



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

Análisis termodinámico, fibrilogénico, y estructural de cadenas ligeras tipo λ VI
implicadas en la amiloidosis AL

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. Oscar Daniel Luna Martínez

TUTOR PRINCIPAL

Dr. Baltazar Becerril Luján
[Instituto de Biotecnología, UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

Dr. Joel Osuna Quintero
[Instituto de Biotecnología, UNAM](#)
Dra. Elba Villegas Villarreal
[Centro de Investigación en Biotecnología, UAEM](#)

Cuernavaca. marzo, 2017



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

ÍNDICE

	Página
Índice de Figuras	III
Índice de Tablas	IV
Lista de Abreviaturas	IV
Agradecimientos	V
Abstract	1
Resumen	2
Introducción	
Enfermedades derivadas del plegamiento anómalo de las proteínas	3
Amiloidosis	3
Asociación preferencial de la región variable de las cadenas ligeras λ VI en la amiloidosis AL	4
Antecedentes Directos del Proyecto	8
Análisis de redes como método de análisis	9
Hipótesis y Objetivos	11
Materiales y Métodos	
Estadística	12
Estructuras cristalográficas para la validación	12
Análisis de redes	12
Grado	13
Eigenvectores	13
Entropía de redes	14
Dinámica molecular	15
Entropía dinámica de la dinámica molecular	16
Mutantes de 6aJL2 y de AR	17
Desnaturalización térmica	17
Dicroísmo circular	18
Cristalización y recolección de datos	18
Determinación estructural y refinamiento	19

Resultados

Parte I. Diferencias y similitudes estructurales de las cadenas ligeras λ VI	
Determinación y comparación de las estructuras cristalográficas de AR y de 6a-R25G contra la estructura de 6aJL2	20
Aplicación de análisis de redes sobre 6aJL2 para identificar los residuos clave involucrados en la estabilización de este dominio	24
Propuesta y validación de una metodología para la jerarquización de residuos empleando proteínas evaluadas termodinámicamente con diferente estructura terciaria	26
Evaluación de las estructuras de AR y de 6aJL2 con la metodología propuesta	37
Parte II.- Análisis termodinámico y estructural de AR	
Selección y análisis de tres residuos importantes en AR	42
Conservación de la estructura secundaria de las mutantes de AR	43
Estabilidad termodinámica de las mutantes de AR	44
Cinéticas de fibrilogénesis y visualización de las fibras por microscopía electrónica	47
Cambios detectables en el extremo N-terminal generados por la ausencia de Arg25	48
Generación de estructuras cuaternarias a partir de las estructuras cristalográficas	49
Discusión	
Amiloidosis AL λ VI	52
AR, la contraparte de 6aJL2	53
Convergencia de las estructuras octaméricas	55
Otros aspectos de las proteínas amiloidogénicas	56
Bibliografía	58
Anexo A. Jerarquización de los residuos de 6aJL2	68
Anexo B. Jerarquización de los residuos de AR	70
Anexo C. Artículos generados	72

ÍNDICE DE FIGURAS

	Página
Figura 1. Proceso de maduración de una célula plasmática	6
Figura 2. Inmunoglobulina	6
Figura 3. Líneas germinales λ	7
Figura 4. Fibrillogénesis <i>in vitro</i> de rV _L 6aJL2, JTO, WIL y R25G	8
Figura 5. Alineamiento de secuencias de cadenas ligeras $\lambda 6$	10
Figura 6. Matriz Hermitiana	13
Figura 7. Matriz estocástica	14
Figura 8. Traslape de estructuras	21
Figura 9. Convergencia entre las metodologías Grado, Eigenectores, y Entropía de Redes	25
Figura 10. Contraste entre el parámetro Grado y valores experimentales	27
Figura 11. RMSD de la cadena principal	28
Figura 12. Visualización de los valores SDRI el extremo N-terminal intrínsecamente desestructurado de la región de interacción de la Janus Kinase.	30
Figura 13. Contraste entre los SDRI y su posición en la secuencia	31
Figura 14. Gradiente de color	32
Figura 15. Contraste de los SDRI contra sus correspondientes valores experimentales	34
Figura 16. Comparación de los SDRI con las cinéticas de desnaturalización y renaturalización del inhibidor de quimiotripsina	35
Figura 17. Análisis del desorden intrínseco y la flexibilidad estructural	36
Figura 18. Comparación de SDRI de 6aJL2 y AR	37
Figura 19. Cambios jerárquicos en AR	40
Figura 20. Cavidades	40
Figura 21. Efectos de la temperatura sobre la fluorescencia intrínseca	43
Figura 22. Espectros en UV lejano en estado nativo, desnaturalizado, y renaturalizado	44
Figura 23. Experimentos de desnaturalización	45
Figura 24. Fibrillogénesis <i>in vitro</i>	47
Figura 25. Microscopía de fibras amiloides	48
Figura 26. Extremos N-terminal	49
Figura 27. Dímeros de Bence-Jones	50
Figura 28. Octámeros	51

ÍNDICE DE TABLAS

	Página
Tabla 1. Colección de datos y estadística del refinamiento estructural	20
Tabla 2. Interacciones intramoleculares	23
Tabla 3. Residuos de 6aJL2 mejor y peor calificados	25
Tabla 4. Estadística de la distribución valores teóricos – SDRI	30
Tabla 5. Estadística de SDRI contra datos experimentales	32
Tabla 6. Jerarquía de residuos de 6aJL2 y AR	38
Tabla 7. Comparación de Jerarquías	39
Tabla 8. Estabilidad termodinámica	46

ABREVIATURAS

ADN	Ácido Desoxirribonucleico
C_m	Concentración media de desnaturizante
GndHCl	Cloruro de Guanidinio
ΔG°	Energía libre de Gibbs
ΔH°	Entalpía
ΔCp	Capacidad calórica
kcal	Kilocalorías
mM	Milimolar
μl	Microlitros
μm	Micrómetros
ns	Nanosegundos
NMR	Resonancia Magnética Nuclear
NPT	Ensamble isotérmico-isobárico.
PBS	Buffer de fosfatos
PDB	<i>Protein DataBase</i>
RMSD	<i>Root Mean Square Deviation</i>
RMSF	<i>Root Mean Square Fluctuation</i>
RSA	Área Expuesta al Solvente del Residuo (<i>Residue Solvent Accessibility</i>)
SPC	Agua tipo <i>Simple Point Charge</i>
SDRI	<i>Structural-Determining Residue Identifier</i>
ThT	Tioflavín T
T_m	Temperatura media de desnaturalización
C_L	Región constante de la cadena ligera
V_L	Región variable de cadena ligera
J_L	Segmento de unión entre C _L y V _L

AGRADECIMIENTOS

AGRADECIMIENTOS ACADÉMICOS

Al IBt, y a la UNAM por la diversidad cultural e intelectual que me permitieron ampliar los horizontes. Al laboratorio donde pasé un poco más de 10 años de vida, especialmente al Dr. Possani por su apoyo y por el café matutino. A mi jefe, Dr. Becerril, por su confianza para desarrollar una metodología para entender el tema de amiloidosis. Al Programa de Apoyo a los Estudios del Posgrado (PAEP) por el apoyo a los congresos. A Gloria, Toño, y Jalil de la Unidad de Docencia por siempre facilitar la tramitología. A CONACYT, por la beca número 177224; y a DGAPA IN 217510 por la financiación del proyecto.

A quienes estuvieron presentes en los comités de candidaturas, tutorales, tutorales ampliados, y a mis sinodales por sus valiosas sugerencias. Al Dr. Joel Osuna por todos los ánimos. Al Dr. Daniel Alejandro Fernández, Dr. Miguel Costas, Dra. Gloria Saab, y al Dr. Ramón Garduño por sus aportaciones y críticas puntuales. Al laboratorio Nacional de Estructura de Macromoléculas, a la Dra. Alejandra Hernández y la Dra. Adela Rodríguez por su apoyo para generar y analizar las estructuras cristalográficas. Al Dr. Vladimir Uversky por su invaluable propuesta.

A la Unidad de Síntesis y Secuenciación del IBt- Paul Gaytán, Eugenio López, Jorge Yáñez, y a Santiago Becerra, por tener los oligos a tiempo así como los momentos de análisis y convivencia. A los miembros de la Unidad de Cómputo- Arturo Ocádiz, Juan Manuel Hurtado, Roberto Rodríguez, y David Castañeda, por todas las atenciones prestadas.

AGRADECIMIENTOS NO ACADÉMICOS

Difícil mencionar a cada una de las personas que me apoyaron, convidaron, toleraron, auspiciaron, hostigaron, y brindaron amistad. A mis coterráneos en el laboratorio que hacían el día a día algo más llevadero como Zamudio, Gina Gurrola, Fredy, Tere, Jimena, Mary, Linda, etc. Con quienes tuve la oportunidad de convivir y hacer amistad en el IBt como Magda, Montse, Anilú, Abraham, etc. A todos los académicos y no académicos del IBt, como el Dr. Rudiño, Dra. Marcela, Adrián, etc. A Carmencici, Miryam, y Rosalba por compartir muchos momentos tanto dentro como fuera de la mesa de trabajo. A Cristina Torres, porque siempre me has ayudado cuando más lo he requerido. A Efrén, por ser un excelente amigo.

A la familia de Miryam, por todo el respaldo en esta última etapa. A mi adorada familia de Chicago, por brindar su aprecio incondicional a pesar de la distancia. A mi familia nuclear- mi papá (☩), mi mamá, mi hermana Nayeli, y a mi hermano Daniel, porque nos hemos mantenido unidos en diversas circunstancias, los quiero mil.

Esta etapa me inspira una de las frases de mi escritor favorito José Saramago: “En el fondo, todos tenemos la necesidad de decir quiénes somos y qué es lo que estamos haciendo, la necesidad de dejar algo hecho, porque esta vida no es eterna y dejar cosas hechas puede ser una forma de eternidad.”

ABSTRACT

Amyloidosis AL is a systemic disease where variable domains of some immunoglobulin light chain families' orderly aggregate into amyloid fibers. The multiorgan dysfunction irretrievably leads to the death of the patient diagnosed with this disease. In our group, we have developed an experimental methodology that has helped us to study the molecular bases associated with the fibrillogenesis of one of the most implicated families in this disease, the germ line $\lambda 6a$. While 6aJL2 is a stable model protein, AR is an unstable variant obtained from a patient who died few months after being diagnosed with AL amyloidosis.

In the first part of this work, a theoretical methodology was proposed and validated in order to characterize globular protein structures, regardless their tertiary or quaternary assembly, through the hierarchization of each residue sequence according to its structural relevance related to thermodynamic parameters. This methodology was subsequently applied to 6aJL2 and AR structures depicting relevant differences between hierarchies, mainly in their anti-aggregation segments.

In the second part of this work, AR was mutated back to 6aJL2 in some selected residues and stabilizing effects were obtained with mutations located within hydrophobic core. Furthermore, a previous hypothesis about the influence of the residue in position 25 has been withdrawn because it was demonstrated its null influence in the folding of AR CDR1 segment. While, monomeric comparison among mutant 6a-R25G, 6aJL2, and AR not only showed both unique and shared interactions, slight modifications in the N-terminal segment might be contributing to enhance well-ordered aggregation. Additionally, the most remarkable data obtained from crystallographic structures was obtained through the generation of *in silico* assembled quaternary structures. Both octameric assembling structures from AR and 6aJL2 converged with the crystallographic octameric ensemble of 6a-R25G, which might be revealing a common assembling process throughout oligomer development, opening new perspectives about the influence of the destabilizing mutations.

RESUMEN

La amiloidosis AL es una enfermedad sistémica en la cual los dominios variables de ciertas familias de cadenas ligeras de inmunoglobulina se agregan ordenadamente en forma de fibras amiloides. La disfuncionalidad multiorgánica conduce irremediablemente a la muerte del paciente que es diagnosticado con esta enfermedad. En nuestro grupo, hemos desarrollado un sistema experimental que nos ha permitido estudiar algunas de las bases moleculares asociadas a la fibrilogenesis de una de las familias de cadenas ligeras mayormente implicadas en esta enfermedad, la línea germinal $\lambda 6a$. Mientras que 6aJL2 es una proteína modelo estable, AR es una variante altamente inestable, la cual fue obtenida de un paciente que falleció a los pocos meses después de haberlo diagnosticado con amiloidosis AL.

En la primera parte de este trabajo se propuso y se validó una metodología que permitiera evaluar las cualidades estructurales de proteínas globulares, irrespectivamente de su estructuración terciaria o cuaternaria, mediante la jerarquización de cada residuo en términos de su aportación para mantener una estabilidad estructural. Esta metodología fue posteriormente aplicada en las estructuras de 6aJL2 y en AR mostrando diferencias relevantes entre las jerarquías, principalmente en los segmentos de anti-agregación.

En la segunda parte de este trabajo, se mutó a AR de regreso a la secuencia de 6aJL2 en ciertos residuos seleccionados. Los resultados revelaron que sólo hubo efectos estabilizantes cuando los residuos localizados en el núcleo hidrofóbico fueron regresados a los correspondientes a la línea germinal. Además, de acuerdo con los resultados obtenidos, una hipótesis previamente planteada acerca del efecto estabilizador de un residuo de Arginina en la posición 25 fue rechazada ya que se demostró una nula influencia en la estabilización de la variante correspondiente a AR. La comparación entre monómeros de 6aJL2, su mutante 6a-R25G, y AR mostraron interacciones únicas y compartidas así como ligeras modificaciones en la estructuración de sus segmentos N-terminal que podrían contribuir con la aceleración de la agregación ordenada. A partir de los datos cristalográficos se generaron *in silico* estructuras octaméricas de 6aJL2 y de AR los cuales formaron ensamblajes coincidentes con en el octámero cristalográfico de 6a-R25G. Estos resultados pudieran sugerir un mecanismo de ensamblaje común durante la formación de oligómeros en el proceso de la formación de fibras, abriendo nuevas perspectivas sobre la influencia de las mutaciones desestabilizantes.

INTRODUCCIÓN

Enfermedades derivadas del plegamiento anómalo de las proteínas

Michel de Montaigne, filósofo humanista renacentista, escribió: “Mas tú no mueres porque estás enfermo, mueres porque eres vivo” (Montaigne, 2015). Bajo esta frase se puede aludir que aquellos que no seremos víctimas de asesinatos o accidentes, falleceremos como consecuencia de alguna de las tantas enfermedades, infecciosas o no, que aquejan a la humanidad. Dentro de todo el conjunto de enfermedades está descrito un subconjunto que engloba a las enfermedades derivadas por el plegamiento anómalo de las proteínas donde una o varias mutaciones conllevan a la pérdida de la función a diferentes niveles celulares (Gregersen, 2006). Por ejemplo, la fibrosis quística es una enfermedad de tipo hereditaria debido a mutaciones en distintas regiones que generan la pérdida de función de un canal transmembranal encargado principalmente de la difusión y el intercambio entre los iones cloro y bicarbonato (O’Sullivan & Freedman, 2009). Aunque órganos como el hígado, el tracto gastrointestinal, o el páncreas, entre otros, resultan seriamente afectados, la acumulación de cloro en los pulmones impide que la función de antibióticos innatos, como la β -defensina 1, actúen correctamente permitiendo que las bacterias se alojen y desarrollen (Goldman *et al.*, 1997). Otra enfermedad derivada por el plegamiento anómalo de las proteínas es la fenilcetonuria. Este padecimiento de tipo hereditario es causado por un error metabólico derivado de mutaciones en el gen de la fenilalanina hidroxilasa, reduciendo o nulificando la actividad enzimática lo cual causa una acumulación de fenilalanina que por su toxicidad afecta al sistema nervioso central (Blau *et al.*, 2010). Muchas de esas mutaciones afectan el dominio catalítico de esta enzima, y en menores porcentajes están afectados también los sitios de unión al cofactor tetrahidrobiopterina, al oxígeno molecular y al hierro, asimismo también las mutaciones afectan al dominio regulatorio y al dominio de homo-tetramerización (Zurflüh *et al.*, 2008). Sin embargo, el grupo más grande de enfermedades derivadas del plegamiento anómalo de proteínas está asociado a las proteínas, o péptidos, que no sólo han perdido su función original sino también se agregan ordenadamente en forma de fibras (Chiti & Dobson, 2006).

Amiloidosis

Las amiloidosis son un grupo de enfermedades derivadas del plegamiento anómalo de las proteínas donde proteínas, o péptidos, se agregan en forma de fibras conocidas como amiloides pudiéndose depositar en uno o varios órganos (Chiti & Dobson, 2006). En las amiloidosis localizadas, en las que solo un órgano es afectado, podemos encontrar enfermedades neurodegenerativas como el Alzheimer, Parkinson, y Huntington. En el grupo de amiloidosis localizadas no neurodegenerativas están enfermedades como la llamada catarata de los ojos y la diabetes tipo 2. El último grupo lo comprenden las amiloidosis sistémicas en las

cuales la agregación amiloide afecta a varios órganos como en el caso de la amilodosis derivada de la lisozima, del fibrinógeno, y la amilodosis de cadenas ligeras de inmunoglobulinas.

Antes de dirigir el siguiente texto hacia una descripción más detallada de la amilodosis derivada de cadenas ligeras de inmunoglobulina, cabe resaltar que las proteínas involucradas en la formación de agregados amiloides carecen de alguna secuencia específica o alguna firma estructural que pudiera simplificar la búsqueda de proteínas susceptibles a la formación de fibras (Knowles *et al.*, 2014). Otra característica importante de las fibras amiloides, derivada de las observaciones de la microscopía electrónica, las describe como rectas, no ramificadas, con diámetros en el orden de los 10 nm y con longitudes desde los 100 nm hasta varias micras (Tycko, 2006). Adicionalmente, las fibras presentan un color verde manzana bajo luz polarizada después de haber sido teñidas con rojo Congo y se observa un aumento en la fluorescencia al ponerlas en contacto con Tioflavin T (Puchtler & Sweat, 1965; Levine, 1993). Por medio del dicroísmo circular y la difracción de rayos X, las fibras muestran un alto contenido de hojas β (Makin & Serpell, 2005).

En general, el resultado más interesante de la agregación amiloide es precisamente la formación de fibras con características estructurales en común a partir de una gran diversidad de secuencias. El efecto de los cambios estructurales puede involucrar desde un apilamiento directo, la interacción entre segmentos específicos, el intercambio de dominios, o hasta la reestructuración completa en el caso de los péptidos desestructurados (Nelson & Eisenberg, 2006). Además, estas rutas de agregación dependen de las mutaciones en la secuencia, la estabilidad termodinámica, y el ambiente químico celular (Fernandez-Escamilla *et al.*, 2004; Meng *et al.*, 2008). Una aproximación inicial para entender las causas de la amilodosis AL, cuya complejidad está acentuada por el grado de diversidad de secuencias, se centra justo en entender la influencia de las mutaciones en el desarrollo de esta enfermedad.

Asociación preferencial de la región variable de las cadenas ligeras λ VI en la amilodosis AL

Las gammapatías monoclonales incluyen un grupo de enfermedades caracterizadas por la proliferación de clonas de células plasmáticas que producen un único tipo de cadena ligera y/o pesada (componente monoclonal) en cantidades excesivas (The International Myeloma Working Group, 2003). La amilodosis AL es una gammapatía monoclonal causada por la agregación de fibras amiloides en forma sistémica, las cuales están conformadas por las regiones variables de cadenas ligeras (V_L) de algunas inmunoglobulinas. Por otro lado, sólo como dato complementario para reflejar la importancia de la amilodosis AL, la amilodosis AH se refiere a la agregación amiloide de fragmentos de cadenas pesadas de inmunoglobulinas aunque ésta es una enfermedad extraordinariamente inusual con una incidencia menor al 1% de las gammapatías monoclonales (Mai *et al.*, 2003; Miyazaki *et al.*, 2008).

La edad promedio en la que se manifiesta la amiloidosis AL es alrededor de los 65 años, siendo las dos terceras partes de los pacientes hombres y menos del 5% de ellos menores a 40 años (Makin & Serpell, 2005). Se ha reportado que la incidencia anual tan sólo en los Estados Unidos y en el Reino Unido puede variar desde 5 hasta 13 personas por cada millón de habitantes, sin contar aquellos casos diagnosticados erróneamente (Weiss *et al.*, 2016; Nienhuis *et al.*, 2016). Esta enfermedad se caracteriza por ser de condición progresiva y sistémica donde cerca del 70% de los pacientes presentan dos o más órganos involucrados siendo los riñones, corazón, hígado, sistema nervioso periférico y el sistema nervioso autónomo los más afectados (Pepys, 2006). Dependiendo de los órganos afectados, las manifestaciones clínicas de la amiloidosis AL llegan a ser muy variadas además de que la sintomatología es poco específica como fatiga, edema y pérdida de peso, entre otras (Pepys, 2006).

Antes de abordar un enfoque orientado hacia la parte estructural sobre la amiloidosis AL, es importante conocer el origen y las características de las inmunoglobulinas. Una célula B plasmática secretora de anticuerpos es originada a partir de una célula totipotencial ubicada en la médula ósea después de varios procesos de maduración en diferentes órganos (Figura 1). Los anticuerpos son generados a partir de los genes de anticuerpos, conocidos como líneas germinales, los cuales maduran a través de un mecanismo conocido como hipermutación somática dándole así una gran versatilidad al sistema inmune adaptativo de los mamíferos. Este proceso garantiza que el organismo pueda reconocer una gran cantidad de antígenos en la naturaleza (Benjamini *et al.*, 2000; Janeway *et al.*, 2001, Clark *et al.*, 2006). La estructura típica de una inmunoglobulina tipo IgG, mostrada en la Figura 2, está compuesta de dos cadenas pesadas y dos cadenas ligeras (Schroeder *et al.*, 2010). A su vez, la cadena ligera puede segmentarse en un dominio constante y en un dominio variable; siendo éste último el dominio implicado en la agregación amiloide (Figura 1B). Los dominios variables de las cadenas ligeras (V_L) pueden ser clasificados en κ ó λ , dependiendo de su secuencia la cual hace referencia al *locus* cromosomal correspondiente (Meffre *et al.*, 2000; Lefranc, 2001a; Lefranc, 2001b; Schroeder & Cavacini, 2010). Dada la naturaleza monoclonal de las V_L amiloidogénicas, se ha conjeturado que cada V_L procede de una célula B plasmática que ha pasado por todo el proceso de maduración celular pero que, hasta la fecha, se desconoce el tipo de discrasia celular que desencadena la sobre-excreción de estas V_L libres (Merlini & Stone, 2006). Una V_L está estructuralmente compuesta por dos regiones, las Regiones Marco (Framework Regions) que sirven como andamiaje para las Regiones Determinantes de Complementariedad (CDRs) las cuales se reconocen y se unen a su respectivo antígeno.

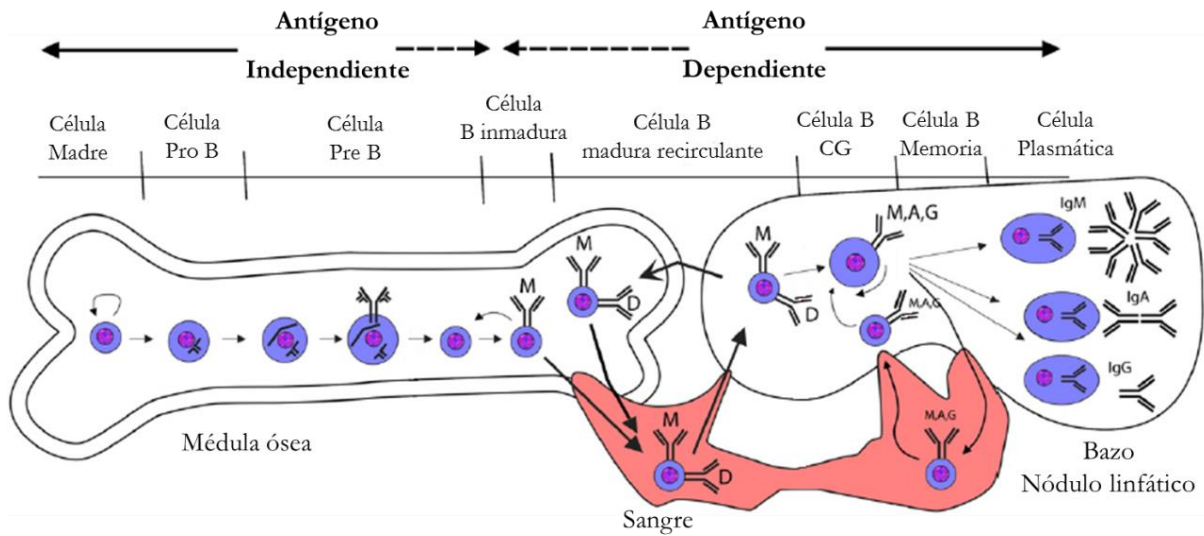


Figura 1. Proceso de maduración de una célula plasmática, adaptado de Schroeder *et al*, 2010. El desarrollo de una célula B mediante el rearreglo de la inmunoglobulina después de su nacimiento en la médula ósea y su estimulación previa antes de su estimulación ante el antígeno. Posterior al desarrollo, la célula B incrementa su dependencia al estímulo del antígeno. Las células B inmaduras abandonan la médula ósea, empiezan a expresar IgD, y circulan en la sangre hasta llevar a los órganos linfoides secundarios y la médula ósea. El encuentro con el antígeno cognado puede causar la transformación celular hacia una célula B de memoria o una célula plasmática. Los centros germinales (CG) son sitios dentro de los nódulos linfoides secundarios donde los linfocitos B proliferan, se diferencian, y mutan sus genes (a través de la hipermutación somática), e intercambian el tipo de anticuerpo (IgG, IgA, o IgM) que producen durante la respuesta inmunológica normal a una infección.

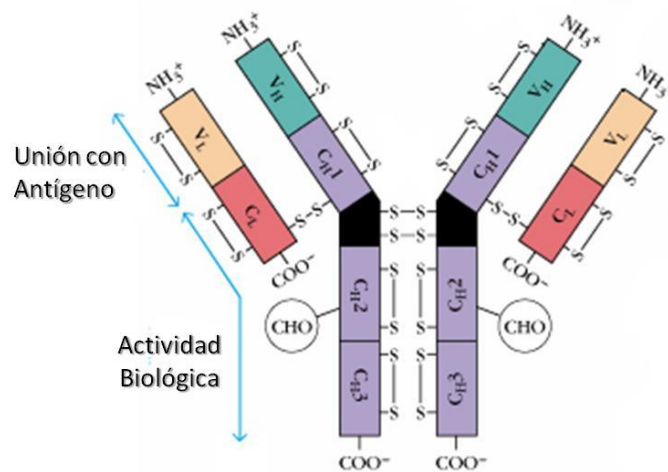


Figura 2. Inmunoglobulina. Estructura general de una inmunoglobulina IgG, adaptado de LeFranc, 2001a. En morado se muestran los dominios constantes de las cadenas pesadas (C_{H1} , C_{H2} , C_{H3}) y en bermellón el dominio constante de las cadenas ligeras (C_L). Los dominios variables de las cadenas ligeras (V_L) en color salmón, y los dominios variables de las cadenas pesadas (V_H) en color verde.

La complejidad de estudiar la amiloidosis AL radica en que las V_L están expuestas al mecanismo de la hipermutación somática, generando así una gran variabilidad de secuencias (Meffre *et al.*, 2000; Pieper *et al.*, 2013). Sin embargo, a pesar de contar con una gran diversidad de líneas germinales, varios análisis de incidencia de amiloidosis mostraron que las $V_L \lambda$ tuvieron una mayor incidencia con respecto a las $V_L \kappa$, en una proporción de 3:1; mientras que el uso preferencial de cadenas ligeras en un individuo sano es 1:2 (Perfetti *et al.*, 2002; Abraham *et al.*, 2003). Entonces, una de las cuestiones más importantes sobre esta enfermedad es conocer la razón del por qué las cadenas ligeras tipo λ tienen una alta incidencia en la amiloidosis AL a pesar de tener un menor uso preferencial. En la Figura 3 se muestra que las $V_L \lambda$ derivadas de los genes $3r$ y $6a$ se encuentran altamente asociadas con la amiloidosis AL (Perfetti *et al.*, 2002). Un dato que llama la atención de la Figura 3 es que la expresión policlonal de la línea germinal $6a$ es considerablemente menor a su incidencia en la amiloidosis AL e, inclusive, menor con respecto a las otras líneas germinales. Otros trabajos también hacen referencia a la alta incidencia de esta línea germinal alcanzando desde un 20% hasta un 40% de las series de pacientes estudiados (Solomon *et al.*, 1982; Solomon & Wiss, 1995; Perfetti *et al.*, 1998). Con esta gran tendencia de las V_L derivadas de la línea germinal $\lambda 6a$ a la agregación amiloide se empezó a formular la siguiente pregunta: ¿qué tiene de especial la línea germinal $\lambda 6a$ que la hace susceptible a la agregación amiloide?

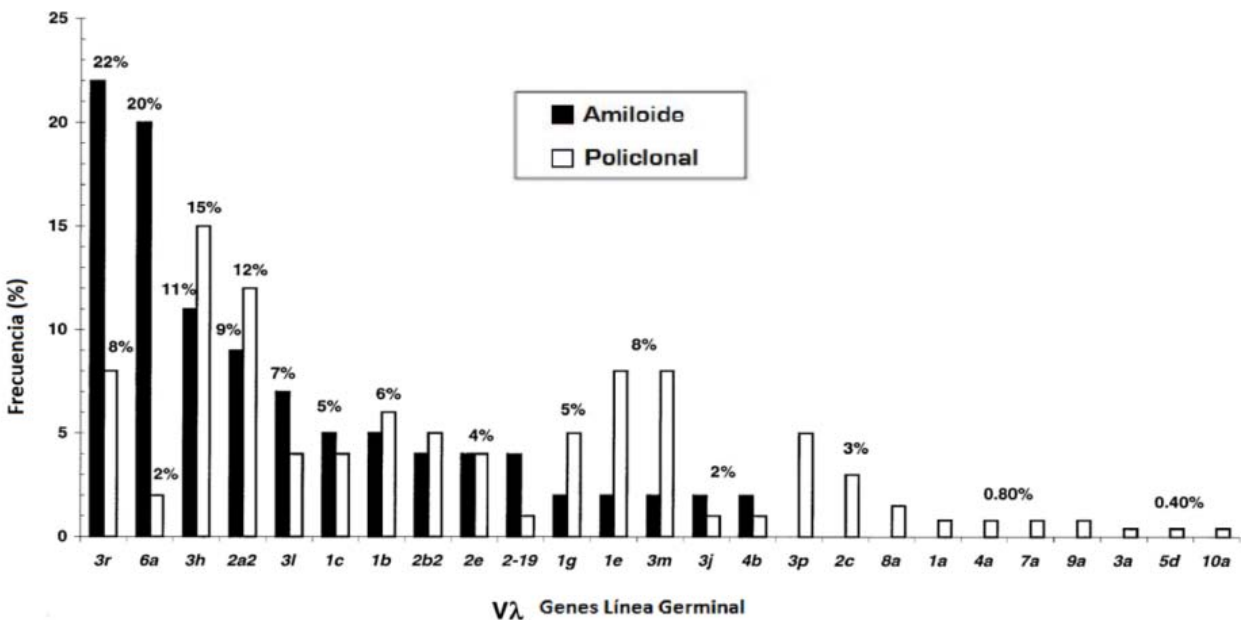


Figura 3. Líneas Germinales λ , tomado de Perfetti *et al.*, 2002. Contraste de la expresión del repertorio de líneas germinales $V\lambda$ de células plasmáticas de individuos sanos y la incidencia de la misma con V_L extraídas de pacientes con amiloidosis AL.

ANTECEDENTES DIRECTOS DEL PROYECTO

La manera en que inició la investigación para explicar la alta tendencia de las V_L derivadas de la línea germinal $\lambda 6a$ con la amiloidosis AL fue con la hipótesis que alude a la amiloidogenicidad intrínseca del gen $\lambda 6a$, implicando que las V_L derivadas sean amiloidogénicas y que durante el proceso de hipermutación somática aumentaría esta tendencia. En nuestro grupo se planteó y se analizó esta hipótesis en el trabajo de Luis del Pozo (del Pozo *et al.*, 2008). En resumen, se sintetizó y se expresó el gen de la línea germinal $6a$ junto con el segmento $J\lambda 2$, denominándose la proteína como $6aJL2$. Para analizar las propiedades biofísicas de esta proteína modelo se determinó su estabilidad termodinámica y se analizó su cinética de formación de fibras, o fibrilogénesis *in vitro*. Los resultados mostraron que $6aJL2$ era más estable que otras V_L derivadas de pacientes (AR, WIL y JTO), las cuales han sido utilizadas para estudiar el efecto de las mutaciones somáticas sobre sus propiedades amiloidogénicas (del Pozo *et al.*, 2008; González-Andrade, 2010). Además, Otros análisis con mutantes sitio específicas de esta proteína mostraron una correlación entre la estabilidad y la tendencia a formar fibras *in vitro* (del Pozo-Yauner *et al.*, 2008; González-Andrade, 2013; del Pozo-Yauner *et al.*, 2014).

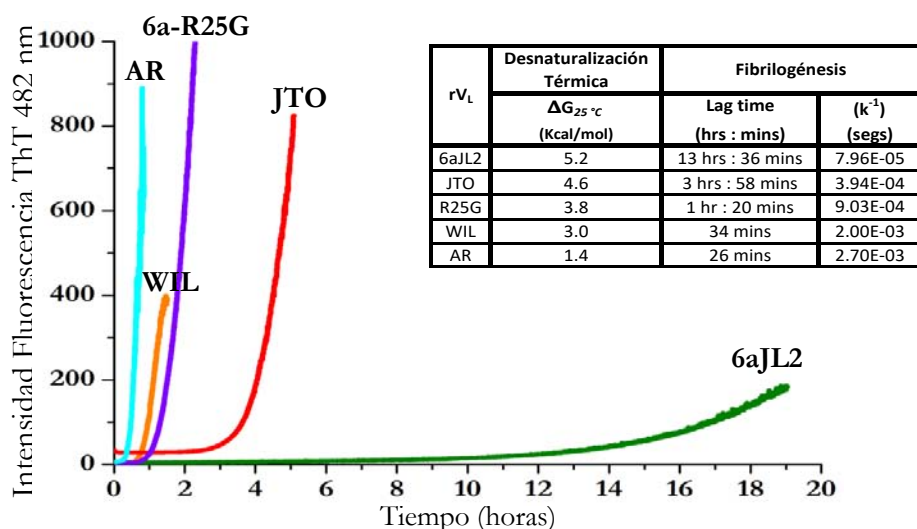


Figura. 4. Fibrilogénesis *in vitro* de rV_L 6aJL2, JTO, WIL y R25G. El experimento se realizó a $37^\circ C$, en un volumen de 3 mL, a una concentración de $100 \mu g/mL$ de proteína en PBS 1X y ThT $20 \mu M$ en agitación constante. Cada curva representa la variación de la intensidad de fluorescencia de ThT a $482 nm$ con respecto al tiempo. **Tabla interna.** Se muestran la ΔG° , el tiempo *lag* y la constante de velocidad de las 5 rV_L .

Desde un punto muy general, todo lo anterior se asemeja al postulado de Jean-Jaques Rosseau que dice que el hombre es bueno por naturaleza y que la sociedad lo pervierte; análogamente se podría decir que $6aJL2$ es estable y resistente a la formación de fibras pero que, durante el proceso de hipermutación somática, hay altas probabilidades de que las mutaciones generadas tengan como consecuencia una pérdida de la estabilidad termodinámica

y una exacerbación de las propiedades fibrilogénicas. Uno de los datos más sorprendentes fue que, a pesar de que las mutaciones modificaron sustancialmente las propiedades termodinámicas y fibrilogénicas de 6aJL2, el traslape de las estructuras cristalográficas de varias mutantes sencillas contra 6aJL2, revela una alta similitud entre las cadenas principales (Hernández-Santoyo *et al.*, 2010; Maya-Martínez *et al.*, 2015). Una visión más global considerando un efecto compensatorio producto de las interacciones de las cadenas laterales, podría indicar cuáles segmentos de estas proteínas son los más afectados.

Análisis de redes como método de análisis

Así como el poema introductorio de la obra de Ernest Hemingway “Por Quién Doblan Las Campanas” que describe que nadie es una isla y que cada hombre es parte de un total, una de las maneras en las que se han estudiado las estructuras proteicas ha sido el considerarlas como una red interconectada donde cada residuo es considerado como un nodo y la relación entre cada par de residuos, ya sea a través de un enlace covalente o no covalente, puede ser considerado como un vértice (Watts & Strogatz, 1998; Vendruscolo *et al.*, 2002; Muppirala *et al.*, 2006). Otro tipo de vértice puede ser definido a partir de la distancia atómica que hay entre cada par de nodos; y uso de las distancias más cercanas ha permitido generar mapas de contactos. (Mirny *et al.*, 1996). En general se ha demostrado que modificaciones en los nodos altamente conectados causan cambios en las propiedades de la red en su conjunto. Por otro lado, estos análisis han mostrado que las redes tienen una mayor tolerancia a la perturbación en los nodos de bajo grado de conectividad (Albert *et al.*, 2000). Estas observaciones permiten suponer que aplicando un análisis de redes a una estructura proteica permitiría identificar residuos estructuralmente importantes.

La principal motivación para aplicar el análisis de redes a las estructuras correspondientes a las $V_L \lambda 6a$ es el identificar mutaciones aparentemente insignificantes pero, que en su conjunto, resultan relevantes especialmente si se altera la conectividad en los nodos importantes. Cabe señalar que el alineamiento de secuencias de V_L no mostró un patrón distintivo entre aquellas que son amiloidogénicas de aquellas que no lo son (Hurle *et al.*, 1994). Asimismo, el traslape de carbonos α entre las diferentes estructuras tampoco mostró algún cambio conformacional evidente (Hernández-Santoyo *et al.*, 2010). Mediante un análisis de redes aplicado a la estructura de 6aJL2 y de otras cadenas ligeras derivadas de la misma línea germinal, se podrían identificar los residuos determinantes de la estabilidad estructural correspondientes a las $V_L \lambda 6a$ además de los ya descritos para las inmunoglobulinas como el Trp35 y el puente disulfuro Cys23-Cys88 (Chothia *et al.*, 1998).

Para este trabajo, se optó por comparar estructuras cristalográficas y así identificar las diferencias que resultasen determinantes. Se seleccionó la cadena ligera AR para ser comparada contra 6aJL2 debido a que esta variante resultó ser la más inestable y la más fibrillogénica de todo el conjunto (González-Andrade *et al.*, 2013). AR tiene una característica particular en su secuencia de aminoácidos y es la presencia de una Gly en la posición 25 haciéndola similar a una mutante de 6aJL2 llamada 6a-R25G (Figura 5). La importancia de esta última variante radica en que hay indicios de que esta secuencia pudiera ser una variante alotípica de 6aJL2 donde una Gly está presente en 25% de las secuencias y en el otro 75%, una Arg ocupa esta posición (del Pozo *et al.*, 2008). Adicionalmente, los experimentos mostraron que la sustitución Arg25Gly disminuyó significativamente la estabilidad del dominio 6aJL2.

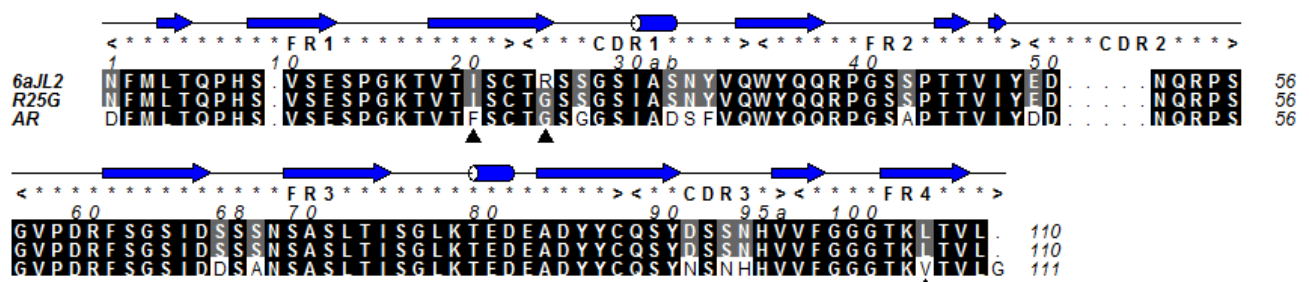


Figura 5. Alineamiento de secuencias de cadenas ligeras $\lambda 6$. Los aminoácidos están coloreados acorde a su conservación relativa usando el programa ALINE (Bond & Schüttelkopf, 2009) y numerados acorde a Kabat (Kabat *et al.*, 1991). La proteína 6aJL2 es comparada contra la mutante sencilla 6a-R25G y con la cadena ligera derivada de paciente AR. Las regiones de los Frameworks (FR) y de las Regiones Complementarias Determinantes (CDR) están marcadas.

En la primera parte de este trabajo se planteó la generación y el análisis de la estructura cristalográfica de AR y de 6a-R25G para su comparación con 6aJL2. Aunque se emplearon varias técnicas que consideran las estructuras de proteína como una red global de interacciones, se optó por generar y validar una propuesta metodológica para jerarquizar cada residuo acorde a su relevancia estructural (Luna-Martínez *et al.*, 2016). Se aplicó esta metodología para identificar las diferencias y similitudes en las estructuras de estas dos cadenas ligeras. En la segunda parte de este trabajo se retomó el análisis termodinámico de AR conducente a la generación de mutantes específicas, además de ampliar el análisis estructural entre 6aJL2, 6a-R25G y AR. Adelantando brevemente una de las cuestiones más importantes de esta parte, fue que se rechazó una hipótesis muy arraigada con respecto a las regiones variables de las cadenas ligeras de tipo λVI .

Sin más por el momento, y habiendo establecido el marco teórico de referencia así como los antecedentes directos de este proyecto, se dará pie a describir brevemente el análisis termodinámico y estructural para entender las propiedades biofísicas y fibrillogénicas de la región variable de las cadenas ligeras tipo λVI .

HIPÓTESIS

Los residuos presentes en el dominio variable de cadenas ligeras correspondiente a la línea germinal 6aJL2 que mantienen una alta conectividad, pueden ser identificados de manera teórica siendo jerarquizables y diferentes con respecto a los residuos de la variante AR.

OBJETIVOS

General

- Analizar las diferencias estructurales entre las estructuras cristalográficas de 6aJL2 y AR para entender las propiedades biofísicas y fibrilogénicas de estos dominios variables de cadenas ligeras.

Específicos

- Determinar la estructura cristalográfica de AR y de 6a-R25G para contrastarla con la estructura de 6aJL2.
- Analizar la estructura cristalográfica de 6aJL2 mediante análisis de redes para identificar los residuos clave involucrados en la estabilización de este dominio variable de cadenas ligeras.
- Proponer y validar una metodología de jerarquización de residuos empleando proteínas evaluadas termodinámicamente con diferente estructura terciaria.
- Evaluar las estructuras de AR y 6aJL2 con la metodología propuesta.

METODOLOGÍA

Estadística

El análisis de componentes principales (PCA) fue realizado usando el programa R (R Core Development Team, 2013). Los análisis termodinámicos fueron realizados por triplicado.

Estructuras cristalográficas para la validación

El código PDB, el tamaño, y la clase de proteína de las estructuras seleccionadas son las siguientes: inhibidor de quimiotripsina (2CI2; 65 residuos; 16% α + 21% β); 6aJL2 (2W0K; chain A, 111 residuos; 5% α + 46% β); apoflavodoxina (1FTG; 168 residuos; 35% α + 19% β); represor dimérico Arc (1ARR; 108 residuos; 62% α + 9% β); dominio de unión al DNA del receptor de estrógeno humano α en complejo con su DNA de reconocimiento (1HCQ; cadenas A, B; 128 residuos; 26% α + 9% β , y cadenas C y D, 56 nucleótidos); proteína de choque térmico de *Bacillus subtilis* (1CSP; 67 residuos; 4% α + 55% β); proteína de choque térmico de *B. caldolyticus* (1C9O, 66 residuos; 4% α + 62% β); y el segmento N-terminal intrínsecamente desordenado del supresor de señalización de citosina 5 de mamífero. (2N34, 70 residuos; 12% α). Se realizaron modificaciones estructurales menores en los archivos del PDB como el relleno de los extremos N-terminal y de las cadenas laterales usando el programa Swiss PDB (Guex & Peitsch, 1997). Además, todos los nodos fueron reenumerados de manera continua para evitar repeticiones; por ejemplo, en el PDB 2CI2, la proteína inicia con el residuo 19. Consecuentemente, todos los residuos fueron sistemáticamente reenumerados para que el primer residuo fuera identificado como el residuo 1 y así sucesivamente. El área expuesta al disolvente fue calculada con el programa NOC (Chen, Cang & Nymeyer, 2007).

Análisis de redes

El análisis de redes se basa en un parámetro fundamental que es la existencia de una relación entre diversos nodos. Para este trabajo se harán las siguientes consideraciones generales para la relación entre nodos (Bocaletti *et al.*, 2006).

- 1) **No dirigidas.** Las uniones no muestran una dirección implicando que cualquier relación representada será simétrica.
- 2) **Jerarquizadas.** Las uniones entre nodos tienen un valor.
- 3) **Simples.** Las uniones múltiples entre el mismo par de vértices no están permitidas.
- 4) **Dispersas.** Para una red no dirigida, el número máximo de uniones (M) es igual a la combinatoria del número de nodos $n(n-1)/2$. Por lo que una red dispersa tendrá una M mucho menor a $n(n-1)/2$.
- 5) **Conectadas.** Cualquier nodo puede relacionarse con otro nodo lejano mediante el trazo de una ruta con un número finito de nodos. No existen nodos aislados.

Grado

Los residuos pueden ser considerados como nodos y los contactos como vértices. Por lo tanto, un vértice puede ser definido cuando dos átomos, no Hidrógenos, de un par de residuos están dentro de una distancia de 5 Å (Li *et al.*, 2008). Para estudiar la topología de una red de contactos entre residuos, se midió el grado del nodo-*i*, K_i , como el número de vecinos del nodo-*i*. Se seleccionaron las cadenas A de las estructuras cristalográficas. En el caso de las estructuras cuaternarias, toda la estructura fue considerada para medir el grado de cada nodo pero únicamente se consideró para las comparaciones sucesivas.

Eigenvectores

Otra medida de centralidad más sofisticada son los eigenvectores. En general, un nodo conectado con nodos reconocidos como influyentes atribuirá un mayor peso a su centralidad que si estuviera conectado con nodos poco influyentes (Newman, 2008). Por lo tanto, la centralidad por eigenvectores puede ser también vista como la suma jerarquizada no solamente de las conexiones directas, sino también las conexiones indirectas. Matemáticamente, una red puede ser representada por una matriz llamada matriz de adyacencia A. Ésta es una matriz hermitiana simétrica $n \times n$ donde n es el número de nodos de la red y los elementos pueden ser las distancias mínimas entre cada residuo (Piziak & Odell, 2006). Dícese que una matriz es hermitiana cuando el elemento a_{ij} es igual al elemento a_{ji} de la matriz transpuesta conjugada. Dicho de otra manera, si \mathbf{A} tiene números imaginarios, $\bar{\mathbf{A}}$ es la matriz con el cambio de signo de los números imaginarios de la matriz \mathbf{A} ; y $\bar{\mathbf{A}}^*$ es la matriz transpuesta de $\bar{\mathbf{A}}$ (Figura 6).

$$\mathbf{A} = \begin{bmatrix} 3 & 2 + i \\ 2 - i & 1 \end{bmatrix}$$

Figura 6. Matriz Hermitiana. La matriz transpuesta conjugada de A es igual a la misma matriz.

Formalmente, si \mathbf{A} es una transformación lineal, un vector no-cero \mathbf{x} multiplicado a la derecha de la matriz \mathbf{A} es entonces el eigenvector derecho de A si existe un valor escalar λ que cumpla con la siguiente ecuación:

$$\mathbf{Ax} = \lambda\mathbf{x} \quad (\text{Ecuación 1})$$

Se dice que el escalar λ es el eigenvalor de A correspondiente al eigenvector \mathbf{x} . El número total de eigenvalores y, por lo tanto, de eigenvectores (excluyendo al vector cero), es igual al número de nodos n por ser \mathbf{A} una matriz simétrica. El conjunto de todos los eigenvalores posibles de \mathbf{A} es llamado espectro de \mathbf{A} . La elección de cuál eigenvalor, y su respectivo eigenvector, como factor de centralidad está sustentada en el teorema de Perron-Frobenius el cual afirma que una matriz cuadrada real (sin elementos con números imaginarios) y con todos sus elementos positivos, tiene un eigenvalor cuyo valor es real y es el mayor de todo el espectro de eigenvalores (Horn & Johnson, 1990; Piziak & Odell, 2006; Newman,

2008). De lo anterior, el eigenvector correspondiente a este eigenvalor tendrá estrictamente componentes positivos los cuales son la medida de la importancia y calidad del nodo. Además del teorema de Perron-Frobenius, otros trabajos han demostrado que la principal clasificación de los nodos por eigenvectores está mediada por el uso del mayor eigenvalor (Bonacich, 1972).

Para redes jerarquizadas, el valor de la unión podría interpretarse como la fuerza de la misma; a mayor valor, mayor fuerza (Newman, 2004). Lo anterior no modifica lo dicho para los eigenvectores como parámetro de clasificación de los nodos pero modificaría la forma de parametrizar las matrices de distancias entre residuos. El artilugio matemático para modificar las matrices de distancias sin afectar la relación es calculando el inverso del valor de la distancia. De esta manera, entre más cercana sea la interacción, mayor será su peso en el cálculo por eigenvectores.

Entropía de redes

En el trabajo de Manke & Demetrius, 2004, proponen una aproximación sistemática al análisis de redes basándose en otras teorías como mecánica estadística, teoría ergódica y teoría de la información. Ellos proponen que los procesos biológicos operan típicamente en un estado estacionario, donde las características macroscópicas observables permanecen constantes por periodos largos de tiempo. Esto no implica que los procesos microscópicos sean estáticos sino que su interacción compleja y continua resulta en un fenotipo estable que puede ser observado experimentalmente. En sus trabajos (Manke *et al.*, 2005; Manke *et al.*, 2006) usan datos provenientes de *knockouts* de genes y experimentos de RNA de interferencia de *Saccharomyces cerevisiae* y de *Caenorhabditis elegans*. El análisis muestra que los *knockouts* de proteínas con gran contribución a la entropía de redes son los más letales. Las matrices reportadas se basan en una matriz estocástica derecha donde la suma de las filas de la matriz cuadrada real tiene el valor de 1 (Figura 7).

$$P = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{1n} \\ p_{21} & p_{22} & \dots & p_{2n} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{bmatrix} \quad \text{donde} \quad \begin{array}{l} \sum_{i=1}^n p_{1i} = 1 \\ \sum_{i=1}^n p_{2i} = 1 \\ \dots \\ \sum_{i=1}^n p_{ni} = 1 \end{array}$$

Figura 7. Matriz estocástica. La sumatoria de uno de los elementos de cada fila es igual a 1.

Se escogieron las matrices de adyacencia del inverso de las distancias de 5 Å de las estructuras. La razón por la cual se escogió esta conversión matemática es que, de manera similar a los eigenvectores, entre mayor sea la cercanía entre los residuos mayor será el impacto en caso de un cambio de la interacción entre estos. Se procedió a analizar esta matriz

estocástica usando la ecuación de la entropía dinámica de Shannon (Ecuación 2). La entropía dinámica de un proceso de Markov caracteriza la diversidad de caminos posibles y está relacionado con la respuesta del sistema a las perturbaciones (Horn & Johnson, 1990). El parámetro π_i es un vector estacionario de probabilidad definido como un vector que no cambia bajo la aplicación de una matriz estocástica. También se define como el eigenvector izquierdo de la matriz estocástica de probabilidad.

$$H = -\sum_{i,j} \pi_i p_{ij} \log p_{ij} \quad (\text{Ecuación 2})$$

El primer dato necesario para el cálculo de entropía es la generación de una matriz estocástica a partir de la matriz de adyacencia del inverso de las distancias. Se suman los valores de las filas y cada valor de la fila se divide entre esta suma. Ahora la matriz del inverso de las distancias es una matriz estocástica. Con el programa Matcad 15TM se calcularon los eigenvectores izquierdos correspondientes y se escogió aquel cuyo valor procedía del mayor eigenvalor. El vector generado es el vector π cuyo valor se multiplica al valor de la entropía de Shannon

Dinámica molecular

Las estructuras cristalográficas fueron preparadas usando el paquete *Protein Preparation Wizard* incluido en el programa Maestro 9.2 (Schrödinger LCC, NY, USA), y se depositaron en una caja de agua de 10 Å las cuales contenían 14513 moléculas de agua de Punto de Carga Sencilla (SPC, por sus siglas en inglés) para 2CI2, 16667 para 2W0K, 20051 para 1FTG, 20990 para 1ARR, 28142 para 1HCQ, 12219 para 1CSP, y 12033 para 1C9O. Se añadieron iones neutralizantes, y otros iones metálicos previamente presentes en la estructura cristalográfica fueron dejados en el mismo lugar. Los cálculos del alineamiento y el análisis de datos fueron hechos con los programas Desmond y Maestro, respectivamente (Maestro-Desmond Interoperability Tools, versión 3.0; Schrödinger, NY, USA). El campo de fuerza OPLS 2005 fue usado para cada dinámica molecular. Las condiciones periódicas cúbicas resultantes para cada proteína fueron (53.9 x 53.9 x 53.9 Å³ para 2CI2; 56.8 x 56.8 x 56.8 Å³ para 2W0K; 60.9 x 60.9 x 60.9 Å³ para 1FTG; 61.2 x 61.2 x 61.2 Å³ para 1ARR; 51.1 x 51.1 x 51.1 Å³ para 1CSP; y 50.8 x 50.8 x 50.8 para 1C9O); dado el tamaño del complejo 1HCQ, se utilizaron fronteras rectangulares (58.5 x 59 x 91 Å³). Cada simulación fue ajustada con un ensamble isotérmico-isobárico (NPT, donde se mantiene la cantidad de sustancia, presión, y temperatura) unido débilmente a un baño térmico a presión constante de 1 atm y un tiempo de relajación de 2 ps, regulado por el baróstato Berendsen (Berensen *et al.*, 1984). Todas las interacciones fueron seguidas usando una distancia de corte de 9 Å, y para distancias de larga distancia (electrostáticas y de van der Waals), se utilizó el método de red de partícula suavizada de Ewald con una tolerancia de 1×10^{-9} (Essman *et al.*, 1995). Para asegurarse que las simulaciones

iniciarán de un mínimo local, se aplicó un algoritmo de alineamiento simultáneo. Este método inició con la simulación a una alta temperatura (400 K) para evitar barreras termodinámicas y conformacionales, seguido de un enfriamiento gradual (alineamiento) para alcanzar regímenes de baja energía. Éste es un método ampliamente usado para la optimización de estructuras provenientes de métodos experimentales, modelado comparativo de proteínas, o el estudio de plegamientos o desnaturalizaciones mediante dinámicas conformacionales de proteínas o péptidos (Mori & Okamoto, 2009). El sistema completo fue calentado a 10 K por 30 ps, 100 K por 100 ps, 300 K por 200 ps, 400 K por 300 ps, 400 K por 500 ps, y luego enfriado a 298 K por 1000 ps. Para asegurarse de que el calentado a 400 K no afectara la estructura de la proteína o proteína/ADN, se comparó el *Root Mean Square Deviation* (RMSD) de los átomos pesados (C, N, O, S, P) provenientes de la comparación entre la estructura del alineamiento y su correspondiente estructura cristalográfica. Si la desviación estándar del RMSD fuese $\leq 1.0 \text{ \AA}$ se supuso que el alineamiento no modificó a la estructura o la desnaturalizó. De hecho, dado que el desplazamiento de los átomos pesados es mayor a 1 \AA se consideró como un cambio conformacional y que se encontró un mínimo local. El paso de la interpolación lineal entre dos puntos adyacentes de tiempo fue empleado. Después del sexto paso, se simularon 25 ns con un tiempo de integración de 1 fs. El análisis total fue realizado usando las coordenadas de la trayectoria y las energías registradas cada 1.2 ps. Un modelo fue extraído cada 0.25 ns de simulación, y el total de 100 modelos fue guardado como un archivo tipo PDB.

Entropía dinámica de la dinámica molecular

Se generó un algoritmo codificado en lenguaje Perl que utiliza al archivo PDB de los 100 modelos. Como primer paso, el primer archivo de cada modelo contiene las coordenadas atómicas, sus correspondientes pesos moleculares, y los nombres de cada nodo (aminoácido o base nucleotídica). Para el segundo paso se calcularon las coordenadas del centro de masa del nodo. En el tercer paso se calculó la distancia entre los centros de masa de cada par de nodos y se normalizaron con la siguiente Ecuación (3)

$$\overline{d}_{ij} = \frac{d_{ij}}{r_{vdW_i} + r_{vdW_j}} \quad \text{Ecuación (3)}$$

donde \overline{d}_{ij} es la distancia normalizada, d_{ij} es la distancia entre los centros de masas del nodo i y el nodo j , y r_{vdW} es el radio de van der Waals del respectivo nodo considerando todos sus átomos. Para los aminoácidos, los valores del radio de van der Waals comprendiendo todo el residuo fueron obtenidos de Darby & Creighton, 1993. Para las bases de nucleótidos, estos valores fueron obtenidos de Voss & Gerstein, 2005. Se calculó la distancia promedio entre cada par de nodos de los 100 modelos extraídos. Dado que los siguientes pasos involucran las propiedades de los eigenvectores y eigenvalores, se calculó el valor inverso de la distancia

promedio normalizada, para que así los valores más pequeños, distancias muy cercanas, tuvieran mayor peso.

