

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

**UTILIZACIÓN DE LA BPTI (BOVINE PANCREATIC TRYPSIN INHIBITOR) COMO
MODELO PARA EL DESARROLLO DE UN MÉTODO PARA LA
CARACTERIZACIÓN DEL ENTORNO ENERGÉTICO DE PROTEÍNAS A 300 K**

HERNÁN ANDRÉS MORALES NAVARRETE
andresmoralesn2@hotmail.com

**PROYECTO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO DE
FÍSICO**

Director: Marco Bayas Rea PhD
marco.bayas@epn.edu.ec

Quito, Marzo 2010

DECLARACIÓN

Yo, Hernán Andrés Morales Navarrete, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentada para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

La Escuela Politécnica Nacional puede hacer uso de los derechos correspondientes a este trabajo, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.

HERNÁN ANDRÉS MORALES NAVARRETE

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado Hernán Andrés Morales Navarrete, bajo mi supervisión.

Marco Bayas Rea , Ph. D.
DIRECTOR

AGRADECIMIENTOS

“Tras un paso otro. Pequeñas distancias que no tienen destino, solo secuencia. Caminaba... De pronto, la distancia crecía. Cada paso se hacía más y más largo. De caminar a volar, pensé. Pero no. No se trataba de volar. Era algo más precioso y sutil: navegar a flor de piso. Derrotar la realidad. Despertar soñando.”

Una vida sin metas no tiene sentido, pero vivir únicamente por alcanzarlas es como morir a pasos. Aprovechar cada coyuntura en la vida y hacer que todo nuestro esfuerzo se vuelque al bien de los demás es lo más importante. La culminación de esta fase de mi vida no es más que eso, el fin de una etapa llena de alegrías y tristezas, de triunfos y derrotas, pero sobretodo de mucho esfuerzo y apoyo de la gente que de una u otra forma compartió este tramo del camino de la vida conmigo. Dios y la vida me dieron sueños para crecer, sabiduría para decidir, fuerza para luchar, alegría para disfrutar y un gran número de valiosísimas personas para compartirlo. La lucha fue compartida y ahora el triunfo también lo es. La victoria y la alegría es de mis padres, quienes con su apoyo incondicional hicieron de esta utopía una realidad; de mi hermano, compañero incondicional de luchas diarias; de mi familia, pilar de mi vida y fuente inagotable de fuerzas y consejos; de mis amig@s, apoyo en los momentos difíciles y compañía en los alegres; del doctor Marco Bayas, participante y animador constante de mi proyecto de titulación; en fin de todas las personas que tal vez con su simple presencia me animaron a seguir adelante.

Hernán Andrés

DEDICATORIA

“Busco un cielo en esta tierra donde poder sembrar mis ganas. Busco un eterno instante dulce, cálido y soleado para cantar mis versos tontos. Busco horizontes nuevos, busco justicias y alegrías; letargos conscientes, sueños nuevos. Busco vida en mis recuerdos. Mi ser en palabras. Busco la muerte como principio. Las aguas de un mar bravío. Busco fronteras para desterrarlas. Busco la burla ingenua del desquiciado. Busco y rebusco en todo este aire. Busco el futuro, busco el presente; lo que seré desde hoy.”

A Dios, fuente y motor de vida

A mis padres, Hernán y María, apoyo incondicional en la lucha diaria

A mi hermano, Diego, compañero en el camino de la vida

A la memoria de quienes siempre estuvieron y siempre permanecerán mi lado :

Pedro Fabián y Ana Beatríz.

Hernán Andrés

ÍNDICE DE CONTENIDO

ÍNDICE DE FIGURAS	I
ÍNDICE DE TABLAS	VI
Resumen	VIII
Abstract	IX
1. Introducción: Contexto Biofísico del problema	1
1.1. Problema del plegamiento proteínico	1
1.1.1. Postulado de Anfinsen	2
1.1.2. Paradoja de Levinthal	2
1.1.3. Intermediarios y plegamiento incorrecto	3
1.2. Modelos para describir el plegamiento proteínico	4
1.3. Entorno Energético (Energy Landscape) de proteínas	5
1.4. La BPTI, Bovine Pancreatic Trypsin Inhibitor, como modelo de estudio de problemas relacionados con el plegamiento proteínico	7
1.4.1. Complejo Tripsina-BPTI	8
1.4.2. Estructura de la BPTI	10
1.4.3. Estudios sobre el mecanismo de plegamiento de la BPTI	12
2. Métodos Computacionales	15
2.1. Dinámica Molecular	15
2.1.1. NAMD: NANoscale Molecular Dynamics	16
2.1.2. Interacciones Moleculares: Funciones de Campo de Fuerza	16

2.1.3.	Modelos de solvente	18
2.1.4.	Condiciones de Frontera	19
2.1.5.	Truncación de interacciones	20
2.1.6.	Algoritmos de integración	22
2.1.7.	Ensamblajes en la Dinámica Molecular	24
2.1.7.1.	Ensamble NVT	24
2.1.8.	Protocolo de simulación	25
2.2.	Análisis Conformacional	27
2.2.1.	Muestreo	28
2.2.1.1.	Dinámica Molecular a altas temperaturas	28
2.2.1.2.	Dinámica Molecular Dirigida	29
2.2.2.	Optimización	31
2.2.2.1.	Minimización	31
2.2.3.	Análisis	32
2.2.3.1.	Medición de disimilitudes	32
2.2.3.2.	Análisis de grupos (Cluster Analysis)	32
2.2.3.3.	Escalamiento Multidimensional	35
2.2.3.3.1.	Análisis de Componentes Principales (ACP)	36
2.2.3.3.2.	Análisis de Coordenada Principal (ACooP)	37

3. Conformaciones espaciales de la BPTI 38

3.1.	Simulaciones de Dinámica Molecular de la BPTI	38
3.1.1.	Construcción de los Sistemas	40
3.1.1.1.	Estructuras	40
3.1.1.2.	Hidratación y neutralización	41
3.1.2.	Minimización	43
3.1.3.	Calentamiento hasta 300 K	44
3.1.4.	Equilibración	46
3.1.5.	Producción: Muestreo y optimización conformacional	47
3.1.5.1.	Estructuras asociadas al estado nativo	48
3.1.5.2.	Estructuras no nativas	49
3.1.5.3.	Estructuras asociadas a transiciones desde los mínimos locales al mínimo global	55
3.2.	Conformaciones asociadas al estado nativo	56
3.2.1.	BPTI natural	57

3.2.2. BPTI mutada	60
3.3. Conformaciones no nativas	63
3.3.1. BPTI natural	63
3.3.2. BPTI mutada	71
4. Entorno Energético de la BPTI	78
4.1. Análisis de Coordenada Principal	78
4.1.1. Estructuras de la BPTI natural	79
4.1.2. Estructuras de la BPTI mutada	84
4.2. Construcción del Entorno Energético	90
4.2.1. Entorno Energético de la BPTI natural	90
4.2.2. Entorno Energético de la BPTI mutada	91
4.2.3. Comparación de los Entorno Energético inherentes a la BPTI natural y mutada	92
5. Conclusiones	95
Anexos	102
Referencias	115

ÍNDICE DE FIGURAS

1.1. Representación tridimensional de Entornos Energéticos	6
1.2. Representación esquemática en bidimensional de un Entorno Energético en forma de embudo rugoso	7
1.3. Complejo Tripsina-BPTI	9
1.4. Zona de contacto BPTI-Tripsina	9
1.5. Estructura primaria, secundaria y terciaria de la BPTI	10
1.6. Disposición espacial de los aminoácidos de la BPTI	11
1.7. Enlaces disulfuro presentes en la BPTI	12
1.8. Esquema del plegamiento de la BPTI según Weissman y Finn	13
2.1. Representación esquemática de las cinco contribuciones al campo de fuerza usada en NAMD	18
2.2. Condiciones Periódicas de frontera	19
2.3. Truncación de interacciones	21
2.4. Diagrama de flujo del protocolo de simulación de Dinámica Molecular	27
2.5. Dinámica Molecular a altas temperaturas	29
2.6. Optimización conformacional: minimización de energía	31
2.7. Métodos de análisis de grupos	34
3.1. Diagrama de flujo del tratamiento realizado a partir de la estructura cristalina de la BPTI	39
3.2. Estructura primaria y espacial de la BPTI natural y mutada	41
3.3. Sistema molecular de la proteína natural	42
3.4. Gráfico de distancias rms entre las estructuras minimizadas de las proteínas natural y mutante	44
3.5. Minimización y calentamiento hasta 300 K de los sistemas moleculares natural y mutado	45
3.6. Gráfico de distancias rms entre las estructuras calentadas a 300 K de las proteínas natural y mutada	45

3.7. Equilibración a 300 K de los sistemas moleculares natural y mutado	47
3.8. Muestreo y optimización de estructuras asociadas al estado nativo	49
3.9. Calentamiento hasta 1000 K	50
3.10. Equilibración a 1000K	51
3.11. Producción y optimización de las estructuras de la muestra 1 de los sistemas moleculares natural y mutado	54
3.12. Simulaciones dirigidas	56
3.13. Representación gráfica de la matriz de disimilitud de las estructuras asociadas al estado nativo de la BPTI natural	57
3.14. Dendograma del análisis de grupos de las estructuras asociadas al estado nativo de la BPTI natural	58
3.15. Gráfico de distancias rms entre las 7 conformaciones y la estructura calentada a 300K minimizada de la proteína natural	60
3.16. Representación gráfica de la matriz de disimilitud de las estructuras asociadas al estado nativo de la BPTI mutada	61
3.17. Dendograma del análisis de grupos de las estructuras asociadas al estado nativo de la BPTI mutada	61
3.18. Gráfico de distancias rms entre las 9 conformaciones y la estructura calentada a 300K minimizada de la proteína mutada	62
3.19. Estructura de la primera conformación de la muestra 1 de la proteína natural	64
3.20. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 1 y la estructura calentada a 300K minimizada de la proteína natural	65
3.21. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 2 y la estructura calentada a 300K minimizada de la proteína natural	66
3.22. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 3 y la estructura calentada a 300K minimizada de la proteína natural	66
3.23. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 4 y la estructura calentada a 300K minimizada de la proteína natural	67
3.24. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 5 y la estructura calentada a 300K minimizada de la proteína natural	67

3.25. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 6 y la estructura calentada a 300K minimizada de la proteína natural	68
3.26. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 7 y la estructura calentada a 300K minimizada de la proteína natural	68
3.27. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 8 y la estructura calentada a 300K minimizada de la proteína natural	69
3.28. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 9 y la estructura calentada a 300K minimizada de la proteína natural	69
3.29. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 10 y la estructura calentada a 300K minimizada de la proteína natural	70
3.30. Estructura de la primera conformación de la muestra 1 de la proteína mutada	71
3.31. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 1 y la estructura calentada a 300K minimizada de la proteína mutada	72
3.32. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 2 y la estructura calentada a 300K minimizada de la proteína mutada	73
3.33. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 3 y la estructura calentada a 300K minimizada de la proteína mutada	73
3.34. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 4 y la estructura calentada a 300K minimizada de la proteína mutada	74
3.35. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 5 y la estructura calentada a 300K minimizada de la proteína mutada	74
3.36. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 6 y la estructura calentada a 300K minimizada de la proteína mutada	75

3.37. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 7 y la estructura calentada a 300K minimizada de la proteína mutada	75
3.38. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 8 y la estructura calentada a 300K minimizada de la proteína mutada	76
3.39. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 9 y la estructura calentada a 300K minimizada de la proteína mutada	76
3.40. Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 10 y la estructura calentada a 300K minimizada de la proteína mutada	77
4.1. Representación esquemática del mínimo global y los mínimos locales asociados	79
4.2. Porcentaje explicado de la proyección de la estructuras naturales en función del índice del valor propio asociado	80
4.3. Energía de las estructuras de la proteína natural en función de las dos primeras coordenadas principales	81
4.4. Energía de las estructuras de las simulaciones dirigidas de la proteína natural en función de las dos primeras coordenadas principales	82
4.5. Energía de las estructuras de las conformaciones de la proteína natural en función de las dos primeras coordenadas principales	83
4.6. Porcentaje explicado de la proyección de la estructuras mutadas en función del índice del valor propio asociado	85
4.7. Energía de las estructuras de la proteína mutada en función de las dos primeras coordenadas principales	87
4.8. Energía de las estructuras de las simulaciones dirigidas de la proteína mutada en función de las dos primeras coordenadas principales	88
4.9. Energía de las estructuras de las conformaciones de la proteína mutada en función de las dos primeras coordenadas principales	89
4.10. Entorno Energético de la BPTI natural	93
4.11. Entorno Energético de la BPTI mutada	94

5.1. Estructuras de las primeras conformación de las muestras 1 de las proteína natural y mutada	97
5.2. Representación tridimensional de los Entornos Energéticos de la BPTI natural y mutante	98
5.3. Proyección de los Entornos Energéticos de la BPTI natural y mutada .	99
5.4. Estructura de las conformaciones naturales con menor energía de la proteína natural y mutada	100

ÍNDICE DE TABLAS

1.1. Número de aminoácidos presentes en la BPTI según su carga	11
3.1. Cargas netas de los sistemas moleculares de las estructuras natural y mutada	42
3.2. Características generales de los sistemas natural y mutado	42
3.3. Producción y optimización de las estructuras de las muestras no nativas	53
3.4. Análisis jerárquico de las estructuras asociadas al estado nativo de la BPTI natural	60
3.5. Análisis jerárquico de las estructuras asociadas al estado nativo de la BPTI mutada	63
3.6. Análisis jerárquico de la muestra 1 de estructuras no nativas de la BPTI natural	65
3.7. Análisis jerárquico de la muestra 2 de estructuras no nativas de la BPTI natural	66
3.8. Análisis jerárquico de la muestra 3 de estructuras no nativas de la BPTI natural	66
3.9. Análisis jerárquico de la muestra 4 de estructuras no nativas de la BPTI natural	67
3.10. Análisis jerárquico de la muestra 5 de estructuras no nativas de la BPTI natural	67
3.11. Análisis jerárquico de la muestra 6 de estructuras no nativas de la BPTI natural	68
3.12. Análisis jerárquico de la muestra 7 de estructuras no nativas de la BPTI natural	68
3.13. Análisis jerárquico de la muestra 8 de estructuras no nativas de la BPTI natural	69
3.14. Análisis jerárquico de la muestra 9 de estructuras no nativas de la BPTI natural	69

3.15. Análisis jerárquico de la muestra 10 de estructuras no nativas de la BPTI natural	70
3.16. Lazos móviles que caracterizan a las conformaciones no nativas de la BPTI natural	70
3.17. Análisis jerárquico de la muestra 1 de estructuras no nativas de la BPTI mutada	72
3.18. Análisis jerárquico de la muestra 2 de estructuras no nativas de la BPTI mutada	73
3.19. Análisis jerárquico de la muestra 3 de estructuras no nativas de la BPTI mutada	73
3.20. Análisis jerárquico de la muestra 4 de estructuras no nativas de la BPTI mutada	74
3.21. Análisis jerárquico de la muestra 5 de estructuras no nativas de la BPTI mutada	74
3.22. Análisis jerárquico de la muestra 6 de estructuras no nativas de la BPTI mutada	75
3.23. Análisis jerárquico de la muestra 7 de estructuras no nativas de la BPTI mutada	75
3.24. Análisis jerárquico de la muestra 8 de estructuras no nativas de la BPTI mutada	76
3.25. Análisis jerárquico de la muestra 9 de estructuras no nativas de la BPTI mutada	76
3.26. Análisis jerárquico de la muestra 10 de estructuras no nativas de la BPTI mutada	77
3.27. Lazos móviles que caracterizan a las conformaciones no nativas de la BPTI mutada	77
4.1. Altura de la barreras energéticas asociadas a las simulaciones dirigidas	86
5.1. Conformaciones accesibles a la BPTI natural y mutada	96
5.2. Características de los Entornos Energéticos de la BPTI natural y mutada	99

Resumen

El Inhibidor de Tripsina Pancreática Bovina (Bovine Pancreatic Trypsin Inhibitor, BPTI) y una mutación puntual de éste que carece de enlaces disulfuro fueron estudiados utilizando simulaciones de Dinámica Molecular. Las estructuras obtenidas de las simulaciones fueron utilizadas para describir las características principales de las conformaciones accesibles a estas proteínas así como sus correspondientes Entornos Energéticos. Utilizando Análisis de Grupos (Cluster Analysis), se examinaron 11 muestras de estructuras de la BPTI natural hidratada y 11 muestras de la mutada con el fin de determinar conformaciones accesibles a estos polipéptidos. Una de las muestras corresponde a estructuras asociadas al estado nativo, mientras que las otras corresponden a estructuras no nativas. Cada muestra se obtuvo trayectorias de Dinámica Molecular de 1 ns a 300 K. Las estructuras fueron analizadas utilizando distancia rms de los carbonos alfa entre ellas, la cual nos brinda una medida global de los cambios conformacionales. La proteína mutada presenta un mayor número de regiones dinámicas específicas. Así, se encontraron nueve conformaciones asociadas al estado nativo de la mutada separadas por una distancia rms de 0.73 Å, mientras que para la proteína natural se encontraron únicamente 7 con una distancia rms mutua de 0.6 Å. Adicionalmente, se encontraron 88 conformaciones no nativas con una distancia rms entre 0.92 y 1.07 Å para la BPTI natural y 90 conformaciones con una distancia rms entre 1.01 y 1.53 Å para la mutada. Utilizando Análisis de Coordenada Principal (ACoP) se construyó una representación de los Entornos Energéticos de estas proteínas. Se encontró que Entornos sustancialmente diferentes caracterizan a cada proteína. El Entorno de la proteína natural presenta un mínimo global pronunciado muy estrecho y dos mínimos locales profundos separados del mínimo global por barreras energéticas del orden los 4.8 aJ (1158.87 kT). Las barreras entre mínimos locales se encuentran alrededor de los 1.6 aJ (386.29 kT). En el caso del Entorno de la proteína mutada, existe un mínimo global más ancho que la BPTI natural y cinco mínimos locales que se encuentran separados del global por barreras entre 2.4 aJ (579.44 kT) and 4.48 aJ (1081.62 kT), y separados entre ellos por barreras más pequeñas del orden de los 1.28 aJ (309.03 kT). Los resultados de este trabajo muestran la importancia de los enlaces disulfuro en la determinación de las conformaciones accesibles a la BPTI y consecuentemente en su función biológica.

Abstract

The protein Bovine Pancreatic Trypsin Inhibitor (BPTI) and a point mutant lacking the disulfide bonds were studied using Molecular Dynamics simulations. The resultant structures were used to describe the main characteristics of the accessible conformations of the proteins as well as the corresponding Energy Landscapes. Using cluster analysis, 11 samples of structures of natural solvated BPTI and 11 samples of mutated solvated BPTI were analyzed to determine accessible conformations of these polypeptides. One of them corresponds to structures associated with the native state and the other ones correspond to non-native conformations. Each sample was obtained from Molecular Dynamics trajectories of 1 ns at 300 K. The structures were analyzed using their pair wise alpha carbons rms distance which gave a global measure of conformational changes. The mutated protein showed the highest number of specific dynamic regions. In fact, nine conformations separated by 0.73 Å rms distance were found for the native state of the mutated protein whereas only seven separated by 0.6 Å rms distance were found for the natural one. Additionally, there were also found 88 non-native conformations with rms distances between 0.92 and 1.07 Å for natural BPTI and 90 conformations with rms distances between 1.01 and 1.53 Å for mutated protein. Using Principal Coordinates Analysis a representation of the Energy Landscapes of these proteins were constructed. It was found that substantially different Landscapes characterize each protein. The natural protein Landscape has a very sharp narrow basin which corresponds to the global minimum and two deep local minima separated from the global minimum by energy barriers around 4.8 aJ (1158.87 kT). The barriers between themselves are around 1.6 aJ (386.29 kT). In the case of the mutant protein Landscape, there is a broader basin than for the natural one and five local minima which are separated from the global minimum by barriers between 2.4 aJ (579.44 kT) and 4.48 aJ (1081.62 kT), and separated between themselves by smaller barriers around 1.28 aJ (309.03 kT). Overall, the results of this report showed the importance of the disulfide bonds in determining the accessible conformations of BPTI and consequently its biological function.

Capítulo 1

Introducción: Contexto Biofísico del problema

1.1. Problema del plegamiento proteínico

Un organismo vivo posee alrededor de 100000 tipos diferentes de proteínas, cada una de las cuales realiza una función biológica específica [1]. Las proteínas se sintetizan dentro de la célula mediante la acción de Ribosomas en base a la información codificada en el ARN, proceso que genera una cadena lineal de aminoácidos propia y única para cada una [2]. Luego de ser sintetizada, una cadena polipeptídica transita por una serie de estados en los que su conformación espacial se modifica hasta alcanzar el estado termodinámicamente estable, con una estructura tridimensional particular, conocida como su estado nativo; el cual determina la funcionalidad biológica de la proteína resultante [3, 4]. El discernimiento de las reglas que determinan cómo una cadena de aminoácidos (estructura primaria), adquiere su estructura tridimensional funcional (estado nativo), se conoce como el problema del plegamiento proteínico y se ha convertido en uno de los mayores retos de la ciencia [2, 3, 5, 6, 7].

Una conformación proteínica es una forma tridimensional de una molécula asociada a un arreglo espacial particular de los diferentes elementos que la conforman [8]. El estado nativo de una proteína corresponde a la conformación con mínima energía libre y se caracteriza por tener una baja cantidad de entropía. El estado desnaturalizado por el contrario, se caracteriza por poseer una alta entropía, ya

que sus interacciones nativas se perdieron y los aminoácidos pueden formar varios arreglos; es complicado definir las características del estado desnaturalizado ya que éste depende de las condiciones de desplegamiento (pH, temperatura, adición de sustancias químicas, etc.) [2, 9].

El problema del plegamiento proteínico ha sido abordado desde muchas perspectivas, se ha realizado un sinnúmero de estudios experimentales y teóricos, sin embargo no se ha logrado conceptualizar en forma clara este fenómeno [2, 5]. La determinación experimental de la transición entre los estados desnaturalizados y nativo de una proteína es muy complicada, debido al gran número de conformaciones que puede adoptar la cadena polipeptídica y a la inestabilidad relativa de los estados intermedios; por ello la utilización de métodos computacionales, como las simulaciones de Dinámica Molecular, es una herramienta clave en investigaciones relacionadas con este problema [2, 7, 10].

1.1.1. Postulado de Anfinsen

Basado en estudios experimentales del proceso de renaturalización a partir de la Ribonucleasa totalmente desnaturalizada [11], Anfinsen postuló lo que se conoce con el nombre de hipótesis termodinámica; que establece que la conformación nativa de una proteína en un ambiente fisiológicamente adecuado (pH, temperatura, solvente, etc.) corresponde al estado termodinámicamente estable y representa el mínimo global de la energía libre de Gibbs; además toda la información necesaria para obtenerla está contenida en la secuencia lineal de aminoácidos [11]. En condiciones adecuadas, la proteína busca espontáneamente la conformación que optimice sus interacciones tanto internas como con el solvente que la rodea [9]. Este control termodinámico del plegado implica que la búsqueda de la conformación nativa debe darse entre una gran cantidad de conformaciones [12].

1.1.2. Paradoja de Levinthal

Si se considera que todas las conformaciones de una proteína son igual de probables, el plegamiento se transformaría en una búsqueda aleatoria del estado nativo, de modo que, el tiempo necesario para encontrar la conformación nativa está dado por el producto del número de conformaciones posibles de la cadena y el tiempo requerido para encontrar una conformación. Para una proteína de 100 aminoácidos, considerando tres orientaciones por residuo existen $3^{100} = 510^{47}$

conformaciones, y si el tiempo para encontrar cada una de éstas está en el orden de 10^{-11} segundos, tenemos que el tiempo necesario para el plegamiento se encontraría alrededor de 10^{52} años. Sin embargo sabemos que en el contexto biológico las proteínas se plegan en cuestión de segundos o menos, esto se conoce como la paradoja de Levinthal [13, 14]. Esta paradoja plantea que el proceso del plegamiento proteínico no es un problema de búsqueda aleatoria de conformaciones, sino que involucra una serie de pasos bien definidos; la solución planteada por Levinthal es que existe un camino específico, determinado por un control cinético, para llegar al estado nativo [15], es decir, existe un número determinado de conformaciones accesibles (cinéticamente cercanas) a la proteína por los cuales ésta debe transitar antes de llegar a su estado nativo.

1.1.3. Intermediarios y plegamiento incorrecto

En principio se consideraba que el proceso de plegamiento consistía únicamente en la transición entre el estado desplegado y el plegado, sin embargo, en base a estudios cinéticos realizados con el fin de explicar la paradoja de Levinthal se ha demostrado la existencia de estados intermediarios en el plegamiento, evidencia que sugiere que el plegamiento sigue un modelo multiestado, es decir, la proteína parte del estado desnaturalizado a los estados intermediarios y finalmente al estado nativo [6]. Un intermediario general en el plegamiento de proteínas es el denominado Glóbulo Fundido, que se caracteriza por tener gran parte de la estructura secundaria y un bajo porcentaje de la terciaria, con respecto a la estructura nativa. No se puede definir a este estado como una única conformación, sino más bien por un conjunto de conformaciones que se convierten entre ellas de forma muy rápida [16].

No todos los intermediarios se hallan en el camino hacia el estado nativo, debido a las diferencias en estabilidad que poseen. Los que se encuentran en el camino hacia la estructura nativa se denominan “on pathway” y nos proveen de información acerca del plegado, mientras que los otros se denominan “off pathway” y nos dan información acerca de los fenómenos del plegamiento incorrecto, que son de gran trascendencia para la biomedicina [7, 16]. Existe una amplia gama de enfermedades relacionadas con el plegamiento incorrecto de proteínas. Están por ejemplo las disfunciones celulares conocidas con el nombre de amiloidosis,

en las que los plegamientos incorrectos conducen a la formación irreversible de agregados fibrilares insolubles llamados amiloides, los cuales interfieren con el desarrollo de diversas funciones celulares. Dentro de este grupo de patologías se encuentran enfermedades como Alzheimer, Parkinson y Creutz-Jakob [2]. También se puede mencionar que según estudios recientes, ciertas proteínas pueden actuar como agentes infecciosos cuando están plegadas incorrectamente, pues además de no realizar su función biológica, tienen la capacidad de inducir cambios en la estructura terciaria de sus vecinas. La encefalopatía espongiiforme bovina o enfermedad de las vacas locas es una enfermedad producida por este tipo de proteínas, que se denominan priones [2]. De lo mencionado anteriormente, puede concluirse que los estudios asociados al problema del plegamiento proteínico no tiene relevancia solo desde el punto de generación de ciencias básica, sino que sus resultados pueden ser utilizados para solucionar problemas biomédicos asociados.

1.2. Modelos para describir el plegamiento proteínico

La paradoja de Levinthal sugiere que una búsqueda aleatoria del estado nativo no es factible, por lo que debe existir un camino específico de conformaciones accesibles en el proceso del plegado. Se han propuesto varios modelos para explicar la ruta del plegado. Entre los modelos clásicos más importantes tenemos:

- a. Modelo de Nucleación –Propagación (1961). Un grupo de aminoácidos cercanos forman uno o más grupos estructurales, alrededor de los cuales continúa la formación de la estructura, es decir, el proceso se inicia con la formación de una estructura secundaria que se propaga gradualmente a lo largo de toda la cadena, finalmente como consecuencia de la formación de las estructuras secundarias se da la estructura terciaria de la proteína [7].
- b. Modelo de Difusión - Colisión (1976). Este modelo considera que los elementos de las estructuras secundarias se forman rápida y simultáneamente en varias partes de la cadena, luego éstos se propagan hasta colapsar y formar la estructura terciaria de la proteína [7].

- c. Modelo del Colapso Hidrofóbico (1985). Se considera al efecto hidrofóbico como la fuerza motora y estabilizadora de la proteína. Según este modelo el primer evento de plegamiento proteínico es el colapso de la cadena polipeptídica guiado por las interacciones hidrofóbicas, luego del cual se da la formación de las estructuras secundarias [7].

- d. Modelo de Nucleación-Condensación (1995). Los dos primeros modelos se basan en la formación inicial de las estructuras secundarias, lo que provoca el arreglo tridimensional (estructura ternaria); el tercero por el contrario, sugiere que la estructura secundaria se da como consecuencia de un arreglo espacial de los residuos hidrofóbicos. Un modelo propuesto que conjuga estas dos visiones clásicas es el de Nucleación-Condensación; el que propone la formación de una estructura pre-nativa compuesta por aminoácidos cercanos, estabilizada mediante interacciones no locales. Este núcleo no se encuentra presente en el estado desnaturalizado y no necesariamente se encuentra formado en su totalidad en el estado nativo [7].

La diversidad de forma y tamaño de las proteínas hace que los mecanismos de plegado de ellas también sean variados, así muchas proteínas pequeñas parecen plegarse siguiendo el modelo de Nucleación-Condensación, sin embargo en sistemas más complejos tanto la estructura secundaria como la ternaria se forman sistemáticamente en la transición por los estados intermedios [7]. El enfoque teórico más general acerca del plegamiento proteínico es el del Entorno Energético (Energy Landscape), que define para cada cadena polipeptídica, una superficie en términos de energía y entropía. Las conformaciones espaciales accesibles a la cadena están representadas en esta superficie, la conformación nativa corresponde al mínimo de la misma [17].

1.3. Entorno Energético (Energy Landscape) de proteínas

La descripción del plegado proteínico se dificulta por el alto grado de complejidad del proceso, sin embargo, gran parte de esta complejidad puede ser descrita mediante un análisis estadístico de las energías conformacionales proteínicas; esto se conoce como la teoría del Entorno Energético (Energy Landscape) [18, 19].

Esta teoría, que muestra la tendencia de la superficie energética hacia el estado nativo y explica satisfactoriamente cuándo y porqué se dan caminos únicos para el plegamiento, es actualmente una de las vías más prometedoras para solucionar de una forma consistente la paradoja de Levinthal [17, 19].

La hipersuperficie definida por la energía de la proteína en función de un arreglo de coordenadas que describen a cada conformación constituye su Entorno Energético (Energy Landscape). Esta superficie detalla las conformaciones espaciales accesibles a la proteína así como sus transiciones a temperatura ambiente [17, 19]. Un cambio conformacional, como el plegado proteínico, está determinado por una o varias trayectorias a lo largo de esta superficie, por lo tanto la concepción de un camino único de plegamiento es reemplazada con la idea de que cada cadena polipeptídica puede llegar al estado nativo por varias vías, mediante un movimiento entre estados intermediarios (que corresponden a mínimos locales) superando barreras energéticas locales (figura 1.1). Dependiendo de las características estadísticas del Entorno Energético, se puede determinar la existencia de los caminos accesibles en el proceso de plegamiento, si el proceso se da entre dos o varios estados, e incluso recabar información acerca del plegamiento incorrecto y la agregación [6, 17].

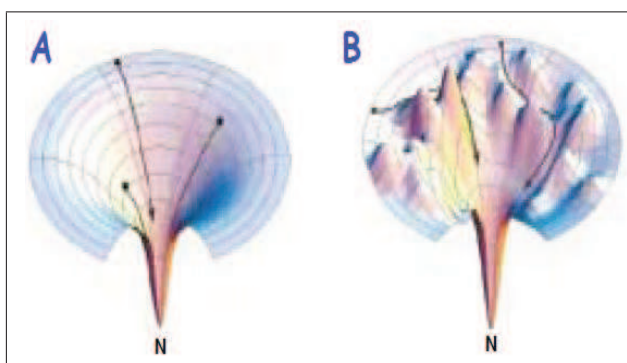


Figura 1.1: Representación de Entornos Energéticos en forma de embudos, adaptada de la referencia [6]. **A:** Entorno Energético ideal, sin rugosidades; el plegamiento es rápido y se da entre dos estados. **B:** Entorno Energético rugoso con presencia de mínimos locales y barreras; este corresponde a un proceso de plegado que sigue el modelo multiestado con varias vías de plegamiento.

Una hipótesis propuesta a partir del concepto del Entorno Energético es que esta hipersuperficie tiene la forma de un embudo multidimensional [19, 20, 21]. La gran diferencia energética entre los estados desnaturalizado y nativo da como resultado una cierta tendencia en el movimiento de la cadena polipeptídica (hacia el estado de menor energía) dentro del espacio conformacional [20]; ésto provoca

que el Entorno Energético subyacente esté representado por una superficie rugosa inclinada hacia un mínimo global (figura 1.2), es decir, todas las estructuras con energía relativamente alta se encuentran guiadas hacia el estado de mínima energía a través de múltiples caminos de intermediarios separados por barreras [19].

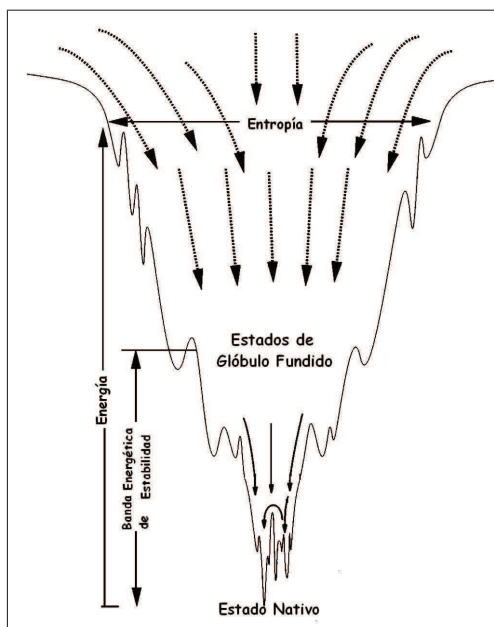


Figura 1.2: Representación esquemática en dos dimensiones de Entornos Energéticos en forma de embudo rugoso, adaptada de la referencia [19]. La altura del embudo representa la energía y el ancho la entropía de la proteína. Todas las conformaciones tienden hacia el estado de mínima energía (nativo). Los estados del Glóbulo fundido representan conformaciones compactas desnaturalizadas con un alto grado de estructuras secundarias presente y están separados del estado nativo por una banda energética que garantiza la estabilidad de este estado.

1.4. La BPTI, Bovine Pancreatic Trypsin Inhibitor, como modelo de estudio de problemas relacionados con el plegamiento proteínico

En 1961 Anfinsen demostró que las interacciones no covalentes (van der Waals, enlaces puente hidrógeno e interacciones iónicas) determinan el proceso del plegamiento de proteínas. Sin embargo, existe un grupo especial de proteínas caracterizadas por tener en su estructura cisteínas formando enlaces disulfuro (enlaces covalentes); estos enlaces covalentes que unen las estructuras secundarias, son muy importantes en el mantenimiento de la estructura terciaria y su formación está íntimamente relacionada con los caminos de plegamiento [22, 23].

Durante el plegamiento proteínico los enlaces disulfuro se rompen y se forman con el fin de acomodarse a las conformaciones más favorables de la proteína [11]. Se han efectuado una serie de estudios de este tipo, siendo los primeros realizados con el Inhibidor de Tripsina Pancreática Bobina (Bovine Pancreatic Trypsin Inhibitor, BPTI) y la Ribonucleasa [23].

La BPTI es una proteína monomérica de 58 aminoácidos, cuya masa molecular es 6512 Daltons [24, 25]. Su función es la inhibición de la acción de las enzimas de la familia Serina Proteasa, las cuales están involucradas en los procesos de digestión y coagulación. Por ello ha sido utilizada en varias aplicaciones farmacológicas relacionadas con enfermedades asociadas a disfunciones de estas enzimas [25, 26]. La relevancia de la BPTI no se da únicamente por su utilidad a nivel biomédico, esta proteína fue el primer modelo utilizado para simulaciones de Dinámica Molecular y ha servido de base para estudios de este tipo, debido a que es una proteína pequeña y altamente estable [27]. Su gran estabilidad se debe a la presencia de tres enlaces disulfuro en su estructura tridimensional; además tiene una inusual resistencia a la digestión enzimática, lo que indica que la molécula es compacta [28, 29]. Estas características, además de su tamaño hacen de la BPTI un sistema modelo para estudios experimentales y teóricos de proteínas [25, 28]: su estructura ha sido investigada mediante cristalografía de rayos X y espectroscopía de resonancia magnética nuclear (NMR) [24, 25].

1.4.1. Complejo Tripsina-BPTI

El cuerpo humano necesita constantemente un suministro de aminoácidos para la nutrición y desarrollo celular. Existen enzimas especializadas encargadas de separar las proteínas en aminoácidos o péptidos de menor tamaño (digestión proteínica). La Tripsina es una de estas y se encuentra en el líquido digestivo que es producido por el páncreas [24, 25].

La Tripsina utiliza principalmente la acción del aminoácido serina para separar los aminoácidos de las proteínas en una posición cercana de los residuos lisina o arginina. Su unión con la proteína se da por medio de la interacción electrostática entre su residuo Asp189 (negativo) y un residuo Lisina o Arginina (positivos) de la proteína y a través de tres aminoácidos (Ser195, His57 y Asp102) forma un sistema de carga que se encarga de romper el enlace peptídico entre la lisina o arginina y el aminoácido contiguo.

La BPTI se enlaza con la Serina de la Tripsina formando un complejo en forma de hongo, la Tripsina representa la cabeza del hongo, y la BPTI su tallo [25]. Como muestra la figura 1.3, la región principal de contacto entre la Tripsina y la BPTI ocurre en la región formada por los residuos Lys15 y Ala16 (región activa); a diferencia de otras proteínas la BPTI tiene una estructura tridimensional en la cual el enlace peptídico entre los residuos Lys15 y Ala16 están lo suficientemente alejados del residuo Ser195 de la Tripsina, de modo que suprime su función y previene así la degradación no controlada de proteínas [24, 25](Figura 1.4).

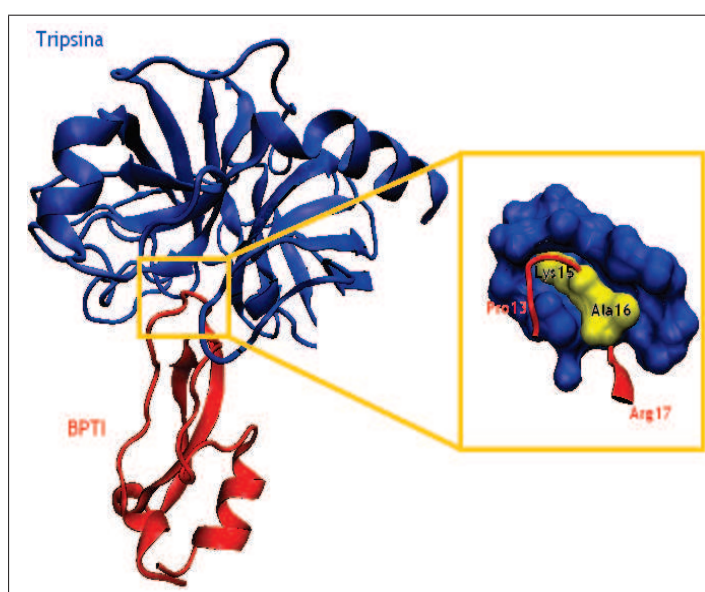


Figura 1.3: Complejo Tripsina-BPTI. En la figura principal muestra el complejo en representación “NewCartoon”. En la ampliación, una representación ”Surf” de la cavidad formada por la Tripsina y la zona activa de la BPTI, donde se aprecia el acoplamiento justo entre ellas.

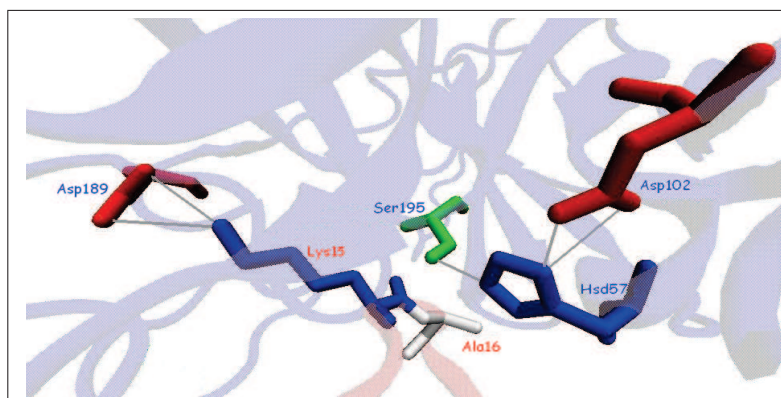


Figura 1.4: Zona de contacto BPTI-Tripsina. Asp189 interacciona eléctricamente con Lys15 para unir la Tripsina con la BPTI. Ser195, His57 y Asp102 (Tripsina) formando un sistema de carga para separa los residuos Lys15 y Ala16 de la proteína. Adaptada de la referencia [24].

1.4.2. Estructura de la BPTI

Esta proteína monomérica tiene una estructura tridimensional compacta con forma de pera, estabilizada por la presencia de un núcleo hidrofóbico (figura 1.6-D), por los enlaces disulfuro (figura 1.7) y los enlaces puente hidrógeno entre las estructuras secundarias. En el presente trabajo, se utilizó la estructura de la forma cristalina III de la BPTI, obtenida mediante Cristalografía de rayos X (archivo pdb 6pti)[28]. Ésta corresponde a la proteína monomérica que contiene 56 aminoácidos y 881 átomos.

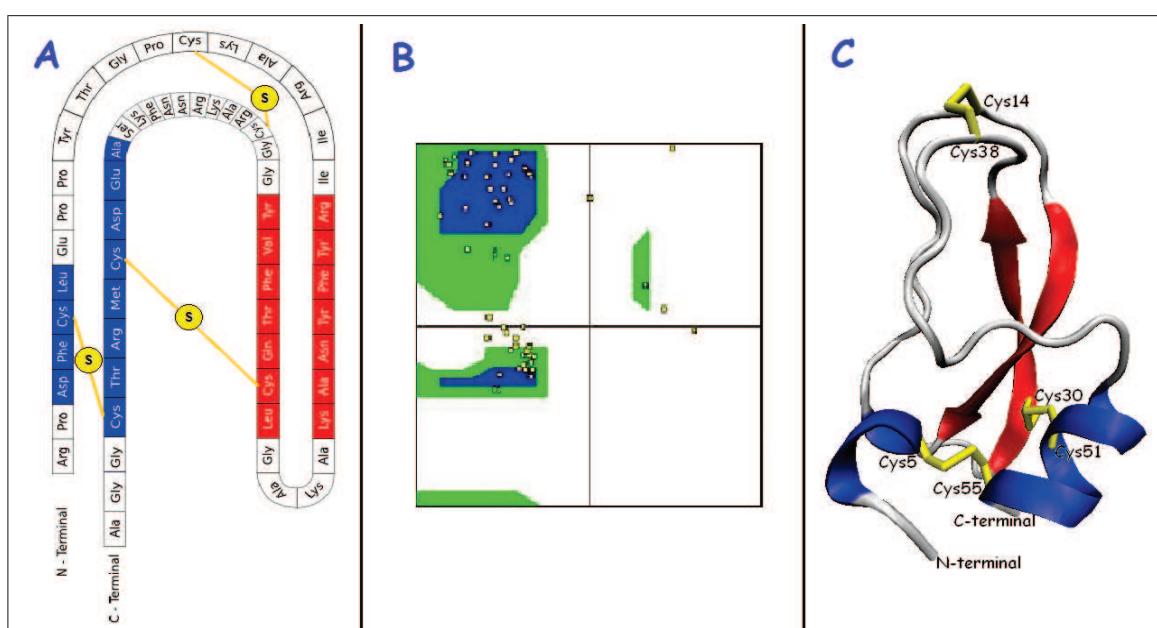


Figura 1.5: **A:** Secuencia de aminoácidos de la proteína BPTI adaptada de la referencia [24]. En el diagrama se muestra la relación entre la estructura primaria, secundaria y terciaria de la BPTI: se presenta la secuencia de aminoácidos, las estructuras secundarias formadas por ellos y las enlaces disulfuro que las unen. **B:** Mapa de Ramachandran de la BPTI: gráfica de los pares de ángulos (φ, ψ) para cada aminoácido, las regiones azul y verde representan pares de ángulos permitidos y parcialmente permitidos para la formación de estructuras secundarias respectivamente. En el cuadrante superior izquierdo están los pares de ángulos que pueden formar hojas β y en el inferior izquierdo los que pueden formar hélices. **C:** Estructura tridimensional de la BPTI. Están representadas en azul las estructuras tipo hélice y en rojo las dos hojas β antiparalelas. Las Cisteínas: Cys5, Cys55, Cys14, Cys38, Cys30, Cys51 que forman los tres enlaces disulfuro que unen las estructuras secundarias de la proteína se encuentra en color amarillo.

