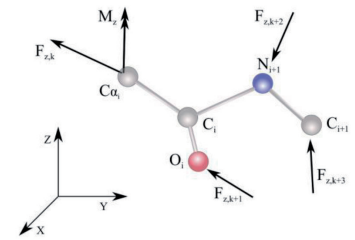


Estudio de las proteínas desde una perspectiva biocinemática: estado del arte



Protein study from a biokinematic perspective: literature review

Mikel Díez Sánchez, Víctor Petuya Arcocha, Alfonso Hernández Frías y Mónica Urizar Arana

Universidad del País Vasco (UPV/EHU). Escuela de Ingeniería de Bilbao. Departamento de Ingeniería Mecánica. c/ Plaza Ingeniero Torres Quevedo, 1 - 48013 Bilbao (España).

DOI: <http://dx.doi.org/10.6036/9033> | Recibido: 05/11/2018 • Inicio Evaluación: 06/11/2018 • Aceptado: 14/02/2019

ABSTRACT

- Proteins are the macromolecules responsible for almost every function living organisms need to exist. These macromolecules, which resemble small mechanisms or robots, were discovered 200 years ago, and first seen in their functional form less than 60 years ago, and they represent one of the more challenging areas of study. Protein study is an incredibly rich and multidisciplinary field, as it needs to be studied from different areas such as biology, physics or engineering. Current research involves great challenges since we are far from understanding how does the interaction between atoms work and the actual technology limitations represent a big obstacle when trying to understand how these nature robots operate. In this paper, a brief state of the art of the methodologies aimed at the study of protein folding or molecular mechanism simulation is presented. The paper is specifically focussed on procedures closely related to engineering and mechanism and machine study, without forgetting some classic methodologies which help understanding the global state of the art.
- Keywords:** proteins, structure, biokinematics, mechanisms.

RESUMEN

Las proteínas son las macromoléculas encargadas de llevar a cabo la gran mayoría de las funciones vitales que los seres vivos requieren para su existencia. Estas macromoléculas, las cuales se asemejan a pequeños mecanismos o robots, fueron descubiertas hace 200 años y observadas en su forma funcional hace menos de 60, y suponen uno de los campos de estudio más prometedores del momento. El ámbito de estudio de las proteínas es increíblemente rico y multidisciplinar, no pudiéndose abordar completamente si no es desde una combinación de la biología y la ingeniería. La investigación actual presenta grandes retos: la complejidad de las fuerzas que gobiernan las uniones atómicas, las cuales todavía estamos comenzando a comprender, y las limitaciones tecnológicas actuales que presentan serios obstáculos a la hora de tratar de comprender mejor el funcionamiento de estos robots de la naturaleza. En este trabajo, se presenta un breve estado del arte de los métodos desarrollados para resolver problemas como el plegado de proteínas o la simulación del mecanismo molecular. El artículo se centra especialmente en aquellos que están más estrechamente relacionados con la ingeniería, pero sin olvidar los métodos clásicos para tratar de comprender un poco mejor el estado actual de la técnica.

Palabras clave: proteínas, estructura, biocinemática, mecanismos.

1. INTRODUCCIÓN

El descubrimiento del genoma humano supuso un hito en el estudio de la genética, especialmente en el área de las proteínas. El genoma proporciona una herramienta de trabajo inigualable, ofreciendo la oportunidad de descubrir nuevas proteínas y, por lo tanto, información sobre los procesos biológicos que tienen lugar dentro de cada célula. Las proteínas son unas de las macromoléculas más importantes de los seres vivos ya que son las responsables de la mayoría de las funciones necesarias para la vida. Dichas funciones van desde la formación de tejidos y la defensa frente a agentes externos, hasta la catalización de reacciones químicas y el transporte de otros compuestos químicos. El estudio de las proteínas y la comprensión de cómo llevan a cabo sus funciones, aporta información de gran utilidad a biólogos y químicos para la ayuda al desarrollo de medicamentos, nuevos materiales, equipamiento médico o nuevos procesos de fabricación [1, 2].

En general, cuando se habla de proteínas es común relacionarlas con áreas como pueden ser la nutrición, la biología, o la física. Sin embargo, si se profundiza un poco en la naturaleza de las mismas, y se analiza su estructura, la forma en la que se crean o incluso cómo realizan sus funciones, resulta que no son tan diferentes de un mecanismo. En su estructura es posible modelizar elementos rígidos, elementos de unión que funcionan como juntas cinemáticas, cadenas completas que se asemejan a robots serie, etc. Esta novedosa forma de entender y estudiar las proteínas ha propiciado la aparición de nuevos métodos, que podrían denominarse *biocinmáticos* [3] y que tratan de aplicar técnicas y conceptos de Robótica y Cinemática y Dinámica de Máquinas al estudio de las proteínas.

En este artículo, se van a describir las problemáticas más relevantes en el estudio de las proteínas que pueden ser abordadas desde áreas como la Teoría de Mecanismos o la Teoría de Estructuras. En primer lugar, se realizará una descripción de los métodos tradicionales aplicados a su estudio. Posteriormente, se expondrán con más detalle los métodos aplicados a dichas problemáticas pero que tienen unas fuertes raíces en la Robótica o la Teoría de Mecanismos. En aras de facilitar la comprensión del artículo, se realizará previamente una breve descripción de lo que es una proteína haciendo hincapié en los detalles que permiten abordar una modelización de ésta desde un enfoque mecánico.

