



DIVISIÓN DE CIENCIAS NATURALES E INGENIERÍA
DEPARTAMENTO DE CIENCIAS NATURALES

PROPUESTA DE PROYECTO DIVISIONAL

Estudio de la estabilidad térmica de proteínas
utilizando técnicas computacionales

Dr. Salomón de Jesús Alas Guardado

Ciudad de México, Septiembre de 2019

Presentación de Proyecto de Investigación al Consejo Divisional de la DCNI

1. Título

Estudio de la estabilidad térmica de proteínas utilizando técnicas computacionales

2. Responsable y participantes del proyecto

Responsable

Dr. Salomón de Jesús Alas Guardado

Profesores y alumnos participantes

Dr. Felipe Aparicio Platas

Dr. Gerardo Pérez Hernández

Alumnos de la DCNI

3. Tipo de proyecto

Investigación básica

4. Áreas temáticas

- a) Físicoquímica
- b) Sistemas Computacionales y Bioinformática

5. Líneas de generación y/o aplicación del conocimiento

- a) Físicoquímica Molecular
- b) Ingeniería y Diseño Molecular

6. Periodo y duración

- Fecha de inicio: Octubre de 2019
- Fecha final: Septiembre de 2023
- Duración: 4 años

7. Proyecto

Introducción

Las proteínas surgen como producto de la información que cada organismo posee en su genoma. Estas macromoléculas cumplen con una función específica en cada sistema y han evolucionado durante miles de años; persistiendo sólo aquellas proteínas que permiten la sobrevivencia del organismo de donde provienen [1].

Existen organismos con la capacidad de vivir en ambientes extremos de temperatura, pH, salinidad, presión, etc. Cuando un organismo desarrolla la capacidad de adaptarse a un ambiente de temperatura mayor a 45 °C es considerado un organismo termófilo. Por otra parte los organismos que tienen una temperatura óptima de crecimiento entre 15 y 35 °C son conocidos como organismos mesófilos [1,2]. Para que los organismos termófilos persistan bajo esas condiciones extremas de temperatura, es necesario que las proteínas sigan realizando su función específica de manera adecuada.

Se sabe que las proteínas realizan su función solo cuando se encuentran en el estado nativo, el cual es un estado tridimensional muy particular en donde la proteína se encuentra plegada y se considera la estructura más estable que puede adquirir. Cuando las condiciones ambientales como temperatura, pH, salinidad, entre otras, cambian de manera considerable la proteína se despliega y pierde la función, pasando del estado nativo a un estado desnaturalizado [2].

De aquí que, el estudio de los factores estructurales y fisicoquímicos que le proporcionan mayor estabilidad a una proteína para mantener su actividad a temperaturas altas ha tenido un gran impacto tanto en la generación de conocimiento en las ciencias básicas como en aplicación industrial. Esto último debido a que al aumentar la estabilidad térmica de una proteína se puede obtener un producto específico con valor agregado. Sin embargo, en la actualidad estos factores son poco conocidos [3].

Una manera de estudiar a tales factores es con ayuda de la simulación computacional, este tipo de herramientas como su nombre lo dice nos permiten simular, en este caso sistemas biológicos, para tratar de comprender y explicar el tipo de interacciones que contribuyen a que una proteína permanezca estable y realice su función al aumentar la temperatura. De manera particular la herramienta de simulación computacional de dinámica molecular (DM) permite estudiar de manera teórica los cambios e interacciones que ocurren en sistemas biológicos, como las proteínas, en un periodo de tiempo determinado.

Es por ello que en este proyecto de investigación se pretende estudiar diferentes proteínas homólogas procedentes de dos tipos de microorganismos, uno mesófilo y el otro termófilo, utilizando DM, con el fin de determinar cuáles son los factores estructurales y fisicoquímicos responsables de aumentar la estabilidad térmica de la proteína.

