



Estimación de la entropía durante la transición conformacional del lazo 36 de la hemaglutinina

Estimation of entropy during the conformational transition of loop 36 of hemagglutinin

Estimativa da entropia durante a transição conformacional da alça 36 da hemaglutinina

Klever David Cajamarca-Sacta ^I

klever.cajamarca@unach.edu.ec

<https://orcid.org/0009-0008-6446-5276>

Laura Esther Muñoz-Escobar ^{II}

laura.munoz@unach.edu.ec

<https://orcid.org/0009-0003-5573-7923>

Sandra Elizabeth Tenelanda-Cudco ^{III}

stenelanda@unach.edu.ec

<https://orcid.org/0000-0001-6215-9517>

Norma Isabel Allauca-Sandoval ^{IV}

normaisabel.allauca@unach.edu.ec

<https://orcid.org/0000-0003-0269-8417>

Correspondencia: klever.cajamarca@unach.edu.ec

Ciencias de la Educación

Artículo de Investigación

* **Recibido:** 10 de marzo de 2024 * **Aceptado:** 09 de abril de 2024 * **Publicado:** 22 de mayo de 2024

- I. Facultad Ciencias de la Educación, Humanas y Tecnologías, Pedagogía de las Ciencias Experimentales Matemáticas y la Física, Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- II. Facultad Ciencias de la Educación, Humanas y Tecnologías, Pedagogía de las Ciencias Experimentales Matemáticas y la Física, Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- III. Facultad Ciencias de la Educación, Humanas y Tecnologías, Pedagogía de las Ciencias Experimentales Matemáticas y la Física, Universidad Nacional de Chimborazo, Riobamba, Ecuador.
- IV. Facultad Ciencias de la Educación, Humanas y Tecnologías, Pedagogía de las Ciencias Experimentales Matemáticas y la Física, Universidad Nacional de Chimborazo, Riobamba, Ecuador.

Resumen

La evolución de la entropía durante la transición entre las conformaciones no-fusogénica y fusogénica del lazo 36 de la cadena HA2 de la Hemaglutinina ha sido estudiada utilizando trayectorias de simulaciones de dinámica molecular y la llamada aproximación quasi-armónica. Para el cálculo de la entropía de las conformaciones intermedias se aplicó un análisis de componentes principales al conjunto de estructuras asociadas a cada conformación. Las varianzas de estas componentes principales pueden considerarse asociadas a un conjunto de osciladores independientes cuya entropía puede calcularse. Los valores de entropía obtenidos se encuentran en el orden de 17 kcal/mol K, los cuales son mayores en un orden de magnitud a los resultados obtenidos en trabajos anteriores. Adicionalmente, se comprobó que el término $-TS$ contribuye de forma no despreciable a la energía libre. El término de energía interna se estimó a partir del promedio sobre el ensamble, aplicado sobre las trayectorias de dinámica molecular de equilibrarían para cada conformación intermedia, como resultado la altura de la barrera energética para la transición se estimó en 189.75 kcal/mol.

Palabras clave: Hemaglutinina; Entropía; Aproximación Quasi-Armónica; Dinámica Molecular.

Abstract

The evolution of entropy during the transition between the non-fusogenic and fusogenic conformations of loop 36 of the HA2 chain of Hemagglutinin has been studied using trajectories of molecular dynamics simulations and the so-called quasi-harmonic approximation. To calculate the entropy of the intermediate conformations, a principal component analysis was applied to the set of structures associated with each conformation. The variances of these principal components can be considered associated with a set of independent oscillators whose entropy can be calculated. The entropy values obtained are in the order of 17 kcal/mol K, which are greater by an order of magnitude than the results obtained in previous works. Additionally, it was proven that the $-TS$ term contributes in a non-negligible way to the free energy. The internal energy term was estimated from the average over the ensemble, applied to the equilibrium molecular dynamics trajectories for each intermediate conformation, as a result the height of the energy barrier for the transition was estimated at 189.75 kcal/mol.

Keywords: Hemagglutinin; Entropy; Quasi-Harmonic Approach; Molecular Dynamics.

Resumo

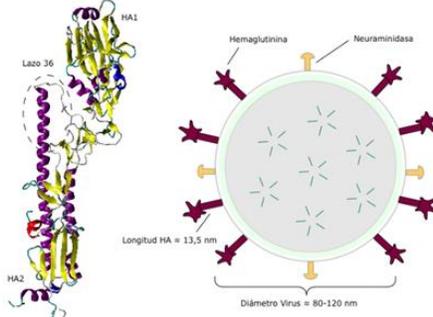
A evolução da entropia durante a transição entre as conformações não fusogênicas e fusogênicas da alça 36 da cadeia HA2 da Hemaglutinina foi estudada usando trajetórias de simulações de dinâmica molecular e a chamada aproximação quase harmônica. Para calcular a entropia das conformações intermediárias, foi aplicada uma análise de componentes principais ao conjunto de estruturas associadas a cada conformação. As variâncias destes componentes principais podem ser consideradas associadas a um conjunto de osciladores independentes cuja entropia pode ser calculada. Os valores de entropia obtidos são da ordem de 17 kcal/mol K, valores que são uma ordem de grandeza superiores aos resultados obtidos em trabalhos anteriores. Adicionalmente, foi comprovado que o termo $-TS$ contribui de forma não desprezível para a energia livre. O termo de energia interna foi estimado a partir da média do conjunto, aplicada às trajetórias de dinâmica molecular de equilíbrio para cada conformação intermediária, como resultado a altura da barreira de energia para a transição foi estimada em 189,75 kcal/mol.

Palavras-chave: Hemaglutinina; Entropia; Abordagem Quase Harmônica; Dinâmica Molecular.

Introducción

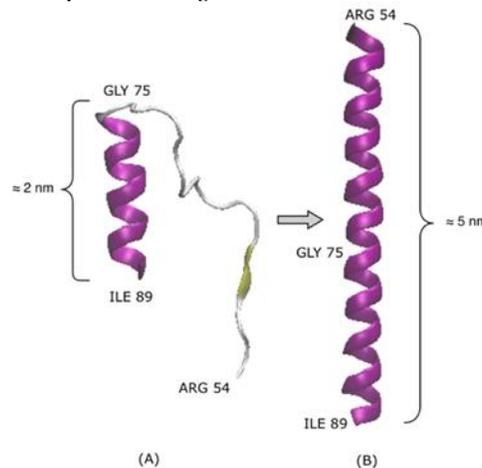
La Hemaglutinina (HA) es una proteína presente en la superficie del virus de la influenza, siendo parte fundamental del mecanismo de adhesión y fusión del virus a la membrana celular (Bullough et al., 1994; Carr y Kim, 1993). La Hemaglutinina está formada por dos subunidades HA1 y HA2. La estructura molecular de la Hemaglutinina ha sido caracterizada completamente de forma experimental (Bullough et al., 1994; Sauter et al., 1992), y la información completa de su composición y estructura molecular se encuentran libremente disponibles en el Protein Data Bank (PDB) (Berman et al., 2000) tanto para la configuración no-fusogénica como para la fusogénica.

