si las conformaciones están separadas por barreras energéticas relativamente altas. Este método consiste en simular múltiples réplicas del mismo sistema a diferentes temperaturas y realizar intercambios al azar del estado completo de dos replicas a intervalos regulares con la probabilidad:

$$P(1 \leftrightarrow 2) = \min\left(1, \exp\left[\left(\frac{1}{k_B T_1} - \frac{1}{k_B T_2}\right)(U_1 - U_2)\right]\right)$$

Donde T_1 y T_2 son las temperaturas de referencia y U_1 y U_2 son las energías potenciales instantáneas de las réplicas 1 y 2 respectivamente. Solo se intentan intercambios para temperaturas vecinas ya que la probabilidad de intercambio

⁵ Adaptado de Sugita y Okamoto (1999) y del manual de usuario de GROMACS 4.5.4

disminuye muy rápidamente con la diferencia de temperatura (**Figura A2**). Después de un intercambio las velocidades son escaladas por $\left(\frac{T_1}{T_2}\right)^{\pm 0.5}$.

De esta manera, los sistemas realizan un camino aleatorio libre unidimensional en el espacio de la temperatura. Este camino aleatorio induce un camino aleatorio en el espacio de energía que permite a las simulaciones escapar de los mínimos locales de energía. Para obtener un desempeño óptimo de REMD, se requiere elegir una distribución adecuada de temperaturas, de manera de asegurar una probabilidad suficiente ($P > 0.1$) y uniforme entre las distintas réplicas, y que la temperatura más alta sea suficiente para evitar quedar atrapado en un mínimo de energía local.

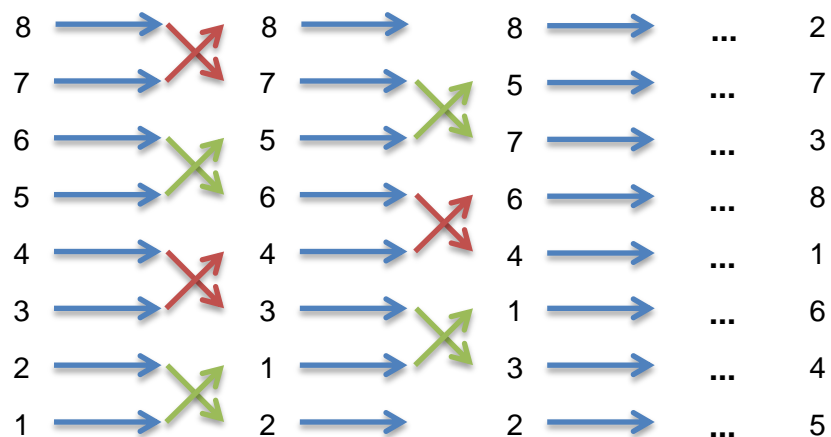


Figura A2. Representación del protocolo de REMD para ocho réplicas.

Cada réplica comienza con la misma estructura inicial pero se simula a una temperatura distinta, siendo la réplica 1 la que comienza a menor temperatura y la réplica 8 la de mayor temperatura. Los intercambios de réplicas se intentan cada cierto número de pasos de simulación. Las flechas verdes representan los intentos de intercambio exitosos, mientras que las flechas rojas representan los intentos de intercambio fallidos.

9.6 Algoritmo de predicción *de novo* de Rosetta⁶

Rosetta es un paquete de software unificado para la predicción de estructuras de proteína y diseño funcional. Como todos los algoritmos de predicción debe desarrollar dos tareas. Primero debe explorar o muestrear el espacio conformacional relevante.

⁶ Adaptado de Kaufmann et al. (2010)

En segundo lugar debe clasificar de forma precisa o evaluar la energía de los modelos estructurales resultantes. Con este fin Rosetta implementa aproximaciones de muestreo Metropolis Monte Carlo guiados por conocimiento, en conjunto con funciones de energía guiadas por conocimiento. Las funciones de energía guiadas por conocimiento suponen que la mayoría de las propiedades moleculares pueden ser derivadas de la información disponible, en este caso del Banco de Datos de Proteínas (PDB).

Rosetta separa el muestreo de la conformación general del refinamiento local para la cadena principal. Se modelan los cambios conformacionales generales por medio de intercambios de conformaciones de fragmentos peptídicos de tres y nueve aminoácidos, obtenidos a partir de una biblioteca de fragmentos (**Figura A3**). Las conformaciones de los péptidos son recogidas del PDB para tramos de secuencia homólogos que capturan el sesgo estructural para la secuencia local. Para el refinamiento local de los modelos de proteínas se utiliza Metropolis Monte Carlo para muestrear los ángulos ϕ y ψ que no desarmen el plegamiento global de la proteína.

Rosetta reduce drásticamente el número de conformaciones muestreadas a través del uso de conformaciones discretas de las cadenas laterales que se observan en el PDB. Estos rotámeros capturan las combinaciones permitidas entre ángulos de torsión de las cadenas laterales así como los ángulos ϕ y ψ de la cadena principal, y por lo tanto reducen tamaño del espacio conformacional.

El algoritmo comienza con una cadena peptídica extendida. La inserción de fragmentos de cadena principal pliega rápidamente la proteína usando una función de energía de baja resolución y métodos de muestreo. Se intentan aproximadamente 30000 inserciones de fragmentos de nueve residuos seguidas por otras 10000 inserciones de fragmentos de tres residuos para generar un modelo de proteína. Usualmente se pliegan entre 20000 – 50000 modelos para cada proteína individual. Los modelos resultantes pueden ser refinados a un nivel atómico.

