



RIDAA
Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes

Chiappinelli, Matías Daniel

Efecto de mutaciones sobre la conservación de la diversidad conformacional de proteínas



Esta obra está bajo una Licencia Creative Commons Argentina.
Atribución - No Comercial - Sin Obra Derivada 2.5
<https://creativecommons.org/licenses/by-nc-nd/2.5/ar/>

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:

Chiappinelli, M. D. (2023). *Efecto de mutaciones sobre la conservación de la diversidad conformacional de proteínas (Trabajo final)*. Universidad Nacional de Quilmes. Departamento de Ciencia y Tecnología. Laboratorio de Química y Biología Computacional. Argentina. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes <http://ridaa.unq.edu.ar/handle/20.500.11807/4526>

Puede encontrar éste y otros documentos en: <https://ridaa.unq.edu.ar>



DEPARTAMENTO DE CIENCIA Y TECNOLOGÍA

UNIVERSIDAD NACIONAL DE QUILMES

EFFECTO DE MUTACIONES SOBRE LA CONSERVACIÓN DE LA DIVERSIDAD CONFORMACIONAL DE PROTEÍNAS

Seminario de Grado de la Licenciatura en Biotecnología

Modalidad: Seminario de Investigación

Estudiante:

Matías Daniel Chiappinelli

Legajo UNQ:

35.481

Director de Seminario de Grado:

Dr. Tadeo Enrique Saldaño

Marzo de 2023

ÍNDICE

1. RESUMEN.....	3
2. INTRODUCCIÓN.....	3
3. HIPÓTESIS Y OBJETIVOS	6
4. MARCO METODOLÓGICO.....	7
4.1 SELECCIÓN DEL CONJUNTO DE PROTEÍNAS	7
4.2 ANÁLISIS DE MODOS NORMALES.....	7
4.2.1 Modelos de Redes Anisotrópicas (ANM)	7
4.2.2 Modelo ENCoM.....	11
4.3 VECTOR DIFERENCIA ENTRE CONFÓRMEROS – PROYECCIÓN SOBRE LA BASE DE LOS MODOS NORMALES	12
4.4 NÚMERO DE PARTICIPACIÓN Y SUBESPACIO DE MODOS NORMALES	12
4.5 SIMULACIÓN DE MUTACIONES PUNTUALES.....	13
4.6 COMPARACIÓN DE LOS SUBESPACIOS DE MODOS NORMALES ENTRE PROTEÍNA <i>WILD-TYPE</i> Y PROTEÍNA MUTADA.....	13
4.7 CARACTERIZACIÓN DE RESIDUOS DINÁMICAMENTE IMPORTANTES EN EL CAMBIO CONFORMACIONAL	14
5. RESULTADOS Y DISCUSIÓN.....	16
5.1 CARACTERIZACIÓN DEL <i>SET</i> DE PROTEÍNAS.....	16
5.2 IDENTIFICACIÓN DE RESIDUOS EN POSICIONES CLAVES	18
5.3 UNG_HUMAN – <i>Uracil-ADN glicosilasa</i>	20
5.4 HutP – <i>Proteína anti-terminación de unión a ARN</i>	24
5.5 MetJ – <i>Represor del regulón met</i>	28
5.6 DsbC – <i>Proteína intercambiadora de tiol:disulfuros</i>	33
5.7 Hsp20 – <i>Heat shock protein</i>	37
6. CONCLUSIÓN.....	42
7. REFERENCIAS BIBLIOGRÁFICAS	43

1. RESUMEN

El presente seminario de investigación presenta una metodología novedosa para identificar los residuos claves para la preservación del estado nativo de las proteínas. La metodología utilizada se basó en el análisis de los cambios conformacionales posibles entre los distintos conformeros en equilibrio dinámico en el estado nativo de varias proteínas, empleando el análisis de modos normales con un potencial sensible al tipo de interacciones específicas entre los aminoácidos. Esta metodología resulta en una mejora respecto a trabajos anteriores, ya que se incorpora la capacidad de distinguir los distintos tipos de mutaciones y se extiende a la posibilidad de estudiar el conjunto de cambios conformacionales posibles entre todos los distintos pares de conformeros existentes en el estado nativo. Los resultados obtenidos mostraron que varios de los residuos claves obtenidos son importantes por su participación en interacciones entre estructuras secundarias, sitios activos y/o cavidades relacionadas con la función biológica, así como por su conservación evolutiva y su relación con patologías reportadas. Este trabajo representa un primer paso prometedor para encarar un estudio más abarcativo, considerando un mayor número de proteínas, análisis más detallados de conservación evolutiva y la variación del impacto sobre los cambios conformacionales entre las diferentes mutaciones específicas por sitio, clasificadas según las propiedades fisicoquímicas de los residuos.

2. INTRODUCCIÓN

Según la teoría neutra de la evolución molecular [Kimura-1974], los residuos más relevantes para la función de las proteínas varían más lentamente que los menos importantes. En base a este concepto se han desarrollado métodos computacionales orientados a la asignación de funciones biológicas de proteínas basados en la secuencia, tales como los métodos de búsqueda de similitud de secuencia, y basados en la estructura, tales como los métodos que utilizan perfiles y modelos ocultos de Markov (*Hidden Markov Models*) [Thornton-2005]. Sin embargo, estos métodos no proporcionan una información completa sobre la naturaleza de la relación secuencia-estructura-función y se requiere información adicional relacionada con la dinámica de las proteínas [Bahar-2004, Benkovic-2006, Orengo-2008, Gerstein-2004, Uniprot-2010, Walsh-2010].

La dinámica de proteínas suele estar comúnmente asociada a la función biológica estableciendo un puente en la relación estructura-función. Actualmente se cuenta con suficiente evidencia teórica y experimental que soporta la preexistencia de poblaciones de conformeros en equilibrio dinámico sobre la superficie de energía libre de una proteína [Wolynes-1991, Tawfik-2009]. Bajo este concepto de dinamismo, el estado nativo de una proteína puede ser representado como una distribución no uniforme de poblaciones de conformeros en equilibrio dinámico [Noussinov-2004]. Esta coexistencia de conformeros con diferentes afinidades por el ligando y/o roles en la función biológica es la base del modelo de selección conformacional o pre-equilibrio. Las barreras energéticas que separan estos conformeros son comúnmente superadas por las fluctuaciones térmicas. Por lo tanto, las vibraciones de las proteínas, en condiciones de equilibrio térmico, modulan la diversidad conformacional y la población relativa de los conformeros. Las diferencias en las propiedades estructurales y dinámicas de los conformeros pueden tener importantes consecuencias funcionales, indicando que la diversidad conformacional de proteínas está íntimamente ligada a su función biológica.

La dinámica vibracional específica de cada confórmero es responsable de los cambios conformacionales que posee la proteína y, por esta razón, las vibraciones asociadas a los cambios conformacionales suelen ser conservadas evolutivamente. Esto permite la posibilidad de identificar posiciones claves en la proteína cuyas mutaciones pueden afectar significativamente estas vibraciones.

Los modelos teóricos utilizados en la actualidad para analizar la flexibilidad y dinámica de proteínas son básicamente dos: dinámica molecular (MD, del inglés *Molecular Dynamic*) y análisis de modos normales (NMA, del inglés *Normal Mode Analysis*) [Mc Cammon-1987, Montgomery-1988, Brooks-1995, Karplus-1995]. Por un lado, los métodos de MD consisten en propagar numéricamente las ecuaciones de movimiento de las macromoléculas. Estos métodos permiten estudiar el movimiento anarmónico de las proteínas embebidas en solvente explícito, lo cual permite identificar el efecto de interacciones proteína-solvente específicas, así como el análisis de entropías conformacionales y barreras de energía libre. Sin embargo, tanto los requerimientos particulares de cada sistema de estudio como su alto costo computacional impiden la utilización sistemática de MD para un gran conjunto de proteínas. Por otro lado, el NMA consiste en una descomposición del movimiento vibracional de la proteína en términos de movimientos armónicos en torno a la posición de equilibrio donde el solvente no suele ser considerado. Cada modo normal representa el desplazamiento relativo de los átomos de la proteína de acuerdo con una frecuencia característica del modo. Esto permite el análisis simultáneo de los movimientos vibracionales a distintas escalas de tiempo.

La implementación tradicional tanto de MD como NMA suele utilizar campos de fuerza de Mecánica Molecular que consideran interacciones específicas entre todos los átomos de la proteína. Estos son potenciales de interacción empíricos entre todos los átomos de la estructura y se basan en la partición de la energía total en contribuciones de distinto tipo, como ser, estiramientos de enlaces, movimientos de flexión angulares, rotaciones en torno a ángulos dihedros, interacciones coulómbicas, dipolo-dipolo, puentes de hidrógeno, y van der Waals [Brooks-2004, Case-2003]. El cálculo del potencial y sus derivadas representa la principal contribución al costo computacional de ambas metodologías. En el caso de NMA debe también considerarse el requerimiento computacional para la diagonalización de la matriz de derivadas segundas del potencial. En el presente trabajo se estudia la dinámica involucrada en el cambio conformacional de un *set* que involucra varias proteínas. Esto impide el uso de MD y NMA a partir de campos de fuerza de Mecánica Molecular que tengan en cuenta todos los átomos de la molécula. Por tal motivo, utilizamos NMA basados en la utilización de potenciales de grano grueso (*coarse grained*) [Bahar-2006]. Estos modelos permiten representar a una proteína como una red tridimensional de resortes entre residuos.

Los métodos basados en NMA que utilizan modelos de red elástica (ENM, del inglés *Elastic Network Model*) [Orozco-2008] han demostrado ser en gran medida eficientes para explorar la dinámica global de proteínas comúnmente asociadas a las transiciones conformacionales [Chirikjian-2009]. Además, se observa que solo unos pocos modos normales de baja frecuencia son necesarios para lograr una buena descripción de la mayoría de los cambios conformacionales asociados a la función biológica de las proteínas [Sanejouand-2001]. En particular, utilizamos el modelo de redes anisotrópicas (ANM, del inglés *Anisotropic Network Model*) [Bahar-2001] que permite la adecuada descripción de los modos de más baja frecuencia en proteínas.

NMA desacopla los movimientos complejos y las fluctuaciones de las proteínas en una combinación lineal de osciladores armónicos independientes, es decir, los modos normales, cada uno de los cuales implica los movimientos concertados de muchos átomos. De esa manera, se pueden identificar movimientos de dominio a gran escala, involucrados en la conexión de los diferentes estados conformacionales relacionados con la función. El efecto de las mutaciones sobre estos modos colectivos y funcionalmente relevantes ha sido previamente estudiado desde diferentes puntos de vista [Bahar-2000], [Thirumalai-2006], [Maguid-2005]. La comprensión molecular de la función biológica requiere la identificación de la red de residuos que participan en la dinámica relacionada con la función, como la unión del sustrato y la liberación de productos, regulaciones alostéricas y plegado.

En los últimos años, el grupo de investigación de la Universidad Nacional de Quilmes, dirigido por el Prof. Sebastián Fernández-Alberti, ha desarrollado un método que permite la identificación de redes de residuos dinámicamente importantes para la diversidad conformacional de proteínas [Fernandez-Alberti-2016]. Brevemente, el método se basa en la previa selección de modos normales involucrados en el cambio conformacional debido a la unión al ligando. El efecto de mutaciones puntuales sobre el subespacio definido por estos modos se analiza introduciendo perturbaciones locales en el modelo de red elástica de la proteína. Esto permitió identificar posiciones claves cuyas mutaciones perturban significativamente a este subespacio vibracional. Estas posiciones claves representan residuos dinámicamente importantes para el cambio conformacional de unión al ligando. Finalmente, se constató que estos residuos dinámicamente importantes se encuentran conservados evolutivamente. Es así como esta línea de estudio permite identificar residuos críticos para la función de las proteínas a través de su participación en la relación secuencia-estructura-función asociada a la preservación de la diversidad conformacional.

3. HIPÓTESIS Y OBJETIVOS

En el contexto anteriormente expresado, el presente seminario propone analizar el efecto de mutaciones sobre la diversidad conformacional de proteínas evaluando su impacto sobre las vibraciones involucradas en los cambios conformacionales.

El objetivo principal de esta investigación es la adaptación y utilización de métodos computacionales que permitan predecir el efecto de mutaciones sobre la dinámica vibracional responsable de la diversidad conformacional de las proteínas con el fin de identificar sitios relevantes para su preservación.

Para tal fin, se proponen los siguientes objetivos específicos:

- Definir los subespacios de modos normales que participan de los cambios conformacionales requeridos para garantizar el equilibrio dinámico entre los distintos conformeros que caracterizan al estado nativo de las proteínas.
- Estudiar la respuesta de los subespacios de modos normales que participan de forma activa ante mutaciones puntuales en los distintos residuos que componen la proteína.
- Explorar las variaciones de los subespacios vibracionales para identificar los residuos claves de las proteínas y relacionarlos con su implicancia funcional.

Con el fin de realizar un análisis más detallado sobre la conservación de la diversidad conformacional y su implicancia funcional, se incluyeron en el análisis de cada proteína datos de conservación evolutiva de los sitios relevantes mediante la búsqueda en la base de datos ConSurf-DB [Ben-Tal-2020], [Ben-Tal-2009] y datos sobre cavidades de las proteínas mediante la búsqueda en la base de datos CaviDB [Parisi-2022].

4. MARCO METODOLÓGICO

El método actual [Fernandez-Alberti-2016], desarrollado por el grupo de trabajo, permite predecir sitios dinámicamente relevantes para preservar el cambio conformacional *apo-holo*. En este trabajo, el método fue adaptado para predecir sitios dinámicamente relevantes para conservar el espacio conformacional representado por conjuntos de confórmers existentes obtenidos de la base de datos de diversidad conformacional (CoDNaS) [Parisi-2016], i.e. no solo el par *apo-holo*. Por otro lado, se incorporó un modelo de cálculo de modos normales que considera las interacciones específicas entre los distintos tipos de residuos. Este permite simular de manera más precisa el efecto de mutaciones respecto a los trabajos anteriores, que consideraban las mutaciones como simples perturbaciones de constantes de resorte locales sin hacer distinción entre los distintos tipos de residuos mutados.

4.1 SELECCIÓN DEL CONJUNTO DE PROTEÍNAS

CoDNaS es una base de datos redundante de distintas estructuras para la misma proteína, curada y ampliamente relacionada con distintas bases de datos. A partir de la misma, explorando familias de proteínas homólogas con diversidad conformacional basada experimentalmente, se seleccionaron *sets* de proteínas que abarcan distintos grados de extensión de la diversidad conformacional. Cada proteína en CoDNaS tiene asociado un valor máximo de la raíz de la desviación cuadrática media (RMSD, del inglés *Root Mean Square Deviation*) derivado de la comparación de todos los confórmers que pertenecen a una proteína dada.

Este RMSD es considerado como un valor de la disimilitud estructural entre un par de proteínas alineadas y se calcula como

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad (\text{Ec. 4.1.1})$$

donde N es la cantidad de residuos y d_i es la distancia euclidiana entre los C α en la posición i-ésima medida en Å (Armstrong).

En el presente trabajo, se ha utilizado el valor de RMSD como medida de la disimilitud estructural entre pares de confórmers de una misma proteína. A partir de ello, se seleccionaron proteínas cuyos confórmers presenten entre sí un RMSD global mayor a 0,8 Å con el fin de garantizar la diversidad conformacional para cada proteína estudiada.

4.2 ANÁLISIS DE MODOS NORMALES

4.2.1 Modelos de Redes Anisotrópicas (ANM)

El modelo teórico utilizado para NMA en este seminario es el Modelo de Redes Anisotrópicas (ANM) debido a que resulta ser una alternativa de bajo costo computacional para NMA. Originalmente, este modelo considera a la proteína como una red tridimensional de resortes con una única constante de fuerza armónica. Los carbonos alfa y beta (C α y C β) del *backbone* proteico son tomados como los nodos de la red. Al considerar solo los

$C\alpha$ y $C\beta$ se reduce significativamente el número de grados de libertad del sistema. Por otro lado, el modelo considera a la estructura cristalográfica como la correspondiente estructura de equilibrio. Por tal motivo el método no requiere la minimización del sistema para la obtención del Hessiano. De este modo, también se reduce significativamente el costo computacional de los cálculos.

Para determinar si dos nodos deben estar unidos por un resorte, se fija un radio de corte r_c en Å tal que

$$\text{si } |r_j^0 - r_i^0| \leq r_c \Rightarrow k_{ij} = \gamma$$

$$\text{si } |r_j^0 - r_i^0| > r_c \Rightarrow k_{ij} = 0$$

donde r_j^0 y r_i^0 son las posiciones de equilibrio de los residuos j e i , respectivamente (Figura X1) y k_{ij} es la constante del resorte cuyo valor depende de la constante de fuerza γ y del radio de corte r_c . Según este planteo, entonces, se contemplan tanto uniones entre dos residuos adyacentes como entre cualquier par de residuos dentro del radio r_c definido. Esto significa que no se discierne entre uniones del tipo covalente, puente de hidrógeno e interacciones de Van der Waals.

Si consideramos dos residuos i y j unidos por un resorte, el potencial elástico armónico al cual están sujetos será:

$$V = \frac{1}{2} k_{ij} (s_{ij} - s_{ij}^0)^2 = \frac{1}{2} k_{ij} \left\{ \left[(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2 \right]^{\frac{1}{2}} - s_{ij}^0 \right\}^2 \quad (\text{Ec. 4.2.1})$$

donde $s_{ij} = r_j - r_i$ y $s_{ij}^0 = r_j^0 - r_i^0$. r_j y r_i son las posiciones instantáneas fuera del equilibrio de los residuos j e i .

