

Una exploración a la relación estructura, dinámica y función en proteínas simétricas

J. C. Higuera Calderón*, KN. Ramos G[†].

Introducción al estado sólido, Departamento de física, Universidad Nacional de Colombia, Bogotá.

4 de mayo de 2021

Resumen

En este trabajo se realizó una comparación de la dinámica en proteínas. En particular se estudió la semejanza en las propiedades vibracionales para proteínas con igual y distinta simetría y función biológica. Se encontró que tanto las simetrías como la función biológica son propiedades que determinan semejanzas de las propiedades vibracionales.

1. Marco teórico

De forma previa a hablar de los aspectos más específicos en los que se centra el proyecto, consideramos importante explicar qué son las proteínas y conceptos relacionados con estas, es decir ¿cómo se producen? y ¿qué funcionalidad tienen?

1.1. Proteínas

Las proteínas son moléculas orgánicas formadas por bloques conocidos como aminoácidos, los cuales están conformados por tres partes, un grupo amino, un grupo carboxilo y una parte variable que se conoce como cadena principal, es la cadena principal lo que diferencia un aminoácido de otro y le confiere propiedades químicas específicas a cada uno.

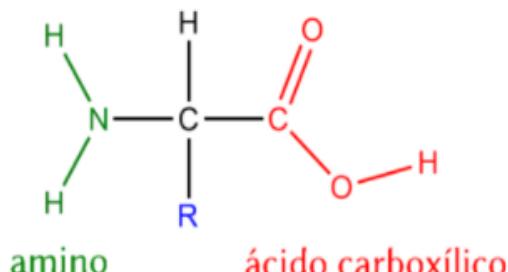


Figura 1: Aminoácido [1]

Los aminoácidos se unen mediante enlaces entre el grupo carboxilo y el grupo amino, estos enlaces se conocen como enlaces peptídicos y permiten a los aminoácidos conformar cadenas lineales, los péptidos. Como consecuencia de las propiedades de los grupos funcionales presentes, los péptidos se pliegan formando estructuras que se conocen como estructuras secundarias, de las cuales las más comunes son la hélice alfa y la conformación beta.

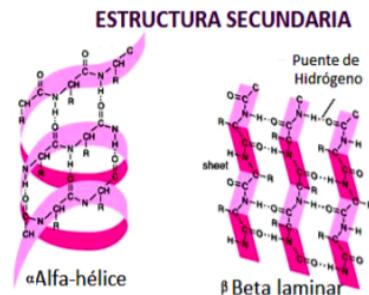


Figura 2: Alfa hélice y conformación beta [2]

A su vez, estas estructuras secundarias se unen y pliegan entre ellas para conformar estructuras terciarias, las cuales corresponden a la distribución tridimensional de todos los átomos que constituyen la proteína, por esta razón es posible afirmar que las funciones biológicas derivan de la estructura terciaria, puesto que la disposición espacial de los grupos funcionales de la proteína condicionan su capacidad de interacción con otras moléculas.

De la discusión anterior se sigue que para producir proteínas con funciones biológicas específicas es necesario darles una forma tridimensional específica, para lo cual se debe disponer ciertos aminoácidos en los péptidos con un orden determinado; el conservar la secuencia de aminoácidos de estas cadenas lineales es la principal función del ADN. A continuación delinearemos el proceso por el cual se producen las cadenas lineales de aminoácidos a partir del ADN

El ADN es una macromolécula bicatenaria de gran extensión formada por nucleótidos, que no son más que

*jchiguerac@unal.edu.co
†kramosg@unal.edu.co

moléculas orgánicas producto de la unión de una desoxirribosa, un grupo fosfato y una base nitrogenada.

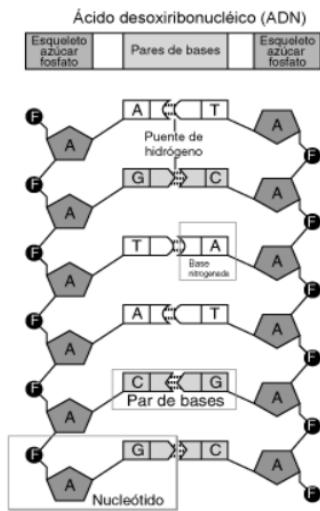


Figura 3: Nucleótidos [3]

Mientras la desoxirribosa y el grupo fosfato son constantes a lo largo de todo el ADN las bases nitrogenadas no, de tal forma que la secuencia de aminoácidos se encuentra codificada en la secuencia de bases nitrogenadas.

En la primera parte de este proceso conocido como transcripción, a partir de enzimas (moléculas capaces de catalizar reacciones químicas) se traduce la secuencia de bases nitrogenadas del ADN al ARN mensajero, el cual tiene la tarea de llevar la información de la secuencia de bases nitrogenadas a los ribosomas, ya que los ribosomas son las estructuras encargadas de la traducción de bases nitrogenadas a aminoácidos; esta traducción asocia tripletas de bases nitrogenadas con uno de los 20 aminoácidos posibles, esto tiene como consecuencia que la traducción es degenerada, es decir tripletas distintas de bases nitrogenadas pueden ser traducidas a un mismo aminoácido, asunto importante en el estudio de mutaciones.

Las proteínas son fundamentales ya que sustentan prácticamente todas las funciones de los seres vivientes, desde las redes de señalización molecular que funcionan en base a redes de interacción proteína - proteína, redes metabólicas en las cuales sus actores principales son proteínas capaces de catalizar reacciones químicas, redes de expresión génica en donde participan proteínas conocidas como factores de transcripción, en el sistema inmune como anticuerpos y también son usadas para fines estructurales.

1.2. Dinámica y función en proteínas

Para estudiar la funcionalidad biológica de proteínas, es importante entender la relación entre su estructura y su función, el puente que conecta estos dos aspectos es la dinámica; así será fundamental entender como la estructura de una proteína se relaciona con su dinámica y a su vez está dinámica con la función biológica.

En el pasado muchos de estos estudios estaban limitados a descripciones cualitativas y limitados a un reducido número de estructuras estáticas, sin embargo con la cuantificación de los datos experimentales está en constante aumento, en especial desde la introducción de la dimensión temporal en los estudios de rayos X y resonancia magnética nuclear[4]. En este contexto se han desarrollado métodos computacionales para complementar los datos estructurales y proveer de modelos para la dinámica de biomoléculas.

Respecto a esta relación entre dinámica y función se ha encontrado evidencia interesante, de la cual aquí mencionaré sólo dos:

1. Fluctuaciones alrededor del equilibrio gobiernan la funcionalidad biológica [5].
2. Dado que la selección natural actúa sobre la función biológica, entonces es posible que la dinámica de proteínas esté más conservada a través de la evolución que la estructura [4].