Se construyó una matriz con estos valores donde $A = (a_{ij}) \geq 0$ de tamaño $N \times N$, donde N es el número de nodos. En este caso, la matriz A es simétrica ($a_{ij} = a_{ji}$) y no dirigida. Siguiendo la estrategia matemática descrita por Demetrius & Manke, 2005, se supuso que el proceso estocástico está dado por una matriz de Markov $P = p_{ij}$ donde $p_{ij} \geq 0$ y $\sum_j p_{ij} = 1$. La distribución estacionaria de la matriz P está descrita por la Ecuación (4),

$$\pi P = \pi \quad \text{(Ecuación 4)}$$

donde π está definido como el eigenvector izquierdo asociado con el eigenvalor de mayor valor resuelto con el programa Mathcad 15TM. La entropía dinámica de este proceso para cada nodo, H_i , está descrito por la Ecuación (5),

$$H_i = -\sum_j \pi_{ij} p_{ij} \log p_{ij} \quad \text{(Ecuación 5)}$$

donde el término $p_{ij} \log p_{ij}$ es la entropía de Shannon.

Mutantes de 6aJL2 y de AR

Tanto la síntesis de mutantes de AR como de las mutantes sencillas de 6aJL2 (6a-R25H, 6a-I30G, 6a-Y36F, y 6a-Q6N) fue realizada usando PCR recursivo (Prodromou & Pearl, 1992). El gen obtenido fue clonado en el vector pSyn 1 (Schier *et al.*, 1995). Todas las contrucciones fueron verificadas usando secuenciación de nucleótidos (Sanger, Nicklen & Coulson, 1977). Todas las variantes fueron expresadas en *Escherichia coli* BL21 (DE3) y purificadas como se describió previamente (Luna-Martínez, 2009). La proteína purificada fue verificada mediante electroforesis SDS-PAGE, y la concentración de la proteína fue determinada espectrofotométricamente a 280 nm en buffer 6.5 M GndHCl en buffer de fosfato de sodio 20 mM, pH 7.5, usando coeficientes de extinción molar calculados con el programa ProtParam, del servidor ExPASy (Gasteiger *et al.*, 2005).

Desnaturalización térmica

Las muestras conteniendo 50 $\mu\text{g/ml}$ de proteína en solución salina de fosfatos (PBS), pH 7.5, fueron colocadas en una celda de 3 ml. Los cambios de la fluorescencia del triptofano fueron medidas usando un espectrofluorímetro LS50B Perkin Elmer con una longitud de onda de excitación de 295 nm (2.5 mm de apertura) y una longitud de onda de emisión de 355 nm (5 mm de apertura). La solución fue calentada usando un baño de agua con un sistema de recirculación conectada al espectrofluorímetro y la temperatura de la solución fue medida con un termistor. Los datos de desnaturalización térmicos fueron analizados como se describió en

Pace *et al.*, 1989, suponiendo un proceso de dos estados y siguiendo la ecuación de van't Hoff (Ecuación 6).

$$\frac{d(\ln K_{app})}{d(1/T)} = -\frac{\Delta H_{app}}{R} \quad (\text{Ecuación 6})$$

donde T es la temperatura, ΔH es la entalpía, y R es la constante de los gases. El valor de ΔG a temperatura ambiente fue determinada usando la ecuación 7,

$$\Delta G(T) = \Delta H_m \left(1 - \frac{T}{T_m}\right) - \Delta C_p \left[(T_m - T) + T \ln\left(\frac{T}{T_m}\right) \right] \quad (\text{Ecuación 7})$$

donde T está a 298 K, y C_p es el cambio en la capacidad calórica asociada con el proceso de desnaturalización, calculado teóricamente con el método descrito por Milardi *et al.*, 1997.

Dicroísmo circular

Los espectros de dicroísmo circular y los experimentos de desnaturalización térmica fueron registrados en un espectropolarímetro JASCO J-715 (JASCO INC., Easton, MD) equipado con un peltier enfriado con agua. Los espectros de UV lejano fueron registrados usando una celda de cuarzo de 0.1 cm de longitud de paso con una concentración de proteína a 400 $\mu\text{g/ml}$ en solución de 50 mM Na_2HPO_4 , 100 mM NaCl, pH 8.0.

Los datos crudos fueron convertidos a elipticidad molar usando la ecuación 8,

$$[\theta] = 100 * \theta / (C * l) \quad (\text{Ecuación 8})$$

donde C es la concentración molar, e l es la longitud de paso de la celda en cm.

Cristalización y recolección de datos

Previo a las pruebas de cristalización, las proteínas de AR y de 6a-R25G fueron mantenidas en 40 mM Tris, 20 mM NaCl, pH 8.2. Ambas proteínas fueron cristalizadas usando la técnica de difusión de vapor con gota colgante a 18 °C. Para 6a-R25G, los mejores resultados fueron obtenidos a partir de gotas compuestas por mezclas de volúmenes iguales de 7 mg/ml de solución de proteína en 100 mM Mes, pH 6.5, con 0.1 M de cacodilato de sodio y 1.4 M de acetato de sodio como precipitantes. Para AR, se usó el mismo agente precipitante pero la proteína fue diluida en TrisHCl 100 mM, pH 7.5. Los datos de difracción fueron colectados con un ánodo rotatorio Rigaku MicroMax-007 HF ($\text{Cu } \kappa \alpha$, $\lambda = 1.5416$) a 100 K en un detector de imagen de placa R-AXIS IV++. La matriz incluyó glicerol al 30% como crioprotector. Los datos fueron indexados e integrados usando XDS (Kabsch, 2010), y las reflexiones fueron combinadas y escaladas usando el programa Scala en CCP4 (Collaborative Computational Project 4, 1994). Los grupos espaciales correctos fueron determinados con Pointless (McCoy, 2006).

Determinación estructural y refinamiento

Las estructuras 3D fueron resueltas usando reemplazo molecular con el programa Phaser (McCoy *et al.*, 2007) en Phenix (Adams *et al.*, 2010), usando las coordenadas atómicas de WIL (PDB 2CD0) como modelo. La inspección de los mapas y la construcción de los modelos fueron hechas en COOT (Emsley & Cowtan, 2004), y los modelos resultantes fueron refinados con phenix.refine (Afonine *et al.*, 2012). Las moléculas de agua fueron adicionadas al modelo al final del refinamiento mediante una búsqueda basada en los picos observados en las diferencias de los mapas con un criterio basado en la distancia. MolProbity fue usado para el análisis de la estereoquímica del modelo y su validación (Chen *et al.*, 2010). La presencia de la mutación fue confirmada a través de las diferencias de los mapas de Fourier usando la estructura nativa como base. Los análisis de la interfaz del dímero fueron hechos usando el servidor PISA (Krissinel & Henrick, 2007), y los alineamientos con 3d-SS (Sumathi *et al.*, 2006). Los modelos finales fueron dibujados con Chimera (Dundas *et al.*, 2006), y las interacciones residuo-residuo fueron llevados a cabo con el servidor PIC (Tina *et al.*, 2007). Las coordenadas atómicas y los factores estructurales de los modelos finales fueron depositadas en el Protein Data Bank, con los números de acceso 5IR3 para AR y 5C9K para 6a-R25G.

RESULTADOS

Parte I. Diferencias y similitudes estructurales de las cadenas ligeras λ VI

Determinación y comparación de las estructuras cristalográficas de AR y de 6a-R25G contra la estructura de 6aJL2

En colaboración con la Dra. Adela Rodríguez Romero y de la Dra. Alejandra Hernández Santoyo del Instituto de Química de la UNAM, se generaron las estructuras cristalográficas de AR y de 6a-R25G. Los datos cristalográficos y la estadística de refinamiento para AR y 6a-R25G se muestran en la Tabla 1. Ambos cristales difractaron a una resolución de 1.7 Å y 1.92 Å respectivamente.

Tabla 1. Colección de datos y estadística del refinamiento estructural.

Proteína (PDB)	AR (5IR3)	6a-R25G (5C9K)
Colección de datos		
Longitud de onda (Å)	1.54	1.54
Temperatura (°K)	100	100
Grupo espacial	I4 ₁ 22	C ₁ 2 ₁
Dimensiones de la celda <i>a, b, c</i> (Å)	a=71.11, b=71.11, c=96.23	a=104.56, b=95.63, c=102.58
Ángulos de la celda	$\alpha=\beta=\gamma=90$	$\alpha=90.0, \beta=96.0, \gamma=90.0$
Coefficiente de Matthews (Å ³ Da ⁻¹)	2.43	2.69
Límite de resolución (Å)	34.81-1.70	34.65-1.92
Reflexiones únicas ^a	13829(1338)	76562(11143)
R _{merge} ^b	0.089 (0.486)	0.073 (0.460)
I/ σ I	11.3 (3.3)	13.1 (3.2)
CC _{1/2} (%)	99.6 (87.4)	99.7(80.0)
Integridad (%)	99.4 (99.4)	100.0 (100.0)
Multiplicidad	4.3 (4.3)	2.9 (2.8)
Wilson B Factor (Å ²)	12.20	19.56
Estadística de refinamiento		
R _{work} /R _{free} ^c (%)	17.6/20.7	17.3/21.9
No. de átomos		
Proteína/solvente/acetato	1048/157/1	6687/1010/4
B-value promedio (Å ²)	18.05	23.0
RMSD de idealidad		
Longitud de enlace (Å)	0.006	0.011
Ángulo de enlace (°)	0.971	1.27
Gráfica de Ramachandran (%)		
Regiones permitidas	97.00	96.31
Regiones no permitidas	3.4	3.69
Regiones atípicas	0.0	0.0

^a Valores en paréntesis corresponden a la última capa de resolución.

^b $R_{\text{merge}} = \sum_b \sum_l (|I_{j,b} - \langle I_b \rangle|) / \sum_b \sum_l \langle I_b \rangle$, donde *b* es el índice único de reflexión, $I_{j,b}$ es la intensidad de la simetría relacionada con la reflexión, e $\langle I_b \rangle$ es la intensidad promedio.

^c $R = \sum_b ||F_o|_b - |F_c|_b| / \sum_b |F_o|_b$ para todas las reflexiones, donde F_o y F_c son los factores estructurales observados y calculados, respectivamente, y *b* define las reflexiones únicas. R_{free} está calculado análogamente para examinar las reflexiones, seleccionadas aleatoriamente, y excluidas del refinamiento.

El reemplazo molecular ejecutado con el programa Phaser, incluido en el programa Phenix (Adams *et al.*, 2002), reportó un reemplazo exitoso acorde a sus respectivos valores LLG y TFZ (Tabla 1). El patrón de difracción mostró que los cristales de AR pertenecían al grupo espacial tetragonal $I4_122$, mientras que los cristales de 6a-R25G pertenecían al grupo espacial monoclinico C_12_1 . El mapa final de densidad electrónica 2Fo-Fc muestra una buena calidad y una buena correlación en las regiones determinantes de complementariedad (CDR) y las regiones marco (FR), por lo que ambas estructuras conservan la estructura canónica de una V_L , es decir, un dominio tipo inmunoglobulina (β sándwich).

Las unidades asimétricas mostraron que AR difractó como un monómero mientras que 6a-R25G difractó como un octámero, los cuales se analizan en la página 49, a diferencia de la estructura dimérica reportada de 6aJL2 (PDB 2W0K). Los coeficientes de Matthews, definido como el volumen del cristal por unidad de masa molecular de proteína, estuvieron en el rango descrito para proteínas globulares del mismo peso molecular (Matthews, 1968; Kantardjieff & Rupp, 2003). De manera sorprendente, el traslape del monómero de AR contra los monómeros A de 6aJL2 y de 6a-R25G mostraron una alta similitud con diferencias de tan solo 0.45 Å y 0.35 Å, respectivamente (Figura 8A). Sin embargo, un efecto notorio fue la pérdida de interacciones generadas por la presencia de la Arg25 de 6aJL2 como la interacción catión- π formada con la Phe2 y varios puentes de hidrógeno a su alrededor (Figura 8B). En la ausencia de la Arg25, se observó que Phe2 se reorientó hacia el interior del núcleo hidrofóbico superior tanto en 6a-R25G como en AR (Figura 8C).

Como la comparación entre cadenas principales no evidenció cambios estructurales dramáticos pero se detectó una ligera rotación en un residuo, se procedió a analizar el efecto de la reorientación de las cadenas laterales y su impacto en la modificación en las interacciones intramoleculares. Se usó el servidor PIC (Tina *et al.*, 2007) para ubicar estas interacciones (hidrofóbicas, puentes salinos, apilamiento de anillos aromáticos, e interacciones catión- π), y en la Tabla 2 se muestran aquellas que resultaron diferentes entre 6aJL2, 6a-R25G, y AR. Para el caso de los puentes de hidrógeno, se seleccionaron los más relevantes.

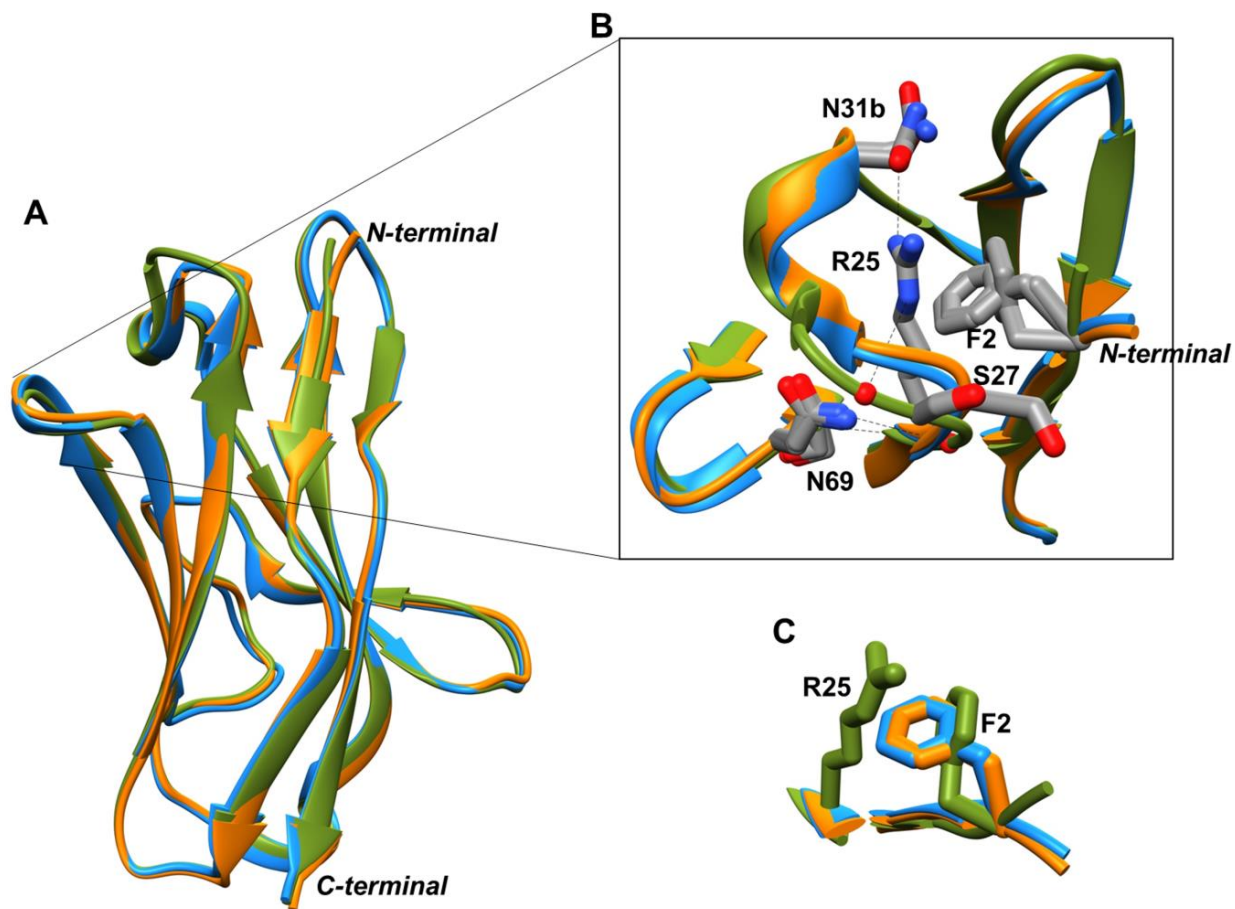


Figure 8. Traslape de estructuras. Se muestra el traslape entre las cadenas principales de 6aJL2 (verde) sobre su mutante sencilla 6a-R25G (azul) y sobre AR (naranja). **A.** Traslape de AR y los monómeros A de 6aJL2, y 6a-R25G. **B.** Las interacciones alrededor de Arg25 en 6aJL2 abarcan los segmentos del CDR1, el extremo N-terminal, y el asa correspondiente al residuo 68. **C.** La mutación Arg25Gly indujo la rotación de Phe2 así como lo muestran las estructuras de 6a-R25G y AR.

Tabla 2. Interacciones intramoleculares. Se describen las interacciones intramoleculares hidrofóbicas, puentes de hidrógeno, iónicas, aromático-aromático, cation- π diferentes entre 6aJL2, 6a-R25G, y AR determinadas con el servidor Protein Interaction Server – PIC usando las distancias interatómicas predefinidas (Tina *et al.*, 2007). En guiones se muestran las interacciones ausentes con respecto a la otra proteína.

Interacciones Hidrofóbicas		
6aJL2	6a-R25G	AR
-	PHE2-ILE30	PHE2-ILE30
-	PRO7-VAL11	-
PRO15-LEU107	-	PRO15-LEU107
-	ILE21-LEU104	PHE21-VAL104
-	ILE21-ILE75	PHE21-ILE75
ALA31-ILE66	-	-
TRP35-TYR86	TRP35-TYR86	-
-	-	ALA43-PRO44
ILE48-PHE62	-	-
VAL106-LEU107	-	-
Puentes de Hidrógeno		
6aJL2	6a-R25G	AR
-	-	GLN6(OE1)-CYS23(SG)
-	GLN6(OE1)-CYS88(SG)	GLN6(OE1)-CYS88(SG)
ARG25(NE)-SER27(O)	-	-
ARG25(NH1)-SER29(O)	-	-
ARG25(NH1)-ASN31b(OD1)	-	-
SER27(N)-ARG25(O)	-	-
SER68a(OG)-ASP67(OD2)	-	-
-	SER68a(N)-ASP67(OD1)	SER68a(N)-ASP67(OD1)
-	SER68b(N)-ASP67(O)	-
SER68b(OG)-ASP67(OD2)	SER68b(OG)-ASP67(OD2)	-
SER68b(OG)-SER70(OG)	SER68b(OG)-SER70(OG)	-
ASN69(ND2)-ARG25(O)	ASN69(ND2)-GLY25(O)	ASN69(ND2)-GLY25(O)
-	SER70(N)-SER68b(O)	SER70(N)-ALA68b(O)
-	SER70(OG)-SER68b(OG)	SER70(OG)-SER68b(OG)
-	GLY99(N)-CYS88(SG)	GLY99(N)-CYS88(SG)
GLY100(N)-GLN6(OE1)	-	-
-	THR102(N)-GLN6(NE2)	THR102(N)-GLN6(NE2)
LEU104(N)-THR102(O)	-	VAL104(N)-THR102(O)
Interacciones iónicas		
6aJL2	6a-R25G	AR
-	GLU13-LYS17	GLU13-LYS17
-	ARG39-GLU81	ARG39-GLU81
ASP92-HIS95a	ASP92-HIS95a	-
Interacciones Anillo aromático – Anillo aromático		
6aJL2	6a-R25G	AR
-	-	PHE21-TRP35
-	-	PHE21-TYR86
Interacciones Cation - π		
6aJL2	6a-R25G	AR
PHE2-ARG25	-	-

Fue interesante observar el impacto que tuvo una mutación sencilla, 6a-R25G, ya que la modificación en las interacciones no fue sólo en el segmento correspondiente sino en todo el dominio generando un efecto a distancia. Por otro lado, en AR, el residuo Phe21 está interactuando hidrofóbicamente con Ile30, Trp35, Ile75, y Tyr86, los cuales en conjunto forman parte del núcleo hidrofóbico característico de estos dominios variables de inmunoglobulina. Sin embargo, usando esta estrategia, un inconveniente muy importante a enfatizar fue la imposibilidad para jerarquizar la relevancia estructural de cada residuo así como el impacto global de las mutaciones de AR y, sobre todo, el efecto global que una mutación sencilla ocasionó como se puede apreciar en 6a-R25G.

Considerando que AR es el extremo opuesto de 6aJL2, desde un punto de vista termodinámico y fibrilogénico, las siguientes partes del trabajo se centrarán en estas dos proteínas. Por lo tanto, se decidió emplear una visión general que abarcara los cambios de orientación de las cadenas laterales aun cuando el traslape entre las cadenas principales fuese similar. Para ello se optó por la estrategia de análisis de redes como metodología descriptiva usando a 6aJL2 como modelo de prueba.

Aplicación de análisis de redes sobre 6aJL2 para identificar los residuos clave involucrados en la estabilización de este dominio

En años recientes se ha desarrollado una nueva estrategia para estudiar las estructuras proteicas vistas como una red social. Esta técnica se basa en la teoría de gráficas cuyo enfoque es analizar las estructuras como si fueran figuras. Las maneras en que se puede transformar la estructura terciaria de una proteína en una gráfica de redes es considerando a cada residuo como un nodo y su relación con otro residuo es representada mediante una línea de unión entre cada vértice (Muppirala *et al.*, 2006). El análisis de redes permite identificar los residuos que presentan un mayor grado de conectividad en toda la molécula. Los residuos más interconectados potencialmente estarían direccionando el plegamiento favoreciendo la estabilización de la estructura nativa (Albert *et al.*, 2000). Las estructuras entonces tendrían nodos “maestros” cuya afectación modificaría sustancialmente las propiedades de toda la red ya que estos se encuentran altamente conectados. Asimismo, habría otros nodos cuyas modificaciones serían compensadas por el resto de la red dado su bajo grado de conectividad (Albert *et al.*, 2000). La aplicación del análisis de redes a una estructura proteica ha permitido localizar residuos estructuralmente importantes que se encuentran altamente conectados.

Dentro de la diversidad de parámetros para caracterizar una red, se usó Grado, Eigenectores, y Entropía de Redes para caracterizar primeramente la estructura de 6aJL2. Grado se definió como el número de vecinos que tiene un residuo dentro de un radio de 5 Å considerando cualquier átomo que no fuese Hidrógeno (Li *et al.*, 2008). La técnica de Eigenectores evalúa la jerarquía de un nodo incrementándola si éste se encuentra aledaño a uno que está altamente conectado a partir de una matriz de contactos (Costa *et al.*, 2007); por

lo que se consideraron todos los átomos que estuvieran a una distancia de 5 Å. La técnica de Entropía de Redes se apoya en la teoría de la información haciendo uso de la entropía de Shannon donde cada uno de los nodos es calificado con un valor probabilístico para afectar a la red (Demetrius & Manke, 2004). Para esta técnica se consideraron todas las distancias entre todos los átomos. En el apéndice A se encuentran enlistados los residuos acorde a estas tres técnicas. Se seleccionaron los quince residuos mejor y peor posicionados de 6aJL2 para cada técnica y en la Tabla 3 se muestra la convergencia de estos residuos entre estas tres técnicas. En la Figura 9 se muestra la ubicación de estos residuos en la estructura.

Tabla 3. Residuos de 6aJL2 mejor y peor calificados. Se muestra la convergencia entre los 15 mejores y los 15 peores residuos de 6aJL2 usando diferentes parámetros de análisis de redes. Se calculó el porcentaje de Área Expuesta al Disolvente (*Residue Solvent Accessibility*, RSA) con el programa NOC.

Convergencia Métodos	Residuos mejor calificados - (%RSA)	Residuos peor calificados - (%RSA)
Grado Eigenvectores Entropía de Redes	Trp35 - (0.6)	Ser42 - (100)
	Tyr86 - (0.0)	Leu107 - (60.8)
	Gln6 - (12.2)	Asn1 - (86.5)
	Ile48 - (0.0)	Pro40 - (87.1)
	Leu73 - (0.0)	Gly41 - (100)
	Cys23 - (0.5)	Ser56 - (100)
	Ile30 - (2.2)	Gly57 - (91.7)
	Ile21 - (0.7)	Ser68a - (100)
		Ser93 - (100)
		Ser94 - (100)

Se observó una buena relación en la valoración de los residuos entre estos tres parámetros mostrando al Trp35 como residuo principal y aquellos reportados como estructuralmente importantes en este tipo de dominios tipo β -sándwich (Chothia *et al.*, 1998). Los residuos ubicados en el asa 40-60, en el asa de los residuos 68, 68a, y 68b, así como en los CDRs, fueron calificados como de baja conectividad. Sin embargo, hubo una consideración que se hizo para seleccionar al parámetro más adecuado y fue la base teórica en las que se sustentan estos tres parámetros. El parámetro Grado es, principalmente, una aproximación de tipo geométrica, por lo que un residuo de mayor tamaño tiene mayor posibilidad de tener más vecinos que aquellos de menor tamaño como es la glicina. Aunque las estrategias de Eigenvectores y la Entropía de Redes usan una estrategia matemática similar para su resolución, el cálculo de eigenvectores a partir de una matriz de contactos, la Entropía de Redes describe el valor resultante del nodo con un valor probabilístico de afectar la red (Demetrius & Manke, 2004). Por lo tanto, acorde a la Entropía de Redes, a mayor valor entrópico del nodo una mayor probabilidad de afectar la red.

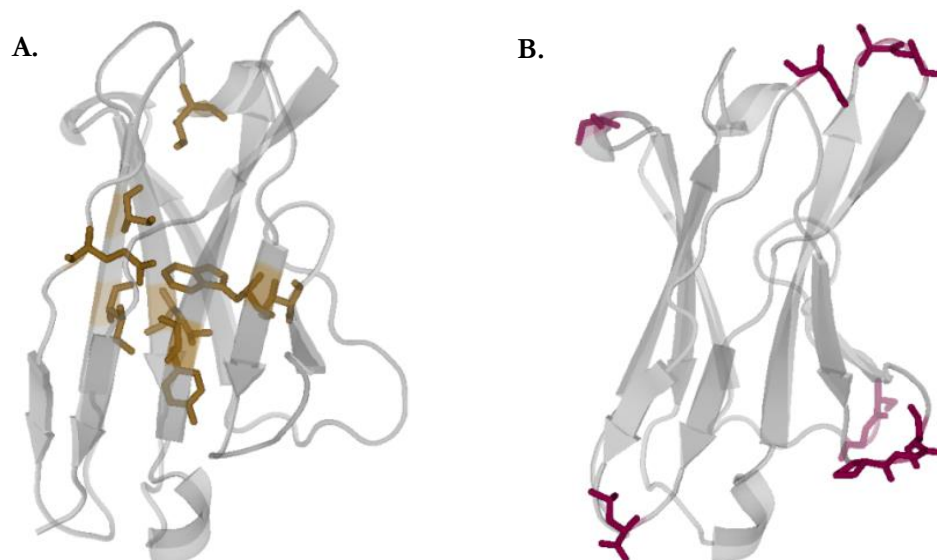


Figura 9. Convergencia entre la metodología de Grado, Eigenectores, y Entropía de Redes. **A.** Ubicación de los residuos mejor calificados marcados en ocre los cuales están apuntando hacia el interior hidrofóbico de la molécula. **B.** Ubicación de los residuos peor puntuados marcados en violeta, los cuales están apuntando hacia el exterior de la molécula, hidrofílicos, y capaces de interactuar con un supuesto antígeno.

Sin embargo, es importante mencionar que un parámetro no tomado en consideración en los análisis de redes es el dinamismo de las estructuras moleculares así como los movimientos que resultan ser críticos para su función (Dror *et al.*, 2012). Por lo tanto, se consideró importante el incorporar los resultados generados de una dinámica molecular con el parámetro de Entropía de Redes.

Propuesta y validación de una metodología para la jerarquización de residuos empleando proteínas evaluadas termodinámicamente con diferente estructura terciaria

Se decidió seleccionar otras proteínas cuya estructura fuese diferente de 6aJL2 y que tuvieran un amplio conjunto de mutaciones puntuales analizadas termodinámicamente. Las proteínas seleccionadas fueron el inhibidor de quimi tripsina (PDB 2CI2), apoflavodoxina (PDB 1FTG), represor dimérico Arc (PDB 1ARR), complejo del dominio de unión α del receptor humano de estrógeno reconociendo su secuencia de ADN (PDB 1HCQ), proteína de choque térmico de *Bacillus subtilis* (PDB 1CSP), y la proteína de choque térmico de *B. caldolyticus* (PDB 1C9O); sus análisis termodinámicos están previamente publicados (Milla *et al.*, 1994; Itzhaki *et al.*, 1995; Perl & Schmid, 2001; Banci *et al.*, 2004; Campos *et al.*, 2004a; Campos *et al.*, 2004b; Wunderlich *et al.*, 2005; Wunderlich & Schmid, 2006; Gribenko & Makhatadze, 2007; Del Pozo-Yauner *et al.*, 2008; Deegan *et al.*, 2010; Hernández-Santoyo *et al.*, 2010; Van den Bedem, 2013; Del Pozo-Yauner *et al.*, 2014). Por el contrario, Aunque el segmento desordenado del supresor de citosina (PDB 2N34) no fue analizado termodinámicamente (Chandrashekar *et al.*, 2015), éste sirvió como modelo de comparación entre proteínas globulares y un dominio intrínsecamente desestructurado.

Primeramente se comparó el parámetro Grado contra los experimentos termodinámicos de mutaciones puntuales de las proteínas globulares (Figura 10). En el caso del dominio de unión al ADN del receptor α de estrógeno se midió la capacidad de reconocimiento del receptor dimérico al ADN. En el caso de 6aJL2, se generaron y se analizaron termodinámicamente varias mutantes puntuales (6a-Q6N, 6a-R25H, 6a-I30G). La Figura 10 muestra la correlación entre el parámetro Grado y el efecto sobre la estabilidad termodinámica de la proteína. Sin embargo, un detalle muy importante de remarcar es que con este parámetro no se puede jerarquizar cada residuo acorde a su importancia estructural ya que varios residuos pueden compartir el mismo número de vecinos.

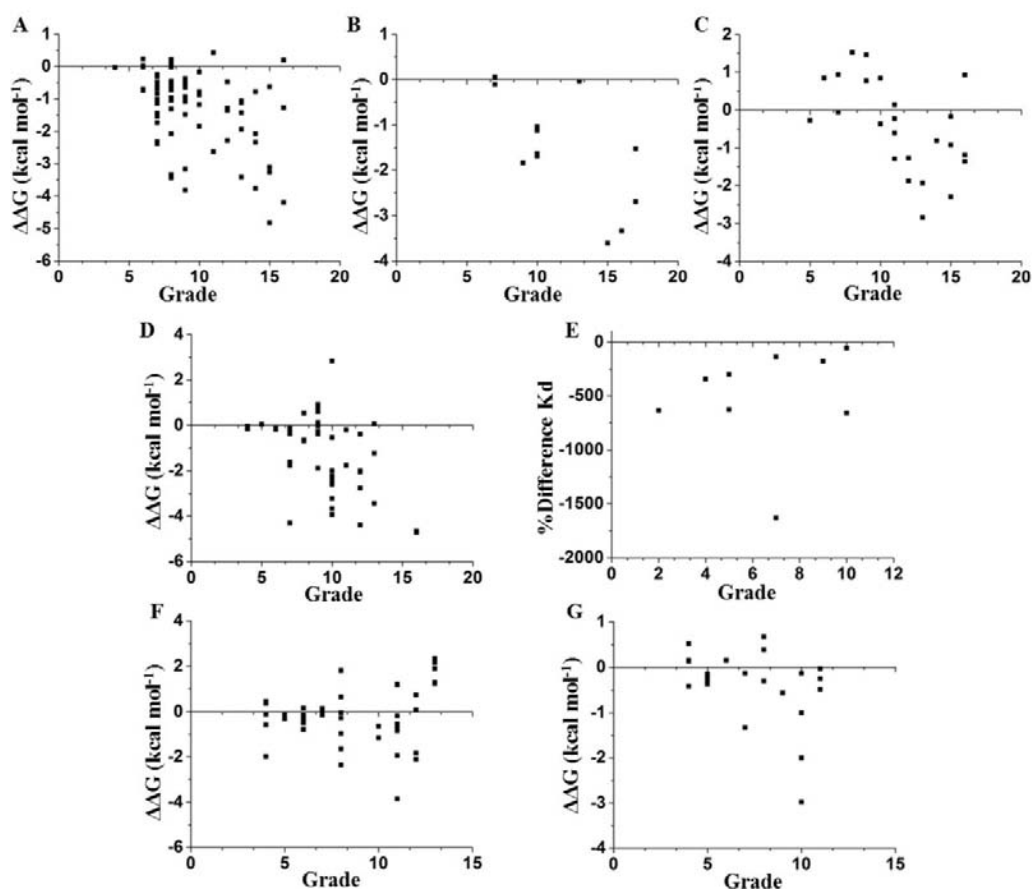


Figura 10. Contraste entre el parámetro Grado y valores experimentales. Cada punto de la gráfica representa una mutación puntual de la proteína en cuestión posicionado en el número de contactos del residuo original. **A.** Inhibidor de quimi tripsina, **B.** 6aJL2, **C.** apoflavodoxina, **D.** represor dimérico Arc, **E.** dominio de unión al DNA del receptor α de estrógeno, **F.** proteína de choque térmico de *B. subtilis*, y **G.** proteína de choque térmico de *B. caldolyticus*.

Se buscó una metodología publicada que conjuntara el análisis de redes y la dinámica molecular pero la búsqueda resultó infructuosa. Por lo tanto, se optó por proponer y validar una metodología que integrara estas dos técnicas para obtener las ventajas de ambas (Luna-Martínez *et al*, 2016). Para dinamizar las proteínas globulares mencionadas anteriormente, éstas

fueron preparadas con el comando *Protein Preparation Wizard* del programa Maestro (Schrödinger LCC, NY, USA), y colocadas en una caja de agua de 10 Å. Mediante el comando Desmond de este mismo programa (Maestro-Desmond Interoperability Tools, version 3.0; Schrödinger, NY, USA), las proteínas fueron colocadas en una caja de agua, calentadas rápidamente, y luego mantenidas a 25 °C durante 25 ns. Todo el proceso se analizó mediante el *Root Mean Square Deviation* (RMSD) de la cadena principal observándose la estabilización del dominio durante este período de tiempo (Figura 11). Un dato interesante observable del RMSD del represor dimérico Arc (Figura 11D), fue que esta proteína registró la mayor diferencia de RMSD con respecto a su estructura inicial debido a que la estructura cristalográfica es de baja calidad acorde al reporte del sitio del PDB. Además, otro factor a considerar es que este represor dimérico está encargado de reconocer a una secuencia de ADN por lo que requeriría mayor grado de libertad para lograrlo. En general, en este marco de tiempo, se puede decir que las moléculas están estabilizadas ya que no se registraron cambios abruptos en el valor de RMSD a diferencia, por ejemplo, de una dinámica molecular donde un β_2 -receptor adrenérgico está reconociendo a un β -bloqueador (Dror *et al.*, 2012).

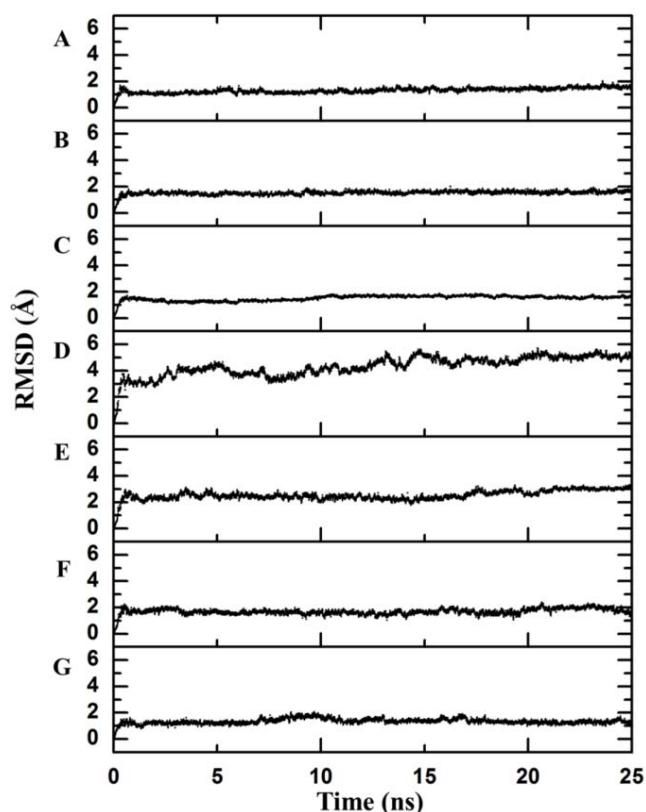


Figura 11. RMSD de la cadena principal. La diferencia de RMSD de la cadena principal entre la estructura cristalográfica y la correspondiente estructura en el momento indicado de la dinámica molecular usando Desmond. Hay un cambio observable en los primeros picosegundos durante el “calentamiento” de la estructura, pero se mantiene estable después de ese momento. **A.** Inhibidor de quimi tripsina, **B.** 6aJL2, **C.** apoflavodoxina, **D.** represor dimérico Arc, **E.** dominio de unión al DNA del receptor α de estrógeno, **F.** proteína de choque térmico de *B. subtilis*, y **G.** proteína de choque térmico de *B. caldolyticus*.

Para conjuntar la dinámica molecular con el análisis de redes, se extrajeron 100 estructuras a lo largo del tiempo de la dinámica, se calculó el promedio de la distancia del centro normalizado de masas entre cada par de residuos, y se construyó una matriz cuadrada con estos valores. Esta matriz se resolvió con la metodología de la Entropía de Shannon cuyo valor resultante se interpreta como el grado de conectividad del residuo en un sentido probabilístico; por lo que, a mayor valor entrópico, mayor probabilidad de perturbar a la red en caso de una perturbación. El valor entrópico de cada nodo fue dividido entre su respectivo valor de fluctuación generado por la dinámica molecular calculado por el *Root Mean Square Fluctuation* (RMSF). Este nuevo parámetro que reúne conectividad y movimiento fue denominado como *Structural-Determining Residue Identifier* (SDRI).

Para el cálculo de los SDRI del segmento desordenado se emplearon las veinte estructuras reportadas (Nicholson & Norton, 2015), y se empleó el RMSD (Figura 12A) y el RMSF (Figura 12B) calculado usando el servidor RCI considerando únicamente los átomos de la cadena principal (Berjanskii & Wishart, 2013). El RMSD de las estructuras del NMR fue superior al de las estructuras sometidas a dinámica molecular, $>10 \text{ \AA}$ (Figura 9A). De estas estructuras se realizó el mismo tratamiento matemático descrito calculándose los valores SDRI coloreándose en gradiente sobre la estructura (Figura 12C).

En la Figura 13 se muestra la distribución de los SDRI con respecto a su posición en la respectiva secuencia. La distribución de los SDRI de las proteínas globulares es muy disímil al del dominio poco estructurado del supresor de citosina. Se puede observar que el comportamiento de los SDRI está mejor distribuido en las proteínas globulares que en el segmento desordenado. Para corroborar el enunciado anterior se realizó un análisis estadístico de la distribución de valores utilizando la simetría, curtosis, y el análisis Shapiro-Wilk como parámetros descriptivos (Tabla 4). Si una distribución es simétrica, existe el mismo número de valores tanto a la derecha como a la izquierda de la media. La curtosis describe la concentración de valores alrededor de la media describiendo si la curva es puntiaguda o roma, siendo valores igual o cercanos a cero cuando se tiene una distribución normal. La prueba Shapiro-Wilk se usa para contrastar que los datos cumplen con una distribución normal (Shapiro & Wilk 1965). Resultó interesante que el segmento desordenado tuviera valores marcadamente disímiles con respecto a los valores de las proteínas globulares, además de que los valores estadísticos resultantes convergieran a una distribución normal.

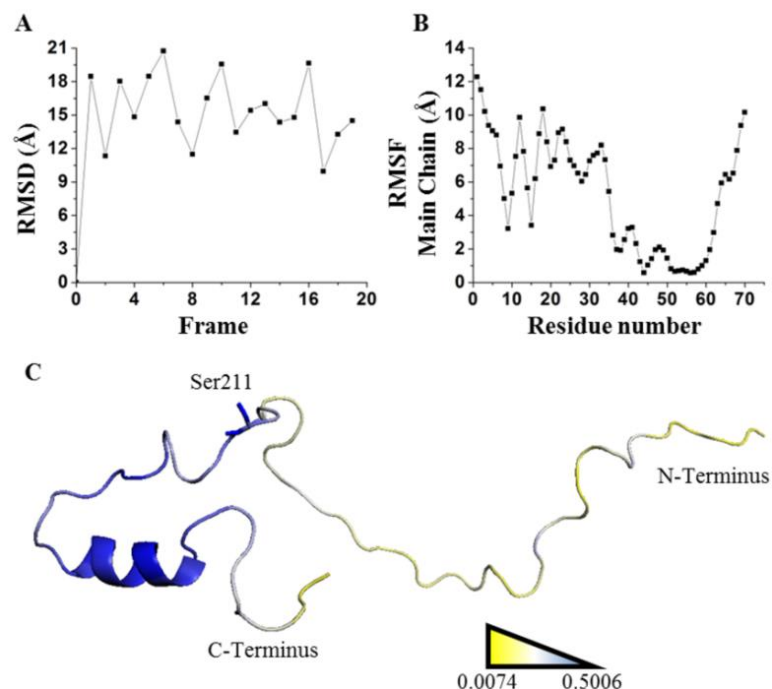


Figura 12. Visualización de los valores SDRI el extremo N-terminal intrínsecamente desestructurado de la región de interacción de la Janus Kinase. **A.** Valores de RMSD al comparar cada estructura contra la primera del PDB 2N34. **B.** Valores de RMSF de cada residuo de la cadena principal mostrando que la horquilla es más rígida que el resto de la proteína. **C.** Gradiente de color donde el valor más alto de SDRI está en azul y el valor más bajo está en amarillo.

Tabla 4. Estadística de la distribución de valores teóricos – SDRIs. Se evaluaron la simetría, la curtosis, y la normalidad de las distribuciones medidas con la prueba Shapiro-Wilk. La distribución de los SDRIs de las proteínas globulares muestra una buena distribución a diferencia del segmento desordenado. El análisis fue hecho usando SigmaPlot 11.0 (Systat Software, San Jose, CA).

Proteína	Simetría	Curtosis	Shapiro-Wilk
Inhibidor de quimiotripsina	0.16	-0.59	0.98
6aJL2	0.74	-0.03	0.95
Apoflavodoxina	0.80	0.12	0.95
Represor dimérico Arc	0.51	0.02	0.97
Complejo Receptor Estrógeno/DNA	-0.11	-0.19	0.99
Proteína de choque térmico de <i>B. subtilis</i>	0.53	-0.67	0.94
Proteína de choque térmico de <i>B. caldolyticus</i>	0.61	-0.54	0.94
Segmento desordenado	1.75	2.15	0.71

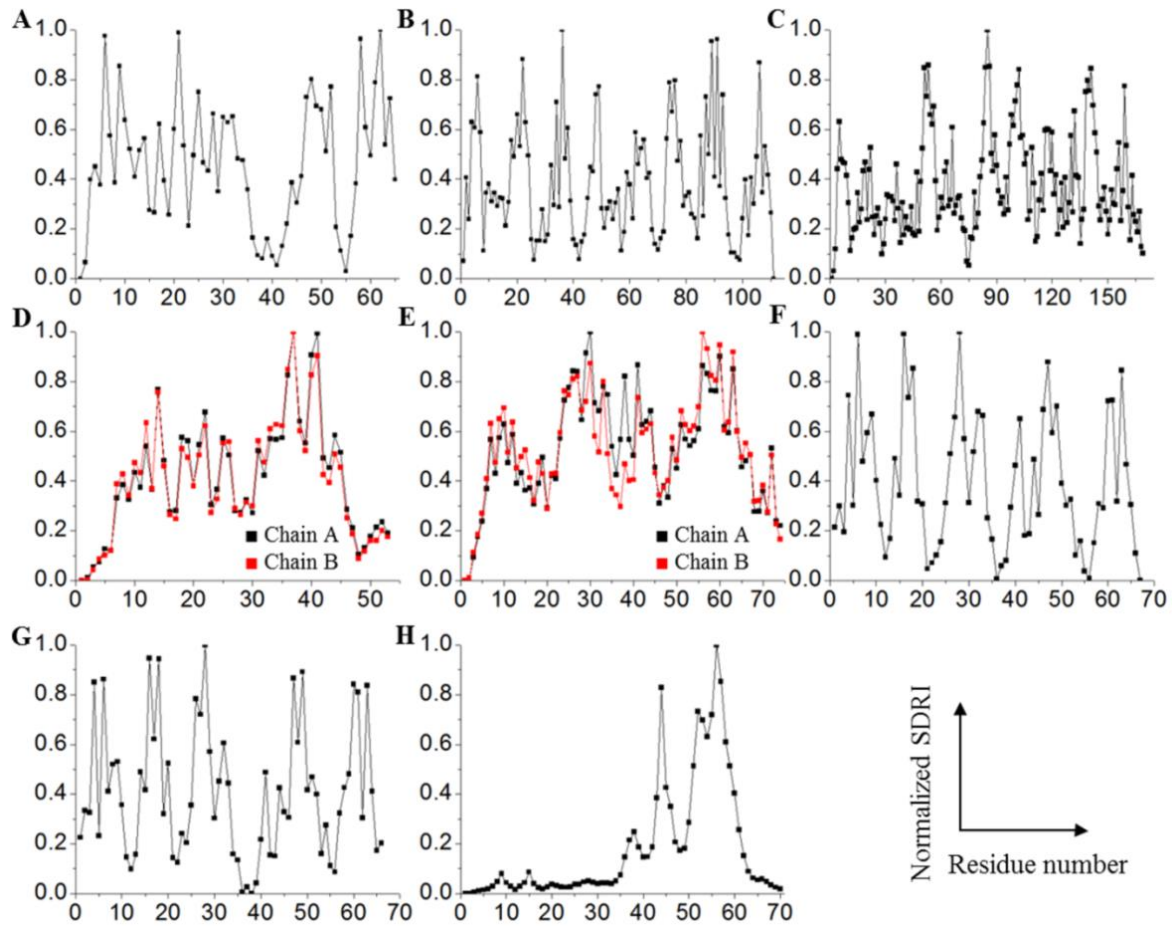


Figura 13. Contraste entre los SDRI y su posición en la secuencia. Cada punto de la gráfica representa el valor normalizado del SDRI proveniente del análisis teórico y su posición acorde a la secuencia. Valores cerca de 0 indican residuos muy flexibles y de baja conectividad y valores cercanos a 1 indican residuos rígidos con alta conectividad. **A.** Inhibidor de quimi tripsina, **B.** 6ajL2, **C.** apoflavodoxina, **D.** represor dimérico Arc, **E.** dominio de unión al DNA del receptor α de estrógeno, **F.** proteína de choque térmico de *B. subtilis*, **G.** proteína de choque térmico de *B. caldolyticus*, y **H.** el segmento desordenado del supresor de citosina.

Al colorear la posición de los SDRI en las estructuras de las proteínas globulares se ubicaron los residuos acorde a su importancia estructural. Al observar la posición de los SDRI en la estructura, se puede notar que los residuos de más alto puntaje recayeron mayoritariamente en residuos apuntando hacia el núcleo hidrofóbico de la molécula (Figura 14). En cambio, los residuos de menor puntaje fueron aquellos de mayor flexibilidad y menor conectividad los cuales fueron ubicados en segmentos encargados de reconocimiento a otras macromoléculas como el asa del inhibidor de quimi tripsina, los CDRs de 6ajL2, el asa superior de la apoflavodoxina, y los extremos amino del represor dimérico Arc. Por lo tanto, el algoritmo propuesto puede ser aplicado a cualquier tipo de estructura terciaria o cuaternaria.

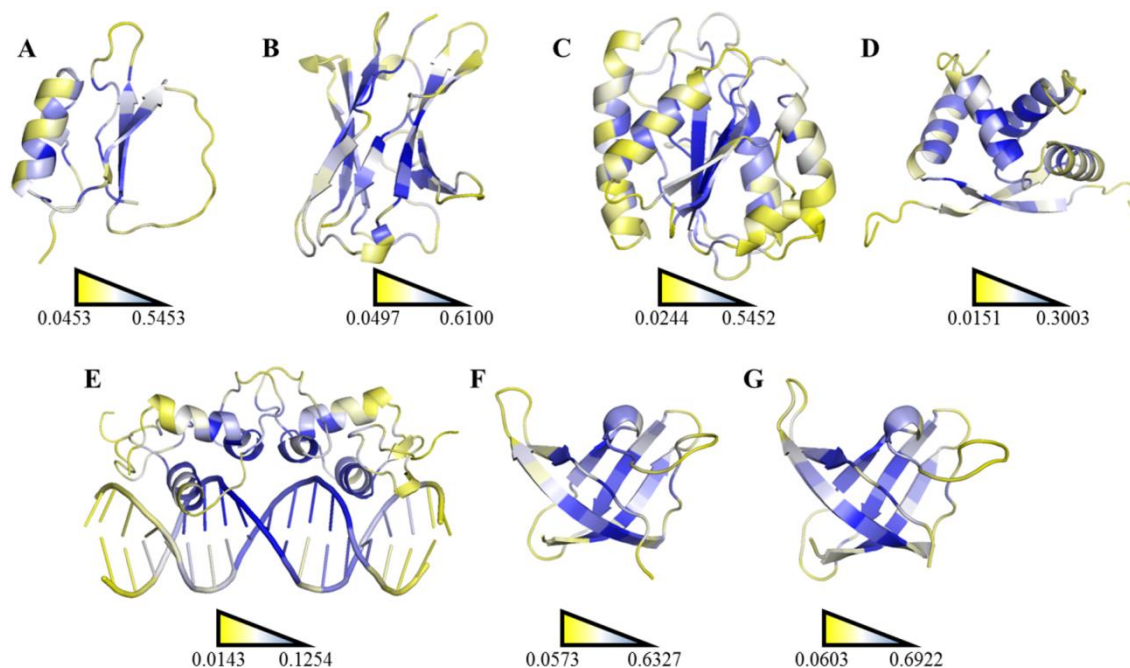


Figura 14. Gradiente de color. El valor más alto de SDRI coloreado en azul representa el nodo más conectado con movimiento restringido, y el SDRI más bajo coloreado en amarillo exhibe al nodo con mayor movimiento y menor conectividad. Las estructuras fueron preparadas usando PyMol. **A.** Inhibidor de quimi tripsina (PDB 2CI2), **B.** 6aJL2 (PDB 2W0K), **C.** apoflavodoxina (PDB 1FTG), **D.** represor dimérico Arc (PDB 1ARR), **E.** dominio de unión al DNA del receptor α de estrógeno (PDB 1HCQ), **F.** proteína de choque térmico de *B. subtilis* (PDB 1CSP), y **G.** proteína de choque térmico de *B. caldolyticus* (PDB 1C9O).

Los SDRIs y los valores termodinámicos derivados de sus respectivas mutaciones sencillas fueron graficadas y analizadas estadísticamente mediante regresión lineal y con el coeficiente de correlación de Pearson (Tabla 5). A pesar de que estos parámetros de estadística lineal mostraban una baja relación entre los SDRIs y los datos experimentales, la inspección visual entre estos datos sugería fuertemente que seguían un patrón (Figura 15).

Tabla 5. Estadística de SDRIs contra datos experimentales. El coeficiente de determinación (R^2) y el coeficiente de correlación de Pearson describen la relación entre los valores teóricos SDRIs y los valores experimentales. El análisis fue realizado usando el programa SigmaPlot11.0 (Systat Software, SanJose, CA).

Proteína	R^2	Coefficiente de Correlación de Pearson
Inhibidor de Quimi tripsina	0.25	-0.50
6aJL2	0.22	-0.54
Apoflavodoxina	0.28	-0.55
Represor dimérico Arc	0.24	-0.50
Complejo Receptor/DNA	0.29	-0.54
Proteína de choque térmico <i>B. subtilis</i>	0.06	-0.12
Proteína de choque térmico <i>B. caldolyticus</i>	0.08	-0.21

Por consiguiente se realizó un análisis estadístico aplicando un Análisis de Componentes Principales (PCA) para identificar patrones de datos aparentemente no relacionados (Abdi & Williams, 2010). Los resultados fueron presentados como *biplots* circulares los cuales permitieron examinar la relación entre las variables SDRI y datos experimentales como vectores (Figura 15). Los *biplots* permiten el análisis de la interacción entre los datos y sus variables, por lo que el término *bi* se refiere a la muestra simultánea entre las filas y las columnas de la tabla, no a la bi-dimensionalidad de la gráfica; por lo *biplots* sólo muestran las dimensiones (variables) que reúnen la máxima cantidad de variación de la tabla. (Kroonenberg, 2007).

Basados en los *biplots* del PCA y dado el ángulo calculado entre ambas variables, se pudo mostrar que el presente método es capaz de estimar el efecto de una mutación sencilla en la estructura de la proteína dependiendo de la importancia de un dado residuo sin importar su posición. Los *biplots* del PCA mostraron una buena correlación entre los datos teóricos y los datos experimentales donde el porcentaje de la variable teórica (DIM1) puede explicar gran parte de los valores experimentales (DIM2) de las proteínas analizadas. Estos resultados confirmaron la hipótesis de que los residuos más conectados y más rígidos son los que tienen una mayor influencia en la estabilidad estructural sin importar el tipo de estructuración. Sólo habría que recalcar que la razón por la cual la proteína de choque térmico de *B. subtilis* mostró un menor valor de relación con respecto a las otras, podría atribuirse a que las mutaciones sencillas fueron dirigidas hacia el aumento de la estabilidad termodinámica guiando los cambios hacia la variante termófila de *B. caldolyticus*.

En general se observó que a mayor valor de SDRI hay una mayor tendencia a desestabilizar sustancialmente la estructura como ocurrió en el caso del represor dimérico Arc (VA22, EA36, IA37, VA41, and FA45), donde estas mutantes mostraron poca o nula cooperación en los experimentos de desnaturalización (Milla, Brown & Sauer, 1994). Estas mutaciones fueron realizadas en residuos con valores altos de SDRI (6, 4, 1, 2, y 19, respectivamente). El siguiente efecto observado fue que aumentar la estabilidad termodinámica ocurrió en las mutaciones realizadas en los residuos con menor valor de SDRI como en el caso de la apoflavodoxina (Figura 15C) y en la proteína de choque térmico de *B. subtilis* (Figura 15F) donde se favoreció la generación de puentes salinos.

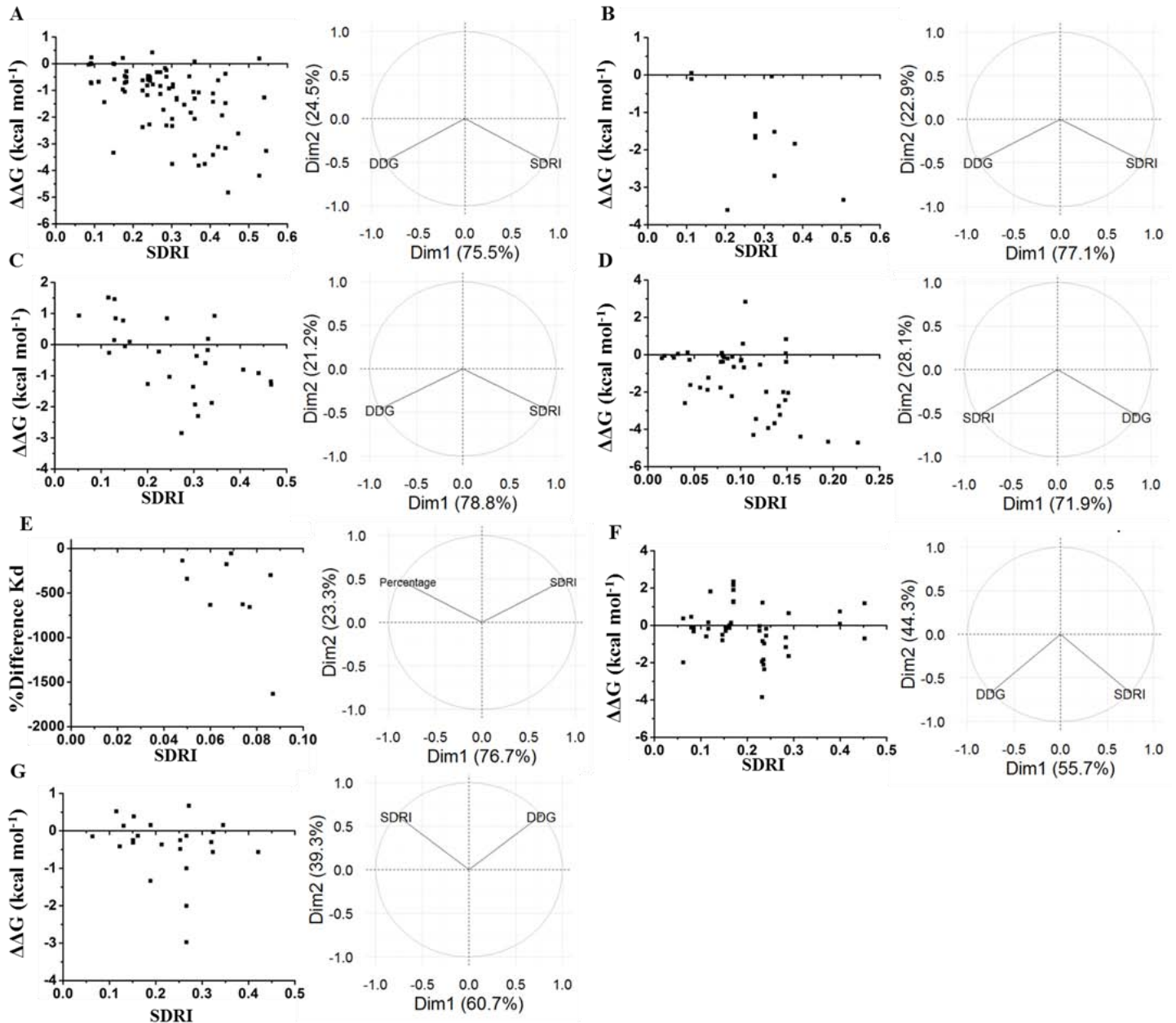


Figura 15. Contraste de los SDRIs contra sus correspondientes valores experimentales. Cada punto de las gráficas representa una mutación sencilla en la proteína de interés. En el caso de los homodímeros únicamente se evaluó la cadena A. Los datos estadísticos se muestran en términos de eigenvectores: Dim1 está referido a los SDRIs y Dim2 está referido a los valores experimentales. Se calculó el cambio de estabilidad termodinámica de la mutante sencilla con respecto a la proteína nativa en términos de $\Delta\Delta G$ (DDG). Para el caso del receptor de estrógeno, se midió la capacidad de reconocimiento del dímero a la secuencia de ADN en términos del porcentaje de diferencia de afinidad de la mutante con respecto a la nativa (%Difference). **A.** Inhibidor de quimi tripsina, **B.** 6aJL2, **C.** apoflavodoxina, **D.** represor dímérico Arc, **E.** dominio de unión al ADN del receptor α de estrógeno, **F.** proteína de choque térmico de *B. subtilis*, y **G.** proteína de choque térmico de *B. caldolyticus*

Continuando con la validación de la metodología, se analizó la relación de los SDRIs con el efecto cinético de renaturalización/desnaturalización que una mutación sencilla puede tener sobre una proteína (Figura 16). Los análisis estadísticos indicaron que los SDRIs tuvieron mayor relación con las cinéticas de desnaturalización (Figura 16A) que con los de la cinética de renaturalización (Figura 16B). El plegamiento es un rearrreglo dinámico de la red, lo cual implicaría que se requerirían otras condiciones y tiempos más largos para tener dinámicas moleculares que involucren los procesos de replegamiento/desplegamiento.