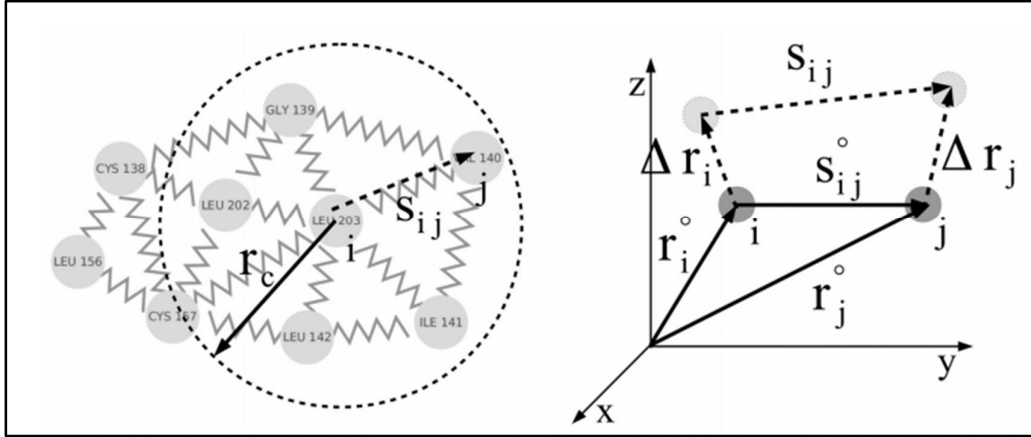


Figura 1. Representación esquemática de los nodos en el ANM. Cada nodo se conecta a sus vecinos mediante resortes, dependiendo del radio de corte r_c . s_{ij}^0 es la distancia de equilibrio entre los sitios i y j , cuyas posiciones de equilibrio son r_i^0 y r_j^0 . Los vectores fluctuación Δr_i , Δr_j y la distancia entre los residuos s_{ij} , se grafican con línea punteada.

La primera y segunda derivada del potencial V con respecto a las componentes de r_i son:

$$\frac{\partial V}{\partial x_i} = -\frac{\partial V}{\partial x_j} = -k_{ij}(x_j - x_i) \left(1 - \frac{s_{ij}^0}{s_{ij}}\right) \quad (\text{Ec. 4.2.2})$$

$$\frac{\partial^2 V}{\partial x_i^2} = \frac{\partial^2 V}{\partial x_j^2} = k_{ij} \left(1 + s_{ij}^0 \frac{(x_j - x_i)^2}{s_{ij}^3} - \frac{s_{ij}^0}{s_{ij}}\right) \quad (\text{Ec. 4.2.3})$$

Ecuaciones análogas se obtienen para las componentes y y z de r_i .

Para el caso en que $s_{ij} = s_{ij}^0$ las Ec. 4.2.2 y 4.2.3 se reducen a:

$$\frac{\partial V}{\partial x_i} = 0 \quad (\text{Ec. 4.2.4})$$

$$\frac{\partial^2 V}{\partial x_i^2} = k_{ij} \frac{(x_j - x_i)^2}{s_{ij}^2} \quad (\text{Ec. 4.2.5})$$

y las segundas derivadas cruzadas:

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{\partial^2 V}{\partial x_j \partial y_i} = -k_{ij} \frac{(x_j - x_i)(y_j - y_i)}{s_{ij}^2} \quad (\text{Ec. 4.2.6})$$

El potencial elástico en la posición de equilibrio s_{ij}^0 , $V(s_{ij}^0) = 0$.

Para el caso general de N residuos unidos por M resortes, las ecuaciones de las segundas derivadas del potencial pueden ser organizadas en una matriz Hessiana $3N \times 3N$, H.

$$H = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & & & H_{2N} \\ \vdots & & & \vdots \\ H_{N1} & H_{N2} & \dots & H_{NN} \end{bmatrix} \quad (\text{Ec. 4.2.7})$$

donde cada superelemento ij-ésimo ($i \neq j$) H_{ij} de H es

$$H_{ij} = \begin{bmatrix} \partial^2 V / \partial x_i \partial x_j & \partial^2 V / \partial x_i \partial y_j & \partial^2 V / \partial x_i \partial z_j \\ \partial^2 V / \partial y_i \partial x_j & \partial^2 V / \partial y_i \partial y_j & \partial^2 V / \partial y_i \partial z_j \\ \partial^2 V / \partial z_i \partial x_j & \partial^2 V / \partial z_i \partial y_j & \partial^2 V / \partial z_i \partial z_j \end{bmatrix} \quad (\text{Ec. 4.2.8})$$

Para el caso de los superelementos H_{ij} , las ecuaciones Ec. 4.2.5 y 4.2.6 pasan a ser,

$$\frac{\partial^2 V}{\partial x_i^2} = k_{ij} \sum_j \frac{(x_j - x_i)^2}{s_{ij}^2} \quad (\text{Ec. 4.2.9})$$

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{\partial^2 V}{\partial x_j \partial y_i} = -k_{ij} \sum_j \frac{(x_j - x_i)(y_j - y_i)}{s_{ij}^2} \quad (\text{Ec. 4.2.10})$$

donde la sumatoria se realiza sobre todos los vecinos del residuo i ($j = 1, M$) conectado por resorte.

Finalmente, los modos normales del ANM se obtienen diagonalizando la matriz H .

$$\tilde{Q} H Q = \Lambda$$

donde Q es la matriz de autovectores Q_k de H , esto es, los modos normales. Λ es la matriz diagonal de autovalores λ_k y $\tilde{Q}$ es la matriz traspuesta de Q . Los autovalores λ_k representan los cuadrados de las frecuencias ω_k de los modos individuales Q_k ,

$$\lambda_k = \omega_k^2 \quad (\text{Ec. 4.2.11})$$

Los modos normales Q_k representan movimientos armónicos colectivos que involucran desplazamientos concertados de un conjunto de átomos y su movimiento puede escribirse como

$$Q_k = A_k \cos(\omega_k t) \quad (\text{Ec. 4.2.12})$$

siendo A_k la amplitud del autovector, asociado a la frecuencia ω_k .

Como consecuencia del sistema de referencia utilizado aparecen seis frecuencias cero en la diagonalización, las cuales están asociadas con las tres direcciones de traslación pura de un cuerpo rígido y los tres ángulos que definen la rotación en un cuerpo rígido. Es decir, existen 6 direcciones de movimiento que corresponden a características de cuerpo rígido exclusivamente y no dan información de las vibraciones internas, quedando entonces $3N - 6$ autovalores distintos de cero.

El conjunto de modos normales $\{Q_k\}_{k=1, 3N-6}$ es un conjunto ortonormal que, como tal, forman una base de los posibles desplazamientos del sistema en el entorno de la conformación de equilibrio. Cuando se expresa un modo normal en el sistema de coordenadas cartesianas describe una fluctuación concertada de los átomos del sistema que vibran con la misma frecuencia. Los perfiles de los modos de baja frecuencia revelan los mecanismos de movimientos colectivos o globales que involucran el desplazamiento de grandes bloques o dominios de las proteínas [Erman-1998, Covell-1999]. Los residuos con movimiento más restringido en dichos modos (sus mínimos) suelen tener un comportamiento crítico, por ejemplo, como centros de movimientos tipo "*bending*", que gobiernan el movimiento concertado de dominios completos [Erman-1998, Covell-1999, Bahar-2000, Bahar-2002a, Bahar-2002b, Covell-2002a, Covell-2002b, Jernigan-2004, Bahar-2003]. Trabajos anteriores evidencian que el ANM reproduce adecuadamente las amplitudes de movimiento relativo de los residuos descritas por los modos de baja frecuencia [Covell-1999, Kneller-1999, Maritan-2004]. Estos modos involucran el desplazamiento global de un gran número de residuos y por lo tanto experimentan un potencial efectivo insensible al detalle de las interacciones individuales específicas de cada par de residuos involucrados [Thirumalai-2006]. Por el contrario, los modos de alta frecuencia definen fluctuaciones más localizadas y su correcta descripción requiere de un

tratamiento diferencial de las interacciones entre pares de residuos específicos. [Erman-1998].

4.2.2 Modelo ENCoM

En el presente trabajo se utilizó el análisis de modos normales mencionado anteriormente con diferencia en la función potencial utilizada. Esta diferencia consistió en utilizar el modelo ENCoM [Najmanovich-2014, Najmanovich-2021], el cual desarrolla una función energética potencial distinta a la mencionada anteriormente (Ec. 4.2.13) y una posterior parametrización con el fin de maximizar la habilidad predictiva del algoritmo en términos de superposición y predicción de mutaciones. Este nuevo potencial ya no sólo considera como nodos los C α y C β , sino que tiene en cuenta otras interacciones, a saber:

$$\begin{aligned}
 V_{ENCoM} &= \sum_{\text{enlaces}} V_1(r, r_0) + \sum_{\text{ángulos}} V_2(\theta, \theta_0) + \sum_{\substack{\text{ángulos} \\ \text{diedros}}} V_3(\phi, \phi_0) + \sum_{i-j < 3} V_4(r_{ij}, r_{ij,0}) \\
 &= \sum_{\text{enlaces}} \alpha_1 (r - r_0)^2 + \sum_{\text{ángulos}} \alpha_2 (\theta - \theta_0)^2 + \sum_{\substack{\text{ángulos} \\ \text{dihedros}}} \alpha_3 (\phi - \phi_0)^2 \\
 &\quad + \sum_{i < j-3} (\beta_{ij} + \alpha_4) \left[5 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 6 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{10} \right]
 \end{aligned}
 \tag{Ec. 4.2.13}$$

donde r corresponde a las posiciones de los residuos en un confórmero y r_0 las posiciones de los residuos en la conformación en equilibrio; θ y θ_0 corresponden a los ángulos de los residuos en un confórmero y en la conformación en equilibrio respectivamente; ϕ y ϕ_0 corresponden a los ángulos diedros de los residuos en un confórmero y en la conformación en equilibrio respectivamente; r_{ij} y $r_{ij,0}$ corresponden a la distancia entre los residuos i y j en un confórmero y en la conformación en equilibrio, respectivamente, cuando la distancia entre ellos es menor a 3 Å. Los términos α son utilizados para la parametrización con $\alpha_i = [10^{-4}, 10^8]$ para $i = [1, 4]$.

La principal diferencia entre este potencial y el de ANM yace en los términos β_{ij} que modulan las interacciones no enlazantes entre pares de aminoácidos según el área superficial en contacto. Específicamente,

$$\beta_{ij} = \sum_k^{N_i} \sum_l^{N_j} \varepsilon_{T(k)T(l)} S_{kl}
 \tag{Ec. 4.2.14}$$

donde $\varepsilon_{T(k)T(l)}$ representa una energía de interacción por pares entre los tipos de átomos $T(k)$ y $T(l)$ de los átomos k y l respectivamente de los aminoácidos i y j que contienen N_i y N_j átomos cada uno.

4.3 VECTOR DIFERENCIA ENTRE CONÓRMEROS – PROYECCIÓN SOBRE LA BASE DE LOS MODOS NORMALES

El vector diferencia (V) entre un par de estructuras (llamadas estructura 1 y 2) de una misma proteína describe el cambio conformacional involucrado en la misma. Para el cálculo de este vector se define cada elemento de este como

$$V_i = \frac{(y_i - x_i)}{\sum_j^{3N} y_j - x_j} \quad (\text{Ec. 4.3.1})$$

donde x_i e y_i son las coordenadas de los Ca de la estructura 1 y 2 respectivamente; y N es el número de residuos.

A continuación, el vector normalizado V se proyecta sobre la base de los modos normales de la estructura 1, según:

$$V_Q = \sum_{k=1}^{3N-6} (V \cdot Q_k) Q_k = \sum_{k=1}^{3N-6} \left(\sum_{j=1}^{3N} (v_j Q_{jk}) \right) Q_k = \sum_{k=1}^{3N-6} c_k Q_k \quad (\text{Ec. 4.3.2})$$

con

$$c_k = \sum_{j=1}^{3N} (v_j Q_{jk}) \quad (\text{Ec. 4.3.3})$$

4.4 NÚMERO DE PARTICIPACIÓN Y SUBESPACIO DE MODOS NORMALES

El hecho de que los modos normales proporcionen una descripción armónica desacoplada de las vibraciones de las proteínas es fundamental para identificar los movimientos vibracionales de equilibrio individuales que participan en los cambios conformacionales. No obstante, la identidad de los modos normales debe rastrearse después de pequeñas perturbaciones y esta no es una tarea sencilla, ya que pueden introducir reordenamientos en su orden de frecuencia. Además, la complejidad de la función energética potencial de una proteína puede hacer que varíen sustancialmente y, eventualmente, que se mezclen fuertemente. Con el fin de minimizar estos efectos, en el presente trabajo no nos ocupamos de los modos normales individuales, sino de los subespacios de modos normales asociados a los cambios conformacionales.

El grado de deslocalización de V_Q (Ec. 4.3.2) entre los diferentes modos normales de un córmoro se puede obtener evaluando el número de participación (M) de modos como:

$$M = \left(\sum_{k=1}^{3N-6} (c_k)^4 \right)^{-1} \quad (\text{Ec. 4.4.1})$$

El número de participación se ha introducido originalmente como metodología para describir la medida de deslocalización espacial de un vector de modos normales [Bell-1970, Elliott-1999, Zúñiga-2011]. En ese caso, M tiene un valor de $3N$ ante una traslación pura, y un valor igual a uno para un modo altamente localizado. Más allá de estos dos extremos, M puede utilizarse para definir la deslocalización en situaciones intermedias. Es decir, este número puede representar una medida de la deslocalización de un vector de modos normales sobre la base de las coordenadas cartesianas atómicas. El valor de M , redondeado al entero más alto y cercano, contiene información sobre el número de modos necesarios para describir la dirección del cambio conformacional. Valores de $M \approx 3N - 6$ reflejan que el cambio conformacional se extiende entre todas las vibraciones del cónfórmero, es decir, el espacio completo de los modos normales es necesario para lograr una buena representación del cambio conformacional. Valores de $M \approx 1$ indica que un único modo normal domina la dirección del cambio conformacional. De esta forma, los primeros M modos ordenados por el índice f_k de valores decrecientes de $(c_k)^2$ (Ec. 4.3.3) definen el subespacio mínimo (S), de modos $\{Q_{f_k}\}_{k=1, M}$, necesarios para lograr una buena descripción del cambio conformacional. De este modo, S contiene los modos normales implicados en el cambio conformacional entre las 2 estructuras analizadas.

4.5 SIMULACIÓN DE MUTACIONES PUNTUALES

La utilización del modelo de potencial ENCoM sensible a mutaciones, permite analizar el impacto de mutar cada aminoácido por los 19 restantes del conjunto de aminoácidos esenciales, generando así un nuevo cónfórmero por cada variante. Las mutaciones se generaron mediante el programa *Modeller* [Sali-2016]. Para cada mutación se utilizó una minimización de dos pasos. El primero optimiza solo el residuo mutado con el resto de la proteína fijada. El segundo paso optimiza los átomos vecinos no mutados. Luego de cada mutación calculamos nuevamente los modos normales según el modelo ENCoM.

4.6 COMPARACIÓN DE LOS SUBESPACIOS DE MODOS NORMALES ENTRE PROTEÍNA *WILD-TYPE* Y PROTEÍNA MUTADA

El procedimiento se basa en la construcción de la matriz Grammiana de los vectores proyección ($p_j^{s^i s}$) de cada elemento correspondiente al subespacio de modos normales S ($3N \times M$) original sobre el subespacio de modos normales después de la mutación, S^i ($3N \times M$). Siendo $\{Q_{f_k}\}_{k=1, M}$ el *set* de modos normales relevantes funcionalmente antes de la mutación y $\{Q_{f_k}^i\}_{k=1, M}$ el *set* de modos normales calculados después de la mutación del residuo i . Donde M es el número de participación para cada proteína, es decir el número de modos normales que contribuye al cambio conformacional. El vector $p_j^{s^i s}$ formado a partir de la proyección de cada vector de modos normales Q_j^i sobre la base del subespacio generado por $\{Q_{f_k}\}_{k=1, M}$ es calculado mediante la siguiente fórmula:

$$p_j^{s^i s} = \sum_{k=1}^M (Q_j^i \cdot Q_k) Q_k$$

(Ec. 4.6.1)

La matriz Grammiana, $\mathbf{G}$ ($M \times M$), formada por el set de vectores $p_j^{s^i s}$ $j=1, M$ se calcula como la matriz del producto interno, cuyas entradas están dadas por

$$G_{kl} = (p_k^{s^i s} \cdot p_l^{s^i s}) \quad (\text{Ec. 4.6.2})$$

La diagonalización de la matriz $\mathbf{G}$,

$$\mathbf{L}_G^T \mathbf{G} \mathbf{L}_G = \Lambda_G \quad (\text{Ec. 4.6.3})$$

nos permite utilizar los autovalores de $\mathbf{G}$, $\{\lambda_k\}_{k=1, M}$, como una medida de la similitud entre los dos subespacios. Dado que todos los autovalores de $\mathbf{G}$ (Grammiano) varían entre 0 y 1 [Krzanowski-1979], se puede definir una similitud de los dos subespacios como

$$\xi^{s^i s} = \frac{\sum_k^M \lambda_k}{M} \quad (\text{Ec. 4.6.4})$$

Por lo tanto, $\xi^{s^i s}$ representa la similitud entre los subespacios de modos activos antes y después de la mutación de un residuo. Un valor de $\xi^{s^i s}$ igual o cercano a 1 determina que la perturbación en ese residuo no modifica el subespacio de modos normales que contribuye al cambio conformacional y, por ende, la mutación no afectaría a la diversidad conformacional. Cuanto menor sea el valor de $\xi^{s^i s}$, mayor será el efecto de la mutación del residuo i -ésimo sobre el subespacio de modos asociados con el cambio conformacional. Es decir, la diversidad conformacional requerida por la proteína se verá afectada.