1.3. Simetrías en proteínas

La simetría hace referencia a características de una estructura matemática que se mantienen invariantes bajo transformaciones de determinado tipo. De manera más formal si tenemos un conjunto K de elementos, junto a un grupo de transformaciones G que pueden actuar sobre los elementos de K , de forma tal que cualquier $g \in G$ cumpla con $(g \in G : k \rightarrow k') \forall k, k' \in K$, entonces se dice que hay una simetría $g \in G$ sobre $k \in K$ si se cumple lo siguiente [6]:

$$\forall g \in G : g(k) = k$$

La ciencia desde sus principios ha formulado teorías para entender el funcionamiento de varios sistemas mediante la comprensión de las simetrías, desde las cuerdas vibrantes hasta los modelos que explican parte de la estructura del sistema solar. En la biología también se ha empleado este uso de simetrías, en especial cuando hablamos de la parte molecular, por ejemplo, una simetría estructural se presenta en la forma de doble hélice del ADN reportadas por Watson y Crick en 1953 en donde esta se puede dividir en dos partes que si bien

presentan diferencias en su composición química poseen una estructura semejante geométricamente. [7]

Se ha encontrado que la gran mayoría de las proteínas solubles y las ligadas a las membranas en los seres vivos forman complejos oligoméricos (Entiéndase por oligómero como una molécula formada por varias componentes idénticas) los cuales presentan varias sub-unidades que resultan idénticas [7]. Como ejemplos de la relevancia de las cadenas ologoméricas en todos los organismos tenemos que: en 1929 Svedberg publicó un artículo bajo el nombre de "Mass and size of protein molecules" [8] en donde se le atribuye a la comprobación de que las proteínas se encuentran compuestas por sub-unidades de carácter discreto; o también podemos ver que en 1991 Goodsell DS cuantificó la prevalencia de proteínas oligoméricas en las células basado en la concentración de proteínas solubles en el organismo *Escherichia coli* en el artículo "Inside a living cell." [9].

Todos los aminoácidos menos la glicina tienen imágenes especulares de su estructura de tal forma que están etiquetados como L-aminoácidos y R-aminoácidos para distinguir estas imágenes. Se ha encontrado que los aminoácidos que forman las proteínas de los organismos son solamente L-aminoácidos, una de las mayores consecuencias de la adopción exclusiva de L-aminoácidos en las proteínas de los organismos es que las proteínas oligoméricas adoptan simetrías enantiomórficas mientras que las simetrías de inversión no se permiten [7]. Ejemplos de esto han sido observados naturalmente en varias proteínas, de hecho, todos los grupos puntuales cristalográficos se han observado; la selección de un grupo particular de simetría puede llegar a tener un efecto notable sobre la funcionalidad y estabilidad de un complejo proteico.

A continuación mostramos los diferentes tipos de proteínas según el grupo en el que se pueden clasificar.

Grupos cíclicos: Los grupos cíclicos tienen un solo eje de simetría rotacional, aquí podemos encontrar el grupo de simetrías C1 en donde están las proteínas monoméricas y el grupo de simetrías C2 en donde están las proteínas diméricas; estas proteínas se encuentran implicadas en funciones que requieren direccionalidad como la interacción de membranas [7], algunos ejemplos:

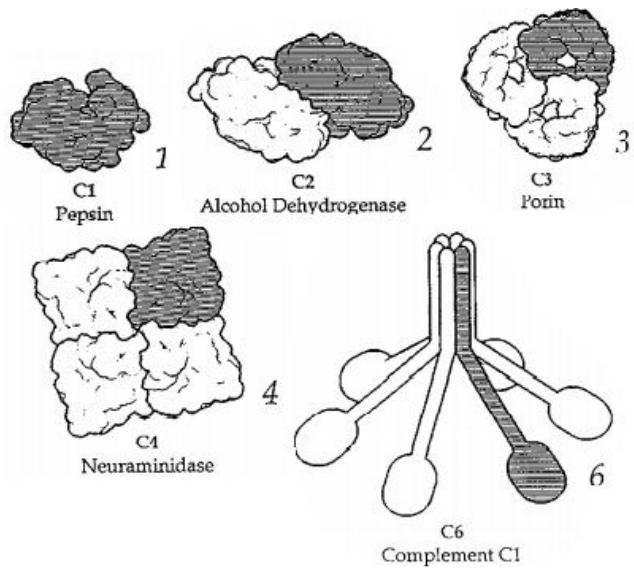


Figura 4: Proteínas pertenecientes al grupo cíclico, tipos C1, C2, C3, C4, C6.

Grupos diedrales: Este grupo es caracterizado porque las proteínas tienen un eje de simetría rotacional y otro perpendicular con simetría $n = 2$, las proteínas que se encuentran dentro de este grupo son comúnmente clasificadas como enzimas solubles citoplasmáticas; las opciones de estabilidad y de interacción de estas moléculas son mayores presentándose con más regularidad que en las proteínas que se encierran dentro del grupo cíclico incluso si tienen el mismo número de sub-unidades [7], algunos ejemplos:

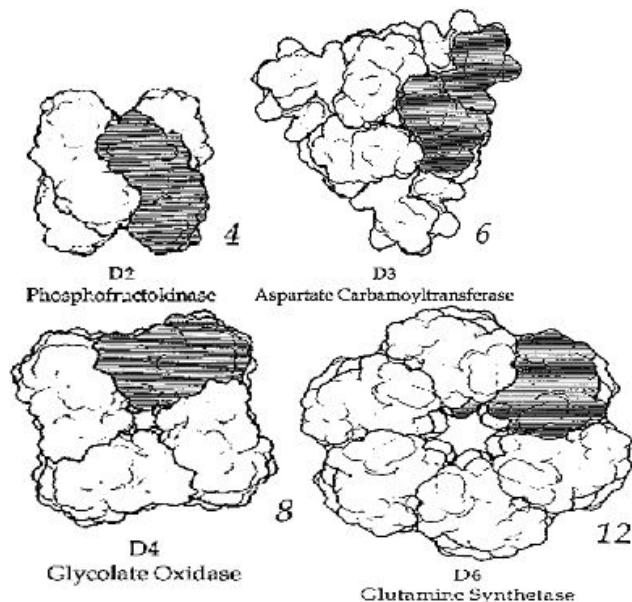


Figura 5: Proteínas pertenecientes al grupo diedral, tipos D4, D6, D8, D12.

Grupos cúbicos: Las simetrías cubicas contienen un eje de simetría de orden $n = 3$ combinado con otro eje rotacional no-perpendicular en donde tenemos tres posibilidades, la primera es la tetraédrica con ejes rotacionales $n = 3$ y $n = 2$, la segunda es la octaédral con ejes rotacionales $n = 3$ y $n = 4$, y por ultimo tenemos la icosaédral en donde hay dos ejes rotacionales con $n = 3$ y $n = 5$. Las proteínas clasificadas dentro del grupo cubico se encuentran en primer lugar especializadas en roles relacionados con el almacenamiento y transporte como lo sugiere Crick y Watson en [10], a continuación tenemos algunos ejemplos:

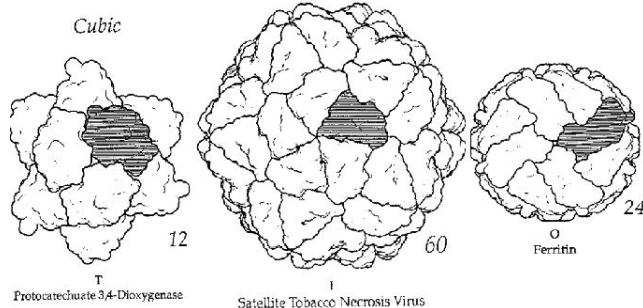


Figura 6: Proteínas pertenecientes al grupo cúbico, tipos C12, C24, C60.