Figura 1: (A) La Hemaglutinina, on sus subunidades HA1 y HA2, el lazo 36 se encuentra en la subunidad HA, imagen generada mediante el software VMD a partir del registro 1HTM del PDB. (B) La Hemaglutinina en la superficie del virus de la influenza



De particular interés resulta la dinámica del cambio conformacional que sufre el llamado Lazo 36, el cual corresponde a la región comprendida entre los residuos 54 y 89 de la cadena HA. El modelo con mayor aceptación para la descripción de la transición conformacional es el conocido como Spring-Loaded (Carr y Kim, 1993) el cual plantea que la estructura nativa de la Hemaglutinina está atrapada en un estado metaestable, y que la disminución del pH la desestabiliza produciendo la transición. El cambio ocurre desde su conformación inicial no-fusogénica a pH 7 hasta una conformación final fusogénica a pH 5. Se ha planteado que el movimiento lineal producido por esta transición conformacional, pueda ser aprovechado como un nanomotor lineal (Banta et al., 2007; Dubey et al., 2003).

Figura 2: Estructura Secundaria del Lazo 36: (A) Conformación no-fusogénica a pH 7, imagen generada a partir del registro 1HGF del PDB utilizando el software VMD. (B) Conformación fusogénica a pH 5, imagen generada a partir del registro 1HTM del PDB,



La transición conformacional del lazo 36 de la Hemaglutinina es un proceso que, por su simplicidad, permite estudiar el plegado proteínico. Los detalles atómicos de la transición accesibles directamente con las simulaciones de dinámica molecular, permiten el cálculo de la energía, así como de la entropía involucradas en el proceso. Esta información permitirá establecer relaciones cuantitativas entre estas dos magnitudes durante un proceso de plegado. Adicionalmente, el entendimiento de los mecanismos atómicos asociados a la transición conformacional del lazo 36 de la Hemaglutinina tiene potenciales aplicaciones biomédicas y biotecnológicas.

En un trabajo precedente se utilizó simulaciones TMD para explorar la transición del lazo 36 (Calderón, 2015). A partir de la trayectoria obtenida se aplicó análisis de grupos para identificar posibles conformaciones intermedias, obteniéndose 13 grupos. A partir de estos datos se realizó una descripción estadística del proceso, la cual estimó la entropía con un método simple de conteo de estados accesibles. Por ello se consideró necesario un análisis alternativo de los datos.

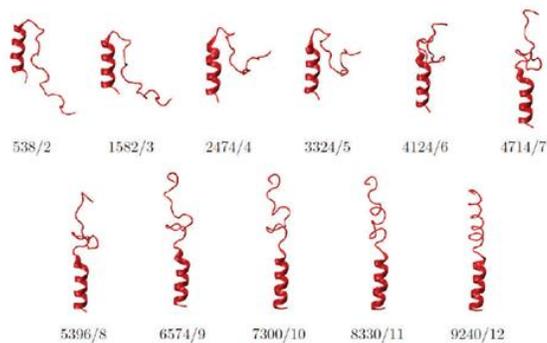
En este trabajo, se realizó un nuevo análisis de los datos, tanto para la trayectoria TMD como para las trayectorias de equilibración de cada conformación intermedia. Para este nuevo análisis en primer lugar se comparó los resultados del método de agrupamiento desarrollado en (Calderón, 2015) con los resultados de un agrupamiento basado en la similitud de estructuras secundarias en el espacio de Ramachandran, posteriormente se estimó la entropía del lazo 36 de la HA utilizando el método de la Aproximación Quasi-Armónica y finalmente caracterizó la barrera energética asociada a la transición conformacional.

Metodología

Estructuras

En el trabajo de Calderón J., se realizó una simulación de dinámica molecular tipo TMD, donde la estructura del lazo 36 en su conformación no-fusogénica se utilizó como estructura inicial, y como estructura final u objetivo se utilizó la estructura fusogénica. Esta simulación produjo una trayectoria compuesta por 10000 estructuras. Las estructuras de la trayectoria TMD fueron sometidas a un análisis de grupos, obteniéndose 13 conformaciones intermedias correspondientes a 11 conformaciones representativas, cuyas estructuras representativas se muestran en la fig. 3, además de las conformaciones correspondientes a la conformación no-fusogénica y fusogénica.

Figura 3: Las 11 estructuras representativas intermedias identificadas en (Calderón, 2015), imagen reproducida del trabajo citado.



Para la simulación de equilibración, se partió de las estructuras representativas de cada conformación, las cuales se dejaron evolucionar mediante dinámica molecular, durante un tiempo total de 6 ns, con condiciones de frontera esféricas. Estas trayectorias son las que se utilizarán para el cálculo de la entropía. Cada trayectoria tiene un total de 12000 estructuras, correspondiendo cada estructura a 20 ps de tiempo en la equilibración.

Mapas de Ramachandran

Los ángulos de Ramachandran (Ramachandran et al., 1968) representan la orientación de los aminoácidos unidos mediante enlaces peptídicos en un polipéptido o proteína. Estos ángulos permiten describir las rotaciones del *backbone o esqueleto* del polipéptido en torno a los enlaces N-C α (llamado ϕ) y C α -C (llamado ψ).

La representación en un plano de ϕ vs ψ , se conoce como mapa de Ramachandran, y permite observar la distribución de los ángulos de torsión en una estructura de la proteína. También proporciona una visión general de las regiones excluidas, que muestran que ciertas rotaciones del polipéptido no están permitidas, debido a las interacciones entre los átomos de los aminoácidos. Finalmente, en el mapa de Ramachandran se pueden identificar regiones correspondientes a motivos de la estructura secundaria de una proteína, tal como las regiones correspondientes a los denominados α -hélices y hojas- β .

