



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



Universidad  
Nacional  
de Quilmes

Saldaño, Tadeo Enrique

# Residuos dinámicamente importantes en 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:*

Saldaño, T. D. (2018) *Residuos dinámicamente importantes en la diversidad conformacional de proteínas. (Tesis de doctorado). Universidad Nacional de Quilmes, Bernal, Argentina. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes*  
<http://ridaa.unq.edu.ar/handle/20.500.11807/794>

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

## **Residuos dinámicamente importantes en la diversidad conformacional de proteínas**

*TESIS DOCTORAL*

**Tadeo Enrique Saldaño**

tadeoenriquesalda@gmail.com

### **Resumen**

Las propiedades de flexibilidad y dinámica de las proteínas han sido desde hace tiempo asociadas a sus funciones biológicas. Estas propiedades han sido originalmente usadas para explicar la heterogeneidad de las propiedades de unión de la seroalbúmina bovina por Kurush en 1950 y su descripción formal fue incluida en el clásico modelo de Monod-Wyman-Changeux de regulación alostérica, también conocido como modelo de pre-equilibrio.

Actualmente está bien establecido que la forma funcional de una proteína, normalmente conocida como el estado nativo, no es única. El concepto de “embudo” propuesto por Wolynes et al. en 1995 ha sido utilizado para explicar los canales de plegamiento de las proteínas considerando un “fondo de potencial rugoso” representando al conjunto de isómeros conformacionales. El grado de la diversidad conformacional de una proteína puede entonces relacionarse con la extensión de la “rugosidad del fondo del embudo”. La distribución de los conformeros en este potencial ha sido previamente asociada al plegamiento, la historia evolutiva, y la presencia de ciertas mutaciones.

Actualmente, los conceptos antes mencionados han derivado en la noción de la preexistencia de poblaciones de conformeros en equilibrio dinámico (dinamismo) sobre la superficie de energía potencial de una proteína. Este equilibrio dinámico soporta la hipótesis de la unión del ligando a una conformación específica de mayor afinidad y ofrece una visión central que permite explicar la función biológica.

La dinámica vibracional intramolecular asociada a cada conformación garantiza las transiciones conformacionales que, debido a su importancia, podrían estar asociadas con rasgos conservados evolutivamente. El análisis de modos normales, basado en un modelo de “grano grueso” de la proteína, puede proporcionar la información necesaria para explorar estas características.

En la primer parte de la tesis presentamos un nuevo procedimiento para identificar residuos en posiciones clave para el mantenimiento de la diversidad conformacional asociada a la unión al ligando. El método se aplicó a un conjunto de 188 pares de estructuras de proteínas, cristalizadas con y sin ligando. En primer lugar, se seleccionaron los modos normales implicados en el cambio conformacional de acuerdo con las distorsiones estructurales introducidas por unión al ligando. El subespacio definido por estos modos se utilizó para analizar el efecto de mutaciones puntuales en la conservación de la diversidad conformacional de la proteína. Definimos las posiciones cuyas mutaciones alteran en mayor medida estos subespacios como posiciones claves, es decir, son residuos dinámicamente importantes que median el cambio conformacional de unión al ligando. Encontramos una correlación negativa entre la conservación evolutiva de los residuos de una proteína y el impacto de sus mutaciones sobre los subespacios de modos normales asociados a la unión del ligando. Estas posiciones se muestran conservadas evolutivamente, en su mayoría son residuos con baja exposición al solvente, alifáticos y localizados en estructuras regulares de la proteína como hoja- $\beta$  y hélice- $\alpha$ .

Durante la segunda etapa del doctorado, combinamos la información obtenida de métodos basados en propiedades estructurales y dinámicas de proteínas con información relacionada

con redes de interacciones de residuos. La representación de la estructura proteica como red facilita la búsqueda de determinantes topológicos, que pueden estar relacionados con residuos funcionalmente importantes.

Cada estructura proteica de nuestro set de estudio se representó como una red de contacto de residuos y se realizó un análisis exhaustivo de las propiedades de la red. Se analizaron tres parámetros topológicos de redes, Grado, Centralidad de Intermediación y Proximidad Central. Particularmente los últimos dos se han previamente utilizado en diversos trabajos para identificar residuos funcionales ya que la Centralidad de Intermediación refleja el control que ejerce un nodo sobre las interacciones de otros nodos en la red y la Proximidad Central se asocia a cuan rápido se propaga la información desde un nodo dado a otros nodos alcanzables en la red.

Estudiamos la correlación de estos parámetros topológicos de redes con nuestro parámetro para definir mutaciones que alteran en mayor medida el subespacio de modos normales asociado a la unión al ligando. Además, analizamos la correlación entre la conservación evolutiva y el área de accesibilidad al solvente, con estos parámetros topológicos. Encontramos una buena correlación entre los efectos de las mutaciones en los subespacios de modos normales asociados a la unión del ligando y los parámetros topológicos de redes. Estos parámetros han sido utilizados para predecir con éxito sitios activos o sitios funcionales en varias proteínas. Esto apoya nuestro método para detectar residuos dinámicamente importantes que median el cambio conformacional de unión al ligando.

Determinamos que los residuos dinámicamente relevantes tienden a estar interconectados, por lo que es posible definir redes de residuos que modulan dinámicamente los cambios conformacionales.

Como etapa final de la tesis, exploramos las posiciones claves que sustentan la estabilidad dinámica de la estructura homotetramérica de la Transtiretina Humana. El estado nativo de la Transtiretina presenta dos sitios de unión a la hormona tiroxina, que son generados por la interfaz dímero-dímero, mediante interacciones débiles. La disociación de la estructura tetramérica es el primer paso del proceso de formación de fibrillas de amiloide. Un gran número de mutaciones puntuales por desestabilizar la estructura cuaternaria del tetrámero muestran efectos pro-amiloidogénicos.

Basándonos en el análisis de modos normales y su respuesta a las perturbaciones locales hemos identificado las posiciones cuyas mutaciones alteran en mayor medida la dinámica de equilibrio del tetrámero de Transtiretina. Encontramos que estas posiciones están localizadas principalmente en las hojas- $\beta$  E y F, del monómero de Transtiretina y el *loop* entre estas estructuras secundarias.

Determinamos que la interfase monómero-monómero es una de las regiones más vulnerables, ya que mutaciones en residuos de esta región conducen a cambios significativos en la dinámica de equilibrio del tetrámero y, por lo tanto, favorece la disociación de la estructura. Además, hemos encontrado que las mutaciones en los residuos localizados en la interfaz dímero-dímero y/o en el sitio de unión a la hormona tiroxina desestabilizan al tetrámero más que el promedio. La unión de la tiroxina estabiliza la estructura tetramérica estableciendo interacciones con los residuos en la interfaz dímero-dímero.

Debido a esto se han propuesto varios compuestos como drogas en la terapia de la amiloidosis por Transtiretina, siendo clave la afinidad por el sitio unión del tetrámero, y por tanto, la contribución a la integridad de la estructura. Comparamos varios compuestos de acuerdo a su efecto sobre las vibraciones asociadas a la unión del ligando. Discutimos y analizamos nuestra comparación de drogas en términos de parámetros y mediciones asociadas a las afinidades de unión al ligando y la estabilización del estado nativo del tetrámero de Transtiretina.

A pesar de la presencia en el tetrámero de dos sitios de unión idénticos, la unión de la tiroxina en solución se caracteriza por una fuerte cooperatividad negativa. La flexibilidad estructural del tetrámero, simulada mediante el análisis de modos normales, expuso patrones vibratorios asimétricos en ambos dímeros. Las fluctuaciones térmicas revelan diferencias en el tamaño y la flexibilidad de las cavidades de unión de los ligandos en la interfaz dímero-dímero. Es decir, pequeñas diferencias estructurales entre los dímeros, o monómeros pueden conducir a diferencias funcionales significativas en la dinámica del tetrámero de Transtiretina.



# **RESIDUOS DINÁMICAMENTE IMPORTANTES EN LA DIVERSIDAD CONFORMACIONAL DE PROTEÍNAS**

Saldaño Tadeo Enrique

Trabajo de Tesis para optar al título de  
Doctor en Ciencia y Tecnología, de la Universidad Nacional de Quilmes

Director de Tesis: Dr. Sebastián Fernández Alberti

Co-director de Tesis: Dr. Gustavo Parisi

Lugar de trabajo: Unidad de Físicoquímica, Departamento de Ciencia y Tecnología  
Universidad Nacional de Quilmes

2018





---

## AGRADECIMIENTOS

A mi esposa María Florencia Vadillo que es la persona que me hace ser un poco mejor cada día y la AMO con toda mi vida.

A Graciela Beatriz Blanco y Enrique Mario Saldaño, mamá y papá, que principalmente me dieron la vida, amor y enseñanzas.

A mis hermanos, Simón Antonio, Joaquín Federico, Jesús Bautista y Matías José que son y van a seguir siendo la alegría de mi vida. De mas esta decir que somos los Hermanos Saldaño!!

A mis suegros Sylvana Armengol y Antonio Vadillo Martinez, porque le dieron la vida a la mujer hermosa que me acompaña en esta vida. También agradezco a sus hermanos Santiago, Caterina, Paz, Luciana y Sol, junto con Martin y Jorge, porque me aceptaron como miembro de su familia.

A mis sobrinos con los cuales juego y me dejan ser un niño mas.

Agradezco a mi director (o jefe como le digo ante mis amigos que no entienden el sistema universitario o académico) Sebastián Fernández Alberti por estos 7 años compartidos, desde mi tesina de grado hasta esta Tesis doctoral.

A mi co-director Gustavo Parisi por la contribución enorme que tuvo en este trabajo, por compartir generosamente sus conocimientos de biología estructural y bioinformática

A todos mis compañeros de grupo Beatriz Rodríguez, Dianelys Ondarse, Javier Ríos, Laura Alfonso Hernández, Nicolás Oldani, Patricio Germán Barletta, Ricardo Frankling Mergarejo y Víctor Manuel Freixas Lemus.

Al *Structural Bionformatic Group* o “los Parisi” Alexander Monzon, Ana Julia Valez Rueda, Cristián Emanuel Guisande Donadio, Guillermo Ignacio Benítez, Julia Marchetti, Marcia Hasenahuer, Nicolas Palopoli y Silvina Fornasari.

Al grupo de la Unidad de Físicoquímica, oficina 31 Agustín Ormazábal, Claudia Gonzalez, Juliana Palma, Gustavo Pierdominicci, Rodrigo Cossio Pérez y Vanesa Racigh.

A mis amigos en estos 10 años dentro de la UNQ, Macarena Siri, Romina G. Armando, Matias Lorch y Luis Federico Veiga, Anastasia Herrero, Nahuel Cejas y Joaquín Urdinez.

A mis dos ingenieros favoritos recibidos en la UNQ Ornella Di Marco y Pablo Gabriel Cadeddu.

A mis amigos de la ciudad de Azul.

Por ultimo quiero agradecer a los jurados que muy amablemente aceptaron participar de esta defensa de tesis.



*Los hermanos sean unidos  
Porque esa es la ley primera  
Tengan unión verdadera  
En cualquier tiempo que sea  
Porque si entre ellos pelean  
Los devoran los de afuera.*

El Gaucho Martín Fierro, 1872

José Hernández



---

## RESUMEN

Las propiedades de flexibilidad y dinámica de las proteínas han sido desde hace tiempo asociadas a sus funciones biológicas. Estas propiedades han sido originalmente usadas para explicar la heterogeneidad de las propiedades de unión de la seroalbúmina bovina por Kurush en 1950 y su descripción formal fue incluida en el clásico modelo de Monod-Wyman-Changeux de regulación alostérica, también conocido como modelo de pre-equilibrio.

Actualmente está bien establecido que la forma funcional de una proteína, normalmente conocida como el estado nativo, no es única. El concepto de “embudo” propuesto por Wolynes et al. en 1995 ha sido utilizado para explicar los canales de plegamiento de las proteínas considerando un “fondo de potencial rugoso” representando al conjunto de isómeros conformacionales. El grado de la diversidad conformacional de una proteína puede entonces relacionarse con la extensión de la “rugosidad del fondo del embudo”. La distribución de los conformeros en este potencial ha sido previamente asociada al plegamiento, la historia evolutiva, y la presencia de ciertas mutaciones. Actualmente, los conceptos antes mencionados han derivado en la noción de la preexistencia de poblaciones de conformeros en equilibrio dinámico (dinamismo) sobre la superficie de energía potencial de una proteína. Este equilibrio dinámico soporta la hipótesis de la unión del ligando a una conformación específica de mayor afinidad y ofrece una visión central que permite explicar la función biológica.

La dinámica vibracional intramolecular asociada a cada conformación garantiza las transiciones conformacionales que, debido a su importancia, podrían estar asociadas con rasgos conservados evolutivamente. El análisis de modos normales, basado en un modelo de “grano grueso” de la proteína, puede proporcionar la información necesaria para explorar estas características.

En la primer parte de la tesis presentamos un nuevo procedimiento para identificar residuos en posiciones clave para el mantenimiento de la diversidad conformacional asociada a la unión al ligando. El método se aplicó a un conjunto de 188 pares de estructuras de proteínas, cristalizadas con y sin ligando. En primer lugar, se seleccionaron los modos normales implicados en el cambio conformacional de acuerdo con las distorsiones estructurales introducidas por unión al ligando. El subespacio definido por estos modos se utilizó para analizar el efecto de mutaciones puntuales en la conservación de la diversidad conformacional de la proteína. Definimos las posiciones cuyas mutaciones alteran en mayor medida estos subespacios como posiciones claves, es decir, son residuos dinámicamente importantes que median el cambio conformacional de unión al ligando. Encontramos una correlación negativa entre la conservación evolutiva de los residuos de una proteína y el impacto de sus mutaciones sobre los subespacios de modos normales asociados a la unión del ligando. Estas posiciones se muestran conservadas evolutivamente, en su mayoría son residuos con baja exposición al solvente, alifáticos y localizados en estructuras regulares de la proteína como hoja- $\beta$  y hélice- $\alpha$ .

Durante la segunda etapa del doctorado, combinamos la información obtenida de métodos basados en propiedades estructurales y dinámicas de proteínas con información relacionada con redes de interacciones de residuos. La representación de la estructura proteica como red facilita la búsqueda de determinantes topológicos, que pueden estar relacionados con residuos funcionalmente importantes. Cada estructura proteica de nuestro set de estudio se representó como una red de contacto de residuos y se realizó un análisis exhaustivo de las propiedades de la red. Se analizaron tres parámetros topológicos de redes, Grado, Centralidad de Intermediación y Proximidad Central. Particularmente los últimos dos se han previamente utilizado en diversos trabajos para identificar residuos funcionales ya que la Centralidad de Intermediación refleja el control que ejerce un nodo sobre las interacciones de otros nodos en la red y la Proximidad Central se asocia a cuan rápido se propaga la información desde un nodo dado a otros nodos alcanzables en la red.

Estudiamos la correlación de estos parámetros topológicos de redes con nuestro parámetro para definir mutaciones que alteran en mayor medida el subespacio de modos normales asociado a la unión al ligando. Además, analizamos la correlación entre la conservación evolutiva y el área de accesibilidad al solvente, con estos parámetros topológicos. Encontramos una buena correlación entre los efectos de las mutaciones en los subespacios de modos normales asociados a la unión del ligando y los parámetros topológicos de redes. Estos parámetros han sido utilizados para predecir con éxito sitios activos o sitios funcionales en varias proteínas. Esto apoya nuestro método para detectar residuos dinámicamente importantes que median el cambio conformacional de unión al ligando. Determinamos que los residuos dinámicamente relevantes tienden a estar interconectados, por lo que es posible definir redes de residuos que modulan dinámicamente los cambios conformacionales.

Como etapa final de la tesis, exploramos las posiciones claves que sustentan la estabilidad dinámica de la estructura homotetramérica de la Transtiretina Humana. El estado nativo de la Transtiretina presenta dos sitios de unión a la hormona tiroxina, que son generados por la interfaz dímero-dímero, mediante interacciones débiles. La disociación de la estructura tetramérica es el primer paso del proceso de formación de fibrillas de amiloide. Un gran número de mutaciones puntuales por desestabilizar la estructura cuaternaria del tetrámero muestran efectos pro-amiloidogénicos. Basándonos en el análisis de modos normales y su respuesta a las perturbaciones locales hemos identificado las posiciones cuyas mutaciones alteran en mayor medida la dinámica de equilibrio del tetrámero de Transtiretina. Encontramos que estas posiciones están localizadas principalmente en las hojas- $\beta$  E y F, del monómero de Transtiretina y el *loop* entre estas estructuras secundarias. Determinamos que la interfase monómero-monómero es una de las regiones más vulnerables, ya que mutaciones en residuos de esta región conducen a cambios significativos en la dinámica de equilibrio del tetrámero y, por lo tanto, favorece la disociación de la estructura. Además, hemos encontrado que las mutaciones en los residuos localizados en la interfaz dímero-dímero y/o en el sitio de unión a la hormona tiroxina desestabilizan al tetrámero más que el promedio. La unión de la tiroxina estabiliza la estructura tetramérica estableciendo interacciones con los residuos en la interfaz dímero-dímero.

Debido a esto se han propuesto varios compuestos como drogas en la terapia de la amiloidosis por Transtiretina, siendo clave la afinidad por el sitio unión del tetrámero, y por tanto, la contribución a la integridad de la estructura. Comparamos varios compuestos de acuerdo a su efecto sobre las vibraciones asociadas a la unión del ligando. Discutimos y analizamos nuestra comparación de drogas en términos de parámetros y mediciones asociadas a las afinidades de unión al ligando y la estabilización del estado nativo del tetrámero de Transtiretina.

A pesar de la presencia en el tetrámero de dos sitios de unión idénticos, la unión de la tiroxina en solución se caracteriza por una fuerte cooperatividad negativa. La flexibilidad estructural del tetrámero, simulada mediante el análisis de modos normales, expuso patrones vibratorios asimétricos en ambos dímeros. Las fluctuaciones térmicas revelan diferencias en el tamaño y la flexibilidad de las cavidades de unión de los ligandos en la interfaz dímero-dímero. Es decir, pequeñas diferencias estructurales entre los dímeros, o monómeros pueden conducir a diferencias funcionales significativas en la dinámica del tetrámero de Transtiretina.





# ÍNDICE GENERAL

AGRADECIMIENTOS	iii
RESUMEN	vii
ÍNDICE GENERAL	xi
1. INTRODUCCIÓN	2
1.1 OBJETIVOS	4
1.2 ABREVIATURAS UTILIZADAS	4
2. ANÁLISIS DE MODOS NORMALES: FUNDAMENTOS	8
2.1 MODOS NORMALES DE VIBRACIÓN	8
2.1.1 Ecuaciones del movimiento vibracional	8
2.1.2 Solución de las ecuaciones de movimiento	11
2.2 MODELO DE RED ANISOTRÓPICA: FUNDAMENTOS	14
2.2.1 Adaptación del modelo: constante de fuerza asociada al tipo de interacción	18
2.2.2 Adaptación del modelo: simulación de mutación puntual	19
2.3 MODELO DE RED ANISOTRÓPICA: APLICACIONES	19
2.3.1 Factores B de temperatura	19
2.3.2 Conjunto de estructuras que representan fluctuaciones térmicas	21
2.3.3 Grado de colectividad	23
3. DINÁMICA VIBRACIONAL ASOCIADA A LA MULTIPLICIDAD CONFORMACIONAL	26
3.1 CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO: ASPECTOS ESTRUCTURALES	26
3.1.1 Determinación de diferencias estructurales (RMSD)	26
3.1.2 Definición del vector diferencia entre confórmeros	26
3.2 CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO: ASPECTOS DINÁMICOS	27
3.2.1 Subespacios vibracionales activos: definición	27
3.2.2 Subespacios vibracionales activos: comparación	29
3.2.3 Dinámica asociada a patrones de flexibilidad relativa	31
3.3 CARACTERIZACIÓN DE RESIDUOS DINÁMICAMENTE IMPORTANTES EN EL CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO	32
3.3.1 Residuos en posiciones claves: definición	32
3.3.2 Análisis de la conservación evolutiva	32
3.3.3 Determinación de estructura secundaria	33
3.3.4 Accesibilidad al solvente	34
3.3.5 Distancia de residuos al sitio activo	34
3.3.6 Ángulo entre elementos de estructura secundaria	35
3.3.7 Redes de contactos inter-atómicos	37

---

3.3.8 Parámetros topológicos de redes	39
4. RESULTADOS	44
4.1 RESIDUOS EN POSICIONES CLAVES PARA EL MANTENIMIENTO DE LA DIVERSIDAD CONFORMACIONAL	44
4.1.1 Determinación del set de proteínas	44
4.1.2 Identificación de residuos en posiciones claves	44
4.1.3 Caracterización de las posiciones claves	49
4.1.4 Ejemplos	55
4.2 REDES DE RESIDUOS EN POSICIONES CLAVES	58
4.2.1 Redes de residuos, conectividad y robustez	58
4.2.2 Ejemplos	59
4.2.3 Relación entre residuos en posiciones claves y parámetros topológicos de redes	62
4.3 EVALUACIÓN DEL EFECTO DE LAS MUTACIONES Y LA UNIÓN DEL LIGANDO A LA DINÁMICA DEL HOMOTETRÁMERO DE TRANSTIRETINA	65
4.3.1 Transtiretina Humana	65
4.3.2 Efecto de las mutaciones en la dinámica de equilibrio del tetrámero	68
4.3.3 Efecto de la unión de drogas a la dinámica de equilibrio del tetrámero	74
4.3.4 Evidencia de asimetría en la dinámica del tetrámero de Transtiretina	80
5. CONCLUSIONES GENERALES	88
APÉNDICE	92
A. LISTA DE AMINOÁCIDOS	92
B. PROTEÍNAS DEL SET DE ESTUDIO	93
C. TEOREMA DE EQUIPARTICIÓN	103
D. CORRELACIÓN DE PEARSON Y DE SPEARMAN	105
E. EJEMPLOS – RIN	107
REFERENCIAS	126

# *Capítulo 1*

## *Introducción*

*We're jamming*

*I wanna jam it with you*

*We're jamming, jamming*

*And I hope you like jamming too...*

Jamming, album Exodus, 1977.

Robert Nesta Marley Booker, Bob Marley

# 1. INTRODUCCIÓN

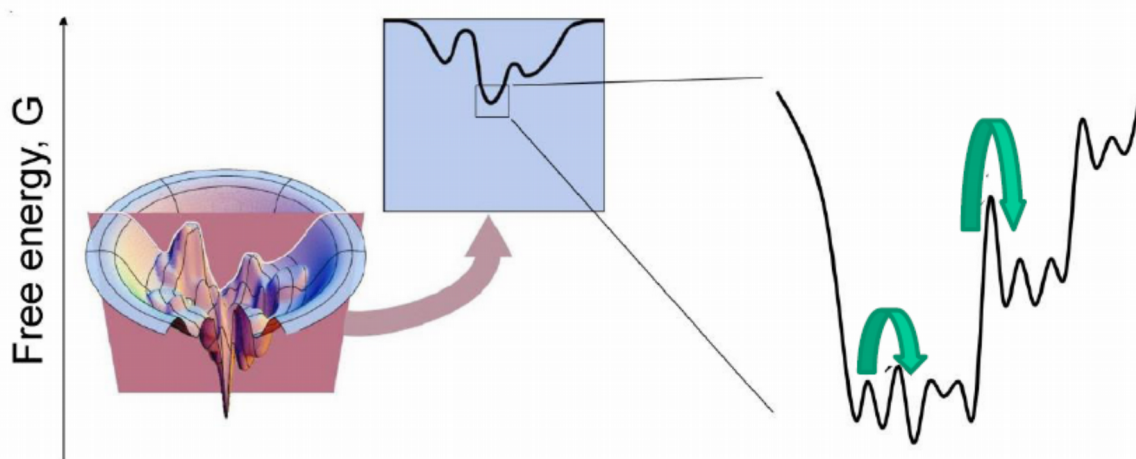
El avance en el desarrollo de nuevos métodos computacionales aplicados a la asignación de funciones biológicas a nuevas secuencias de proteínas representa un desafío crucial en la era post-genómica. Los protocolos estándar de anotación funcional consisten en transferir la anotación obtenida a partir de bases de datos. Esta transferencia se realiza generalmente mediante búsquedas de similitud secuencial que permiten capturar proteínas homólogas. Los métodos más comúnmente usados, tales como las búsquedas de similitud secuencia-secuencia, han sido reemplazados en los últimos años por métodos más poderosos que permiten la detección incluso de homólogos remotos. Como ejemplo de estos últimos podemos mencionar los métodos que utilizan perfiles y modelos de cadena de Markov Escondida (*Hidden Markov Models*) [Gutteridge 2005]. Con la finalidad de aumentar la sensibilidad en la detección de relaciones evolutivas distantes, se han incorporado métodos que utilizan el reconocimiento del plegamiento en la asignación funcional [Pazos 2004]. La hipótesis principal subyacente a estos procedimientos es que proteínas homólogas conservan su función biológica. Esta es la razón de la existencia de muchas bases de datos de familias de proteínas con anotaciones funcionales (por ejemplo PFAM [Bateman 2004], SMART [Letunic 2004], PROSITE [Hulo 2004]). Sin embargo, el reconocimiento por similitud obtenido en búsquedas usando estas bases de datos puede resultar en una proteína con diferente función dentro de una misma familia. Estas adaptaciones funcionales, pueden resultar por eventos de duplicación de genes, diferencias en dominios estructurales, mutaciones de posiciones funcionales claves, o especiación. Debido a esto, la pertenencia de una dada secuencia a una familia de proteínas indicaría, en el mejor de los casos, una amplia gama de funciones que la proteína puede realizar.

Con el objetivo de realizar asignaciones funcionales más específicas, sería importante contar con métodos para predecir el grado de diversificación funcional dentro de una misma familia y por lo tanto evaluar la existencia de subfamilias [Lassmann 2005]. El estudio de la diversificación funcional de familias en subfamilias ha brindado gran información sobre el origen de la adaptación funcional y la relativa promiscuidad por sustratos, sin embargo esto solo se ha aplicado a conjuntos de proteínas bien caracterizados [Thomas 2003]. La divergencia funcional dentro de una misma familia no es un hecho particular, sino que muchas familias han evolucionado para acomodar una amplia gama de funciones específicas contenidas en cada subfamilia [Khersonsky 2010].

Distintos esfuerzos se han realizado con el objeto de aumentar la especificidad en la asignación funcional. Así, distintas propiedades estructurales asociadas a la función tales como propiedades superficiales, composición de hendiduras y cavidades, han sido consideradas y relacionadas a la conservación evolutiva relativa de distintos residuos [Ferre 2004, Binkowski 2004]. El éxito de estos métodos se ha visto perjudicado por la existencia de distintas conformaciones posibles de una misma proteína, dificultando el reconocimiento de una tendencia que permita asociar

características estructurales particulares con la función [Gutteridge 2005]. Sin embargo, y de particular importancia para nuestro proyecto, dicha diversidad conformacional puede originarse tanto debido a la diversificación funcional de la proteína como a distorsiones estructurales asociadas a movimientos requeridos en la función biológica.

Actualmente está bien establecido que la forma funcional de una proteína, normalmente conocida como el estado nativo, no es única. El concepto de “embudo” [Wolynes 1995] ha sido utilizado para explicar el plegamiento de las proteínas considerando un “fondo de potencial rugoso” (ver **Figura 1.1**) donde cada uno de los picos de esta rugosidad representa a cada uno de los conformeros del estado nativo [Tsai 1999]. El grado de la diversidad conformacional de una proteína puede entonces relacionarse con la extensión de la “rugosidad del fondo del embudo” la que se ha relacionado con el tipo de plegamiento [Keskin 2000], la historia evolutiva [Maguid 2006], y la presencia de ciertas mutaciones [Sinha 2001]. Estos conformeros en equilibrio muestran afinidades diferenciales por ligandos como el sustrato o moduladores alostéricos. La función biológica no reside únicamente en el conformero mas afín por el sustrato, sino en el conjunto de conformeros. Actualmente, los conceptos antes mencionados han derivado en la noción de la preexistencia de poblaciones de conformeros en equilibrio dinámico (dinamismo) sobre la superficie de energía potencial de una proteína [Frauenfelder 1991, Kumar 2000]. Este equilibrio dinámico soporta la hipótesis de la unión del ligando a una conformación específica de mayor afinidad y ofrece una visión central que permite explicar la función biológica.



**Figura 1.1 “Fondo de potencial rugoso”.** El grado de la diversidad conformacional de una proteína puede entonces relacionarse con la extensión de la “rugosidad del fondo del embudo”. Si el potencial es liso, entonces las proteínas nativas deberían tener pequeñas fluctuaciones, pero si los potenciales son rugosos, entonces las proteínas nativas podrían fluctuar a diferentes conformaciones. Pequeños cambios en la energía podrían provocar grandes cambios en la estructura.

En trabajos previos, se ha detectado que los distintos conformeros poseen un patrón de sustitución específico [Juritz 2012] y a su vez que el dinamismo de una proteína se conserva en la evolución [Maguid 2008]. De esta forma nuestra, principal hipótesis es que los aminoácidos o red de

aminoácidos vinculados al dinamismo de las proteínas se conservan evolutivamente. Debido a la estrecha relación entre la dinámica de la proteína y su función, la detección de estas redes de residuos ayudarían a mejorar la definición de familias y/o subfamilias funcionales.

## 1.1 OBJETIVOS

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 ausencia de ligandos, en equilibrio dinámico sobre la superficie de energía libre de una proteína. Esto ofrece una visión central que permite explicar la función biológica. Si la función de una proteína se encuentra fuertemente relacionada a su dinamismo [Tokuriki 2009], sitios estructurales determinantes de estos movimientos pueden ser considerados como sitios funcionales.

En este contexto, el objetivo principal de la tesis es el desarrollo de métodos que permitan identificar y caracterizar residuos o redes de residuos dinámicamente importantes para el mantenimiento de la diversidad conformacional de una proteína, esperando poder considerar estos residuos como elementos asociados a la función biológica.

Bajo estas hipótesis planteadas, nuestros objetivos específicos son: desarrollar métodos, basados en la dinámica de los diferentes conformeros de una proteína, que permitan identificar residuos importantes para la conservación de su diversidad conformacional; combinar la información obtenida a partir de características estructurales y dinámicas de las proteínas con información relacionada con su conservación evolutiva y adaptar e implementar métodos que permitan establecer redes que conecten estos residuos en los distintos conformeros de una proteína. El proyecto representa un primer paso hacia una meta ulterior de desarrollar métodos de asignación funcional de proteínas que contemplen el requerimiento de la conservación evolutiva de su diversidad conformacional.

## 1.2 ABREVIATURAS UTILIZADAS

1D - Unidimensional

3D - Tridimensional

ANM - Modelo de red anisotrópica (*Anisotropic Network Model*)

ACP - Proteína Transportadora de Acilo (*Acyl-Carrier Protein*)

ARS - Área Relativa de accesibilidad al Solvente

CIB1 - Proteína 1 de Unión a Calcio e Integrina (*Calcium and integrin-binding protein 1*)

ENM - Modelos de redes elásticas (*Elastic Network Model*)

ENCoM - Modelos de redes de contactos elásticos (*Elastic Network Atom Contact Model*)

DSSP - Por sus siglas en inglés, *Dictionary of protein secondary structure*

Factores B - Factores de temperatura o de Debye-Waller

FAC - Miocardiopatía Amiloide Familiar (*Familial Amyloid Cardiomyopathy*)  
FAP - Polineuropatía Amiloidótica Familiar (*Familial Amyloidotic Polyneuropathy*)  
GNM - Modelo de red gaussiana (*Gaussian Network Model*)  
HSSP - Por sus siglas en inglés, *Homology-derived Secondary Structure of Proteins*  
IAC - Contacto Inter-atómico (*Inter-Atomic Contact*)  
lb - Hace referencia a con ligando (*ligand-bound*)  
lf - Hace referencia a sin ligando (*ligand-free*)  
MD - Dinámica molecular (*Molecular Dynamics*)  
NMA - Análisis de modos normales (*Normal Mode Analysis*)  
NCC – Número de Componentes Conectados (*Number of Connected Component*)  
PDB - *The Protein Data Bank*  
pdb - Por extensión, el código o el archivo de cada proteína en PDB  
PHPT1 - Fosfohistidina Fosfatasa 1 (*Phosphohistidine phosphatase 1*)  
RBP - Proteína de unión al retinol (*Retinol-binding protein*)  
RIN - Por sus siglas en inglés, *Residue Interaction Networks*  
RING 2.0 - Por sus siglas en inglés, *Residue Interaction Network Generator*  
RMN - Resonancia magnética nuclear  
RMSD - Raíz de la desviación cuadrática media (*Root mean square deviation*)  
RSA - Accesibilidad Relativa al Solvente  
RX - Rayos X  
SAS - Superficie de Accesibilidad al Solvente  
SS - Separación de Secuencia (*Sequence Separation*)  
SSA - Amiloidosis Senil (*Senile Systemic Amyloidosis*)  
SSE - Elemento de estructura secundaria (*Secondary Structure Element*)  
T4 - Hormona Tiroxina  
TTR - Transtiretina Humana





# *Capítulo 2*

## *Análisis de Modos*

### *Normales: fundamentos*

*—Adiós —dijo el zorro—. He aquí mi secreto,  
que no puede ser más simple: sólo con el corazón  
se puede ver bien; lo esencial es invisible para los ojos.  
—Lo esencial es invisible para los ojos - repitió  
el principito para acordarse.*

El Principito, 1943.

Antoine Marie Jean-Baptiste Roger Conde de Saint-Exupéry

## 2. ANÁLISIS DE MODOS NORMALES: FUNDAMENTOS

Los movimientos de las proteínas -en condiciones cercanas a las fisiológicas- involucran un conjunto de conformaciones confinadas al entorno del plegamiento del estado nativo. Los cambios conformacionales que las conectan son en general alcanzados mediante desplazamientos colectivos de grandes grupos de residuos o dominios estructurales. Estos desplazamientos involucran movimientos concertados a baja escala relativa de frecuencias dentro de la proteína. Su relevancia funcional ha sido demostrada en numerosos trabajos anteriores [Brooks 1983, Seno 1990, Case 1994, Janezic 1995a, Janezic 1995b, Jernigan 1999, Tama 2001, Isin 2002].

La dinámica vibracional en torno al plegamiento nativo de una proteína puede ser descripta, en buena aproximación, por modelos lineales como el análisis de modos normales (NMA). Esta idea surgió inicialmente hace más de 40 años [McCammon 1976]. En el estudio de un movimiento tipo "*bending*" de la lisozima, se encontró que el potencial era aproximadamente parabólico, y se trató el movimiento relativo de dos dominios como un oscilador armónico angular. La frecuencia obtenida para la oscilación coincidió con la frecuencia del primer modo normal calculada años después [Brooks 1985]. Desde entonces el NMA se ha utilizado para explorar la dinámica conformacional de las proteínas [Go 1983, Brooks 1983] y ha demostrado ser un método físicamente plausible y eficiente como herramienta matemática y computacional. El interés en el uso de NMA se renovó en los últimos 17 años, principalmente debido a la introducción de modelos muy simples, basados en la mecánica de redes de polímeros [Tirion 1996, Hinsen 1999a, Tama 2001, Atilgan 2001, Li 2002].

### 2.1 MODOS NORMALES DE VIBRACIÓN

#### 2.1.1 Ecuaciones del movimiento vibracional

Si consideramos una molécula poliatómica formada por  $N$  átomos que vibran en torno a cierta posición de equilibrio para describir sus movimientos vibracionales conviene utilizar un sistema de coordenadas fijo en la molécula, esto es que se traslada y gira con ella. Sean  $x_\alpha$ ,  $y_\alpha$  y  $z_\alpha$  las coordenadas cartesianas del átomo  $\alpha$  respecto al sistema fijo en la molécula y sean  $x_\alpha^0$ ,  $y_\alpha^0$  y  $z_\alpha^0$  los valores en el punto de equilibrio. La energía cinética vibracional viene dada por la expresión

$$T = \frac{1}{2} \sum_{\alpha=1}^N m_\alpha \left[ \left( \frac{dx_\alpha}{dt} \right)^2 + \left( \frac{dy_\alpha}{dt} \right)^2 + \left( \frac{dz_\alpha}{dt} \right)^2 \right] \quad (\text{Ec. 2.1})$$

En este punto, conviene introducir las denominadas coordenadas cartesianas ponderadas en masa definidas como

$$\begin{aligned}
q_1 &= \sqrt{m_1}(x_1 - x_1^0), & q_2 &= \sqrt{m_1}(y_1 - y_1^0), & q_3 &= \sqrt{m_1}(z_1 - y_1^0) \\
q_4 &= \sqrt{m_2}(x_2 - x_2^0), & q_5 &= \sqrt{m_2}(y_2 - y_2^0), & q_6 &= \sqrt{m_2}(z_2 - y_2^0) \\
&\dots \\
q_{3N-2} &= \sqrt{m_N}(x_N - x_N^0), & q_{3N-1} &= \sqrt{m_N}(y_N - y_N^0), & q_{3N} &= \sqrt{m_N}(z_N - y_N^0)
\end{aligned} \tag{Ec. 2.2}$$

Derivando estas expresiones con respecto al tiempo y sustituyendo los resultados en la Ec. 2.1 la energía cinética puede escribirse como

$$T = \frac{1}{2} \sum_{i=1}^{3N} \left( \frac{dq_i}{dt} \right)^2 \tag{Ec. 2.3}$$

Los movimientos vibracionales de la molécula están controlados por su superficie de energía potencial  $V$ . Esta superficie de potencial depende, en general, de las  $3N$  coordenadas cartesianas o de las  $3N$  coordenadas de desplazamiento ponderadas  $q_i$ , es decir

$$V = V(q_1, \dots, q_{3N}) \tag{Ec.2.4}$$

Entonces podemos expresar la energía potencial como un desarrollo en serie de Taylor de las coordenadas  $q_i$  en torno a sus valores de equilibrio, que de acuerdo con las definiciones de la Ec. 2.2 todos son nulos ( $q_i^0 = 0$ ). Escribimos así

$$V = V|_0 + \sum_{i=1}^{3N} \left( \frac{\partial V}{\partial q_i} \right) \Big|_0 q_i + \frac{1}{2!} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right) \Big|_0 q_i q_j + \frac{1}{3!} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \sum_{k=1}^{3N} \left( \frac{\partial^3 V}{\partial q_i \partial q_j \partial q_k} \right) \Big|_0 q_i q_j q_k + \dots \tag{Ec. 2.5}$$

El subíndice 0 hace referencia a la conformación de equilibrio, que corresponde a un mínimo en la energía potencial, por lo tanto

$$\left( \frac{\partial V}{\partial q_i} \right) \Big|_0 = 0 \tag{Ec. 2.6}$$

Además, como el potencial está definido a menos de una constante, se puede definir nulo para la conformación de equilibrio de referencia, con lo cual

$$V|_0 = 0 \tag{Ec. 2.7}$$

Finalmente, asumiendo que los desplazamientos del sistema son suficientemente pequeños, podemos despreciar los términos de tercer orden y superiores en el desarrollo de Taylor, reduciendo el potencial a la forma cuadrática en las fluctuaciones, llamada aproximación armónica del potencial:

$$V = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right) \Big|_0 q_i q_j \tag{Ec. 2.8}$$

si definimos

$$k_{ij} = \left( \frac{\partial^2 V}{\partial q_i \partial q_j} \right) \Big|_0 \tag{Ec. 2.9}$$

donde  $k_{ij} = k_{ji}$ , nos queda

$$V = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} k_{ij} q_i q_j \quad (\text{Ec. 2.10})$$

Las constantes  $k_{ij}$  son las denominadas constantes de fuerza armónica para las coordenadas de desplazamiento ponderadas.

Para averiguar cómo cambian las coordenadas vibracionales con el tiempo, y cómo vibra por tanto la molécula, hemos de resolver las ecuaciones del movimiento vibracionales. De acuerdo con la segunda ley de Newton estas ecuaciones vienen dadas por las expresiones

$$F_{x,\alpha} = m_\alpha \frac{d^2 x_\alpha}{dt^2} \quad F_{y,\alpha} = m_\alpha \frac{d^2 y_\alpha}{dt^2} \quad F_{z,\alpha} = m_\alpha \frac{d^2 z_\alpha}{dt^2} \quad (\text{Ec. 2.11})$$

donde  $F_{x,\alpha}$ ,  $F_{y,\alpha}$  y  $F_{z,\alpha}$  son las componentes cartesianas de la fuerza que actúa sobre el átomo  $\alpha$ . La energía potencial esta relacionada con las fuerzas mediante la expresión general  $F_i = -\partial V / \partial x_i$ , así que podemos escribir las ecuaciones del movimiento en función de la energía potencial de la siguiente manera

$$\begin{aligned} m_\alpha \frac{d^2 x_\alpha}{dt^2} &= -\frac{\partial V}{\partial x_\alpha} \\ m_\alpha \frac{d^2 y_\alpha}{dt^2} &= -\frac{\partial V}{\partial y_\alpha} \\ m_\alpha \frac{d^2 z_\alpha}{dt^2} &= -\frac{\partial V}{\partial z_\alpha} \end{aligned} \quad (\text{Ec. 2.12})$$

Cambiamos ahora en estas ecuaciones las coordenadas cartesianas por las coordenadas de desplazamiento ponderadas. Tomamos para ello la ecuación del movimiento de la coordenada  $x_1$ , que viene dada por

$$m_1 \frac{d^2 x_1}{dt^2} = -\frac{\partial V}{\partial x_1} \quad (\text{Ec. 2.13})$$

Derivando dos veces con respecto al tiempo la ecuación de transformación  $q_1 = \sqrt{m_1}(x_1 - x_1^0)$  obtenemos

$$\frac{d^2 q_1}{dt^2} = \sqrt{m_1} \frac{d^2 x_1}{dt^2} \quad (\text{Ec. 2.14})$$

y usando la regla de la cadena en la derivada del potencial nos queda

$$\frac{\partial V}{\partial x_1} = \frac{\partial V}{\partial q_1} \frac{dq_1}{dx_1} = \sqrt{m_1} \frac{\partial V}{\partial q_1} \quad (\text{Ec. 2.15})$$

Usando estas dos ultimas ecuaciones en la Ec. 2.13 obtenemos

$$\frac{d^2 q_1}{dt^2} = -\frac{\partial V}{\partial q_1} \quad (\text{Ec. 2.16})$$

Las ecuaciones de movimiento restantes se transforman de un modo análogo y podemos escribirlas todas de forma compacta como

$$\frac{d^2 q_u}{dt^2} + \frac{\partial V}{\partial q_u} = 0 \quad u = 1, \dots, 3N \quad (\text{Ec. 2.17})$$

Ahora debemos sustituir en la Ec 2.17, la expresión obtenida para la energía potencial  $V$  (Ec. 2.10).

Pero primero desarrollamos la derivada de  $V$  con respecto a  $q_u$

$$\begin{aligned} \frac{\partial V}{\partial q_u} &= \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} k_{ij} \left[ \frac{\partial q_i}{\partial q_u} q_j + q_i \frac{\partial q_j}{\partial q_u} \right] = \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} k_{ij} [\delta_{iu} q_j + \delta_{ju} q_i] \\ &= \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} k_{ij} \delta_{iu} q_j + \frac{1}{2} \sum_{i=1}^{3N} \sum_{j=1}^{3N} k_{ij} \delta_{ju} q_i = \frac{1}{2} \sum_{j=1}^{3N} k_{uj} q_j + \frac{1}{2} \sum_{i=1}^{3N} k_{iu} q_i \\ &= \frac{1}{2} \sum_{j=1}^{3N} k_{uj} q_j + \frac{1}{2} \sum_{j=1}^{3N} k_{ju} q_j = \sum_{j=1}^{3N} k_{uj} q_j \end{aligned} \quad (\text{Ec. 2.18})$$

donde hemos tenido en cuenta que las coordenadas  $q_i$  son independientes entre sí. Sustituyendo esta expresión en la Ec. 2.17 obtenemos finalmente

$$\frac{d^2 q_u}{dt^2} + \sum_{j=1}^{3N} k_{uj} q_j = 0 \quad u = 1, \dots, 3N \quad (\text{Ec. 2.19})$$

## 2.1.2 Solución de las ecuaciones de movimiento

La Ec. 2.19 representa un sistema de ecuaciones diferenciales acopladas cuyas incógnitas son las coordenadas de desplazamiento ponderadas  $q_u$ . Cada una de estas ecuaciones diferenciales contiene a todas las funciones  $q_u$ , es decir la derivada segunda de cada coordenada  $q_u$  depende de las variaciones de todas las coordenadas con el tiempo. Estos acoplamientos complican la resolución del sistema de ecuaciones diferenciales y podemos tratar de eliminarlos realizando un cambio de variables adecuado. Llamemos  $Q_i$  a las nuevas variables que buscamos, y supongamos que las coordenadas de desplazamiento ponderadas  $q_u$  estén relacionadas con ellas mediante las combinaciones lineales

$$q_u = \sum_{i=1}^{3N} l_{ui} Q_i \quad (\text{Ec. 2.20})$$

donde  $l_{ui}$  son coeficientes que especificaremos mas adelante. En notación matricial estas ecuaciones de transformación se expresan de la forma

$$q = LQ \quad (\text{Ec. 2.21})$$

siendo  $q$  el vector columna que contiene las coordenadas de desplazamiento ponderadas  $q_u$ ,  $Q$  es el vector columna que contiene las nuevas coordenadas  $Q_i$  y  $L$  es la matriz de coeficientes  $l_{ui}$ . La Ec. 2.8 para la función de energía potencial en notación matricial viene dada, a su vez por

$$V = \frac{1}{2} \tilde{q} U q \quad (\text{Ec. 2.22})$$

donde  $\tilde{q}$  es el vector fila que se forma al trasponer el vector columna  $q$  y  $U$  es la matriz cuadrada

$$U = \begin{bmatrix} \frac{\partial^2 V}{\partial q_1^2} & \frac{\partial^2 V}{\partial q_1 \partial q_2} & \dots & \frac{\partial^2 V}{\partial q_1 \partial q_N} \\ \frac{\partial^2 V}{\partial q_2 \partial q_1} & \frac{\partial^2 V}{\partial q_2^2} & \dots & \frac{\partial^2 V}{\partial q_2 \partial q_N} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 V}{\partial q_N \partial q_1} & \frac{\partial^2 V}{\partial q_N \partial q_2} & \dots & \frac{\partial^2 V}{\partial q_N^2} \end{bmatrix} \quad (\text{Ec. 2.23})$$

Combinando las Ec. 2.21 y Ec. 2.22 obtenemos la expresión

$$V = \frac{1}{2} \tilde{q} U q = \frac{1}{2} (\widetilde{LQ}) U L Q = \frac{1}{2} \tilde{Q} \tilde{L} U L Q \quad (\text{Ec. 2.24})$$

donde usamos la propiedad  $(\widetilde{LQ}) = \tilde{Q} \tilde{L}$ .

Debemos ahora encontrar la matriz de coeficientes de transformación  $L$ , en álgebra lineal a esta se la define como la matriz de autovectores de  $U$  y se obtiene a partir de la diagonalización de la misma. La ecuación matricial de valores propios de  $U$  esta dada por

$$U L = L \Lambda \quad (\text{Ec. 2.25})$$

donde  $\Lambda$  es la matriz diagonal que contiene los autovalores  $\lambda_i$ . Multiplicando por la izquierda por la matriz inversa de  $L$  obtenemos

$$L^{-1} U L = L^{-1} L \Lambda = I \Lambda = \Lambda \quad (\text{Ec. 2.26})$$

$I$  es la matriz unidad. Puesto que la matriz  $U$  es una matriz real y simétrica, su matriz de autovectores es ortonormal. Una matriz ortonormal es aquellas cuya inversa es igual a su traspuesta. En nuestro caso por tanto tenemos que  $L^{-1} = \tilde{L}$ , y podemos escribir la Ec. 2.26 de la forma

$$\tilde{L} U L = \Lambda \quad (\text{Ec. 2.27})$$

Usando esta expresión en la Ec. 2.24 nos queda

$$V = \frac{1}{2} \tilde{Q} \Lambda Q \quad (\text{Ec. 2.28})$$

y desarrollando los productos matriciales, teniendo en cuenta que la matriz  $\Lambda$  es diagonal, obtenemos

$$V = \frac{1}{2} \sum_{i=1}^{3N} \lambda_i Q_i^2 \quad (\text{Ec. 2.29})$$

Conseguimos así escribir la función de energía potencial como una sumatoria de las nuevas variables  $Q_i$ .