1.4. Ventajas funcionales de las proteínas con simetrías

Autores como Klotz en trabajos relacionados con la estructura cuaternaria de algunas proteínas [11] han comentado acerca de una ventaja selectiva que dirige la evolución de proteínas de carácter monomérico hacia un carácter oligomérico debido a que se observa que estas ultimas son aún más comunes. Esta pregunta es abordada por Monod en una publicación en donde se habla de la simetría en proteínas y de cómo esta se relaciona con ciertas funciones biológicas [12], aquí Monod menciona que estas fuerzas que dirigen la evolución de proteínas monoméricas en oligomericas están relacionadas con la estabilidad que se produce cuando el complejo es oligomérico; dejando de un lado el detalle del por qué se producen estructuras proteicas con cierta simetría y de carácter oligomérico abordamos cuestiones relacionadas con la funcionalidad y forma de estas ultimas debido a que son las más representativas cuando se habla de simetría en proteínas:

Debido a que las proteínas de carácter oligomérico son consideradas estructuras relativamente largas en comparación con las estructuras monoméricas empezaremos hablando de las ventajas que poseen las proteínas largas en comparación con las que poseen cadenas de aminoácidos más cortas; **primera ventaja**, muchas proteínas tienen un función que requiere el ensamblaje de muchos aminoácidos para garantizar la estabilidad de la estruc-

tura; **segunda ventaja**, las interacciones alostéricas en donde surgen cambios estructurales en los receptores de las proteínas surgen de forma selectiva prefiriendo cadenas oligoméricas en vez de cadenas monoméricas; **tercera ventaja**, como Monod comentó en [12] las proteínas más largas poseen ventajas en cuanto a su estabilidad reflejando una menor tasa de desnaturización (proceso en el cual surgen cambios estructurales en las proteínas haciendo que pierdan su estructura, plegamiento, y debida funcionalidad), la estabilidad de una proteína implica un delicado equilibrio entre la estabilización entálpica debida a muchas interacciones débiles no unidas y el efecto paralelo de varios factores entrópicos de movilidad conformacional y solvatación como sugiere Lumry R y Rajender S. en [13] (1970), también se observa que en proteínas pequeñas las compensaciones entrópicas de orden estructural no son suficientes para compensar el costo entrópico de las restricciones conformacionales a nivel molecular; la **cuarta ventaja** está relacionada con la reducción de la área superficial de la proteína, esto resulta importante para que la proteína sea más estable una vez esté interactuando con un solvente, y este objetivo puede cumplirse de forma más eficaz cuando la proteína en cuestión es más larga y se puede plegar para impedir que gran parte de su superficie no entre en contacto con el solvente, lo cual resulta más difícil de lograr cuando se trata de proteínas monoméricas.

A continuación mencionaremos las ventajas del por qué la construcción de proteínas encasillables dentro de un tipo de simetrías resulta favorable.

La estabilidad de las proteínas oligoméricas es una consecuencia de dos factores, el primero tiene que ver con la especificidad de las interfaces proteína-proteína lo cual favorece los complejos simétricos y segundo, tenemos que el máximo número de interacciones intersubunidades aparecen en complejos simétricos. La razón por la cual varias estructuras oligoméricas han revelado que los sitios en donde se dan interacciones proteína-proteína es porque estas interacciones son compuestas por superficies de dos dimensiones que son altamente específicas y direccionales, la direccionalidad de la que se habla aquí es la relacionada con las interfaces proteína-proteína que asegura que toda proteína homooligomérica sea simétrica. [7]

Para los fines funcionales correctos de una proteína se exige que esta sea capaz de evitar agregados no deseados, en este caso el hecho de que la proteína tenga una simetría bien definida impide la aparición de proteínas mutantes con agregados al volverlas menos estables una vez esto sucede, se ha observado que varios desordenes como la enfermedad de Alzheimer están altamente relacionados con la aparición de proteínas con agregados como los priones.

Por ultimo, debemos mencionar que Wolynes en su artículo "The energy landscapes and motions of proteins"

[14] habla acerca de la eficiencia del plegamiento que presentan las proteínas simétricas, en donde afirma que estas estructuras generan menos barreras de energía cinética a diferencia de las proteínas no simétricas las cuales debido a su estructura no se les permite reducir el número de barreras cinéticas, además de que basándose en analogías con los centros atómicos, argumenta que la energía "landscape" disponible para realizar los plegamientos en complejos simétricos podría variar menos que en los complejos asimétricos.

la especificidad de las interfaces proteína-proteína favorece los complejos simétricos

1.5. Análisis de modos normales en proteínas

Como se mencionó anteriormente, se han desarrollado distintos métodos para producir modelos dinámicos a partir de los datos estructurales, uno de los más comunes son los de dinámica molecular, en estos se modela la dinámica de la proteína a partir de las leyes de Newton, sin embargo este tipo de simulaciones no pasan de los microsegundos, por lo cual es necesario un método que nos permita comparar la dinámica de proteínas de una forma sistemática y computacionalmente eficiente, es aquí donde entra el análisis de modos normales, esta técnica es del mismo estilo que la usada por Einstein para modelar las oscilaciones de los núcleos en una red cristalina alrededor de su posición de equilibrio; en esencia son aplicaciones del teorema de pequeñas oscilaciones. La idea consiste en que una proteína al ser ligeramente perturbada de su conformación de equilibrio, oscilará alrededor de esta conformación en virtud de una fuerza restauradora, la justificación de la aplicación de este tipo de modelos se encuentra en [15] donde muestran que los movimientos globales de los péptidos pueden ser producidos por modelos armónicos que contienen dos términos, uno asociado con vibraciones rápidas descrito por el cálculo estándar de modos normales en un mínimo local y un término correspondiente a movimiento Browniano en un potencial armónico efectivo.

Un sistema se dice en equilibrio cuando las fuerzas generalizadas actuando sobre él son iguales a cero, Si etiquetamos como \mathbf{q}^0 a las coordenadas generalizadas en la posición de equilibrio, la ecuación de la energía potencial puede ser descrita por una serie en términos de \mathbf{q} :

$$V(\mathbf{q}) = V(\mathbf{q}^0) + \left(\frac{\partial V}{\partial q_n} \right)^0 \eta_i + \frac{1}{2} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right) \eta_i \eta_j + \dots$$

donde q_i y q_j corresponden a las configuraciones instantáneas de las componentes i,j y $\eta_i = q_i - q_i^0$.

El primer término en la serie representa el mínimo del

potencial el cual podemos tomar como cero, el segundo término debe ser cero dado que estamos en un mínimo local, así el potencial se escribe

$$V(\mathbf{q}) = \frac{1}{2} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right) \eta_i \eta_j = \frac{1}{2} \eta_i V_{ij} \eta_j,$$

donde V_{ij} es la matriz Hessiana que contiene las segundas derivadas del potencial respecto a las componentes del sistema.

Si consideramos la energía cinética

$$T(\mathbf{q}) = \frac{1}{2} \mathbf{M} \frac{d^2 \eta_i}{dt^2},$$

donde \mathbf{M} es una matriz diagonal que contiene las masas de cada partícula, la ecuación de movimiento se puede escribir como:

$$\frac{1}{2} \mathbf{M} \frac{d^2 \eta_i}{dt^2} + \frac{1}{2} \eta_i V_{ij} \eta_j = 0$$

La cual tiene como solución:

$$\eta_i = a_{ik} \cos(w_k t + \delta_k)$$

Donde a_{ik} es la amplitud de oscilación, w_k la frecuencia de oscilación y δ_k la fase. Sustituyendo esta solución en la ecuación de movimiento es posible reescribirla como:

$$\mathbf{V}\mathbf{A} = \Lambda\mathbf{A}$$

Donde \mathbf{A} es la matriz que contiene los autovectores de la matriz Hessiana \mathbf{V} y Λ es una matriz diagonal que contiene los autovalores.