El polígono que encierra a la región correspondiente a un α -hélice en el mapa de Ramachandran está determinada por los puntos indicados en la siguiente tabla (Brändén y Tooze, 1999)

Tabla 1: Puntos del polígono que encierra la región correspondiente a un α -hélice en el mapa de Ramachandran (Brändén y Tooze, 1999).

ϕ	ψ
[grados]	[grados]
]]
-180,0	-34,9
-164,3	-42,9
-133,0	-42,9
-109,4	-32,2
-106,9	-21,4

-44,3	-21,4
-44,3	-71,1
-180,0	-71,1

Aproximación Quasi-Armónica

Dado que no existe un método exacto para calcular la entropía de una biomolécula a partir de los datos de dinámica molecular, es necesario realizar una estimación de la misma. La estrategia que se utilizó en este trabajo para estimar la entropía, es la llamada Aproximación Quasi-Armónica (QHA) (Karplus y Kushick, 1981), la cual provee un límite superior a la entropía conformacional de una biomolécula en términos de osciladores armónicos independientes asociados a sus grados de libertad, identificados con un análisis de componentes principales (Jolliffe, 1986).

El Análisis de Componentes Principales es una técnica estadística, utilizada para analizar datos multivariantes. Los objetivos prácticos de la utilización del análisis de componentes principales han sido enumerados en (Jeffers, 1967), siendo de interés examinar la correlación entre las variables de un conjunto de datos, ya que para la aplicación de la aproximación quasi-armónica para el cálculo de la entropía es necesario tener variables no-correlacionadas.

Componentes principales de una macromolécula derivadas de trayectorias de dinámica molecular

A partir de trayectorias de simulaciones de dinámica molecular, en las cuales se han eliminado los átomos del solvente, se puede plantear una matriz de trayectoria R la cual tiene dimensiones $3N_a \times n_f$ siendo N_a el número de átomos y n_f la cantidad de trayectorias de simulación que se consideran para el análisis, es decir:

A partir de trayectorias de simulaciones de dinámica molecular, en las cuales se han eliminado los átomos del solvente, se puede plantear una matriz de trayectoria \mathbf{R} la cual tiene dimensiones $3N_a \times n_f$ siendo N_a el número de átomos y n_f la cantidad de trayectorias de simulación que se consideran para el análisis, es decir:

$$\mathbf{R} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \\ \vdots \\ \mathbf{r}_i \\ \vdots \\ \mathbf{r}_{nf} \end{pmatrix} \quad (1)$$

donde \mathbf{r}_i corresponde a la estructura i -ésima, la cual contiene las coordenadas de cada átomo, arregladas en un vector con $3N_a$ componentes:

$$\mathbf{r}_i = (x_1^i, y_1^i, z_1^i, \dots, x_{N_a}^i, y_{N_a}^i, z_{N_a}^i) \quad (2)$$

Para considerar las masas de los átomos ubicados en cada posición \mathbf{r}_i , se aplicó a la matriz de trayectoria \mathbf{R} la transformación (Schlitter, 1993):

$$\mathbf{R}' = \mathbf{M}^{1/2} \mathbf{R} \quad (3)$$

donde la matriz $\mathbf{M}^{1/2}$ de dimensiones $3N_a \times 3N_a$ viene dada por la matriz diagonal conformada por la raíz cuadrada de las masas de cada átomo.

Sobre esta matriz \mathbf{R}' se aplica el análisis de componentes principales, para lo cual primero se obtiene la matriz de covarianza de \mathbf{R}' :

$$\mathbf{C} = \frac{1}{n} \mathbf{R}' \mathbf{R}'^T \quad (4)$$

Se aplica la descomposición de valor singular (SVD) a la matriz \mathbf{C} , obteniendo un conjunto de $3N_a$ autovalores λ , los cuales debido a la ponderación realizada a través de la matriz \mathbf{M} contienen valores de masa que son combinación de las originales (Schlitter, 1993; Schafer et al., 2000), por lo cual los autovalores pueden representarse de la forma:

$$\lambda_i = m_{eff} \sigma_{PCA}^2 \quad (6)$$

donde cada autovalor λ_i es la desviación estándar de las componentes principales obtenidas, las cuales debido a la ponderación realizada contienen un factor m_{eff} el cual es una función de las masas atómicas $m_1 \dots m_{3N_a}$.

Frecuencias Asociadas a las Componentes Principales

Es un resultado conocido que para un sistema con desviación estándar nula $\langle x \rangle = 0$ y varianza conocida $\langle x^2 \rangle$, el valor máximo de la entropía se presenta cuando las energías y las varianzas para

cada estado son proporcionales, excepto por una constante aditiva. Esta condición es satisfecha por el Hamiltoniano de un oscilador armónico (Schlitter, 1993; Schafer et al., 2000), es decir:

$$H = \frac{1}{2} \left(\frac{p^2}{m} + m\omega^2 x^2 \right) \quad (7)$$

donde p es el momento canónico asociado a x .

Para este tipo de sistemas es de interés encontrar la frecuencia a partir de la varianza $\langle x^2 \rangle$. Para ello se puede utilizar la siguiente aproximación (Schlitter, 1993):

$$m\omega^2 \langle x^2 \rangle = k_B T \quad (8)$$

Esta expresión se cumple también en el régimen clásico, debido al teorema de equipartición.

Una generalización propuesta para el caso n-dimensional es a través de la matriz de covarianza (Schlitter, 1993), de forma que la expresión anterior se puede representar como:

$$m\omega_n^2 \sigma_n^2 = k_B T \quad (9)$$

Las varianzas σ_n^2 pueden estimarse con simulaciones de dinámica molecular del sistema. Este tipo de simulaciones de dinámica molecular, se realizan usualmente en un sistema de coordenadas cartesiano, por lo cual varias de sus coordenadas estarán correlacionadas. Con el fin de obtener una transformación de coordenadas, las cuales sean independientes, se puede aplicar una transformación ortogonal, tal como el análisis de coordenadas principales, las cuales además están ordenadas desde la que contiene la mayor cantidad de varianza del sistema, hasta la que contiene la menor, de forma que:

$$m_{eff} \omega_i^2 \sigma_{i(PC)}^2 = k_B T \quad (10)$$

Esta expresión corresponde a la hipótesis fundamental de la aproximación Quasi-Armónica, la cual considera que los autovalores obtenidos a partir del análisis de componentes principales, corresponden a un conjunto de osciladores armónicos cuánticos no-correlacionados (Meirovitch et al., 2009; Barón et al., 2009).