En este punto podemos definir que la matriz  $L$  es la matriz de autovectores que permite el cambio de coordenadas, relacionándolas mediante los coeficientes  $l_{ui}$ , Ec. 2.20 o

$$Q_i = \sum_{u=1}^{3N} l_{ui} q_u \quad (\text{Ec. 2.30})$$

Ahora debemos reescribir las ecuaciones de movimiento (Ec.2.19) respecto a las nuevas coordenadas  $Q_i$ . Para esto primero derivamos dos veces respecto al tiempo la Ec. 2.20, obteniendo

$$\frac{d^2 q_u}{dt^2} = \sum_{i=1}^{3N} l_{ui} \frac{d^2 Q_i}{dt^2} \quad (\text{Ec. 2.31})$$

y usamos, en segundo lugar, la regla de la cadena para desarrollar las derivadas de la energía potencial como

$$\frac{\partial V}{\partial q_u} = \sum_{i=1}^{3N} \frac{\partial V}{\partial Q_i} \frac{\partial Q_i}{\partial q_u} = \sum_{i=1}^{3N} \frac{\partial V}{\partial Q_i} l_{ui} \quad (\text{Ec. 2.32})$$

donde hemos utilizado las Ec. 2.30 para determinar las derivadas  $\frac{\partial Q_i}{\partial q_u}$ . Sustituyendo las Ec. 2.31 y Ec. 2.32 en las ecuaciones del movimiento (Ec. 2.19) y sacando como factor común los coeficientes  $l_{ui}$ , nos queda

$$\sum_{i=1}^{3N} l_{ui} \left( \frac{d^2 Q_i}{dt^2} + \frac{\partial V}{\partial Q_i} \right) = 0 \quad u = 1, \dots, 3N \quad (\text{Ec. 2.33})$$

Puesto que los coeficientes  $l_{ui}$  son, en general, diferentes de cero, esta sumatoria sólo se anula cuando lo hacen los factores que multiplican a dichos coeficientes, es decir, cuando se cumple que

$$\frac{d^2 Q_i}{dt^2} + \frac{\partial V}{\partial Q_i} = 0 \quad i = 1, \dots, 3N \quad (\text{Ec. 2.34})$$

Estas son las ecuaciones del movimiento que buscamos en función de las coordenadas  $Q_i$ . Derivando la energía potencial dada por la ecuación Ec. 2.29 respecto a  $Q_i$  obtenemos

$$\frac{\partial V}{\partial Q_i} = \sum_{k=1}^{3N} \lambda_k Q_k \frac{\partial Q_k}{\partial Q_i} = \sum_{k=1}^{3N} \lambda_k Q_k \delta_{ki} = \lambda_i Q_i \quad (\text{Ec. 2.35})$$

y sustituyendo este resultado en las ecuaciones del movimiento (Ec. 2.34) nos queda finalmente

$$\frac{d^2 Q_i}{dt^2} + \lambda_i Q_i = 0 \quad i = 1, \dots, 3N \quad (\text{Ec. 2.36})$$

Estas ecuaciones están completamente desacopladas, es decir la derivada segunda de cada variable  $Q_i$  con respecto al tiempo depende únicamente de ella misma.

Las coordenadas  $Q_i$  introducidas para desacoplar las ecuaciones vibratorias del movimiento de la molécula se denominan coordenadas normales de vibración. Las ecuaciones del movimiento (Ec. 2.36) para estas coordenadas son similares a las de un oscilador armónico unidimensional de masa reducida y constante de fuerza  $\lambda_i$ . Las soluciones clásicas de la Ec. 2.36 pueden escribirse de la forma

$$Q_i(t) = B_i \text{sen}(\lambda_i^{1/2} t + b_i) \quad (\text{Ec. 2.37})$$

donde  $B_i$  y  $b_i$  son las amplitudes y fases, que dependen de las condiciones iniciales. Una vez obtenidas las coordenadas normales  $Q_i$  podemos determinar las coordenadas de desplazamiento ponderadas  $q_u$  sustituyendo la Ec. 2.37 en la Ec. 2.20, quedándonos

$$q_u = \sum_{i=1}^{3N} A_{ui} \text{sen}(\lambda_i^{1/2} t + b_i) \quad u = 1, \dots, 3N \quad (\text{Ec. 2.38})$$



donde  $A_{ui} = l_{ui} B_i$ . Estas son las soluciones generales de las ecuaciones del movimiento vibracional para las coordenadas de desplazamiento ponderadas. Puesto que dichas ecuaciones forman un sistema de  $3N$  ecuaciones diferenciales de segundo orden, sus soluciones generales admiten un total de  $6N$  constantes arbitrarias a especificar, que son las  $3N$  amplitudes  $B_i$  y los  $3N$  factores de fase  $b_i$ . Los valores de estas constantes se calculan fijando las condiciones iniciales del sistema, es decir, a partir de los  $3N$  valores iniciales (a tiempo  $t = 0$ ) de las coordenadas  $q_u$  y los  $3N$  valores iniciales de las correspondientes velocidades.

Si suponemos que las condiciones iniciales son tales que todas las amplitudes  $B_i$  son iguales a cero salvo una, la  $B_m$ . La única constante  $A_{ui}$  que no se anula en la Ec. 2.38 es la  $A_{um}$ , de modo que las soluciones para las coordenadas  $q_u(t)$  se reducen a

$$q_u = A_{um} \text{sen}(\lambda_m^{1/2} t + b_m) \quad (\text{Ec. 2.39})$$

En este caso todos los átomos vibran en fase con la misma frecuencia  $\nu_m$  asociada a la constante de fuerza normal  $\lambda_m$ , es decir

$$\nu_m = \frac{\sqrt{\lambda_m}}{2\pi} \quad (\text{Ec. 2.40})$$

Usando la Ec. 2.37, estas soluciones particulares pueden expresarse como

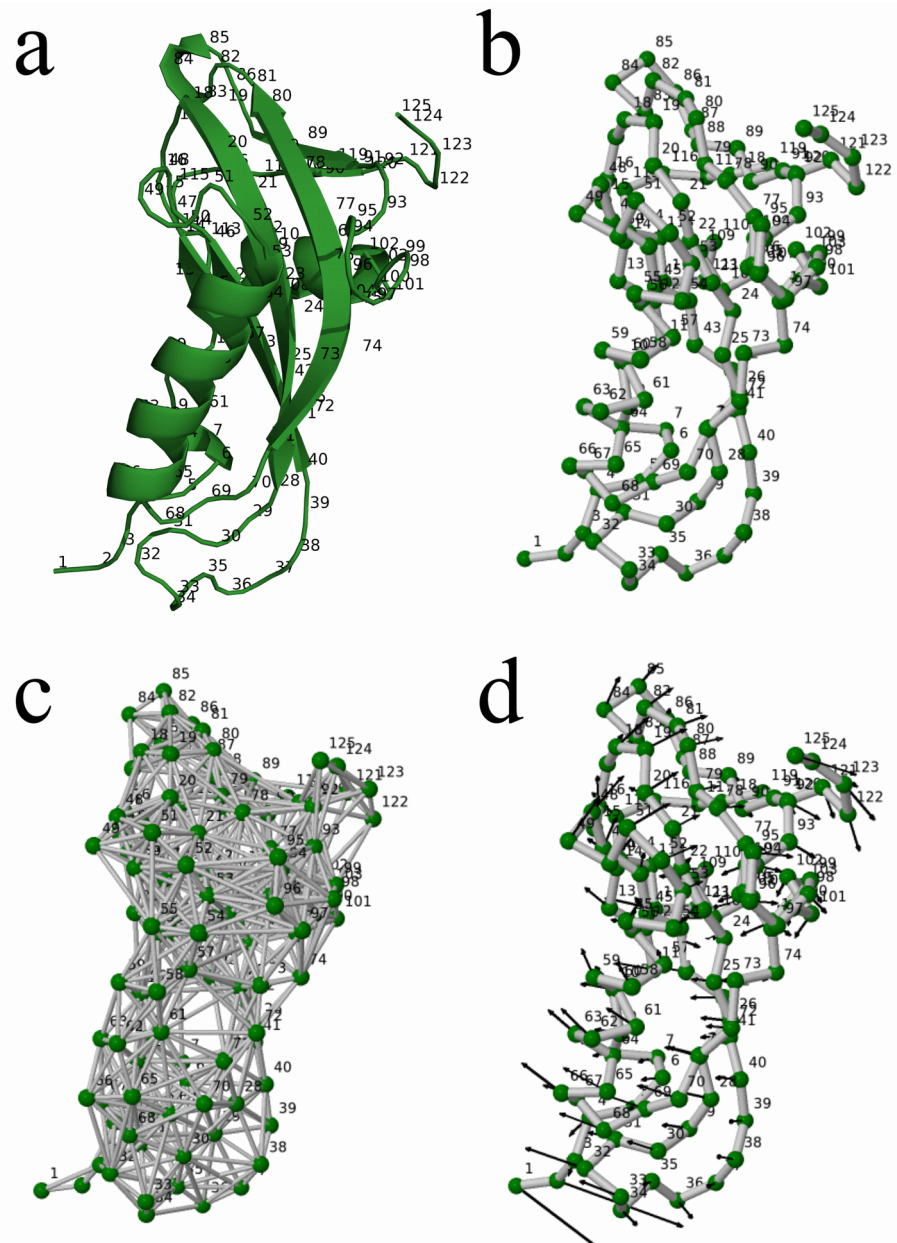
$$q_u = l_{um} Q_m(t) \quad u = 1, \dots, 3N \quad (\text{Ec. 2.41})$$

Las vibraciones de este tipo son los denominados modos normales de vibración.

## 2.2 MODELO DE RED ANISOTRÓPICA: FUNDAMENTOS

En esencia, hay dos tipos diferentes de modelos de redes elásticas (ENM, por sus siglas en inglés *Elastic Network Model*), que difieren por su dimensionalidad. El Modelo de Red Gaussiana (GNM, por sus siglas en inglés *Gaussian Network Model*), propuesto por Ivet Bahar, Burak Erman y Turkan Haliloglu en 1997 [Bahar 1997a, Haliloglu 1997], es un modelo unidimensional. Mientras que el modelo de Tirion, más tarde llamado Modelo de Red Anisotrópica (ANM, por sus siglas en inglés *Anisotropic Network Model*) [Atilgan 2001], es tridimensional.

ANM es una de las variantes de bajo costo computacional a los cálculos de NMA que considera a la proteína como una red de partículas (o nodos) interconectadas por resortes armónicos (ver **Figura 2.1**). Estos últimos están afectados por una única constante de fuerza  $\gamma$ , la cual determina un potencial armónico universal para el modelo [Tirion 1996]. El número de grados de libertad del sistema se reduce al considerar solo los  $C\alpha$  provenientes de la estructura cristalográfica obtenida de la *Protein Data Bank* (PDB). Es decir, se tiene en cuenta un solo nodo por residuo. 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 de la matriz Hessiana. De este modo se reduce muchísimo el costo computacional de los cálculos.



**Figura 2.1** a) Estructura sin ligando de la proteína Fosfohistidina Fosfatasa 1 (PHPT1, código pdb 2AI6). b) Posiciones de los  $C\alpha$  de la cadena principal. c) Modelo de red elástica que considera a todos los  $C\alpha$  dentro de un radio determinado como nodos interactuantes. d) Dirección de desplazamiento de los  $C\alpha$  correspondiente al primer modo normal derivado del Modelo de Red Anisotrópica (ANM).

Para determinar si dos  $C\alpha$  deben estar unidos por un resorte, se fija un radio de corte  $r_c$  (ver

**Figura 2.2**) 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 2.2**) y  $k_{ij}$

es la constante del resorte cuyo valor depende de la constante de fuerza  $\gamma$  y del radio de corte  $r_c$ .

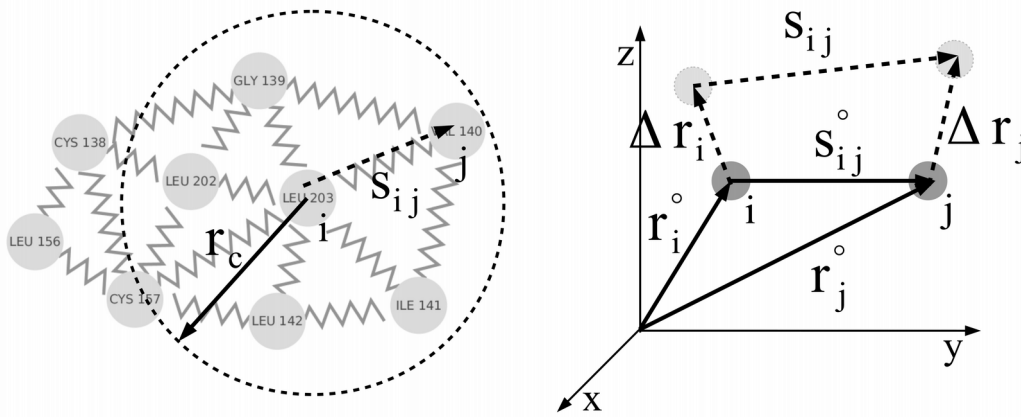
Según este planteo, entonces, se contemplan tanto uniones entre dos residuos adyacentes como entre

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 (**Figura 2.2**), el potencial elástico armónico al cual están sujetos será:

$$V = \frac{1}{2} k_{ij} (s_{ij} - s_{ij}^0)^2 = \frac{1}{2} k_{ij} \left( \left[ (x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2 \right]^{\frac{1}{2}} - s_{ij}^0 \right)^2 \quad (\text{Ec. 2.42})$$

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 2.2**).



**Figura 2.2** 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  $\Delta r_i$ ,  $\Delta r_j$  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:

$$\frac{\partial V}{\partial x_i} = -\frac{\partial V}{\partial x_j} = -k_{ij} (x_j - x_i) \left( 1 - \frac{s_{ij}^0}{s_{ij}} \right) \quad (\text{Ec. 2.43})$$

$$\frac{\partial^2 V}{\partial x_i^2} = \frac{\partial^2 V}{\partial x_j^2} = k_{ij} \left( 1 + s_{ij}^0 \frac{(x_i - x_j)^2}{s_{ij}^3} - \frac{s_{ij}^0}{s_{ij}} \right) \quad (\text{Ec. 2.44})$$

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

Para el caso en que  $s_{ij} = s_{ij}^0$  las Ec. 2.43 y 2.44 se reducen a:

$$\frac{\partial V}{\partial x_i} = 0 \quad (\text{Ec. 2.45})$$

$$\frac{\partial^2 V}{\partial x_i^2} = k_{ij} \frac{(x_j - x_i)^2}{s_{ij}^2} \quad (\text{Ec. 2.46})$$

y las segundas derivadas cruzadas:

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{\partial^2 V}{\partial x_j \partial y_i} = -k_{ij} \frac{(x_j - x_i)(y_j - y_i)}{s_{ij}^2} \quad (\text{Ec. 2.47})$$

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.

$$H = \begin{bmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \cdots & \mathbf{H}_{1N} \\ \mathbf{H}_{21} & & & \mathbf{H}_{2N} \\ \vdots & & & \vdots \\ \mathbf{H}_{N1} & \mathbf{H}_{12} & \cdots & \mathbf{H}_{NN} \end{bmatrix} \quad (\text{Ec. 2.48})$$

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

$$\mathbf{H}_{ij} = \begin{bmatrix} \partial^2 V / \partial x_i \partial x_j & \partial^2 V / \partial x_i \partial y_j & \partial^2 V / \partial x_i \partial z_j \\ \partial^2 V / \partial y_i \partial x_j & \partial^2 V / \partial y_i \partial y_j & \partial^2 V / \partial y_i \partial z_j \\ \partial^2 V / \partial z_i \partial x_j & \partial^2 V / \partial z_i \partial y_j & \partial^2 V / \partial z_i \partial z_j \end{bmatrix} \quad (\text{Ec. 2.49})$$

Para el caso de los superelementos  $\mathbf{H}_{ii}$ , las ecuaciones Ec. 2.46 y 2.47 pasan a ser,

$$\frac{\partial^2 V}{\partial x_i^2} = k_{ij} \sum_j \frac{(x_j - x_i)^2}{s_{ij}^2} \quad (\text{Ec. 2.50})$$

$$\frac{\partial^2 V}{\partial x_i \partial y_j} = \frac{\partial^2 V}{\partial x_j \partial y_i} = -k_{ij} \sum_j \frac{(x_j - x_i)(y_j - y_i)}{s_{ij}^2} \quad (\text{Ec. 2.51})$$

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,  $\Lambda$  es la matriz diagonal de autovalores  $\lambda_k$  y es  $\tilde{Q}$  la traspuesta de Q. Los autovalores  $\lambda_k$  representan los cuadrados de las frecuencias  $\omega_k$  de los modos individuales,

$$\lambda_k = \omega_k^2 \quad (\text{Ec. 2.52})$$

Y podemos definir a los autovectores  $Q_k$  como

$$Q_k = A_k \cos(\omega_k t) \quad (\text{Ec. 2.53})$$

siendo  $A_k$  la amplitud del autovector, asociado a la frecuencia  $\omega_k$ .

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 una rotación en un cuerpo rígido. O sea, existen 6 direcciones de movimiento (de las  $3N$ ) que corresponden a características de cuerpo rígido exclusivamente, y no dan información de las vibraciones internas.

Los modos son vectores ortonormales, por lo tanto el conjunto de los  $3N$  modos normales 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 cooperativos o globales que

involucran el desplazamiento de grandes bloques o dominios de las proteínas [Bahar 1998, Bahar 1999a]. 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 [Bahar 1998, Jernigan 1999, Bahar 1999a, Keskin 2000, Temiz 2002, Isin 2002, Keskin 2002a, Keskin 2002b, Wang 2004, Xu 2003]. Trabajos anteriores evidencian que el ANM reproduce adecuadamente las amplitudes de movimiento relativo de los residuos descritas por los modos de baja frecuencia [Bahar 1999a, Hinsen 1999a, Micheletti 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 [Zheng 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. [Bahar 1998].

### 2.2.1 Adaptación del modelo: constante de fuerza asociada al tipo de interacción

Como se mencionó en la Sección 2.2 el método de ANM utiliza una única constante de fuerza  $\gamma$  para todas las interacciones. Sin embargo, en ésta tesis se ha modificado parcialmente el método de manera de poder obtener una visión biológica más adecuada que contemple la información de tipos de interacciones entre residuos. Esta información es incorporada en el método de ANM mediante constantes de fuerza diferenciales. Esto es, se utilizaron constantes de fuerza entre residuos que representaran aproximadamente el orden de magnitud relativo a cada tipo de unión, tomando como referencia la unión covalente. Se aplicaron las siguientes reglas para determinar el valor de la constante [Saldaño 2016],

si  $|i-j| = 1 \Rightarrow k_{ij} = \gamma$

de otra manera

si  $|r_j^0 - r_i^0| \leq r_c$  entonces

si  $i$  e  $j$  están conectados mediante enlace peptídico o puente disulfuro  $\Rightarrow k_{ij} = \gamma$

si  $i$  e  $j$  interactúan mediante puente de hidrógeno o puente salino  $\Rightarrow k_{ij} = \gamma \times 0.1$

sino  $\Rightarrow k_{ij} = \gamma \times 0.01$

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

Nuevamente vemos que las constantes del resorte  $k_{ij}$  del modelo depende de la constante de fuerza  $\gamma$ . Determinamos el valor de  $\gamma$  de manera tal que maximice el ajuste entre los perfiles de los factores B teóricos y los factores B experimentales.

Para determinar las interacciones entre residuos se utilizó, en una primera instancia una adaptación del software CSU [Sobolev 1999] y luego se mejoró la metodología al utilizar el programa RING 2.0 (ver Capítulo 3, sección 3.3.7) [Martin 2011, Piovesan 2016].

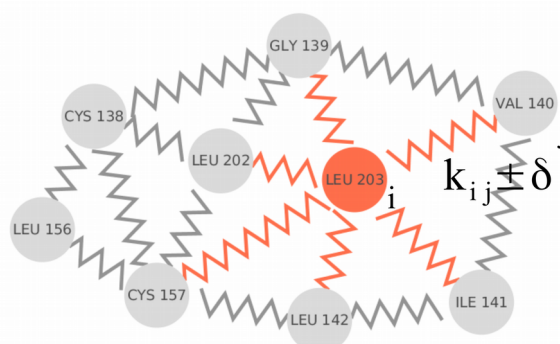
Para el estudio de la dinámica de equilibrio de la proteína Transtiretina, adaptamos el modelo de ANM de manera que para cada residuo se considere como nodo los átomos de la cadena principal, el C $\beta$  y un centro de masa de la cadena lateral.

## 2.2.2 Adaptación del modelo: simulación de mutación puntual

Para la simulación de una mutación puntual se desarrolló una variante del método de ANM. Las mutaciones fueron simuladas mediante perturbaciones locales del valor de la constante del resorte utilizada en el cálculo de modos normales [Zheng 2005, Zheng 2006, Zheng 2009a, Zheng 2009b Saldaño 2016].

La mutación puntual un residuo  $i$  se simula mediante la perturbaciones a las interacciones en las que participa este residuo. Esto se realizó cambiando las constantes del resorte  $k_{ix}$  que conectan  $i$  con otros residuos ( $x$ ) por una pequeña cantidad  $\delta$ . De esta forma la constante para todas las interacciones del residuo  $i$  sería  $k_{ix} \pm \delta$  (**Figura 2.3**).

Luego de la perturbación y con la nuevas constantes del resorte calculamos nuevamente los modos normales del ANM.



**Figura 2.3** Representación esquemática de los nodos en el ANM, cada nodo se conecta a sus vecinos mediante resortes. En este caso se muestran en rojo los resortes del nodo  $i$  con cinco nodos, entre ellos el nodo  $j$ . La mutación puntual un residuo  $i$  se simula mediante la introducción de perturbaciones a las interacciones en las que participa este residuo. Por ejemplo, la constante del resorte para la interacción entre los residuos  $i, j$  sería  $k_{ij} \pm \delta$ .

## 2.3 MODELO DE RED ANISOTRÓPICA: APLICACIONES

### 2.3.1 Factores B de temperatura

Las fluctuaciones cuadráticas medias  $\langle \Delta r_i^2 \rangle$  de los residuos se pueden determinar experimentalmente mediante los factores B de temperatura asociados a la determinación cristalográfica por RX o por diferencias cuadráticas medias entre los diferentes modelos de RMN. Como tales, suelen ser utilizadas para verificar la validez del modelo ANM en la proteína de estudio y optimizar los modelos y métodos computacionales. A partir de la publicación original del Bahar et al. [Bahar 1997a], varias aplicaciones demostraron que las fluctuaciones predichas mediante ENM, en

especial ANM, concuerdan con los factores B experimentales [Bahar 1997b, Bahar 1999c, Keskin 2000, Atilgan 2001, Kundu 2002].

Los factores B experimentales informados en las estructuras del PDB brindan información sobre la movilidad de los átomo. Representan la disminución de la intensidad en la difracción debido tanto al desorden dinámico causado por la vibración, dependiente de la temperatura, de los átomos como al desorden estático, que está relacionado con la orientación de la molécula. El valor del Factor B determinado experimentalmente no es una cantidad absoluta; en su lugar, depende de otros factores, como la resolución general de la estructura, los contactos de cristal y, lo que es más importante, los procedimientos de refinamiento específicos [Schlessinger 2005]. El factor B o también llamado factor cristalográfico de Debye-Waller del átomo  $i$ ,  $B_i$ , se espera que esté relacionado con las fluctuaciones de sus coordenadas atómicas a través de [Sanejouand 2013],

$$B_i = \frac{8\pi^2}{3} \langle \Delta r_i^2 \rangle \quad (\text{Ec. 2.54})$$

siendo  $\langle \Delta r_i^2 \rangle = \langle \Delta x_i^2 + \Delta y_i^2 + \Delta z_i^2 \rangle$ , el valor de las fluctuaciones cuadráticas media de cada residuo.

A partir del calculo de modos normales, se pueden definir los Factores B teóricos,  $B_i^{\text{teo}}$  como

$$B_i^{\text{teo}} = \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \quad (\text{Ec. 2.55})$$

Estos Factores  $B_i^{\text{teo}}$  deben ser escalados con el fin de compararlos con los Factores B experimentales,  $B_i^{\text{exp}}$

$$B_i^{\text{teo,escalado}} = B_i^{\text{teo}} \frac{\sum_i^N B_i^{\text{exp}}}{\sum_i^N B_i^{\text{teo}}} = \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \frac{\sum_i^N B_i^{\text{exp}}}{\sum_i^N B_i^{\text{teo}}} \quad (\text{Ec. 2.56})$$

donde  $B_i^{\text{teo}}$  es el factor B teórico del residuo  $i$  y  $B_i^{\text{exp}}$  es el factor B experimental del mismo residuo obtenido del archivo pdb. Conociendo el valor del Factor B escalado del residuo  $i$  ( $B_i^{\text{teo,escalado}}$ ) e usando la Ec. 2.54 podemos calcula su valor de fluctuaciones cuadráticas media como

$$\langle \Delta r_i^2 \rangle = \frac{3}{8\pi^2} B_i^{\text{teo,escalado}} \quad (\text{Ec. 2.57})$$

Para el cálculo de la correlaciones de movimiento entre dos residuos se tiene en cuenta que [Bahar 1997a, Tama 2000, Atilgan 2001],

$$\langle \Delta r_i \cdot \Delta r_j \rangle = \sum_k^{3N-6} [\Delta r_i \cdot \Delta r_j]_k = \frac{3k_B T}{\gamma} \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{jk}] \quad (\text{Ec. 2.58})$$

donde se expresan las correlaciones cruzadas entre las fluctuaciones de los residuos como una sumatoria sobre los modos no nulos.  $k_B$  es la constante de Boltzmann y  $T$  es la temperatura absoluta (300 Kelvin). Para el caso en que  $i = j$ , la Ec. 2.55 queda

$$\langle \Delta r_i^2 \rangle = \frac{3k_B T}{\gamma} \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \quad (\text{Ec. 2.59})$$

Reemplazando  $\langle \Delta r_i^2 \rangle$  de la Ec. 2.54 por la Ec. 2.59 obtenemos

$$B_i = \frac{8 \pi^2 k_B T}{\gamma} \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \quad (\text{Ec. 2.60})$$

De la Ec. 2.59 obtenemos la expresión que define a la constante de fuerza  $\gamma$

$$\gamma = \frac{3 k_B T}{\langle \Delta r_i^2 \rangle} \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \quad (\text{Ec. 2.61})$$

Si desarrollamos esta expresión, reemplazando  $\langle \Delta r_i^2 \rangle$  por el termino de la Ec. 2.57 llegamos a

$$\gamma = \frac{8 \pi^2 k_B T}{B_i^{\text{teo, escalado}}} \sum_k^{3N-6} [\lambda_k^{-1} Q_{ik} \widetilde{Q}_{ik}] \quad (\text{Ec. 2.62})$$

Luego cambiando  $B_i^{\text{teo, escalado}}$  por el termino de la Ec. 2.56 obtenemos

$$\gamma = 8 \pi^2 k_B T \frac{\sum_i^N B_i^{\text{teo}}}{\sum_i^N B_i^{\text{exp}}} \quad (\text{Ec. 2.63})$$

Siendo  $k_B T = 4,11 \times 10^{-21} \text{ J} = 4,11 \times 10^{-21} \frac{\text{Kg m}^2}{\text{S}^2} = 0,411 \frac{\text{Kg \AA}^2}{\text{S}^2}$ , la constante  $\gamma$  posee unidades de

$$\left[ \frac{\text{Kg}}{\text{S}^2} \right].$$

Una de las formas de estudiar si el modelo de ANM describe de forma correcta las fluctuaciones de los residuos es comparar los factores B experimentales ( $B_i^{\text{exp}}$ ) con los factores B teóricos ( $B_i^{\text{teo, escalado}}$ ) de la siguiente manera

$$\rho^{\text{teo-exp}} = \frac{\sum_{i=1}^N (B_i^{\text{teo, escalado}} - \langle B^{\text{teo, escalado}} \rangle) (B_i^{\text{exp}} - \langle B^{\text{exp}} \rangle)}{\sqrt{\sum_{i=1}^N (B_i^{\text{teo, escalado}} - \langle B^{\text{teo, escalado}} \rangle)^2} \sqrt{\sum_{i=1}^N (B_i^{\text{exp}} - \langle B^{\text{exp}} \rangle)^2}} \quad (\text{Ec. 2.64})$$

donde  $\langle B^{\text{teo, escalado}} \rangle$  y  $\langle B^{\text{exp}} \rangle$  son el factor B teórico promedio y el factor B experimental promedio respectivamente.

La determinación del radio de corte esta relacionada con los factores de temperatura ya que como se ve en la Ec. 2.62, el  $B_i^{\text{teo, escalado}}$  esta relacionado del valor de la constante de fuerza  $\gamma$  y a su vez  $\gamma$  depende del radio de corte  $r_c$  entre los residuos. De esta manera al variar el radio  $r_c$ , puede aumentar o disminuir la correlación entre los factores de temperatura.

### 2.3.2 Conjunto de estructuras que representan fluctuaciones térmicas

En la Sección 2.1.2, Ec. 2.29 conseguimos escribir la función de energía potencial como una



sumatoria de las coordenadas normales de vibración  $Q_k$ . A continuación, definimos la Energía Potencial promedio asociada a un modo normal  $\langle V_k \rangle$  como

$$\langle V_k \rangle = \frac{1}{2} \lambda_k \langle Q_k^2 \rangle \quad (\text{Ec. 2.65})$$

En la Sección 2.2, Ec. 2.53 definimos  $Q_k$  como  $Q_k = A_k \cos(\omega_k t)$ , por los que  $Q_k^2 = A_k^2 \cos^2(\omega_k t)$ . Si consideramos el promedio temporal del coseno cuadrado  $\langle \cos^2(\omega_k t) \rangle$  igual a  $\frac{1}{2}$ , entonces  $\langle Q_k^2 \rangle = \frac{A_k^2}{2}$  y reemplazando este termino en la Ec. 2.65 obtenemos

$$\langle V_k \rangle = \frac{1}{4} \lambda_k A_k^2 \quad (\text{Ec. 2.66})$$

El teorema de equipartición de la energía para un oscilador armónico establece que, en equilibrio, cada grado de libertad contribuye  $\frac{1}{2} k_B T$  a la energía promedio por molécula (Apéndice C) [Tipler 2008]. Por esto podemos expresar la Energía Potencial promedio de un oscilador armónico como

$$\langle V_k \rangle = \frac{1}{2} k_B T \quad (\text{Ec. 2.67})$$

donde  $k_B$  es la constante de Boltzmann y  $T$  es la temperatura absoluta. Combinando la Ec. 2.66 y Ec. 2.67 establecemos que

$$\frac{1}{4} \lambda_k A_k^2 = \frac{1}{2} k_B T \quad (\text{Ec. 2.68})$$

En la Sección 2.3.1 comentamos que  $\gamma$  (y  $r_c$ ) nos permite ajustar mejor a las fluctuaciones de los residuos. Considerando esto decidimos escalar los autovalores  $\lambda_k$

$$\lambda_k^{\text{escalado}} = \gamma \lambda_k \quad (\text{Ec. 2.69})$$

Finalmente, reemplazando  $\lambda_k$  en la Ec. 2.68 por  $\lambda_k^{\text{escalado}}$  y despejando  $A_k$  definimos la amplitud de cada modo normal, a  $T$ , como

$$A_k = \left( \frac{2 k_B T}{\lambda_k^{\text{escalado}}} \right)^{1/2} \quad (\text{Ec. 2.70})$$

Se puede generar un conjunto de estructuras que representen distorsiones térmicas obtenidas a partir de la estructura original de un archivo tipo pdb eligiendo valores aleatorios de amplitud  $a_k$  dentro del intervalo  $[-A_k : A_k]$ .

Desplazamientos aleatorios  $\delta r$  de las coordenadas de los átomos en la dirección de cada uno de los modos normales se obtienen mediante

$$\delta r_i = \frac{1}{\sqrt{m_i}} \sum_k^{3N-6} Q_{ik} a_k \quad (\text{Ec. 2.71})$$

siendo  $a_k(\text{Å})$  el valor aleatorio de amplitud correspondiente al modo normal  $Q_k$  y  $Q_{ik}$  es el elemento  $i$ -ésimo del mismo modo [Mahajan 2017].  $m_i$  es la masa de átomo correspondiente que se considero igual a 1.

### 2.3.3 Grado de colectividad

El grado de colectividad  $\kappa$  de un modo normal  $Q$  es una medida del número de residuos que son desplazados significativamente en la fluctuación descrita por ese modo [Tama 2001, Batista 2010]. Para cada modo normal  $Q_k$ , de longitud  $3N$  con elementos  $Q_{ik}$ , el grado de colectividad  $\kappa_k$ , se define como

$$\kappa_k = \frac{1}{N} \exp\left(-\sum_{i=1}^N (Q_{ik}^r)^2 \ln(Q_{ik}^r)^2\right) \quad (\text{Ec. 2.72})$$

siendo  $(Q_{ik}^r)^2 = (Q_{ik}^x)^2 + (Q_{ik}^y)^2 + (Q_{ik}^z)^2$  y  $(Q_{ik}^j)^2$  ( $j = x, y, z$ ) son los componentes del  $i$ -ésimo  $C\alpha$  en el modo normal  $k$ .  $\kappa_k \approx N^{-1}$  significa el modo normal representa el movimiento de un bajo numero de residuos y  $\kappa_k \approx 1$  representa que el modo normal describe el desplazamiento de un gran numero de aminoácidos.



*Capítulo 3*

*Dinámica vibracional*

*asociada a la*

*multiplicidad*

*conformacional*

*Cualquier tecnología lo suficientemente  
avanzada es indistinguible de la magia*

3° ley de Clarke sobre el avance científico,  
Perfiles del futuro (*Profiles of the future*), 1973.  
Arthur C. Clarke

### 3. DINÁMICA VIBRACIONAL ASOCIADA A LA MULTIPLICIDAD CONFORMACIONAL

Los modos normales implicados en el cambio conformacional se seleccionaron en base a su correspondiente superposición con las distorsiones estructurales introducidas por unión del ligando. En esta sección, se describe el procedimiento implementado para definir el subespacio compuesto por estos modos y la metodología utilizada para la caracterización de los residuos en posiciones claves.

#### 3.1 CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO: ASPECTOS ESTRUCTURALES

##### 3.1.1 Determinación de diferencias estructurales (RMSD)

La disimilitud estructural entre un par de proteínas alineadas se calcula generalmente como la raíz de la desviación cuadrática media, (RMSD por sus siglas en inglés, *Root Mean Square Deviation*)

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N d_i^2}{N}} \quad (\text{Ec. 3.1})$$

donde  $N$  es la cantidad de residuos y  $d_i$  es la distancia euclídeana entre los  $C\alpha$  en la posición  $i$ -ésima medida en Å. En la presente tesis, hemos utilizado el valor de RMSD como medida de disimilitud estructural entre pares de conformeros con y sin ligando de una misma proteína.

##### 3.1.2 Definición del vector diferencia entre conformeros

El vector diferencia  $\mathbf{V}$ , entre un par de estructuras con y sin ligando retiene la dirección del cambio estructural observado a partir de la unión del ligando. Definimos cada elemento del vector  $\mathbf{V}$  como

$$v_i = \frac{(y_i - x_i)}{\sum_j y_j - x_j} \quad (\text{Ec. 3.2})$$

donde  $x_i$  e  $y_i$  son las coordenadas de los  $C\alpha$  de la estructura sin ligando y con ligando respectivamente;  $N$  es el número de  $C\alpha$  de la estructura.

## 3.2 CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO: ASPECTOS DINÁMICOS

### 3.2.1 Subespacios vibracionales activos: definición

En primer lugar, el par de estructuras con y sin ligando se superpone minimizando el RMSD. El vector diferencia  $\mathbf{V}$  entre ambas estructuras retiene la dirección del cambio estructural observado tras la unión del ligando. Sin embargo, muchas proteínas contienen regiones no estructuradas o demasiado flexibles, cuyas coordenadas no están bien definidas experimentalmente. Con el fin de reducir la posibilidad de que nuestros resultados puedan verse sesgados por cualquier distorsión estructural que no esté directamente relacionada con la unión al ligando, usamos un factor *Gaussian-weighting* [Damm 2006] para reescribir la Ec. 3.2, del vector diferencia  $\mathbf{V}$ , de manera que sus elementos estén ponderados por los factores B correspondientes,

$$v_i = \frac{(y_i - x_i) e^{-\frac{B_i^{lf} + B_i^{lb}}{w}}}{\sum_j \left( (y_j - x_j) e^{-\frac{B_j^{lf} + B_j^{lb}}{w}} \right)^2} \quad (\text{Ec. 3.3})$$

Siendo  $B_i^{lf}$  el valor de factor B teórico por residuo de la estructura sin ligando y  $B_i^{lb}$  el valor correspondiente a la estructura con ligando ( $lb$  y  $lf$  hacen referencia a con ligando por *ligand-bound* y sin ligando por *ligand-free*). Se eligió un valor del factor de escala,  $w$ , igual a 0,01; aunque resultados similares a los presentados en el Capítulo 4, se obtuvieron con valores de  $w$  igual a 0,001 y 0,1.

A continuación, el vector normalizado  $\mathbf{V}$  se proyecta sobre la base de los modos normales de la estructura sin ligando

$$\mathbf{V}_Q = \sum_{k=1}^{3N-6} (\mathbf{V} \cdot \mathbf{Q}_k) \mathbf{Q}_k = \sum_{k=1}^{3N-6} \left( \sum_{j=1}^{3N} (v_j Q_{jk}) \right) \mathbf{Q}_k = \sum_{k=1}^{3N-6} c_k \mathbf{Q}_k \quad (\text{Ec. 3.4})$$

con

$$c_k = \sum_{j=1}^{3N} (v_j Q_{jk}) \quad (\text{Ec. 3.5})$$

El grado de deslocalización de  $\mathbf{V}_Q$  entre los diferentes modos normales del confórmero sin ligando se puede obtener evaluando el número de participación ( $M$ ) de modos como:

$$M = \left( \sum_{k=1}^{3N-6} (c_k)^4 \right)^{-1} \quad (\text{Ec. 3.6})$$

El número de participación se ha introducido originalmente como metodología para describir la medida de la deslocalización espacial de un vector de modos normales [Bell 1970, Taraskin 1999, Kalstein 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, que este número puede representar una medida de la deslocalización de un vector de modos normales sobre la base de las coordenadas cartesianas atómicas. En el presente trabajo, se amplió este concepto para aplicarlo a la deslocalización del vector  $\mathbf{V}_Q$  (Ec. 3.4), que tiene en cuenta las distorsiones estructurales introducidas por la unión del ligando, sobre la base de los modos normales del confórmero libre. 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 confórmero sin ligandos, 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. 3.5) definen el subespacio mínimo ( $\mathbf{S}$ ), de modos  $\{Q_{f_k}\}_{k=1,M}$ , necesarios para lograr una buena descripción del cambio conformacional. De este modo,  $\mathbf{S}$  contiene los modos normales implicados en el cambio conformacional por unión al ligando. Denominamos como modos activos al conjunto de modos normales que componen el subespacio  $\mathbf{S}$ . Es decir, el tamaño y la composición de los subespacios  $\mathbf{S}$ , asociados al cambio conformacional son definidos por  $M$  y el conjunto de  $M$  modos activos, respectivamente.

El efecto de la mutación puntual de un residuo  $i$  en el subespacio  $\mathbf{S}$  de modos normales asociados a la unión al ligando se simuló introduciendo perturbaciones a las interacciones en las que participa el residuo  $i$  (ver Capítulo 2, sección 2.2.2). De esta manera, se obtiene un nuevo conjunto de modos normales  $\{Q_k^i\}_{k=1,3N-6}$ . Para definir el nuevo subespacio  $\mathbf{S}^i$  es necesario establecer una correspondencia uno-a-uno entre el conjunto de modos normales original y los modos normales luego de la perturbación ya que la perturbación de las interacciones elásticas locales pueden provocar cambios en el orden de energía de los modos. Por esto, la asignación de los modos perturbados en función del criterio de orden de energía se vuelve inútil. La correspondencia entre ambos conjuntos de modos,  $\{Q_k^i\}_{k=1,3N-6}$  y  $\{Q_k\}_{k=1,3N-6}$ , se basa en seleccionar los valores más altos de solapamientos entre los modos [Atilgan 2001, Kalstein 2010]. Las superposiciones máximas se obtienen a través de la maximización de la traza del cuadrado de la matriz de solapamiento  $\mathbf{O}$  cuyos elementos se definen como el producto escalar

$$O_{kk'}^i = Q_k^t \cdot Q_{k'}^i \quad (\text{Ec. 3.7})$$

Esto se realiza seleccionando aquellos elementos de la matriz  $\mathbf{O}$ , uno para cada fila, y cada uno perteneciente a una columna diferente (o viceversa), que maximizan la suma de sus valores al cuadrado. Este tipo de problema, que es muy común en Economía y se lo conoce como el problema de Costo Mínimo (o *min-cost* en inglés), surge al querer distribuir  $N$  trabajos distintos entre  $N$  personas distintas de manera tal que cada una haga un trabajo distinto minimizando el costo global de todas las

tareas. Formalmente, sea  $\mathcal{C} = \{c_{ij}\}$ , con  $c_{ij} \geq 0 \quad \forall i, j$ , la matriz de costos  $\mathcal{N} \times \mathcal{N}$  y  $f_i$  una función que computa la permutación de las columnas  $j$ . El problema de asignación de Costo Mínimo consiste en encontrar el conjunto de permutaciones que minimicen la traza  $z$ , de la matriz  $\mathcal{C}$ , definida como

$$z = \sum_{i=1}^{\mathcal{N}} c_{i, f_i} \quad (\text{Ec. 3.8})$$

y puede ser resuelto de forma eficiente utilizando para ello el algoritmo conocido como Húngaro [Carpaneto 1988, Kalstein 2010]. Puesto que este algoritmo busca minimizar la traza de la matriz  $\mathcal{C}$ , para resolver nuestro problema en particular lo aplicamos a la matriz cuyos elementos son  $c_{ij} = -(O_{\kappa\kappa}^i)^2$ .

### 3.2.2 Subespacios vibracionales activos: comparación

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

$$p_j^{S^iS} = \sum_{k=1}^M (Q_j^i \cdot Q_k) Q_k \quad (\text{Ec. 3.9})$$

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

$$G_{k1} = (p_k^{S^iS} \cdot p_1^{S^iS}) \quad (\text{Ec. 3.10})$$

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

$$\mathbf{L}_G^T \mathbf{G} \mathbf{L}_G = \Lambda_G \quad (\text{Ec. 3.11})$$

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  $\mathbf{G}$  varía entre 0 y 1 [Krzanowski 1979], se puede definir una medida de la similitud de los dos subespacios como

$$\zeta^{S^iS} = \frac{\sum_k \lambda_k}{M} \quad (\text{Ec. 3.12})$$

Por lo tanto  $\zeta^{S^iS}$  representa la similitud entre los subespacio de modos activos antes y después de la perturbación de un residuo. Un valor de  $\zeta^{S^iS}$  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 lo tanto la mutación no afectaría a la diversidad conformacional. Cuanto menor sea el valor de  $\zeta^{S^i S}$ , mayor será el efecto de la mutación del residuo  $i$ -ésimo sobre el subespacio de modos asociados con el cambio conformacional. Es decir, que la diversidad conformacional requerida por la proteína se verá afectada.

Para cada residuo obtuvimos dos valores de similitud de subespacio, ya que realizamos dos perturbaciones,  $k_{ij}+\delta$  y  $k_{ij}-\delta$ , con  $\delta=0.05$ . De esta forma para cada residuo analizamos el efecto del aumento ( $k_{ij}+\delta$ ) de las constantes de fuerza en las interacciones en las que participa y el efecto de la disminución ( $k_{ij}-\delta$ ) de las mismas constantes de fuerza (ver Capítulo 2, sección 2.2.2, **Figura 2.2**). Luego calculamos el promedio a partir de los dos valores de similitud de subespacio,

$$\langle \zeta^{S^i S} \rangle = \frac{\zeta_{k_{ij}+\delta}^{S^i S} + \zeta_{k_{ij}-\delta}^{S^i S}}{2} \quad (\text{Ec. 3.13})$$

Para determinar el valor final de respuesta a la perturbación de cada residuo, se utilizó un modelo semejante al de *coarse-grained* de [Zheng 2009b], donde el valor para cada residuo  $i$  se define como

$$\langle \zeta^{S^i S} \rangle^\dagger = \frac{\langle \zeta^{S^i S} \rangle + \sum_j \langle \zeta^{S^i S} \rangle}{1+n_j} \quad (\text{Ec. 3.14})$$

para todo  $j$  que cumpla con la condición  $|r_j^0 - r_i^0| < 7 \text{ \AA}$ , donde  $r_i^0$  y  $r_j^0$  son las posiciones de equilibrio de los residuos  $i$  y  $j$ ;  $n_j$  es igual al número de residuos que cumplen con esta condición. De esta forma, analizamos regiones espaciales en lugar de residuos individuales.

El valor de  $\langle \zeta^{S^i S} \rangle^\dagger$  aumenta con la dimensionalidad del subespacio  $S$ . Esto dificulta la comparación estadística de  $\langle \zeta^{S^i S} \rangle^\dagger$  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  $\langle \zeta^{S^i S} \rangle^\dagger$  como:

$$Z_{\text{score}}^{S^i} = \frac{\langle \zeta^{S^i S} \rangle^\dagger - \overline{\langle \zeta^{S^i S} \rangle^\dagger}}{\sigma^S} \quad (\text{Ec. 3.15})$$

donde  $\overline{\langle \zeta^{S^i S} \rangle^\dagger}$  y  $\sigma^S$  son el promedio y la desviación estándar de  $\langle \zeta^{S^i S} \rangle^\dagger$  evaluados sobre todos los residuos de la proteína.

Para el análisis de la perturbación sobre el espacio vibracional de una mutación puntual en la proteína Transtretina (TTR) modificamos la metodología para la comparación del espacio formado por los modos normales. En este punto cabe aclarar que no definimos un subespacio de modos normales sino que evaluamos la estabilidad de la estructura tetramérica de TTR utilizando la totalidad de los modos normales. Para simular mutaciones puntuales en el tetrámero de TTR, perturbamos simultáneamente las constantes de fuerza del  $i$ -ésimo residuo en cada monómero. El nuevo conjunto

de modos normales perturbados  $\{Q_k^i\}_{k=1,4 \times 3N-6}$  ( $N$  es el número de residuos de cada monómero) se comparó con los modos normales no perturbados  $\{Q_k\}_{k=1,4 \times 3N-6}$  de la siguiente manera.

En primer lugar, calculamos la matriz de solapamiento  $\mathbf{O}$  (Ec 3.7). Luego definimos la similitud  $R^i$  entre estos conjuntos de modos normales como

$$R^i = \frac{\sum_{k=1}^{4 \times 3N} (O_k^{i,\max})^{1/2}}{4 \times 3N} \quad (\text{Ec. 3.17})$$

con  $O_k^{i,\max}$  definido como el valor máximo de  $(O_{kk'}^i)^2$  entre las proyecciones  $O_{kk'}^i$  ( $k'=1, 4 \times 3N-6$ ) del modo  $Q_{k'}^i$  sobre los modos  $\{Q_k\}_{k=1,4 \times 3N-6}$ .

Cuanto menor es el valor de  $R^i$  mayor es el efecto que la mutación en el residuo  $i$  genera en la dinámica de equilibrio del tetrámero de TTR. Finalmente evaluamos el efecto de la mutación del residuo  $i$  en la dinámica de equilibrio del tetrámero de TTR normalizando  $R^i$  como

$$Z_{\text{score}}^{R^i} = \frac{R^i - \bar{R}}{\sigma^R} \quad (\text{Ec. 3.18})$$

donde  $\bar{R}$  y  $\sigma^R$  son el promedio y la desviación estándar de la distribución de  $R^i$  sobre todos los residuos. De esta manera los residuos con valores negativos grandes de  $Z_{\text{score}}^{R^i}$  son residuos cuyas mutaciones alteran significativamente la dinámica de equilibrio del tetrámero de TTR.

### 3.2.3 Dinámica asociada a patrones de flexibilidad relativa

En la sección 3.2.1 describimos la metodología mediante la cual determinamos los modos normales que describen el cambio conformacional asociado con la unión al ligando. De manera similar podemos seleccionar los modos normales que describen el patrón de flexibilidad de cada proteína. Para esto definimos el vector  $\mathbf{B}^{\text{lf}}$ ; cuyos elementos  $B_i^{\text{lf}}$  corresponden a los factores  $B$  asociados a cada  $i$ -ésimo residuo de la estructura sin ligando (lf). Luego proyectamos el vector  $\mathbf{B}^{\text{lf}}$  sobre la base de los modos normales del mismo conformero.

$$\mathbf{B}_Q^{\text{lf}} = \sum_{k=1}^{3N-6} (\mathbf{B}^{\text{lf}} \cdot Q_k) Q_k = \sum_{k=1}^{3N-6} \left( \sum_{j=1}^{3N} (B_j^{\text{lf}} Q_{jk}) \right) Q_k = \sum_{k=1}^{3N-6} b_k Q_k \quad (\text{Ec. 3.19})$$

con

$$b_k = \sum_{j=1}^{3N} (B_j^{\text{lf}} Q_{jk}) \quad (\text{Ec. 3.20})$$

De esta manera, el número de participación (ver Sección 3.2.1) de modos asociados al patrón de flexibilidad,  $M_B$  se define como

$$M_B = \left( \sum_{k=1}^{3N-6} (b_k)^4 \right)^{-1} \quad (\text{Ec. 3.21})$$

Con una interpretación equivalente al número de participación, descrita en la Sección 3.2.1, los primeros  $M_B$  modos ordenados por valor de  $b_k^2$  en forma decreciente definen el subespacio mínimo  $S_B$  de modos requeridos para lograr una buena descripción del patrón de flexibilidad. Es decir,  $S_B$  contiene los modos normales relevantes para el cálculo de los factores B teóricos de la conformación sin ligando.

### 3.3 CARACTERIZACIÓN DE RESIDUOS DINÁMICAMENTE IMPORTANTES EN EL CAMBIO CONFORMACIONAL POR UNIÓN AL LIGANDO

#### 3.3.1 Residuos en posiciones claves: definición

Definimos las posiciones claves como aquellas dentro del 5% de los valores más bajos de  $Z_{score}^S$  para cada estructura en el conjunto de proteínas estudiadas. También consideramos la definición de las posiciones como aquellas dentro del 1% o 10%, pero no obtuvimos diferencias cualitativas en nuestros resultados. En este punto, cabe resaltar que nuestra definición de posiciones claves se basa en valores relativos del cambio en las direcciones de los modos normales que participan de la unión al ligando y no en las diferencias estructurales absolutas de las estructuras de ambos conformeros. El número de residuos en posiciones claves ( $N_{PC}$ ) obtenido para cada par de conformeros del set de proteínas de estudio se proporciona en el Apéndice B.