1.6. Modelo de red elástica

A pesar de que el análisis de modos normales es menos costoso computacionalmente que la dinámica molecular, sigue siendo complicado realizar estos cálculos para proteínas que contienen varios miles de átomos. El primer problema es que la estructura a estudiar debe ser energéticamente minimizada para asegurar de que la conformación de la que se parte es realmente una conformación de equilibrio relativa al campo de fuerza elegido, frecuentemente este proceso distorsiona la estructura, llevando a que el método parte de una estructura diferente a la experimentalmente determinada. El segundo problema es que diagonalizar la matriz hessiana para todos los átomos aún sigue siendo bastante costoso computacionalmente, por esto se han desarrollado métodos de grano grueso, uno de los más usados es el modelo de red elástica.

La base de la aplicación de este modelo en proteínas recae en dos aproximaciones, primero la sustitución de los aminoácidos por masas puntuales conectadas con resortes, la ubicación de las masas puntuales se hace en el carbono alpha el cual se encuentra central en la estructura del aminoácido, tal como se observa en la figura 1. Este modelo disminuye considerablemente la dimensionalidad del problema, pero como consecuencia nos reduce el detalle de la descripción. Segundo en vez de trabajar con los complicados campos de fuerza semi empíricos usados en el análisis de modos normales estándar, se usará un potencial armónico simple:

$$V(\mathbf{q}) = \sum_{d_{ij} < R_c} C (d_{ij} - d_{ij}^0)^2$$

Donde d_{ij} corresponde a la distancia entre las masas i y j, d_{ij}^0 es la distancia en la conformación de equilibrio y C es la constante del resorte asumida igual para cada par. Es importante resaltar que en este método ya no es necesario minimizar la energía de la estructura. R_c corresponde a una distancia desde la cual ya no habrá interacción entre las masas, en la práctica se escoge entre $7A^0 - 8A^0$.

Algunas de las cantidades que nos interesan calcular a partir de este modelo son las frecuencias asociadas a cada modo, las energías de deformación y las fluctuaciones, esta última medida se define como la suma de los desplazamientos a lo largo de cada modo (excluyendo los modos triviales) pesados por el reciproco de los autovalores, los cuales corresponden al cuadrado de la frecuencia del modo correspondiente.

1.7. Alineamiento y comparación

Para poder tener un marco en el cual comparar las fluctuaciones de las proteínas es necesario realizar primero un procedimiento de alineamiento de secuencias de aminoácidos, esto corresponde a una asignación aminoácido-aminoácido en las dos secuencias, donde el principal propósito del alineamiento es identificar aminoácidos homólogos, es decir, aminoácidos codificados por bases en la misma posición del genoma.

Este alineamiento se hará con el algoritmo MUSTANG: A Multiple Structural Alignment Algorithm presentado en [16].

2. Objetivo General

Este trabajo tiene dos objetivos generales. Uno es de aprendizaje, buscamos profundizar más en la biología y la física de las proteínas. El segundo objetivo es técnico, a partir de estos conocimientos plantearemos una

pregunta sencilla que nos permita familiarizarnos con algunas técnicas y conceptos de la biofísica y biología molecular.

De forma particular:

Explorar la relación entre simetrías de una proteína y su dinámica a partir del modelo de red elástica; en este contexto se realizará una comparación de la dinámica de distintas proteínas en base a su simetría y función biológica.

3. Metodología

A continuación se presenta la metodología que se siguió para desarrollar la exploración y comparación planteadas en el objetivo general.

3.1. Elección de proteínas

Se buscaron en el Protein Data Bank proteínas con dos funciones biológicas distintas, metabolismo y antígeno; posteriormente se buscaron simetrías que estuvieran presentes en proteínas con estas dos funcionalidades, así se eligieron como simetrías la cíclica C2 y la Dihedral D2. Para seleccionar entre estas proteínas se tuvo en cuenta que el método usado es de grano grueso, por lo cual es posible que en la aplicación de la aproximación se pierda parte de la simetría de las proteínas, para evitar esto usamos dos criterios cualitativos, por un lado observamos la estructura de las proteínas antes y después de la aproximación, eligiendo así proteínas las cuales mantuvieran claramente su simetría luego de la aproximación, por otro lado siguiendo el razonamiento de [15] simetrías en la estructura deberían reflejarse en simetrías en las vibraciones de los aminoácidos, en base a estos dos criterios elegimos las proteínas.

Las proteínas elegidas y sus clasificaciones fueron:

Tabla 1: Proteínas elegidas, filas simetrías columnas función biológica

	Metabolismo	Antígeno
Dihedral D2	1ugy, 1ovs	1uld, 4k79
Cíclica C2	7c68, 1v6x	6jys, 6orh

En el anexo se presentan imágenes con la forma tridimensional de cada una de las proteínas, sus ejes de simetría y sus fluctuaciones.

3.2. Cálculos

Para realizar los cálculos en base a los métodos expuestos en el marco teórico se utilizó el software presentado

en [4], el cual hace uso de la aplicación Molecular Modeling ToolKit escrita en Python. De aquí se obtuvieron las frecuencias y energías de deformación para cada proteína individual, los alineamientos de las secuencias de aminoácidos y las fluctuaciones para los residuos alineados.

Posteriormente se usaron estos datos para calcular la diferencia en fluctuaciones para los aminoácidos alineados, a estas diferencias se les calculó el promedio y la varianza. Del alineamiento se calcularon los porcentajes de semejanza en la secuencia de aminoácidos. Esto cálculos se hicieron en el lenguaje Python, haciendo uso de las librerías Pandas y Numpy.

4. Resultados

4.1. Frecuencias

Se presentan las frecuencias al cuadrado para los primeros 56 modos normales para cada proteína.

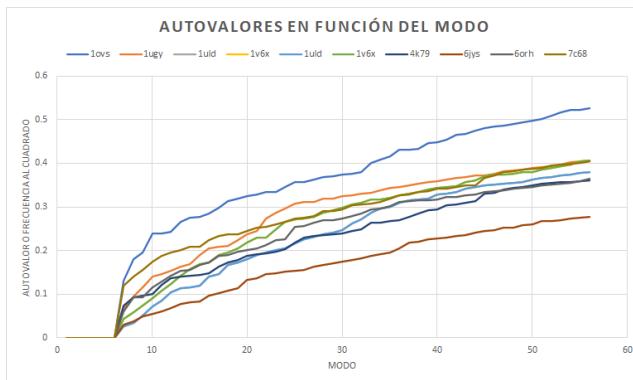


Figura 7: Frecuencias al cuadrado para los primeros 56 modos normales para cada proteína

Resalta el hecho de que a grandes valores de frecuencias las curvas de frecuencia en función del modo se estabilizan, si comparamos la gráfica en este punto con la tabla 1 se evidencia que las 4 curvas con mayor frecuencia en este punto corresponden a la función biológica metabolismo, mientras las 4 curvas con menor frecuencia corresponden a la función antígeno.

Note que mientras la función biológica divide claramente las curvas en el punto donde ya se estabilizaron, la simetría no, sin embargo dentro de las divisiones marcadas por la función biológica si parece haber influencia en la simetría, en el sentido de que proteínas con misma simetría tienden a converger sus curvas. Las dos curvas que más difieren corresponden precisamente a dos proteínas que no coinciden ni en simetría ni en función biológica.

4.2. Energías de deformación

Las energías de deformación corresponden a la energía potencial acumulada al deformarse el sistema respecto a su posición de equilibrio. A continuación una gráfica que presenta las energías de deformación para los primeros 20 modos.