Desarrollando la ecuación anterior, para cada autovalor se puede obtener una frecuencia ω :

$$\omega_i = \sqrt{\frac{k_B T}{m_{eff} \sigma_{i(PC)}^2}} = \sqrt{\frac{k_B T}{\lambda_i}} \quad (11)$$

La última ecuación asigna un valor de frecuencia a cada autovalor obtenido mediante el análisis de componentes principales.

Estimación de la Entropía

Las trayectorias disponibles, se encuentran en un formato binario (DCD), por lo cual se exportaron a coordenadas del tipo x,y,z , utilizando el software VMD (Humphrey et al., 1996), las coordenadas exportadas se organizaron en una matriz, utilizando el lenguaje de programación Python (Van Rossum y Drake Jr., 1995), la estructura de la matriz es:

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n_a} \\ r_{21} & r_{22} & r_{23} & \cdots & r_{2n_a} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{n_s1} & r_{n_s2} & r_{n_s3} & \cdots & r_{n_s n_a} \end{bmatrix} \quad (13)$$

donde $r_{ij} = (x_{ij}, y_{ij}, z_{ij})$, i indica el número de estructura y j el número de átomo, de forma que la matriz tiene dimensiones $3n_a \times n_s$, siendo n_a el número de átomos, y n_s el número de estructuras en cada trayectoria de equilibración.

Sobre la trayectoria de dinámica molecular de equilibración representada por esta matriz se aplicó el análisis de componentes principales (PCA), mediante un script en R (R Core Team, 2015). A partir de los autovalores (desviaciones estándar) de las componentes principales obtenidas, se encontraron las frecuencias correspondientes a través de la Aproximación Quasi-Armónica, con estas frecuencias se estimó la entropía mediante la expresión (Schlitter, 1993; Schafer et al., 2000):

$$S_Q = k_B \sum_{i=1}^{3N} \left[\frac{1}{k_B T} \left(\frac{\hbar}{e^{\frac{\hbar \omega_i}{k_B T}} - 1} \right) - \ln \left(1 - e^{-\frac{\hbar \omega_i}{k_B T}} \right) \right] \quad (12)$$

Energía libre de gibbs y barrera energética

Para estimar la energía libre de Gibbs ΔG , se requiere de dos componentes, la diferencia de energía interna ΔU , y la contribución de la entropía $-T\Delta S$. La diferencia de energía interna puede obtenerse a partir de los valores de energía obtenidos durante las simulaciones de dinámica molecular, y la contribución de la entropía se puede obtener a partir del valor calculado con el método expuesto en la sección anterior.

Para describir la barrera energética, se probaron coordenadas de reacción, que describan el comportamiento energético durante la transición. Como candidato a coordenada de reacción se utilizó la primera coordenada principal resultante de la aplicación del PCA. Otra posible coordenada es el RMSD entre las posiciones de los átomos en las estructuras, calculado con

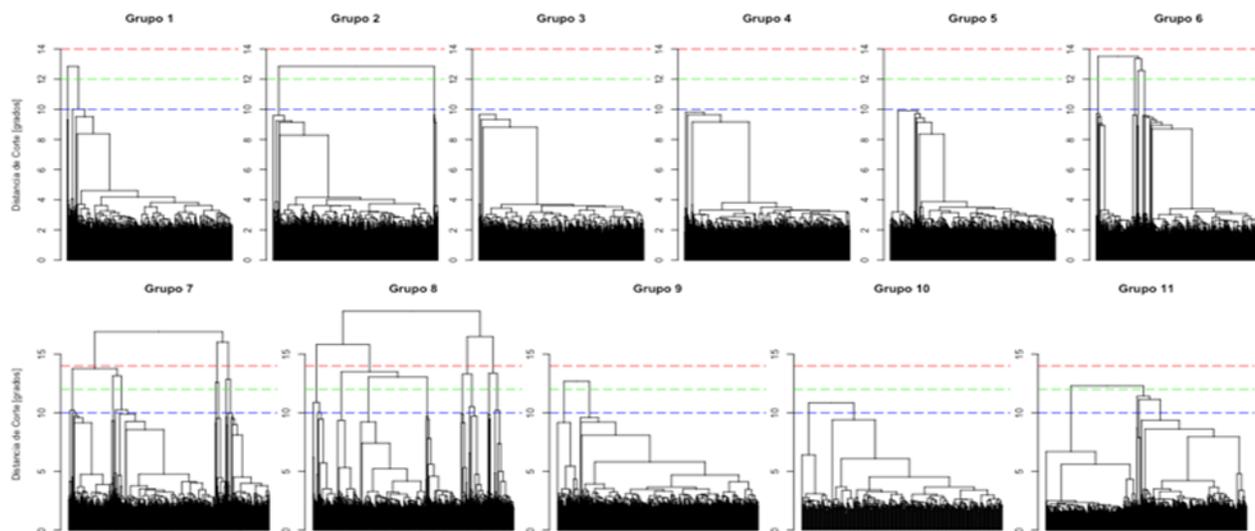
respecto a la estructura inicial (no-fusogénica), y a la estructura final (fusogénica), el RMSD se calculó mediante la herramienta "RMSD Calculator" disponible en el software VMD (Humphrey et al., 1996).

Resultados y discusión

Análisis de Grupos en el Espacio de Ramachandran

Para explorar la posibilidad de la existencia de grupos adicionales, se realizó un nuevo análisis de grupos sobre cada conjunto de estructuras de la trayectoria TMD clasificadas dentro de un grupo de acuerdo al criterio del RMSD, utilizando como parámetro la distancia euclídea de los ángulos de torsión entre las estructuras del grupo. De forma que se obtuvieron 11 dendogramas, correspondientes a cada grupo, los cuales se muestran en la fig. 4.

Figura 4: Dendogramas para cada grupo.



A distancias de corte representativas, tales como 10, 12 y 14 grados, se obtiene un número de grupos comparable para cada dendograma. Los resultados de este análisis se reportan en la siguiente tabla.

Tabla 2: Número de grupos para distancias de corte específicas.

Grupo	Distancia de Corte		
	14	12	10
Número de Grupos			
1	1	2	2
2	1	2	2
3	1	1	1
4	1	1	1
5	1	1	1
6	1	5	5
7	3	7	9
8	4	4	11
9	1	2	2
10	1	1	2
11	1	2	5

Se puede comprobar que los grupos 3,4 y 5 no tienen subgrupos para las distancias de corte consideradas, por ello se puede afirmar que están bien definidos de acuerdo a los dos criterios utilizados (RMSD y distancia en el espacio de Ramachandran). Por otra parte, los grupos 7 y 8 son los que más subgrupos presentan para las distancias de corte consideradas.