#### 3.3.2 Análisis de la conservación evolutiva

Para cada par de conformeros de nuestro set de proteínas de estudio, obtuvimos alineamientos de la base de datos HSSP (por sus siglas en inglés, *Homology-derived Secondary Structure of Proteins*). Esta es una base de datos que combina información estructural y secuencial. Para cada proteína de estructura 3D conocida de la base de datos PDB, HSSP contiene una alineación de secuencias múltiples de todos los homólogos disponibles y un perfil de secuencia característico de la familia [Schneider 1997]. Se utilizaron secuencias del alineamiento cuyo número de residuos era igual o mayor al 80% del tamaño de la secuencia aminoacídica de la proteína en estudio.

Para el análisis de la conservación relativa de cada aminoácido, de cada proteína del set de estudio, utilizamos el programa AL2CO [Pei 2001]. Para el cálculo de la conservación, se considero el método *Henikoff entropy measure*. El programa realiza dos pasos para estimar la conservación de una posición en un alineamiento múltiple de secuencias. En el primer paso, se estima la frecuencia del aminoácido ( $f_i^A$ ) en la posición seleccionada

$$f_i^A = \frac{\sum_{k=1}^{n(i)} \delta(A, k, i) w_k}{\sum_{k=1}^{n(i)} w_k} \quad (\text{Ec. 3.22})$$

donde consideramos  $\delta(A, k, i) = 1$  si el aminoácido A esta en la secuencia k en la posición i, en caso contrario  $\delta(A, k, i) = 0$ .  $w_k$  es la ponderación de la secuencia k. Para el calculo de  $w_k$  el programa utiliza el método Henikoff-Henikoff modificado, al igual que el que se aplica en PSI-BLAST [Henikoff 1994, Altschul 1997]. La idea de la ponderación es resaltar diferencias entre pares de secuencias en la alineación. De esta manera dos secuencias cercanas con alta similitud influyen en las frecuencias de aminoácidos menos que un par de secuencias divergentes. Por lo tanto, el peso atribuido a secuencias similares dentro de una de una gran familia es menor que el peso de una secuencia divergente. En el calculo de la ponderación una posición se descarta si es invariante o si contiene *gaps* en al menos el 50 % de las secuencias [Pei 2001].

En el segundo paso, el índice de conservación  $c_i$ , se calcula utilizando las frecuencias  $f_i^A$ ,

$$c_i = \sum_{A=1}^{20} f_i^A \ln(f_i^A) \quad (\text{Ec. 3.23}).$$

Para determinar el valor final de conservación  $C_i$ , se utilizo un modelo semejante al de *coarse-grained* de [Zheng 2009b], donde el valor para cada residuo i se define a partir de

$$C_i = \frac{c_i + \sum_j c_j}{1 + n_j} \quad (\text{Ec. 3.24})$$

para todo j que cumpla con la condición  $|r_j^0 - r_i^0| < 7 \text{ \AA}$ , donde  $r_i^0$  y  $r_j^0$  son las posiciones de equilibrio de los residuos i y j;  $n_j$  es igual al número de residuos que cumplen con esta condición. De esta forma; al igual que para la respuesta a la perturbación puntual; analizamos regiones espaciales en lugar de residuos individuales. Para cada proteína del set se normalizaron los valores de  $C_i$  como:

$$Z_{\text{score}}^C = \frac{C_i - \bar{C}_i}{\sigma^C} \quad (\text{Ec. 3.25})$$

donde  $\bar{C}_i$  y  $\sigma^C$  son el promedio y la desviación estándar de  $C_i$  por proteína.

### 3.3.3 Determinación de estructura secundaria

Para la determinación de la estructura secundaria de los aminoácidos se utilizó el programa DSSP (por sus siglas en ingles, *Dictionary of protein secondary structure*) [Kabsch 1983, Touw 2015]. El algoritmo funciona mediante el cálculo de la asignación de estructura secundaria más probable para la estructura 3D de una proteína. Esto se hace mediante la lectura de la posición de los átomos en una proteína seguido por el cálculo de la energía de enlace de H entre todos ellos. Los dos mejores enlaces de H para cada átomo se utilizan entonces para determinar la clase de estructura secundaria más probable para cada residuo en la proteína.

Se empieza identificando los enlaces de hidrógeno intra-catenarios de la proteína utilizando una definición puramente electrostática, suponiendo cargas parciales de  $-0,42 e$  al oxígeno del carbonilo y de  $0,20 e$  al hidrógeno de la amida. Así un puente de hidrógeno es identificado si  $E$  en la siguiente ecuación es inferior a  $-0,5$  kcal/mol

$$E = q_1 q_2 \left( \frac{1}{r_{\text{ON}}} + \frac{1}{r_{\text{CH}}} - \frac{1}{r_{\text{OH}}} - \frac{1}{r_{\text{CN}}} \right) f \quad (\text{Ec. 3.26})$$

con  $q_1 = 0,42 e$  y  $q_2 = 0,20 e$ , siendo  $e$  la carga del electrón.  $r_{AB}$  es la distancia interatómica entre A y B, y esta definida en angstroms (Å).  $f$  es el factor dimensional y tiene un valor de 332 kcal/mol.

Basado en esto, se asignan ocho tipos de estructura secundaria. El patrón estructural para el programa es el ***n*-turn** (T). Este se define a partir de la presencia de un enlace de hidrógeno entre un CO (i) y un NH (i+n), donde i es el número de aminoácido y n puede ser 3,4 o 5. Las hélices se definen como un mínimo de 2 ***n*-turn** consecutivos, definiéndose tres tipos diferentes de hélices, **hélice- $\alpha$**  (H), **hélice-3** o **hélice-3<sub>10</sub>** (G) y **hélice-5** o **hélice- $\pi$**  (I). En donde los enlaces de H se dan entre el residuo i y i+n, siendo n igual a 4, 3 o 5 respectivamente. Dos tipos de hojas- $\beta$ ; **hoja- $\beta$  unitaria** (B), donde se generan 2 enlaces de hidrógeno entre dos pares de tres residuos, [ i-1, i, i+1 ] y [ j-1, j, j+1 ]. Y **hoja- $\beta$  extendida** (E), a partir de dos o más hojas- $\beta$  unitarias consecutivas. Las regiones de alta curvatura, donde el ángulo entre los carbonos alpha [ i-1, i ] y [ i, i+1 ] es menor a  $70^\circ$ , se las designa como **bend** (S). A las estructuras que no apliquen a ninguna de las reglas definidas, se las designa como **loops** (N) [Kabsch 1983].

### 3.3.4 Accesibilidad al solvente

La superficie de accesibilidad al solvente (SAS) de los residuos la calculamos mediante el programa NACCESS [Hubbard 1993]. Este programa calcula la superficie de accesibilidad al solvente para cada átomo de una molécula mediante la simulación de esferas de tamaño dado en torno a una superficie de van der Waals para cada átomo [Lee 1971]. Puesto que el tamaño de las cadenas laterales de los distintos aminoácidos no es el mismo, una aproximación es considerar el Área Relativa de accesibilidad al Solvente (ARS), expresando la SAS absoluta calculada para un aminoácido X como un porcentaje de la SAS que se observaría si ese mismo aminoácido X estuviese entre dos Alaninas (Ala-X-Ala) o entre dos Glicinas (Gly-X-Gly), en conformaciones extendidas, de esta manera se expone el residuo X en el tripéptido tanto como normalmente sería posible en una proteína. Por lo que un valor bajo de ARS implica que el residuo se encuentra cubierto o rodeado, impidiendo que la accesibilidad sea la misma que la que tendría en el tripéptido. Debido a ángulos de enlace inusuales, las longitudes de enlace y la geometría distorsionada en proteínas reales, estos valores pueden exceder de 100% (especialmente para la prolina donde el programa utiliza un tripéptido Ala-Pro-Ala verdadero, construido de una proteína real). Asimismo, puede ser que se tenga grandes valores de RSA, cuando los residuos están previos a una rotura de cadena o al final de la misma, tal como en el

extremo amino o carboxilo terminal de la proteína. Los tripéptidos Ala-X-Ala o Gly-X-Gly se construyen usando el paquete de gráficos moleculares QUANTA (Software de modelado molecular y simulaciones dinámicas. *Accelrys Software Inc.*) [Hubbard 1993].

### 3.3.5 Distancia de residuos al sitio activo

Establecer interacciones entre proteínas y ligandos biológicamente relevantes es un paso importante hacia la comprensión de las funciones de las proteínas. Un ligando presente en una proteína se considera biológicamente relevante si interactúa con la proteína y desempeña ciertas funciones biológicas, como inhibidor, activador u análogo de sustrato. La mayoría de los métodos de predicción de sitios de unión al ligando usan las estructuras de proteínas del PDB como patrones. Sin embargo, no todos los ligandos presentes en el PDB son biológicamente relevantes, ya que diferentes moléculas pequeñas (por ejemplo glicerol, etilenglicol) se usan a menudo como aditivos para resolver las estructuras de las proteínas [Yang 2013]. Yang et al. desarrollaron un procedimiento para evaluar la relevancia biológica de los ligandos presentes en las estructuras de PDB, esto con el fin de facilitar el estudio de *docking* proteína-ligando y la detección virtual de ligandos. De esta manera generaron la base de datos BioLip, la cual es una base de datos curada semi-manualmente con información de interacciones proteína-ligando biológicamente relevante de la base de datos de PDB [Yang 2013]. Cada entrada en BioLiP contiene una lista completa de anotaciones. Particularmente nos interesamos en la información sobre residuos de unión al ligando, afinidad de unión al ligando y residuos del sitio catalítico (información conectada con la base *Catalytic Site Atlas* [Furnham 2014]).

A partir de la identificación de los residuos que componen el sitio activo de la proteína, se realizó el cálculo del centro de masa del sitio activo. Las coordenadas del mismo se obtuvieron a partir del siguiente cálculo

$$x_{CM} = \frac{\sum_{i=1}^N m_i x_i}{m_T} \quad y_{CM} = \frac{\sum_{i=1}^N m_i y_i}{m_T} \quad z_{CM} = \frac{\sum_{i=1}^N m_i z_i}{m_T} \quad (\text{Ec. 3.27})$$

Donde N es el número de residuos que componen el sitio activo.  $m_T$  es la suma de la masa de los N residuos que conforman el sitio activo;  $m_i$  es la masa del residuo i, que al igual que en cálculo de modos normales se consideró igual a uno (1).  $x_i$ ,  $y_i$  y  $z_i$  son las coordenadas cartesianas del i-ésimo residuo que conforma el sitio activo, obtenidas a partir del pdb de la proteína.

A continuación, se calculó la distancia euclídeana de cada residuo respecto al centro de masa del sitio activo de la proteína

$$d_{i,CM} = \sqrt{(x_i - x_{CM})^2 + (y_i - y_{CM})^2 + (z_i - z_{CM})^2} \quad (\text{Ec. 3.28})$$

### 3.3.6 Ángulo entre elementos de estructura secundaria

A partir de la información obtenida mediante el programa DSSP (Sección 3.3.3), determinamos los elementos de estructura secundaria de cada proteína de nuestro set de estudio (SSE, por *Secondary Structure Element*). Posteriormente, seleccionamos pares de SSEs que presenten interacciones entre sus residuos, esto es, interacción inter-SSEs. Con el fin de analizar modificaciones en la orientación relativa entre estos pares de SSEs durante el cambio conformacional, realizamos el siguiente procedimiento.

Primero determinamos de los ejes principales de inercia mediante la construcción de la matriz de Tensor de Inercia  $\mathbf{I}$  y su posterior diagonalización. Los elementos diagonales de la matriz  $\mathbf{I}$ , son

$$\begin{aligned} I_{xx} &= \sum_{i=1}^N m_i [(y_i - y_{CM})^2 + (z_i - z_{CM})^2] \\ I_{yy} &= \sum_{i=1}^N m_i [(x_i - x_{CM})^2 + (z_i - z_{CM})^2] \\ I_{zz} &= \sum_{i=1}^N m_i [(x_i - x_{CM})^2 + (y_i - y_{CM})^2] \end{aligned} \quad (\text{Ec. 3.29})$$

mientras que los elementos fuera de la diagonal se definen como

$$I_{xy} = - \sum_{i=1}^N m_i (x_i - x_{CM})(y_i - y_{CM}) \quad (\text{Ec. 3.30})$$

donde  $N$  es el número de residuos que componen el primer SSE (SSE- $I$ );  $m_i$  es la masa del residuo  $i$ , que al igual que en calculo de modos normales, se considero igual a uno (1).  $x_i$ ,  $y_i$  y  $z_i$  son las coordenadas cartesianas del  $i$ -ésimo residuo que conforma el SSE- $I$ ;  $x_{CM}$ ,  $y_{CM}$ ,  $z_{CM}$  son las coordenadas del centro de masa del SSE- $I$ .

Los momentos principales y los ejes principales de inercia son los autovalores y autovectores obtenidos al diagonalizar la matriz  $\mathbf{I}$  [Browner 1992]. Los autovectores  $v_1$ ,  $v_2$  y  $v_3$  satisfacen la ecuación

$$\mathbf{I} v_k = \lambda_k^\dagger v_k \quad (\text{Ec. 3.31})$$

para  $k = 1, 2$  y  $3$ ; donde  $\lambda_k^\dagger$  son los momentos principales.

Luego rotamos la SSE- $I$  para poner los ejes cartesianos en coincidencia con los ejes principales de inercia. Para esto utilizamos la matriz de autovectores  $v_k$  definiendo las nuevas coordenadas como

$$\begin{aligned} x_i' &= x_i v_{11} + y_i v_{12} + z_i v_{13} \\ y_i' &= x_i v_{21} + y_i v_{22} + z_i v_{23} \\ z_i' &= x_i v_{31} + y_i v_{32} + z_i v_{33} \end{aligned} \quad (\text{Ec. 3.32})$$

Igualmente aplicamos esta rotación sobre el segundo SSE (SSE- $II$ ). De esta manera obtuvimos nuevas coordenadas cartesianas para SSE- $I$  e SSE- $II$ .

A partir de las estructuras SSE- $I$  y SSE- $II$  reorientadas, generamos dos nuevas matrices de inercia  $\mathbf{I}'$  y  $\mathbf{I}''$  que fueron diagonalizadas para obtener nuevos ejes principales de inercia. Tal como se

implementó, la determinación de los ejes principales describe la orientación del SSE-II en relación con el SSE-I. La disposición de un eje principal con respecto a otro se describe mediante el ángulo entre los ejes principales. Por ejemplo calculamos el ángulo entre el autovector  $v_1^I$  y el autovector  $v_1^{II}$  como

$$\theta_1 = \arccos \frac{v_1^I v_1^{II}}{|v_1^I| |v_1^{II}|} \quad (\text{Ec. 3.33})$$

Mediante ecuaciones análogas para los autovectores  $v_2$  y  $v_3$  obtenemos  $\theta_2$  y  $\theta_3$  [Browner 1992, Brylinski 2008].

Finalmente, calculamos el modulo de la diferencia entre los ángulo, para cada eje de coordenada, entre las estructuras con (Ib) y sin (If) ligando.

$$\Delta \theta_k = |\theta_k^{If} - \theta_k^{Ib}| \quad (\text{Ec. 3.34})$$

con k igual a 1, 2 o 3. Elegimos la mayor diferencia entre estos ángulos como una medida cuantitativa de las diferencias de orientación relativa de las SSE.

### 3.3.7 Redes de contactos inter-atómicos

A pesar del número creciente de estructuras de proteínas publicadas y de la idea de que la función de una proteína se basa en su estructura tridimensional, hay un acceso limitado a programas automáticos que utilicen la estructura de las proteínas para identificar residuos críticos para su función biológica. Esto es porque la identificación de residuos funcionales en proteínas es un problema complejo, incluso cuando se cuenta con estructuras atómicas detalladas [Jones 2004]. Esto resulta más complicado aun cuando no se conocen estructuras de homólogos con residuos funcionales caracterizados. Solo una pequeña fracción de todas las proteínas conocidas ha sido bioquímicamente bien estudiada. Por lo tanto, para la mayoría de las proteínas con estructuras tridimensionales resueltas, necesitamos métodos *de novo* para predecir los residuos funcionales.

La combinación de datos de conservación evolutiva, junto con información de la estructura, ha sido exitosa para predecir sitios de unión de ligandos y sitios activos en diversas proteínas [Armon 2001]. Sin embargo, algunas proteínas con estructuras conocidas no tienen homólogos determinados. Para este conjunto de proteínas, es clave tener una forma precisa de identificar residuos críticos en ausencia de ortólogos conocidos es clave. Utilizando solo datos estructurales, los residuos funcionales se pueden identificar mediante el cálculo energético de la estructura y propiedades de ionización [Elcock 2001, Ondrechen 2001]. En un enfoque diferente, estos residuos podrían caracterizarse por sus propiedades estructurales o motivos estructurales recurrentes. [Bartlett 2002, Wangikar 2003].

En las últimas décadas, el análisis de redes se ha utilizado para modelar diversos sistemas, como la Red Mundial de Internet (*World Wide Web*), los sistemas sociales y los sistemas biológicos (por ejemplo, interacciones proteína-proteína o la red metabólica celular) [Oltvai 2002]. El concepto de red recientemente se ha aplicado para analizar y predecir la dinámica de proteínas [Böde 2007], el plegamiento [Dokholyan 2002, Vendruscolo 2002] y residuos funcionales [Amitai 2004, Thibert 2005,



Cusack 2007]. Las redes proporcionan una manera eficaz y sencilla de describir las interacciones entre aminoácidos y ofrecen una novedosa idea para explorar la estructura y función de proteínas.

La representación de la estructura de la proteína como red facilita la búsqueda de determinantes topológicos que pueden ser asociados a residuos funcionalmente importantes. Sería razonable suponer que los residuos centrales en la estructura proteica deberían desempeñar un papel importante en la transmisión de la información entre los aminoácidos. Estos residuos centrales se han asociado con residuos claves o críticos en el plegamiento de proteínas [Dokholyan 2002, Vendruscolo 2002] o la funcionalidad [Thibert 2005, Cusack 2007] y con residuos del sitio activo dentro de familias de enzimas [Amitai 2004]. En otros trabajos, la centralidad de los residuos se presenta como una característica topológica conservada dentro de las familias de proteínas [Del Sol 2006].

En esta tesis, la construcción de las redes de interacciones entre residuos (RIN, por sus siglas en inglés, *Residue Interaction Networks*) se llevo a cabo mediante el programa RING 2.0 (por sus siglas en inglés, *Residue Interaction Network Generator*) [Martin 2011, Piovesan 2016]. Dentro de una RIN, cada residuo se representa como un nodo, y estos nodos están conectados, mediante enlaces o *linkers*, en función de la interacción entre ellos.

En este punto cabe aclarar que inicialmente trabajamos con el programa CSU [Sobolev 1999] para la determinación de tipo de interacciones entre residuos, principalmente para la adaptación del Modelo de ANM (Capítulo 2, sección 2.2.1). Pero este solo nos proporcionaba información detallada sobre interacciones de tipo Puente de Hidrógeno.

RING 2.0 es una nueva herramienta web diseñada para la visualización y el análisis de estructuras de proteínas en términos de interacciones fisico-químicas, la información evolutiva y energética. El algoritmo de RING 2.0 genera la red en dos pasos. El primero identifica una lista de pares de residuo entre los que pueden existir interacción, basado simplemente en la distancia entre los mismos. El segundo caracteriza cada contacto identificando el tipo específico de interacción.

El primer criterio a la hora de definir las redes es el grado de simplicidad al considerar un residuo. RING 2.0 permite diferentes estrategias en este sentido. En la opción *Closest* se consideran para medir la distancia todos los átomos del par de residuos; esta opción es conveniente para pDBs con buena resolución donde se pueda confiar en las coordenadas descriptas para las cadenas laterales. La segunda posibilidad, llamada *Lollipop*, realiza el cálculo de la distancia entre los centros de masa de los dos residuos que interactúan. Además, el algoritmo comprueba que las cadenas laterales no apuntan en dirección opuesta. Las dos posibilidades restantes, son consideradas de tipo “grano grueso” (*Coarse Grained*) ya que la distancia entre residuos se calcula considerando solo C $\alpha$  o solo carbonos beta, C $\beta$ .

Inicialmente para RING2.0 todas las interacciones definidas se denominan genéricamente como contacto interatómico (IAC, por sus siglas en inglés, *Inter-Atomic Contact*). En el segundo paso el algoritmo identifica el tipo de interacción para cada par. Teniendo en cuenta que el cálculo de RING se basa en criterios geométricos, cada par de residuos puede formar múltiples interacciones; esto da

lugar al segundo parámetro para definir el tipo de red. Se puede considerar una sola interacción por residuo, definida como la de mayor energía posible. O generar una red de múltiples interacciones por residuo, pero sólo una por tipo de interacción.

El algoritmo, identifica 6 tipos diferentes de interacción, además de la interacción genérica (IAC) que simplemente indica un contacto basado en la distancia entre residuos. El cálculo de interacciones tipo puente de hidrógeno está inspirado en el método aplicado por el programa DSSP (Sección 3.3.3). Teniendo en cuenta un par de átomos, donante/aceptor (D/A), dos reglas definen un enlace de hidrógeno. Primero, la distancia, entre A y D es menor o igual a 3,5 Å; segundo el ángulo H-D-A ( $\theta$ ), formado por el hidrógeno del donante, el átomo donante y el átomo aceptor, es menor o igual a 63°. Las interacciones de Van der Waals se identifican midiendo la distancia entre la superficie de dos átomos, considerando 0,5 Å como umbral de distancia. Sólo se consideran pares de átomos de carbono-carbono o carbono-azufre para una interacción válida y la distancia se calcula teniendo en cuenta el radio de Van der Waals, de estos átomos. En el caso de Glutamina y la Asparagina, el programa considera interacciones de Van der Waals entre átomos de nitrógeno u oxígeno con carbono. Se establece un Puentes Disulfuro cuando la distancia entre los átomos de azufre de Cisteínas es menor o igual a 2,5 Å. Los Puentes salinos ocurren entre residuos con cargas opuestas. Esta interacción iónica es posible si la distancia entre los centros de masa de los grupos cargados es menor o igual a 4,0 Å.

Las últimas dos interacciones que identifica el programa son, interacción  $\pi$ - $\pi$  y catión- $\pi$ . Las interacciones catión- $\pi$  son posibles entre los aminoácidos cargados positivamente y una cadena lateral aromática. Las interacciones de apilamiento o  $\pi$ - $\pi$  se evalúan entre residuos aromáticos (His, Tyr, Trp, Phe). Estos dos tipos de interacciones no fueron considerados para nuestro trabajo. Por último, RING 2.0 no define enlaces peptídicos, por lo que se los agregó al archivo de salida del programa.

Dentro del programa se debe configurar el valor de Separación de Secuencia (SS por sus siglas en inglés *Sequence Separation*), que determina la separación mínima, a nivel de secuencia, entre dos residuos para determinar el tipo de interacción.

Para cada par de estructuras generamos una red total, considerando todos los residuos; una red de residuos en posiciones claves y una red de otros residuos, considerando únicamente los residuos fuera de las posiciones claves. Para analizar la robustez de las redes de residuos en posiciones claves, elegimos residuos aleatoriamente para generar redes al azar. Para cada proteína, se consideró el número de residuos en posiciones claves (N) y la distancia entre ellos. Como primer paso seleccionamos un residuo aleatoriamente, luego el segundo residuo aleatorio debe estar a la misma distancia que los primeros dos residuos en posiciones claves. Para el tercer residuo, la distancia al segundo debe ser la misma a la distancia entre el segundo y el tercer residuo en posición clave. De esta manera se eligió sistemáticamente los N aminoácidos que formaran la red de residuos aleatorios.

### 3.3.8 Parámetros topológicos de redes

En esta tesis analizamos y correlacionamos nuestro método de identificación de residuos claves para el mantenimiento de la diversidad conformacional asociada a la unión al ligando con información obtenida del análisis de redes. Se estudió especialmente parámetros topológicos de redes que se hayan utilizado anteriormente para identificar residuos funcionales [Dokholyan 2002, Vendruscolo 2002, Amitai 2004, Thibert 2005, Del Sol 2006, Cusack 2007].

Los parámetros topológicos de redes se calcularon mediante un complemento del programa Cytoscape, llamado *NetworkAnalyzer* [Assenov 2008, Doncheva 2012]. El primer parámetro que calcula el programa es el número de conexiones de cada nodo, llamado Grado o *Degree*. En base a esta información, se determina el Número de Componentes Conectados (NCC por sus siglas en inglés *Number of Connected Component*). Dentro de una red todos los nodos que están conectados entre sí forman una Componente Conectado, por lo que el NCC indica la conectividad de una red, un número NCC pequeño sugiere una conectividad mayor. Si no existe un nodo aislado o un grupo de nodos aislados el valor de NCC es igual a uno (1), es decir que todos los nodos se encuentran interconectados, generando una sola red.

El camino por el cual dos nodos  $n$  y  $m$  se unen, se determina a partir del número enlaces o *linkers* que los conectan. Pueden existir varios caminos que conecten ambos nodos, por lo tanto para cada nodo  $n$  se puede definir el camino mas corto que lo conecte con el nodo  $m$ , llamado Longitud del Camino Mínimo o *Shortest Path Length*,  $L(n, m)$ . Finalmente la Longitud Media del Camino Mínimo o *Average Shortest Path Length*,  $\langle L(n, m) \rangle$ , es el camino mas corto promedio entre  $n$  y cualquier otro nodo. Si  $n$  es un nodo aislado, el valor de este parámetro es cero.

El valor de Proximidad Central o *Closeness Centrality*,  $Cc_n$ , se define como la inversa de la Longitud Media del Camino Mínimo,

$$Cc_n = 1 / \langle L(n, m) \rangle \quad (\text{Ec. 3.35})$$

Es una medida de la rapidez con la que la información se extiende desde un nodo a otros nodos accesibles en la red. Este valor para cada nodo es un número entre 0 y 1, donde nodos con valores de  $Cc_n$  cercanos a 1 interactúan directamente, o a través de pocos intermediarios, con la mayoría de los nodos de la red; mientras que el  $Cc_n$  de un nodo aislado es igual a 0. Por ejemplo el valor de Proximidad Central del nodo  $b$  ( $Cc_b$ ) en la **Figura 3.1** se calcula como:

$$Cc_b = \frac{1}{[(L(b, a) + L(b, c) + L(b, d) + L(b, e)) / 4]} = \frac{4}{(1 + 1 + 1 + 2)} = \frac{4}{5} = 0,8$$

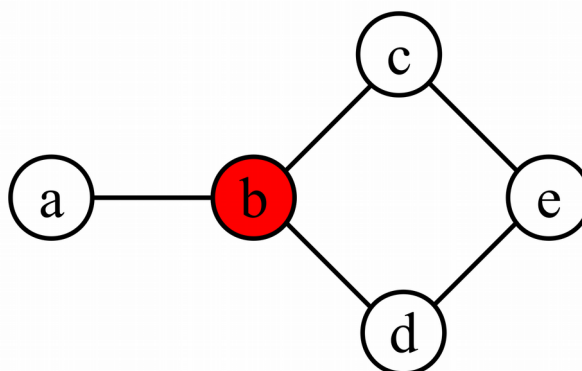
El parámetro Centralidad de Intermediación o *Betweenness Centrality*,  $B_n$ , refleja una medida del control que un nodo ejerce sobre las interacciones de otros nodos en la red. Permite distinguir los nodos que unen “comunidades” o subredes densas, en lugar de los nodos que se encuentran dentro de la misma. Se calcula como,

$$B_n = \sum_{s \neq n \neq t} \frac{L(s,t)_n}{L(s,t)} \quad (\text{Ec. 3.36})$$

Donde  $s$  y  $t$  son nodos en la red, diferentes de  $n$ .  $L(s,t)$  denota la Longitud del Camino Mínimo entre los nodos  $s$  y  $t$ ;  $L(s,t)_n$  es la Longitud de Camino Mínimo del nodo  $s$  al  $t$ , pasando si o si por el nodo  $n$ . El valor  $B_n$  para cada nodo es normalizado por  $(N-1)(N-2)/2$ ,  $N$  es el número total de nodos en el componente conectado al que  $n$  pertenece; Por lo tanto, el  $B_n$  de cada nodo es un número entre 0 y 1. Los nodos con valores de  $B_n$  mayores son aquellos que conectan *clusters* de nodos, no necesariamente poseen número alto de contactos, pero si son claves para relacionar nodos separados de una red. Por ejemplo el valor de Centralidad de Intermediación del nodo  $b$  ( $B_b$ ), en la **Figura 3.1** se calcula como:

$$B_b = \left( \frac{L(a,c)_b}{L(a,c)} + \frac{L(a,d)_b}{L(a,d)} + \frac{L(a,e)_b}{L(a,e)} + \frac{L(c,d)_b}{L(c,d)} + \frac{L(c,e)_b}{L(c,e)} + \frac{L(d,e)_b}{L(d,e)} \right) / 6$$

$$= \left( \left(\frac{1}{1}\right) + \left(\frac{1}{1}\right) + \left(\frac{2}{2}\right) + \left(\frac{1}{2}\right) + 0 + 0 \right) / 6 \approx 0,583$$



**Figura 3.1** Ejemplo de red con 5 nodos y 5 *linkers*. Se distingue al nodo  $b$  en rojo ya que es que utilizamos para los cálculos de  $Cc_b$  y  $B_b$ .

En trabajos anteriores, se ha postulado que la cercanía entre residuos dentro de una red permitía caracterizar sitios funcionales dentro de una proteína [Del Sol 2006]. Amitai et al., llevaron a cabo una predicción a gran escala de residuos del sitio activo y examinaron la relación entre valores de Proximidad Central por residuo, con otras funciones de los mismos [Amitai 2004]. Además, estudios anteriores han relacionado parámetros de centralidad de una red de interacción de aminoácidos con la capacidad de tolerar la sustitución de diversos aminoácidos dentro de la red. En particular, se ha demostrado que el parámetro de Centralidad de Intermediación destaca residuos importantes para el plegado de una proteína [Vendruscolo 2002]. Por lo tanto, el valor de Centralidad de Intermediación podría estar asociado con la capacidad de un residuo para desempeñar un papel crítico en la función de una proteína [Cusack 2007].



# *Capítulo 4*

## *Resultados*

*1° Ley. Un robot no hará daño a un ser humano o, por inacción, permitir que un ser humano sufra daño.*

*2° Ley. Un robot debe hacer o realizar las órdenes dadas por los seres humanos, excepto si estas órdenes entrasen en conflicto con la 1ª Ley.*

*3° Ley. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la 1ª o la 2ª Ley.*

Círculo vicioso (*Runaround*), 1942.

Isaac Asimov

## 4. RESULTADOS

### 4.1 RESIDUOS EN POSICIONES CLAVES PARA EL MANTENIMIENTO DE LA DIVERSIDAD CONFORMACIONAL

#### 4.1.1 Determinación del set de proteínas

El set de estructuras de proteínas de estudio fue seleccionado a partir de pares de conformeros con y sin ligando de la base de datos de la diversidad conformacional (CoDNaS, por sus siglas en inglés, *Conformational Diversity in the Native State of proteins*) [Monzon 2016]. Esta base de datos es una colección de estructuras redundantes para la misma proteína, obtenidas a partir de diferentes protocolos experimentales. CoDNaS permite conectar sus datos con información fisicoquímica y biológica, permitiendo explorar cuales son los diferentes parámetros que modulan la diversidad conformacional de proteínas. El valor de RMSD, teniendo en cuenta solo los C $\alpha$ , se considera como una medida de la diversidad conformacional. En la tesis, elegimos pares de estructuras de la misma proteína cuya única diferencia en la estimación de la estructura es la presencia o ausencia de ligando. Ante la existencia de más de una estructura cristalizada se priorizó que la diferencia estructural entre conformeros sea máxima de acuerdo al valor de RMSD.

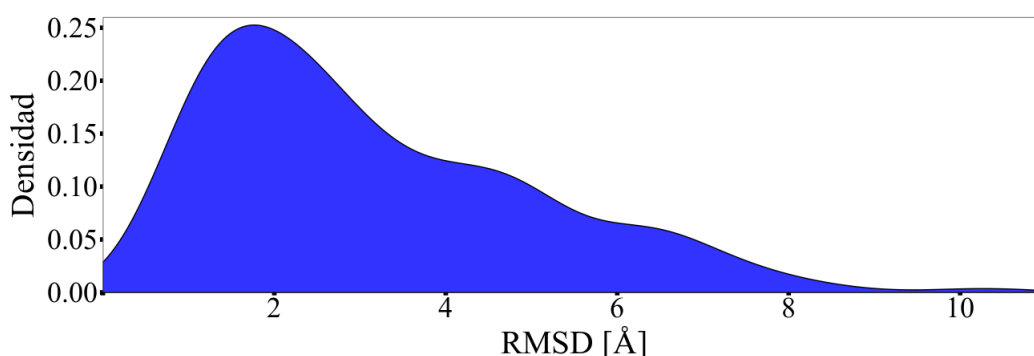
Aplicamos varios filtros en el conjunto de datos original para obtener un set bien curado: (i) solo utilizamos estructuras con una resolución  $< 4 \text{ \AA}$ , (ii) las estructuras no podían contener huecos o regiones sin definir dentro del archivo de pdb, (iii) la correlación entre los factores B experimentales y teóricos debía ser mayor a 0,4 (ver Capítulo 2, sección 2.3.1, Ec. 2.64), (iv) para el análisis de la conservación evolutiva, dentro de la base de datos HSSP, debíamos contar con más de 100 estructuras homólogas por proteínas.

Se obtuvo un total de 188 pares de estructuras sin y con ligado. La **Figura 4.1** muestra la distribución de los valores de RMSD obtenidos sobre todos los pares del conjunto de datos. La lista de los pares con su correspondiente código pdb se proporciona en el Apéndice B.

#### 4.1.2 Identificación de residuos en posiciones claves

Varios estudios previos han demostrado que los cambios conformacionales asociados a la unión del ligando suelen estar dominados por sólo unos pocos modos normales de baja frecuencia [Tama 2001, Zheng 2006, Atilgan 2001, Bahar 2005, Delarue 2002]. En nuestro caso, el número de modos normales que componen el subespacio  $S$  asociado al cambio conformacional se definió a partir del cálculo del número de participación  $M$  (ver Capítulo 3, sección 3.2.1, Ec. 3.6). La **Figura 4.2a** muestra la distribución de la fracción de modos normales involucrados en el cambio conformacional calculado como  $M/3N-6$  obtenidos sobre todos los pares de estructuras en nuestro conjunto de

proteínas. El valor promedio es  $0,15 \pm 0,09$ , confirmando que mediante un número bajo de modos normales se puede representar el espacio vibracional de una proteína asociado a la unión del ligando. Sin embargo, esto no se cumple siempre [Petrone 2006], ya que se pueden apreciar en la distribución valores mayores a 0,5. La **Figura 4.2a** también muestra la comparación entre las distribuciones de valores  $M/3N-6$  y  $M_B/3N-6$  (ver Capítulo 3, sección 3.2.3). Se puede apreciar, que se necesita un mayor número modos normales para lograr una buena descripción de los patrones de flexibilidad, respecto al número de modos necesarios para describir el cambio asociado a la unión del ligando.



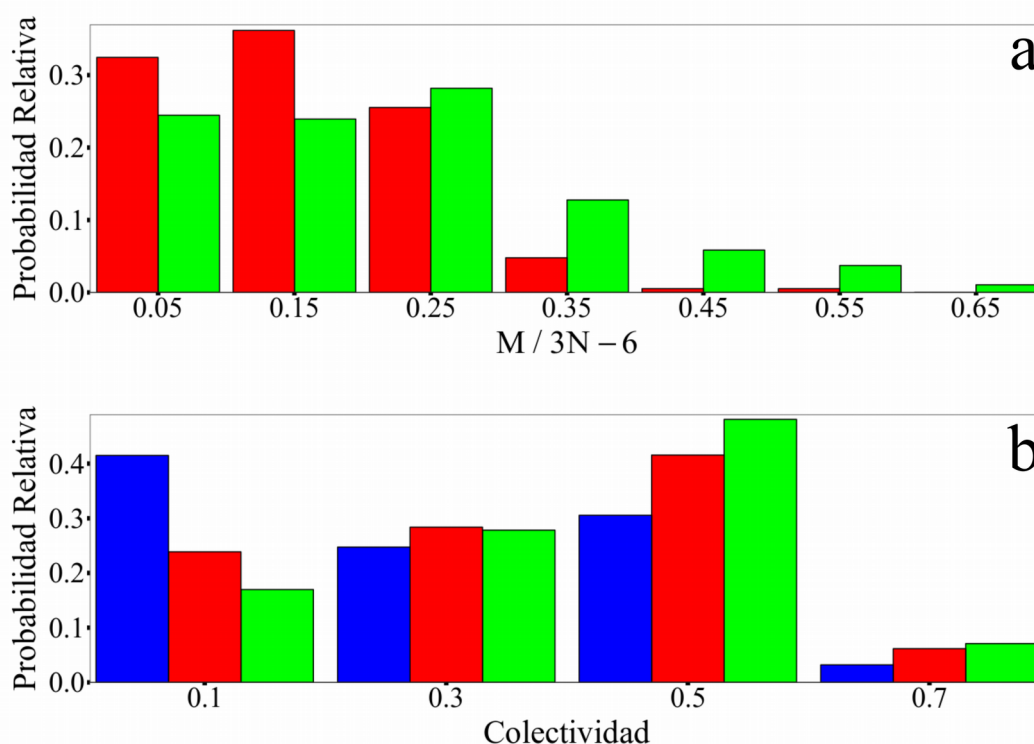
**Figura 4.1 Distribución de los valores de RMSD para todos los pares del set de estudio.** Realizamos la distribución de Densidad mediante el paquete ggplot2 de R y el *software Rstudio* [Wickham 2016]. La aplicación del suavizado tipo núcleo (*kernel smoothing*) a la distribución de frecuencia de un conjunto de datos produce la estimación de Densidad Núcleo (*Kernel Density Estimate*). Esta es una alternativa útil al histograma ya que que no requiere la elección arbitraria del ancho de la ventana y presenta un resultado suavizado.

La composición de los subespacios  $S$  (ver Capítulo 3, sección 3.2.1) de cada par del set de estudio se muestra en la **Figura 4.2b** como distribución del grado de colectividad  $\kappa_k$  (ver Capítulo 2, sección 2.3.3). En general, los modos normales implicados en el cambio conformacional representan movimientos vibratorios más colectivos que el resto de modos. El máximo de la distribución en 0,5 indica que, en promedio, la mitad de los residuos participan en los desplazamientos concertados descritos por cada uno de estos modos. Además, la **Figura 4.2b** muestra la distribución del grado de colectividad para modos que pertenecen al subespacio  $S_B$  (ver Capítulo 3, sección 3.2.3). La comparación con los modos normales que participan en el cambio conformacional indica que los modos involucrados en el patrón de flexibilidad son sólo ligeramente mas colectivos. Este resultado concuerda con estudios previos que demuestran que los cambios conformacionales se asocian comúnmente a los modos normales de baja frecuencia [Tama 2001, Petrone 2006]. A pesar de ello, la participación en el cambio conformacional de modos normales localizados está lejos de ser insignificante [Petrone 2006].

También hemos explorado la dependencia del subespacio  $S$ , asociado a la unión del ligando, con el RMSD entre conformeros y el tamaño de la proteína. Para ello, hemos considerado el número de participación,  $M$  y el valor promedio de colectividad de los modos que pertenecen al subespacio  $S$ .



Obtuvimos coeficientes de correlación de Spearman de 0,03 ( $p$ -value = 0,007) y -0,14 ( $p$ -value <  $2,2 \times 10^{-18}$ ), entre el promedio de la colectividad y el RMSD y el tamaño de las proteínas respectivamente. También, encontramos una correlación insignificante de 0,09 ( $p$ -value = 0,23) entre el número  $M$  y el RMSD. Sólo se obtuvo una correlación significativa de 0,49 ( $p$ -value =  $7,3 \times 10^{-13}$ ) entre  $M$  y el tamaño de la proteína.

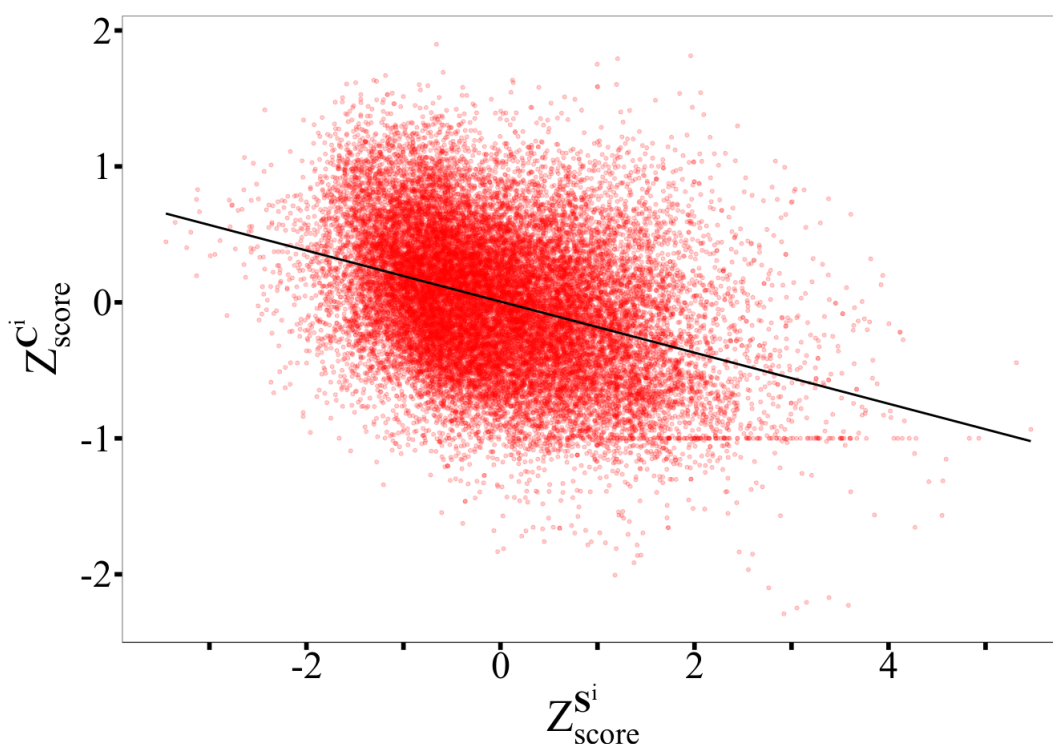


**Figura 4.2** (a) Distribución de la fracción de modos normales involucrados en el cambio conformacional calculado como  $M/3N$  (rojo), y la fracción de modos normales que participan en el patrón de flexibilidad calculado como  $M_B/3N$  (verde); obtenido sobre todos los pares de estructuras en nuestro conjunto de datos. (b) Distribución del grado de colectividad,  $\kappa_k$ , para cada modo normal que participa en el cambio conformacional (rojo), cada modo normal que participa en el perfil de flexibilidad (factor B) (verde) y para el resto de los modos normales (azul).

Como se mencionó antes, la diversidad conformacional del estado nativo desempeña un papel central en la modulación de la función de la proteína. La coexistencia de conformeros con diferentes afinidades por el ligando en un equilibrio dinámico es la base para el modelo de pre-equilibrio. Los movimientos internos del conformero sin ligando deben contener los cambios conformacionales y garantizar la transición del conformero libre al conformero unido. Por lo tanto, el efecto de las mutaciones en el subespacio  $S$  de modos normales asociados a la unión del ligando debe correlacionar con la conservación evolutiva de los sitios correspondientes.

Para investigar esto, la **Figura 4.3** muestra la relación entre el efecto de una mutación puntual sobre las vibraciones involucradas en la unión al ligando  $Z_{score}^S$  (definido en el Capítulo 3, sección 3.2.2) y la conservación evolutiva  $Z_{score}^C$  (definido en el Capítulo 3, sección 3.3.2). Obtuvimos un

coeficiente de correlación de Spearman  $\rho = -0,36$  con un  $p$ -value menor a  $2,2 \times 10^{-16}$ . Esto significa que cuanto mayor es el impacto producido por las mutaciones sitio-específicas sobre el subespacio, mayor es la conservación evolutiva específica del sitio.



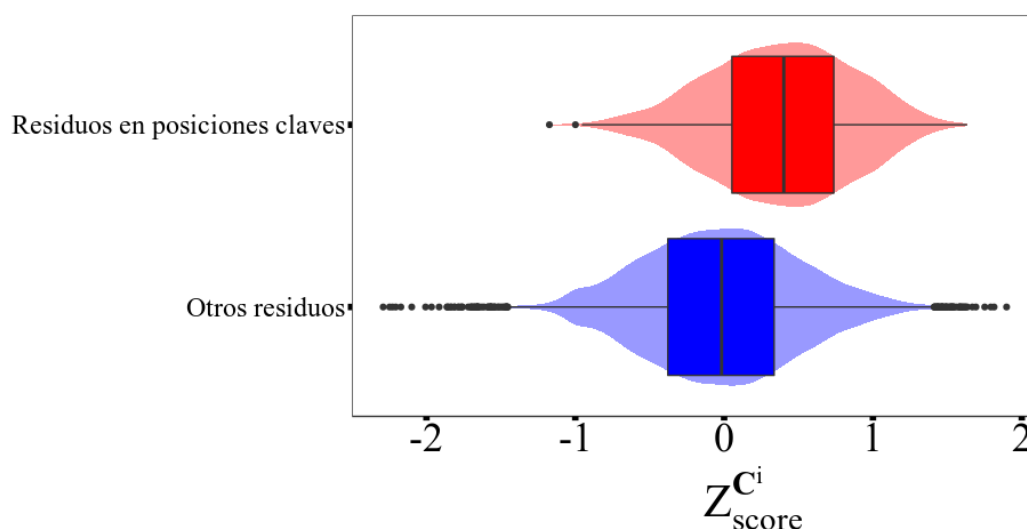
**Figura 4.3 Efecto de las mutaciones sobre las vibraciones involucradas en la unión al ligando ( $Z_{score}^{Si}$ ) frente a la conservación evolutiva ( $Z_{score}^{Ci}$ ).** Se incluye la línea de regresión lineal y el coeficiente de correlación  $\rho$  se muestra en la esquina superior derecha.

Con el fin de analizar la dependencia de  $Z_{score}^{Si}$  con el tamaño de las proteínas y el RMSD entre conformeros, consideramos distintos subconjuntos de nuestro set de proteínas. Los subconjuntos se separaron según: (a)  $RMSD < \overline{RMSD}$ ; (b)  $RMSD > \overline{RMSD}$ ; (c)  $tamaño < \overline{tamaño}$ ; (d)  $tamaño > \overline{tamaño}$ , siendo el  $\overline{RMSD} = 2,0 \text{ \AA}$  (Figura 4.1) y  $\overline{tamaño} = 80$ . Obtuvimos coeficientes de correlación de Spearman de -0,32, -0,35, -0,30 y -0,34 para los subconjuntos (a) - (d), respectivamente. En todos los casos, se obtuvo un  $p$ -value  $< 2,2 \times 10^{-16}$ . A pesar de que nuestro resultado no está fuertemente influenciado ni por el tamaño de la proteína, ni por el RMSD, se observa una ligera dependencia. Es decir, observamos mejores correlaciones para proteínas grandes y proteínas con mayores distorsiones estructurales (RMSD) producidas por la unión del ligando.

Nuestra metodología nos permite identificar las posiciones claves para la conservación evolutiva de la diversidad conformacional de la proteína necesaria para la unión del ligando. Es decir, podemos determinar las posiciones cuyas mutaciones alteran en mayor medida el subespacio  $S$  que contiene los modos normales que describen la transición conformacional entre la estructura sin ligando y la estructura unida al ligando. Para cada par de estructura con y sin ligando, definimos las

posiciones claves como aquellas dentro del 5% de los valores más bajos de  $Z_{score}^S$ . El cambio de la definición de las posiciones usando el 1% o 10% no modifica cualitativamente nuestro resultado.

En la **Figura 4.4**, analizamos la conservación evolutiva de estos residuos en posiciones claves en relación con el resto de los residuos. La distribución de  $Z_{score}^C$ , para los residuos en posiciones claves se encuentra desplazada hacia valores mayores, lo que indica que estos residuos se encuentran conservados evolutivamente. La diferencia entre ambas distribuciones se validó estadísticamente realizando una Prueba o Test de Kolmogorov-Smirnov, donde se obtuvo un valor de 0,31 con un  $p-value = 2,2 \times 10^{-16}$ .



**Figura 4.4** Distribución de los valores de conservación evolutiva,  $Z_{score}^C$ , obtenida para los residuos en posiciones claves (rojo) y para el resto de los residuos (azul). Los rectángulos inferiores y superiores del gráfico de caja o *box-plot*, corresponden al primer y tercer cuartil, y la banda negra dentro de la caja es la mediana (segundo cuartil). El *violin-plot*, debajo de la representación de caja, muestra la distribución de la variable.

En este punto es importante destacar que el objetivo del presente trabajo no es explicar completamente la conservación evolutiva de los residuos a través de su relevancia en la diversidad conformacional de proteínas. Trabajos anteriores demuestran que la conservación evolutiva de la secuencia es consecuencia de múltiples factores como la estructura, la dinámica y / o características funcionales [Liu 2012, Morcos 2013, Juritz 2013, Parisi 2015].