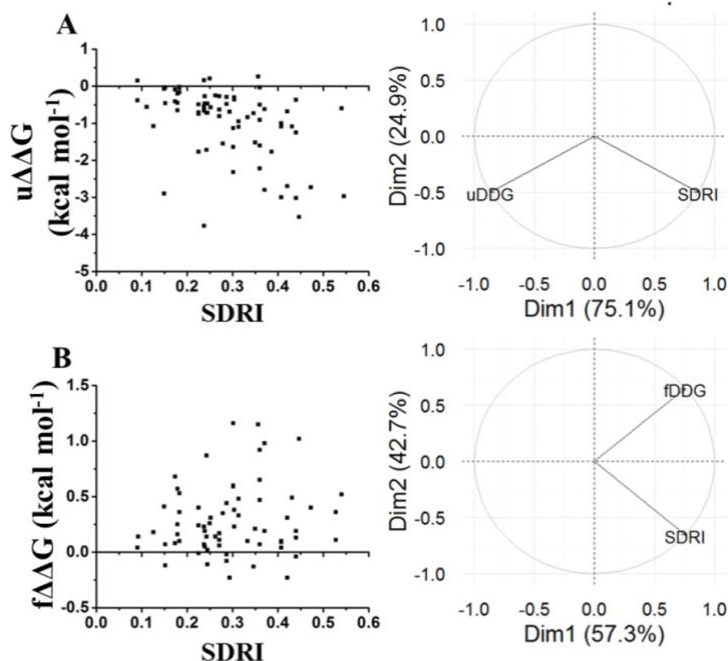


Figura 16. Comparación de los SDRIs con las cinéticas de desnaturalización y renaturalización del inhibidor de quimioproteasa. Las cinéticas de desnaturalización mostraron una mejor correlación con los SDRIs como se muestra en los biplots circulares. DIM1 se refiere a los valores SDRIs y DIM2 a los valores experimentales, $\Delta\Delta G$ (DDG). **A.** $\Delta\Delta G$ de la cinética de desnaturalización, mostrado como uDDG en el *biplot* circular; **B.** $\Delta\Delta G$ de la cinética de renaturalización, mostrado como fDDG en el *biplot* circular.

Para finalizar la validación de la metodología, otro punto importante no contemplado inicialmente fue la ubicación de las regiones desestructuradas en cualquier región. En colaboración con el Dr. Vladimir Uversky se hizo una comparación de los SDRIs con los servidores PONDR® VSL2B y FlexPred. FlexPred predice las fluctuaciones absolutas por residuo de una estructura tridimensional usando los factores B de la proteína en cuestión (Jamroz, Kolinski & Kihara, 2012). La propensidad del desorden intrínseco del predictor PONDR® VSL2B es uno de los predictores más exactos (Fan & Kurgan, 2014; Peng *et al.*, 2005; Peng & Kurgan, 2012). Los resultados de estas comparaciones, mostrados en la Figura 17, muestran claramente que estas tres herramientas computacionales pueden “ver” diferentes, aunque relacionadas, propiedades de una proteína. Hay que hacer notar que se usó la función (1-SDRI) para representar los valores y resaltar el desorden intrínseco en lugar de la rigidez

estructural. Estos resultados sugieren que hay una convergencia entre los valores de flexibilidad estructural calculados a partir de la estructura cristalográfica y la propensión intrínseca de una proteína a preservar el desorden. Además, pareciera que los residuos esenciales para la preservación de la estabilidad global de la estructura de la proteína están típicamente localizados en los segmentos más ordenados y menos flexibles.

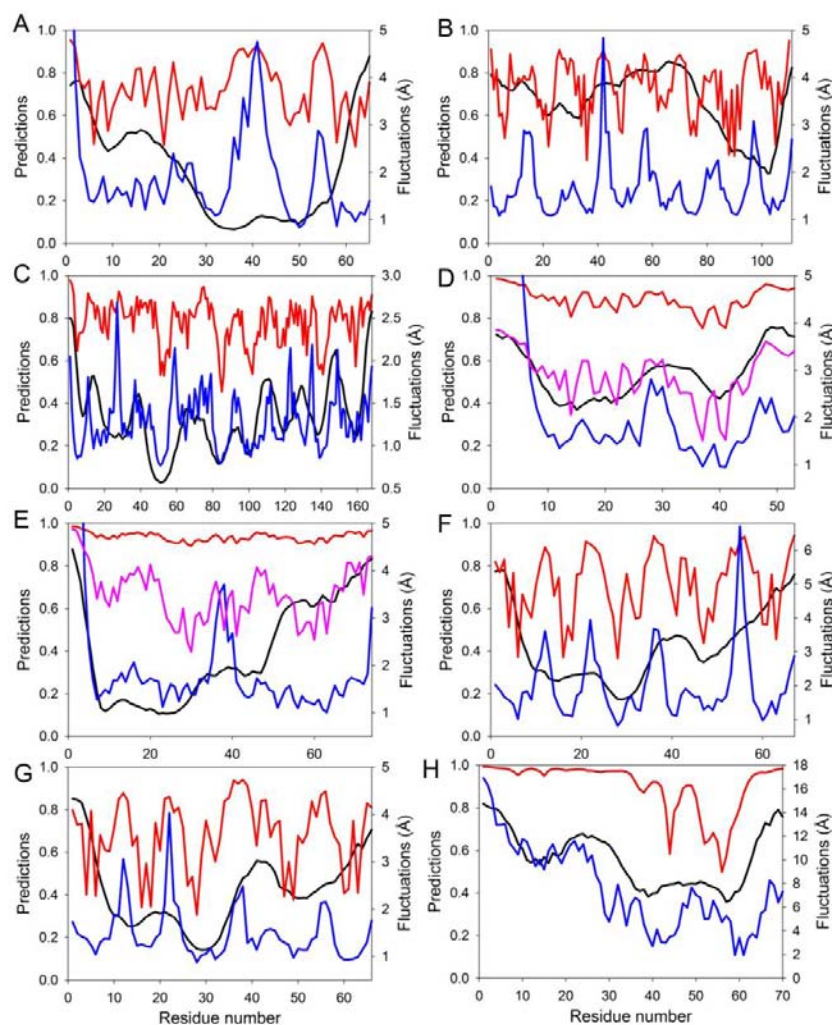


Figura 17. Análisis del desorden intrínseco y la flexibilidad estructural. Las distribuciones de la propensión al desorden intrínseco evaluado con PONDRA® VSL2B (curvas en negro), flexibilidad estructural estimada con FlexPred (curvas en azul), y la función (1-SDRI) que jerarquiza los residuos (curvas rojas y rosas) para las proteínas evaluadas. **A.** Inhibidor de quimi tripsina, **B.** 6aJL2, **C.** apoflavodoxina, **D.** represor dimérico Arc, **E.** dominio de unión al DNA del receptor α de estrógeno, **F.** proteína de choque térmico de *B. subtilis*, **G.** proteína de choque térmico de *B. caldolyticus*, y **H.** el segmento desordenado del supresor de citosina. La propensión al desorden intrínseco y la función (1-SDRI) están normalizadas entre valores de 0 y 1. El represor dimérico Arc y el complejo estructural del receptor de estrógeno resultaron en una reducción dramática de la amplitud de la función (1-SDRI). Las figuras correspondientes (D y E) también incluyen las curvas expandidas (1-SDRI) coloreadas en rosa.

Describiendo un poco los espectros, la inspección visual de las gráficas mostradas en la Figura 17 sugiere que en la mayoría de los casos, la propensión al desorden intrínseco genera bandas anchas las cuales definen la apariencia global de las curvas, mientras que la función (1-SDRI) y FlexPred agregan una resolución más fina. Adicionalmente, para algunas regiones, hay poca convergencia entre estas tres herramientas que puede ser debido a las consideraciones particulares de cada herramienta. No obstante, todas estas observaciones importantes sugieren que el desorden intrínseco predicho por la secuencia de aminoácidos sirve como un fondo importante para definir la flexibilidad global de la estructura terciaria de la proteína, que también es modulada por interacciones a distancia en una molécula plegada. Finalmente, este resultado sugiere que el desorden intrínseco de una proteína está contenido a partir de la estructura primaria pero la estructura terciaria sólo expone ciertos segmentos desordenados para llevar a cabo su función.

Por lo tanto, la metodología teórica propuesta fue examinada y validada con diferentes proteínas cuyas estructuras fuesen diferentes a la de 6aJL2 (Luna-Martínez *et al.*, 2016). La metodología puede jerarquizar cada nodo, aminoácido o nucleótido, acorde a su importancia estructural independientemente del tipo de estructura terciaria o cuaternaria.

Evaluación de las estructuras de AR y 6aJL2 con la metodología propuesta

Una vez validada la metodología se procedió a aplicarla sobre las estructuras de 6aJL2 y AR. Los SDRIs fueron normalizados – el valor más alto es igual a 1 y el valor más bajo igual a 0, y comparados para observar los cambios de comportamiento (Figura 18). Aunque el patrón de distribución de los SDRIs entre AR y 6aJL2 describe un comportamiento similar entre estas estructuras, hay algunos cambios en los valores de algunos residuos que estarían afectando la importancia de otros residuos centrales ubicados a su alrededor.

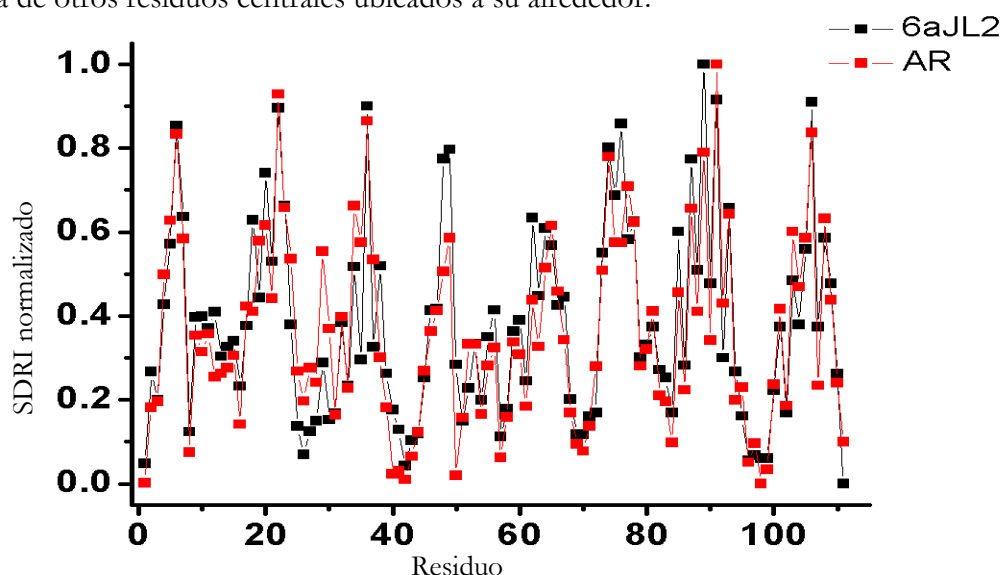


Figura 18. Comparación de SDRIs de 6aJL2 y AR. Los valores SDRIs de cada residuo de cada estructura fueron normalizados. En negro se muestran los SDRIs de 6aJL2, en rojo los SDRIs de AR.

Para todas las estructuras, cada residuo se jerarquizó acorde a los SDRIs donde el valor más alto se asignó al residuo más importante. En el anexo A se muestra la jerarquización total de cada uno de los residuos de AR y de 6aJL2 en base a sus SDRIs, y en la Tabla 6 se muestran los 15 residuos mejor y peor calificados para cada estructura. Los residuos mejor calificados entre ambas estructuras son similares convergiendo con residuos propios del núcleo hidrofóbico. La diferencia más notoria entre ambas estructuras se observó en los residuos peor calificados indicando que la flexibilidad y la conectividad cambiaron en varios segmentos de AR.

Tabla 6. Jerarquía de residuos de 6aJL2 y AR. Se muestran los 15 residuos con los valores más altos de SDRIs y los 15 residuos con los valores más bajos.

Proteína	6aJL2		AR	
Residuos mejor puntuados SDRI	1. Tyr86 2. Cys88 3. Thr102 4. Trp35 5. Cys23 6. Leu73 7. Gln6 8. Ala71	9. Ile48 10. Val47 11. Ala84 12. Ile21 13. Ser72 14. Thr24 15. Ser90	1. Cys88 2. Cys23 3. Trp35 4. Thr102 5. Gln6 6. Tyr86 7. Ala71 8. Thr74	9. Val33 10. Thr24 11. Ala84 12. Ser90 13. Val104 14. Thr5 15. Ile75
Residuos peor puntuados SDRI	97. Gly28 98. His8 99. Ser42 100. Ser68 101. Ser68a 102. Ser56 103. Ser42 104. Ser27	105. Ser94 106. His95a 107. Asn95 108. Ser93 109. Asn1 110. Gly41 111. Leu107	97. His95 98. Asp1 99. Gly41 100. Tyr49 101. Arg39 102. Pro40 103. His95a 104. Ser93	105. Ser56 106. Ser42 107. His8 108. Ser68a 109. Asp68 110. Asn94 111. Glu81

Como cada residuo representa un nodo y cada nodo tiene una posición dentro de su respectiva jerarquía, se comparó la posición de cada nodo de AR con respecto a la posición en 6aJL2. Los cambios positivos apuntan a que el nodo aumentó de importancia dada la disminución de su flexibilidad y un incremento en el número de interacciones a su alrededor. Por el contrario, los cambios negativos apuntan a una pérdida de su importancia por el aumento en la flexibilidad del nodo así como una pérdida considerable en el número de contactos a su alrededor. En la Tabla 7 se muestran los diez cambios de jerarquía, tanto positivos como negativos, los cuales sugieren cambios compensatorios, o descompensatorios, con respecto a 6aJL2.

Tabla 7. Comparación de Jerarquías. Se tomó a la jerarquía de 6aJL2 como base para ser comparada contra AR (Apéndice A). Los cambios indican la variación en la importancia del respectivo nodo indistintamente de la presencia de una mutación en esa posición.

Comparación	Cambio Positivo	Cambio Negativo
6aJL2 vs AR	1. Ala31Ala (45)	101. Leu73Leu (-19)
	2. Gln34Gln (41)	102. Arg61Arg (-20)
	3. Ile30Ile (40)	103. Val47Val (-21)
	4. Tyr36Tyr (33)	104. Phe62Phe (-21)
	5. Gly28Gly (30)	105. Arg39Arg (-22)
	6. Ser26Ser (26)	106. Lys103Lys (-22)
	7. Asp51Asp (25)	107. Val19Val (-26)
	8. Gln89Gln (25)	108. Glu13Glu (-28)
	9. Ser27Gly (23)	109. Gln37Gln (-33)
	10. Asn69Asn (23)	110. Tyr49Tyr (-41)

Si se analiza el efecto residuo por residuo, la Tabla 7 pareciera no tener tanta relevancia pero toma otra perspectiva cuando se conjunta toda la información y se pone en una perspectiva estructural. Globalmente se observa que los cambios positivos y los cambios negativos tuvieron lugar en conjunto aunque en distintas regiones de AR (Figura 19). Las regiones más afectadas fueron el extremo N-terminal y el asa 40-60 las cuales han sido descritas como dominios de anti-agregación (Richardson & Richardson, 2001). Por lo que se podría inferir que una perturbación en este tipo de dominios podría desembocar en un mayor número de interacciones intermoleculares inadecuadas. En contraste, el dominio del CDR1 de AR mejoró su conectividad y se redujo su flexibilidad, aumentando así su posición jerárquica acorde a los SDRIs. Una probable explicación al aparente aumento de rigidez del CDR1 de AR podría ser atribuido a un efecto protector de este segmento hacia la cavidad ocasionada por la ausencia de la Arg25, la cual es considerablemente mayor a la de 6aJL2, 160.3 \AA^3 y 23.4 \AA^3 respectivamente (Figura 20).

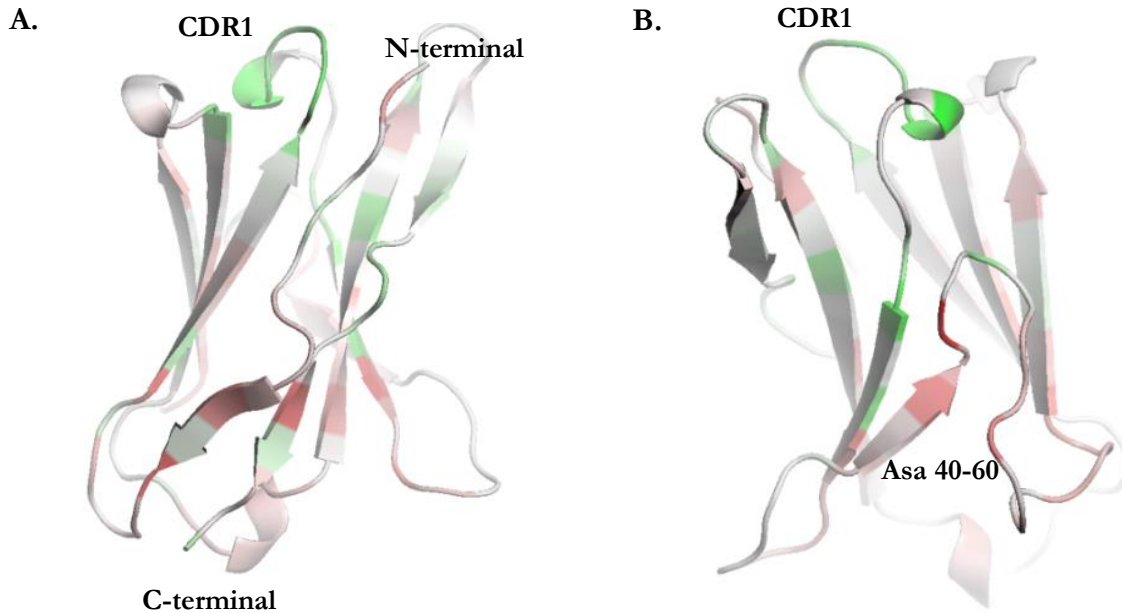


Figura 19. Cambios de jerarquía de AR. La jerarquía de residuos de 6aJL2 fue tomada como base para ser comparada con la jerarquía de residuos de AR. En gradiente de color, en verde están representados los residuos que aumentaron de jerarquía, en rojo los que disminuyeron, y en blanco los que no tuvieron cambio. **A.** Vista frontal de AR mostrándose el CDR1 y los extremos amino y carboxilo terminales, **B.** Vista posterior de AR mostrándose el CDR1 y el asa 40-60 rica en prolínas.

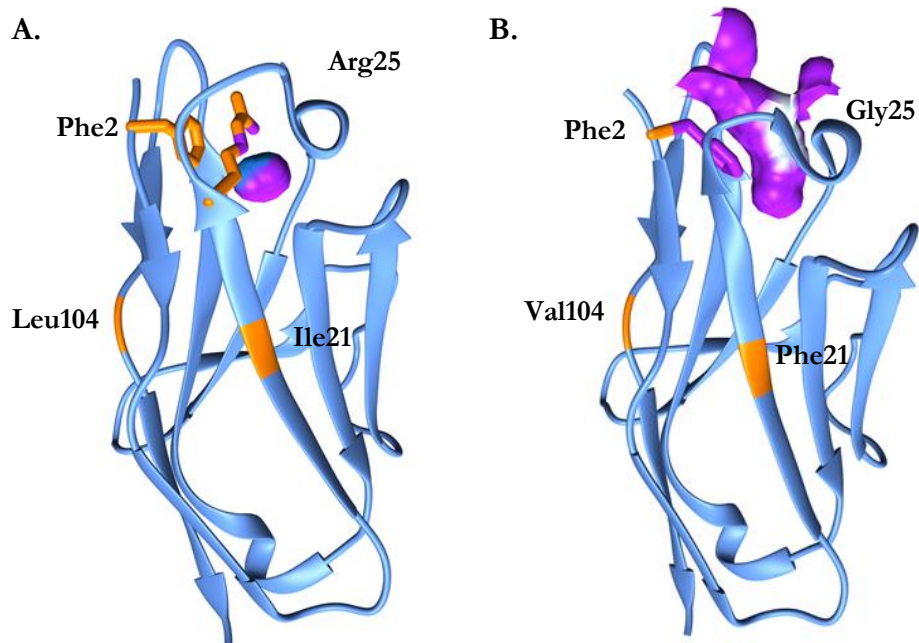


Figura 20. Cavidades. En naranja se muestran los residuos 2, 21, 25 y 104, y en morado se muestra la cavidad ubicada en la región del CDR1 determinada con el servidor CASTp (Dundas *et al.*, 2006). **A.** 6aJL2, **B.** AR

Sintetizando la primera parte de este trabajo, se propuso y se validó una metodología capaz de jerarquizar los nodos (aminoácidos o nucleótidos) en términos de la importancia estructural asociada a la estabilidad termodinámica independientemente de su tipo de estructuración. El requisito más importante para la aplicación de esta metodología es que el dominio a analizar debe de ser globular para obtener información útil. En el caso de un péptido intrínsecamente desestructurado no se generó información sustancial más allá de lo visualmente evidente.

Una ventaja que presentó esta metodología fue que, al comparar dos estructuras con alto grado de identidad, las diferencias fueron aún más evidentes aunque debemos tomar en cuenta que esto no excluye el uso de otras herramientas computacionales y técnicas experimentales para afinar la información generada. Al conjuntar los parámetros de conectividad y flexibilidad, cada residuo fue jerarquizado revelando su importancia estructural. La comparación entre las jerarquías de los residuos de AR y 6aJL2 mostraron cambios globales y locales. Por ejemplo, un residuo conservado entre ambas proteínas podría haber cambiado drásticamente su importancia estructural como en el caso de Ala31 que aumentó 45 posiciones su importancia jerárquica, o en el caso de Tyr49 que decreció 41 posiciones (ver Tabla 7). Los segmentos de AR donde se acumularon las pérdidas de jerarquías se ubicaron en segmentos descritos como importantes para evitar la agregación intermolecular (segmento N-terminal y el asa 40-60). La modificación de la interacción de estos segmentos podría ser apreciada bajo simulaciones a tiempos más largos en condiciones desnaturalizantes.

Por otro lado, el segmento que se rigidizó en AR fue el CDR1 (Figura 19). Hay que considerar que el aumento de rigidez de este segmento no implica que el resto del dominio se estabilice, así como lo muestran los datos termodinámicos. Una explicación al incremento de rigidez del CDR1 de AR es que podría estar protegiendo la cavidad generada por la ausencia de la Arginina en la posición 25 (Figura 20). Esta cavidad, que es considerablemente mayor que la ubicada en 6aJL2 en el mismo segmento, no sólo afectó la reorientación de la cadena lateral de la Phe2 sino también estaría alterando la reorientación de las cadenas de otros residuos sin alterar la orientación de la cadena principal.

Parte II.- Análisis termodinámico y estructural de AR

En la siguiente parte de este trabajo, el objetivo fue entender los efectos de los residuos que presentaron resultados desestabilizantes en todo el dominio de AR mediante la generación de mutaciones hacia la línea germinal, 6aJL2. Sólo hay que enfatizar que la importancia de estudiar estos detalles radica en el hecho de que el paciente “AR” falleció cinco meses después de ser diagnosticado y que ésta fue la primera cadena amiloidogénica λ VI reportada (Husby *et al.*, 1974; Sletten *et al.*, 1981). Como se mencionó anteriormente, AR tiene una identidad de secuencia cercana al 85% con respecto a 6aJL2 (Figura 5).

Selección y análisis de tres residuos importantes en AR

Comparado contra 6aJL2, seis mutaciones en AR fueron localizadas en los *frameworks* (FRs)- N1D, I21F, S43A, E68D, S68bA, y L104V; y ocho mutaciones fueron localizadas en las Regiones Determinantes de Complementariedad (CDRs), segmentos que reconocen el antígeno- R25G, G27S, S31bD, N32S, Y33F, E50D, D92N, S94NS, y N95H (Figura 5). Las mutaciones en los FRs son menos frecuentes porque el proceso de hipermutación somática está concentrado en los “*hot-spots*” del DNA de los CDRs (Monson *et al.*, 2000; Clark *et al.*, 2006). Destacadamente, las posiciones 21 y 104 han sido descritas como pertenecientes al núcleo hidrofóbico central junto con el Trp35 (Chothia *et al.*, 1998; Williams *et al.*, 1996). La relevancia del Trp35 es que también puede ser usado como reportero de la compactación del núcleo hidrofóbico ya que su fluorescencia está mediada tanto por la hidrofobicidad local como por la cercanía del puente disulfuro Cys23-Cys88 (Lakowicz J.R., 2006; Walters *et al.*, 2009).

Dada la gran diferencia de fluorescencia intrínseca de 6aJL2 y de AR a dos diferentes temperaturas (Figura 21), considerando que la eficiencia cuántica del triptófano no se ve afectada por la temperatura ya que este residuo se encuentra ubicado al interior de la proteína (Leach, 1969), y que los valores de T_m de la desnaturalización térmica son de 49.9 °C y 33 °C respectivamente (González-Andrade *et al.*, 2013), sospechamos que el núcleo hidrofóbico de AR no tiene una compactación adecuada. Además, la baja estabilidad termodinámica de AR calculada por González-Andrade *et al.*, de $\Delta G^{25^\circ\text{C}} = 1.7 \text{ kcal mol}^{-1}$ podría ser atribuida a esta baja compactación.

Derivado de estas observaciones, nos enfocamos en los residuos que estuvieran afectando el núcleo hidrofóbico central de AR y que fueran diferentes de los de 6aJL2. Por lo tanto, se propusieron las mutantes AR-F21I y AR-V104L. Por otro lado, la mutación 6a-R25G pareciera una variante alotípica de la línea germinal *6a* (del Pozo *et al.*, 2008). Esta sustitución disminuyó sustancialmente la estabilidad termodinámica de 6aJL2 y, por ende, aumentó su propensión a la formación de fibras. Resulta lógico hipotetizar que la mutante AR-G25R tendrá efectos estabilizantes en esta proteína.

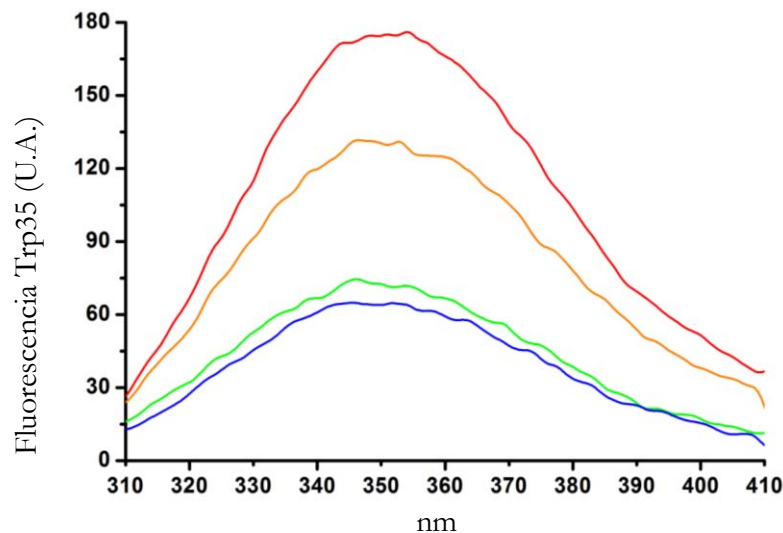


Figura 21. Efectos de la temperatura sobre la fluorescencia intrínseca. Tanto 6aJL2 como AR fueron incubadas a la temperatura indicada durante 10 minutos. 50 $\mu\text{g/ml}$ de proteína fue excitada a 295 nm y se registró el espectro de emisión entre 310 y 410 nm. En línea roja se muestra a AR incubada a 37 °C y en naranja a 25 °C; en línea verde se muestra a 6aJL2 incubada a 37 °C y en azul a 25 °C.

Entonces, las mutaciones que se realizaron en AR tanto en versión sencilla como en su combinatoria fueron AR-F21I, AR-G25R, y AR-V104L. Las mutantes dobles y triples fueron nombradas acorde a la posición de los residuos mutados; por ejemplo, la triple mutante fue nombrada como AR-21/25/104. Todas las mutantes fueron expresadas y purificadas así como se describió en la sección de Materiales y Métodos. Contrario a lo que se esperaba, AR-G25R mostró prácticamente una nula expresión heteróloga y la variante AR-25/104 mostró una baja expresión, alrededor de 100 μg de proteína purificada por litro de cultivo. No obstante, la variante AR-21/104 generó más de 1.5 mg de proteína purificada.

Conservación de la estructura secundaria de las mutantes de AR

Se analizó la calidad de la estructura secundaria mediante dicroísmo circular de UV lejano (Figura 22). Todos los espectros de las mutantes de AR, 6aJL2, y la mutante 6a-R25G, mostraron un mínimo alrededor de 215 nm el cual es típico de una estructura rica en estructura tipo β . No obstante, no todas las mutantes de AR mostraron un mínimo marcado en 230 nm, característico de residuos aromáticos ópticamente activos, y este mínimo podría deberse a un rearrreglo estructural interesante ya que sólo las mutantes conteniendo la mutación F21I tenían este mínimo. Otros espectros colectados fueron obtenidos a través de un calentamiento gradual de las muestras hasta los 70 °C. Posteriormente fueron enfriadas hasta 20 °C para así obtener los espectros de las proteínas desnaturalizadas y renaturalizadas. Estos espectros fueron muy similares entre sí, en su respectiva variante, sugiriendo que el proceso de desplegamiento es reversible.

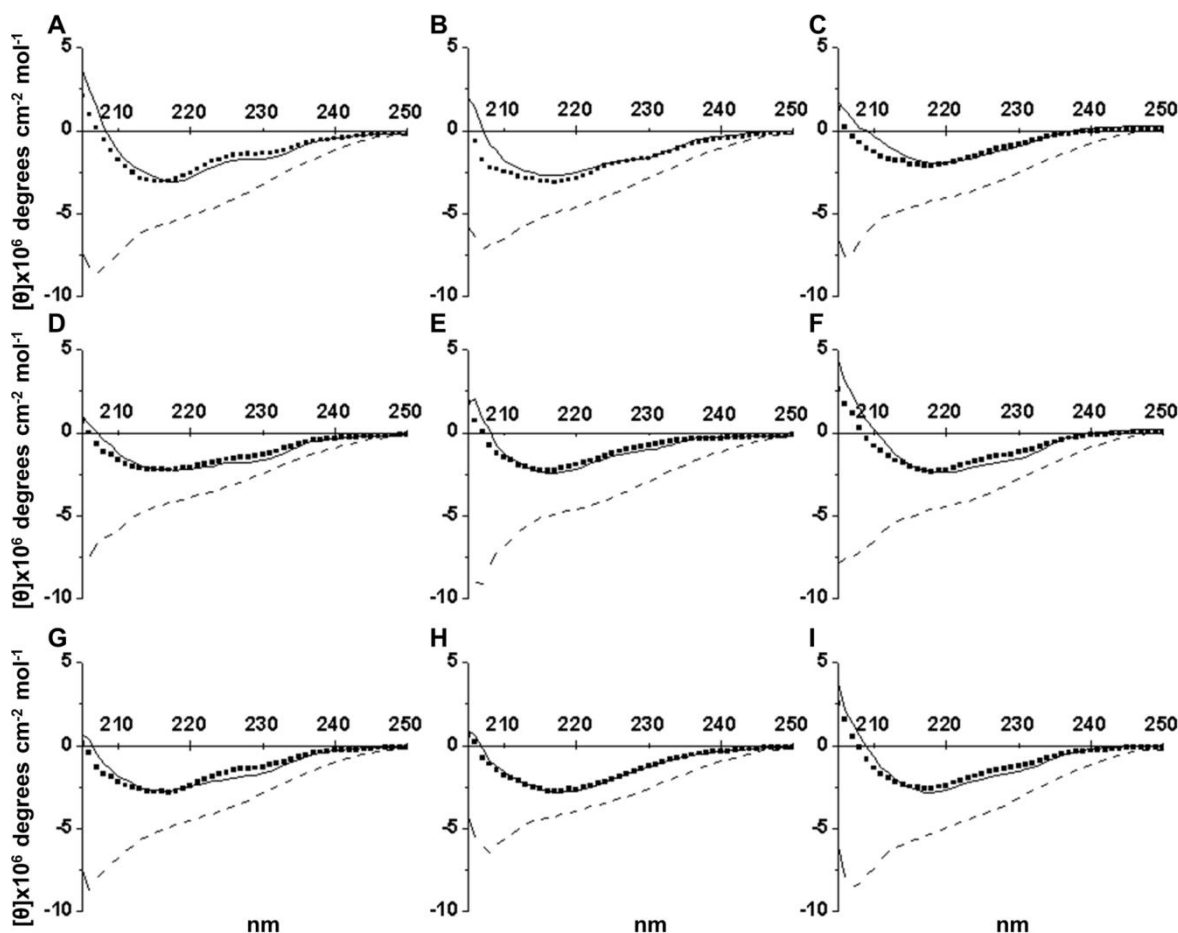


Figura 22. Espectros en UV lejano en estado nativo, desnaturizado, y renaturalizado. Las líneas continuas muestran los espectros de las proteínas nativas obtenidos a 20 °C; los espectros de las proteínas desnaturizadas térmicamente a una velocidad de 1 K/min hasta 70 °C y mantenidas a esta temperatura durante 10 minutos se muestran en líneas con puntuado discontinuo (---); y los espectros de las proteínas renaturalizadas a 20 °C después del ciclo de calentamiento se muestran en líneas con cuadros (■). A. 6aJL2, B. 6a-R25G, C. AR, D. AR-F21I, E. AR-V104L, F. AR-21/25, G. AR-21/104, H. AR-25/104, e I. AR-21/25/104.

Estabilidad termodinámica de las mutantes de AR

Las proteínas fueron analizadas mediante experimentos por desnaturación química y térmica (Figura 23, Tabla 8). Notablemente, exceptuando la mutante AR-G25R, las proteínas siguieron un proceso de desnaturación de dos estados. AR-F21I fue la mutante sencilla menos afectada reflejándose en la mejoría en los valores de la estabilidad termodinámica (ΔG°), el punto medio desnaturación (C_m), y en la temperatura media (T_m). Aunque las mutantes sencillas AR-F21I y AR-V104L mostraron cambios de entalpía similares (ΔH°); AR-V104L no contribuyó importantemente a la estabilización como ocurrió con la mutante AR-F21I. La mutación G25R, como la mutante AR-21/25, difícilmente contribuyó a la estabilidad comparada con la mutante AR-F21I a pesar del notorio incremento en la ΔH° .

La mutante AR-25/104 mostró un ligero incremento en los valores ΔH° y de $\Delta\Delta G^\circ$ comparado con la mutante AR-V104L; sin embargo, en términos del aumento de la estabilidad de $\Delta\Delta G^\circ$, no fue significativo a comparación de la mutante AR-F21I.

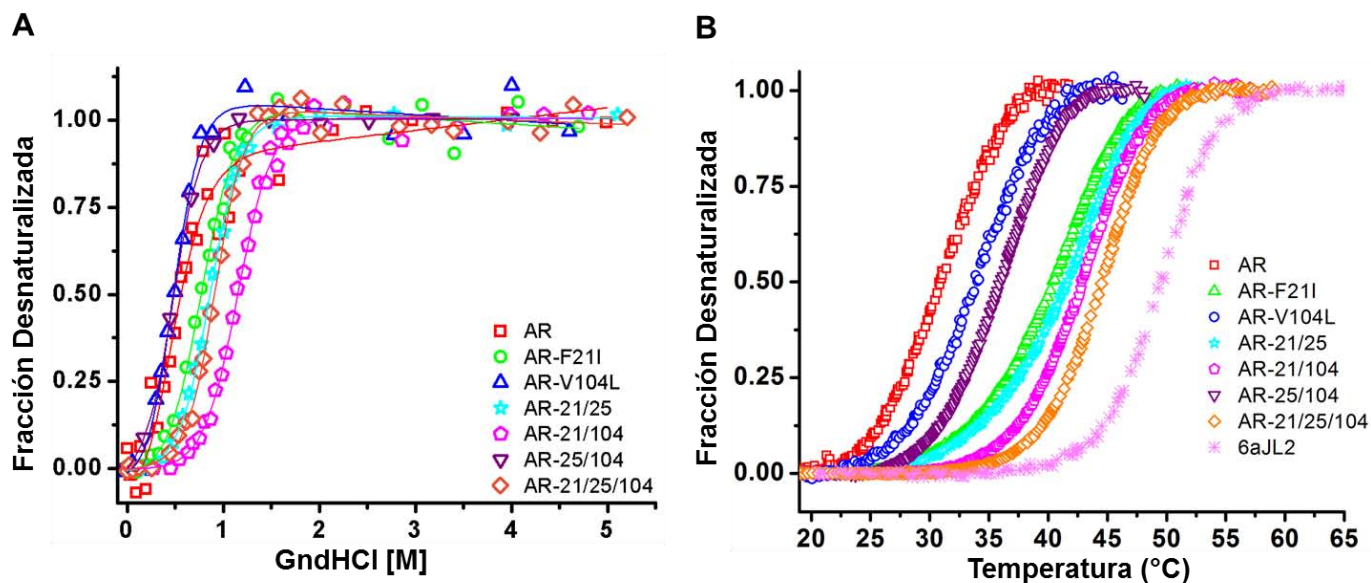


Figura 23. Experimentos de desnaturalización. Los parámetros termodinámicos calculados están mostrados en la Tabla 8. **A.** Fracción de proteína desnaturalizada en función de la concentración del cloruro de guanidina. Las líneas sólidas representan el ajuste a un modelo de dos estados. **B.** Fracción de proteína desnaturalizada en función de la temperatura, calentada a una velocidad de 1 °C/min.

Aunque la mutante AR-21/104 mostró parámetros termodinámicos térmicos similares a la mutante AR-21/25, los parámetros derivados de los experimentos químicos muestran mejores valores en la mutante AR-21/104 sugiriendo un restablecimiento en el núcleo hidrofóbico. Notablemente, la triple mutante exhibió valores de estabilidad térmica similares a los de éstas dobles mutantes, indicando que la mutación Gly25Arg no tuvo una aportación sustancial al mejoramiento de la estabilidad termodinámica. Entonces, el anillo aromático de la Phe21 está perturbando enormemente interacciones importantes entre los residuos que conforman el núcleo hidrofóbico. Aun así, ni la más estable de las variantes de AR pudo igualar los parámetros termodinámicos de los de 6aJL2 (Tabla 8), sugiriendo que otros cambios no contemplados en este trabajo estarían previniendo la completa estabilización de AR.

Tabla 8. Estabilidad termodinámica. Los experimentos se realizaron por triplicado y el error mostrado es la desviación estándar. Para la desnaturalización térmica, se muestran los valores calculados a partir de las curvas de desnaturalización. Los valores de $\Delta\Delta G$ fueron calculados considerando a AR como la proteína nativa; los valores positivos indican una mayor estabilidad que la de AR. La mutante AR-G25R fue una mutación deletérea y sus valores termodinámicos no fueron determinados. ^a Datos tomados de del Pozo-Yauner, 2008. ^b $\Delta\Delta G = (T_m \text{Mutante} - T_m \text{WildType}) * (\Delta H^\circ \text{WildType} / T_m \text{WildType})$ (Pace *et al.*, 1989). ^c $\Delta\Delta G = m \text{WildType} * (C_m \text{Mutante} - C_m \text{WildType})$ (Pace *et al.*, 1989)

Parámetros Termodinámicos	AR (□)	AR-F2II (Δ)	AR-V104L (○)	AR-21/25 (★)	AR-21/104 (●)	AR-25/104 (▽)	AR-21/25/104 (◇)	6aJL2 ^a (*)	6a-R25G ^a
ΔH_m° kcal mol ⁻¹	51.8 ± 3.2	50.6 ± 4.0	58.3 ± 1.9	70.4 ± 1.4	69.5 ± 2.5	62.5 ± 0.2	69.4 ± 0.2	86.2	77.2
T_m °C	31.6 ± 0.3	40.4 ± 2.2	35.0 ± 0.3	43.5 ± 0.4	44.8 ± 0.4	36.8 ± 0.9	44.7 ± 0.1	49.9	44.2
$\Delta G^{25^\circ\text{C}}$ kcal mol ⁻¹	1.2 ± 0.1	2.5 ± 0.2	1.9 ± 0.1	4.1 ± 0.1	4.3 ± 0.2	2.3 ± 0.2	4.3 ± 0.1	5.2	3.8
$\Delta\Delta G^\circ$ ^b kcal mol ⁻¹	-	1.5	0.6	2.0	2.2	0.9	2.2	3.1	2.1
C_m [M]	0.4 ± 0.1	0.8 ± 0.1	0.4 ± 0.1	0.8 ± 0.1	1.2 ± 0.1	0.5 ± 0.1	1.0 ± 0.1	1.4	1.0
$-m$ kcal mol ⁻¹	4.0 ± 0.2	3.2 ± 0.2	3.5 ± 0.5	3.5 ± 0.5	3.1 ± 0.4	4.5 ± 0.6	3.7 ± 0.3	3.6	3.8
$\Delta G^\circ_{\text{H}_2\text{O}}$ kcal mol ⁻¹	1.6 ± 0.3	2.6 ± 0.2	1.5 ± 0.2	2.9 ± 0.5	3.8 ± 0.3	2.1 ± 0.2	3.5 ± 0.2	5.2	3.7
$\Delta\Delta G^\circ$ ^c kcal mol ⁻¹	-	1.5	0.1	1.7	3.2	0.2	2.1	4.0	2.4
t_{lag} h	0.4	1.0	0.5	2.7	10.5	0.9	7.5	14.2	2.0
Growth rate h ⁻¹	11	7.3	7.3	1.8	1.0	6.8	1.2	0.5	1.8

Cinéticas de fibrillogénesis y visualización de las fibras por microscopía electrónica

Los valores termodinámicos fueron comparados con la formación de fibras *in vitro*, experimentos conocidos como fibrillogénesis *in vitro* (Figura 24). Fue sorprendente observar que AR formó fibras en menos de 1 hora a 37 °C, mostrando el tiempo *lag* más corto y la velocidad de incorporación más rápida. Fue igualmente sorprendente que estos tiempos aumentaron cuando los experimentos se realizaron a 25 °C (Figura 24A).

Las mutantes también exhibieron tiempos *lag* muy diferentes entre sí pero acordes a su estabilidad termodinámica, con la excepción de la triple mutante AR-21/25/104 (Figura 24B). Aquellas con cinéticas de fibrillogénesis similares a las de AR fueron las mutantes menos estables como AR-V104L y AR-25/104, mostrando que la mutación V104L por sí sola no pudo contrarrestar el efecto destabilizante de la Phe21. Por ende, las mutantes conteniendo la mutación F21I de AR mostraron una mayor resistencia a la formación de fibras. Interesantemente, la mutante AR-21/104 fue la mutante menos fibrillogénica.

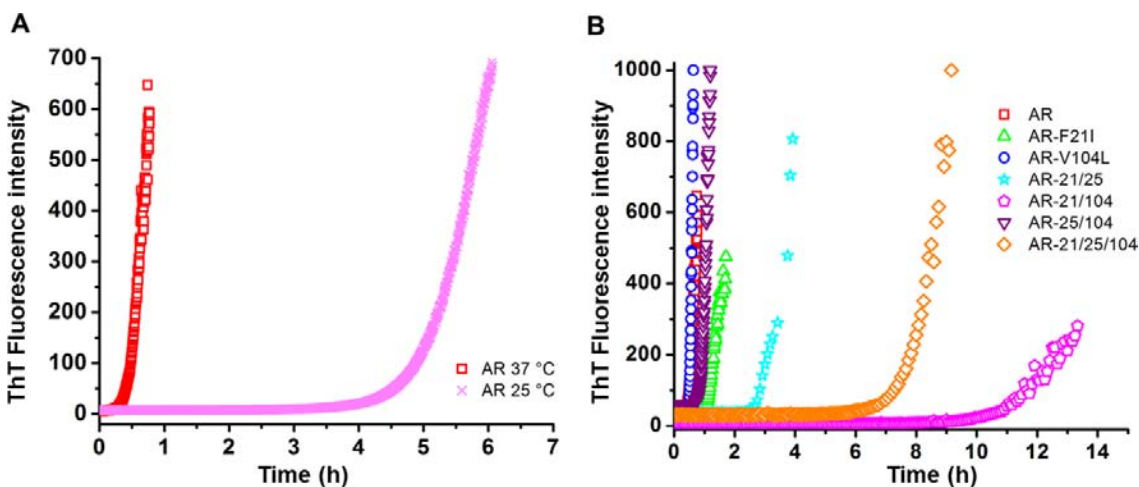


Figura 24. Fibrillogénesis *in vitro*. Las muestras conteniendo 100 $\mu\text{g ml}^{-1}$ de proteína fueron incubadas en PBS 1X a 37 °C, a menos que otra temperatura esté indicada, y bajo agitación continua. La formación de fibras fue monitoreada mediante los cambios de fluorescencia del Tioflavín T. Los experimentos se repitieron por triplicado. **A.** AR incubada a 37 °C y a 25 °C mostrando diferentes tiempos *lag*. **B.** Fibrillogénesis de las mutantes de AR mostrando diferentes tiempos *lag* y diferentes velocidades de extensión de fibras, incubadas a 37 °C.

En colaboración con la Dra. Rosana Sánchez del Instituto de Biotecnología de la UNAM, la microscopía electrónica mostró cambios en la distribución poblacional entre fibras amiloides y lo que aparentemente son agregados amorfos (Figura 25). Como se puede apreciar en la microscopía de AR (Figura 25A), se encontró una gran cantidad de fibras definidas, pequeñas, y delgadas. Conforme la mutante de AR es más estable, la población de fibras va disminuyendo tanto en cantidad como en definición, al igual que van en aumento las regiones ligeramente turbias de las cuales podría decirse que toman la apariencia de agregados amorfos (Figura 25 B-G). En pocas palabras, podríase decir que las cinéticas de formación de fibras *in vitro* fueron afectadas conforme el aumento de la estabilidad.

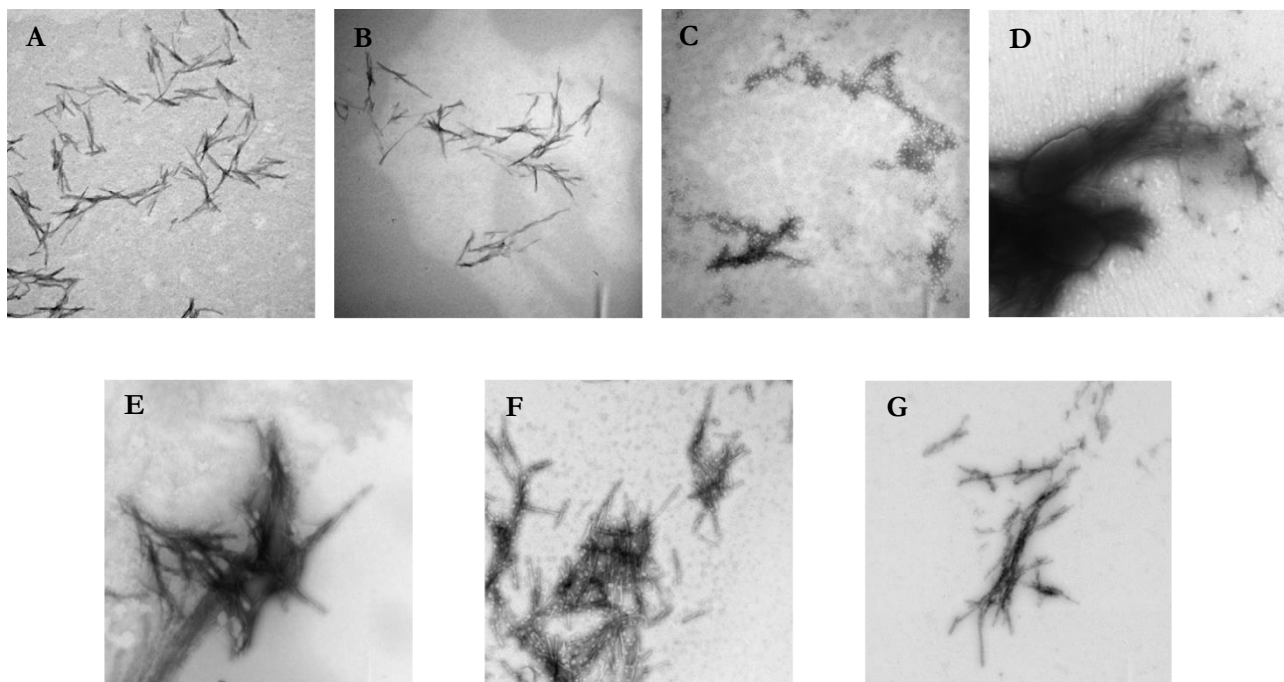


Figura 25. Microscopía de fibras amiloides. La microscopía electrónica de las fibras, posterior a una incubación de 24 horas, se realizó a una magnificación de 50 K de muestras a una concentración de 200 µg ml⁻¹. **A.** AR, **B.** AR-F21I, **C.** AR-V104L, **D.** AR-25/104, **E.** AR-21/104, **F.** AR-21/25, y **G.** AR-21/25/104

Cambios detectables en el extremo N-terminal generados por la ausencia de Arg25

Debido a que previamente se había descrito en la Figura 8 una alta similitud estructural entre las cadenas principales de 6aJL2, su variante 6a-R25G, y AR, este resultado permitía suponer que la influencia que podría estar teniendo la posición 25 sobre la estructura en su conjunto resultaría sutil. Conjuntamente, la evidencia experimental derivada de las mutantes de AR sugiere que la arginina, proveniente de la mutación Gly25Arg, no se está acomodando adecuadamente y que, de alguna manera, el ambiente químico de AR en esta región está influyendo en una correcta orientación. Como la Phe2 mostró un cambio en su orientación, se conjeturó sobre la presencia de cambios mínimos en el extremo N-terminal. Asimismo, previamente se ha demostrado que mutar la Phe2 por otros residuos afecta la estabilidad del dominio de 6aJL2 (del Pozo *et al.*, 2015).

Por esas razones, se consideró importante poner atención en el segmento N-terminal entre los residuos 1 al 33 (Figura 26). Puesto que 6aJL2 cristalizó como un dímero, se compararon ambos monómeros, y se observó que la hebra β1 estaba ausente en el monómero A pero presente en el monómero B (Figura 26A). Aun así, esta hebra de 6aJL2 resultó ser ligeramente más corta con respecto al monómero A de 6a-R25G, y, al mismo tiempo se observó que la hebra β3 aumentó de longitud en 6a-R25G (Figura 26B). La similitud del segmento N-terminal entre 6a-R25G y el de AR es notorio (Figura 26C). Empero todas estas

modificaciones en las hebras β_1 y β_3 entre estas variantes no afectó en lo más mínimo la longitud de la hebra β_2 (Figura 26D).

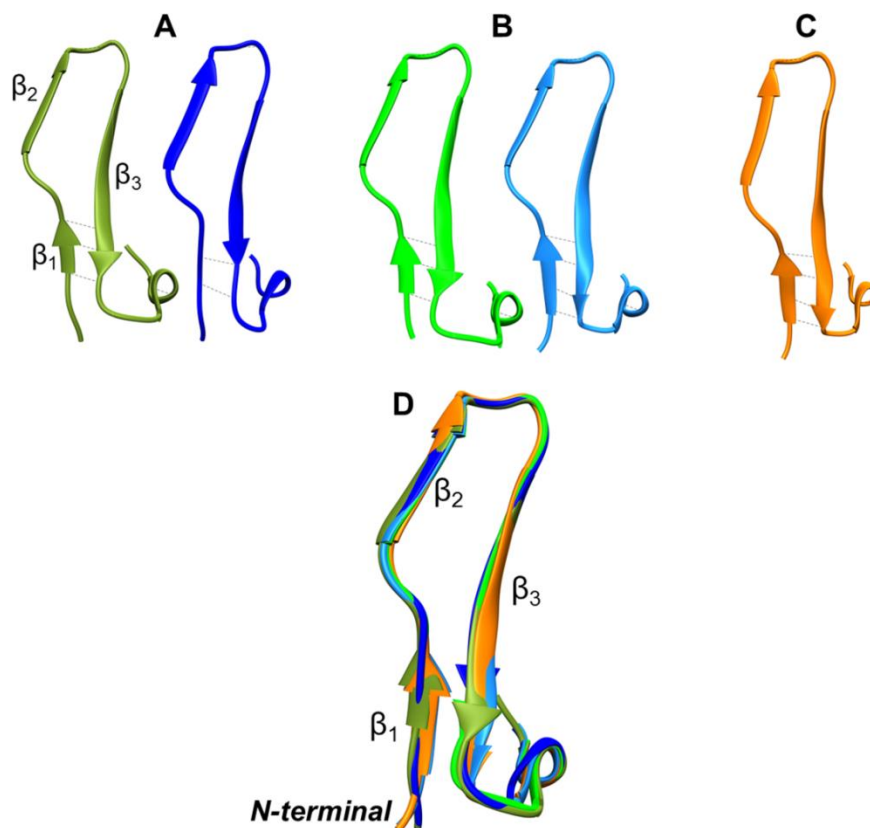


Figura 26. Extremos N-terminal. La longitud entre las hebras β_1 y β_3 y el número de puentes hidrógeno uniendo estas hebras son diferentes entre 6aJL2, 6a-R25G, y AR. **A.** En color oliva se muestra el monómero A de 6aJL2, y en azul rey el mismo segmento del monómero B. **B.** En verde se muestra este segmento del monómero A de 6aJL2, y en cian al mismo segmento del monómero A de 6a-R25G. **C.** En naranja se muestra este segmento de AR. **D.** Traslape de estos segmentos del monómero A de 6aJL2, 6a-R25G, y AR.

Todas estas modificaciones en la estructura secundaria (Figura 22), y de las interacciones entre residuos (Tabla 2), a pesar de la alta homología estructural, dieron pauta a hipotetizar si todo esto modificaría la estructuración cuaternaria. Por lo que se decidió analizar la estructuración cuaternaria más simple que, para este caso, son los dímeros de Bence-Jones.

Generación de estructuras cuaternarias a partir de las estructuras cristalográficas

Siguiendo sus propios parámetros de simetría se construyeron dímeros de Bence-Jones de 6aJL2, 6a-R25G, y AR y sus áreas de interfase usando el servidor PISA (Krissinel & Henrick, 2007). El traslape entre todos estos dímeros mostró una alta homología (Figura 27); asimismo, el área de interfase para los tres dímeros resultó similar con un valor de 623 Å. Sin embargo, el número de puentes de hidrógeno y de puentes salinos localizados en la interfase resultó diferente. En 6aJL2 se encontraron 6 puentes de hidrógeno y ningún puente salino; en

6a-R25G, 5 puentes de hidrógeno y 2 puentes salinos; y en AR, 4 puentes de hidrógeno y 4 puentes salinos. Lo anterior sugiere que, en caso de formarse en la orientación adecuada, los dímeros de Bence-Jones de AR tendrían interacciones muy fuertes en el área de interfase.