Entropía conformacional

Frecuencias de los osciladores armónicos

En la fig. 5, se muestran las frecuencias de los osciladores armónicos en la aproximación Quasi-Armónica, correspondientes a las componentes principales, para cada una de las conformaciones intermedias.

Figura 5: Frecuencias de los osciladores armónicos de la aproximación quasi-armónica, correspondientes a las componentes principales.

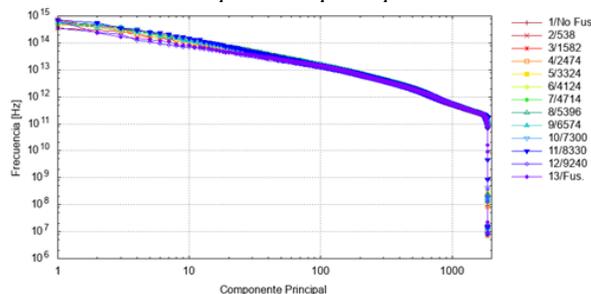
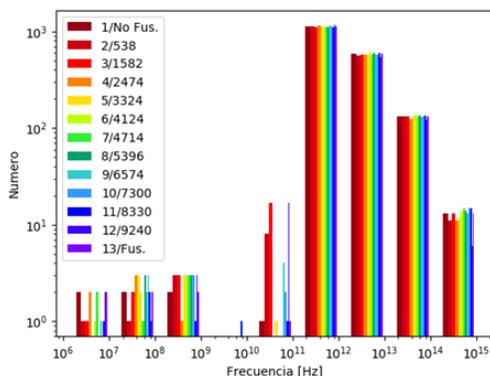


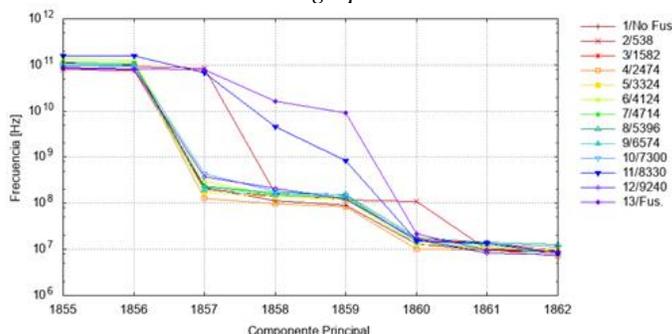
Figura 6: Histograma de las frecuencias de los osciladores armónicos correspondientes a las componentes principales, cada color está asociado a una conformación.



A partir de los histogramas, se puede observar que las frecuencias de los osciladores armónicos, se encuentran mayoritariamente en el rango entre 10^{11} [Hz] y 10^{14} [Hz]. Considerando que la frecuencia de oscilación de las vibraciones moleculares se encuentran en el orden de entre 10^{13} [Hz] y 10^{14} [Hz] (Barth, 2007; Dill y Broomberg, 2010) se puede observar que las frecuencias obtenidas son cercanas a este rango.

Por otra parte, en los histogramas se puede apreciar la presencia de frecuencias en un rango bajo, las cuales corresponden a las últimas frecuencias obtenidas mediante la aproximación quasi-armónica, como se puede observar en la fig. 7.

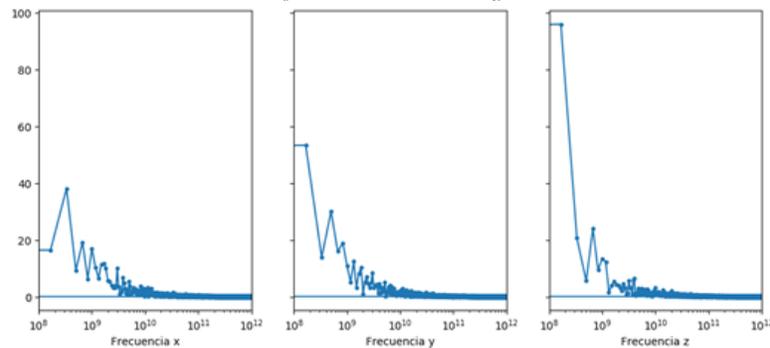
Figura 7: Frecuencias de los últimos osciladores armónicos, correspondientes a las últimas componentes principales, los 6 últimos valores presentan una variación significativa con respecto a los anteriores, para todos los grupos.



Para explicar la presencia de estas frecuencias, en primera instancia se analizó la posibilidad de existencia de oscilaciones del centro de masa de la molécula, para lo cual se calcularon las

posiciones del centro de masa a lo largo de la trayectoria TMD. Para caracterizar estas oscilaciones del centro de Masa, se procedió a obtener la transformada de Fourier mediante el método FFT de las posiciones del centro de masa, considerando además que dado que se tiene en total 12000 estructuras correspondientes a 6 [ns] de tiempo de simulación, el tiempo de muestreo sería de 500 [ps], lo cual nos permite obtener el rango de frecuencias para la FFT. Los resultados de este análisis se muestran en la fig. 8.

Figura 8: Transformada de Fourier de las coordenadas del centro de masa para la trayectoria de equilibración de la conformación no-Fusogénica

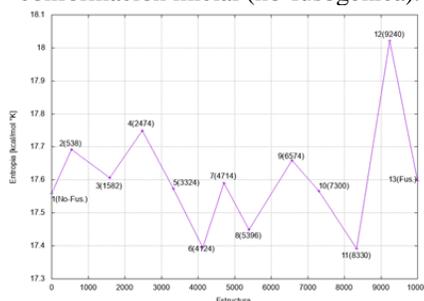


A partir de la figura anterior, se puede observar que la mayor componente de frecuencia se encuentra en el rango entre 10^8 [Hz] y 10^9 [Hz], para cada coordenada analizada, lo cual permite explicar 3 de los 6 valores de la fig. 7.

Entropía Absoluta

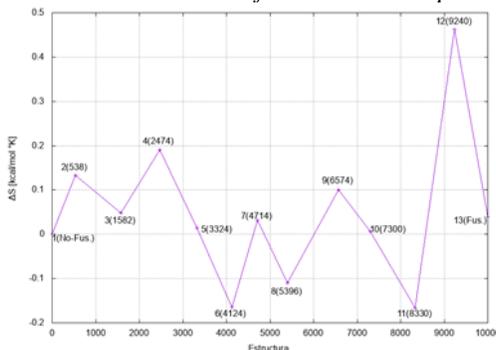
Para estimar la entropía, a partir de las frecuencias de los osciladores armónicos de la aproximación quasi-armónica, se utilizó la ecuación (ec. 12). Este procedimiento se realizó para cada conformación intermedia, obteniéndose en total 13 valores de entropía, los cuales se encuentran en el orden de 17 [kcal/mol °K].