Los resultados presentados en las **Figuras 4.3** y **4.4** enfatizan que la diversidad conformacional del estado nativo de una proteína es sólo uno de los muchos aspectos que modulan la función de la misma. Por este motivo, los residuos dinámicamente relevantes, o las regiones asociadas a la diversidad conformacional, se encuentran evolutivamente más conservados que otros residuos. A pesar de la existencia de múltiples fuentes de conservación evolutiva, destacamos la correlación entre nuestra medida del efecto de mutaciones sobre dinámica involucrada con la unión al ligando y la divergencia evolutiva. Los  $p-values$  obtenidos en el análisis de las **Figuras 4.3** y **4.4** cuantifican la

significación estadística de nuestros resultados, lo que indica que los datos observados son inconsistentes con el supuesto de que la hipótesis nula es verdadera.

### 4.1.3 Caracterización de las posiciones claves

Realizamos diferentes estudios para caracterizar los residuos situados en posiciones claves. En primer lugar, se analizó la incidencia de los diferentes tipos de aminoácidos, definida como

$$I_{\alpha} = \frac{v_{\alpha}^{pc}}{v_{\alpha}} \quad (\text{Ec. 4.1})$$

donde  $v_{\alpha}^{pc}$  es la frecuencia en la que el aminoácido  $\alpha$  es detectado como un residuo en una posición clave, y  $v_{\alpha}$  la frecuencia correspondiente del mismo aminoácido  $\alpha$  en el resto de las posiciones. Un valor de  $I_{\alpha} > 1$  indica que la frecuencia del aminoácido  $\alpha$  dentro de las posiciones claves es mayor en relación con su frecuencia observada en el conjunto de proteínas estudiadas. La **Tabla 4.1** muestra estos valores.

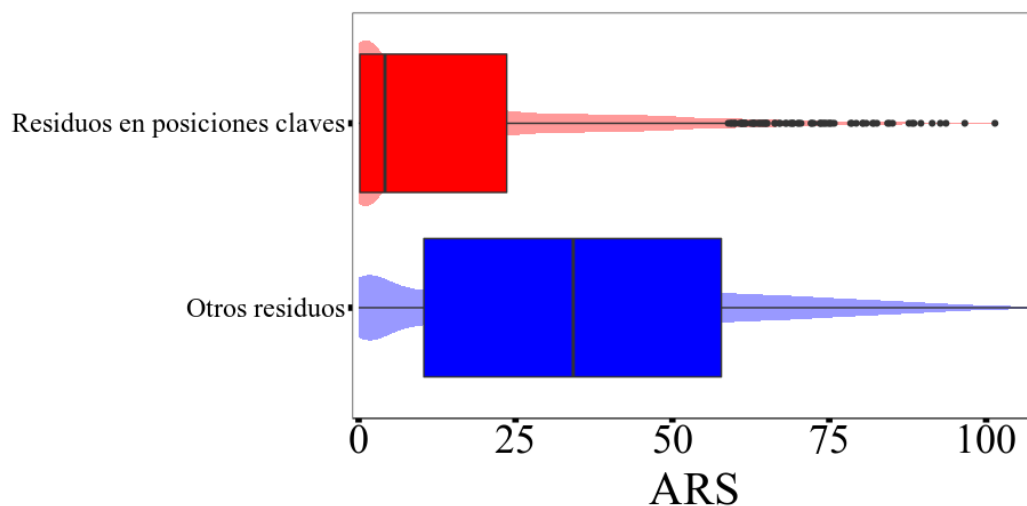
**Tabla 4.1** Incidencia de los residuos en posiciones claves.

Cys	2.412	Tyr	0.856
Trp	1.626	Gly	0.837
Val	1.625	Gln	0.807
Ile	1.577	His	0.803
Phe	1.569	Ser	0.792
Leu	1.432	Asp	0.706
Met	1.197	Glu	0.685
Asn	0.932	Lys	0.669
Ala	0.889	Arg	0.527
Thr	0.883	Pro	0.438

Cys presenta el mayor valor de  $I_{\alpha}$  debido principalmente a la importancia de los enlaces de puentes disulfuro. Luego, aminoácidos no polares, como Val, Ile, Leu, Met, Trp y Phe, están entre los residuos observados con mayor frecuencia en las posiciones claves detectadas. Esto está de acuerdo con la comparación de la distribución del Área Relativa de accesibilidad al Solvente (ARS), calculada usando el programa NACCESS, para residuos en posiciones claves respecto al resto de residuos en la proteína (**Figura 4.5**). A partir de los valores de ARS, se encontró que las posiciones clave se encuentran, en general, en el interior de la estructura proteica.

En este punto, resulta interesante analizar las correlaciones entre  $Z_{score}^s$ ,  $Z_{score}^c$ , ARS y el número de contactos entre aminoácidos (Grado o *Degree*), para cada residuo de nuestro conjunto de proteínas calculado utilizando el programa RING. Por un lado, el coeficiente de correlación de Pearson entre los valores de  $Z_{score}^s$  y ARS dió como resultado un valor de 0,48, mientras que el valor correspondiente entre  $Z_{score}^s$  y el *Degree* fue de -0,46. Por otro lado, se obtuvieron correlaciones de

-0,27 entre  $Z_{score}^c$  y ARS; y de 0,23 entre  $Z_{score}^c$  y el Grado o *Degree*. Se ve claramente que mientras ARS y el número de contactos por residuo correlacionan fuertemente con el valor de  $Z_{score}^s$ , ambos se correlacionan débilmente con el valor de  $Z_{score}^c$ . Considerando nuestra correlación de -0,36 entre  $Z_{score}^s$  y  $Z_{score}^c$  concluimos que este valor no puede explicarse por una simple evaluación de la ARS y el número de contactos por residuo.



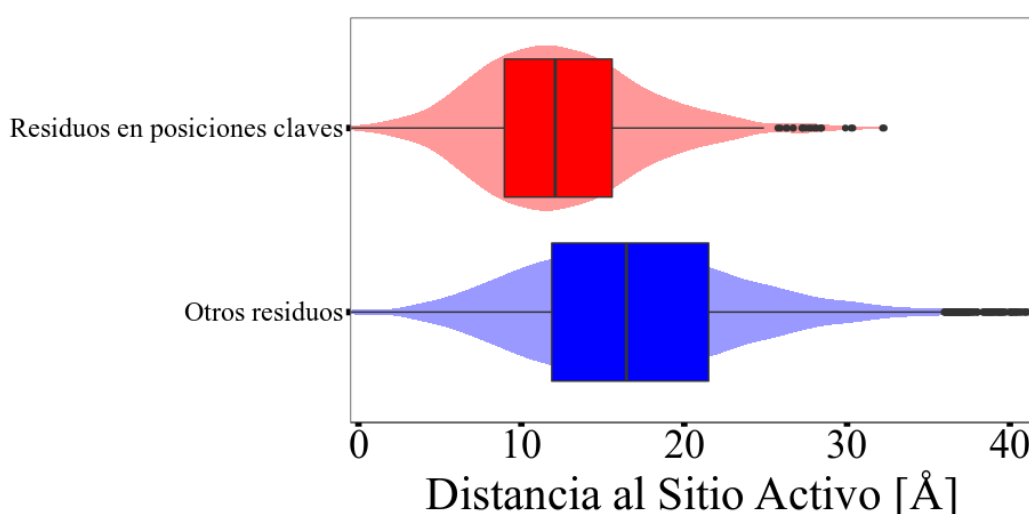
**Figura 4.5** Área Relativa de accesibilidad al Solvente (ARS) para los residuos en posiciones claves (rojo) y el resto de las posiciones en la proteína (azul).

También analizamos la relación entre el valor de  $Z_{score}^s$  o  $Z_{score}^c$ , y el RMSD por residuo ( $RMSD^i$ ) entre las estructuras con y sin ligando. La fuerte correlación de 0,4 entre el valor de  $Z_{score}^s$  y el  $RMSD^i$  indica que las mutaciones en posiciones con poca divergencia estructural entre los conformeros probablemente tengan un fuerte impacto sobre las vibraciones asociadas al cambio conformacional generado por la unión del ligando. Sin embargo, se obtuvo una correlación muy débil de -0,16 entre el valor de  $Z_{score}^c$  y el  $RMSD^i$ . Es decir, no todos los residuos con baja movilidad relativa durante el cambio conformacional se encuentran conservados evolutivamente.

Utilizamos la base de datos BioLip [Yang 2013] para obtener información relativa al sitio activo de cada proteína en el set de estudio. A partir de esta información, se calculó la distancia de los residuos en posiciones claves respecto al centro de masa del sitio activo de la proteína. La **Figura 4.6** muestra la distribución de estas distancias para los residuos considerados claves y para el resto de residuos de cada estructura. Se determinó que, en general, los residuos en posiciones claves están más cerca del sitio activo sin formar parte de él; solo el 10% de las posiciones claves corresponden a residuos que forman parte del sitio activo.

Estudios previos han demostrado que los residuos del sitio activo están frecuentemente relacionados con residuos que provocan cambios conformacionales asociados a la unión del ligando

[Ming 2005, Yang 2005, Ming 2006, Zheng 2009b]. El coeficiente de correlación de Pearson entre los valores de  $Z_{score}^S$  y la distancia al centro de masa de los sitios activos es de 0,39 con un  $p$ -value de  $2,2 \times 10^{-16}$ . Las transiciones conformacionales deberían introducir cambios en el sitio activo que conducen a cambios significativos en la afinidad para el ligando. A pesar de ello, los residuos del sitio activo sólo comprenden una pequeña fracción de los residuos claves predichos; esto concuerda con los resultados previos obtenidos por Zheng et al. [Zheng 2009b]. Por lo tanto, la mayoría de los residuos en posiciones claves, conservados evolutivamente, no están directamente asociados a la catálisis enzimática.



**Figura 4.6** Distribución de las distancias de los residuos en posiciones claves (rojo) y del resto de los residuos (azul) al centro de masa del sitio activo de la proteína.

A continuación, analizamos los elementos de estructura secundaria (SSE, por sus siglas en inglés) del que forman parte los residuos en posiciones claves. Para esto, utilizamos el programa DSSP [Kabsch 1983, Touw 2015] que define ocho tipos de elementos de estructura secundaria: **H** (hélice- $\alpha$ ), **G** (hélice-3<sub>10</sub>), **I** (hélice- $\pi$ ), **B** (hoja- $\beta$  unitaria), **E** (hoja- $\beta$  extendida), **T** ( $n$ -turn), **S** ( $bend$ ) y **N** ( $loop$ ). La **Tabla 4.2** muestra los valores de la incidencia,  $I_{SSE,X}$ , de los residuos claves en los diferentes SSE, definida como

$$I_{SSE,X} = \frac{v_{SSE,X}^{pc}}{v_{SSE,X}} \quad (\text{Ec. 4.2})$$

donde  $v_{SSE,X}^{pc}$  es la frecuencia de las posiciones claves en el SSE X; con  $X = H, B, E, G, I, T$  o  $S$ . Y  $v_{SSE,X}$  la frecuencia correspondiente en el resto de los residuos. Un valor de  $I_{SSE,X} > 1$  indica que la frecuencia con la que los residuos en posiciones claves se encuentran dentro del SSE X es mayor en relación con su frecuencia observada en el conjunto de proteínas estudiadas. Se determinó que las posiciones claves son más frecuentemente en elementos de estructura del tipo hoja- $\beta$  extendida y hélice- $\alpha$ .

**Tabla 4.2** Incidencia de diferentes residuos en posiciones claves relacionado con los SSE.

E	2.087
H	1.146
B	1.019
G	0.608
N	0.547
S	0.355
T	0.349
I	0.000

Nuestra medida de las distorsiones estructurales introducidas por la unión al ligando está dada por el vector diferencia  $\mathbf{V}$ , cuyos elementos están ponderados por los factores B correspondientes (ver Capítulo 3, sección 3.2.1). De esta forma, evitamos que nuestros resultados puedan verse sesgados por cualquier distorsión estructural que no esté directamente relacionada con la unión del ligando. *Loops* y otras regiones flexibles se descartan inherentemente mientras se resaltan dominios y regiones de tipo bisagra (*hinge*). Por lo tanto, se ponderó particularmente dos tipos de residuos con bajos factores B. Por un lado, se destacaron los residuos que presentan grandes contribuciones al cambio conformacional. Estos residuos experimentan grandes distorsiones estructurales tras la unión del ligando sin presentar una flexibilidad significativa o incertidumbres en sus coordenadas en el conjunto de conformaciones del estado nativo de la proteína sin ligando. Estos residuos son arrastrados por los movimientos colectivos que se desencadenan cuando el conjunto de posiciones en equilibrio cambian hacia la conformación con ligando de la proteína. Por otro lado, también se destacaron residuos que apenas se mueven entre las conformaciones sin y con ligando. Estos residuos se localizan en regiones bien definidas del tipo bisagra, sin conectar elementos de estructura secundaria (SSE) o dominios de forma secuencial, como *loops*; participando quizás como pivotes a través de contactos entre SSE o entre dominios. De esta forma las mutaciones introducidas en este último tipo de residuos pueden afectar fuertemente los movimientos vibratoriales implicados en cambios conformacionales por la unión del ligando. Para confirmar esto analizamos la incidencia de contactos inter-SSE definida como

$$I_{\text{SSE},X} = \frac{v_{\text{inter-SSE},X,Y}^{\text{pc}}}{v_{\text{inter-SSE},X,Y}} \quad (\text{Ec. 4.3})$$

donde  $v_{\text{inter-SSE},X,Y}^{\text{pc}}$  es la frecuencia de los residuos en posiciones claves que participan en contactos SSE, entre los residuos localizados en X, siendo  $X = \mathbf{E}$  y  $\mathbf{H}$ ; e  $Y = \mathbf{E}, \mathbf{B}, \mathbf{H}, \mathbf{G}, \mathbf{S}, \mathbf{T}, \mathbf{N}$  y  $\mathbf{I}$ ; y  $v_{\text{inter-SSE},X,Y}$  la frecuencia correspondiente en el resto de los residuos. La **Tabla 4.3** muestra estos valores. Observamos una gran incidencia de contactos inter-SSE de residuos en posiciones clave,

confirmando nuestra hipótesis de que estos residuos participan de contactos entre SSE bien estructuradas, como hélice- $\alpha$  (**H**), hoja- $\beta$  extendida (**E**).

**Tabla 4.3** Incidencia de residuos en posiciones claves que participan en contactos inter-SSE.

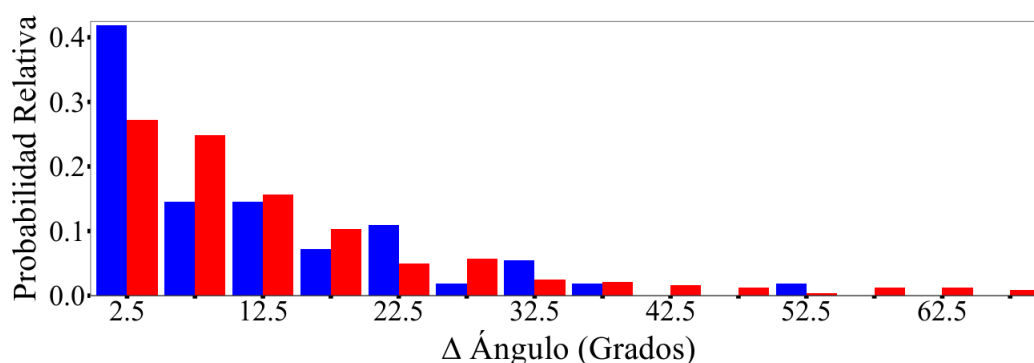
E-H	2.83096	H-B	2.70408
E-N	1.92404	H-E	2.19460
E-E	1.89671	H-H	1.62143
E-B	1.88451	H-S	0.86975
E-S	1.71122	H-T	0.83943
E-G	1.56960	H-G	0.81984
E-T	1.03616	H-N	0.76340
E-I	0.00000	H-I	0.00000

Nuestro análisis no depende de la información de la secuencia de proteínas, ni del análisis de la conservación evolutiva y el mapeo estructural de la información filogenética como métodos de rastreo evolutivo. No intentamos competir con métodos desarrollados previamente para la predicción de sitios de unión a ligandos [Lichtarge 1996, Glaser 2003]. La funcionalidad de los residuos en posiciones claves no está necesariamente relacionada con las interacciones proteína-ligando o la actividad catalítica, sino con la diversidad conformacional asociada a la unión al ligando. Por lo tanto, no se espera que todas las mutaciones que presenten efectos sobre la afinidad por el sustrato y la actividad catalítica puedan asociarse a nuestra definición de residuos en posiciones claves que involucran aminoácidos asociados a un aspecto muy particular de la funcionalidad de la proteína, que son las vibraciones asociadas a distorsiones estructurales introducidas por unión del ligando. Para analizar eso, hemos comparado nuestros resultados con información de datos experimentales provistos por la base de datos UniProt [UniProt 2015]. Esta base de datos proporciona una visión general y completa de la información disponible sobre proteínas, incluida información relacionada con la función, la actividad catalítica y mutaciones con efectos reportados sobre la afinidad por el sustrato y la actividad catalítica. UniProt contiene información sobre 185 mutaciones para 43 proteínas de nuestro set; solo 13 de estas mutaciones, distribuidas en 11 proteínas corresponden a residuos en posiciones claves. Este resultado es esperado ya que, como se informó anteriormente, solo ~10% de los residuos en posiciones claves corresponden a residuos del sitio activo. Es decir, nuestros residuos claves predichos no coinciden con los residuos catalíticos. Teniendo en cuenta que nuestro procedimiento permite la identificación de regiones espaciales claves en lugar de residuos individuales, hemos ampliado nuestro análisis para incluir residuos que están en contacto directo con los residuos en posiciones claves. Para esto utilizamos información obtenida mediante el programa RING 2.0 [Piovesan 2016]. Encontramos que 98 de las mutaciones informadas en UniProt concuerdan con nuestros hallazgos. Es decir, el 53% de las mutaciones con evidencia experimental relacionada con



la afinidad del ligando y la catálisis enzimática coinciden o están en estrecho contacto con los residuos en posiciones claves para el mantenimiento de la diversidad conformacional asociada a la unión del ligando.

Con el fin de analizar más a fondo el papel de las posiciones clave como pivotes entre SSE, llevamos a cabo una metodología previamente desarrollada para investigar los movimientos de dominios entre confórmers con y sin ligando [Brylinski 2008]. Considerando un residuo clave que pertenece a un SSE X y que realiza un contacto inter-SSE con un SSE Y, se calculó la diferencia entre los ángulos formados por cada eje de inercia correspondiente al SSE X y el mismo eje de la SSE Y, para las estructuras con y sin ligando. Elegimos la mayor diferencia entre estos ángulos como una medida cuantitativa de las diferencias de orientación relativa de las SSE. (ver Capítulo 3, sección 3.3.6 o [Browner 1992, Brylinski 2008]). Nuestros resultados, presentados en la **Figura 4.7**, indican que las SSE conectadas a través de un residuo clave presentan movimientos angulares más grandes en comparación con aquellas en las que ningún residuo clave participa en el contacto entre SSE.

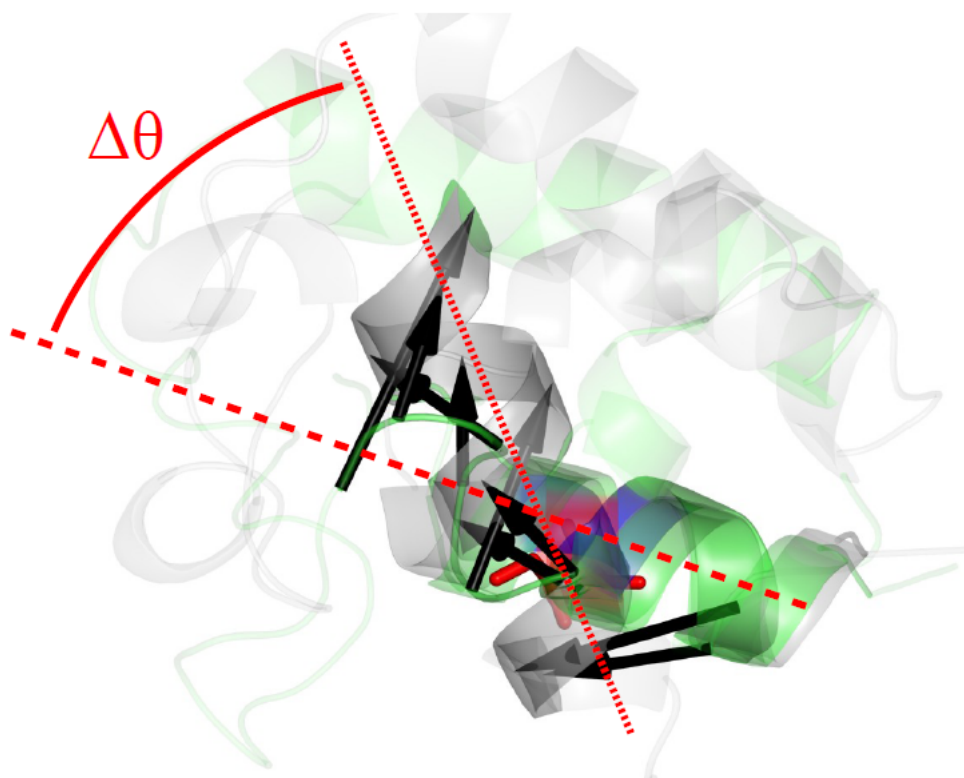


**Figura 4.7** Distribución de la mayor diferencia entre los ángulos formados por los ejes de inercia de SSE individuales, conectados a través de un residuo en posición clave (rojo), y a través de otros residuos (azul).

Es interesante observar que la **Figura 4.7** relaciona los residuos en posiciones claves con distorsiones estructurales generadas por la unión del ligando. Las diferencias de los movimientos angulares se obtienen directamente de las coordenadas PDB de las estructuras. Por lo tanto, el uso de un método de “grano grueso”, basado en la descripción de la proteína como una red elástica de carbonos  $\alpha$ , no sesga estos desplazamientos relativos entre las SSE.

Para aclarar el papel que los contactos inter-SSE mediados por residuos en posiciones claves tienen en la transición conformacional mediada por la unión del ligando, la **Figura 4.8** muestra el caso de la Proteína Transportadora de Acilo (ACP por sus siglas en inglés, *Acyl-Carrier Protein*) de *Escherichia coli* como ejemplo de un residuo en posición clave que participa de un contacto inter-SSE H-H. ACP es una proteína de 77 aminoácidos implicada en la síntesis de ácidos grasos (códigos pdb 1ACP y 2FAE, para estructuras sin y con ligando, respectivamente [Kim 1990, Roujeinikova 2007]). La **Figura 4.8** muestra el residuo clave I69 localizado en la hélice- $\alpha$  H4 (Q66-H75). Este residuo

interactúa con V7 perteneciente a la hélice- $\alpha$  H1 (E4-Q14). Las flechas indican las direcciones en las que los residuos se mueven durante el cambio conformacional al unirse el ligando. El ángulo  $\Delta\theta$  indica el cambio en la orientación relativa entre H1 y H4, con I69 participando como pivote a través del contacto inter-SSE con V7.

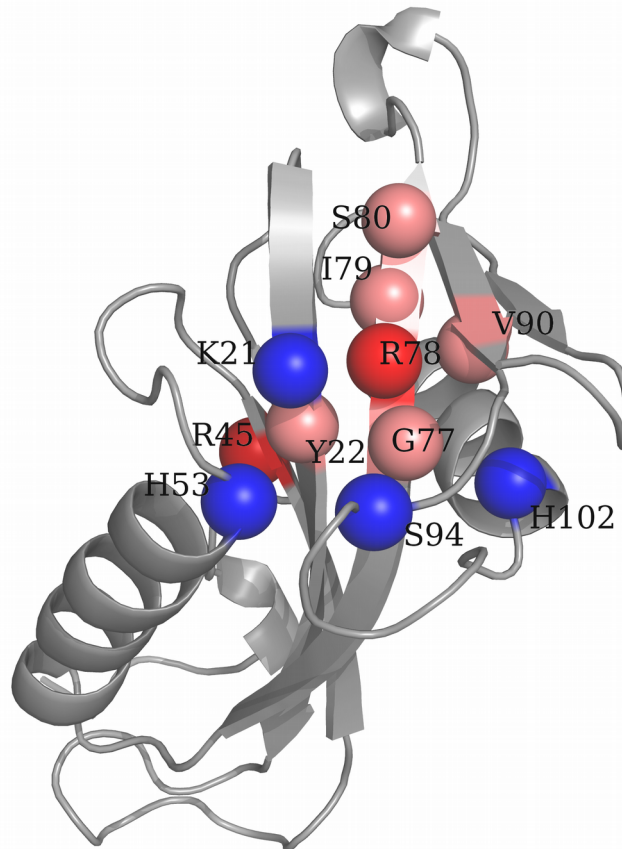


**Figura 4.8 Cambio en la orientación relativa entre dos  $\alpha$ -hélices de la proteína transportadora de acilo (ACP) de *Escherichia coli*.** La estructura sin ligando (código pdb: 1ACP, cadena A) y la estructura con el ligando unido (código pdb: 2FAE, cadena B) se representan en verde y gris, respectivamente. El residuo de posición clave I69 (rojo) participa de un contacto intra-SSE H-H con V7. Las flechas indican las direcciones en las que los residuos se mueven durante el cambio conformacional al unirse al ligando.  $\Delta\theta = \theta - \theta'$ , siendo  $\theta$  el ángulo entre la hélice- $\alpha$  H4 (Q66-H75) y hélice- $\alpha$  H1 (E4-Q14) para la estructura sin ligando y  $\theta'$  el ángulo correspondiente a la estructura con ligando.

#### 4.1.4 Ejemplos

Con el objeto de ejemplificar nuestro resultado, presentamos un par de casos seleccionados de nuestro set. El primer ejemplo es la proteína Fosfohistidina Fosfatasa 1 (PHPT1, por sus siglas en inglés, *Phosphohistidine phosphatase 1*), cuyos códigos pdb son 2AI6 y 2OZWf para las estructuras sin y con ligando, respectivamente [Gong 2009]. Esta enzima de 125 aminoácidos desempeña un papel importante en la transducción de señales y otras funciones celulares; el sitio activo está ubicado entre la hélice- $\alpha$  H1 y el *loop* L5. La **Figura 4.9** muestra la estructura PHPT1 en su forma libre de ligando. Se han identificado siete residuos en posiciones claves, es decir sitios dinámicamente importantes que median el cambio conformacional de unión a ligando: Y22, R45, G77, R78, I79, S80, V90. Todos conservados evolutivamente. De acuerdo con la información proporcionada por la base de datos

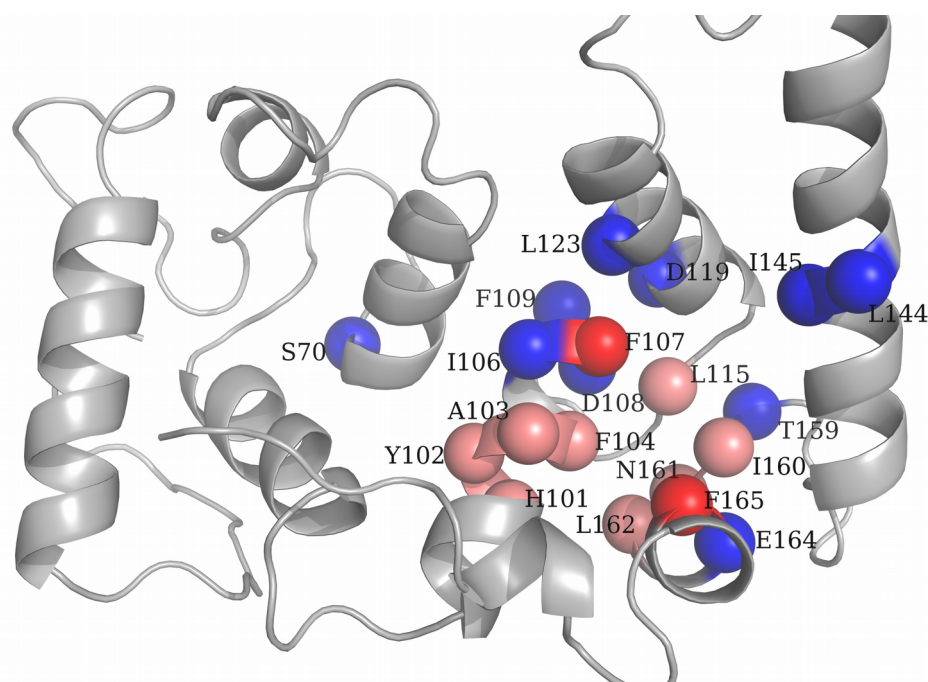
UniProt, las mutaciones en K21, R45, H53, R78, S94 y H102 tienen efectos sobre la afinidad por el sustrato y la actividad catalítica. En la **Figura 4.9**, se indican los residuos en posiciones claves y los residuos identificados por UniProt con mutaciones. Como se puede ver, la mayoría de los residuos en posiciones claves corresponden o están en contacto con residuos cuyas mutaciones alteran la afinidad por el sustrato y la actividad catalítica, confirmadas experimentalmente.



**Figura 4.9 Estructura sin ligando de la proteína Fosfohistidina Fosfatasa 1 (PHPT1, código pdb 2A16).** Los residuos se colorearon de la siguiente manera: residuos en posiciones claves (rosa), residuos identificados por UniProt, cuyas mutaciones afectan la afinidad por el sustrato y la actividad catalítica (azul) y los residuos en posiciones claves e identificados también por UniProt (rojo).

El segundo ejemplo que ilustra nuestros hallazgos corresponde a la Proteína 1 de Unión a Calcio e Integrina (CIB1, por sus siglas en inglés, *Calcium and integrin-binding protein 1*), código pdb 1DGU para la estructura sin ligando y 1Y1A para el conformero unido al ligando [Hwang 2000, Blamey 2005]. Esta enzima tiene 183 residuos; y se une al dominio citoplasmático  $\alpha$ IIB de 20 residuos de la integrina  $\alpha$ IIB $\beta$ 3 plaquetaria. Actúa como un regulador de señalización global en una amplia variedad de proteínas en las células, principalmente en las plaquetas. La **Figura 4.10** muestra la estructura sin ligando de CIB1. Se han seleccionado diez residuos en posiciones claves, todos conservados evolutivamente, H101, Y102, A103, F104, F107, L115, I160, N161, L162, F165. Como se ha señalado anteriormente, nuestro procedimiento permite la identificación de regiones espaciales H101-F107 e I160-F165 en lugar de residuos individuales. Los residuos que presentan evidencia

experimental de mutaciones que afectan la unión a ligando y la actividad catalítica son: S78, I106-F109, D119, L123, L144, I145, T159, E164 y F165. Todos estos residuos se indican en la **Figura 4.10**.



**Figura 4.10 Estructura sin ligando de la Proteína 1 de Unión a Calcio e Integrina (CIB1, código pdb 1DGU).** Los residuos se colorearon de la siguiente manera: residuos en posiciones claves (rosa), residuos identificados por UniProt, cuyas mutaciones afectan la afinidad por el sustrato y la actividad catalítica (azul) y los residuos en posiciones claves e identificados también por UniProt (rojo).

Es importante destacar que los efectos sobre la afinidad por el sustrato y la actividad catalítica no están necesariamente asociados a los efectos sobre la diversidad conformacional de la proteína. Nuestros residuos en posiciones claves están asociados a un aspecto muy particular de la funcionalidad de la proteína, las vibraciones asociadas a las distorsiones estructurales introducidas por la unión del ligando. A pesar de eso, ambas regiones espaciales H101-F107 e I160-F165, poseen residuos con evidencia experimental obtenida de UniProt.

Finalmente, el efecto de las mutaciones en los residuos en posiciones claves se ha analizado utilizando el modelo de red de contactos elástico ENCoM (por sus siglas en inglés, *Elastic Network Atom Contact Model*) [Frappier 2014, Frappier 2015] y el algoritmo FoldX [Guerois 2002, Schymkowitz 2005]. Para ambos casos, PHPT1 y CIB1, la diferencia predicha en los cambios de energía libre ( $\Delta\Delta G$ ), evaluadas con ENCoM y FoldX, indican que las mutaciones en los residuos en posiciones claves corresponden a mutaciones desestabilizadoras, es decir, mutaciones que afectan la estabilidad debido a una disminución en la entropía del estado plegado. El  $\Delta\Delta G$  promedio considerando todas las mutaciones posibles en cada residuo en posición clave fue de 2.0 kcal/mol y 1.3 kcal/mol para PHPT1 y CIB, respectivamente. Al seleccionar las mutaciones más desestabilizantes,  $\Delta\Delta G_{\max}$ , para cada residuo de posición clave, obtuvimos un promedio de 4.8 kcal/mol para PHPT1 y de 3.6 kcal/mol para CIB1. Es decir, en ambos casos, los residuos en la

posiciones claves implican aminoácidos cuyas mutaciones pueden afectar drásticamente la estructura de la proteína [Tokuriki 2007].

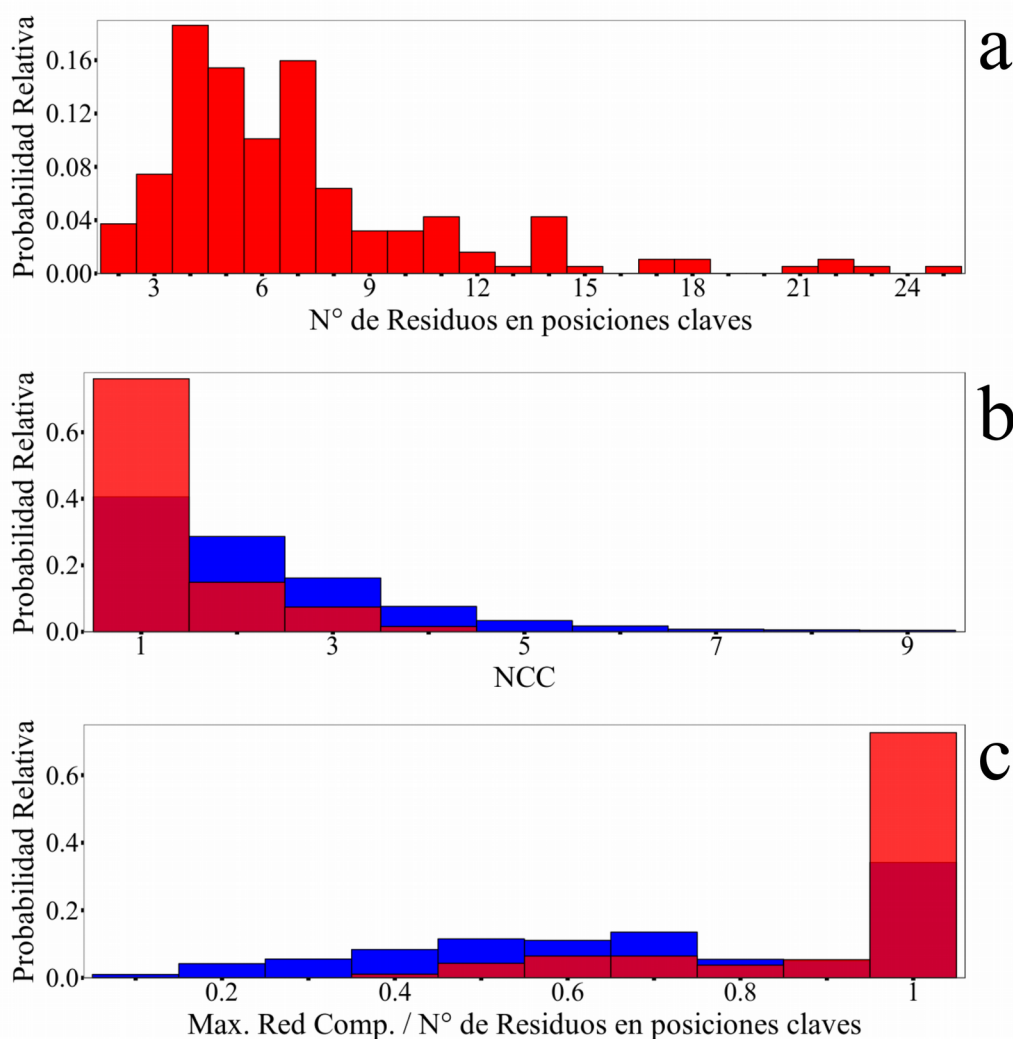
## 4.2 REDES DE RESIDUOS EN POSICIONES CLAVES

### 4.2.1 Redes de residuos, conectividad y robustez

A partir de la red total de cada proteína, filtramos la información de manera tal de generar una subred de interacciones entre residuos en posiciones claves (**Redes de posiciones claves**) y una subred de interacciones entre residuos seleccionados al azar (**Redes random**). En la **Figura 4.11a** presentamos la distribución de residuos en posiciones claves para el conjunto de proteínas estudiadas, esto es relevante para conocer el número posible de residuos que van a conformar las redes.

Uno de nuestros objetivos es determinar si los residuos en posiciones claves se encuentran conectados formando redes de interacciones entre ellos. Para esto, analizamos el Número de Componentes Conectados, NCC (ver Capítulo 3, sección 3.3.8), ya que este simple parámetro sirve para analizar la conectividad de una red. Considerando un SS (ver Capítulo 3, sección 3.3.7) igual a 1, encontramos que, para la mayoría de los pares del set de proteínas los residuos en posiciones claves, se encuentran interconectados. En la **Figura 4.11b** se puede apreciar que las Redes random (azul) se encuentran menos interconectadas, es decir que estos residuos se encuentran separados en un mayor número de *clusters*. Para las Redes de posiciones claves (rojo) el valor promedio de NCC es de  $1,3 \pm 0,7$ , mientras que para las Redes random el valor promedio es de  $2,3 \pm 1,7$ . Estos valores sostienen nuestra suposición de que los residuos en posiciones claves muestran un patrón de conectividad más alto que el resto de los residuos.

A modo de analizar la robustez de las redes estudiamos la proporción de interacciones entre residuos que se mantienen luego del cambio conformacional debido a la unión al ligando. En base a las redes de cada conformero, definimos el número de residuos que se mantienen dentro de un Componente Conectado (ver Capítulo 3, sección 3.3.8) para ambos conformeros y llamamos a esto Máxima Red Compartida. Es decir que si para un conformero sin ligando cinco de sus residuos se encuentran interconectados y para el conformero con ligando solo cuatro están conectados, el valor de Máxima Red Compartida es igual a cuatro. En la **Figura 4.11c**, presentamos la distribución del valor de Máxima Red Compartida dividido el número de residuos en posiciones claves por proteína. En rojo se muestra la distribución correspondientes a las Redes de posiciones claves, para la cual el valor promedio fue de  $0,9 \pm 0,2$ . La distribución obtenida a partir de las Redes random se presentó en azul y el valor promedio de la misma fue de  $0,7 \pm 0,3$ . Estos valores promedios indican que las Redes de posiciones claves se conservan en mayor medida luego del cambio conformacional asociado con la unión al ligando.



**Figura 4.11** (a) Distribución del número de residuos en posiciones claves para cada proteína del set de estudio. (b) Distribución del Número de Componentes Conectados (NCC), para Redes de posiciones claves (rojo) y Redes random (azul). (c) Distribución de la máxima red compartida por las estructuras sin y con ligando, sobre el total de residuos en posiciones claves por proteína, esto también puede ser visto como la fracción de red compartida. En rojo se presenta la distribución de las Redes de posiciones claves y en la azul la distribución correspondiente a las Redes random.

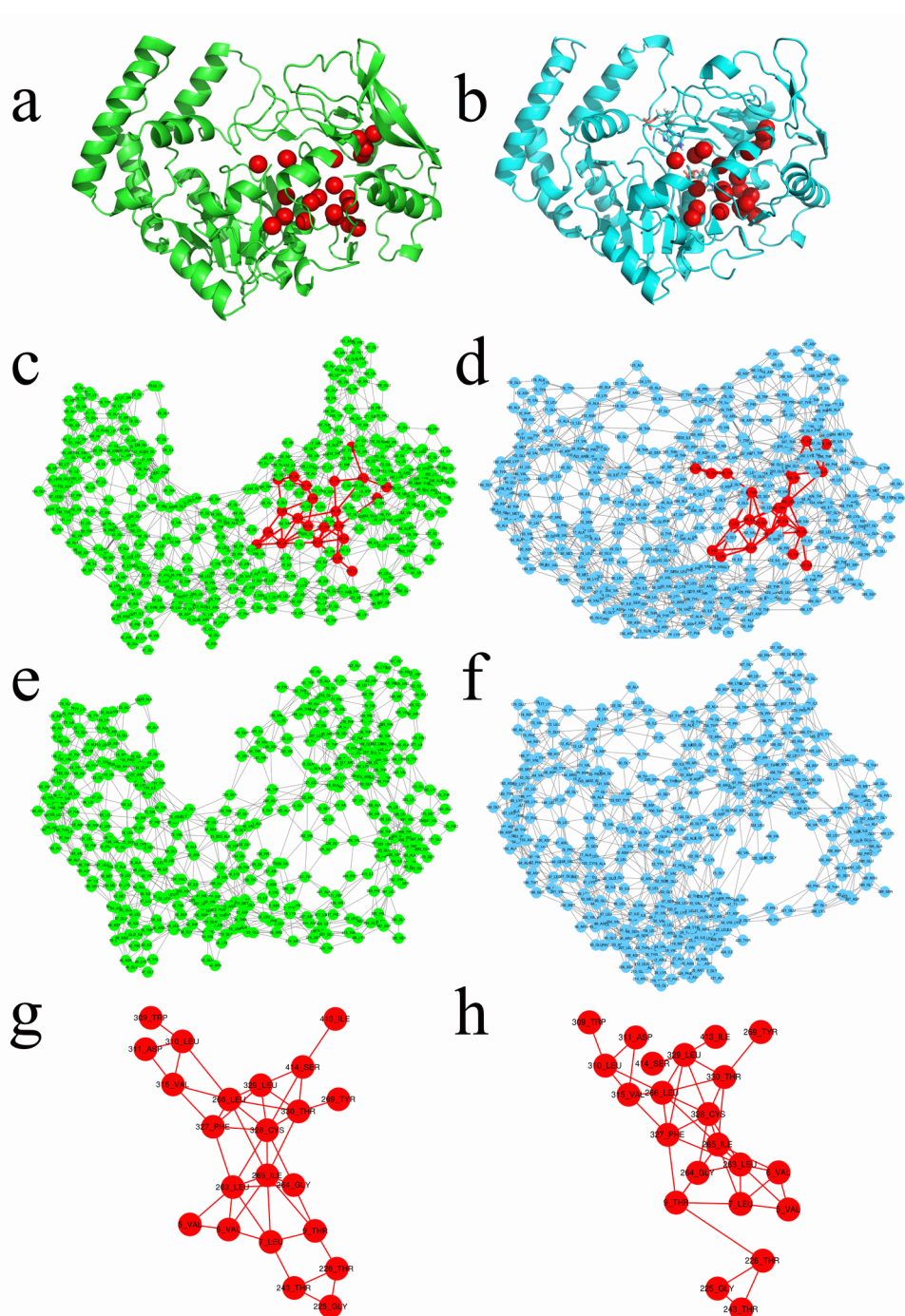
## 4.2.2 Ejemplos

En las **Figuras 4.12 y 4.13** se muestran ejemplos obtenidos de nuestro set de pares de proteínas analizando diferentes tipos de redes.

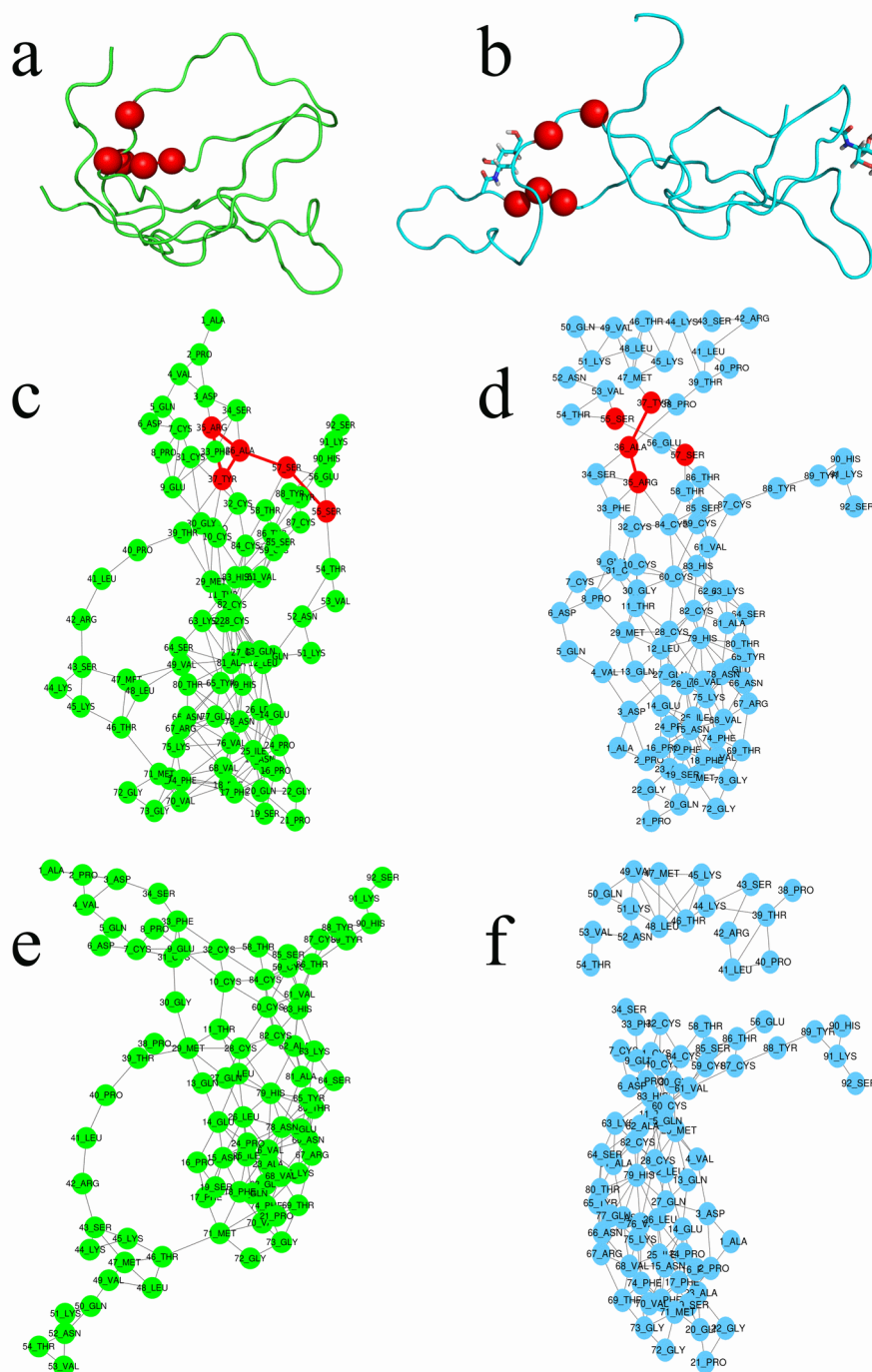
El primer ejemplo corresponde a la Adenilosuccinato Sintetasa (*Adenylosuccinate Synthetase*). En **Figura 4.12a** presentamos la estructura 3D de confórmero sin ligando, código de pdb son 1ADE [Silva 1995]. La representación 3D del confórmero con ligando, código de pdb 1NHT [Poland 1997] se presentó en **Figura 4.12b**. En ambas figuras los residuos en posiciones claves se muestran como esferas rojas. La red total de cada confórmero se presento en **Figura 4.12c y Figura 4.12d**; donde se resalto en rojo a los residuos en posiciones claves y sus contactos. En las redes en las que no se consideran los residuos en posiciones claves, **Figuras 4.12e, 4.12f**, se puede apreciar el cambio en las



conexiones, principalmente para el conformero sin ligando en la zona superior derecha. Finalmente en las Figuras 4.12g y 4.12h, se presentaron las Redes posiciones claves.



**Figura 4.12** Estructura y RINs de la proteína Adenylosuccinato Sintetasa (*Adenylosuccinate Synthetase*). Estructura sin ligando, código de pdb 1ADE (a, c, e, g); estructura con ligando, código de pdb 1NHT (b, d, f, h). (a, b) Estructura visualizada mediante Pymol, en rojo se resaltan los residuos en posiciones claves. (c, d) Red total de residuos, nuevamente los residuos en posiciones claves se muestran en rojo. (e, f) Red de interacciones entre residuos fuera de las posiciones claves. (g, h) Red de interacciones entre los residuos en posiciones claves. En el Apéndice E se presentan las redes en forma individual.



**Figura 4.13 Estructura y RINs de la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( $\alpha$ -Subunit of Human Chorionic Gonadotropin).** Estructura sin ligando, código de pdb 1DZ7 (a, c, e, g); estructura con ligando, código de pdb 1E9J (b, d, f, h). (a, b) Estructura visualizada mediante Pymol, en rojo se resaltan los residuos en posiciones claves. (c, d) Red total de residuos, nuevamente los residuos en posiciones claves se muestran en rojo. (e, f) Red de interacciones entre residuos fuera de las posiciones claves. En el Apéndice E se presentan las redes en forma individual.

El segundo ejemplo utilizado es la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( $\alpha$ -Subunit of Human Chorionic Gonadotropin). En **Figura 4.13a** presentamos la estructura 3D de conformero sin ligando, código de pdb 1DZ7 [Vliegthart 1999]. La representación 3D del conformero con ligando, código de pdb 1E9J [Erbel 2000] se presentó en **Figura 4.13b**. En ambas



figuras los residuos en posiciones claves se muestran como esferas rojas. En las **Figuras 4.13c y 4.13d**, se representó la red total de cada conformero, resaltando en rojo a los residuos en posiciones claves y sus interacciones. En las redes en las que se desestiman los residuos en posiciones claves, **Figura 4.13e** y especialmente en la **Figura 4.13f**, se ve el cambio en las conexiones entre residuos, donde se generan dos *clusters* de residuos.