La figura 1.5-A muestra su estructura primaria, es decir, la secuencia lineal de aminoácidos que la componen y cómo están organizados en la formación de las diferentes estructuras secundarias presentes. La estructura secundaria de la BPTI está formada por dos hélices: la primera que va desde el residuo Asp3 al Leu6, la segunda que se encuentra cerca al terminal carboxilo va del Ala48 al Cys55; dos hojas β anti paralelas, la primera desde Ile18 al Asn24 y la segunda desde Leu29 al Tyr35 [24].

Debido a que es una proteína soluble en agua, los residuos no polares (hidrofóbicos) se sitúan en su interior, lo que le da gran estabilidad y es una de las causas de lo compacto de su disposición espacial [30]. Como muestra la tabla 1.1, aproximadamente el 45 % de los aminoácidos de la BPTI son no polares. La figura 1.6-D presenta los residuos no polares: Phe4, Cys5, Leu6, Pro8, Phe22, Tyr23, Cys30, Phe33, Asn43, Phe45, Cys51 y Cys55, formando un núcleo hidrofóbico en el interior de la proteína; lo que la vuelve compacta y hace que su desnaturalización sea más difícil de lograrse. Los aminoácidos polares forman el 30 % y los cargados el 25 % de la proteína (Figuras 1.6-A, 1.6-B y 1.6-C).

Tipo de Aminoácido	Número de residuos presentes en la BPTI
Polares	17
No polares	25
Cargados Ácidos	4
Cargados Básicos	10

Tabla 1.1: Número de aminoácidos presentes en la BPTI según su carga.

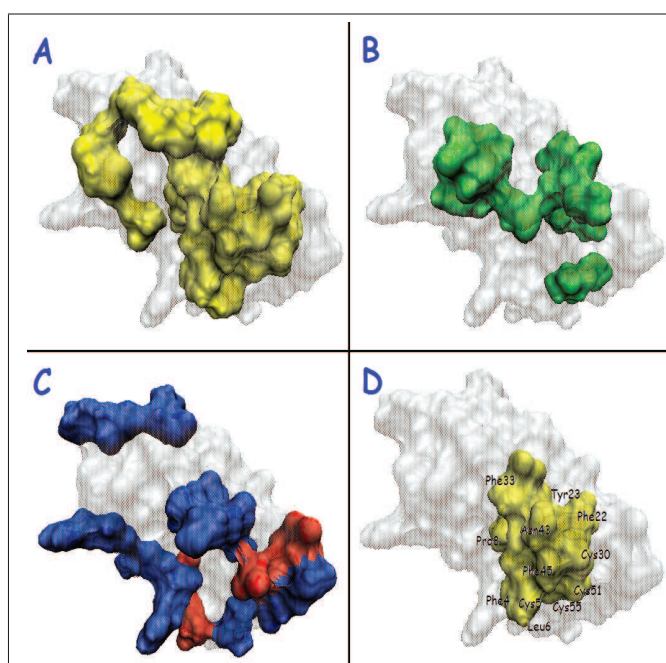


Figura 1.6: Disposición espacial de los aminoácidos de la BPTI, representación “Surf”, identificando los residuos: **A:** no polares (amarillo), **B:** polares (verdes) y **C:** cargados (ácidos-rojo, básicos-azul). **D:** Núcleo hidrofóbico formada en el interior de la proteína.

La presencia de tres enlaces disulfuro formados por seis Cisteínas: Cys5-Cys55, Cys14-Cys38, y Cys30-Cys51 aumentan la estabilidad de la BPTI ya que unen los elementos de su estructura secundaria [24]. En la figura 1.7 se observa como el enlace disulfuro formado por Cys5-Cys55 une la hélice α con la hélice β_{10} , el formado por Cys30-Cys51 une con la hélice β_{10} con una hoja β y el formado por Cys14-Cys38 une los lazos formados entre los residuos Glu7-Arg17 y Gly36-Ala40.

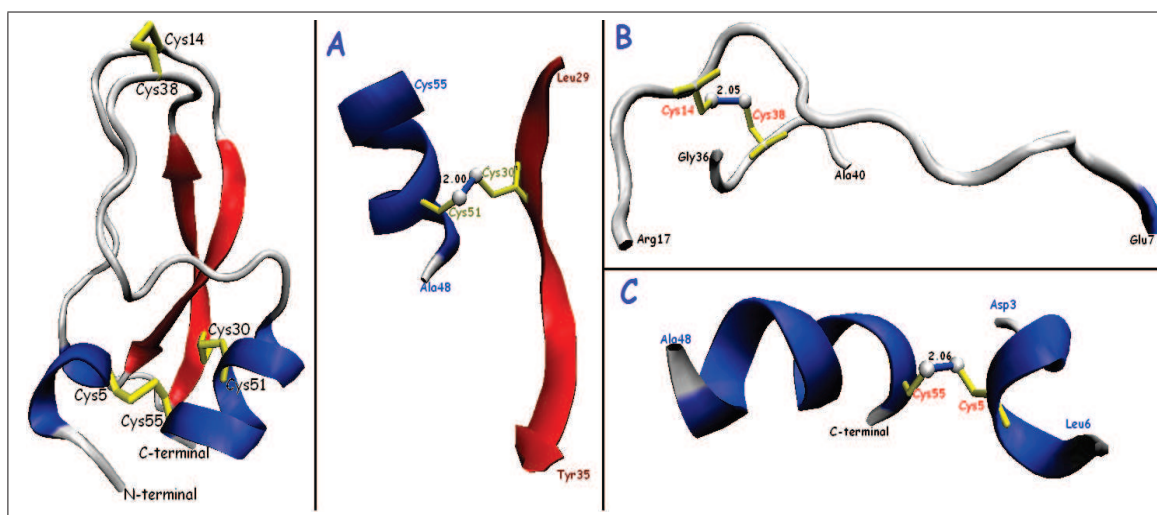


Figura 1.7: Enlaces disulfuro presentes en la estructura ternaria de la BPTI. **A:** Enlace disulfuro entre las Cys30 y Cys51 que tiene una longitud de 2.00 Å. **B:** Enlace disulfuro entre las Cys14 y Cys38 que tiene una longitud de 2.05 Å. **C:** Enlace disulfuro entre las Cys5 y Cys55 que tiene una longitud de 2.06 Å.

1.4.3. Estudios sobre el mecanismo de plegamiento de la BPTI

Las primeras investigaciones realizadas por Creighton mostraron que los procesos tanto de plegamiento como de desnaturalización de esta proteína, además de ser reversibles, se realizan a través de un mecanismo único [4]. Utilizando técnicas experimentales, Creighton aisló y estudió los intermediarios que contenían enlaces disulfuro y describió el proceso de plegamiento en términos de estos. Existen setenta y cinco posibles intermediarios (producto de las diferentes posibles combinaciones de las seis Cisteínas en la formación de enlaces disulfuro), de los cuales se identificaron siete intermediarios predominantes, dos contenían un enlace disulfuro y cinco dos enlaces disulfuro; no se identificaron intermediarios que posean tres enlaces disulfuro [4, 6, 31, 32].

Con la utilización de técnicas más avanzadas para atrapar y aislar intermediarios, Weissman y Finn [33] reexaminaron el modelo de plegamiento de la BPTI propuesto

por Creighton y mostraron la existencia de solo cinco intermediarios, dos de ellos contenían un enlace disulfuro y tres de ellos dos, además todos ellos poseían únicamente enlaces disulfuro nativos (presente en la estructura nativa) (Figura 1.8). La existencia de intermediarios que contienen únicamente enlaces disulfuro nativos muestran que las interacciones covalentes no específicas que estabilizan la estructura ternaria de la proteína tienen un papel preponderante en los caminos de plegamiento seguidos por la proteína [33].

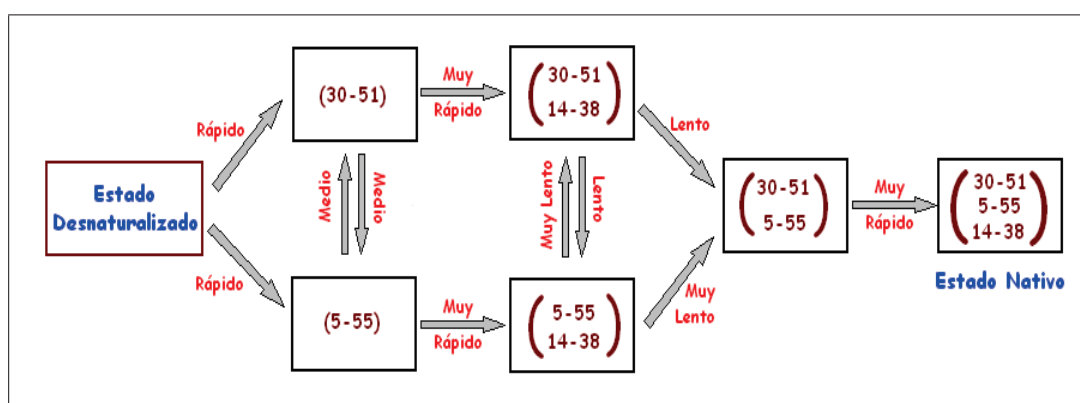


Figura 1.8: Esquema del plegamiento de la BPTI según Weissman y Finn, adaptado de la referencia [33]. Proceso del plegamiento de la BPTI, desde de su estado desnaturalizado hasta su estado nativo, pasando por cinco intermediarios. El intermediario (5-55, 14-38) corresponde a un estado intermediario cinéticamente atrapado, el término muy rápido se refiere a tiempos del orden de los milisegundos y el término muy lento a tiempos del orden de meses.

Una serie de estudios, tanto experimentales como teóricos, se han realizado posteriormente con el fin de corroborar los modelos antes descritos o para estudiar el papel de cierto tipo de aminoácidos en los intermediarios del plegamiento. Se ha investigado acerca de los efectos de mutaciones puntuales en los intermediarios predominantes [34, 35, 36], y se han realizado trabajos teóricos que reproduzcan los resultados de los modelos experimentales antes mencionados [31, 37]; sin embargo aún no existe una descripción completa y consistente del plegado de esta proteína. En 1995, Troyer y Cohen realizaron un estudio con el fin de describir el Entorno Energético de la BPTI [38]. Mediante simulaciones de Dinámica Molecular, describieron el mínimo global del Entorno Energético de esta proteína, a través de las conformaciones espaciales que lo caracterizan. Se determinó la existencia de 7 conformaciones accesibles, como resultado del estudio de un nanosegundo de simulación de Dinámica Molecular a 298 K mediante técnicas de análisis de grupos. Las simulaciones se realizaron utilizando solvente explícito, en particular se colocó a la proteína en una caja de agua.

A temperatura ambiente y bajo condiciones fisiológicas adecuadas, una cadena polipeptídica se mueve a través de múltiples conformaciones que están representadas por mínimos locales en la superficie multidimensional del Entorno Energético y separadas entre sí por barreras de energía. Estos estados conformacionales representan el promedio de un amplio número de estructuras ligeramente diferentes alrededor de las cuales la cadena transita [39, 40]. En este contexto, es importante el desarrollo de un método de análisis que permita estudiar de manera sistemática las conformaciones de las cadenas polipeptídicas obtenidas mediante técnicas computacionales como las simulaciones de Dinámica Molecular dentro del concepto del Entorno Energético. Esto contribuirá al desarrollo de futuras investigaciones relacionadas con el problema del plegado proteínico. En el presente trabajo, se desarrollará este método y particularmente se lo aplicará a la proteína BPTI (Bovine Pancreatic Trypsin Inhibitor) y a una estructura obtenida luego de realizar mutaciones que alteren los enlaces disulfuro presentes esta. El propósito es confrontar sus resultados con trabajos como el realizado por Troyer y Cohen, y expandérselos al estudio no solo del mínimo global sino de mínimos locales alejados; además de analizar cuál es el efecto de enlaces disulfuro en las conformaciones accesibles y cómo ésto se refleja en el Entorno Energético de la proteína.

Capítulo 2

Métodos Computacionales

2.1. Dinámica Molecular

La simulación de Dinámica Molecular es una técnica con la cual se obtiene la evolución temporal de un arreglo de átomos que interactúan entre sí, mediante la resolución de sus ecuaciones de movimiento [41, 42, 43]. Las interacciones se evalúan por medio de campos de fuerza, los cuales pueden obtenerse mediante cálculos y observaciones experimentales [42, 44, 45]. El fundamento de las simulaciones de Dinámica Molecular es el estudio de la conducta promedio de un sistema de muchas partículas mediante el cálculo numérico de la evolución temporal natural del sistema y el cálculo del valor medio de la cantidad de interés en un tiempo suficientemente grande [43].

Particularmente, la simulación de Dinámica Molecular es una herramienta importante para entender la base física de la estructura y la función biológica de las macromoléculas. Estas simulaciones proveen detalles concernientes al movimiento individual de los átomos en función del tiempo, que pueden ser utilizados para responder preguntas específicas acerca del sistema estudiado [46]. Actualmente una de las más importantes aplicaciones de la Dinámica Molecular es el análisis conformacional de macromoléculas (proteínas, polímeros, etc.) y el estudio de la Dinámica y Termodinámica de estos sistemas biológicos.

2.1.1. NAMD: NAnoscale Molecular Dynamics

Las simulaciones de Dinámica Molecular proveen información muy valiosa acerca del comportamiento de sistemas multiatómicos, como los biológicos. El gran avance en este ámbito, ha llevado a la creación de varios paquetes computacionales que realizan este tipo de trabajo; los primeros en desarrollarse fueron X-PLOR [58] y CHARMM (Chemistry at HARvard Molecular Mechanics) [49] para computadores en serie, sin embargo existía un alto costo computacional con simulaciones de sistemas biomoleculares grandes, por lo que se desarrollaron paquetes que se adapten a computadores en paralelo, reduciendo así de manera sustancial los costos computacionales [42].

NAMD [44] es una plataforma de Dinámica Molecular de alto rendimiento diseñada para la realización de simulaciones de sistema biomoleculares con muchos átomos (100000 o más), que trabaja tanto con sistemas de clusters computacionales como con procesadores individuales. Este paquete de simulación computacional trabaja con los campos de fuerza, parametrizaciones y formatos de AMBER (Assisted Model Building with Energy Refinement)[50] y CHARMM. Junto con la plataforma de visualización VMD (Visual Molecular Dynamics) [47], forman una herramienta muy eficiente para el estudio de sistemas biomoleculares como ADN, proteínas, complejos moleculares, etc. [44, 51]. A continuación se describirán los fundamentos básicos de la Dinámica Molecular, conceptos y algoritmos generales utilizados en este tipo de simulaciones, así como los específicos utilizados por NAMD.

2.1.2. Interacciones Moleculares: Funciones de Campo de Fuerza

Las trayectorias de un grupo de átomos que interactúan entre sí, están determinadas por las ecuaciones de movimiento de Newton [43, 44]:

$$m_i \ddot{\vec{r}}_i = -\frac{\partial}{\partial \vec{r}_i} U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N), \quad i = 1, 2, \dots, N \quad (2.1)$$

Cada átomo experimenta una fuerza determinada por la energía potencial U , que en la Dinámica Molecular se representa a través de un campo de fuerza, que expresa

la interacción entre los átomos del sistema [41, 44, 45, 52, 53]. La función de campo de fuerza $U(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_N)$ expresa las interacciones de este átomo con el resto de átomos del sistema. Para la selección de los campos de fuerza, debemos tomar en cuenta dos aspectos: la exactitud y la eficiencia computacional. Los campos de fuerza contienen varias contribuciones, éstas están asociadas con las desviaciones de los enlaces y ángulos de sus valores de equilibrio (o referencia), interacciones electrostáticas, e interacciones de van der Waals [54]. La plataforma NAMD utiliza una función de energía potencial que incluye cinco contribuciones [44]:

$$U = U_{\text{enlace}} + U_{\text{ángulo}} + U_{\text{torsión}} + U_{\text{vdW}} + U_{\text{Coulomb}} \quad (2.2)$$

Con:

$$U_{\text{enlace}} = \sum_{\text{enlaces } i} k_i^{\text{enlaces}} (r_i - r_{01})^2 \quad (2.3)$$

$$U_{\text{ángulo}} = \sum_{\text{ángulos } i} k_i^{\text{ángulo}} (\theta_i - \theta_{01})^2 \quad (2.4)$$

$$U_{\text{torsión}} = \sum_{\text{torsión } i} \begin{cases} k_i^{\text{ángulo}} [1 + \cos(n_i \phi_i - \gamma_i)] & , n_i \neq 0 \\ k_i^{\text{ángulo}} (\phi_i - \gamma_i)^2 & , n_i = 0 \end{cases} \quad (2.5)$$

$$U_{\text{vdW}} = \sum_i \sum_{j>i} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.6)$$

$$U_{\text{Coulomb}} = \sum_i \sum_{j>i} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.7)$$

La primera contribución (ec 2.3) corresponde a la interacción entre todos los pares de átomos con enlaces covalentes. Mediante un potencial armónico describe el estiramiento o contracción del enlace; las fuerzas entre átomos con enlaces covalentes son muy fuertes y la energía necesaria para desviarlos significativamente de su equilibrio es considerable. La segunda contribución (ec 2.4) tiene su origen en la flexión angular asociada a las variaciones de los ángulos entre cada par de enlaces covalentes que comparten un átomo en el vértice, también está modelizada mediante un potencial armónico; la energía necesaria para deformar cada ángulo y sacarlo del equilibrio es menor que la necesaria para estirar o contraer un enlace. La tercera contribución (ec 2.5), como se muestra en la figura 2.1, describe la variación del ángulo diedro, formado entre átomos unidos por tres enlaces covalentes, que se encuentra sometido a una torsión. La cuarta

contribución (ec 2.6), corresponde a las fuerzas de van der Waals, aproximada por el potencial de Lennard-Jones. Finalmente, la quinta contribución (ec 2.7), corresponde a las interacciones electrostáticas, modelado mediante el potencial Coulómbico [41, 44, 54]. Figura 2.1.

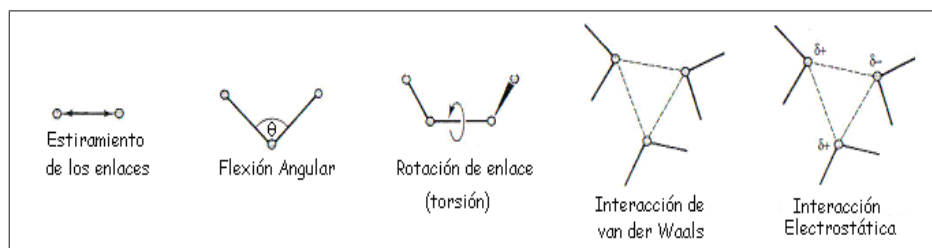


Figura 2.1: Representación esquemática de las cinco contribuciones al campo de fuerza usada en NAMD.

La determinación de las constantes en las tres primeras contribuciones se hace mediante técnicas empíricas y cálculos mecánico cuánticos y se encuentran en los archivos de parámetro del campo de fuerza [55, 56]; en NAMD se utilizan parametrizaciones con especificaciones de campos de fuerza CHARMM y AMBER [44].

2.1.3. Modelos de solvente

Para la realización de simulaciones de Dinámica Molecular de biomoléculas es necesario considerar el efecto del solvente en el cual se encuentra [27, 55, 57]. Este tratamiento se puede realizar de dos formas, utilizando el modelo del solvente implícito o solvente explícito. El modelo más estudiado y utilizado es el del solvente explícito y consiste en colocar a la biomolécula en una esfera o una caja del solvente requerido, es decir, rodear la estructura de moléculas solvente ubicadas aleatoriamente [57], de este modo en la simulación se toman en cuenta las interacciones del solvente con la de la estructura estudiada, estudiándose así el comportamiento de todos los átomos del sistema (biomolécula más solvente) en el tiempo; una de las desventajas de este modelo es el alto costo computacional debido a la presencia de un alto número de átomos, por ende de interacciones presentes en el sistema. En el caso del modelo del efecto del solvente implícito, la ventaja es que se suprimen errores en el cálculo estadístico existente cuando el número de moléculas de solvente es demasiado grande; todas las formulaciones para el cálculo de las interacciones electrostática incluyen la permitividad en el vacío ϵ_0 , sin embargo cuando se quiere añadir los efectos del solvente sin incluir

explícitamente las moléculas de este, el modelo más simple a seguir es incrementar el valor de la permitividad relativa hasta el valor necesario (dependiendo del solvente que quiera simularse), sin embargo existen modelos mucho más complejos de solvente implícito, ya que la variación en las interacciones electrostáticas no es el único de los efectos del solvente que debe considerarse [12, 54, 57]. NAMD no posee un modelo de solvente implícito incluido en su programación.

2.1.4. Condiciones de Frontera

En la simulación de un sistema limitado en el espacio, es muy importante el tratamiento que se dé a los efectos de superficie [41, 54]. Al considerar un sistema limitado por superficies, los átomos cercanos a ésta tienen menos vecinos que los que se encuentran el interior, esto produce consecuencias llamadas efectos de superficie [41]. En las simulaciones de Dinámica Molecular, sin importar cuan grande sea el sistema con el que se trabaja, se utiliza un número relativamente pequeño de partículas; además ya que el sistema está limitado espacialmente, el número de átomos que se encuentran cerca de la superficie es considerable [53, 54]. Debido a estas dos premisas la relación entre el número de átomos cercanos a la superficie y el número total es mayor a la real, lo cual puede causar un incremento de los efectos superficiales (en relación a los efectos reales) y por ende alteraciones en las mediciones de las propiedades del sistema [41, 53].

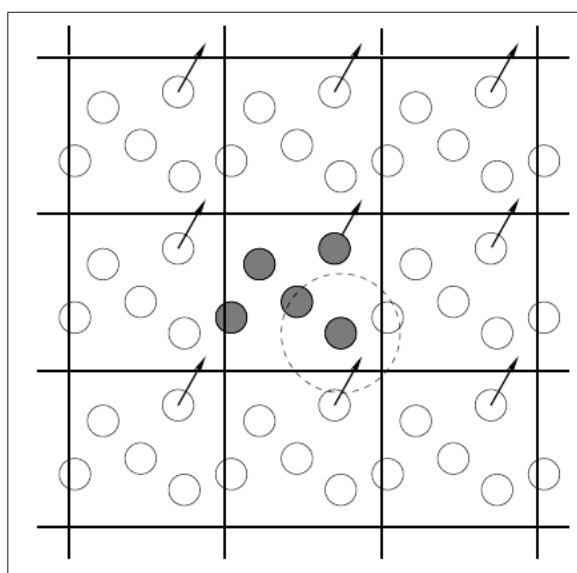


Figura 2.2: Condiciones Periódicas de frontera. Movimiento de un átomo fuera de la caja base, y movimiento de un átomo imagen para reemplazarlo. Se muestra además el criterio de las mínimas imágenes, donde el cálculo de las interacciones se dan dentro de una distancia de corte determinada de modo se tome en cuenta únicamente la interacción de cada átomo con la imagen mas cercana. Tomado de la referencia [48].

Para el tratamiento de los efectos de superficie se utilizan las condiciones de frontera, entre las más utilizadas son las condiciones de frontera periódicas. En este caso se coloca a los átomos dentro de una caja y réplicas de ésta rodeándola, es decir, la caja es replicada mediante traslación hasta llenar todo el espacio. Cada átomo del sistema en realidad representa un arreglo infinito de átomos ubicados en las diferentes réplicas de la caja. En la simulación, si un átomo sale de la caja base es reemplazado por una imagen que ingresa de la caja vecina [41, 44, 53, 54] (Figura 2.2). El fundamento de esta técnica es que cada átomo de la caja no interactúa únicamente con los átomos de ésta, sino también con sus imágenes en las cajas cercanas, esto es que las interacciones atraviesan las fronteras de la caja con lo que se eliminan los efectos de superficie del sistema y la posición de la frontera no tiene efectos en el sistema [41]. Con esto parecería que el número de interacciones se incrementaría, sin embargo, tomando en cuenta que el potencial es de corto alcance, podemos utilizar el “criterio de las imágenes mínimas”: cada átomo interactúa solo con el átomo o imagen más cercana en la red periódica [53, 54].

No siempre se utilizan condiciones de frontera periódicas, ya que cuando se trabaja con sistemas no homogéneos o fuera del equilibrio, pueden causar errores. La manera más sencilla de realizar una simulación que tome en cuenta tanto los efectos del solvente como los efectos de superficie, es colocando una capa fina de moléculas de solvente de alrededor de la estructura, en este caso el número de moléculas de solvente es menor que las necesarias para simulaciones con condiciones periódicas. De este modo los efectos de frontera de la estructura se transfieren de la interfaz molécula-vacío a la interfaz solvente-vacío, obteniendo mejores resultados en el tratamiento de la biomolécula [54]. Dentro de este tipo de condiciones de frontera tenemos las esféricas y las cilíndricas, en las cuales se fuerza a los átomos del sistema a conservar una forma determinada mediante la aplicación de un potencial armónico al sistema [54, 55].

2.1.5. Truncación de interacciones

Las interacciones de van der Waals se dan entre todos los átomos no enlazados del sistema, por lo que el número de términos que deben ser calculados es del orden del cuadrado del número de átomos existentes; para sistemas muy grandes el cálculo exacto de todas las interacciones no es práctico (muy costoso computacionalmente). Considerando que este tipo de interacciones decaen

rápidamente con el incremento de la distancia entre los átomos que interaccionan, se suelen realizar una serie de restricciones con el fin de reducir estos costos computacionales, como calcular únicamente las interacciones entre partículas que se encuentran dentro de un radio determinado y utilizar el criterio de las imágenes mínimas [53, 54].

En NAMD se establece para las interacciones de van der Waals una distancia de corte (cutoff); el cálculo de este tipo de interacciones se da de modo que se toman en cuenta únicamente las interacciones entre átomos que se encuentran dentro de esta distancia (Figura 2.3). Para esto se almacena, para cada átomo, una lista de los vecinos no enlazados con los cuales interacciona (átomos cuya distancia a el es menor o igual que la distancia de corte); esta lista se actualiza periódicamente en cada ciclo. La utilización de este método produce discontinuidades de la función de energía; para eliminarlas se aplican funciones tipo interruptor, éstas funciones polinómicas de la distancia, suavizan el decaimiento de la interacción desde una distancia (switchdist) dada hasta llegar a cero en la distancia de corte (cutoff), como se muestra en la figura 2.3-B. Para agilizar los procesos de cálculo computacional, NAMD utiliza una distancia dentro de la cual se buscan los átomos que interaccionan de forma no-enlazante (pairlistdist); de este modo la búsqueda no se realiza en todo el sistema sino solo dentro del radio antes mencionado [44].

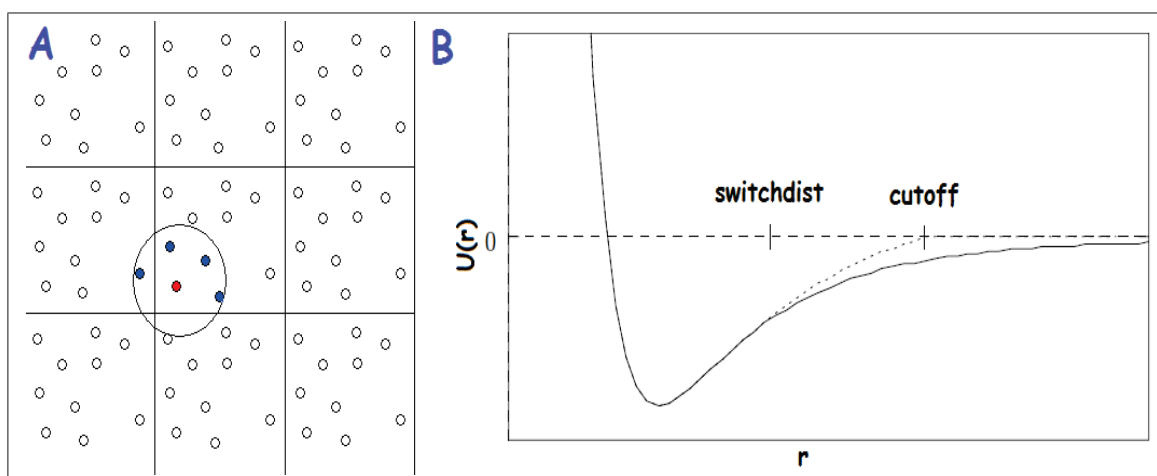


Figura 2.3: **A:** Representación del Potencial Truncado a la distancia de corte adaptada de la referencia [54]. El cálculo de las interacciones se da únicamente entre los átomos dentro de la circunferencia. Un sistema con condiciones de frontera periódicas, se muestra el potencial de truncado además del criterio de mínimas imágenes, para el átomo rojo únicamente se calculan las interacciones con los átomos azules. **B:** Gráfica del potencial del Lennard-Jones con (línea entrecortada) y sin (línea continua) la aplicación de las distancias switchdist y cuoff. Adaptada de la referencia [42].

2.1.6. Algoritmos de integración

Los sistemas biológicos son caóticos, es decir la trayectoria del sistema en el espacio depende muy sensiblemente de las condiciones iniciales; condiciones iniciales diferentes, aunque sean cercanas, producen que las trayectorias diverjan exponencialmente en su evolución temporal (inestabilidad de Lyapunov) [43, 44]. El fin de las simulaciones de Dinámica Molecular no es predecir con exactitud lo que va a ocurrir con el sistema conociendo sus condiciones iniciales, su principal interés está en las predicciones estadísticas [43, 53]. Por esto la elección del algoritmo para la integración numérica de las ecuaciones de movimiento es muy importante en las simulaciones [34].

Para simulaciones con energía, número de partículas y volumen constantes (ensamble NVE), lo importante del algoritmo de integración no es únicamente cuán preciso es, sino también su eficiencia, reversibilidad del tiempo y como conserva las propiedades dinámicas fundamentales como la energía y el momento [54]; el algoritmo de Verlet es rápido, su conservación de energía es buena y la precisión de las trayectorias generadas por éste también lo es [43], por ello en Dinámica Molecular es el algoritmo de integración más comúnmente usado [41].

El desarrollo del algoritmo de Verlet, parte de la expansión de Taylor de tercer orden para la posición $\vec{r}_n(t)$ [40, 42, 43, 48]:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \dot{\vec{r}}(t)\Delta t + \frac{\ddot{\vec{r}}(t)\Delta t^2}{2!} + \frac{\overset{\cdot\cdot\cdot}{\ddot{\vec{r}}}(t)\Delta t^3}{3!} + O(\Delta t^4) \quad (2.8)$$

Sabemos que: $\dot{\vec{r}}(t) = \vec{v}(t)$, además: $\ddot{\vec{r}}(t) = \vec{a}(t) = \frac{\vec{f}(t)}{m}$, siendo m la masa de la partícula. Reemplazando esto en la ecuación (2.8) tenemos:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{\vec{f}(t)}{2m}\Delta t^2 + \frac{\overset{\cdot\cdot\cdot}{\ddot{\vec{r}}}(t)\Delta t^3}{3!} + O(\Delta t^4) \quad (2.9)$$

Análogamente:

$$\vec{r}(t - \Delta t) = \vec{r}(t) - \vec{v}(t)\Delta t + \frac{\vec{f}(t)}{2m}\Delta t^2 - \frac{\overset{\cdot\cdot\cdot}{\ddot{\vec{r}}}(t)\Delta t^3}{3!} + O(\Delta t^4) \quad (2.10)$$

Sumando las expresiones (2.9) y (2.10), obtenemos:

$$\vec{r}(t + \Delta t) + \vec{r}(t - \Delta t) = 2\vec{r}(t) + \frac{\vec{f}(t)}{m}\Delta t^2 + O(\Delta t^4) \quad (2.11)$$

Como se ve el error en el cálculo de las nuevas posiciones es el orden de Δt^4 , por lo tanto el error del algoritmo es del orden de Δt^4 , en consecuencia este algoritmo es bastante preciso y por ello comunmente utilizado en la Dinámica Molecular. Un problema con este algoritmo es que las velocidades no se generan directamente, sin embargo se pueden obtener del conocimiento de la trayectoria [41, 43]:

$$\vec{r}(t + \Delta t) - \vec{r}(t - \Delta t) = 2\vec{v}(t)\Delta t + O(\Delta t^3) \quad (2.12)$$

$$(2.13)$$

$$\therefore \vec{v}(t) = \frac{\vec{r}(t + \Delta t) - \vec{r}(t - \Delta t)}{2\Delta t} + O(\Delta t^3) \quad (2.14)$$

NAMD utiliza una variación de este método, llamado el algoritmo de la velocidad de Verlet para simulaciones de ensambles NVE [44], el cual obtiene las posiciones y velocidades en un paso dado $(\vec{r}_{n+1}, \vec{v}_{n+1})$, conociendo las posiciones y velocidades en el paso anterior (\vec{r}_n, \vec{v}_n) y suponiendo que la fuerza $\vec{f}_n = \vec{f}_n(\vec{r}_n)$ está calculada [41, 43]. El algoritmo de las velocidades de Verlet permite calcular las posiciones y las velocidades en el mismo instante. Para el desarrollo de este algoritmo escribimos la expansión de Taylor de segundo orden:

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \vec{v}(t)\Delta t + \frac{\vec{f}(t)}{2m}\Delta t^2 \quad (2.15)$$

El cálculo de las velocidades se hace mediante la ecuación:

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{f}(t + \Delta t) + \vec{f}(t)}{2m}\Delta t \quad (2.16)$$

Es decir, primero se calculan las nuevas posiciones y a partir de éstas las nuevas fuerzas, para poder calcular las nuevas velocidades.

2.1.7. Ensamblen en la Dinámica Molecular

Los algoritmos antes mencionados se aplican cuando se considera la conservación de la energía total del sistema, es decir, un ensamble microcanónico o ensamble NVE (número de partículas, volumen y energía total constantes) [41]. Aunque tradicionalmente la Dinámica Molecular se ejecuta en este tipo de ensambles, existen también otras posibilidades como son los ensambles NVT (número de partículas, volumen y temperatura constantes) y NPT (número de partículas, volumen y presión constantes) [54], las cuales son más relevantes, ya que están asociadas a la energía libre de Helmholtz (NVT) y la energía libre de Gibbs (NPT) [55]. Para poder realizar simulaciones de Dinámica Molecular de estos ensambles, se deben modificar las ecuaciones del movimiento, de modo que esto nos lleve a generar una correcta distribución del ensamble [43]. En el presente trabajo se utilizará el ensamble NVT.

2.1.7.1. Ensamble NVT

Utilizamos simulaciones a temperatura constante cuando queremos estudiar como cambia la conducta de un sistema con la temperatura por ejemplo el plegamiento proteínico [54, 55]. Existen dos métodos para analizar la conducta de la Dinámica Molecular a temperatura constante [53]: El primer método consiste en hacer una selección aleatoria de las velocidades de una distribución de Maxwell-Boltzman, periódicamente. Esto equivale a poner al sistema en contacto con un reservorio de calor ocasionalmente. Este es un método simple de implementar y muy fiable [53, 55]. El segundo método consiste en poner el sistema en contacto con un reservorio de calor e introducir una variable de este dentro de las ecuaciones de movimiento; bajo estas condiciones la probabilidad de encontrar el sistema en un estado de energía dado esta determinado por la distribución de Boltzmann [41, 53].

NAMD utiliza un acoplamiento estocástico [44, 53] con un reservorio mediante la ecuación de movimiento de Langevin para generar la distribución de Boltzmann para el ensamble canónico (NVT) [44]. La ecuación del movimiento de Langevin tiene la forma:

$$m \dot{\vec{v}} = \vec{f}(\vec{r}) - \gamma \vec{v} + \sqrt{\frac{2\gamma k_B T}{m}} \vec{R}(t) \quad (2.17)$$

Donde, m es la masa, \vec{v} es la velocidad, \vec{f} la fuerza, γ es el coeficiente de fricción y $\vec{R}(t)$ es la fuerza aleatoria, que obedece a una distribución Gaussiana y no está relacionada ni con la posición, ni la velocidad ni la fuerza que actúa sobre las partículas. Para la integración de esta ecuación NAMD utiliza el método Brunger-Brooks-Karplus, que es una extensión del método de Verlet para las ecuaciones de Langevin [44]. La relación de recurrencia para la posición dada por este método es:

$$\vec{r}_{n+1} = \vec{r}_n + \frac{1 - \frac{\gamma\Delta t}{2}}{1 + \frac{\gamma\Delta t}{2}}(\vec{r}_n - \vec{r}_{n+1}) + \frac{1}{1 + \frac{\gamma\Delta t}{2}}\Delta t^2 \left[\frac{\vec{f}(\vec{r}_n)}{m} + \sqrt{\frac{2\gamma k_B T}{m}} \vec{Z}_n \right] \quad (2.18)$$

Donde, \vec{Z}_n es un arreglo de variables Gaussianas aleatorias con valor medio 0 y varianza 1.

2.1.8. Protocolo de simulación

El punto de partida de las simulaciones es la asignación de los valores iniciales. Las coordenadas iniciales del sistema se adquieren, generalmente, de la base de datos "Protein Data Bank"[58] y son obtenidas mediante técnicas experimentales como cristalografía de rayos X y Resonancia Magnética Nuclear (NMR); en caso de que no estén determinadas las coordenadas de los hidrógenos (propio de la cristalografía de rayos X), éstos deben ser adicionados adecuadamente antes de la simulación. Para obtener resultados realistas, es necesario considerar el efecto del solvente implementándolo al sistema sea explícita o implícitamente [57]. Finalmente el sistema debe ser eléctricamente neutro para evitar inestabilidades en las simulaciones.

Los parámetros de simulación se determinan analizando las características del sistema con el que se va a trabajar, es decir, el tipo de ensamble a simularse: NVE, NVT o NPT. Para ello se deben especificar las condiciones del número de partículas (N), volumen (V), presión (P), temperatura (T), energía total (E) que van a ser utilizadas. Además se deben especificar las condiciones de frontera que utilizarán, y las especificaciones en cálculos, como truncaciones de potencial e interacciones de largo alcance [57]. Son claves en las simulaciones de Dinámica Molecular, la elección del tamaño adecuado de la simulación así como de los pasos

de tiempo entre cada integración de las ecuaciones del movimiento. En general el tiempo total de la simulación debe ser 10 veces mayor al tiempo en que ocurre el fenómeno estudiado [57] y el tamaño de los pasos entre las integraciones debe ser menor que el tiempo en que suceden los movimiento más rápidos en el sistema con el fin de reproducir estos y evitar así inestabilidades, en las biomoléculas los movimientos más rápidos son los relacionados con vibraciones de los enlaces que conectan los hidrógenos con átomos pesados y se dan en tiempos del orden de los femtosegundos [55].

Luego de que las cordenadas iniciales del sistema han sido definidas, su energía debe minimizarse, de esta forma los átomos son reubicados de modo que la geometría sea lo más óptima posible (mínimo energético) [55]. Posteriormente, el sistema se calienta desde 0 K hasta la temperatura deseada mediante un protocolo que reasigna gradualmente las velocidades de los átomos del sistema hasta llegar a dicha temperatura.

Una vez que el sistema se encuentra a la temperatura deseada se procede a equilibrarlo. La equilibración se realiza hasta que los parámetros representativos del sistema se estabilicen (alrededor de 0.5 nanosegundo para biomoléculas como proteínas y ADN). Finalmente, luego de equilibrado el sistema se llega a la fase de producción, en la cual se genera una trayectoria (estructuras en el tiempo); el tiempo de duración de esta fase depende del proceso que quiera estudiarse [55, 57].

Como resultado de cada paso del protocolo (minimización, calentamiento, equilibración y producción) el programa NAMD arroja una serie de archivos con información concerniente a cada proceso: un archivo de texto que contiene la información de la evolución de la energía (total y cada uno de los aportes por separado), presión y temperatura del sistema en el tiempo; otros binarios que contienen las coordenadas y velocidades finales del sistema así como uno que almacena las coordenadas de los átomos del sistema cada cierto número de pasos especificado previamente. La figura 2.4 muestra un esquema del protocolo, desde la construcción del sistema hasta la producción de las trayectorias de Dinámica Molecular.