Figura 9: Entropía absoluta obtenida mediante la Aproximación Quasi-Armónica, el valor máximo de entropía 18.021 [kcal/mol °K] ocurre para la conformación 12 cuya estructura representativa es la 9240, en tanto que disminuye para la conformación fusogénica, aproximándose al valor 17.597 [kcal/mol °K] semejante al de la conformación inicial (no-fusogénica).



A partir de los valores obtenidos de entropía absoluta, se puede determinar la diferencia de entropía entre conformaciones, obteniéndose 12 valores de diferencia de entropía, y tomando como valor de referencia el primer valor de entropía, el cual corresponde a la conformación no-fusogénica.

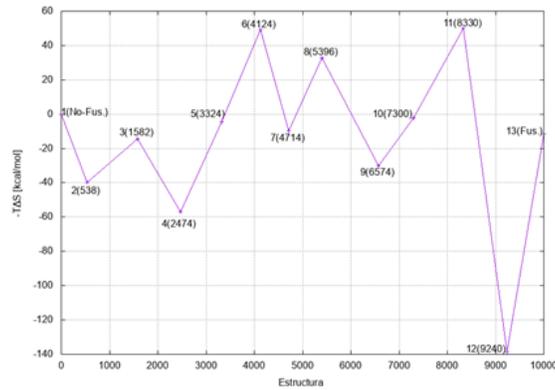
Figura 10: Diferencia de entropía entre conformaciones, tomando como referencia la entropía de la conformación inicial (no-fusogénica), el valor máximo de la diferencia de entropía es de 0.4622 [kcal/mol °K]



Energía Libre

A partir de la estimación de la diferencia de entropía, cuyo resultado se presentó en la sección anterior, se puede determinar el término de contribución de la entropía a la energía libre $-T\Delta S$, dada la temperatura $T=300 K$ a la que se realizó la dinámica molecular.

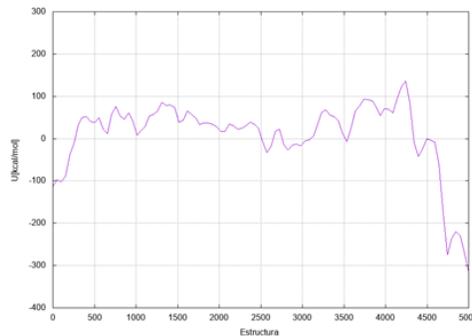
Figura 11. Contribución de la entropía a la energía libre.



Energía interna en la trayectoria TMD

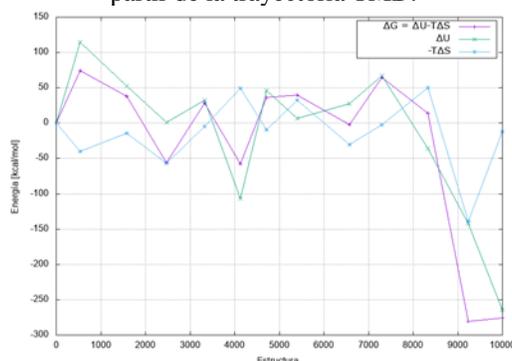
La energía interna se estimó utilizando el plugin “Namd-Energy” incluido en el software VMD (Humphrey et al., 1996), sobre la trayectoria TMD al igual que en (Calderón, 2015). Como resultado se obtienen 10000 valores de energía, correspondientes a cada estructura de la trayectoria, como se muestra en la siguiente figura:

Figura 12. Energía interna U , estimada a partir de la trayectoria TMD.



Para el análisis se tomó el valor de la energía interna para la estructura representativa de cada grupo, añadiéndole la contribución de la entropía, para así obtener la energía libre de Gibbs; los resultados de esta aproximación se muestran en la siguiente figura:

Figura 13: Energía libre de Gibbs, estimada utilizando la energía interna de la estructura representativa, obtenida a partir de la trayectoria TMD.



Se puede observar, que el valor máximo de energía libre 74 [kcal/mol] se presenta para el grupo $I(538)$, y se mantiene relativamente estable, después se reduce drásticamente, hasta el mínimo $-280.77 \text{ [kcal/mol]}$ en la estructura correspondiente al grupo $II(9240)$, después incrementa levemente hasta la estructura correspondiente a la conformación fusogénica.

Estimación utilizando las trayectorias de equilibración para cada conformación intermedia

Para estimar la energía interna, se utilizaron las trayectorias de equilibración para cada grupo, suponiendo que todas las estructuras son equivalentes y por tanto pertenecientes a una conformación, de forma que cada trayectoria se puede considerar como un ensamble, y por tanto la energía interna se puede estimar mediante la siguiente expresión (Thijssen, 2007):

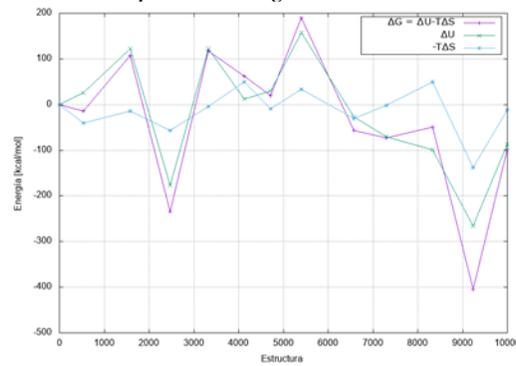
$$\langle E \rangle = \frac{1}{M} \sum_{n=1}^M E^n \quad (14)$$

donde M es el total de estructuras en el ensamble (trayectoria), y n es la n -ésima estructura.

Para el análisis, resulta conveniente tomar un subconjunto de estructuras de la trayectoria de equilibración, las que tengan un valor de RMSD estable, para lo cual resulta conveniente analizar el RMSD de la trayectoria de equilibración de la estructura no-Fusogénica, por tanto, se decidió tomar las últimas 2000 estructuras de las trayectorias de equilibración para cada grupo para el cálculo de la energía interna, considerando además que $U = \langle E \rangle$.

En la siguiente figura, se muestra la energía libre de Gibbs ΔG además de las componentes que la constituyen: la energía interna ΔU , y la contribución de la entropía $-T\Delta S$.