El valor de $\xi^{s^i s}$ aumenta con la dimensionalidad del subespacio $\mathbf{S}$. Esto dificulta la comparación estadística de $\xi^{s^i s}$ obtenido a partir de subespacios de distintas dimensionalidades. Con el fin de resolver este problema, para cada proteína del set de datos se normalizaron los valores de $\xi^{s^i s}$ como:

$$\mathbf{Z}_{\text{score}}^{s^i} = \frac{\xi^{s^i s} - \overline{\xi^{s^i s}}}{\sigma^S} \quad (\text{Ec. 4.6.5})$$

donde $\overline{\xi^{s^i s}}$ y σ^S son el promedio y la desviación estándar de $\xi^{s^i s}$ evaluados sobre todos los residuos de la proteína.

4.7 CARACTERIZACIÓN DE RESIDUOS DINÁMICAMENTE IMPORTANTES EN EL CAMBIO CONFORMACIONAL

Finalmente, definimos los residuos dinámicamente relevantes o residuos claves como aquellos dentro del 10% de los valores más bajos de $\mathbf{Z}_{\text{score}}^{s^i}$ para cada par de estructuras en el conjunto de proteínas estudiadas.

Los residuos dinámicamente relevantes obtenidos para los distintos pares de estructuras de cada proteína fueron caracterizados, utilizando la base de datos UniProt, desde el punto

de vista de sus propiedades fisicoquímicas, la presencia de mutaciones anotadas en la base de datos y sus efectos en la funcionalidad de la proteína, y las interacciones, tanto en la propia estructura (participación en cavidades, sitios catalíticos, estructuras secundarias relevantes, etc.) como con otras subunidades y en la interacción con ligandos, con el fin de entender su relevancia en la función biológica de la proteína. Para tal fin, se incluyó el análisis de conservación evolutiva de dichos residuos utilizando la base de datos ConSurf-DB [Ben-Tal-2020], [Ben-Tal-2009]. Esta base de datos utiliza un modelo no probabilístico basado en la realización de alineamientos múltiples de secuencia, la utilización del método de vecinos cercanos (*Neighbor-Joining*) [Nei-1987] para la construcción de árboles filogenéticos a partir de los cuales se obtienen las tasas evolutivas para cada posición en la secuencia de la proteína, y la creación de una matriz de sustitución a partir de una metodología Bayesiana [Pupko-2004] para obtener grados discretos de conservación en una escala del 1 al 9, donde 1 es altamente variable y 9 es altamente conservado, para cada residuo de la proteína.

5. RESULTADOS Y DISCUSIÓN

5.1 CARACTERIZACIÓN DEL SET DE PROTEÍNAS

Partiendo de la base de datos de diversidad conformacional CoDNaS, se procedió a una depuración utilizando distintos criterios. En primer lugar, se seleccionaron aquellas proteínas que poseyeran un mínimo de 3 estructuras cristalizadas con el fin de garantizar la diversidad conformacional. En segundo lugar, se descartaron aquellas proteínas en las cuales alguna de las estructuras presentaba discontinuidades en su secuencia de aminoácidos. Esto conduciría a la existencia de estructuras que, correspondiendo a la misma proteína, tuvieran distinto número de residuos. En tercer lugar, se seleccionaron las estructuras que no presentaran mutaciones o, en su defecto, si la mayoría de las estructuras cristalizadas poseían un mismo tipo de mutación, se descartaron aquellas que no tenían esa mutación. En cuarto lugar, se eliminaron estructuras cuyo valor de RMSD con alguna/s de las demás estructuras pertenecientes a la misma proteína sea menor a 0.8 Å. Esto es, consideramos que valores de RMSD menores a 0.8 Å implican la inexistencia de un cambio conformacional significativo en la proteína. Esto condujo a la selección de un conjunto de 7 proteínas representativas con un número variable de conformeros cada una (Figura 2.a).

En el histograma de RMSD (Å) que se muestra en la Figura 2.b es posible notar que la distribución presenta un máximo para RMSD cercanos a 1,5 Å, una media de 2,1 Å y una desviación estándar de 1,0 Å.

El valor del número de participación (*p-number*), indica el número mínimo de modos normales necesarios para una correcta descripción del vector diferencia V entre conformeros. Este permite definir el subespacio de modos activos correspondientes al cambio conformacional observado en los distintos pares de estructuras. La distribución de *p-number* obtenidos para las proteínas analizadas presenta un promedio de 36 con una desviación estándar de 30. En la Figura 2.c se muestra la distribución del porcentaje que representan los valores de *p-number* sobre la cantidad de modos normales. De esta manera se observa el porcentaje del espacio total que representa el subespacio de modos activos. Cabe destacar que en ningún caso el subespacio de modos activos supera el 20% del total de modos normales y que, en la mayoría de los casos, el número de participación involucra sólo un 5% del total de modos normales, lo cual comprueba que es necesario solamente unos pocos modos normales para describir el movimiento de un grupo colectivo de átomos en la proteína.

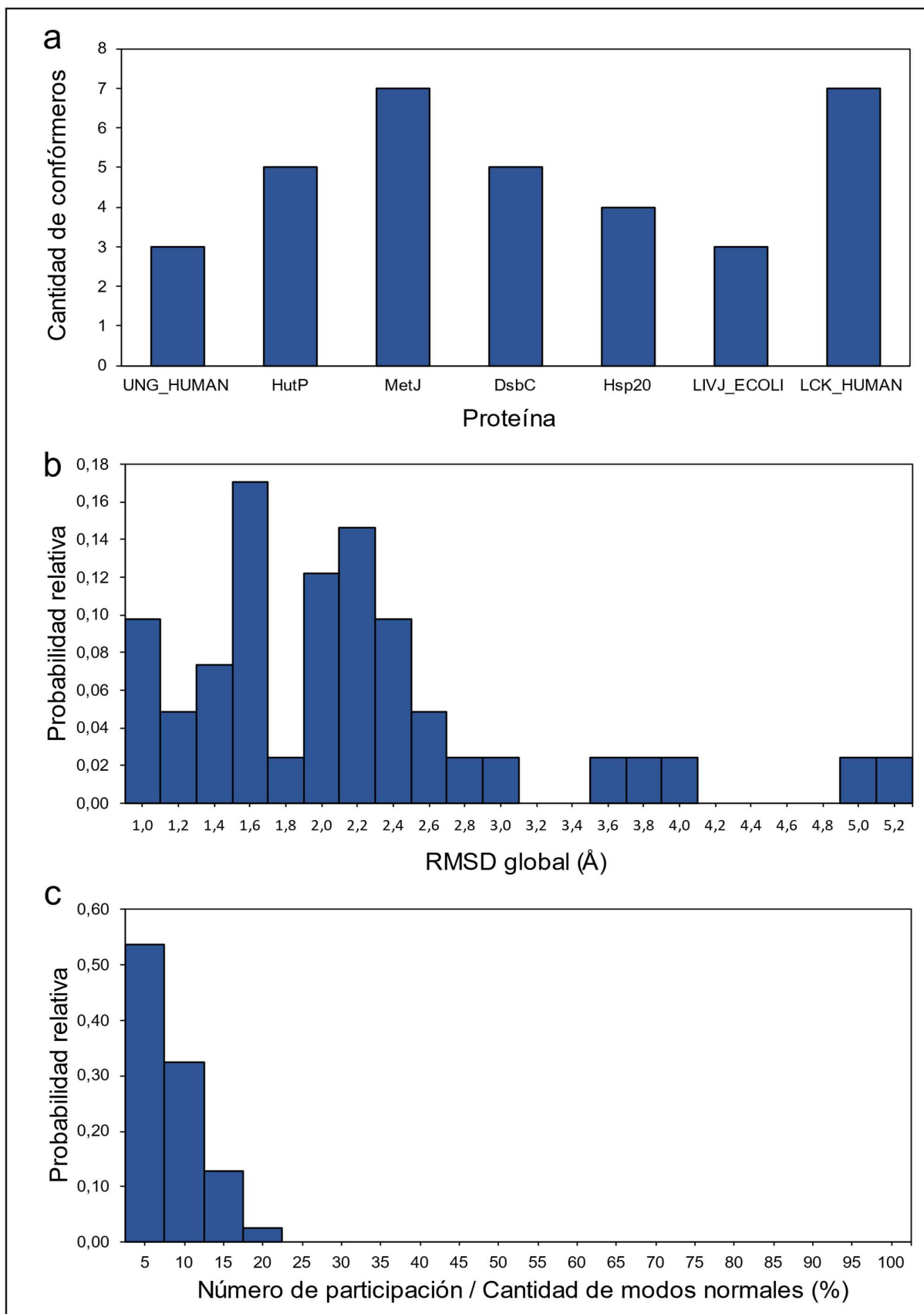


Figura 2. Caracterización del conjunto de proteínas. a) Número de conformeros por proteína. **b)** Distribución de RMSD global entre conformeros. **c)** Distribución de número de participación sobre la cantidad de modos normales.

5.2 IDENTIFICACIÓN DE RESIDUOS EN POSICIONES CLAVES

El cálculo de los modos normales obtenidos para el sistema presentando la mutación puntual de un residuo es comparado con el cálculo de modos normales de la proteína original. Ambos *sets* de modos definen distintos subespacios de modos activos, cuya comparación permitió cuantificar la respuesta a la mutación del residuo considerado. Los valores relativos de las respuestas a mutaciones en los distintos residuos que componen una proteína permitieron definir el conjunto de residuos activos.

Con el fin de ejemplificar los resultados que se obtuvieron a partir del procedimiento desarrollado en este seminario de investigación, se presentan los resultados obtenidos para 5 de las 7 proteínas que forman el conjunto de proteínas de estudio. En la Figura 3 es posible observar el porcentaje de coincidencia de los residuos claves obtenidos para cada proteína según la cantidad de cambios conformacionales en los que se encuentran como residuos relevantes para mantener la diversidad conformacional. Tomando en cuenta el porcentaje de coincidencia de los residuos claves obtenidos para todos los cambios conformacionales posibles de cada proteína, se obtuvo una media del 7 ± 6 %; mientras que tomando en cuenta el porcentaje de coincidencia de los residuos claves obtenidos para todos los cambios conformacionales menos uno, se obtuvo una media del 11 ± 14 %. A fin de abarcar un número óptimo de residuos activos presentes en el mayor número posible de cambios conformacionales, se seleccionaron hasta el 12% de los residuos activos totales que coinciden en el mayor número de cambios conformacionales. Esto es, residuos activos presentes en una cantidad de cambios conformacionales menor, no fueron considerados por tratarse de residuos específicos relacionados con un bajo número de cambios conformacionales. Para cada una de las proteínas analizadas, se detalla a continuación la respuesta a las mutaciones y el *set* de residuos activos relacionados con su implicancia estructural y funcional.

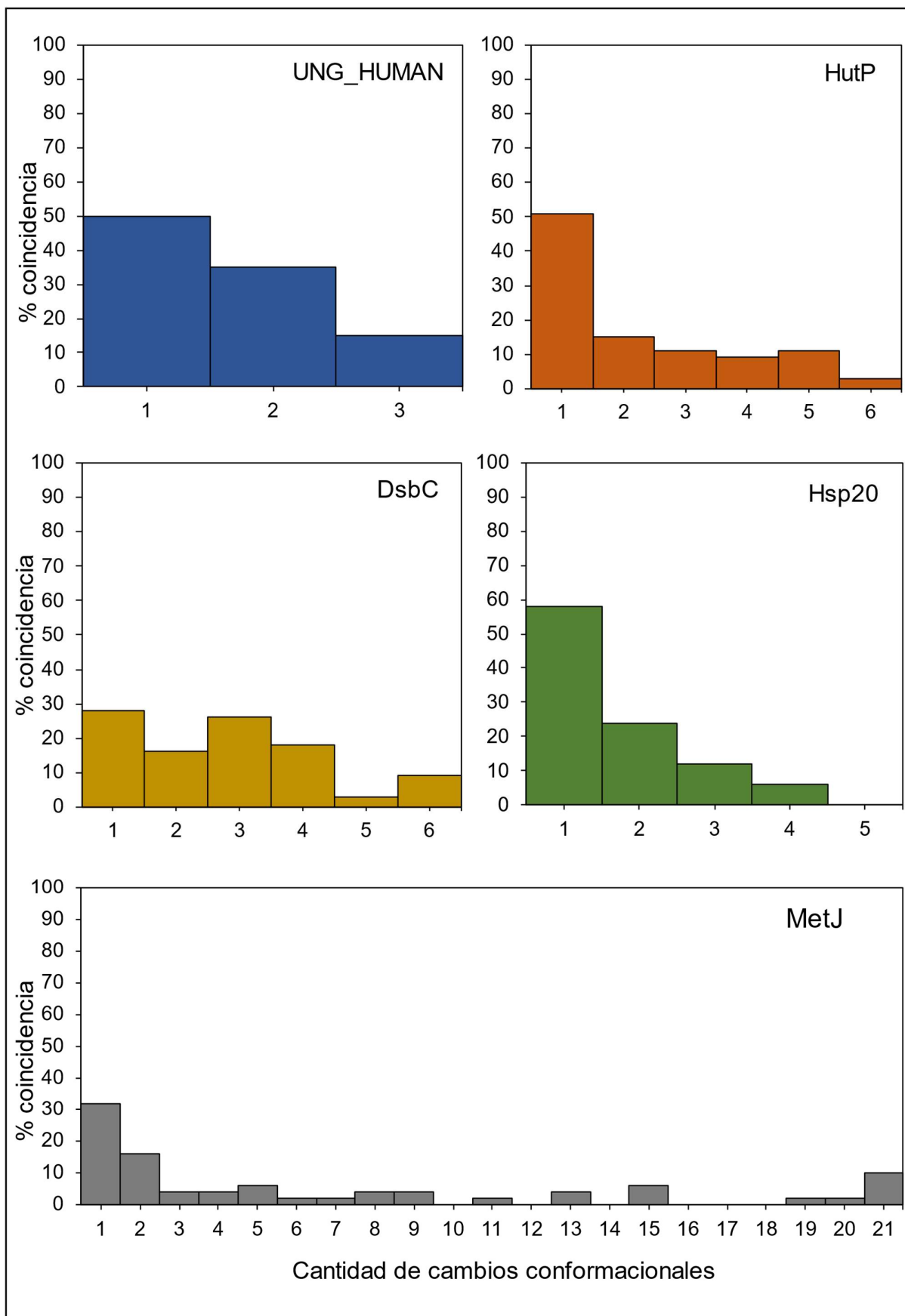


Figura 3. Porcentaje de coincidencia de residuos claves según la cantidad de cambios conformacionales de cada proteína.

5.3 UNG_HUMAN – Uracil-ADN glicosilasa

La proteína UNG_HUMAN (PDB ID: 1AKZ), también conocida como uracil-ADN glicosilasa, tiene como función principal eliminar los residuos de uracilo (U) del ADN que pueden surgir como resultado de la incorporación incorrecta de nucleótidos de dUMP por parte de la ADN polimerasa o debido a la desaminación de 5-metilcitocinas (C) en *Homo sapiens*. La enzima AID (del inglés *activation-induced cytidine deaminase*) es expresada por las células B de *H. sapiens* y cambia citocinas por uracilos en el ADN, lo cual puede llevar a un *mismatch* que será reparado por la maquinaria de corrección de errores de la célula o a una escisión de bases mediada por UNG_HUMAN. Esto permite que el mecanismo de reparación del ADN de la célula introduzca mutaciones al reparar ese error. Algunas mutaciones pueden darse en las secuencias de las cadenas pesadas y livianas de los paratopes, llevando a la generación de un BCR (receptor de células B) de mayor afinidad (hipermutación somática). Por otra parte, debido a que en las regiones génicas cercanas a los dominios constantes del BCR se encuentran secuencias con alto contenido C-G, cuando alguna de estas regiones se encuentre en transcripción activa ocurrirán varios cambios de C por U y esas bases serán removidas. Esto genera múltiples *nicks* en ambas cadenas de dichas regiones génicas que conducen a la acción del mecanismo de reparación de la doble hebra (DSBR, del inglés *Double Strand Break Repair*), el cual recombina las regiones cercanas a los sitios de lesión. De esta forma, se produce el cambio de una clase de receptor a otra (cambio de clase) [Qiu-2020].

Debido a la relevancia de la función de la proteína UNG_HUMAN en la diversidad secundaria y el cambio de clase del receptor de las células B de los seres humanos, se sabe que mutaciones en el gen de la proteína pueden generar un síndrome de inmunodeficiencia poco común caracterizado por niveles séricos de IgM normales o elevados con ausencia de IgG, IgA e IgE, lo cual resulta en una profunda susceptibilidad a infecciones bacterianas [Durandy-2003].

A partir de la simulación de los efectos de mutaciones en los subespacios de modos normales, se calcularon los coeficientes de correlación de Spearman entre los valores normalizados de ξ^{is} (Ec. 4.6.5), promediados sobre el conjunto de las 19 mutaciones posibles para cada sitio i , de los distintos cambios conformacionales (Figura 4.a). Esto indica un alto grado de coincidencia del impacto relativo de las distintas mutaciones por sitio para los distintos cambios conformacionales. Sin embargo, en la Figura 4.b es posible observar que el cambio conformacional que involucra las estructuras 2HXM_A y 1AKZ_A es más susceptible a las mutaciones. Por otra parte, los valores de RMSD globales (Figura 4.d), así como los valores de RMSD posicionales (Figura 4.c) entre cada par de estructuras involucradas evidencian la disimilitud estructural entre ellas y, por ende, la existencia de 3 cambios conformacionales identificados para UNG_HUMAN.