1.1. PROTEÍNAS: MECANISMOS EN MINIATURA

Una proteína está compuesta por aminoácidos, de los cuales existen 20 diferentes. Cuando estos aminoácidos se unen crean la "cadena" de la proteína siendo el número y orden de éstos lo que determina el tipo de proteína constituida. En la Figura 1, se muestra un ejemplo de un aminoácido, en este caso el ácido glutámico. Todos los aminoácidos comparten 4 átomos que son siempre iguales: un átomo de nitrógeno, dos átomos de carbono y un átomo de oxígeno (Figura 1 (a)). Sin embargo, poseen una cadena de átomos única que los diferencia del resto de aminoácidos. Esta cadena se conoce como cadena secundaria y siempre nace del carbono central denominado *carbono alpha*, $C\alpha$. Cuando una proteína se crea, lo hace uniendo los aminoácidos uno después de otro, formando lo que se denomina la cadena principal de la proteína (ver Figura 1 (b)).

Cuando esta cadena principal se forma, la proteína adopta la configuración con la que lleva a cabo su función. Esta configuración se denomina estructura terciaria de la proteína.

En una proteína pueden distinguirse dos tipos de movimientos. Por un lado, los átomos poseen un movimiento vibratorio alrededor de su posición de equilibrio y, por otro, un movimiento de rotación alrededor de los ejes definidos por los enlaces entre átomos. Para el estudio del primero, los enlaces se asemejarían a resortes definidos entre los átomos que unen. Sin embargo, si se quiere abordar el estudio desde una perspectiva global del movimiento de la proteína, estos movimientos vibratorios pueden ser despreciados [4]. En este caso, se pueden definir dos conjuntos de grados de libertad dentro de una proteína: los grados de libertad de rotación de la cadena principal, y, denominados ángulos diédricos (ver Figura 1(b)), y los grados de libertad de rotación de las cadenas secundarias. De este modo, la cadena principal de la proteína puede entenderse como un robot serie con cientos o miles de grados de libertad, de la que a su vez surgen multitud de pequeños robots serie de 2 a 5 grados de libertad correspondientes a las cadenas secundarias. Esta modelización, propuesta por varios autores [3, 5], es la que se utiliza en los estudios basados en un enfoque *biocinémático*.

1.2. PROBLEMAS PLANTEADOS EN EL ESTUDIO DE LAS PROTEÍNAS

El estudio de una proteína se puede abordar desde muy diferentes áreas, cada una enfocada a un aspecto, función o característica particular de la misma. Algunos de los problemas que se pueden estudiar son considerados la "Piedra de Rosetta" de las proteínas, ya que, en caso de dar solución a los mismos, el conocimiento sobre la biología molecular crecería de forma exponencial. De este modo, se pueden destacar cinco problemas diferentes.

El primero, y más importante, sería el plegado de proteínas (*protein folding*) [6, 7]. El objetivo fundamental de este problema es obtener la forma funcional (estructura terciaria) de una proteína conociendo únicamente la cadena de aminoácidos que la conforma (ver Figura 2). Los métodos experimentales, y gran parte de los métodos analíticos actuales, sólo son capaces de obtener la configuración plegada. Sin embargo, el proceso por el cual las proteínas alcanzan dicha configuración final es desconocido. Además, en los casos en los que las proteínas no son solubles, los métodos experimentales no pueden ser aplicados. El plegado de proteínas es uno de los retos más complejos que existen hoy en día en la biología estructural, tanto por su complejidad teórica, como por la imposibilidad de resolverse de forma directa probando todas las configuraciones posibles para una cadena de aminoácidos debido al elevado coste computacional [8,9].

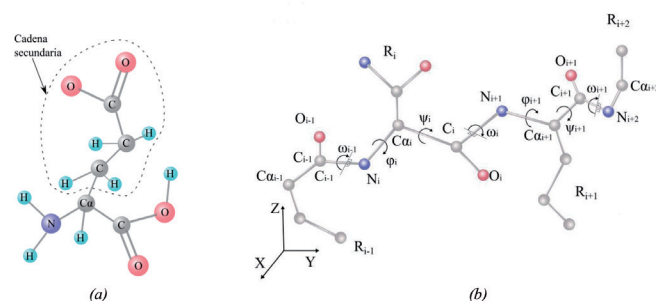


Fig. 1: (a): Aminoácido ácido glutámico. (b) Sección de la cadena principal de una proteína

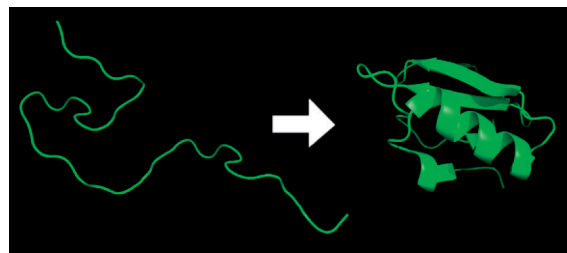


Fig. 2: Proteína desplegada y plegada en su forma funcional

Otro problema estrechamente relacionado con el plegado de proteínas, sería la simulación del mecanismo molecular (*molecular mechanism simulation*) [10, 11]. En este caso, se analiza el cambio de forma que sufren algunas proteínas cuando llevan a cabo su función. Este proceso sucede, en función de la proteína objeto de estudio, para permitir el acceso de algún compuesto químico, para unirse con otras proteínas o realizar otro tipo de acciones (ver Figura 3). Al igual que en el plegado de proteínas, si bien los métodos experimentales son capaces de obtener la configuración inicial y final del movimiento, no permiten observar el movimiento de forma continua. Sin embargo, obtener la simulación del movimiento entre estos dos estados, es fundamental para comprender el funcionamiento de la proteína. No obstante, en comparación con el plegado de proteínas, el coste computacional del problema es significativamente menor.