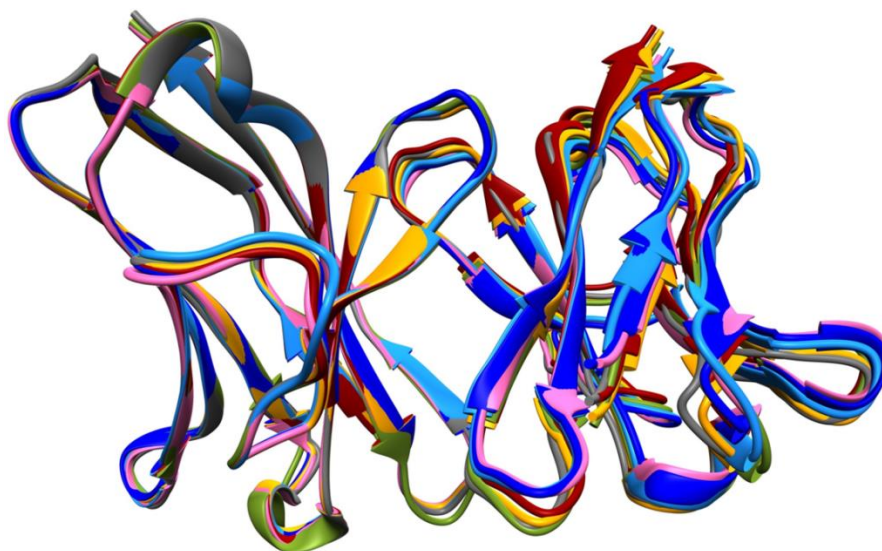


Figura 27. Dímeros de Bence-Jones. Estructuras diméricas generadas con el servidor PISA. El dímero en color gris pertenece a AR; en verde, 6aJL2. Los dímeros de 6a-R25G estuvieron organizados de la siguiente manera: en naranja se encuentra interactuando los monómeros A-G; en cian, los monómeros F-D; en rojo, los monómeros C-E; en azul rey, los monómeros H-B; y en rosa, los monómeros B-H.

Estos resultados se vuelven más interesantes cuando se incrementa el tamaño de la celda unitaria siguiendo sus parámetros de simetría sin seleccionar la generación de dímeros de Bence-Jones. Usando el programa Chimera UCSF (Pettersen *et al.*, 2004), se generaron las estructuras octaméricas de 6aJL2 y de AR dado que la variante 6a-R25G cristalizó naturalmente como octámero (Figura 28). Considerando que AR presenta 14 mutaciones y una glicina extra comparada contra 6aJL2, el fenómeno más interesante fue que sus presuntas estructuras octaméricas se asemejaron al octámero cristalográfico de 6a-R25G (Figura 28A-C). Como el octámero de 6a-R25G no requirió el fenómeno de intercambio de segmentos o intercambio de dominios en esta multimerización, este resultado sugiere que el apilamiento directo es el mecanismo preferido para la agregación de las cadenas ligeras tipo λ VI (Nelson & Eisenberg, 2006). Al rotar el octámero de 6a-R25G hacia una proyección transversal (Figura 28D), se puede observar que la organización de las cadenas principales permite un pequeño hueco donde están apuntando las cadenas laterales de aminoácidos polares como His8, Gln38, Ser42, y Tyr87 junto con Pro7 (Figura 28E). Estos mismos aminoácidos están presentes en los octámeros de 6aJL2 y de AR. De acuerdo con todos estos datos estructurales, podría sugerirse que el reajuste del extremo N-terminal podría considerarse irrelevante pero quizás se trate de un cambio significativo que facilite la asociación intermolecular.

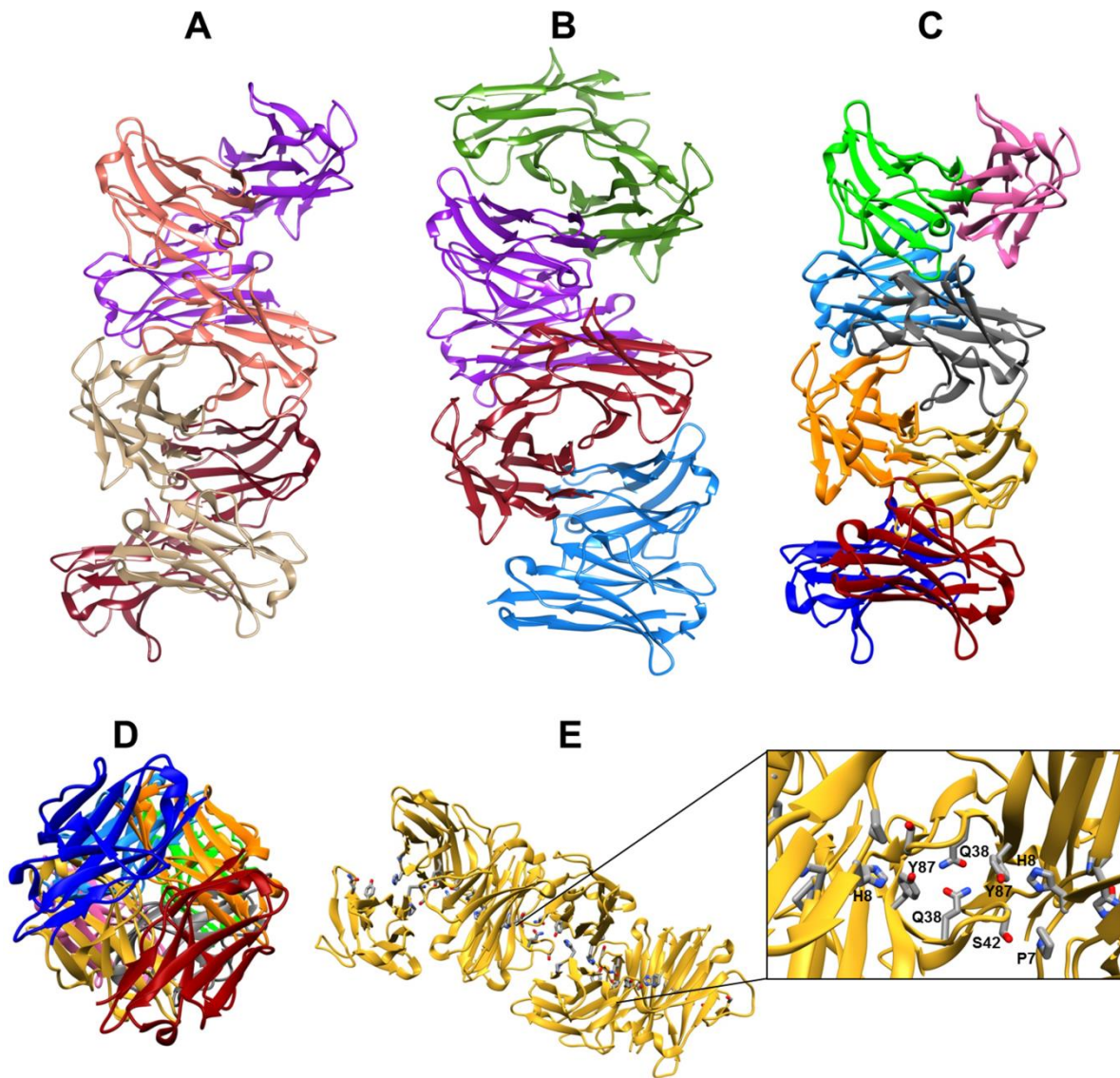


Figura 28. Octámeros. Las estructuras homo-octaméricas fueron generadas usando el programa Chimera UCSF siguiendo los parámetros de simetría del cristal reportado de **A.** 6aJL2 (PDB 2W0K), y **B.** AR (PDB 51R3). **C.** Octámero cristalográfico de 6a-R25G (PDB 5C9K). **D.** Eje transversal del octámero de 6a-R25G, y **E.** residuos localizados en el interior del hueco.

DISCUSIÓN

Amiloidosis AL λ VI

Muchos de nuestros esfuerzos en el laboratorio se han enfocado a explicar la alta asociación de la línea germinal $6a$ con la amiloidosis AL, visto todo desde una perspectiva termodinámica y estructural. Aun cuando se haya construido y expresado a 6aJL2 junto con muchas mutantes sencillas, el proceso de hipermutación somática incrementa el grado de dificultad ya que durante este proceso difícilmente se produce una región variable de cadena ligera con una sola mutación. Se requerirían más de dos mil mutantes sencillas y un número exorbitante de mutaciones en su combinatoria a lo largo de los 111 residuos que comprende esta proteína.

Para solventar estas disyuntivas experimentales se han tomado otras alternativas como, por ejemplo, el analizar otras líneas germinales tipo λ , específicamente tipo λ III (Villalba-Velázquez *et al.*, 2015). Una de las conclusiones del trabajo de Villalba *et al.* fue que 6aJL2 resultó la línea germinal menos estable comparada con las otras líneas germinales aun cuando esta línea germinal es más estable que las V_L provenientes de ella. En conjunto con los otros trabajos donde se han generado mutaciones puntuales de 6aJL2 (del Pozo-Yauner *et al.*, 2008; Hernández-Santoyo *et al.*, 2010), esto nos orilla a suponer que esta línea germinal tiene una alta tendencia a la disminución de la estabilidad termodinámica. Desde un punto de vista evolutivo, hay indicios de que la línea germinal $V_L\lambda 6$ es una de las líneas germinales ancestrales ya que se han reportado genes ortólogos compartidos entre humanos, ratones, y tiburones (Marchalonis *et al.*, 1998). Sabiendo que el rearrreglo de cadenas variables tipo κ ocurre antes que las tipo λ (Korsmeyer *et al.* 1981), y que a pesar que la línea germinal $6a$ es de bajo uso preferencial en paciente sanos pero que en estado libre tiene una alta probabilidad de derivar en amiloidosis AL (Perfetti *et al.*, 2002), podría plantearse que el uso de esta línea germinal estaría restringido.

Otro de los datos más interesantes de la caracterización de 6aJL2 fue que su cinética de fibrilogénesis *in vitro* llegó a la máxima velocidad de incorporación y el menor tiempo *lag* al incubarse a la concentración media de desnaturizante, apuntando a que esta proteína en un estado parcialmente desnaturizado es más susceptible a la agregación (Blancas-Mejía *et al.*, 2009). Si bien se ha constatado la presencia de posibles segmentos pro-fibrilogénicos en la secuencia de 6aJL2 usando predictores como TANGO (Villalba-Velázquez *et al.*, 2015), estos no tienen las características adecuadas para favorecer la agregación amiloide. Por ejemplo, en la construcción para la línea germinal $\lambda 3m$, 3mJL2, se encontraron secuencias pro-fibrilogénicas pero esta línea germinal no está implicada en la amiloidosis AL ni tampoco desarrolló fibras *in vitro* (Villalba-Velázquez *et al.*, 2015). En lo que a 6aJL2 concierne, el segmento pro-fibrilogénico detectado en el segmento pro-fibrilogénico mejor calificado está

localizado entre los residuos 35-40, varios de los cuales se ubican en los canales interiores de los octámeros (Figura 28).

Sin embargo, a pesar de todos estos análisis, es difícil extrapolar el efecto que una mutación puede tener en todo el dominio puesto que algunas mutaciones pueden ser compensatorias y otras desestabilizantes. Adicionalmente, analizar a 6aJL2 bajo la estrategia de mutaciones puntuales conllevaría un trabajo por demás arduo evaluando cada una de las dos mil variantes y, por demás extenuante, el análisis de la combinatoria global. Si bien, hay que recordar que toda esta área de investigación inició formalmente con el análisis termodinámico y mutagénico de una V_L derivada de paciente, relacionándose estos parámetros con experimentos de fibrillogénesis *in vitro* (Hurle *et al.*, 1994). Los resultados indican que una menor estabilidad termodinámica correspondía a una mayor tendencia a la formación de fibras, sugiriendo que esto podría servir como diagnóstico en la evolución de la enfermedad. Así como V_L REI, que es de isotipo κ I, muchas otras V_L s han sido caracterizadas para entender el por qué esta susceptibilidad a la agregación ordenada de fibras.

Consecuentemente, se seleccionó a AR, una V_L no caracterizada en el inicio de esta investigación con una característica en particular: la presencia de una glicina en la posición 25 la cual la hace análoga a la mutante 6a-R25G, una variante presumiblemente alotípica de 6aJL2. Otras V_L s tipo λ VI, como WIL y JTO, habían sido analizadas previamente pero éstas presentan una Arg25 en su secuencia (Wall *et al.*, 1999). Quedaba entonces la cuestión del papel que tenía esta mutación Arg25Gly en un contexto estructural.

AR, la contraparte de 6aJL2

Antes de analizar las implicaciones, comparaciones, o cualquier otra disertación derivadas del análisis termodinámico, fibrillogénico, o del análisis de redes, un dato no mencionado en la sección de resultados, porque sólo fueron resultados negativos pero que es necesario hacer hincapié, fue el intento de subclonar a AR y a sus mutantes del vector de baja expresión pSyn1 al vector de alta expresión pET22b para así solventar la cuestión de la baja expresión heteróloga. No se obtuvieron transformantes de *E. coli* aun utilizándose diferentes cepas como DH5 α , TG1, W3110, y XL-10Gold; así como tampoco de los procesos de subclonación ya que se usaron controles y se secuenció el vector pET22b. Únicamente usando el método PCR CloneJet de ThermoFischer se obtuvieron colonias; sin embargo, las secuencias mostraron pérdidas de nucleótidos entre el sitio de unión al ribosoma (rbs) y el péptido señal (pelB) ubicado río arriba de los sitios de clonación los sitios de clonación. No se pudo detectar proteína recombinante en *E. coli* BL21 o TG1 aun reduciendo la cantidad de inductor (IPTG). En conclusión, estos resultados indican que tanto AR como sus variantes pudieron haber resultado tóxicas para *E. coli*.

Volviendo a los resultados termodinámicos, AR es la V_L más inestable y la más fibrilogénica reportada hasta el momento. No es sorprendente, entonces, que el paciente haya fallecido en un tiempo significativamente menor al tiempo promedio después del diagnóstico considerando que la detección se haya realizado “oportunamente”. Si existiese un banco de valores termodinámicos de estas V_L s amiloidogénicas, se podría pronosticar un tiempo aproximado de supervivencia no considerando intervención médica como el trasplante de órganos afectados, el tratamiento con algún fármaco como heparán, o inclusive con quimioterapia (Gillmore & Hawkins, 2006).

Ahora bien, las estructuras secundarias y las estructuras cristalográficas de AR y de 6a-R25G mostraron que su estructuración era similar y que sus cadenas principales estaban altamente conservadas (Figuras 8 y 28); Cabe mencionar que estos resultados son frecuentes así como lo descrito en las mutantes de JTO (Wall *et al.*, 2004), las V_L s tipo κI AL-09 y AL-103 (DiCostanzo *et al.*, 2012), las variantes analizadas de la línea germinal λIII (Villalba-Velázquez, 2015), e inclusive con varias mutantes de 6aJL2 (Hernández-Santoyo *et al.*, 2010). Uno de los puntos más dramáticos durante este proyecto fue el rechazo de una hipótesis muy enraizada en el grupo de investigación sobre la modificación en la estructura secundaria del CDR1 propiciada por el residuo ubicado en la posición 25 (del Pozo-Yauner *et al.*, 2006). Este dato no sólo modificó la visión del papel del CDR1 sino que nos obligó a considerar los efectos compensatorios de la rotación de las cadenas laterales en un entorno global (Tabla 2).

Durante estos años de investigación se generaron dos maneras para entender a AR, desde una perspectiva teórica y otra experimental. Brevemente, la perspectiva teórica involucró la propuesta de reunir el análisis de redes con las dinámicas moleculares, así como su validación con proteínas caracterizadas pero completamente ajenas a una V_L (Luna-Martínez *et al.*, 2016). Los datos teóricos correspondieron en buena medida con los datos termodinámicos así como con la ubicación de segmentos desestructurados (Figuras 12 y 14). A pesar de que cada proteína llevaría su respectivo vasto análisis, el objetivo de esta metodología era analizar a la estructura de manera global sin necesidad de emplear otras aproximaciones como la incorporación de evolución de la comparación filogenética de la secuencia o un análisis del cambio de función. Esta metodología tiene ventajas a futuro como, por ejemplo, el análisis de una estructura de una proteína constituida por aminoácidos no comunes o con modificaciones post-traduccionales *sui generis* de un organismo (Liu & Schultz, 2010). De esta manera, si y sólo si esta proteína *de novo* contase con una estructura determinada experimentalmente, se podría analizar la influencia que cada nodo (residuo o nucleótido) está teniendo sobre la estructura.

Retornando al tema de la amiloidosis AL, el aplicar esta metodología en 6aJL2 y en AR, éstas mostraron diferencias no esperadas como que AR tuviera mayor número de interacciones en el CDR1 y una disminución de éstas en los extremos N-terminal y en el asa de prolina 40-60 (Figura 19). Aparentemente esto va en contrasentido de lo dicho para el CDR1; sin

embargo, este segmento estructural podría estar funcionando como segmento protector al impedir la exposición del núcleo hidrofóbico superior (Figura 20). Esto podría explicar los resultados termodinámicos del por qué la mutación Gly25Arg no se comportó acorde a lo pronosticado. En otras palabras, el resultado de la rigidización del CDR1 de AR podría explicar el por qué las variantes de AR con la mutación Gly25Arg no incrementaron la estabilidad termodinámica del dominio. Con ello, no sólo implicaría que AR tuviera un ambiente químico diferente en el CDR1 con respecto al de 6aJL2, sino también que la aparente rigidización impidiera un correcto acomodo de la arginina.

Los resultados termodinámicos de las mutantes del núcleo hidrofóbico de AR indicaron que este dominio contiene un núcleo perturbado por un anillo aromático (Tabla 6). Sin embargo, considerando que AR es una V_L producto de los procesos de maduración de una célula B, hubiese sido interesante conocer la V_H contraparte que le permitiese contrarrestar los efectos de esta baja estabilidad. Además, conociendo la V_H y la V_L , hasta se podría indagar qué antígeno o antígenos serían capaces de reconocer. Ahora bien, aun cuando AR es la variante más inestable, la estructura de AR *per se* muestra interacciones compensatorias repartidas en toda la estructura con respecto a 6aJL2.

Las cinéticas de fibrilogénesis de las mutantes de AR tuvieron muy buena relación con los datos termodinámicos pero las microscopías electrónicas mostraron un aumento en la población de agregados amorfos (Figura 25), sugiriendo que la estabilización estructural de AR estaría modificando las rutas de agregación. Éste es un dato que podría parecer evidente; sin embargo, en el caso de las mutantes de la línea germinal 3rJL2, las fibras dejaron de ser las típicas fibras rectas mostrando una curvatura no reportada previamente (Villalba *et al.*, 2015). Sabiendo que una mutación podría diversificar las propiedades estructurales de una fibra además de las propiedades termodinámicas de la variante, se ha propuesto que los amiloides pueden ser utilizados como biomateriales (Peralta *et al.*, 2015).

Convergencia de las estructuras octaméricas

Contrastando la rigidización del CDR1, los segmentos que se flexibilizaron en AR (Figura 19) fueron segmentos descritos como importantes para evitar la agregación intermolecular como el extremo N-terminal y el asa de prolina 40-60 (Richardson & Richardson, 2001). Cristalográficamente, también se encontraron cambios en el extremo N-terminal ya que la longitud de las hebras $\beta 1$ y $\beta 3$ aumentó en AR (Figura 26). Cambios mínimos podrían desencadenar cambios sustanciales a nivel de interacción intermolecular. Al generar estructuras dimericas de 6aJL2, 6a-R25G, y de AR, el área de interfaz también resultó diferente (Figura 27). Interesantemente, la convergencia de las estructuras octaméricas de 6aJL2 y AR con el octámero cristalográfico de 6a-R25G sugeriría fuertemente una ruta de agregación en común para las V_L s λVI (Figura 28).

La relevancia de tener un octámero cristalográfico, como el de 6a-R25G, abre las puertas a generar estrategias racionales que permitan explicar los mecanismos de reconocimiento celular y los mecanismos plausibles una vez que los oligómeros se inserten en la membrana. Se ha demostrado que los oligómeros pueden inducir toxicidad celular a través de la generación de un tipo de canales iónicos (Demuro *et al.*, 2005; Kagan *et al.*, 2011). Conjuntamente, la interacción podría ser dependiente del tipo de receptores celulares o de la composición de las balsas lipídicas específicas (Morante *et al.*, 2015). Aunado a estos experimentos, hay evidencia que muestra que la internalización de la V_L es dependiente del tipo de V_L en cuestión y que, a su vez, disparan diferentes procesos de internalización celular y transformaciones fenotípicas (Keeling *et al.*, 2004; Teng *et al.*, 2014). Desde un punto de vista personal, la investigación relacionada con la amiloidosis AL debiera estar ahora más enfocada a los efectos celulares que generan estas V_L para así generar prognosis más acertadas así como también terapias personalizadas.

Otros aspectos de las proteínas amiloidogénicas

Contrario a lo señalado en este trabajo acerca de que las proteínas amiloidogénicas están excepcionalmente involucradas en enfermedades, está la contraparte de aquellos amiloides con funcionalidad biológica los cuales están distribuidas en muchos organismos conocidas como el amiloma (Chiti & Dobson, 2006; Goldschmidt *et al.*, 2010). En los humanos está descrita la proteína Pmel17, la cual está involucrada en la disminución del efecto de varios procesos citotóxicos como el daño por UV y el estrés oxidativo (Fowler *et al.*, 2006). Claro está que la biosíntesis de esta proteína es un proceso muy controlado ya que sus precursores son altamente citotóxicos (Fowler *et al.*, 2007). Entonces, las proteínas amiloides también tendrían un objetivo biológico ya que, evolutivamente, podrían considerarse como “las ovejas negras” de las proteínas.

Para finalizar, una de las cuestiones que siempre quedaron en duda era la utilidad biológica del péptido β -amiloide, el cual está directamente involucrado en el Alzheimer, ya que se desconocía la función biológica positiva. En fechas recientes, este péptido se relacionó con dos efectos biológicos *in vivo* muy importantes para los organismos; el primero a mencionar es la relación de este péptido con el sueño y el segundo es su relación como péptido antimicrobiano. Se analizó la generación de este péptido en ratones el cual resultó constitutivo durante el día, fungiendo como un supuesto neurotransmisor; pero que, durante la noche, las neuronas se compactan para que el líquido intersticial funja como líquido limpiador eliminando los desechos neuronales (Xie *et al.*, 2013). En otro trabajo en el cual usando ratones, el nemátodo *Caenorhabditis elegans*, y cultivos celulares, se observó un efecto antimicrobiano contra la meningitis ocasionada por *Salmonella entérica* serotipo Typhimurium (Kumar *et al.*, 2016).

Conjuntando todo lo mencionado acerca de las propiedades antimicrobianas de los amiloides, los diversos mecanismos de agregación, la alta regulación celular para producir amiloides funcionales, y las modificaciones en la geometría de las fibras influenciada por los efectos termodinámicos, da la impresión de que las proteínas amiloides tienen una alta versatilidad pero que deben ser controladas ya que pueden causar enfermedades crónico-degenerativas.

BIBLIOGRAFÍA

- Abdi, H., and Williams, L.J. (2010) Principal component analysis. *WTREs Computational Statistics*. **2**: 433-459. DOI: 10.1002/wics.101
- Abraham, R.S., Geyer, S.M., Price-Troska, T.L., Allmer, C., Kile, R.A., Gertz, M.A. and Fonseca, R. (2003) Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain-associated amyloidosis (AL). *Blood* **101**: 3801 –3808. DOI: 10.1182/blood-2002-09-2707
- Adams, P.D., Afonine, P.V., Bunkóczi, V.B., Chen, I.W., Davis, N., Echols, J.J., Headd, L.W., Hung, G.J., Kapral, R.W., Grosse-Kunstleve, A.J., McCoy, N.W., Moriarty, R., Oeffner, R.J., Read, D.C., Richardson, J.S., Richardson, T.C., Terwilliger, T.C., and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213-221. DOI: 10.1107/S0907444909052925
- Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol Crystallogr* **68**: 352-367. DOI: 10.1107/S0907444912001308
- Albert, R., Jeong, H., and Barabasi, A. (2000) Error and attack tolerance of complex networks *Nature*. **406**: 378 – 382. DOI: 10.1038/35019019
- Banci, L., Bertini, I., Calderone, V., Cramaro, F., Del Conte, R., Fantoni, A., Mangani, S., Quattrone, A., and Viezzoli, M.S. (2004) A prokaryotic superoxide dismutase paralog lacking two Cu ligands: from largely unstructured in solution to ordered in the crystal. *Proc Natl Acad Sci U.S.A.* **102**: 7541-7546. DOI: 10.1073/pnas.0502450102
- Benjamini, E., Coico, R., Sunshine, G. (2000) Immunology A Short Course. 4th Edition John Wiley & Sons Inc. ISBN 0-471-34890-2
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak JR. (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**: 3684–3690 DOI: 10.1063/1.448118
- Berjanskii, M.V., and Wishart, D.S. (2013) A simple method to measure protein side-chain mobility using NMR chemical shifts. *J Am Chem Soc.* **135**: 14536-14539. DOI: 10.1021/ja407509z
- Blau, N., van Spronsen, F.J., and Levy, H.L. (2010) Phenylketonuria. *Lancet*. **376**: 1417-14127. DOI: 10.1016/S0140-6736(10)60961-0.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D.-U. (2006) Complex networks: Structure and dynamics. *Physics Reports*. **424**: 175-308. DOI: 10.1016/j.physrep.2005.10.009
- Bonacich, P. (1972) Factoring and Weighting approaches to status scores and clique identification. *J Math Sociol* **2**: 113 – 120. DOI: 10.1080/0022250X.1972.9989806
- Bond, C.S., and Schüttelkopf, A.W. (2009) ALINE: a WYSIWYG protein sequence alignment editor for publication-quality alignments. *Acta Crystallogr. D Biol. Crystallogr.* **65**: 510–512. DOI: 10.1107/S09074449090007835

- Campos, L.A., Bueno, M., Lopez-Llano, J., Jiménez, M.A., and Sancho, J. (2004a) Structure of stable protein folding intermediates by equilibrium ϕ -analysis: the apoflavodoxin thermal intermediate. *J Mol Biol.* **344**: 239-255. DOI: 10.1016/j.jmb.2004.08.081
- Campos, L.A., Garcia-Mira, M.M., Godoy-Ruiz, R., Sanchez-Ruiz, J.M., and Sancho J. (2004b) Do proteins always benefit from a stability increase? Relevant and residual stabilization in a three-state protein by charge optimization. *J Mol Biol.* **344**: 223-237. DOI: 10.1016/j.jmb.2004.09.047
- Chandrashekar, I.R., Mohanty, B., Linossi, E.M., Dagley, L.F., Leung, E.W.W., Murphy, J.M., Babon, J.J., Nicholson, S.E., and Norton, R.S. (2015) Structure and functional characterization of the conserved JAK interaction region in the intrinsically disordered N-terminus of SOCS5. *Biochemistry.* **54**: 4672-4682. DOI: 10.1021/acs.biochem.5b00619
- Chen, M.E., Cang, H.X., and Nymeyer, H.Dr. (2007) Nymeyer's Group, Inst. Mol. Biol., Florida State University. Available online at <http://noch.sourceforge.net> (accessed 27 Nov 2015).
- Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**: 12-21. DOI: 10.1107/S0907444909042073
- Chiti, F., and Dobson, C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**: 333-366. DOI: 10.1146/annurev.biochem.75.101304.123901
- Chothia, C., Gelfand, I., and Kister, A. (1998) Structural determinants in the sequences of immunoglobulin variable domain. *J Mol Biol.* **278**: 457 – 479. DOI: 10.1006/jmbi.1998.1653
- Clark, L.A., Ganesan, S., Papp, S., and van Vlijmen, H.W.T. (2006) Trends in antibody sequence changes during the somatic hypermutation process. *J Immun.* **177**: 333–340. DOI: 10.4049/jimmunol.177.1.333
- Collaborative Computational Project 4. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **50**: 760-763. DOI: 10.1107/S0907444994003112
- Costa, L.F., Rodrigues, F.A., Travieso, G., and Villas-Boas, P.R. (2007) Characterization of complex networks: a survey of measurements. *Advances in Physics.* **56**: 167-242. DOI: 10.1080/00018730601170527
- Darby, N.J., and Creighton TE. (1993) Protein structure. Oxford: IRL Press at Oxford University Press. p. 4. ISBN 019963310X
- Deegan, B.J., Seldeen, K.L., McDonald, C.B., Bhat, V., and Farooq, A. (2010) Binding of the ER α nuclear receptor to DNA is coupled to proton uptake. *Biochemistry.* **49**: 5978-5988. DOI: 10.1021/bi1004359
- del Pozo-Yauner, L., Ortiz, E., Sánchez, R., Sánchez-López, R., Güerrecá, L., Murphy, C.L., Allen, A., Wall, J.S., Fernández-Velasco, D.A., Solomon, A., and Becerril, B. (2008) Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains. *Proteins.* **72**: 684-692. DOI: 10.1002/prot.21934.
- del Pozo-Yauner, L., Wall, J.S., González-Andrade, M., Sánchez-López, R., Rodríguez-Ambríz, S.L., Pérez-Carreón, J.I., Ochoa-Leyva, A., and Fernández-Velasco, D.A. (2014) The N-terminal strand

- modulates immunoglobulin light chain fibrillogenesis. *Biochem Biophys Res Commun.* 443: 495-499. DOI: 10.1016/j.bbrc.2013.11.123
- Demetrius, L., and Manke, T. (2004) Robustness and network evolution – an entropic principle. *Physica A* **346**: 682-696. DOI: 10.1016/j.physa.2004.07.011
- Demuro A., Mina E., Kaye R., Milton S.C., Parker I., and Glabe C.G. (2005) Calcium dysregulation and membrane disruption as a ubiquitous neurotoxic mechanism of soluble amyloid oligomers. *J Biol. Chem.* **280**: 17294 – 17300. DOI: 10.1074/jbc.M500997200
- DiCostanzo, A.C., Thompson, J.R., Peterson, F.C., Volkman, B.F., and Ramirez-Alvarado, M. (2012) Tyrosine residues mediate fibril formation in a dynamic light chain dimer interface. *J. Biol. Chem.* **33**: 27997-28006. DOI: 10.1074/jbc.M112.362921
- Dror, R., Mirks, R.M., Grossman, J.P., Huafe, X., and Shaw, D.E. (2012) Biomolecular simulation: A computational microscope for Molecular Biology. *Annu Rev Biophys.* **14**: 429-452. DOI: 10.1146/annurev-biophys-042910-155245
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucl Acids Res.* **34**: W116-W118. DOI: 10.1093/nar/gkl282
- Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research.* **34**: W116-W118. DOI: 10.1093/nar/gkl282.
- Emsley, P., and Cowtan, K.. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60: 2126-2132. DOI: 10.1107/S0907444904019158
- Essman, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen LG. (1995) A smooth particle mesh Ewald method. *J Chem Phys* **103**: 8577–85933. DOI: 10.1063/1.470117
- Fan, X., Kurgan, L. (2014) Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *J Biomol Struct Dyn.* **32**: 448-464. DOI: 10.1080/07391102.2013.775969
- Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotech.* **22**: 1302-1306. DOI: 10.1016/j.jmb.2008.06.062.
- Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., and Kelly, J.W. (2006) Functional amyloid formation within mammalian tissue. *PLoS Biology* **4**: e6. DOI: 10.1371/journal.pbio.0040006.
- Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.F. (2007) Functional amyloids – from bacteria to humans.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005) Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. Humana Press. pp. 571–607. DOI: 10.1385/1-59259-890-0:571
- Gillmore, J.D., and Hawkins, P.N. (2006) Drug insight: emerging therapies for amyloidosis. *Nat Clin Pract* **2**: 268-270. DOI:10.1038/ncpneph0169

- Goldman, M.J., Anderson, G.M., Stolzenberg, E.D., Kari, U.P., Zasloff, M., and Wilson J.M. (1997) Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* **88**: 553–560. DOI: 10.1016/S0140-6736(09)60327-5
- Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *Proc Natl Acad Sci USA* **10**: 3487-3492. DOI: 10.1073/pnas.0915166107
- González-Andrade, M., Becerril-Luján, B., Sánchez-López, R., Ceceña-Álvarez, H., Pérez-Carreón, J.I., Ortiz-Surí, E., Fernández-Velasco, D.A., and del Pozo-Yauner, L. (2013) Mutational and genetic determinants of λ 6 light chain amyloidogenesis. *FEBS J.* **280**: 6173-6183. DOI: 10.1111/febs.12538
- Gregersen, N. (2006) Protein misfolding disorders: Pathogenesis and intervention. *J Inherit Metab Dis.* **29**: 456-470. DOI: 10.1007/s10545-006-0301-4
- Gribenko, A.V., and Makhatadze, G.I. (2007) Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J Mol Biol.* **366**: 842-856. DOI: 10.1016/j.jmb.2006.11.061
- Guex, N., and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**: 2714–2723. DOI: 10.1002/elps.1150181505
- Hernández-Santoyo, A., del Pozo-Yauner, L., Fuentes-Silva, D., Ortiz, E., Rudiño-Piñera, E., Sánchez-López, R., Horjales, E., Becerril, B., and Rodríguez-Romero, A. (2010) A single mutation at the sheet switch region results in conformational changes favoring lambda6 light-chain fibrillogenesis. *J Mol Biol.* **396**: 280–292. DOI: 10.1016/j.jmb.2009.11.038
- Horn, R.A. and Johnson, C.R. (1990) Matrix Analysis. Cambridge University Press. ISBN: 9780521386326
- Hurle, M.R., Helms, L.R., Li, L., Chan, W., and Wetzel, R. (1994) A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci USA.* **91**: 5446:5450. DOI: 10.1073/pnas.91.12.5446
- Husby G., Natvig J.B., and Sletten K. (1974) New, third class of amyloid fibril protein. *The Journal of Experimental Medicine.* **139**: 773 – 778.
- Jamroz, M., Kolinski, A., and Kihara, D. (2012) Structural features that predict real-value fluctuations of globular proteins. *Proteins: Structure, Function, and Bioinformatics.* **80**: 1425-1435. DOI: 10.1002/prot.24040
- Janeway, C.H., Travers, P., Walport, M., and Shlomchik, M. (2001) Immunobiology The immune System in Health and Disease. 5th Edition Garland Publishing. ISBN-10: 0-8153-3642-X
- Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991) Sequences of proteins of Immunological Interest, Vol. 5th, National Institutes of Health, Bethesda, MD. ISBN: 094137565X
- Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol Crystallogr* **66**: 125-132 DOI: 10.1107/S0907444909047337
- Kagan B.L., Jang H., Capone R., Teran-Arce F., Ramachandran S., Lal R., and Nussinov R. (2012) Antimicrobial properties of amyloid peptides. *Mol. Pharm.* **9**: 708-717. DOI: 10.1021/mp200419b

- Kantardjieff, K.A., and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA and protein-nucleic acid complex crystals. *Protein Science*. **12**: 1865–1871 DOI: 10.1110/ps.0350503
- Keeling, J., Teng, J., and Herrera, G.A. (2004) AL-amyloidosis and light-chain deposition disease light chains induce divergent phenotypic transformations of human mesangial cells. *Lab. Inv.* **84**: 1322-1338. DOI: 10.1038/labinvest.3700161.
- Knowles, T.P.J., Vendruscolo, M., and Dobson, C.M. (2014) The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol.* **15**:384-96. DOI: 10.1038/nrm3810.
- Korsmeyer, S.J., Hieter, P.A., Ravetch, J.V., Poplack, D.G., Waldmann, T.A., and Leder, P. (1981) Developmental hierarchy of immunoglobulin gene rearrangements in human leukemic pre-B-cells. *Proc Natl Acad Sci USA.* **78**: 7096-7100. DOI: 10.1073/pnas.78.11.7096
- Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol.* **372**: 774 – 797. DOI: 10.1016/j.jmb.2007.05.022.
- Kronenberg, P. (2007) Applied multiway data analysis – Appendix B Biplots and their interpretation. John Wiley & Sons, Inc. DOI: 10.1002/9780470238004
- Kumar, D.K.V., Choi, S.H., Washicosky, K.J., Eimer, W.A., Tucker, S., Ghofrani, J., Lefkowitz, A., McColl, G., Goldstein, L.E., Tanzi, R.E., and Moir, R.D. (2016) Amyloid- β peptide protects against microbial infection in mouse and worm models of Alzheimer’s disease. *Science Transl Med* **8**: 340ra72. DOI: 10.1126/scitranslmed.aaf1059.
- Lakowicz, J.R. (2006). Principles of fluorescence spectroscopy (3rd ed.). New York, NY: LLC, Springer Science + Business Media. DOI: 10.1007/978-0-387-46312-4
- Leach, S. (1969) Physical principles and techniques of protein chemistry, Part A. Chapter 4. Fluorescence of Proteins. pp. 225. Academic Press, New York. ISBN: 978-0-12-440101-3
- Lefranc, M.P. (2001a) Nomenclature of the Human Immunoglobulin Kappa (IG κ) genes. *Exp Clin Immun.* **18**: 161 – 174. DOI: 10.1159/000049195
- Lefranc, M.P. (2001b) Nomenclature of the Human Immunoglobulin Lambda (IG λ) genes. *Exp Clin Immun.* **18**: 242 – 254. DOI: 10.1159/000049203
- LeVine, H.III. (1993) Thioflavine T interaction with synthetic Alzheimer’s disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci.* **2**: 404–410. DOI: 10.1002/pro.5560020312
- Li J., Wang, J., and Wang, W (2008) Identifying folding nucleus based on residue contact networks of proteins. *Proteins* **71**: 1899–1907. DOI: 10.1002/prot.21891
- Liu, C.C., and Schultz, P.G. (2010) Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**: 413-444. DOI: 10.1146/annurev.biochem.052308.105824
- Luna-Martínez, O.D. (2009) Estudio de las bases estructurales y moleculares asociadas a las propiedades biofísicas y fibrilógicas *in vitro* de la rV_L AR. (Tesis de Maestría) Universidad Nacional Autónoma de México.

- Luna-Martínez, O.D., Vidal-Limón, A., Villalba-Velázquez, M.I., Sánchez-Alcalá R, Garduño-Juárez R, Uversky VN, & Becerril B. (2016) Simple approach for ranking structure determining residues. *PeerJ*. **4**:e2136. DOI: 10.7717/peerj.2136
- Mai, H.L., Sheikh-Hamad, D., Herrera, G.A., Gu, X., and Truong, L.D. (2003) Immunoglobulin heavy chain can be amyloidogenic: morphologic characterization including immunoelectron microscopy. *Am J Surg Pathol*. **27**: 541-545.
- Makin, O.S. and Serpell, L.C. (2005) Structures for amyloid fibrils. *FEBS J*. **272**: 5950–5961. DOI: 10.1111/j.1742-4658.2005.05025.x
- Manke T., Demetrius L., and Vingron, M. (2005) Lethality and Entropy of Protein Interaction Networks. *Genome Inform* **16**: 159 – 163.
- Manke T., Demetrius L., and Vingron, M. (2006) An entropic characterization of protein interaction networks and cellular robustness. *J R Soc Interface* **3**: 843 – 850. DOI: 10.1098/rsif.2006.0140
- Marchalonis, J.J., Schluter, S.F., Bernstein, R.M., and Hohman, V.S. (1998) Antibodies of sharks: revolution and evolution. *Immu Rev*. **166**: 103-122. DOI: 10.1111/j.1600-065X.1998.tb01256.x
- Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*. **33**: 491–497 DOI: 10.1016/0022-2836(68)90205-2
- Maya-Martinez, R., Gil-Rodriguez, P., and Amero, C. (2015) Solution structure of 6aJL2 and 6aJL2-R24G amyloidogenic light chain proteins. *Biochem Biophys Res Commun*. **456**: 695 – 699. DOI: 10.1016/j.bbrc.2014.12.044
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674. DOI: 10.1107/S0021889807021206
- Meffre, E., Casellas, R., and Nussenzweig, M.C. (2000) Antibody regulation of B cell development. *Nat Immunol*. **1**: 379-385. DOI: 10.1038/80816
- Meng, X., Fink, A.L., and Uversky, V.N. (2008) The effect of membranes on the *in vitro* fibrillation of an amyloidogenic light-chain variable-domain SMA. *J Mol Biol*. **381**: 989-999. DOI: 10.1016/j.jmb.2008.06.062
- Merlini, G., and Stone, M.J. (2006) Dangerous small B-cell clones. *Blood*. **108**: 2520-2530. DOI: 10.1182/blood-2006-03-001164
- Milardi, D., la Rosa, C., Fasone, S., and Grasso D (1997) An alternative approach in the structure-based predictions of the thermodynamics of protein unfolding. *Biophys Chem* **69**: 43-51. DOI: 10.1016/S0301-4622(97)00071-9.
- Milla, M.E., Brown, B.M., and Sauer, R.T. (1994) Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nat Structl & Mol Biol*. **1**:518-523. DOI: 10.1038/nsb0894-518
- Mirny L., and Domany E. (1996). Protein fold recognition and dynamics in the space of contact maps. *Amino Acids*. **410**: 391-410. DOI: 10.1002/(SICI)1097-0134(199612)26:4<391::AID-PROT3>3.0.CO;2-F
- Miyazaki, D., Yazaki, M., Gono, T., Kametani, F., Tsuchiya, A., Matsuda, M., Takenaka, Y., Hosh, Y.2nd, and Ikeda, S. (2008) AH amyloidosis associated with an immunoglobulin heavy chain variable region (VH1) fragment: a case report. *Amyloid*. **15**: 125-128. DOI: 10.1080/13506120802006229.

- Monson, N.L., Dörner, T., and Lipsky, P.E. (2000) Targeting and selection in human V λ rearrangements. *European Journal of Immunology*. **30**: 1597 – 1605
- Montaigne, M. (2015) Ensayos sobre educación / Michel de Montaigne. Capítulo XIII, pp. 156. Edición de Joan Lluís Llinàs; traducido por Ediciones Cátedra. Edición digital. Madrid: Biblioteca Nueva. ISBN 8416170274
- Morante K., Caaveiro J.M.M., Tanaka K., González-Mañas J.M., and Tsumoto K. (2015) A pore-forming toxin requires a specific residue for its activity in membranes with particular physicochemical properties. *J Biol. Chem.* **290**: 10850 – 10861. DOI: 10.1074/jbc.M114.615211
- Mori, T., and Okamoto, Y. (2009) Folding simulations of gramicidin A into the β -helix conformations: Simulated annealing molecular dynamics study. *J Cheml Phys* **131**: 165103. DOI: 10.1063/1.3247578
- Muppirala, U.K., Li, Z. (2006) A simple approach for protein structure discrimination based on the network pattern of conserved hydrophobic residues. *Protein Eng Des Sel.* **19**: 265 – 275. DOI: 10.1093/protein/gzl009
- Nelson, R., Eisenberg, D. (2006) Recent atomic models of amyloid fibril structure. *Curr Opi Struct Biol.* **16**: 260 – 265. DOI: 10.1016/j.sbi.2006.03.007
- Newman, M.E.J. (2004) Analysis of weighted networks. *Phys Rev E* **70**: 056131. DOI: 10.1103/PhysRevE.70.056131
- Newman, M.E.J. (2008) The mathematics of networks. The New Palgrave Encyclopedia of Economics. Blume L.E. & Durlaf S.N. Palgrave Macmillan
- Nicholson, S.E., and Norton, R.S. (2015) Structure and functional characterization of the conserved JAK interaction region in the intrinsically disordered N-Terminus of SOCS5. *Biochemistry.* **54**:4672–4682. DOI: 10.1021/acs.biochem.5b00619
- Nienhuis, H.L.A., Bijzet, J., and Hazenberg, B.P.C. (2016) The prevalence and management of systemic amyloidosis in Western countries. *Kidney Dis.* **2**: 10-19. DOI: 10.1159/000444206
- O’Sullivan, B.P., and Freedman, S.D. (2009) Cystic fibrosis. *Lancet.* **373**: 1891-1904. DOI: 10.1016/S0140-6736(09)60327-5
- Pace, C.N., Shirley, B.A., and Thompson, J.A. (1989) Measuring the conformational stability of a protein. In Protein Structure: A Practical Approach (Creighton, T. E., ed) pp. 311–330, IRL Press, Oxford, UK
- Peng, K., Vucetic, S., Radivojac, P., Brown, C.J., Dunker, A.K., and Obradovic, Z. (2005) Optimizing long intrinsic disorder predictors with protein evolutionary information. *J Bioinf Comp Biol.* **3**: 35-60. DOI: 10.1142/S0219720005000886
- Peng, Z.L., and Kurgan, L. (2012) Comprehensive comparative assessment of in-silico predictors of disordered regions. *Curr Prot Pept Sci.* **13**: 6-18. DOI: 10.2174/138920312799277938
- Pepys, M.B. (2006) Amyloidosis. *Annu Rev Med.* **57**: 223-241. DOI: 10.1146/annurev.med.57.121304.131243
- Peralta, M.D., Karsai, A., Ngo, A., Sierra, C., Fong, K.T., Hayre, N.R., Mirzaee, N., Ravikumar, K.M., Kluber, A.J., Chen, X., Liu, G.Y., Toney, M.D., Singh, R.R., and Cox, D.L. (2015) Engineering amyloid

- fibrils from β -solenoid proteins for biomaterials applications. *ACS Nano* **27**: 449-463. DOI: 10.1021/nn5056089
- Perfetti, V., Casarini, S., Palladini, G., Vignarelli, M.C., Klersy, C., Diegoli, M., Ascari, E., and Merlini, G. (2002) Analysis of $V\lambda$ - $J\lambda$ expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r(III) as a new amyloid-associated germline gene segment. *Blood*. **100**: 948-953. DOI: 10.1182/blood-2002-01-0114
- Perfetti, V., Ubbiali, P., Vignarelli, M.C., Diegoli, M., Fasani, R., Stoppini, M., Lisa, A., Mangione, P., Obici, L., Arbustini, E., and Merlini, G. (1998) Evidence that amyloidogenic light chains undergo antigen-driven selection. *Blood*. **91**: 2948-2954.
- Perl, D., and Schmid, F.X. (2001) Electrostatic stabilization of a thermophilic cold shock protein. *J Mol Biol*. **313**: 343-357. DOI: 10.1006/jmbi.2001.5050
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. **25**: 1605-1612. DOI: 10.1002/jcc.20084
- Pieper, K., Grimbacher, B., and Eibel, H. (2013) B-cell biology and development. *J Allergy Clin Immunol*. **131**: 959-971. DOI: 10.1016/j.jaci.2013.01.046
- Piziak, R., and Odell, P.L. (2006) Matrix Theory: From Generalized Inverses to Jordan Form. Taylor & Francis Inc.
- Prodromou, C., and Pearl, L.H. (1992) Recursive PCR: a novel technique for total gene synthesis. *Prot Eng* **5**: 827-829. DOI: 10.1093/protein/5.8.827
- Puchtler, H. and Sweat, F. (1965) Congo red as a stain for fluorescence microscopy of amyloid. *J Histochem Cytochem*. **13**: 693-694. DOI: 10.1177/13.8.693
- R Core Development Team (2013) R: A language and environment for statistical computing Vienna, Austria. ISBN: 3-900051-07-0. Available online at <http://R-project.org> (accessed 27 Nov 2015).
- Richardson, J.S., and Richardson, D.C. (2001) Natural β -sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc Natl Acad Sci of U.S.A.* **99**: 2754-2759. DOI: 10.1073/pnas.052706099.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467.
- Schier, R., Marks, J.D., Wolf, E.J., Apell, G., Wong, C., McCartney, J.E., Bookman, M.A., Huston, J.S., Houston, L.L., Weiner, L.M., and Adams, G.P. (1995) In vitro and in vivo characterization of a human anti-c-e-erbB-e single-chain Fv isolated from a filamentous phage antibody library. *Immunotechnology* **1**: 73-81. DOI: 10.1016/1380-2933(95)00007-0
- Schroeder, H.W., and Cavacini, L. (2010) Structure and function of immunoglobulins. *J Allergy Clin Immunol*. **125**: S41-S52. DOI: 10.1016/j.jaci.2009.09.046
- Shapiro, S.S.; and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*. **52**: 591-611. DOI:10.1093/biomet/52.3-4.591
- Sletten, K., Natvig, J.B., Husby, G. and Juul, J. (1981) The complete amino acid sequence of a prototype immunoglobulin- λ light-chain-type amyloid-fibril protein AR. *The Biochemical Journal*. **195**: 561-572. DOI: 10.1042/bj1950561.

- Solomon, A., and Weiss, D.T. (1995) Structural and functional properties of human λ -light chain variable-region subgroups. *Clin Diagn Lab Immunol.* **2**: 387-394. DOI:
- Solomon, A., Frangione, B., and Franklin, E.C. (1982) Bence Jones proteins and light chains of immunoglobulins: Preferential association of the $V_{\lambda VI}$ subgroup of human light chains with amyloidosis AL(λ). *J Clin Invest.* **70**: 453-460. DOI: 10.1172/JCI110635
- Sumathi, K., Ananthakshmi, P., Roshan, M.N.A., and Sekar, K. (2006) 3dSS: 3D structural superposition. *Nucl Acids Res* **34**: W128-W132. DOI: 10.1093/nar/gkl036
- Teng, J., Turbat-Herrera, E.A., and Herrera, G.A. (2014) Extrusion of Amyloid Fibrils to the Extracellular Space in Experimental Mesangial AL-Amyloidosis: Transmission and Scanning Electron Microscopy Studies and Correlation with Renal Biopsy Observations. *Ultrastructural Pathology.* **38**: 104-115. DOI: 10.3109/01913123.2013.861568.
- The International Myeloma Working Group (2003) Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *British Journal of Haematology.* **121**: 749–757. DOI: 10.1046/j.1365-2141.2003.04355.x
- Tina, K.G., Bhadra, R., and Srinivasan, N. (2007) PIC: protein interactions calculator. *Nucleic Acids Res* **35**: W473-W476. DOI: 10.1093/nar/gkm423
- Tycko, R. (2006) Molecular structure of amyloid fibrils: insights from solid-state NMR. *Quarterly Reviews of Biophysics.* **39**: 1-55. DOI: 10.1017/S0033583506004173
- Van den Bedem, H., Bhabha, G., Yan, K., Wright, P.E., and Fraser, J.S. (2013) Automated identification of functional dynamic contact networks from X-ray crystallography. *Nat Methods.* **10**: 896-902. DOI: 10.1038/nmeth.2592
- Vendruscolo, M., Dokholyan, N.V., Paci, E., and Karplus, M. (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys Rev E Stat Nonlin Soft Matter Phys.* **65**: 06910-1 – 06910-4. DOI: 10.1103/PhysRevE.65.061910
- Voss, N.R., and Gerstein, M. (2005) Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *J Mol Biol* **346**: 477–492. DOI: 10.1016/j.jmb.2004.11.072
- Wall, J.S., Gupta, V., Wilkerson, M., Schell, M., Loris, R., Adams, P., Solomon, A., Stevens, F., and Dealwis, C. (2004) Structural basis of light chain amyloidogenity: comparison of the thermodynamic properties, fibrillogenic potential and tertiary structural features of four $V_{\lambda 6}$ proteins. *J. Mol. Recog.* **17**: 323-331. DOI: 10.1002/jmr.681
- Walters, J., Milan, S., and Clark, A.C. (2009) Practical approaches to protein folding and assembly: Spectroscopic strategies in thermodynamics and kinetics. *Methods in Enzymology.* **455**:1-39. DOI: 10.1016/S0076-6879(08)04201-8.
- Watts, D.J., and Strogatz, S.H. (1998) Collective dynamics of ‘small-world’ networks. *Nature.* **393**: 440-442. DOI: 10.1038/30918
- Weiss, B.M., Wong, S.W., and Comenzo, R.L. (2016) Beyond the plasma cell: emerging therapies for immunoglobulin light chain amyloidosis. *Blood.* **127**: 2275-2280. DOI: 10.1182/blood-2015-11-681650.

Wunderlich, M., Martin, A., Schmid, F.X. (2005) Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions. *J Mol Biol.* **347**: 1063-1076. DOI: 10.1016/j.jmb.2005.02.014

Wunderlich, M., Schmid, F.X. (2006) The correlation between protein stability and dipole moment: a critical test. *Prot Eng Des Sel.* **19**: 355-358. DOI: 10.1093/protein/gzl019

Xie, L., Hang, H., Xu, Q., Chen, M.J., Liao, Y., Thiyagarajan, M., O'Donnell, J., Christensen, D.J., Nicholson, C., Iliff, J.J., Takano, T., Deane, R., and Nedergaard, M. (2013) Sleep drives metabolite clearance from the adult brain. *Science* **342**: 373-377. DOI: 10.1126/science.1241224.

Zurflüh, M.R., Zschocke, J., Lindner, M., Feillet, F., Chery, C., Burlina, a., Stevens, R.C., Thöny, B., and Blau, N. (2008) Molecular genetics of tetrahydrobiopterin responsive phenylalanine hydroxylase deficiency. *Hum Mutat.* **29**: 167–175. DOI: 10.1002/humu.20637

Anexo A. Jerarquización de los residuos de 6aJL2. La numeración de los residuos está en base a Chothia *et al.*,1998. Al valor más alto de SDRI le corresponde la posición más alta en la jerarquía. SDRI's tomados de Luna-Martínez *et al.*, 2016.

Residuo	SDRI x 10 ²	Jerarquía
Asn1	7.55	109
Phe2	18.47	71
Met3	15.09	83
Leu4	26.54	38
Thr5	33.74	24
Gln6	47.89	7
Pro7	36.99	16
His8	11.30	98
Ser9	25.04	45
Val11	25.09	44
Ser12	23.69	54
Glu13	25.62	43
Ser14	20.33	62
Pro15	21.48	60
Gly16	22.17	57
Lys17	16.75	78
Thr18	23.97	50
Val19	36.62	18
Thr20	27.31	37
Ile21	42.21	12
Ser22	31.67	28
Cys23	49.98	5
Thr24	38.27	14
Arg25	24.15	48
Ser26	11.99	95
Ser27	8.57	104
Gly28	11.34	97
Ser29	12.61	94
Ile30	19.54	66
Ala31	12.75	92
Ser31a	13.54	89
Asn31b	24.33	47
Tyr32	16.81	77
Val33	31.03	30
Gln34	19.93	65
Trp35	50.21	4
Tyr36	21.46	61
Gln37	31.11	29
Gln38	18.27	72
Arg39	13.93	85
Pro40	11.57	96
Gly41	7.22	110
Ser42	10.28	103
Ser43	11.07	99
Pro44	17.79	74
Thr45	25.78	42
Thr46	26.02	40
Val47	43.94	10
Ile48	45.02	9
Tyr49	19.33	67

Residuo	SDRI x 10 ²	Jerarquía
Glu50	12.62	93
Asp51	16.54	79
Asn52	21.78	58
Gln53	15.12	82
Arg54	22.63	56
Pro55	25.84	41
Ser56	10.71	102
Gly57	14.01	84
Val58	23.34	55
Pro59	24.63	46
Asp60	17.37	76
Arg61	36.90	17
Phe62	27.55	35
Ser63	35.65	20
Gly64	33.58	25
Ser65	26.43	39
Ile66	27.43	36
Asp67	15.17	81
Ser68	11.00	100
Ser68a	10.97	101
Ser68b	13.17	91
Asn69	13.57	88
Ser70	32.64	27
Ala71	45.25	8
Ser72	39.53	13
Leu73	48.13	6
Thr74	34.26	23
Ile75	36.45	19
Ser76	20.20	63
Gly77	21.70	59
Leu78	23.88	51
Lys79	18.72	69
Thr80	17.79	75
Glu81	13.60	87
Asp82	35.20	21
Glu83	19.26	68
Ala84	43.86	11
Asp85	30.60	31
Tyr86	55.21	1
Tyr87	29.00	34
Cys88	50.99	2
Gln89	20.14	64
Ser90	38.02	15
Tyr91	18.50	70
Asp92	13.18	90
Ser93	7.93	108
Ser94	8.48	105
Asn95	8.11	107
His95a	8.12	106
Val96	16.22	80

Cont. Anexo A. Jerarquización de los residuos de 6aJL2. La numeración de los residuos está en base a Chothia *et al.*,1998. Al valor más alto de SDRI le corresponde la posición más alta en la jerarquía. SDRI's tomados de Luna-Martínez *et al.*, 2016.

Residuo	SDRI x 10²	Jerarquía
Val97	23.87	52
Phe98	13.60	86
Gly99	29.39	32
Gly100	24.13	49
Gly101	33.10	26
Thr102	50.69	3
Lys103	23.83	53
Leu104	34.48	22
Thr105	29.03	33
Val106	18.24	73
Leu107	5.10	111

Anexo B. Jerarquización de los residuos de AR. La numeración de los residuos está en base a Chothia *et al.*,1998. Al valor más alto de SDRI le corresponde la posición más alta en la jerarquía. SDRI's tomados de Luna-Martínez *et al.*, 2016. Se comparó el cambio de jerarquía con respecto a la de 6aJL2.