### 4.2.3 Relación entre residuos en posiciones claves y parámetros topológicos de redes

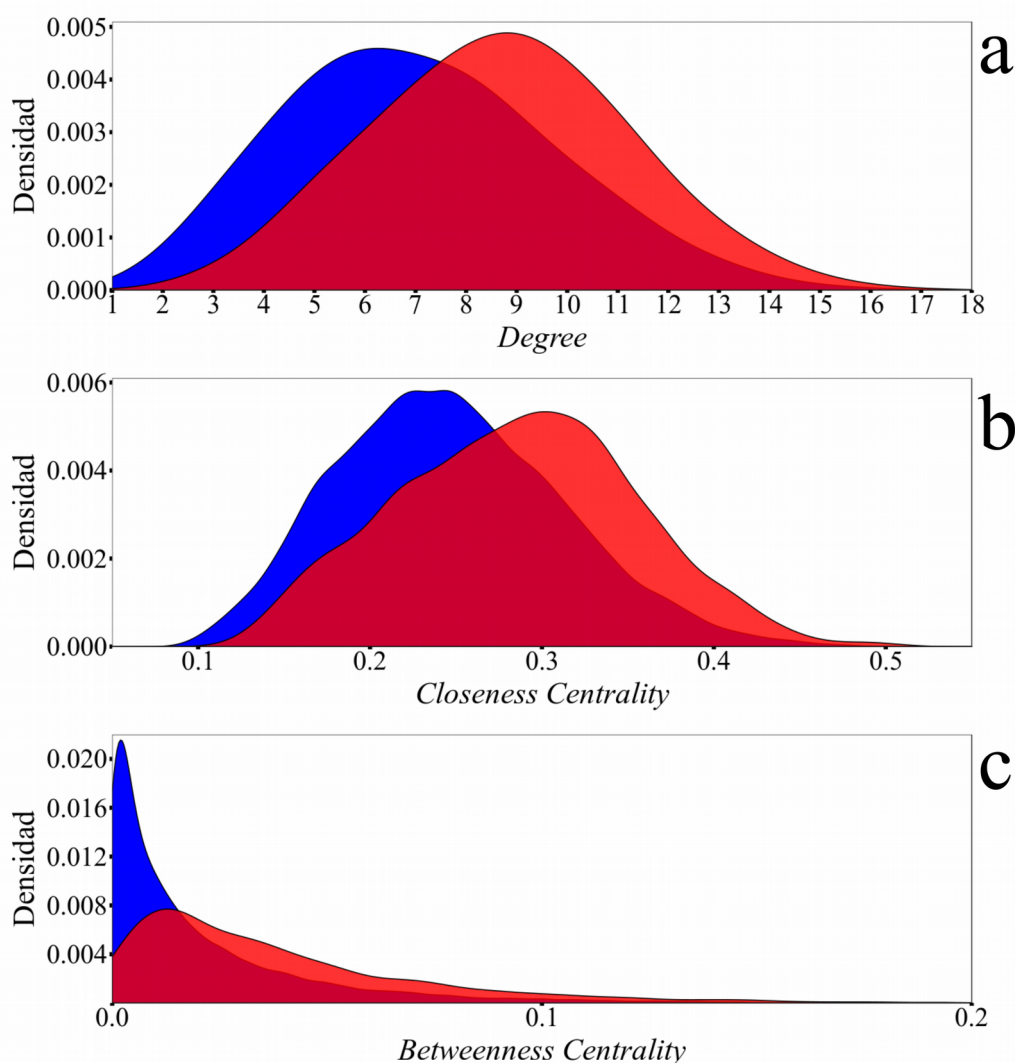
Para cada estructura de nuestro set de proteínas de estudio, mediante el programa RING, se generó una red de interacciones del tipo *Closest Atom*, considerando sólo la interacción de mayor energía entre residuos (ver Capítulo 3, sección 3.3.7). Luego utilizamos el programa Cytoscape y el complemento *NetworkAnalyzer* para calcular de los parámetros topológicos de redes.

El valor de Grado o *Degree* se determina a partir del número de interacciones que posee un residuo. Durante la caracterización de los residuos en posiciones claves, presentada en la sección anterior comentamos que la correlación entre el valor de  $Z_{score}^i$  y el número de contactos por residuo fue de -0,46. La **Figura 4.14a** afirma esta correlación ya que se aprecia que la mayoría de los residuos en posiciones claves tienen valores de Grado o *Degree* mayores que el resto de los residuos.

En diversos trabajos, se ha demostrado que residuos del sitio activo y residuos conservados evolutivamente tienen altos valores de Proximidad Central o *Closeness Centrality*,  $Cc_n$  [Amitai 2004, Thibert 2005, Fajardo 2013]. Residuos con altos valores de  $Cc_n$  interactúan directamente o mediante unos pocos intermediarios con los demás residuos de la proteína. En la **Figura 4.14b** se puede apreciar que los residuos en posiciones claves poseen valores más altos de  $Cc_n$ , respecto a las posiciones donde las mutaciones no afectaron significativamente la dinámica de la estructura. El hecho de que los residuos en posiciones claves presentan valores de  $Cc_n$  mayores sugiere que pueden difundir y recibir (de manera efectiva) señales del resto de la proteína. Esta capacidad es la clave de la importancia de estos residuos ya que podríamos definirlos como responsables de las vibraciones asociadas a distorsiones estructurales introducidas por unión al ligando.

La función de la proteína resulta de la estructura tridimensional adoptada por la secuencia de la proteína y, por lo tanto, la estructura tridimensional de una proteína debe ser considerada para identificar residuos críticos [Benkovic 2003]. En ese sentido, un residuo crítico para una estructura proteica es también crítico para la función de la proteína. Por lo tanto, la referencia a residuos críticos para la función de proteína incluye ambos tipos de residuos: residuos críticos para la estructura de la proteína y / o residuos críticos para su función biológica (por ejemplo, catálisis) [Cusack 2007]. El valor de Centralidad de Intermediación o *Betweenness Centrality*,  $B_n$ , destaca residuos importantes

para el plegado de una proteína [Vendruscolo 2002]. Por lo tanto, este valor podría estar asociado con la capacidad de un residuo para desempeñar un papel crítico en la función de una proteína.



**Figura 4.14** Distribución de los parámetros topológicos de redes obtenidos para los residuos en posiciones claves (rojo) y el resto de los residuos (azul). (a) Distribución de los valores de Grado o *Degree*, el cual representa el número de contactos por residuo. (b) Distribución de los valores de Proximidad Central o *Closeness Centrality*. (c) Distribución de los valores de Centralidad de Intermediación o *Betweenness Centrality*. La diferencia entre ambas poblaciones se validó estadísticamente mediante el test estadístico de Kolmogorov-Smirnov. Se obtuvieron valores de 0,26; 0,25 y 0,31 para las distribuciones en **a**, **b** y **c** respectivamente; todos con un *p-value* igual a  $2,2 \times 10^{-16}$ . Realizamos las distribuciones de Densidad mediante el paquete *ggplot2* de R y el *software Rstudio* [Wickham 2016]. La aplicación del suavizado tipo núcleo (*kernel smoothing*) a la distribución de frecuencia de un conjunto de datos produce la estimación de Densidad Núcleo (*Kernel Density Estimate*). Esta es una alternativa útil al histograma ya que que no requiere la elección arbitraria del ancho de la ventana y presenta un resultado suavizado.

En la **Figura 4.14c**, se muestra que para varios de los residuos en posiciones claves los valores de  $B_n$  son mayores que para el resto de los residuos; esto puede estar asociado a residuos que cumplen funciones como pivote, es decir que pueden ser considerados como puntos de inflexión que

median arreglos o cambios en la estructura proteica, en nuestro caso, asociados a la unión del ligando. Este resultado concuerda con nuestro análisis de interacciones intra-SSE y los movimientos angulares entre estos SSE.

En este punto, resulta interesante analizar las correlaciones entre  $Z_{score}^{S^i}$ ,  $Z_{score}^{C^i}$ ,  $B_n$  y  $Cc_n$ . Por un lado, el coeficiente de correlación de Pearson entre los valores de  $Z_{score}^{S^i}$  y  $B_n$  fue de -0,3; y el coeficiente entre  $Z_{score}^{S^i}$  y  $Cc_n$  fue de -0,6; en ambos casos se obtuvo un  $p$ -value  $< 2,2 \times 10^{-16}$ . Por otro lado, se obtuvieron correlaciones de 0,20 entre  $Z_{score}^{C^i}$  y  $B_n$ ; y de 0,34 entre  $Z_{score}^{C^i}$  y  $Cc_n$ . Se ve claramente que  $B_n$  y  $Cc_n$  correlacionan mejor con el valor de  $Z_{score}^{S^i}$ .

La identificación de residuos críticos en proteínas es importante tanto para la modulación de la función proteica como para la clasificación de proteínas. Con este fin, las secuencias de proteínas constituyen la primera y más abundante fuente de datos para inferir la función de la proteína y, por lo tanto, la mayoría de los métodos computacionales diseñados para identificar residuos críticos se basan en el análisis de secuencias de proteínas. Amitai et al y Thiebert et al. llegan a la conclusión de que combinando información de conservación evolutiva con parámetros topológicos de redes, en especial  $Cc_n$ , mejoran la identificación de residuos funcionales [Amitai 2004, Thibert 2005]. Dado que nuestro parámetro  $Z_{score}^{S^i}$  posee una mayor correlación con estos mismos parámetros topológicos de redes, también podemos concluir que la combinación entre estos parámetros mejoraría la detección de posiciones claves relacionadas con la dinámica de proteínas.

En resumen, adaptamos e implementamos métodos de generación de redes de interacciones entre aminoácidos, que nos permitieron establecer conexiones entre los residuos en posiciones claves y de esta manera definimos redes de residuos importantes para la preservación de la diversidad conformacional de una proteína. Sugerimos que nuestro nuevo enfoque para la identificación de residuos o redes de residuos claves para el mantenimiento de la diversidad conformacional no sólo complementa otros métodos conocidos sino que también conduce a una comprensión más profunda de la relación entre la estructura de la proteína y sitios funcionales específicos. Esta relación es probablemente una consecuencia directa de la robustez funcional necesaria de las proteínas.

### 4.3 EVALUACIÓN DEL EFECTO DE LAS MUTACIONES Y LA UNIÓN DEL LIGANDO A LA DINÁMICA DEL HOMOTETRÁMERO DE TRANSTIRETINA

#### 4.3.1 Transtiretina Humana

La amiloidosis es definida como un grupo de enfermedades en las que se destruyen funciones orgánicas como resultado de depósitos de proteínas solubles como fibras estables e insolubles, conocidas como fibras de amiloide. La transtiretina (TTR) es una de las proteínas humanas cuyo agregado o precipitado genera fibras de amiloide asociadas a distintas condiciones patológicas: Amiloidosis Senil (SSA por sus siglas en inglés, *Senile Systemic Amyloidosis*), Polineuropatía Amiloidótica Familiar (FAP, por sus siglas en inglés, *Familial Amyloidotic Polyneuropathy*) y Miocardiopatía Amiloide Familiar (FAC, por sus siglas en inglés, *Familial Amyloid Cardiomyopathy*) [Miroy 1996, Wojtczak 1996, Hörnberg 2000, Sant'Anna 2016]. SSA y FAC son causadas por el agregado de TTR normal y mutada, respectivamente, preferentemente en el corazón. FAP se caracteriza por la formación de depósitos de TTR mutante en los nervios periféricos, autónomos, y el corazón, pero también en otros sitios como el pulmón, el túnel carpiano y el intestino. Se han reportado más de 100 variantes diferentes de TTR; existen mutaciones raras dentro de la TTR, donde la formación de fibras se desarrolla en el sistema nervioso central, dando como resultado depósitos de fibras en las leptomeninges, en el parénquima cerebral y en los ojos. FAP es causada por una serie de mutaciones en TTR, incluyendo L55P y V30M. La mutación de FAC más frecuente es V122I [Hammarstrom 2003, Bulawa 2012, Saelices 2015, Sant'Anna 2016, Iakovleva 2016].

TTR es una proteína homotetramérica, del plasma humano. Cada monómero está compuesto de 127 residuos de aminoácidos. Esta se expresa, predominantemente, en el hígado adulto, el plexo coroideo del cerebro y la retina. En suero, TTR puede unirse a la hormona tiroxina (T4) y además transporta la proteína de unión al retinol (RBP, por sus siglas en inglés). El estado nativo de TTR presenta dos sitios de unión a la tiroxina, que son generados por la interfaz dímero-dímero, mediante interacciones débiles. La disociación del tetrámero alrededor de esta interfaz genera dímeros que se disocian rápidamente en monómeros. La posterior desnaturalización parcial del monómero, que probablemente involucra la disociación de cadenas  $\beta$ , promueve la formación de oligómeros solubles y fibras amiloides, siendo este proceso favorable termodinámicamente. Por lo tanto, se espera que la estabilización del tetrámero de TTR contribuya a inhibir la fibrilogénesis de TTR [Miroy 1996, Redondo 2000, Quintas 2001, Cardoso 2002, Miller 2004, Johnson 2005, Choi 2010, Connelly 2010, Kolstoe 2010, Bulawa 2012, Palaninathan 2012, Zanotti 2013].

La unión de la hormona tiroxina T4 inhibe la amiloidogénesis causada por TTR estableciendo interacciones con los residuos en la interfaz dímero-dímero. Solo entre un 10-25% de TTR disponible en el plasma sanguíneo, tiene unido la hormona T4, porque la proteína Globulina de Unión a Tiroxina

(TBG, por sus siglas en inglés, *Thyroxine-binding globulin*) es la mayor transportadora de T4 en el plasma. Es así que el nivel de T4 dentro de la sangre, disponible para la unión de TTR, es demasiado bajo para estabilizar eficazmente el tetrámero. La administración externa de la hormona para alcanzar la saturación completa de todas las moléculas de TTR *in vivo* no puede realizarse sin efectos secundarios adversos [Iakovleva 2016]. Se sabe que los compuestos que se unen al sitio de unión, hidrofóbico de T4 estabilizan el tetrámero y enlentecen la agregación o formación de fibras [Bulawa 2012, Zanotti 2013, Cianci 2015, Florio 2015, Iakovleva 2016, Sant'Anna 2016]. Es por esto que se han desarrollado distintos agentes estabilizadores alternativos, y el fármaco Tafamidis (Vyndaqel <sup>TM</sup>) [Bulawa 2012] ha sido aprobado para uso clínico. Un estabilizador TTR eficaz debe tener alta afinidad por el sitio de unión a la hormona y mostrar una alta selectividad para su *target* en el cuerpo humano. Debe además tener una incidencia baja de efectos secundarios y ser resistente a la degradación enzimática *in vivo* [Adamski-Werner 2004, Bulawa 2012, Iakovleva 2016, Sant'Anna 2016]. Explorar nuevas clases de compuestos con cierta afinidad por el tetrámero TTR es la base de la búsqueda de candidatos prometedores en la terapia contra la enfermedad.

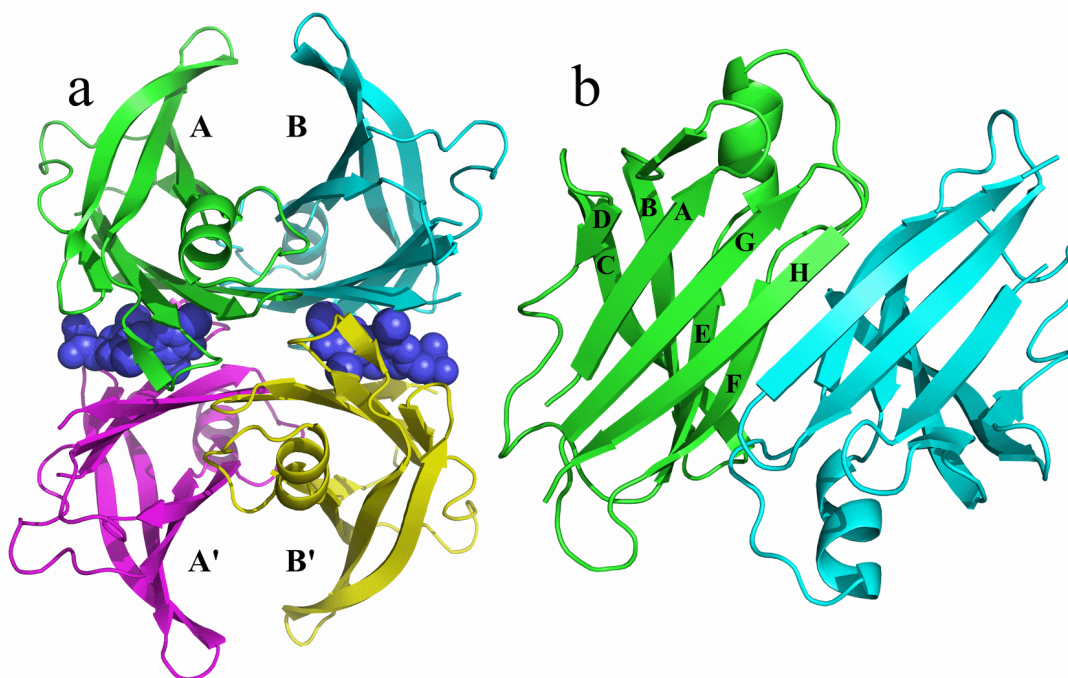
Las mutaciones puntuales que desestabilizan el tetrámero TTR aumentan la propensión a la formación de fibrillas [Johnson 2005, Hurshman Babbes 2008, Cendron 2009, Connelly 2010]. La comprensión a nivel molecular de la dinámica de equilibrio del tetrámero de TTR requiere la identificación de la red de residuos cuyas mutaciones afectan directamente a su estabilidad estructural y dinámica relativa. Adaptando nuestro método desarrollado para definir residuos en posiciones claves asociados con el cambio conformacional por unión al ligando, identificamos los residuos en posición clave cuyas mutaciones alteran significativamente la dinámica de equilibrio del tetrámero de TTR.

Las conformaciones con y sin ligando del tetrámero de TTR coexisten como mínimos locales dentro del potencial de energía de TTR. Los cambios conformacionales implicados en la unión al ligando se consiguen mediante su dinámica de vibración intramolecular. Exploramos la posibilidad de que la alta afinidad relativa de ciertos ligandos a TTR pueda ser una consecuencia de cambios significativos introducidos en el subespacio de modos normales del tetrámero implicado en los cambios conformacionales responsables de la transición entre el confórmero libre y el confórmero unido al ligando. Si la unión del ligando modifica estos modos normales, será más difícil que las fluctuaciones térmicas superen la barrera energética que conlleva a la disociación del complejo tetramérico. En esta tesis presentamos un procedimiento para definir y comparar subespacios de modo normal asociados a la unión al ligando. Adaptamos este método para analizar la estabilidad de la dinámica del complejo TTR-ligando y comparamos varios ligandos de acuerdo con su efecto sobre las vibraciones asociadas a la unión del ligando con el fin de contribuir con nuevas características dinámicas complementarias a futuras exploraciones de potenciales fármacos.

La estructura tetramérica de la TTR unida a la hormona T4 se representa en la **Figura 4.15**. Las hormonas T4 se localizan en las cavidades hidrofóbicas de cada interfaz dímero-dímero. Los dímeros están compuestos de dos monómeros (A-A' y B-B' respectivamente). Cada monómero está

formado por dos hojas- $\beta$ , cada una formada por cuatro hebras antiparalelas, y una hélice- $\alpha$  corta. Las interfases monómero-monómero se estabilizan mediante enlaces puente de hidrógeno entre las hojas- $\beta$  H y F.

La lista de estructuras de TTR utilizadas para el análisis, se indica en la **Tabla 4.4**. Aplicamos varios filtros a las estructuras TTR disponibles en la base de datos PDB para obtener un conjunto de datos bien curado: (i) estructuras cristalizadas con resolución  $< 1,7 \text{ \AA}$ , (ii) estructuras completas entre los residuos 11 y 124, (iii) estructuras sin mutaciones, (iv) estructuras para las cuales la correlación entre factores B experimentales y teóricos es mayor a  $0,6 \text{ \AA}$ , (v) estructuras con ligandos donde se haya demostrado que inhiben la formación de fibras, (vi) estructuras con ligandos cuyas afinidades relativas se han demostrado experimentalmente usando ensayos de unión competitiva por Calorimetría Isotérmica de Titulación (ITC, por sus siglas en inglés, *Isothermal Titration Calorimetry*).



**Figura 4.15 (a)** Estructura de la Transtiretina humana unida a la hormona tiroxina (código pdb: 2rox). Las moléculas de T4, dentro de las cavidades de unión, se muestran como esferas. **(b)** Estructura del dímero A-B. Las hojas- $\beta$  de cada monómero se identifican mediante letras (A-H) como fue sugerido por Blake et al, en 1978 [Blake 1978].

**Tabla 4.4** Estructuras de TTR consideradas para el análisis.

TTR - ligando	Código pdb
Sin ligando	1f4l
T4	2rox
T4-resveratrol	5cr1
Tafamidis	3tct
Diflunisal	4i89
TBBPA	5hjh
Apigenin	4wo0
CHF5074	4i85
Daidzein	5al8
Quercetin	4wnj
Pterostilbene	4wns
Genistein	5akv
Resveratrol-3-O-sulfate	5al0

### 4.3.2 Efecto de las mutaciones en la dinámica de equilibrio del tetrámero

La dinámica vibracional intramolecular de una proteína está intrínsecamente relacionada con su estabilidad estructural. Por lo tanto, los cambios en la dinámica de equilibrio del tetrámero de TTR pueden estar asociados a cambios en su estabilidad relativa. En base a esto propusimos que las mutaciones que afecten la dinámica del tetrámero desestabilizarían al mismo, aumentando la propensión de la formación de fibrillas de TTR [Saldaño 2017].

Mediante nuestra metodología, simulamos mutaciones puntuales para cada residuo de la estructura de tetrámero de TTR sin ligando con el fin de identificar los residuos en posiciones claves para el mantenimiento de la estabilidad de la dinámica del tetrámero. El valor de  $Z_{score}^{R^i}$  (ver Capítulo 3, sección 3.2.2) es una medida del cambio en la dinámica de equilibrio del tetrámero de TTR introducida por la mutación en el  $i$ -ésimo residuo. Cuanto menor sea el valor de  $Z_{score}^{R^i}$ , mayor será el efecto de la mutación del residuo sobre la estabilidad del tetrámero. Los residuos en posiciones claves se definieron como aquellos dentro del 10% de los residuos con menor valor de  $Z_{score}^{R^i}$ . En la **Tabla 4.5** se analizan estos residuos; la mayoría de ellos están localizados en la hoja- $\beta$  E, la hoja- $\beta$  F y el *loop* entre estas. Casi la mitad de los residuos claves se localizaron en la interfase monómero-monómero, revelando que esta es una de las regiones más vulnerables a las mutaciones que generan cambios significativos en la dinámica de equilibrio del tetrámero. En la **Tabla 4.5** también mostramos que las posiciones definidas como claves coinciden en su mayoría con mutaciones informadas

previamente y asociadas a la amiloidosis por fibrillas de TTR. Particularmente, la posición V30, con cuatro mutaciones diferentes en esta posición y con  $Z_{score}^{R^i} = -1,17$ , ha sido identificada como una de las mutaciones más destacadas de la Polineuropatía Amiloidótica Familiar [Sipe 1994, Miroy 1996, Johnson 2005]. La **Figura 4.16** muestra la localización de los residuos claves dentro de la estructura de TTR.

Los residuos de las posiciones claves se han analizado utilizando el algoritmo FoldX [Guerois 2002, Schymkowitz 2005], que predice diferencias en los cambios de energía libre,

$$\Delta \Delta G = \Delta G_{mutant} - \Delta G_{wild\ type} \quad (\text{Ec. 4.4})$$

donde  $\Delta G_{mutant}$  es el cambio energía libre de la estructura mutada y  $\Delta G_{wild\ type}$  es el cambio energía libre de la estructura nativa. El valor promedio de  $\Delta \Delta G$  sobre todas las mutaciones posibles obtenido para cada residuo clave se muestra en la **Tabla 4.5**. Acorde con nuestras predicciones, encontramos que para la mayoría de los residuos en posiciones claves, las modificaciones son fuertemente desestabilizante ( $\Delta \Delta G > 2\text{kcal / mol}$ ) [Tokuriki 2007].

**Tabla 4.5** Residuos en posiciones claves para el tetrámero de TTR.

Residuo	$Z_{score}^{R^i}$	Elemento de estructura secundaria	Tipo	Nº IAC Intra Monómero	Nº IAC Inter Monómero	Mutación	$\langle \Delta \Delta G \rangle$
His 88	-1,87	Loop EF	*	7	1	His → Arg	-1,03
Tyr 69	-1,58	Hoja-β E		13		Tyr → His; Tyr → Ile	18,03
Thr 75	-1,52	Hélice-α		9			8,18
Phe 95	-1,46	Hoja-β F	*	9	2		20,56
Phe 87	-1,44	Loop EF	*	4	6	Phe → Met	18,53
Ile 73	-1,38	Hoja-β E		11		Ile → Val	17,49
Glu 89	-1,33	Loop EF	*	5	3	Glu → Gln; Glu → Lys	3,55
Val 71	-1,32	Hoja-β E		9		Val → Ala	16,81
Val 93	-1,22	Hoja-β F	*	6	2	Val → Met	9,74
Val 30	-1,17	Hoja-β B		7		Val → Leu; Val → Met Val → Ala; Val → Gly	23,46
Asp 74	-1,15	Hoja-β E		7		Asp → His (No Amiloidogenica)	7,63

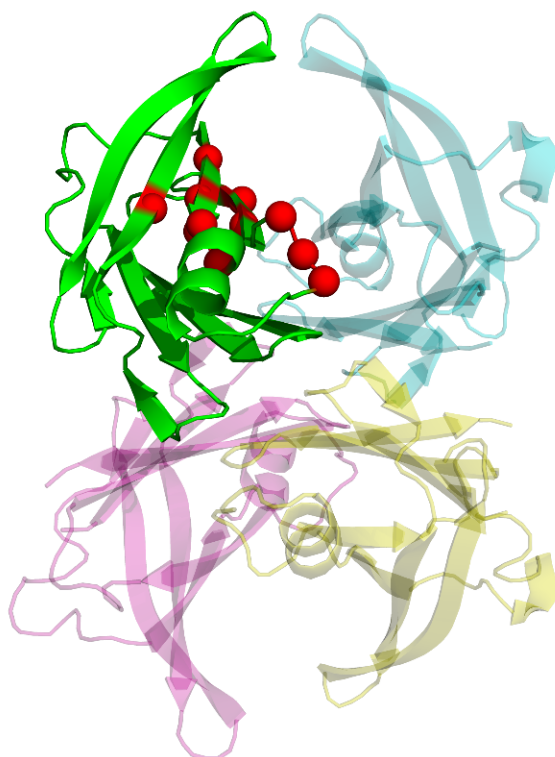
**Tipo:**\* interfase monómero-monómero [Hörnberg 2000]

# interfase dímero-dímero [Hörnberg 2000]

† Cavidad de unión del ligando [Wojtczak 1996, Tokuriki 2007, Florio 2015]

**Mutación:** <http://www.amyloidosismutations.com/mut-attr.php>

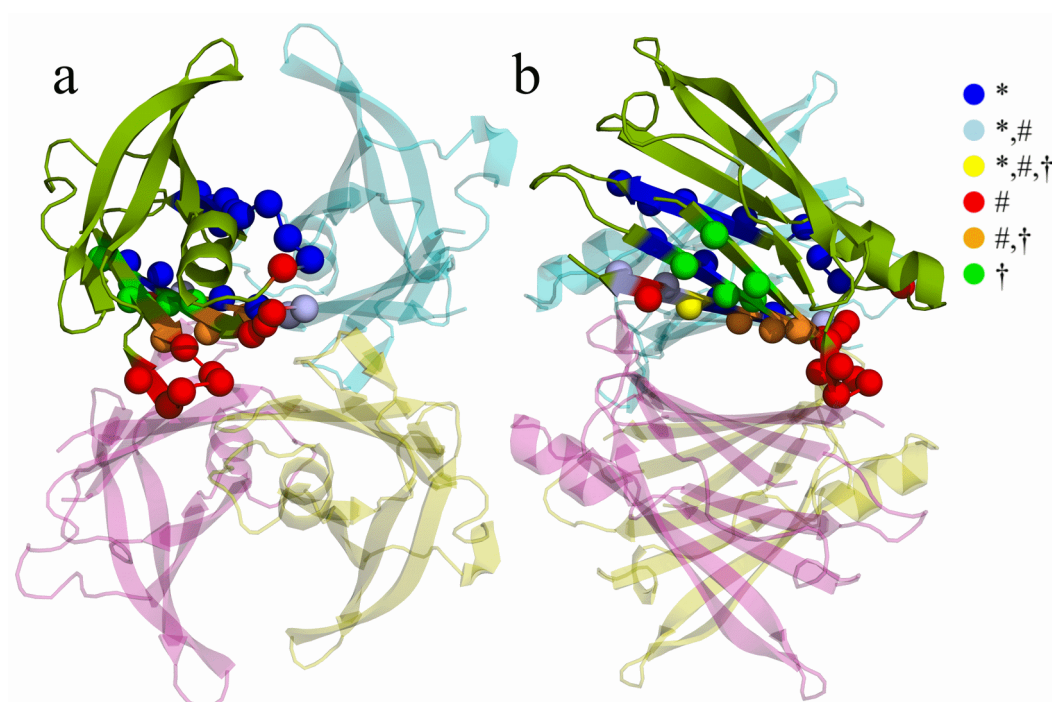




**Figura 4.16 Residuos en posiciones claves, mostrados como esferas.** Sus localizaciones se muestran en un solo monómero.

La **Tabla 4.6** muestra los valores de  $Z_{score}^{R^i}$  de residuos previamente identificados como parte del sitio de unión de la hormona tiroxina y las interfaces monómero-monómero y dímero-dímero [Wojtczak 1996, Hörnberg 2000, Adamski-Werner 2004, Zhang 2016]. Todos los residuos que forman la cavidad de unión del ligando presentan valores negativos de  $Z_{score}^{R^i}$ , lo que indica que las mutaciones en estos residuos desestabilizan al tetrámero más que el promedio. Se observan valores negativos de  $Z_{score}^{R^i}$  para los residuos de la interfaz dímero-dímero dentro del fragmento L110-V122 (hoja- $\beta$  G, loop-GH, y hoja- $\beta$  H), y los residuos L17, D18 (hoja- $\beta$  A); mientras que para el segmento A19-S23 (loop AB) y el residuo S85 (loop EF) obtuvimos valores positivos de  $Z_{score}^{R^i}$ .

Finalmente, todos los residuos de la interfase monómero-monómero también presentan valores negativos de  $Z_{score}^{R^i}$ . Es decir, las mutaciones en cualquiera de los residuos del segmento F87-V122 perturban las vibraciones del tetrámero más que el promedio. La **Figura 4.17** muestra la localización de todos estos residuos dentro de la estructura de la TTR.



**Figura 4.17** Residuos de la interfaz monómero-monómero, dímero-dímero y de la cavidad de unión del ligando, mostrados como esferas. Sus localizaciones se muestran sólo en un monómero. (\* interfaz monómero-monómero, # interfaz dímero-dímero, † cavidad de unión del ligando).

**Tabla 4.6** Valores de  $Z_{score}^{R^i}$  de los residuos de la interfaz monómero-monómero, dímero-dímero y de la cavidad de unión del ligando.

Residuo	$Z_{score}^{R^i}$	Elemento de estructura secundaria	Tipo	Nº IAC Intra Monómero	Nº IAC Inter Monómero	Mutación
Met 13	-0,16	Hoja- $\beta$ A	†	9		Met $\rightarrow$ Ile (No Amiloidogénica)
Lys 15	-0,45	Hoja- $\beta$ A	†	8		
Leu 17	-0,21	Hoja- $\beta$ A	#,†	8	2	
Asp 18	-0,30	Hoja- $\beta$ A	#	8	1	Asp $\rightarrow$ Asn; Asp $\rightarrow$ Gly; Asp $\rightarrow$ Glu
Ala 19	0,15	Loop AB	#	7	3	Ala $\rightarrow$ Asp
Val 20	0,18	Loop AB	#	6	2	Val $\rightarrow$ Ile
Arg 21	0,53	Loop AB	#	6	1	
Gly 22	0,19	Loop AB	#	4	1	
Ser 23	0,14	Loop AB	#	5	1	Ser $\rightarrow$ Asn
Ser 85	0,31	Loop EF	#	4	1	
Phe 87	-1,44	Loop EF	*	4	6	
His 88	-1,87	Loop EF	*	7	1	
Glu 89	-1,33	Loop EF	*	5	3	Glu $\rightarrow$ Gln; Glu $\rightarrow$ Lys
His 90	-0,92	Loop EF	*	5	1	His $\rightarrow$ Asp; His $\rightarrow$ Asn

						(No Amiloidogenica)
Glu 92	-0,53	Hoja-β F	*	6	4	Gln → Lys
Val 93	-1,22	Hoja-βF	*	6	2	Val → Met
Val 94	-1,08	Hoja-β F	*	6	4	Val → Ala
Phe 95	-1,46	Hoja-β F	*	9	2	
Thr 96	-0,93	Hoja-β F	*	4	3	
Tyr 105	-1,04	Hoja-β G	*	14	1	
Thr 106	-0,37	Hoja-β G	†	4		
Ile 107	-1,01	Hoja-β G	*	9	1	Ile → Val; Ile → Phe Ile → Met
Ala 108	-0,43	Hoja-β G	†	4		
Ala 109	-0,61	Hoja-β G	†	6		Ala → Ser (No Amiloidogenica) Ala → Thr (No Amiloidogenica)
Leu 110	-0,06	Hoja-β G	#,†	6	2	
Ser 112	-0,21	Hoja-β G	#	7	1	Ser → Ile
Pro 113	-0,62	Loop GH	#	9	1	
Tyr 114	-1,07	Loop GH	*,#	6	6	Tyr → His; Tyr → Cys
Ser 115	-0,18	Hoja-β H	*,#	5	4	
Tyr 116	-0,70	Hoja-β H	*	6	6	Tyr → Ser
Ser 117	0,44	Hoja-β H	#,†	3	2	
Thr 118	-0,33	Hoja-β H	*	5	1	
Thr 119	-0,58	Hoja-β H	*, #, †	3	5	Thr → Met (No Amiloidogenica)
Ala 120	-0,91	Hoja-β H	*, #	4	2	Ala → Ser
Val 121	-0,54	Hoja-β H	#	4	2	
Val 122	-0,40	Loop	*, #	5	3	Val → Ile; Val → del; Val → Ala

**Tipo:** \* interfase monómero-monómero [Hörnberg 2000]

# interfase dímero-dímero [Hörnberg 2000]

† Cavidad de unión del ligando [Wojtczak 1996, Tokuriki 2007, Florio 2015]

**Mutación:** <http://www.amyloidosismutations.com/mut-attr.php>

Con el fin de resaltar la robustez de nuestros resultados, es importante mencionar que los resultados mostrados en la **Tabla 4.5** se reprodujeron utilizando una segunda estructura de TTR sin ligando, código de pdb 2qgb. El RMSD entre las estructuras 1f41 y 2qgb es de 0,47 Å. A pesar de los cambios en el orden, se obtuvieron resultados cualitativos similares para ambas estructuras. Considerando los residuos dentro del 20% con valores de  $Z_{score}^R$  más bajos, más del 75% de estos residuos se reproducen en ambas estructuras y ~ 90% coinciden en +/- un residuo. La lista completa de

los residuos dentro del 20% con valores de  $Z_{score}^{R^i}$  más bajos obtenidos usando ambas estructuras sin ligando se proporciona en las **Tablas 4.7** y **4.8**.

**Tabla 4.7** Residuos dentro del 20% con menor valor de  $Z_{score}^{R^i}$  para la estructura código de pdb 1f41.

Residuo	$Z_{score}^{R^i}$	Elemento de estructura secundaria	Tipo	N° IAC Intra Monómero	N° IAC Inter Monómero	Mutación
His 88	-1,87	Loop EF	*	7	1	His → Arg
Tyr 69	-1,58	Hoja-β E		13		Tyr → His; Tyr → Ile
Thr 75	-1,52	Hélice-α		9		
Phe 95	-1,46	Hoja-β F	*	9	2	
Phe 87	-1,44	Loop EF	*	4	6	Phe → Met
Ile 73	-1,38	Hoja-β E		11		Ile → Val
Glu 89	-1,33	Loop EF	*	5	3	Glu → Gln; Glu → Lys
Val 71	-1,32	Hoja-β E		9		Val → Ala
Val 93	-1,22	Hoja-β F	*	6	2	Val → Met
Val 30	-1,17	Hoja-β B		7		Val → Leu; Val → Met Val → Ala; Val → Gly
Asp 74	-1,15	Hoja-β E		7		Asp → His (No Amiloidogenica)
Glu 72	-1,11	Hoja-β E		6		Glu → Gly
Leu 111	-1,10	Hoja-β G		11		Leu → Met
Lys 70	-1,09	Hoja-β E		7	1	Lys → Asn
Val 94	-1,08	Hoja-β F	*	6	4	Val → Ala
Tyr 114	-1,07	Loop GH	*, #	6	6	Tyr → His; Tyr → Cys Tyr → Cys
Tyr 105	-1,04	Hoja-β G	*	14	1	
Ile 107	-1,01	Hoja-β G	*	9	1	Ile → Val; Ile → Phe Ile → Met
Tyr 78	-0,99	Hélice-α		12		Tyr → Phe
Val 28	-0,98	Hoja-β B		9		Val → Met
Ala 91	-0,98	Hoja-β F		5		Ala → Ser
Thr 96	-0,93	Hoja-β F	*	4	3	

**Tipo:**\* interfase monómero-monómero [Hörnberg 2000]

# interfase dímero-dímero [Hörnberg 2000]

† Cavidad de unión del ligando [Wojtczak 1996, Tokuriki 2007, Florio 2015]

**Mutación:** <http://www.amyloidosismutations.com/mut-attr.php>

**Tabla 4.8** Residuos dentro del 20% con menor valor de  $Z_{score}^{R^i}$  para la estructura código de pdb 2qgb.

Residuo	$Z_{score}^{R^i}$	Elemento de estructura secundaria	Tipo	Nº IAC Intra Monómero	Nº IAC Inter Monómero	Mutación
Tyr 105	-1,6	Hoja-β G	*	14	1	
Phe 87	-1,54	Loop EF	*	4	6	Phe → Met
Thr 75	-1,49	Hélice-α		9		
His 88	-1,45	Loop EF	*	7	1	His → Arg
Tyr 69	-1,33	Hoja-β E		13		Tyr → His; Tyr → Ile
Ile 73	-1,24	Hoja-β E		11		Ile → Val
Pro 86	-1,23	Loop EF		7		
Phe 95	-1,19	Hoja-β F	*	9	2	
Thr 120	-1,16	Hoja-β H	*, #	4	2	Ala → Ser
Val 93	-1,15	Hoja-β F	*	6	2	Val → Met
Trp 79	-1,1	Hélice-α		13		
Ile 107	-1,08	Hoja-β G	*	9	1	Ile → Val; Ile → Phe Ile → Met
Asp 74	-1,02	Hoja-β E		7		Asp → His (No Amiloidogénica)
Val 71	-0,97	Hoja-β E		9		Val → Ala
Val 14	-0,93	Hoja-β A		8		
Thr 96	-0,92	Hoja-β F	*	4	3	
Glu 89	-0,9	Loop EF	*	5	3	Glu → Gln; Glu → Lys
Val 121	-0,89	Hoja-β H	#	4	2	
Leu 111	-0,88	Hoja-β G		11		Leu → Met
Val 94	-0,87	Hoja-β F	*	6	4	Val → Ala
Tyr 114	-0,86	Loop GH	*, #	6	6	Tyr → His; Tyr → Cys Tyr → Cys
Lys 76	-0,84	Hélice-α		8	1	

**Tipo:**\* interfase monómero-monómero [Hörnberg 2000]

# interfase dímero-dímero [Hörnberg 2000]

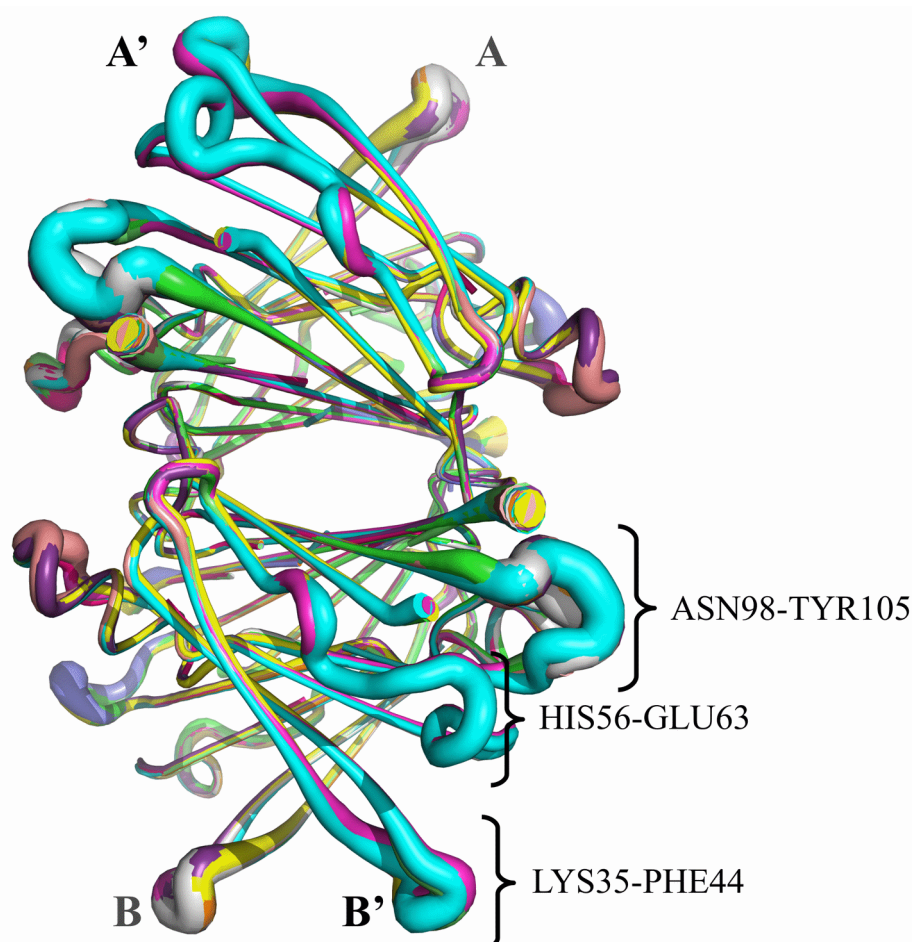
† Cavidad de unión del ligando [Wojtczak 1996, Tokuriki 2007, Florio 2015]

**Mutación:** <http://www.amyloidosismutations.com/mut-attrib.php>

### 4.3.3 Efecto de la unión de drogas a la dinámica de equilibrio del tetramero

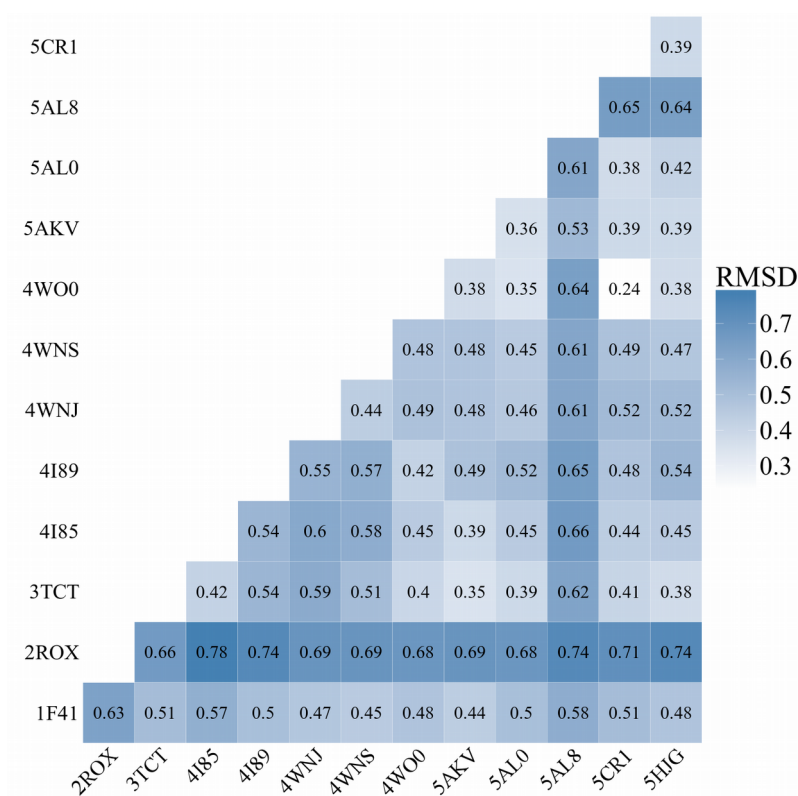
Hemos realizado un análisis comparativo de los efectos de varios fármacos y compuestos (ver **Tabla 4.4**) sobre las vibraciones implicadas en los cambios conformacionales asociados a la formación / disociación del complejo tetramérico sin ligando. La **Figura 4.18** muestra las estructuras

superpuestas listadas en la **Tabla 4.4**. En la figura se resaltan aquellas regiones que presentan mayor flexibilidad en base a los valores de Factores B experimentales. Ninguno de los residuos de las posiciones claves enumerados en la **Tabla 4.5** pertenece a una de estas regiones.



**Figura 4.18** Representación de las 13 estructuras de TTR superpuestas, enumeradas en la **Tabla 4.4**. El ancho de la representación corresponde a los valores de los factores B experimentales. También se indican los segmentos con mayor flexibilidad.

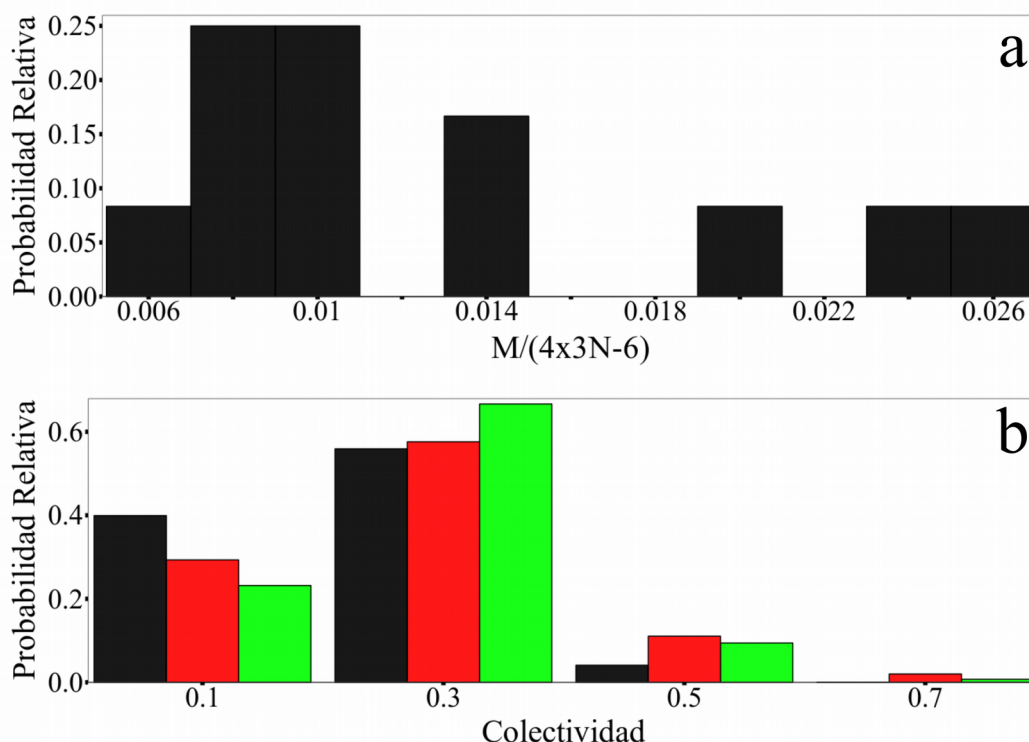
Las distorsiones estructurales introducidas por la unión del ligando se analizan en términos de RMSD. Una información detallada de los valores de RMSD entre todas las estructuras de nuestro conjunto de datos se muestra en la **Figura 4.19**. Los valores bajos de RMSD indican que la estructura libre de ligando del tetrámero de TTR no se ve afectada en gran medida por la unión del ligando. La estructura del tetrámero unida a la hormona T4 (código pdb 2ROX) es la que presenta las principales diferencias estructurales con el resto de las estructuras, debido principalmente a las diferencias estructurales entre la hormona T4 y los otros compuestos. A pesar de ello, los valores de RMSD entre las estructuras parecen no seguir una relación clara respecto a las similitudes químicas entre los ligandos, perteneciendo en su mayor parte a la gran familia de polifenoles.