En relación al análisis de la estructura de la proteína, existen dos problemáticas: el cierre de lazo (*loop closure*) y el análisis estructural. El cierre de lazo [12] surge como consecuencia de la limitación actual de los métodos experimentales. Éstos, en ocasiones, no dan como resultado la estructura completa de la proteína, sino que dejan partes de la estructura de la proteína sin definir. El objetivo de este análisis sería completar dichas partes indefinidas generando una estructura final completa. Por otro lado, el análisis estructural de una proteína trata de determinar qué zonas de ésta pueden tener movimiento relativo entre ellas y cuáles están rigidamente unidas. Los resultados de este análisis pueden utilizarse como punto de partida para la simulación del mecanismo molecular, reduciendo así el número de grados de libertad representativos implicados el proceso.

Finalmente, el último problema que puede ser considerado como afín al campo de los mecanismos o las estructuras, es la unión de proteínas (*protein docking*). Cuando se habla de proteínas se tiende a pensar en una única proteína, independiente, realizando su función. La realidad es que, en muchos casos, varias proteínas se unen para formar un compuesto mayor. De nuevo, los procesos por los que se producen estas uniones, o incluso en algunos casos cuál es la zona en la que se van a unir dos proteínas, son desconocidos.

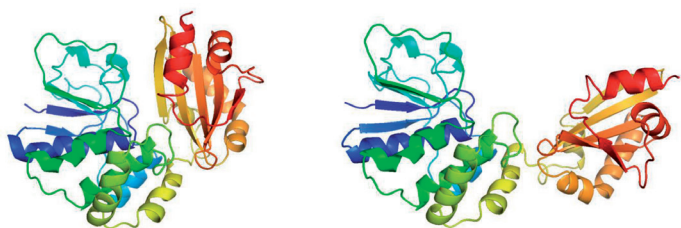


Fig. 3: Estado inicial y final de la proteína 1BPTI

Una vez definidas brevemente las problemáticas actuales, se va a proceder a describir algunos de los métodos existentes para el estudio de éstas.

2. MÉTODOS PARA EL ESTUDIO DE LAS PROTEÍNAS

2.1. MÉTODOS CLÁSICOS PARA EL ESTUDIO DE LAS PROTEÍNAS

Sin duda alguna, uno de los campos más importantes dentro de las proteínas es el estudio y obtención de las estructuras proteicas. Para definir el funcionamiento de una proteína es necesario conocer su estructura terciaria, la cual se puede obtener, como se ha comentado, bien mediante métodos experimentales o bien utilizando métodos analíticos.

Hoy en día, se utilizan principalmente dos métodos experimentales para la obtención de la estructura terciaria de una proteína. En concreto, la Resonancia Magnética Nuclear (RMN) y la Cristalografía de Rayos X. El principio de la Cristalografía de Rayos X es sencillo [13]. Se hace incidir un haz de fotones sobre un cristal ordenado, formado por la proteína a estudiar. El adjetivo de "ordenado" se refiere a que todas las proteínas que forman el cristal han de estar dispuestas en la misma orientación. Cuando el haz de fotones de RX atraviesa el cristal, son difractados por los planos de las estructuras cristalinas de las proteínas ordenadas. Esta difracción de los RX es medida por un detector, que puede ser electrónico, o una simple película de Rayos X. El resultado obtenido se muestra en la Figura 4. A partir de ese patrón plano, se extrapolan las posiciones de los átomos de la proteína siendo posible deducir su estructura.

La Resonancia Magnética Nuclear utiliza el principio de la resonancia [14]. Al introducir los átomos en un campo magnético y aplicar una energía, los átomos vibran a una determinada frecuencia, la cual es posible medir. De este modo, se detectan las frecuencias y las posiciones de los átomos de la proteína. Al igual que con la cristalografía de rayos X, no se obtiene una imagen tridimensional, sino que se obtiene únicamente una representación plana de las posibles posiciones de los átomos. Ambos métodos requieren de un postprocesado donde se trata de generar la estructura tridimensional coherente con el patrón experimental obtenido. La Cristalografía de Rayos X y la Resonancia Magnética Nuclear ofrecen resoluciones entre los 1-4 Å y son actualmente los referentes a la hora de obtener estructuras proteicas. La mayor desventaja de estos métodos es que, tanto por la tecnología disponible como por las propias características de algunas proteínas, no son capaces de obtener ciertas estructuras.