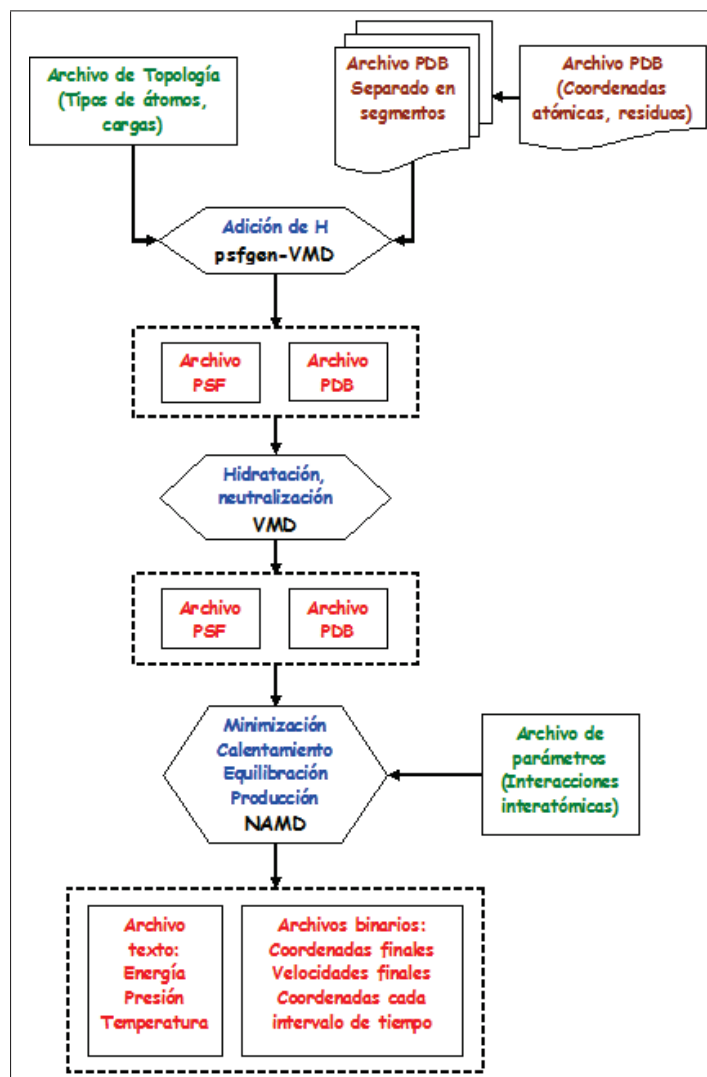


Figura 2.4: Diagrama de flujo del protocolo de simulación de Dinámica Molecular. Para la construcción del sistema, se utiliza la plataforma VMD y para los procesos de del sistema, minimización, calentamiento, equilibración y producción es necesaria la plataforma de simulación NAMD.

2.2. Análisis Conformacional

Las propiedades físicas, químicas así como las funciones biológicas de una molécula dependen de la conformación que ésta adopte, es decir, de la forma tridimensional asociada a un arreglo espacial particular de los diferentes elementos que la conforman [16, 54]. El análisis conformacional es el estudio de las conformaciones accesibles a una molécula y la influencia en sus propiedades [55]. Este tipo de análisis es muy común en el diseño de drogas y en el estudio del plegamiento proteínico por la versatilidad y resultados que muestra.

En el contexto de las simulaciones de Dinámica Molecular, el análisis conformacional se refiere al estudio de las estructuras obtenidas mediante las simulaciones con el fin de extraer información acerca del sistema. Este análisis permite identificar los cambios drásticos en las estructuras (transiciones conformacionales) y enfocarse en el estudio de grupos de estructuras similares (conformaciones) entre los cuales estos se producen [54, 55]. Estos estudios consisten básicamente de tres pasos: muestreo, optimización y análisis.

2.2.1. Muestreo

En una simulación lo suficientemente larga, la molécula podría adoptar todas sus conformaciones accesibles. Sin embargo esto no es posible, ya que fenómenos que implican cambios conformacionales como el plegamiento proteínico se dan en el orden de los segundos, y en el mejor de los casos, el poder computacional actual no va más allá de producir simulaciones de fracciones de microsegundos. Existen sin embargo varios métodos sistemáticos para explorar este espacio conformacional, basados en estrategias de muestreo cuyo fin es encontrar un conjunto amplio de estructuras que representen la mayor cantidad de conformaciones posibles [55]. Utilizando Dinámica Molecular, las dos estrategias utilizadas son realizar simulaciones a altas temperaturas y simulaciones de Dinámica Molecular Dirigidas (Targeted Molecular Dynamics).

2.2.1.1. Dinámica Molecular a altas temperaturas

Como se mencionó antes, una trayectoria de Dinámica Molecular suficientemente larga a temperatura ambiente podría describir toda el espacio conformacional de una molécula, en cambio las estructuras resultantes de simulaciones del orden de los nanosegundo están confinadas una pequeña región de este espacio, ya que a temperatura ambiente la probabilidad de que la molécula cruce altas barreras de energía en esos tiempos es demasiado pequeña, como se muestra en la figura 2.5. Una solución para obtener una muestra amplia de conformaciones es utilizar el comportamiento de moléculas a altas temperaturas ya que bajo estas condiciones es más sencillo que la molécula supere barreras energéticas y pase a nuevos estados (conformaciones) [55, 59]. Usualmente se utilizan temperaturas alrededor de los 1000 K para simulaciones de péptidos y proteínas, ya que es suficientemente grande como para que la moléculas cruce altas barreras

energéticas, pero está por debajo de la temperatura a la cual se da la transición *cis-trans* en los péptidos [59]. Ya que en la mayoría de campos de fuerza para biomoléculas, los enlaces covalentes están modelados con un potencial armónico, éstos no se rompen espontáneamente por el aumento de temperatura, por lo que el esqueleto (backbone) de la estructura no cambia lo que implica que no existen cambios químicos en ella [55].

Como muestra la figura 2.5 las estructuras obtenidas con esta estrategia no están asociadas al Entorno Energético de la proteína a 300K sino a 1000K, por ello es necesario un tratamiento antes de utilizarlas para definir el espacio conformacional a 300 K. La estrategia utilizada consiste en el enfriamiento suave de las estructuras hasta alcanzar un estado estable a 300 K que posea menor energía. La clave de este método está en el enfriamiento paulatino (muy lento) al que es sometido el sistema, el cual se realiza de forma análoga al calentamiento, es decir, mediante el reescalamiento de las velocidades de los átomos del sistema. Este enfriamiento es capaz de llevar a la estructura a un mínimo energético lejano a la estructura inicial (en el Entorno Energético a 300 K), superando barreras energéticas[54, 55].

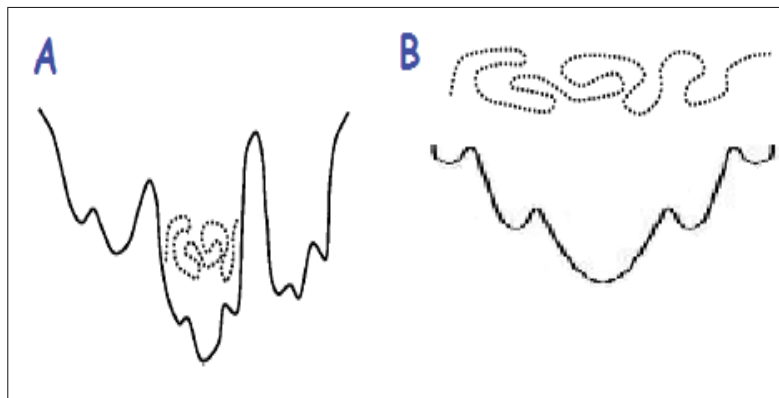


Figura 2.5: **A:** Representación esquemática de una simulación de Dinámica Molecular a temperatura ambiente, las estructuras resultantes están confinadas a un pequeño espacio de la superficie energética debido a la presencia de barreras de energía altas que no pueden ser vencidas. **B:** Representación esquemática de una simulación de Dinámica Molecular a altas temperaturas, la molécula puede superar varias barreras de energía. Adaptado de la referencia [55].

2.2.1.2. Dinámica Molecular Dirigida

La técnica de Dinámica Molecular Dirigida (Targeted Molecular Dynamics) es utilizada para encontrar la secuencia de conformaciones que toma una molécula entre dos estados conocidos, esto se realiza mediante la resolución de las ecuaciones de Newton con un potencial físico y una restricción holonómica. La

aplicación de esta restricción a cada átomo de la estructura se da de modo que la distancia rms con el estado final decrece continuamente.

En coordenadas cartesianas la distancia rms d_{ij} entre dos estructuras i, j en una molécula determinada se define como el mínimo del funcional:

$$d_{ij} = \sqrt{\frac{1}{N} \sum_{k=1}^N \left| \vec{r}_k^{(i)} - \vec{r}_k^{(j)} \right|^2} \quad (2.19)$$

donde N es el número de átomos de la molécula, $\vec{r}_k^{(i)}$, $\vec{r}_k^{(j)}$ son las coordenadas cartesianas del átomo k de las estructuras i, j respectivamente. El mínimo de la ecuación (2.19) se obtiene mediante una óptima superposición de las dos estructuras.

En este tipo de simulaciones los átomos del sistema son guiados desde un estado inicial hasta un final por medio de la aplicación de fuerzas direccionadas; en cada paso de integración la estructura es alineada con la estructura objetivo y se calcula la distancia rms entre ellas, así la fuerza aplicada se da de modo que la distancia rms decrece [42]. El potencial (restricción holonómica) para las simulaciones dirigidas está dado por:

$$U_{TMD} = \frac{k}{2N} [d_{rms}(t) - d_{rms}^*(t)]^2 \quad (2.20)$$

donde k es la constante de la restricción, N el número de átomos del sistema, $d_{rms}(t)$ es la distancia rms instantánea entre la estructura en el tiempo t y la estructura objetivo (previo alineamiento) y $d_{rms}^*(t)$ evoluciona linealmente desde un valor inicial (d_{rms} entre las estructuras inicial y objetivo) hasta un valor final especificado.

La Dinámica Molecular Dirigida es ampliamente utilizada en estudios relacionados con cambios conformacionales ya que nos provee de un modelo bastante aproximado en tiempos computacionalmente manejables, además sus resultados son equivalentes a los que se obtienen con simulaciones de Dinámica Molecular tradicionales [60] Esto nos brinda la posibilidad de estudiar la región del espacio conformacional comprendida entre dos estados, por ende hallar información acerca de la barreras energéticas que las separan.

2.2.2. Optimización

Mediante las estrategias mencionadas podemos encontrar una amplia muestra de conformaciones, sin embargo muchas de las estructuras generadas con estos métodos no corresponden a conformaciones accesibles (localmente estables), sino a conformaciones transitorias [55, 38]. Antes de realizar el análisis de las estructuras es conveniente llevar todas estas conformaciones al mínimo local más cercano para de este modo obtener únicamente estructuras asociadas a conformaciones estables.

2.2.2.1. Minimización

Mediante el método de minimización de energía las conformaciones caen en el mínimo local energético más cercano, es decir, simplemente busca el estado con menor energía, como se muestran en la figura 2.6. El protocolo de minimización busca optimizar la geometría de la estructura de modo que las interacciones internas den como resultado la estructura con mínima energía pero cerca a la inicial, este método por su naturaleza no permite que se superen barreras energéticas. Cada estructura se dirige hacia el mínimo local más cercano [55, 38]. Existen varios algoritmos de minimización de energía, entre los que podemos nombrar: Descenso por pasos (SD), Newton-Raphson, Gradiente Conjugado (CG), etc. NAMD utiliza el Algoritmo del Gradiente Conjugado [42].

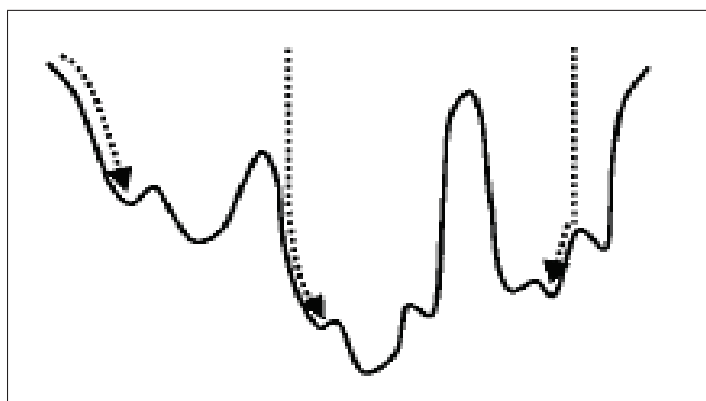


Figura 2.6: Métodos de optimización conformacional. Representación esquemática de la Minimización de energía, en la cual las estructuras caen a los mínimos locales más cercanos sin superar ninguna barrera energética. Adaptado de la referencia [55].

2.2.3. Análisis

El análisis de la muestra puede estar dirigido al estudio de características generales de la molécula como la flexibilidad o a la identificación de tendencias en la muestra de estructuras; pueden también enfocarse en la identificación y caracterización de grupos de estructuras con energías bajas que potencialmente pueden utilizarse en el diseño de drogas o para el estudio de las conformaciones accesibles durante el plegado proteínico, etc. [55]

2.2.3.1. Medición de disimilitudes

Antes de iniciar el análisis se debe realizar una comparación cuantitativa entre las estructuras de la muestra mediante una medición del grado de disimilitud entre ellas. Comúnmente el grado de disimilitud entre dos estructuras se cuantifica mediante la distancia rms (Root mean square distance o distancia cuadrática media).

Por lo regular el cálculo de la distancia rms se hace únicamente de los átomos pesados o del esqueleto (backbone) de la estructura, ya que éstos son los únicos importantes en los cambios conformacionales de molécula. Los resultados de las distancias rms entre todas las estructuras de la muestra definen la matriz de disimilitud Δ , cuyos elementos (d_{ij}) , son las distancias rms entre las conformaciones i, j [55].

2.2.3.2. Analisis de grupos (Cluster Analysis)

La matriz de disimilitud Δ contiene información sobre la distancia relativa entre las estructuras, sin embargo casi nunca provee información por sí sola, por lo que requiere de un tratamiento adicional con el fin de obtener datos acerca del comportamiento de la molécula. Una de las técnicas utilizadas para el agrupamiento de estructuras es el análisis de grupos (cluster analysis), el cual considera la similitud estructural entre las conformaciones para la formación de conglomerados de estructuras [55]. El propósito de esta técnica multivariante es formar grupos de individuos compactos respecto a ciertas características, es decir, que los individuos que forman cada grupo sean similares entre sí (respecto a las características determinadas) y diferentes a los de los otros grupos; esto permite descubrir asociaciones de estructuras que no son evidentes a priori, por lo que sus resultados pueden contribuir a la construcción de esquemas de clasificación, a la identificación de características de los grupos, etc.

Para la construcción de los grupos de individuos existen dos métodos generales, en el uno la clasificación resultante tiene un número creciente de clases anidadas y se denomina método jerárquico, mientras que en el segundo las clases resultantes no son anidadas y se denomina método no jerárquico. Para utilizar el método no jerárquico es necesario conocer el número de grupos a priori, en cambio, en los jerárquicos no se requiere esa información; estos últimos además se utilizan comúnmente con fines exploratorios y sus resultados se suelen afinar utilizando métodos no jerárquicos, es decir, ambos son complementarios. La selección del método a utilizarse para el análisis de grupos depende de las condiciones particulares del problema a analizarse, no existe un criterio determinado para escoger cual de los algoritmos es mejor y generalmente esta selección se hace valorando cual de ellos refleja de mejor manera lo que se está buscando. En el estudio conformacional de proteínas, en general, no se tiene un conocimiento previo del número de grupos que desean obtenerse, por lo que es conveniente la utilización de métodos jerárquicos.

Los métodos jerárquicos, a su vez se dividen en aglomerativos y divisivos; en los primeros se forman grupos, realizando fusiones de los individuos mediante un criterio de distancia entre ellos; existe una serie de estos métodos que se diferencian entre si por la forma de definir la distancia entre los grupos; en los segundos, por el contrario, se parte de un grupo amplio (todos los individuos) que se va particionando hasta formar los grupos requeridos. La distancia entre grupos define el nivel de la formación de un nuevo grupo y se denomina distancia de aglomeración. Los resultados de todas las clasificaciones jerárquicas se presentan mediante un diagrama bidimensional llamado dendograma, el cual muestra las sucesivas divisiones o fusiones (aglomeraciones) realizadas en el proceso de análisis, en cual los nodos representan la formación de un nuevo grupo y la altura el nivel de agrupación (distancia de aglomeración). Los mejores resultados se obtienen con los aglomerativos.

Entre los métodos jerárquicos más comúnmente utilizados tenemos: el del vecino más próximo (single linkage), del vecino mas lejano (complete linkage), del grupo promedio (group average), del centroide y de Ward [61, 62]. Éstos siguen un algoritmo similar, la diferencia radica en la forma cómo toman en cuenta las distancias entre los grupos que se forman. El del vecino más próximo o de la mínima

distancia considera como distancia de aglomeración a la menor de las distancias existentes entre los miembros más cercanos de distintos grupos (Figura 2.7-A); de esta forma se obtienen grupos bastante homogéneos pero permite agrupación de individuos muy lejanos tendiendo a construir grupos muy grandes y sin sentido. El del vecino más lejano o de la distancia máxima considera, para unir grupos, la mayor de las distancias existentes entre los miembro más lejanos de distinto grupos (Figura 2.7-B); los grupos formados son más compactos (no permite la agrupación de individuos muy alejados) pero heterogéneos. El del grupo promedio considera, para la unión de grupos, el promedio de las distancias entre todos los individuos de los grupos (Figura 2.7-C); crea grupos de tamaño intermedio (mas grandes que los generados por el método del vecino más próximo pero mas pequeños que lo producidos con el método del vecino más lejano) y provee una buena representación gráfica. En el método de Ward no se calcula distancias entre grupos, sino que se forma los grupos de modo que se maximice la homogeneidad dentro de éste; pese a ser eficiente tiende a crear grupos muy pequeños. En la práctica es conveniente realizar el análisis con varios de estos métodos, comparar su consistencia y seleccionar el que provea una solución interpretable [54, 61, 62].

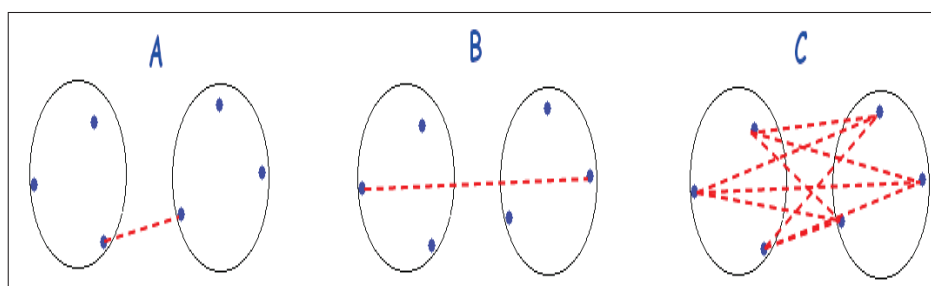


Figura 2.7: Representación esquemática de algunos métodos de análisis de grupos aglomerativo y cómo se consideran las distancias entre grupo en cada uno de ellos. **A:** Método del vecino más cercano (single linkage). **B:** Método del vecino más lejano (complete linkage). **C:** Método del grupo promedio (group average).

Una vez obtenido el resultado del análisis de grupos se puede determinar el número óptimo de grupos. Sin embargo no existen criterios objetivos para dicha selección; si se escogen muy pocos grupos estos serán heterogéneos y si se seleccionan demasiados su interpretación se vuelve muy complicada. La elección del número óptimo de grupos es particular para cada caso y para hacerla es conveniente considerar las características iniciales de los individuos, el dendograma y la matriz de disimilitud [61] y puede ser sujeta a varias interpretaciones; no existe un criterio universal aceptado. Particularmente en el estudio conformacional es conveniente y ampliamente utilizada la representación gráfica de la matriz de disimilitud Δ

para determinar regiones que representan grupos de estructuras [55, 38]. Otra de la opciones utilizadas es mediante el análisis de la distancia de aglomeración, considerando que saltos grandes de esta distancia determinan la distancia de corte; otra forma es utilizando el criterio de Mojena [63], el cuál define una distancia máxima de aglomeración tolerable α , bajo la cual se da la formación del número óptimo de grupos: $\alpha \geq \bar{\alpha} + ks$, donde $\bar{\alpha}$ es la media de las distancia de aglomeración y s la desviación estándar.

2.2.3.3. Escalamiento Multidimensional

El problema intrínseco del análisis de las conformaciones accesibles a una molécula es la alta dimensionalidad que presenta el espacio conformacional; una molécula de N átomos tiene un espacio conformacional $(3N-6)$ dimensional. Sin embargo, en la práctica, son suficientes únicamente un número pequeño de dimensiones para describir de una forma apropiada el Entorno Energético inherente a la proteína [55].

En general, el escalamiento dimensional de estos espacios se puede realizar mediante dos técnicas: el análisis de componente principal (PCA) y el análisis de coordenada principal (PCoorA). Los datos iniciales para esta clase de análisis pueden ser de varios tipos: datos brutos y distancias o disimilitudes entre pares de objetos.

Los datos brutos corresponden a una matriz $\mathbf{X}_{n \times p}$ que contiene los valores de p variables de n individuos; en el caso particular del análisis conformacionales los individuos corresponden a las estructuras obtenidas luego del muestreo y la optimización y las variables a las coordenadas de los átomos de estas.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & & x_{np} \end{bmatrix}$$

Las distancias o disimilitudes se almacenan en una matriz $\Delta_{n \times n}$ como la descrita en el paratado 2.2.3.1., esto es una matriz simétrica que contiene la disimilitud entre objetos. la diagonal principal contiene únicamente ceros.

$$\Delta = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1n} \\ d_{21} & 0 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & & 0 \end{bmatrix} \quad d_{ii} = 0; \quad d_{ij} = d_{ji}; \quad i, j = 1, \dots, n$$

La reducción de la dimensionalidad de espacios manteniendo las distancias interconformacionales se realiza proyectando los datos multidimensionales en un subespacio óptimo de menor dimensionalidad [54, 55, 62]. Este análisis fue introducido en el estudio de proteínas por Karplus y se ha constituido en uno de los métodos más utilizados en aplicaciones relacionadas al agrupamiento y visualización de conformaciones [55]. Los análisis de grupos pueden realizarse con cualquier paquete de estadística multivariante, en el presente trabajo se utilizó para estos el paquete DarWin5 [68].

2.2.3.3.1. Análisis de Componentes Principales (ACP) Este análisis consiste en encontrar transformaciones ortogonales de las variables originales con el fin de obtener un nuevo conjunto de variables no correlacionadas, denominadas Componentes Principales, que son combinaciones lineales de las variables originales. El fin de ACP es recoger mayor parte de la variabilidad de los datos en unas pocas variables (las primeras), obteniéndose de esta manera una reducción de la dimensión en los mismos. El propósito fundamental de la técnica consiste en la reducción de la dimensión de los datos con el fin de simplificar el problema en estudio [62].

Las columnas matriz $X_{n \times p}$ generan un espacio p dimensional, por lo que la transformación lineal que se debe obtener es una rotación en el espacio p dimensional; el espacio generado por las primeras q componentes es por lo tanto, un subespacio vectorial q -dimensional del espacio p -dimensional original. En principio se pueden obtener tantas componentes principales como variables originales, por tanto existen p vectores $Y_i, n \times 1$, que contienen los valores de las componentes principales para cada uno de los individuos; la primera componente principal (Y_1) contiene la varianza de mayor tamaño de la muestra, la segunda mayor varianza corresponde a la segunda componente (Y_2) y así sucesivamente. Para realizar esta transformación lineal es necesario construir la matriz de covarianza $p \times p$, $C = X^T X$, que expresa la relación entre las variables y posee una base completa de vectores propios.

2.2.3.3.2. Análisis de Coordenada Principal (ACooP) De igual manera que en el del ACP, el problema fundamental del ACooP es la obtención de los vectores y valores propios de un espacio vectorial, que en el campo del cálculo matricial se da bajo el problema de la diagonalización de una matriz cuadrada. La diferencia radica en la matriz de la cual se debe obtener éstos; el ACooP parte de la matriz de distancias $\Delta_{n \times n}$ la cual expresa la relación (distancias o disimilitudes) entre los individuos [64, 65]. En principio este método se desarrollo con el fin de obtener un grupo de coordenadas a partir de una serie de observaciones cualitativas, sin embargo, si se utiliza variables cuantitativas este método es análogo al ACP [69].

Para el ACooP se parte de un espacio cualquiera en el que las disimilaridades entre individuos se miden mediante una distancia cualquiera que no tiene que tener una interpretación física concreta y el resultado final son las coordenadas de los individuos en un espacio euclídeo; en el análisis conformacional es usual utilizar como distancia de disimilitud entre estructuras la distancia rms almacenada en la matriz Δ . Para garantizar que esta matriz Δ tenga un valor propio igual a cero, ésta debe ser centrada, es decir, la suma de todos los elementos de cada fila y columna debe ser cero; en caso de no serlo, se deben realizar las transformaciones adecuadas antes de realizar el análisis. Los valores propios resultantes dan el porcentaje de proyección de la distribución original (porcentaje explicado) y los vectores propios corresponden a las coordenadas de los puntos originales en el nuevo arreglo de ejes [62, 64, 65].

La dualidad entre estos dos métodos fue probada por Gower en 1996, es decir, los valores y vectores propios resultantes de ambos métodos son los mismos [66]. Ambas técnicas han sido ampliamente utilizadas tanto para el estudio de trayectoria de Dinámica Molecular como para la construcción de Entornos Energéticos; sin embargo una de la ventajas del ACooP en este tipo de análisis es que la matriz base de éste (matriz de distancia o de disimilitud Δ) muestra la relación entre las estructuras, la cual es más importante que la relación entre variables en estos casos [64]. Se utilizó el programa DarWin5 [68] porque posee con un costo computacional aceptable y provee gran versatilidad para el manejo de sus resultados.

Capítulo 3

Conformaciones espaciales de la BPTI

3.1. Simulaciones de Dinámica Molecular de la BPTI

Las simulaciones de Dinámica Molecular aplicadas a la caracterización de conformaciones proteínicas han permitido ampliar el conocimiento obtenido con técnicas experimentales [39, 67]. En este contexto se puede definir una conformación como el promedio de una serie de estructuras semejantes obtenidas de la Dinámica Molecular, a las cuales corresponde una región de baja energía de la superficie del Entorno Energético, por lo tanto una transición entre conformaciones implica superar una barrera de energía. Estos movimientos conformacionales en las proteínas se dan en varios niveles, desde cambios en los ángulos de torsión hasta reordenamiento de las estructuras secundarias, cada uno con tiempos que van desde los nano segundos hasta los segundos. La descripción completa del espacio conformacional de una proteína a temperatura ambiente no es computacionalmente alcanzable; técnicas como las simulaciones a altas temperaturas y las simulaciones dirigidas pueden ser utilizadas con el fin de obtener una muestra de estructuras que describa de la manera más detallada su espacio conformacional. A continuación se describirán los resultados de la aplicación de estas técnicas al caso de la BPTI natural y de una estructura obtenida sustituyendo las cisteínas 5, 14 y 30 por alaninas, aterando de este modo los enlaces disulfuro presentes en la proteína.

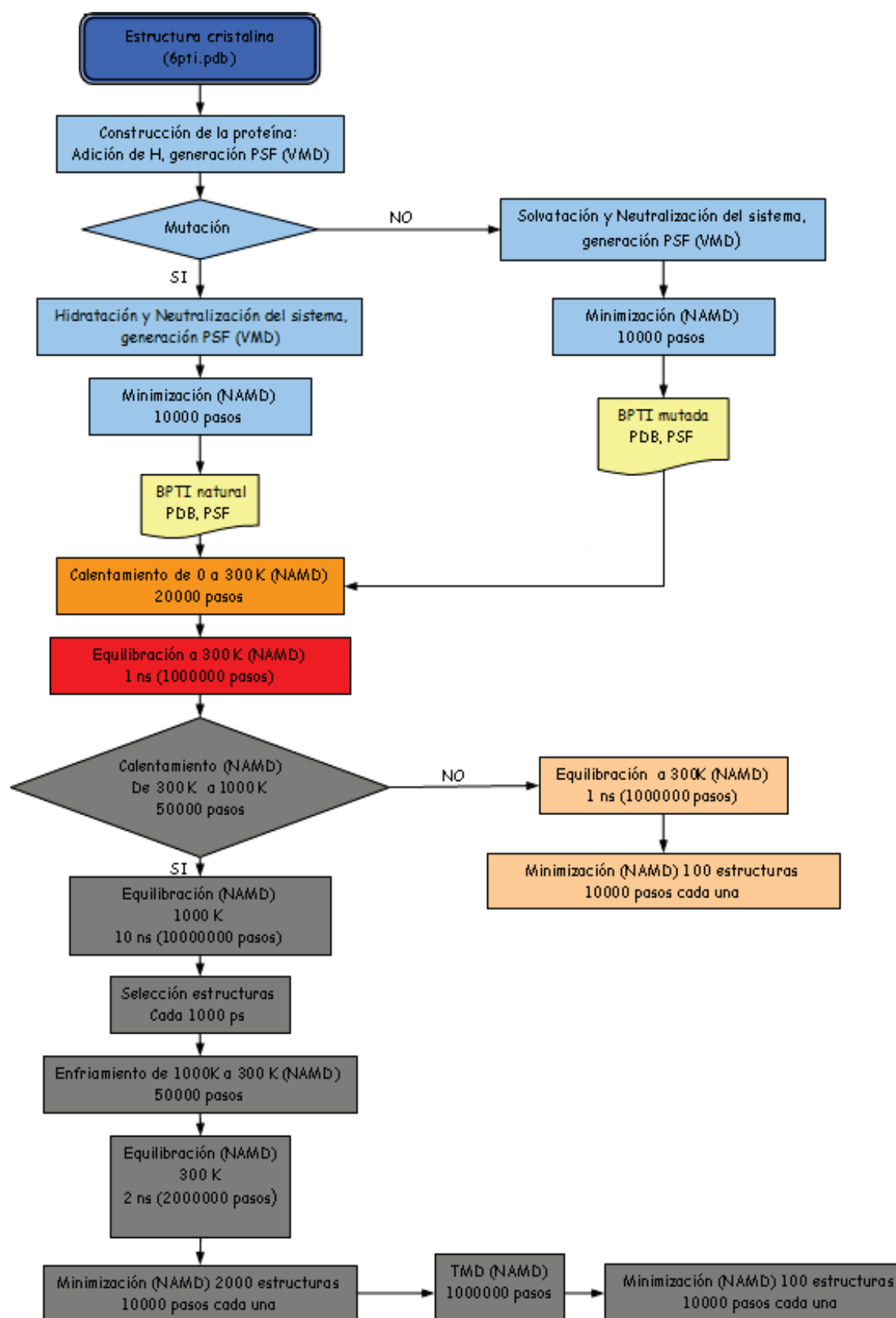


Figura 3.1: Diagrama de flujo del tratamiento realizado a partir de la estructura cristalina de la BPTI hasta obtener las muestras del espacio conformacional asociadas al estado nativo y no nativo, tanto para la proteína natural como para la mutada.

El esquema de la figura 3.1 muestra el proceso seguido, desde la construcción de los sistemas natural y mutado, hasta la producción de las muestras de las estructuras asociadas a sus estados nativos y de estructuras no nativas. Todas las simulaciones de dinámica molecular (minimización, equilibración, calentamiento, enfriamiento, simulaciones dirigidas) se realizaron con los parámetros de campo de fuerza CHARMM22 [56] y tomando en cuenta todos los átomos de sistema (proteína, solvente e iones) explícitamente. Se utilizaron condiciones de frontera esféricas con un potencial armónico de constante 10 y exponente 2, el cual atrae a los átomos del sistema al centro de la esfera, una distancia de interrupción (switchdist) de 8 Å, una distancia de corte (cutoff) de 12 Å y distancia de búsqueda (pairlistdist) de 13.5 Å. Se consideraron todos los enlaces que involucran hidrógenos rígidos (no existen vibraciones) y se incluyeron los cálculos de todas las fuerzas entre átomos fijos. Para esto se utilizó la plataforma de simulación NAMD 2.6 para Windows [44].

3.1.1. Construcción de los Sistemas

Las coordenadas iniciales de la BPTI se extrajeron del archivo pdb 6pti, que corresponde a la estructura de la forma cristalina III del Inhibidor de Tripsina Pancreático Bovino, obtenida mediante Cristalografía de rayos X [28] y corresponden a una proteína monomérica que contiene 56 aminoácidos.

3.1.1.1. Estructuras

Debido a la naturaleza de la técnica con la que fue obtenida la estructura (cristalografía de rayos X), existen unas pocas moléculas de agua rodeando la proteína, además no están presentes los átomos de hidrógeno. Por medio del programa VMD, se removieron las moléculas de agua y se creó un nuevo archivo pdb que contiene únicamente la información correspondiente a las coordenadas de los átomos de la proteína. La adición de átomos de hidrógeno y la asignación de conectividad interatómica (enlaces químicos) se realizó con el paquete para construcción de estructuras "PSFGEN 1.0" que está incluido en VMD y el archivo de topología top_all22_prot_lipid.inp [56]. Para la construcción de la proteína mutada, se aplicó este procedimiento y posteriormente se realizaron las mutaciones, cambiando las cisteínas 5, 14 y 30 por alaninas (figura 3.2) con el paquete "MUTANT 1.2" incorporado en VMD. (Anexos A y B)

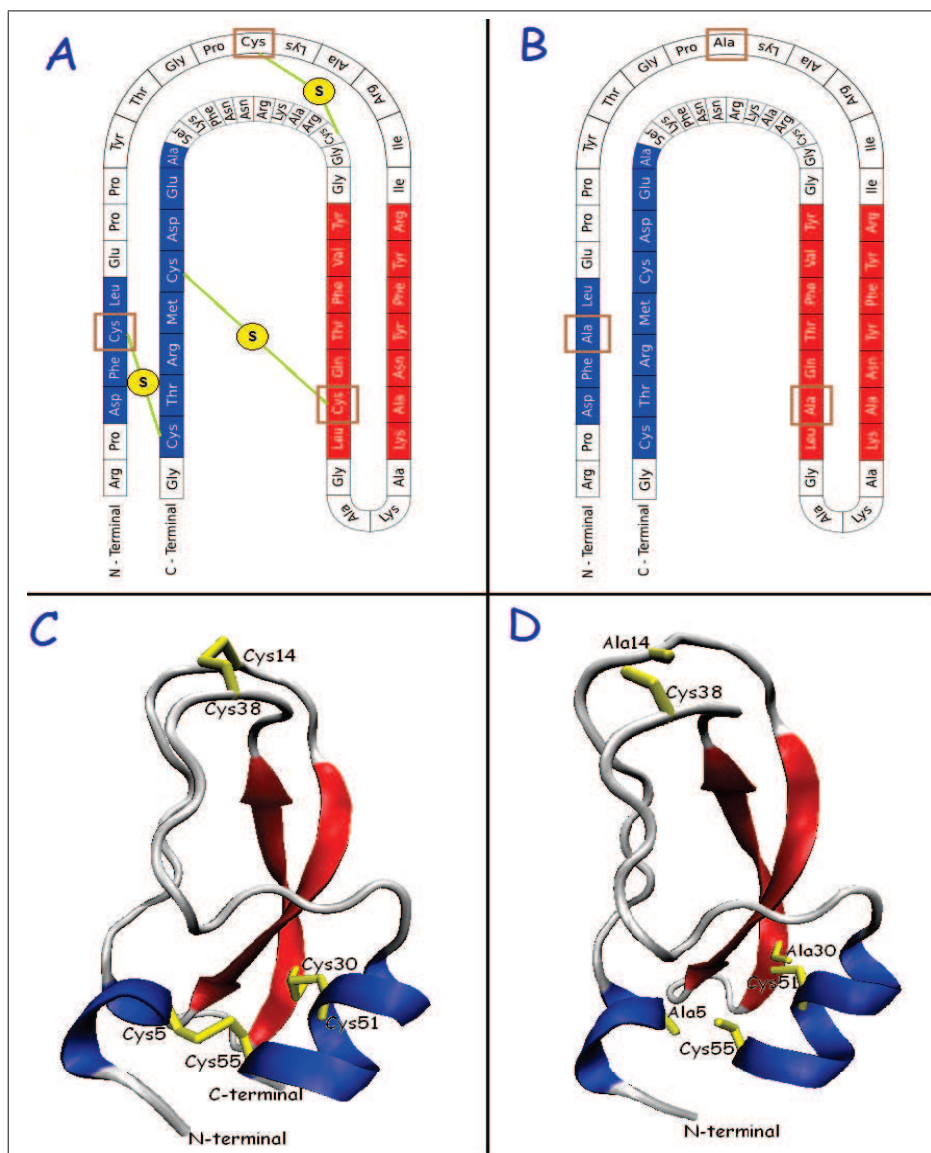


Figura 3.2: A, B: Estructura primaria de la BPTI natural y mutada respectivamente. C, D: Estructura espacial de las proteínas natural y mutada obtenidas a partir del archivo PDB 6pti. En rojo las hojas β antiparalelas, en azul las hlices.

3.1.1.2. Hidratación y neutralización

La BPTI es una proteína soluble en agua, para tomar en cuenta los efectos del solvente en la dinámica de la estructura. Se utilizó el modelo de solvente explícito, se colocó cada proteína en una esfera de agua, con la utilidad “SOLVATE” del VMD. Para evitar inestabilidades durante las simulaciones es necesario que los sistemas sean electricamente neutros [54]; el cálculo de la carga neta de los sistemas anteriores arrojó valores positivos (Tabla 3.1); por lo que fue necesario utilizar la utilidad “AUTOIONIZE” de VMD para neutralizarlos; se añadieron 4 iones de Sodio y 10 iones de Cloro aleatoriamente en el solvente (Tabla 3.1). Obteniendo

de este modo dos sistemas, conformados por la proteína, el solvente (agua) y los iones (Tabla 3.2); los radios de las esferas de agua son de 22.270 Å y 22.274 Å para la natural y mutada respectivamente. La figura 3.3 muestra el sistema natural preparado de esta forma.

Sistema	Carga Neta del sistema hidratado [e]	Carga Neta del sistema hidratado y neutralizado [e]
BPTI natural	6.000000708	6.62E-07
BPTI mutada	6.000000656	5.08E-07

Tabla 3.1: Cargas netas de los sistemas moleculares de las estructuras no mutada y mutada, antes y después de la neutralización (adición de iones), calculadas en el programa VMD.

Sistema	Número de residuos				Número de átomos				Número de enlaces	Núm. ángulos entre enlaces	Número de torsiones
	Agua	Proteína	Iones	Total	Agua	Proteína	Iones	Total			
Natural	1190	56	14	1260	3570	875	14	4459	3269	2780	2329
Mutado	1192	56	14	1262	3576	878	14	4468	3273	2785	2338

Tabla 3.2: Características generales de los sistemas obtenidos con las estructuras natural y mutada.

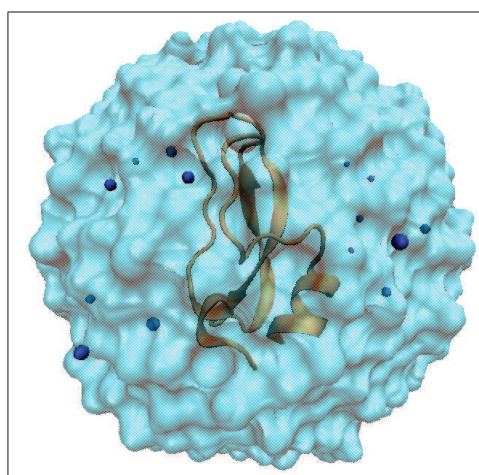


Figura 3.3: Sistema molecular de la proteína natural. Representación “NewCartoon” de la proteína, “Surf” del solvente y “CPK” de los iones.

3.1.2. Minimización

Los átomos de hidrógeno añadidos así como las moléculas de agua se colocan en posiciones aleatorias, por lo que un protocolo de minimización energética es necesario, a fin de que la proteína adopte la estructura tridimensional óptima. Se realizó la minimización de los sistemas preparados, corriendo 10000 pasos de 1fs con el método del Gradiente Conjugado para minimización de energía y salvando las coordenadas de los átomos del sistema cada 1000 pasos en un archivo binario, las energías del sistema son guardadas cada 100 pasos en un archivo de texto (Anexo C).

Durante este proceso, el radio de giro disminuye de 11.245 Å a 11.217 Å para la natural y de 11.242 Å a 11.133 Å para la mutada, debido que la estructura busca tener una geometría óptima más compacta. La d_{rms} respecto a la estructura cristalina respectiva aumenta hasta llegar a 0.533 Å y 0.619 Å para la natural y mutada (Figura 3.5), debido a que las estructuras cambian en busca del mínimo local más cercano a las estructuras iniciales, los valores son pequeños. La energía interna total de la proteína natural varió de -2.69 attojoules (aJ) hasta -8.15 aJ y la de la mutada de -2.21 aJ a -8.41 aJ; como muestra la figura 3.5, la energía de ambos sistemas tiende hacia un valor constante a partir del paso 2000, por lo que los 10000 pasos ejecutados son suficientes como para encontrar el mínimo local más cercano.

Para realizar una comparación cuantitativa entre las estructuras minimizadas de las proteínas natural y mutada se contruyó un gráfico de distancias rms entre ambas. Luego de alinear las estructuras natural y mutada, se calcula la distancia rms de grupos de 7 residuos y se asigna esta d_{rms} al residuo intermedio del grupo. De este modo se pueden identificar las regiones en la cuales se acentúan las diferencias entre ambas estructuras. La figura 3.4-A presenta tres picos entre los residuos Pro8 y Arg17, Tyr23 y Cys30/Ala30, y Val34 y Asn43 que corresponden a las regiones donde radica la diferencia entre las estructuras natural y mutada, como se muestra en la figura 3.4-B); en dichas regiones encontramos, en la proteína natural, cisteínas formando enlaces disulfuro. La distancia rms total entre ambas es de 0.532 Å, lo que indica que son estructuras bastante similares, su diferencia se da por efecto de la presencia o no de los enlaces disulfuro y resalta en la regiones donde se encuentran estos.

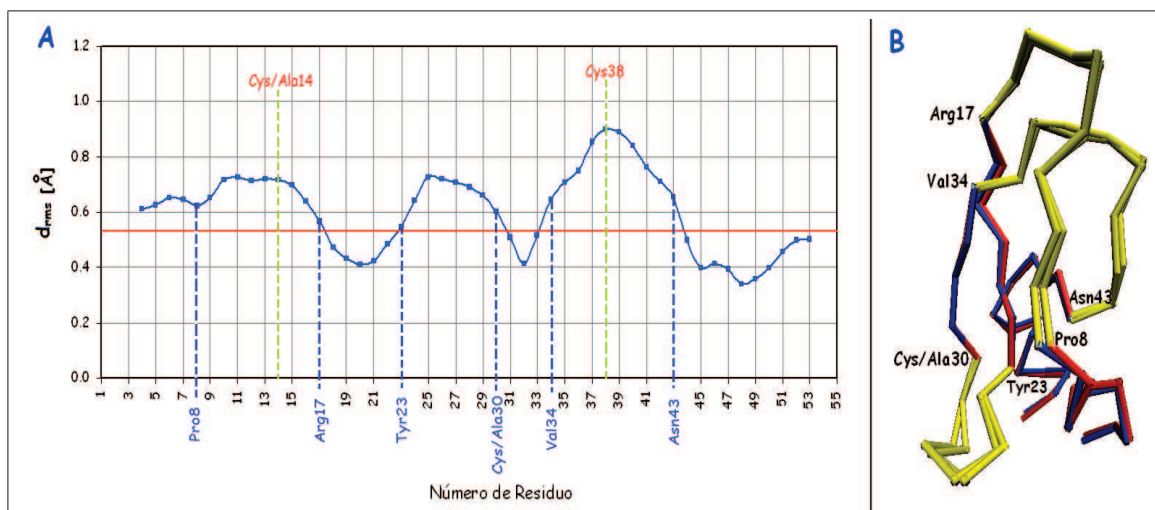


Figura 3.4: **A:** Gráfico de distancias rms entre las estructuras minimizadas de las proteínas natural y mutante, la línea roja representa la d_{rms} total entre las dos estructuras. **B:** Representación “Trace” de las estructuras comparadas: en rojo la natural, azul la mutada y amarillo las regiones con mayor diferencia de d_{rms} .