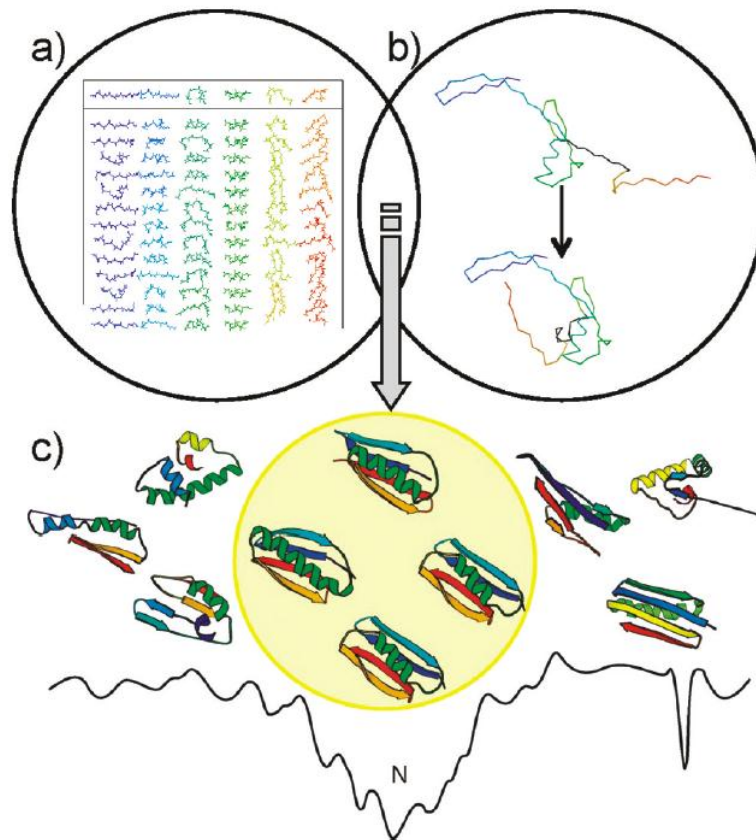


Figura A3. Algoritmo de plegamiento *de novo*.

Rosetta comienza desde bibliotecas de fragmentos con ángulos (φ y ϕ) dependiente de secuencia que capturan el espacio de conformación local accesible a una secuencia. La combinación de fragmentos diferentes de las bibliotecas pliega la proteína a través de la optimización de contactos no locales. La función de energía de baja resolución representada en el panel (c) suaviza la hiperficie de energía rugosa, resultando en un mínimo amplio y profundo para la conformación nativa. Una minimización Metropolis Monte Carlo dirige la estructura hacia el mínimo global. Figura adaptada de Kaufmann et al., 2010.

El protocolo de plegamiento *de novo* de proteínas se adaptó para permitir la predicción precisa de proteínas homoméricas. Para ello se desarrolló el protocolo “Rosetta Fold and Dock” que combina el protocolo de predicción *de novo* y de acoplamiento proteína-proteína. Este nuevo protocolo alterna entre ciclos de inserción simétrica de fragmentos y acoplamiento de cuerpo rígido entre los monómeros parcialmente ensamblados.

“Fold and Dock” supone que los elementos de estructura secundaria de un homómero son simétricos e inserta los mismos fragmentos para cada subunidad. Como la interfaz entre un homómero se encuentra en gran medida enterrada, el

acoplamiento durante el plegamiento permite proteger y estabilizar esta región durante el proceso de plegamiento, lo que mejora enormemente la precisión.

En la predicción de estructura *de novo* se utiliza comúnmente el proceso de agrupamiento de los mejores modelos y su clasificación por tamaño de grupo, siendo el grupo de mayor tamaño el que se encuentra teóricamente más cerca del plegamiento nativo, esto se basa en la observación de que los pozos de energía más profundos son también los más amplios.

La función de energía de baja resolución trata las cadenas laterales como centroides. Esta función de energía modela la solvatación, electrostática, puentes de hidrógeno entre hebras β e impedimentos estéricos. Los efectos de solvatación son modelados como la probabilidad de observar un aminoácido en particular rodeado por un número dado de carbonos α dentro de una distancia de corte dependiente de aminoácido. Las interacciones electrostáticas son modeladas como la probabilidad de observar una distancia dada entre dos centroides de aminoácidos. Puentes de hidrógeno entre las hebras β se evalúa en base al arreglo geométrico relativo de los fragmentos de hebra. Se penaliza la superposición de átomos de la cadena principal y los centroides de las cadenas laterales y de este modo se provee el componente repulsivo de una fuerza de van der Waals. Un término de radio de giro se usa para modelar el efecto de atracción de van der Waals. Todos los perfiles de probabilidad fueron derivados de los cristales disponibles en el PDB usando estadística Bayesiana. La baja resolución de esta función de energía basada en centroides suaviza la hiperficie de energía a costa de su precisión. La hiperficie suavizada de energía permite mantener una baja energía para las estructuras que están cercanas al mínimo global, incluso con defectos estructurales que una función de energía atómica penalizaría rígidamente.

La función de energía de alta resolución permite un detalle a nivel atómico, para ello combina el potencial Lennard-Jones 6-12 para describir las fuerzas de van der Waals, una aproximación de solvatación, un potencial de puentes de hidrógeno dependiente de la orientación, un término electrostático basado en conocimiento, y un término de

energía libre interna para los aminoácidos dependiente de conformación basado en conocimiento.