Figura 14: Energía libre de Gibbs, utilizando el promedio obtenido a partir de las trayectorias de equilibración para la energía interna.

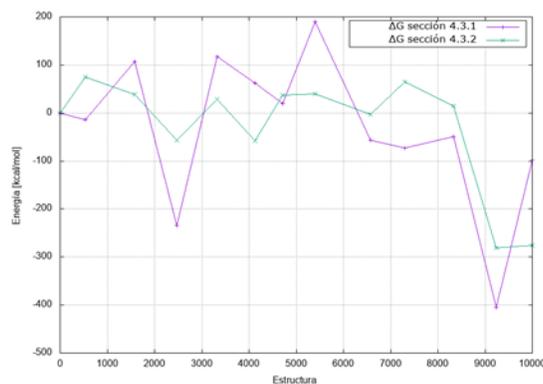


A partir de la figura, se puede comprobar que el valor máximo de energía libre $189.75 [kcal/mol]$ se presenta para el grupo 6(5396), y se tienen dos valores mínimos, uno de $-234.55 [kcal/mol]$ correspondiente al grupo 3(2474), y un mínimo global de $-404.72 [kcal/mol]$ correspondiente al grupo 11(9240), el cual por lo tanto corresponde a la conformación intermedia más estable.

Comparación de los valores de Energía Libre obtenidos mediante los dos métodos

La energía libre obtenida directamente a partir del valor de la energía libre de la estructura representativa en la trayectoria TMD, fluctúa considerablemente en comparación con la obtenida a partir del promedio sobre el ensamble generado a partir de las trayectorias de equilibración.

Figura 15: Comparación de la energía libre de Gibbs, obtenida mediante los dos métodos anteriores.



Barrera Energética

A continuación se presentan los valores de energía libre asociados a las diferentes conformaciones en función de la primera coordenada principal del PCA, el $RMSD_0$, y el $RMSD_f$.

Figura 16: PCA, obtenido de la aplicación del análisis de componentes principales a las estructuras de la trayectoria TMD, se ha aplicado la transformación $PCA' = \max(PCA) - PCA$ de forma que se muestre el incremento de la variable.

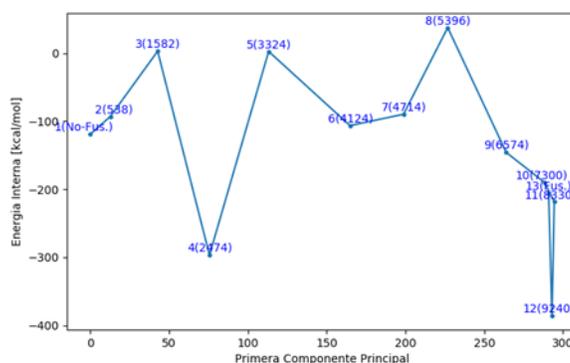
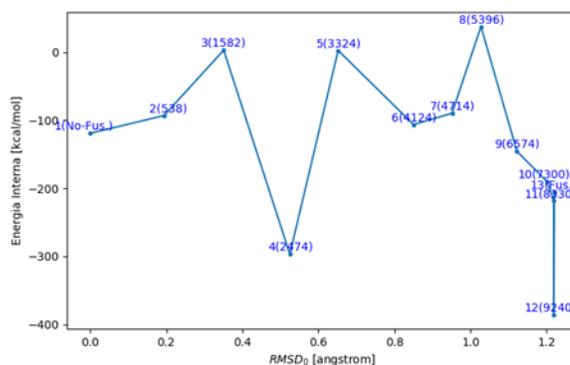
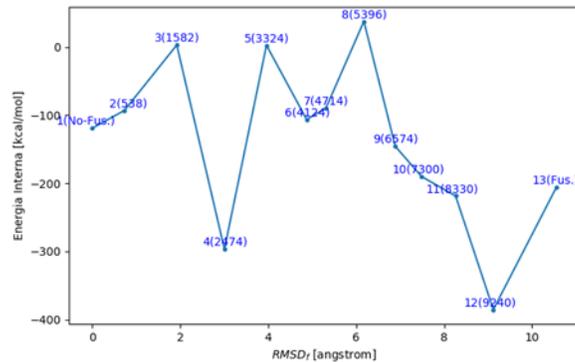


Figura 17: $RMSD_0$, este parámetro fue calculado con respecto a la estructura no-fusogénica.



Tanto para el PCA como para el $RMSD_0$ como coordenadas de reacción, se puede ver el mínimo local de energía libre correspondiente a la cuarta conformación intermedia, en tanto que las conformaciones {10, 11, 12 y 13} se encuentran muy próximas en el eje correspondiente tanto al PCA como al $RMSD_0$, por lo cual se puede ver una correlación entre las dos posibles coordenadas de reacción.

Figura 18: $RMSD_f$ calculado con respecto a la estructura fusogénica, se ha aplicado la transformación: $RMSD'_f = \max(RMSD_f) - RMSD_f$, de forma que se muestre el incremento de la variable.



Por otra parte, para el $RMSD_f$ se puede observar que las conformaciones {10, 11, 12 y 13} están mejor espaciadas, y se manifiesta una posible transición natural entre las conformaciones {8, 9, 10, 11, 12}.

Conclusiones

En este artículo se ha investigado la evolución de la entropía para la transición conformacional del lazo 36 de la Hemaglutinina, para lo cual primero se evaluó la pertinencia de las conformaciones determinadas mediante análisis de grupos en el espacio de Ramachandran, posteriormente se obtuvieron frecuencias a través de la aproximación quasi-armónica con las cuales se pudo estimar la entropía, y a la vez estimar la contribución de entrópica de la energía libre, y con ello describir la barrera energética asociada a la transición conformacional.

El análisis de grupos en el plano de Ramachandran para cada grupo, muestra que los grupos 3, 4 y 5 no tienen subgrupos para las distancias de corte consideradas, por ello se puede afirmar que están bien definidos de acuerdo a los dos criterios utilizados ($RMSD$ y distancia en el espacio de Ramachandran). Por otra parte los grupos 7 y 8 son los que más subgrupos presentan para las distancias de corte consideradas y reafirma la posibilidad de existencia de grupos adicionales.