En la Figura 4.e es posible observar que no existe correlación lineal entre los valores de perturbación promedio y RMSD posicional para ninguno de los cambios conformacionales analizados, lo cual indica que el impacto que las mutaciones de cada residuo tienen sobre el subespacio vibracional involucrado en un cambio conformacional no está asociado al grado de desplazamiento estructural de ese residuo durante dicho cambio conformacional.

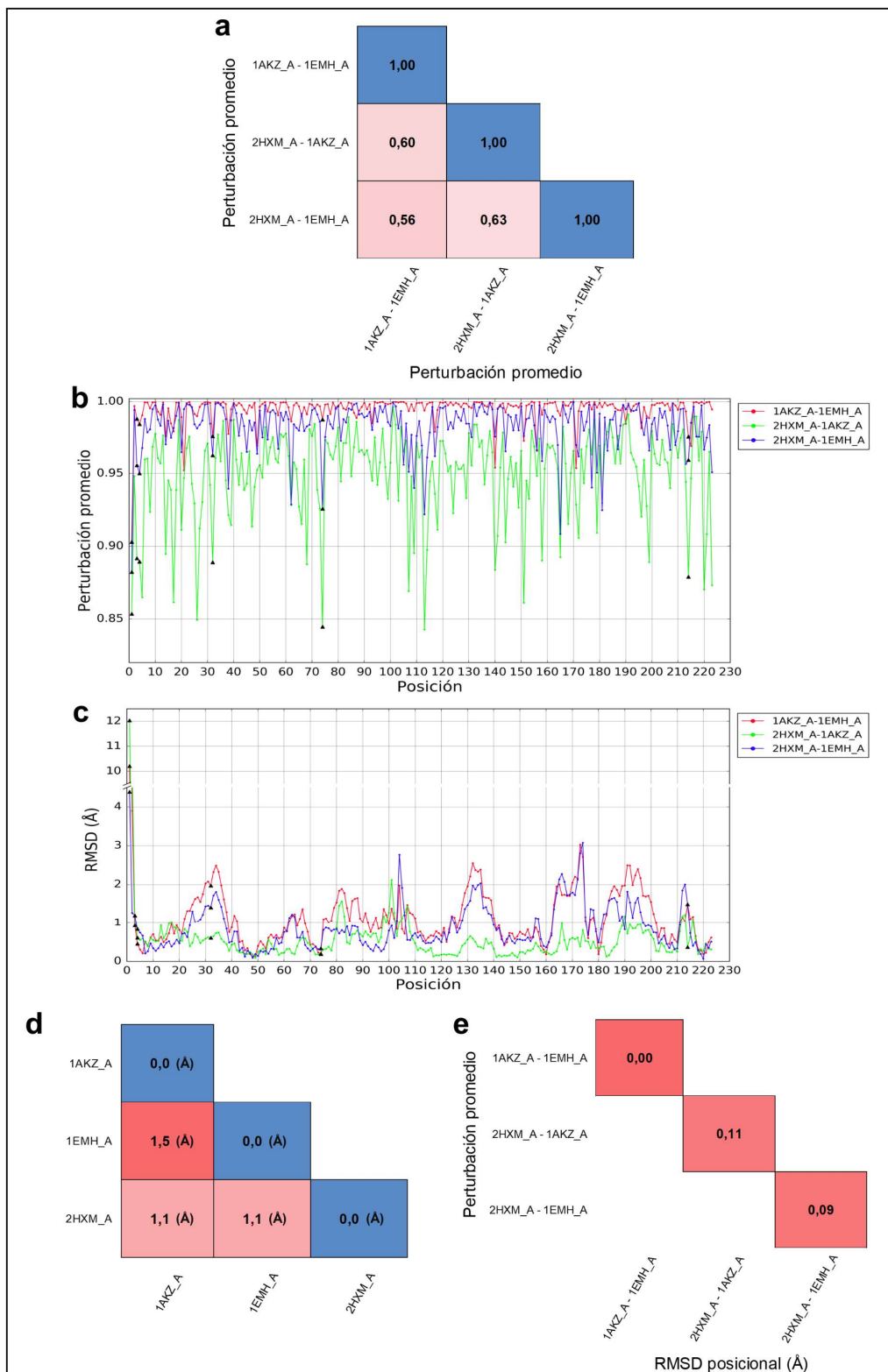


Figura 4. Caracterización de UNG_HUMAN. **a)** Coeficientes de correlación entre los valores de perturbación promedio de los modos normales para los distintos cambios conformacionales. **b)** Distribución de los valores de perturbación, promediados sobre el conjunto de las 19 mutaciones posibles, por posición para los distintos cambios conformacionales; se destacan con triángulos negros los valores de perturbación de los residuos claves. **c)** Distribución de los valores de RMSD posicionales para los distintos cambios conformacionales; se destacan con triángulos negros los valores de RMSD posicionales de los residuos claves. **d)** Valores de RMSD globales entre los distintos conformeros. **e)** Coeficientes de correlación entre los valores de perturbación promedio y RMSD posicional para cada cambio conformacional.

A partir de la metodología empleada, se lograron obtener 22 residuos claves por conformero para UNG_HUMAN, resultando 6 de ellos presentes en todos los cambios conformacionales analizados, a saber: M82, F84, F85, R113, G155, G295, según numeración en PDB.

Tal como puede observarse en la Figura 5.a, los residuos claves M82, F84 y F85 forman parte de una estructura de *loop* al final de la cadena aminoacídica, poseen propiedades hidrofóbicas y los últimos dos forman parte de una cavidad en la proteína según la base de datos CaviDB (Figura 5.b). A pesar de ser residuos claves para todos los cambios conformacionales analizados y sin presentar correlación entre el RMSD posicional y los valores de perturbación, este grupo de residuos poseen un RMSD posicional elevado respecto al resto de los residuos de la proteína (Figura 4.c).

Por otra parte, los residuos claves R113, G155 y G295 poseen anotaciones de mutaciones en la base de datos UniProt, reportadas por Durandy *et al.* como causantes del síndrome de inmunodeficiencia hiper-IgM de tipo 5 (HIGM5). Estos residuos se ubican en regiones distantes entre sí dentro de la proteína formando parte de distintas estructuras secundarias. El residuo R113 es un aminoácido cargado positivamente que forma parte de la misma cavidad mencionada anteriormente, siendo parte de una hélice- α , y posee interacciones del tipo puente de hidrógeno con otra hélice- α adyacente (Figura 5.c). Estas interacciones podrían ser la causa de la relevancia del residuo R113 en la diversidad conformacional de la proteína y, por ende, en su función biológica. El residuo G155, por su parte, posee características polares y se encuentra formando parte de un *loop* en el *core* de la proteína, cuyo RMSD posicional es significativamente inferior a los demás *loops*. A su vez, este residuo forma parte de un sitio activo ubicado en el interior de un túnel formado por la estructura secundaria de la proteína (Figura 5.e y 5.f), siendo una posición altamente conservada según la base de datos ConSurf. Estas observaciones concuerdan con la teoría neutra de la evolución molecular y la hipótesis de que, mediante esta metodología, es posible identificar residuos dinámicamente relevantes para preservar la diversidad conformacional. Finalmente, el residuo G295 también posee características polares y forma parte del extremo de una hélice- α con bajo RMSD posicional (Figura 5.d), pero no forma parte de ningún sitio activo ni se encuentra conservado evolutivamente. Sin embargo, la mutación anotada para esta posición en UniProt indica causa de HIGM5, con lo cual este residuo también es relevante para la funcionalidad de la proteína.

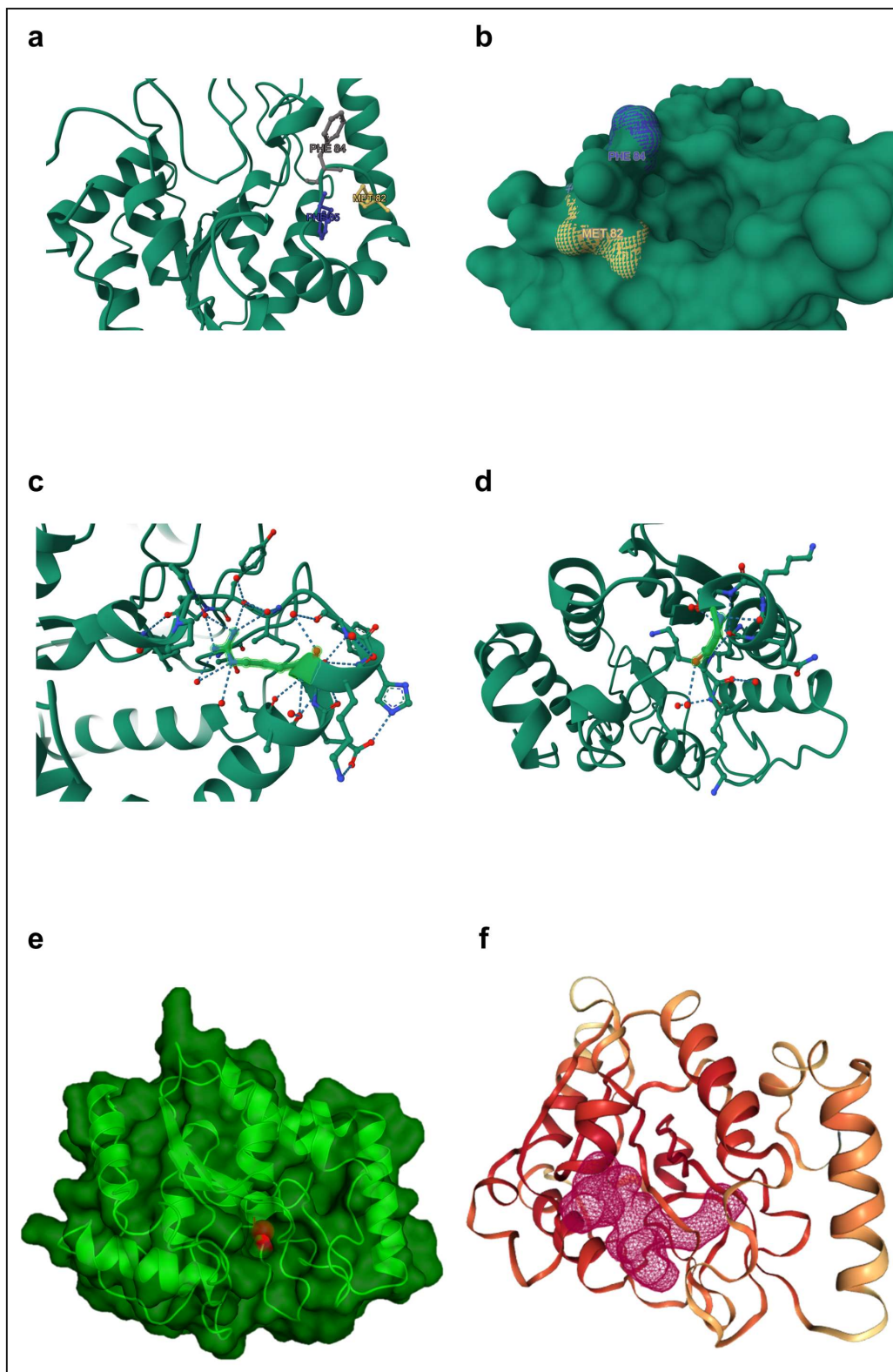


Figura 5. Residuos claves e interacciones estructurales en UNG_HUMAN. **a)** Residuos claves M82, F84 y F85. **b)** Visualización de la superficie que conforma la cavidad en la que participan los residuos claves F84 y F85. **c)** Residuo clave R113 y sus interacciones con otros residuos. **d)** Residuo clave G295 y sus interacciones con otros residuos. **e)** Visualización de la superficie que conforma la cavidad en la que se encuentra el sitio activo, en rojo el residuo clave G155; imagen realizada con el visualizador PyMOL. **f)** Cavidad en forma de túnel en la que se encuentra el residuo clave G155; imagen tomada de la base de datos CaviDB. Las imágenes a), b) y c) fueron realizadas con el visualizador GLmol [Rose-2021].

5.4 HutP – Proteína anti-terminación de unión a ARN

Las enzimas Hut de *Bacillus subtilis* permiten el transporte y degradación de histidina para ser utilizada como fuente de carbono, energía y nitrógeno cuando este aminoácido se encuentra disponible en el medio en altas concentraciones [Bender-2012]. Dichas enzimas se encuentran codificadas en el genoma de *B. subtilis* en forma del operón *hutPHUIG* (Figura 6.a). Este operón es regulado por un terminador de la transcripción situado entre los marcos abiertos de lectura de HutP y HutH (Figura 6.b), impidiendo, así, la transcripción de los genes que codifican para las enzimas de transporte y degradación de histidina. Ante la acumulación de histidina intracelular, un hexámero de HutP (PDB ID: 1WPS) se une a este aminoácido obteniendo una conformación que le permite la unión al ion Mg^{+2} . Luego de la unión de estos dos ligandos, la proteína HutP obtiene una conformación que le permite la unión a dos regiones ricas en G-C de *hutC*. La unión de HutP a la horquilla terminadora de la transcripción *hutC* permite la desestabilización de esta y, consecuentemente, la continuación de la transcripción del resto del operón *hutPHUIG*.

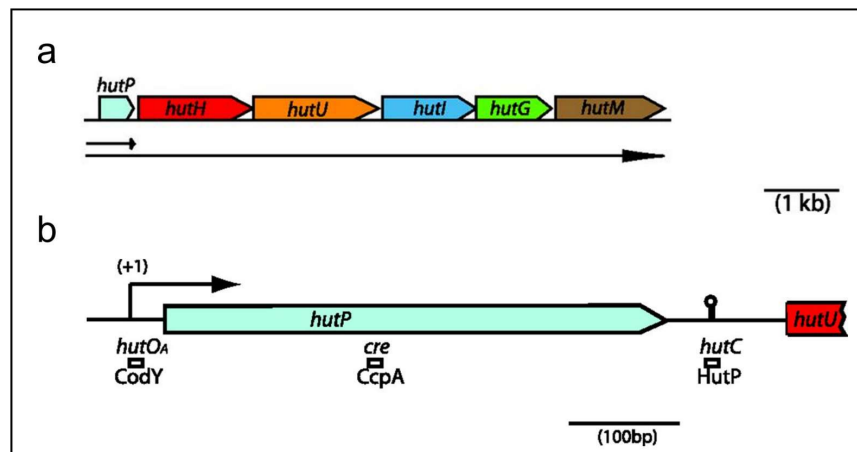


Figura 6. a) Operón *hutPHUIG*: *hutP*, proteína reguladora anti-terminación de unión a ARN; *hutH*, histidasa; *hutU*, urocanasa; *hutI*, propionato de imidazolona hidrolasa; *hutG*, formiminoglutamato hidrolasa; *hutM*, transportador de histidina. **b) Mecanismo de regulación de la transcripción del operón *hutPHUIG*:** luego del sitio de inicio de la transcripción (región operadora *hutO_A*) se encuentra la proteína reguladora anti-terminación de unión a ARN (*hutP*); *downstream* a ésta se encuentra la región de unión de *hutP* (*hutC*) y, a continuación, el resto de los marcos abiertos de lectura del operon *hutPHUIG*. Figura obtenida de Bender R. A. (2012).

El valor de perturbación, promediado sobre las 19 mutaciones posibles por sitio para todos los cambios conformacionales posibles, evidencia correlación entre varios de estos cambios conformacionales, pero no en la totalidad de los casos. (Figura 7.a). Esto indica que las mutaciones afectan de forma heterogénea al conjunto de cambios conformacionales, lo cual también es posible observar en la Figura 7.b, demostrando la utilidad de abarcar la mayor parte posible del espacio conformacional de la proteína para el análisis del efecto de mutaciones sobre su dinamismo. Por otra parte, los valores de RMSD globales (Figura 7.d), así como los valores de RMSD posicionales (Figura 7.c) entre cada par de estructuras involucradas evidencian una fuerte disimilitud estructural entre ellas y, por ende, la existencia de 6 cambios conformacionales identificados para HutP.

Al igual que en UNG_HUMAN, la Figura 7.e muestra que no existe correlación lineal entre los valores de perturbación promedio y RMSD posicional para ninguno de los cambios conformacionales analizados.