Antecedentes

El estudio de estabilidad térmica de proteínas se puede realizar mediante desnaturalización térmica y química; siguiendo el proceso con técnicas de dicroísmo circular, fluorescencia o calorimetría diferencial de barrido. Al utilizar calorimetría se obtienen tres parámetros: la temperatura de fusión (T_m), el cambio de entalpía a la temperatura de fusión (ΔH_m) y el cambio de la capacidad calorífica (ΔC_p) entre los estados nativo y desnaturalizado de la proteína [4].

Utilizando los datos anteriores se puede realizar una curva de estabilidad térmica como la mostrada en la Figura 1. En donde se observa la temperatura en la cual la proteína se encuentra más estable (T_s) y la temperatura de fusión (T_m), que es la temperatura en la cual el 50% de la proteína se encuentra en estado nativo y el otro 50% en el estado desnaturalizado [4]. A partir de este punto la proteína pierde su funcionalidad.

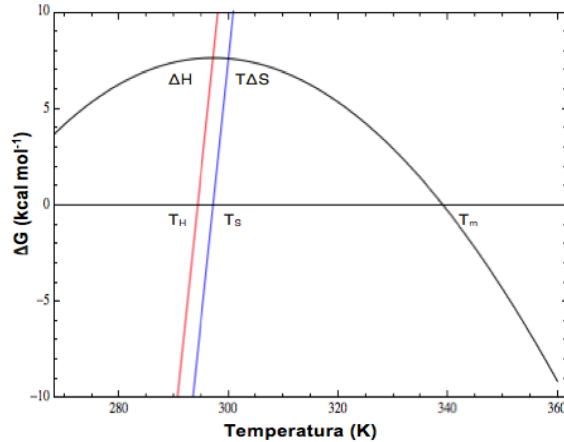


Figura 1. Curva de estabilidad térmica de una proteína: T_S es la temperatura de máxima estabilidad ($\Delta S = 0$), T_H es la temperatura en donde $\Delta H = 0$ y T_m es la temperatura de fusión ($\Delta G = 0$).

En particular, las proteínas de organismos termófilas tienen que aumentar su T_m para mantener la funcionalidad a una temperatura más alta. Esto se observa, por ejemplo, en el comportamiento de la proteína Metil guanina metil transferasa (MGMT) procedente de un organismo termófilo y su homóloga procedente de un organismo mesófilo (Figura 2).

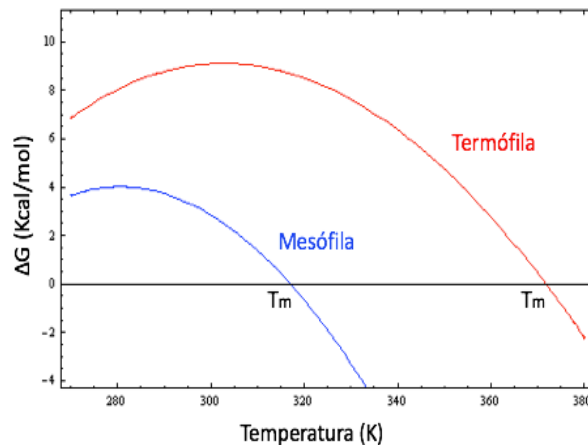


Figura 2. Comparación de las curvas de estabilidad térmica de la MGMT para los microorganismos *Escherichia coli* (línea en color azul) y *Thermococcus kodakarensis* (línea en color rojo). Se puede observar que la proteína MGMT del microorganismo termófilo permanece estable a temperaturas más altas.

Otra estrategia para comprender la estabilidad térmica de proteínas es el uso de simulaciones de dinámica molecular (DM), la cual es una técnica que se utiliza para estudiar las propiedades estructurales y dinámicas de sistemas biomoleculares [5]. Con esta técnica computacional, es posible estudiar proteínas, con su medio y las condiciones bioquímicas como las que se establecen en un laboratorio experimental.

Con las simulaciones de DM se puede obtener información acerca de la desviación de movimientos estructurales, los cuales consisten en medir por ejemplo: desviación de la raíz

cuadrada media (RMSD), fluctuación de la raíz cuadrada media (RMSF), radio de giro (Rg), análisis de la componente principal (PCA), entre otros. Además, se pueden hacer mediciones fisicoquímicas como: puentes salinos, puentes de hidrógeno (proteína-proteína y proteína-solvente), puentes disulfuro, entre otros. También se pueden hacer otras mediciones estructurales como: área accesible al solvente (ASA), contactos nativos, estructura secundaria, etc.