3.1.3. Calentamiento hasta 300 K

Para el calentamiento se partió de los sistemas producidos luego de la minimización de los sistemas a 0K; aumentando su temperatura linealmente hasta 300K en 20000 pasos de 1 fs; la reasignación aleatoria de velocidades para los átomos del sistema se realizó cada 25 pasos, de modo que la temperatura total del sistema se incrementó en 0.375 K cada 25 pasos, desde 0K hasta 300K. (Anexo D)

El calentamiento de los sistemas produce una variación de las estructuras producto del aumento de la energía interna de los sistemas. Como muestra la figura 3.5; sin embargo, la variación de los parámetros radio de giro y d_{rms} es bastante pequeña, es decir, las proteínas de ambos sistemas no cambiaron drásticamente de forma pero energéticamente sí sufrieron un aumento muy grande de energía, producto del aumento de temperatura. La figura 3.5 muestra la energía, radio de giro y la distancia rms de las proteínas natural y mutada en función del número de pasos de simulación. Durante el calentamiento, el radio de giro aumentó de 11.217 Å a 11.221 Å para la nativa y de 11.133 Å a 11.216 Å para la mutada; la distancia rms respecto a la estructura cristalina llegó a 0.626 Å y 0.966 Å para la natural y mutada respectivamente; la energía de la proteína natural aumentó hasta -3.94 aJ (-952.17 kT) y la de la mutada hasta -4.14 aJ (-1000.24 kT).

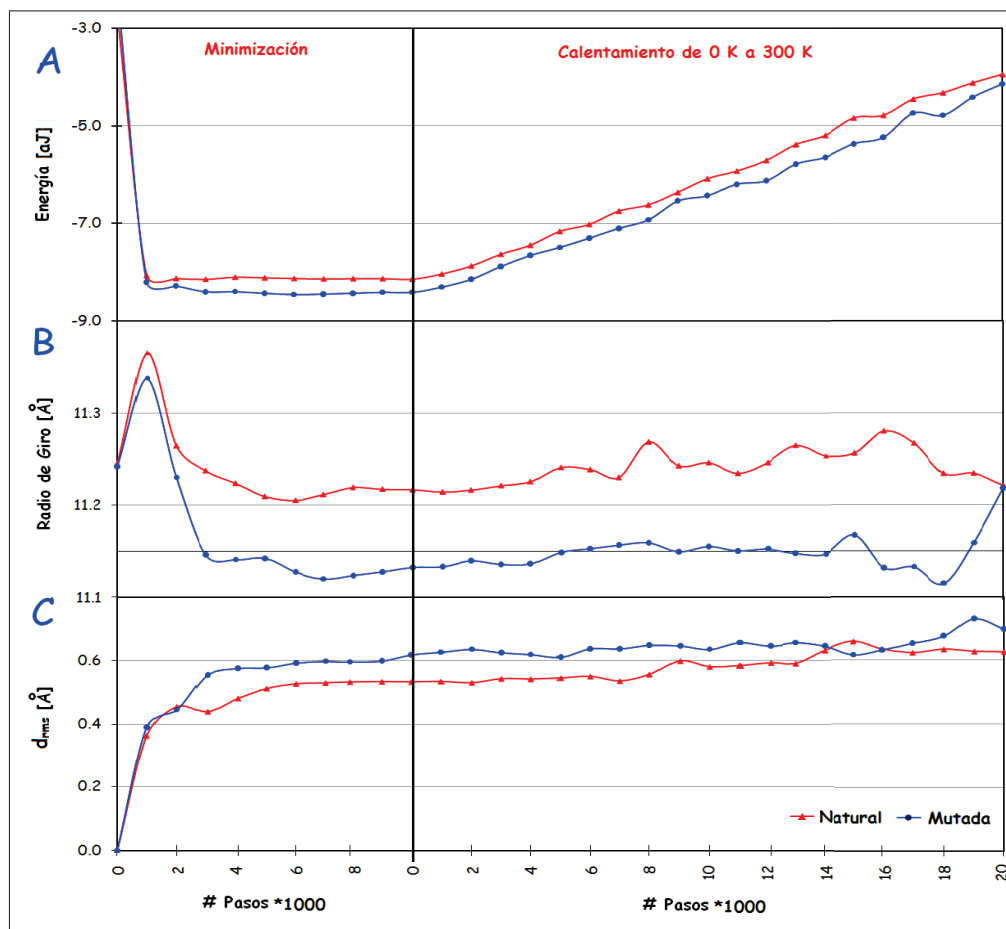


Figura 3.5: En el panel izquierdo la minimización y en el derecho el calentamiento de 0 K a 300 K de los sistemas moleculares natural y mutado. Evolución de **A:** la energía, **B:** radio de giro y **C:** d_{rms} de la proteína de cada sistema en función del número de pasos de simulación.

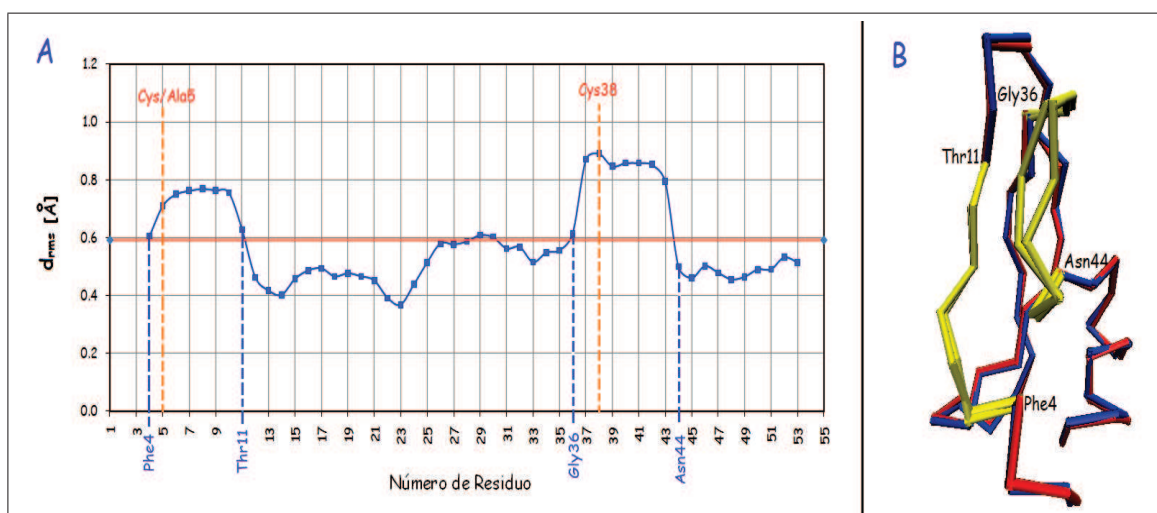


Figura 3.6: **A:** Gráfico de distancias rms entre las estructuras calentadas a 300 K de las proteínas natural y mutada, la línea roja representa la d_{rms} total entre las dos estructuras. **B:** Representación "Trace" de las proteínas: en rojo la natural, en azul la mutada y en amarillo las regiones con mayor diferencia de d_{rms} .

Para identificar las diferencias entre las estructuras de las proteínas natural y mutada luego del calentamiento hasta 300K, se realizó un gráfico de distancias rms como el descrito en el apartado 3.1.2. En la figura 3.6-A, podemos identificar la presencia de dos picos, entre los residuos Ph4 y Thr11, y Gly36 y Asn44; lo que nos indica que después del calentamiento se acentuó la diferencia en la región donde se encuentra el residuo 38, además se marcó una diferencia notable en la región cercana al residuo 5, ambas corresponden a cisteínas en la estructura natural (figura 3.6). La distancia rms total entre ambas es de 0.591 Å, es decir, el calentamiento acentuó la diferencia global entre las estructuras y particularmente en las regiones donde se encuentran las cisteínas antes mencionadas.

3.1.4. Equilibración

Los sistemas resultantes del calentamiento a 300K se equilibraron por 1ns, con los parámetros generales mencionados en el apartado 3.1. Los 1000000 pasos de 1 fs se realizaron a temperatura constante, utilizando la dinámica de Langevin a 300 K con un coeficiente de fricción $\gamma = 5ps^{-1}$ (ec 2.17) en todos los átomos del sistema excepto los hidrógenos. (Anexo E).

Se analizaron la energía, radios de giro y distancia rms de las proteínas (nativa y mutada) en función del tiempo de simulación (figura 3.7). A partir de los 800 ps de equilibración el valor del radio de giro tanto de la estructura no mutada como de la mutada, se torna estable, tomando valores entre 11.2 y 11.4 Å; la distancia rms respecto la respectiva estructura calentada a 300K, incrementa lentamente hasta los 800 ps, donde los valores se tornan estables entre 1.2 y 1.4 Å para la natural y entre 0.85 y 1.0 Å para la mutada; la energía en cambio tiende a estabilizarse a partir de 750 ps tomando valores entre los -4.2 y -3.5 aJ (-1014.49 y -845.91 kT) para la natural y -4.6 y -4.0 aJ (-1111.11 y -966.18 kT) para la mutada. Por ende, se puede considerar que tanto el sistema de la proteína nativa como el de la mutada llegan al equilibrio después de aproximadamente 800 ps.

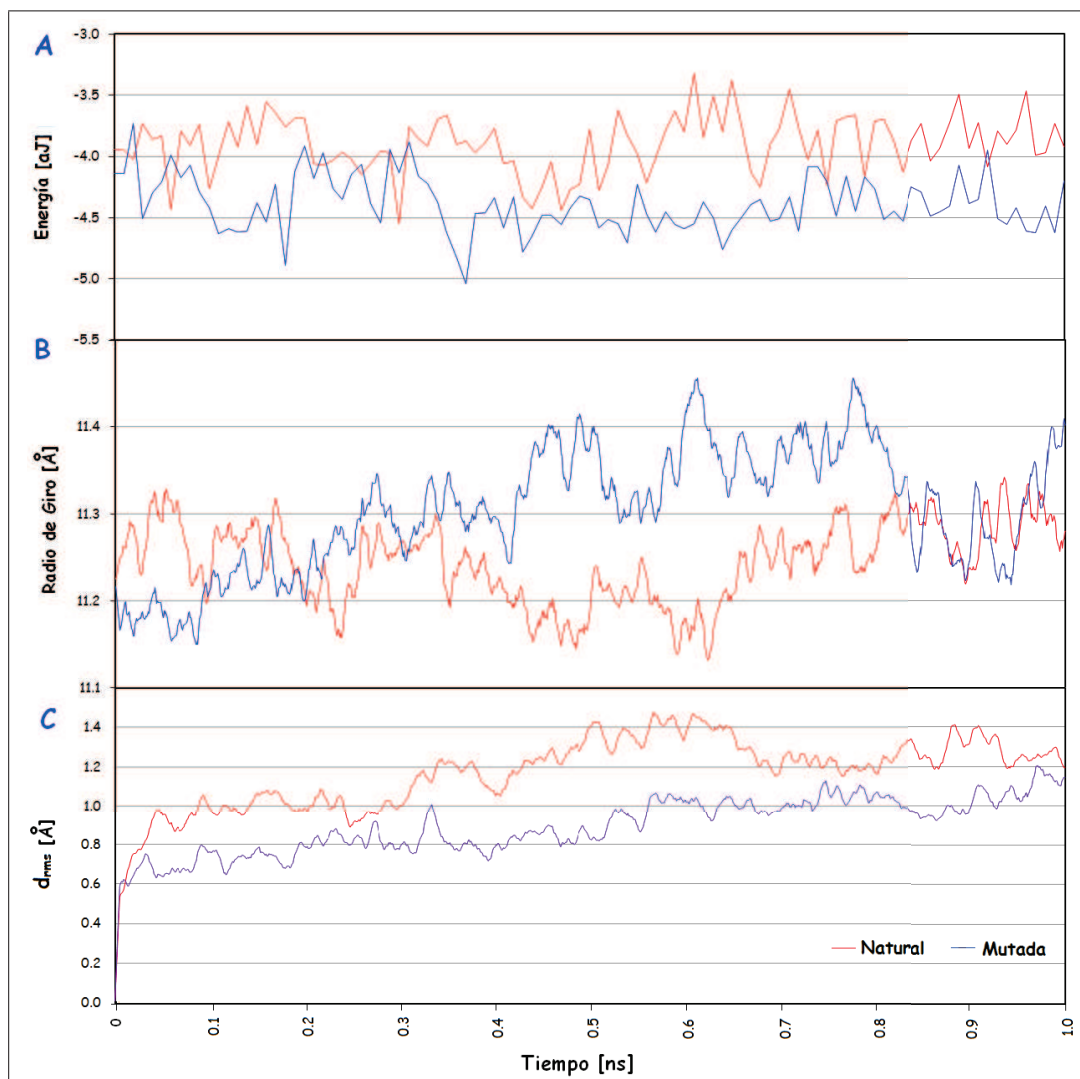


Figura 3.7: Equilibración a 300 K de los sistemas moleculares natural y mutada. Evolución de la energía, radio de giro y d_{rms} de la proteína de cada sistema en función del número del tiempo de simulación. **A:** Energía de la proteína correspondiente a 100 estructuras generadas cada 10 ps. **B:** El radio de giro fue calculado de la proteína. **C:** d_{rms} del esqueleto de la proteína respecto con respecto a la estructura resultante del proceso del calentamiento a 300 K.

3.1.5. Producción: Muestreo y optimización conformacional

La figura 3.1 muestra esquemáticamente el procedimiento realizado para tres grupos de estructuras para cada uno de los sistemas, natural y mutado. El primero se obtuvo mediante simulaciones de Dinámica Molecular a 300K, para el segundo se utilizó simulaciones a altas temperaturas y el tercero se obtuvo utilizando simulaciones dirigidas; la optimización de las estructuras de los grupos se la realizó mediante protocolos de minimización de energía. Los tres grupos de muestras que se generaron terminaron conformados de la siguiente manera: uno por estructuras asociadas al estado nativo (mínimo global), otro por estructuras no

nativas (asociadas a mínimos locales) y finalmente uno con estructuras asociadas a transiciones desde los mínimos locales al mínimo global. A continuación se especifican dichos procedimientos, los resultados y características generales de las muestras obtenidas a partir de los sistemas natural y mutado.

3.1.5.1. Estructuras asociadas al estado nativo

Las estructuras asociadas al estado nativo se obtuvieron mediante simulaciones de Dinámica Molecular a 300 K. Las simulaciones se realizaron a partir de los sistemas producidos después de un nanosegundo de equilibración y utilizando los parámetros generales descritos en el apartado 3.1. Se simularon 1000000 pasos de 1 fs cada uno utilizando la dinámica de Langevin con los parámetros mencionados en el apartado 3.1.3. (protocolo de equilibración) y los resultados (coordenadas) fueron salvados cada 1000 pasos en un archivo binario.

Para la conformación del grupo se tomaron 100 estructuras, escogidas de la trayectoria original, una cada 10 ps y fueron numeradas en orden temporal de generación en la simulación. Como se mencionó en el apartado 2.2.2. es necesario optimizar la muestra, es decir, desplazar cada estructura al mínimo local más cercano; este proceso se lo realizó con el programa NAMD, utilizando el protocolo de Minimización descrito en el apartado 3.2.1. para cada estructura.

La figura 3.8, muestra cómo el proceso de optimización cambia los parámetros que caracterizan una estructura como son el radio de giro, distancia rms y su energía. Los valores del radio de giro de las estructuras luego de la minimización disminuyen alrededor de 0.1 Å tanto para las estructuras de muestra de la natural como para las de la mutada, la d_{rms} , luego de este proceso casi no varían respecto a las estructuras obtenidas directamente de la Dinámica Molecular; es decir, las propiedades que caracterizan la forma general de la estructura no cambian sustancialmente con la minimización, en cambio la energía de las estructuras varía drásticamente luego de la minimización; los valores disminuyen alrededor de 4.2 aJ (1014.49 kT) para cada estructura de ambas muestras.

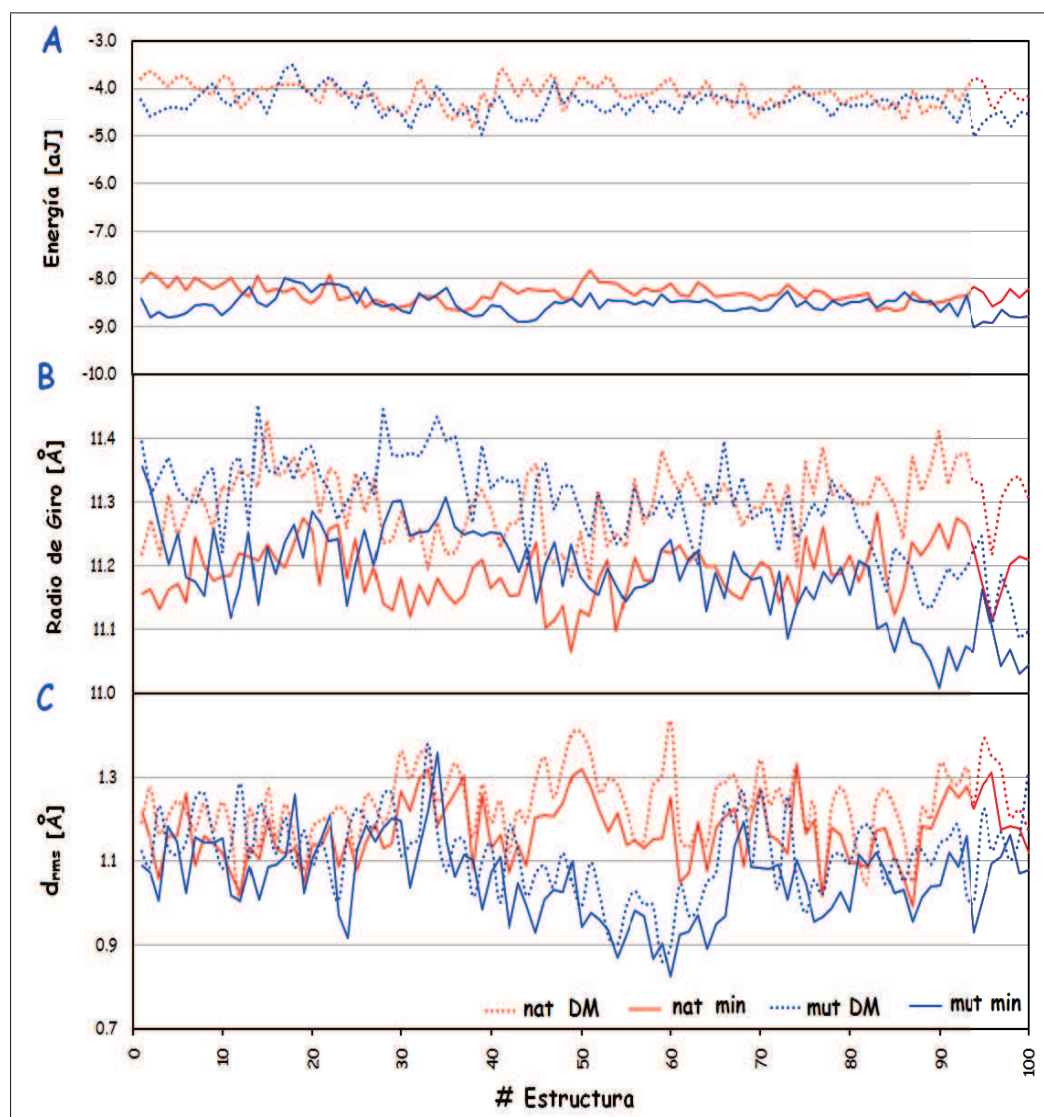


Figura 3.8: **A:** Energía de las estructuras de la muestra obtenida a partir de los sistemas de la proteína natural y mutada directamente de la Dinámica Molecular (nat MD y mut MD) y luego de ser minimizadas (nat min y mut min). **B:** Radio de giro de las estructuras. **C:** d_{rms} del esqueleto (backbone) de las estructuras respecto a la estructura calentada a 300K para las obtenidas directamente de la Dinámica Molecular y con la calentada a 300K minimizada para las estructuras optimizadas.

3.1.5.2. Estructuras no nativas

Luego de la equilibración de 1ns a 300 K se siguieron los siguientes pasos: Calentamiento a 1000 K, equilibración a 1000 K durante 10ns, selección de 10 estructuras, enfriamiento a 300 K de cada una, equilibración a 300K durante 2ns de cada estructura y finalmente optimización de las estructuras producidas.

Calentamiento hasta 1000 K: Partiendo de los sistemas resultantes la equilibración a 300K durante un nanosegundo, se calentó cada sistema hasta 1000K utilizando un protocolo similar al expuesto en el apartado 3.1.2. Las condiciones de frontera, restricciones, parámetros del campo de fuerza y parámetros de integración son los mismos que los utilizados para el calentamiento de 0K a 300K y el aumento de la temperatura del sistema se lo realizó en 50000 pasos de 1fs, reasignando las velocidades de los átomos cada 25 pasos, es decir, realizando un incremento de temperatura de de 0.35K cada ciclo, desde 300 K hasta 1000 K. (Anexo F)

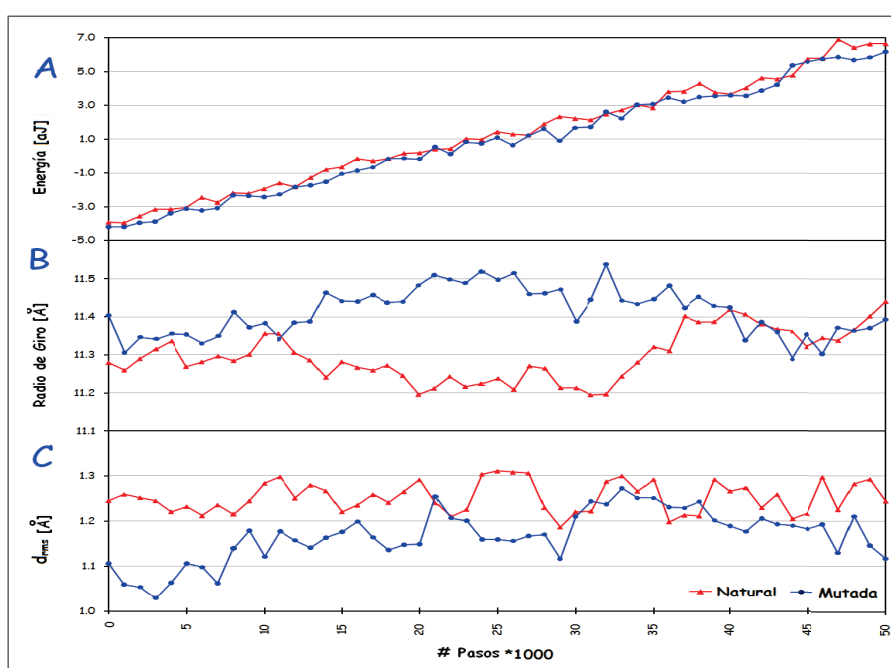


Figura 3.9: Calentamiento de 300 K a 1000 K de los sistemas moleculares natural y mutado. Evolución de **A:** la energía, **B:** radio de giro y **C:** d_{rms} de la proteína de cada sistema en función del número de pasos de simulación.

Como muestra la figura 3.9, pese a la cantidad de energía que el sistema adquiere por aumento de temperatura, los parámetros característicos de la proteína no sufren mayores cambios, el radio de giro está entre 11.2-11.45 Å y 11.3-11.55 Å, y la distancia rms respecto a la estructura calentada a 300K fluctúa entre 1.2-1.3 Å y 1.0-1.25 Å para los sistemas de las estructuras natural y mutada respectivamente. Debido a la rapidez del proceso de calentamiento la proteína no cambió considerablemente de forma, pese al cambio de energía. La energía aumenta alrededor de 10.6 aJ (2560.39 kT) en ambos sistemas, llegando hasta 6.65 aJ (1605.31 kT) para la natural y 6.17 aJ (1489.37 kT) para la mutada.

Equilibración a 1000K: En esta fase se producen las estructuras, que han logrado superar las barreras de potencial gracias al aumento de temperatura, mediante una simulación de equilibración del sistema utilizando un ensamble NVT a 1000K. Se realizó la simulación, para el sistema natural y para el mutado. Partiendo de los sistemas generados luego del calentamiento hasta 1000K, se procedieron a realizar 10000000 pasos de 1fs (10ns) de equilibración a 1000K utilizando la dinámica de Langevin; las condiciones de frontera, restricciones y demás parámetros de la simulación son los mismos que los utilizados en el protocolo de equilibración del sistema descrito en el apartado 3.1.3. Los resultados de las coordenadas fueron almacenados cada 1000 pasos, obteniéndose de este modo 10000 estructuras.

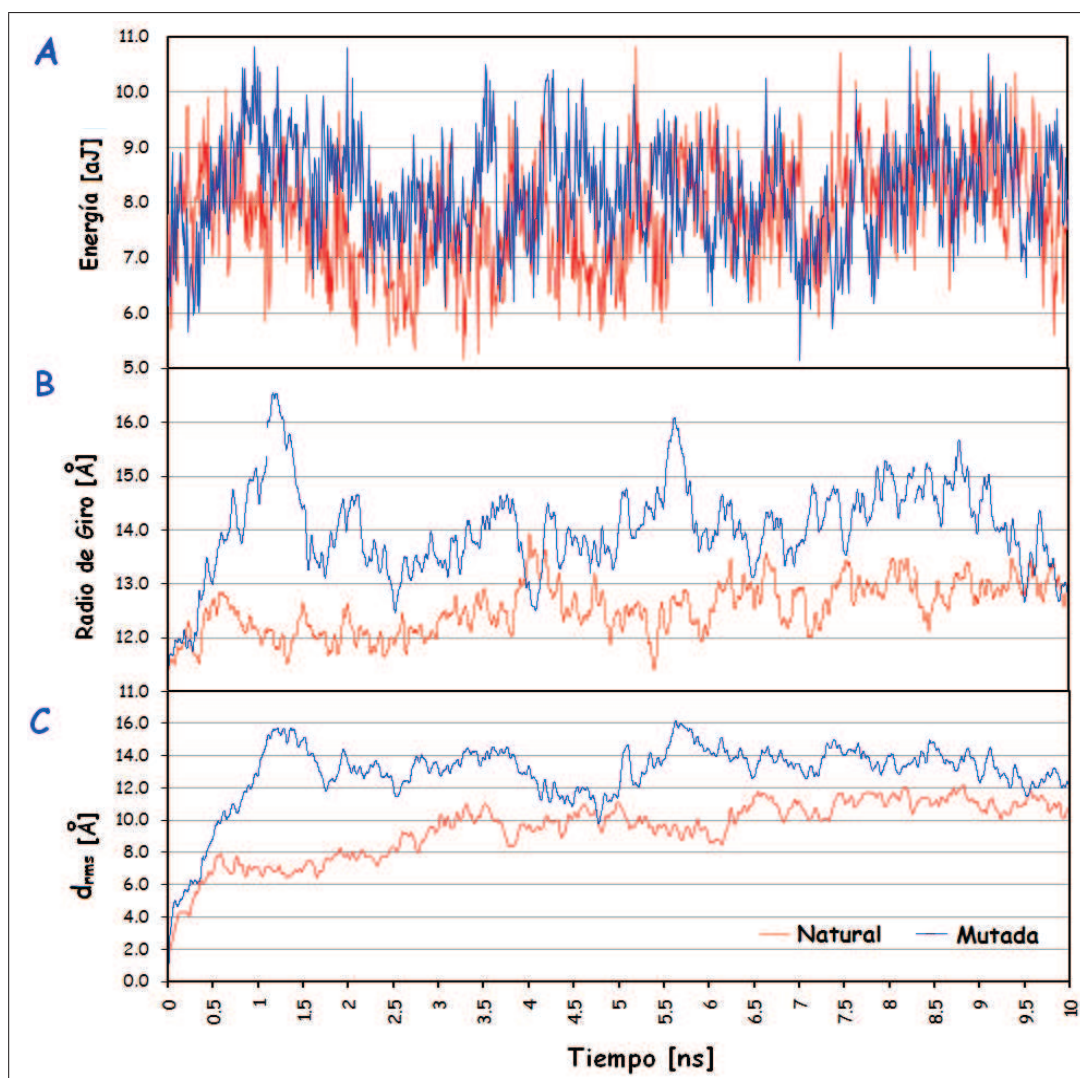


Figura 3.10: Equilibración a 1000K de los sistemas moleculares natural y mutado. Evolución de **A:** la energía, **B:** radio de giro y **C:** d_{rms} de la proteína de cada sistema en función del tiempo de simulación. El cálculo tanto el radio de giro como la RMSD se lo hizo de 10000 estructuras (generadas cada 1 ps de equilibración). La energía presentada en la gráfica corresponde a 1000 estructuras generadas cada 10ps.

Como muestra la figura 3.10, tanto el radio de giro como la distancia rms sufren drásticos cambios (lo que no se dio en ninguno de los procesos anteriores), ésto sugiere que la proteína sufre cambios conformacionales, superando barreras de energía, por lo que es necesario un análisis más detallado de las estructuras. Los valores de energía fluctúan entre las 5.1 y 10.9 aJ (1231.88 y 2632.85 kT).

Selección de estructuras: Las estructuras generadas a altas temperatura en sí no son conformaciones accesibles a la BPTI a 300K, necesitan ser enfriadas a esta temperatura. Para cumplir con los objetivos del presente trabajo se seleccionaron 10 estructuras para ser enfriadas; en principio cada una de éstas está asociada a una conformación accesible a 300K. Las estructuras fueron seleccionadas cada 1ns de la trayectoria.

Enfriamiento a 300K: Las diez estructuras se enfriaron utilizando un protocolo similar al usado para el calentamiento; con las mismas condiciones de frontera, restricciones, parámetros del campo de fuerza y de integración que en las anteriores simulaciones, se realizó la disminución de la temperatura de los sistemas en 50000 pasos de 1fs, reasignando las velocidades de los átomos cada 25 pasos, es decir, disminuyendo la temperatura 0.35K cada ciclo, desde 1000 K hasta 300 K. (Anexo G)

Equilibración a 300K: En principio, cada una de las diez estructuras obtenida a altas temperaturas y enfriada a 300 K está asociada a una conformación accesible a la BPTI a esta temperatura. Para obtener una muestra que describa de mejor manera cada una de estas regiones, se realizó primero la equilibración de cada estructura durante 1ns. Las simulaciones de equilibración durante 1ns a 300 K, se realizaron utilizando la dinámica de Langevin mediante el protocolo de equilibración utilizado en el apartado 3.1.3.

Producción de estructuras: Las estructuras obtenidas durante el primer nanosegundo de equilibración corresponden a transiciones a un estado conformacional estable (equilibrio), por lo que fue necesario realizar un segundo nanosegundo de equilibración para cada estructura con el fin de obtener muestras en la que se puedan identificar conformaciones estables de la proteína asociadas a mínimos locales. Las simulaciones de equilibración durante 1ns a 300K, se realizaron utilizando la dinámica de Langevin mediante el protocolo de equilibración

utilizado en el apartado 3.1.3. Se extrajeron 100 estructuras de cada una de las 10 trayectorias de Dinámica Molecular de 1ns, una estructura cada 10ps de simulación; obteniéndose de esta forma una muestra de 100 estructuras alrededor de cada mínimo. La muestra i -ésima tomó como punto inicial la estructura del i -ésimo nanosegundo de la equilibración a 1000 K. De este modo se obtuvo un grupo de muestras de 1000 estructuras a 300K.

Optimización de las estructuras: Como antes, es necesario realizar la optimización de las estructuras de cada muestra. Este proceso se realizó mediante la minimización de energía de las estructuras producidas en la Dinámica Molecular, utilizando el protocolo del apartado 3.1.1.3. con los mismos parámetros de de fuerza y de integración de las simulaciones anteriores.

Número de muestra	Energía promedio [aJ]				Radio de giro promedio [Å]				d_{rms} promedio [Å]			
	Natural		Mutada		Natural		Mutada		Natural		Mutada	
	Producción	Optimización	Producción	Optimización	Producción	Optimización	Producción	Optimización	Producción	Optimización	Producción	Optimización
1	-2.60	-6.65	-0.90	-4.91	11.95	11.81	13.67	13.48	6.89	6.82	11.86	11.70
2	-2.78	-6.89	-1.14	-5.18	12.25	12.11	13.16	12.97	7.52	7.44	13.39	13.20
3	-3.94	-8.07	-2.49	-6.56	11.97	11.83	13.54	13.35	9.87	9.79	13.70	13.46
4	-2.08	-6.10	-3.57	-7.68	13.10	12.94	12.56	12.41	9.50	9.43	13.02	12.85
5	-3.25	-7.35	-1.64	-5.70	12.88	12.71	13.33	13.14	10.80	10.68	11.55	11.37
6	-2.54	-6.67	-2.95	-7.09	12.07	11.90	13.64	13.45	8.78	8.68	14.26	14.04
7	-1.50	-5.57	-3.36	-7.47	12.36	12.19	14.03	13.82	11.13	11.03	14.40	14.17
8	-1.02	-5.12	-2.33	-6.42	13.46	13.29	13.99	13.77	11.37	11.26	13.80	13.62
9	-2.41	-6.47	-2.75	-6.78	12.22	12.06	13.94	13.73	10.31	10.19	12.41	12.24
10	-2.25	-6.30	-3.05	-7.14	12.32	12.16	12.29	12.12	10.34	10.20	12.14	11.99

Tabla 3.3: Producción y optimización de las estructuras de las muestras no nativas de los sistemas natural y mutado. Valores promedio de los parámetros energía, radio de giro y d_{rms} de las estructuras de cada una de las muestras.

La tabla 3.3 nos muestra cómo la minimización de energía no afecta mayormente la forma de la estructura; el radio de giro disminuye aproximadamente 0.16 Å y 0.19 Å, y la d_{rms} , 0.1 Å y 0.2 Å, para la natural y mutada respectivamente luego de la minimización energética; por tanto, luego de la refinación de la muestra vamos a tener un conjunto de estructuras similares a las iniciales en forma, pero con distinta energía. Este proceso hace que la energía de la proteína cambie drásticamente,

disminuyendo alrededor de 4.1 aJ (990.34 kT) en cada estructura tanto para las del sistema natural como para las del mutado. Tanto la tabla 3.3 presenta, en general, valores promedio de radio de giro y d_{rms} menores para la estructura natural que para la mutada, lo que indica que las estructuras generadas a partir de la proteína natural son más estables que las obtenidas a partir de la mutada, producto de la presencia de los enlaces disulfuro, los que condicionan los cambios en la forma de la estructura proteína, haciendo más difícil su deformación. La figura 3.11 corresponde a la muestra 1 de los sistemas natural y mutado.

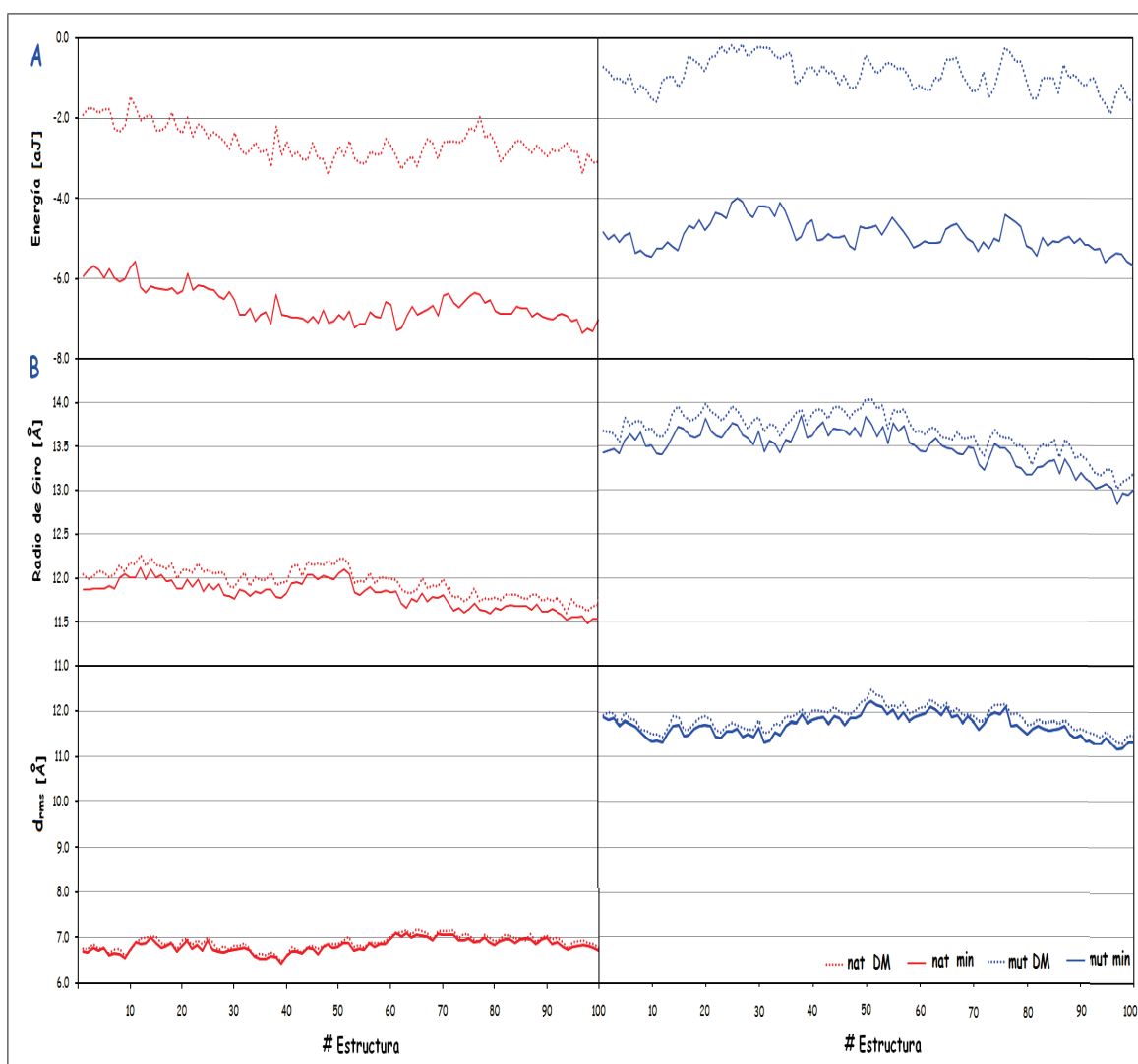


Figura 3.11: Producción de las estructuras de la muestra 1 de los sistemas moleculares natural y mutado. Evolución de **A:** la energía, **B:** radio de giro y **C:** d_{rms} de la proteína de cada sistema en función del tiempo de simulación. La d_{rms} se calculó respecto a la estructura calentada a 300K para las obtenidas directamente de la Dinámica Molecular y a la calentada a 300K minimizada para las estructuras optimizadas..

3.1.5.3. Estructuras asociadas a transiciones desde los mínimos locales al mínimo global

Los grupos de estructuras de la muestra a altas temperaturas corresponden a mínimos locales de la superficie del Entorno Energético, por lo que proveen información acerca de dichas regiones, sin embargo, la información de las barreras de energía no está contenida en dicha muestra; las barreras energéticas corresponden simplemente a regiones donde no están presentes estructuras de la muestra. Con el fin de obtener información acerca de estas barreras y para obtener un mapeo más completo del Entorno Energético de la BPTI, se utilizaron simulaciones de Dinámica Molecular Dirigidas; esta técnica nos permite encontrar la secuencia de estructuras intermedias entre dos estados conocidos.

Se realizaron diez simulaciones de este tipo, partiendo de la estructura producto de la equilibración (1ns) minimizada de cada muestra hasta la estructura obtenida de la equilibración directa (1ns) a 300K minimizada. Se usó temperatura constante (300K) empleando la dinámica de Langevin y los parámetros de campo de fuerza y de integración de las simulaciones anteriores; debido a que en este tipo de simulaciones no se pueden utilizar otro tipo de restricciones, no se colocaron en el protocolo ni restricciones de átomos rígidos ni condiciones de frontera esféricas. Cada simulación dirigida se realizó en 100000 pasos de 1fs, con una constante elástica para la fuerza dirigida de $200 \text{ kcal/mol}\text{\AA}^2$ (138870.43 pN/nm); la d_{rms} inicial fue calculada por NAMD de las coordenadas iniciales y la final es 0. (Anexo H). Para completar la formación de la muestra de las estructuras producto de la Dinámica Dirigida, es necesario realizar una minimización de energía de cada estructura, para ello se utilizó el protocolo del apartado 3.2.1. para cada una de las estructuras.

La figura 3.12 muestra de la energía, radio de giro y distancia rms de las estructuras desde la inicial respectiva hasta llegar a la estructura objetivo. Estas estructuras de la simulación describen una barrera de energía, lo que implica un transición entre dos estados asociados a mínimos en Entorno Energético. En general no se pueden comparar directamente los resultados obtenidos de cada simulación dirigida de los sistema natural y mutado, ya que las estructuras iniciales de cada uno son esencialmente diferentes, como muestra la figura 3.12. Sin embargo al hacer una comparación global de todas, se encontró que las barreras de energía asociadas son ligeramente mayores para la natural que para la mutada (tabla 4.1).