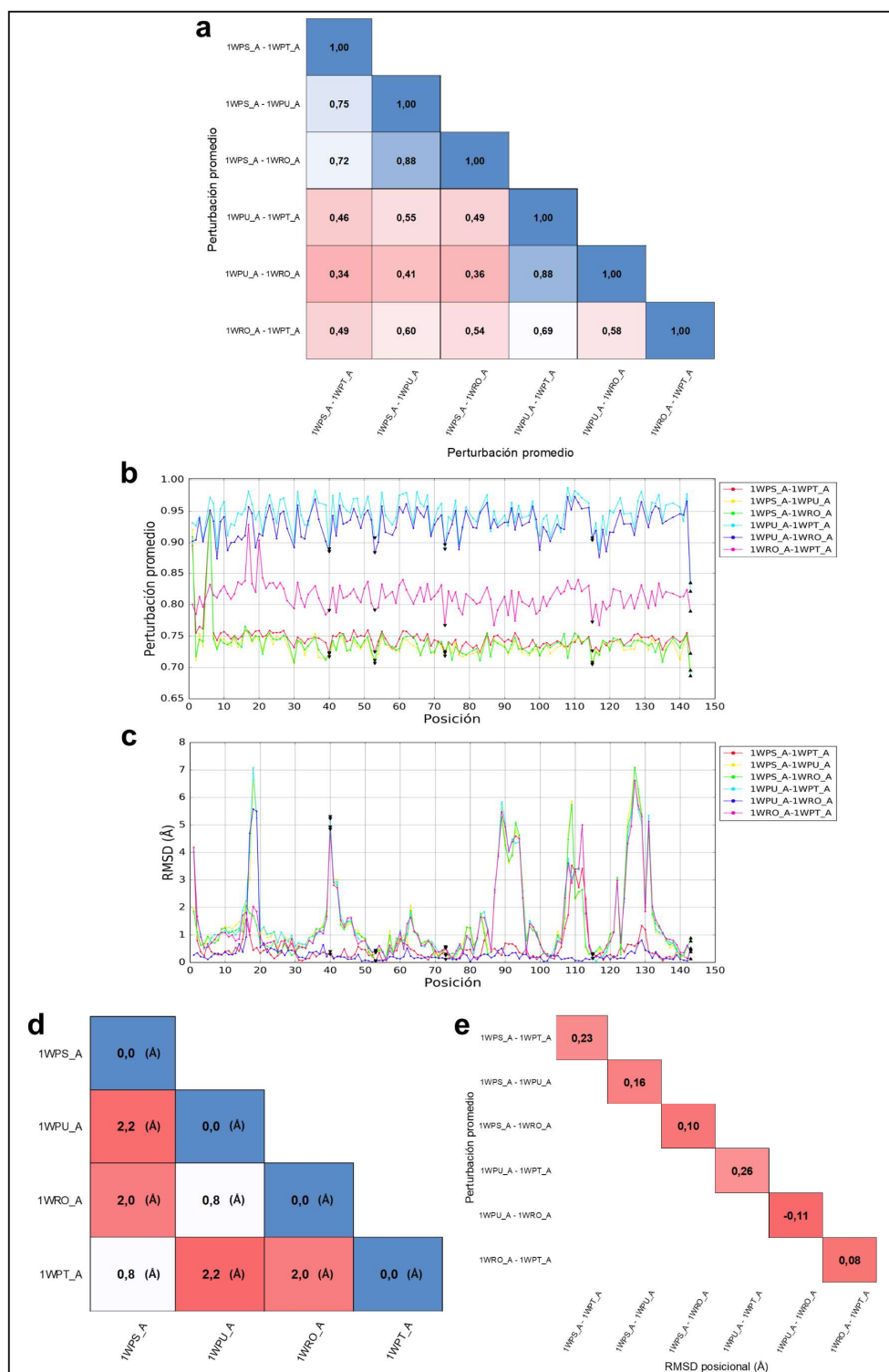


Figura 7. Caracterización de HutP. **a)** Coeficientes de correlación entre valores de perturbación promedio de los modos normales para los distintos cambios conformacionales. **b)** Distribución de los valores de perturbación promedio por posición para los distintos cambios conformacionales; se destacan con triángulos negros hacia arriba los valores de perturbación de los residuos claves que se comparten para los 6 cambios

conformacionales y con triángulos negros hacia abajo los que se comparten para 5 de los 6 cambios conformacionales. **c)** Distribución de los valores de RMSD posicionales para los distintos cambios conformacionales; se destacan con la misma configuración de triángulos negros que en b) los valores de RMSD posicionales de los residuos claves. **d)** Valores de RMSD globales entre los distintos conformeros. **e)** Coeficientes de correlación entre los valores de perturbación promedio y RMSD posicional para cada cambio conformacional.

A partir de la metodología empleada, se lograron obtener 13 residuos claves por conformero para HutP, resultando 1 de ellos (residuo I148) en una posición relevante presente en todos los cambios conformacionales analizados, y 4 de ellos (residuos M45, K58, T78 y I120) en posiciones relevantes para 5 de los 6 cambios conformacionales analizados. Todos estos residuos se encuentran altamente conservados evolutivamente según la base de datos ConSurf.

Tal como puede observarse en la Figura 8.a, el residuo clave I148 corresponde a un aminoácido de carácter hidrofóbico que forma parte del final de una lámina- β en el extremo C-terminal de la proteína, formando una interacción de interfase del tipo puente de hidrógeno con una hélice- α de la misma cadena, probablemente estabilizando ambas estructuras secundarias. Además, este residuo forma parte de una cavidad, según la base de datos CaviDB (Figura 8.b). Por otra parte, la lámina- β y la hélice- α con las que dicho residuo clave interacciona forman parte de la unión del ligando histidina, que se sitúa en la interfase entre dos cadenas de la proteína interaccionando con ambas cadenas (Figura 8.c). Según la relevancia a nivel estructural y funcional de este residuo clave, resulta coherente con la hipótesis planteada en este trabajo que el mismo se encuentre altamente conservado, tal como se observó en la base de datos de conservación evolutiva ConSurf.

Por otro lado, el residuo clave M45 corresponde a un aminoácido de carácter hidrofóbico que se ubica en un *loop* entre una hélice- α y una lámina- β interaccionando con varias estructuras mediante puentes de hidrógeno (Figura 8.d). Posiblemente, notando el alto valor de RMSD posicional que posee este residuo (Figura 7.c), se trate de una región con alta libertad de movimiento para un gran conjunto de residuos, siendo la lámina- β la región de la proteína que posee las interacciones con el ARNm y permitiendo, así, parte del movimiento que produce la relajación de la horquilla en el ARN. Por otra parte, el residuo clave K58 corresponde a un aminoácido de características básicas que se ubica en una hélice- α interaccionando con otra hélice- α adyacente que contiene al residuo clave T78 (Figura 8.e). Este último, por su parte, corresponde a un aminoácido de carácter polar que se ubica en el centro de la hélice- α mencionada. Por último, el residuo clave I120 corresponde a un aminoácido de carácter hidrofóbico ubicado en la lámina- β de la proteína interaccionando con otros residuos dentro de esta propia estructura (Figura 8.f).

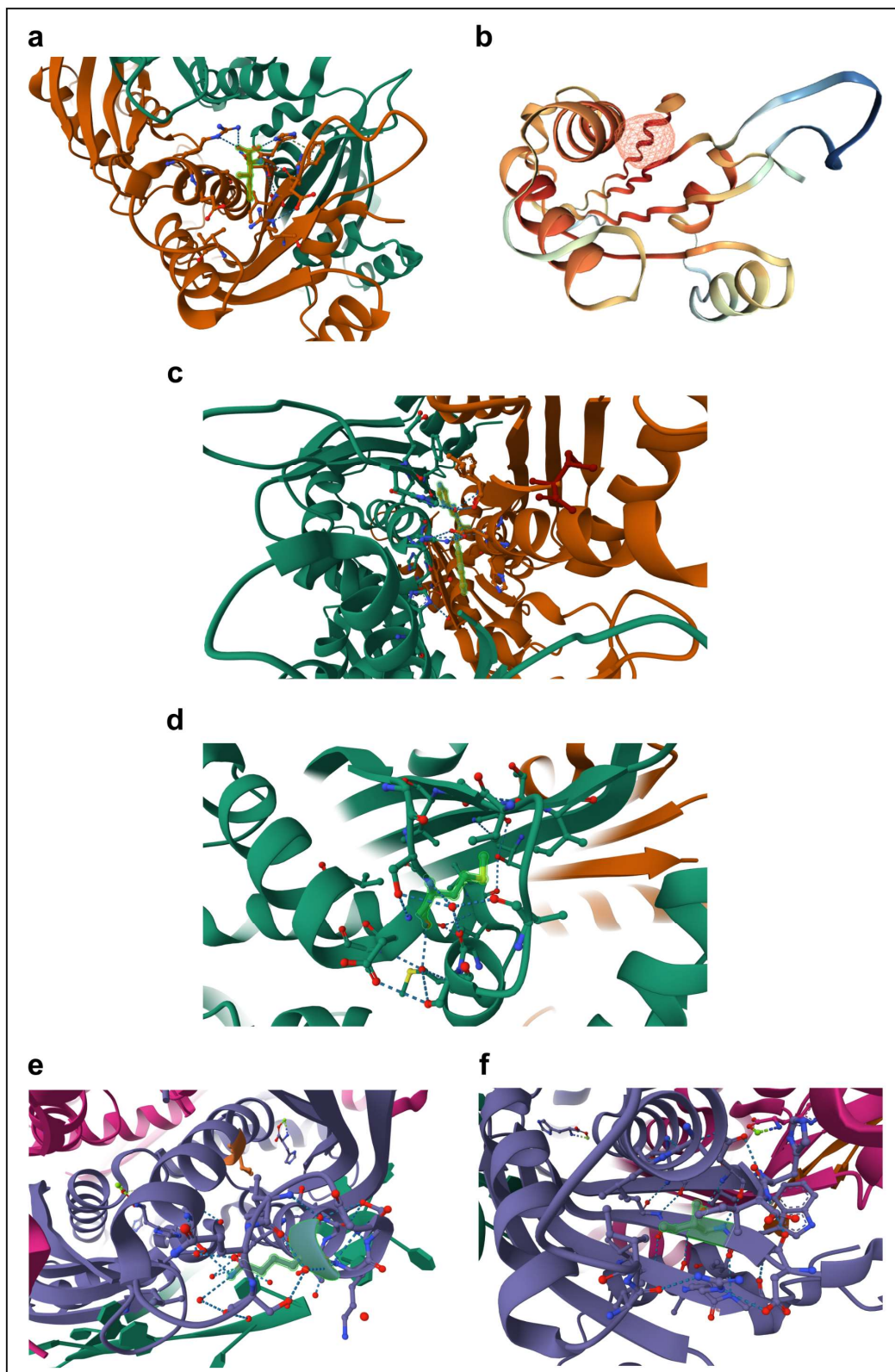


Figura 8. Residuos claves e interacciones estructurales en HutP. **a)** Residuo clave I148 y sus interacciones con otros residuos. **b)** Cavidad en la que se encuentra el residuo clave I148; imagen tomada de la base de datos CaviDB. **c)** Interacciones con el ligando histidina; en verde se resalta el ligando y en rojo se resalta el residuo clave I148. **d)** Residuo clave M45 y sus interacciones con otros residuos. **e)** En verde se resalta el residuo clave K58 y en naranja el residuo clave T78, ambos con sus interacciones con otros residuos. **f)** Residuo clave I120 y sus interacciones con otros residuos. Las imágenes **a**, **c**, **d**, **e** y **f** fueron realizadas con el visualizador GLmol.

5.5 MetJ – Represor del regulón *met*

La proteína MetJ (PDB ID: 1MJK) consiste en un homodímero que, al unirse al ligando S-adenosilmetionina (SAM), adopta una conformación que le permite unirse a una secuencia palindrómica repetitiva de 8 pares de bases (AGACGTCT), conocida como “caja *met*”, presente en distintos operadores en el genoma de *Escherichia coli*. De esta forma, el represor MetJ regula de forma global varios genes ubicados en distintas regiones del genoma relacionados a la biosíntesis del aminoácido metionina, ya que estos genes codifican enzimas involucradas en la biosíntesis de este aminoácido y en la transformación de metionina a SAM [Phillips-1992, Reitzer-2014]. A su vez, la expresión de esta proteína es autorregulada, ya que en la región promotora del gen *metJ* también se encuentra una caja *met* [Zakin-1984]. Este mecanismo de regulación ocurre cuando *E. coli* crece en un medio con alta concentración de metionina.

A partir de la simulación de mutaciones en los subespacios de modos normales, se observó una fuerte correlación entre los distintos cambios conformacionales (Figura 9.a y 9.b). Esto indica que las mutaciones afectan de forma homogénea al conjunto de conformeros analizados para esta proteína. Por otra parte, al igual que en las proteínas anteriores, los valores de RMSD globales (Figura 9.d), así como los valores de RMSD posicionales (Figura 9.c) entre cada par de estructuras involucradas evidencian la disimilitud estructural entre ellas y, por ende, la existencia de 21 cambios conformacionales identificados para MetJ. Tampoco en esta proteína se observa correlación lineal entre los valores de perturbación promedio y RMSD posicional para ninguno de los cambios conformacionales analizados (Figura 9.e)

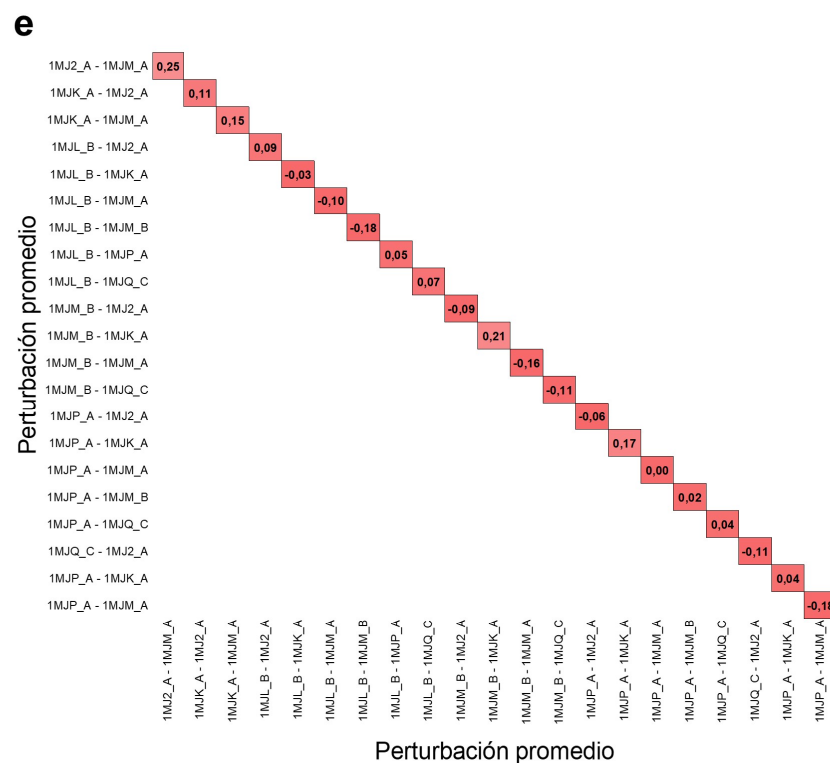
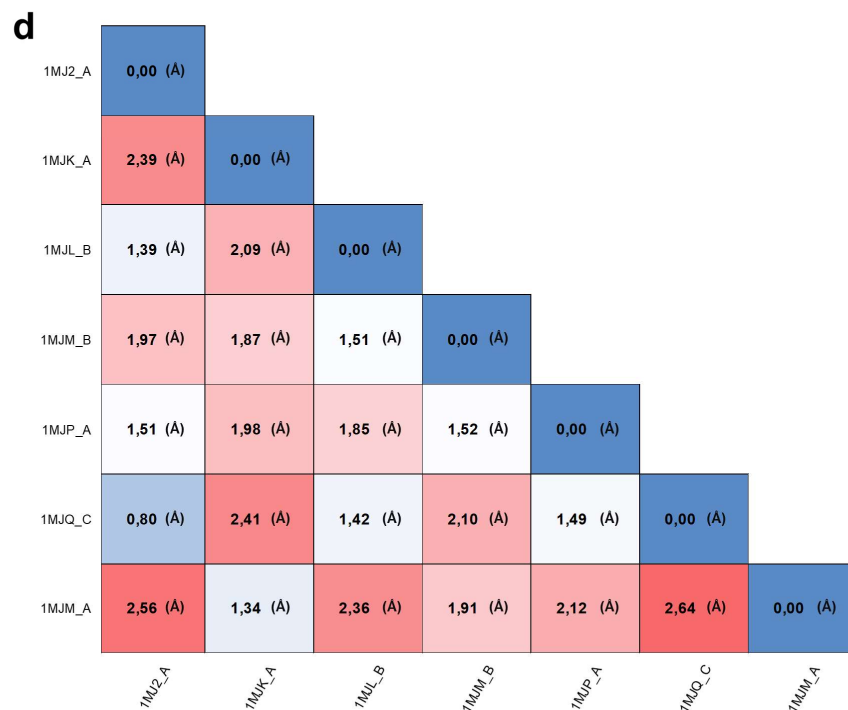


Figura 9. Caracterización de MetJ. (Continuación) **d)** Valores de RMSD globales entre los distintos conformeros. **e)** Coeficientes de correlación entre los valores de perturbación promedio y RMSD posicional para cada cambio conformacional.

A partir de la metodología empleada, se lograron obtener 16 residuos claves por conformero para MetR, resultando 5 de ellos (residuos W3, G5, G15, G67, Y104) en posiciones relevantes presente en todos los cambios conformacionales analizados, 1 de ellos (residuo G96) en una posición relevante presente en 20 de los 21 cambios conformacionales, y 1 de ellos (residuo R42) en una posición relevante para 19 de los 21 cambios conformacionales analizados.

El residuo clave W3 corresponde a un aminoácido de carácter hidrofóbico, altamente conservado, presente en el extremo N-terminal de la proteína interaccionando con una hélice- α que participa en la unión a SAM (Figura 10.a). Adyacente a este residuo se encuentra el residuo clave G5 que corresponde a un aminoácido de carácter polar que no posee interacciones relevantes desde el punto de vista funcional ni estructural, lo cual es coherente con su grado de conservación evolutiva promedio. Por otra parte, el residuo clave G15 corresponde también a un aminoácido de carácter polar que se encuentra en una estructura de *loop* con alto RMSD posicional (Figura 9.c) y posee interacciones con el esqueleto de fosfatos del ADN (Figura 10.b). Debido a la relevancia funcional de este residuo es de esperar que se encuentre altamente conservado evolutivamente, tal como lo confirma la base de datos ConSurf. De la misma manera, el residuo clave G67 corresponde a un aminoácido de carácter polar relevante desde el punto de vista funcional, ya que interacciona mediante puente de hidrógeno con el ligando SAM (Figura 10.c), siendo altamente conservado evolutivamente. Por otro lado, el residuo clave Y104 corresponde a otro aminoácido de carácter polar, presente en el extremo C-terminal de la proteína, que mantiene interacciones del tipo hidrofóbicas y puentes de hidrógeno con otros *loops* y se encuentra altamente conservado evolutivamente.