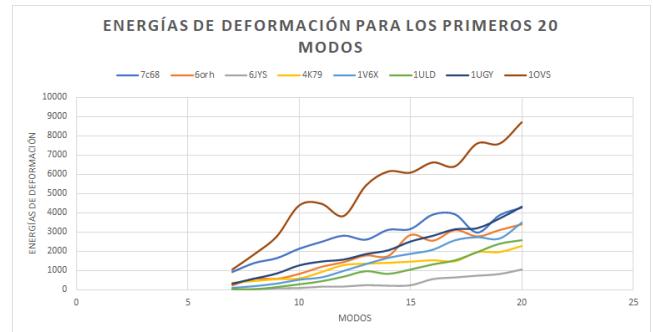


Figura 8: Energías de deformación para los primeros 20 aminoácidos

Acá se mantienen las conclusiones hechas sobre las frecuencias a excepción de 6orh que rompe el patrón de división según función biológica. Note que las proteínas que más difieren son las mismas en este caso.

4.3. Comparaciones: alineamiento y fluctuaciones

A continuación se presentarán tablas con las comparaciones de fluctuaciones hechas en base al alineamiento, esto junto a la medida del porcentaje de semejanza de las secuencias de aminoácidos de cada proteína y gráficos de fluctuaciones en aminoácidos alineados para cada pareja de proteínas.

4.3.1. Proteínas con igual simetría e igual función biológica

Tabla 2: Porcentaje de semejanza, promedio y diferencia de fluctuaciones para aminoácidos alineados en proteínas con igual simetría e igual función biológica

Proteínas	Semejanza	Fluctuaciones	Varianza
1ovs-1ugy	17.97 %	0.64	1.61
1uld-4k79	16.32 %	0.45	1.11
6jys-6orh	23.39 %	0.35	1.31
7c68-1v6x	30.43 %	0.14	0.05

En esta tabla se evidencia que las comparaciones hechas entre las proteínas elegidas con simetría cíclica C2 poseen mayor semejanza que las elegidas con simetría dihedral D2. Así mismo se evidencia que estas últimas

difieren más en sus fluctuaciones. Note que para las funciones biológicas no parece haber un patrón claro.

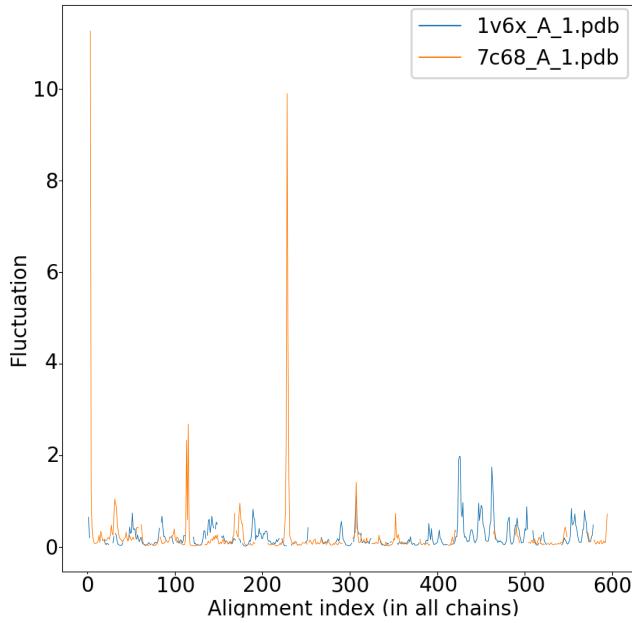


Figura 9: Comparación de fluctuaciones para los residuos alineados de 7C68 Y 1V6X. Eje X aminoácidos alineados, eje y fluctuaciones.

4.3.2. Proteínas con igual simetría y distinta función biológica

Tabla 3: Porcentaje de semejanza promedio y diferencia de fluctuaciones para aminoácidos alineados en proteínas con igual simetría y distinta función biológica

Proteínas	Semejanza	Fluctuaciones	Varianza
1v6x-6orh	36.39 %	0.16	0.16
1v6x-6jys	23.31 %	0.32	1.12
6jys-7c68	20.05 %	0.30	1.10
6orh-7c68	28.85 %	0.20	1.13
1ovs-1uld	17.75 %	0.18	0.13
1ovs-4k79	14.86 %	0.31	0.43
1ugy-1uld	18.20 %	1.06	10.50
1ugy-4k79	15.22 %	0.44	0.40

Las primeras cuatro comparaciones corresponden a proteínas con simetría cíclica C2 y las ultimas cuatro a simetría dihedral D2. Se evidencia una mayor semejanza y menor diferencia promedio de fluctuaciones para proteínas con la simetría cíclica,

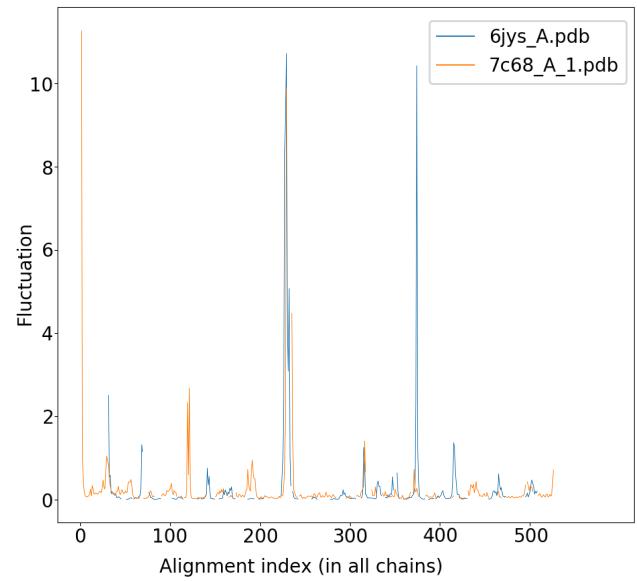


Figura 10: Comparación de fluctuaciones para los residuos alineados de 6JYS y 7C68. Eje X aminoácidos alineados, eje Y fluctuaciones.

4.3.3. Proteínas con distinta simetría e igual función biológica

Tabla 4: Porcentaje de semejanza, promedio y diferencia de fluctuaciones para aminoácidos alineados en proteínas con distinta simetría e igual función biológica

Proteínas	Semejanza	Fluctuaciones	Varianza
1ovs - 1v6x	17.87 %	0.81	14.25
1ovs - 7c68	16.30 %	0.68	6.50
1ugy - 1v6x	15.50 %	0.67	1.90
1ugy - 7c68	16.10 %	0.71	1.93
1uld-6jys	16.98 %	0.82	26.15
1uld-6orh	17.79 %	0.55	9.76
4k79-6jys	18.12 %	0.43	0.89
4k79-6orh	21.47 %	0.26	0.30

Las primeras cuatro comparaciones corresponden a proteínas con función biológica metabolismo, las últimas cuatro a la función biológica antígeno. En esta tabla no parece haber una diferencia clara en las comparaciones para cada función biológica.