Dentro de los métodos considerados como métodos clásicos también existen procedimientos que utilizan un enfoque analítico. En este ámbito, destacan los métodos *ab initio* y los métodos homólogos [15]. Al igual que en el caso de los métodos experimentales, ambos procedimientos tratan de resolver el problema del plegado de proteínas. Los métodos *ab initio* [16], aplican la hipótesis de que la estructura terciaria de la proteína se corresponde

con el mínimo potencial energético. Para ello, se utilizan procesos iterativos que implementan algoritmos de búsqueda como los métodos de Monte Carlo [17]. Estos métodos son muy costosos computacionalmente, llegando incluso a necesitar meses de cálculo en un ordenador convencional para obtener la estructura terciaria. Además, el proceso de minimización de las funciones potenciales presenta multitud de mínimos locales que pueden hacer que el algoritmo obtenga una estructura errónea [18]. Pese a todas estas desventajas, los métodos *ab initio* tienen el potencial de obtener la estructura terciaria de una proteína partiendo únicamente de su secuencia de aminoácidos o estructura primaria.

Por otro lado, los métodos homólogos [19] parten de la hipótesis de que, pese a existir un gran número de proteínas diferentes, las estructuras terciarias que éstas pueden adoptar son limitadas [20]. A partir de esta hipótesis, los métodos clasificados dentro de esta familia obtienen la estructura terciaria de la proteína a partir de la comparación con otras cuya estructura primaria es análoga a la de la proteína objeto de estudio. Al no requerir la minimización de una función potencial, los métodos homólogos son mucho menos costosos computacionalmente. Sin embargo, estos métodos presentan un gran inconveniente, ya que no pueden resolver estructuras de las que no dispongan de una referencia previa.

Nuevos enfoques como los presentados en [21] tratan de combinar estos procedimientos descritos para obtener métodos más eficientes computacionalmente y capaces de obtener la estructura de cualquier proteína. Sin embargo, en la actualidad, el 96% de las estructuras almacenadas en el Banco de Datos de Proteínas (*Protein Data Bank*, PDB) [22] son obtenidas experimentalmente por lo que los métodos analíticos aún tienen mucho camino por recorrer.

Se destacan tres métodos: análisis de modos y frecuencias [23-25], planificación de trayectorias [26-30] y método de la minimización cinetostática. De estos, el método de la minimización cinetostática se presenta a continuación, encontrándose la explicación de los dos restantes en el material suplementario disponible en la web.

2.2. MÉTODO DE LA MINIMIZACIÓN KINETOSTÁTICA - KINETOSTATIC COMPLIANCE METHOD

Tanto el método de la minimización cinetostática (kinetostatic compliance method) [5] como el método de los ángulos diédricos adaptativos (adaptive dihedral angle approach), que se explicará a continuación, hacen uso de una modelización simplificada de la proteína. Dicha modelización considera como elementos rígidos los enlaces entre los átomos, manteniendo de esta forma constantes su longitud y el ángulo relativo entre dos enlaces. Por lo tanto, los grados de libertad se reducen a los denominados ángulos diédricos de la cadena principal y a los giros de las cadenas secundarias. La modelización se asemeja a la presentada en la Figura 1 de apartado 1.1. Este método se utiliza para abordar el problema del plegado de proteínas tratando de reducir el coste computacional mediante el uso de esta modelización.

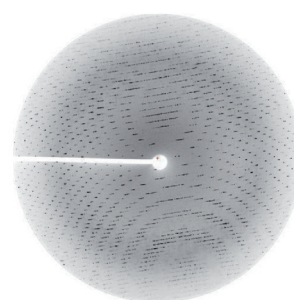


Fig. 4: Patrón de difracción de un cristal de una proteína [13]

Para realizar la simulación, el método de la minimización cinetostática utiliza la información del campo potencial para, derivando la expresión, obtener unas fuerzas resultantes equivalentes aplicadas sobre los átomos de la proteína. Una vez calculadas las fuerzas, se obtiene un sistema de momentos equivalentes aplicados en los ángulos diédricos, a partir de los cuales se asigna un incremento angular a aplicar. El proceso se repite hasta minimizar el valor de los momentos calculados en los ángulos diédricos. En la Figura 5, se muestra un ejemplo. En la Figura podemos ver las fuerzas, F_z , que se aplican sobre varios átomos de la cadena principal obtenidas a partir de la información del campo potencial, así como el momento equivalente, M_z , aplicado en el $C\alpha$ que generaría el movimiento para la siguiente iteración.

El método ha ido evolucionando considerando no sólo los grados de libertad de la cadena principal, sino también los de las cadenas secundarias [31]. En este caso, se plantean diferentes alternativas como considerar únicamente los grados de libertad de las cadenas secundarias en algunos momentos de la simulación, o escalar las cadenas para, progresivamente ir devolviéndolas a su tamaño original, minimizando así su efecto global sobre el resto de la proteína. Más recientemente, los desarrollos se han ido centrando en optimizar los algoritmos de evaluación y procesamiento utilizando técnicas de computación paralela y aceleración por GPU [32]. En el artículo citado se puede observar el gran impacto que la programación y el hardware utilizado para la simulación tienen en el coste computacional de los diferentes métodos. Se observan desde tiempos de simulación de varios segundos por iteración en las implementaciones más simples, hasta décimas de segundo en las implementaciones más optimizadas y sobre estaciones de cálculo dedicadas.