**Figura 4.19** Información detallada de los valores RMSD entre todas las estructuras de nuestro conjunto de datos.

Por otra parte, la **Figura 4.20a** muestra la distribución de la fracción de modos normales, calculados a partir de la estructura sin ligando, implicados en las transiciones conformacionales inducidas por la unión del ligando al tetrámero de TTR. Se calculó un número de participación ( $M$ ), que define el tamaño del subespacio de modos normales, por cada uno de los ligandos. La fracción se calculó a partir del valor del número  $M$  (ver Capítulo 3, sección 3.2.1) sobre el total de modos normales,  $M/(4 \times 3N - 6)$ , siendo  $N$  el número de  $C\alpha$  de la estructura monomérica. Los valores  $\langle M \rangle = 104 \pm 54$  y  $\langle M/(4 \times 3N - 6) \rangle = 0,01 \pm 0,007$  indican que una fracción significativamente pequeña del espacio vibracional participa en el proceso de unión al ligando. Los modos normales pueden ser analizados en términos de sus colectividades. La **Figura 4.20b** señala que los modos implicados en la transición conformacional inducida por el ligando están compuestos por vibraciones dentro de toda la gama de frecuencias, mostrando una distribución de colectividades ligeramente superior a otros modos. La distribución de colectividad de los modos relacionados con el patrón de flexibilidad del tetrámero de TTR también se muestra en la **Figura 4.20b**. Podemos ver que los subespacios de los modos normales asociados a la unión del ligando no implican modos más colectivos que los que describen la flexibilidad de la proteína.





**Figura 4.20 (a)** Distribución de la fracción de modos normales implicados en la transición conformacional inducida por la unión del ligando, calculada como  $M/(4 \times 3N - 6)$ ; **(b)** Distribución del grado de colectividad,  $\kappa_k$ , para cada modo normal que participa en el cambio conformacional (rojo), para cada modo normal que describe el patrón de flexibilidad, según los valores de factores B (verde) y para el resto de los modos (negro).

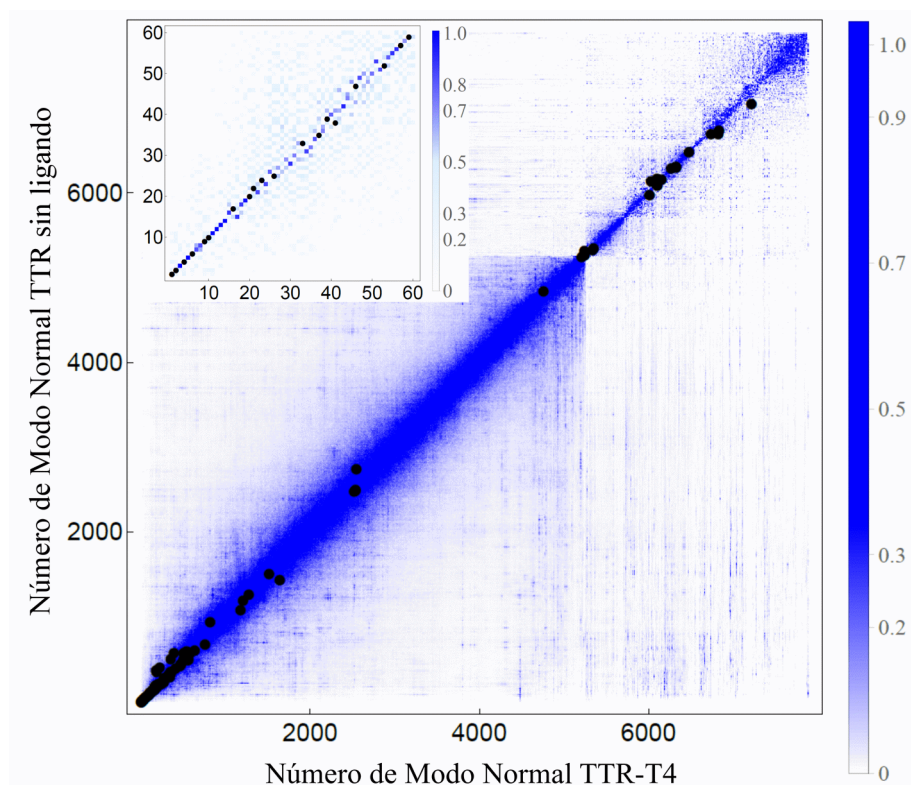
Para analizar la composición del subespacio de modos normales asociados a la unión del ligando, la **Figura 4.21** representa la matriz de solapamiento entre los modos normales de la estructura TTR sin ligando y la TTR unida al T4, destacando los modos que pertenecen al subespacio de unión del ligando. Como puede verse, estos modos normales no están uniformemente distribuidos en toda la gama de modos. La mayoría de estos modos se localizan en el rango de baja frecuencia y solo una contribución menor corresponde a los modos de alta frecuencia. Esto concuerda con estudios previos que asocian modos normales de baja frecuencia con cambios conformacionales observados por unión del ligando [Tama 2001, Petrone 2006].

Para el análisis del cambio de las vibraciones internas del tetrámero asociadas a la unión del ligando, utilizamos nuestro método de comparación de subespacios (ver Capítulo 3, sección 3.2.2, ecuaciones 3.9 a 3.12). En este caso se compararon los subespacios de modos normales calculados a partir de la estructura sin ligando,  $S$ , contra cada uno de los subespacios de modos normales generados a partir de la estructura unida al ligando,  $S^{\text{ligando}}$  [Saldaño 2017]. De esta manera se obtuvo un valor de comparación para cada ligando,  $\zeta^{S^{\text{ligando}}S}$ . Luego se normalizo cada valor de  $\zeta^{S^{\text{ligando}}S}$

$$Z_{\text{score}}^{S^{\text{ligando}}S} = \frac{\langle \zeta^{S^{\text{ligando}}S} \rangle - \langle \zeta^{\text{ref}} \rangle}{\sigma^{\text{ref}}} \quad (\text{Ec. 4.5})$$



donde  $\bar{\zeta}^{\text{ref}}$  y  $\sigma^{\text{ref}}$  son la media y la desviación estándar de una distribución de referencia, construida a partir de la comparación de  $\mathbf{S}^{\text{ligando}}$  con subespacios  $\mathbf{S}^{\text{ref}}$  obtenidos mediante una selección aleatoria de  $M$  modos normales de la estructura sin ligando. Se consideró una ventana de  $\pm 10$  modos centrados en cada modo del subespacio  $\mathbf{S}$  original.



**Figura 4.21** Matriz de superposición entre los modos normales de la estructura de TTR sin ligando y los modos normales de la estructura de TTR con el ligando T4 unido. Se resaltó a los modos que pertenecen al subespacio de unión al ligando mediante puntos negros. El recuadro muestra la matriz de superposición para modos dentro del rango de baja frecuencia.

En la **Figura 4.22** mostramos la comparación de nuestro conjunto de ligandos de acuerdo con su efecto sobre las vibraciones asociadas a la unión al ligando. Con el fin de relacionar la afinidad de la unión con los cambios introducidos sobre las vibraciones, analizamos nuestros valores de  $Z_{\text{score}}^{\text{ligandoS}}$  en términos de parámetros experimentales y mediciones asociadas a la afinidad de unión del ligando y la estabilización del estado nativo del tetrámero. Los ligandos que presentan valores bajos de  $Z_{\text{score}}^{\text{ligandoS}}$  introducen cambios significativos en las vibraciones asociadas a los cambios conformacionales del tetrámero que conllevan a la disociación del mismo. Como consecuencia, el complejo tetrámero-ligando es más estable y por lo tanto se inhibe la fibrillogénesis.

Realizando ensayos de unión competitiva, se ha demostrado que los polifenoles Pterostilbene y Quercetin tienen preferencia por sitios de unión diferentes en TTR respecto a la hormona T4 [Cianci 2015]. Además, los experimentos revelan que, a concentraciones iguales, el ligando Quercetin es

capaz de desplazar al Pterostilbene [Cianci 2015], y esto concuerda con el valor de  $Z_{score}^{ligandos}$  inferior para Quercetin respecto al Pterostilbene. Aunque ni Pterostilbene ni Quercetin son capaces de desplazar al T4, el polifenol Apigenin lo desplaza, pero sólo a concentraciones muy altas [Cianci 2015]. Estos resultados indican que Apigenin presenta menor afinidad que T4. Esto está de acuerdo con nuestro mayor valor de  $Z_{score}^{ligandos}$  para Apigenin respecto a T4. La afinidad de la TTR por el Tafamidis ha demostrado ser más alta que la de los polifenoles y T4, algo que  $Z_{score}^{ligandos}$  sólo puede confirmar para el caso de Pterostilbene y Quercetin. Sin embargo, Apigenin, T4 y Tafamidis están cerca en orden de  $Z_{score}^{ligandos}$ .

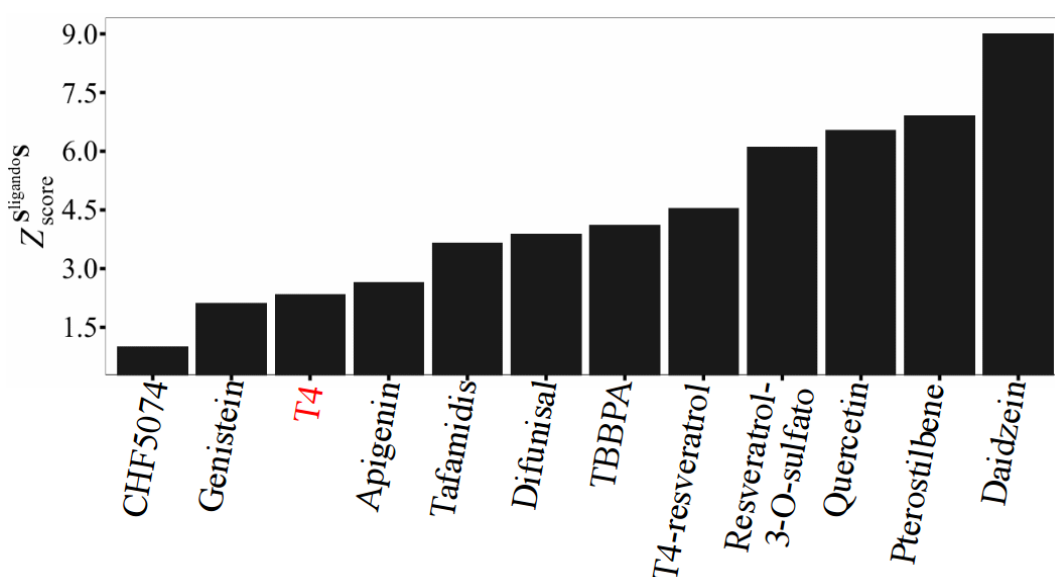


Figura 4.22 Valores de  $Z_{score}^{ligandos}$  para los diferentes ligandos del set.

Según experimentos de competición utilizando Resveratrol y T4 radiomarcado como sonda, el Resveratrol no es capaz de desplazar el T4 unido. Además, los experimentos de fluorescencia muestran que metabolitos del Resveratrol, como el Resveratrol-3-O-sulfato, tienen afinidades más bajas para TTR [Cianci 2015]. En buen acuerdo con estos resultados experimentales, obtenemos que  $Z_{score}^{ligandos}$  (T4) <  $Z_{score}^{ligandos}$  (Resveratrol) <  $Z_{score}^{ligandos}$  (Resveratrol-3-O-sulfato). Además, los ensayos fluorométricos de unión competitiva, utilizando Resveratrol como sonda fluorescente, indican que el ligando CHF5074 presenta una elevada afinidad de unión para TTR, de acuerdo con su bajo valor de  $Z_{score}^{ligandos}$ .

Las constantes de disociación del ligando ( $K_d$ ) determinadas por ITC [Iakovleva 2015], indican que la  $K_d$  (Diflunisal) = 580 nM es mayor que  $K_d$  (Apigenina = 250 nM, lo cual concuerda con nuestro resultado,  $Z_{score}^{ligandos}$  (Apigenin) <  $Z_{score}^{ligandos}$  (Diflunisal). Además, los experimentos de ITC también informan valores de  $K_d$  (TBBPA) = 20 nM y  $K_d$  (Tafamidis) = 3 nM [Iakovleva 2016] que

conduce al siguiente orden de afinidad por TTR, tafamidis > TBBPA > diflunisal. Los valores de  $Z_{score}^{S_{ligando}^S}$  (TBBPA) >  $Z_{score}^{S_{ligando}^S}$  (Diflunisal) no concuerdan con estos resultados, a pesar de que ambos compuestos están cerca en orden de  $Z_{score}^{S_{ligando}^S}$ .

La capacidad relativa de Genistein, Apigenin y Daidzen para desplazar la hormona T4, radiomarcada, unida a TTR indica el orden de las afinidades de unión para TTR como Genistein > Apigenin > Daidzen, y esto concuerda con el orden dado en la **Figura 4.22**. Entre ellos, sólo Genistein y Apigenin son capaces de desplazar al Resveratrol unido a TTR mientras que Daidzen resulto un ligando débil que exhibe una baja afinidad de unión. Es decir, las afinidades de unión a TTR varían según el siguiente orden Genistein > Apigenin > Resveratrol > Daidzen y nuestro el orden concuerda siendo  $Z_{score}^{S_{ligando}^S}$  (Genistein) <  $Z_{score}^{S_{ligando}^S}$  (Apigenin) <  $Z_{score}^{S_{ligando}^S}$  (Resveratrol) <  $Z_{score}^{S_{ligando}^S}$  (Daidzen).

En este punto es interesante mencionar que los valores de  $Z_{score}^{S_{ligando}^S}$  no se correlacionan ni con los valores de RMSD; ni con los valores de  $\Delta IACs$  entre la estructura sin ligando y las demás estructuras. El parámetro IACs (*Inter Atomic Contacts*), hace referencia al numero promedio de interacciones por residuo o  $\langle Degree \rangle$  de cada estructura, por lo que  $\Delta IACs$  es la diferencia de  $\langle Degree \rangle$  entre la estructura sin ligando contra cada estructura unida al ligando. Obtuvimos coeficientes de correlación de Spearman de -0,22 entre  $Z_{score}^{S_{ligando}^S}$  y RMSD; y de -0,07 entre  $Z_{score}^{S_{ligando}^S}$  y  $\Delta IACs$ . Por lo tanto,  $Z_{score}^{S_{ligando}^S}$  no depende de estas características estructurales. Es decir que los ligandos que introducen mayores distorsiones estructurales o cambios en las interacciones entre residuos no corresponden a los que modifican en mayor medida la dinámica de equilibrio del tetrámero.

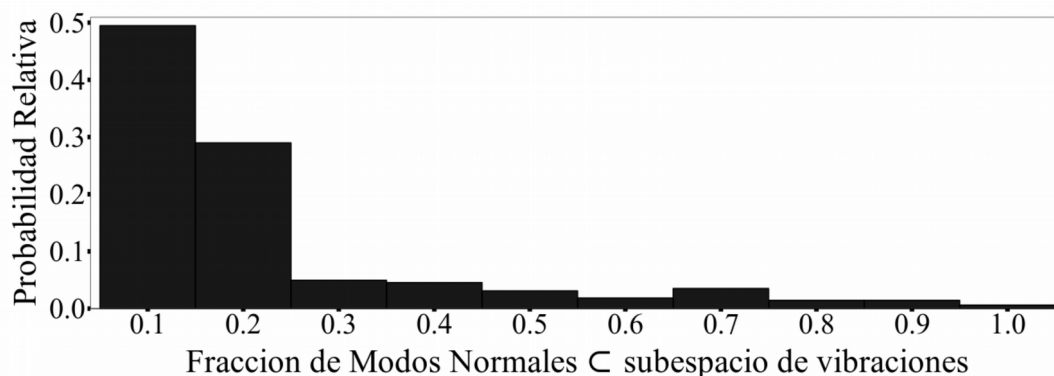
### 4.3.4 Evidencia de asimetría en la dinámica del tetrámero de Transtiretina

A pesar de la presencia en el tetrámero de TTR de dos sitios de unión idénticos, los cuales aparecen ocupados al cristalizarla aparentando una afinidad similar por T4 [Wojtczak 1996], su unión en solución se caracteriza por una fuerte cooperatividad negativa, con aproximadamente dos órdenes de magnitud de diferencia en los valores de Kd para la primera y la segunda molécula de T4 unidas a TTR [Ferguson 1975]. Recientemente, se ha obtenido evidencia adicional que demuestra la heterogeneidad del sitio de unión de TTR tanto en solución, usando el Resveratrol como un ligando fluorescente [Florio 2015], y en estructuras cristalizadas [Cianci 2015].

Actualmente, dentro del *Protein Data Bank*, se pueden encontrar más de 240 estructuras cristalinas de TTR formando complejos con una variedad de ligandos químicamente diferentes, cuya unión a menudo exhibe una cooperatividad negativa. Sin embargo, la base molecular del comportamiento cooperativo y de la heterogeneidad de los sitios de unión de T4 aún no se ha aclarado.

Es por esto que en esta sección demostraremos cómo el análisis de modo normal se ha utilizado para iluminar aspectos asimétricos de la dinámica del tetrámero de TTR.

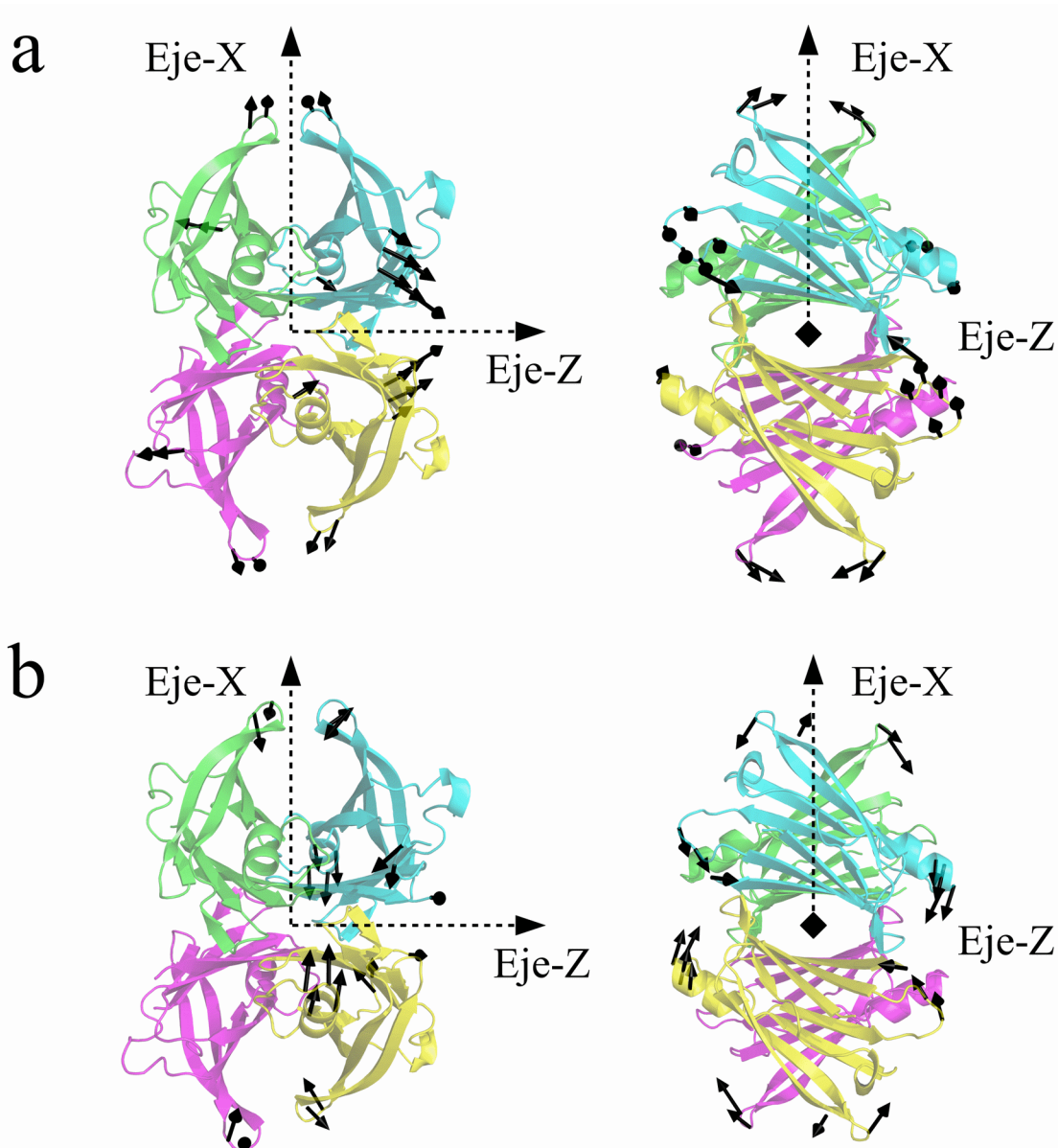
Inicialmente, con el fin de identificar aspectos dinámicos comunes, comparamos los subespacios asociados a la unión de cada ligando al par de sitios de unión del tetrámero, utilizando el subespacio obtenido a partir la unión de T4 como referencia [Zanotti 2017]. La **Figura 4.23** muestra la distribución de la fracción de modos compartidos entre los subespacio y el subespacio de unión a T4.



**Figura 4.23** Distribución de la fracción de modos normales compartidos entre cada subespacio asociado a la unión de un ligando y el subespacio asociado a la unión del T4.

Mientras que la mayoría de los modos normales cambian entre los ligandos, dos modos normales (n° 22 y n° 25) se encuentran en todos los subespacios asociados a la unión del ligando; en la **Figura 4.24** se representan estos modos. Ambos modos presentan una simetría rotacional doble con respecto al eje Z. Implican una combinación de movimientos de tipo *Shear* (modo n° 22) y *Hinge* o bisagra (modo n° 25) entre dímeros en una sola de las cavidades de unión a T4. Estos movimientos relativos se encuentran principalmente localizados en el *loop*-EF, *loop*-FG y *loop*-BC que afectan solo a una de las dos interfaces dímero-dímero [Saldaño 2017]. Estudios previos [Cianci 2015] muestran que los ligandos de TTR presentan diferentes capacidades para saturar los dos sitios de unión de T4, lo que conduce a la presencia de un sitio de unión principal con afinidad clara por el ligando y un segundo sitio con menor afinidad. Nuestros resultados complementan esta evidencia de asimetría estructural, la cual se asocia con una cooperatividad negativa de unión al ligando [Ferguson 1975, Cheng 1977].

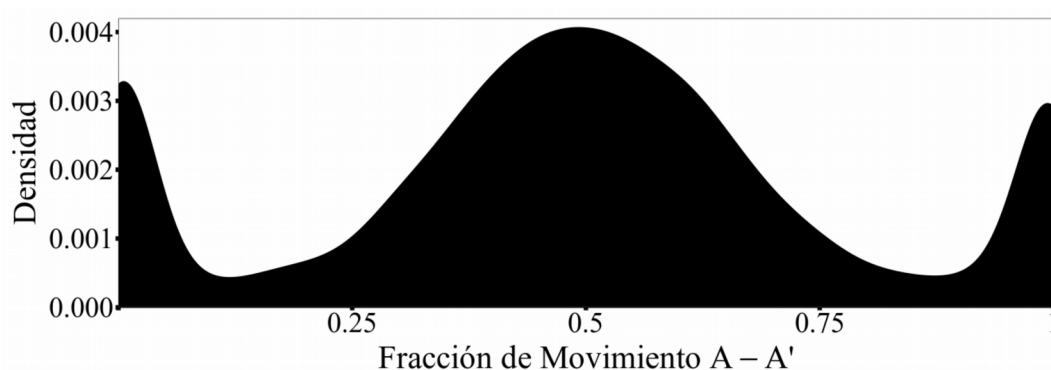
Mediante el análisis de modos normales, hemos analizado las diferencias en las movilidades de los residuos en las subunidades A-A' y B-B'. Para este propósito, se calculó la fracción de desplazamientos relativos de los C $\alpha$  de la subunidad A-A', para cada modo normal del tetrámero de TTR sin ligando (código de pdb 1F41). La distribución de estos valores se muestra en la **Figura 4.25**. El pico en valores de  $\sim 1$  corresponde a modos normales completamente localizados en el dímero A-A', mientras que los modos normales localizados en el dímero B-B' están representados por el pico en  $\sim 0$ . El máximo en  $\sim 0,5$  indica que la mayoría de los modos normales se distribuyen por igual entre ambos dímeros. Sin embargo, la distribución no es completamente simétrica.



**Figura 4.24** Modos normales 22 (a) y 25 (b), participan en los cambios conformacionales para todos los ligandos considerados en nuestro análisis.

Si bien la mayoría de los modos normales se encuentran deslocalizados entre las subunidades A-A' y B-B' (**Figura 4.25**), varios modos están principalmente localizados en uno de ellos. Para analizar más a fondo este hallazgo, los modos normales del tetrámero de TTR se clasificaron de la siguiente manera: (1) modos normales simétricos: vibraciones deslocalizadas entre las subunidades A-A' y B-B', con fracciones de movimientos en la subunidad A-A' dentro del rango [0.45: 0.55] (**Figura 4.25**) y (2) modos asimétricos: modos localizado preferentemente en una subunidad, con fracción de movimientos en la subunidad A-A'  $< 0,45$  o  $> 0,55$ . Los modos (2) se pueden clasificar como: (2a) modos asimétricos por diferencias en amplitudes relativas: modos que implican movimientos similares con diferentes amplitudes en cada subunidad; (2b) modos asimétricos por pares: modos que muestran

diferentes movimientos en cada subunidad, pero para los cuales existe un modo “contraparte” relacionado por simetría rotacional doble, es decir, que implica movimientos equivalentes, pero en la otra subunidad; y (2c) modos completamente asimétricos: modos asimétricos que representan desplazamientos relativos en una subunidad sin una contraparte en la otra subunidad. Siguiendo esta clasificación, hemos encontrado que solo el 18.5%, 1.1% y 16.4% de los modos corresponden a los tipos (1), (2a) y (2b) respectivamente, mientras que el 64% de los modos son completamente asimétricos (2c). Este resultado indica que la mayoría de las vibraciones del tetrámero TTR no presentan una simetría rotacional doble respecto al eje cristalográfico que separa las subunidades A-A' y B-B'. Además, solo algunos de ellos presentan vibraciones que se replican en ambas subunidades. Por lo tanto, se espera que estas asimetrías en los patrones vibracionales de las subunidades A-A' y B-B' se reflejen en diferentes propiedades dinámicas relevantes para la unión al ligando.



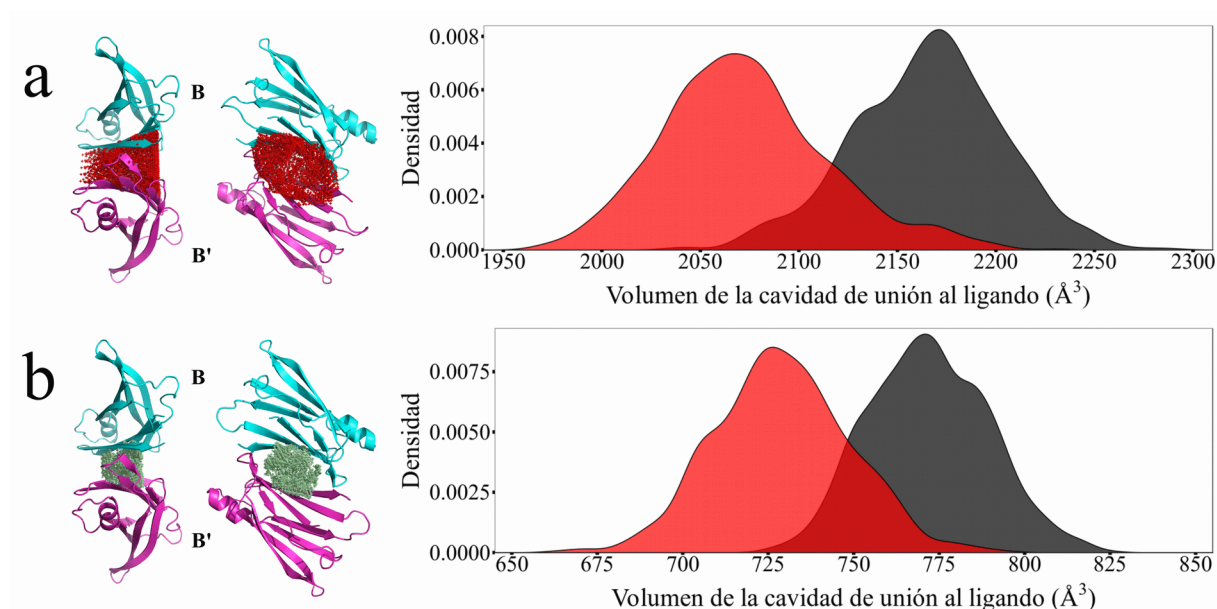
**Figura 4.25 Distribución de la fracción de desplazamientos relativos de los Ca de la subunidad A-A' evaluados en cada modo normal del tetrámero de TTR sin ligando.** Realizamos las distribuciones de Densidad mediante el paquete ggplot2 de R y el *software Rstudio* [Wickham 2016]. La aplicación del suavizado tipo núcleo (*kernel smoothing*) a la distribución de frecuencia de un conjunto de datos produce la estimación de Densidad Núcleo (*Kernel Density Estimate*). Esta es una alternativa útil al histograma ya que que no requiere la elección arbitraria del ancho de la ventana y presenta un resultado suavizado.

Con el fin de analizar aspectos funcionales de las asimetrías estructurales y dinámicas entre las subunidades A-A' y B-B', se calcularon los volúmenes de las cavidades en las que se produce la unión del ligando en ambas interfaces dímero-dímero, para un gran número de estructuras generadas (ver Capítulo 2, sección 2.3.2) representando distorsiones térmicas de la estructura original del tetrámero de TTR sin ligando. Los volúmenes se calcularon mediante un algoritmo [Barletta 2017] que combina el algoritmo de Casco Convexo o Envolverte convexa (*Convex Hull*) [Barber 1996] y las triangulaciones de Delaunay [Hert 2017]. Las cavidades se analizaron considerando los 33 residuos por subunidad que recubren las mismas, enumerados en la **Tabla 4.9**, o teniendo en cuenta solo 10 residuos que interactúan con diferentes ligando [Tomar 2012]. La **Figura 4.26** representa la distribución resultante de los volúmenes de la cavidad de unión al ligando para cada cavidad en las interfaces A-A' y B-B'. Se puede apreciar que, las fluctuaciones térmicas revelan diferencias en el tamaño y la flexibilidad de las cavidades en cada interfaz dímero-dímero. Esto se observa para ambos

tipos de cavidades, las definidas usando todos los residuos que recubren la cavidad o solo aquellos residuos que interactúan con el ligando.

**Tabla 4.9** Residuos que definen la cavidad de unión del ligando de TTR. Los residuos que interactúan con el ligando se resaltan en rojo.

LEU 12	GLU 54	LEU 111
MET 13	LEU 55	SER 112
VAL 14	HIS 56	SER 115
<b>LYS 15</b>	GLY 57	TYR 116
VAL 16	ARG 104	<b>SER 117</b>
<b>LEU 17</b>	TYR 105	<b>THR 118</b>
ASP 18	<b>THR 106</b>	<b>THR 119</b>
SER 50	ILE 107	ALA 120
GLU 51	<b>ALA 108</b>	<b>VAL 121</b>
SER 52	<b>ALA 109</b>	VAL 122
GLY 53	<b>LEU 110</b>	THR 123



**Figura 4.26** Cavidades de unión al ligando y sus correspondientes fluctuaciones térmicas: las cavidades del sitio de unión del ligando se definieron de acuerdo con (a) los 33 residuos por subunidad enumerados en la **Tabla 4.9**, y (b) 10 residuos, los cuales tienen interacción con diferentes ligandos, resaltados en rojo en la **Tabla 4.9**. Las distribuciones de los volúmenes las cavidades del dímero A-A' y del dímero B-B', se presentan en negro y rojo, respectivamente. Los volúmenes fueron calculados para un gran número de estructuras generadas (ver Capítulo 2, sección 2.3.2) que representan las distorsiones térmicas de la estructura sin ligando del tetrámero de TTR (código pdb 1F41). Realizamos las distribuciones de Densidad mediante el paquete ggplot2 de R y el *software Rstudio* [Wickham 2016]. La aplicación del suavizado tipo núcleo (*kernel smoothing*) a la distribución de frecuencia de un conjunto de datos produce la estimación de Densidad Núcleo (*Kernel Density Estimate*). Esta es una alternativa útil al histograma ya que que no requiere la elección arbitraria del ancho de la ventana y presenta un resultado suavizado.

A pesar de la presencia en el tetrámero de dos sitios de unión idénticos, la unión de la tiroxina en solución se caracteriza por una fuerte cooperatividad negativa. Nuestros datos proporcionan evidencia de una significativa flexibilidad estructural y asimetría en la dinámica del tetrámero TTR, y

---

estas características pueden ser asociadas con su comportamiento cooperativo. La unión de un ligando a una de las dos cavidades, la más favorable en el momento de la unión, posiblemente congela la conformación del tetrámero en un estado ligeramente asimétrico, dejando el otro sitio de unión en una conformación menos favorable para la unión de una segunda molécula. La segunda  $K_d$  es generalmente más grande que la primera, pero la unión todavía tiene lugar, lo que sugiere que la perturbación del segundo sitio de unión es relativamente pequeña.





# *Capítulo 5*

## *Conclusiones generales*

*“...el “Eternauta” me llamó él, para explicar en una sola palabra mi condición de navegante del tiempo, de viajero de la eternidad. Mi triste y desolada condición de peregrino de los siglos...”.*

El Eternauta, 1957-1959.

Hector German Oesterheld

Francisco Solano López

## 5. CONCLUSIONES GENERALES

La diversidad conformacional de una proteína implica un equilibrio dinámico entre conformeros con distintas afinidades por el ligando. Los movimientos internos de la proteína garantizan la interconversión entre ellos. Debido a su relevancia para la función de la proteína, la diversidad conformacional asociada a la unión del ligando debe conservarse evolutivamente. En esta tesis, hemos presentado un novedoso procedimiento para identificar las posiciones clave cuyas mutaciones tienen un efecto significativo en los modos normales de vibración implicados en el cambio conformacional asociado a la unión del ligando. Hemos aplicado nuestro método a un conjunto refinado de estructuras proteicas, considerando para cada proteína el conformero con y sin ligando.

Con el fin de evitar mezclas de modos normales y/o reordenamientos de los mismos debido a cambios en la frecuencia generados por la unión del ligando, no analizamos modos normales individuales sino que utilizamos subespacios de modos normales. Los modos dentro de estos subespacios describen las vibraciones asociadas al cambio conformacional por la unión del ligando. Hemos descrito un procedimiento para definir y comparar estos subespacios. Nuestra definición de posiciones claves, es decir, las posiciones que son dinámicamente importantes para la unión de ligando, se basa en el efecto de las mutaciones en estos subespacios.

Encontramos una correlación negativa entre los efectos de mutaciones sitio-específicas sobre los subespacios de modos normales asociados a la unión del ligando y la conservación evolutiva de estos sitios. Los residuos cuyas mutaciones alteran en mayor medida el subespacio se definen como posiciones claves. Se encontró que los residuos en posiciones claves son en su mayoría alifáticos, localizados en elementos de estructuras secundarias bien definidas como hojas- $\beta$  y hélice- $\alpha$ . Además, parecen participar como residuos pivote a través de contactos entre estos elementos de estructura secundaria.

Los residuos en posiciones claves se identificaron utilizando subespacios de vibraciones colectivas que participan de un cambio conformacional específico. Estas vibraciones colectivas son descritas frecuentemente por modos normales de baja frecuencia, que implican movimientos concertados de residuos que pueden localizarse en regiones separadas de la estructura proteica. Por lo tanto, el método no se ve afectado por ningún sesgo que pueda sobreestimar el efecto de los residuos localizados cerca del sitio de unión al ligando. Hemos demostrado que sólo el 10% de los residuos en posiciones claves corresponden a residuos del sitio activo. Es decir, los residuos del sitio activo sólo comprenden una pequeña fracción de los residuos claves predichos.

Nuestros residuos en posiciones claves están asociados a un aspecto muy particular de la funcionalidad de la proteína, es decir, a las vibraciones asociadas a las distorsiones estructurales introducidas por la unión al ligando. En este sentido, el análisis proporciona información distinta y

complementaria respecto a estudios basados en la identificación de sitios activos secuenciales y estructurales mediante similitudes entre proteínas homólogas.

El método desarrollado en esta tesis no está restringido solo a identificar residuos de posición clave cuyas mutaciones afectan directamente la afinidad por sustrato. Puede aplicarse directamente para identificar residuos en posición clave cuyas mutaciones afecten: la constante oligomerización, la estabilidad, las interacciones inter-proteínas o respuestas alostéricas.

Debido a que parte de la funcionalidad de una proteína reside en su capacidad de realizar las transiciones conformacionales requeridas, pensamos que nuestro método para estimar las posiciones claves relacionadas con la dinámica de proteínas podría ayudarnos a mejorar nuestra comprensión de la relación estructura-función, así como la diversificación funcional durante la evolución.

En esta tesis combinamos la información obtenida de métodos basados en propiedades estructurales y dinámicas de proteínas con información relacionada con redes de interacción de residuos. Encontramos una buena correlación entre los efectos mutaciones sitio-específicas en los subespacios de los modos normales asociados a la unión del ligando y tres parámetros topológicos de redes. Estos parámetros han sido previamente propuestos, prediciendo con éxito sitios activos o sitios funcionales en varias proteínas [Amitai 2004, Thibert 2005]. Observamos que nuestros residuos claves tienden a estar interconectados, por lo que pudimos definir redes de residuos que modulan los cambios conformacionales.

Sugerimos que nuestro método para definir los residuos claves, conduce a una comprensión más profunda de la relación entre la estructura una proteína y sus sitios funcionales específicos. Esta relación es probablemente una consecuencia directa de la robustez funcional necesaria de las proteínas.

En la parte final de la tesis, aplicamos y adaptamos nuestra metodología al estudio específico del efecto de mutaciones y de la unión del ligando sobre la dinámica de equilibrio de la estructura tetramérica de la Transtiretina Humana (TTR). Exploramos los residuos en posiciones claves que afectan a las vibraciones del tetrámero de TTR. La mayoría de ellos están localizados en la hoja- $\beta$  E, hoja- $\beta$  F y el loop EF, con mutaciones amiloidogénicas previamente reportadas. Se identificó a la interfase monómero-monómero como una de las regiones más vulnerables a mutaciones que conducen a cambios significativos en la dinámica del tetrámero. Además, hemos encontrado que mutaciones en los residuos localizados en la interfaz dímero-dímero y/o en el sitio de unión a la hormona T4 desestabilizan al tetrámero más que el promedio.

Mediante el análisis del efecto de mutaciones en la dinámica de equilibrio del tetrámero TTR. Se espera que este aspecto particular, entre otras características estructurales, esté asociado a la estabilidad de la estructura cuaternaria.

Teniendo en cuenta la diversidad de razones que llevan a una mutación a ser pro-amiloide, no esperamos una correlación entre las mutaciones amiloidogénicas informadas y el valor  $Z_{score}^i$ , que

sólo cuantifica las perturbaciones de la dinámica de equilibrio del tetrámero TTR. Sin embargo, los valores  $Z_{score}^S$  no sólo representan una contribución complementaria para explorar potenciales mutaciones, sino también permiten entender sobre los efectos de algunas de las mutaciones informadas.

Hemos comparado varios fármacos y compuestos de acuerdo con su efecto sobre las vibraciones asociadas a la unión del ligando. Se encontró que una parte significativamente pequeña del espacio vibracional participa de las transiciones conformacionales inducidas por la unión del ligando. Hemos discutido los efectos de los ligandos particularmente en términos de parámetros experimentales, mediciones asociadas a la afinidad por el ligando y la estabilización del estado nativo del tetrámero. En general, el efecto de los ligandos sobre la dinámica de equilibrio del tetrámero sin ligando concuerda con hallazgos de grupos experimentales obtenidos mediante ensayos de unión competitiva.

Se encontraron dos modos normales presentes en todos los subespacios determinados por los diferentes ligandos. Ambos implican movimientos entre dímeros en sólo una de las cavidades de unión a T4, evidenciando la unión asimétrica del ligando, relacionada con su cooperatividad negativa.

Nuestros datos proporcionan evidencia de una significativa flexibilidad estructural y asimetría en la dinámica del tetrámero TTR, y estas características pueden ser asociadas con su comportamiento cooperativo. Como consecuencia de las asimetrías en los patrones vibracionales de ambas subunidades A-A' y B-B', las fluctuaciones térmicas conducen a diferencias en el tamaño y la flexibilidad de las cavidades de unión del ligando (véase la **Figura 4.26**). Estas diferencias son mayores entre las cavidades, definidas por todos los residuos que delimitan la misma, que entre las cavidades pequeñas, definidas solo por los residuos que interactúan con el ligando. Por lo tanto, nuestros resultados indican posibles diferencias en la unión del ligando, la entrada y difusión del ligando a lo largo de ambas cavidades.

Consideramos que los resultados obtenidos en esta tesis representan un aporte complementario que puede ayudar a futuros desarrollos de nuevos fármacos contra la amiloidosis asociada a la Transteritina.

# *Apéndice*

**A. LISTA DE AMINOÁCIDOS**

<b>Nombre</b>	<b>Código de tres letras</b>	<b>Código de una letra</b>
Alanina	Ala	A
Cisteína	Cys	C
Aspartato	Asp	D
Glutamato	Glu	E
Fenilalanina	Phe	F
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Lisina	Lys	K
Leucina	Leu	L
Metionina	Met	M
Asparagina	Asn	N
Prolina	Pro	P
Glutamina	Gln	Q
Arginina	Arg	R
Serina	Ser	S
Treonina	Thr	T
Valina	Val	V
Triptófano	Trp	W
Tirosina	Tyr	Y

**B. PROTEÍNAS DEL SET DE ESTUDIO**

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1ACP	A	Solución NMR	2FAE	B	Difracción de Rayos X	4,69	77	4	13	
1ADE	B	Difracción de Rayos X	1NHT	A	Difracción de Rayos X	2,44	431	22	8	0,83
1AEL	A	Solución NMR	1URE	A	Solución NMR	2,69	131	7	13	
1AFH	A	Solución NMR	1FK3	A	Difracción de Rayos X	2,27	93	5	13	
1AKZ	A	Difracción de Rayos X	1SSP	E	Difracción de Rayos X	1,50	223	12	14	0,68
1APC	A	Solución NMR	1QPU	A	Solución NMR	4,00	106	6	13	
1ASY	A	Difracción de Rayos X	1ASZ	B	Difracción de Rayos X	0,92	490	25	14	0,41
1AW0	A	Solución NMR	2AW0	A	Solución NMR	1,05	72	4	13	
1B4M	A	Solución NMR	1EII	A	Solución NMR	3,17	134	7	13	
1BMG	A	Difracción de Rayos X	3PWV	E	Difracción de Rayos X	1,03	98	5	7	0,64
1BV2	A	Solución NMR	1UVB	A	Difracción de Rayos X	1,96	91	5	13	
1BVM	A	Solución NMR	2BPP	A	Difracción de Rayos X	2,03	123	7	19	
1BYI	A	Difracción de Rayos X	1DAF	A	Difracción de Rayos X	0,93	224	12	20	0,57
1C54	A	Solución NMR	1RGH	B	Difracción de Rayos X	1,49	96	5	17	
1CBI	A	Difracción de Rayos X	1CBR	B	Difracción de Rayos X	1,64	136	7	13	0,7
1CGS	H	Difracción de Rayos X	2CGR	H	Difracción de Rayos X	3,91	214	11	13	0,48
1CMS	A	Difracción de Rayos X	1CZI	E	Difracción de Rayos X	1,12	323	17	13	0,65
1COE	A	Solución NMR	1V6P	B	Difracción de Rayos X	2,83	62	4	13	
1CXO	A	Solución NMR	1TGX	C	Difracción de Rayos X	3,07	60	4	13	

\* N, se refiere al número de residuos.

† N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.



Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1DGU	A	Solución NMR	1Y1A	B	Difracción de Rayos X	16,88	183	10	13	
1DZ7	A	Solución NMR	1E9J	A	Solución NMR	16,42	92	5	13	
1E0L	A	Solución NMR	2RLY	W	Solución NMR	4,23	37	2	13	
1E8J	A	Solución NMR	2DSX	A	Difracción de Rayos X	1,66	52	3	13	
1EAL	A	Solución NMR	1EIO	A	Solución NMR	2,39	127	7	13	
1ED0	A	Solución NMR	1OKH	B	Difracción de Rayos X	1,13	46	3	13	
1F3C	B	Solución NMR	1F95	B	Solución NMR	3,71	89	5	13	
1F3Y	A	Solución NMR	1JKN	A	Solución NMR	5,08	165	9	13	
1F54	A	Solución NMR	1F55	A	Solución NMR	6,63	77	4	13	
1F6P	D	Difracción de Rayos X	1F74	C	Difracción de Rayos X	2,03	293	15	12	0,58
1F70	A	Solución NMR	1J7O	A	Solución NMR	3,98	76	4	15	0,78
1F7E	A	Solución NMR	1FF7	A	Solución NMR	5,16	46	3	13	
1FES	A	Solución NMR	2GGP	A	Solución NMR	2,11	73	4	13	
1FKR	A	Solución NMR	1J4R	D	Difracción de Rayos X	1,93	107	6	13	
1FMF	A	Solución NMR	1ID8	A	Solución NMR	4,74	137	7	13	
1FRA	A	Solución NMR	3EBX	A	Difracción de Rayos X	1,57	62	4	20	0,49
1FSF	A	Difracción de Rayos X	1FQO	B	Difracción de Rayos X	1,11	266	14	9	0,88
1FVQ	A	Solución NMR	2GGP	B	Solución NMR	1,42	72	4	13	
1G7H	C	Difracción de Rayos X	1LKR	B	Difracción de Rayos X	1,90	129	7	19	0,64
1GAU	A	Solución NMR	1GAT	A	Solución NMR	1,24	60	4	13	

\* N, se refiere al número de residuos.

<sup>†</sup> N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1GFD	A	Solución NMR	1IO6	A	Solución NMR	1,68	59	3	18	0,91
1GH1	A	Solución NMR	1CZ2	A	Solución NMR	3,04	90	5	13	
1GQN	A	Difracción de Rayos X	1QFE	B	Difracción de Rayos X	1,26	252	13	9	0,66
1GQZ	A	Difracción de Rayos X	2GKE	A	Difracción de Rayos X	2,45	274	14	10	0,54
1GWP	A	Solución NMR	1AFV	A	Difracción de Rayos X	3,55	151	8	13	
1HDN	A	Solución NMR	2JEL	P	Difracción de Rayos X	1,59	85	5	13	
1HEV	A	Solución NMR	1Q9B	A	Difracción de Rayos X	1,31	43	3	13	
1HPJ	A	Solución NMR	1HPK	A	Solución NMR	2,02	79	4	13	
1HSI	B	Difracción de Rayos X	1HSH	D	Difracción de Rayos X	0,99	99	5	18	0,6
1I56	A	Solución NMR	1EL1	A	Difracción de Rayos X	1,95	130	7	19	0,89
1I6C	A	Solución NMR	1I8G	B	Solución NMR	4,68	39	2	13	
1IGP	A	Difracción de Rayos X	2AU6	A	Difracción de Rayos X	1,12	175	9	15	0,63
1IJA	A	Solución NMR	2KID	A	Solución NMR	4,54	148	8	13	
1IXA	A	Solución NMR	1EDM	B	Difracción de Rayos X	1,68	39	2	13	
1JFJ	A	Solución NMR	1JFK	A	Solución NMR	14,73	134	7	13	
1JJX	A	Solución NMR	1FE3	A	Difracción de Rayos X	1,12	131	7	13	
1JM4	B	Solución NMR	1WUM	A	Solución NMR	5,33	118	6	13	
1JOR	A	Solución NMR	1JOQ	A	Solución NMR	6,36	149	8	13	
1JSP	B	Solución NMR	2L84	A	Solución NMR	6,10	121	7	13	
1JYU	A	Difracción de Rayos X	1JYR	A	Difracción de Rayos X	12,78	96	5	9	0,58

\* N, se refiere al número de residuos.

† N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1K2H	A	Solución NMR	1ZFS	B	Solución NMR	6,59	93	5	13	
1KEM	H	Difracción de Rayos X	1KEL	H	Difracción de Rayos X	1,17	218	11	15	0,54
1KUL	A	Solución NMR	1ACZ	A	Solución NMR	4,20	108	6	13	
1KUN	A	Solución NMR	2KNT	A	Difracción de Rayos X	2,50	58	3	13	
1KVI	A	Solución NMR	1KVJ	A	Solución NMR	3,53	79	4	13	
1L3H	A	Solución NMR	1ICF	I	Difracción de Rayos X	2,66	65	4	13	
1LIP	A	Solución NMR	1JTB	A	Solución NMR	3,30	91	5	13	
1LMZ	A	Solución NMR	1P7M	A	Solución NMR	3,81	187	10	13	
1LQI	A	Solución NMR	2ASC	A	Difracción de Rayos X	1,28	65	4	13	
1LRE	A	Solución NMR	2FYL_A	A	Solución NMR	3,52	81	5	13	
1LZ4	A	Difracción de Rayos X	208L	A	Difracción de Rayos X	1,36	130	7	19	0,47
1MJM	B	Difracción de Rayos X	1MJO	D	Difracción de Rayos X	2,22	104	6	19	0,52
1MO7	A	Solución NMR	1MO8	A	Solución NMR	4,93	213	11	13	
1MPH	A	Solución NMR	1BTN	A	Difracción de Rayos X	2,42	106	6	10	0,79
1MR4	A	Solución NMR	4AAZ	B	Difracción de Rayos X	1,58	47	3	13	
1MUT	A	Solución NMR	1PUN	A	Solución NMR	4,98	129	7	13	
1MX7	A	Solución NMR	1MX8	A	Solución NMR	2,25	134	7	13	
1NEA	A	Solución NMR	1IQ9	A	Difracción de Rayos X	2,09	61	4	13	
1NGZ	A	Difracción de Rayos X	1N7M	H	Difracción de Rayos X	1,52	213	11	7	0,66
1NM4	A	Solución NMR	1OT4	A	Solución NMR	3,61	102	6	13	

\* N, se refiere al número de residuos.