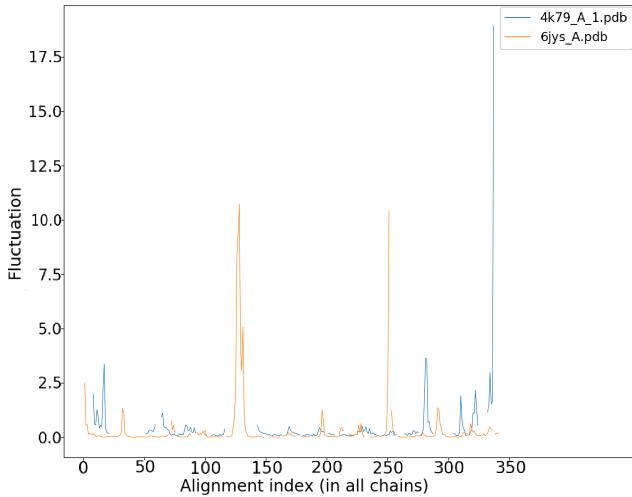


Figura 11: Comparación de fluctuaciones para los residuos alineados de 4K79 y 6JYS. Eje X aminoácidos alineados, eje Y fluctuaciones.

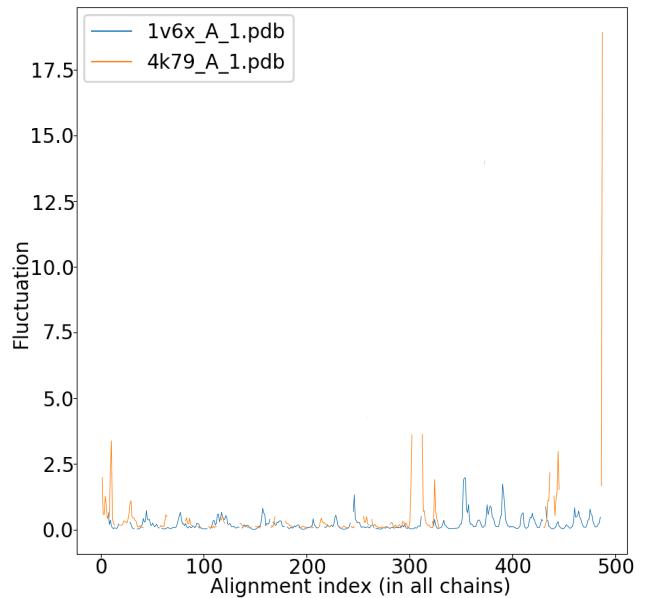


Figura 12: Comparación de fluctuaciones para los residuos alineados de 4K79 y 1V6X. Eje X aminoácidos alineados, eje Y fluctuaciones.

5. Discusión

4.3.4. Proteínas con distinta simetría y distinta función biológica

Tabla 5: Porcentaje de semejanza, promedio y diferencia de fluctuaciones para aminoácidos alineados en proteínas con distinta simetría e igual función biológica

Proteínas	Semejanza	Fluctuaciones	Varianza
1uld - 1v6x	18.01 %	0.25	0.10
1v6x - 4k79	19.15 %	0.27	0.28
7c68 - 1uld	16.36 %	1.06	32.21
7c68 - 4k79	19.39 %	0.42	2.42
1ovs-6jys	15.62 %	0.56	5.90
1ovs-6orh	16.37 %	0.89	15.86
1ugy-6jys	14.90 %	0.64	1.82
1ugy-6orh	16.16 %	0.64	1.88

Note que ningún porcentaje de semejanza llega al 20 % como si ocurría en los anteriores casos, por otro lado todos los porcentajes de semejanza son cercanos entre sí, sin embargo las varianzas si presentan grandes variaciones.

Las comparaciones hechas en este trabajo se pueden dividir en dos, Por un lado las provenientes de datos calculados para las proteínas individualmente que posteriormente fueron comparados y por otro lado las surgidas de los cálculos de fluctuaciones y semejanza de secuencias para las proteínas luego del proceso de alineamiento. En el primer caso se evidencia que tanto para la comparación de frecuencias como de energías de deformación la función biológica es más determinante de semejanzas que la simetría. En el segundo caso tenemos lo opuesto, tanto en las comparaciones de diferencia de fluctuaciones promedio como de semejanza en la secuencia de aminoácidos la simetría parece más determinante de semejanza que la función biológica.

Desde la teoría se espera que la dinámica de las proteínas, en este caso sus propiedades vibracionales dependan tanto de propiedades estructurales como la simetría, como de su función biológica; las comparaciones hechas en este trabajo confirman esto. Sin embargo es relevante notar que dependiendo el procedimiento de comparación de las propiedades vibracionales, en este caso primero calculando por separado y luego alineando, la simetría o la función biológica se alzan como la característica más relevante para determinar semejanza en la dinámica de las proteínas.

Finalmente es importante tener en cuenta que en este trabajo sólo se tomaron 8 proteínas, las cuales no son una muestra estadística representativa. Por otra parte dado que las búsquedas se hicieron en el protein da-

ta bank, se usó la definición de función biológica del mismo, lo cual oscurece aún más la claridad de las conclusiones.

Referencias

- [1] Wikipedia. Aminoácido. <https://es.wikipedia.org/wiki/Amino%C3%A1cido>, 2021.
- [2] Mabel Sanquea. Estructura y función de las proteínas. https://www.amandalabarca.cl/blog1/images/docs_2020/12-tercero-medio/13-17-julio/13-17-julio-dif-biologia-proteinas-funcion.pdf, 2020.
- [3] Wikipedia. Nucleotido. <https://es.wikipedia.org/wiki/Nucle%C3%B3tido>, 2021.
- [4] Sandhya T Tiwari, Edvin Fuglebak, Siv M Hollup, Lars Skjaerven, Tristan Cagnolini, Svenn H Grindhaug, Kidane M Tekle, and Nathalie Reuter. Webnm@ v2.0: Web server and services for comparing protein flexibility. *BMC Bioinformatics*, 15(427), 2014.
- [5] Katherine Henzler-Wildman and Dorothee Kern. Dynamic personalities of proteins. *Nature*, 450(7172):964–972, 2007.
- [6] Robert M Wald. General relativity(book). *Chicago, University of Chicago Press, 1984, 504 p*, 1984.
- [7] David S Goodsell and Arthur J Olson. Structural symmetry and protein function. *Annual review of biophysics and biomolecular structure*, 29(1):105–153, 2000.
- [8] THE SVEDBERG. Mass and size of protein molecules. *Nature*, 123(3110):871–871, 1929.
- [9] David S Goodsell. Inside a living cell. *Trends in biochemical sciences*, 16:203–206, 1991.
- [10] Francis HC Crick, James D Watson, et al. Structure of small viruses. *Nature*, 177:473–5, 1956.
- [11] IM Klotz, NR Langebman, and DW Dahmalla. Quaternary structure of proteins. *Annual review of biochemistry*, 39(1):25–62, 1970.
- [12] J. Monod. On symmetry and function in biological systems. In André Lwoff and Agnes Ullmann, editors, *Selected Papers in Molecular Biology by Jacques Monod*, pages 701–713. Academic Press, 1978. <https://doi.org/10.1016/B978-0-12-460482-7.50061-0>.
- [13] Rufus Lumry and Shyamala Rajender. Enthalpy–entropy compensation phenomena in water solutions of proteins and small molecules: a ubiquitous property of water. *Biopolymers: Original Research on Biomolecules*, 9(10):1125–1227, 1970.
- [14] Hans Frauenfelder, Stephen G Sligar, and Peter G Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.
- [15] Konrad Hinsen, Andrei-Jose Petrescu, Serge Delleur, Marie-Claire Bellissent-Funel, and Gerald R Kneller. Harmonicity in slow protein dynamics. *Chemical Physics*, 261(1-2):25–37, 2000.
- [16] Arun S Konagurthu, James C Whisstock, Peter J Stuckey, and Arthur M Lesk. Mustang: a multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics*, 64(3):559–574, 2006.