3. UNA NUEVA PROPUESTA BIOCINEMÁTICA: MÉTODO DE LOS ÁNGULOS DIÉDRICOS ADAPTATIVOS - ADAPTIVE DIHEDRAL ANGLE APPROACH

El denominado Método de los ángulos diédricos adaptativos, desarrollado en el grupo de investigación COMPMECH de la UPV/EHU (www.ehu.es/compmech), con la colaboración del centro de investigación CIC bioGUNE (www.cicbiogune.es), se centra en el problema de simulación del mecanismo molecular [33, 34]. El procedimiento propuesto calcula y simula el movimiento entre las posiciones inicial y final de la proteína, las cuales son obtenidas experimentalmente. La modelización propuesta se basa, al igual que el método de la minimización cinetostática, en el enfoque presentado en el apartado 1.1.

Una característica particular de este método es que combina tres algoritmos diferentes: un procedimiento de normalización de la estructura de la proteína, un procedimiento para la detección de estructuras secundarias y un algoritmo para realizar el movimiento de la proteína. A continuación, se explicarán las principales características de dichos procesos.

3.1. NORMALIZACIÓN DE LA ESTRUCTURA DE LA PROTEÍNA

En el procedimiento desarrollado se parte de las posiciones inicial y final de la proteína obtenidas experimentalmente. Debido a los movimientos de vibración de los átomos, en las imágenes experimentales, la longitud de los enlaces atómicos o los valores de los ángulos entre enlaces no son iguales en las posiciones inicial y final. Esto concuerda con la naturaleza de una proteína. Sin embargo, el modelo propuesto considera los enlaces como barras de longitud constante, por lo que es necesario realizar un proceso

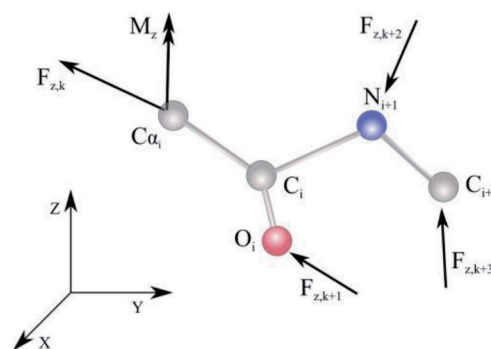


Fig. 5: Fuerzas calculadas en los átomos y momento equivalente aplicado en el $C\alpha$

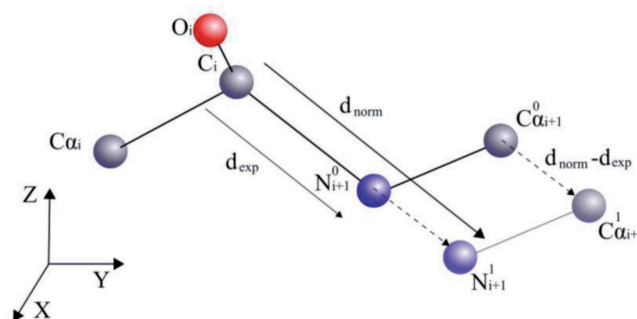


Fig. 6: Aplicación de la restricción de distancia entre el átomo de C del aminoácido i y el de N del aminoácido $i+1$. Obsérvese que los siguientes átomos también se desplazan la misma magnitud (en este caso, sólo se ha representado el siguiente átomo de la cadena)

de normalización que asegure que las dos configuraciones (inicial y final) son geométricamente iguales sin modificar la estructura de la proteína.

El procedimiento de normalización actúa en dos fases diferentes. En la primera, se impone a los átomos pertenecientes a un mismo plano peptídico que se encuentren realmente en dicho plano. En una segunda fase, el proceso hace que todas las longitudes de los enlaces sean iguales en ambas posiciones de la proteína. Este orden asegura que al finalizar el proceso de normalización no existan colisiones entre átomos. Para lograr esto, la normalización de las longitudes de los enlaces entre átomos se realiza imponiendo restricciones de distancia, comenzando en un extremo de la proteína y finalizando en el otro. Cuando se aplica una restricción de distancia entre dos átomos, se desplazan todos los átomos siguientes de la proteína en la misma dirección y con la misma magnitud. En la Figura 6, se ha representado la aplicación de una restricción de distancia en un enlace en una parte de la cadena de la proteína. Como se puede observar, se está aplicando una restricción de distancia al átomo N_{i+1} imponiendo la distancia de enlace normalizada (d_{norm}) y los átomos siguientes de la cadena se desplazan para no alterar la estructura de la proteína.

3.2. DETECCIÓN DE ESTRUCTURAS SECUNDARIAS

De cara a reducir el coste computacional asociado al proceso de simulación, dentro de esta propuesta se incluye un algoritmo de detección de estructuras secundarias. Para ello, se utilizan los datos estadísticos y teóricos extraídos de [35]. En dicho trabajo, se muestran dos estudios diferentes. Por un lado, se presentan los valores teóricos que tienen los ángulos diédricos para formar una estructura secundaria. Por otro lado, se realiza un estudio en el que se muestra la probabilidad de que un aminoácido pertenezca a una estructura secundaria en función de los valores de sus ángulos diédricos.