Por otra parte, el residuo clave G96, presente en 20 de los 21 cambios conformacionales, corresponde a un aminoácido de carácter polar ubicado al final de una hélice- α , posiblemente dándole estabilidad a la misma, pero sin interacciones relevantes desde el punto de vista funcional. Este residuo se encuentra moderadamente conservado evolutivamente. Mientras que el residuo clave R42, presente en 19 de los 21 cambios conformacionales, corresponde a un aminoácido de carácter básico ubicado en una hélice- α y presenta múltiples interacciones del tipo puente de hidrógeno con la propia estructura secundaria, con otra estructura de *loop* de la proteína y con el ligando SAM (Figura 10.d). Debido a la relevancia desde el punto de vista estructural y funcional, es de esperar que este residuo se encuentre conservado evolutivamente, tal como lo indica la base de datos ConSurf.

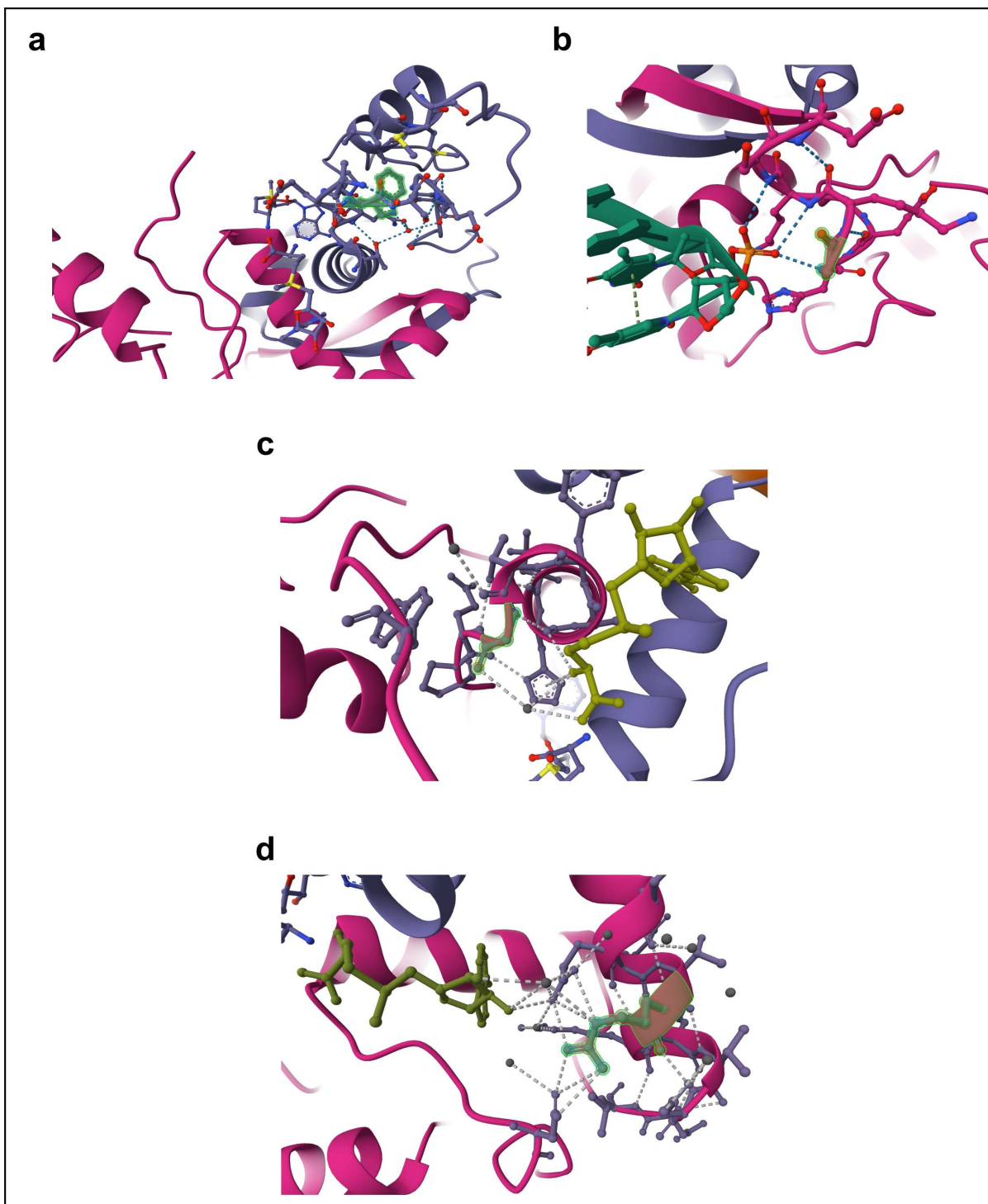


Figura 10. Residuos claves e interacciones estructurales en MetJ. a) Residuo clave W3 y sus interacciones con otros residuos. b) Residuo clave G15 y sus interacciones con el esqueleto de fosfatos del ADN (resaltado en verde). c) Residuo clave G67 y sus interacciones con otros residuos y con el ligando SAM (resaltado en amarillo). d) Residuo clave R42 y sus interacciones con otros residuos y con el ligando SAM (resaltado en amarillo). Todas las imágenes fueron realizadas con el visualizador Glmol a partir de la estructura cristalizada con ligando 1CMA.

5.6 DsbC – Proteína intercambiadora de tiol:disulfuros

DsbC (PDB ID: 1JZD) consiste en una disulfuro isomerasa que se encuentra en el periplasma de *E. coli*. Se trata de un homodímero con forma de V similar a la forma que adoptan las inmunoglobulinas. Como el ambiente periplásmico es oxidativo, proteínas con grupos tioles libres secretadas a este espacio subcelular serán expuestas a la sulfenilación de sus cisteínas y, posteriormente, a la oxidación irreversible de estos grupos funcionales a ácido sulfínico y sulfónico. DsbG y DsbC son parte de un sistema reductor periplásmico que controla el nivel de sulfenilación de cisteína y proporciona equivalentes reductores para rescatar proteínas secretadas oxidativamente dañadas. Actúa transfiriendo su enlace disulfuro a otras proteínas y se reduce en el proceso. Finalmente, DsbC es reoxidada por DsbD para volver a tener su sitio catalítico activo [Collet-2009, Metcalf-2002].

A partir de la simulación de mutaciones en los subespacios de modos normales, se observó una fuerte correlación entre los distintos cambios conformacionales (Figura 11.a y 11.b). Esto indica que las mutaciones afectan de forma homogénea al conjunto de conformeros analizados para esta proteína. Por otra parte, los valores de RMSD posicionales (Figura 11.c), así como los valores de RMSD globales (Figura 11.d) entre cada par de estructuras involucradas evidencian la disimilitud estructural entre ellas y, por ende, la existencia de 6 cambios conformacionales identificados para DsbC.

En la Figura 11.e es posible observar que no existe correlación lineal entre los valores de perturbación promedio y RMSD posicional para ninguno de los cambios conformacionales analizados, lo cual quiere decir que las diferencias en los subespacios de modos normales se deben efectivamente a la mutación simulada y no a diferencias estructurales entre los conformeros de la proteína.

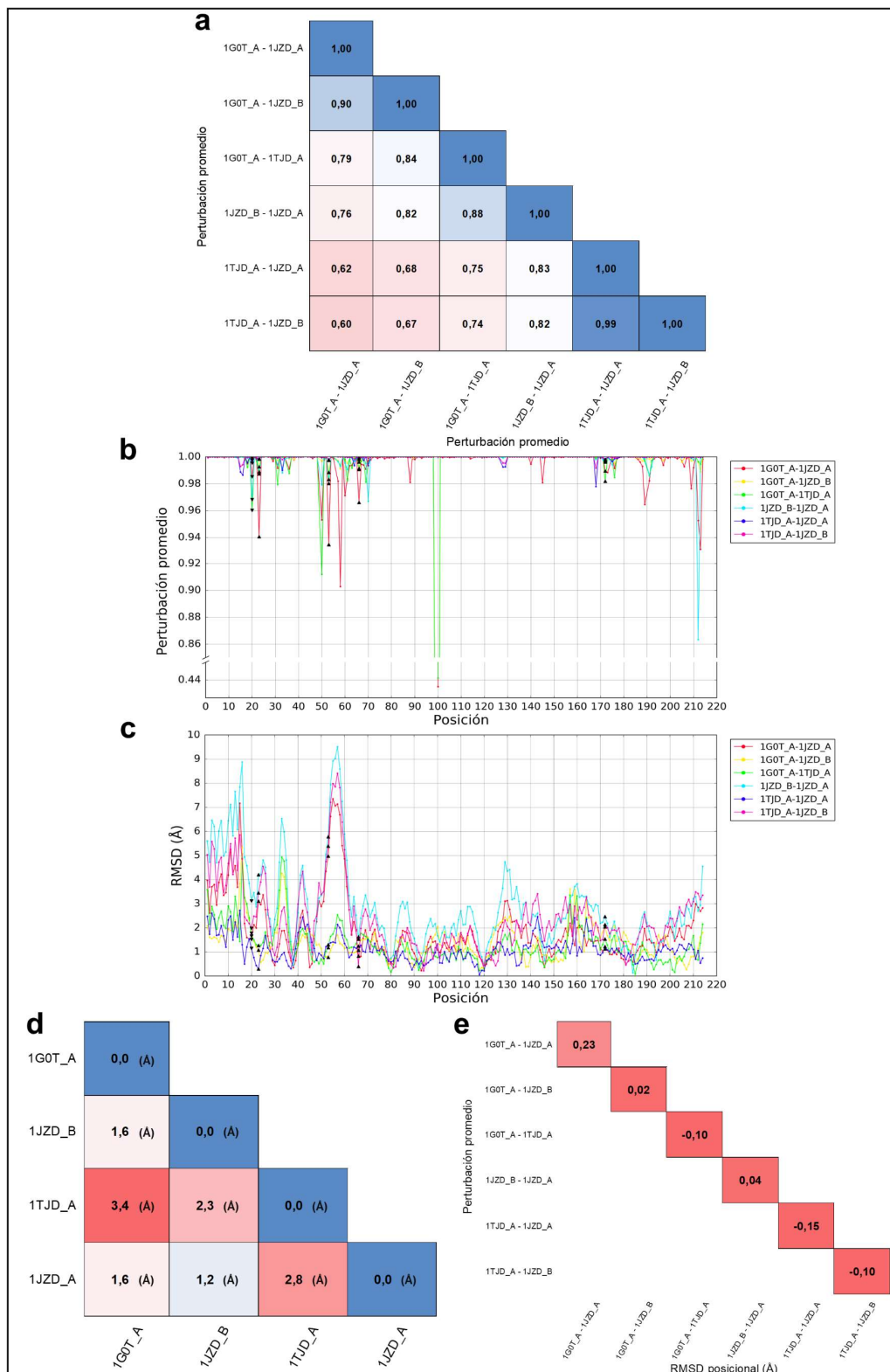


Figura 11. Caracterización de DsbC. **a)** Coeficientes de correlación entre valores de perturbación promedio de los modos normales para los distintos cambios conformacionales. **b)** Distribución de los valores de perturbación promedio por posición para los distintos cambios conformacionales; se destacan con triángulos negros los valores de perturbación de los residuos claves. **c)** Valores de RMSD globales entre los distintos conformeros. **d)** Distribución de los valores de RMSD posicionales para los distintos cambios conformacionales; se destacan con triángulos negros los valores de RMSD posicionales de los residuos claves. **e)** Coeficientes de correlación entre los valores de perturbación promedio y RMSD posicional para cada cambio conformacional.

A partir de la metodología empleada y teniendo en cuenta que las estructuras utilizadas para esta proteína contenían una mutación en la posición 121 anotada en la base de datos UniProt, se lograron obtener 20 residuos claves por confórmero para DsbC, resultando 4 de ellos (residuos P23, D53, M66, A172) en posiciones relevantes presente en todos los cambios conformacionales analizados y 1 de ellos (residuo Q20) en una posición relevante presente en 5 de los 6 cambios conformacionales.

El residuo clave P23 (Figura 12.a) corresponde a un aminoácido de carácter hidrofóbico que forma parte de un *loop*. Si bien forma parte de una cavidad (Figura 12.c), este residuo no posee interacciones con otras estructuras y, como es de esperar, se encuentra muy poco conservado evolutivamente. Por otro lado, formando parte de la misma cavidad, se ubica el residuo clave D53 (Figura 12.b) correspondiente a un aminoácido de carácter ácido que forma parte de un *loop* de alto RMSD posicional (Figura 11.c), cuya estructura posee varios puntos de interacción con la otra cadena del homodímero. Todos los residuos que conforman dicho *loop* fueron obtenidos como residuos claves para alguno o varios confórmeros, siendo el residuo D53 el que coincide para todos los confórmeros de la proteína.

Por otra parte, el residuo clave M66 (Figura 12.d) corresponde a un aminoácido de carácter hidrofóbico ubicado en una hélice- α que interacciona con una lámina- β dentro de la misma cadena. Mientras que el residuo clave Q20 (Figura 12.e) corresponde a un aminoácido de carácter polar que se encuentra en una lámina- β y mantiene interacciones con esta estructura posiblemente estabilizándola. Tanto el residuo M66 como el residuo Q20 forman parte de una misma cavidad (Figura 12.g) que tiene interacción con el ligando MES (ácido 2-etanosulfónico), según anotaciones en la base de datos PDBe-KB.

Finalmente, el residuo clave A172 (Figura 12.f) corresponde a un aminoácido de carácter hidrofóbico que se ubica en una hélice- α , manteniendo interacciones con dicha estructura y formando parte de un sitio activo con interacción directa con el ligando MES.

A pesar de la relevancia estructural y funcional de los residuos claves mencionados, ninguno se encuentra conservado evolutivamente según la base de datos ConSurf. Resulta interesante utilizar otros métodos de conservación evolutiva para el análisis de esta proteína.

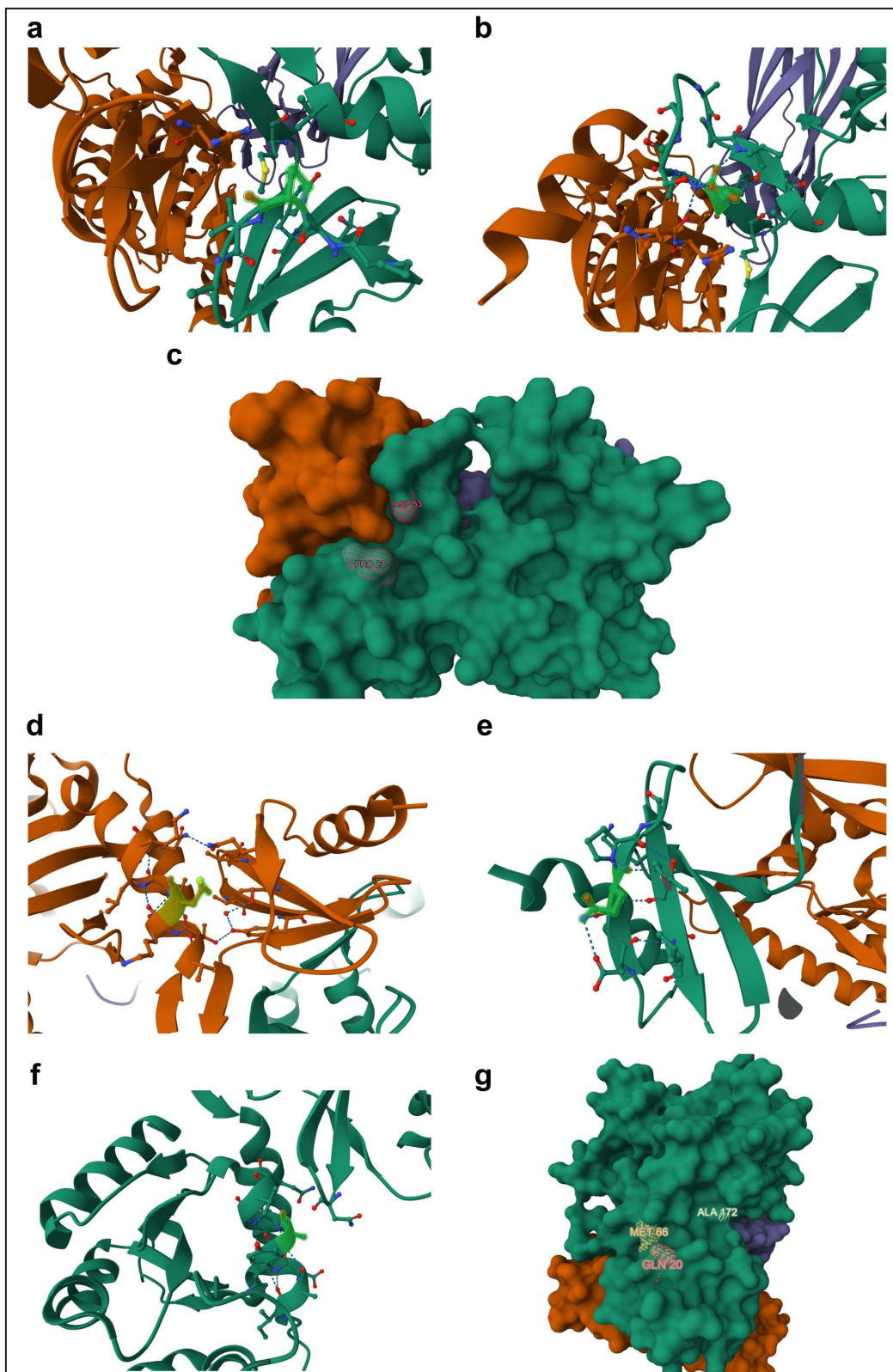


Figura 12. Residuos claves e interacciones estructurales en DsbC. a) Residuo clave P23 y sus interacciones con otros residuos. b) Residuo clave D53 y sus interacciones con otros residuos. c) Visualización de la superficie que conforma la cavidad en la que participan los residuos claves P23 y D53. d) Residuo clave M66 y sus interacciones con otros residuos. e) Residuo clave Q20 y sus interacciones con otros residuos. f) Visualización de la superficie que conforma la cavidad en la que participan los residuos claves M66 y Q20 y el sitio activo dentro de la misma, del cual forma parte el residuo clave A172. Todas las imágenes fueron realizadas con el visualizador GLmol.