Con la DM se han hecho estudios de proteínas específicas; estas simulaciones han permitido observar qué tipos de interacciones intermoleculares proporcionan mayor estabilidad térmica a una proteína. Se ha observado que existe una tendencia a cambios en la conformación y un aumento en los puentes salinos que le confieren estabilidad térmica cuando la temperatura cambia, sin embargo esto no ocurre de manera general para todas las proteínas [3,4]. Por tal motivo se propone realizar análisis estructurales y fisicoquímicos, como los mencionados, de diferentes proteínas homólogas procedentes de microorganismos termófilos y mesófilos para tratar de entender el por qué las proteínas de organismos termofilos presentan mayor estabilidad a aumentos de temperatura.

Metodología

1. Para llevar a cabo la selección de las proteínas, primero es necesario buscar en la literatura proteínas que sean de tipo reversible, esto es, que su desplegamiento sea de un estado, esto con el fin de obtener los datos de calorimetría diferencial de barrido: la temperatura de fusión (T_m), el cambio de entalpía a la temperatura de fusión (ΔH_m) y el cambio de la capacidad calorífica (ΔC_p) entre los estados nativo y desnaturalizado de la proteína [4].
2. Con los datos de calorimetría diferencial de barrido se procederá a realizar las curvas de estabilidad térmica tanto para las proteínas procedentes de los organismos termófilos como de las proteínas homólogas procedentes de los organismos mesófilos. Esto con el fin de tener el comportamiento de las curvas de estabilidad para cada par de proteínas.
3. Para realizar el análisis de las estructuras primarias, secundarias y terciarias de las proteínas se utilizarán herramientas bioinformáticas como BLAST (Basic Local Alignment Tool), VMD (Visual Molecular Dynamics), PyMol, MOE (Molecular Operating Environment), entre otras. Previamente las estructuras de cada proteína se obtendrán de los archivos del Protein Data Bank (PDB). En caso de que no exista la estructura en el PDB, la estructura se modelará con la ayuda de algún software como el SWISS-PROT.
4. Las simulaciones de dinámica molecular se realizaran con el programa de alto rendimiento GROMACS (GRONingen MACHine for Chemical Simulations) versión 2016.3, utilizando el campo de fuerza AMBER99SB. Para llevar a cabo tales simulaciones se usa el protocolo siguiente:
 - a) Se toman las coordenadas de la estructura proteica del código PDB correspondiente. La cual se ha obtenido por difracción de rayos X (RX) o por Resonancia Magnética Nuclear (RMN), o en caso de no existir la estructura, ésta se modela con algún software (SWISS-PROT por ejemplo).
 - b) Se deben de agregar los átomos de hidrógeno necesarios para protonar a la proteína. Esto se puede realizar utilizando el software PROPKA.

- c) Se utiliza solvente explícito, por tanto la proteína se debe solvatar con agua. El modelo de agua que se pretende utilizar es SPC/E (Extended Simple Point Charge). El cual es adecuado para trabajar con el campo de fuerza propuesto.
- d) Se deben de agregar iones de Na^+ y Cl^- para neutralizar el sistema.
- e) Se establecen las medidas de la celda de simulación, la cual puede ser un cubo de $L_x \times L_y \times L_z \text{ \AA}^3$ o alguna otra forma geométrica como un dodecaedro.
- f) Se establecen condiciones de contorno periódicas. A fin de evitar efectos de tamaño finito.
- g) Etapa de minimización de la energía del sistema: se relaja la estructura de la proteína a fin de evitar contactos inapropiados provenientes del código PDB (resolución en la difracción de RX o la RMN), además, así se tienen las moléculas de agua y iones orientadas adecuadamente. Se pretende utilizar el algoritmo de *descenso más rápido* [6].
- h) Etapa de equilibración del sistema: se utiliza una función que considera el termostato de *rescalamiento de velocidades* para el acoplamiento de temperatura [7], el cual es una extensión del termostato de Berendsen pero utilizando una fuerza aleatoria, de tal forma que la distribución de la energía cinética cambia y por tanto corrige la temperatura indirectamente. Se propone que la escala de tiempo del termostato (τ) sea de $\tau = 0.1 \text{ ps}$.