Basándose en los resultados de dicho estudio, el proceso propuesto utiliza únicamente los valores de los ángulos diédricos para la detección de los distintos tipos de estructuras secundarias (hélices- α u hojas- β) presentes en la proteína. El proceso recorre la cadena de la proteína calculando los valores de los ángulos diédricos, y cuando encuentra tres aminoácidos consecutivos con ángulos que pueden pertenecer a una estructura secundaria, los asocia como parte de ella y va extendiendo la estructura por la cadena hasta que encuentra un aminoácido con unos valores de los ángulos diédricos que no se corresponden con una estructura secundaria. Tanto el número de aminoácidos necesarios para detectar una estructura secundaria, como el valor de la tolerancia respecto del valor teórico de los ángulos diédricos ($\pm 30^\circ$ en línea con los resultados mostrados en [46]) se han seleccionado para obtener la mejor correlación entre los resultados del procedimiento y los datos experimentales. Pese a la aparente simplicidad del procedimiento, su eficacia queda demostrada como muestran los resultados de la Figura 7, donde se ha aplicado este procedimiento a 7 proteínas diferentes.

Proteína	% de aminoácidos situados correctamente
1zac	96.6
1k9p	96.6
3cln	97.2
1k20	93.48
2peq	100
4fkx	82.35
3sza	86.6

Fig. 7: Resultados de proceso de detección de estructuras secundarias en 7 proteínas diferentes

3.3. MOVIMIENTO DE LA PROTEÍNA

El método de los ángulos diédricos adaptativos, a diferencia de otros métodos, no realiza ninguna minimización del valor de la energía potencial de la proteína ni utiliza ninguna fuerza asociada a la misma para obtener el movimiento de la proteína. Por ello, se logra uno de los objetivos principales que es tener un coste computacional muy bajo. El método se centra en obtener posibles trayectorias comprobando que, durante el proceso de simulación, no se generen estructuras biológicamente no posibles. Para esto, el método realiza en cada iteración una búsqueda de los incrementos óptimos a aplicar en cada uno de los ángulos diédricos. Dichos incrementos se aplican de forma secuencial en los ángulos diédricos para obtener una nueva posición. Los valores óptimos de los incrementos vienen determinados por el incremento total de cada ángulo diédrico entre la posición inicial y final, el número de posiciones intermedias y el valor de la energía potencial. Para cada incremento, se evalúa su efecto sobre el valor de la energía potencial de la proteína y, en base al valor inicial de ésta y el máximo cambio en la energía permitido en cada etapa de la simulación, se ajusta el valor del incremento a aplicar. Asimismo, el procedimiento va determinando si un grado de libertad es bloqueado en alguna etapa de la simulación o incluso si se invierte su sentido de giro para tratar de buscar un movimiento más realista. En la Figura 8, se puede ver el pseudocódigo del algoritmo para una iteración de la simulación del movimiento de la cadena principal de la proteína.

Durante el cálculo del movimiento de la proteína, también se tiene en cuenta el movimiento de las cadenas secundarias. Para

```
1: foreach Grado de libertad  $i$  de la proteína do
2:   Girar el grado de libertad  $i$  ( $\Delta\psi_i||\Delta\phi_i$ )
3:    $E_i^k \leftarrow$  Cálculo de la energía potencial después del giro del grado de libertad  $i$ 
4:    $\Delta E_i^k = E_i^k - E_{i-1}^k \leftarrow$  Valor del incremento de energía asociado al grado de libertad  $i$ 
5: end foreach
6:  $E^k \leftarrow$  Cálculo de la energía potencial después del paso  $k$ 
7: while  $\left[\left(E^k - E^0\right) / E^0\right] \geq \epsilon^k$  do
8:   Deshacer el giro del grado de libertad  $i$  ( $-\Delta\psi_i||-\Delta\phi_i$ ) asociado al mayor incremento  $\Delta E_i^k$ 
9:   Girar el grado de libertad  $i$  con un incremento angular reducido ( $\Delta\psi_i/h||\Delta\phi_i/h$ )
10:   $E^k \leftarrow$  Valor de la energía potencial
11:  if  $\left[\left(E^k - E^0\right) / E^0\right] \geq \epsilon^k$  then
12:    Deshacer el giro del grado de libertad  $i$  ( $-\Delta\psi_i/h||-\Delta\phi_i/h$ )
13:  end if
14:   $\Delta E_i^k = 0$ 
15:  if El grado de libertad  $i$  ha sido bloqueado más de  $m$  veces then
16:     $\Delta\psi_i = -\Delta\psi_i||\Delta\phi_i = -\Delta\phi_i$ 
17:  end if
18:  if El grado de libertad  $i$  ha girado en sentido contrario  $n$  veces then
19:     $\Delta\psi_i = -\Delta\psi_i||\Delta\phi_i = -\Delta\phi_i$ 
20:  end if
21:   $E^k \leftarrow$  Cálculo de la energía potencial después del paso  $k$ 
22: end while
```

Fig. 8: Pseudocódigo del algoritmo principal del método de los ángulos diédricos adaptativos

ello, únicamente se considera un grado de libertad por cadena secundaria. Después de cada iteración de movimiento de la cadena principal, el método trata de reducir el valor de la energía potencial de cara a la siguiente iteración girando las cadenas secundarias de la proteína. El proceso es similar: se calcula la variación que se genera en la energía potencial con el giro de cada cadena secundaria para luego aplicar únicamente aquellos giros que reduzcan dicho valor. El coste computacional del método varía en gran medida en función de si se incluyen o no estos grados de libertad secundarios variando desde unas pocas horas a varios días de simulación en un Pentium Core 2 Duo con 4Gb de RAM.