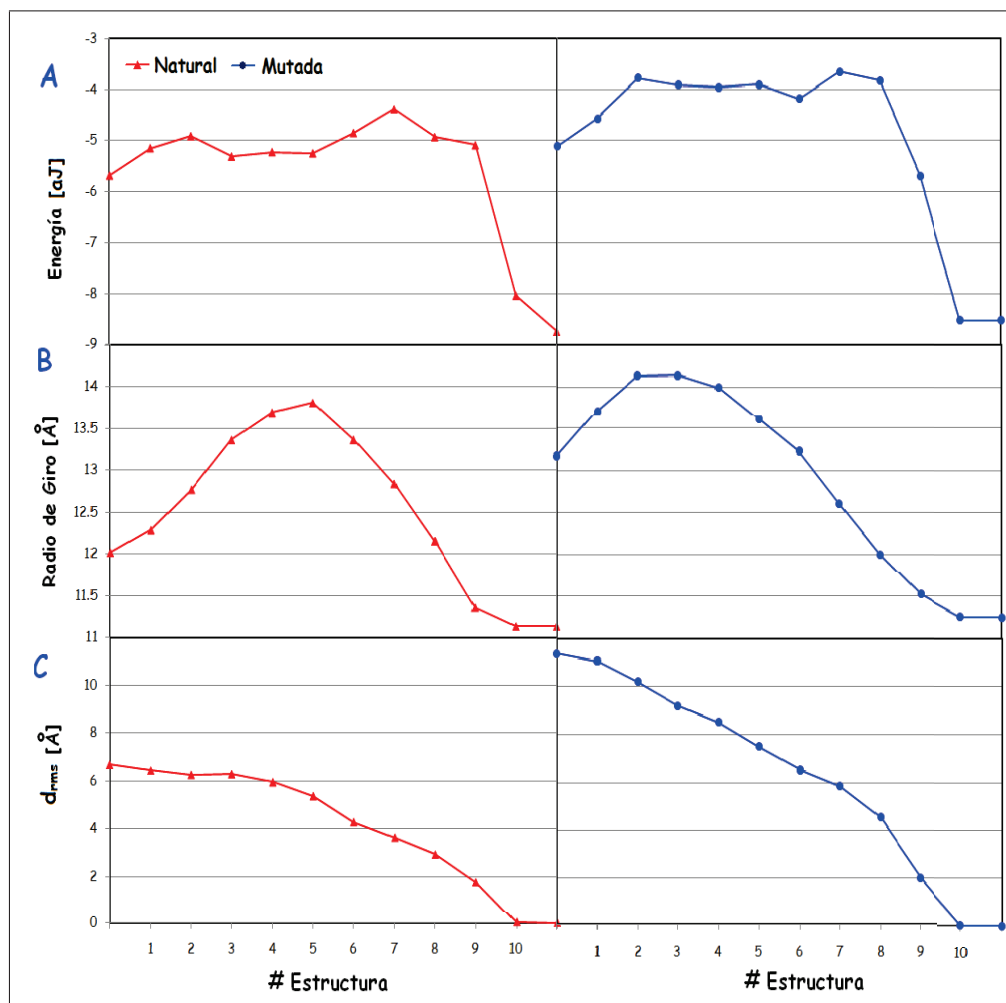


Figura 3.12: Simulación dirigida desde la muestra 1. Evolución de la energía (A), radio de giro (B) y d_{rms} (C) de la proteína natural (izquierda) y mutada (derecha) durante las simulaciones Dirigidas. La i -ésima simulación (TMD i) inicia con la estructura producida luego de la equilibración durante 1ns de la i -ésima muestra hasta la estructura de la equilibración de 1ns directa a 300K minimizada.

3.2. Conformaciones asociadas al estado nativo

Las estructuras resultantes de las simulaciones de Dinámica Molecular a 300 K fueron organizadas mediante un análisis de grupos jerárquico que permitió identificar tanto los estados conformacionales como la organización jerárquica de éstos dentro de otros estados definidos por el incremento de la distancia criterio [38]. De esta forma se identificaron las características generales de las conformaciones de la BPTI en la vecindad del estado nativo, así como también se determinaron sus distancias rms, radios de giro y energías. Para la agrupación de las estructuras asociadas al mínimo global se utilizó como base la matriz disimilitud Δ descrita en el apartado 2.2.3.1. Para su cálculo, primero se

superpusieron todas las estructuras de modo que la distancia rms entre cada par de estructuras sea mínima, utilizando la utilidad “alinear” del VDM; se tomó en cuenta únicamente las carbonos alfa de las estructuras. El mismo estudio fue realizado con la proteína mutada para examinar cómo las mutaciones que alteran los enlaces disulfuro afectan en los estados conformacionales de la BPTI. Todos los análisis estadísticos se realizaron utilizando el programa DARWin5[68].

3.2.1. BPTI natural

El grupo de 100 estructuras asociadas al mínimo global provenientes de 1ns de Dinámica Molecular a 300K fue analizado con el fin de confrontar dichos resultados con los obtenidos por Troyer y Cohen [38]. La figura 3.13 es una representación gráfica de la matriz de disimilitud , definida por las distancias rms de los carbonos alfa de las estructuras. Las zonas azules y verdes corresponden a grupos de estructuras cuya d_{rms} mutua es bajo ($< 0.6 \text{ \AA}$). Se presentan regiones considerables de estructuras similares (cuya d_{rms} sea menor que 0.60 \AA); esta evaluación visual se constata formalmente con el análisis de grupos jerárquico cuyo resultado se presenta en la figura 3.14.

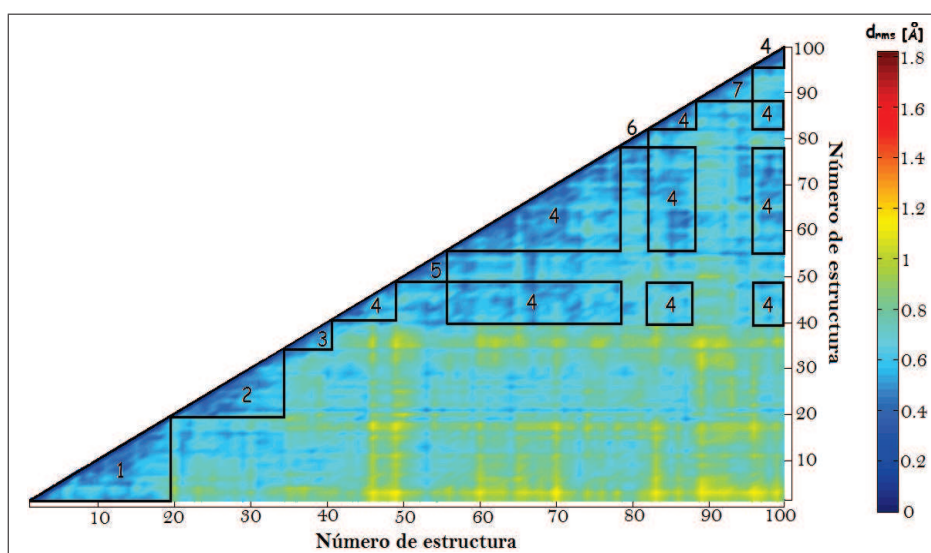


Figura 3.13: Representación gráfica de la matriz de disimilitud de las estructuras asociadas al estado nativo de la BPTI natural. Identificación de los 7 grupos obtenidos en el análisis de grupos; las regiones representadas corresponde a los 7 grupos formados tomando como distancia criterio 0.6 \AA .

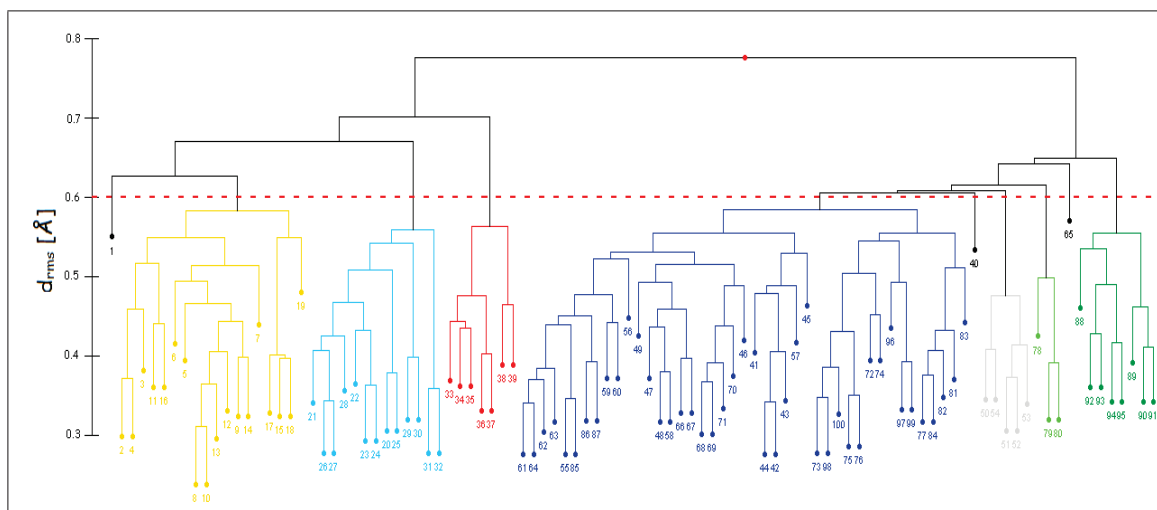


Figura 3.14: Dendrograma del análisis de grupos de las estructuras asociadas al estado nativo de la BPTI natural. Resultado del análisis de grupos jerárquico aglomerativo con el algoritmo del grupo promedio. La línea roja punteada a una d_{rms} de 0.60 Å separa las estructuras en 7 grupos.

La elección tanto del algoritmo de agrupamiento como de la distancia criterio para la selección de grupos es subjetiva y depende del problema específico en el que se esté trabajando, por lo que se utilizó como guía para la selección de ellos el análisis cualitativo de la matriz de disimilitud. El análisis de grupos partió de la matriz de distancias rms de los carbonos alfa entre las estructuras; este análisis se realizó utilizando varios algoritmos (el del vecino más próximo, del vecino más lejano, del grupo promedio y de Ward), siendo el del grupo promedio (average linkage) el que arrojó los resultados que más concuerdan con los obtenidos de la visualización de la matriz de disimilitud (figura 3.13) y con los obtenidos por Troyer y Cohen [38]. Troyer y Cohen encontraron que existen 7 conformaciones para la BPTI; la formación de siete conformaciones coincide con la aplicación del criterio de Mojena con una constante $k = 1.6$; la distancia de corte para la formación de estos grupos es 0.60 Å, semejante a la encontrada por Troyer y Cohen. Pese a que a esta distancia máxima de aglomeración da la formación de 10 grupos, 3 de ellos corresponden a estructuras solas, que no forman parte de ningún grupo a este nivel por lo tanto no están asociadas a conformaciones.

Los grupos formados por las estructuras minimizadas están constituidos por estructuras que se encuentran dentro de los intervalos: 20-190; 200-320; 330-390; 410-490, 550-770, 810-870, 960-1000; 500-540; 780-800; y 880-950 ps están representados en el dendrograma de la figura 3.14 y en la tabla 3.4. La distancia entre los elementos de cada cluster (intracluster) se encuentra entre 0.31 y 0.6 Å, y la distancia entre grupos (intercluster) entre 0.7 y 1.6 Å.

La proteína no regresa a un grupo en el que previamente estuvo hasta los 540 ps, esto implica que las barreras inherentes a las transiciones entre estos mínimos locales son muy grandes como para ser superadas en este intervalo de tiempo; sin embargo a partir de este tiempo, la proteína tiende a regresar a un grupo determinado, las barreras de este con las nuevas conformaciones son lo suficientemente pequeñas como para ser superadas varias veces en este intervalo de tiempo (aproximadamente 460 ps).

Cada uno de los grupos obtenidos en el análisis de grupos corresponde a un estado conformacional de la proteína; estas conformaciones difieren entre si en prácticamente toda la cadena de aminoácidos, sin embargo, existen diferencias más pronunciadas que las caracterizan. Con el propósito de identificar y caracterizar los cambios en cada conformación, se obtuvo la estructura promedio de cada grupo (promedio de las coordenadas de cada átomo de las estructuras que forman el grupo) y se minimizó su energía siguiendo el protocolo de minimización del apartado 3.1.2., obteniendo de este modo una estructura que representa y caracteriza a cada grupo (conformación) tanto en los parámetros asociados a la forma (radio de giro, d_{rms}) como energéticamente.

La identificación de los cambios conformacionales mediante un gráfico de distancias rms entre cada conformación (estructura promedio minimizada) y la estructura calentada a 300 K minimizada. Como muestra la gráfica 3.15, las principales diferencias entre las conformaciones se encuentran en los lazos formados entre los residuos Thr11 – Tyr21, Try21 – Gln31 y Gln31 - Asn44 por lo que las conformaciones obtenidas se pueden definir en términos del movimiento de estos lazos. El lazo no sufre cambios en su forma entre las conformaciones, éste simplemente pivotea sobre sus esquinas ángulos pequeños, por ello la d_{rms} de los residuos “vértice” (Thr11, Tyr21, Gln31, Asn44) es pequeña. El primer lazo corresponde a la zona activa de la proteína, además en él se encuentra una de las cisteína (Cys14) que forma un enlace disulfuro, al igual que en los otros dos lazos (Cys30 y Cys38). El núcleo hidrofóbico propio de la estructura nativa, prácticamente no cambia en estas conformaciones, el radio de giro de este toma valores entre 6.767 Å a 6.92 Å en las conformaciones descritas; estos valores son muy cercanos al que toma en la proteína calentada a 300K y minimizada (6.766 Å).

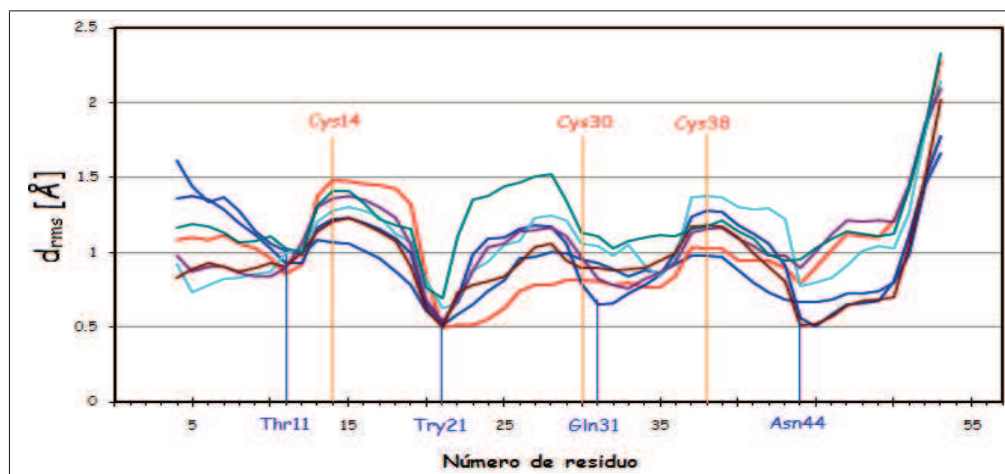


Figura 3.15: Gráfico de distancias rms entre las 7 conformaciones y la estructura calentada a 300K minimizada de la proteína natural.

Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
				Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
0.60	7	18	2 - 19	-8.14 ± 0.16	1.13 ± 0.06	11.20 ± 0.04	-8.20	1.12	11.23
		13	20 - 32	-8.43 ± 0.19	1.17 ± 0.06	11.19 ± 0.05	-8.52	1.24	11.26
		7	33 - 39	-8.53 ± 0.13	1.23 ± 0.08	11.17 ± 0.03	-8.39	1.16	11.24
		43	41 - 49, 55 - 77, 81 - 87, 96 - 100	-8.32 ± 0.15	1.16 ± 0.08	11.18 ± 0.04	-8.04	1.25	11.53
		5	50 - 54	-8.03 ± 0.11	1.24 ± 0.06	11.15 ± 0.05	-8.01	1.23	11.14
		3	78 - 80	-8.41 ± 0.04	1.15 ± 0.04	11.20 ± 0.02	-8.41	1.12	11.20
		8	88 - 95	-8.38 ± 0.12	1.24 ± 0.04	11.24 ± 0.03	-8.45	1.38	11.21

Tabla 3.4: Análisis jerárquico de las estructuras asociadas al estado nativo de la BPTI natural.

La tabla 3.4 presenta en forma resumida las características de los grupos. Los valores de radio de giro, d_{rms} y energía de las estructuras promedio de cada grupo son bastante aproximados al promedio de los valores de estos parámetros de las estructuras de cada grupo, por lo que se puede caracterizar de forma adecuada a cada uno de los grupos por medio de sus estructura promedio. Los valores similares de la distancia rms de las estructuras promedio no implican que las estructuras sean semejantes entre si, sino que tienen el mismo grado de disimilitud con la estructura calentada a 300K minimizada.

3.2.2. BPTI mutada

La figura 3.16 es una representación gráfica de la matriz de disimilitud, definida por las distancias rms de los carbonos alfa de las estructuras. Para el análisis de grupos se utilizó, al igual que para el sistema de la estructura natural, el algoritmo de agrupación jerárquica del grupo promedio, cuyos resultados son los más semejantes a los del análisis visual de la matriz de disimilitud (Figura 3.16). Para la selección del número óptimo de grupo se aplicó el criterio de Mojena con

$k=1.6$, encontrando una distancia de corte de 0.73 \AA . La figura 3.16 muestra la identificación de los nueve grupos formados en el análisis en la matriz de disimilitud. El dendograma de la figura 3.17, muestra esquemáticamente la formación de los nueve grupos de estructuras a un nivel de 0.73 \AA .

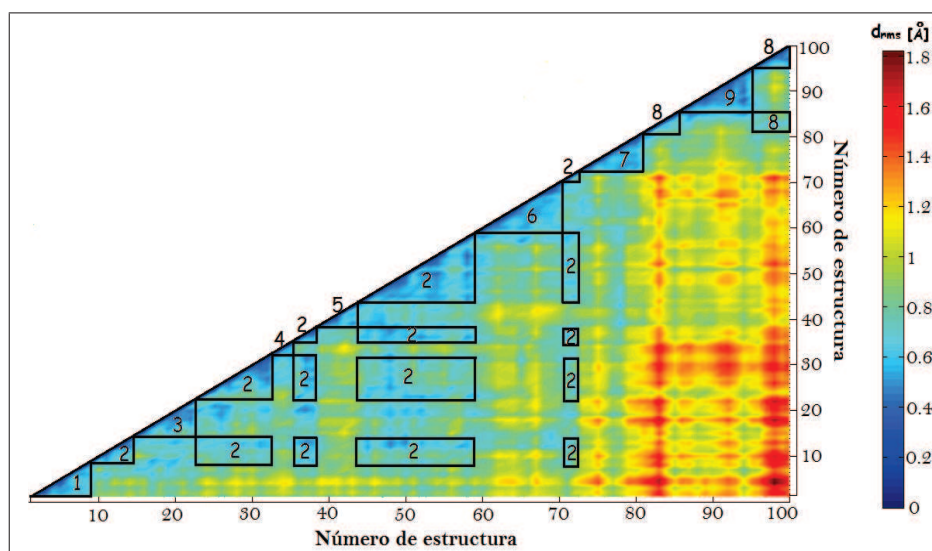


Figura 3.16: Representación gráfica de la matriz de disimilitud de las estructuras asociadas al estado nativo de la BPTI mutada. Identificación de los 9 grupos obtenidos en el análisis de grupos; las regiones representadas corresponde a los 7 grupos formados tomando como distancia criterio 0.73 \AA .

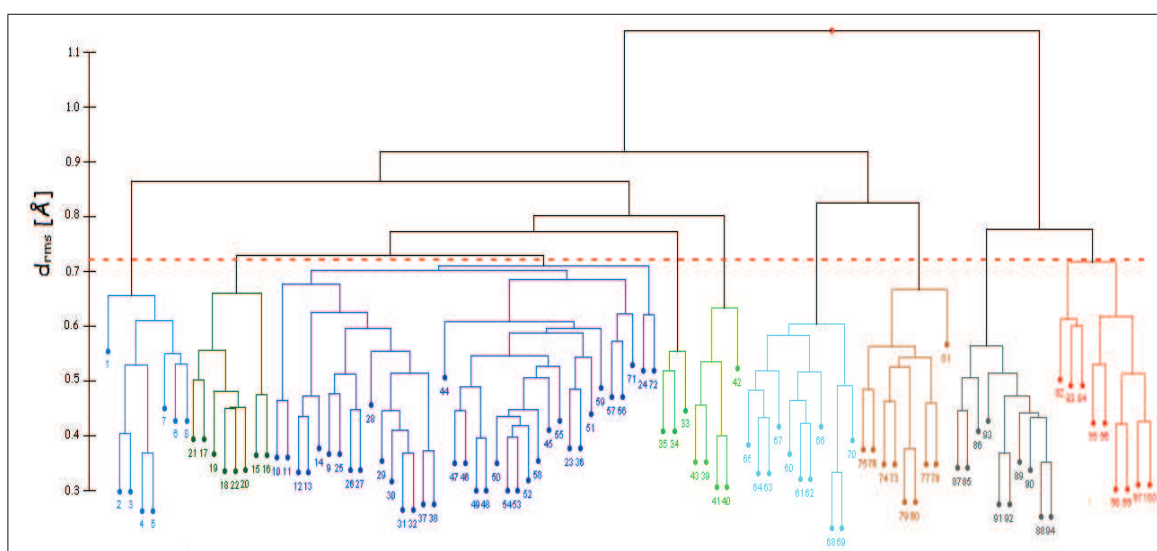


Figura 3.17: Dendograma del análisis de grupos de las estructuras asociadas al estado nativo de la BPTI mutada. Resultado del análisis de grupos jerárquico aglomerativo con el algoritmo del grupo promedio. La línea roja punteada a una d_{rms} de 0.73 \AA separa las estructuras en 9 grupos.

Los grupos formados por las estructuras minimizadas no están constituidos por estructuras continuas en el tiempo, es decir, la proteína regresa a conformaciones que ya visitó anteriormente; estos regresos a conformaciones antes sugieren que

las barreras que separan los mínimos asociados a las conformaciones no son muy altas y por tanto fáciles de superar. Los grupos están formados por estructuras dentro de los intervalos de tiempo: 10-80; 90-320, 360-380, 440-590, 710-720; 330- 350; 390-430; 600-700; 730-810; 820-840, 950-1000; y 850-9400 ps y están representados tanto en dendograma de la Figura 3.17 como en la matriz de disimilitud de la Figura 3.16 y en la tabla 3.5. La distancia entre los elementos de cada cluster (intracluster) se encuentra entre 0.33 y 0.65 Å, y la distancia entre grupos (intercluster) entre 0.7 y 1.65 Å.

La gráfica 3.18 muestra que las principales diferencias entre las nueve conformaciones de la BPTI mutada se encuentran en los lazos formados entre los Thr11 – Tyr21, Try21 – Gln31 y Gln31 - Asn44; es decir, las conformaciones de la proteína mutada coinciden en su determinación con las de la natural; la mutación acentuó la dinámica de la proteína en las regiones mencionadas.

Durante los cambios entre conformaciones estos lazos casi no cambian su forma sino que pivotean sobre sus “vértices” (residuos inicial y final de cada lazo descrito), este movimiento es mucho más pronunciado en el lazo formado entre los residuos Tyr21 – Gln31 (figura 3.18). El núcleo hidrofóbico propio de la estructura mutada nativa, prácticamente no cambia en estas conformaciones, el radio de giro de este toma valores entre 6.735 Å a 6.900 Å en las conformaciones descritas; estos valores son muy cercanos al que toma en la proteína calentada a 300K y minimizada (6.904 Å).

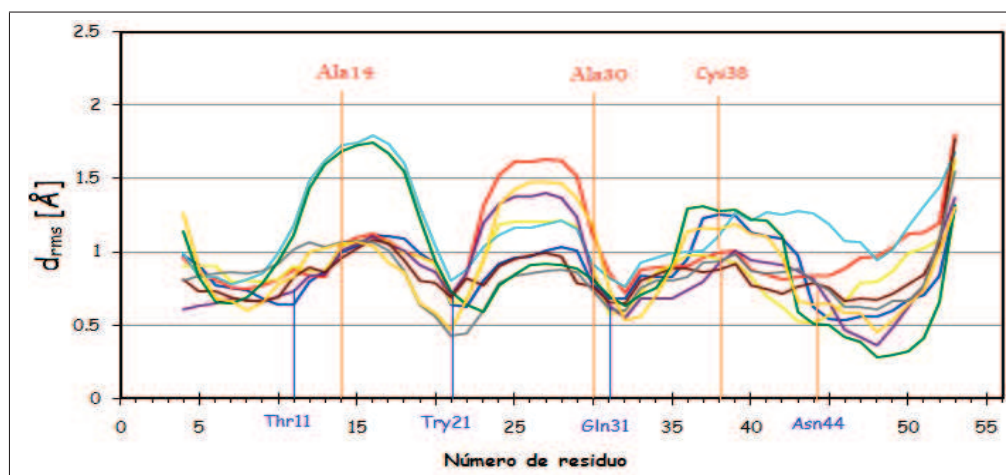


Figura 3.18: Gráfico de distancias rms entre las 9 conformaciones y la estructura calentada a 300K minimizada de la proteína mutada.

Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
				Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
0.73	9	8	1 - 8	-8.67 ± 0.15	1.10 ± 0.07	11.26 ± 0.06	-8.73	0.97	11.21
		37	9 - 14, 23 - 32, 36 - 38, 44 - 59, 71 - 72	-8.51 ± 0.21	1.04 ± 0.10	11.26 ± 0.05	-8.48	1.17	11.28
		8	15 - 22	-8.20 ± 0.21	1.13 ± 0.08	11.27 ± 0.03	-8.25	1.03	11.35
		0	33 - 35	-8.31 ± 0.05	1.24 ± 0.11	11.34 ± 0.03	-8.37	1.25	11.26
		5	39 - 43	-8.72 ± 0.09	1.03 ± 0.07	11.30 ± 0.03	-8.77	0.95	11.22
		11	60 - 70	-8.56 ± 0.10	1.00 ± 0.11	11.26 ± 0.03	-8.53	1.00	11.23
		9	73 - 81	-8.51 ± 0.11	1.02 ± 0.06	11.25 ± 0.05	-8.76	0.96	11.15
		9	82 - 84, 95 - 100	-8.70 ± 0.11	1.09 ± 0.04	11.15 ± 0.09	-8.53	1.10	11.09
		10	85 - 94	-8.55 ± 0.20	1.04 ± 0.07	11.13 ± 0.04	-8.81	1.05	11.07

Tabla 3.5: Análisis jerárquico de las estructuras asociadas al estado nativo de la BPTI mutada.

La tabla 3.5 presenta los principales resultados del análisis de grupos y las características principales de las conformaciones obtenidas. La estructura mutada sufre más cambios que la natural, lo que se ve reflejado no solo en los movimientos de los lazos que caracterizan a las conformaciones, sino también por los valores de los parámetros asociados a las estructuras (radio de giro, d_{rms} y energía).

3.3. Conformaciones no nativas

Las conformaciones descritas en el apartado 3.2. están asociadas al estado nativo de la proteína, es decir, a las cercanías del mínimo global de la superficie del Entorno Energético inherente a cada estructura. En este apartado se describirán las conformaciones accesibles a la proteína tanto natural como mutada asociadas a mínimos locales, es decir, estados no nativos. Para la identificación de estas conformaciones se partió de las muestras de estructuras descritas en el apartado 3.1.4.2. En principio cada una de las muestras corresponde a un mínimo local diferente, por lo tanto las conformaciones identificadas se encuentran en sus alrededores. Para estos análisis se utilizaron los mismos procedimientos y criterios que los usados en la identificación de las conformaciones asociadas al mínimo global, ésto es: un análisis de grupos jerárquico aglomerativo, partiendo de la matriz de disimilitud de las d_{rms} mutuas entre estructuras, con el método del grupo promedio; para la selección del número óptimo de grupos se utilizó el criterio de Mojena con una constante $k=1.6$. Todos los análisis estadísticos se realizaron utilizando el programa DARWin5[68].

3.3.1. BPTI natural

Las tablas 3.6 a 3.15, muestran los principales resultados del análisis de las muestras de la estructura natural. Las distancias de corte fueron calculadas con el

criterio de Mojena con $k=1.6$. Para todas la muestras, estos valores son mayores que el encontrado para la muestra asociada al mínimo global (0.6 \AA), esto implica que las estructuras que forman cada grupo son menos homogéneas (menor similitud entre estructuras). Cada grupo de estructuras corresponde a una conformación accesible a la proteína, este número varía entre 7 y 11 para las muestras, es decir, la proteína se vuelve más dinámica en las cercanías de los mínimos locales que en la cercanía al mínimo global, además, se puede presumir que las barreras energéticas en esos entornos son más pequeñas y por ende más fáciles de superar.

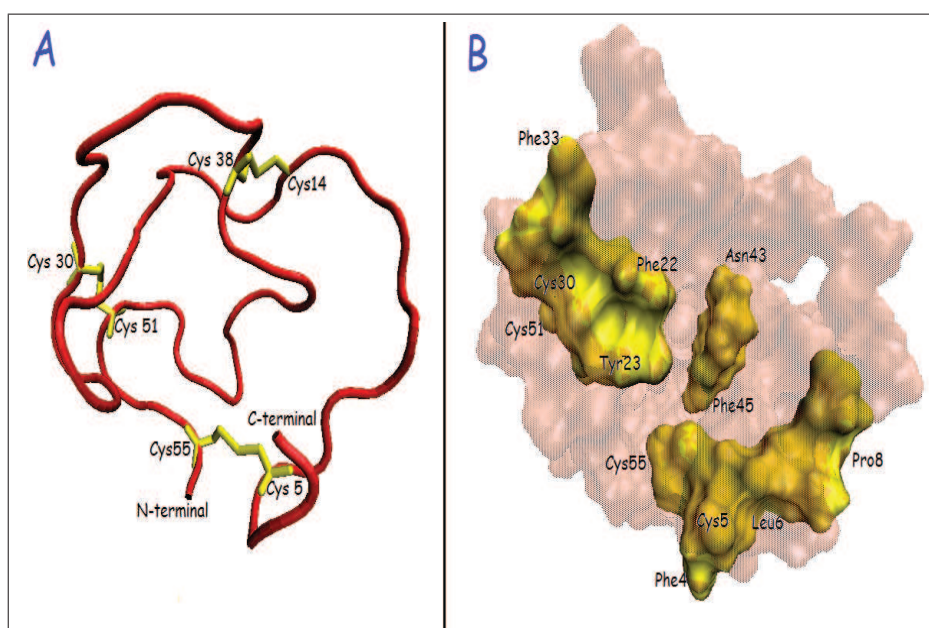


Figura 3.19: Estructura de la primera conformación de la muestra 1 de la proteína natural. **A:** Estructura tridimensional. **B:** disposición espacial de los residuos que formaban el núcleo hidrofóbico.

Considerando los valores promedio y las desviaciones estándar de los parámetros radio de giro, distancia rms y energía de cada grupo, y los obtenidos de las estructuras promedio minimizadas correspondientes, podemos considerar que estas pueden caracterizar de una forma consistente a cada grupo, por ende cada conformación. Los valores de radio de giro y d_{rms} de estas conformaciones son mucho mayores que los encontrados para las conformaciones asociadas al mínimo global, lo que implica que la estructura de la proteína ha cambiado sustancialmente; como muestra la figura 3.19, la estructura secundaria de la proteína se ha perdido totalmente, solo encontramos presentes lazos unidos básicamente por los enlaces disulfuro, los residuos que forman el núcleo hidrofóbico (en el estado nativo) se encuentran espacialmente separados lo que implica que este núcleo tampoco se forma en las conformaciones.

En las figuras 3.20 a 3.29 que muestran las gráficas de distancias rms entre las conformaciones de cada muestra y la estructura caliente a 300 K minimizada, se observa ciertos lazos que poseen mayor movilidad. Así se puede caracterizar a las conformaciones dentro de cada muestra en función de la movilidad de estos lazos, los resultados están expuestos en la tabla 3.16. Las cisteínas que forman los enlaces disulfuro, tienden en general a tener poca movilidad, en general se ubican en los “vértices” de los lazos móviles o en sus cercanías; cuando estos residuos presentan cierta movilidad también la presenta la correspondiente cisteínas con la que forman el enlace disulfuro. El movimiento de las cisteínas está restringido por la presencia de los enlaces disulfuro. Los parámetros radio de giro y d_{rms} de las conformaciones son similares dentro de cada muestra y bastante diferentes entre muestras, es decir, las estructuras de cada muestra son muy similares y difieren por movimientos ligeros en ciertos lazos, como muestra la tabla 3.16, en cambio las estructuras de diferentes muestras varían drásticamente, existe una distribución espacial de los residuos muy diferente. Los lazos que caracterizan pivotan sobre sus esquinas formando pequeños ángulos.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
1	1.03	8	8	1 - 8	-5.87 ± 0.14	11.90 ± 0.05	6.69 ± 0.06	-5.94	11.85	6.71
			3	9 - 11	-5.78 ± 0.22	12.02 ± 0.02	6.72 ± 0.17	-5.79	11.95	6.74
			5	12 - 16	-6.25 ± 0.06	12.05 ± 0.06	6.87 ± 0.08	-5.99	12.01	6.86
			8	17 - 24	-6.22 ± 0.15	11.93 ± 0.05	6.80 ± 0.08	-6.41	11.88	6.79
			10	25 - 34	-6.60 ± 0.29	11.85 ± 0.05	6.73 ± 0.08	-6.25	11.84	6.64
			9	35 - 43	-6.89 ± 0.20	11.87 ± 0.07	6.59 ± 0.08	-6.84	11.89	6.59
			19	44 - 62	-6.99 ± 0.19	11.93 ± 0.11	6.83 ± 0.11	-6.98	11.93	6.75
38	63 - 100	-6.81 ± 0.25	11.65 ± 0.08	6.92 ± 0.10	-6.72	11.69	6.95			

Tabla 3.6: Análisis jerárquico de la muestra 1 de estructuras no nativas de la BPTI natural.

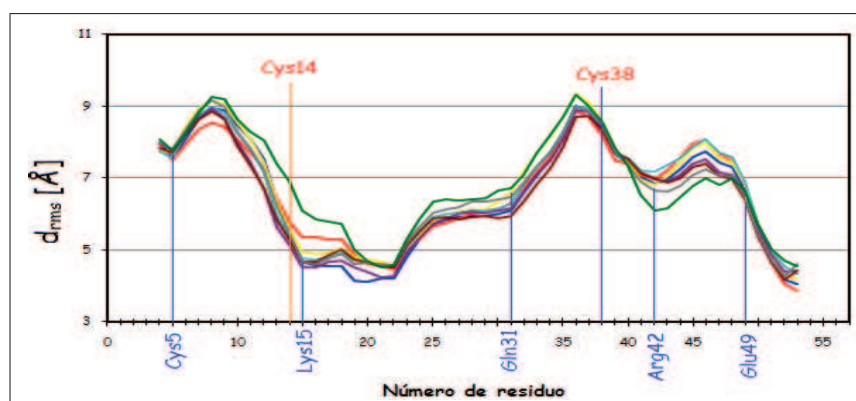


Figura 3.20: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 1 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
2	1.04	10	12	1 - 12	-6.68 ± 0.26	12.11 ± 0.07	7.75 ± 0.09	-6.75	12.08	7.82
			5	13 - 17	-6.70 ± 0.16	12.21 ± 0.08	7.70 ± 0.10	-6.78	12.21	7.65
			10	18 - 27	-6.87 ± 0.19	12.25 ± 0.06	7.69 ± 0.05	-6.92	12.23	7.69
			13	28 - 40	-6.83 ± 0.14	12.18 ± 0.05	7.38 ± 0.10	-6.86	12.13	7.40
			8	41 - 48	-6.67 ± 0.18	12.15 ± 0.08	7.42 ± 0.11	-6.58	12.07	7.34
			17	49 - 65	-6.87 ± 0.16	12.07 ± 0.06	7.34 ± 0.09	-7.08	12.07	7.43
			4	66 - 69	-6.86 ± 0.22	12.03 ± 0.08	7.31 ± 0.07	-6.87	12.02	7.26
			15	70 - 84	-7.03 ± 0.18	12.00 ± 0.07	7.18 ± 0.06	-7.01	12.08	7.21
10	85 - 94	-7.12 ± 0.17	12.08 ± 0.08	7.45 ± 0.07	-7.09	12.14	7.55			
6	95 - 100	-7.27 ± 0.10	12.13 ± 0.02	7.27 ± 0.07	-7.43	12.10	7.30			

Tabla 3.7: Análisis jerárquico de la muestra 2 de estructuras no nativas de la BPTI natural.

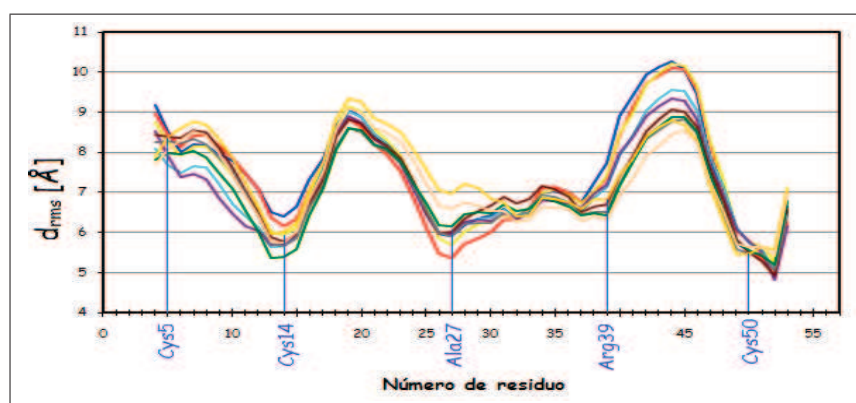


Figura 3.21: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 2 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
3	0.92	10	7	1 - 7	-8.05 ± 0.12	11.67 ± 0.05	9.98 ± 0.05	-7.96	11.68	10.05
			8	8 - 15	-8.09 ± 0.19	11.73 ± 0.04	9.97 ± 0.06	-7.99	11.81	10.02
			6	16 - 21	-8.18 ± 0.12	11.82 ± 0.06	10.00 ± 0.11	-7.91	11.81	10.00
			3	22 - 24	-8.41 ± 0.07	11.91 ± 0.05	9.87 ± 0.08	-8.47	11.91	9.83
			8	25 - 32	-8.19 ± 0.10	11.80 ± 0.03	10.01 ± 0.08	-8.10	11.83	10.01
			4	33 - 36	-7.96 ± 0.14	11.87 ± 0.05	9.86 ± 0.07	-7.91	11.89	9.86
			27	37 - 63	-8.10 ± 0.13	11.89 ± 0.05	9.70 ± 0.08	-8.07	11.92	9.76
			8	64 - 71	-8.13 ± 0.09	11.81 ± 0.09	9.87 ± 0.06	-8.15	11.85	9.89
15	72 - 86	-7.97 ± 0.14	11.83 ± 0.05	9.68 ± 0.08	-8.08	11.85	9.70			
14	87 - 100	-7.96 ± 0.17	11.80 ± 0.06	9.62 ± 0.10	-8.05	11.77	9.62			

Tabla 3.8: Análisis jerárquico de la muestra 3 de estructuras no nativas de la BPTI natural.

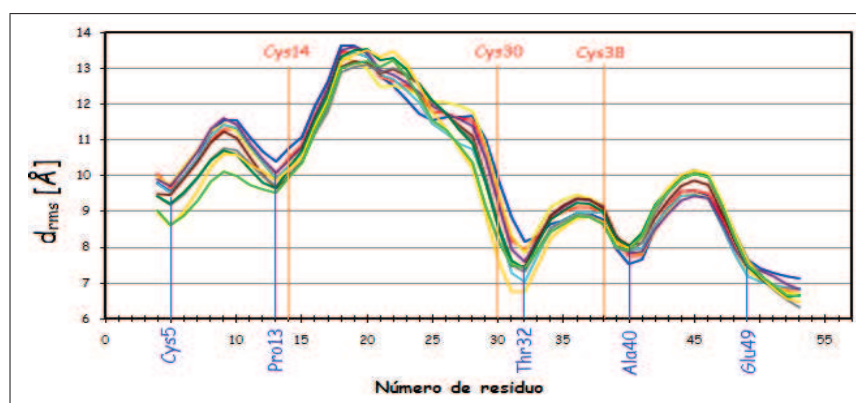


Figura 3.22: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 3 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
4	1.21	8	19	1 - 19	-6.03 ± 0.22	12.95 ± 0.08	9.45 ± 0.11	-6.16	12.97	9.47
			9	20 - 28	-5.98 ± 0.18	12.94 ± 0.14	9.31 ± 0.13	-6.14	12.95	9.35
			22	29 - 50	-6.04 ± 0.12	13.03 ± 0.08	9.64 ± 0.13	-6.23	13.11	9.64
			20	51 - 70	-6.14 ± 0.18	12.96 ± 0.11	9.47 ± 0.10	-5.94	13.00	9.56
			8	71 - 78	-6.24 ± 0.17	12.74 ± 0.07	9.38 ± 0.08	-6.27	12.70	9.42
			10	79 - 88	-6.27 ± 0.22	12.90 ± 0.08	9.38 ± 0.11	-6.49	12.90	9.39
			8	89 - 96	-6.12 ± 0.15	12.97 ± 0.04	9.19 ± 0.11	-6.19	12.92	9.24
4	97 - 100	-6.11 ± 0.17	12.88 ± 0.05	9.02 ± 0.04	-6.06	12.87	9.07			

Tabla 3.9: Análisis jerárquico de la muestra 4 de estructuras no nativas de la BPTI natural.

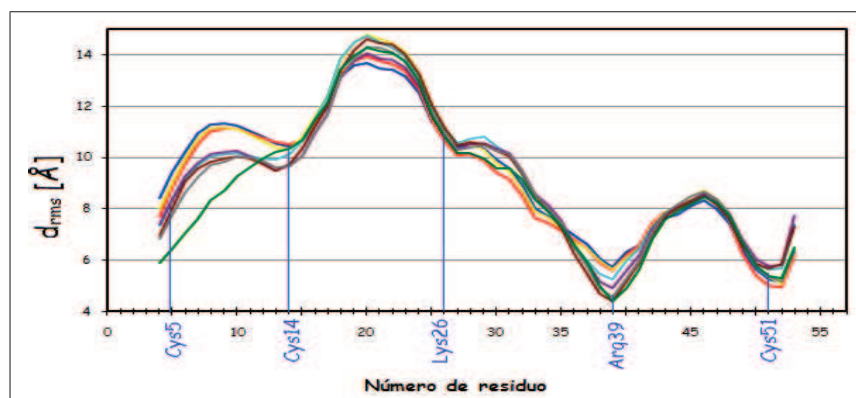


Figura 3.23: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 4 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
5	1.07	9	10	1 - 10	-7.45 ± 0.08	12.61 ± 0.09	10.71 ± 0.07	-7.57	12.61	10.75
			7	11 - 17	-7.38 ± 0.21	12.70 ± 0.10	10.96 ± 0.08	-7.36	12.70	10.92
			7	18 - 24	-7.44 ± 0.12	12.84 ± 0.08	10.95 ± 0.09	-7.64	12.74	10.87
			14	25 - 38	-7.50 ± 0.20	12.95 ± 0.13	11.13 ± 0.11	-7.42	13.07	11.23
			7	39 - 45	-7.56 ± 0.07	12.85 ± 0.03	10.94 ± 0.07	-7.52	12.90	10.83
			9	46 - 54	-7.46 ± 0.18	12.72 ± 0.06	10.71 ± 0.14	-7.71	12.79	10.84
			11	55 - 65	-7.28 ± 0.17	12.70 ± 0.14	10.55 ± 0.15	-7.35	12.78	10.59
			19	66 - 84	-7.21 ± 0.11	12.62 ± 0.08	10.41 ± 0.11	-7.22	12.54	10.34
16	85 - 100	-7.15 ± 0.14	12.53 ± 0.06	10.35 ± 0.04	-7.19	12.56	10.29			

Tabla 3.10: Análisis jerárquico de la muestra 5 de estructuras no nativas de la BPTI natural.