Residuo	SDRI x 10 ²	Jerarquía	Cambio
Asp1	7.84	110	-1
Phe2	18.00	86	-15
Met3	18.79	82	1
Leu4	35.85	32	6
Thr5	43.17	14	10
Gln6	54.74	5	2
Pro7	40.70	21	-5
His8	11.95	101	-3
Ser9	27.69	50	-5
Val11	25.48	59	-15
Ser12	27.88	49	5
Glu13	22.12	71	-28
Ser14	22.58	70	-8
Pro15	23.34	66	-6
Gly16	24.98	61	-4
Lys17	15.73	93	-15
Thr18	31.61	40	10
Val19	30.89	44	-26
Thr20	40.40	22	15
Phe21	42.51	16	-4
Ser22	32.63	36	-8
Cys23	60.16	2	3
Thr24	44.90	10	4
Gly25	38.02	27	21
Ser26	22.86	69	26
Gly27	18.84	81	23
Gly28	23.33	67	30
Ser29	21.36	72	22
Ile30	38.98	26	40
Ala31	28.59	47	45
Asp31a	16.97	90	-1
Ser31b	30.22	46	1
Phe32	20.59	77	0
Val33	45.12	9	21
Gln34	40.17	24	41
Trp35	56.53	3	1
Tyr36	37.88	28	33
Gln37	24.73	62	-33
Gln38	17.99	87	-15
Arg39	9.07	107	-22
Pro40	9.46	106	-10
Gly41	8.30	109	1
Ser42	11.39	102	1
Ala43	14.66	95	4
Pro44	22.89	68	6
Thr45	28.25	48	-6
Thr46	31.01	42	-2
Val47	36.26	31	-21

Residuo	SDRI x 10 ²	Jerarquía	Cambio
Tyr49	8.84	108	-41
Asp50	16.62	92	1
Asp51	26.54	54	25
Asn52	26.47	55	3
Gln53	17.06	89	-7
Arg54	23.61	63	-7
Pro55	26.01	57	-16
Ser56	11.23	103	-1
Gly57	16.65	91	-7
Val58	26.79	53	2
Pro59	25.10	60	-14
Asp60	18.14	85	-9
Arg61	32.49	37	-20
Phe62	26.19	56	-21
Ser63	36.81	29	-9
Gly64	42.47	17	8
Ser65	33.61	34	5
Ile66	27.08	51	-15
Asp67	17.26	88	-7
Asp68	13.07	99	1
Ser68a	12.11	100	1
Ala68b	15.48	94	-3
Asn69	23.54	65	23
Ser70	36.41	30	-3
Ala71	51.74	7	1
Ser72	40.18	23	-10
Leu73	40.14	25	-19
Thr74	47.71	8	15
Ile75	42.91	15	4
Ser76	23.57	64	-1
Gly77	25.80	58	1
Leu78	30.97	43	8
Lys79	19.61	79	-10
Thr80	18.77	83	-8
Glu81	13.28	97	-10
Asp82	33.48	35	-14
Glu83	20.36	78	-10
Ala84	44.79	11	0
Asp85	30.88	45	-14
Tyr86	52.32	6	-5
Tyr87	27.03	52	-18
Cys88	64.18	1	1
Gln89	32.01	39	25
Ser90	44.02	12	3
Tyr91	18.97	80	-10
Asn92	20.74	76	14
Ser93	10.59	104	4
Asn94	13.18	98	7

Cont. Anexo B. Jerarquización de los residuos de AR. La numeración de los residuos está en base a Chothia et al.,1998. Al valor más alto de SDRI le corresponde la posición más alta en la jerarquía. SDRI's tomados de Luna-Martínez et al., 2016. Se comparó el cambio de jerarquía con respecto a la de 6aJL2.

Residuo	SDRI x 10 ²	Jerarquía	Cambio
His95	7.71	111	-4
His95a	9.65	105	1
Val96	21.12	74	6
Val97	31.24	41	11
Phe98	18.24	84	2
Gly99	41.65	18	14
Gly100	34.21	33	16
Gly101	40.83	19	7
Thr102	54.95	4	-1
Lys103	20.97	75	-22
Val104	43.39	13	9
Thr105	32.44	38	-5
Val106	21.33	73	0
Leu107	13.39	96	15



Simple approach for ranking structure determining residues

Oscar D. Luna-Martínez¹, Abraham Vidal-Limón², Miryam I. Villalba-Velázquez¹, Rosalba Sánchez-Alcalá¹, Ramón Garduño-Juárez³, Vladimir N. Uversky^{4,5,6} and Baltazar Becerril¹

¹Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

²División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica, San Luis Potosí, Mexico

³Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

⁴Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, University of South Florida, Tampa, FL, United States

⁵Institute for Biological Instrumentation, Russian Academy of Sciences, Puschino, Moscow Region, Russia

⁶Laboratory of Structural Dynamics, Stability and Folding of Proteins, Russian Academy of Sciences, St. Petersburg, Russia

ABSTRACT

Mutating residues has been a common task in order to study structural properties of the protein of interest. Here, we propose and validate a simple method that allows the identification of structural determinants; i.e., residues essential for preservation of the stability of global structure, regardless of the protein topology. This method evaluates all of the residues in a 3D structure of a given globular protein by ranking them according to their connectivity and movement restrictions without topology constraints. Our results matched up with sequence-based predictors that look up for intrinsically disordered segments, suggesting that protein disorder can also be described with the proposed methodology.

Subjects Biochemistry, Biophysics, Computational Biology, Molecular Biology

Keywords Molecular dynamics, Shannon dynamical entropy, Intrinsic disorder, Thermodynamics

INTRODUCTION

It is widely known that highly conserved residues in multiple sequence alignments of proteins are considered as key residues essential for protein structure and function including protein folding, phylogeny and evolutionary processes (*Phillips, Janies & Wheeler, 2000; Edgar & Batzoglou, 2006; Pierri, Parisi & Porcelli, 2010*). Nevertheless, neutral mutations can accomplish great modifications to the structural properties of the protein by modifying its thermodynamic stability or biological function as seen in extremophiles in order to adapt to their environment (*Jaenicke & Böhm, 1998; Rothschild & Mancinelli, 2001; Reed et al., 2013*). Proteins can be considered as molecular entities with a tridimensional structure which are exposed to solvent and, frequently, interacting with other macromolecules. Each amino acid residue in a protein and each nucleotide base in a nucleic acid can be seen as a node and due to their high connectivity they represent hubs. Therefore, network analyses applied on protein structures attempt to assess the location of essential hubs

Submitted 7 December 2015

Accepted 25 May 2016

Published 22 June 2016

Corresponding author

Baltazar Becerril,
baltazar@ibt.unam.mx

Academic editor

Rudiger Ettrich

Additional Information and
Declarations can be found on
page 18

DOI 10.7717/peerj.2136

© Copyright

2016 Luna-Martínez et al.

Distributed under

Creative Commons CC-BY 4.0

OPEN ACCESS

that are necessary for the preservation of the stability of the global structure (*Vendruscolo et al., 2002*). Several mathematical approaches have been described which have helped to identify these hubs, each one employing different considerations such as geometrical approximations, thresholds, and information theory (*Costa et al., 2007; Böde et al., 2007*). Other approaches combine network parameters with protein properties, e.g., Relative Solvent Accessibility (RSA), to optimize the hierarchization of each residue within the structure (*Li, Wang & Wang, 2008*). In addition, the structural relevance of each residue can be assessed by ranking the theoretical scores obtained from mathematical approximations applied over protein structures (*Greene & Higman, 2003; Amitai et al., 2004*).

Because protein structures are not rigid molecular assemblies, X-ray structure determinations provide a “snapshot” of a “ground state,” which is assumed to represent the lowest energy conformation in a crystal lattice (*Karplus & Kuriyan, 2005; Rodríguez-Rodríguez et al., 2012*). Molecular dynamics simulations is an *in silico* tool that can provide information on detailed atomic motions at different time-scales, which have been increased through the development of more powerful hardware (*Kepleis et al., 2009; Dror et al., 2012*). Taking advantage of these technological advances, we sought to demonstrate that joining this important parameter with network analyses will allow the compilation of a simple method for ranking structure determining residues involved in protein stabilization. The problem of classifying each node according to its structural relevance is far from being trivial because of various reasons. For example, some residues that are not located at the hydrophobic core are known to have a long-distance effect on the structure, even in the case of a neutral mutation (*Tokuriki et al., 2008; Pace et al., 2011*). Our hypothesis is that the rigidity and the connectivity of each residue, irrespectively of its solvent exposure, can be associated to a specific theoretical score that can be used as a ranking parameter, where highly connected residues with restricted movement should have the greatest effects on the overall stability of globular proteins. Among all network descriptors, we selected Shannon dynamical entropy as a connectivity parameter (*Demetrius & Manke, 2004; Costa et al., 2007*). In this study, some well-known protein structures were unbiased chosen with the goal of developing a simple method to identify, in a hierarchical way, those amino acid residues that are determinant in the structural stability of a protein. This approach basically combines molecular dynamics with a network analysis based on the Shannon dynamical entropy.

MATERIALS AND METHODS

Statistics

Principal component analysis was performed using the R software (*R Core Development Team, 2013*). The thermal unfolding experiments of the 6aJL2 mutants were repeated in triplicate.

Crystallographic structures

The PDB codes, size, and protein class of the selected structures are as follows: chymotrypsin inhibitor (2CI2; 65 residues; 16% α + 21% β); 6aJL2 (2W0K; chain A, 111 residues; 5% α + 46% β); apoflavodoxin (1FTG; 168 residues; 35% α + 19% β); dimeric Arc repressor (1ARR; 108 residues; 62% α + 9% β); DNA-binding domain of human estrogen receptor

α in complex with the estrogen response element DNA duplex (1HCQ; chains A, B; 128 residues; 26% α + 9% β , and chains C and D, 56 nucleotides); cold shock protein from *Bacillus subtilis* (1CSP; 67 residues; 4% α + 55% β); cold shock protein from *B. caldolyticus* (1C9O, 66 residues; 4% α + 62% β); and the largely disordered N-terminus of suppressor of cytokine signaling 5 mammalian suppressor of Janus Kinase interaction region (2N34, 70 residues; 12% α). Minor structural modifications were performed on PDB files such as fulfilling N-terminal domains and incomplete lateral chains by using Swiss PDB software (Guex & Peitsch, 1997). Furthermore, all nodes were renumbered in continuous order avoiding repetitions; i.e., in PDB 2CI2 the first residue begins as residue 19. Consequently, all residues were systematically renumbered so the first residue was identified as residue 1 and so forth. Solvent-exposed area was calculated with the NOC software (Chen, Cang & Nymeyer, 2007).

Grade

Residues can be regarded as nodes and contacts as edges. Hence, an edge was defined when any two non-hydrogen atoms from a pair of residues are within distance of 5 Å. To study the topology of the residue contact network, we measured the degree of node-*i*, K_i , as the number of neighbors of node-*i*. Chain A of crystallographic structures was selected. In the case of quaternary structures, all the structure was considered to measure the grade of each node but only chain A was selected for further comparisons.

Molecular dynamics

The query protein was prepared using the Protein Preparation Wizard in Maestro 9.2 package (Schrödinger LCC, NY, USA) and included in a 10-Å water box that contained 14513 SPC-type water molecules for 2CI2, 16667 for 6aJL2, 20051 for 1FTG, 20990 for 1ARR, 28142 for 1HCQ, 12219 for 1CSP, and 12033 for 1C9O. Neutralizing ions were added, and other metal ions already present in the protein structure were left at the same place. The simulated annealing calculations and data analysis were conducted using the Desmond and Maestro programs, respectively (Maestro-Desmond Interoperability Tools, version 3.0; Schrödinger, NY, USA). The OPLS_2005 force field was used for every molecular dynamics simulation. Cubic periodic boundary conditions were used for most of the proteins ($53.9 \times 53.9 \times 53.9 \text{ \AA}^3$ for 2CI2, $56.8 \times 56.8 \times 56.8 \text{ \AA}^3$ for 6aJL2, $60.9 \times 60.9 \times 60.9 \text{ \AA}^3$ for 1FTG, $61.2 \times 61.2.9 \times 61.2 \text{ \AA}^3$ for 1ARR, $51.1 \times 51.1 \times 51.1 \text{ \AA}^3$ for 1CSP, and $50.8 \times 50.8 \times 50.8 \text{ \AA}^3$ for 1C9O); due to the size of the complex 1HCQ, rectangular cuboid boundary conditions were used ($58.5 \times 59 \times 91 \text{ \AA}^3$ for 1HCQ). Each simulation was adjusted with an NPT ensemble by weak coupling to an external bath temperature at constant pressure of 1 atm and relaxation time of 2 ps, regulated by Berendsen barostat (Berendsen et al., 1984). All short-range interactions were computed using a 9 Å cutoff, and for long-range interactions (electrostatic and Van der Waals), a smooth particle mesh Ewald method with a tolerance of 1×10^{-9} was applied (Essmann et al., 1995). To ensure that our simulations started from local minima, a simulated annealing algorithm was performed. This method started the simulation at high temperature (400 K) to overcome thermodynamic and conformational barriers, followed by gradual

cooling (annealing) to reach low energy regimes. It is widely used for the optimization of structures from experimental methods, comparative protein modeling, or studying the conformational dynamics of protein or peptide folding and unfolding (Mori & Okamoto, 2009). The full system was heated at 10 K for 30 ps, 100 K for 100 ps, 300 K for 200 ps, 400 K for 300 ps, and 400 K for 500 ps and then cooled to 298 K for 1,000 ps. To ensure that heating at 400 K did not affect protein or protein/DNA structure, the RMSD of heavy atoms (C, N, O, S, P) derived from the annealed structure was compared against corresponding crystallographic structure. If RMSD standard deviation was $\leq 1.0 \text{ \AA}$, then it was assumed that the annealing algorithm did not change the whole structure or denature it. In fact, since the displacement of heavy atoms above 1 \AA is considered as a conformational change, the algorithm can be trusted in finding the local minimum. A lineal interpolation step between two adjacent time points was employed. After the sixth step, a production of 25 ns was achieved with an integration time of 1 fs. The entire analysis was performed using trajectory coordinates, and the energies were written to a disk every 1.2 ps. A frame was extracted every 0.25 ns throughout the simulation, and the overall frames were saved as a PDB file.

Dynamical entropy

An analytical algorithm encoded in the Perl language was developed to generate three files using the 100-frames PDB-file as the template. In the first step, a single file is generated for each model, and this single file contains the atom coordinates, the corresponding molecular weight, and the name of the node (amino acid or nucleotide base). The second step calculates the coordinates of the center of mass of the node. The third step calculates the distance between the center of mass of each pair of nodes and their normalized distance described by Eq. (1),

$$\overline{d}_{ij} = \frac{d_{ij}}{r_{vdW_i} + r_{vdW_j}} \quad (1)$$

where \overline{d}_{ij} is the normalized distance, d_{ij} is the distance between the center of mass of node i and that of node j , and r_{vdW} is the Van der Waals radius of the respective node considering all their atoms. For amino acid residues, values of their Van der Waals radii comprehending the whole residue were obtained from Darby & Creighton (1993). For nucleotide bases, values were obtained from Voss & Gerstein (2005). The mean distance between each pair of nodes of the 100 structures sampled was calculated. Since the next steps involve eigenvector and eigenvalue measurement properties, the mean value of the inverse normalized distance was calculated, so the largest weight represents the closest distance between a pair of nodes and the smallest weight represents the longest distance.

A weighted adjacency matrix $A = (a_{ij}) \geq 0$ of size $N \times N$, where N is the number of nodes, was constructed. In this case, matrix A is symmetric ($a_{ij} = a_{ji}$) and undirected. Following the mathematical strategy described in Demetrius & Manke (2004) we now assume that the stochastic process is given by a Markov Matrix $P = p_{ij}$ where $p_{ij} \geq 0$ and $\sum_j p_{ij} = 1$. The stationary distribution of matrix P is described by Eq. (2),

$$\pi P = \pi \quad (2)$$

where π is defined as the left-hand eigenvector associated with the largest eigenvalue solved with Mathcad 15TM software. The dynamical entropy of this process of each node, H_i , is described by Eq. (3),

$$H_i = - \sum_j \pi_i p_{ij} \log p_{ij} \quad (3)$$

where the term $p_{ij} \log p_{ij}$ is the standard Shannon entropy.

6aJL2 mutants

The synthesis of single-point mutants of 6aJL2 (Arg25His, Ile30Gly, Tyr36Phe, and Gln6Asn) was performed using recursive PCR (*Prodromou & Pearl, 1992*). The obtained DNAs were cloned into the pSyn1 expression vector (*Schier et al., 1995*). All of the constructions were verified by nucleotide sequencing (*Sanger, Nicklen & Coulson, 1977*). The variants were expressed in *Escherichia coli* BL21 (DE3) and purified as described previously (*Del Pozo-Yauner et al., 2008*). The protein purity was verified using SDS-PAGE electrophoresis, and the protein concentration was determined spectrophotometrically at 280 nm in 6.5 M GdnHCl and 20 mM sodium phosphate buffer, pH 7.5, using molar extinction coefficients calculated from the amino acid sequence using the ProtParam software, which is available at the ExpASy website (*Artimo et al. al., 2012*).

Unfolding

Samples containing 50 $\mu\text{g/ml}$ of protein in phosphate-buffered saline (PBS), pH 7.5, were placed into a 3-ml quartz cuvette. Changes in the tryptophan fluorescence were measured using a LS50B Perkin Elmer spectrofluorometer with an excitation wavelength of 295 nm (2.5-mm bandwidth) and an emission wavelength of 355 nm (5-mm bandwidth). The temperature was increased from 298 to 350 K at a rate of 1 K/min and then samples were cooled to 298 K at the same rate. The data were analyzed using the thermal unfolding Eq. (4), which was obtained from *Eftink (1995)*.

$$F_{Trp} = \frac{(y_n + m_n T) + (y_d + m_d T) e^{\left(\frac{\Delta H_m}{RT_m} - \frac{\Delta H_m}{RT}\right)}}{1 + e^{\left(\frac{\Delta H_m}{RT_m} - \frac{\Delta H_m}{RT}\right)}} \quad (4)$$

where F_{Trp} is the tryptophan fluorescence, T is the temperature, T_m is the temperature of the midpoint, ΔH_m is the enthalpy at T_m , y_n and m_n describe the pre-transition phase, and y_d and m_d describe the post-transition phase. Non-linear regression was performed using the OriginPro8TM software. The change in the Gibbs energy of the wild-type versus the mutant ($\Delta\Delta G$) was calculated for temperature- and denaturant-induced unfolding processes using the following equations.

Thermal unfolding (*Becktel & Schellman, 1987*), Eq. (5):

$$\Delta\Delta G = \frac{\Delta H_{mWT}}{T_{mWT}} (T_{mMUT} - T_{mWT}) \quad (5)$$

where WT refers to the wild-type values and MUT refers to the mutant values of the melting temperature T_m , and ΔH_m is the enthalpy value.

Chemical unfolding (Creighton, 1990), Eq. (6):

$$\Delta\Delta G = m_{WT}(C_{mMUT} - C_{mWT}) \quad (6)$$

where m_{WT} is the transition slope of the wild-type, and C_m is the denaturant concentration at which $\Delta G = 0$.

RESULTS AND DISCUSSION

Many of the residues located at the hydrophobic core are considered essential (deleterious if mutated), but it is hard to identify which one will be lethal through simple visual inspection of a protein structure. In most cases reported so far, mutations increase the amount of unstable conformers, rendering a protein more susceptible to external perturbations; and this is more likely if the mutations are located at the inner core of the protein (Jackson *et al.*, 1993; Baldwin & Matthews, 1994; Lei & Duan, 2004; Kumar & Nussinov, 2001; Reed *et al.*, 2013). Additionally, evaluating the role of each position by replacing the corresponding residue with any other amino acid becomes a more challenging task as the protein becomes bigger. Therefore, we sought to generate a simple method to assess proteins independently of their structural complexity (secondary, tertiary or even quaternary structure).

We selected several proteins with diverse topologies that have been thermodynamically characterized to evaluate the effects of the incorporation of some single-point mutations near neutral pH. Chymotrypsin inhibitor, 6aJL2, apoflavodoxin, arc repressor, estrogen receptor/DNA estrogen/response element complex (ERC), cold shock protein of *B. caldolyticus* (Csp C), cold shock protein of *B. subtilis* (Csp S), and the largely disordered N-terminus domain of Janus Kinase interaction region were the representative proteins chosen for this study (see Methods section for PDB details). Not only these proteins were selected because they exhibit different folds, but also because they were analyzed using distinct experimental procedures that provided different thermodynamic parameters covering two-state and three state unfolding pathways. Most of the selected proteins have been thermodynamically evaluated elsewhere through denaturation experiments using single-point mutants (Milla, Brown & Sauer, 1994; Itzhaki, Otzen & Fersht, 1995; Perl & Schmid, 2001; Banci *et al.*, 2004; Campos *et al.*, 2004a; Campos *et al.*, 2004b; Wunderlich, Martin & Schmid, 2005; Wunderlich & Schmid, 2006; Gribenko & Makhataдзе, 2007; Del Pozo-Yauner *et al.*, 2008; Deegan *et al.*, 2010; Hernández-Santoyo *et al.*, 2010; Van den Bedem, 2013; Del Pozo-Yauner *et al.*, 2014). $\Delta\Delta G$ value represents the effect of a determined mutation on the stability compared to the wild-type protein structure. On the other hand, the estrogen receptor complex (ERC) was assessed by determining the affinity constant (K_D) against its target DNA. The percentage difference in terms of K_D determines the variation level in affinity between the wild-type ER homodimer and its mutant homodimer when binding to its target DNA sequence. Both, percentages lower than 100 in K_D and negative $\Delta\Delta G$ values indicate less stable structures.

We compared these experimental data with a network parameter- Grade (see File S1). This parameter is defined as the number of neighbors of each node when any two non-hydrogen atoms from residues i and j are within a cutoff distance (Li, Wang & Wang,

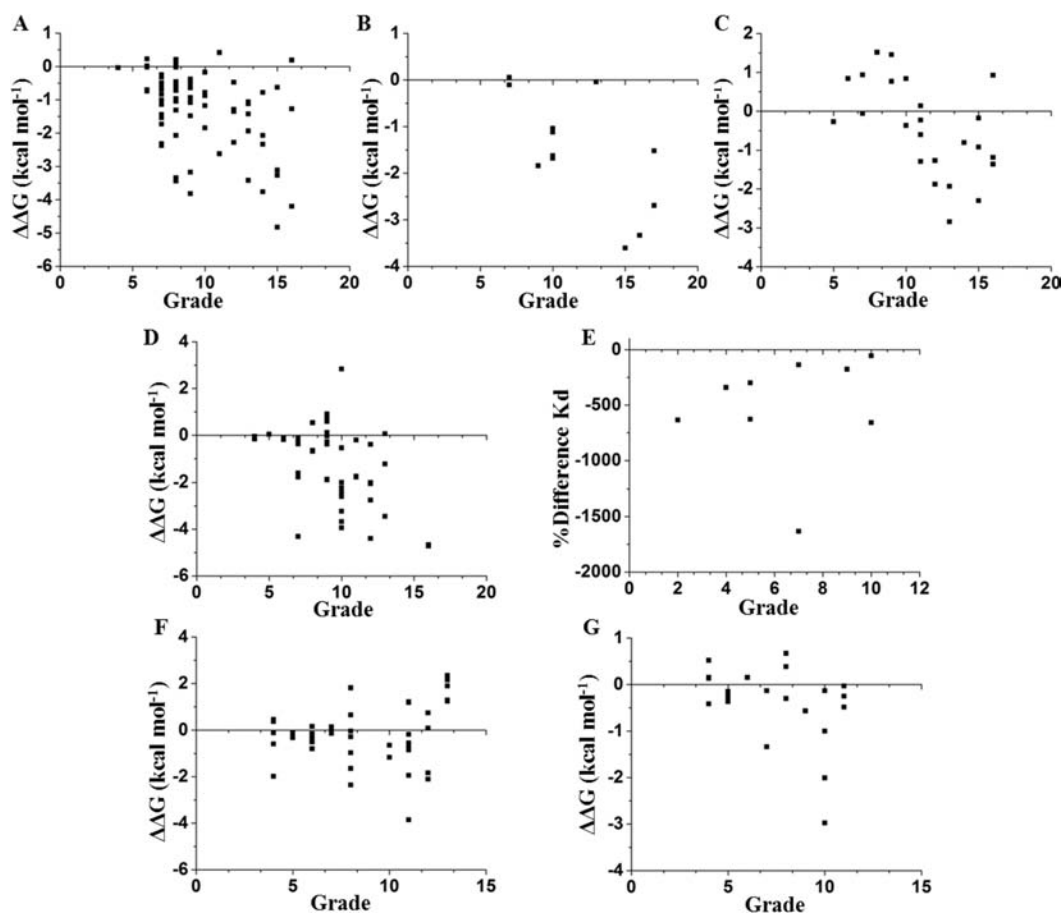


Figure 1 Contrasting of network parameter grade against experimental values. Each point in the graphics represents scores determined from the number of contacts surrounding the residue where the mutation was performed (see [File S1](#)). (A) Chymotrypsin inhibitor; (B) 6aJL2; (C) apoflavodoxin; (D) arc repressor; (E) DNA-binding domain of the estrogen receptor w ; (F) cold shock protein from *B. subtilis*; and (G) cold shock protein from *B. caldolyticus*.

2008). According to [Fig. 1](#), there is a higher probability to affect the stability if the selected node is highly surrounded in a cutoff distance of 5 Å. However, this parameter does not provide a full hierarchization of each node; some nodes can share same grade-value but exhibit a different impact on the stability. Another caveat is that size matters, because bigger residues, such as tryptophan or phenylalanine, tend to make more interactions than smaller residues, such as glycine or alanine (see [File S1](#)). Furthermore, proteins are not rigid structural assemblies since most of their contacts are in a dynamical condition. Thus, we suggest that the molecular dynamics of a protein should be gathered with more precise network analyses in order to properly assess the influence of determined mutations on the stability of the protein structure.

As a first step to assess this assumption, the crystallographic structures were subjected to 25 ns of molecular dynamics simulations as described in the Methods section. The completion of this process was confirmed by inspecting the Root Mean Square Deviation (RMSD) values of the main chain indicating the convergence to a stationary structural

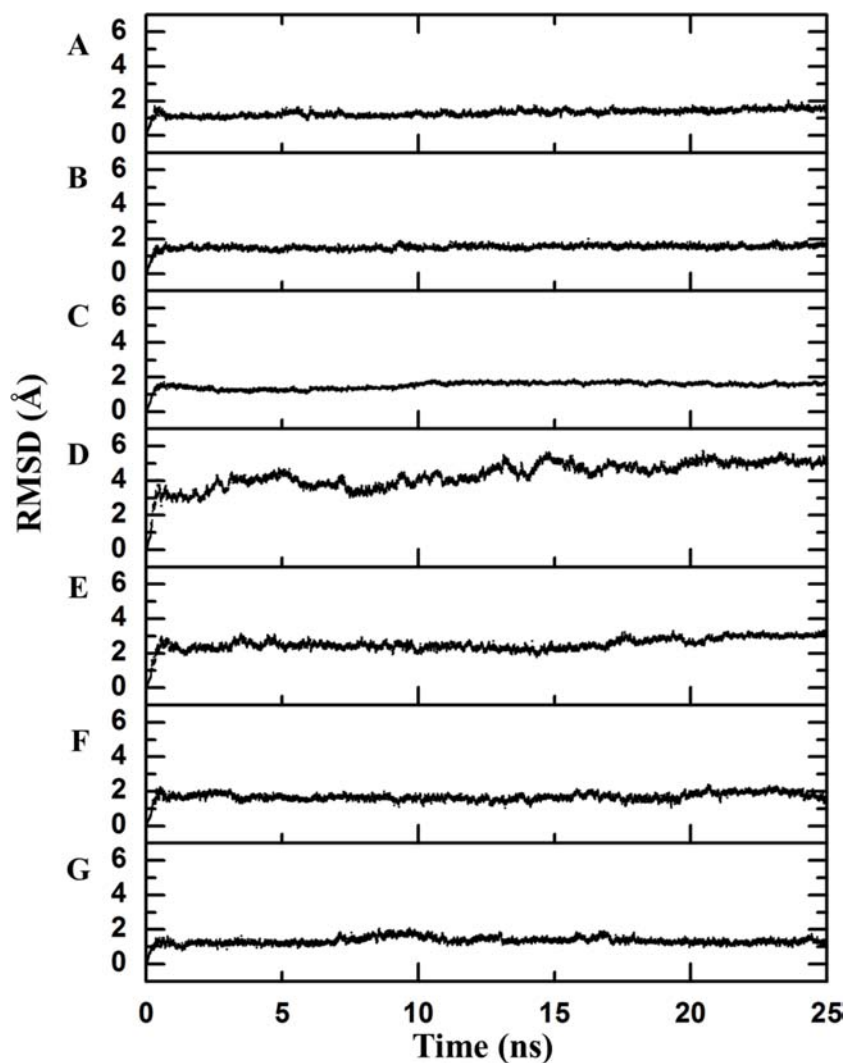


Figure 2 Backbone RMSD. RMSD difference between the backbone of the crystallographic structure and the corresponding structure present at the indicated time of molecular dynamics; this difference was calculated using Desmond. There is an observable change after the first few nanoseconds of the dynamics simulations, during which the structure is “heated,” but the protein remains stable after that moment. (A) Chymotrypsin inhibitor; (B) 6aJL2; (C) apoflavodoxin; (D) arc repressor; (E) DNA-binding domain of the estrogen receptor; (F) cold shock protein from *B. subtilis*; and (G) cold shock protein from *B. caldolyticus*.

movement at 300 K (Fig. 2). Every 0.25 ns, a structure was extracted by means of generating one hundred frames over all the simulation time in order to obtain a good representation of the possible movements of both the lateral and the main chains. Next, a network analysis was performed for each frame. In accordance with definitions, each protein residue and each nucleotide base were considered as individual nodes. The center of mass was calculated based on all of the atoms of the node. The interaction strength between a pair of nodes was measured in terms of the distance between their mass centers which was subsequently normalized by the summation of their Van der Waals radii considering all their atoms, as

Table 1 Statistics of theoretical scores—SDRIs. Skewness, kurtosis, and the normality of the SDRI distribution measured with Shapiro–Wilk test. SDRI distribution of globular proteins shows a good distribution unlike unstructured peptide. Analysis was performed using SigmaPlot11.0 software (Systat Software, San Jose, CA).

Protein	Skewness	Kurtosis	Shapiro–Wilk
Chymotrypsin inhibitor	0.16	−0.59	0.98
6aJL2	0.74	−0.03	0.95
Apoflavodoxin	0.80	0.12	0.95
Arc repressor	0.51	0.02	0.97
Complex estrogen/Receptor	−0.11	−0.19	0.99
Cold shock protein BS	0.53	−0.67	0.94
Cold shock protein BC	0.61	−0.54	0.94
Unstructured peptide	1.75	2.15	0.71

stated in Materials and Methods section (values less than 1 imply that the interaction is very strong).

We selected dynamical Shannon entropy as an approach to estimate connectivity being aware that the results describe probabilistic values. The next step in the calculation process included the use of eigenvector properties for which, the greater value, the more important interaction. The inverse value of the mean normalized distance for the 100 frames was calculated and rounded up to four decimals. At this point, we would like to emphasize that this strategy makes a cutoff distance unnecessary. Then, each of these values was incorporated into a square matrix that was, in turn, converted into a row stochastic matrix. Assuming that the microscopic process of the network is Markovian, the matrix was solved consistently using a dynamical entropy equation corresponding to a Markov process (Demetrius & Manke, 2004; Costa et al., 2007). Each node is now associated with a dynamical entropic value (H_i), which can be interpreted as a connectivity parameter with a probabilistic character. We also presume that the movement restriction of each node, which is represented by its Root Mean Square Fluctuation (RMSF) value derived from the molecular dynamics, is associated with its dynamical entropic value. Therefore, each node was scored by dividing its entropic value by its respective RMSF value, and this score indicates the relative importance of each node within its respective structure (see File S2). These theoretical scores were identified as structural-determining residue identifier (SDRI). By normalizing SDRIs and plotting them against their sequence, SDRIs were distributed throughout the structure (Fig. 3). Distribution statistics- skewness, kurtosis, and Shapiro–Wilk test showed that globular proteins are near normal distribution unlike the unstructured peptide which is described hereinafter (Table 1). Interestingly, structures in complex—Arc repressor and Complex Receptor DNA, showed fewer residues with low SDRI values compared to the other analyzed structures, suggesting that both are notably less disordered.

We compared the distribution of the non-normalized SDRIs obtained in our study with scores from the structure-based flexibility as well as the sequence-based intrinsic disorder predisposition of the query proteins. The structural flexibility of these proteins was obtained

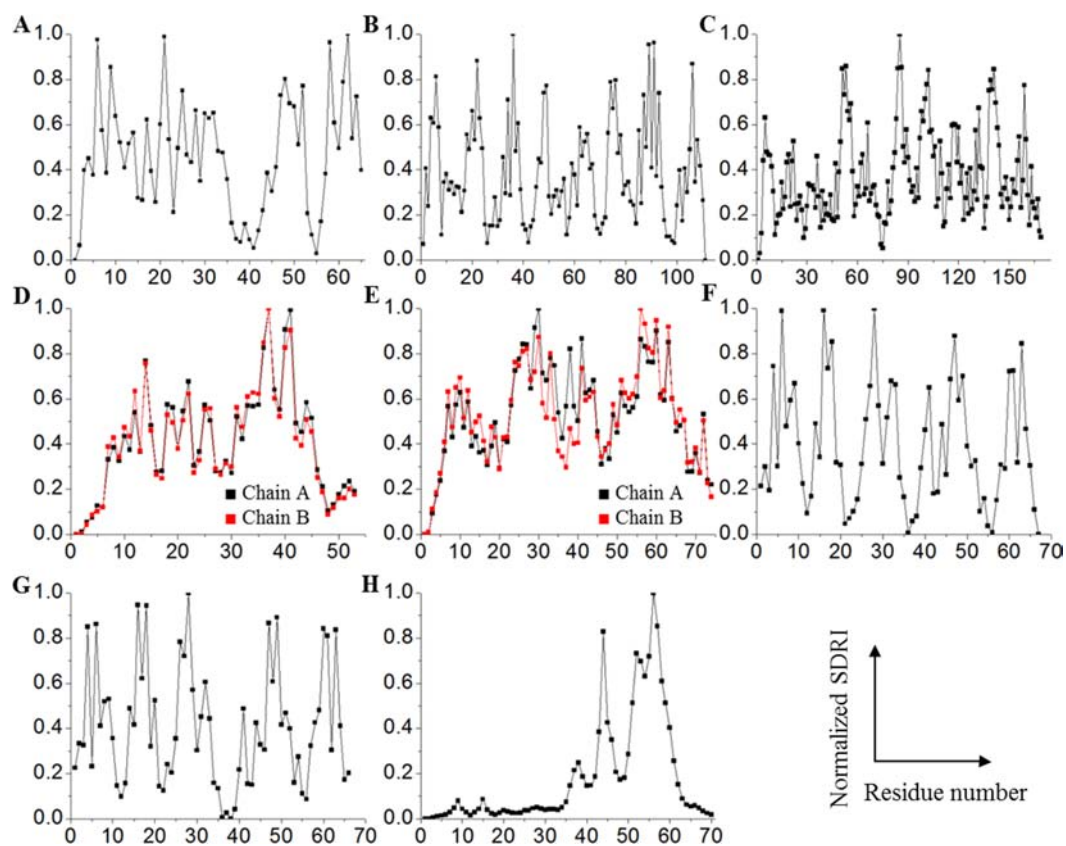


Figure 3 Contrasting of SDRI values against their corresponding sequence. Each point in the graphics represents normalized SDRI values from the theoretical analysis according to their sequence position (see [File S2](#)). Values near 0 indicate flexible and lesser connected residues while values near 1 indicate rigid and well-connected residues: (A) Chymotrypsin inhibitor; (B) 6aJL2; (C) apoflavodoxin; (D) arc repressor; (E) complex of estrogen receptor α /DNA estrogen response element; (F) cold shock protein from *B. subtilis*; (G) cold shock protein from *B. caldolyticus*; and (H) the JAK interaction region of SOCS5.

by utilizing the FlexPred tool that predicts the absolute fluctuations per-residue from a three-dimensional structure using the B-factors of a query protein ([Jamroz, Kolinski & Kihara, 2012](#)). The intrinsic disorder propensities per-residue of these proteins was obtained by using PONDR[®] VSL2B predictor, which is one of the more accurate stand-alone disorder predictors ([Fan & Kurgan, 2014](#); [Peng et al., 2005](#); [Peng & Kurgan, 2012](#)). Results of these comparisons are shown in [Fig. 4](#) and clearly illustrated that these three computational tools, SDRI, FlexPred, and PONDR[®] VSL2B, can “see” different, although related, features in a protein. Note that we used (1-SDRI) function instead of SDRI when representing the SDRI values to compare data from these three tools “in phase” to highlight intrinsic disorder instead of structural rigidity. These results suggested that there is a good agreement between the structural flexibility calculated from the protein crystal structure and the propensity of a protein to preserve disorder. Furthermore, it seemed that residues essential for the preservation of the stability of global protein structure are typically located within highly ordered and less flexible domains. In terms of spectral analysis, the visual inspection of plots shown in [Fig. 4](#) suggested that, in many cases, the propensity for intrinsic

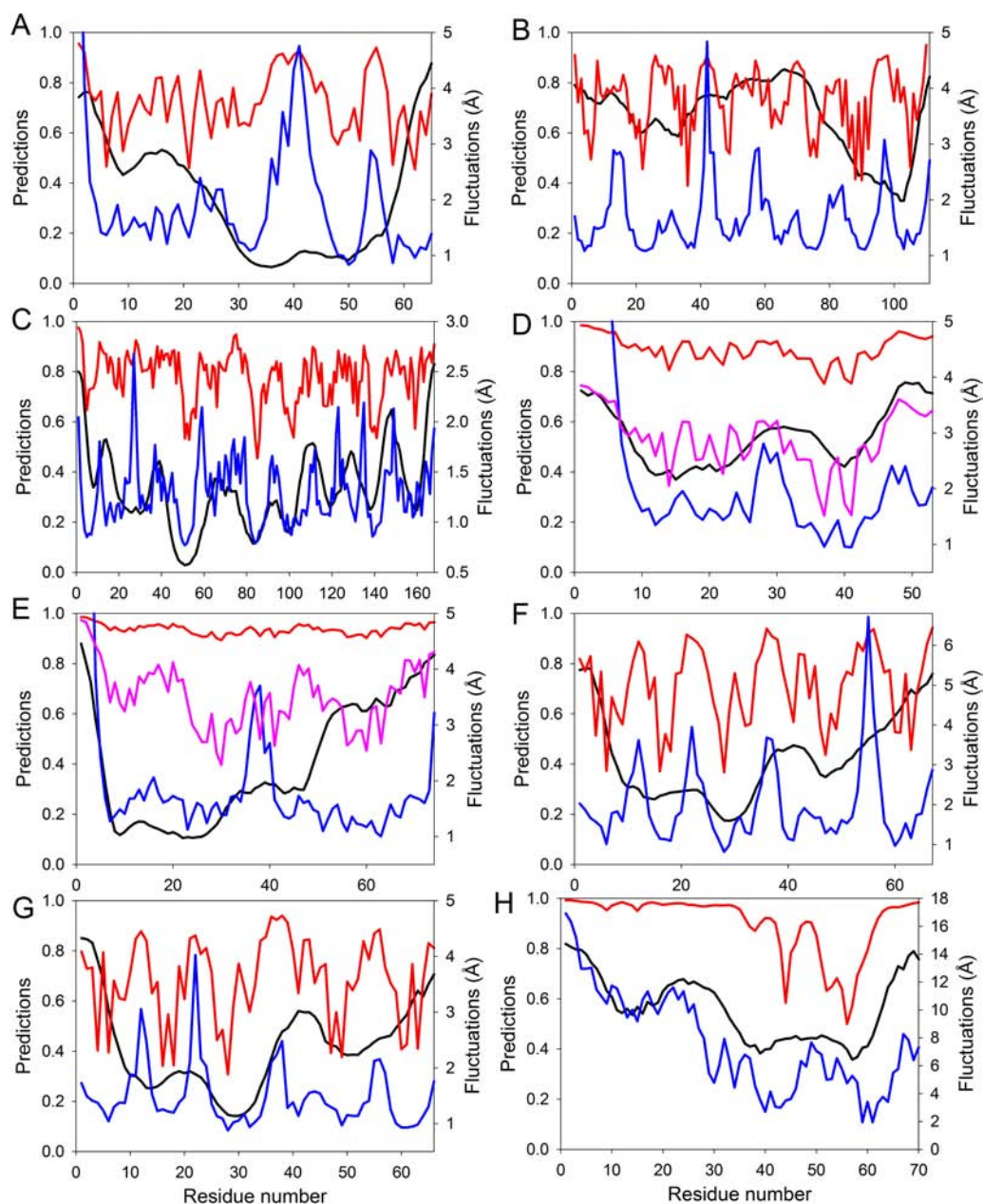


Figure 4 Evaluating intrinsic disorder and structural flexibility. Distributions of predicted intrinsic disorder propensity evaluated by PONDR[®] VSL2B (black curves), predicted structural flexibility estimated by FlexPred (blue curves) and (1-SDRI) function ranking structure determining residues (red or pink curves) for a set of query proteins: (A) Chymotrypsin inhibitor; (B) 6aJL2; (C) apoflavodoxin; (D) arc repressor; (E) DNA-binding domain of the estrogen receptor α ; (F) cold shock protein from *B. subtilis*; (G) cold shock protein from *B. caldolyticus*; and (H) the JAK interaction region of SOCS5. Propensities for intrinsic disorder and (1-SDRI) function are scaled from 0 to 1. Since formation of two complexes—dimeric Arc repressor and a complex between the estrogen receptor DNA-binding domain and the DNA estrogen response element—resulted in a dramatic reduction of the amplitude of the (1-SDRI) function, corresponding plots (D and E) also include expanded (1-SDRI) curves shown in pink.

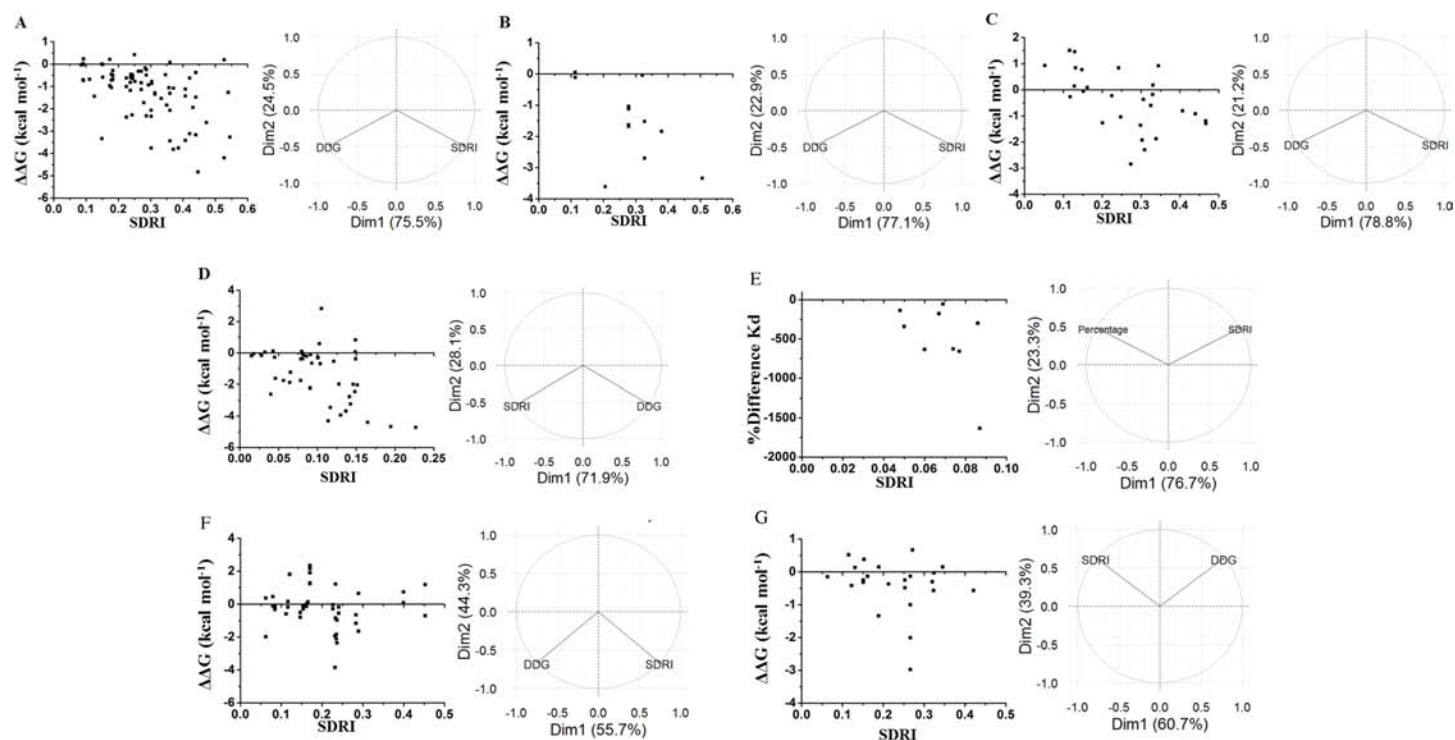


Figure 5 Contrasting of SDRIs against their corresponding experimental values. Each point in the graphics represents a single-mutation experiment compared with the scores from the theoretical analysis (see [File S3](#)). In the case of homodimers, only the values for chain A are shown. Statistical data is shown in terms of eigenvectors: DIM1 refers to SDRIs ($H_p \text{RMSF}_i^{-1}$), and DIM2 refers to the experimental values, $\Delta\Delta G$ (DDG) or KD percentage difference. (A) Chymotrypsin inhibitor; (B) 6aJL2; (C) apoflavodoxin; (D) arc repressor; (E) complex of estrogen receptor α /DNA estrogen response element; (F) cold shock protein from *B. subtilis*; and (G) cold shock protein from *B. caldolyticus*.

disorder results in broad bands that define global appearance of the curves, whereas the outputs of (1-SDRI) and FlexPred add fine structural resolution to the resulting plots. Additionally, for some regions, noticeable disagreements can be found among the outputs of these three tools, which can be attributed to the particular considerations of each tool. Nevertheless, these important observations suggested that intrinsic disorder propensity, predicted from amino acid sequence, serves as an important background defining global flexibility of a protein 3D-structure which is fine-tuned by long-distance interactions taking place in a folded molecule.

A more efficient way to exploit SDRIs was to visualize each node according to its theoretical score; the higher SDRI the more important is the residue to preserve the structure. In homodimeric domains (arc repressor and ERC) both monomers showed the same distribution. In our analysis, the residues with highest SDRI values corresponded to structure-determinant residues ([Fig. 5](#)). Complementarily, heat maps based on these values facilitated the localization of essential residues or segments involved in a protein's biological function ([Fig. 6](#)). Despite that linear statistics parameters (R-squared and Pearson Correlation Coefficient) showed low correlation between SDRIs and experimental values ([Table 2](#)), visual inspection of [Fig. 5](#) strongly suggested that a pattern was followed. Therefore, we performed a statistical scrutiny applying Principal Component Analyses

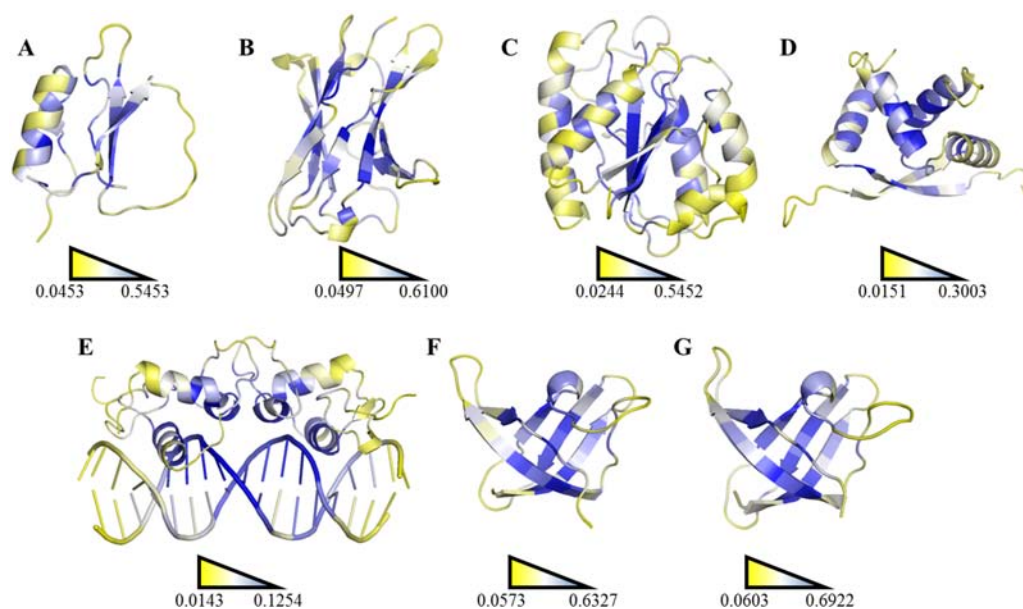


Figure 6 Ranking heat map. The highest SDRI, representing a highly connected node with restricted movement, is shown in blue, and the lowest theoretically scored node, which is barely connected and exhibits a high degree of movement, is represented in yellow (see [File S2](#)). The structures were prepared using the PyMOL software. (A) Chymotrypsin inhibitor (PDB entry 2CI2); (B) 6aJL2 (PDB entry 2W0K); (C) apoflavodoxin (PDB entry 1FTG); (D) arc repressor (PDB entry 1ARR); (E) complex of estrogen receptor α /DNA estrogen response element (PDB entry 1HCQ); (F) cold shock protein from *B. subtilis*, and (G) cold shock protein from *B. caldolyticus*.

Table 2 Statistics of SDRIs versus experimental results. Coefficient of determination (R^2) and Pearson correlation coefficient analysis describe mild correlation between SDRIs and experimental values. Analysis was performed using SigmaPlot11.0 software (Systat Software, San Jose, CA).

Protein	R^2	Pearson correlation coefficient
Chymotrypsin inhibitor	0.25	-0.50
6aJL2	0.22	-0.54
Apoflavodoxin	0.28	-0.55
Arc repressor	0.24	-0.50
Complex estrogen/Receptor	0.29	-0.54
Cold shock protein BS	0.06	-0.12
Cold shock protein BC	0.08	-0.21

(PCA) to identify data patterns of apparently uncorrelated variables ([Abdi & Williams, 2010](#)). Results were represented as circular biplots allowing us to examine the correlation between variables SDRI and experimental data as vectors (see [Fig. 5](#) and [Fig. S2](#)). Based on biplots from PCA results and due to the angle between variables SDRIs and thermodynamic data, it can be emphasized that the present method is able to estimate the effect of a single-point mutation on protein structure depending on the importance of a given residue irrespective of its position. The results confirmed our hypothesis that the most connected and the most rigid residues are the most influential on the structural stability of the protein

despite their involvement in any kind of structural organization. For example, the worst effects in Arc repressor were observed in five mutants (VA22, EA36, IA37, VA41, and FA45) that showed little or no cooperation in denaturation experiments (Milla, Brown & Sauer, 1994). Remarkably, these mutations were performed on residues that matched with high SDRI values (6, 4, 1, 2, and 19, respectively).

The next comparison was between the SDRI of two structures that have different thermodynamic stability in spite of their high sequence and high structural homology. We selected the Cold shock protein from thermophilic *B. subtilis* (Csp S) and from hyperthermophilic *B. caldolyticus* (Csp C) bacteria. Csp C and Csp S share a sequence homology >80%, but the hyperthermophilic variant, Csp C, is more stable than its thermophilic counterpart. Perl and Schmidt generated mutants in Csp S by directing them to Csp C sequence (see Fig. S1) (Perl & Schmid, 2001). Our results show that Csp S stabilizing mutations were performed over low connected and highly flexible residues, residues with low SDRI. Most stabilizing mutations of Csp S, that were directed to the sequence of the hyperthermophilic variant- Csp C, were those incorporated on the surface bonding flexible residues through the formation of salt bridges. Likewise, Tokuriki and Tawfik reported that mutations on surface residues of their analyzed proteins resulted in low destabilizing effects while mutations on core residues caused stronger destabilizing effects (Tokuriki & Tawfik, 2009).

We are particularly interested in the characterization of 6aJL2, an immunoglobulin light chain variable domain, based on the fact that 6a is the most implicated germ line in AL amyloidosis disease (Comenzo et al., 2001). Destabilizing mutations of 6aJL2 enhances its propensity to generate protein fibers. Strikingly, the crystallographic structures of destabilizing mutants exhibited a low RMSD difference when overlapped against the wild-type structure (Hernández-Santoyo et al., 2010). To experimentally demonstrate our hypothesis, we selected four residues with low RSA values but different SDRI to perform single-point mutations: Gln6Asn, Arg24His, Tyr36Phe, and Ile29Gly (SDRI 0.5055, 0.3273, 0.3274, and 0.2057, respectively). Despite that the greatest destabilization impact was detected when the size of the lateral chain was minimized, as observed with mutant Ile29Gly (RSA 2.2%, $\Delta\Delta G = -3.61 \text{ kcal mol}^{-1}$), mutant Gln6Asn (RSA 12%, $\Delta\Delta G = -3.34 \text{ kcal mol}^{-1}$) was more relevant due to the neutrality of the change. The lateral chain size-reduction by one methylene had a remarkable impact on the protein stability. We selected residue Gln6 in 6aJL2 because it is a highly conserved residue in immunoglobulin light chains sequence alignments (Williams et al., 1996). Opposite to this neutral mutation, Tyr36Phe did not affect the structure suggesting that the hydroxyl group of the tyrosine is not playing a relevant role (RSA 0.6%, $\Delta\Delta G = -0.04 \text{ kcal mol}^{-1}$). While Arg24His (RSA 10%, $\Delta\Delta G = -2.70 \text{ kcal mol}^{-1}$) performed in this work was more destabilizing than reported mutation Arg24Gly ($\Delta\Delta G = -1.52 \text{ kcal mol}^{-1}$) (Del Pozo-Yauner et al., 2008), other effects should be considered like Phe2 reorientation to the upper hydrophobic core to compensate the absence of the Arg24 guanidinium group (Del Pozo-Yauner et al., 2014). Interestingly, Phe2 is not among the residues with higher SDRI in 6aJL2 pointing out that compensatory effects might be attributable to the flexibility of the lateral chain.

This prompted us to assess about the role of residues or protein segments with the lowest SDRIs. Since proteins can interact with other proteins and other macromolecular, important residues not only maintain the connectivity along the tertiary structure but also maintaining quaternary structure as seen in the complex of the estrogen receptor with DNA by modifying the affinity (Deegan *et al.*, 2010). We found that the most flexible and unconnected regions were associated with active site functions, as in the case of apoflavodoxin, to which the cofactor, flavin mononucleotide, binds (Genzor *et al.*, 1996). In chymotrypsin inhibitor, the larger loop, which is flexible and unconnected, harbors the active site (Jackson *et al.*, 1993). However, not all mutations can enhance the stability while preserving the original function of a protein. In the case of T4 lysozyme, some mutants were found to be more stable but resulted in losses of the protein's original function (Shoichet *et al.*, 1995). If the purpose is to modify the function of a protein, potential mutations should be assessed by other means, such as evolutionary multiple sequence alignment (Alexander *et al.*, 2009; Halabi *et al.*, 2009). If the aim is only to increase the protein stability, a good approach could be locating the less structurally important residues and generating changes that benefit the formation of salt bridges as shown in apoflavodoxin. Mutations localized on the surface and designed to establish salt bridges were able to increase the overall stability in apoflavodoxin (Campos *et al.*, 2004b). Remarkably, these mutations were performed on the low-scored SDRI residues.

Since folding requires certain flexibility degree, we analyzed the relationship between kinetic data with SDRIs. Other experiments performed in chymotrypsin inhibitor were folding/unfolding kinetics (Fig. 7) (Itzhaki, Otzen & Fersht, 1995). In this case, the SDRIs showed a better correlation with the $\Delta\Delta G$ unfolding kinetic values (75% from the PCA analysis) than the $\Delta\Delta G$ folding kinetic values (57% from the PCA analysis). We would like to reiterate that folding is a dynamic rearrangement of the network because longer times and other conditions are required for better simulations of the protein folding/unfolding pathways.

Our last validation was performed over an intrinsically disordered peptide, resolved by NMR in solution, which only showed a small structured portion of the peptide (Chandrashekar *et al.*, 2015). Before applying the network strategy described here, the RMSD of each one of the 20 frames was calculated by comparing them against the first frame (Fig. 8A). It was evident that there is a remarkable structural motion freedom at the N-terminus, even higher than those obtained by molecular dynamics of the other proteins described here (Figs. 2, 4 and 8B). In order to follow the methodology proposed here, RMSF values of the main chain of each residue were calculated using RCI server (Berjanskii & Wishart, 2013). An important difference compared to globular proteins was the distribution statistics of the SDRIs, having the highest value dispersion dissimilar to a normally distributed population (Table 1 and Fig. 4H). The most connected and less flexible residues were located in the hairpin, shown in Fig. 8C, which contributes to the scaffolding functionality allowing phosphorylation of Ser211. Also, the plasticity of the disordered N-terminus would enable this peptide to bind multiple components of the signaling pathway in which it is involved (Chandrashekar *et al.*, 2015).