<sup>†</sup> N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1NTI	A	Solución NMR	1ACA	A	Solución NMR	3,22	86	5	9	0,7
1NTR	A	Solución NMR	1KRX	A	Solución NMR	4,55	124	7	13	
1NVO	B	Solución NMR	1OVR	D	Difracción de Rayos X	2,45	48	3	13	
1O1U	A	Solución NMR	1O1V	A	Solución NMR	2,25	127	7	13	
1OIB	B	Difracción de Rayos X	1IXG	A	Difracción de Rayos X	3,06	321	17	19	0,61
1OMT	A	Solución NMR	1PPF	I	Difracción de Rayos X	2,81	56	3	13	
1OQ3	A	Solución NMR	1OQ6	A	Solución NMR	2,51	76	4	14	0,87
1ORM	A	Solución NMR	1QJ8	A	Difracción de Rayos X	6,13	148	8	13	
1P8G	A	Solución NMR	1K0V	A	Solución NMR	2,89	73	4	13	
1PCF	A	Difracción de Rayos X	2C62	B	Difracción de Rayos X	1,18	66	4	9	0,69
1PDB	A	Difracción de Rayos X	1YHO	A	Difracción de Rayos X	1,66	186	10	18	0,42
1PFL	A	Difracción de Rayos X	1FIL	A	Solución NMR	1,90	139	7	13	
1PIT	A	Solución NMR	5PTI	A	Difracción de Rayos X	2,37	58	3	10	0,62
1PON	A	Solución NMR	1CTD	B	Solución NMR	4,96	34	2	13	
1PV6	B	Difracción de Rayos X	2CFQ	A	Difracción de Rayos X	1,16	417	21	17	0,65
1QFN	A	Solución NMR	1GRX	A	Solución NMR	2,65	85	5	13	
1QKD	A	Difracción de Rayos X	5EBX	A	Difracción de Rayos X	0,83	62	4	20	0,42
1R57	A	Solución NMR	2H5M	A	Solución NMR	5,71	102	6	13	
1RCH	A	Solución NMR	1RDD	A	Difracción de Rayos X	2,16	155	8	13	
1RK7	A	Solución NMR	1KMG	A	Solución NMR	4,93	153	8	13	

\* N, se refiere al número de residuos.

<sup>†</sup> N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1S6O	A	Solución NMR	1S6U	A	Solución NMR	2,71	76	4	13	
1SBT	A	Difracción de Rayos X	2SBT	A	Difracción de Rayos X	1,05	275	14	13	
1SF1	B	Solución NMR	4AIY	L	Solución NMR	6,66	30	2	13	
1SKT	A	Solución NMR	1TNQ	A	Solución NMR	7,19	90	5	13	
1SPY	A	Solución NMR	2KRD	C	Solución NMR	6,48	89	5	9	0,74
1SYM	B	Solución NMR	1XYD	B	Solución NMR	6,69	92	5	13	
1TFU	A	Difracción de Rayos X	3UC5	A	Difracción de Rayos X	1,06	157	8	8	0,73
1TGK	A	Difracción de Rayos X	1TGJ	A	Difracción de Rayos X	0,84	112	6	7	0,49
1TJD	A	Difracción de Rayos X	1EEJ	B	Difracción de Rayos X	2,08	216	11	19	0,65
1TL5	A	Solución NMR	2K1R	B	Solución NMR	1,97	68	4	12	0,86
1TVQ	A	Difracción de Rayos X	1TW4	B	Difracción de Rayos X	1,25	125	7	7	0,52
1TW0	A	Difracción de Rayos X	1TXC	B	Difracción de Rayos X	0,90	157	8	7	0,56
1U0N	A	Difracción de Rayos X	1AUQ	A	Difracción de Rayos X	0,68	207	11	7	0,75
1U37	A	Solución NMR	1U38	A	Solución NMR	1,22	89	5	13	
1URF	A	Solución NMR	2RMK	B	Solución NMR	5,71	81	5	13	
1URP	D	Difracción de Rayos X	2DRI	A	Difracción de Rayos X	4,20	271	14	9	0,49
1USG	A	Difracción de Rayos X	1USI	C	Difracción de Rayos X	7,14	345	18	7	0,72
1W4U	A	Solución NMR	1UR6	A	Solución NMR THEORETIC AL MODEL	3,01	147	8	8	0,49
1WCR	A	Solución NMR	2LRL	C	Solución NMR	2,04	103	6	11	0,83

\* N, se refiere al número de residuos.

<sup>†</sup> N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
1XOA	A	Solución NMR	2H6X	B	Difracción de Rayos X	1,37	108	6	13	
1XSA	A	Solución NMR	1XSC	A	Solución NMR	4,76	153	8	13	
1Y3K	A	Solución NMR	1Y3J	A	Solución NMR	5,15	77	4	12	0,89
1YJR	A	Solución NMR	1YJT	A	Solución NMR	3,24	75	4	9	0,72
1YP5	A	Difracción de Rayos X	1ZUY	A	Difracción de Rayos X	1,04	58	3	7	0,63
1Z15	A	Difracción de Rayos X	1Z17	A	Difracción de Rayos X	6,45	344	18	9	0,82
1Z7T	A	Solución NMR	1Z6H	A	Solución NMR	2,78	72	4	13	
2A8F	B	Difracción de Rayos X	2AIB	B	Difracción de Rayos X	1,05	98	5	9	0,68
2AI6	A	Solución NMR	2OZW	A	Solución NMR	3,09	125	7	13	
2BCB	A	Solución NMR	1B1G	A	Solución NMR	2,10	75	4	13	
2BNH	A	Difracción de Rayos X	1DFJ	I	Difracción de Rayos X	1,51	456	23	20	0,61
2CG7	A	Difracción de Rayos X	2CG6	A	Difracción de Rayos X	6,25	90	5	14	0,45
2CJO	A	Solución NMR	1ROE	A	Solución NMR	4,55	97	5	9	0,54
2CRS	A	Solución NMR	1H0J	A	Difracción de Rayos X	2,65	60	4	13	
2D9E	A	Solución NMR	2RS9	B	Solución NMR	5,30	121	7	13	
2EZW	A	Solución NMR	3IM3	A	Difracción de Rayos X	5,84	50	3	13	
2F63	A	Solución NMR	1EQM	A	Difracción de Rayos X	4,62	158	8	13	0,67
2FHM	A	Solución NMR	2HLT	A	Solución NMR	1,90	91	5	13	
2FKX	A	Solución NMR	3MR8	O	Difracción de Rayos X	10,28	88	5	13	
2FUI	A	Solución NMR	2FUU	A	Solución NMR	7,65	62	4	13	

\* N, se refiere al número de residuos.

† N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	γ
2HQI	A	Solución NMR	1AFJ	A	Solución NMR	2,71	72	4	13	
2IN2	A	Solución NMR	2B0F	A	Solución NMR	2,14	182	10	13	
2JU3	A	Solución NMR	2JU8	A	Solución NMR	3,17	127	7	13	
2JWW	A	Solución NMR	1RTP	1	Difracción de Rayos X	2,79	109	6	13	
2K0Y	A	Solución NMR	2KG9	A	Solución NMR	3,25	86	5	10	0,76
2K3S	B	Solución NMR	2LLQ	A	Solución NMR	4,97	67	4	13	
2K43	A	Solución NMR	2K8R	A	Solución NMR	3,93	133	7	20	0,73
2K45	A	Solución NMR	1BYN	A	Solución NMR	2,59	128	7	14	0,57
2K6Y	A	Solución NMR	2K70	A	Solución NMR	3,77	120	7	7	0,76
2K7C	A	Solución NMR	2K7D	A	Solución NMR	8,43	72	4	13	
2KAX	A	Solución NMR	2KAY	B	Solución NMR	5,08	92	5	9	0,68
2KE0	A	Solución NMR	2KO7	A	Solución NMR	4,85	117	6	13	
2KM2	B	Solución NMR	1GZW	B	Difracción de Rayos X	2,03	134	7	13	
2KT2	A	Solución NMR	2KT3	A	Solución NMR	1,75	69	4	13	
2KWH	A	Solución NMR	2KWI	B	Solución NMR	4,08	56	3	13	
2KXL	A	Solución NMR	2K0G	A	Solución NMR	6,09	142	8	13	
2KYC	A	Solución NMR	2KYF	A	Solución NMR	2,48	108	6	13	
2L28	A	Solución NMR	2HQP	A	Solución NMR	3,42	162	9	13	
2L50	A	Solución NMR	2L51	B	Solución NMR	4,56	102	6	8	0,86
2L68	A	Solución NMR	2LKK	A	Solución NMR	1,58	126	7	13	

\* N, se refiere al número de residuos.

† N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	γ
2LAO	A	Difracción de Rayos X	1LAH	E	Difracción de Rayos X	4,75	238	12	11	0,56
2LHS	A	Solución NMR	2BEM	A	Difracción de Rayos X	1,52	170	9	13	
2LKC	A	Solución NMR	2LKD	A	Solución NMR	7,47	178	9	12	0,68
2LPK	A	Solución NMR	2LL8	A	Solución NMR	7,84	101	6	13	
2LPV	A	Solución NMR	3UQI	A	Difracción de Rayos X	1,98	108	6	13	
2LTW	A	Solución NMR	2LAY	A	Solución NMR	2,98	36	2	13	
2LVI	A	Solución NMR	2LVK	A	Solución NMR	4,23	77	4	13	
2LYS	A	Solución NMR	1UTX	A	Difracción de Rayos X	5,96	66	4	13	
2MCG	2	Difracción de Rayos X	1A8J	H	Difracción de Rayos X	12,97	215	11	10	0,53
2NLN	A	Solución NMR	1RRO	A	Difracción de Rayos X	3,90	108	6	13	
2NX5	Q	Difracción de Rayos X	3MV7	A	Difracción de Rayos X	1,80	276	14	8	0,72
2OMJ	A	Solución NMR	2OS6	A	Solución NMR	6,60	89	5	13	
2OMZ	B	Difracción de Rayos X	2OMU	B	Difracción de Rayos X	0,99	104	6	20	0,6
2P3M	A	Solución NMR	2VBT	A	Difracción de Rayos X	2,78	136	7	13	
2QMV	A	Solución NMR	4PRG	D	Difracción de Rayos X	4,20	270	14	13	
2R90	D	Difracción de Rayos X	2R2K	A	Difracción de Rayos X	1,32	171	9	19	0,7
2RR0	A	Solución NMR	2RQZ	A	Solución NMR	4,28	38	2	13	
2UZ5	A	Solución NMR	2VCD	A	Solución NMR	3,37	137	7	13	
3P2J	A	Difracción de Rayos X	3KK0	A	Difracción de Rayos X	1,61	191	10	19	0,71
3PHV	A	Difracción de Rayos X	2QNP	B	Difracción de Rayos X	1,38	99	5	13	

\* N, se refiere al número de residuos.

† N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.



Sin ligando	Cadena	Método experimental	Con ligando	Cadena	Método experimental	RMSD [Å]	N*	N <sub>PC</sub> <sup>†</sup>	Radio de corte	$\gamma$
3RDJ	A	Difracción de Rayos X	3GPE	A	Difracción de Rayos X	1,31	137	7	18	0,54
4AKE	B	Difracción de Rayos X	2ECK	B	Difracción de Rayos X	6,95	214	11	18	0,8
4ICB	A	Difracción de Rayos X	1IGV	A	Difracción de Rayos X	2,35	75	4	8	0,79
4ULL	A	Solución NMR	2C5C	J	Difracción de Rayos X	2,52	69	4	13	
8PTI	A	Difracción de Rayos X	2FTM	B	Difracción de Rayos X	2,01	58	3	10	0,46

\* N, se refiere al número de residuos.

<sup>†</sup> N<sub>PC</sub>, se refiere al número de residuos en posiciones claves.

### C. TEOREMA DE EQUIPARTICIÓN [Tipler 2008]

La derivación general del Teorema de Equipartición implica conocimientos de Mecánica Estadística más allá del alcance de esta tesis. Sin embargo, se puede demostrar el Teorema de Equipartición a partir de casos especiales, particularmente considerando Osciladores Armónicos. Un Oscilador Armónico consiste en una partícula que se mueve en una dimensión bajo la acción de una fuerza recuperadora elástica, proporcional a la constante del resorte ( $\kappa$ ) y a la distancia al punto de equilibrio ( $x$ ).

La energía cinética de la partícula en cualquier instante es  $\frac{1}{2}m v_x^2$  y su la energía potencial es  $\frac{1}{2}\kappa x^2$ , por lo que la energía total  $E$ , es

$$E = \frac{1}{2}m v_x^2 + \frac{1}{2}\kappa x^2 \quad (\text{Ec. C.1})$$

Boltzmann derivó una relación de distribución que hizo posible la predicción del número probable de partículas que ocuparán cada uno de los estados de energía disponibles en dicho sistema en equilibrio térmico. La distribución de Boltzmann

$$f_B(E) = A e^{-E/k_B T} \quad (\text{Ec. C.2})$$

es la función de distribución fundamental de la física estadística clásica, donde  $A$  llamado Factor de Boltzmann es una constante de normalización cuyo valor depende del sistema particular que se está considerando y  $k_B$  es la constante de Boltzmann.

La probabilidad de que un oscilador tenga energía  $E$  correspondiente a  $v_x$  en  $dv_x$  y  $x$  en  $dx$  será

$$f_B(E) dx dv_x = A e^{-(m v_x^2/2k_B T + \kappa x^2/2k_B T)} dx dv_x \quad (\text{Ec. C.3})$$

La constante  $A$  se determina a partir de la condición de normalización donde la probabilidad de encontrar a la partícula en algún punto del espacio debe ser igual a 1, es decir

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} A e^{-(m v_x^2/2k_B T + \kappa x^2/2k_B T)} dx dv_x = 1 \quad (\text{Ec. C.4})$$

$$A \int_{-\infty}^{+\infty} e^{-(m v_x^2/2k_B T)} dv_x \int_{-\infty}^{+\infty} e^{-(\kappa x^2/2k_B T)} dx = 1$$

Cada una de las integrales en la Ec. C.4 es de la misma forma y al resolverlas obtenemos

$$A = \left( \frac{m}{2\pi k_B T} \right)^{1/2} \left( \frac{\kappa}{2\pi k_B T} \right)^{1/2} \quad (\text{Ec. C.5})$$

La energía promedio de un oscilador está dada por

$$\langle E \rangle = \int \int E f_B(E) dx dv_x = A \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left( \frac{1}{2}m v_x^2 + \frac{1}{2}\kappa x^2 \right) e^{-(m v_x^2/2k_B T)} e^{-(\kappa x^2/2k_B T)} dx dv_x \quad (\text{Ec. C.6})$$

$$\begin{aligned} \langle E \rangle = & A \int_{-\infty}^{+\infty} \frac{1}{2} m v_x^2 e^{-(m v_x^2 / 2 k_B T)} dv_x \int_{-\infty}^{+\infty} e^{-(\kappa x^2 / 2 k_B T)} dx \\ & + A \int_{-\infty}^{+\infty} e^{-(m v_x^2 / 2 k_B T)} dv_x \int_{-\infty}^{+\infty} \frac{1}{2} \kappa x^2 e^{-(\kappa x^2 / 2 k_B T)} dx \end{aligned} \quad (\text{Ec. C.7})$$

Definimos el primer término en la Ec. C.7 como la energía cinética promedio del oscilador

$\left\langle \frac{1}{2} m v_x^2 \right\rangle$  y el segundo término como la energía potencial promedio  $\left\langle \frac{1}{2} \kappa x^2 \right\rangle$  del oscilador. Al resolver

las integrales y multiplicar por A obtenemos

$$\langle E \rangle = \left\langle \frac{1}{2} m v_x^2 \right\rangle + \left\langle \frac{1}{2} \kappa x^2 \right\rangle = \frac{1}{2} k_B T + \frac{1}{2} k_B T = k_B T \quad (\text{Ec. C.8})$$

De la Ec. C.8 se determina que tanto la energía cinética promedio como la energía potencial

promedio dependen solo de la temperatura absoluta y que cada valor promedio es igual a  $\frac{1}{2} k_B T$ .

Este resultado obtenido considerando un Oscilador Armónico es un caso especial del Teorema de Equipartición de la Energía, el cual establece que en equilibrio cada grado de libertad

contribuye  $\frac{1}{2} k_B T$  a la energía promedio  $\langle E \rangle$  por molécula.

## D. CORRELACIÓN DE PEARSON Y DE SPEARMAN

Una medida de correlación es una variable aleatoria que se usa para muestras bivariadas, es decir datos consistentes en pares de números. La muestra bivariada de tamaño  $n$  se representa  $(X, Y) = (X_1, Y_1), \dots, (X_n, Y_n)$ . Una medida de la correlación entre  $X$  e  $Y$  debe cumplir las siguientes condiciones:

- i.* La medida de correlación debe estar en el rango  $[-1, 1]$ .
- ii.* Si los menores valores de  $X$  tienden a estar apareados con los menores valores de  $Y$  y los mayores valores de  $X$  con los mayores valores de  $Y$ , la correlación es positiva y la medida debe tender a 1.
- iii.* Si los menores valores de  $X$  tienden a estar apareados con los mayores valores de  $Y$  y viceversa, la correlación es negativa y la medida debe tender a -1.
- iv.* Si los valores de  $X$  tienden a estar aleatoriamente apareados con los valores de  $Y$ , por ejemplo en el caso particular de  $X$  e  $Y$  independientes, se dice que  $X$  y  $Y$  no están correlacionados y la medida debe tender a cero.

La medida de correlación más usual es el coeficiente  $r$  de producto de momentos de Pearson

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X}) \sum_{i=1}^n (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (\text{Ec. D.1})$$

donde la barra superior representa los valores medios de cada muestra. El coeficiente  $r$  es el cociente entre la covarianza de las muestras y el producto de sus desviaciones estándar. Da una medida de la asociación lineal entre  $X$  e  $Y$ , es decir que  $r = 1$  si el gráfico  $Y$  *versus*  $X$  es una recta con pendiente positiva y  $r = -1$  si es una recta con pendiente negativa. El coeficiente  $r$  puede calcularse para cualquier clase de datos numéricos, independientemente de la escala de medida o del tipo de distribución de las muestras. Sin embargo,  $r$  es una variable aleatoria y como tal tiene una función de distribución, que depende de la función de distribución bivariada de las muestras  $(X, Y)$ . Por este motivo,  $r$  tiene valor estadístico si se asume una distribución normal de las muestras, pero no es aplicable como estadístico en pruebas no paramétricas, a menos que se conozca la distribución de los datos [Conover 1980].

Existen otras medidas de correlación definidas de tal manera que sus funciones de distribución no dependen de la función distribución bivariada de  $(X, Y)$  si  $X$  e  $Y$  son independientes y continuas. Muchas de estas medidas están basadas en el rango que se asigna a los datos. En la muestra bivariada de tamaño  $n$ :  $(X, Y) = (X_1, Y_1), \dots, (X_n, Y_n)$ , se define el rango  $R(X_i)$  comparando  $X_i$  con los demás valores de  $X$ . Es decir,  $R(X_i) = 1$  si  $X_i$  es el menor de los valores de  $X$ ,  $R(X_i) = 2$  si es el segundo menor valor de  $X$ ,  $X_i = n$  si  $X_i$  es el mayor valor de la muestra  $X$ . De manera análoga se definen los rangos  $R(Y_i)$ .

La medida de correlación de rangos  $\rho$  definida por Spearman en 1904 se calcula

$$\rho = \frac{\sum_{i=1}^n R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left(\sum_{i=1}^n R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)\left(\sum_{i=1}^n R(Y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right)}} \quad (\text{Ec. D.2})$$

El coeficiente  $\rho$  de Spearman se obtiene reemplazando los datos por sus rangos y calculando el coeficiente  $r$  de Pearson entre los rangos  $R(X)$  y  $R(Y)$  [Conover 1980].

La función de distribución del coeficiente  $\rho$  no depende de las distribuciones de las muestras, por lo cual  $\rho$  se suele usar como estadístico para pruebas de hipótesis no paramétricas de independencia entre dos variables aleatorias. La distribución nula se obtiene de tablas de distribución de  $\rho$ , o para valores grandes de  $n$ , se puede aproximar el cuantil  $p$ -ésimo de  $\rho$  según

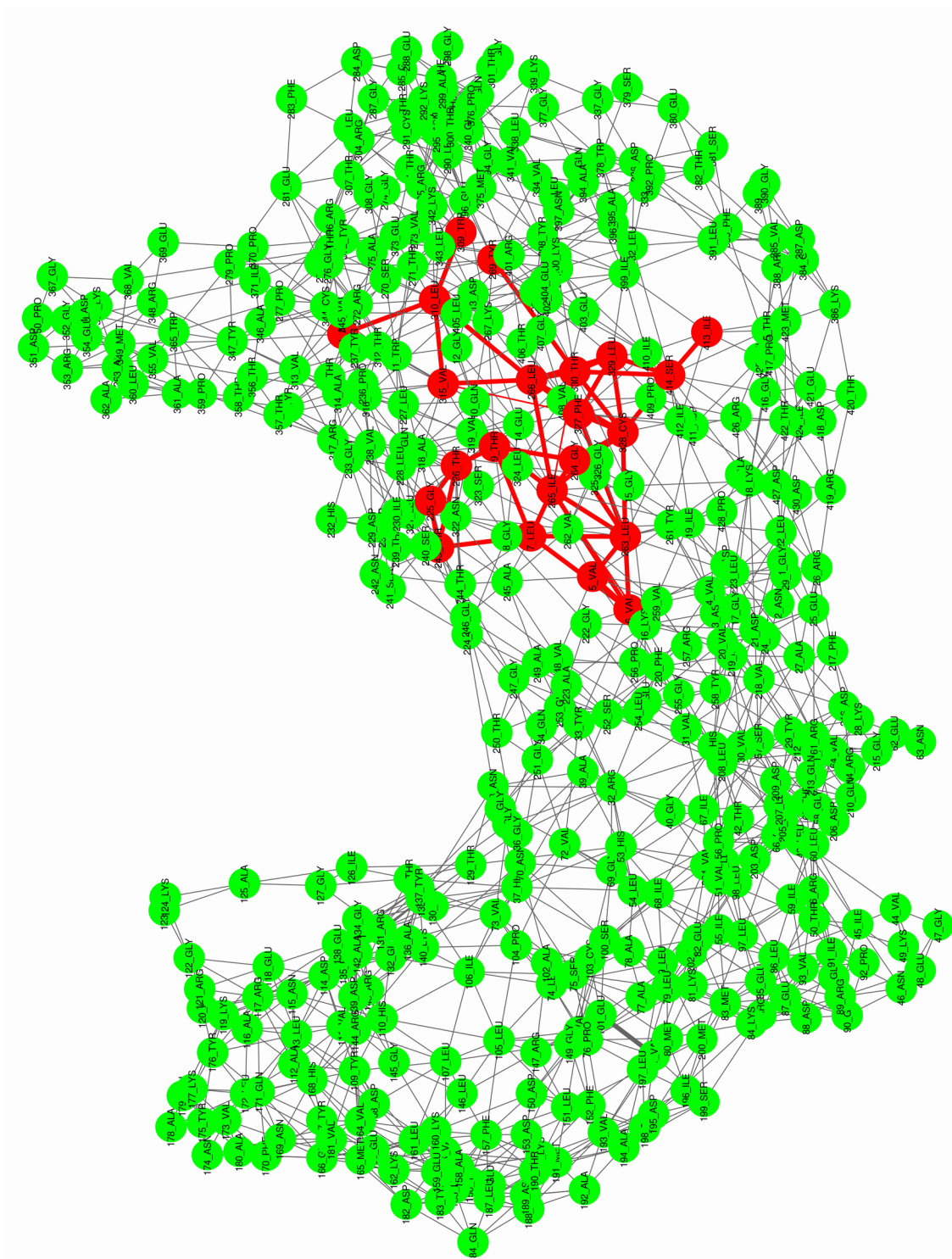
$$\omega_p = \frac{z_p}{\sqrt{n-1}} \quad (\text{Ec. D.3})$$

donde  $z_p$  es el cuantil  $p$ -ésimo de la distribución normal estandarizada. La hipótesis nula de independencia entre  $X$  e  $Y$  se rechaza con nivel  $\alpha$  si

- i.*  $|\rho| > 1 - \alpha/2$  (prueba de dos colas).
- ii.*  $\rho < -\omega_{1-\alpha}$  (prueba de cola inferior para correlación negativa).
- iii.*  $\rho > \omega_{1-\alpha}$  (prueba de cola superior para correlación positiva).

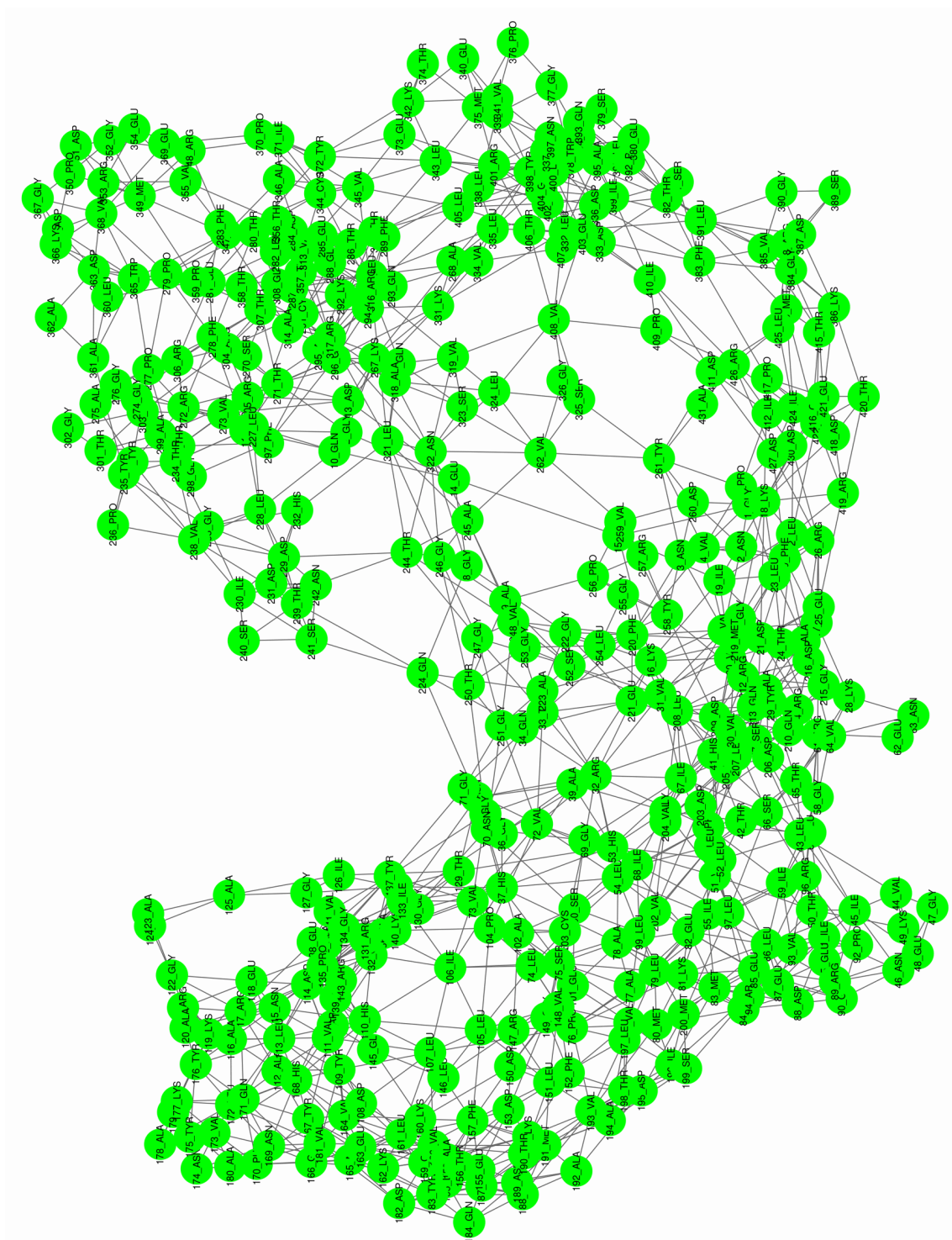
## E. EJEMPLOS - RIN

En esta sección se presentan de forma individual las redes de interacciones presentadas en la Figura 4.12 y Figura 4.13.



**Figura E.1** Red de interacciones entre residuos (RIN) de la proteína Adenilosuccinato Sintetasa (*Adenylosuccinate Synthetase*). Estructura sin ligando, código de pdb 1ADE. Red total de residuos, los residuos en posiciones claves se muestran en rojo (Figura 4.12c, Capítulo 4, sección 4.2.2).

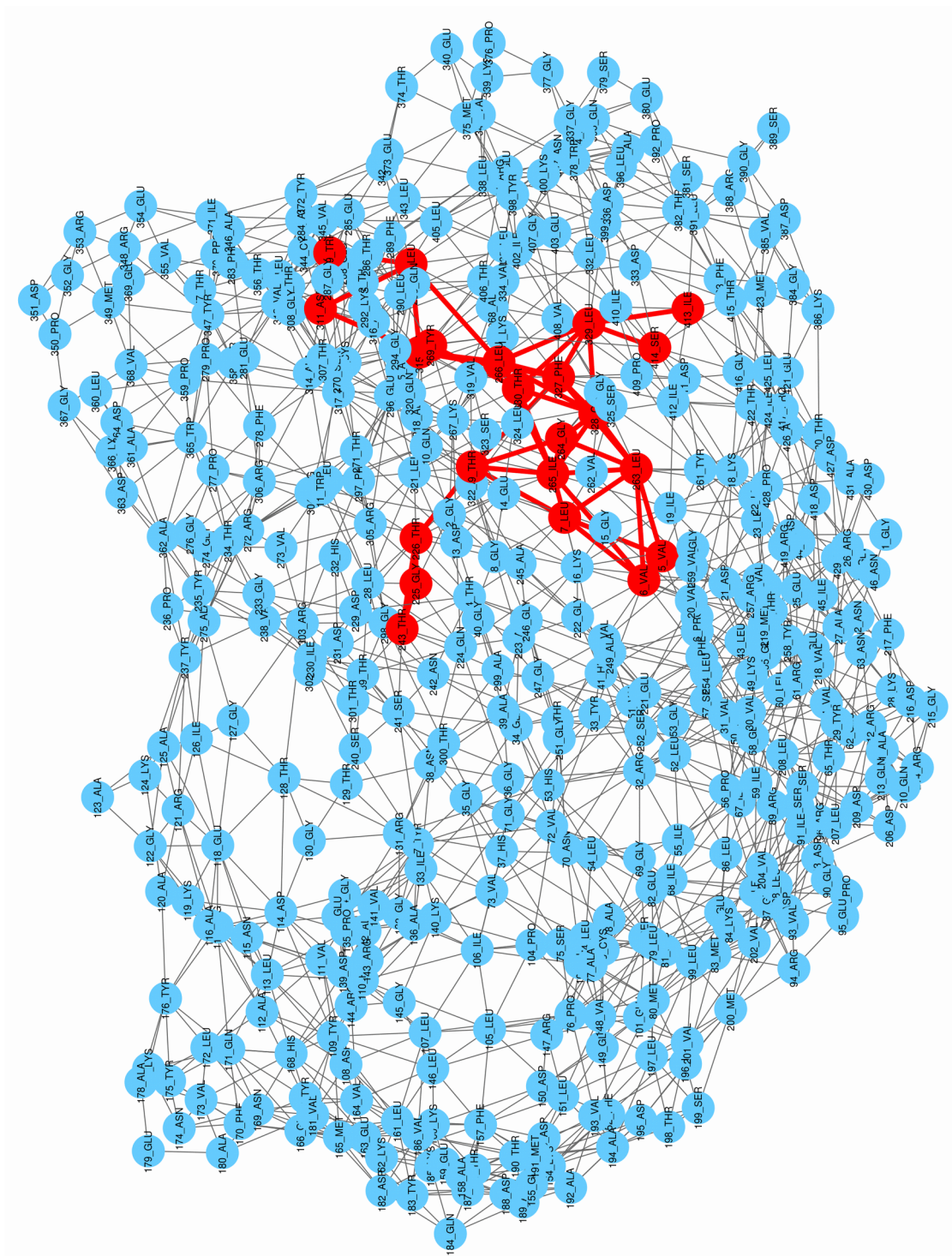




**Figura E.2** RIN de la proteína Adenilosuccinato Sintetasa (*Adenylosuccinate Synthetase*). Estructura sin ligando, código de pdb 1ADE. Red de interacciones entre residuos fuera de las posiciones claves (**Figura 4.12e**, Capítulo 4, sección 4.2.2).

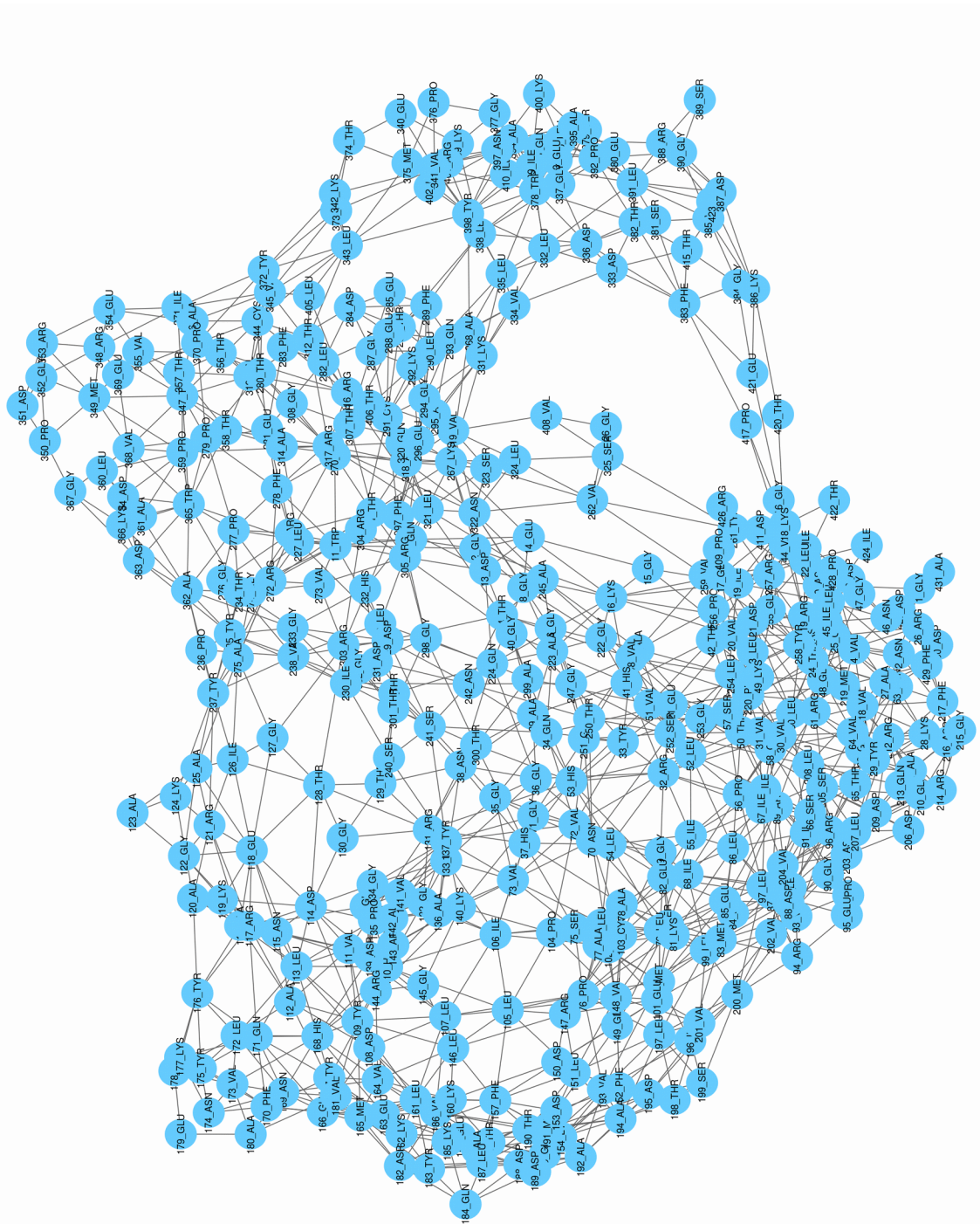






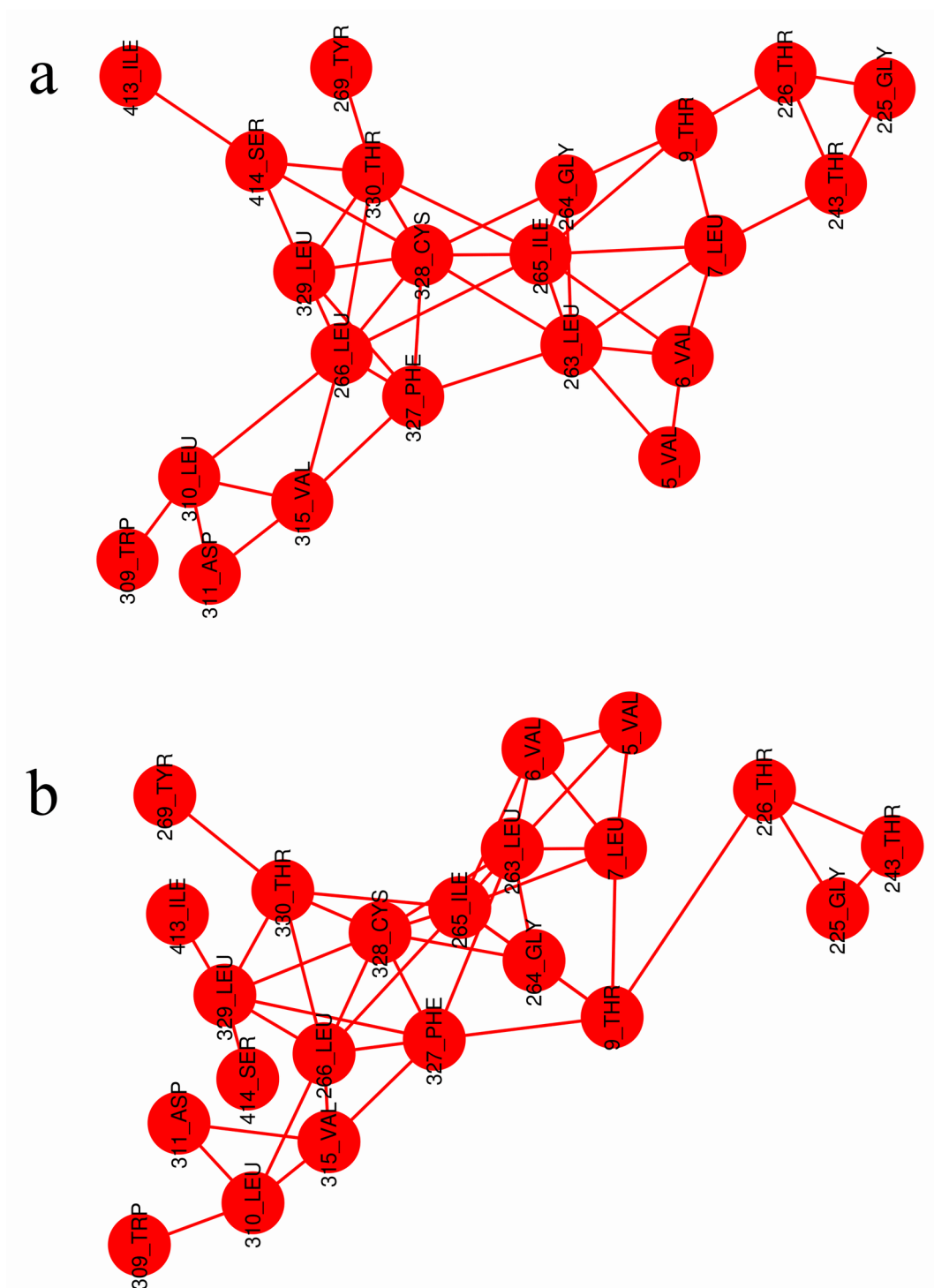
**Figura E.3 RIN de la proteína Adenylosuccinato Sintetasa (*Adenylosuccinate Synthetase*).** Estructura estructura con ligando, código de pdb 1NHT. Red total de residuos, los residuos en posiciones claves se muestran en rojo (**Figura 4.12d**, Capítulo 4, sección 4.2.2).





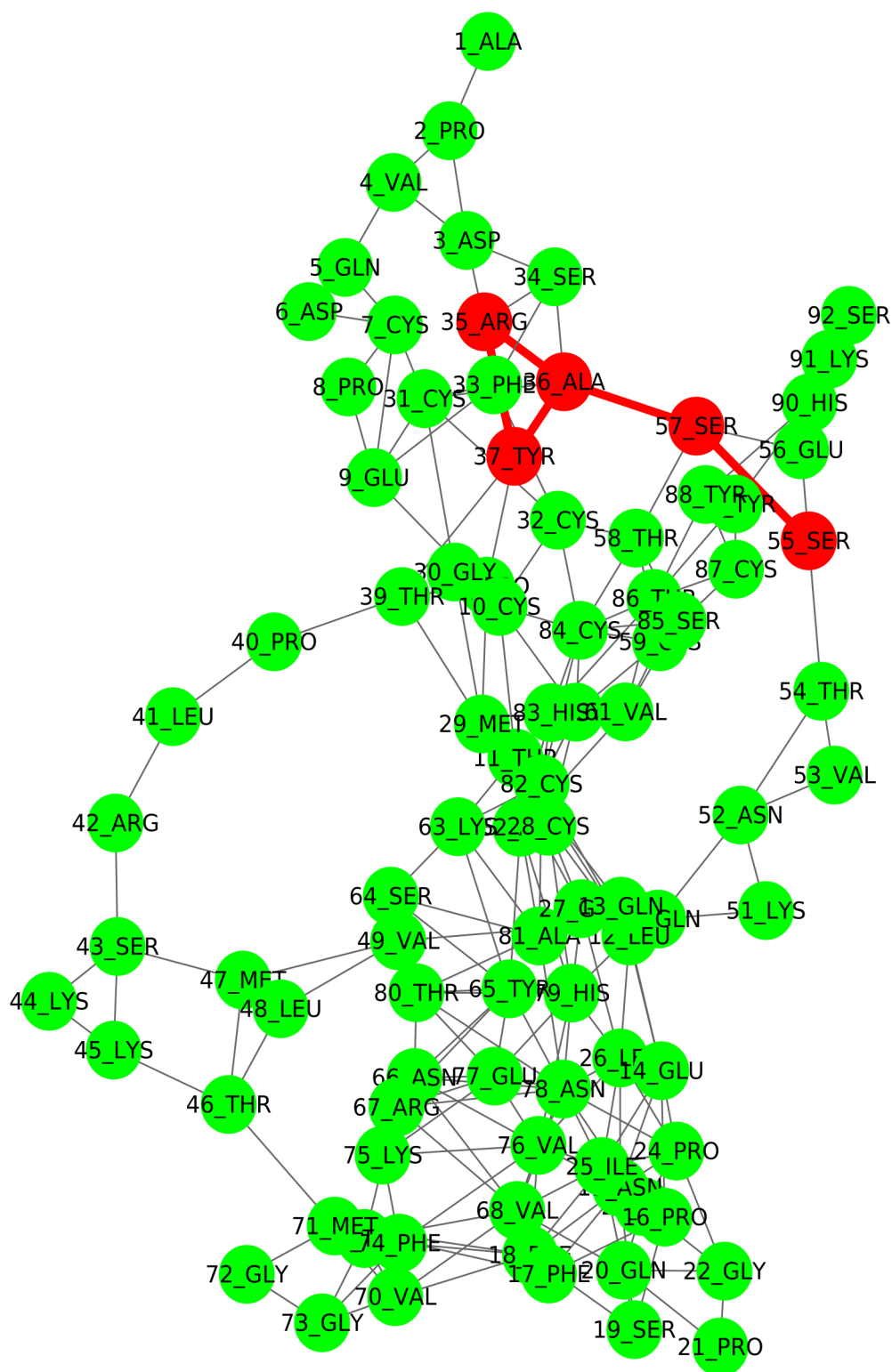
**Figura E.4 RIN** de la proteína Adenilosuccinato Sintetasa (*Adenylosuccinate Synthetase*). Estructura con ligando, código de pdb 1NHT. Red de interacciones entre residuos fuera de las posiciones claves (**Figura 4.12f**, Capítulo 4, sección 4.2.2).





**Figura E.5 RINs de la proteína Adenylosuccinato Sintetasa (*Adenylosuccinate Synthetase*).** (a) Red de interacciones entre los residuos en posiciones claves de la estructura sin ligando, código de pdb 1ADE (Figura 4.12g, de la sección 4.2.2). (b) Red de interacciones entre los residuos en posiciones claves de la estructura con ligando, código de pdb 1NHT (Figura 4.12h, Capítulo 4, sección 4.2.2).

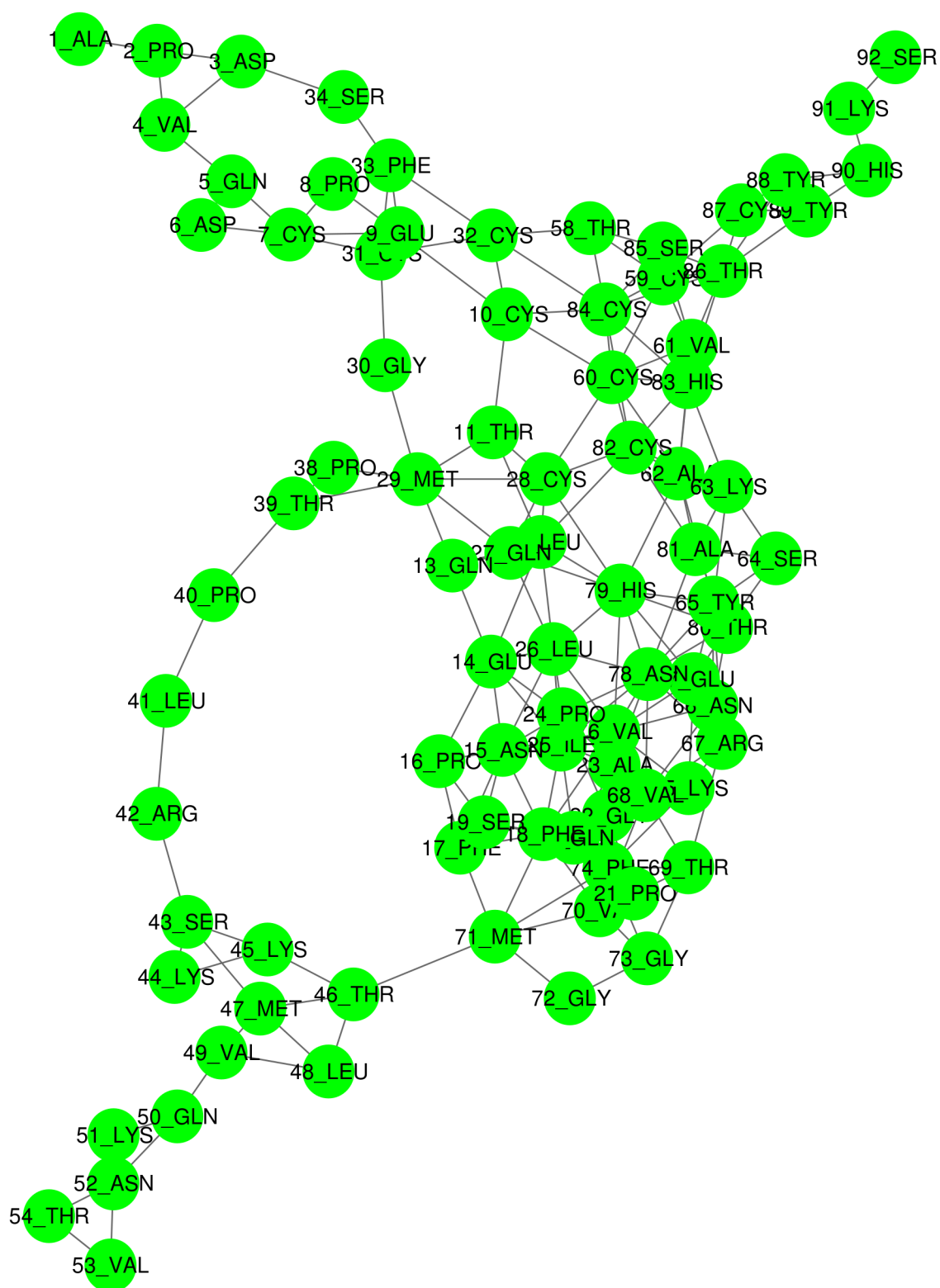




**Figura E.6** Red de interacciones entre residuos (RIN) de la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( $\alpha$ -Subunit of Human Chorionic Gonadotropin). Estructura sin ligando, código de pdb 1DZ7. Red total de residuos, los residuos en posiciones claves se muestran en rojo (**Figura 4.13c**, Capítulo 4, sección 4.2.2).