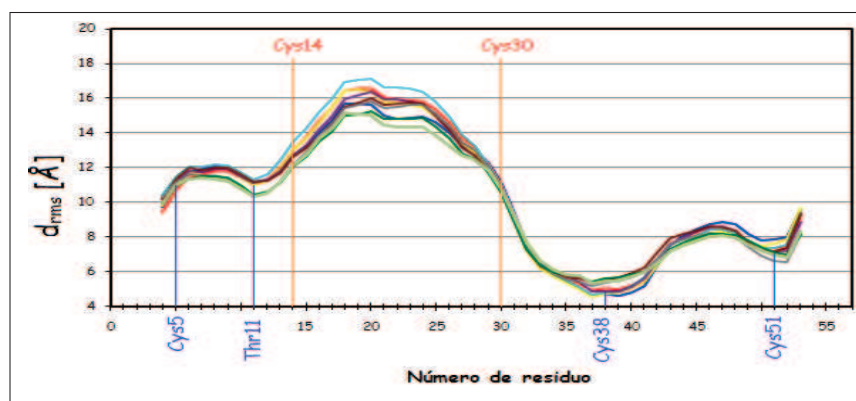


Figura 3.24: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 5 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
6	0.96	8	21	1 - 8 , 12 - 24	-7.17 ± 0.14	11.80 ± 0.06	8.51 ± 0.07	-7.42	11.73	8.50
			3	9 - 11	-7.11 ± 0.04	11.66 ± 0.03	8.59 ± 0.04	-7.09	11.68	8.61
			21	25 - 45	-7.16 ± 0.22	11.82 ± 0.07	8.50 ± 0.08	-7.19	11.84	8.57
			13	46 - 58	-6.52 ± 0.19	11.91 ± 0.05	8.70 ± 0.07	-6.44	11.86	8.77
			16	59 - 74	-6.17 ± 0.23	11.96 ± 0.05	8.77 ± 0.07	-6.15	11.92	8.66
			12	75 - 86	-6.17 ± 0.22	12.00 ± 0.05	8.88 ± 0.04	-6.42	11.98	8.92
			9	87 - 95	-6.21 ± 0.19	12.00 ± 0.06	8.88 ± 0.06	-6.18	11.98	8.80
5	96 - 100	-6.30 ± 0.07	12.10 ± 0.09	9.05 ± 0.04	-6.24	12.13	9.15			

Tabla 3.11: Análisis jerárquico de la muestra 6 de estructuras no nativas de la BPTI natural.

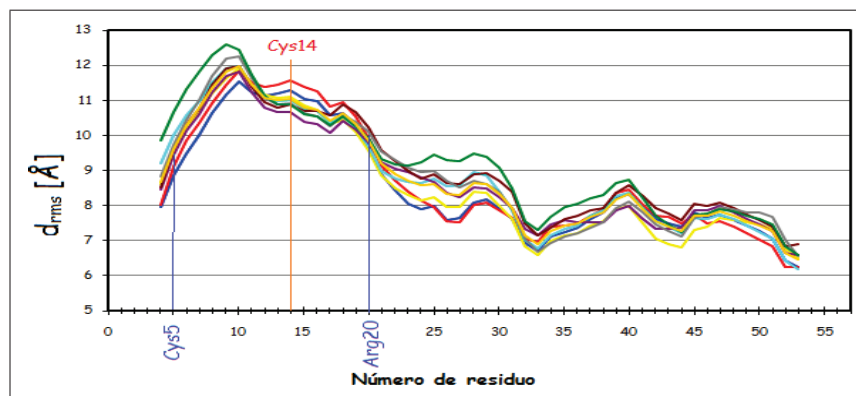


Figura 3.25: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 6 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
7	1.01	6	11	1 - 11	-5.40 ± 0.15	12.53 ± 0.09	11.14 ± 0.09	-5.21	12.51	11.14
			19	12 - 30	-5.54 ± 0.14	12.26 ± 0.09	10.93 ± 0.10	-5.55	12.24	10.98
			21	31 - 51	-5.72 ± 0.19	12.24 ± 0.07	11.05 ± 0.11	-5.73	12.26	11.09
			4	52 - 55	-5.78 ± 0.10	12.14 ± 0.05	11.00 ± 0.06	-5.85	12.06	10.96
			34	56 - 89	-5.52 ± 0.21	12.06 ± 0.07	10.98 ± 0.09	-5.75	12.16	11.14
			11	90 - 100	-5.59 ± 0.26	12.08 ± 0.06	11.20 ± 0.06	-5.38	12.19	11.28

Tabla 3.12: Análisis jerárquico de la muestra 7 de estructuras no nativas de la BPTI natural.

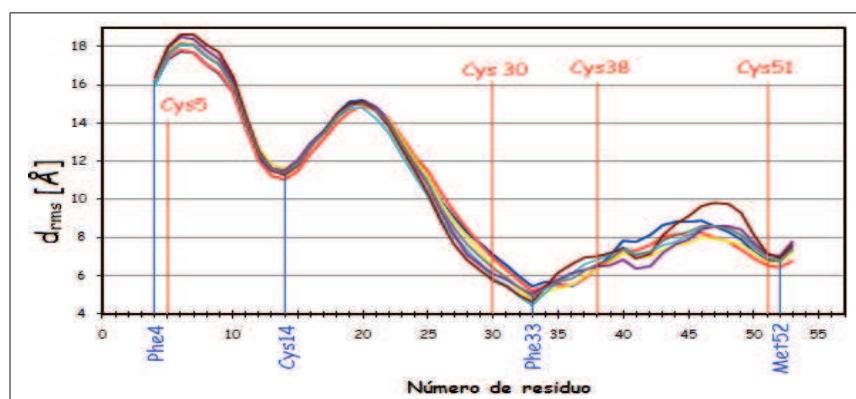


Figura 3.26: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 7 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
8	1.09	10	6	1 - 6	-5.20 ± 0.17	13.32 ± 0.01	11.31 ± 0.06	-5.14	13.31	11.24
			12	7 - 18	-5.03 ± 0.17	13.25 ± 0.07	11.25 ± 0.09	-5.24	13.20	11.31
			7	19 - 25	-5.05 ± 0.24	13.11 ± 0.06	11.09 ± 0.05	-5.00	13.14	11.07
			6	26 - 31	-4.99 ± 0.27	13.21 ± 0.09	11.27 ± 0.11	-4.62	13.20	11.23
			3	32 - 34	-4.95 ± 0.09	13.26 ± 0.10	11.46 ± 0.05	-5.26	13.35	11.52
			22	35 - 56	-5.15 ± 0.23	13.24 ± 0.08	11.21 ± 0.07	-5.00	13.30	11.22
			13	57 - 69	-5.09 ± 0.20	13.27 ± 0.07	11.24 ± 0.08	-4.94	13.24	11.30
			5	70 - 74	-5.26 ± 0.26	13.30 ± 0.05	11.29 ± 0.05	-5.16	13.25	11.22
			22	75 - 86 , 91 - 100	-5.19 ± 0.17	13.50 ± 0.09	11.31 ± 0.07	-4.98	13.69	11.46
4	87 - 90	-5.11 ± 0.20	13.46 ± 0.04	11.27 ± 0.03	-5.21	13.40	11.17			

Tabla 3.13: Análisis jerárquico de la muestra 8 de estructuras no nativas de la BPTI natural.

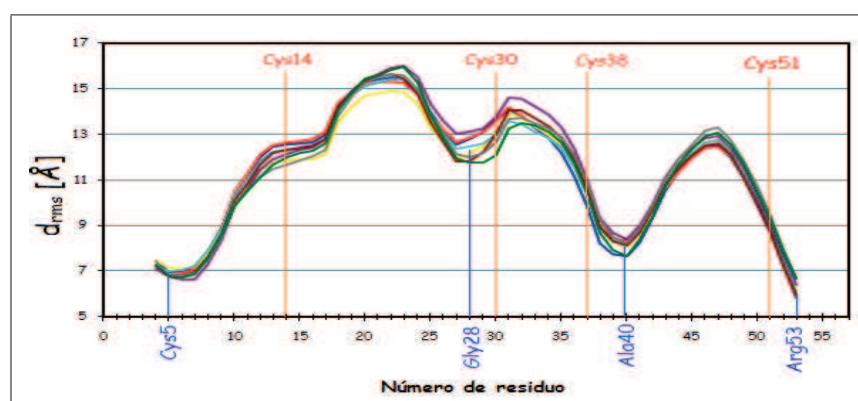


Figura 3.27: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 8 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
9	1.00	11	20	1 - 20	-6.40 ± 0.26	11.98 ± 0.05	10.54 ± 0.08	-6.25	11.91	10.54
			8	21 - 28	-6.49 ± 0.13	12.04 ± 0.08	10.40 ± 0.04	-6.52	12.10	10.42
			14	29 - 32 , 38 - 47	-6.79 ± 0.17	12.03 ± 0.07	10.40 ± 0.06	-6.77	11.98	10.51
			5	33 - 37	-6.75 ± 0.19	11.91 ± 0.03	10.41 ± 0.06	-6.85	11.94	10.36
			5	48 - 52	-6.34 ± 0.21	12.08 ± 0.05	10.09 ± 0.12	-6.24	12.06	10.04
			8	53 - 60	-6.39 ± 0.19	12.11 ± 0.04	10.00 ± 0.06	-6.43	12.12	9.98
			9	61 - 69	-6.62 ± 0.12	12.11 ± 0.06	9.96 ± 0.05	-6.86	12.09	10.01
			10	70 - 79	-6.41 ± 0.18	12.21 ± 0.04	10.01 ± 0.12	-6.64	12.13	10.02
			9	80 - 88	-6.27 ± 0.24	12.02 ± 0.07	9.83 ± 0.09	-6.24	12.02	9.73
			8	89 - 96	-6.15 ± 0.15	12.00 ± 0.05	9.84 ± 0.05	-6.37	12.07	9.88
			4	97 - 100	-6.44 ± 0.07	12.02 ± 0.06	9.92 ± 0.12	-6.44	12.01	9.86

Tabla 3.14: Análisis jerárquico de la muestra 9 de estructuras no nativas de la BPTI natural.

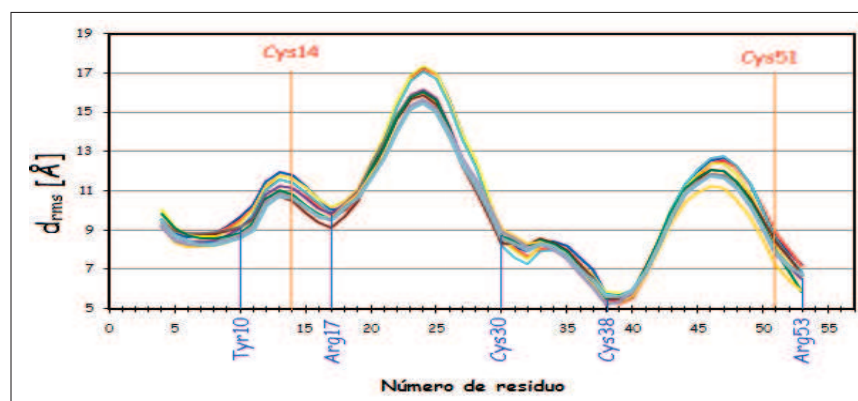


Figura 3.28: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 9 y la estructura calentada a 300K minimizada de la proteína natural.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
10	1.01	8	22	1 - 22	-6.33 ± 0.22	12.03 ± 0.07	9.97 ± 0.06	-6.39	12.08	9.99
			7	23 - 29	-5.97 ± 0.17	11.94 ± 0.06	9.91 ± 0.08	-6.11	11.94	9.87
			19	30 - 48	-6.16 ± 0.19	12.03 ± 0.07	10.11 ± 0.11	-6.25	12.07	10.15
			14	49 - 62	-6.53 ± 0.25	12.16 ± 0.05	10.27 ± 0.07	-6.71	12.22	10.26
			8	63 - 70	-6.53 ± 0.25	12.26 ± 0.13	10.44 ± 0.09	-6.62	12.24	10.39
			19	71 - 89	-6.27 ± 0.15	12.36 ± 0.06	10.39 ± 0.08	-6.47	12.31	10.34
			2	90 - 91	-6.43 ± 0.17	12.54 ± 0.04	10.59 ± 0.08	-6.43	12.48	10.51
9	92 - 100	-6.28 ± 0.30	12.32 ± 0.06	10.33 ± 0.06	-6.51	12.33	10.35			

Tabla 3.15: Análisis jerárquico de la muestra 10 de estructuras no nativas de la BPTI natural.

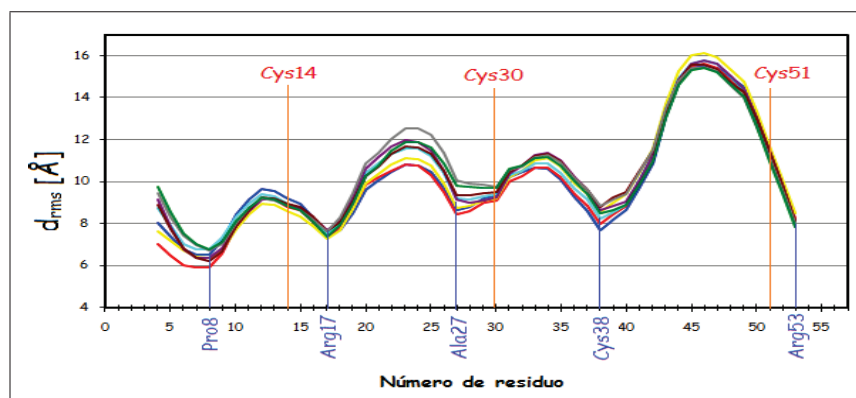


Figura 3.29: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 10 y la estructura calentada a 300K minimizada de la proteína natural.

Número de muestra	1	2	3	4	5
Lazos con mayor movilidad	Cys5-Lys15 Gln31-Arg42 Arg42-Glu49	Cy5-Cys14 Cys14-Ala27 Arg39-Cys50	Cys5-Pro13 Pro13-Thr32 Thr32-Ala40 Ala40-Glu49	Cys5-Cy14 Cys14-Lys26 Arg39-Cys51	Cys5-Thr11 Thr11-Csy38 Cys38-Cys51
Número de muestra	6	7	8	9	10
Lazos con mayor movilidad	Cys5-Arg20	Phe4-Cys14 Cys14-Phe33 Phe33-Met52	Cys5-Gly28 Gly28-Ala40 Ala40-Arg53	Tyr10-Arg17 Arg17-Cys30 Cys38-Arg53	Pro8-Arg17 Arg17-Ala27 Ala27-Cys38 Cys38-Arg53

Tabla 3.16: Lazos móviles que caracterizan a las conformaciones no nativas de la BPTI natural

3.3.2. BPTI mutada

Las tablas 3.17 a 3.26, muestran los principales resultados del análisis sobre las muestras de estructuras no nativas del sistema mutado. Las distancias de corte encontradas con el criterio de Mojena son aproximadamente 0.5 Å mayores que las obtenidas para las muestras de la natural, es decir, las estructuras dentro de cada grupo son más diferentes entre ellas que en los grupos formados por las estructuras naturales; la proteína tiene mayor movilidad debido a la no existencia de enlaces disulfuro que restrinjan su movimiento. El número de conformaciones accesibles (grupos de estructuras) varía entre 8 y 11 para cada muestra, es decir, el número de conformaciones para la proteína mutada es apenas mayor que para la natural, sin embargo, las conformaciones para la natural se caracterizan por ligeros cambios en lazos, en cambio para la mutada los movimientos son más drásticos, lo que implica una diferencia más pronunciada en el reordenamiento tridimensional de la cadena de aminoácidos por ende mayor diferencia de energía entre conformaciones.

Como muestra la figura 3.30, la proteína ha perdido totalmente su estructura tridimensional, no existen estructuras secundarias y los residuos que formaban el núcleo hidrofóbico se encuentran dispersos. La estructura es mucho más desplegada que la obtenida para el sistema natural, esto como consecuencia de la no presencia de los enlaces disulfuro.

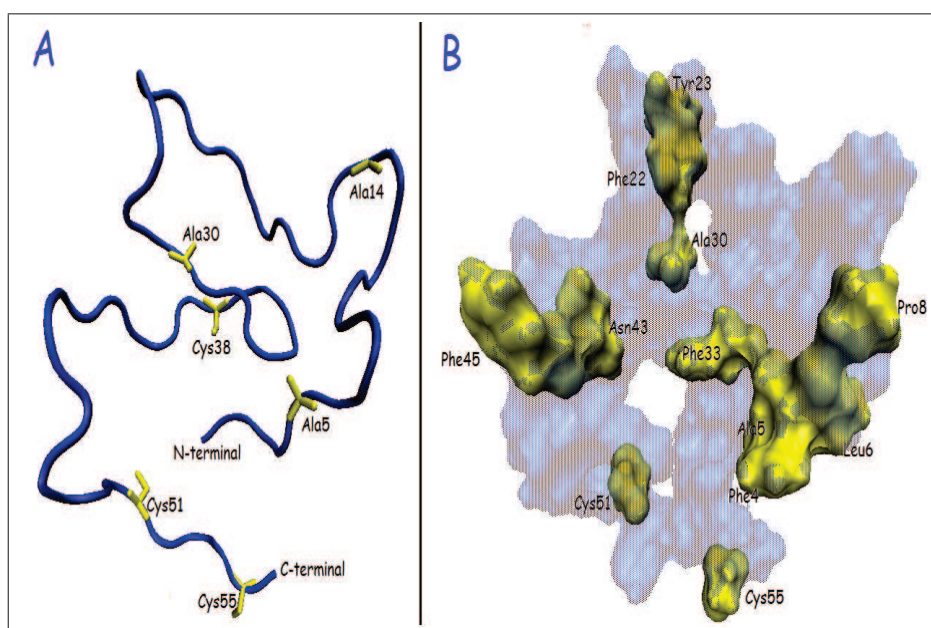


Figura 3.30: Estructura de la primera conformación de la muestra 1 de la proteína mutada. **A:** Estructura tridimensional. **B:** disposición espacial de los residuos que forman el núcleo hidrofóbico.

Se realizó las gráficas de las distancias rms entre las conformaciones de cada muestra y la estructura calentada a 300 K minimizada, como se muestra en las figuras de 3.31 a 3.40, encontrándose que las conformaciones de cada muestra pueden ser descritas por el movimiento de lazos determinados que poseen una mayor movilidad (tabla 3.27).

En general los valores que toman los parámetros radio de giro y d_{rms} de las conformaciones son mayores que los existentes para la proteína natural. Esto implica que la proteína mutada tiene una mayor dinámica que la natural. Al contrario de lo que sucede en la natural, las cisteínas que formaban los enlaces disulfuro y las alaninas colocadas en la mutación, tienen en general mucha mayor movilidad.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
1	1.42	10	5	1 - 5	-4.96 ± 0.10	13.46 ± 0.06	11.81 ± 0.07	-5.18	13.44	11.77
			11	6 - 16	-5.22 ± 0.19	13.57 ± 0.11	11.54 ± 0.16	-5.40	13.53	11.48
			12	17 - 28	-4.43 ± 0.27	13.67 ± 0.07	11.58 ± 0.10	-4.28	13.67	11.60
			9	29 -37	-4.41 ± 0.30	13.55 ± 0.09	11.57 ± 0.16	-4.53	13.55	11.65
			5	38 - 42	-4.84 ± 0.23	13.72 ± 0.11	11.86 ± 0.07	-4.83	13.69	11.75
			11	43 - 53	-4.91 ± 0.19	13.69 ± 0.07	11.96 ± 0.18	-5.01	13.68	12.01
			13	54 - 66	-4.91 ± 0.24	13.55 ± 0.11	11.97 ± 0.10	-4.88	13.62	12.11
			10	67 - 76	-4.97 ± 0.28	13.41 ± 0.09	11.87 ± 0.14	-5.23	13.42	11.90
			12	77 - 88	-4.99 ± 0.28	13.27 ± 0.07	11.64 ± 0.07	-5.14	13.21	11.64
			12	89 - 100	-5.34 ± 0.21	13.04 ± 0.09	11.33 ± 0.10	-5.36	13.12	11.42

Tabla 3.17: Análisis jerárquico de la muestra 1 de estructuras no nativas de la BPTI mutada.

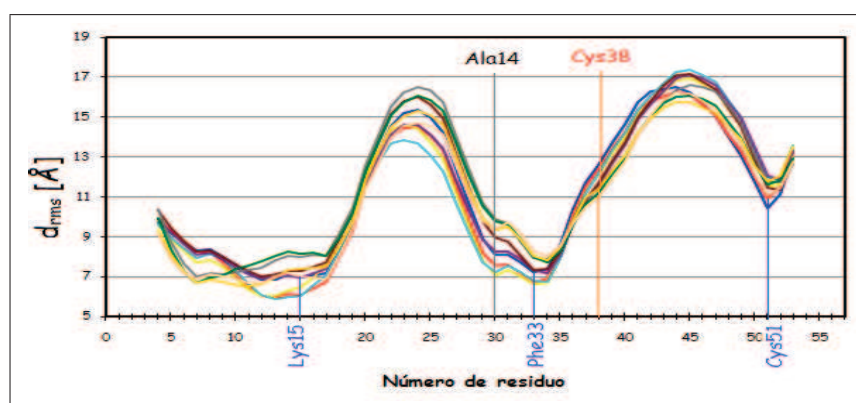


Figura 3.31: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 1 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
2	1.53	10	8	1 - 8	-6.01 ± 0.24	13.08 ± 0.07	13.26 ± 0.12	-6.49	13.13	13.20
			9	9 - 17	-5.84 ± 0.18	13.28 ± 0.11	13.64 ± 0.13	-5.92	13.38	13.77
			9	18 - 26	-5.73 ± 0.16	13.30 ± 0.06	13.62 ± 0.11	-5.86	13.24	13.60
			9	27 - 35	-5.24 ± 0.28	13.13 ± 0.11	13.26 ± 0.10	-5.62	13.25	13.36
			10	36 - 45	-5.14 ± 0.12	12.94 ± 0.12	13.21 ± 0.10	-4.94	13.00	13.33
			12	46 - 57	-4.64 ± 0.22	12.96 ± 0.07	13.24 ± 0.10	-5.08	12.83	13.21
			9	58 - 66	-5.08 ± 0.25	12.73 ± 0.06	13.06 ± 0.11	-5.02	12.76	13.15
			13	67 - 79	-5.03 ± 0.25	12.66 ± 0.09	13.06 ± 0.14	-5.70	12.57	13.03
			13	80 - 92	-5.00 ± 0.23	12.88 ± 0.10	12.95 ± 0.09	-4.69	12.93	12.98
			8	93 - 100	-4.42 ± 0.16	12.97 ± 0.07	12.82 ± 0.10	-4.58	12.88	12.74

Tabla 3.18: Análisis jerárquico de la muestra 2 de estructuras no nativas de la BPTI mutada.

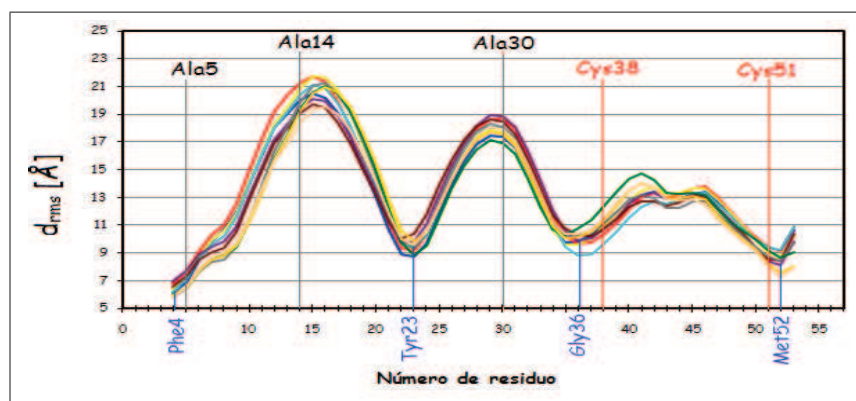


Figura 3.32: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 2 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
3	1.46	11	5	1 - 5	-7.08 ± 0.06	12.94 ± 0.11	13.26 ± 0.12	-6.88	12.91	13.19
			13	6 - 18	-6.79 ± 0.28	13.14 ± 0.10	13.44 ± 0.10	-6.99	13.24	13.47
			5	19 - 23	-6.71 ± 0.21	13.13 ± 0.13	13.73 ± 0.20	-6.76	13.12	13.81
			15	24 - 38	-6.51 ± 0.25	13.30 ± 0.12	13.57 ± 0.10	-6.27	13.39	13.69
			2	39 - 40	-6.15 ± 0.05	13.28 ± 0.01	13.47 ± 0.01	-6.23	13.23	13.45
			16	41 - 56	-6.35 ± 0.22	13.46 ± 0.10	13.59 ± 0.17	-6.45	13.47	13.58
			7	57 - 63	-6.26 ± 0.21	13.46 ± 0.07	13.86 ± 0.07	-6.49	13.52	13.93
			8	64 - 71	-6.28 ± 0.12	13.64 ± 0.11	13.64 ± 0.17	-6.34	13.58	13.63
			10	72 - 81	-6.69 ± 0.37	13.55 ± 0.08	13.52 ± 0.14	-6.40	13.67	13.69
			7	82 - 88	-6.45 ± 0.25	13.46 ± 0.06	13.25 ± 0.09	-6.64	13.44	13.23
						12	89 - 100	-6.77 ± 0.16	13.24 ± 0.09	12.88 ± 0.10

Tabla 3.19: Análisis jerárquico de la muestra 3 de estructuras no nativas de la BPTI mutada.

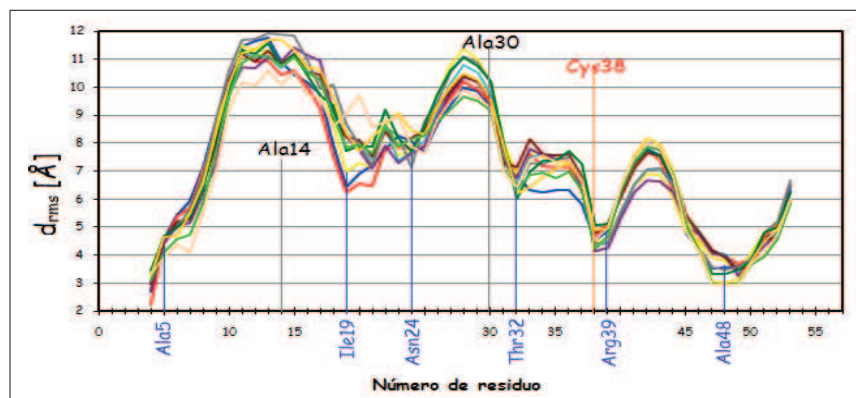


Figura 3.33: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 3 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
4	1.01	8	5	1 - 5	-6.40 ± 0.20	12.27 ± 0.04	12.96 ± 0.07	-6.62	12.28	13.00
			4	6 - 9	-7.05 ± 0.26	12.34 ± 0.06	12.95 ± 0.08	-7.11	12.41	13.03
			6	10 - 15	-7.12 ± 0.37	12.37 ± 0.08	13.01 ± 0.05	-7.28	12.41	13.11
			14	16 - 29	-7.38 ± 0.40	12.43 ± 0.06	13.06 ± 0.08	-7.68	12.37	13.03
			2	30 - 31	-7.58 ± 0.00	12.46 ± 0.06	13.23 ± 0.05	-7.68	12.44	13.25
			17	32 - 48	-7.83 ± 0.16	12.53 ± 0.05	12.89 ± 0.10	-8.05	12.55	12.92
			28	49 - 53, 70 - 78, 87 - 100	-7.96 ± 0.13	12.37 ± 0.08	12.77 ± 0.07	-7.61	12.55	12.81
24	54 - 69, 79 - 86	-7.96 ± 0.09	12.40 ± 0.04	12.70 ± 0.08	-7.47	12.53	12.78			

Tabla 3.20: Análisis jerárquico de la muestra 4 de estructuras no nativas de la BPTI mutada.

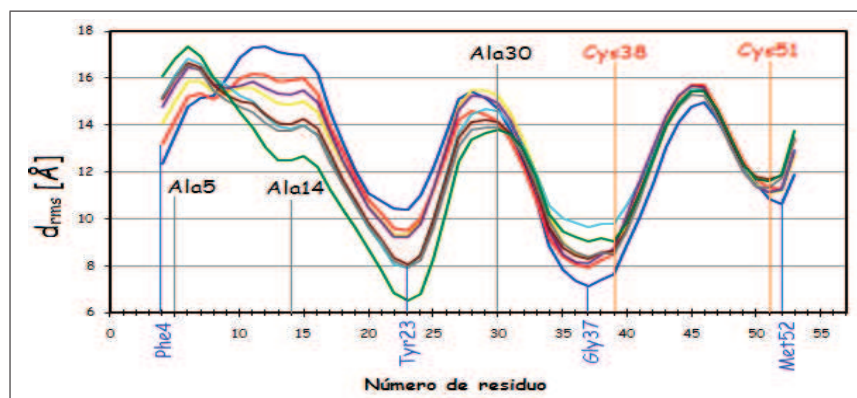


Figura 3.34: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 4 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
5	1.31	8	12	1 - 12	-5.55 ± 0.33	13.40 ± 0.07	11.52 ± 0.07	-5.74	13.49	11.56
			10	13 - 22	-5.77 ± 0.15	13.28 ± 0.10	11.40 ± 0.08	-5.73	13.32	11.56
			10	23 - 32	-5.97 ± 0.24	13.08 ± 0.06	11.44 ± 0.08	-6.02	13.09	11.44
			9	33 - 41	-5.77 ± 0.26	13.13 ± 0.09	11.21 ± 0.08	-5.99	13.08	11.09
			8	42 - 49	-5.76 ± 0.19	13.16 ± 0.12	11.29 ± 0.15	-5.69	13.23	11.38
			8	50 - 57	-5.63 ± 0.16	13.28 ± 0.05	11.37 ± 0.15	-5.80	13.31	11.37
			16	58 - 73	-5.80 ± 0.29	13.15 ± 0.10	11.34 ± 0.09	-5.66	13.23	11.44
27	74 - 100	-5.56 ± 0.19	12.93 ± 0.13	11.34 ± 0.11	-5.31	13.05	11.33			

Tabla 3.21: Análisis jerárquico de la muestra 5 de estructuras no nativas de la BPTI mutada.

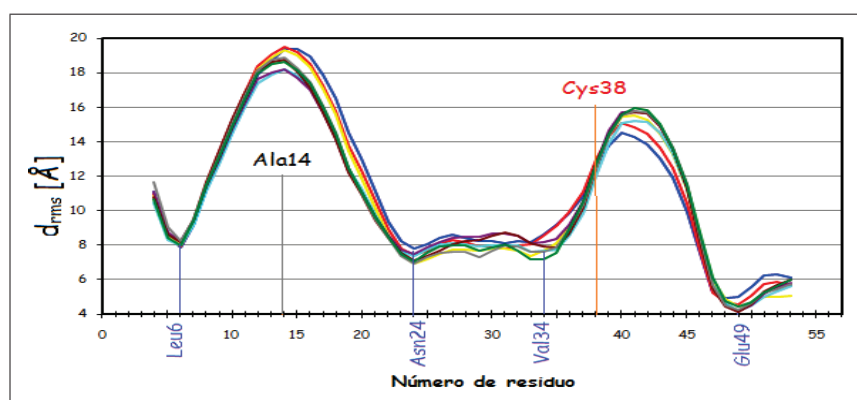


Figura 3.35: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 5 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
6	1.29	10	5	1 - 5	-6.61 ± 0.16	13.78 ± 0.12	14.51 ± 0.12	-6.70	13.79	14.48
			6	6 - 11	-6.44 ± 0.20	13.64 ± 0.09	14.36 ± 0.13	-6.59	13.60	14.35
			6	12 - 17	-6.77 ± 0.22	13.50 ± 0.11	14.16 ± 0.10	-6.80	13.53	14.10
			15	18 - 32	-7.34 ± 0.18	13.45 ± 0.07	14.21 ± 0.07	-7.24	13.36	14.07
			7	33 - 39	-7.32 ± 0.15	13.49 ± 0.08	14.08 ± 0.14	-7.58	13.49	14.14
			8	40 - 47	-7.16 ± 0.11	13.56 ± 0.14	14.17 ± 0.19	-7.21	13.62	14.18
			15	48 - 62	-7.31 ± 0.15	13.27 ± 0.08	13.77 ± 0.09	-7.47	13.34	13.85
			17	63 - 79	-6.91 ± 0.26	13.20 ± 0.08	13.78 ± 0.12	-6.76	13.22	13.79
			10	80 - 89	-7.09 ± 0.19	13.57 ± 0.09	14.12 ± 0.12	-7.44	13.47	13.98
			11	90 - 100	-7.27 ± 0.19	13.59 ± 0.08	13.92 ± 0.07	-7.53	13.60	13.95

Tabla 3.22: Análisis jerárquico de la muestra 6 de estructuras no nativas de la BPTI mutada.

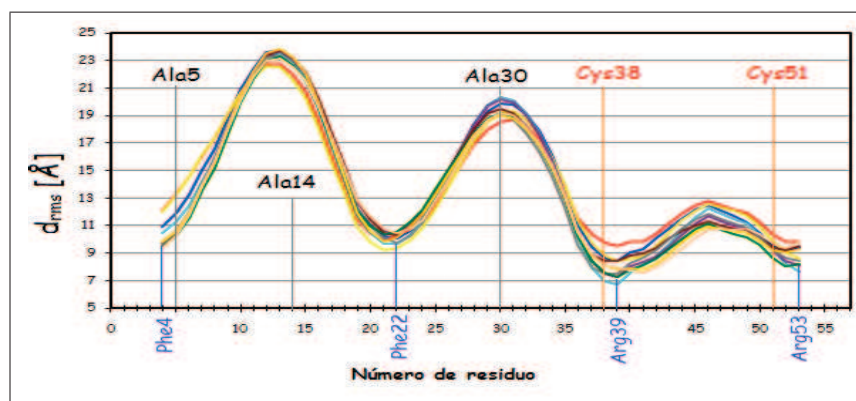


Figura 3.36: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 6 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parámetros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
7	1.40	8	10	1 - 10	-7.54 ± 0.20	13.40 ± 0.10	14.05 ± 0.18	-7.64	13.42	14.07
			7	11 - 17	-7.80 ± 0.18	13.39 ± 0.12	13.62 ± 0.15	-7.90	13.34	13.55
			12	18 - 29	-7.77 ± 0.21	13.58 ± 0.12	13.81 ± 0.12	-7.86	13.59	13.81
			24	30 - 53	-7.33 ± 0.34	13.70 ± 0.14	14.08 ± 0.16	-7.09	13.71	14.11
			12	54 - 65	-7.37 ± 0.17	14.12 ± 0.09	14.30 ± 0.10	-7.40	14.07	14.31
			7	66 - 72	-7.34 ± 0.15	14.16 ± 0.15	14.47 ± 0.16	-7.25	14.19	14.44
			11	73 - 83	-7.34 ± 0.18	14.23 ± 0.06	14.55 ± 0.05	-7.29	14.30	14.70
			17	84 - 100	-7.50 ± 0.18	13.97 ± 0.13	14.41 ± 0.15	-7.84	14.02	14.47

Tabla 3.23: Análisis jerárquico de la muestra 7 de estructuras no nativas de la BPTI mutada.

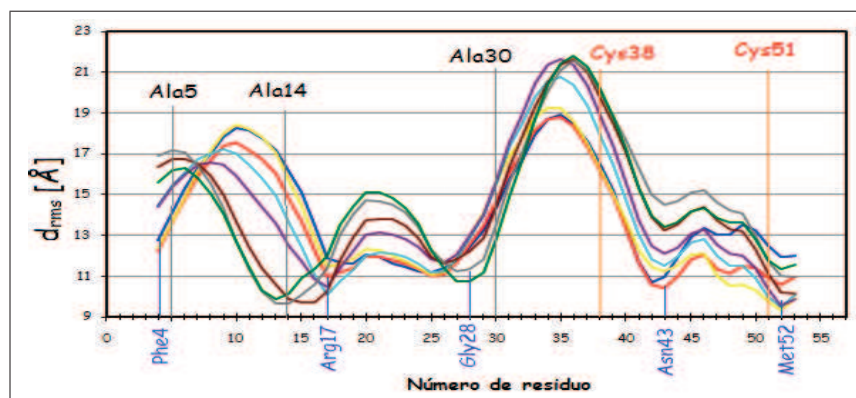


Figura 3.37: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 7 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
8	1.25	8	9	1 - 9	-6.46 ± 0.25	14.01 ± 0.08	13.78 ± 0.08	-6.69	14.02	13.78
			14	10 - 23	-6.45 ± 0.18	14.08 ± 0.11	13.70 ± 0.10	-6.39	14.09	13.74
			14	24 - 37	-6.13 ± 0.19	14.01 ± 0.11	13.46 ± 0.11	-6.07	14.04	13.43
			9	38 - 46	-6.00 ± 0.20	13.96 ± 0.07	13.60 ± 0.09	-6.05	14.05	13.68
			8	47 - 54	-5.97 ± 0.15	13.76 ± 0.10	13.58 ± 0.08	-6.05	13.80	13.67
			11	55 - 65	-6.32 ± 0.30	13.59 ± 0.09	13.55 ± 0.12	-6.00	13.62	13.56
			21	66 - 86	-6.75 ± 0.12	13.46 ± 0.10	13.60 ± 0.08	-6.75	13.52	13.70
			14	87 - 100	-6.74 ± 0.15	13.59 ± 0.11	13.70 ± 0.11	-6.29	13.63	13.76

Tabla 3.24: Análisis jerárquico de la muestra 8 de estructuras no nativas de la BPTI mutada.

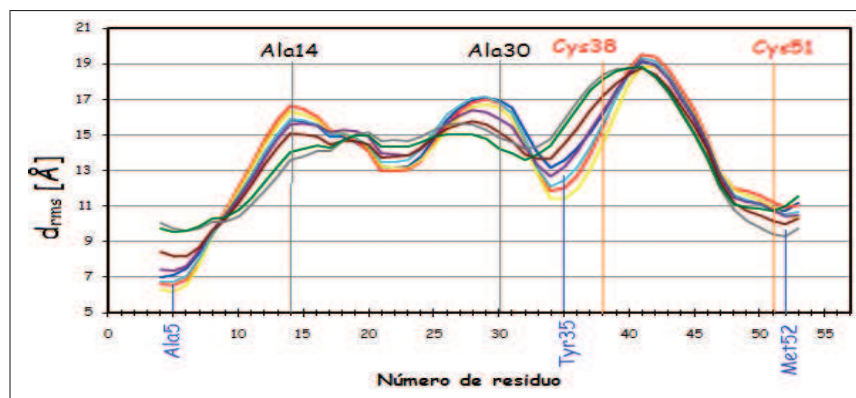


Figura 3.38: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 8 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
9	1.47	9	12	1 - 12	-6.78 ± 0.14	13.26 ± 0.08	12.15 ± 0.09	-6.97	13.27	12.18
			7	13 - 19	-6.73 ± 0.18	13.51 ± 0.10	12.25 ± 0.09	-6.96	13.50	12.19
			15	20 - 34	-6.73 ± 0.18	13.66 ± 0.12	12.09 ± 0.19	-6.90	13.66	11.96
			6	35 - 40	-7.00 ± 0.10	13.80 ± 0.11	12.23 ± 0.11	-7.19	13.79	12.38
			7	41 - 47	-6.79 ± 0.11	13.83 ± 0.11	11.97 ± 0.13	-6.79	13.86	12.05
			13	48 - 60	-6.86 ± 0.23	13.77 ± 0.11	12.10 ± 0.13	-6.97	13.80	11.99
			21	61 - 81	-6.69 ± 0.34	13.84 ± 0.08	12.42 ± 0.12	-6.68	13.89	12.39
			7	82 - 88	-6.60 ± 0.15	13.97 ± 0.07	12.57 ± 0.12	-6.81	13.96	12.57
			12	89 - 100	-6.91 ± 0.20	13.99 ± 0.08	12.29 ± 0.16	-6.97	14.10	12.32

Tabla 3.25: Análisis jerárquico de la muestra 9 de estructuras no nativas de la BPTI mutada.

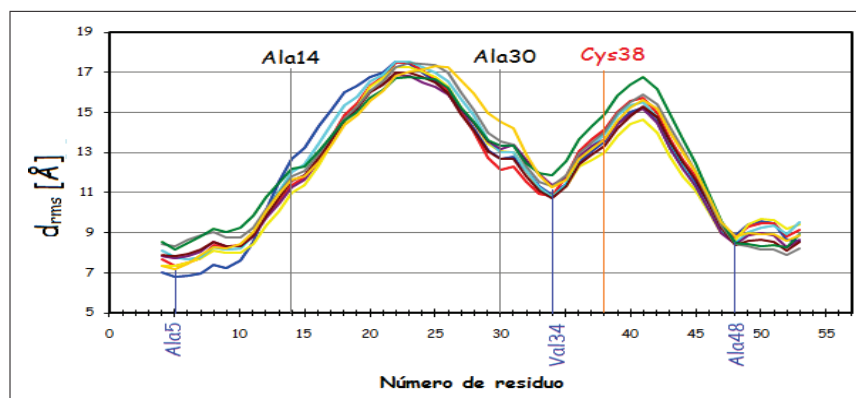


Figura 3.39: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 9 y la estructura calentada a 300K minimizada de la proteína mutada.

Muestra número	Distancia de corte [Å]	Número de grupos	Número de estructuras por grupo	Estructuras de cada grupo	Parametros promedio del grupo			Estructura promedio minimizada		
					Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]	Energía [aJ]	d_{rms} [Å]	Radio de giro [Å]
10	1.20	8	12	1 - 12	-6.54 ± 0.25	12.39 ± 0.06	12.52 ± 0.09	-6.66	12.40	12.50
			12	13 - 24	-6.81 ± 0.19	12.35 ± 0.11	12.32 ± 0.11	-6.88	12.34	12.25
			7	25 - 31	-6.72 ± 0.20	12.15 ± 0.09	12.16 ± 0.10	-6.65	12.08	12.13
			34	32 - 65	-7.08 ± 0.20	12.09 ± 0.07	11.88 ± 0.14	-7.32	12.10	12.01
			7	66 - 72	-7.27 ± 0.15	12.04 ± 0.05	11.79 ± 0.11	-7.37	11.98	11.83
			16	73 - 88	-7.60 ± 0.16	11.99 ± 0.06	11.70 ± 0.07	-7.72	11.97	11.67
			9	89 - 97	-7.67 ± 0.13	11.95 ± 0.06	11.81 ± 0.07	-7.35	11.95	11.85
3	97 - 100	-8.18 ± 0.10	11.88 ± 0.05	12.03 ± 0.06	-8.21	11.89	12.00			

Tabla 3.26: Análisis jerárquico de la muestra 10 de estructuras no nativas de la BPTI mutada.