4. CONCLUSIONES

En el presente artículo, se ha realizado una amplia revisión bibliográfica de un área de estudio que actualmente no ha sido muy explotada desde el punto de vista de la ingeniería mecánica, como es el estudio de las proteínas. Se han presentado las principales problemáticas que pueden ser abordadas desde la perspectiva del área de conocimiento de la ingeniería, las cuales, además, son fundamentales para avanzar en el conocimiento de la formación y función de las proteínas dentro de los seres vivos. Se han presentado también las principales metodologías que ya se han aplicado al estudio de estos problemas y que tienen unas fuertes raíces en la Teoría de Mecanismos y la Teoría de Estructuras. Si bien ningún método ha logrado resolver de manera eficiente ninguno de los problemas asociados a las proteínas, sí se ha demostrado que estos problemas pueden ser abordados de forma satisfactoria desde diferentes enfoques, como el biocinématico, y con diferentes modelizaciones de la estructura de una proteína.

Actualmente, y con la aparición de nuevas técnicas de computación paralela o inteligencia artificial, estos mismos métodos pueden ser revisados y actualizados para valorar si su optimización o implantación mediante otros algoritmos puede ser relevante para la resolución de las áreas de estudio aquí presentadas.

REFERENCIAS

- [1] Lexa K.W., Carlson H.A. Protein Flexibility in Docking and Surface Mapping. *Q. Rev. Biophys.* 2012 Aug; 45(3): 301–343. DOI: <http://doi.org/10.1017/S0033583512000066>
- [2] Carlson H.A. Protein flexibility is an important component of structure-based drug discovery. *Curr Pharm Des.* 2002;8(17):1571–8. DOI: <http://doi.org/10.2174/1381612023394232>
- [3] Díez M., Petuya V., Martínez-Cruz LA., Hernández A. A biokinematic approach for the computational simulation of proteins molecular mechanism. *Mechanism and Machine Theory* 46, 1854–1868, 2011. DOI: <https://doi.org/10.1016/j.mechmachtheory.2011.07.013>
- [4] Zhang M., Kavraki L.E. A New Method for Fast and Accurate Derivation of Molecular Conformations. *J. Chem. Inf. Comput. Sci.*, 2002, 42 (1), pp 64–70. DOI: <http://doi.org/10.1021/ci010327z>
- [5] Kazerounian K., Fellow A., Latif K., Alvarado C. Protofold: A successive kinetostatic compliance method for protein conformation prediction. *Journal of Mechanical Design*, 127:712, 2005. DOI: <http://doi.org/10.1115/1.1867502>
- [6] Anfinsen C. B. The formation and stabilization of protein structure. *Biochemical Journal*, 128(4): 737–749, 1972. PMID: PMC1173893
- [7] Anfinsen C. B. Nobel Laureate in Chemistry 1972. Principles that govern the folding of protein chains. The Nobel Lecture. In *Science Volume 181*, pp 223–230. American Association for the Advancement of Science, 1973
- [8] Levinthal C. Are there pathways for protein folding? *Journal of Medical Physics*, 1968. DOI: <http://doi.org/10.1051/jcp/1968650044>
- [9] Zwanzig R., Szabo A., Bagchi B. Levinthal's paradox. *PNAS*, 89 (1) 20–22 1992. PMID: PMC48166
- [10] Asbury C. L., Fehr A. N., Block S. M. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science*, 302(5653):2130–2134, 2003. DOI: <https://doi.org/10.1126/science.1092985>
- [11] Lucas M., Encinar J.A., Arribas E.A., Oyenarte I., García I.G., Kortazar D., Fernández J.A., Mato J.M., Martínez-Chantar M.L., Martínez-Cruz LA. Binding of S-Methyl-5-Thioadenosine and S-Adenosyl-L-Methionine to Protein MJO100 Triggers an Open-to-Closed Conformational Change in Its CBS Motif Pair. *Journal of molecular Biology*, 396(3):800–820, 2010. DOI: <http://doi.org/10.1016/j.jmb.2009.12.012>
- [12] Wedemeyer W.J., Scheraga H.A. Exact analytical loop closure in proteins using polynomial equations. *Journal of computational chemistry*, 20(8):819–844, 1999. DOI: [http://doi.org/10.1002/\(SICI\)1096-987X\(199906\)20:8<819::AID-JCC10020801001>3.0.CO;2-Y](http://doi.org/10.1002/(SICI)1096-987X(199906)20:8<819::AID-JCC10020801001>3.0.CO;2-Y)
- [13] Woolfson M. M. An Introduction to X-ray Crystallography. Cambridge University Press, 2 edition, January 1997. DOI: <https://doi.org/10.1017/CBO9780511622557>
- [14] Roberts J. D. Nuclear Magnetic Resonance: Applications to Organic Chemistry. McGraw-Hill Series 1959
- [15] Lesk A. M. Introduction to Protein Science: Architecture, Function, and Genomics. Oxford University Press, USA, 2 edition, May 2010. ISBN: 9780198716846
- [16] Marx, D., Hutter, J. Ab Initio Molecular Dynamics. Basic Theory and Advanced Methods. Cambridge Univ Pr, April 2009. ISBN-10: 0521898633