5.7 Hsp20 – *Heat shock protein*

La proteína Hsp20 (PDB ID: 4YL9) pertenece a la superfamilia de las *Heat shock proteins* y se encuentra en el organismo *Sulfolobus solfataricus* que pertenece al dominio *Archaea*. Se trata de una proteína que funciona como chaperona para evitar la agregación de proteínas mal plegadas durante el estrés térmico. En el extremo N-terminal se encuentra una hélice- α mediante la cual interaccionan los tetrámeros de Hsp20. El cuerpo de la proteína está compuesto por dos láminas- β enfrentadas que conforman la región por la cual se une a las proteínas mal plegadas, siendo la región estructural más conservada de las Hsp. Esta región además interviene en la formación de dímeros, los cuales son estables en solución y se pueden unir entre sí formando oligómeros simétricos que pueden contener 4-18 dímeros. Estos oligómeros se pueden organizar en distintas conformaciones para unirse a distintas proteínas malformadas. La región C-terminal de la proteína tiene un motivo Ile-X-Ile que participa en la formación de los oligómeros mediante interacciones hidrofóbicas con un surco hidrofóbico de otro dímero [Yun-2015].

A partir de la simulación de mutaciones en los subespacios de modos normales, no se obtuvo una correlación entre los distintos cambios conformacionales (Figura 13.a). Esto indica que las mutaciones afectan de forma heterogénea al conjunto de conformeros analizados, lo cual también es posible observar en la Figura 13.b, demostrando la utilidad de abarcar la mayor parte posible del espacio conformacional de la proteína para el análisis del efecto de mutaciones sobre la misma. Por otra parte, los valores de RMSD globales (Figura 13.d), así como los valores de RMSD posicionales (Figura 13.c) entre cada par de estructuras involucradas evidencian una fuerte disimilitud estructural entre ellas y, por ende, la existencia de 5 cambios conformacionales identificados para Hsp20.

En la Figura 13.e es posible observar que no existe correlación lineal entre los valores de perturbación promedio y RMSD posicional para ninguno de los cambios conformacionales analizados, lo cual quiere decir que las diferencias en los subespacios de modos normales se deben efectivamente a la mutación simulada y no a diferencias estructurales entre los conformeros de la proteína.

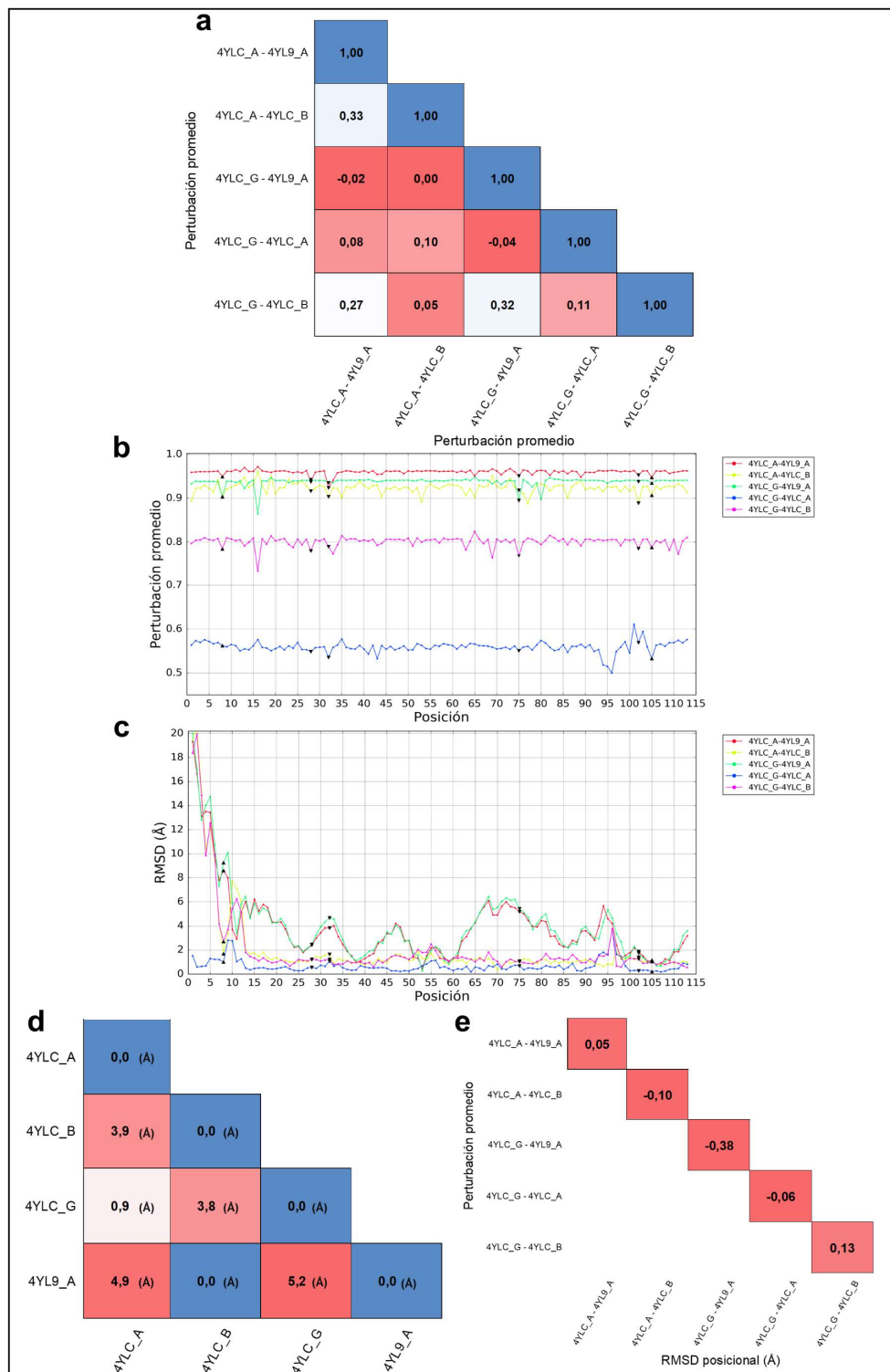


Figura 13. Caracterización de Hsp20. a) Coeficientes de correlación entre valores de perturbación promedio de los modos normales para los distintos cambios conformacionales. **b)** Distribución de los valores de perturbación promedio por posición para los distintos cambios conformacionales; se destacan con triángulos negros hacia arriba los valores de perturbación de los residuos claves que se comparten para 4 de los 5 cambios conformacionales y con triángulos negros hacia abajo los que se comparten para 3 de los 6 cambios conformacionales. **c)** Distribución de los valores de RMSD posicionales para los distintos cambios conformacionales; se destacan con la misma configuración de triángulos negros que en b) los valores de RMSD posicionales de los residuos claves. **d)** Valores de RMSD globales entre los distintos conformeros. **e)** Coeficientes de correlación entre los valores de perturbación promedio y RMSD posicional para cada cambio conformacional.

Se obtuvieron 11 residuos claves por confórmero para Hsp20, resultando 2 de ellos (residuos G10 y G107) en posiciones relevantes presentes en 4 de los 5 cambios conformacionales analizados, y 4 de ellos (residuos M30, G34, A77, Y104) en posiciones relevantes presentes en 3 de los 5 cambios conformacionales analizados.

Según las observaciones realizadas por Cai-Hong *et al.*, se consideran los residuos 16 y 20 como residuos claves para la actividad de la proteína, debido a sus propiedades hidrofóbicas, y se encuentran presentes en una estructura de hélice- α . Esta estructura puede presentar una relajación en los residuos 16-17 que disparan un cambio conformacional que resulta en un movimiento del tipo *bending* y la disrupción de la hélice- α para favorecer conformaciones más flexibles que presenten mayor actividad (Figura 14.a). Si bien la metodología empleada en el presente trabajo no reveló estas posiciones como residuos claves, sí se obtuvieron residuos cercanos en las posiciones 15 y 18 de características ácidas y básicas, y considerados como residuos claves para 1 ó 2 cambios conformacionales respectivamente (Figura 14.b). A su vez, el residuo clave G10, de carácter polar, también se encuentra dentro de la hélice- α estableciendo interacciones con la cadena de otro dímero cuando la proteína se encuentra formando un tetramero. Cabe destacar que el residuo G10 forma parte de la misma cavidad que el residuo 15, según la base de datos CaviDB. Asimismo, todos los residuos que conforman la hélice- α no se encuentran conservados desde el punto de vista evolutivo, lo cual resulta comprensible al observar la alta flexibilidad que requiere esta estructura para formar distintos oligómeros que permitan el reconocimiento de distintas proteínas mal plegadas.

Por otra parte, Cai-Hong *et al.* destacaron que, tanto la mutación del residuo A102D que lleva al reemplazo del aminoácido original con propiedades hidrofóbicas por un aminoácido de carácter ácido, así como la delección de los últimos residuos donde se encuentra el motivo Ile-X-Ile (residuos 111-113), conllevan a la disminución en la actividad de la proteína debido a la incapacidad de formar oligómeros mayores que octámeros, lo cual reduce las combinaciones posibles para el reconocimiento de proteínas mal plegadas. Mediante nuestra metodología, se obtuvo el residuo A102 (Figura 14.c) como un residuo clave presente en el cambio conformacional dado por los confórmeros 4YLC_G-4YLC_A, siendo el cambio conformacional que presentó mayor perturbación del subespacio de modos normales (Figura 13.b). Este residuo forma parte de 2 cavidades y se encuentra altamente conservado evolutivamente (Figura 14.d). Además, se obtuvieron los residuos L109 y I111 (Figura 14.e) como residuos claves del cambio conformacional dado por 4YLC_G-4YLC_B, siendo otro cambio conformacional que también implicó grandes perturbaciones del subespacio de modos normales. Estos últimos residuos no forman parte de ninguna cavidad, pero también se encuentran conservados evolutivamente.

El residuo clave G107 corresponde a un aminoácido de carácter polar que se ubica en un giro- β e interacciona con la propia lámina- β , probablemente manteniendo su estructura. Además, la estructura de la cual forma parte posee interacciones con el giro- β análogo de la otra cadena perteneciente al mismo dímero (Figura 14.f). Este residuo se encuentra altamente conservado según la base de datos ConSurf. Por otro lado, el residuo clave Y104 corresponde a un aminoácido de carácter polar que se ubica en la misma lámina- β que el residuo mencionado anteriormente. Este residuo, que se encuentra levemente conservado, mantiene interacciones con un *loop* dentro de la misma cadena, probablemente manteniendo las dos láminas- β enfrentadas (Figura 14.g).

El residuo clave M30 corresponde a un aminoácido de carácter hidrofóbico que se ubica en la misma lámina- β mencionada anteriormente interaccionando directamente, mediante

puentes de hidrógeno, con el residuo clave A77 ubicado en un *loop* de la otra cadena perteneciente al mismo dímero (Figura 14.h). Este último residuo también corresponde a un aminoácido de carácter hidrofóbico y, de la misma forma que el residuo anterior, ninguno de los dos se encuentra conservado evolutivamente. Por último, el residuo clave G34 corresponde a un aminoácido de carácter polar que también se ubica en la misma lámina- β ya mencionada y no posee interacciones con residuos cercanos. Este residuo tampoco se encuentra conservado evolutivamente.

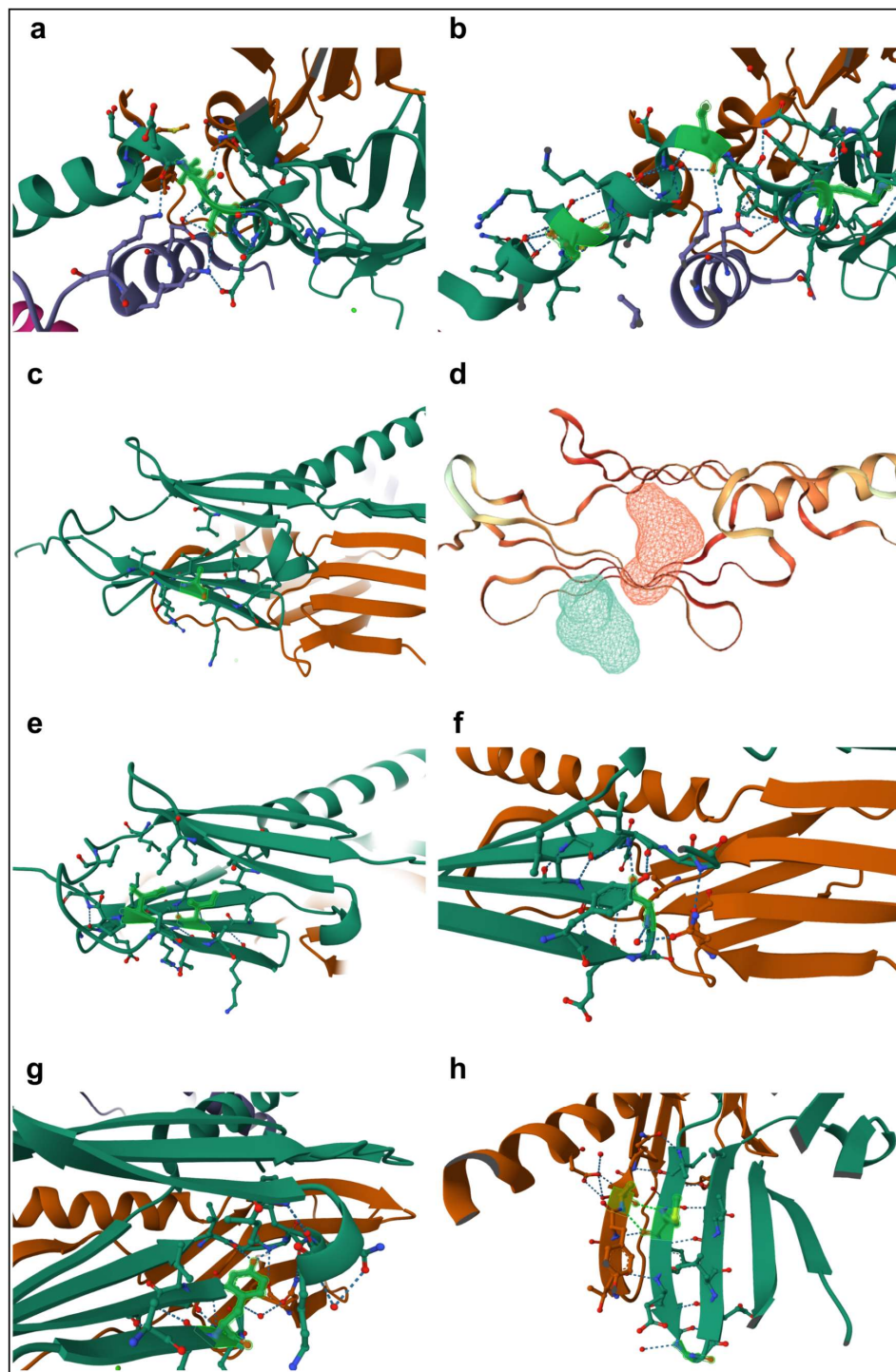


Figura 14. Residuos claves e interacciones estructurales en Hsp20. **a)** Relaxación de la hélice- α en residuos 16 y 17 resaltados; imagen realizada a partir de la estructura cristalizada 4YLB. **b)** Residuos claves G10, 15 y 18, resaltados de izquierda a derecha, y sus interacciones con otros residuos; imagen realizada a partir de la estructura cristalizada 4YLB. **c)** Residuo clave A102 y sus interacciones con otros residuos; imagen realizada a partir de la estructura cristalizada 4YLB. **d)** Cavidades en las que se encuentra el residuo clave A102; imagen tomada de la base de datos CaviDB. **e)** Resaltados de izquierda a derecha, residuos claves I113 y I111 y sus interacciones con otros residuos. **f)** Residuo clave G107 y sus interacciones con otros residuos. **g)** Residuo clave Y104 y sus interacciones con otros residuos. **h)** Resaltados de izquierda a derecha, residuos claves A77, M30 y G34 y sus interacciones con otros residuos. Las imágenes a), b) y c) fueron realizadas a partir de la estructura cristalizada 4YLB, mientras que el resto de las imágenes fueron realizadas con la estructura cristalizada 4YL9. Todas las imágenes, excepto la imagen **d**, fueron realizadas con el visualizador Glmol.