Anexos

Anexo A. Forma tridimensional de proteínas y fluctuaciones

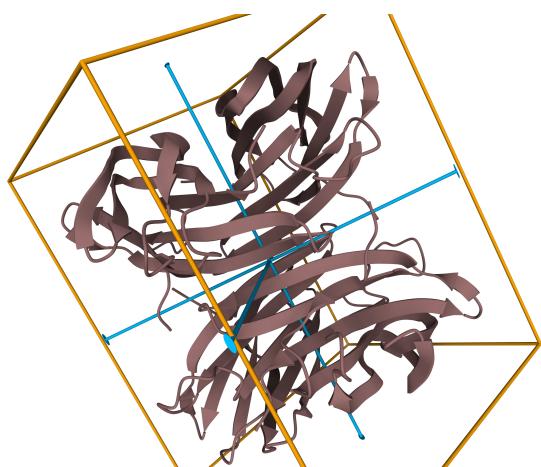


Figura Anexo A.1: Proteína 1OVS, función biológica: Metabolismo, simetría global: Dihedral D2.

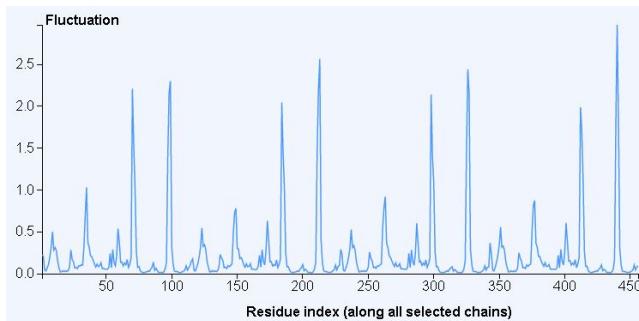


Figura Anexo A.2: Fluctuaciones en 1OVS.

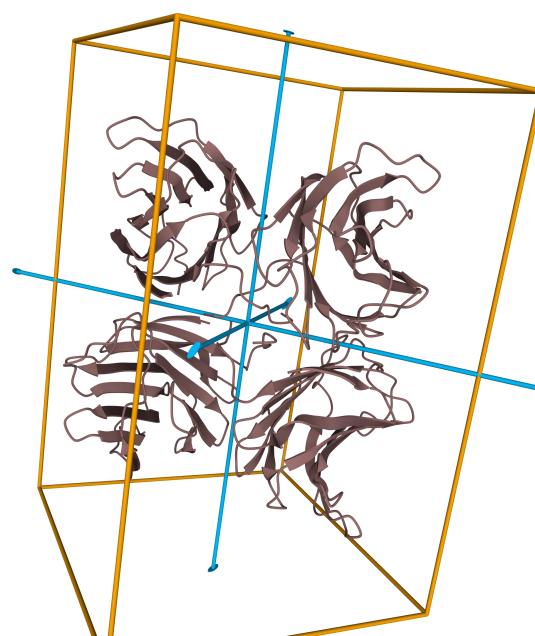


Figura Anexo A.5: Proteína 1ULD, función biológica: Antígeno, simetría global: Dihedral D2.

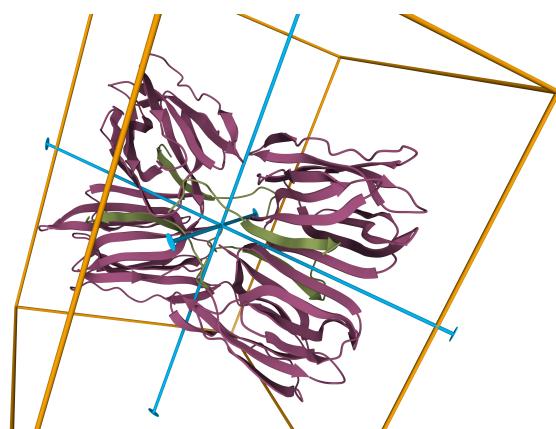


Figura Anexo A.3: Proteína 1UGY, función biológica: Metabolismo, simetría global: Dihedral D2.

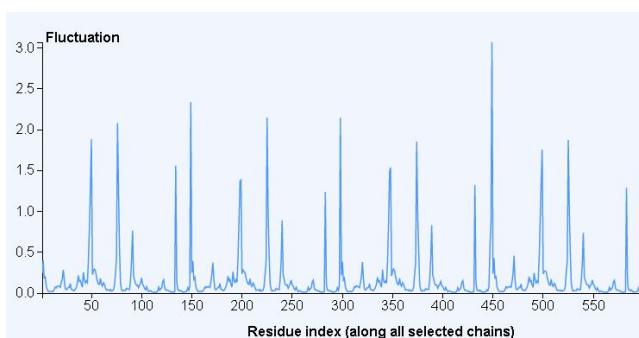


Figura Anexo A.4: Fluctuaciones en 1UGY.

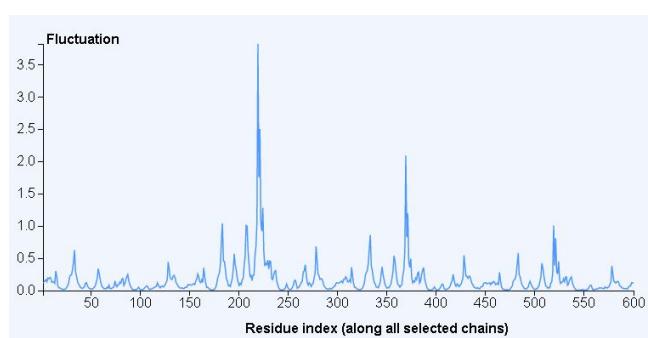


Figura Anexo A.6: Fluctuaciones en 1ULD.

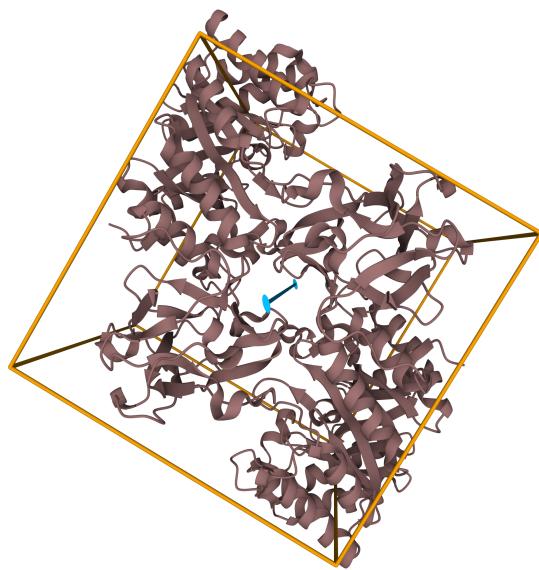


Figura Anexo A.7: Proteína 1V6X, función biológica: Metabolismo, simetría global: Cíclica C2.

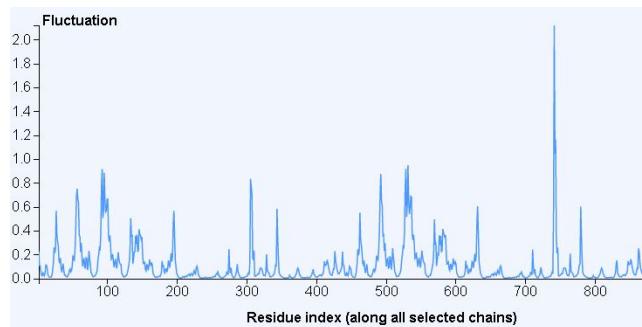


Figura Anexo A.8: Fluctuaciones en 1V6X.