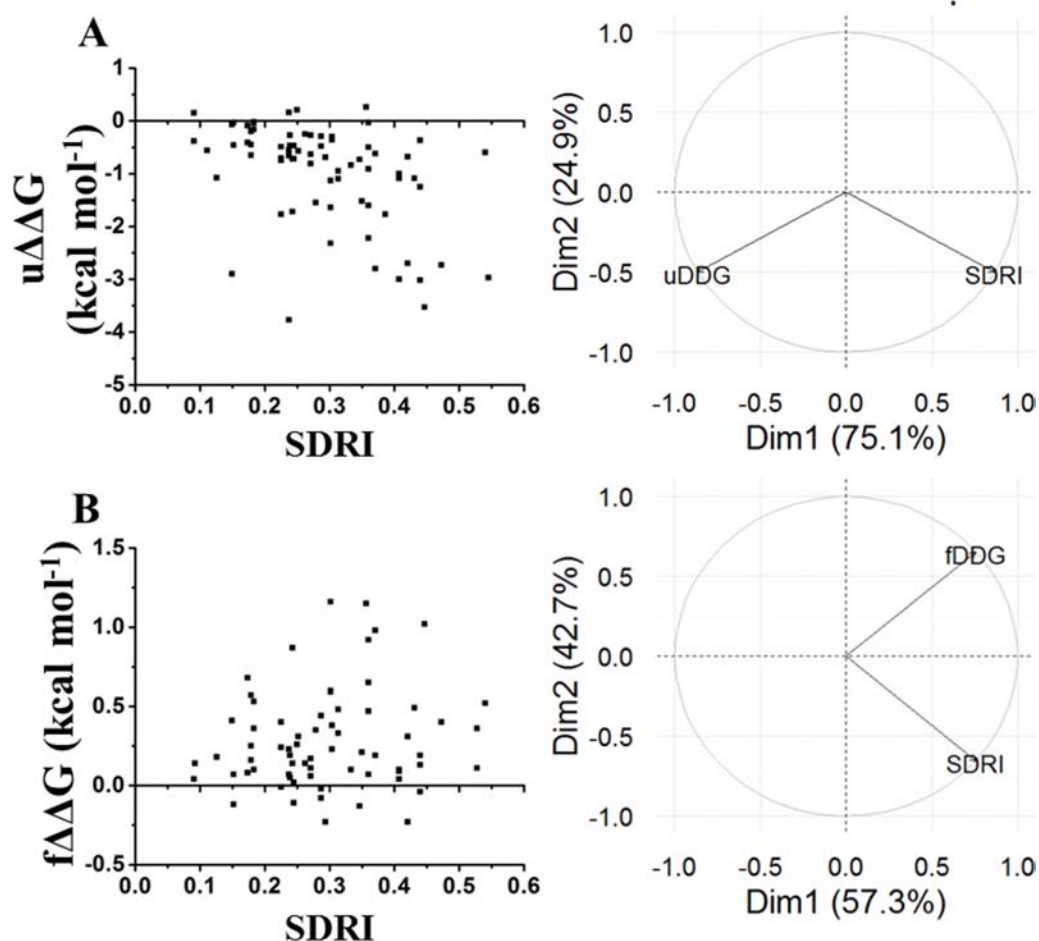


Figure 7 Comparison of SDRIs with the kinetic unfolding and kinetic refolding processes of the chymotrypsin inhibitor. Unfolding kinetics exhibited a better correlation with SDRIs as shown in circular bi-plots. DIM1 refers to SDRIs ($H_p \cdot RMSF_i^{-1}$), and DIM2 refers to the experimental values, $\Delta\Delta G$ (DDG) or KD percent difference (see File S3). (A) $\Delta\Delta G$ Unfolding kinetics, $uDDG$ in circular biplot; (B) $\Delta\Delta G$ Refolding kinetics, $fDDG$ in circular biplot.

It should be noticed that applying this methodology on proteins with high structural motion might not provide enough information to predict which residues will interact with ligands. Unstructured peptides can be analyzed by other means (Kosol *et al.*, 2013; Shaw *et al.*, 2010). We simulated the unbound estrogen receptor and, despite the drastic change on the ranking position and the SDRIs, it is not evident which segments of the estrogen receptor dimer will recognize precise DNA sequence (see File S4). Furthermore, we decided to extend 6aJL2 simulation time to 50 ns and we only detected subtle, but not significant, changes on SDRIs for this globular protein (see File S4). Thus, appropriate molecular dynamics simulation accomplished on globular proteins is an essential step for this methodology. Moreover, our results agreed with sequence-based predictors that look up for the intrinsically disordered segments regardless protein complexity. SDRI values displayed a versatile mathematical parameter since function (1-SDRI) might be applicable to highlight disordered segments. Such disordered segments may increase the

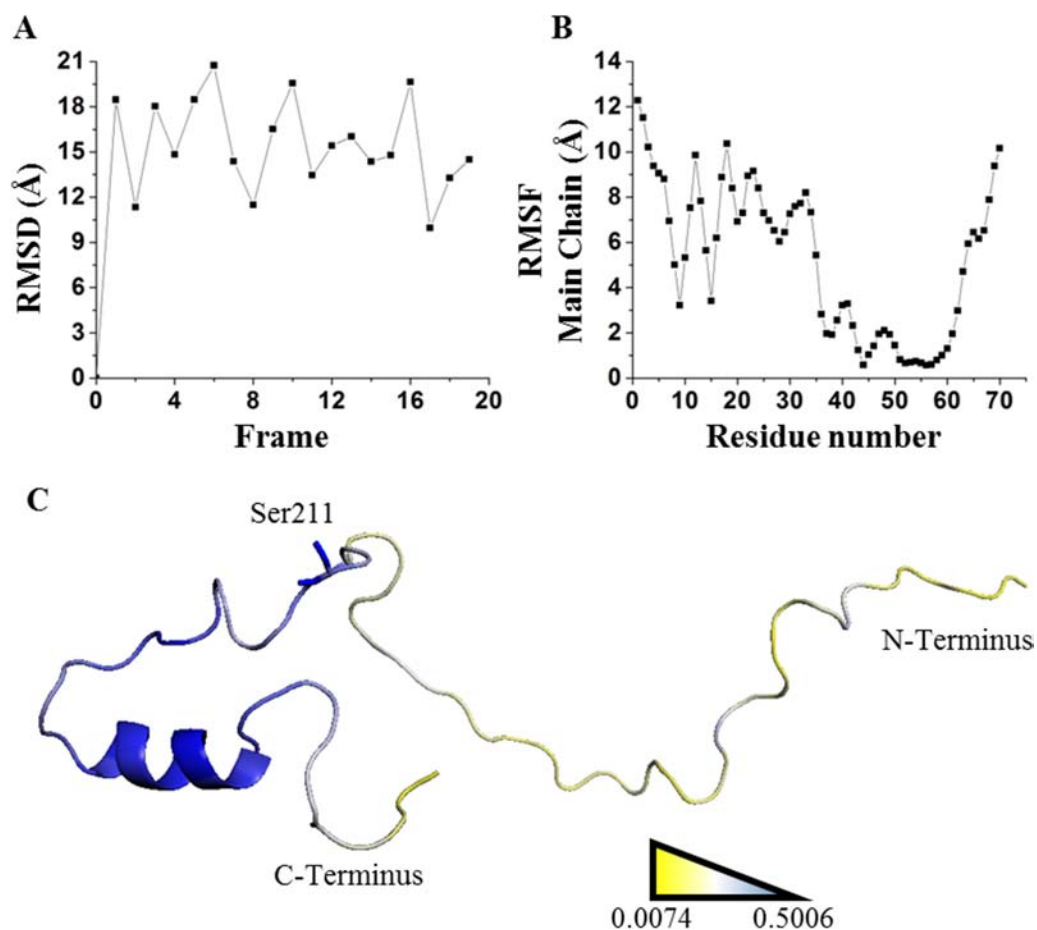


Figure 8 Applying SDRI approach over an unstructured peptide. Largely disordered N-terminus of suppressor of cytokine signaling 5 mammalian suppressor of Janus Kinase interaction region. (A) RMSD values by comparing each frame against first frame from PDB 2N34. (B) RMSF values of each residue showing that hairpin domain is more rigid than the rest of the protein. (C) Ranking heat map where the highest SDRI is shown in blue and the lowest SDRI is represented in yellow (see [File S2](#)).

capability of organisms to tolerate environmental challenges by diversifying the properties of their proteins to recognize several molecular partners such as cofactors, DNA, or other proteins. Higher SDRI values imply higher probabilities to modify protein stability, but there is a higher tendency to destabilization. As reported by *Tokuriki & Tawfik (2009)*, one of the evolutionary implications of protein destabilization is that other functions or adaptations may be achieved. Finally, under the scenario of analyzing uncommon foldings or even structures bearing non-natural amino acids, our method might be suitable to assess structure stability since it does not depend on previous information such as an evolutionary multiple sequence alignment. In conclusion, we have validated a method for the analysis of globular proteins by ranking each one of their residues according to their structural relevance from a theoretical score- SDRI.

ACKNOWLEDGEMENTS

We thank J Osuna-Quintero, JL Martínez-Morales, and M Aldana-González for the valuable discussions and advice, C Torres-Duarte for critically reviewing the manuscript, the Oligonucleotide Synthesis Unit of the Institute of Biotechnology, T Olamendi-Portugal for the DNA sequencing, and G Corzo-Burguete for facilitating the use of the computer hardware that was used for the molecular dynamics simulations

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

ODLM and MIVV were supported by CONACYT 177224 and 185179, respectively. This work was supported by grants from DGAPA IN 217510. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

CONACYT: 177224, 185179.

DGAPA IN: 217510.

Competing Interests

Vladimir N. Uversky is an Academic Editor for PeerJ.

Author Contributions

- Oscar D. Luna-Martínez conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Abraham Vidal-Limón performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Miryam I. Villalba-Velázquez performed the experiments, analyzed the data, reviewed drafts of the paper.
- Rosalba Sánchez-Alcalá performed the experiments.
- Ramón Garduño-Juárez analyzed the data, contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Vladimir N. Uversky analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Baltazar Becerril analyzed the data, wrote the paper, reviewed drafts of the paper.

Data Availability

The following information was supplied regarding data availability:

The raw data has been supplied as [Supplemental Information](#).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2136#supplemental-information>.

REFERENCES

- Abdi H, Williams LJ. 2010. Principal component analysis. *WIREs Computational Statistics* 2:433–459 DOI 10.1002/wics.101.
- Alexander PA, He Y, Chen Y, Orban J, Bryan PN. 2009. A minimal sequence code for switching protein structure and function. *Proceedings of the National Academy of Sciences of the United States of America* 106:21149–21154 DOI 10.1073/pnas.0906408106.
- Amitai G, Shemesh A, Sitbon E, Shklar M, Netanelly D, Venger I, Pietrokovski S. 2004. Network analysis of protein structures identifies functional residues. *Journal of Molecular Biology* 344:1135–1146 DOI 10.1016/j.jmb.2004.10.055.
- Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, Grosdidier A, Hernandez C, Ioannidis V, Kuznetsov D, Liechti R, Moretti S, Mostaguir K, Redaschi N, Rossier G, Xenarios I, Stockinger H. 2012. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* 40:W597–W603 DOI 10.1093/nar/gks400.
- Baldwin EP, Matthews BW. 1994. Core-packing constraints, hydrophobicity and protein design. *Current Opinion Biotechnology* 5:396–402 DOI 10.1016/0958-1669(94)90048-5.
- Banci L, Bertini I, Calderone V, Cramaro F, Del Conte R, Fantoni A, Mangani S, Quattrone A, Viezzoli MS. 2004. A prokaryotic superoxide dismutase paralog lacking two Cu ligands: from largely unstructured in solution to ordered in the crystal. *Proceedings of the National Academy of Sciences of the United States of America* 102:7541–7546 DOI 10.1073/pnas.0502450102.
- Becktel WJ, Schellman JA. 1987. Protein stability curves. *Biopolymers* 26:1859–1877 DOI 10.1002/bip.360261104.
- Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola AS, Haak JR. 1984. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* 81:3684–3690 DOI 10.1063/1.448118.
- Berjanskii MV, Wishart DS. 2013. A simple method to measure protein side-chain mobility using NMR chemical shifts. *Journal of the American Chemical Society* 135:14536–14539 DOI 10.1021/ja407509z.
- Böde C, Kovács IA, Szalay MS, Palotai R, Korcsmáros T, Csermely P. 2007. Network analysis of protein dynamics. *FEBS Letters* 581:2776–2782 DOI 10.1016/j.febslet.2007.05.021.
- Campos LA, Bueno M, Lopez-Llano J, Jiménez MA, Sancho J. 2004a. Structure of stable protein folding intermediates by equilibrium ϕ -analysis: the apoflavodoxin thermal intermediate. *Journal of Molecular Biology* 344:239–255 DOI 10.1016/j.jmb.2004.08.081.
- Campos LA, Garcia-Mira MM, Godoy-Ruiz R, Sanchez-Ruiz JM, Sancho J. 2004b. Do proteins always benefit from a stability increase? Relevant and residual stabilization in a three-state protein by charge optimization. *Journal of Molecular Biology* 344:223–237 DOI 10.1016/j.jmb.2004.09.047.

- Chandrashekar IR, Mohanty B, Linossi EM, Dagley LF, Leung EWW, Murphy JM, Babon JJ, Nicholson SE, Norton RS. 2015.** Structure and functional characterization of the conserved JAK interaction region in the intrinsically disordered N-terminus of SOCS5. *Biochemistry* **54**:4672–4682 DOI [10.1021/acs.biochem.5b00619](https://doi.org/10.1021/acs.biochem.5b00619).
- Chen ME, Cang HX, Nymeyer H. 2007.** *Nymeyer's group. Inst. Mol. Biol.*, Florida State University. Available at <http://noch.sourceforge.net> (accessed 27 November 2015).
- Comenzo RL, Zhang Y, Martinez C, Osman K, Herrera GA. 2001.** The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig V(L) germ line gene use and clonal plasma cell burden. *Blood* **98**:714–720 DOI [10.1182/blood.V98.3.714](https://doi.org/10.1182/blood.V98.3.714).
- Costa LF, Rodrigues FA, Travieso G, Villas Boas PR. 2007.** Characterization of complex networks: a survey of measurements. *Advances in Physics* **56**:167–242 DOI [10.1080/00018730601170527](https://doi.org/10.1080/00018730601170527).
- Creighton TE. 1990.** *Protein structure: a practical approach*. Oxford: IRL Press.
- Darby N, Creighton TE. 1993.** Protein structure. In: *Focus*. New York: IRL Press, Oxford University Press Inc., p. 4.
- Deegan BJ, Seldeen KL, McDonald CB, Bhat V, Farooq A. 2010.** Binding of the ER α nuclear receptor to DNA is coupled to proton uptake. *Biochemistry* **49**:5978–5988 DOI [10.1021/bi1004359](https://doi.org/10.1021/bi1004359).
- Del Pozo-Yauner L, Ortiz E, Sánchez R, Sánchez-López R, Güereca L, Murphy CL, Allen A, Wall JS, Fernández-Velasco DA, Solomon A, Becerril B. 2008.** Influence of the germline sequence on the thermodynamic stability and fibrillogenicity of human lambda 6 light chains. *Proteins Structure Function and Bioinformatics* **72**:684–692 DOI [10.1002/prot.21934](https://doi.org/10.1002/prot.21934).
- Del Pozo-Yauner L, Wall JS, González Andrade M, Sánchez-López R, Rodríguez-Ambríz SL, Pérez-Carreón JI, Ochoa-Leyva A, Fernández-Velasco DA. 2014.** The N-terminal strand modulates immunoglobulin light chain fibrillogenesis. *Biochemical and Biophysical Research Communications* **443**:495–499 DOI [10.1016/j.bbrc.2013.11.123](https://doi.org/10.1016/j.bbrc.2013.11.123).
- Demetrius L, Manke T. 2004.** Robustness and network evolution—an entropic principle. *Physica A* **346**:682–696 DOI [10.1016/j.physa.2004.07.011](https://doi.org/10.1016/j.physa.2004.07.011).
- Dror RO, Dirks RM, Grossman JP, Xu H, Shaw DE. 2012.** Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics* **41**:429–452 DOI [10.1146/annurev-biophys-042910-155245](https://doi.org/10.1146/annurev-biophys-042910-155245).
- Edgar RC, Batzoglou S. 2006.** Multiple sequence alignment. *Current Opinion in Structural Biology* **16**:368–373 DOI [10.1016/j.sbi.2006.04.004](https://doi.org/10.1016/j.sbi.2006.04.004).
- Eftink MR. 1995.** Use of multiple spectroscopic methods to monitor equilibrium unfolding of proteins. *Methods in Enzymology* **259**:487–512 DOI [10.1016/0076-6879\(95\)59058-7](https://doi.org/10.1016/0076-6879(95)59058-7).
- Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. 1995.** A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**:8577–85933 DOI [10.1063/1.470117](https://doi.org/10.1063/1.470117).

- Fan X, Kurgan L. 2014.** Accurate prediction of disorder in protein chains with a comprehensive and empirically designed consensus. *Journal of Biomolecular Structure and Dynamics* **32**:448–464 DOI [10.1080/07391102.2013.775969](https://doi.org/10.1080/07391102.2013.775969).
- Genzor CG, Perales-Alcón A, Sancho J, Romero A. 1996.** Closure of a tyrosine/tryptophan aromatic gate leads to a compact fold in apoflavodoxin. *Nature Structural & Molecular Biology* **3**:329–332 DOI [10.1038/nsb0496-329](https://doi.org/10.1038/nsb0496-329).
- Greene LH, Higman VA. 2003.** Uncovering network systems within protein structures. *Journal of Molecular Biology* **334**:781–791 DOI [10.1016/j.jmb.2003.08.061](https://doi.org/10.1016/j.jmb.2003.08.061).
- Gribenko AV, Makhatadze GI. 2007.** Role of the charge–charge interactions in defining stability and halophilicity of the CspB proteins. *Journal of Molecular Biology* **366**:842–856 DOI [10.1016/j.jmb.2006.11.061](https://doi.org/10.1016/j.jmb.2006.11.061).
- Guex N, Peitsch MC. 1997.** SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714–2723 DOI [10.1002/elps.1150181505](https://doi.org/10.1002/elps.1150181505).
- Halabi N, Rivoire O, Leibler S, Ranganathan R. 2009.** Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**:774–786 DOI [10.1016/j.cell.2009.07.038](https://doi.org/10.1016/j.cell.2009.07.038).
- Hernández-Santoyo A, Del Pozo Yauner L, Fuentes-Silva D, Ortiz E, Rudiño-Piñera E, Sánchez-López R, Horjales E, Becerril B, Rodríguez-Romero A. 2010.** A single mutation at the sheet switch region results in conformational changes favoring lambda6 light-chain fibrillogenesis. *Journal of Molecular Biology* **396**:280–292 DOI [10.1016/j.jmb.2009.11.038](https://doi.org/10.1016/j.jmb.2009.11.038).
- Itzhaki LS, Otzen DE, Fersht AR. 1995.** The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation–condensation mechanism for protein folding. *Journal of Molecular Biology* **254**:260–288 DOI [10.1006/jmbi.1995.0616](https://doi.org/10.1006/jmbi.1995.0616).
- Jackson SE, Moracci M, ElMasry N, Dobson CM, Fersht AR. 1993.** Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**:11259–11269 DOI [10.1021/bi00093a001](https://doi.org/10.1021/bi00093a001).
- Jaenicke R, Böhm G. 1998.** The stability of proteins in extreme environments. *Current Opinion in Structural Biology* **8**:738–748 DOI [10.1016/S0959-440X\(98\)80094-8](https://doi.org/10.1016/S0959-440X(98)80094-8).
- Jamroz M, Kolinski A, Kihara D. 2012.** Structural features that predict real-value fluctuations of globular proteins. *Proteins: Structure, Function, and Bioinformatics* **80**:1425–1435 DOI [10.1002/prot.24040](https://doi.org/10.1002/prot.24040).
- Karplus M, Kuriyan J. 2005.** Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **102**:6679–6685 DOI [10.1073/pnas.0408930102](https://doi.org/10.1073/pnas.0408930102).
- Kepleis JL, Lindorff-Larsen K, Dror RO, Shaw DE. 2009.** Long-timescale molecular dynamics simulations of protein structure and function. *Current Opinion in Structural Biology* **19**:120–127 DOI [10.1016/j.sbi.2009.03.004](https://doi.org/10.1016/j.sbi.2009.03.004).
- Kosol S, Contreras-Martos S, Cedeño C, Tompa P. 2013.** Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**:10802–10828 DOI [10.3390/molecules180910802](https://doi.org/10.3390/molecules180910802).

- Kumar S, Nussinov R. 2001.** How do thermophilic proteins deal with heat? *Cellular and Molecular Life Sciences* **58**:1216–1233 DOI [10.1007/PL00000935](https://doi.org/10.1007/PL00000935).
- Lei H, Duan Y. 2004.** The role of plastic β -hairpin and weak hydrophobic core in the stability and unfolding of a full sequence design protein. *The Journal of Chemical Physics* **121**:12104–12111 DOI [10.1063/1.1822916](https://doi.org/10.1063/1.1822916).
- Li J, Wang J, Wang W. 2008.** Identifying folding nucleus based on residue contact networks of proteins. *Proteins: Structure, Function, and Bioinformatics* **71**:1899–1907 DOI [10.1002/prot.21891](https://doi.org/10.1002/prot.21891).
- Milla ME, Brown BM, Sauer RT. 1994.** Protein stability effects of a complete set of alanine substitutions in Arc repressor. *Nature Structural & Molecular Biology* **1**:518–523 DOI [10.1038/nsb0894-518](https://doi.org/10.1038/nsb0894-518).
- Mori T, Okamoto Y. 2009.** Folding simulations of gramicidin A into the β -helix conformations: simulated annealing molecular dynamics study. *The Journal of Chemical Physics* **131**:165103 DOI [10.1063/1.3247578](https://doi.org/10.1063/1.3247578).
- Pace CN, Fu H, Fryar KL, Landua J, Trevino SR, Shirley BA, McNutt Hendricks M, Iimura S, Gajiwala K, Scholtz JM, Grimsley GR. 2011.** Contribution of hydrophobic interactions to protein stability. *Journal of Molecular Biology* **408**:514–528 DOI [10.1016/j.jmb.2011.02.053](https://doi.org/10.1016/j.jmb.2011.02.053).
- Peng ZL, Kurgan L. 2012.** Comprehensive comparative assessment of in-silico predictors of disordered regions. *Current Protein & Peptide Science* **13**:6–18 DOI [10.2174/138920312799277938](https://doi.org/10.2174/138920312799277938).
- Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z. 2005.** Optimizing long intrinsic disorder predictors with protein evolutionary information. *Journal of Bioinformatics and Computational Biology* **3**:35–60 DOI [10.1142/S0219720005000886](https://doi.org/10.1142/S0219720005000886).
- Perl D, Schmid FX. 2001.** Electrostatic stabilization of a thermophilic cold shock protein. *Journal of Molecular Biology* **313**:343–357 DOI [10.1006/jmbi.2001.5050](https://doi.org/10.1006/jmbi.2001.5050).
- Phillips A, Janies D, Wheeler W. 2000.** Multiple sequence alignment in phylogenetic analysis. *Molecular Phylogenetics and Evolution* **16**:317–330 DOI [10.1006/mpev.2000.0785](https://doi.org/10.1006/mpev.2000.0785).
- Pierri CL, Parisi G, Porcelli V. 2010.** Computational approaches for protein function prediction: a combined strategy from multiple sequence alignment to molecular docking-based virtual screening. *Biochimica et Biophysica Acta* **1804**:1695–1712 DOI [10.1016/j.bbapap.2010.04.008](https://doi.org/10.1016/j.bbapap.2010.04.008).
- Prodromou C, Pearl LH. 1992.** Recursive PCR: a novel technique for total gene synthesis. *Protein Engineering* **5**:827–829 DOI [10.1093/protein/5.8.827](https://doi.org/10.1093/protein/5.8.827).
- R Core Development Team. 2013.** *R: a language environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at <http://R-project.org> (accessed 27 November 2015).
- Reed CJ, Lewis H, Trejo E, Winston V, Caryn E. 2013.** Protein adaptations in archaeal extremophiles. *Archaea* **2013**: Article 373275 DOI [10.1155/2013/373275](https://doi.org/10.1155/2013/373275).
- Rodríguez-Rodríguez ER, Ledezma-Candanoza LM, Contreras-Ferrat LG, Olamendi-Portugal T, Possani LD, Becerril B, Riaño-Umbarila L. 2012.** A single mutation

- in framework 2 of the heavy variable domain improves the properties of a diabody and a related single chain antibody. *Journal of Molecular Biology* **423**:337–350 DOI [10.1016/j.jmb.2012.07.007](https://doi.org/10.1016/j.jmb.2012.07.007).
- Rothschild LJ, Mancinelli RL. 2001.** Life in extreme environments. *Nature* **409**:1092–1101 DOI [10.1038/35059215](https://doi.org/10.1038/35059215).
- Sanger F, Nicklen S, Coulson AR. 1977.** DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**:5463–5467 DOI [10.1073/pnas.74.12.5463](https://doi.org/10.1073/pnas.74.12.5463).
- Schier R, Marks JD, Wolf EJ, Apell G, Wong C, McCartney JE, Bookman MA, Huston JS, Houston LL, Weiner LM, Adams GP. 1995.** *In vitro* and *in vivo* characterization of a human anti-c-e-erbB-e single-chain Fv isolated from a filamentous phage antibody library. *Immunotechnology* **1**:73–81 DOI [10.1016/1380-2933\(95\)00007-0](https://doi.org/10.1016/1380-2933(95)00007-0).
- Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Yibing S, Wriggers W. 2010.** Atomic-level characterization of the structural dynamics of proteins. *Science* **330**:341–346 DOI [10.1126/science.1187409](https://doi.org/10.1126/science.1187409).
- Shoichet BK, Baase WA, Kuroki R, Matthews BW. 1995.** A relationship between protein stability and protein function. *Proceedings of the National Academy of Sciences of the United States of America* **92**:452–456 DOI [10.1073/pnas.92.2.452](https://doi.org/10.1073/pnas.92.2.452).
- Tokuriki N, Stricher F, Serrano L, Tawfik DS. 2008.** How protein stability and new functions trade off. *PLoS Computational Biology* **4**(2):e1000002 DOI [10.1371/journal.pcbi.1000002](https://doi.org/10.1371/journal.pcbi.1000002).
- Tokuriki N, Tawfik DS. 2009.** Stability effects of mutations and protein evolvability. *Current Opinion in Structural Biology* **19**:596–604 DOI [10.1016/j.sbi.2009.08.003](https://doi.org/10.1016/j.sbi.2009.08.003).
- Van den Bedem H, Bhabha G, Yan K, Wright PE, Fraser JS. 2013.** Automated identification of functional dynamic contact networks from X-ray crystallography. *Nature Methods* **10**:896–902 DOI [10.1038/nmeth.2592](https://doi.org/10.1038/nmeth.2592).
- Vendruscolo M, Dokholyan NV, Paci E, Karplus M. 2002.** Small-world view of the amino acids that play a key role in protein folding. *Physical Review E* **65**:061910 DOI [10.1103/PhysRevE.65.061910](https://doi.org/10.1103/PhysRevE.65.061910).
- Voss NR, Gerstein M. 2005.** Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more tightly. *Journal of Molecular Biology* **346**:477–492 DOI [10.1016/j.jmb.2004.11.072](https://doi.org/10.1016/j.jmb.2004.11.072).
- Williams SC, Frippiat JP, Tomlinson IM, Ignatovich O, Lefranc MP, Winter G. 1996.** Sequence evolution of the human germline V λ repertoire. *Journal of Molecular Biology* **264**:220–232 DOI [10.1006/jmbi.1996.0636](https://doi.org/10.1006/jmbi.1996.0636).
- Wunderlich M, Martin A, Schmid FX. 2005.** Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions. *Journal of Molecular Biology* **347**:1063–1076 DOI [10.1016/j.jmb.2005.02.014](https://doi.org/10.1016/j.jmb.2005.02.014).
- Wunderlich M, Schmid FX. 2006.** The correlation between protein stability and dipole moment: a critical test. *Protein Engineering, Design and Selection* **19**:355–358 DOI [10.1093/protein/gzl019](https://doi.org/10.1093/protein/gzl019).

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 2CI2		Grade (Number of Neighbors)
Residue		
ASN	1	2
LEU	2	4
LYS	3	9
THR	4	7
GLU	5	8
TRP	6	16
PRO	7	7
GLU	8	9
LEU	9	11
VAL	10	11
GLY	11	7
LYS	12	11
SER	13	10
VAL	14	11
GLU	15	7
GLU	16	8
ALA	17	13
LYS	18	12
LYS	19	8
VAL	20	12
ILE	21	16
LEU	22	12
GLN	23	8
ASP	24	9
LYS	25	15
PRO	26	7
GLU	27	7
ALA	28	9
GLN	29	7
ILE	30	9
ILE	31	8
VAL	32	7
LEU	33	10
PRO	34	7
VAL	35	9
GLY	36	6
THR	37	7
ILE	38	4

Protein 2CI2		Grade (Number of Neighbors)
Residue		
VAL	39	7
THR	40	6
MET	41	4
GLU	42	7
TYR	43	7
ARG	44	7
ILE	45	9
ASP	46	9
ARG	47	14
VAL	48	15
ARG	49	12
LEU	50	14
PHE	51	14
VAL	52	13
ASP	53	8
LYS	54	4
LEU	55	4
ASP	56	10
ASN	57	10
ILE	58	16
ALA	59	10
GLU	60	8
VAL	61	9
PRO	62	15
ARG	63	10
VAL	64	13
GLY	65	9

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 2C12	Grade (Number of Neighbors)	Guanidinium chloride denaturation		
		m UF (kcal mol ⁻¹)	C _m (M)	ΔΔG° (kcal mol ⁻¹)
Wild Type (WT)	-	1.9	4	-
KA3	9	1.77	3.72	-0.53
KM3	9	1.88	3.66	-0.65
TA4	7	1.83	3.56	-0.84
TV4	7	1.67	3.83	-0.32
TG4	7	1.95	3.4	-1.14
PA7	7	3.46	3.19	-1.54
EA8	9	1.75	3.76	-0.46
EQ8	9	1.91	3.68	-0.61
LA9	11	2.15	2.62	-2.62
KA12	11	1.52	4.22	0.42
SG13	10	2.1	3.59	-0.78
SA13	10	1.86	3.54	-0.87
EQ15	7	1.87	3.85	-0.29
ED15	7	1.83	3.73	-0.51
EN15	7	2.01	3.64	-0.68
EQ16	8	1.84	3.76	-0.46
ED16	8	2.2	3.62	-0.72
EN16	8	2.18	3.45	-1.05
AG17	13	1.8	3.44	-1.06
KA18	12	1.73	3.75	-0.48
KG18	12	1.99	2.8	-2.28
KA19	8	1.61	4.11	0.21
KG19	8	1.73	3.49	-0.97
VA20	12	2.01	3.75	-0.48
IV21	16	1.99	3.33	-1.27

Mutant 2C12	Grade (Number of Neighbors)	Guanidinium chloride denaturation		
		m UF (kcal mol ⁻¹)	C _m (M)	ΔΔG° (kcal mol ⁻¹)
LA22	12	1.93	3.32	-1.29
LG22	12	1.68	3.29	-1.35
QA23	8	1.77	3.99	-0.02
QG23	8	1.81	3.69	-0.59
DA24	9	1.87	3.51	-0.93
KA25	15	1.57	3.67	-0.63
KG25	15	1.75	2.36	-3.12
PA26	7	2.07	3.09	-1.73
EA27	7	1.78	3.83	-0.32
IV30	9	1.99	3.43	-1.08
IA30	9	2.07	1.99	-3.82
IV31	8	1.77	4.04	0.08
IA31	8	2.22	2.91	-2.07
IG31	8	2.13	2.19	-3.44
IT31	8	1.97	3.31	-1.31
LA32	7	2.11	2.78	-2.32
LI32	7	1.7	3.87	-0.25
LV32	7	1.76	3.74	-0.49
PA33	10	1.79	3.91	-0.17
VT34	7	1.65	3.47	-1.01
VA34	7	1.7	3.67	-0.63
VG34	7	2	2.75	-2.38
TS36	6	1.66	3.99	-0.02
TV36	6	1.67	3.61	-0.74
TA36	6	1.5	4.12	0.23
IA38	4	1.59	3.98	-0.04
VA39	7	2.14	3.24	-1.44

Reference. J. Mol. Biol. (1995) 254: 260-288

$$\Delta\Delta G^\circ = m_{WT} * (C_m^{MUT} - C_m^{WT})$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 2C12	Grade (Number of Neighbors)	Guanidinium chloride denaturation		
		m UF (kcal mol ⁻¹)	C _m (M)	ΔΔG° (kcal mol ⁻¹)
TA40	6	1.89	3.63	-0.70
TD40	6	1.87	4.01	0.02
EA42	7	1.78	3.64	-0.68
RA44	7	1.78	3.7	-0.57
DA46	14	1.83	3.59	-0.78
VA48	15	1.76	1.46	-4.83
LA50	14	1.98	2.02	-3.76
FL51	14	1.95	2.91	-2.07
FV51	14	2.24	2.77	-2.34
FA51	14	2.17	2.02	-3.76
VA52	13	2.15	2.98	-1.94
DA53	8	1.99	2.24	-3.34
DN53	8	1.6	4	0.00
ND57	10	1.8	3.38	-1.18
NA57	10	1.79	3.57	-0.82
IV58	16	1.82	4.1	0.19
IA58	16	1.93	1.79	-4.20
AG59	10	2.19	3.03	-1.84
VT61	9	1.63	3.8	-0.38
VA61	9	1.92	3.22	-1.48
VG61	9	2.61	2.33	-3.17
PA62	15	1.8	2.28	-3.27
VT64	13	1.79	3.41	-1.12
VA64	13	1.95	3.25	-1.43
VG64	13	2.1	2.2	-3.42

Reference. J. Mol. Biol. (1995) 254: 260-288

$$\Delta\Delta G^\circ = m_{WT} \cdot (C_m^{MUT} - C_m^{WT})$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 6aJL2		Grade (Number of Neighbors)
Residue		
ASN	1	5
PHE	2	10
MET	3	8
LEU	4	15
THR	5	8
GLN	6	16
PRO	7	9
HIS	8	7
SER	9	6
VAL	10	11
SER	11	7
GLU	12	12
SER	13	8
PRO	14	8
GLY	15	7
LYS	16	10
THR	17	9
VAL	18	12
THR	19	9
ILE	20	14
SER	21	12
CYX	22	15
THR	23	11
ARG	24	17
SER	25	7
SER	26	7
GLY	27	6
SER	28	8
ILE	29	15
ALA	30	5
SER	31	6
ASN	32	12
TYR	33	9
VAL	34	14
GLN	35	13
TRP	36	22
TYR	37	13
GLN	38	15

Protein 6aJL2		Grade (Number of Neighbors)
Residue		
GLN	39	11
ARG	40	10
PRO	41	4
GLY	42	4
SER	43	6
SER	44	7
PRO	45	7
THR	46	8
THR	47	10
VAL	48	14
ILE	49	16
TYR	50	12
GLU	51	8
ASP	52	13
ASN	53	7
GLN	54	8
ARG	55	11
PRO	56	9
SER	57	4
GLY	58	4
VAL	59	9
PRO	60	6
ASP	61	7
ARG	62	12
PHE	63	15
SER	64	9
GLY	65	11
SER	66	10
ILE	67	11
ASP	68	9
SER	69	7
SER	70	4
SER	71	6
ASN	72	14
SER	73	13
ALA	74	14
SER	75	13
LEU	76	16

Protein 6aJL2		Grade (Number of Neighbors)
Residue		
THR	77	13
ILE	78	14
SER	79	10
GLY	80	9
LEU	81	15
LYS	82	9
THR	83	7
GLU	84	7
ASP	85	16
GLU	86	12
ALA	87	10
ASP	88	10
TYR	89	19
TYR	90	14
CYX	91	13
GLN	92	14
SER	93	14
TYR	94	13
ASP	95	10
SER	96	4
SER	97	4
ASN	98	7
HIS	99	7
VAL	100	7
VAL	101	13
PHE	102	10
GLY	103	10
GLY	104	7
GLY	105	9
THR	106	14
LYS	107	12
LEU	108	15
THR	109	8
VAL	110	11
LEU	111	6

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 6aJL2	Grade (Number of Neighbors)	Thermal denaturation			Reference
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
Wild Type (WT)	-	323.05	86.2	-	Biochem. Biophys. Res. Commun. (2014) 443: 495-499
Phe2Ser	10	316.95	73.6	-1.63	
Phe2Leu	10	318.85	76.1	-1.12	
Phe2Pro	10	316.75	73.7	-1.68	
Phe2Trp	10	319.15	81.3	-1.04	
Arg24Gly	17	317.35	77.2	-1.52	Proteins (2008) 72: 684-692
Pro7Ser	9	316.15	79.5	-1.84	J. Mol. Biol. (2010) 396: 280-292
His8Pro	7	323.25	84.8	0.05	
His8Ser	7	322.65	84.4	-0.11	
Arg24His	17	312.95 ± 1.15	56.47 ± 5.65	-2.70	This work
Ile29Gly	15	309.53 ± 0.53	60.74 ± 2.55	-3.61	
Tyr37Phe	13	322.90 ± 0.31	87.23 ± 3.06	-0.04	
Gln6Asn	16	310.55 ± 0.30	10.39 ± 3.21	-3.34	

$$\Delta\Delta G^\circ = (T_m\text{MUT} - T_m\text{WT}) * (\Delta H^\circ\text{WT} / T_m\text{WT})$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1FTG		Grade (Number of Neighbors)
Residue		
SER	1	10
LYS	2	7
LYS	3	9
ILE	4	15
GLY	5	11
LEU	6	13
PHE	7	16
TYR	8	15
GLY	9	13
THR	10	11
GLN	11	8
THR	12	5
GLY	13	7
LYS	14	12
THR	15	15
GLU	16	10
SER	17	8
VAL	18	14
ALA	19	13
GLU	20	9
ILE	21	10
ILE	22	16
ARG	23	12
ASP	24	7
GLU	25	9
PHE	26	15
GLY	27	9
ASN	28	4
ASP	29	7
VAL	30	11
VAL	31	12
THR	32	11
LEU	33	10
HIS	34	10
ASP	35	10
VAL	36	12
SER	37	6
GLN	38	6

Protein 1FTG		Grade (Number of Neighbors)
Residue		
ALA	39	9
GLU	40	8
VAL	41	11
THR	42	6
ASP	43	11
LEU	44	15
ASN	45	9
ASP	46	7
TYR	47	15
GLN	48	9
TYR	49	9
LEU	50	15
ILE	51	16
ILE	52	14
GLY	53	11
CYS	54	16
PRO	55	12
THR	56	14
TRP	57	12
ASN	58	9
ILE	59	6
GLY	60	4
GLU	61	6
LEU	62	12
GLN	63	13
SER	64	6
ASP	65	11
TRP	66	16
GLU	67	10
GLY	68	8
LEU	69	13
TYR	70	12
SER	71	7
GLU	72	9
LEU	73	14
ASP	74	9
ASP	75	7
VAL	76	11

Protein 1FTG		Grade (Number of Neighbors)
Residue		
ASP	77	7
PHE	78	14
ASN	79	8
GLY	80	6
LYS	81	16
LEU	82	11
VAL	83	13
ALA	84	11
TYR	85	17
PHE	86	19
GLY	87	12
THR	88	13
GLY	89	12
ASP	90	13
GLN	91	14
ILE	92	8
GLY	93	5
TYR	94	13
ALA	95	11
ASP	96	8
ASN	97	10
PHE	98	20
GLN	99	16
ASP	100	13
ALA	101	10
ILE	102	18
GLY	103	12
ILE	104	11
LEU	105	16
GLU	106	14
GLU	107	10
LYS	108	11
ILE	109	14
SER	110	11
GLN	111	8
ARG	112	13
GLY	113	10

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1FTG		Grade (Number of Neighbors)
Residue		
GLY	114	10
LYS	115	8
THR	116	10
VAL	117	12
GLY	118	12
TYR	119	14
TRP	120	16
SER	121	8
THR	122	10
ASP	123	4
GLY	124	7
TYR	125	15
ASP	126	6
PHE	127	11
ASN	128	7
ASP	129	11
SER	130	12
LYS	131	9
ALA	132	12
LEU	133	12
ARG	134	12
ASN	135	4
GLY	136	8
LYS	137	11
PHE	138	17
VAL	139	15
GLY	140	11
LEU	141	15
ALA	142	11
LEU	143	15
ASP	144	14
GLU	145	11
ASP	146	8
ASN	147	9
GLN	148	12
SER	149	8
ASP	150	5
LEU	151	8

Protein 1FTG		Grade (Number of Neighbors)
Residue		
THR	152	13
ASP	153	8
ASP	154	8
ARG	155	17
ILE	156	15
LYS	157	10
SER	158	9
TRP	159	20
VAL	160	13
ALA	161	9
GLN	162	11
LEU	163	14
LYS	164	12
SER	165	9
GLU	166	10
PHE	167	15
GLY	168	6
LEU	169	7

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 1FTG	Grade (Number of Neighbors)	Thermal denaturation							Reference
		Tm1 (K)	$\Delta H^{\circ}1$ (kcal mol ⁻¹)	$\Delta \Delta G^{\circ}1$ (kcal mol ⁻¹)	Tm2 (K)	DH ^o 2 (kcal mol ⁻¹)	$\Delta \Delta G^{\circ}2$ (kcal mol ⁻¹)	Total $\Delta \Delta G^{\circ}$ (kcal mol ⁻¹)	
Wild Type (WT)	-	312	22	-	326	60	-	-	J. Mol. Biol. (2004) 344: 223-237
E20K	9	309.3	32.7	-0.19	331.2	66.2	0.96	0.77	
E40K	8	316.5	27	0.32	332.5	59.6	1.20	1.51	
E72K	9	312.8	2	0.06	333.6	68.4	1.40	1.46	
D75K	7	314.8	28.3	0.20	330	54.3	0.74	0.93	
D126K	6	320	21.8	0.56	327.5	39.7	0.28	0.84	
D150K	5	305.8	25.1	-0.44	326.9	48	0.17	-0.27	
WT	-	317.3	33.9	-	329	52.7	-	-	J. Mol. Biol. (2004) 344: 239-255
L6A	13	305.5	28.7	-1.26	319.1	23.8	-1.59	-2.85	
I22V	16	311	33	-0.67	324.7	37.3	-0.69	-1.36	
V31A	12	312.6	30.9	-0.50	324.2	41.5	-0.77	-1.27	
D43A	11	317.7	34.1	0.04	329.6	49.5	0.10	0.14	
I51V	16	313.2	31.2	-0.44	324.3	38.8	-0.75	-1.19	
I52V	14	315.9	39.8	-0.15	324.9	54.5	-0.66	-0.81	
S71A	7	316.3	34.5	-0.11	329.3	55	0.05	-0.06	
A84G	11	313.3	34.4	-0.43	323.6	31.9	-0.86	-1.29	
N97A	10	314.9	28.6	-0.26	328.3	48.7	-0.11	-0.37	
Q99A	16	322.5	52.7	0.56	331.3	56.2	0.37	0.92	
I104V	11	313.3	29.5	-0.43	327.9	54.5	-0.18	-0.60	
S110A	11	315.6	32.5	-0.18	328.7	44.3	-0.05	-0.23	
V117A	12	309.3	36.9	-0.85	322.6	33.7	-1.03	-1.88	
T122S	10	323.2	35.5	0.63	330.3	50.7	0.21	0.84	
V139A	15	310.9	29	-0.68	327.5	50.8	-0.24	-0.92	
L143A	15	315.8	33.3	-0.16	328.9	50	-0.02	-0.18	
I156V	15	308.8	25.3	-0.91	320.3	29.9	-1.39	-2.30	
V160A	13	309.1	27.7	-0.88	322.4	49.2	-1.06	-1.93	

$$\Delta \Delta G = (Tm_{MUT} - Tm_{WT}) * (\Delta H_{WT} / Tm_{WT})$$

$$\text{Total } \Delta \Delta G = \Delta \Delta G1 + \Delta \Delta G2$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1ARR		Grade (Number of Neighbors)	
Chain	Residue		
A	MET	1	6
	LYS	2	4
	GLY	3	4
	MET	4	5
	SER	5	7
	LYS	6	9
	MET	7	8
	PRO	8	10
	GLN	9	7
	PHE	10	13
	ASN	11	9
	LEU	12	12
	ARG	13	9
	TRP	14	16
	PRO	15	10
	ARG	16	9
	GLU	17	7
	VAL	18	12
	LEU	19	12
	ASP	20	8
	LEU	21	10
	VAL	22	15
	ARG	23	11
	LYS	24	9
	VAL	25	13
	ALA	26	13
	GLU	27	9
	GLU	28	9
	ASN	29	10
	GLY	30	7
	ARG	31	10
	SER	32	7
	VAL	33	10
	ASN	34	9
	SER	35	9
	GLU	36	15
	ILE	37	16

Protein 1ARR		Grade (Number of Neighbors)		
Chain	Residue			
A	TYR	38	12	
	GLN	39	9	
	ARG	40	16	
	VAL	41	14	
	MET	42	10	
	GLU	43	10	
	SER	44	12	
	PHE	45	13	
	LYS	46	6	
	LYS	47	9	
	GLU	48	10	
	GLY	49	7	
	ARG	50	11	
	ILE	51	13	
	GLY	52	8	
	ALA	53	7	
	B	MET	54	5
		LYS	55	4
		GLY	56	4
		MET	57	5
SER		58	6	
LYS		59	9	
MET		60	8	
PRO		61	10	
GLN		62	7	
PHE		63	11	
ASN		64	8	
LEU		65	11	
ARG		66	9	
TRP		67	16	
PRO		68	10	
ARG		69	12	
GLU		70	8	
VAL		71	11	
LEU		72	11	
ASP		73	8	
LEU		74	11	

Protein 1ARR		Grade (Number of Neighbors)	
Chain	Residue		
B	VAL	75	16
	ARG	76	11
	LYS	77	8
	VAL	78	14
	ALA	79	13
	GLU	80	9
	GLU	81	11
	ASN	82	11
	GLY	83	7
	ARG	84	12
	SER	85	7
	VAL	86	12
	ASN	87	10
	SER	88	9
	GLU	89	16
	ILE	90	16
	TYR	91	12
	GLN	92	9
	ARG	93	18
	VAL	94	13
	MET	95	10
	FLU	96	10
	SER	97	12
	PHE	98	13
	LYS	99	6
	LYS	100	10
	GLU	101	7
	GLY	102	6
	ARG	103	10
	ILE	104	13
GLY	105	6	
ALA	106	5	

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 1ARR	Grade (Number of Neighbors Chain A)	Thermal denaturation			Reference
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
WT-st6	-	332.15	62.3	-	Nat. Struc. Biol (1994) 1: 518-523
PA8	10	347.25	80.2	2.83	
SA35	9	336.55	57.1	0.83	
NA11	9	335.25	57.2	0.58	
KA6	9	332.75	59	0.11	
RA16	9	332.65	60	0.09	
VA25	13	332.45	58.3	0.06	
MA4	5	332.35	59.4	0.04	
EA27	9	331.95	61.6	-0.04	
KA2	4	331.85	57.2	-0.06	
QA9	7	331.55	59.5	-0.11	
GA3	4	331.25	60	-0.17	
MA1	6	331.15	59.7	-0.19	
SA5	7	330.65	53.3	-0.28	
RA13	9	330.45	51.3	-0.32	
EA17	7	330.15	62.1	-0.38	
VA18	12	330.05	63.1	-0.39	
EA43	10	329.25	59.4	-0.54	
MA7	8	328.65	52.6	-0.66	
DA20	8	328.45	56.9	-0.69	
LA19	12	321.45	48.7	-2.01	
FA10	13	313.75	34.8	-3.45	
WT-st11	-	331.05	58.5	-	
NA34	9	336.15	57.1	0.90	
QA39	9	334.55	61.6	0.62	
GA52	8	334.05	57.7	0.53	

$$\Delta\Delta G = (T_m \text{MUT} - T_m \text{WT}) * (\Delta H^\circ \text{WT} / T_m \text{WT})$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 1ARR	Grade (Number of Neighbors Chain A)	Thermal denaturation			Reference
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
KA46	6	330.25	52.9	-0.14	Nat. Struc. Biol (1994) 1: 518-523
RA23	11	329.85	56.5	-0.21	
KA24	9	329.45	59.9	-0.28	
EA28	9	328.85	52.1	-0.39	
IA51	13	324.05	47.4	-1.24	
GA49	7	321.85	45.8	-1.63	
GA30	7	321.05	48.8	-1.77	
RA50	11	321.05	41.9	-1.77	
KA47	9	320.35	41.2	-1.89	
PA15	10	319.75	38.1	-2.00	
SA44	12	319.45	41	-2.05	
NA29	10	318.45	44.2	-2.23	
VA33	10	317.25	39.7	-2.44	
EA48	10	316.35	29.3	-2.60	
LA12	12	315.45	40.4	-2.76	
LA21	10	312.75	36.7	-3.23	
RA31	10	310.25	34.2	-3.68	
MA42	10	308.75	26.1	-3.94	
SA32	7	306.65	24.4	-4.31	
YA38	12	306.15	24.1	-4.40	
WA14	16	304.65	25.3	-4.67	
RA40	16	304.35	13.6	-4.72	
VA22	9	UNF	-	-	
EA36	15	UNF	-	-	
IA37	16	UNF	-	-	
VA41	14	UNF	-	-	
FA45	13	UNF	-	-	

$$\Delta\Delta G = (T_m\text{MUT} - T_m\text{WT}) * (\Delta H^\circ\text{WT} / T_m\text{WT})$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1HCQ			Grade (Number of Neighbors)	Protein 1HCQ			Grade (Number of Neighbors)	Protein 1HCQ			Grade (Number of Neighbors)
Chain	Residue			Chain	Residue			Chain	Residue		
A	MET	1	2	A	HIS	38	5	B	MET	75	7
	LYS	2	8		ASN	39	4		LYS	76	4
	GLU	3	10		ASP	40	5		GLU	77	11
	THR	4	11		TYR	41	9		THR	78	10
	ARG	5	9		MET	42	10		ARG	79	12
	TYR	6	6		CYS	43	13		TYR	80	13
	CYS	7	5		PRO	44	10		CYS	81	7
	ALA	8	5		ALA	45	11		ALA	82	5
	VAL	9	5		THR	46	11		VAL	83	8
	CYS	10	5		ASN	47	7		CYS	84	7
	ASN	11	7		GLN	48	10		ASN	85	12
	ASP	12	10		CYS	49	10		ASP	86	6
	TYR	13	13		THR	50	10		TYR	87	12
	ALA	14	10		ILE	51	10		ALA	88	7
	SER	15	7		ASP	52	8		SER	89	8
	GLY	16	6		LYS	53	10		GLY	90	7
	TYR	17	5		ASN	54	14		TYR	91	13
	HIS	18	4		ARG	55	14		HIS	92	16
	TYR	19	2		ARG	56	6		TYR	93	10
	GLY	20	8		LYS	57	11		GLY	94	10
	VAL	21	10		SER	58	13		VAL	95	14
	TRP	22	9		CYS	59	15		TRP	96	12
	SER	23	9		GLN	60	15		SER	97	10
	CYS	24	6		ALA	61	15		CYX	98	12
	GLU	25	5		CYS	62	11		GLU	99	18
	GLY	26	5		ARG	63	16		GLY	100	14
	CYS	27	5		LEU	64	17		CYX	101	11
	LYS	28	5		ARG	65	11		LYS	102	12
	ALA	29	7		LYS	66	15		ALA	103	13
	PHE	30	11		CYS	67	16		PHE	104	10
	PHE	31	12		TYR	68	14		PHE	105	7
	LYS	32	10		GLU	69	13		LYS	106	9
	ARG	33	7		VAL	70	11		ARG	107	9
	SER	34	7		GLY	71	10		SER	108	14
	ILE	35	5		MET	72	8		ILE	109	5
	GLN	36	4		MET	73	6		GLN	110	5
	GLY	37	2		LYS	74	4		GLY	111	3

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1HCQ		Grade (Number of Neighbors)	Protein 1HCQ		Grade (Number of Neighbors)	
Chain	Residue		Chain	Residue		
B	HIS	112	4			
	ASN	113	4			
	ASP	114	6			
	TYR	115	9			
	MET	116	9			
	CYS	117	13			
	PRO	118	10			
	ALA	119	12			
	THR	120	11			
	ASN	121	8			
	GLN	122	10			
	CYS	123	10			
	THR	124	10			
	ILE	125	9			
	ASP	126	8			
	LYS	127	10			
	ASN	128	14			
	ARG	129	14			
	ARG	130	7			
	LYS	131	11			
	SER	132	13			
	CYS	133	15			
	GLN	134	15			
	ALA	135	14			
	CYS	136	11			
	ARG	137	16			
	LEU	138	17			
	ARG	139	11			
	LYS	140	15			
	CYS	141	16			
	TYR	142	14			
	GLU	143	12			
VAL	144	11				
GLY	145	8				
MET	146	7				
MET	147	6				
LYS	148	7				
C	DC	149	4			
	DC	150	4			
	DA	151	10			
	DG	152	9			
	DG	153	12			
	DT	154	13			
	DC	155	8			
	DA	156	5			
	DC	157	8			
	DA	158	7			
	DG	159	12			
	DT	160	7			
	DG	161	12			
	DA	162	7			
	DC	163	9			
	DC	164	7			
	DT	165	12			
	DG	166	18			
	DC	167	10			
	D	DC	168	10		
		DA	169	14		
		DG	170	13		
		DG	171	10		
		DT	172	13		
		DC	173	18		
		DA	174	13		
		DC	175	12		
		DT	176	12		
DG		177	14			
DT		178	10			
DG		179	6			
DA		180	9			
DC		181	8			
DC		182	13			
DT		183	6			
DG		184	4			

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Mutant 1HCQ	Grade (Number of Neighbors Chain A)	Dimer recognition		Reference
		Kd (nM)	Kd Difference (%)	
Wild Type (WT)	-	43	-	Biochemistry (2010) 49: 5978-5988
D12A	10	67	-56	
H18A	4	190	-342	
E25A	5	172	-300	
S15A	7	102	-137	
Y19A	2	316	-635	
S23A	9	119	-177	
K28A	5	313	-628	
K32A	10	326	-658	
R33A	7	745	-1633	

$$\% K_d \text{ Difference} = (K_d\text{WT} - K_d\text{MUT}) / K_d\text{WT} * 100$$

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1CSP		Grade (Number of Neighbors)
Residue		
MET	1	7
LEU	2	10
GLU	3	13
GLY	4	12
LYS	5	11
VAL	6	14
LYS	7	9
TRP	8	8
PHE	9	16
ASN	10	8
SER	11	7
GLU	12	4
LYS	13	6
GLY	14	10
PHE	15	11
GLY	16	10
PHE	17	12
ILE	18	17
GLU	19	11
VAL	20	12
GLU	21	5
GLY	22	4
GLN	23	7
ASP	24	6
ASP	25	8
VAL	26	13
PHE	27	11
VAL	28	15
HIS	29	10
PHE	30	10
SER	31	5
ALA	32	10
ILE	33	13
GLN	34	9
GLY	35	5
GLU	36	4
GLY	37	4
PHE	38	5

Protein 1CSP		Grade (Number of Neighbors)
Residue		
LYS	39	8
THR	40	8
LEU	41	12
GLU	42	7
GLU	43	7
GLY	44	6
GLN	45	11
ALA	46	11
VAL	47	15
SER	48	12
PHE	49	15
GLU	50	10
ILE	51	12
VAL	52	8
GLU	53	6
GLY	54	5
ASN	55	4
ARG	56	4
GLY	57	7
PRO	58	8
GLN	59	10
ALA	60	13
ALA	61	9
ASN	62	10
VAL	63	14
THR	64	9
LYS	65	11
GLU	66	8
ALA	67	4

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1CSP	Grade (Number of Neighbors)	Thermal denaturation			Reference
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
Csp S WT	-	53.6	46.13	-	JMB (2001) 313, 343-357
E3R	13	69.6	52.34	2.26	
E3L	13	62.7	47.80	1.28	
A46E	11	48.6	39.67	-0.71	
E66L	8	66.4	53.54	1.81	
Csp S WT	-	53.8	46.13	-	JMB (2005) 347, 1063-1076
E3K	13	70.4	48.52	2.34	
K65I	11	62.4	48.04	1.21	
E66K	8	66.7	54.02	1.82	
E43S	7	54.7	44.46	0.13	
A46K	11	62.2	51.63	1.19	
Csp S WT	-	52.3	39.20	-	JMB (2007) 366, 842-856
E3R	13	70.2	50.19	2.16	
E3Q	13	62.6	46.61	1.24	
K7E	9	UNF	-	-	
K7Q	9	UNF	-	-	
N10D	8	57.7	43.26	0.65	
D25Q	8	44.2	29.16	-0.98	
S48E	12	53	38.00	0.08	
E50Q	10	42.6	34.18	-1.17	
R56Q	4	55.3	42.54	0.36	
Csp S WT*	-	55.3	38.24	-	JMB (2007) 366, 842-856
E3R*	13	71.6	49.95	1.90	
K5E*	11	22.2	14.58	-3.85	
K5Q*	11	38.6	29.88	-1.94	

$$\Delta\Delta G^\circ = (T_m\text{MUT} - T_m\text{WT}) * (\Delta H^\circ\text{WT} / T_m\text{WT})$$

* means that the protein was cut from MHHHHHML N-terminal sequence

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1CSP	Grade (Number of Neighbors)	Thermal denaturation			Reference
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
E12K**	4	50.2	37.76	-0.59	JMB (2007) 366, 842-856
K13E*	6	52.7	36.57	-0.30	
K13Q*	6	54	41.35	-0.15	
E19K*	11	50.5	37.28	-0.56	
E19Q*	11	53.7	39.91	-0.19	
V20Q*	12	39.5	29.64	-1.84	
V20E*	12	UNF	-	-	
V20K*	12	37.2	25.81	-2.11	
E21K*	5	54.3	43.26	-0.12	
E21Q*	5	52.5	34.89	-0.33	
D24K*	6	50.9	34.18	-0.51	
D24N*	6	48.4	31.07	-0.80	
D25K*	8	35	13.38	-2.36	
K39E*	8	52.8	35.85	-0.29	
K39Q*	8	54.9	35.37	-0.05	
E42K*	7	55.3	42.78	0.00	
E42Q*	7	54.1	35.61	-0.14	
E43K*	7	56.5	46.13	0.14	
S48K*	12	61.6	41.59	0.73	
E50K*	10	49.7	31.55	-0.65	
E53K*	6	53.7	40.15	-0.19	
E53Q*	6	56.6	43.74	0.15	
N55K*	4	54.3	37.76	-0.12	
N55D*	4	59.2	44.22	0.45	
K56E*	4	38.2	24.14	-1.99	
K65Q*	11	48	31.31	-0.85	

$$\Delta\Delta G^\circ = (T_m\text{MUT} - T_m\text{WT}) * (\Delta H^\circ\text{WT} / T_m\text{WT})$$

* means that the protein was cut from MHHHHHHML N-terminal sequence

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1C9O		Grade (Number of Neighbors)
Residue		
MET	1	6
GLN	2	8
ARG	3	10
GLY	4	11
LYS	5	12
VAL	6	15
LYS	7	9
TRP	8	8
PHE	9	16
ASN	10	8
ASN	11	8
GLU	12	4
LYS	13	5
GLY	14	10
TYR	15	11
GLY	16	10
PHE	17	12
ILE	18	17
GLU	19	9
VAL	20	12
GLU	21	5
GLY	22	4
GLY	23	5
SER	24	6
ASP	25	9
VAL	26	12
PHE	27	11
VAL	28	15
HIS	29	9
PHE	30	9
THR	31	4
ALA	32	10
ILE	33	13
GLN	34	8
GLY	35	5
GLU	36	5
GLY	37	4

Protein 1C9O		Grade (Number of Neighbors)
Residue		
PHE	38	6
LYS	39	11
THR	40	8
LEU	41	12
GLU	42	7
GLU	43	7
GLY	44	6
GLN	45	11
GLU	46	11
VAL	47	15
SER	48	11
PHE	49	15
GLU	50	9
ILE	51	10
VAL	52	9
GLN	53	7
GLY	54	7
ASN	55	4
ARG	56	4
GLY	57	7
PRO	58	9
GLN	59	11
ALA	60	14
ALA	61	9
ASN	62	10
VAL	63	14
VAL	64	8
LYS	65	10
LEU	66	7

File S1. Network parameter Grade compared to experimental data. For each residue of the protein, the number of neighbors surrounding heavy atoms (C, O, N, S, P), in a distance of 5 Å was measured. Grade values were compared to experimental data.