Las frecuencias obtenidas a través de la aproximación quasi-armónica se encuentra en el rango entre 10^6 y 10^{15} [Hz], con la mayoría de frecuencias en el rango entre 10^{11} y 10^{14} [Hz], lo cual permite establecer una relación directa entre las frecuencias encontradas y el comportamiento molecular, ya que coinciden con el rango de las frecuencias de vibración molecular. Por otra parte, se comprobó que las frecuencias de menor valor, corresponden a modos colectivos de vibración,

entre los cuales se analizó la oscilación del centro de masa, y se comprobó que esta oscilación tiene un máximo de frecuencia en torno a 10^8 [Hz]. De esta forma se puede verificar que la aproximación quasi-armónica, produce resultados consistentes con los resultados experimentales (Bullough et al., 1994).

La estimación de la entropía a través de la aproximación quasi-armónica, provee valores de entropía del orden de $S \approx 17$ [kcal/mol °K], y valores de diferencia de entropía con un máximo de $\Delta S = 0.8$ [kcal/mol °K], lo cual produce una contribución energética de la entropía máxima de $-T\Delta S = -240$ [kcal/mol], este valor difiere en 2 órdenes de magnitud del valor reportado en trabajos anteriores (Calderón, 2015); sin embargo la distribución de los valores obtenidos, siguen una tendencia similar, con un cambio abrupto de $-T\Delta S$ en la conformación *I2* cuya estructura representativa es la 9240.

Para la descripción de la barrera energética se probaron 3 posibles coordenadas de reacción, tanto para la primera coordenada principal como para el $RMSD_0$ se tiene un comportamiento similar, por lo cual se puede ver una posible correlación entre las dos coordenadas de reacción. Por otra parte, la coordenada que mejor describe el sistema es $RMSD_f$, ya que permite apreciar la barrera energética durante todas las etapas de la transición.

En el futuro, sería relevante utilizar las estructuras representativas encontradas mediante análisis de grupos, considerando como parámetros tanto el RMSD como la distancia euclídea en el plano de Ramachandran, ya que los resultados de utilizar los dos criterios se pueden considerar como complementarios. Esto permitiría además encontrar valores adicionales de entropía, utilizando los mismos datos de simulaciones de dinámica molecular para cada conformación, ya que se podría aplicar la aproximación quasi-armónica sobre particiones de la trayectoria en subconjuntos de estructuras, que correspondan a los grupos adicionales. Otro aspecto a tomar en cuenta es que, en la aproximación quasi-armónica, la componente de frecuencias bajas contribuyen mayoritariamente a la entropía, por ello resultaría relevante ampliar el análisis espectral presentado en la sección 3.2.1, de modo que incluya no solo la oscilación del centro de masa, sino también otros aspectos, por ejemplo, la rotación de la molécula.

Referencias

1. Banta, Scott et al. (2007). «Engineering protein and peptide building blocks for nanotechnology». *Journal of Nanoscience and Nanotechnology*, 7(2), 387-401.

2. Baron, Riccardo et al. (2009). «Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: correction terms and convergence properties». *Journal of Chemical Theory and Computation*, 5(12), 3150-3160.
3. Barth, Andreas (2007). «Infrared spectroscopy of proteins». *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 1767(9), 1073-1101.
4. Berman, Helen M et al. (2000). «The protein data bank». *Nucleic Acids Research*, 28(1), 235-242.
5. Brändén, Carl-Ivar y John Tooze (1999). «Introduction to Protein Structure». Taylor & Francis.
6. Bullough, Per A et al. (1994). «Structure of influenza haemagglutinin at the pH of membrane fusion». *Nature*, 371(6492), 37-43.
7. Calderón, Jaime Rodrigo (2015). «Descripción estadística del proceso de plegamiento del lazo 36 de la cadena AH2 de la Hemaglutinina». Tesis EPN.
8. Carr, Chavela M. y Peter S. Kim (1993). «A spring-loaded mechanism for the conformational change of influenza hemagglutinin». *Cell*, 73(4), 823-832.
9. Dill, Ken y Sarina Bromberg (2010). «Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience. » Garland Science.
10. Dubey, A et al. (2003). «Viral protein linear (VPL) nano-actuators», Third IEEE Conference on Nanotechnology 2003, 1, 140-143.
11. Humphrey, William et al. (1996). «VMD: visual molecular dynamics». *Journal of molecular graphics*, 14(1), 33-38.
12. Jeffers, J. N. R. (1967). «Two Case Studies in the Application of Principal Component Analysis». *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 16(3), 225-236.
13. Jolliffe, Ian T. (1986). «Principal Component Analysis. » Springer.
14. Karplus, Martin y Joseph N Kushick (1981). «Method for estimating the configurational entropy of macromolecules». *Macromolecules*, 14(2), 325-332.
15. Meirovitch, Hagai et al. (2009). «Methods for calculating the entropy and free energy and their application to problems involving protein flexibility and ligand binding». *Current Protein and Peptide Science*, 10(3), 229-243.

16. Morales, Hernán y Marco Bayas (2013). «Efecto de mutaciones puntuales en las transiciones conformacionales de un nanomotor biológico». *Revista Politécnica*, 32(3), 63-70.
17. Numata, J et al. (2007). «Conformational entropy of biomolecules: beyond the quasi-harmonic approximation. » *Genome informatics. International Conference on Genome Informatics*, 18, 192.
18. Phillips, James C et al. (2005). «Scalable molecular dynamics with NAMD». *Journal of computational chemistry*, 26(16), 1781-1802.
19. R Core Team (2015). «R: A Language and Environment for Statistical Computing. » R Foundation for Statistical Computing. Vienna, Austria.
20. Ramachandran, G. y V. Sasisekharan (1968). «Conformation of polypeptides and proteins». *Advances in protein chemistry*, 23, 283-437.
21. Sauter, Nicholas K. et al. (1992). «Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography». *Biochemistry*, 31(40), 9609-9621.
22. Schäfer, Heiko et al. (2000). «Absolute entropies from molecular dynamics simulation trajectories». *The Journal of Chemical Physics*, 113(18), 7809-7817.
23. Schlitter, Jürgen (1993). «Estimation of absolute and relative entropies of macromolecules using the covariance matrix». *Chemical Physics Letters*, 215(6), 617-621.
24. Thijssen, Jos (2007). «Computational physics». Cambridge university press.
25. Van Rossum, Guido y Fred L Drake Jr (1995). «Python reference manual». Centrum voor Wiskunde en Informatica Amsterdam.
26. Zhou, Yu et al. (2014). «Exploring the early stages of the pH-induced conformational change of influenza hemagglutinin». *Proteins: Structure, Function, and Bioinformatics*, 82(10), 2412-2428.