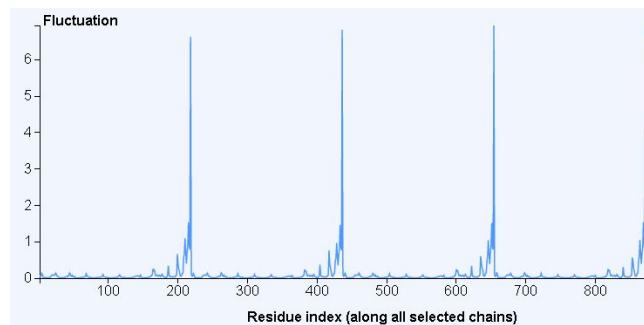


Figura Anexo A.10: Fluctuaciones en 4K79.

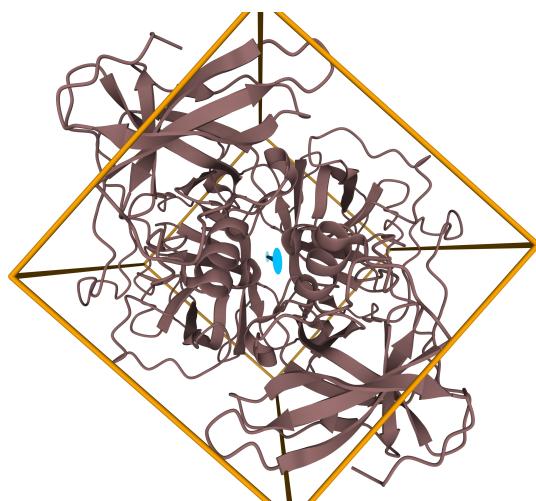


Figura Anexo A.11: Proteína 6JYS, función biológica: Antígeno, simetría global: Cíclica C2.

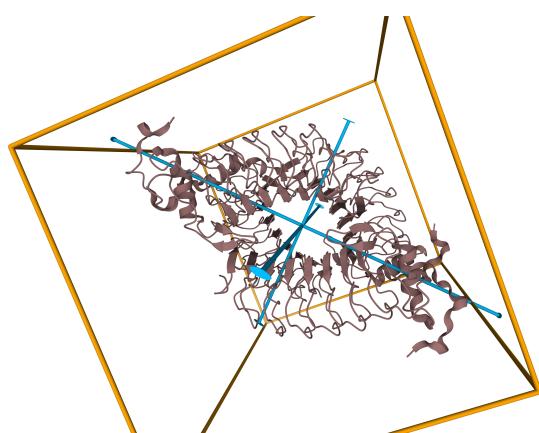


Figura Anexo A.9: Proteína 4k79, función biológica: Antígeno, simetría global: Dihedral D2.

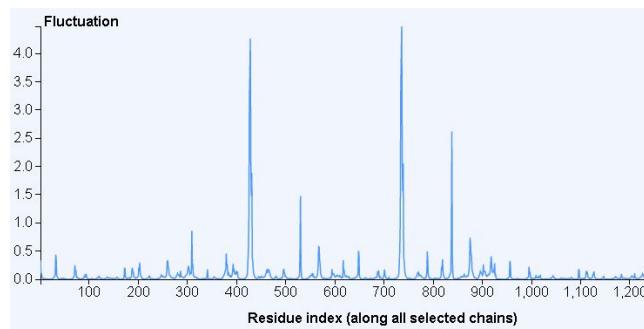


Figura Anexo A.12: Fluctuaciones en 6JYS.

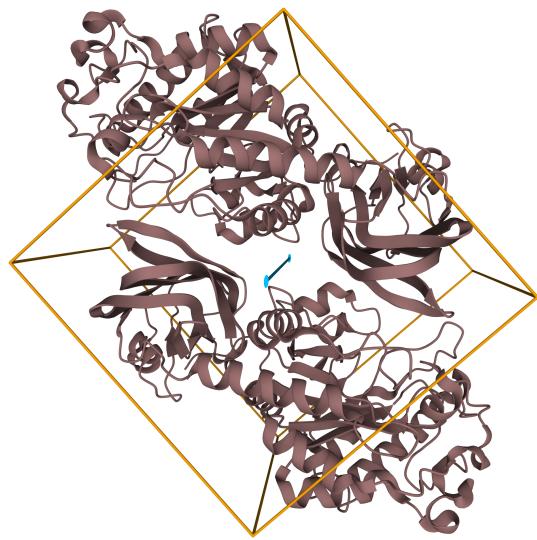


Figura Anexo A.13: Proteína 6ORH, función biológica: Antígeno, simetría global: Cíclica C2.

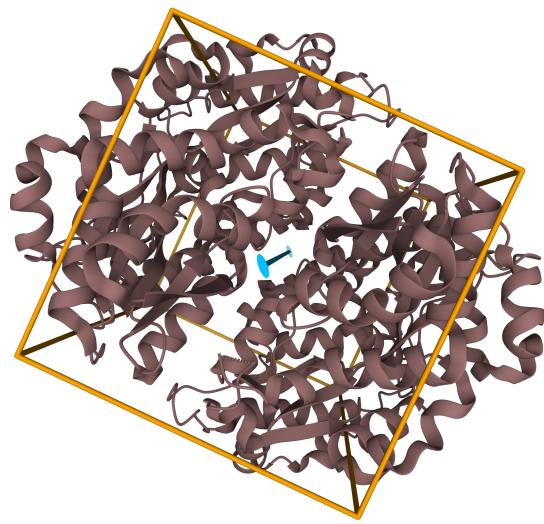


Figura Anexo A.15: Proteína 7C68, función biológica: Metabolismo, simetría global: Cíclica C2.

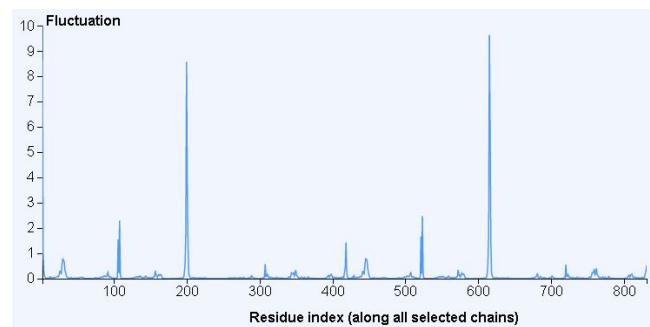


Figura Anexo A.16: Fluctuaciones en 7C68.

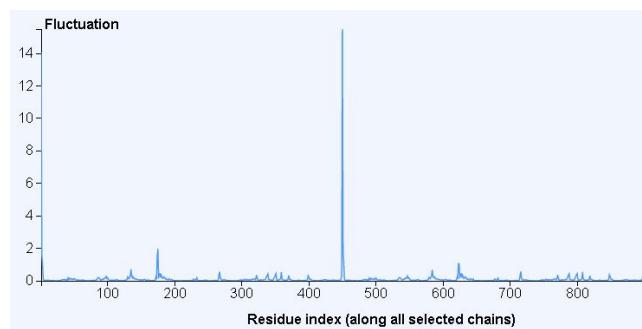


Figura Anexo A.14: Fluctuaciones en 6ORH.