6. CONCLUSIÓN

En el presente seminario de investigación presentamos una metodología que permitió identificar residuos claves para preservar el estado nativo de las proteínas, definido como el conjunto de confórmeros en equilibrio dinámico entre sí. La metodología presenta avances importantes respecto a la previamente desarrollada en trabajos anteriores. Por un lado, se incorporó la capacidad de distinguir los distintos tipos posibles de mutaciones utilizando un potencial sensible al tipo de interacciones específicas entre los distintos aminoácidos. Por otro lado, la metodología se ha extendido a la posibilidad de estudiar el conjunto de cambios conformacionales posibles entre todos los distintos pares de confórmeros existentes en el estado nativo, cuando los trabajos previos sólo se limitaron a cambios conformacionales de tipo *apo-holo*. Esta nueva metodología permite una comparación más directa con los efectos de mutaciones reportados experimentalmente y, a su vez, una mayor exploración de alternativas de residuos claves.

Los resultados obtenidos mostraron que varios de los residuos claves, identificados como tales en la mayoría de los cambios conformacionales, resultan ser a) residuos cuyas mutaciones han sido reportadas experimentalmente como causantes de patologías; b) presentan un alto grado de conservación evolutiva; c) participan de interacciones entre elementos de distintas estructuras secundarias; d) participan de sitios activos y/o cavidades relacionadas con la función biológica.

Por estos motivos, consideramos que la presente tesina representa un primer paso prometedor para encarar un estudio más abarcativo. Esto es, a) considerando un mayor número de proteínas con diversidad conformacional reportada; b) incorporando análisis detallados de correlación entre el grado de conservación de los residuos cuyas mutaciones afectan los subespacios vibracionales involucrados en los cambios conformacionales del estado nativo y la conservación evolutiva; y c) incluyendo análisis detallados de la variación del impacto sobre los cambios conformacionales entre las diferentes mutaciones específicas por sitio, clasificadas según las propiedades fisicoquímicas de los residuos.

7. REFERENCIAS BIBLIOGRÁFICAS

- [Bahar-2000] Keskin, O., Jernigan, R. L., & Bahar, I. (2000). *Proteins with similar architecture exhibit similar large-scale dynamic behavior*. Biophysical journal, 78(4), 2093–2106. [https://doi.org/10.1016/S0006-3495\(00\)76756-7](https://doi.org/10.1016/S0006-3495(00)76756-7)
- [Bahar-2001] Atilgan, A. R., Durell, S. R., Jernigan, R. L., Demirel, M. C., Keskin, O., & Bahar, I. (2001). *Anisotropy of fluctuation dynamics of proteins with an elastic network model*. Biophysical journal, 80(1), 505–515. [https://doi.org/10.1016/S0006-3495\(01\)76033-X](https://doi.org/10.1016/S0006-3495(01)76033-X)
- [Bahar-2002a] Temiz N. A. and Bahar I. *Inhibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase*. (2002) Proteins 49(1):61-70.
- [Bahar-2002b] Isin B., Doruker P. and Bahar I. *Functional motions of influenza virus hemagglutinin: a structure-based analytical approach*. (2002) Biophys. J. 82(2):569-81.
- [Bahar-2003] Xu, C., Tobi, D., & Bahar, I. (2003). *Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T<-->R2 transition*. Journal of molecular biology, 333(1), 153–168. <https://doi.org/10.1016/j.jmb.2003.08.027>
- [Bahar-2004] Chen, S. C., & Bahar, I. (2004). *Mining frequent patterns in protein structures: a study of protease families*. Bioinformatics (Oxford, England), 20 Suppl 1(Suppl 1), i77–i85. <https://doi.org/10.1093/bioinformatics/bth912>
- [Bahar-2006] Cui Q., Bahar I. (2006). *Normal Mode Analysis: Theory and applications to biological and chemical Systems*. Chapman and Hall/CRC, Taylor and Francis Group. <https://doi.org/10.1201/9781420035070>
- [Bell-1970] Bell RJ, Dean P, Hibbins-Butter DC. *Localization of normal modes*. 1970; 3:2111–8.
- [Bender-2012] Bender R. A. (2012). *Regulation of the histidine utilization (hut) system in bacteria*. Microbiology and molecular biology reviews: MMBR, 76(3), 565–584. <https://doi.org/10.1128/MMBR.00014-12>
- [Benkovic-2006] Hammes-Schiffer, S., & Benkovic, S. J. (2006). *Relating protein motion to catalysis*. Annual review of biochemistry, 75, 519–541. <https://doi.org/10.1146/annurev.biochem.75.103004.142800>
- [Ben-Tal-2009] Goldenberg, O., Erez, E., Nimrod, G., & Ben-Tal, N. (2009). *The ConSurf-DB: pre-calculated evolutionary conservation profiles of protein structures*. Nucleic acids research, 37(Database issue), D323–D327. <https://doi.org/10.1093/nar/gkn822>
- [Ben-Tal-2020] Ben Chorin, A., Masrati, G., Kessel, A., Narunsky, A., Sprinzak, J., Lahav, S., Ashkenazy, H., & Ben-Tal, N. (2020). *ConSurf-DB: An accessible repository for the evolutionary conservation patterns of the majority of PDB proteins*. Protein science: a publication of the Protein Society, 29(1), 258–267. <https://doi.org/10.1002/pro.3779>
- [Brooks-1995] Janežič, D. and Brooks, B.R. (1995). *Harmonic analysis of large systems. II. Comparison of different protein models*. J. Comput. Chem., 16: 1543-1553. <https://doi.org/10.1002/jcc.540161210>
- [Brooks-2004] Patel, S., Mackerell, A. D., Jr, & Brooks, C. L., 3rd (2004). *CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular*

- dynamics simulations using a nonadditive electrostatic model*. Journal of computational chemistry, 25(12), 1504–1514. <https://doi.org/10.1002/jcc.20077>
- [Case-2003]** Ponder, J. W., & Case, D. A. (2003). *Force fields for protein simulations*. Advances in protein chemistry, 66, 27–85. [https://doi.org/10.1016/s0065-3233\(03\)66002-x](https://doi.org/10.1016/s0065-3233(03)66002-x)
- [Chirikjian-2009]** Schuyler, A. D., Jernigan, R. L., Qasba, P. K., Ramakrishnan, B., & Chirikjian, G. S. (2009). *Iterative cluster-NMA: A tool for generating conformational transitions in proteins*. Proteins, 74(3), 760–776. <https://doi.org/10.1002/prot.22200>
- [Collet-2009]** Depuydt, M., Leonard, S. E., Vertommen, D., Denoncin, K., Morsomme, P., Wahni, K., Messens, J., Carroll, K. S., & Collet, J. F. (2009). *A periplasmic reducing system protects single cysteine residues from oxidation*. Science (New York, N.Y.), 326(5956), 1109–1111. <https://doi.org/10.1126/science.1179557>
- [Covell-1999]** Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R., & Covell, D. G. (1999). *Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function*. Journal of molecular biology, 285(3), 1023–1037. <https://doi.org/10.1006/jmbi.1998.2371>
- [Covell-2002a]** Keskin, O., Durell, S. R., Bahar, I., Jernigan, R. L., & Covell, D. G. (2002). *Relating molecular flexibility to function: a case study of tubulin*. Biophysical journal, 83(2), 663–680. [https://doi.org/10.1016/S0006-3495\(02\)75199-0](https://doi.org/10.1016/S0006-3495(02)75199-0)
- [Covell-2002b]** Keskin, O., Bahar, I., Flatow, D., Covell, D. G., & Jernigan, R. L. (2002). *Molecular mechanisms of chaperonin GroEL-GroES function*. Biochemistry, 41(2), 491–501. <https://doi.org/10.1021/bi011393x>
- [Durandy-2003]** Imai, K., Slupphaug, G., Lee, W. I., Revy, P., Nonoyama, S., Catalan, N., Yel, L., Forveille, M., Kavli, B., Krokan, H. E., Ochs, H. D., Fischer, A., & Durandy, A. (2003). *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination*. Nature immunology, 4(10), 1023–1028. <https://doi.org/10.1038/ni974>
- [Elliott-1999]** Taraskin S. N. and Elliott S. R. *Anharmonicity and localization of atomic vibrations in vitreous silica*. (1999) Phys Rev B.; 59(13):8572–85. <https://doi.org/10.1103/PhysRevB.59.8572>
- [Erman-1998]** Bahar I., Atilgan A. R., Demirel M. C., Erman B. (1998) *Vibrational Dynamics of Folded Proteins: Significance of Slow and Fast Motions in Relation to Function and Stability*. Phys. Rev. Lett. 80(12), 2733-2736. <https://doi.org/10.1103/PhysRevLett.80.2733>
- [Fernandez-Alberti-2016]** Saldaño, T. E., Monzon, A. M., Parisi, G., & Fernandez-Alberti, S. (2016). *Evolutionary Conserved Positions Define Protein Conformational Diversity*. PLoS computational biology, 12(3), e1004775. <https://doi.org/10.1371/journal.pcbi.1004775>
- [Gerstein-2004]** Goh, C. S., Milburn, D., & Gerstein, M. (2004). *Conformational changes associated with protein-protein interactions*. Current opinion in structural biology, 14(1), 104–109. <https://doi.org/10.1016/j.sbi.2004.01.005>
- [Jernigan-2004]** Wang, Y., Rader, A. J., Bahar, I., & Jernigan, R. L. (2004). *Global ribosome motions revealed with elastic network model*. Journal of structural biology, 147(3), 302–314. <https://doi.org/10.1016/j.jsb.2004.01.005>

- [Karplus-1995]** Brooks, B.R., Janežič, D. and Karplus, M. (1995). *Harmonic analysis of large systems. I. Methodology*. J. Comput. Chem., 16: 1522-1542.
<https://doi.org/10.1002/jcc.540161209>
- [Kimura-1974]** Kimura, M., & Ohta, T. (1974). *On some principles governing molecular evolution*. Proceedings of the National Academy of Sciences of the United States of America, 71(7), 2848–2852. <https://doi.org/10.1073/pnas.71.7.2848>
- [Kneller-1999]** Hinsen K. and Kneller G. R. (1999). *A simplified force field for describing vibrational protein dynamics over the whole frequency range*. J. Chem. Phys. 24:10766–10769. <https://doi.org/10.1063/1.480441>
- [Krzanowski-1979]** Krzanowski, W.J. (1979). *Between-Groups Comparison of Principal Components*. Journal of the American Statistical Association, 74, 703-707.
- [Maguid-2005]** Maguid, S., Fernandez-Alberti, S., Ferrelli, L., & Echave, J. (2005). *Exploring the common dynamics of homologous proteins*. Application to the globin family. Biophysical journal, 89(1), 3–13. <https://doi.org/10.1529/biophysj.104.053041>
- [Maritan-2004]** Micheletti, C., Carloni, P., & Maritan, A. (2004). *Accurate and efficient description of protein vibrational dynamics: comparing molecular dynamics and Gaussian models*. Proteins, 55(3), 635–645. <https://doi.org/10.1002/prot.20049>
- [Mc Cammon-1987]** McCammon J. A. and Harvey S. C. (1987). *Dynamics of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge. [https://doi.org/10.1016/0307-4412\(89\)90164-7](https://doi.org/10.1016/0307-4412(89)90164-7)
- [Metcalf-2002]** Haebel, P. W., Goldstone, D., Katzen, F., Beckwith, J., & Metcalf, P. (2002). *The disulfide bond isomerase DsbC is activated by an immunoglobulin-fold thiol oxidoreductase: crystal structure of the DsbC-DsbD α complex*. The EMBO journal, 21(18), 4774–4784. <https://doi.org/10.1093/emboj/cdf489>
- [Montgomery-1988]** Brooks III C. L., Karplus M., Montgomery Pettitt B. (1988). *Advances in Chemical Physics Proteins: A theoretical perspective of dynamics, structure, and thermodynamics*. John Wiley and sons. Vol. LXXI.
- [Najmanovich-2014]** Frappier V, Najmanovich RJ (2014). *A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations*. PLoS Comput Biol 10(4): e1003569. doi:10.1371/journal.pcbi.1003569
- [Najmanovich-2021]** Mailhot, O., & Najmanovich, R. (2021). *The NRGTEN Python package: an extensible toolkit for coarse-grained normal mode analysis of proteins, nucleic acids, small molecules and their complexes*. Bioinformatics (Oxford, England), btab189. Advance online publication. <https://doi.org/10.1093/bioinformatics/btab189>
- [Nei-1987]** Saitou, N., & Nei, M. (1987). *The neighbor-joining method: a new method for reconstructing phylogenetic trees*. Molecular biology and evolution, 4(4), 406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>
- [Noussinov-2004]** Gunasekaran, K., Ma, B., & Nussinov, R. (2004). *Is allostery an intrinsic property of all dynamic proteins?* Proteins, 57(3), 433–443. <https://doi.org/10.1002/prot.20232>

[Orengo-2008] Redfern, O. C., Dessailly, B., & Orengo, C. A. (2008). *Exploring the structure and function paradigm*. Current opinion in structural biology, 18(3), 394–402. <https://doi.org/10.1016/j.sbi.2008.05.007>

[Orozco-2008] Emperador, A., Carrillo, O., Rueda, M., & Orozco, M. (2008). *Exploring the suitability of coarse-grained techniques for the representation of protein dynamics*. Biophysical journal, 95(5), 2127–2138. <https://doi.org/10.1529/biophysj.107.119115>

[Parisi-2016] Monzon, A. M., Rohr, C. O., Fornasari, M. S., & Parisi, G. (2016). *CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state*. Database: the journal of biological databases and curation, 2016, baw038. <https://doi.org/10.1093/database/baw038>

[Parisi-2022] Rueda, A. J. V., Bulgarelli, F. L., Palopoli, N., & Parisi, G. (2022). *CaviDB: a database of cavities and their features in the structural and conformational space of proteins*. BioRxiv, 2022.08.02.502569. <https://doi.org/10.1101/2022.08.02.502569>

[Phillips-1992] Somers, W. S., & Phillips, S. E. (1992). *Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by beta-strands*. Nature, 359(6394), 387–393. <https://doi.org/10.1038/359387a0>

[Pupko-2004] Mayrose, I., Graur, D., Ben-Tal, N., & Pupko, T. (2004). *Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior*. Molecular biology and evolution, 21(9), 1781–1791. <https://doi.org/10.1093/molbev/msh194>

[Qiu-2020] Chi, X., Li, Y., & Qiu, X. (2020). *V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation*. Immunology, 160(3), 233–247. <https://doi.org/10.1111/imm.13176>

[Reitzer-2014] Reitzer, L. (2014). *Amino Acid Synthesis. In Reference Module in Biomedical Sciences*. Elsevier. <https://doi.org/10.1016/b978-0-12-801238-3.02427-2>

[Rose-2021] Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S. K., Koča, J., & Rose, A. S. (2021). *Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures*. Nucleic acids research, 49(W1), W431–W437. <https://doi.org/10.1093/nar/gkab314>

[Sali-2016] Webb, B., & Sali, A. (2016). *Comparative Protein Structure Modeling Using MODELLER*. Current protocols in bioinformatics, 54, 5.6.1–5.6.37. <https://doi.org/10.1002/cpbi.3>

[Sanejouand-2001] Tama, F., & Sanejouand, Y. H. (2001). *Conformational change of proteins arising from normal mode calculations*. Protein engineering, 14(1), 1–6. <https://doi.org/10.1093/protein/14.1.1>

[Tawfik-2009] Tokuriki, N., & Tawfik, D. S. (2009). *Protein dynamism and evolvability*. Science (New York, N.Y.), 324(5924), 203–207. <https://doi.org/10.1126/science.1169375>

The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.

[Thirumalai-2006] Zheng, W., Brooks, B. R., & Thirumalai, D. (2006). *Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations*. Proceedings of the National Academy of Sciences of the United States of America, 103(20), 7664–7669. <https://doi.org/10.1073/pnas.0510426103>

[Thornton-2005] Watson, J. D., Laskowski, R. A., & Thornton, J. M. (2005). *Predicting protein function from sequence and structural data*. Current opinion in structural biology, 15(3), 275–284. <https://doi.org/10.1016/j.sbi.2005.04.003>

[Uniprot-2010] Hinz, U., & UniProt Consortium (2010). *From protein sequences to 3D-structures and beyond: the example of the UniProt knowledgebase*. Cellular and molecular life sciences: CMLS, 67(7), 1049–1064. <https://doi.org/10.1007/s00018-009-0229-6>

[Walsh-2010] Sleator, R. D., & Walsh, P. (2010). *An overview of in silico protein function prediction*. Archives of microbiology, 192(3), 151–155. <https://doi.org/10.1007/s00203-010-0549-9>

[Wolynes-1991] Frauenfelder, H., Sligar, S. G., & Wolynes, P. G. (1991). *The energy landscapes and motions of proteins*. Science (New York, N.Y.), 254(5038), 1598–1603. <https://doi.org/10.1126/science.1749933>

[Yun-2015] Liu, L., Chen, J. Y., Yang, B., Wang, F. H., Wang, Y. H., & Yun, C. H. (2015). *Active-State Structures of a Small Heat-Shock Protein Revealed a Molecular Switch for Chaperone Function*. Structure (London, England: 1993), 23(11), 2066–2075. <https://doi.org/10.1016/j.str.2015.08.015>

[Zakin-1984] Saint-Girons, I., Duchange, N., Cohen, G. N., & Zakin, M. M. (1984). *Structure of autoregulation of the metJ regulatory gene in Escherichia coli*. Journal of Biological Chemistry, 259(22), 14282–14285. [https://doi.org/10.1016/s0021-9258\(18\)89890-8](https://doi.org/10.1016/s0021-9258(18)89890-8)

[Zúñiga-2011] Kalstein A., Fernández-Alberti S., Bastida A., Soler M. A., Farag M.H., Zúñiga J. (2011). *Vibrational dynamics of polyatomic molecules in solution: assignment, time evolution and mixing of instantaneous normal modes*. Theor Chem Acc.; 128(4–6):769–82.