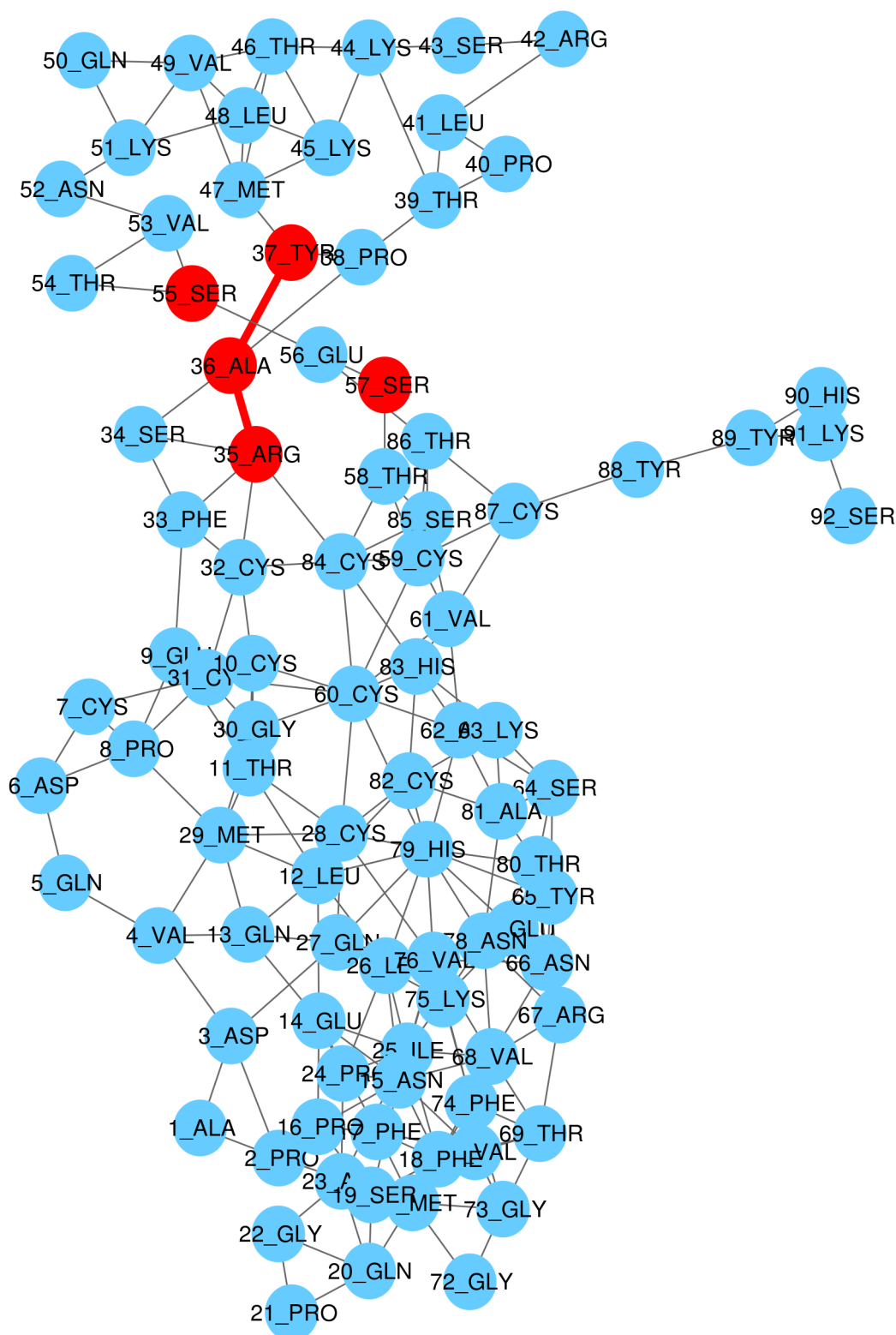






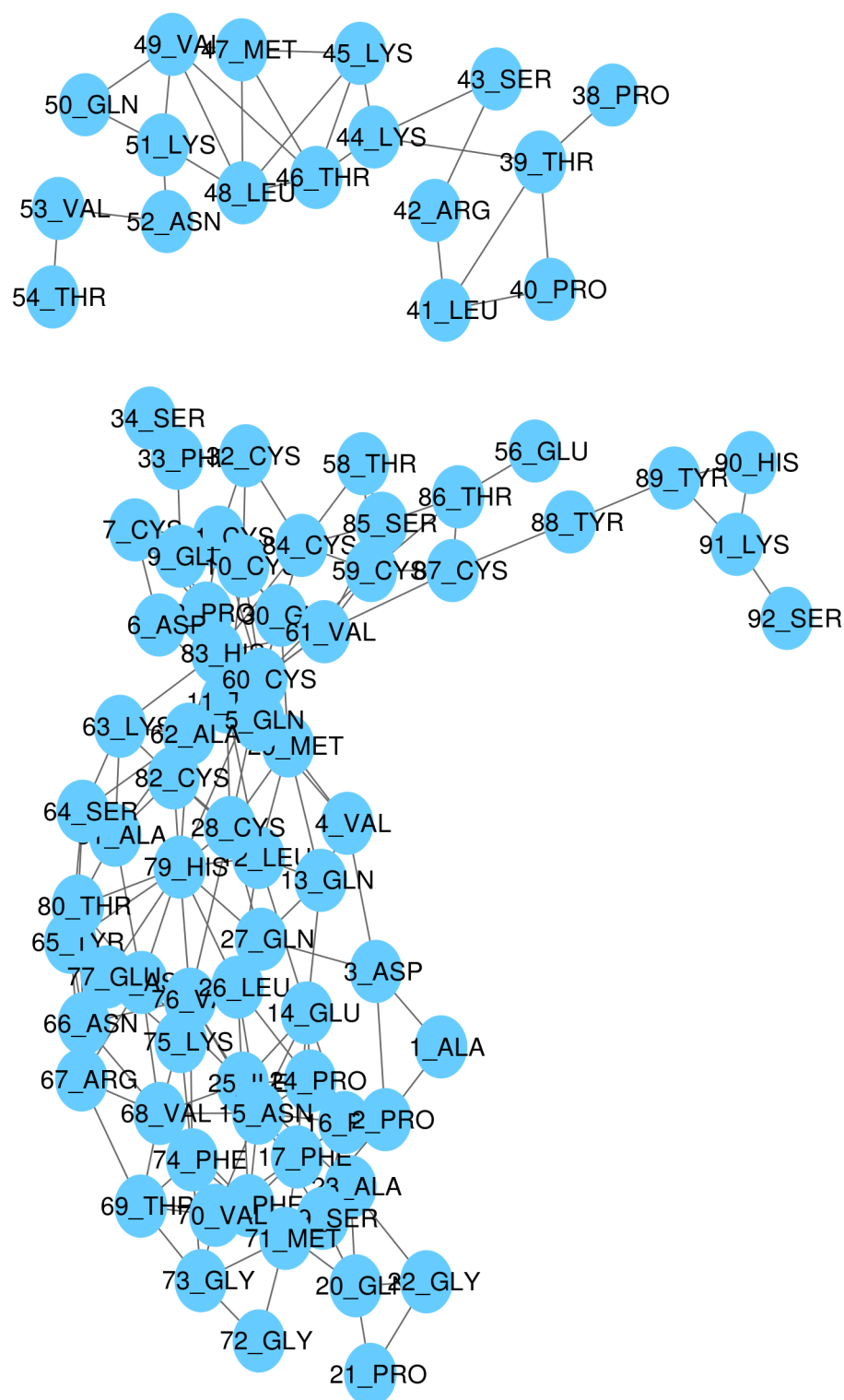
**Figura E.7** RIN de la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( $\alpha$ -Subunit of Human Chorionic Gonadotropin). Estructura sin ligando, código de pdb 1DZ7. Red de interacciones entre residuos fuera de las posiciones claves (Figura 4.13e, Capítulo 4, sección 4.2.2).





**Figura E.8** RIN de la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( *$\alpha$ -Subunit of Human Chorionic Gonadotropin*). Estructura con ligando, código de pdb 1E9J. Red total de residuos, los residuos en posiciones claves se muestran en rojo (**Figura 4.13d**, Capítulo 4, sección 4.2.2).





**Figura E.9** RIN de la subunidad- $\alpha$  de la Gonadotropina Coriónica humana ( $\alpha$ -Subunit of Human Chorionic Gonadotropin). Estructura con ligando, código de pdb 1E9J. Red de interacciones entre residuos fuera de las posiciones claves (**Figura 4.13f**, Capítulo 4, sección 4.2.2).



# *Referencias*



- [Adamski-Werner 2004] Adamski-Werner S. L., Palaninathan S. K., Sacchettini J. C. and Kelly J. W. *Diflunisal Analogues Stabilize the Native State of Transthyretin. Potent Inhibition of Amyloidogenesis.* (2004) *J Med Chem*, 47:355–374.
- [Altschul 1997] Altschul S. F., Madden T. L., Schäffer A. A., Zhang J., Zhang Z., Miller W. and Lipman D. J. *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* (1997) *Nucleic Acids Res.* 25:3389-3402.
- [Amitai 2004] Amitai G., Shemesh A., Sitbon E., Shklar M., Netanel D., Venger I. and Pietrokovski S. *Network analysis of protein structures identifies functional residues.* (2004) *J. Mol. Biol.*, 344(4), 1135- 1146.
- [Armon 2001] Armon A., Graur D., and Ben-Tal N. *ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information.* (2001) *J. Mol. Biol.* 307, 447–463.
- [Assenov 2008] Assenov Y., Ramirez F., Schelhorn S. E., Lengauer T. and Albrecht M. *Computing topological parameters of biological networks.* (2008) *Bioinformatics* 24, 282–284 .
- [Atilgan 2001] Atilgan A.R., Durell S.R., Jernigan R.L., Demirel M.C., Keskin O. and Bahar I. *Anisotropy of Fluctuation Dynamics of Proteins With an Elastic Network Model.* *Biophys. J.* 80, 505-515 (2001).
- [Bahar 1997a] Bahar I., Atilgan A. R. and Erman B. *Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential.* (1997) *Fold. Des.* 2(3):173-81.
- [Bahar 1997b] Bahar I. and Jernigan R. L. *Inter-residue potentials in globular proteins and the dominance of highly specific hydrophilic interactions at close separation.* (1997) *J. Mol. Biol.* 266(1):195-214.
- [Bahar 1998] Bahar I., Atilgan A. R., Demirel M. C. and Erman B. *Vibrational dynamics of folded proteins: significance of slow and fast modes in relation to function and stability.* (1998) *Phys. Rev. Lett.* 80:2733–2736.
- [Bahar 1999a] Bahar I., Erman B., Jernigan R. L., Atilgan A. R. and Covell D. G. *Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function.* (1999) *J. MoBiol.* 285(3):1023-37.
- [Bahar 1999b] Bahar I. and Jernigan R. L. *Cooperative fluctuations and subunit communication in tryptophan synthase.* (1999) *Biochemistry* 38(12):3478-90.
- [Bahar 1999c] Bahar I. *Dynamics of proteins and biomolecular complexes: inferring functional motions from structure.* (1999) *Rev. Chem. Eng.* 15, 319 .
- [Bahar 2005] Bahar I. and Rader A. J. *Coarse-grained normal mode analysis in structural biology.* (2005) *Curr. Opin. Struct. Biol.* 15(5):586-92.
- [Barber 1996] Barber C. B., Dobkin D. P. and Huhdanpaa H. *The Quickhull algorithm for convex hulls.* (1996) *AcmTransactions on Mathematical Software*;22: 469–483.doi:10.1145/235815.235821

- [Barletta 2017] Barletta G. P. and Fernandez-Alberti S. *Proteins fluctuations and cavity changes relationship*. (2017) *J. Chem. Theory Comput* (in Press).
- [Bartlett 2002] Bartlett G. J., Porter C. T., Borkakoti N. and Thornton J. M. *Analysis of catalytic residues in enzyme active sites*. (2002) *J. Mol. Biol.* 324, 105–121.
- [Bateman 2004] Bateman A., Coin L., Durbin R., Finn R. D., Hollich V., Griffiths-Jones S., Khanna A., Marshall M., Moxon S., Sonnhammer, E. L. et al. *The Pfam protein families database*. (2004) *Nucleic Acids Res*, 32, D138-141.
- [Batista 2010] Batista P. R., Robert C. H., Maréchal J. D., Hamida-Rebai M. B., Pascutti P. G., Bisch P. M. and Perahia D. *Consensus modes, a robust description of protein collective motions from multiple-minima normal mode analysis—application to the HIV-1 protease*. (2010) *Phys. Chem. Chem. Phys.* 12, 2850–2859 (doi:10.1039/b919148h).
- [Bell 1970] Bell RJ, Dean P, Hibbins-Butter DC. *Localization of normal modes*. 1970; 3:2111–8.
- [Benkovic 2003] Benkovic S. J and Hammes-Schiffer S. *A perspective on enzyme catalysis*. (2003) *Science*, 301: 1196–1202.
- [Binkowski 2004] Binkowski T. A., Freeman P. and Liang J. *pvSOAR: detecting similar surface patterns of pocket and void surfaces of amino acid residues on proteins*. (2004) *Nucleic Acids Res*, 32, W555-558.
- [Blake 1978] Blake C. C., Geisow M. J., Oatley S. J., Rérat B. and Rérat C. *Structure of prealbumin: secondary, tertiary and quaternary interactions determined by Fourier refinement at 1.8 Å*. (1978) *Journal of Molecular Biology*; 121: 339-356. PMID: 671542
- [Blamey 2005] Blamey C. J., Ceccarelli C., Naik U. P. and Bahnson B. J. *The crystal structure of calcium- and integrin-binding protein 1: Insights into redox regulated functions*. (2005) *Protein Sci.*; 14: 1214–1221 PMID: 15840829
- [Böde 2007] Böde C., Kovacs I. A., Szalay M. S., Palotai R., Korcsmaros T. and Csermely P. *Network analysis of protein dynamics*. (2007) *FEBS Lett.* 581, 2776–2782.
- [Brooks 1983] Brooks B. and Karplus M. *Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor*. (1983) *Proc. Natl. Acad. Sci. USA* 80, 6571.
- [Brooks 1985] Brooks B. R. and Karplus M. *Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme*. (1985) *Proc. Natl. Acad. Sci. USA* 82, 4995-4999.
- [Brooks 1995] Brooks B. R., Janezic D. and Karplus M. *Harmonic analysis of large systems*. (1995) *I. Methodology* 12:1522-1542.
- [Browner 1992] Browner M. F., Fauman E. B. and Fletterick R. J. *Tracking conformational states in allosteric transitions of phosphorylase*. (1992) *Biochemistry*; 31(46):11297–304. PMID: 1445867

- [Brylinski 2008] Brylinski M. and Skolnick J. *What is the relationship between the global structures of apo and holo proteins?* (2008) *Proteins Struct Funct Bioinforma.*; 70:363–77.
- [Bulawa 2012] Bulawa C. E., Connelly S., DeVit M., Wang L., Weigel C., Fleming J. A., Packman J., Powers E. T., Wiseman R. L., Foss T. R., Wilson I. A., Kelly J. W. and Labaudiniere R. *Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade.* (2012) *Proc. Natl. Acad. Sci.*, 109:9629–9634.
- [Cardoso 2002] Cardoso I., Goldsbury C. S., Muller S. A., Olivieri V., Wirtz S., Damas A. M., Aebi U. and Saraiva M. J. *Transthyretin fibrillogenesis entails the assembly of monomers: a molecular model for in vitro assembled transthyretin amyloid-like fibrils.* (2002) *J Mol Biol.*; 317: 683–695. <https://doi.org/10.1006/jmbi.2002.5441> PMID: 11955017.
- [Carpaneto 1988] Carpaneto G., Martello S. and Toth P. *Algorithms and codes for the assignment problem.* (1988) *Ann Oper. Res.* 13:193–223
- [Case 1994] Case D. A. *Normal mode analysis of protein dynamics.* (1994) *Curr. Opin. Struct. Biol.* 4:285-290.
- [Cendron 2009] Cendron L., Trovato A., Seno F., Folli C., Alfieri B., Zanotti G. and Berni R. *Amyloidogenic Potential of Transthyretin Variants INSIGHTS FROM STRUCTURAL AND COMPUTATIONAL ANALYSES.* (2009) *J Biol Chem.*; 284: 25832±25841. PMID: 19602727.
- [Cheng 1977] Cheng S. Y., Pages R. A., Saroff H. A., Edelhoch H. and Robbins J. *Analysis of thyroid hormone binding to human serum prealbumin by 8-anilinonaphthalene-1-sulfonate fluorescence.* (1977) *Biochemistry*; 16:3707±3713. PMID: 889816
- [Choi 2010] Choi S., Reixach N., Connelly S., Johnson S. M., Wilson I. A. and Kelly J. W. *A substructure combination strategy to create potent and selective transthyretin kinetic stabilizers that prevent amyloidogenesis and cytotoxicity.* (2010) *J Am Chem Soc.*; 132: 1359±1370. <https://doi.org/10.1021/ja908562q> PMID: 20043671
- [Cianci 2015] Cianci M., Folli C., Zonta F., Florio P., Berni R. and Zanotti G. *Structural evidence for asymmetric ligand binding to transthyretin.* (2015) *Acta Crystallogr. Sect. D. Biol. Crystallogr.*, 71:1582–1592.
- [Connelly 2010] Connelly S., Choi S., Johnson S. M., Kelly J. W. and Wilson I. A. *Structure-based design of kinetic stabilizers that ameliorate the transthyretin amyloidoses.* (2010) *Curr Opin Struct Biol.*; 20: 54±62. <https://doi.org/10.1016/j.sbi.2009.12.009> PMID: 20133122
- [Conover 1980] Conover W. J. *Practical Nonparametric Statistics.* (1989) Wiley and Sons, New York.
- [Cusack 2007] Cusack M. P., Thibert B., Bredesen D. E. and del Rio G. *Efficient Identification of Critical Residues Based Only on Protein Structure by Network Analysis.* (2007) *PLoS ONE* 2(5): e421. doi:10.1371/journal.pone.0000421.
- [Damm 2006] Damm K. L. and Carlson H. A. *Gaussian-weighted RMSD superposition of proteins: a structural comparison for flexible proteins and predicted protein structures.* (2006) *Biophys J.*; 90(12):4558–73. PMID:16565070

- [Del Sol 2006] Del Sol A., Fujihashi H., Amoros D., and Nussinov R. *Residue centrality, functionally important residues, and active site shape: Analysis of enzyme and non-enzyme families.* (2006) Protein Science; 15(9), 2120–2128. <http://doi.org/10.1110/ps.062249106>
- [Delarue 2002] Delarue M. and Sanejouand Y. H. *Simplified normal mode analysis of conformational transitions in DNAdependent polymerases: The Elastic Network Model.* (2002) J Mol Biol.; 320(5):1011–24. PMID: 12126621
- [Demirel 1998] Demirel M.C, et al. *Identification of kinetically hot residues in proteins.* (1998) Protein Sci 7, 2522.
- [Dokholyan 2002] Dokholyan N. V., Li L., Ding F. and Shakhnovich E. I. *Topological determinants of protein folding.* (2002) Proc. Natl. Acad. Sci. USA, 99(13), 8637-8641.
- [Doncheva 2012] Doncheva N. T., Assenov Y., Domingues F. S. and Albrecht, M. *Topological analysis and interactive visualization of biological networks and protein structures.* (2012) Nat. Protocols; 4 (7), 670-685.
- [Elcock 2001] Elcock A. H. *Prediction of functionally important residues based solely on the computed energetics of protein structure.* (2001) J. Mol. Biol. 312, 885–896.
- [Erbel 2000] Erbel P. J. A., Karimi-Nejad Y., van Kuik J. A., Boelens R., Kamerling J. P., and Vliegthart J. F. G. *Effects of the N-Linked Glycans on the 3D Structure of the Free  $\alpha$ -Subunit of Human Chorionic Gonadotropin.* (2000) Biochemistry 39 (20), 6012-6021. DOI: 10.1021/bi992786n
- [Fajardo 2013] Fajardo J. E. and Fiser A. *Protein structure based prediction of catalytic residues.* (2013) BMC Bioinformatics, 14:63 doi:10.1186/1471-2105-14-63
- [Ferre 2004] Ferre F., Ausiello G., Zanzoni A. and Helmer-Citterich M. *SURFACE: a database of protein surface regions for functional annotation.* (2004) Nucleic Acids Res, 32, D240-244.
- [Ferguson 1975] Ferguson R.N., Edelhoop H., Saroff H. A., Robbins J. and Cahnmann H. J. *Negative cooperativity in the binding of thyroxine to human serum prealbumin. Preparation of tritium-labeled 8-anilino-1-naphthalenesulfonic acid.* (1975) Biochemistry;14: 282–289.
- [Florio 2015] Florio P., Folli C., Cianci M., Del Rio D., Zanotti G., Berni R. *Transthyretin binding heterogeneity and antiamyloidogenic activity of natural polyphenols and their metabolites.* (2015) J Biol Chem., 290:29769–29780.
- [Frappier 2014] Frappier V. and Najmanovich R. J. *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.* (2014) PLOS Computational Biology; 10, e1003569. Doi: 10.1371/journal.pcbi.1003569 PMID: 24762569
- [Frappier 2015] Frappier V., Chartier M. and Najmanovich R. J. *ENCoM server: exploring protein conformational space and the effect of mutations on protein function and stability.* (2015) Nucleic Acids Research.; doi: 10.1093/nar/gkv343

- [Frauenfelder 1991] Frauenfelder H., Sligar S. G. and Wolynes P. G. *The energy landscapes and motions of proteins*. (1991) *Science* 254:1598–1603.
- [Furnham 2014] Furnham N., Holliday G. L., de Beer T. A. P., Jacobsen J. O. B., Pearson W. R. and Thornton J. M. *The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes*. (2014) *Nucleic Acids Research*.42 (Database issue):D485-D489. doi:10.1093/nar/gkt1243.
- [Glaser 2003] Glaser G. *ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information*. (2003) *Bioinformatics*; 19:163–164. PMID: 12499312
- [Go 1983] Go N. and Nishikawa T. *Dynamics of a small globular protein in terms of low-frequency normal modes*. (1983) *Proc. Natl. Acad. Sci. USA* 80, 3696.
- [Goldstein 1980] Goldstein H. *Classical Mechanics*. (1980) Addison-Wesley Pub. Co., Reading, MA.
- [Gong 2009] Gong W., Li Y., Cui G., Hu J., Fang H., Jin C. and Xia B. *Solution structure and catalytic mechanism of human protein histidine phosphatase I*. (2009) *Biochem. J.*; 418: 337–344. doi: 10.1042/BJ20081571 PMID: 18991813
- [Guerois 2002] Guerois R., Nielsen J. E. and Serrano L. *Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations*. (2002) *J Mol Biol.*; 320: 369±387. [https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4) PMID: 12079393
- [Gutteridge 2005] Gutteridge A and Thornton J. *Conformational changes observed in enzyme crystal structures upon substrate binding*. (2005) *J Mol Biol.*; 346(1):21–8. PMID: 15663924
- [Haliloglu 1997] Haliloglu T., Bahar I. and Erman B. *Gaussian dynamics of folded proteins*. (1997) *Phys. Rev. Lett.* 79:3090–3093.
- [Hammarstrom 2003] Hammarstrom P., Wiseman R. L., Powers E. T. and Kelly J. W. *Prevention of Transthyretin Amyloid Disease by Changing Protein Misfolding Energetics*. (2003) *Science* (80- ), 299:713–716.
- [Hert 2017] Hert S. and Seel M. *dD Convex Hulls and Delaunay Triangulations*. In *CGAL User and Reference Manual*. CGAL Editorial Board, 4.10 edition, 2017.
- [Henikoff 1994] Henikoff S. and Henikoff J. G. *Position-based sequence weights*. (1994) *J Mol Biol.*; 243(4):574–8. PMID:7966282
- [Hinsen 1998] Hinsen K. A. *Analysis of domain motions by approximate normal mode calculations*. (1998) *Proteins* 33:417–429.
- [Hinsen 1999a] Hinsen K. and Kneller G. R. *A simplified force field for describing vibrational protein dynamics over the whole frequency range*. (1999) *J. Chem. Phys.* 24:10766–10769.
- [Hinsen 1999b] Hinsen K., Thomas A. and Field M. J. *Analysis of domain motions in large proteins*. (1999) *Proteins* 34:369–382.
- [Hinsen 2000] Hinsen K. A. and Kneller G. R. *Projection methods for the analysis of complex motions in macromolecules*. (2000) *Mol. Sm.* 23:275– 292.



- [Hörnberg 2000] Hörnberg A., Eneqvist T., Olofsson A., Lundgren E. and Sauer-Eriksson A. E. *A comparative analysis of 23 structures of the amyloidogenic protein transthyretin*. (2000) *J Mol Biol*, 302:649–669.
- [Hubbard 1993] Hubbard S.J. and Thornton J.M. *NACCESS*. (1993) Computer Program, Department of Biochemistry and Molecular Biology, University College London.
- [Hulo 2004] Hulo N., Sigrist C. J., Le Saux V., Langendijk-Genevaux P. S., Bordoli L., Gattiker A., De Castro E., Bucher P. and Bairoch A. *Recent improvements to the PROSITE database*. (2004) *Nucleic Acids Res*, 32, D134-137.
- [Hurshman Babbes 2008] Hurshman Babbes A. R., Powers E. T. and Kelly J. W. *Quantification of the Thermodynamically Linked Quaternary and Tertiary Structural Stabilities of Transthyretin and Its Disease-Associated Variants: The Relationship between Stability and Amyloidosis*. (2008) *Biochemistry*; 6969±6984. <https://doi.org/10.1021/bi800636q> PMID: 18537267
- [Hwang 2000] Hwang P. M. and Vogel H. J. *Structures of the platelet calcium- and integrin-binding protein and the alphaIIb integrin cytoplasmic domain suggest a mechanism for calcium-regulated recognition; homology modelling and NMR studies*. (2000) *J.Mol.Recog.*; 13: 83–92.
- [Iakovleva 2015] Iakovleva I., Brännström K., Nilsson L., Gharibyan A. L., Begum A., Anan I., et al. *Enthalpic Forces Correlate with the Selectivity of Transthyretin-Stabilizing Ligands in Human Plasma*. (2015) *J Med Chem.*; 58: 6507±6515. PMID: 26214366. <https://doi.org/10.1021/acs.jmedchem.5b00544>.
- [Iakovleva 2016] Iakovleva I., Begum A., Brännström K., Wijsekera A., Nilsson L., Zhang J., Andersson P. L., Sauer-Eriksson A. E. and Olofsson A. *Tetrabromobisphenol a is an efficient stabilizer of the transthyretin tetramer*. *PLoS One* 2016, 11:1–16.
- [Isin 2002] 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.
- [Janezic 1995a] Janezic D. and Brooks B. R. *Harmonic analysis of large systems. II. Comparison of different protein models*. (1995) *J. Comput. Chem.* 16: 1543-1553.
- [Janezic 1995b] Janezic D., Venable R. M. and Brooks B. R. *Harmonic analysis of large systems. III. Comparison with molecular dynamics*. (1995) *J. Comput. Chem.* 16: 1554-1566.
- [Jernigan 1999] Jernigan R. L., Demirel M. C. and Bahar I. *Relating structure to function through the dominant slow modes of motion of DNA topoisomerase II*. (1999) *International Journal of Quantum Chemistry* 75:301-312.
- [Johnson 2005] Johnson S. M., Wiseman R. L., Sekijima Y., Green N. S., Adamski-werner S. L. and Kelly J.W. *Native State Kinetic Stabilization as a Strategy To Ameliorate Protein Misfolding Diseases: A Focus on the Transthyretin Amyloidoses*. (2005) *Acc. Chem. Res.*, 38, 911-921.
- [Jones 2004] Jones S. and Thornton J. M. *Searching for functional sites in protein structures*. (2004) *Curr. Opin. Chem. Biol.* 8, 3–7.

- [Juritz 2012] Juritz E., Palopoli N., Fornasari M., Fernandez-Alberti S. and Parisi G. *Protein conformational diversity modulates sequence divergence*. (2012) *Mol. Biol. Evol.*, doi: 10.1093/molbev/mss080.
- [Kabsch 1983] Kabsch W. and Sander C. *Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features*. (1983) *Biopolymers*, 22 2577-2637.
- [Kalstein 2010] Kalstein A., Fernández-Alberti S., Bastida A., Soler M. A., Farag M.H., Zúñiga J., et al. *Vibrational dynamics of polyatomic molecules in solution: assignment, time evolution and mixing of instantaneous normal modes*. (2010) *Theor Chem Acc.*; 128(4–6):769–82.
- [Keskin 2000] Keskin O., Jernigan R.L. and Bahar, I. *Proteins with similar architecture exhibit similar large-scale dynamic behavior*. (2000) *Biophys J*, 78, 2093-2106.
- [Keskin 2002a] Keskin O., Durell S. R., Bahar I., Jernigan R. L. and Covell D. G. *Relating molecular flexibility to function: a case study of tubulin*. (2002) *Biophys. J.* 83(2):663-80.
- [Keskin 2002b] Keskin O., Bahar I., Flatow D., Covell D. G. and Jernigan R. L. *Molecular mechanisms of chaperonin GroEL-GroES function*. (2002) *Biochemistry* 41(2):491-501 .
- [Khersonsky 2010] Khersonsky O. and D. S. Tawfik. *Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective*. (2010) *Annu. Rev. Biochem.*; 79:11.1–11.35
- [Kim 1990] Kim Y. and Prestegard J. H. *Refinement of the NMR structures for acyl carrier protein with scalar coupling data*. (1990) *Proteins*; 8: 377–385 PMID: 2091027
- [Kolstoe 2010] Kolstoe S. E. and Wood S. P. *Drug targets for amyloidosis*. (2010) *Biochem Soc Trans.*; 38: 466±470. <https://doi.org/10.1042/BST0380466> PMID: 20298204
- [Krzanowski 1979] Krzanowski W. J. *Between-Groups Comparison of Principal Components*. (1979) *J Am Stat Assoc.*; 74 (367):703–7.
- [Kumar 2000] Kumar S., Ma B., Tsai C. J., Sinha N. and Nussinov R. *Folding and binding cascades: dynamic landscapes and population shifts*. (2000) *Protein Sci*, 9, 10-19.
- [Kundu 2002] Kundu S., Melton J. S., Sorensen D. C. and Phillips G. N. *Dynamics of Proteins in Crystals: Comparison of Experiment With Simple Models*. (2002) *Biophys. J.* 83:723-732.
- [Kurush 1950] Karush, F. *Heterogeneity of the binding sites of bovine serum albumin*. (1950) *J Am Chem Soc*, 72, 2705-2713.
- [Lassmann 2005] Lassmann T. and Sonnhammer E. L. *Automatic assessment of alignment quality*. (2005) *Nucleic Acids Research.*;33(22):7120-7128. doi:10.1093/nar/gki1020.
- [Lee 1971] Lee B. and Richards F. M. *The Interpretation of Protein Structures: Estimation of Static Accessibility*. (1971) *J Mol Biol.* 55, 379-490.
- [Letunic 2004] Letunic I., Copley R. R., Schmidt S., Ciccarelli F. D., Doerks T., Schultz J., Ponting C. P. and Bork P. *SMART 4.0: towards genomic data integration*. (2004) *Nucleic Acids Res*, 32, D142-144.

- [Li 2002] Li G. H. and Cui Q. A. *A coarse-grained normal mode approach for macromolecules: an efficient implementation and application to Ca<sup>2+</sup>-ATPase*. (2002) *Biophys. J.* 83, 2457.
- [Lichtarge 1996] Lichtarge O., Bourne H. R. and Cohen F. E. *An evolutionary trace method defines binding surfaces common to protein families*. (1996) *J Mol Biol.*; 257:342–358. PMID: 8609628
- [Liu 2012] Liu Y and Bahar I. *Sequence Evolution Correlates with Structural Dynamics*. (2012) *Mol Biol Evol.*; 29 (9):2253–2263 doi: 10.1093/molbev/mss097 PMID: 22427707.
- [Lubchenko 2005] Lubchenko V., Wolynes P. G. and Frauenfelder H. *Mosaic energy landscapes of liquids and the control of protein conformational dynamics by glass-forming solvents*. (2005) *J Phys Chem B* 109:7488–7499.
- [Maguid 2006] Maguid S., Fernandez-Alberti S., Parisi G. and Echave J. *Evolutionary conservation of protein backbone flexibility*. (2006) *J Mol Evol*, 63, 448-457.
- [Maguid 2008] Maguid S., Fernández-Alberti S. and Echave J. *Evolutionary conservation of protein vibrational dynamics*. (2008) *Gene*, Volume 422, Issues 1–2, Pages 7-13, ISSN 0378-1119, <https://doi.org/10.1016/j.gene.2008.06.002>.
- [Mahajan 2017] Mahajan S. and Sanejouand Y.-H. *Jumping between protein conformers using normal modes*. (2017) *J. Comput. Chem.*, 38, 1622–1630. DOI: 10.1002/jcc.24803
- [Martin 2011] Martin A. J. M., Vidotto M., Boscaroli F., Di Domenico T., Walsh I. and Tosatto S. C. E. *RING: networking interacting residues, evolutionary information and energetics in protein structures*. (2011) *Bioinformatics*; 27 (14):2003–5. doi: 10.1093/bioinformatics/btr191 PMID: 21493660
- [McCammon 1976] McCammon J. A., Gelin B. R., Karplus M. and Wolynes P. G. *The hinge-bending mode in lysozyme*. (1976) *Nature* 262, 235-326.
- [McCammon 1987] McCammon J. A. and Harvey S. C. *Dynamics of Proteins and Nucleic Acids*. (1987) Cambridge University Press, Cambridge.
- [Micheletti 2002a] Micheletti C., Lattanzi G. and Maritan A. *Elastic properties of proteins: insight on the folding process and evolutionary selection of native structures*. (2002) *J. Mol. Biol.* 321(5):909-21.
- [Micheletti 2002b] Micheletti C. et al. *Crucial stages of protein folding through a solvable model: predicting target sites for enzyme-inhibiting drugs*. (2002) *Protein Sci.* 11, 1878.
- [Micheletti 2004] Micheletti C., Carloni P. and Maritan A. *Accurate and efficient description of protein vibrational dynamics: comparing Molecular Dynamics and Gaussian models*. (2004) *Proteins* 55:635-645.
- [Miller 2004] Miller S. R., Sekijima Y. and Kelly J. W. *Native state stabilization by NSAIDs inhibits transthyretin amyloidogenesis from the most common familial disease variants*. (2004) *Lab Invest.*, 84:545–552.



- [Ming 2005] Ming D. M. and Wall M. E. *Quantifying allosteric effects in proteins*. (2005) *Proteins*; 59(2): 687–707.
- [Ming 2006] Ming D. M. and Wall M. E. *Interactions in native binding sites cause a large change in protein dynamics*. (2006) *J. Mol. Biol.*, 358(2): 213–223
- [Miroy 1996] Miroy G. J., Lai Z., Lashuel H. A. and Peterson S. A., Strang C. and Kelly J. W. *Inhibiting transthyretin amyloid fibril formation via protein stabilization*. (1996) *Proc Natl Acad Sci U S A*, 93:15051–15056.
- [Monzon 2013] Monzon A. M., Juritz E., Fornasari M. S. and Parisi G. *CoDNaS: A database of conformational diversity in the native state of proteins*. (2013) *Bioinformatics*; 29(19):2512–4. doi: 10.1093/bioinformatics/btt405. PMID: 23846747
- [Monzon 2016] Monzon A. M., Rohr C. O., Fornasari M. S. and Parisi G. *CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state*. (2016) *Database: The Journal of Biological Databases and Curation*; 2016:baw038. doi:10.1093/database/baw038.
- [Monzon 2017] Monzon A.M., Zea D. J., Fornasari M.S., Saldaño T. E., Fernandez-Alberti S., Tosatto S. C. E., and Parisi G. *Conformational diversity analysis reveals three functional mechanisms in proteins*. (2017) *PLoS Comput Biol* 13(2): e1005398.
- [Morcos 2013] Morcos F., Jana B., Hwa T. and Onuchic J. N. *Coevolutionary signals across protein lineages help capture multiple protein conformations*. (2013) *Proc Natl Acad Sci USA*; 110(51):20533–8. doi: 10.1073/pnas.1315625110 PMID: 24297889.
- [Oltvai 2002] Oltvai Z. N. and Barabasi A. L. *Systems biology. Life's complexity pyramid*. (2002) *Science*, 298(5594):763-4.
- [Ondrechen 2001] Ondrechen M. J., Clifton J. G. and Ringe D. *THEMATICS: a simple computational predictor of enzyme function from structure*. (2001) *Proc. Natl Acad. Sci. USA*, 98, 12473–12478.
- [Palaninathan 2012] Palaninathan S. K. *Nearly 200 X-Ray Crystal Structures of Transthyretin: What Do They Tell Us About This Protein and the Design of Drugs for TTR Amyloidoses?* (2012) *Curr Med Chem.*; 19: 2324±2342. PMID: 22471981
- [Parisi 2015] Parisi G., Zea D. J., Monzon A. M. and Marino-Buslje C. *Conformational diversity and the emergence of sequence signatures during evolution*. (2015) *Curr Opin Struct Biol.*; 32:58–65 doi: 10.1016/j.sbi.2015.02.005 PMID: 25749052.
- [Pazos 2004] Pazos F. and Sternberg M. J. E. *Automated prediction of protein function and detection of functional sites from structure*. (2004) *Proc Natl Acad Sci U S A*; 101(41):14754–9. PMID: 15456910.
- [Pei 2001] Pei J. and Grishin N. V. *AL2CO: calculation of positional conservation in a protein sequence alignment*. (2001) *Bioinformatics*; 17(8):700–12. PMID: 11524371.

- [Petrone 2006] Petrone P. and Pande V. S. *Can conformational change be described by only a few normal modes?* (2006) *Biophys J.*; 90(5):1583–93. PMID: 16361336.
- [Piovesan 2016] Piovesan D., Minervini G., Tosatto S. C. E. *The RING 2.0 web server for high quality residue interaction networks.* (2016) *Nucleic Acids Research*, 44 (W1), pp. W367-74.
- [Poland 1997] Poland B. W., Bruns C., Fromm H. J. and Honzatko R. B. *Entrapment of 6-thiophosphoryl-IMP in the active site of crystalline adenylosuccinate synthetase from Escherichia coli.* (1997) *J Biol Chem.* Jun 13;272(24):15200-5.
- [Quintas 2001] Quintas A., Vaz D.C., Cardoso I., Saraiva M. J and Brito RM. *Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants.* (2001) *J Biol Chem.*; 276: 27207–27213. <https://doi.org/10.1074/jbc.M101024200> PMID: 11306576
- [Rader 2004] Rader A. J. and Bahar I. *Folding core predictions from network models of proteins.* (2004) *Polymer* 45, 659 .
- [Redondo 2000] Redondo C., Damas A. M. and Saraiva M. J. *Structural evidence for native state stabilization of a conformationally labile amyloidogenic transthyretin variant by fibrillogenesis inhibitors.* (2000) *Biochem J.*; 348: 167–172.
- [Roujeinikova 2007] Roujeinikova A., Simon W. J., Gilroy J., Rice D. W., Rafferty J. B. and Slabas A. R. *Structural Studies of Fatty Acyl-(Acyl Carrier Protein) Thioesters Reveal a Hydrophobic Binding Cavity that Can Expand to Fit Longer Substrates.* (2007) *J.Mol.Biol.*; 365: 135–145. PMID: 17059829
- [Saelices 2015] Saelices L., Johnson L. M., Liang W. Y., Sawaya M. R., Cascio D., Ruchala P., Whitelegge J., Jiang L., Riek R. and Eisenberg D. S. *Uncovering the mechanism of aggregation of human transthyretin.* (2015) *J Biol Chem.*; 290:28932–28943.
- [Saldaño 2016] Saldaño T. E., Monzon A. M., Parisi G. and Fernandez-Alberti S. *Evolutionary Conserved Positions Define Protein Conformational Diversity.* (2016) *PLOS Computational Biology* 12(3): e1004775. <https://doi.org/10.1371/journal.pcbi.1004775>.
- [Saldaño 2017] Saldaño T. E., Zanotti G., Parisi G., Fernandez-Alberti S. *Evaluating the effect of mutations and ligand binding on transthyretin homotetramer dynamics.* (2017) *PLoS ONE* 12(7): e0181019. <https://doi.org/10.1371/journal.pone.0181019>
- [Sanejouand 2013] Sanejouand Y.-H. *Elastic Network Models: Theoretical and Empirical Foundations.* (2013) *Methods in molecular biology (Clifton, N.J.)*. 924. 601-16. 10.1007/978-1-62703-017-5\_23.
- [Sant’Anna 2016] Sant’Anna R., Gallego P., Robinson L. Z., Pereira-Henriques A., Ferreira N., Pinheiro F., Esperante S., Pallares I., Huertas O., Rosário Almeida M., Reixach N., Insa R., Velazquez-Campoy A., Reverter D., Reig N. and Ventura S. *Repositioning tolcapone as a potent inhibitor of transthyretin amyloidogenesis and associated cellular toxicity.* (2016) *Nature Communications*; 7:10787. doi:10.1038/ncomms10787.

- [Schneider 1997] Schneider R., de Daruvar A. and Sander C. *The HSSP database of protein structure —sequence alignments*. (1997) *Nucleic Acids Res.*; 25 (1): 226-230. doi: 10.1093/nar/25.1.226
- [Schymkowitz 2005] Schymkowitz J., Borg J., Stricher F., Nys R., Rousseau F. and Serrano L. *The FoldX web server: An online force field*. (2005) *Nucleic Acids Res.*; 33: 382±388. <https://doi.org/10.1093/nar/gki387> PMID: 15980494.
- [Seno 1990] Seno Y. and Go N. *Deoxymyoglobin studied by the conformational normal mode analysis*. (1990) *J. Mol. Biol.* 216:95–109.
- [Silva 1995] Silva M. M., Poland B. W., Hoffman C. R., Fromm H. J and Honzatko R.B. *Refined crystal structures of unligated adenylosuccinate synthetase from Escherichia coli*. (1995) *J Mol Biol. Dec 1;254(3):431-46*.
- [Sinha 2001] Sinha N. and Nussinov R. *Point mutations and sequence variability in proteins: redistributions of preexisting populations*. (2001) *Proc Natl Acad Sci U S A*, 98, 3139-3144.
- [Sipe 1994] Sipe J. D. *Amyloidosis*. (1994) *Crit Rev Clin Lab Sci.*; 31: 325±354. <https://doi.org/10.3109/10408369409084679> PMID: 7888076
- [Sobolev 1999] Sobolev V., Sorokine A., Prilusky J., Abola E. E. and Edelman M. *Automated analysis of interatomic contacts in proteins*. (1999) *Bioinformatics*; 15(4):327–32. PMID: 10320401
- [Tama 2000] Tama F., Gadea F. X., Marques O., Sanejouand Y.-H. Building-block approach for determining low- frequency normal modes of macromolecules. *Proteins Struct Funct Bioinforma.* 2000; 41(1):1–7.
- [Tama 2001] Tama F. and Sanejouand Y.-H. *Conformational change of proteins arising from normal mode calculations*. (2001) *Protein Eng.* 14(1):1-6.
- [Taraskin 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.
- [Temiz 2002] 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.
- [Thibert 2005] Thibert B., Bredesen D. E. and del Rio G. *Improved prediction of critical residues for protein function based on network and phylogenetic analyses*. (2005) *BMC Bioinformatics*; 6:213.
- [Thomas 2003] Thomas P. D., Campbell M. J., Kejariwal A., Mi H., Karlak B., Daverman R., Diemer K., Muruganujan A. and Narechania A. *PANTHER: A Library of Protein Families and Subfamilies Indexed by Function*. (2003) *Genome Research*;13(9):2129-2141. doi:10.1101/gr.772403.
- [Tipler 2008] Tipler P. A. Llewellyn R. A. *Chapter 8, More Chapter8. A derivation of the Equipartition Theorem*. *Modern Physics - Fifth Edition*. (2008) W. H. Freeman and Company.

- [Tirion 1996] Tirion M.M. *Large amplitude elastic motions in proteins from a single-parameter, atomic analysis.* (1996) Phys. Rev. Lett. 77:1905–1908.
- [Tokuriki 2007] Tokuriki N., Stricher F., Schymkowitz J., Serrano L. and Tawfik D. S. *The Stability Effects of Protein Mutations Appear to be Universally Distributed.* (2007) J Mol Biol.; 369: 1318±1332. <https://doi.org/10.1016/j.jmb.2007.03.069> PMID: 17482644.
- [Tokuriki 2009] Tokuriki N. and Tawfik D. S. *Protein dynamism and evolvability.* (2009) Science, 324, 203-207.
- [Tomar 2012] Tomar D., Khan T., Singh R. R., Mishra S., Gupta S., Surolia A. and Salunke D. M. *Crystallographic Study of Novel Transthyretin Ligands Exhibiting Negative-Cooperativity between Two Thyroxine Binding Sites.* (2012) PLoS ONE. Public Library of Science;7: e43522. doi:10.1371/journal.pone.0043522
- [Touw 2015] Touw W. G., Baakman C., Black J., te Beek T. A. H., Krieger E., Joosten R. P. and Vriend G. *A series of PDB related databases for everyday needs.* (2015) Nucleic Acids Research; January; 43(Database issue): D364-D368.
- [Tsai 1999] Tsai C.J., Kumar S., Ma B.Y., Nussinov R. *Folding Funnels, Binding Funnels, and Protein Function.* Protein Sci. 8, 1181-1190 (1999).
- [UniProt 2015] The UniProt Consortium. *UniProt: a hub for protein information.* (2015) Nucleic Acids Res.; 43: D204-D212. doi: 10.1093/nar/gku989 PMID: 25348405
- [Vliegthart 1999] Vliegthart J. F. G., Erbel P. J. A., Karimi-Nejad, Y., Beer T. de; Boelens R. and Kamerling, J. P. *Solution structure of the  $\alpha$ -subunit of human chorionic gonadotropin.* (1999) European Journal of Biochemistry, volume 260, issue 2, pp. 490 – 498
- [Vendruscolo 2002] Vendruscolo M., Dokholyan N. V., Paci E. and Karplus M. *Small- world view of the amino acids that play a key role in protein folding.* (2002) Phys. Rev. E; 65(6)
- [Wang 2004] Wang Y., Rader A. J., Bahar I. and Jernigan R. L. *Global ribosome motions revealed with elastic network model.* (2004) J. Struct. Biol. 147(3):302-14.
- [Wangikar 2003] Wangikar P. P., Tendulkar A. V., Ramya S., Mali D. N. and Sarawagi S. *Functional sites in protein families uncovered via an objective and automated graph theoretic approach.* (2003) J. Mol. Biol. 326, 955–978.
- [Wickham 2016] Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* (2016) Springer-Verlag New York.
- [Wojtczak 1996] Wojtczak A., Cody V., Luft J. R. and Pangborn W. *Structures of human transthyretin complexed with thyroxine at 2.0 Å resolution and 3',5'-dinitro-N-acetyl-L-thyronine at 2.2 Å resolution.* (1996) Acta Crystallogr D Biol Crystallogr; 52(Pt 4):758–765.
- [Wolynes 1995] Wolynes P. G., Onuchic J. N. and Thirumalai, D. *Navigating the folding routes.* (1995) SCIENCE-NEW YORK THEN WASHINGTON-, 1619-1619.
- [Xu 2003] Xu C., Tobi D. and Bahar I. *Allosteric changes in protein structure computed by a simple mechanical model: hemoglobin T -> R2 transition.* (2003) J. Mol. Biol. 333:153–168.

- [Yang 2005] Yang L. W. and Bahar I. *Coupling between catalytic site and collective dynamics: A requirement for mechanochemical activity of enzymes*. (2005) *Structure*; 13(2):893–904;
- [Yang 2013] Yang J., Roy A. and Zhang Y. *BioLiP: A semi-manually curated database for biologically relevant ligand protein interactions*. (2013) *Nucleic Acids Res.*; 41(D1):1096–103.
- [Zanotti 2013] Zanotti G., Cendron L., Folli C., Florio P., Imbimbo B. P. and Berni R. *Structural evidence for native state stabilization of a conformationally labile amyloidogenic transthyretin variant by fibrillogenesis inhibitors*. (2013) *FEBS Lett*; 587:2325–2331.
- [Zanotti 2017] Zanotti G., Vallese F., Ferrari A., Menozzi I., Saldaño T. E., Fernandez-Alberti S., and Berni R. *Structural and dynamics evidence for scaffold asymmetric flexibility of the human Transthyretin tetramer*. (2017) *PLoS ONE* 12(12): e0187716. <https://doi.org/10.1371/journal.pone.0187716>.
- [Zhang 2016] Zhang J., Begum A., Brännström K., Grundström C., Iakovleva I., Olofsson A., et al. *Structure-Based Virtual Screening Protocol for in Silico Identification of Potential Thyroid Disrupting Chemicals Targeting Transthyretin*. (2016) *Environ Sci Technol.*; 50: 11984±11993. <https://doi.org/10.1021/acs.est.6b02771> PMID: 27668830.
- [Zheng 2005] Zheng W., Brooks B. R., Doniach S. and Thirumalai D. *Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved*. (2005) *Structure*; 13(4):565–77. PMID:15837195.
- [Zheng 2006] Zheng W., Brooks B.R. and Thirumalai D. *Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations*. (2006) *Proc Natl Acad Sci U S A*; 103 (20):7664–9. PMID: 16682636
- [Zheng 2009a] Zheng W. and Thirumalai D. *Coupling between normal modes drives protein conformational dynamics: illustrations using allosteric transitions in myosin II*. (2009) *Biophys J.*; 96(6):2128–37. doi: 10.1016/j. bpj.2008.12.3897 PMID: 19289039
- [Zheng 2009b] Zheng W. and Tekpinar M. *Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model*. (2009) *BMC Struct Biol.*; 9:45. doi: 10.1186/1472-6807-9-45 PMID: 19591676