Para tener mayor efectividad en el control de la temperatura el acoplamiento debe de ocurrir entre dos grupos de partículas: el solvente y la proteína. La temperatura inicial se asigna mediante una distribución de velocidades de Maxwell-Boltzmann y se proponen simular 100 – 300 ps en un ensamble NVT. El paso de simulación se realiza cada 2 fs.

Se propone establecer un intervalo de temperaturas, las cuales se obtendrán de las curvas de estabilidad. Las simulaciones se harán de forma independiente, esto es, para cada punto de temperatura se aplicará este protocolo.

Para equilibrar la presión se acopla una función barostática con un valor de 1 atm. El baróstato que se propone utilizar es el de Parrinello-Rahman [8], el cual modifica los valores de los vectores de la celda de simulación de manera uniforme, esto con el fin de regular el volumen y la presión de la celda. El coeficiente de compresibilidad a utilizar sería de $4.5 \times 10^{-5} \text{ bar}^{-1}$. El tiempo de simulación durante la equilibración del baróstato se pretende que sea de 1 – 3 ns.

Para ambas equilibraciones (termostato y baróstato) los cálculos de las interacciones no enlazantes y coulombicas para cada uno de los átomos se realizan con un radio de corte de interacción de 1 nm. Los átomos vecinos dentro de este radio de corte se obtienen de la lista de átomos vecinos, que se implementa utilizando el *algoritmo de Verlet* [9]. La lista se actualiza cada 10 fs.

El tratamiento de las interacciones electrostáticas se hace por sumas de Ewald, mediante el método de la *mallita de partículas de Ewald* [10]. Además, se utiliza el algoritmo de LINCS para restricciones de enlace [11].

- i) Etapa de producción: una vez obtenida la equilibración del sistema se elimina la traslación del centro de masa de la proteína y el solvente en cada paso de simulación, con el fin de evitar desviaciones que pudieran afectar el control de temperatura. Durante esta etapa se pretenden simular 50 – 500 millones de pasos con un paso de integración de 2 fs lo que equivale a 100 – 1000 ns (1 ms) y guardar

tanto las coordenadas como las energías cada 5000 pasos. El algoritmo integrador que se pretende utilizar es el de *salto de rana* [12].

- j) Se propone realizar 2 réplicas por cada simulación por cada temperatura, variando la semilla inicial de la distribución de velocidades de Maxwell-Boltzmann durante la etapa de equilibración, con el objetivo de evaluar la reproducibilidad de las simulaciones y con el fin de explorar un mayor espacio conformacional de la proteína en cuestión.
- k) Una vez realizadas las simulaciones de dinámica molecular se realizarán los análisis de los factores estructurales: RMSD, RMSF, Rg, ASA, PCA, contactos nativos, etc., y los factores fisicoquímicos: puentes de hidrógenos (proteína-proteína y proteína-solvente), contactos de grupos polares y no polares, puentes salinos, etc.

5. Proteínas que se proponen analizar

Proteína	Organismo mesófilo	Organismo termófilo	Código PDB
Histidina fosfotransportadora	Bacillus subtilis	Bacillus staerothermophilus	2HPR, 1Y4Y
Isopropil malato deshidrogenasa	Escherichia coli	Thermus thermophilus	1CM7, 1OSI
RNasa H	Escherichia coli	Thermus thermophilus	1JXB, 1RIL
Ssh10b	Escherichia coli	Sulfolobus shibatae	1ARQ, 1Y9X

Hipótesis

Las interacciones intermoleculares son las responsable de dar mayor estabilidad térmica a las proteínas procedentes de organismos termófilos en comparación con proteínas homólogas procedentes de organismos mesófilos, sin embargo, estas interacciones dependerán del tipo de proteína.