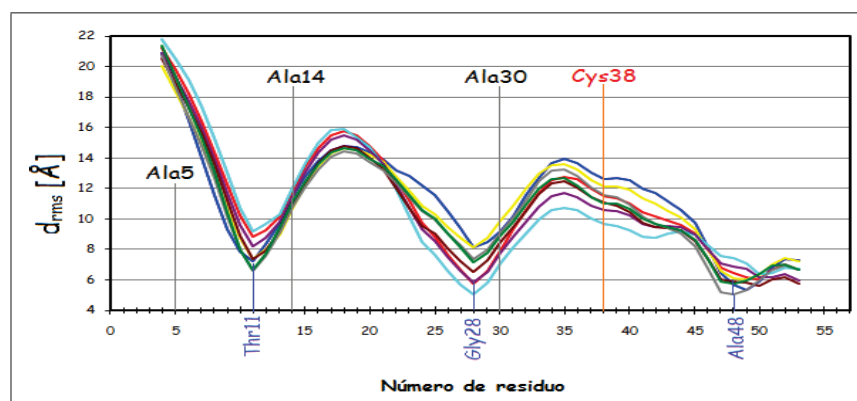


Figura 3.40: Gráfico de distancias rms entre las conformaciones obtenidas de la muestra 10 y la estructura calentada a 300K minimizada de la proteína mutada.

Número de muestra	1	2	3	4	5
Lazos con mayor movilidad	Lys15-Phe33 Phe33-Cys51	Phe4-Tyr23 Tyr23-Gly36 Gly36-Met52	Ala5-Ile19 Asn24-Thr32 Thr32-Arg39 Arg39-Ala48	Phe4-Tyr23 Tyr23-Gly37 Gly37-Met52	Len6-Asn24 Val34-Gln49
Número de muestra	6	7	8	9	10
Lazos con mayor movilidad	Phe4-Phe22 Phe22-Arg39 Arg39-Arg53	Phe4-Arg17 Arg17-Gly28 Gly28-Asn43 Asn43-Met52	Ala5-Tyr35 Tyr35-Met52	Ala5-Val34 Val34-Ala48	Arg1-Thr11 Thr11-Gly28 Gly28-Ala48

Tabla 3.27: Lazos móviles que caracterizan a las conformaciones no nativas de la BPTI mutada

Capítulo 4

Entorno Energético de la BPTI

4.1. Análisis de Coordenada Principal

En general el Entorno Energético de una proteína es una superficie muy irregular caracterizada por poseer un alto grado de rugosidad, asociada a gran cantidad de mínimos locales. En primera instancia a cada una de las muestras descritas en el capítulo anterior le corresponde un mínimo local en el Entorno, como muestra la figura 4.1-A, sin embargo, para un estudio más detallado de dicha superficie se analizaron las conformaciones accesibles a la proteína (determinadas en el capítulo 3); ya que cada una de las conformaciones a su vez determinan mínimos en la superficie del Entorno Energético (figura 4.1-B). Las conformaciones asociadas al estado nativo muestran la presencia de mínimos locales cercanos al global, mientras las conformaciones no nativas corresponden a mínimos locales más alejados. Los Entornos Energéticos de las proteínas natural y mutada se construyeron utilizando no solo las estructuras del análisis conformacional, sino también las estructuras del primer nanosegundo de equilibración de cada muestra y las estructuras de las simulaciones Dirigidas, en total 2300 estructuras. Para poder describir de una forma adecuada cada una de estas estructuras se realizó un Análisis de Coordenada Principal partiendo de la matriz de disimilitud Δ . Primero se superpusieron todas las estructuras de modo que la distancia rms entre cada par de estructuras sea mínima, utilizando la utilidad "alineat" incorporada en el programa VDM y se tomó en cuenta únicamente las carbonos alfa de las estructuras. Todos los análisis estadísticos se realizaron utilizando el programa DARWin5[68].

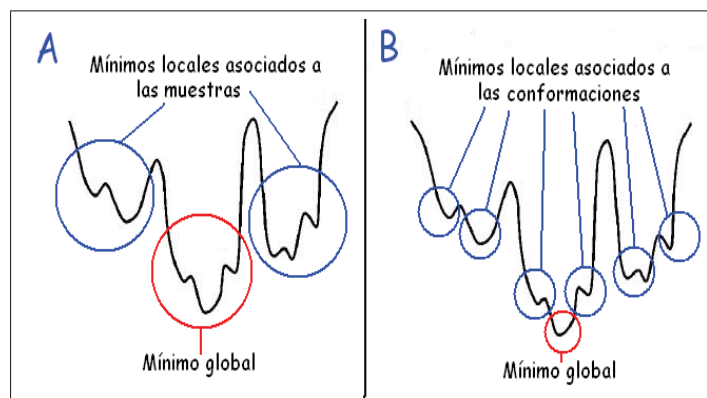


Figura 4.1: Representación esquemática del mínimo global y los mínimos locales asociados **A:** a las muestras de estructuras **B:** a las conformaciones accesibles a la proteína.

4.1.1. Estructuras de la BPTI natural

La figura 4.2 muestra el porcentaje explicado asociado a los valores propios correspondientes a cada coordenada principal obtenidos del ACooP. El porcentaje de proyección de la distribución original contenido en cada coordenada esta representado por este porcentaje explicado, por ende el valor propio más grande lleva mayor información y la proyección es mejor sobre él. Las dos primeras coordenadas principales obtenidas del análisis contienen aproximadamente el 57 % de la información en varianza de los datos originales; en términos prácticos este porcentaje de información es suficiente como para identificar las distancias entre los individuos de la muestra (en este caso estructuras de la BPTI natural) [59, 69].

Para la construcción del Entorno Energético asociado a la proteína es necesario asociar a cada una de la estructuras su energía. Se realizó una traslación de las coordenadas obtenidas del ACooP, de modo que la estructura con menor energía de toda la muestra se encuentre ubicada en el origen de coordenadas. En la figura 4.3-A se encuentran representada la energía de cada una de las estructuras de las muestras asociadas al estado nativo y las no nativas en función de las dos primeras coordenadas generalizadas.

Mientras que en la figura 4.4-A están representadas aquellas obtenidas a partir de las simulaciones dirigidas. La formación de grupos (correspondientes a cada una de las muestras) así como la convergencia de las estructuras obtenidas de la dinámica dirigida hacia la muestra de las asociadas al estado nativo es evidente en las figuras 4.3 y 4.4.

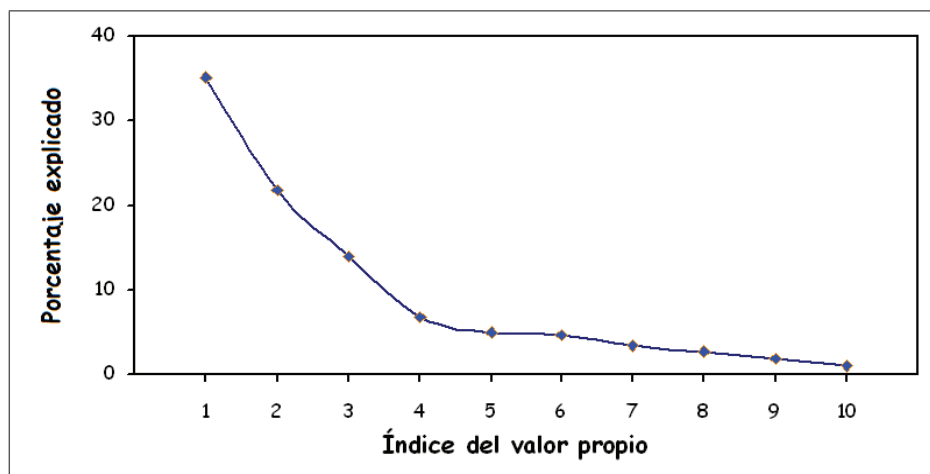


Figura 4.2: Porcentaje explicado de la proyección de las estructuras naturales en función del índice del valor propio asociado. El índice de los valores propios se colocó de modo que el uno corresponde al mayor, y los siguientes en orden descendente. La primera coordenada principal está asociada al mayor valor propio, la segunda coordenada al segundo y así sucesivamente.

Las figuras 4.3-C y 4.3-D muestran la representación de la energía de la BPTI natural en función de cada una de las dos coordenadas principales de las estructuras de la muestra (no se encuentran representadas las obtenidas en la dinámica dirigida); las estructuras de las muestras obtenidas son similares entre sí (se encuentran cercanas en la representación del ACooP y sus valores de energía son cercanos también), por lo que se las puede asociar a mínimos locales en la superficie del Entorno Energético.

La figura 4.4, muestra la energía de las estructuras obtenidas de la dinámica dirigida en función de las dos coordenadas principales que las caracterizan; se puede ver claramente la convergencia desde cada una de las muestra hacia la asociada al estado nativo, además es evidente la existencia de barreras de energéticas entre los mínimos locales y el global. Estas barreras están en el orden de 3.39 ± 1.09 aJ (818.84 ± 263.29 kT); tomando valores desde 1.30 aJ (314.10 kT), para la simulación realizada desde la muestra 1 al mínimo global, hasta 4.81 aJ (116.84 kT), para la que se realizó desde la muestra 3; los valores de la altura de las barreras energéticas asociadas a estas simulaciones están descritas en la tabla 4.1.

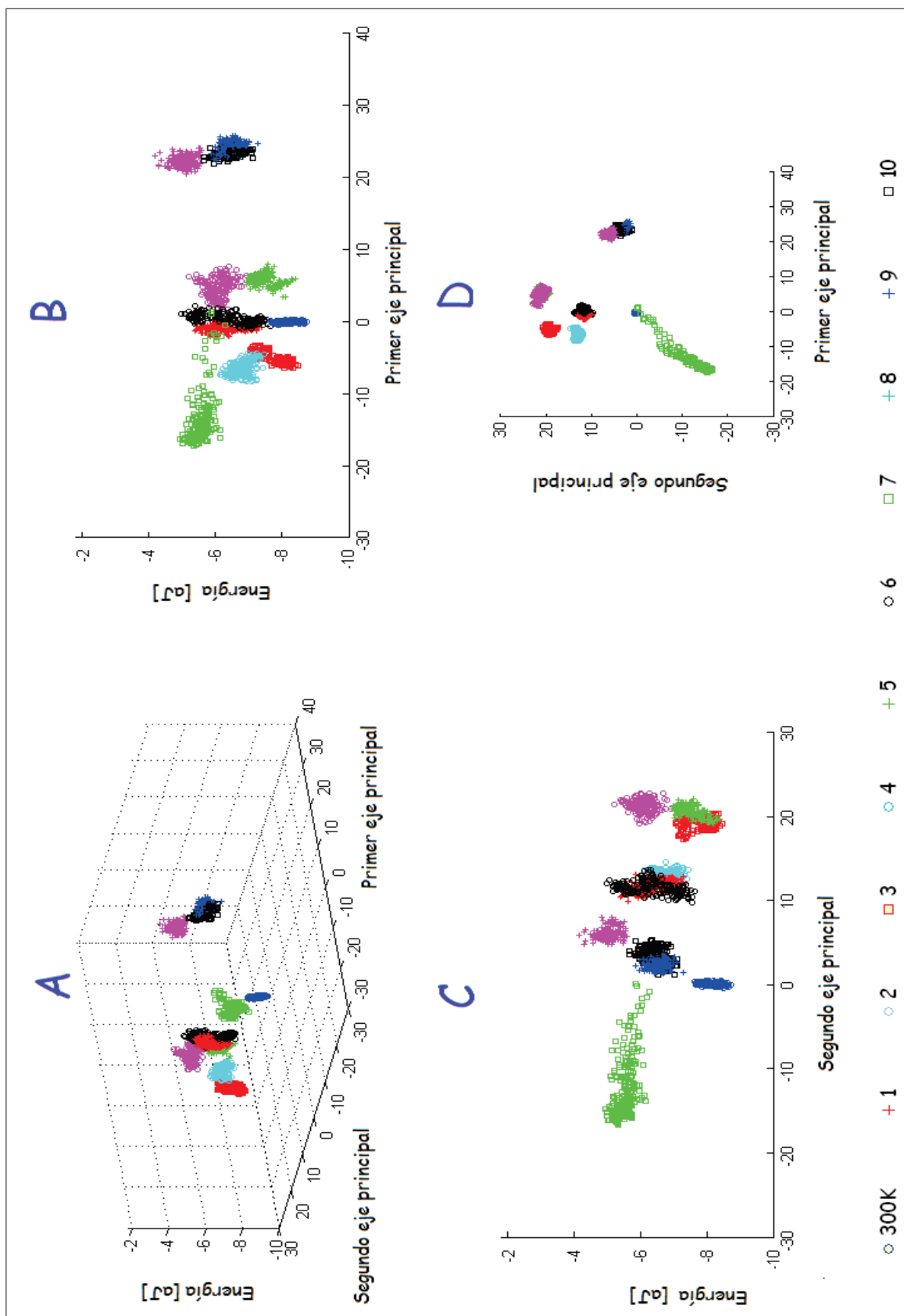


Figura 4.3: A: Energía de las estructuras de la proteína natural (asociadas al mínimo global y a mínimos locales) en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

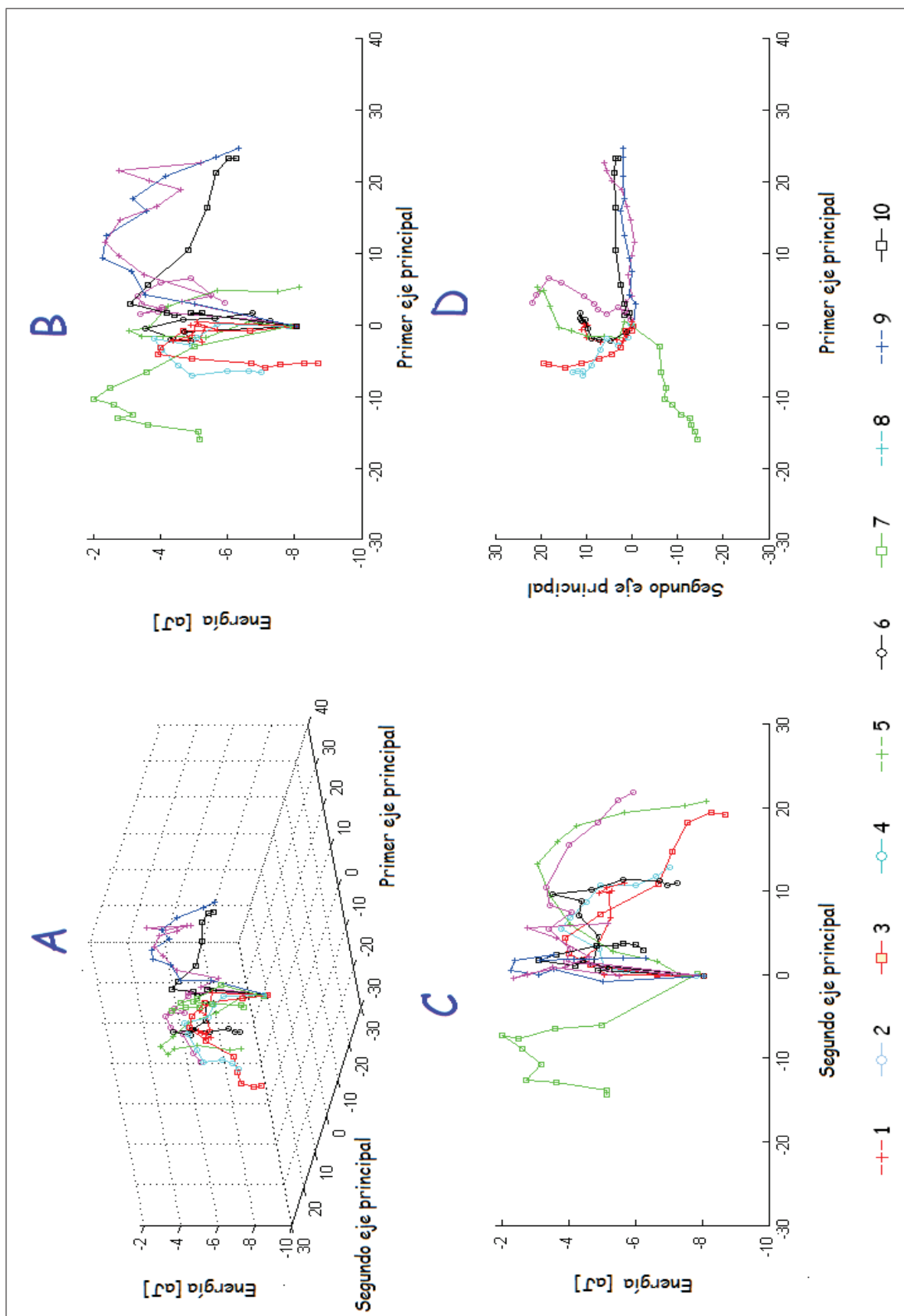


Figura 4.4: A: Energía de las estructuras de las simulaciones dirigidas de la proteína natural (asociadas transiciones desde mínimos locales al mínimos global) en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

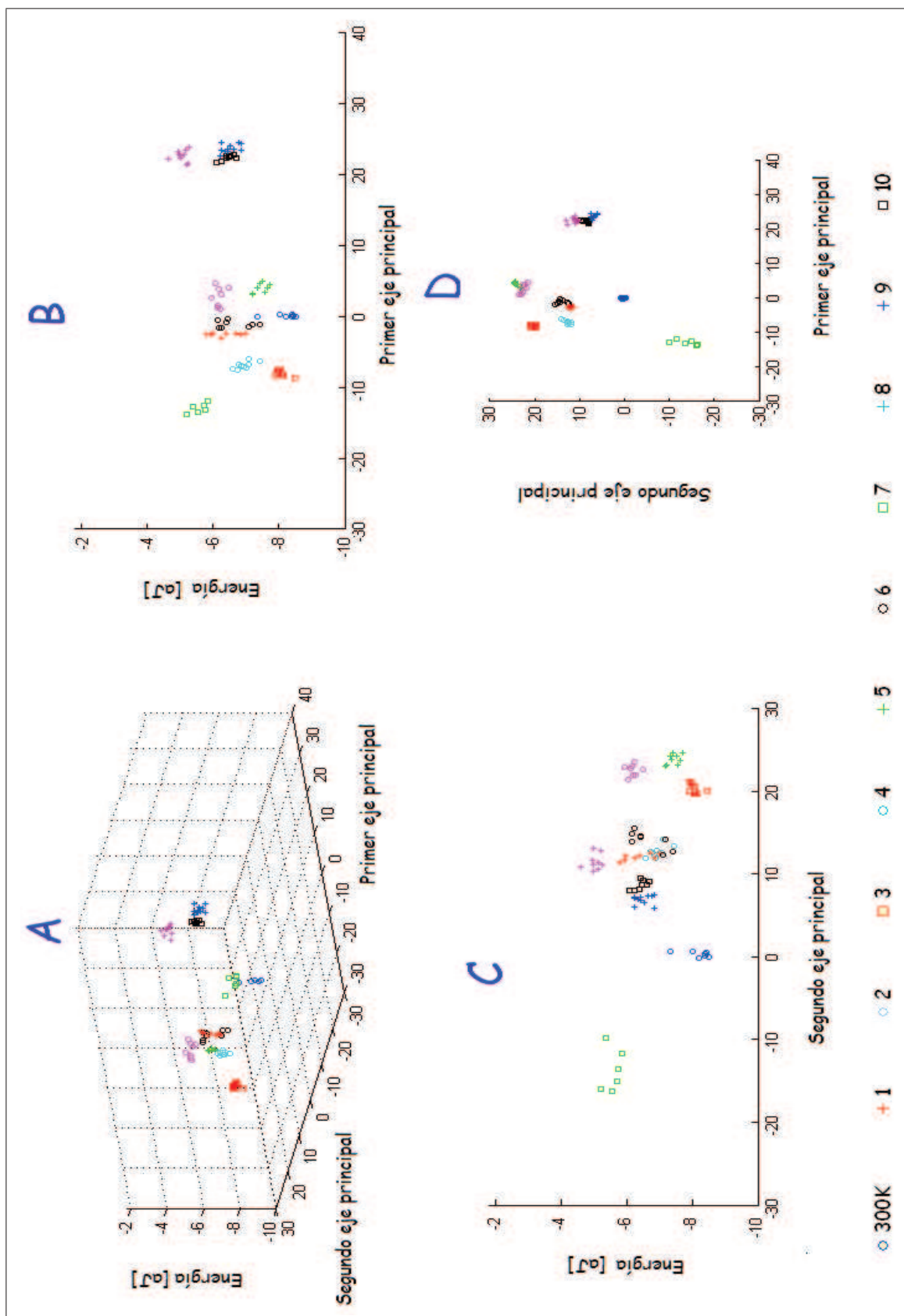


Figura 4.5: A: Energía de las conformaciones de la proteína natural en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

La figura 4.5 muestra los resultados del Análisis de Coordenada Principal realizado con las conformaciones nativas y no nativas de la BPTI natural, 95 estructuras promedio minimizadas producto del análisis de grupos. Las dos primeras coordenadas principales obtenidas del análisis contienen aproximadamente el 55 % de la información en varianza de los datos originales, valor muy próximo al obtenido del análisis de la muestra total de estructuras.

Las estructuras promedio (conformaciones) se ubican en regiones cercanas a las que se encuentran las muestras correspondientes de donde se obtuvieron, tanto en el plano correspondiente a las coordenadas principales como al tomar en cuenta su energía; lo que implica que esta reducción de la muestra es válida, es decir, se podría describir aproximadamente de la misma forma el espacio conformacional, y consecuentemente el Entorno Energético, tanto con las conformaciones obtenidas del análisis de grupos como con el grupo completo de estructuras.

4.1.2. Estructuras de la BPTI mutada

El procedimiento descrito en el apartado 4.1.1. se aplicó también para la estructuras obtenidas a partir del sistema de la proteína mutada, es decir, partiendo de la matriz de disimilitud obtenida entre las 2300 estructuras, se realizó un Análisis de Coordenada Principal de estas. La figura 4.6 muestra los porcentajes explicados asociados a cada valor propio obtenidos del ACooP en función del índice de cada valor propio asociado a las proyecciones en los eje principales de coordenadas, es decir, el porcentaje de proyección de la distribución original contenido en cada coordenada. Las dos coordenadas principales obtenidas del análisis contienen aproximadamente el 44 % de la información en varianza de los datos originales; en términos prácticos con este porcentaje de información se puede identificar las distancias entre los individuos de la muestra (en este caso estructuras de la BPTI mutante), aunque presenta menos precisión que el análisis realizado para la proteína natural.

La figura 4.7, muestra la representación de la energía de las estructuras asociadas al mínimo global y a mínimos locales (2200 estructuras) en función de las dos primeras coordenadas principales trasladadas. La traslación se realizó de modo que la estructura con menor energía de la muestra se encuentre en el origen de coordenadas. A diferencia de lo encontrado para la proteína natural, las estructuras

de cada muestra no presentan un comportamiento tan homogéneo, es decir, no se distingue tan claramente (como para la natural) cada muestra. Además, las dimensiones de los mínimos locales asociados a estas muestras tienen mayores dimensiones (en el sistema de coordenadas principales) que los obtenidos para la estructura no mutada. Las figuras 4.7-C y 4.7-D muestran que, al igual que para la natural, las estructuras asociadas al estado nativo poseen menor energía que las otras muestras.

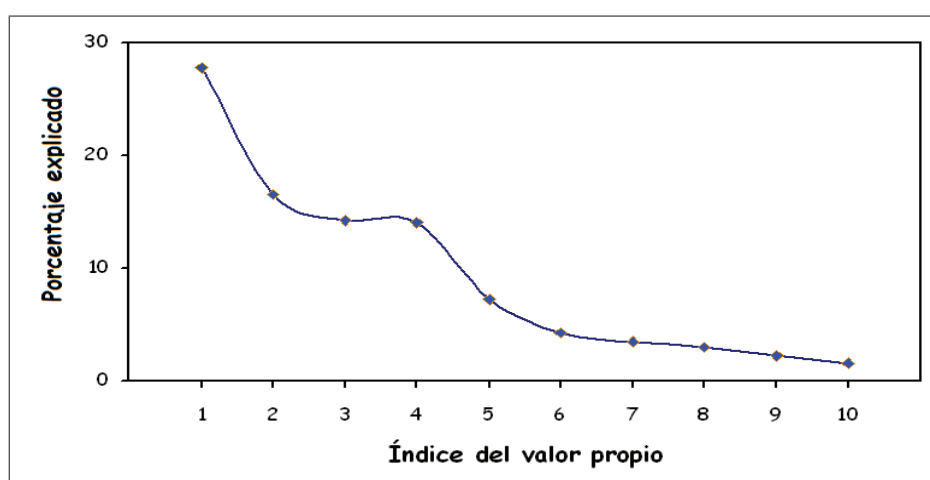


Figura 4.6: Porcentaje explicado de la proyección de las estructuras mutadas en función del índice del valor propio asociado. El índice de los valores propios se colocó de modo que el uno corresponde al mayor, y los siguientes en orden descendente. La primera coordenada principal está asociada al mayor valor propio, la segunda coordenada al segundo y así sucesivamente.

En la figura 4.8 es evidente la convergencia de las estructuras producto de la simulación dirigidas hacia el mínimo global, así como la presencia de barreras de energía entre éste y los mínimos locales correspondientes. Estas barreras están en el orden de 3.03 ± 0.91 aJ (731.88 ± 219.81 kT); y toman valores desde 1.47 aJ (355.07 kT), para la simulación realizada desde la muestra 1 al mínimo global, hasta 4.49 aJ (1084.54 kT), para la realizada desde la muestra 6; los valores de la altura de las barreras energéticas asociadas a estas simulaciones están descritas en la tabla 4.1. En promedio las barreras energéticas presentes en el sistema mutado son menores que las presentes en el sistema natural en aproximadamente 0.36 aJ (86.96 kT).

En la figura 4.9 se describen los resultados del Análisis de Coordenada Principal realizado con las conformaciones nativas y no nativas de la BPTI mutada, 99 estructuras promedio minimizadas producto del análisis de grupos. Las dos

coordenadas principales obtenidas del análisis contienen aproximadamente el 41 % de la información en varianza de los datos originales, valor muy próximo al obtenido del análisis de la muestra total de estructuras; sin embargo estos valores son menores que los obtenidos con las estructuras del sistema no mutado. Las estructuras promedio se ubican en regiones cercanas a las que se encuentran las muestras correspondientes de donde se obtuvieron, tanto en las coordenadas principales como al tomar en cuenta su energía; lo que implica que esta reducción de la muestra es válida, es decir, se podría describir aproximadamente de la misma forma el espacio conformacional, y consecuentemente el Entorno Energético, tanto con las conformaciones obtenidas del análisis de grupos como con el grupo completo de estructuras.

Simulación Dirigida	Barreras Energéticas	
	[kT]	
	Natural	Mutada
1	314.01	355.07
2	780.19	891.30
3	1161.84	707.71
4	630.43	618.36
5	1224.64	637.68
6	900.97	1084.54
7	758.45	971.01
8	685.99	541.06
9	978.26	862.32
10	765.70	659.42
PROM	818.84	731.88
DESV EST	263.29	219.81

Tabla 4.1: Altura de la barreras energéticas asociadas a las simulaciones dirigidas. El número simulación dirigida se asigno tomando en cuenta el número de muestra al que pertenece la estatura que se tomo como punto de partida para la simulación, esto es, la i -ésima simulación dirigida corresponde a la realizada desde la estructura de la i -ésima muestra hasta el mínimo global.

Luego de analizar los parámetros con los que se puede describir el Entorno Energético asociado a las proteína natural y mutada (coordenadas principales y energía), se puede confirmar la existencia de una región en el Entorno Energético, caracterizada por poseer estructuras con menor energía que todas la además y que se encuentra separada de otras regiones (mínimos locales) por barreras energéticas. Además, estos mínimos locales pueden caracterizarse de una forma adecuada mediante las conformaciones descritas en el capítulo 3.

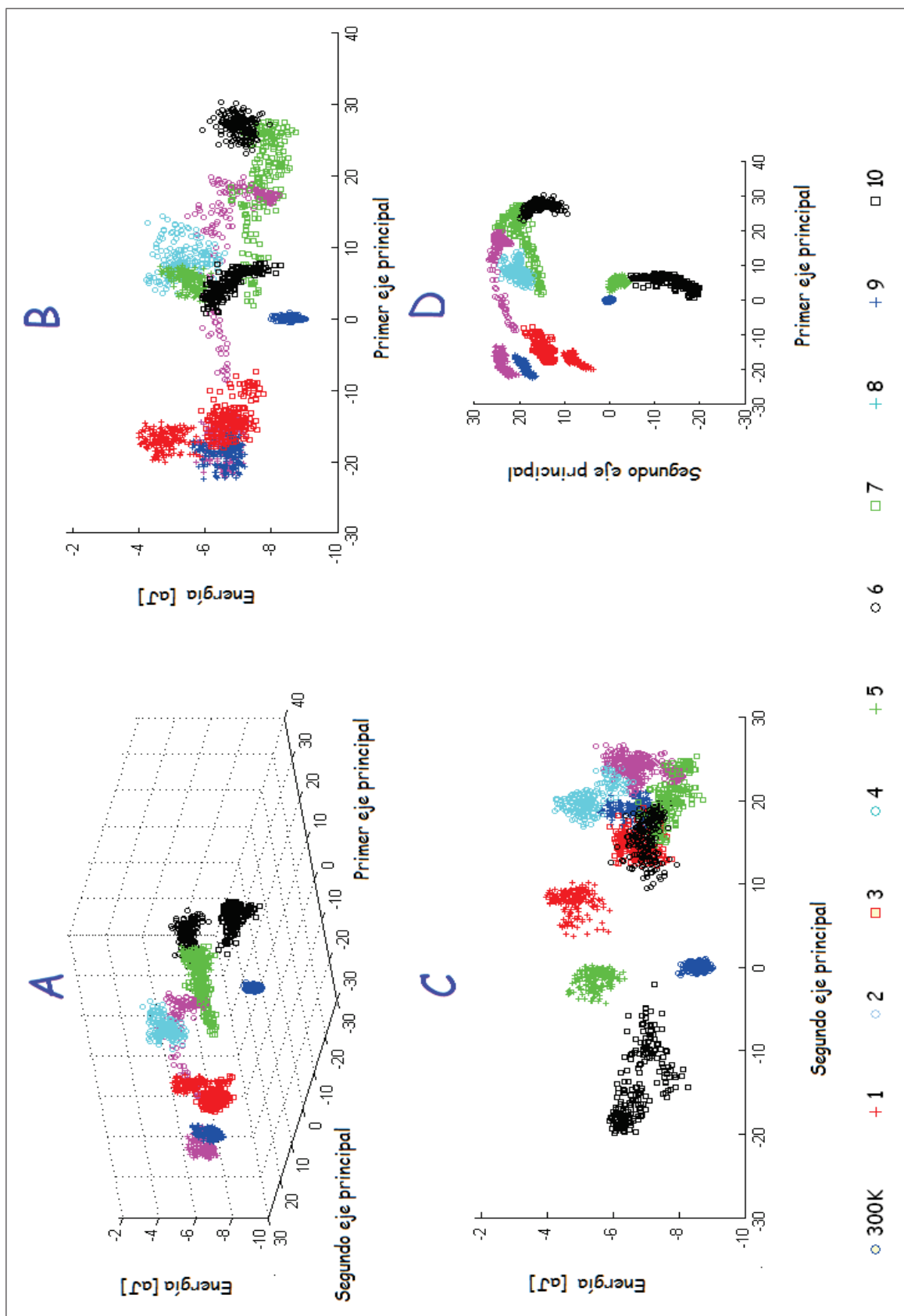


Figura 4.7: A: Energía de las estructuras de la proteína mutada (asociadas al mínimo global y a mínimos locales) en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

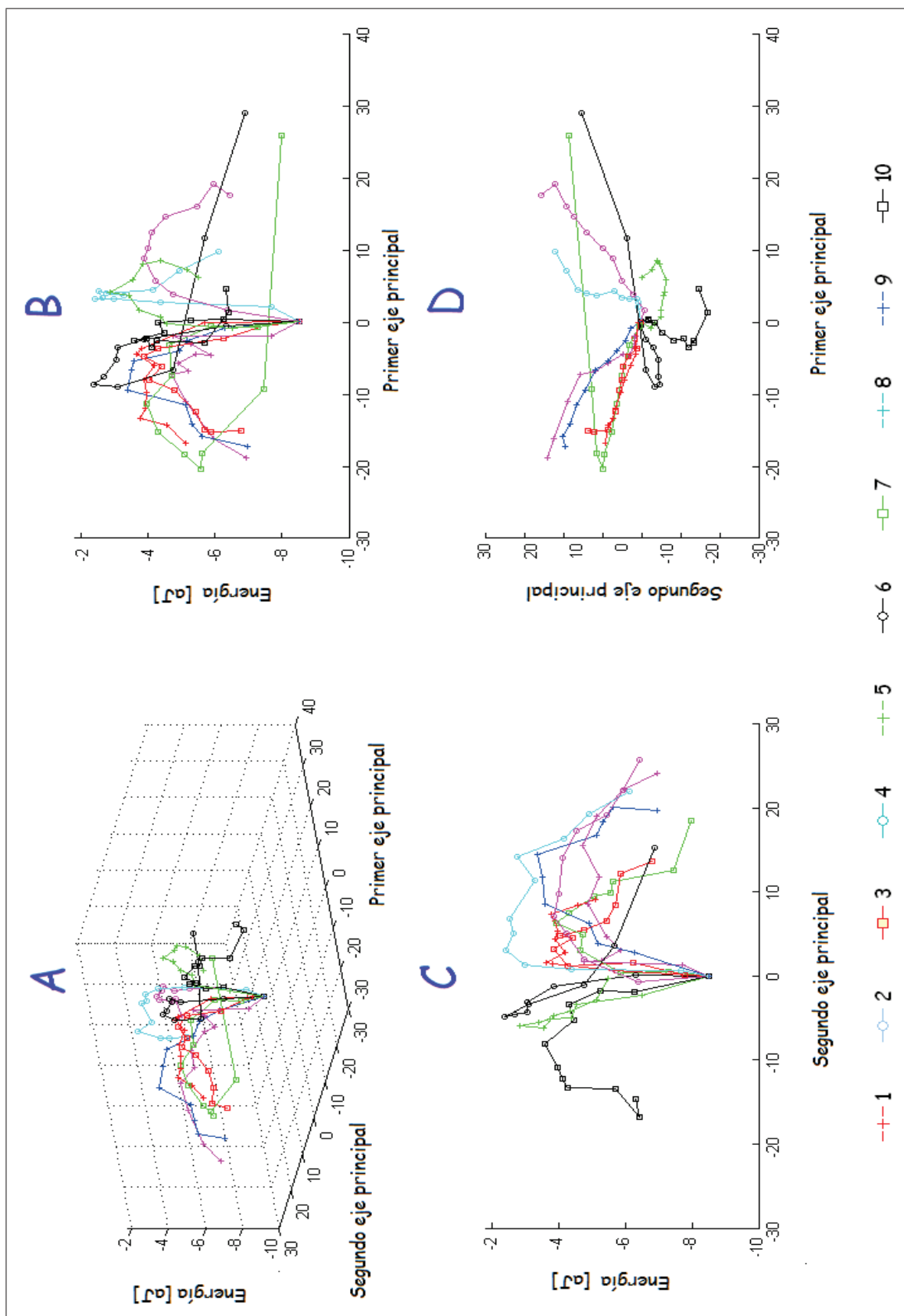


Figura 4.8: A: Energía de las estructuras de las simulaciones dirigidas de la proteína mutada (asociadas transiciones desde mínimos locales al mínimos global) en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

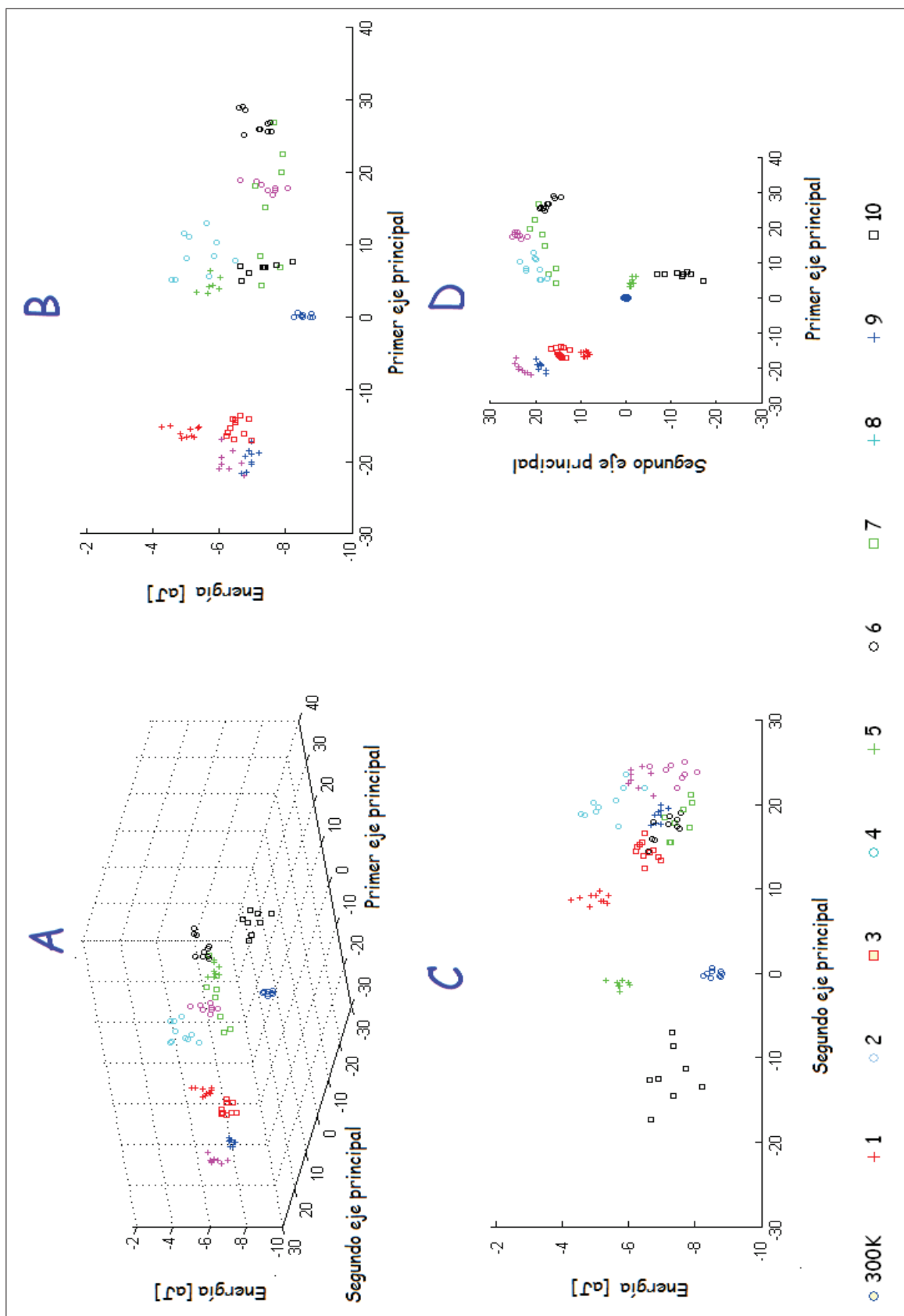


Figura 4.9: A: Energía de las conformaciones de la proteína mutada en función de las dos primeras coordenadas principales. B, C, D: proyecciones en los planos correspondientes.

4.2. Construcción del Entorno Energético

Para la construcción de esta superficie tanto para la proteína natural como para mutada, se utilizaron los resultados del ACooP obtenidos en los apartados 4.1.1. y 4.1.2. con el fin de describir las estructuras de cada muestra, y las energías asociadas a cada una de ellas. La construcción de la superficie se realizó con el programa MATLAB 7.7.0.471, utilizando la aplicación “griddata”, con el método “cubic”; esta aplicación realiza una interpolación cúbica basada en triangulación de puntos. Esta interpolación busca encontrar la superficie que mejor se acople a un arreglo tridimensional de puntos de la forma $z = z(x, y)$, donde x , y , z son las coordenadas de los puntos.

Primero se aplica una triangulación de Delaunay (formación de una red de triángulos que cumplen con la condición de que la circunferencia circunscrita de cada triángulo de la red no debe contener ningún vértice de otro triángulo) a los datos; luego se encuentra un polinomio cúbico en cada triángulo para interpolar los datos; este método produce una superficie suavizada óptima que se acopla de la mejor forma a los puntos originales[70]. En nuestro caso, las coordenadas x, y corresponden a las obtenidas del ACooP y z es la energía asociada a cada estructura.

4.2.1. Entorno Energético de la BPTI natural

La figura 4.10, muestra la superficie del Entorno Energético de la BPTI natural, construida partir de las coordenadas principales del ACooP y las energías asociadas a cada una de las 2300 estructuras de la muestra. Los valores de energía de las estructuras es encuentran entre -8.80 aJ (2125.60 kT) y -1.92 aJ (463.77 kT).

En general la superficie presenta un alto grado de rugosidad, como se puede ver en las figuras 4.10-A, 4.10-B y 4.10-C, es decir, está caracterizado por un alto número de conformaciones proteínicas. Se observa una profunda y estrecha cavidad, que corresponde al mínimo global de la superficie y está determinado por las estructuras asociadas al estado nativo de la BPTI natural. El mínimo global tiene forma de embudo y está separado de los mínimos locales asociados a las muestras por barreras energéticas entre 1.3 aJ (314.01 kT) y 4.49 aJ (1084.54 kT).

El Entorno Energético presenta dos mínimos locales profundos alrededor de los -7.68 aJ (-1855.07 kT) y -8.0 aJ (-1932.37 kT) respectivamente; estos mínimos que corresponden a las muestras 3 y 5 (figuras 4.3 y 4.10) se encuentran bastante alejados del mínimo global y están separados de este por barreras de energéticas del orden 4.80 y 5.07 aJ (1159.42 y 1224.64 kT). Se puede observar también la presencia de tres mínimos locales menos pronunciados, del orden de los -6.72 aJ (-1623.19 kT), alejados del global que corresponde a las estructuras de las muestras 2, 1-6, y 9-10 y separados del mínimo global por barreras del orden de los 3.2 aJ (772.95 kT) y entre sí (entre ellos y los mínimos mas pronunciados) por pequeñas barreras del orden de los 1.6 aJ (386.47 kT). En general, los mínimos presentes en la superficie son bastante anchos y la parte interior de cada uno posee varias rugosidades, lo que indica que los intermediarios asociados a estos están determinados por un grupo de estructuras ligeramente diferentes y no por una sola estructura bien definida.

4.2.2. Entorno Energético de la BPTI mutada

La figura 4.11, muestra la superficie del Entorno Energético de la BPTI mutada, construida partir del las coordenadas principales del ACooP y las energías asociadas a cada una de las 2300 estructuras de la muestra. Los valores de energía del estructuras se encuentran entre -9.12 aJ (-2202.90 kT) y -2.4 aJ (-579.71 kT).

En general la superficie presenta un alto grado de rugosidad, mayor que el presente en la inherente a la proteína natural, como se puede ver en las figuras 4.11-A, 4.11-B y 4.11-C, es decir, está caracterizado por un alto número de conformaciones proteínicas. El mínimo global de la superficie tiene un valor aproximado de -9.12 aJ (-2202.90 kT) y está determinado por las estructuras asociadas al estado nativo de la BPTI mutada. Este mínimo global tiene forma de embudo y está separado de los mínimos locales asociados a las muestras por barreras energéticas entre 1.47 y 5.06 aJ (355.07 y 1222.22 kT); en comparación con la BPTI natural, este es mucho más ancho y ligeramente más profundo.