- [17] Xia Y., Huang E.S., Levitt M., Samudrala, R. Ab initio construction of protein tertiary structures using a hierarchical approach. *Journal of molecular Biology*, 300(1):171–85, 2000. DOI: <http://doi.org/10.1006/jmbi.2000.3835>
- [18] Chaplin M. Do we underestimate the importance of water in cell biology? *Nature Reviews: Molecular Cell Biology*, 7(11):861–866, 2006. DOI: <https://doi.org/10.1038/nrm2021>
- [19] Orry A.J., Abagyan R. editors. Homology Modeling: Methods and Protocols (Methods in Molecular Biology). Humana Press, 2012 edition, 2012. DOI: 10.1007/978-1-61779-588-6
- [20] Wolf Y.I., Grishin N.V. Estimating the number of protein folds and families from complete genome data. *Journal of molecular Biology*, 299(4): 897–905, 2000. DOI: <https://doi.org/10.1006/jmbi.2000.3786>
- [21] Kunszt P., Malmström L., Fantini N., Sudholt W., Lautenschlager M., Reifler R., Ruckstuhl S. Accelerating 3D Protein Modeling Using Cloud Computing: Using Rosetta as a Service on the IBM SmartCloud. *IEEE Seventh International Conference on e-Science Workshops*, 166–169, 2011. DOI: 10.1109/eScienceW.2011.12
- [22] Berman H., Henrick K., Nakamura H., Markley J.L. The worldwide protein data bank (www.pdb): ensuring a single, uniform archive of pdb data. *Nucl. Acids Res.*, 35(suppl1): D301–303, 2007. DOI: <https://doi.org/10.1093/nar/gkl971>
- [23] Kavraki L.E. Protein-Ligand Docking, Including Flexible Receptor-Flexible Ligand Docking. 2007.
- [24] Song G., Amato NM. A motion-planning approach to folding: from paper craft to protein folding. *Robotics and Automation, IEEE Transactions on*, 20(1):60–71, 2004. DOI: 10.1109/ROBOT.2001.932672
- [25] Novinskaya A., Devaurs D., Moll M., Kavraki L.E. Defining Low-Dimensional Projections to Guide Protein Conformational Sampling. *J Comput Biol.* 2017 Jan;24(1):79–89. DOI: <http://doi.org/10.1089/cmb.2016.0144>.
- [26] Go N., Noguti T., Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proceedings of the National Academy of Sciences of the United States of America*, 80:3696–3700, 1983. PMID:6574507
- [27] Brooks B., Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proceedings of the National Academy of Sciences of the United States of America*, 80:6571–6575, 1983. PMID:6579545
- [28] Tirion M.M. Large amplitude elastic motions in proteins from a single parameter, atomic analysis. *Physical Review Letters*, 77(9):1905–1908, 1996. DOI: <http://doi.org/10.1103/PhysRevLett.77.1905>
- [29] López-Blanco J.R., Aliaga J.I., Quintana-Ortí E.S., Chacón P. iMODS: internal coordinates normal mode analysis server. *Nucleic Acids Res.* 2014 Jul; 42: W271–6. DOI: <http://doi.org/10.1093/nar/gku339>. Epub 2014 Apr 25.
- [30] Schuyler A.D., Jernigan R.L., Qasba P.K., Ramakrishnan B., Chirikjian G.S. Iterative cluster-NMA: A tool for generating conformational transitions in proteins. *Proteins: Structure, Function, and Bioinformatics*, 74(3):760–776, 2009. DOI: <http://doi.org/10.1002/prot.22200>
- [31] Bohnenkamp P., Kazerounian K., Ilies H.T. Structural Prediction of Peptide Based Nano Systems via Progressive Landscape Evolution. *Proceedings 12th IFToMM world congress*, 1–6, 2007.
- [32] Tavousi P., Behandish M., Ilies H.T., Kazerounian K. Protofold II: Enhanced Model and Implementation for Kinetostatic Protein Folding. *ASME Transactions, Journal of Nanotechnology in Engineering and Medicine*, 6(3), p.034601, 2016. DOI: <http://doi.org/10.1115/1.4032759>
- [33] Díez M., Petuya V., Martínez-Cruz LA., Hernández A. Biokinematic protein simulation by an adaptive dihedral angle approach. *Mechanism and Machine Theory* (69), 105–114, 2013. DOI: <https://doi.org/10.1016/j.mechmachtheory.2013.05.007>
- [34] Díez M., Petuya V., Martínez-Cruz LA., Hernández A. Insights into mechanism kinematics for protein motion simulation. *Bmc bioinformatics*, 15:184, 2014. DOI: <https://doi.org/10.1186/1471-2105-15-184>
- [35] Frishman D., Argos P. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4):566–579, 1995. DOI: <https://doi.org/10.1002/prot.340230412>

AGRADECIMIENTOS

Los autores desean agradecer el soporte financiero recibido por parte del Gobierno a través del Ministerio de Economía y Competitividad (Proyectos DPI2015-67626-P y DPI2015-64450-R (MINECO/FEDER, UE)) y el apoyo al grupo de investigación, a través del proyecto con Ref. IT949-16, del Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco.

MATERIAL SUPLEMENTARIO

https://www.revistadyna.com/documentos/pdfs/_adic/9033-1.pdf