Objetivo general

Determinar los factores estructurales y fisicoquímicos que le proporcionan estabilidad térmica a proteínas de organismos termófilos utilizando dinámica molecular clásica y herramientas bioinformáticas.

Objetivos particulares

1. Seleccionar diferentes proteínas procedentes de organismo termófilos y sus homólogas procedentes de organismos mesófilos.
2. Realizar las curvas de estabilidad térmica de las proteínas seleccionadas.
3. Analizar las estructuras primarias y secundarias de las proteínas seleccionadas utilizando herramientas bioinformáticas.
4. Analizar factores estructurales de las proteínas seleccionadas utilizando dinámica molecular clásica.

5. Analizar factores fisicoquímicos de las proteínas seleccionadas utilizando dinámica molecular clásica.

8. Metas

Metas científicas

- a) Obtención de proyectos terminales de 4 alumnos de licenciatura (uno por año).
- b) Obtención del grado de maestría de un alumno.
- c) Obtención del grado de doctor de un alumno.
- d) Publicación de un artículo de investigación por año en revistas que cuenten con arbitraje internacional.
- e) Participación en eventos especializados a nivel nacional e internacional.

Metas académicas

- a) Formación de recursos humanos a nivel licenciatura y posgrado con capacidad para realizar trabajo multidisciplinario.
- b) Publicar un artículo de divulgación por alumno de posgrado en alguna revista pertinente a fin de dar a conocer, a un público no especializado, la temática de la investigación realizada en este proyecto.
- c) Impartir conferencias, seminarios y/o talleres en torno a la temática del proyecto.

9. Recursos necesarios

- Equipo de supercómputo de alto rendimiento.
- Licencias de software.
- Discos duros externos.

10. Infraestructura actual

- Laboratorio de Supercómputo y Visualización en paralelo (LSVP), Yoltla, UAM Iztapalapa. En la actualidad se tiene acceso al laboratorio con 5 cuentas; si es necesario se pueden solicitar nuevas cuentas de usuarios.
- Una supercomputadora personal CereBro32 con tarjeta GPU de 2560 cores.
- Una supercomputadora personal con 32 procesadores (CPU).
- Una supercomputadora personal con 36 procesadores (CPU).
- Computadoras personales y mobiliario.

11. Presupuesto

Año	Rubro	Monto
1	• Una supercomputadora personal CereBro32 con tarjeta GPU de 4352 cores.	\$72,000.00
	• Adquisición de 2 discos duros externos de 1 TB	\$6,000.00
2	• Adquisición de un disco duro externo de 1 TB	\$3,000.00
	• Hospedaje y viáticos para asistencia a un congreso	\$8,000.00

	nacional	\$5,000.00
	• Pasajes para asistencia a un congreso nacional	
3	• Adquisición de un disco duro externos de 1 TB	\$3,000.00
	• Hospedaje y viáticos para asistencia a un congreso internacional	\$20,000.00
	• Pasajes para asistencia a un congreso internacional	\$20,000.00
	• Inscripción a congreso internacional	\$10,000.00
4	• No se ejercerá presupuesto	
Total		\$147,000.00

12. Fuentes de financiamiento externas

Fuente: CONACYT

Convocatoria: Investigación Científica Básica 2017-2018

Formalización: 05 de septiembre de 2019

Presupuesto: \$1,430,854.00

Nota: en espera de la primera ministración.

13. Planeación

Año 1			
Proteína de estudio: Histidina fosfotransportadora			
Actividad	T1	T2	T3
Revisión bibliográfica.	X	X	X
Obtención de las curvas de estabilidad térmica.	X		
Análisis de las estructuras primarias y secundarias utilizando herramientas bioinformáticas.	X	X	
Realización de las simulaciones de dinámica molecular.		X	X
Análisis de resultados.		X	X
Escritura de resultados.			X
Publicación de resultados en revistas y en congresos.			X