Protein 1C9O	Grade (Number of Neighbors)	Thermal denaturation			References
		T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG° (kcal mol ⁻¹)	
Csp C WT	-	76.9	58.56	-	JMB (2001) 313, 343-357
Q2L	8	80.9	57.84	0.67	
R3E	10	59.1	45.65	-2.98	
R3L	10	70.9	55.69	-1.00	
R3K	10	76.1	53.54	-0.13	
R3A	10	64.9	51.39	-2.01	
N11S	8	79.2	58.80	0.38	
Y15S	11	76.7	56.64	-0.03	
E21A	5	75	57.84	-0.32	
G23Q	5	74.7	59.51	-0.37	
S24D	6	77.8	61.66	0.15	
T31S	4	77.8	59.75	0.15	
E46A	11	75.4	57.60	-0.25	
Q53E	7	76.1	58.80	-0.13	
V64T	8	75.1	58.32	-0.30	
L66E	7	68.9	50.19	-1.34	
Csp C WT	-	76.9	58.56	-	Prot Eng Des Sel (2006) 19, 355-358
R56E	4	80	60.95	0.52	
H29E	9	73.5	105.88	-0.57	
E46K	11	74	106.84	-0.49	
E21K	5	75.4	49.00	-0.25	
N55K	4	77.7	55.21	0.13	
E36K	5	76	45.65	-0.15	
E50K	9	73.5	43.26	-0.57	
E12K	4	74.4	52.34	-0.42	

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2CI2	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
ASN	1	100	0.1568	3.46	0.0453	65
LEU	2	72.9	0.1726	2.20	0.0783	62
LYS	3	36.5	0.2167	0.89	0.2445	38
THR	4	39.8	0.2135	0.79	0.2710	34
GLU	5	42	0.2208	0.94	0.2342	44
TRP	6	0	0.2674	0.50	0.5339	3
PRO	7	47.2	0.2099	0.63	0.3331	22
GLU	8	46.9	0.2098	0.88	0.2389	41
LEU	9	1.7	0.2507	0.53	0.4727	5
VAL	10	45.2	0.2238	0.61	0.3644	17
GLY	11	56.7	0.1874	0.61	0.3065	26
LYS	12	36	0.2249	0.90	0.2499	37
SER	13	100	0.2086	0.69	0.3037	27
VAL	14	8.5	0.2401	0.73	0.3277	23
GLU	15	54.3	0.2026	1.11	0.1830	48
GLU	16	52.2	0.2004	1.12	0.1787	49
ALA	17	0	0.2310	0.65	0.3565	19
LYS	18	43.4	0.2367	0.98	0.2426	40
LYS	19	63.1	0.2077	1.20	0.1736	50
VAL	20	34.8	0.2236	0.65	0.3463	21
ILE	21	0	0.2600	0.48	0.5402	2
LEU	22	44.1	0.2234	0.71	0.3132	25
GLN	23	78.4	0.1882	1.24	0.1517	52
ASP	24	42.5	0.2040	0.69	0.2935	29
LYS	25	5.5	0.2356	0.56	0.4207	9

Protein 2CI2	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
PRO	26	70.7	0.1891	0.68	0.2788	33
GLU	27	49.5	0.2013	0.77	0.2621	35
ALA	28	9.6	0.2161	0.57	0.3777	14
GLN	29	57.8	0.2190	0.99	0.2205	46
ILE	30	25	0.2485	0.67	0.3704	16
ILE	31	41.2	0.2429	0.67	0.3599	18
VAL	32	47.6	0.2322	0.62	0.3721	15
LEU	33	20.9	0.2417	0.84	0.2867	31
PRO	34	52.3	0.2070	0.73	0.2832	32
VAL	35	51.9	0.2090	0.93	0.2250	45
GLY	36	79.8	0.1644	1.29	0.1277	55
THR	37	40.3	0.1874	2.03	0.0922	59
ILE	38	93.1	0.1739	2.04	0.0851	61
VAL	39	41.8	0.1916	1.52	0.1256	56
THR	40	69.5	0.1664	1.83	0.0907	60
MET	41	98.9	0.1553	2.14	0.0724	63
GLU	42	50.3	0.1776	1.60	0.1109	57
TYR	43	68.3	0.1973	1.27	0.1558	51
ARG	44	45.5	0.2167	0.91	0.2388	42
ILE	45	60.5	0.1988	1.01	0.1977	47
ASP	46	23.1	0.2067	0.82	0.2516	36
ARG	47	14.6	0.2508	0.61	0.4103	10
VAL	48	0	0.2648	0.59	0.4466	6
ARG	49	19.2	0.2592	0.66	0.3927	12
LEU	50	0.3	0.2760	0.72	0.3860	13
PHE	51	16.8	0.2623	0.87	0.3015	28

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2CI2		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
VAL	52	9.5	0.2394	0.56	0.4309	8
ASP	53	39.9	0.1974	1.32	0.1493	53
LYS	54	95.6	0.1784	1.76	0.1015	58
LEU	55	77.7	0.1842	3.05	0.0605	64
ASP	56	47.5	0.2003	1.53	0.1312	54
ASN	57	27.6	0.2138	0.90	0.2372	43
ILE	58	2.4	0.2612	0.50	0.5275	4
ALA	59	44	0.2097	0.60	0.3498	20
GLU	60	39.7	0.2316	0.79	0.2925	30
VAL	61	27.9	0.2416	0.55	0.4397	7
PRO	62	1.3	0.2608	0.48	0.5453	1
ARG	63	47.3	0.2502	0.79	0.3150	24
VAL	64	18	0.2372	0.58	0.4074	11
GLY	65	8.2	0.2070	0.85	0.2445	39

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2W0K	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
ASN	1	86.5	0.1430	1.6	0.0898	110
PHE	2	6.7	0.1898	0.7	0.2779	44
MET	3	62.5	0.1791	1.0	0.1839	80
LEU	4	9.4	0.2170	0.5	0.4032	16
THR	5	54.7	0.1933	0.5	0.3904	18
GLN	6	12.2	0.2218	0.4	0.5055	6
PRO	7	37.7	0.1959	0.5	0.3799	20
HIS	8	79.3	0.1760	1.6	0.1129	102
SER	9	100	0.1731	0.7	0.2430	55
VAL	10	17	0.1992	0.8	0.2634	48
SER	11	100	0.1667	0.7	0.2237	63
GLU	12	28.7	0.1737	0.7	0.2431	54
SER	13	100	0.1546	0.7	0.2135	69
PRO	14	46.8	0.1550	0.7	0.2332	57
GLY	15	55.2	0.1433	0.6	0.2304	60
LYS	16	53.9	0.1676	1.0	0.1692	82
THR	17	47.1	0.1787	0.8	0.2217	64
VAL	18	5.7	0.2050	0.6	0.3617	25
THR	19	52.1	0.2002	0.6	0.3249	33
ILE	20	0.7	0.2346	0.6	0.4207	15
SER	21	100	0.2044	0.6	0.3483	27
CYX	22	0.5	0.2217	0.4	0.5445	4

Protein 2W0K	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
THR	23	37.4	0.2012	0.5	0.4027	17
ARG	24	10.1	0.2034	0.6	0.3273	31
SER	25	100	0.1621	1.2	0.1385	93
SER	26	100	0.1499	1.6	0.0919	109
GLY	27	57.1	0.1495	1.1	0.1351	95
SER	28	100	0.1657	1.2	0.1348	96
ILE	29	2.2	0.2103	1.0	0.2057	73
ALA	30	64.4	0.1663	1.2	0.1333	97
SER	31	100	0.1622	1.1	0.1493	87
ASN	32	22.3	0.1960	0.6	0.3055	37
TYR	33	47.7	0.1908	0.9	0.2157	67
VAL	34	1.8	0.2219	0.5	0.4477	13
GLN	35	8.7	0.2180	1.0	0.2107	70
TRP	36	0.6	0.2575	0.4	0.6100	1
TYR	37	16.9	0.2205	0.7	0.3204	34
GLN	38	14.1	0.2160	0.6	0.3896	19
GLN	39	23.9	0.1953	0.9	0.2257	61
ARG	40	37.3	0.1892	1.4	0.1381	94
PRO	41	87.1	0.1590	1.3	0.1255	100
GLY	42	100	0.1287	1.4	0.0927	107
SER	43	100	0.1521	1.1	0.1327	98
SER	44	100	0.1528	1.0	0.1491	88
PRO	45	46.8	0.1752	0.8	0.2310	59
THR	46	44.5	0.1894	0.6	0.3009	38

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2W0K		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
THR	47	33.9	0.1976	0.7	0.2911	39
VAL	48	3.2	0.2219	0.5	0.4647	10
ILE	49	0	0.2280	0.5	0.4828	9
TYR	50	37	0.1974	0.9	0.2082	72
GLU	51	41.8	0.1796	1.1	0.1640	83
ASP	52	19.9	0.1955	0.9	0.2082	71
ASN	53	49.7	0.1875	0.8	0.2242	62
GLN	54	42.3	0.1832	1.0	0.1829	81
ARG	55	39.3	0.1934	0.9	0.2206	65
PRO	56	21.7	0.1807	0.7	0.2514	51
SER	57	100	0.1472	1.3	0.1125	103
GLY	58	91.7	0.1427	0.9	0.1549	86
VAL	59	20.1	0.1881	0.6	0.2898	40
PRO	60	34.7	0.1765	0.7	0.2621	49
ASP	61	71	0.1596	0.9	0.1858	78
ARG	62	12.4	0.2004	0.5	0.3798	21
PHE	63	2.9	0.2214	0.7	0.3072	36
SER	64	100	0.1956	0.6	0.3437	29
GLY	65	17.5	0.1875	0.5	0.3634	24
SER	66	100	0.1936	0.7	0.2759	46
ILE	67	34.5	0.2055	0.7	0.2880	41
ASP	68	39.4	0.1745	1.1	0.1602	84
SER	69	100	0.1524	1.2	0.1279	99
SER	70	100	0.1407	1.2	0.1152	101

Protein 2W0K		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
SER	71	100	0.1586	1.1	0.1407	92
ASN	72	17.9	0.1782	1.1	0.1554	85
SER	73	100	0.1927	0.5	0.3653	23
ALA	74	0.8	0.2094	0.4	0.4915	8
SER	75	100	0.2100	0.5	0.4262	14
LEU	76	0	0.2425	0.5	0.4964	7
THR	77	21.8	0.2095	0.7	0.3152	35
ILE	78	0	0.2251	0.6	0.3603	26
SER	79	100	0.1783	0.8	0.2138	68
GLY	80	31.3	0.1594	0.7	0.2355	56
LEU	81	2.9	0.1992	0.8	0.2440	52
LYS	82	38.3	0.1801	0.9	0.1955	75
THR	83	70.8	0.1600	0.9	0.1862	77
GLU	84	57.1	0.1668	1.2	0.1409	91
ASP	85	2.2	0.1968	0.5	0.3723	22
GLU	86	38.7	0.1865	1.0	0.1905	76
ALA	87	4.5	0.1967	0.4	0.4596	12
ASP	88	30.4	0.2011	0.6	0.3298	30
TYR	89	0	0.2427	0.4	0.5847	3
TYR	90	21.2	0.2244	0.8	0.2794	43
CYX	91	0.1	0.2265	0.4	0.5887	2
GLN	92	2.3	0.2156	0.8	0.2585	50
SER	93	100	0.2007	0.4	0.4640	11
TYR	94	34.6	0.1944	0.8	0.2311	58

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2W0K		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
ASP	95	6.9	0.1692	1.1	0.1488	89
SER	96	100	0.1399	1.3	0.1081	105
SER	97	100	0.1414	1.3	0.1085	104
ASN	98	69.8	0.1444	1.5	0.0974	106
HIS	99	57.1	0.1641	1.8	0.0923	108
VAL	100	55.5	0.1747	0.9	0.1857	79
VAL	101	9.6	0.1944	0.7	0.2738	47
PHE	102	48.3	0.1999	1.4	0.1471	90
GLY	103	5.4	0.1852	0.7	0.2776	45
GLY	104	75.9	0.1752	0.8	0.2185	66
GLY	105	11.8	0.1869	0.6	0.3251	32
THR	106	0.3	0.2175	0.4	0.5365	5
LYS	107	34.6	0.1994	0.8	0.2438	53
LEU	108	0	0.2168	0.6	0.3483	28
THR	109	35.2	0.1781	0.6	0.2841	42
VAL	110	14.2	0.1773	0.9	0.1985	74
LEU	111	60.8	0.1507	3.0	0.0497	111

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
SER	1	100	0.1221	5.00	0.0244	169
LYS	2	80.6	0.1436	3.55	0.0405	168
LYS	3	52	0.1587	1.83	0.0866	162
ILE	4	1	0.1824	0.72	0.2544	52
GLY	5	0	0.1694	0.48	0.3539	20
LEU	6	0	0.2008	0.73	0.2737	43
PHE	7	0.2	0.2010	0.75	0.2679	47
TYR	8	6.7	0.1989	0.75	0.2662	48
GLY	9	0	0.1679	0.70	0.2408	63
THR	10	22.1	0.1677	0.91	0.1835	92
GLN	11	59.3	0.1508	1.82	0.0829	163
THR	12	71.1	0.1419	1.29	0.1102	154
GLY	13	29.4	0.1391	1.09	0.1277	141
LYS	14	20.2	0.1702	1.31	0.1300	138
THR	15	0.8	0.1835	0.90	0.2046	78
GLU	16	32.3	0.1658	1.16	0.1425	132
SER	17	100	0.1569	0.92	0.1703	105
VAL	18	4.2	0.1877	0.75	0.2501	56
ALA	19	0	0.1742	0.65	0.2685	45
GLU	20	41.9	0.1611	1.09	0.1478	126
ILE	21	38.9	0.1741	0.69	0.2530	54
ILE	22	0.4	0.1942	0.65	0.2987	38
ARG	23	28.5	0.1716	1.12	0.1535	120
ASP	24	62.9	0.1454	1.24	0.1171	149

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
GLU	25	32.5	0.1565	1.01	0.1554	118
PHE	26	4.2	0.1761	1.02	0.1727	102
GLY	27	21	0.1320	0.94	0.1404	134
ASN	28	95.2	0.1303	1.71	0.0760	165
ASP	29	61.2	0.1324	1.37	0.0969	160
VAL	30	11	0.1558	1.05	0.1491	123
VAL	31	2.5	0.1734	0.86	0.2008	80
THR	32	21.6	0.1629	0.83	0.1951	83
LEU	33	31.2	0.1765	0.90	0.1957	82
HIS	34	21.5	0.1764	0.94	0.1870	90
ASP	35	22.9	0.1611	1.10	0.1464	129
VAL	36	5	0.1784	0.68	0.2641	50
SER	37	100	0.1455	0.85	0.1716	103
GLN	38	80.9	0.1415	1.41	0.1002	158
ALA	39	23	0.1496	0.81	0.1840	91
GLU	40	62.6	0.1431	1.23	0.1159	150
VAL	41	8.2	0.1559	1.01	0.1539	119
THR	42	55.5	0.1385	1.06	0.1302	137
ASP	43	25.1	0.1518	1.18	0.1284	140
LEU	44	0.7	0.1806	1.03	0.1753	100
ASN	45	46.5	0.1497	1.25	0.1194	146
ASP	46	74.9	0.1434	1.24	0.1154	151
TYR	47	12.9	0.1781	0.72	0.2475	59
GLN	48	47.5	0.1607	1.30	0.1240	144

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
TYR	49	32	0.1781	0.78	0.2273	70
LEU	50	0.2	0.2008	0.67	0.2978	39
ILE	51	0	0.2100	0.45	0.4664	5
ILE	52	0.4	0.2149	0.53	0.4069	13
GLY	53	0	0.1862	0.40	0.4712	2
CYS	54	0	0.2013	0.55	0.3681	18
PRO	55	3.5	0.1922	0.55	0.3474	22
THR	56	3.6	0.1852	0.48	0.3849	16
TRP	57	18.7	0.1855	0.81	0.2292	69
ASN	58	31.7	0.1546	1.23	0.1256	142
ILE	59	73.7	0.1478	0.97	0.1524	121
GLY	60	86.5	0.1474	0.76	0.1949	85
GLU	61	61.2	0.1588	0.93	0.1715	104
LEU	62	5.9	0.1832	0.74	0.2487	58
GLN	63	5.7	0.1809	0.67	0.2680	46
SER	64	100	0.1479	0.84	0.1759	99
ASP	65	31.6	0.1609	0.85	0.1899	88
TRP	66	0.4	0.2078	0.61	0.3416	24
GLU	67	37.7	0.1590	0.98	0.1617	114
GLY	68	50.9	0.1368	0.77	0.1775	97
LEU	69	7.8	0.1761	0.91	0.1934	86
TYR	70	25.6	0.1766	0.89	0.1978	81
SER	71	100	0.1388	0.92	0.1517	122
GLU	72	43.3	0.1453	1.12	0.1292	139

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
LEU	73	0.4	0.1715	1.38	0.1246	143
ASP	74	45.8	0.1346	2.20	0.0611	166
ASP	75	74.6	0.1301	2.49	0.0523	167
VAL	76	1.8	0.1558	1.40	0.1111	153
ASP	77	54	0.1429	1.32	0.1080	155
PHE	78	0.3	0.1801	0.87	0.2058	77
ASN	79	62.1	0.1388	1.06	0.1311	136
GLY	80	60.8	0.1343	0.83	0.1611	115
LYS	81	4.3	0.1755	0.74	0.2377	65
LEU	82	21.1	0.1805	0.66	0.2727	44
VAL	83	0.8	0.1963	0.56	0.3507	21
ALA	84	0	0.1923	0.41	0.4669	4
TYR	85	0.5	0.2190	0.40	0.5452	1
PHE	86	0.5	0.2196	0.47	0.4695	3
GLY	87	0.4	0.1858	0.65	0.2864	42
THR	88	1.7	0.1942	0.78	0.2497	57
GLY	89	1.3	0.1650	0.51	0.3254	30
ASP	90	15.6	0.1599	0.61	0.2612	51
GLN	91	4.6	0.1620	0.77	0.2091	75
ILE	92	64.9	0.1427	0.78	0.1823	95
GLY	93	64.2	0.1295	0.66	0.1950	84
TYR	94	25	0.1668	1.04	0.1596	116
ALA	95	22.6	0.1423	0.60	0.2367	66
ASP	96	53.1	0.1404	0.84	0.1676	108

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
ASN	97	20.1	0.1602	0.52	0.3062	35
PHE	98	0.5	0.1911	0.52	0.3680	19
GLN	99	0.5	0.2028	0.59	0.3449	23
ASP	100	5.6	0.1803	0.45	0.3969	14
ALA	101	3.9	0.1853	0.43	0.4302	9
ILE	102	0.3	0.2144	0.46	0.4624	7
GLY	103	0.2	0.1698	0.53	0.3201	33
ILE	104	26.6	0.1832	0.56	0.3251	31
LEU	105	0.2	0.1990	0.75	0.2643	49
GLU	106	6.8	0.1855	0.65	0.2870	41
GLU	107	62.7	0.1615	1.09	0.1479	125
LYS	108	16.4	0.1699	1.03	0.1657	110
ILE	109	1.8	0.1873	0.63	0.2988	37
SER	110	100	0.1564	0.70	0.2247	72
GLN	111	62.5	0.1459	1.43	0.1022	157
ARG	112	25.6	0.1584	1.41	0.1126	152
GLY	113	23.2	0.1384	0.73	0.1889	89
GLY	114	12.6	0.1516	0.62	0.2454	60
LYS	115	58.5	0.1700	1.02	0.1671	109
THR	116	17.2	0.1767	0.53	0.3354	27
VAL	117	9.4	0.1853	0.55	0.3388	25
GLY	118	1.6	0.1680	0.50	0.3360	26
TYR	119	45	0.1856	0.74	0.2505	55
TRP	120	16.3	0.1942	0.59	0.3308	28

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
SER	121	100	0.1521	0.76	0.2012	79
THR	122	21.7	0.1541	0.64	0.2422	61
ASP	123	79.9	0.1392	1.19	0.1173	148
GLY	124	51.2	0.1300	0.77	0.1683	107
TYR	125	13.7	0.1758	0.78	0.2248	71
ASP	126	64.4	0.1344	1.02	0.1315	135
PHE	127	18.4	0.1626	0.69	0.2365	67
ASN	128	66.1	0.1321	0.93	0.1418	133
ASP	129	50.6	0.1450	0.79	0.1831	93
SER	130	100	0.1603	0.49	0.3241	32
LYS	131	52.1	0.1719	1.05	0.1636	113
ALA	132	0	0.1761	0.47	0.3756	17
LEU	133	24.9	0.1717	0.71	0.2404	64
ARG	134	36.7	0.1768	0.75	0.2363	68
ASN	135	85	0.1333	1.36	0.0977	159
GLY	136	49.2	0.1303	0.88	0.1482	124
LYS	137	45	0.1677	0.99	0.1697	106
PHE	138	0.2	0.1939	0.47	0.4162	12
VAL	139	0	0.1964	0.45	0.4397	8
GLY	140	0	0.1790	0.43	0.4179	11
LEU	141	0.1	0.2077	0.45	0.4648	6
ALA	142	0.2	0.1831	0.47	0.3878	15
LEU	143	0.5	0.2006	0.61	0.3300	29
ASP	144	0.1	0.1710	0.59	0.2899	40

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
GLU	145	27.9	0.1697	0.96	0.1765	98
ASP	146	55.5	0.1432	0.98	0.1468	128
ASN	147	41.3	0.1448	0.76	0.1905	87
GLN	148	15	0.1580	0.73	0.2161	73
SER	149	100	0.1431	0.87	0.1646	112
ASP	150	76.6	0.1324	1.13	0.1174	147
LEU	151	36.6	0.1560	1.06	0.1471	127
THR	152	4.2	0.1673	0.80	0.2102	74
ASP	153	59.8	0.1499	0.82	0.1824	94
ASP	154	67.1	0.1478	0.82	0.1798	96
ARG	155	9.6	0.1806	0.71	0.2544	53
ILE	156	1.9	0.1875	0.61	0.3094	34
LYS	157	52.2	0.1609	1.11	0.1451	130
SER	158	100	0.1503	0.72	0.2085	76
TRP	159	0	0.2055	0.48	0.4275	10
VAL	160	0.6	0.1782	0.59	0.3029	36
ALA	161	50.7	0.1454	0.83	0.1742	101
GLN	162	27	0.1613	1.53	0.1056	156
LEU	163	0	0.1872	0.78	0.2408	62
LYS	164	26.3	0.1628	1.03	0.1581	117
SER	165	100	0.1369	0.95	0.1445	131
GLU	166	38.7	0.1590	1.30	0.1223	145
PHE	167	3.3	0.1805	1.09	0.1650	111
GLY	168	71.6	0.1258	1.37	0.0916	161

Protein 1FTG		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
LEU	169	56.1	0.1455	1.87	0.0777	164

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1ARR		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking whole structure	Chain A Residue ranking position
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
MET	1	77.4	0.1203	7.97	0.0151	106	53
LYS	2	80.3	0.1218	6.82	0.0179	105	52
GLY	3	70	0.1128	4.01	0.0281	102	51
MET	4	70.6	0.1314	4.04	0.0325	100	50
SER	5	100	0.1388	3.09	0.0449	95	47
LYS	6	71.9	0.1476	3.44	0.0429	97	48
MET	7	85.5	0.1805	1.95	0.0927	69	31
PRO	8	77.1	0.1766	1.68	0.1052	59	26
GLN	9	97.2	0.1835	2.01	0.0911	70	32
PHE	10	79.2	0.2176	1.87	0.1167	54	24
ASN	11	72.3	0.1875	1.83	0.1026	62	28
LEU	12	52.9	0.2212	1.57	0.1413	40	17
ARG	13	77.6	0.1965	1.96	0.1004	64	30
TRP	14	46.3	0.2417	1.24	0.1946	12	5
PRO	15	58.1	0.1989	1.56	0.1278	50	22
ARG	16	68.4	0.1851	2.32	0.0798	78	38
GLU	17	71.7	0.1773	2.19	0.0809	76	36
VAL	18	42	0.2131	1.43	0.1492	30	9
LEU	19	29.7	0.2169	1.48	0.1462	37	14
ASP	20	41	0.1805	1.74	0.1036	60	27
LEU	21	59.3	0.2065	1.45	0.1426	39	16
VAL	22	15.5	0.2207	1.27	0.1732	20	6
ARG	23	47.5	0.1972	2.29	0.0861	74	34
LYS	24	63.7	0.1846	1.83	0.1008	63	29
VAL	25	30.8	0.1998	1.34	0.1488	32	11
ALA	26	0.2	0.1853	1.40	0.1327	46	20
GLU	27	69.9	0.1668	2.07	0.0806	77	37
GLU	28	76.6	0.1628	2.07	0.0787	79	39
ASN	29	44.3	0.1744	1.92	0.0906	71	33
GLY	30	81.4	0.1470	1.88	0.0784	80	40
ARG	31	34	0.2077	1.52	0.1367	44	18
SER	32	100	0.1851	1.62	0.1140	56	25
VAL	33	37.8	0.2152	1.45	0.1483	33	12
ASN	34	79.7	0.2117	1.44	0.1473	35	13

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1ARR		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking whole structure	Chain A Residue ranking position
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
SER	35	100	0.1990	1.34	0.1490	31	10
GLU	36	11.6	0.2245	1.08	0.2077	9	4
ILE	37	28	0.2444	0.98	0.2484	5	1
TYR	38	63.2	0.2260	1.37	0.1647	23	7
GLN	39	23	0.2079	1.44	0.1441	38	15
ARG	40	42.5	0.2338	1.03	0.2267	8	3
VAL	41	47.9	0.2321	0.94	0.2476	6	2
MET	42	44.9	0.2116	1.63	0.1299	47	21
GLU	43	51.2	0.1919	1.59	0.1210	53	23
SER	44	100	0.1930	1.27	0.1514	29	8
PHE	45	45.7	0.2148	1.59	0.1352	45	19
LYS	46	80.7	0.1779	2.18	0.0818	75	35
LYS	47	61.2	0.1731	2.69	0.0644	86	43
GLU	48	77	0.1398	3.51	0.0399	99	49
GLY	49	67.9	0.1311	2.86	0.0458	94	46
ARG	50	83.7	0.1646	2.92	0.0564	90	45
ILE	51	39.5	0.1901	2.91	0.0653	85	42
GLY	52	41.5	0.1676	2.39	0.0702	83	41
ALA	53	96.5	0.1608	2.69	0.0597	89	44
MET	54	75.7	0.1564	8.68	0.0180	104	-
LYS	55	80.4	0.1422	7.47	0.0190	103	-
GLY	56	68.8	0.1405	4.65	0.0302	101	-
MET	57	71.5	0.1828	4.30	0.0425	98	-
SER	58	100	0.1467	3.14	0.0467	93	-
LYS	59	67.6	0.1550	2.99	0.0519	91	-
MET	60	86.5	0.1918	1.50	0.1278	49	-
PRO	61	70.2	0.1787	1.28	0.1390	42	-
GLN	62	98.3	0.1792	1.56	0.1145	55	-
PHE	63	77.5	0.2214	1.46	0.1520	28	-
ASN	64	75.9	0.1953	1.39	0.1405	41	-
LEU	65	60	0.2274	1.15	0.1969	10	-
ARG	66	81.1	0.1973	1.61	0.1224	52	-
TRP	67	42.3	0.2376	1.03	0.2315	7	-
PRO	68	59.7	0.1869	1.26	0.1480	34	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1ARR		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking whole structure	Chain A Residue ranking position
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
ARG	69	72.5	0.1886	2.03	0.0927	68	-
GLU	70	71.9	0.1715	1.95	0.0880	73	-
VAL	71	44.8	0.2068	1.23	0.1675	21	-
LEU	72	28.4	0.2145	1.36	0.1582	26	-
ASP	73	38.2	0.1824	1.46	0.1254	51	-
LEU	74	55.8	0.2087	1.30	0.1605	25	-
VAL	75	19.6	0.2232	1.15	0.1936	14	-
ARG	76	51.5	0.1974	2.07	0.0952	66	-
LYS	77	63.3	0.1868	1.69	0.1106	57	-
VAL	78	34.5	0.2078	1.19	0.1740	19	-
ALA	79	0.7	0.1915	1.09	0.1755	18	-
GLU	80	71.7	0.1728	1.73	0.0998	65	-
GLU	81	81.3	0.1722	1.85	0.0931	67	-
ASN	82	37.8	0.1765	1.65	0.1070	58	-
GLY	83	82.5	0.1510	1.47	0.1026	61	-
ARG	84	28.4	0.2139	1.21	0.1768	17	-
SER	85	100	0.1891	1.24	0.1522	27	-
VAL	86	35.7	0.2175	1.14	0.1901	15	-
ASN	87	80.5	0.2137	1.10	0.1951	11	-
SER	88	100	0.2032	1.05	0.1937	13	-
GLU	89	11.7	0.2287	0.89	0.2578	3	-
ILE	90	31.2	0.2446	0.81	0.3003	1	-
TYR	91	62.5	0.2282	1.21	0.1881	16	-
GLN	92	25.7	0.2120	1.28	0.1655	22	-
ARG	93	37.1	0.2358	0.94	0.2512	4	-
VAL	94	45.2	0.2321	0.85	0.2732	2	-
MET	95	44.7	0.2119	1.53	0.1384	43	-
FLU	96	51.6	0.1940	1.50	0.1289	48	-
SER	97	100	0.1951	1.21	0.1617	24	-
PHE	98	50.8	0.2147	1.47	0.1465	36	-
LYS	99	74.7	0.1779	2.00	0.0890	72	-
LYS	100	64.2	0.1727	2.44	0.0709	82	-
GLU	101	75.1	0.1396	3.26	0.0429	96	-
GLY	102	37.7	0.1344	2.63	0.0511	92	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1ARR		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking whole structure	Chain A Residue ranking position	
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
ARG	103	B	80.7	0.1685	2.64	0.0637	88	-
ILE	104		47.7	0.1709	2.68	0.0638	87	-
GLY	105		56.7	0.1615	2.16	0.0748	81	-
ALA	106		84.4	0.1701	2.52	0.0675	84	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
MET	1	100	0.1143	7.99	0.0143	184	148	74
LYS	2	63.6	0.1141	7.41	0.0154	182	146	73
GLU	3	95.1	0.1177	5.10	0.0231	180	144	72
THR	4	35.6	0.1242	4.06	0.0306	173	141	71
ARG	5	43.4	0.1436	3.96	0.0363	168	136	68
TYR	6	44.8	0.1529	3.15	0.0485	145	117	58
CYS	7	1.6	0.1693	2.54	0.0666	88	67	34
ALA	8	34	0.1567	2.90	0.0540	128	101	50
VAL	9	0	0.1853	2.76	0.0672	84	63	30
CYS	10	1.2	0.1767	2.44	0.0724	65	47	22
ASN	11	65.9	0.1621	2.79	0.0580	116	91	45
ASP	12	19.9	0.1639	2.39	0.0685	82	62	29
TYR	13	62.3	0.1531	3.04	0.0503	138	111	54
ALA	14	20.2	0.1472	2.71	0.0544	126	100	49
SER	15	100	0.1389	2.91	0.0478	146	118	59
GLY	16	15.5	0.1340	2.75	0.0487	144	116	57
TYR	17	62	0.1455	3.40	0.0427	157	127	63
HIS	18	18.6	0.1565	3.11	0.0503	139	112	55
TYR	19	30.8	0.1634	2.72	0.0600	109	85	43
GLY	20	37.7	0.1278	3.10	0.0412	161	130	64
VAL	21	6.1	0.1580	2.97	0.0533	131	104	52
TRP	22	37.9	0.1582	3.04	0.0520	136	109	53
SER	23	0	0.1647	2.46	0.0670	85	64	31
CYS	24	3.5	0.1740	2.14	0.0812	43	30	16
GLU	25	51.4	0.1658	1.93	0.0859	36	23	12
GLY	26	30.9	0.1638	1.78	0.0920	24	13	7
CYS	27	0	0.1834	1.99	0.0919	25	14	8
LYS	28	31.4	0.1765	2.39	0.0740	58	41	20
ALA	29	23.8	0.1640	1.66	0.0987	12	6	2
PHE	30	3.2	0.1956	1.84	0.1065	6	2	1
PHE	31	0.7	0.1814	2.26	0.0803	44	31	17
LYS	32	43.7	0.1644	2.13	0.0772	51	36	18
ARG	33	52.4	0.1740	2.01	0.0865	33	21	11
SER	34	100	0.1618	1.94	0.0833	41	28	15

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A Residue
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
ILE	35	34	0.1577	2.46	0.0640	94	72	38
GLN	36	73.1	0.1432	2.67	0.0537	129	102	51
GLY	37	47.7	0.1364	2.05	0.0667	87	66	33
HIS	38	94.1	0.1763	1.96	0.0900	30	19	10
ASN	39	44.9	0.1588	2.39	0.0666	89	68	35
ASP	40	77.1	0.1554	2.56	0.0608	108	84	42
TYR	41	25.9	0.1905	2.02	0.0943	18	9	4
MET	42	75.1	0.1795	2.49	0.0721	66	48	23
CYS	43	22.3	0.1804	2.46	0.0733	62	45	21
PRO	44	71.5	0.1867	2.42	0.0772	52	37	19
ALA	45	60.4	0.1645	2.92	0.0564	119	94	47
THR	46	65.5	0.1543	3.59	0.0430	156	126	62
ASN	47	56.3	0.1630	3.31	0.0493	141	114	56
GLN	48	75.1	0.1609	3.56	0.0452	150	122	61
CYS	49	27.6	0.1780	2.81	0.0632	97	75	40
THR	50	53	0.1695	3.03	0.0559	120	95	48
ILE	51	7.7	0.1913	2.66	0.0720	67	49	24
ASP	52	24.3	0.1756	2.63	0.0667	86	65	32
LYS	53	50.2	0.1782	2.77	0.0643	93	71	37
ASN	54	79.4	0.1744	2.64	0.0661	91	70	36
ARG	55	48.7	0.1937	2.74	0.0706	73	55	26
ARG	56	14.1	0.2026	2.15	0.0940	19	10	5
LYS	57	70.9	0.1973	2.17	0.0911	27	16	9
SER	58	100	0.1859	2.19	0.0848	37	24	13
CYS	59	2.3	0.1918	2.27	0.0846	39	26	14
GLN	60	44.8	0.1987	2.04	0.0975	13	7	3
ALA	61	19.2	0.1737	2.44	0.0713	70	52	25
CYS	62	0.6	0.1799	2.60	0.0691	80	60	28
ARG	63	7.5	0.2066	2.23	0.0928	22	11	6
LEU	64	13.9	0.1846	2.64	0.0699	77	58	27
ARG	65	30.1	0.1761	3.12	0.0564	118	93	46
LYS	66	25.2	0.1798	3.07	0.0587	114	89	44
CYS	67	0	0.1702	2.79	0.0611	106	83	41
TYR	68	53.7	0.1585	3.98	0.0398	164	133	66
GLU	69	67.2	0.1452	3.63	0.0400	163	132	65

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A	
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$				
VAL	70	A	38.6	0.1549	3.26	0.0475	147	119	60
GLY	71		31.4	0.1306	3.31	0.0395	166	135	67
MET	72		5.4	0.1740	2.74	0.0635	96	74	39
MET	73		65.5	0.1423	3.93	0.0362	169	137	69
LYS	74		91.1	0.1407	4.06	0.0347	171	139	70
MET	75	B	100	0.1146	7.94	0.0144	183	147	-
LYS	76		87.1	0.1140	7.38	0.0155	181	145	-
GLU	77		83.7	0.1157	4.60	0.0252	178	143	-
THR	78		50.8	0.1263	3.97	0.0318	172	140	-
ARG	79		40.6	0.1394	3.50	0.0398	165	134	-
TYR	80		49.6	0.1551	2.93	0.0530	132	105	-
CYS	81		1.2	0.1714	2.32	0.0739	59	42	-
ALA	82		30.1	0.1588	2.68	0.0592	113	88	-
VAL	83		0.1	0.1874	2.48	0.0757	53	38	-
CYS	84		3.6	0.1789	2.24	0.0798	46	33	-
ASN	85		59.5	0.1600	2.55	0.0629	99	77	-
ASP	86		20	0.1638	2.20	0.0743	56	40	-
TYR	87		62	0.1531	2.68	0.0571	117	92	-
ALA	88		22.5	0.1473	2.40	0.0613	104	81	-
SER	89		100	0.1452	2.27	0.0639	95	73	-
GLY	90		14.9	0.1337	2.50	0.0534	130	103	-
TYR	91		61.8	0.1432	3.20	0.0447	152	124	-
HIS	92		17.9	0.1540	2.60	0.0593	112	87	-
TYR	93		31	0.1611	2.94	0.0549	124	98	-
GLY	94		45.6	0.1276	3.06	0.0417	160	129	-
VAL	95		5.7	0.1579	2.89	0.0546	125	99	-
TRP	96		35.8	0.1582	2.88	0.0550	123	97	-
SER	97		16.7	0.1644	2.34	0.0704	74	56	-
CYX	98		2	0.1739	2.02	0.0862	34	22	-
GLU	99	51.1	0.1659	1.96	0.0847	38	25	-	
GLY	100	30.9	0.1638	1.80	0.0908	28	17	-	
CYX	101	0	0.1832	2.00	0.0918	26	15	-	
LYS	102	29.3	0.1785	2.26	0.0790	48	34	-	
ALA	103	20.1	0.1640	1.99	0.0823	42	29	-	
PHE	104	3.6	0.1955	2.02	0.0966	16	8	-	

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
PHE	105	1.8	0.1768	2.56	0.0692	79	59	-
LYS	106	42.6	0.1644	2.61	0.0630	98	76	-
ARG	107	54	0.1828	2.04	0.0898	31	20	-
SER	108	100	0.1578	2.53	0.0625	101	78	-
ILE	109	46	0.1513	3.07	0.0492	142	115	-
GLN	110	69.3	0.1453	3.11	0.0468	149	121	-
GLY	111	65.2	0.1280	3.02	0.0424	158	128	-
HIS	112	38.5	0.1612	2.76	0.0585	115	90	-
ASN	113	82.3	0.1555	2.97	0.0523	134	107	-
ASP	114	82.3	0.1532	2.91	0.0527	133	106	-
TYR	115	27.4	0.1882	2.25	0.0836	40	27	-
MET	116	74.6	0.1840	2.61	0.0704	75	57	-
CYS	117	14.6	0.1805	2.52	0.0717	68	50	-
PRO	118	73.3	0.1890	2.56	0.0738	60	43	-
ALA	119	59.1	0.1646	2.98	0.0551	122	96	-
THR	120	63.7	0.1544	3.30	0.0468	148	120	-
ASN	121	60.1	0.1631	3.28	0.0497	140	113	-
GLN	122	70.2	0.1653	3.17	0.0522	135	108	-
CYS	123	29.4	0.1781	2.60	0.0686	81	61	-
THR	124	61.4	0.1695	2.83	0.0600	110	86	-
ILE	125	12.3	0.1913	2.43	0.0786	49	35	-
ASP	126	20.5	0.1735	2.37	0.0733	61	44	-
LYS	127	47.1	0.1741	2.45	0.0711	71	53	-
ASN	128	63.5	0.1724	2.37	0.0729	63	46	-
ARG	129	51.1	0.1916	2.39	0.0802	45	32	-
ARG	130	11.9	0.2026	1.87	0.1085	4	1	-
LYS	131	68.5	0.1953	1.91	0.1022	8	4	-
SER	132	100	0.1859	2.02	0.0921	23	12	-
CYS	133	2.5	0.1918	2.12	0.0904	29	18	-
GLN	134	43.9	0.1989	1.92	0.1036	7	3	-
ALA	135	18	0.1738	2.44	0.0714	69	51	-
CYS	136	0.1	0.1822	2.45	0.0744	55	39	-
ARG	137	8.8	0.2067	2.05	0.1009	10	5	-
LEU	138	9	0.1825	2.58	0.0708	72	54	-
ARG	139	32.3	0.1762	2.88	0.0612	105	82	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
LYS	140	27.4	0.1819	2.74	0.0665	90	69	-
CYS	141	0	0.1701	2.74	0.0621	102	79	-
TYR	142	52.5	0.1626	3.66	0.0444	154	125	-
GLU	143	71.8	0.1474	3.29	0.0448	151	123	-
VAL	144	34.3	0.1549	3.06	0.0505	137	110	-
GLY	145	32.5	0.1306	3.23	0.0404	162	131	-
MET	146	3.2	0.1716	2.77	0.0619	103	80	-
MET	147	66.2	0.1378	3.85	0.0358	170	138	-
LYS	148	98.8	0.1407	4.69	0.0300	175	142	-
DC	149	-	0.1010	3.31	0.0305	174	-	-
DC	150	-	0.1137	2.55	0.0446	153	-	-
DA	151	-	0.1269	1.81	0.0700	76	-	-
DG	152	-	0.1316	1.75	0.0752	54	-	-
DG	153	-	0.1321	1.82	0.0728	64	-	-
DT	154	-	0.1324	1.52	0.0870	32	-	-
DC	155	-	0.1327	1.68	0.0792	47	-	-
DA	156	-	0.1442	2.42	0.0596	111	-	-
DC	157	-	0.1489	1.73	0.0861	35	-	-
DA	158	-	0.1779	1.64	0.1085	5	-	-
DG	159	-	0.1867	1.66	0.1124	2	-	-
DT	160	-	0.1839	1.47	0.1254	1	-	-
DG	161	-	0.1789	1.60	0.1117	3	-	-
DA	162	-	0.1632	1.68	0.0970	14	-	-
DC	163	-	0.1386	1.87	0.0741	57	-	-
DC	164	-	0.1229	1.86	0.0660	92	-	-
DT	165	-	0.1139	2.59	0.0440	155	-	-
DG	166	-	0.1032	3.74	0.0276	177	-	-
DC	167	-	0.0966	3.86	0.0250	179	-	-
DC	168	-	0.1093	2.93	0.0374	167	-	-
DA	169	-	0.1225	2.49	0.0491	143	-	-
DG	170	-	0.1316	2.43	0.0542	127	-	-
DG	171	-	0.1386	2.27	0.0610	107	-	-
DT	172	-	0.1436	2.07	0.0693	78	-	-
DC	173	-	0.1374	2.19	0.0628	100	-	-
DA	174	-	0.1419	2.10	0.0674	83	-	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1HCQ		Solv-ent Access	Dynamical entropy	Movement restriction	SDRI	Ranking Whole structure	Ranking Dimer structure	Ranking Chain A
Residue	Chain	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$			
DC	175	-	0.1508	1.61	0.0935	21	-	-
DT	176	-	0.1598	1.57	0.1020	9	-	-
DG	177	-	0.1754	1.87	0.0938	20	-	-
DT	178	-	0.1841	1.90	0.0969	15	-	-
DG	179	-	0.1788	1.89	0.0945	17	-	-
DA	180	-	0.1674	1.69	0.0988	11	-	-
DC	181	-	0.1494	1.93	0.0773	50	-	-
DC	182	-	0.1317	2.36	0.0559	121	-	-
DT	183	-	0.1204	2.85	0.0422	159	-	-
DG	184	-	0.1097	3.66	0.0299	176	-	-

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1CSP		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
MET	1	72.2	0.1973	1.09	0.1813	47
LEU	2	34.7	0.2150	0.93	0.2302	41
GLU	3	39.2	0.2044	1.20	0.1703	48
GLY	4	12.2	0.2382	0.49	0.4860	7
LYS	5	45.9	0.2222	0.96	0.2319	40
VAL	6	2.5	0.2500	0.40	0.6266	3
LYS	7	36.1	0.2159	0.65	0.3334	25
TRP	8	39	0.2176	0.55	0.3989	19
PHE	9	13.4	0.2406	0.54	0.4417	14
ASN	10	33	0.2030	0.70	0.2892	28
SER	11	100	0.1961	1.05	0.1872	46
GLU	12	99	0.1627	1.45	0.1121	59
LYS	13	63.7	0.1820	1.18	0.1539	51
GLY	14	7.7	0.2213	0.65	0.3401	23
PHE	15	26	0.2216	0.87	0.2547	30
GLY	16	0	0.2544	0.40	0.6283	2
PHE	17	28.7	0.2406	0.50	0.4805	8
ILE	18	0	0.2635	0.48	0.5489	5
GLU	19	47.9	0.2205	0.92	0.2408	32
VAL	20	8.9	0.2219	0.94	0.2349	37
GLU	21	77.4	0.1756	2.08	0.0846	63
GLY	22	90.1	0.1692	1.72	0.0985	61
GLN	23	56.2	0.1890	1.63	0.1157	58
ASP	24	75.7	0.1917	1.31	0.1466	54

Protein 1CSP		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
ASP	25	41	0.2153	0.91	0.2367	35
VAL	26	5.4	0.2501	0.71	0.3507	22
PHE	27	33.4	0.2444	0.56	0.4354	16
VAL	28	0	0.2687	0.42	0.6327	1
HIS	29	31.1	0.2299	0.60	0.3858	20
PHE	30	40.4	0.2166	0.91	0.2381	34
SER	31	100	0.2168	0.61	0.3552	21
ALA	32	10.1	0.2457	0.55	0.4484	13
ILE	33	4.6	0.2457	0.56	0.4391	15
GLN	34	41.3	0.2115	1.05	0.2018	45
GLY	35	63.5	0.1912	1.25	0.1528	52
GLU	36	88.7	0.1607	2.65	0.0605	66
GLY	37	60.7	0.1695	1.85	0.0914	62
PHE	38	75.6	0.1777	1.71	0.1040	60
LYS	39	35.6	0.2111	0.93	0.2266	42
THR	40	22.4	0.2174	0.67	0.3231	27
LEU	41	5.2	0.2451	0.57	0.4318	17
GLU	42	66.6	0.2032	1.25	0.1619	50
GLU	43	76.2	0.1958	1.19	0.1652	49
GLY	44	42.5	0.2073	0.61	0.3372	24
GLN	45	19.7	0.2239	1.07	0.2102	44
ALA	46	15.9	0.2250	0.50	0.4526	12
VAL	47	0	0.2544	0.45	0.5628	4
SER	48	100	0.2360	0.59	0.3997	18

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1CSP		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
PHE	49	0	0.2532	0.55	0.4610	11
GLU	50	39.5	0.2131	0.75	0.2831	29
ILE	51	32.4	0.2190	0.94	0.2319	39
VAL	52	51.8	0.2092	0.85	0.2451	31
GLU	53	70	0.1785	1.53	0.1163	57
GLY	54	58.6	0.1897	1.27	0.1494	53
ASN	55	100	0.1698	2.14	0.0795	64
ARG	56	96	0.1782	2.86	0.0624	65
GLY	57	26.3	0.1988	1.38	0.1443	55
PRO	58	54.2	0.2176	0.92	0.2354	36
GLN	59	44.5	0.2346	1.04	0.2249	43
ALA	60	4	0.2607	0.55	0.4725	10
ALA	61	18.3	0.2436	0.51	0.4749	9
ASN	62	42.8	0.2206	0.92	0.2407	33
VAL	63	0.1	0.2590	0.48	0.5432	6
THR	64	38.8	0.2294	0.70	0.3260	26
LYS	65	45.5	0.2180	0.94	0.2330	38
GLU	66	52.7	0.1953	1.62	0.1209	56
ALA	67	100	0.1676	2.92	0.0573	67

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1C9O	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
MET	1	70	0.1950	0.96	0.2028	46
GLN	2	36.9	0.2093	0.77	0.2715	35
ARG	3	65.8	0.2051	0.77	0.2662	37
GLY	4	4.8	0.2422	0.41	0.5978	7
LYS	5	33.1	0.2231	1.08	0.2062	45
VAL	6	1.8	0.2474	0.41	0.6047	6
LYS	7	37.8	0.2153	0.67	0.3196	31
TRP	8	39.7	0.2169	0.56	0.3886	19
PHE	9	14.7	0.2399	0.61	0.3952	17
ASN	10	31	0.2007	0.70	0.2857	33
ASN	11	49.3	0.1924	1.26	0.1527	56
GLU	12	90.4	0.1622	1.32	0.1225	61
LYS	13	57.7	0.1813	1.14	0.1597	53
GLY	14	6.8	0.2207	0.60	0.3693	20
TYR	15	30.3	0.2174	0.67	0.3242	28
GLY	16	0	0.2536	0.39	0.6579	2
PHE	17	28.1	0.2416	0.53	0.4530	13
ILE	18	0	0.2662	0.41	0.6563	3
GLU	19	45.8	0.2215	0.85	0.2620	39
VAL	20	18.1	0.2294	0.59	0.3920	18
GLU	21	91.4	0.1941	1.28	0.1510	57
GLY	22	100	0.1769	1.27	0.1391	59
GLY	23	45.9	0.1937	0.91	0.2129	44
SER	24	100	0.2007	1.06	0.1891	48

Protein 1C9O	Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position	
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
ASP	25	41.9	0.2161	0.76	0.2844	34
VAL	26	7	0.2509	0.45	0.5556	11
PHE	27	28.3	0.2452	0.48	0.5150	12
VAL	28	0	0.2678	0.39	0.6922	1
HIS	29	33.3	0.2291	0.54	0.4211	16
PHE	30	40.4	0.2141	0.85	0.2515	42
THR	31	68.7	0.2127	0.62	0.3455	24
ALA	32	11.3	0.2434	0.55	0.4429	15
ILE	33	5.8	0.2435	0.72	0.3401	25
GLN	34	45.7	0.2092	1.30	0.1606	52
GLY	35	66.8	0.1958	1.35	0.1451	58
GLU	36	98.9	0.1621	2.55	0.0637	65
GLY	37	53.7	0.1691	2.19	0.0771	64
PHE	38	72.9	0.1948	3.23	0.0603	66
LYS	39	31.6	0.1969	2.27	0.0867	63
THR	40	21.1	0.2183	1.10	0.1976	47
LEU	41	2.7	0.2443	0.66	0.3686	21
GLU	42	68.4	0.2026	1.28	0.1586	54
GLU	43	77.5	0.1935	1.25	0.1552	55
GLY	44	55.4	0.2084	0.63	0.3284	27
GLN	45	29.5	0.2234	0.83	0.2682	36
GLU	46	33.3	0.2087	0.82	0.2531	40
VAL	47	0	0.2521	0.42	0.6072	5
SER	48	100	0.2354	0.53	0.4449	14

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 1C9O		Solvent Access	Dynamical entropy	Movement restriction	SDRI	Residue ranking position
Residue		% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$	
PHE	49	0	0.2542	0.41	0.6233	4
GLU	50	37.6	0.2125	0.66	0.3232	29
ILE	51	50.3	0.2200	0.62	0.3569	23
VAL	52	39	0.2104	0.67	0.3126	32
GLN	53	90.6	0.1841	1.14	0.1615	51
GLY	54	40.7	0.1902	0.81	0.2343	43
ASN	55	98.4	0.1728	1.32	0.1307	60
ARG	56	87.4	0.1932	1.68	0.1150	62
GLY	57	23.9	0.2008	0.76	0.2647	38
PRO	58	60.6	0.2149	0.65	0.3295	26
GLN	59	35	0.2369	0.65	0.3641	22
ALA	60	8.3	0.2618	0.44	0.5924	8
ALA	61	22.6	0.2428	0.42	0.5723	10
ASN	62	35.6	0.2182	0.86	0.2525	41
VAL	63	0	0.2582	0.44	0.5897	9
VAL	64	37.5	0.2271	0.71	0.3202	30
LYS	65	40.5	0.2118	1.25	0.1696	50
LEU	66	62.1	0.1951	1.03	0.1888	49

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2N34		Solvent Access	Dynamical entropy	Movement restriction Main Chain	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
ARG	1	100	0.0907	12.3	0.0074	70
SER	2	100	0.0897	11.5	0.0078	69
LEU	3	76.5	0.1040	10.2	0.0102	68
ARG	4	80.9	0.1233	9.4	0.0131	67
GLN	5	86.6	0.1291	9.1	0.0143	66
ARG	6	89.1	0.1420	8.8	0.0161	64
LEU	7	64.1	0.1485	6.9	0.0214	55
GLN	8	75.1	0.1552	5.0	0.0310	37
ASP	9	68.6	0.1512	3.2	0.0470	30
THR	10	85.7	0.1532	5.3	0.0288	41
VAL	11	79.2	0.1614	7.5	0.0214	54
GLY	12	56.6	0.1484	9.9	0.0151	65
LEU	13	87.9	0.1688	7.8	0.0216	53
CYS	14	84	0.1614	5.6	0.0286	42
PHE	15	80.3	0.1701	3.4	0.0499	29
PRO	16	60.9	0.1649	6.2	0.0266	47
MET	17	78.5	0.1654	8.9	0.0186	61
ARG	18	78.7	0.1706	10.4	0.0165	62
THR	19	74.9	0.1704	8.4	0.0203	56
TYR	20	76.6	0.1765	6.9	0.0255	50
SER	21	100	0.1664	7.3	0.0228	52
LYS	22	93	0.1804	8.9	0.0202	58
GLN	23	89.2	0.1795	9.2	0.0196	59

Protein 2N34		Solvent Access	Dynamical entropy	Movement restriction Main Chain	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
SER	24	100	0.1703	8.4	0.0203	57
LYS	25	90.3	0.1905	7.3	0.0261	49
PRO	26	76.2	0.1822	7.0	0.0261	48
LEU	27	82	0.1960	6.5	0.0300	39
PHE	28	85	0.1954	6.0	0.0324	35
SER	29	100	0.1883	6.4	0.0292	40
ASN	30	68.6	0.1981	7.3	0.0273	45
LYS	31	66.5	0.2115	7.6	0.0279	43
ARG	32	93.4	0.2139	7.7	0.0277	44
LYS	33	68.3	0.2185	8.2	0.0267	46
ILE	34	54.7	0.2336	7.3	0.0319	36
HIS	35	69.4	0.2356	5.4	0.0434	31
LEU	36	96	0.2236	2.8	0.0794	26
SER	37	100	0.2221	2.0	0.1129	18
GLU	38	40.8	0.2477	1.9	0.1292	17
LEU	39	59.6	0.2540	2.6	0.0994	21
MET	40	47.8	0.2545	3.2	0.0791	27
LEU	41	81.5	0.2626	3.3	0.0800	25
GLU	42	79.7	0.2310	2.3	0.0997	20
LYS	43	61.2	0.2438	1.2	0.1977	13
CYS	44	55.6	0.2408	0.6	0.4159	3
PRO	45	68.9	0.2252	1.0	0.2176	11
PHE	46	37	0.2546	1.4	0.1799	14

File S2. SDRI. For each residue of the protein, it is shown its solvent access, Shannon dynamical entropy, RMSF, and its ranking position according to the theoretical scores.