Año 2			
Proteína de estudio: Isopropil malato deshidrogenasa			
Actividad	T1	T2	T3
Revisión bibliográfica.	X	X	X
Obtención de las curvas de estabilidad térmica.	X		
Análisis de las estructuras primarias y secundarias utilizando herramientas bioinformáticas.	X	X	
Realización de las simulaciones de dinámica molecular.		X	X
Análisis de resultados.		X	X
Escritura de resultados.			X
Publicación de resultados en revistas y en congresos.			X

Año 3			
Proteína de estudio: RNasa H			
Actividad	T1	T2	T3
Revisión bibliográfica	X	X	X
Obtención de las curvas de estabilidad térmica.	X		
Análisis de las estructuras primarias y secundarias utilizando herramientas bioinformáticas.	X	X	
Realización de las simulaciones de dinámica molecular.		X	X
Análisis de resultados.		X	X
Escritura de resultados.			X
Publicación de resultados en revistas y en congresos.			X

Año 4			
Proteína de estudio: Ssh10b			
Actividad	T1	T2	T3
Revisión bibliográfica.	X	X	X
Obtención de las curvas de estabilidad térmica.	X		
Análisis de las estructuras primarias y secundarias utilizando herramientas bioinformáticas.	X	X	
Realización de las simulaciones de dinámica molecular.		X	X
Análisis de resultados.		X	X
Escritura de resultados.			X
Publicación de resultados en revistas y en congresos.			X

Resultados

1. Entregables

- a) Formación de cuatro alumnos de licenciatura, un alumno de maestría y un alumno de doctorado.
- b) Publicación de un artículo de investigación por año en revistas que cuenten con arbitraje internacional.
- c) Publicación de un artículo de divulgación por alumno de posgrado en revistas que cuenten con arbitraje.
- d) Asistencia y participación de los alumnos de licenciatura a un congreso nacional.
- e) Asistencia y participación de cada alumno de posgrado a congresos nacionales e internacionales.

2. Esperados

- a) Consolidación del Laboratorio de Físicoquímica y Simulación Molecular.
- b) Consolidación de la línea de investigación: Estudio de estabilidad de proteínas.

14. Visto bueno de la Jefa del Departamento donde manifiesta la viabilidad de la propuesta en función de financiamiento, infraestructura y recursos humanos necesarios.

Bibliografía

1. Lehninger, A. L. (2004). Principios de bioquímica lehninger. 4a ed. Ediciones Omega, S. A.
2. Jaenicke, R., Závodszky, P. (1990). Proteins under extreme physical conditions. *FEBS Letters*. 268 (2), 344-349.
3. Schellman, J. A. (1987). The thermodynamic stability of proteins. Institute of Molecular Biology, University of Oregon, Eugene, Oregon 97403. pp. 115-137.
4. Razvi, A., Scholtz, J. M. (2006). Lessons in stability from thermophilic proteins. *Protein Science: A Publication of the Protein Society*. 15 (7), 1569-1578.
5. Manjunath, K., Sekar, K. (2013). Molecular Dynamics Perspective on the Protein Thermal Stability: A Case Study Using SAICAR Synthetase. *J. Chem. Inf. Model*. 53 (9), 2448-2461.
6. Petrova, S. S., Solov'Ev, A. D. (1997). The Origin of the Method of Steepest Descent. *Hist. Math*. 24, 361-375.
7. Bussi, G., Donadio, D., Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys*. 126, 14101.
8. Parrinello, M., Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys*. 52, 7182-7190.
9. Páll, S., Hess, B. (2013). A flexible algorithm for calculating pair interactions on SIMD architectures. *Comput. Phys. Commun*. 184, 2641-2650.
10. Darden, T., York, D., Pedersen, L. (1993). Particle mesh Ewald: An $N\text{-log}(N)$ method for Ewald sums in large systems. *J. Chem. Phys*. 98, 10089-10092.
11. Hess, B., Bekker, H., Berendsen, H. J. C., Fraaije, J. G. E. M. (1997). LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem*. 18, 1463-1472.
12. Berendsen, H. J. C., Postma, J. P. M., Van Gunsteren, W. F., Dinola, A., Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys*. 81, 3684-3690.