Protein 2N34		Solvent Access	Dynamical entropy	Movement restriction Main Chain	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
PRO	47	51	0.2130	1.9	0.1092	19
ALA	48	92	0.1955	2.1	0.0930	23
GLY	49	94.5	0.1856	1.9	0.0964	22
SER	50	100	0.2145	1.4	0.1483	15
ASP	51	68.8	0.2114	0.8	0.2607	9
LEU	52	58.5	0.2415	0.7	0.3683	4
ALA	53	12.6	0.2430	0.7	0.3517	6
GLN	54	66.6	0.2349	0.7	0.3181	7
LYS	55	63.2	0.2401	0.7	0.3617	5
TRP	56	16.4	0.2857	0.6	0.5006	1
HIS	57	64.4	0.2556	0.6	0.4282	2
LEU	58	65.9	0.2435	0.8	0.3077	8
ILE	59	50.7	0.2604	1.0	0.2600	10
LYS	60	45.3	0.2691	1.3	0.2067	12
GLN	61	35.5	0.2613	2.0	0.1338	16
HIS	62	83	0.2463	3.0	0.0827	24
THR	63	71.3	0.2433	4.7	0.0517	28
ALA	64	26.2	0.2266	5.9	0.0381	32
PRO	65	89.7	0.2218	6.4	0.0344	34
VAL	66	58	0.2194	6.2	0.0356	33
SER	67	100	0.1987	6.5	0.0305	38
PRO	68	88.9	0.1930	7.9	0.0245	51
HIS	69	95.2	0.1835	9.4	0.0196	60

Protein 2N34		Solvent Access	Dynamical entropy	Movement restriction Main Chain	SDRI	Residue ranking position
Residue	% RSA	H_i	RMSF (Å)	$H_i * RMSF_i^{-1}$		
SER	70	100	0.1647	10.2	0.0162	63

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Mutant 2C12	Guanidinium chloride denaturation			Theoretical score ($H_i^*RMSF_i^{-1}$)
	m UF (kcal mol ⁻¹)	Cm (M)	$\Delta\Delta G$ (kcal mol ⁻¹)	
Wild Type (WT)	1.9	4	-	-
KA3	1.77	3.72	-0.53	0.2445
KM3	1.88	3.66	-0.65	0.2445
TA4	1.83	3.56	-0.84	0.2710
TV4	1.67	3.83	-0.32	0.2710
TG4	1.95	3.4	-1.14	0.2710
PA7	3.46	3.19	-1.54	0.3331
EA8	1.75	3.76	-0.46	0.2389
EQ8	1.91	3.68	-0.61	0.2389
LA9	2.15	2.62	-2.62	0.4727
KA12	1.52	4.22	0.42	0.2499
SG13	2.1	3.59	-0.78	0.3037
SA13	1.86	3.54	-0.87	0.3037
EQ15	1.87	3.85	-0.29	0.1830
ED15	1.83	3.73	-0.51	0.1830
EN15	2.01	3.64	-0.68	0.1830
EQ16	1.84	3.76	-0.46	0.1787
ED16	2.2	3.62	-0.72	0.1787
EN16	2.18	3.45	-1.05	0.1787
AG17	1.8	3.44	-1.06	0.3565
KA18	1.73	3.75	-0.48	0.2426
KG18	1.99	2.8	-2.28	0.2426
KA19	1.61	4.11	0.21	0.1736
KG19	1.73	3.49	-0.97	0.1736
VA20	2.01	3.75	-0.48	0.3463

Mutant 2C12	Guanidinium chloride denaturation			Theoretical score ($H_i^*RMSF_i^{-1}$)
	m UF (kcal mol ⁻¹)	Cm (M)	$\Delta\Delta G$ (kcal mol ⁻¹)	
IV21	1.99	3.33	-1.27	0.5402
LA22	1.93	3.32	-1.29	0.3132
LG22	1.68	3.29	-1.35	0.3132
QA23	1.77	3.99	-0.02	0.1517
QG23	1.81	3.69	-0.59	0.1517
DA24	1.87	3.51	-0.93	0.2935
KA25	1.57	3.67	-0.63	0.4207
KG25	1.75	2.36	-3.12	0.4207
PA26	2.07	3.09	-1.73	0.2788
EA27	1.78	3.83	-0.32	0.2621
IV30	1.99	3.43	-1.08	0.3704
IA30	2.07	1.99	-3.82	0.3704
IV31	1.77	4.04	0.08	0.3599
IA31	2.22	2.91	-2.07	0.3599
IG31	2.13	2.19	-3.44	0.3599
IT31	1.97	3.31	-1.31	0.3599
LA32	2.11	2.78	-2.32	0.2867
LI32	1.7	3.87	-0.25	0.2867
LV32	1.76	3.74	-0.49	0.2867
PA33	1.79	3.91	-0.17	0.2832
VT34	1.65	3.47	-1.01	0.2250
VA34	1.7	3.67	-0.63	0.2250
VG34	2	2.75	-2.38	0.2250
TS36	1.66	3.99	-0.02	0.0922
TV36	1.67	3.61	-0.74	0.0922

Reference. J. Mol. Biol. (1995) 254: 260-288

$$\Delta\Delta G = m_{WT}*(Cm_{MUT}-Cm_{WT})$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Mutant 2CI2	Guanidinium chloride denaturation			Theoretical score ($H_i^*RMSF_i^{-1}$)
	m UF (kcal mol ⁻¹)	Cm (M)	$\Delta\Delta G$ (kcal mol ⁻¹)	
TA36	1.5	4.12	0.23	0.0922
IA38	1.59	3.98	-0.04	0.0851
VA39	2.14	3.24	-1.44	0.1256
TA40	1.89	3.63	-0.70	0.0907
TD40	1.87	4.01	0.02	0.0907
EA42	1.78	3.64	-0.68	0.1109
RA44	1.78	3.7	-0.57	0.2388
DA46	1.83	3.59	-0.78	0.2516
VA48	1.76	1.46	-4.83	0.4466
LA50	1.98	2.02	-3.76	0.3860
FL51	1.95	2.91	-2.07	0.3015
FV51	2.24	2.77	-2.34	0.3015
FA51	2.17	2.02	-3.76	0.3015
VA52	2.15	2.98	-1.94	0.4309
DA53	1.99	2.24	-3.34	0.1493
DN53	1.6	4	0.00	0.1493
ND57	1.8	3.38	-1.18	0.2372
NA57	1.79	3.57	-0.82	0.2372
IV58	1.82	4.1	0.19	0.5275
IA58	1.93	1.79	-4.20	0.5275
AG59	2.19	3.03	-1.84	0.3498
VT61	1.63	3.8	-0.38	0.4397
VA61	1.92	3.22	-1.48	0.4397
VG61	2.61	2.33	-3.17	0.4397
PA62	1.8	2.28	-3.27	0.5453

Mutant 2CI2	Guanidinium chloride denaturation			Theoretical score ($H_i^*RMSF_i^{-1}$)
	m UF (kcal mol ⁻¹)	Cm (M)	$\Delta\Delta G$ (kcal mol ⁻¹)	
VT64	1.79	3.41	-1.12	0.4074
VA64	1.95	3.25	-1.43	0.4074
VG64	2.1	2.2	-3.42	0.4074

Reference. J. Mol. Biol. (1995) 254: 260-288

$\Delta\Delta G = mWT*(CmMUT-CmWT)$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 2W0K	Thermal unfolding			Theoretical score (H,*RMSF; ⁻¹)	Reference
	T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG (kcal mol ⁻¹)		
Wild Type (WT)	323.05	86.2	-	-	Biochem. Biophys. Res. Commun. (2014) 443: 495-499
Phe2Ser	316.95	73.6	-1.63	0.2779	
Phe2Leu	318.85	76.1	-1.12	0.2779	
Phe2Pro	316.75	73.7	-1.68	0.2779	
Phe2Trp	319.15	81.3	-1.04	0.2779	
Arg24Gly	317.35	77.2	-1.52	0.3273	Proteins (2008) 72: 684-692
Pro7Ser	316.15	79.5	-1.84	0.3799	J. Mol. Biol. (2010) 396: 280-292
His8Pro	323.25	84.8	0.05	0.1129	
His8Ser	322.65	84.4	-0.11	0.1129	
Arg24His	312.95 ± 1.15	56.47 ± 5.65	-2.70	0.3273	This work
Ile29Gly	309.53 ± 0.53	60.74 ± 2.55	-3.61	0.2057	
Tyr37Phe	322.90 ± 0.31	87.23 ± 3.06	-0.04	0.3204	
Gln6Asn	310.55 ± 0.30	10.39 ± 3.21	-3.34	0.5055	

$$\Delta\Delta G = (T_m\text{MUT} - T_m\text{WT}) * (\Delta H_{WT} / T_m\text{WT})$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1FTG	Thermal unfolding							Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	Reference
	Tm1 (K)	$\Delta H^\circ 1$ (kcal mol ⁻¹)	$\Delta \Delta G1$ (kcal mol ⁻¹)	Tm2 (K)	$\Delta H^\circ 2$ (kcal mol ⁻¹)	$\Delta \Delta G2$ (kcal mol ⁻¹)	Total $\Delta \Delta G^\circ$ (kcal mol ⁻¹)		
Wild Type (WT)	312	22	-	326	60	-	-	-	J. Mol. Biol. (2004) 344: 223-237
E20K	309.3	32.7	-0.19	331.2	66.2	0.96	0.77	0.1478	
E40K	316.5	27	0.32	332.5	59.6	1.20	1.51	0.1159	
E72K	312.8	2	0.06	333.6	68.4	1.40	1.46	0.1292	
D75K	314.8	28.3	0.20	330	54.3	0.74	0.93	0.0523	
D126K	320	21.8	0.56	327.5	39.7	0.28	0.84	0.1315	
D150K	305.8	25.1	-0.44	326.9	48	0.17	-0.27	0.1174	
WT	317.3	33.9	-	329	52.7	-	-	-	J. Mol. Biol. (2004) 344: 239-255
L6A	305.5	28.7	-1.26	319.1	23.8	-1.59	-2.85	0.2737	
I22V	311	33	-0.67	324.7	37.3	-0.69	-1.36	0.2987	
V31A	312.6	30.9	-0.50	324.2	41.5	-0.77	-1.27	0.2008	
D43A	317.7	34.1	0.04	329.6	49.5	0.10	0.14	0.1284	
I51V	313.2	31.2	-0.44	324.3	38.8	-0.75	-1.19	0.4664	
I52V	315.9	39.8	-0.15	324.9	54.5	-0.66	-0.81	0.4069	
S71A	316.3	34.5	-0.11	329.3	55	0.05	-0.06	0.1517	
A84G	313.3	34.4	-0.43	323.6	31.9	-0.86	-1.29	0.4669	
N97A	314.9	28.6	-0.26	328.3	48.7	-0.11	-0.37	0.3062	
Q99A	322.5	52.7	0.56	331.3	56.2	0.37	0.92	0.3449	
I104V	313.3	29.5	-0.43	327.9	54.5	-0.18	-0.60	0.3251	
S110A	315.6	32.5	-0.18	328.7	44.3	-0.05	-0.23	0.2247	
V117A	309.3	36.9	-0.85	322.6	33.7	-1.03	-1.88	0.3388	
T122S	323.2	35.5	0.63	330.3	50.7	0.21	0.84	0.2422	
V139A	310.9	29	-0.68	327.5	50.8	-0.24	-0.92	0.4397	
L143A	315.8	33.3	-0.16	328.9	50	-0.02	-0.18	0.3300	
I156V	308.8	25.3	-0.91	320.3	29.9	-1.39	-2.30	0.3094	
V160A	309.1	27.7	-0.88	322.4	49.2	-1.06	-1.93	0.3029	

$$\Delta \Delta G^\circ = (T_m \text{MUT} - T_m \text{WT}) * (\Delta H^\circ \text{WT} / T_m \text{WT})$$

$$\text{Total } \Delta \Delta G = \Delta \Delta G1 + \Delta \Delta G2$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1ARR	Thermal unfolding			Theoretical score ($H_i^*RMSF_i^{-1}$)
	T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG (kcal mol ⁻¹)	
WT-st6	332.15	62.3	-	-
PA8	347.25	80.2	2.83	0.1052
SA35	336.55	57.1	0.83	0.1490
NA11	335.25	57.2	0.58	0.1026
KA6	332.75	59	0.11	0.0429
RA16	332.65	60	0.09	0.0798
VA25	332.45	58.3	0.06	0.1488
MA4	332.35	59.4	0.04	0.0325
EA27	331.95	61.6	-0.04	0.0806
KA2	331.85	57.2	-0.06	0.0179
QA9	331.55	59.5	-0.11	0.0911
GA3	331.25	60	-0.17	0.0281
MA1	331.15	59.7	-0.19	0.0151
SA5	330.65	53.3	-0.28	0.0449
RA13	330.45	51.3	-0.32	0.1004
EA17	330.15	62.1	-0.38	0.0809
VA18	330.05	63.1	-0.39	0.1492
EA43	329.25	59.4	-0.54	0.1210
MA7	328.65	52.6	-0.66	0.0927
DA20	328.45	56.9	-0.69	0.1036
LA19	321.45	48.7	-2.01	0.1462
FA10	313.75	34.8	-3.45	0.1167
WT-st11	331.05	58.5	-	-
NA34	336.15	57.1	0.90	0.1473
QA39	334.55	61.6	0.62	0.1441
GA52	334.05	57.7	0.53	0.0702
KA46	330.25	52.9	-0.14	0.0818

Protein 1ARR	Thermal unfolding			Theoretical score ($H_i^*RMSF_i^{-1}$)
	T _m (K)	ΔH° (kcal mol ⁻¹)	ΔΔG (kcal mol ⁻¹)	
RA23	329.85	56.5	-0.21	0.0861
KA24	329.45	59.9	-0.28	0.1008
EA28	328.85	52.1	-0.39	0.0787
IA51	324.05	47.4	-1.24	0.0653
GA49	321.85	45.8	-1.63	0.0458
GA30	321.05	48.8	-1.77	0.0784
RA50	321.05	41.9	-1.77	0.0564
KA47	320.35	41.2	-1.89	0.0644
PA15	319.75	38.1	-2.00	0.1278
SA44	319.45	41	-2.05	0.1514
NA29	318.45	44.2	-2.23	0.0906
VA33	317.25	39.7	-2.44	0.1483
EA48	316.35	29.3	-2.60	0.0399
LA12	315.45	40.4	-2.76	0.1413
LA21	312.75	36.7	-3.23	0.1426
RA31	310.25	34.2	-3.68	0.1367
MA42	308.75	26.1	-3.94	0.1299
SA32	306.65	24.4	-4.31	0.1140
YA38	306.15	24.1	-4.40	0.1647
WA14	304.65	25.3	-4.67	0.1946
RA40	304.35	13.6	-4.72	0.2267
VA22	UNF	-	-	0.1732
EA36	UNF	-	-	0.2077
IA37	UNF	-	-	0.2484
VA41	UNF	-	-	0.2476
FA45	UNF	-	-	0.1352

$$\Delta\Delta G^\circ = (T_m \text{MUT} - T_m \text{WT}) * (\Delta H^\circ \text{WT} / T_m \text{WT})$$

$$\text{Total } \Delta\Delta G = \Delta\Delta G1 + \Delta\Delta G2$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1HCQ	Dimer recognizing DNA		Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	Reference
	K_d (nM)	K_d Difference (%)		
Wild Type (WT)	43	-	-	Biochemistry (2010) 49: 5978-5988
D12A	67	-56	0.0685	
H18A	190	-342	0.0503	
E25A	172	-300	0.0859	
S15A	102	-137	0.0478	
Y19A	316	-635	0.0600	
S23A	119	-177	0.0670	
K28A	313	-628	0.0740	
K32A	326	-658	0.0772	
R33A	745	-1633	0.0865	

$$\% K_d \text{ Difference} = (K_d \text{WT} - K_d \text{MUT}) / K_d \text{WT} * 100$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1CSP	Thermal unfolding			Theoretical score ($H_i^*RMSF_i^{-1}$)	References
	Tm (K)	ΔH° (kcal mol ⁻¹)	$\Delta\Delta G$ (kcal mol ⁻¹)		
Csp S WT	53.6	46.13	-	-	JMB (2001) 313, 343-357
E3R	69.6	52.34	2.26	0.1703	
E3L	62.7	47.80	1.28	0.1703	
A46E	48.6	39.67	-0.71	0.4526	
E66L	66.4	53.54	1.81	0.1209	
Csp S WT	53.8	46.13	-	-	JMB (2005) 347, 1063-1076
E3K	70.4	48.52	2.34	0.1703	
K65I	62.4	48.04	1.21	0.2330	
E66K	66.7	54.02	1.82	0.1209	
E43S	54.7	44.46	0.13	0.1652	
A46K	62.2	51.63	1.19	0.4526	
Csp S WT	52.3	39.20	-	-	JMB (2007) 366, 842-856
E3R	70.2	50.19	2.16	0.1703	
E3Q	62.6	46.61	1.24	0.1703	
K7E	UNF	-	-		
K7Q	UNF	-	-		
N10D	57.7	43.26	0.65	0.2892	
D25Q	44.2	29.16	-0.98	0.2367	
S48E	53	38.00	0.08	0.3997	
E50Q	42.6	34.18	-1.17	0.2831	
R56Q	55.3	42.54	0.36	0.0624	
Csp S WT*	55.3	38.24	-		JMB (2007) 366, 842-856
E3R*	71.6	49.95	1.90	0.1703	
K5E*	22.2	14.58	-3.85	0.2319	
K5Q*	38.6	29.88	-1.94	0.2319	
N10K*	41.1	27.49	-1.65	0.2892	
E12K**	50.2	37.76	-0.59	0.1121	

$$\Delta\Delta G^\circ = (Tm_{MUT} - Tm_{WT}) * (\Delta H^\circ_{WT} / Tm_{WT})$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1CSP	Thermal unfolding			Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	Reference
	Tm (K)	ΔH° (kcal mol ⁻¹)	$\Delta \Delta G$ (kcal mol ⁻¹)		
K13E*	52.7	36.57	-0.30	0.1539	JMB (2007) 366, 842-856
K13Q*	54	41.35	-0.15	0.1539	
E19K*	50.5	37.28	-0.56	0.2408	
E19Q*	53.7	39.91	-0.19	0.2408	
V20Q*	39.5	29.64	-1.84	0.2349	
V20E*	UNF				
V20K*	37.2	25.81	-2.11	0.2349	
E21K*	54.3	43.26	-0.12	0.0846	
E21Q*	52.5	34.89	-0.33	0.0846	
D24K*	50.9	34.18	-0.51	0.1466	
D24N*	48.4	31.07	-0.80	0.1466	
D25K*	35	13.38	-2.36	0.2367	
K39E*	52.8	35.85	-0.29	0.2266	
K39Q*	54.9	35.37	-0.05	0.2266	
E42K*	55.3	42.78	0.00	0.1619	
E42Q*	54.1	35.61	-0.14	0.1619	
E43K*	56.5	46.13	0.14	0.1652	
S48K*	61.6	41.59	0.73	0.3997	
E50K*	49.7	31.55	-0.65	0.2831	
E53K*	53.7	40.15	-0.19	0.1163	
E53Q*	56.6	43.74	0.15	0.1163	
N55K*	54.3	37.76	-0.12	0.0795	
N55D*	59.2	44.22	0.45	0.0795	
K56E*	38.2	24.14	-1.99	0.0624	
K65Q*	48	31.31	-0.85	0.2330	

$$\Delta \Delta G^\circ = (T_m \text{MUT} - T_m \text{WT}) * (\Delta H^\circ \text{WT} / T_m \text{WT})$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Protein 1C9O	Thermal unfolding			Theoretical score ($H_i^*RMSF_i^{-1}$)	Reference
	Tm (K)	ΔH° (kcal mol ⁻¹)	$\Delta \Delta G$ (kcal mol ⁻¹)		
Csp C WT	76.9	58.56	-		JMB (2001) 313, 343-357
Q2L	80.9	57.84	0.67	0.2715	
R3E	59.1	45.65	-2.98	0.2662	
R3L	70.9	55.69	-1.00	0.2662	
R3K	76.1	53.54	-0.13	0.2662	
R3A	64.9	51.39	-2.01	0.2662	
N11S	79.2	58.80	0.38	0.1527	
Y15S	76.7	56.64	-0.03	0.3242	
E21A	75	57.84	-0.32	0.1510	
G23Q	74.7	59.51	-0.37	0.2129	
S24D	77.8	61.66	0.15	0.1891	
T31S	77.8	59.75	0.15	0.3455	
E46A	75.4	57.60	-0.25	0.2531	
Q53E	76.1	58.80	-0.13	0.1615	
V64T	75.1	58.32	-0.30	0.3202	
L66E	68.9	50.19	-1.34	0.1888	
Csp C WT	76.9	58.56	-		Prot Eng Des Sel (2006) 19, 355-358
R56E	80	60.95	0.52	0.1150	
H29E	73.5	105.88	-0.57	0.4211	
E46K	74	106.84	-0.49	0.2531	
E21K	75.4	49.00	-0.25	0.1510	
N55K	77.7	55.21	0.13	0.1307	
E36K	76	45.65	-0.15	0.0637	
E50K	73.5	43.26	-0.57	0.3232	
E12K	74.4	52.34	-0.42	0.1225	

$$\Delta \Delta G^\circ = (Tm_{MUT} - Tm_{WT}) * (\Delta H^\circ_{WT} / Tm_{WT})$$

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Thermodynamic data from unfolding kinetics		
Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G_{\text{H}_2\text{O}}$ (kcal mol ⁻¹)
KA2	0.2445	-0.72
KM2	0.2445	-0.47
TA3	0.2710	-0.63
TV3	0.2710	-0.27
TG3	0.2710	-0.81
PA6	0.3331	-0.84
EA7	0.2389	-0.27
LA8b	0.4727	-2.73
KA11	0.2499	0.21
SG12	0.3037	-0.36
SA12	0.3037	-0.3
EQ14	0.1830	-0.04
ED14	0.1830	-0.16
EN14	0.1830	-0.02
EQ15	0.1787	-0.2
ED15	0.1787	-0.45
EN15	0.1787	-0.65
AG16	0.3565	0.26
KA17	0.2427	-0.72
KG17	0.2427	-1.72
KA18	0.1736	-0.09
KG18	0.1736	-0.41
VA19b	0.3463	-0.73
IV20b	0.5402	-0.6
LA21	0.3132	-0.95
LG21	0.3132	-1.1

Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G_{\text{H}_2\text{O}}$ (kcal mol ⁻¹)
QA22	0.1517	-0.03
QG22	0.1517	-0.46
DA23	0.2935	-0.69
KA24	0.4207	-0.68
KG24	0.4207	-2.7
PA25	0.2788	-1.55
EA26	0.2621	-0.25
IV29b	0.3704	-0.62
IA29b	0.3704	-2.8
IV30	0.3599	-0.03
IA30	0.3599	-0.91
IG30	0.3599	-2.22
IT30	0.3599	-0.5
LA32	0.3599	-1.6
LI32	0.2867	-0.29
LV32	0.2867	-0.48
VT34	0.2250	-0.75
VA34	0.2250	-0.49
VG34	0.2250	-1.77
TV36	0.2250	-0.71
VA38	0.1257	-1.08
TA39c	0.0907	-0.38
TD39c	0.0907	0.15
EA41c	0.1109	-0.56
RA43	0.2388	-0.47
DA45	0.2516	-0.57
VA47b	0.4466	-3.53
LA49b	0.3860	-1.77

Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G_{\text{H}_2\text{O}}$ (kcal mol ⁻¹)
FL50	0.3015	-1.13
FV50	0.3015	-1.64
FA50	0.3015	-2.32
VA51b	0.4309	-1.09
DA52	0.1493	-2.9
DN52	0.1493	-0.07
ND56	0.2372	-0.66
NA56	0.2372	-0.55
IV57b	0.2372	0.16
IA57b	0.2372	-3.77
AG58	0.3498	-1.52
VT60	0.4397	-0.37
VA60	0.4397	-1.25
VG60	0.4397	-3.02
PA61	0.5453	-2.97
VT62	0.4074	-1
VA82	0.4074	-1.09
VG82	0.4074	-3

File S3. Experimental values from thermodynamic experiments. All experiments are described in references. In Csp S, mutants marked with an asterisk (*) means that the recombinant protein was assessed with a Histidine tag.

Thermodynamic data from refolding kinetics		
Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G$ (kcal mol ⁻¹)
KA2	0.2445	-0.11
KM2	0.2445	0.02
TA3	0.2710	0.11
TV3	0.2710	0.17
TG3	0.2710	0.06
PA6	0.3331	0.1
EA7	0.2389	0.19
LA8b	0.4727	0.4
KA11	0.2499	0.26
SG12	0.3037	0.23
SA12	0.3037	0.38
EQ14	0.1830	0.36
ED14	0.1830	0.1
EN14	0.1830	0.53
EQ15	0.1787	0.25
ED15	0.1787	0.16
EN15	0.1787	0.57
AG16	0.3565	1.15
KA17	0.2427	0.14
KG17	0.2427	0.87
KA18	0.1736	0.08
KG18	0.1736	0.68
VA19b	0.3463	-0.13
IV20b	0.5402	0.52
LA21	0.3132	0.33
LG21	0.3132	0.48
QA22	0.1517	-0.12

Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G$ (kcal mol ⁻¹)
QG22	0.1517	0.07
DA23	0.2935	-0.23
KA24	0.4207	-0.23
KG24	0.4207	0.31
PA25	0.2788	0.35
EA26	0.2621	0.14
IV29b	0.3704	0.19
IA29b	0.3704	0.98
IV30	0.3599	0.07
IA30	0.3599	0.65
IG30	0.3599	0.92
IT30	0.3599	0.47
LA32	0.2867	0.44
LI32	0.2867	-0.08
LV32	0.2867	-0.02
VT34	0.2250	0.24
VA34	0.2250	-0.01
VG34	0.2250	0.4
TV36	0.0922	0.14
VA38	0.1257	0.18
TA39c	0.0907	0.04
TD39c	0.2388	0.05
EA41c	0.2516	0.31
RA43	0.4466	1.02
DA45	0.3015	0.59
VA47b	0.3015	0.6
LA49b	0.3015	1.16
FL50	0.4309	0.49

Mutant 2CI2	Theoretical score ($H_i^* \text{RMSF}_i^{-1}$)	$\Delta\Delta G$ (kcal mol ⁻¹)
FV50	0.1493	0.41
FA50	0.2372	0.23
VA51b	0.2372	0.07
DA52	0.5275	0.11
DN52	0.5275	0.36
ND56	0.3498	0.21
NA56	0.4397	0.19
IV57b	0.4397	-0.04
IA57b	0.4397	0.13
AG58	0.4074	0.09
VT60	0.4074	0.04
VA60	0.4074	0.1

File S4. Additional simulations. It is shown the RMSD, RMSF, and the hierarchy position of each residue in two different conditions. In Figure R1, we show the differences between the complex estrogen receptor bound with DNA after 25 ns molecular dynamics simulation and the estrogen receptor unbound from DNA after 50 ns simulation. In Figure R2, we show the convergence between simulating 25 ns and extending simulation to 50 ns of 6aJL2.

1- Human receptor estrogen alpha (1HCQ PDB code) is in complex with DNA in MD simulation. Does the experiment done in complex with DNA, too? It would be more informative if you could compare the crucial residues in this protein in the presence and absence of DNA.

Experiments detailed in Ref. Deegan *et al.*, 2010, where DNA was bound to the estrogen receptor and some of its mutants, concluded that mutants decreased the DNA recognition by estrogen receptor. To compare the crucial residues in the absence of DNA bound to the receptor, we performed a 50 ns molecular dynamics simulation under same conditions detailed in Materials and Methods. Results showed that the unbound dimer has different motion from motions of DNA-bound dimer as shown in RMSD and RMSF plots (Fig. R1A and Fig. R1B, respectively). By comparing normalized theoretical scores of the residues between these two simulations, there are notable changes on the contribution of each residue into the dimer (Fig. R1C).

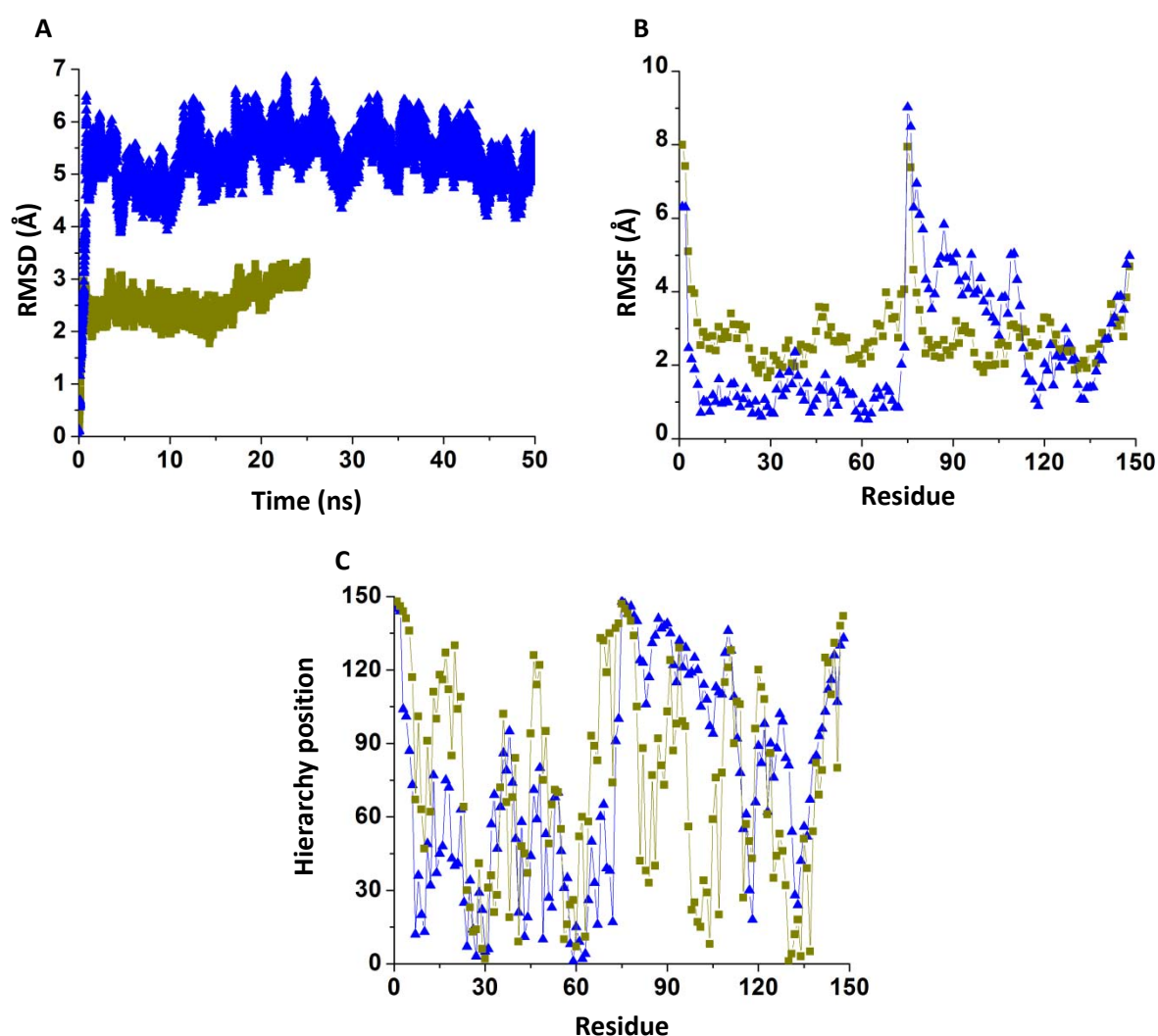


Fig. R1. Comparison between free estrogen dimer and complexed with DNA. In yellow is shown the difference of the complex estrogen receptor bound with DNA after 25 ns molecular dynamics simulation; and in blue, the estrogen receptor unbound from DNA after 50 ns simulation. **A.** RMSD difference. **B.** RMSF values of each residue. **C.** Hierarchy position of each residue.

File S4. Additional simulations. It is shown the RMSD, RMSF, and the hierarchy position of each residue in two different conditions. In Figure R1, we show the differences between the complex estrogen receptor bound with DNA after 25 ns molecular dynamics simulation and the estrogen receptor unbound from DNA after 50 ns simulation. In Figure R2, we show the convergence between simulating 25 ns and extending simulation to 50 ns of 6aJL2.

All the results of simulating estrogen receptor bound to DNA generated data which were remarkably different from the results obtained for the dimer unbound to DNA. However, as stated in the text, we cannot predict which residues are going to recognize precisely interaction with its respective ligand.

2- Even though RMSD of most proteins show equilibrated state by 25ns simulation time, however it is still not enough time to consider as equilibration. I recommend authors select one of the proteins and continue the simulation for at least another 25ns and compare it with non-extended simulation.

We extended 6aJL2 (PDB 2W0K) simulation and compared theoretical scores as shown below. Trajectory maintained the same RMSD difference (Fig. R2A). Slight, but not significant changes could be detected in RMSF values (Fig. R2B). Also, theoretical scores as well as hierarchy were not dramatically affected by extending another 25 ns (Fig. R2C).

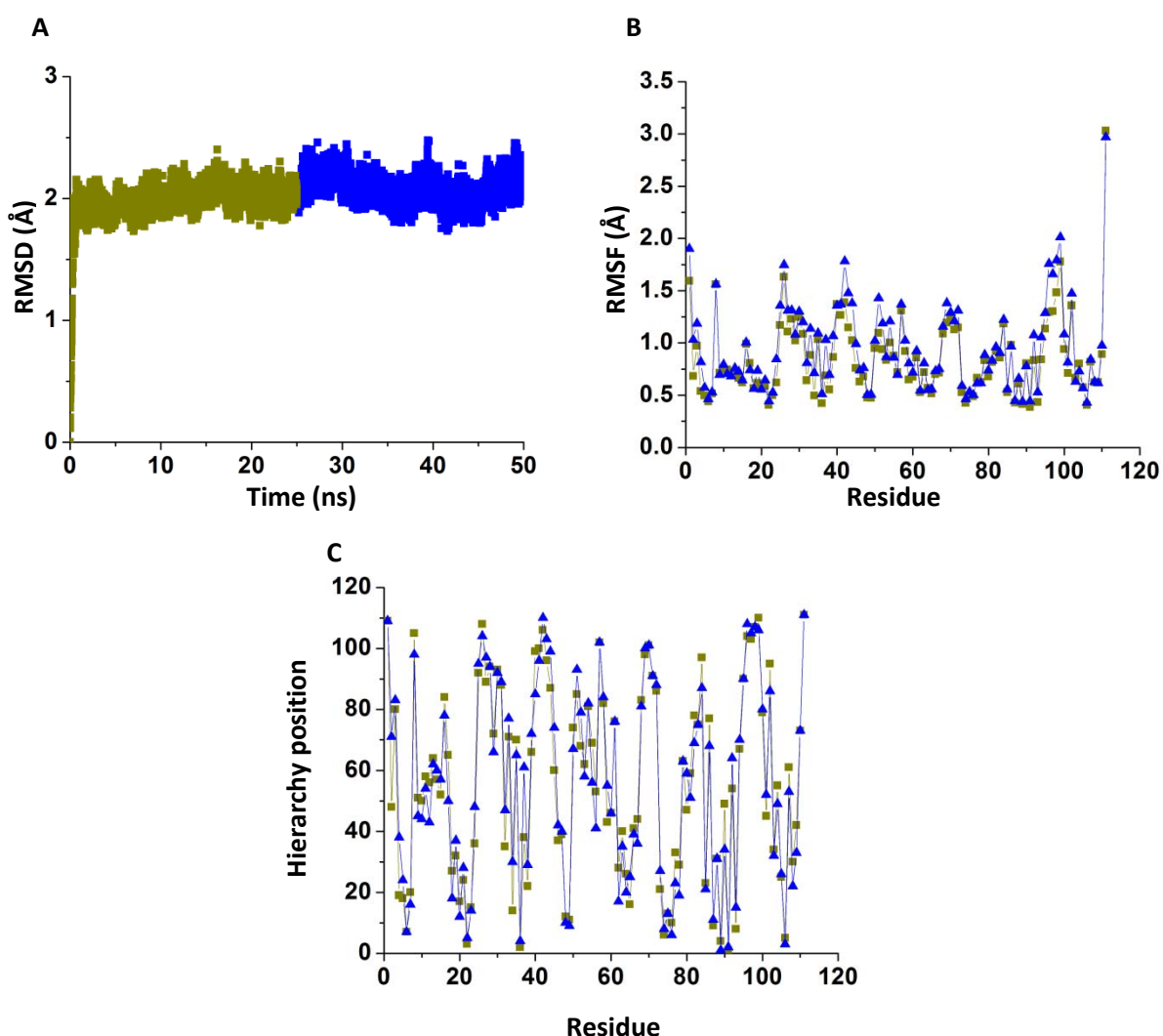


Fig. R2. Comparison between molecular time simulations. In yellow is shown the structure 6aJL2 after 25 ns molecular dynamics simulation; in blue, 6aJL2 after extending simulation to 50 ns simulation. **A.** RMSD difference. **B.** RMSF values of each residue. **C.** Hierarchy position of each residue.

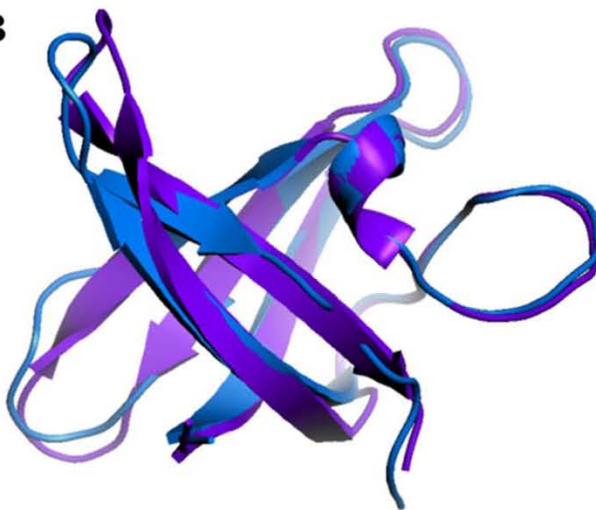
By extending simulation time 25 ns more, hierarchy position of each residue was not considerably modified.

Fig. S1. Overlapping between homologous structures. (A) Sequence aligning between cold shock proteins from *B. caldolyticus* (PDB-1C9O) and *B. subtilis* (PDB-1CSP). In green are remarked identical residues; in pink, similar residues; and in blue, sequence mismatch. (B) Despite that having different thermodynamic parameters, overlapping cold shock proteins shows high structural homology with a main chain RMSD of 0.5 Å calculated with PyMOL software

A

```
1C9O MQRGKVKWFNNEKGYGFIEVEGGSDVVFVHFTAIQEGEGFKTLEEGQAVSFEIVEGGNRGPQAANVTKE  
1CSP .LE.....S...F.....QD.....S.....E.....Q.....V.L
```

B



Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
1  #!/usr/bin/perl
2  # Simple Approach for Ranking Structure Determining Residues
3  open (ARCH, "100_estructuras.pdb") or die "Cannot open file \n";
4
5  $Model = 1;
6  while ($linea = <ARCH>)
7  {
8      if ($linea =~ "^MODEL")
9      {
10         $Num_res = 0;
11         $Comparar_num_res = 0;
12         $Continuidad_conteo = 0;
13         do {
14             shift (@baseline_molecule);
15             $vaciar_push = @baseline_molecule;
16         } until ($vaciar_push == 0);
17     do
18     {
19         shift (@baseline_VM_molecule);
20         $vaciar_push = @baseline_VM_molecule;
21     } until ($vaciar_push == 0);
22     do {
23         shift (@baseline_residue);
24         $vaciar_push = @baseline_residue;
25     } until ($vaciar_push == 0);
26     do {
27         shift (@baseline_VM_residue);
28         $vaciar_push = @baseline_VM_residue;
29     } until ($vaciar_push == 0);
30     do {
31         shift (@Num_res);
32         $vaciar_push = @Num_res;
33     } until ($vaciar_push == 0);
34     $Archsal = "Model_$Model.pl";
35     open (ARCHSAL, ">$Archsal") || die "Cannot open file \n";
36     print ARCHSAL "#Model_$Model\n";
37     $baseline_molecule = "molecule = (";
38     @baseline_molecule = split ("", $baseline_molecule);
39     unshift (@baseline_molecule, "@");
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
40     $baseline_PM_molecule = "PM_molecule = (";
41     @baseline_PM_molecule = split ("", $baseline_PM_molecule);
42     unshift (@baseline_PM_molecule, "@");
43     $baseline_VM_molecule = "VM_molecule = (";
44     @baseline_VM_molecule = split ("", $baseline_VM_molecule);
45     unshift (@baseline_VM_molecule, "@");
46     }
47 if ($linea =~ /^TER")
48     {
49     $Continuidad_conteo = $Num_res;
50     }
51 if ($linea =~ /^ENDMDL")
52     {
53     pop @baseline_residue;
54     push (@baseline_residue, "");\n");
55     $Temporal_3 = join ("",@baseline_residue);
56     print ARCHSAL "$Temporal_3";
57     pop @baseline_PM_residue;
58     push (@baseline_PM_residue, "");\n");
59     $Temporal_33 = join ("",@baseline_PM_residue);
60     print ARCHSAL "$Temporal_33";
61     push (@baseline_VM_residue, "");\n");
62     $Temporal_33 = join ("",@baseline_VM_residue);
63     print ARCHSAL "$Temporal_33";
64     $n=1;
65     do    {
66         print ARCHSAL '$count_atm_res_[',($n-1),"] = ",($count_atm_res_[$n]-1), ",\n" ;
67         $n++;
68     } until ($n> $Num_res);
69     $n= 0;
70     do    {
71         $count_atm_res_[$n] = 0;
72         $n++;
73     } until ($n> $Num_res);
74     pop @baseline_molecule;
75     push (@baseline_molecule, "");\n");
76     $Temporal_4 = join ("",@baseline_molecule);
77     print ARCHSAL "$Temporal_4";
78     pop @baseline_PM_molecule;
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
79     push (@baseline_PM_molecule, "");\n");
80     $Temporal_44 = join (",@baseline_PM_molecule);
81     print ARCHSAL "$Temporal_44";
82     pop @baseline_VM_molecule;
83     push (@baseline_VM_molecule, "");\n");
84     $Temporal_44 = join (",@baseline_VM_molecule);
85     print ARCHSAL "$Temporal_44";
86     print ARCHSAL '$',"Model = $Model;\n";
87     $matrix_salida = "matrix_center_mass_$Model.csv";
88     @matrix_salida = split ("", $matrix_salida);
89     unshift (@matrix_salida, '>');
90     unshift (@matrix_salida, "");
91     unshift (@matrix_salida, ',');
92     unshift (@matrix_salida, "");
93     unshift (@matrix_salida, "");
94     unshift (@matrix_salida, "");
95     unshift (@matrix_salida, ',');
96     unshift (@matrix_salida, "");
97     push (@matrix_salida, "");
98     push (@matrix_salida, ',');
99     push (@matrix_salida, "");
100    push (@matrix_salida, "");
101    push (@matrix_salida, "");
102    push (@matrix_salida, ',');
103    push (@matrix_salida, "");
104    $matrix_salida = join (",@matrix_salida);
105    print ARCHSAL '$Num_res = ',($Num_res-1),";\n";
106    print ARCHSAL "\n";
107    print ARCHSAL "open (ARCHSAL, "","",>center_mass_model_$Model.pl","",") || die "","", "Cannot open file ",'\',"n",","",";\n";
108    print ARCHSAL '$',"j= 0;\n";
109    print ARCHSAL " do\n";
110    print ARCHSAL " {\n";
111    print ARCHSAL '          $',"SMX = 0; ", '$',"SMY = 0; ", '$',"SMZ = 0; ", '$',"SMT = 0;\n";
112    print ARCHSAL '          $',"k= 0;\n";
113    print ARCHSAL " do\n";
114    print ARCHSAL " {\n";
115    print ARCHSAL '          $',"SMX= ", '$',"SMX+(",$',"PM_molecule[",$',"j][", '$',"k]*", '$',"molecule[",$',"j][", '$',"k][0]);\n";
116    print ARCHSAL '          $',"SMY= ", '$',"SMY+(",$',"PM_molecule[",$',"j][", '$',"k]*", '$',"molecule[",$',"j][", '$',"k][1]);\n";
117    print ARCHSAL '          $',"SMZ= ", '$',"SMZ+(",$',"PM_molecule[",$',"j][", '$',"k]*", '$',"molecule[",$',"j][", '$',"k][2]);\n";
```


Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
157         print ARCHSAL "           {\n";
158         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.01;\n";
159         print ARCHSAL "           }\n";
160         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","GLU","")\n";
161         print ARCHSAL "           {\n";
162         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 2.96;\n";
163         print ARCHSAL "           }\n";
164         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","GLY","")\n";
165         print ARCHSAL "           {\n";
166         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 2.25;\n";
167         print ARCHSAL "           }\n";
168         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","HIS","")\n";
169         print ARCHSAL "           {\n";
170         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.04;\n";
171         print ARCHSAL "           }\n";
172         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","ILE","")\n";
173         print ARCHSAL "           {\n";
174         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.09;\n";
175         print ARCHSAL "           }\n";
176         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","LEU","")\n";
177         print ARCHSAL "           {\n";
178         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.09;\n";
179         print ARCHSAL "           }\n";
180         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","LYS","")\n";
181         print ARCHSAL "           {\n";
182         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.18;\n";
183         print ARCHSAL "           }\n";
184         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","MET","")\n";
185         print ARCHSAL "           {\n";
186         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.09;\n";
187         print ARCHSAL "           }\n";
188         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","PHE","")\n";
189         print ARCHSAL "           {\n";
190         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 3.18;\n";
191         print ARCHSAL "           }\n";
192         print ARCHSAL "           elsif (",$',"VM_molecule[",$',"j][0] eq "","PRO","")\n";
193         print ARCHSAL "           {\n";
194         print ARCHSAL '           $',"VM_molecule[",$',"j][0] = 2.78;\n";
195         print ARCHSAL "           }\n";
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```

196     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"SER"',"'')\n";
197     print ARCHSAL "         {\n";
198     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 2.59;\n";
199     print ARCHSAL "         }\n";
200     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"THR"',"'')\n";
201     print ARCHSAL "         {\n";
202     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 2.81;\n";
203     print ARCHSAL "         }\n";
204     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"TRP"',"'')\n";
205     print ARCHSAL "         {\n";
206     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 3.39;\n";
207     print ARCHSAL "         }\n";
208     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"TYR"',"'')\n";
209     print ARCHSAL "         {\n";
210     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 3.23;\n";
211     print ARCHSAL "         }\n";
212     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"VAL"',"'')\n";
213     print ARCHSAL "         {\n";
214     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 2.93;\n";
215     print ARCHSAL "         }\n";
216     print ARCHSAL "         # Voss NR, Gerstein M. 2005. Calculation of standard atomic volumes for RNA and comparison with proteins: RNA is packed more
217     tightly. Journal of Molecular Biology 346:477–492. DOI: 10.1016/j.jmb.2004.11.072\n";
218     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"DG"',"'')\n";
219     print ARCHSAL "         {\n";
220     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 3.27;\n";
221     print ARCHSAL "         }\n";
222     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"DA"',"'')\n";
223     print ARCHSAL "         {\n";
224     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 3.22;\n";
225     print ARCHSAL "         }\n";
226     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"DC"',"'')\n";
227     print ARCHSAL "         {\n";
228     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 3.02;\n";
229     print ARCHSAL "         }\n";
230     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"DU"',"'')\n";
231     print ARCHSAL "         {\n";
232     print ARCHSAL '         $',"VM_molecule[",$',"j][0] = 2.98;\n";
233     print ARCHSAL "         }\n";
234     print ARCHSAL "         elsif (",$',"VM_molecule[",$',"j][0] eq ",'"',"DT"',"'')\n";

```


Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
274     {
275         @linea_larga_atomo = split ("", $linea);
276         $baseline_coordinate = "coordinate_";
277         @baseline_coordinate = split ("", $baseline_coordinate);
278         unshift (@baseline_coordinate, "@");
279         do      {
280             shift (@Num_res);
281             $vaciar_push = @Num_res;
282         } until ($vaciar_push == 0);
283     $n=20;
284     do      {
285         $n++$n;
286         } until ($linea_larga_atomo[$n] =~ '\d');
287     do      {
288         push (@Num_res, "$linea_larga_atomo[$n]");
289         $n++$n;
290         } until ($linea_larga_atomo[$n] =~ '\s');
291     $Num_res = join ("",@Num_res);
292     $Num_res = $Num_res + $Continuidad_conteo;
293     push (@baseline_coordinate, $Num_res);
294     push (@baseline_coordinate, "_");
295     $n=11;
296     do      {
297         $n++$n;
298         } until ($linea_larga_atomo[$n] =~ '\w');
299     if ($linea_larga_atomo[$n] eq 'N')
300         {
301             $PM_atomo = 14;
302         }
303     elsif ($linea_larga_atomo[$n] eq 'C')
304         {
305             $PM_atomo = 12;
306         }
307     elsif ($linea_larga_atomo[$n] eq 'H')
308         {
309             $PM_atomo = 1;
310         }
311     elsif ($linea_larga_atomo[$n] eq 'O')
312         {
```


Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
313         $PM_atomo = 16;
314     }
315     elsif ($linea_larga_atomo[$n] eq 'S')
316     {
317         $PM_atomo = 32;
318     }
319     elsif ($linea_larga_atomo[$n] eq 'P')
320     {
321         $PM_atomo = 31;
322     }
323     elsif ($linea_larga_atomo[$n] eq 'ZN')
324     {
325         $PM_atomo = 65;
326     }
327     elsif ($linea_larga_atomo[$n] eq 'NA')
328     {
329         $PM_atomo = 23;
330     }
331     do {
332         push (@baseline_coordinate, "$linea_larga_atomo[$n]");
333         $n++;
334     } until ($linea_larga_atomo[$n] =~ '\W');
335     @Temporal_1 = @baseline_coordinate;
336     unshift (@Temporal_1, "");
337     if ($Comparar_num_res == 0)
338     {
339         $baseline_residue = "residue_";
340         @baseline_residue = split ("", $baseline_residue);
341         push (@baseline_residue, $Num_res);
342         @Temporal_2 = @baseline_residue;
343         unshift (@Temporal_2, "@");
344         push (@baseline_molecule, @Temporal_2);
345         push (@baseline_molecule, ",");
346         unshift (@baseline_residue, "@");
347         push (@baseline_residue, "= (" , @Temporal_1);
348         push (@baseline_residue, ",");
349         $baseline_PM_residue = "PM_residue_";
350         @baseline_PM_residue = split ("", $baseline_PM_residue);
351         push (@baseline_PM_residue, $Num_res);
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
352         @Temporal_22 = @baseline_PM_residue;
353         unshift (@Temporal_22, "\\@");
354         push (@baseline_PM_molecule, @Temporal_22);
355         push (@baseline_PM_molecule, ",");
356         unshift (@baseline_PM_residue, "@");
357         push (@baseline_PM_residue, "= (", $PM_atomo);
358         push (@baseline_PM_residue, ",");
359         $baseline_VM_residue = "VM_residue_";
360         @baseline_VM_residue = split ("", $baseline_VM_residue);
361         push (@baseline_VM_residue, $Num_res);
362         @Temporal_22 = @baseline_VM_residue;
363         unshift (@Temporal_22, '\\@');
364         push (@baseline_VM_molecule, @Temporal_22);
365         push (@baseline_VM_molecule, ",");
366         unshift (@baseline_VM_residue, '@');
367         push (@baseline_VM_residue, '= ');
368         $n=14;
369         do {
370             $n++;
371             } until ($linea_larga_atomo[$n] =~ '\w');
372         do {
373             push (@baseline_VM_residue, "$linea_larga_atomo[$n]");
374             $n++;
375             } until ($linea_larga_atomo[$n] =~ '\W');
376         push (@baseline_VM_residue, "");
377         $Comparar_num_res = $Num_res;
378         $count_atm_res_[$Num_res]++;
379     }
380     elsif ($Num_res == $Comparar_num_res)
381     {
382         $Temporal_3 = join ("", @Temporal_1);
383         push (@baseline_residue, "$Temporal_3");
384         push (@baseline_residue, ",");
385         push (@baseline_PM_residue, "$PM_atomo");
386         push (@baseline_PM_residue, ",");
387         $count_atm_res_[$Num_res]++;
388     } else
389     {
390         pop @baseline_residue;
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
391     push (@baseline_residue, "");\n");
392     $Temporal_3 = join ("",@baseline_residue);
393     print ARCHSAL "$Temporal_3";
394     pop @baseline_PM_residue;
395     push (@baseline_PM_residue, "");\n");
396     $Temporal_33 = join ("",@baseline_PM_residue);
397     print ARCHSAL "$Temporal_33";
398     push (@baseline_VM_residue, "");\n");
399     $Temporal_33 = join ("",@baseline_VM_residue);
400     print ARCHSAL "$Temporal_33";
401     $baseline_residue = "residue_";
402     @baseline_residue = split ("", $baseline_residue);
403     unshift (@baseline_residue, "@");
404     push (@baseline_residue, $Num_res);
405     @Temporal_2 = @baseline_residue;
406     unshift (@Temporal_2, "\\");
407     push (@baseline_molecule, @Temporal_2);
408     push (@baseline_molecule, ",");
409     push (@baseline_residue, " (",@Temporal_1);
410     push (@baseline_residue, ",");
411     $baseline_PM_residue = "PM_residue_";
412     @baseline_PM_residue = split ("", $baseline_PM_residue);
413     push (@baseline_PM_residue, $Num_res);
414     @Temporal_22 = @baseline_PM_residue;
415     unshift (@Temporal_22, "\\@");
416     push (@baseline_PM_molecule, @Temporal_22);
417     push (@baseline_PM_molecule, ",");
418     unshift (@baseline_PM_residue, "@");
419     push (@baseline_PM_residue, " (",$PM_atomo);
420     push (@baseline_PM_residue, ",");
421     $baseline_VM_residue = "VM_residue_";
422     @baseline_VM_residue = split ("", $baseline_VM_residue);
423     push (@baseline_VM_residue, $Num_res);
424     @Temporal_22 = @baseline_VM_residue;
425     unshift (@Temporal_22, "\\@");
426     push (@baseline_VM_molecule, @Temporal_22);
427     push (@baseline_VM_molecule, ",");
428     unshift (@baseline_VM_residue, '@');
429     push (@baseline_VM_residue, '=');
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
430     $n=14;
431     do {
432         $n=++$n;
433         } until ($linea_larga_atomo[$n] =~ '\w');
434     do {
435         push (@baseline_VM_residue, "$linea_larga_atomo[$n]");
436         $n=++$n;
437         } until ($linea_larga_atomo[$n] =~ '\W');
438     push (@baseline_VM_residue, "");
439     $count_atm_res_[$Num_res]++;
440     $Comparar_num_res = $Num_res;
441 }
442 push (@baseline_coordinate, " = (");
443 $n=29;
444 do {
445     $n=++$n;
446     } until ($linea_larga_atomo[$n] =~ '\S');
447 do {
448     push (@baseline_coordinate, "$linea_larga_atomo[$n]");
449     $n=++$n;
450     } until ($linea_larga_atomo[$n] =~ '\s');
451 push (@baseline_coordinate, ", ");
452 $n=38;
453 do {
454     $n=++$n;
455     } until ($linea_larga_atomo[$n] =~ '\S');
456 do {
457     push (@baseline_coordinate, "$linea_larga_atomo[$n]");
458     $n=++$n;
459     } until ($linea_larga_atomo[$n] =~ '\s');
460 push (@baseline_coordinate, ", ");
461 $n=46;
462 do {
463     $n=++$n;
464     } until ($linea_larga_atomo[$n] =~ '\S');
465 do {
466     push (@baseline_coordinate, "$linea_larga_atomo[$n]");
467     $n=++$n;
468     } until ($linea_larga_atomo[$n] =~ '\s');
```

Supplemental F5. Frame' extraction. Perl code to calculate the distance, normalized distance, and the inverse value of the normalized distance between each pair of nodes, for each frame.

```
469         push (@baseline_coordinate, "");
470         $nombre_corto = join ("",@baseline_coordinate);
471         print ARCHSAL (" $nombre_corto \n");
472     }
473 }
474 close ARCHSAL;
475 close ARCH;
476 @filesModel = glob "Model*.pl";
477 foreach $a (@filesModel){
478     system ($a);
479 }
480 $base2_CMmolecule = "molecule = (";
481 @base2_CMmolecule = split ("", $base2_CMmolecule);
482 unshift (@base2_CMmolecule, "@");
483 $base2_VMmolecule = "VM_molecule = (";
484 @base2_VMmolecule = split ("", $base2_VMmolecule);
485 unshift (@base2_VMmolecule, "@");
486 $n = 1;
487 do {
488     $base2_CMresidue = "CM_residue_$n,";
489     @base2_CMresidue = split ("", $base2_CMresidue);
490     unshift (@base2_CMresidue, '@');
491     push @base2_CMmolecule, @base2_CMresidue;
492
493     $base2_VMresidue = "VM_residue_$n,";
494     @base2_VMresidue = split ("", $base2_VMresidue);
495     unshift (@base2_VMresidue, '@');
496     push @base2_VMmolecule, @base2_VMresidue;
497     $n++;
498 } until ($n>65);
499 pop @base2_CMmolecule;
500 push(@base2_CMmolecule, "");\n";
501 $Temporal_55 = join(" ",@base2_CMmolecule);
502 pop @base2_VMmolecule;
503 push(@base2_VMmolecule, "");\n";
504 $Temporal_66 = join(" ",@base2_VMmolecule);
505 @filesCM = glob "center*.pl";
506 foreach $b (@filesCM){
507     open (ARCHSAL, '>>', $b);
```




Stabilizing an amyloidogenic λ 6 light chain variable domain

Journal:	<i>The FEBS Journal</i>
Manuscript ID	FJ-16-1024
Manuscript Type:	Regular Paper
Date Submitted by the Author:	29-Nov-2016
Complete List of Authors:	Luna-Martínez, Oscar; Universidad Nacional Autonoma de Mexico, Molecular Medicine & Bioprocesses, Institute of Biotechnology Hernández-Santoyo, Alejandra; Universidad Nacional Autonoma de Mexico, Chemistry of Biomacromolecules, Institute of Chemistry Villalba-Velázquez, Miryam; Universidad Nacional Autonoma de Mexico, Molecular Medicine & Bioprocesses, Institute of Biotechnology Sánchez-Alcalá, Rosalba; Universidad Nacional Autonoma de Mexico, Molecular Medicine & Bioprocesses, Biotechnology Institute Fernandez-Velasco, Alejandro; Universidad Nacional Autonoma de Mexico, Biochemistry, Faculty of Medicine Becerril, Baltazar; Universidad Nacional Autonoma de Mexico, Molecular Medicine & Bioprocesses, Institute of Biotechnology
Key Words:	