El Entorno Energético presenta un mínimo local profundo alrededor de los -8.0 aJ (-1932.37 kT), asociado a la muestra 3, (figuras 4.7 y 4.11) que se encuentra bastante alejado del mínimo global y está separado de este por una barrera energética del orden 2.92 aJ (705.31 kT). También se puede observar la presencia

de cinco mínimos locales ligeramente menos pronunciados, del orden de las - 7.36 aJ (1777.78 kT); alejados del global que corresponde a las estructuras de las muestras 3,5, 8, 4-6-7, y 10; el primero está bastante cerca del mínimo global pero se encuentra separado de este por una barrera energética de aproximadamente 2.64 aJ (637.38 kT), los otros tres están bastante alejados del global y separados de éste por barreras entre 2.4 y 4.48 aJ (579.71 y 1082.13 kT). Las barreras energéticas que separan los mínimos locales están en el orden de los 1.28 aJ (309.18 kT), es decir, en promedio menores que las presente en la proteína natural. La parte interior de cada mínimo posee varias rugosidades, lo cual es consistente con el hecho de que los intermediarios asociados a estos están determinados por un grupo de estructuras ligeramente diferentes y no por una sola estructura bien definida.

4.2.3. Comparación de los Entorno Energético inherentes a la BPTI natural y mutada

Como muestran las figuras 4.10 y 4.11, la superficie del Entorno Energético asociada a la BPTI mutada presenta un grado de rugosidad mucho mayor que para la estructura natural; la mayor dinámica que posee la estructura mutada por la no presencia de los enlaces disulfuro, hace que esta posea un mayor número de conformaciones accesibles, por ende mayor número de mínimos en su Entorno energético. El mínimo global de la estructura natural es más angosto y ligeramente menos profundo que el de la mutada. Las barreras energéticas presentes en el Entorno de la BPTI mutada son de menor tamaño, esto implica que las transiciones entre conformaciones son más factibles que para esta estructura .

Las estructuras producidas a partir del sistema mutado están dispersas en una cantidad mayor del espacio conformacional (determinado en este caso por las dos coordenadas principales). Las estructuras de cada muestra tienen un mayor grado de similitud en la proteína natural, por ello los grupos de estructuras correspondientes son más compactos. En el Entorno Energético de la natural se tiende a formar cavidades estrechas y profundas, mientras que en el de la mutada se forman cavidades más anchas y menos profundas. Esto muestra que la dinámica de la proteína natural está restringida por la presencia de los enlaces disulfuro de su estructura.

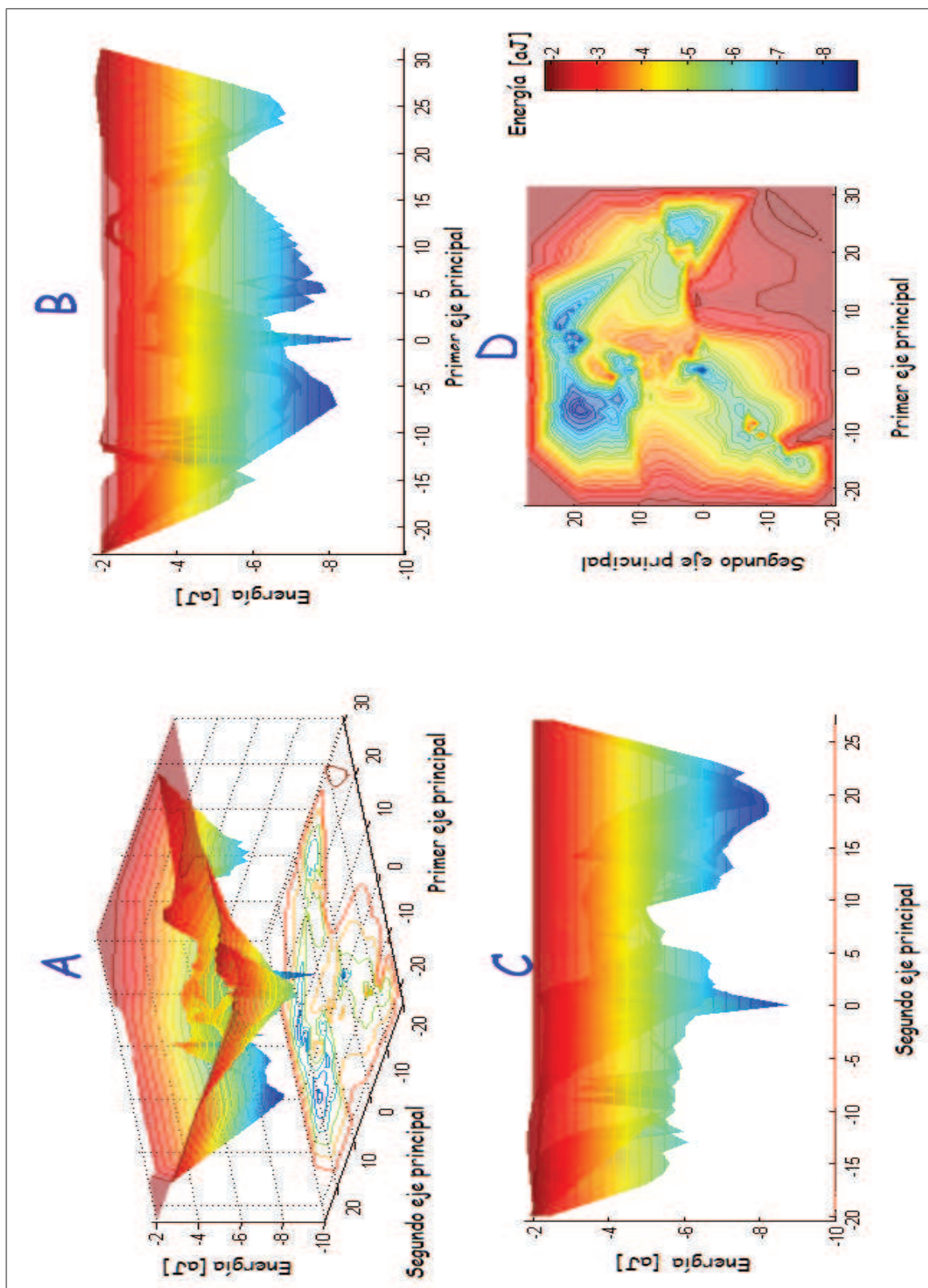


Figura 4.10: Entorno Energético de la BPTI natural. **A:** Representación tridimensional. **B, C:** Proyección en los planos energía en función de la primera y segunda coordenadas respectivamente. **D:** Proyección en el plano de las coordenadas principales.

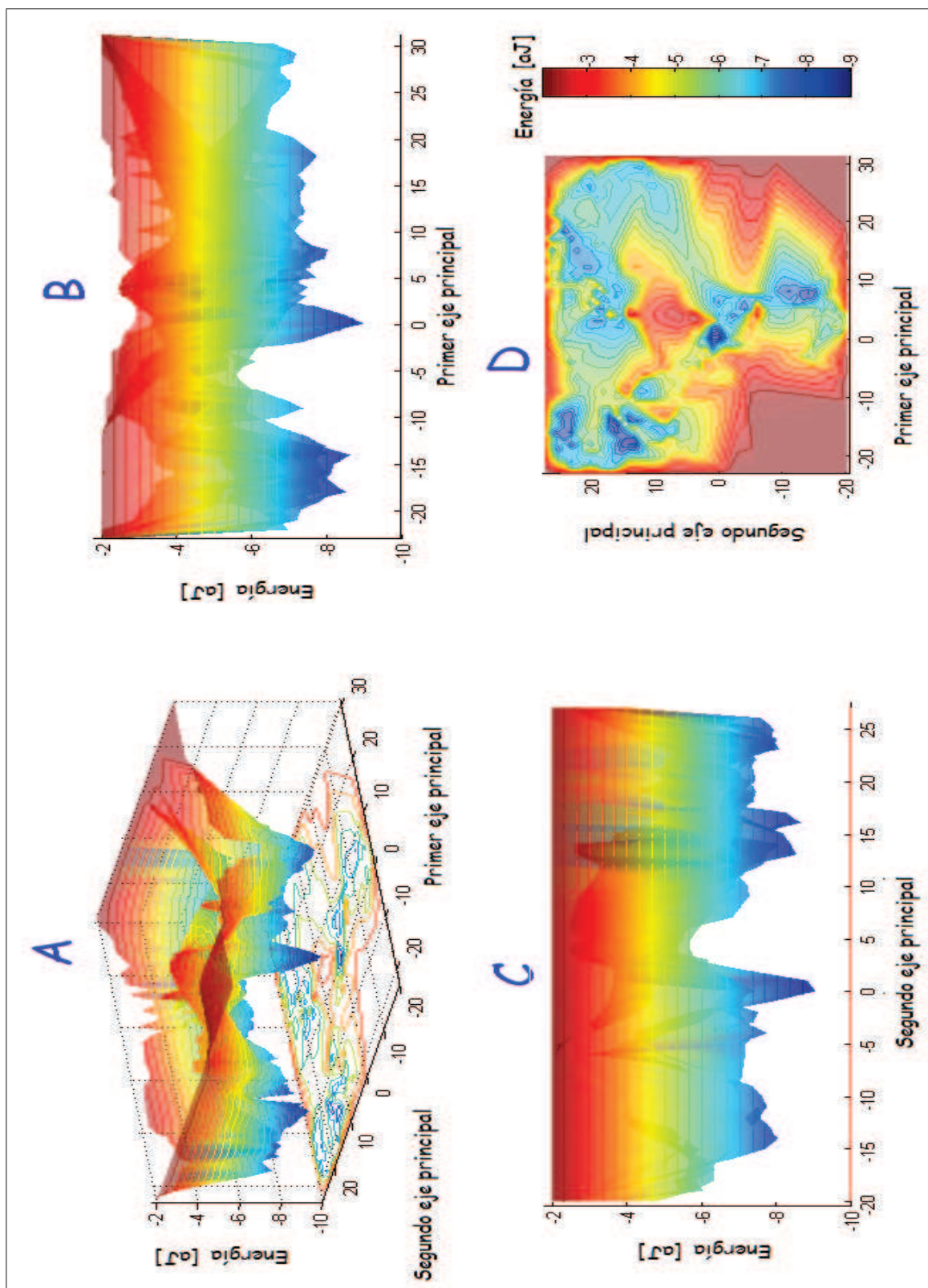


Figura 4.11: Entorno Energético de la BPTI mutada. **A:** Representación tridimensional. **B, C:** Proyección en los planos energía en función de la primera y segunda coordenadas respectivamente. **D:** Proyección en el plano de las coordenadas principales.

Capítulo 5

Conclusiones

Con el propósito de desarrollar un método para caracterizar el Entorno Energético de proteínas a 300 K se analizaron conformaciones obtenidas mediante simulaciones de Dinámica Molecular de la proteína BPTI (Bovine Pancreatic Trypsin Inhibitor) y de una estructura obtenida luego de realizar mutaciones que alteren los enlaces disulfuro presentes en esta proteína. Debido a que estos enlaces dan estabilidad a la proteína, se estudió como su ausencia influye en las conformaciones accesibles a esta y por ende en el Entorno energético asociado. Ya que la descripción completa del espacio conformacional de una proteína a temperatura ambiente no es computacionalmente alcanzable, se utilizaron técnicas como simulaciones a altas temperaturas y simulaciones dirigidas con el fin obtener una descripción amplia de los estados conformacionales tanto de la proteína en su estado natural como de su estructura mutada.

Se adquirieron tres grupos de estructuras utilizando diferentes técnicas de Dinámica Molecular. El primer grupo se obtuvo mediante simulaciones a 300 K y las estructuras resultantes están asociadas al estado nativo de la proteína; el segundo grupo de estructuras se obtuvo utilizando el comportamiento de proteínas a altas temperaturas, se calentó el sistema a 1000 K y se lo equilibró a esta temperatura, luego se seleccionaron ciertas estructuras que fueron enfriadas de nuevo a 300 K y mediante simulaciones a 300 K se obtuvo un grupo de estructuras asociadas a conformaciones no nativas; finalmente un tercer grupo de estructuras se adquirió mediante simulaciones dirigidas, estas están asociadas a transiciones entre estados conformacionales y proveen información acerca de las barreras energéticas

presentes en el Entorno Energético. La optimización de todas las estructuras se realizó mediante protocolos de minimización energética y las simulaciones fueron realizadas con la plataforma NAMD [44].

Muestra	Natural		Mutada	
	Número de conformaciones	Lazos con mayor movilidad	Número de conformaciones	Lazos con mayor movilidad
300 K	7	Thr11-Tyr21 Tyr21-Gln31 Gln31-Asn44	9	Thr11-Tyr21 Tyr21-Gln31 Gln31-Asn44
1	8	Cys5-Lys15 Gln31-Arg42 Arg42-Glu49	10	Lys15-Phe33 Phe33-Cys51
2	10	Cys5-Cys14 Cys14-Ala27 Arg39-Cys50	10	Phe4-Tyr23 Tyr23-Gly36 Gly36-Met52
3	10	Cys5-Pro13 Pro13-Thr32 Thr32-Ala40 Ala40-Glu49	11	Ala5-Ile19 Asn24-Thr32 Thr32-Arg39 Arg39-Ala48
4	8	Cys5-Cy14 Cys14-Lys26 Arg39-Cys51	8	Phe4-Tyr23 Tyr23-Gly37 Gly37-Met52
5	9	Cys5-Thr11 Thr11-Cys38 Cys38-Cys51	8	Len6-Asn24 Val34-Gln49
6	8	Cys5-Arg20	10	Phe4-Phe22 Phe22-Arg39 Arg39-Arg53
7	6	Phe4-Cys14 Cys14-Phe33 Phe33-Met52	8	Phe4-Arg17 Arg17-Gly28 Gly28-Asn43 Asn43-Met52
8	10	Cys5-Gly28 Gly28-Ala40 Ala40-Arg53	8	Ala5-Tyr35 Tyr35-Met52
9	11	Tyr10-Arg17 Arg17-Cys30 Cys38-Arg53	9	Ala5-Val34 Val34-Ala48
10	8	Pro8-Arg17 Arg17-Ala27 Ala27-Cys38 Cys38-Arg53	8	Arg1-Thr11 Thr11-Gly28 Gly28-Ala48

Tabla 5.1: Conformaciones accesibles a la BPTI natural y mutada. Número de conformaciones por muestra y lazos móviles que las caracterizan

En general, se puede observar que la no presencia de los enlaces disulfuro provoca que la proteína se vuelva más dinámica, es decir, pueda visitar más conformaciones; esto implica además que las barreras de energía son más fáciles de superar para la proteína mutada. Mediante el análisis de grupos (cluster analysis) se obtuvieron las conformaciones accesibles tanto para la BPTI natural como para la mutada. La BPTI natural visita 7 conformaciones asociadas al estado nativo en el intervalo de simulación de 1ns, cada una de estas conformaciones está caracterizada por la movilidad de ciertos lazos específicos, como se muestra en la tabla 5.1; las conformaciones tienen una distancia rms de aproximadamente 0.6 Å entre ellas; resultados que coinciden con los obtenidos por Troyen y Cohen [38]. La BPTI mutada visita 9 conformaciones asociadas a su estado nativo en el mismo intervalo de tiempo; cada conformación tiene una distancia rms de

aproximadamente 0.73 Å con respecto a las otras y están caracterizadas por la movilidad de los mismos lazos que la BPTI natural (tabla 5.1); la movilidad de estos lazos esta presente en la estructura natural pero se incrementa en la mutada.

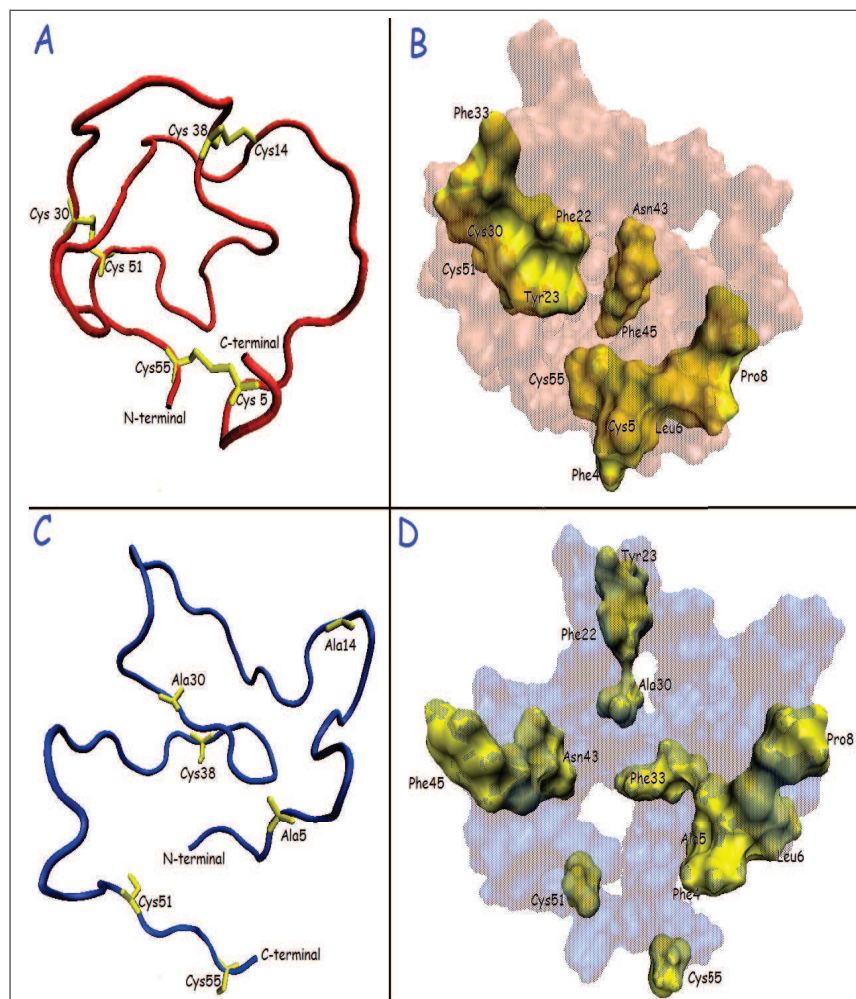


Figura 5.1: Estructuras de las primeras conformación de las muestras 1 de las proteína natural y mutada respectivamente. **A, C:** Estructuras tridimensionales. **B,D:** Disposición espacial de los residuos que forman el núcleo hidrofóbico.

Fueron encontradas además 88 conformaciones no nativas para la BPTI natural, la distancia rms mutua está entre 0.92 y 1.07 Å, y están caracterizadas por poseer lazos con mayor movilidad; para la mutante se encontraron 90 conformaciones con una distancia rms mutua entre 1.01 y 1.53 Å y también caracterizadas por la movilidad de ciertos lazos, como se muestra en la tabla 5.1. Las conformaciones no nativas, obtenidas tanto para la proteína natural como para la mutada, no poseen estructura secundaria, simplemente forman lazos (figura 5.1) y la movilidad de estos es característica en cada muestra (tabla 5.1). En la estructura natural las cisteínas que forman los enlaces disulfuro tienden a tener muy poca movilidad o su movimiento esta restringido al de la cisteína con la que forman el enlace, lo que

no ocurre en la estructura mutada, donde las cisteínas (sin enlace disulfuro) y las alaninas (mutaciones) se encuentran en las regiones móviles de la proteína .

El mayor número de mínimos presentes en la superficie del Entorno Energético indica que la proteína mutada es capaz de visitar más conformaciones que la natural, resultados que coinciden con los obtenidos del análisis de grupos; además las barreras energéticas son menores por lo que transiciones entre conformaciones son más factibles. La figura 5.2 muestra una visión general de los Entornos Energéticos de la BPTI natural y mutada; en la natural, la presencia de un mínimo muy pronunciado y altas barreras energéticas podría asociarse a un plegamiento lento hasta alcanzar un estructura altamente estable (la barrera energética a superarse para salir del estado nativo esta en el orden de los 1159.42 kT); en la proteína mutada, en cambio la presencia de un mínimo global ancho rugoso y barreras pequeñas fáciles de superar podrían asociarse con un plegado rápido pero hacia una estructura no muy estable (en las cercanías del estado nativo la proteína puede adquirir vías conformaciones fácilmente), como se puede observar en la figura 5.3. Los valores de la altura de la barreras energéticas (transiciones entre mínimos) están en el rango de la energía de fotones de luz ultravioleta, por lo que la irradiación con fotones de este rango podrían provocar cambios conformacionales en la proteína que causen un mal funcionamiento biológico.

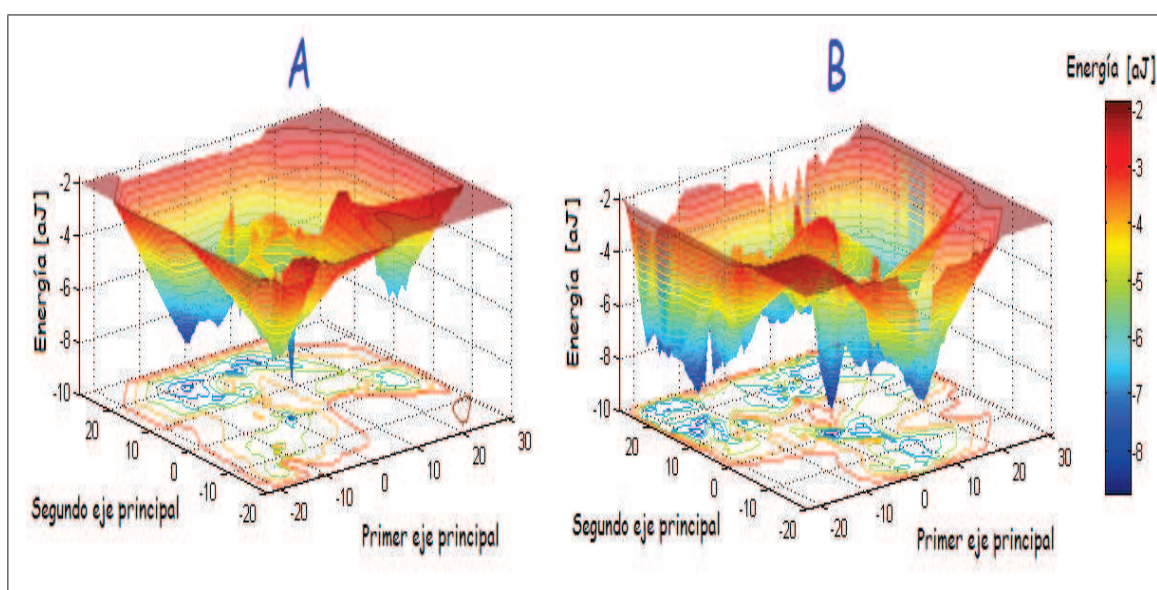


Figura 5.2: **A:** Representación tridimensional del Entorno Energético de la BPTI natural. **B:** Representación tridimensional del Entorno Energético de la BPTI mutada.

Las dos proteínas, BPTI natural y mutada, están caracterizadas por Entornos Energéticos sustancialmente diferentes (figuras 5.2 y 5.3). Mientras el de la natural presenta un mínimo global pronunciado muy estrecho (-8.8 aJ (-2125.60 kT)) y dos mínimos locales bastante profundos separados por barreras energéticas del orden los 4.8 aJ (1159.42 kT), las barreras entre mínimos locales están en el orden de los 1.6 aJ (386.47 kT); el Entorno de la mutante presenta un mínimo global más ancho y ligeramente menor que la natural (-9.12 aJ (-2202.90 kT)) y cinco mínimos locales que se encuentran separados del global por barreras entre 2.4 y 4.48 aJ (579.71 y 1082.13 kT), y entre ellos por barreras más pequeñas del orden de los 1.28 aJ (309.18 kT), como se muestra en la tabla 5.2.

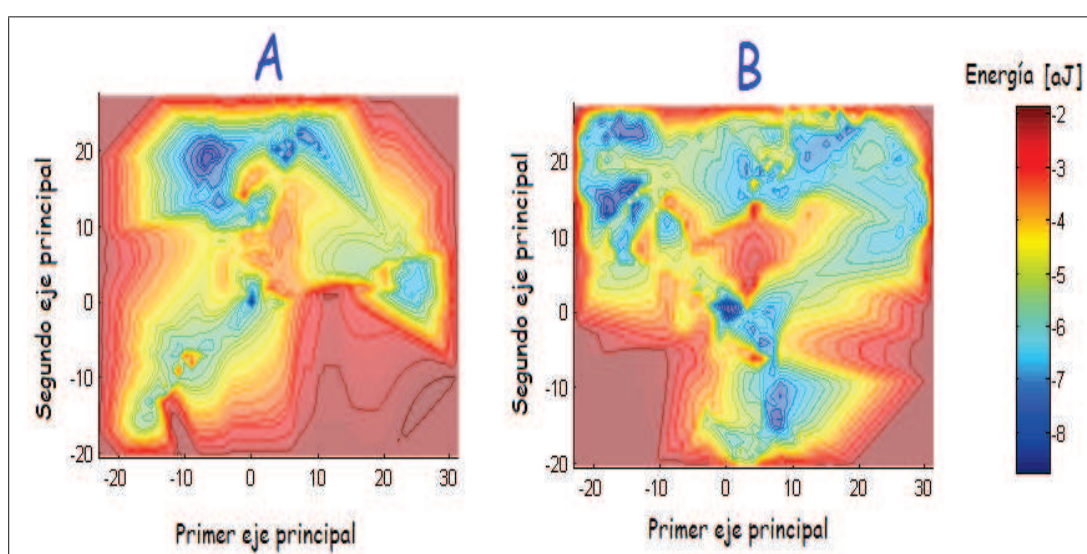


Figura 5.3: **A:** Proyección del Entorno Energético de la BPTI natural. **B:** Proyección del Entorno Energético de la BPTI mutada.

Característica	Entorno Energetico	
	Natural	Mutada
Mínimo Global [kT]	2125.60	2202.90
Número de mínimos locales	2	5
Altura barreras Energéticas entre los mínimos locales y el global	Entre 773.12 y 1159.92 [kT]	Entre 579.71 y 1082.13 [kT]
Altura barreras energéticas entre mínimos globales	386.47 [kT]	309.18 [kT]

Tabla 5.2: Características de los Entornos Energéticos de la BPTI natural y mutada

La ausencia de los enlaces disulfuro en la estructura de de la BPTI altera la disposición espacial de los aminoácidos que forman la región activa de esta proteína. La figura 5.4 muestra las estructuras que representan las conformaciones nativas con menor energía, tanto de la proteína natural como de la mutada; se observa que el lazo que forma la zona activa se encuentra deformado en la conformación de la proteína mutada, lo que provocaría que esta estructura no esté en la capacidad de realizar su función biológica; este comportamiento es similar para todas las conformaciones nativas de la proteína mutada, contrario a lo que sucede en la natural, donde la zona activa conserva la forma en la que puede realizar su función biológica. Haciéndose de este modo notoria la importancia de los enlaces disulfuro en la función biológica de la BPTI. La ausencia de estos enlaces además provoca una disminución en las barreras energéticas del Entorno de la BPTI (figura 5.3) facilitando de este modo la transición a estados desnaturalizado, haciendo menos estable a la proteína.

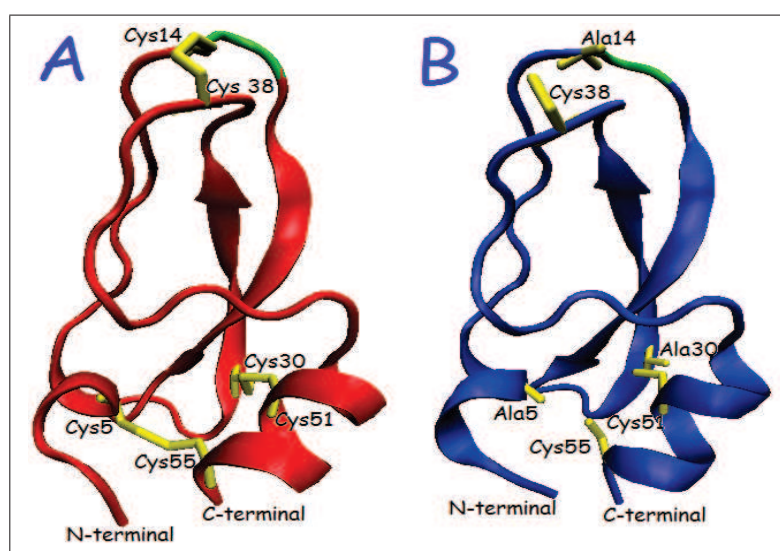


Figura 5.4: A: Estructura de las conformaciones naturales con menor energía de la proteína natural y mutada.

Las diferencias entre los Entornos Energéticos de proteínas reflejan diferencias en los mecanismos de plegamiento de estas [59], por lo tanto una descripción completa y detallada de este solucionaría una serie de problemas relacionados con el plegamiento proteínico. Mediante técnicas de análisis multivariante como el Análisis de Coordenada Principal se pudo obtener una representación tridimensional de los Entornos Energéticos de proteínas a 300K, brindando de este modo una herramienta útil para su estudio. La utilización de simulaciones de Dinámica Molecular dentro del concepto del Entorno Energético brindan la posibilidad de

estudiar de una manera sistemática la dinámica de una proteína bajo condiciones determinadas. En particular la BPTI es un buen modelo para ilustrar este concepto ya que provee facilidad en el tratamiento computacional (simulaciones y análisis accesibles en el rango de los nanosegundos) para obtener una gran cantidad de información y sus resultados pueden ser contrastados con otros estudios tanto teóricos como experimentales.

ANEXO

Anexo A: Preparación del sistema BPTI natural

Sript para preparar el sistema: monómero natural dentro de una esfera de agua

Crear PDB 6pti sola, sin moléculas de agua

```
set 6pti [atomselect top protein]
$6pti writepdb 6pti_sola.pdb
```

Crear PDB y PSF de la 6pti con Hidrogenos

```
package require psfgen
topology top_all22_prot.inp
pdbalias residue HOH TIP3
pdbalias atom ILE CD1 CD
segment U {pdb 6pti_sola.pdb}
coordpdb 6pti_sola.pdb U
guesscoord
writepdb 6pti.pdb
writepsf 6pti.psf
```

Colocar la BPTI natural (6pti) en una esfera de agua de radio suficiente como para cubrirla totalmente

```
set molname 6pti
mol new ${molname}.psf
mol addfile ${molname}.pdb
```

Determina el centro de masa de la molécula y guardas las coordenadas

```
set cen [measure center [atomselect top all] weight mass]
  set x1 [lindex $cen 0]
  set y1 [lindex $cen 1]
  set z1 [lindex $cen 2]
  set max 0
```

Determina la distancia desde el centro de masa al átomo más lejano de la proteína

```
foreach atom [[atomselect top all] get index] {
set pos [lindex [[atomselect top "index $atom" get {x y z}] 0]
  set x2 [lindex $pos 0]
  set y2 [lindex $pos 1]
  set z2 [lindex $pos 2]
set dist [expr pow(($x2-$x1)*($x2-$x1) + ($y2-$y1)*($y2-$y1) +
```

```

($z2-$z1)*($z2-$z1),0.5)]
  if {$dist > $max} {set max $dist}
}
mol delete top

```

Surme gir la molécula en una caja de agua con un forrado suficiente (15 A).

```

package require solvate
solvate ${molname}.psf ${molname}.pdb -t 15 -o del_water

```

```

resetpsf
package require psfgen
mol new del_water.psf
mol addfile del_water.pdb
readpsf del_water.psf
coordpdb del_water.pdb

```

Determinar que moléculas de agua no son necesarias y deben borrarse

```

set wat [atomselect top "same residue as {water and ((x-$x1)*(x-$x1) +
(y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set del [atomselect top "water and not same residue as {water and
((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
  set seg [$del get segid]
  set res [$del get resid]
  set name [$del get name]
  for {set i 0} {$i < [llength $seg]} {incr i} {
delatom [lindex $seg $i] [lindex $res $i] [lindex $name $i]
  }
  writepsf ${molname}_ws.psf
  writepdb ${molname}_ws.pdb
  mol delete top
  mol new ${molname}_ws.psf
  mol addfile ${molname}_ws.pdb
  mol delete top

```

Calcular la carga neta del sistema

```

set sel [atomselect top all]
  set netCharge0 [eval "vecadd [$sel get charge]"]
  $sel delete

```

Neutralizar el sistema: añadimos 10 iones Cl y 4 iones Na. Y calcular la carga neta del Nuevo sistema

```

resetpsf
package require psfgen

```



```

package require autoionize
autoionize -psf 6pti_ws.psf -pdb 6pti_ws.pdb -nna 4 -ncl 10
set sel [atomselect top all]
set netCharge [eval "vecadd [$sel get charge]"]
$sel delete

```

Escribe archivo con centro, radio y carga final del sistema

```

set outfile [open datos.txt w];
puts $outfile "La carga neta del sistema es: ${netCharge0}e"
puts $outfile "El centrode masa de la esfera es: [measure
center [atomselect top all] weight mass]"
puts $outfile "El radio de la esfera es: $max"
puts $outfile "La carga neta del sistema luego
de añadir los iones es: ${netCharge}e"
close $outfile

```

Anexo B: Preparación del sistema BPTI mutada

Sript para preparar el sistema: monómero mutada dentro de una esfera de agua (III)

Crear PDB 6pti sola, sin moléculas de agua

```

set 6pti [atomselect top protein]
$6pti writepdb 6pti_sola.pdb

```

Crear PDB y PSF de la 6pti con Hidrogenos

```

package require psfgen
topology top_all22_prot.inp
pdbalias residue HOH TIP3
pdbalias atom ILE CD1 CD
segment U {pdb 6pti\_sola.pdb}
coordpdb 6pti\_sola.pdb U
guesscoord
writepdb 6pti.pdb
writepsf 6pti.psf

```

Realizar la mutación: Cambiar residuos Cys5, cys14 y Cys 30 por Ala

```

resetpsf
package require psfgen 1.0
package provide mutant 1.2
topology top_all22_prot.inp
pdbalias residue HOH TIP3

```

```

pdbalias residue HIS HSD
pdbalias atom ILE CD1 CD
pdbalias atom GLY OXT OT1
segment X {pdb 6pti.pdb
  mutate 5 ALA
  mutate 14 ALA
  mutate 30 ALA}
coordpdb 6pti.pdb X
guesscoord
writepdb 6pti_mut.pdb
writepsf 6pti_mut.psf

```

Colocar la BPTI mutada (6pti) en una esfera de agua de radio suficiente como para cubrirla totalmente

```

set molname 6pti_mut
mol new ${molname}.psf
mol addfile ${molname}.pdb

```

Determina el centro de masa de la molécula y guardas las coordenadas

```

set cen [measure center [atomselect top all] weight mass]
set x1 [lindex $cen 0]
set y1 [lindex $cen 1]
set z1 [lindex $cen 2]
set max 0

```

Determina la distancia desde el centro de masa al átomo más lejano de la proteína

```

foreach atom [[atomselect top all] get index] {
set pos [lindex [[atomselect top "index $atom"] get {x y z}] 0]
set x2 [lindex $pos 0]
set y2 [lindex $pos 1]
set z2 [lindex $pos 2]
set dist [expr pow(($x2-$x1)*($x2-$x1) + ($y2-$y1)*($y2-$y1) +
($z2-$z1)*($z2-$z1),0.5)]
if {$dist > $max} {set max $dist}
}
mol delete top

```

Surmegir la molécula en una caja de agua con un forrado suficiente (15 A).

```

package require solvate
solvate ${molname}.psf ${molname}.pdb -t 15 -o del_water
resetpsf
package require psfgen

```

```

mol new del_water.psf
mol addfile del_water.pdb
readpsf del_water.psf
coordpdb del_water.pdb

```

Determinar que moléculas de agua no son necesarias y deben borrarse

```

set wat [atomselect top "same residue as {water and ((x-$x1)*(x-$x1) +
(y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set del [atomselect top "water and not same residue as {water and
((x-$x1)*(x-$x1) + (y-$y1)*(y-$y1) + (z-$z1)*(z-$z1))<($max*$max)}"]
set seg [$del get segid]
set res [$del get resid]
set name [$del get name]
for {set i 0} {$i < [llength $seg]} {incr i} {
delatom [lindex $seg $i] [lindex $res $i] [lindex $name $i]
}
writepsf ${molname}_ws.psf
writepdb ${molname}_ws.pdb
mol delete top
mol new ${molname}_ws.psf
mol addfile ${molname}_ws.pdb
mol delete top

```

Calcular la carga neta del sistema

```

set sel [atomselect top all]
set netCharge0 [eval "vecadd [$sel get charge]"]
$sel delete

```

Neutralizar el sistema: añadimos 10 iones Cl y 4 iones Na. Y calcular la carga neta del Nuevo sistema

```

resetpsf
package require psfgen
package require autoionize
autoionize -psf 6pti_mut_ws.psf -pdb 6pti_mut_ws.pdb -nna 4 -ncl 10
set sel [atomselect top all]
set netCharge [eval "vecadd [$sel get charge]"]
$sel delete

```

Escribe archivo con centro, radio y carga final del sistema

```

set outfile [open datos.txt w];
puts $outfile "La carga neta del sistema es: ${netCharge0}e"
puts $outfile "El centrode masa de la esfera es: [measure
center [atomselect top all] weight mass]"

```

```

    puts $outfile "El radio de la esfera es: $max"
puts $outfile "La carga neta del sistema luego
de añadir los iones es: ${netCharge}e"
    close $outfile

```

Anexo C: Protocolo Minimización

#Datos de entrada

```

structure          6pti.psf
coordinates        6pti.pdb
set outputname     6pti_min
firsttimestep      0
temperatura        0

```

Parámetros del Campo de Fuerza

```

paraTypeCharmm     on
parameters         par_all122_prot.inp
exclude            scaled1-4
1-4scaling         1.0
cutoff             12.
switching          on
switchdist         8.
pairlistdist       13.5

```

Parámetros de integración

```

timestep           1.0
stepspercycle      20

```

Datos de salida

```

outputName         $outputname
dcdfreq            1000
outputEnergies     100

```

Minimización

```

minimize           10000

```

Anexo D: Protocolo de calentamiento de 0K a 300K

#Datos de entrada

```

structure      ../Estructuras/6pti.psf
coordinates    ../Estructuras/6pti.pdb
bincoordinates ../Minimizacion/6pti_min.coor
set outputname 6pti_cal
firsttimestep  0

```

Parámetros del Campo de Fuerza

```

paraTypeCharmm  on
parameters      ../Estructuras/par_all22_prot.inp
exclude         scaled1-4
1-4scaling      1.0
cutoff          12.
switching       on
switchdist      8.
pairlistdist    13.5

```

Parámetros de integración

```

timestep        1.0
rigidBonds      all
nonbondedFreq   1
fullElectFrequency 2
stepspercycle   20

```

Datos de salida

```

outputName      $outputname
dcdfreq         1000
outputEnergies  100

```

Condiciones de frontera

```

extendedSystem  ../Minimizacion/6pti_min.xsc
sphericalBC     on
sphericalBCcenter 7.77798271179, 5.73942756653, 3.54496049881
sphericalBCr1   22.2698405881
sphericalBCk1   10
sphericalBCexp1 2

```

Restricciones

```

fixedAtoms      on
fixedAtomsCol   B
fixedAtomsForces on

```

Protocolo de calentamiento

```

temperature          0
reassignFreq        25
reassignIncr         0.375
reassignHold         300

```

Calentamiento

```
run      20000
```

Anexo E: Protocolo de Equilibración a 300K

#Datos de entrada

```

structure            ../Estructuras/6pti.psf
coordinates          ../Estructuras/6pti.pdb
bincoordinates       ../Calentamiento/6pti_cal.coor
binvelocities        ../Calentamiento/6pti_cal.vel
set outputname       6pti_eq1
firsttimestep        0

```

Parámetros del Campo de Fuerza

```

paraTypeCharmm on
parameters           ../Estructuras/par_all22_prot.inp
exclude              scaled1-4
1-4scaling           1.0
cutoff                12.
switching             on
switchdist           8.
pairlistdist         13.5

```

Parámetros de integración

```

timestep             1.0
rigidBonds           all
nonbondedFreq        1
fullElectFrequency   2
stepspercycle        20

```

Datos de salida

```

outputName           $outputname
dcdfreq              1000
outputEnergies       100

```

Condiciones de frontera

```

extendedSystem      ../Calentamiento/6pti_min.xsc
sphericalBC         on
sphericalBCcenter   7.77798271179, 5.73942756653, 3.54496049881
sphericalBCr1       22.2698405881
sphericalBCk1       10
sphericalBCexp1     2

```

Restricciones

```

fixedAtoms          on
fixedAtomsCol       B
fixedAtomsForces    on

```

Control de Temperatura constante

```

langevin            on
langevinDamping     5
langevinTemp        300
langevinHydrogen    off

```

Equilibración

```

run                 100000

```

Anexo F: Protocolo de calentamiento de 300K a 1000K

Datos de entrada

```

structure           ../Estructuras/6pti.psf
coordinates         ../Estructuras/6pti.pdb
bincoordinates      ../Equilibración300/6pti_min.coor
set outputname      6pti_cal
firsttimestep       0

```

Parámetros del Campo de Fuerza

```

paraTypeCharmm     on
parameters          ../Estructuras/par_all22_prot.inp
exclude            scaled1-4
1-4scaling         1.0
cutoff              12.
switching          on
switchdist         8.
pairlistdist       13.5

```

Parámetros de integración

```
timestep          1.0
rigidBonds        all
nonbondedFreq    1
fullElectFrequency 2
stepspercycle    20
```

Datos de salida

```
outputName        $outputname
dcdfreq           1000
outputEnergies    100
```

Condiciones de frontera

```
extendedSystem    ../Minimizacion/6pti_min.xsc
sphericalBC       on
sphericalBCcenter 7.77798271179, 5.73942756653, 3.54496049881
sphericalBCr1     22.2698405881
sphericalBCk1     10
sphericalBCexp1   2
```

Restricciones

```
fixedAtoms        on
fixedAtomsCol     B
fixedAtomsForces  on
```

Protocolo de calentamiento

```
temperature        300
reassignFreq       25
reassignIncr       0.35
reassignHold       1000
```

Calentamiento

```
\run              50000
```

Anexo G: Protocolo de enfriamiento de 1000K a 300K

#Datos de entrada

```
structure          ../Estructuras/6pti.psf
coordinates        ../Altas temperaturas/6pti_1ns.pdb
set outputname     6pti_1ns_enf
firsttimestep      0
```

Parámetros del Campo de Fuerza


```
paraTypeCharmm on
parameters      ../Estructuras/par_all22_prot.inp
exclude         scaled1-4
1-4scaling      1.0
cutoff          12.
switching       on
switchdist      8.
pairlistdist    13.5
```

Parámetros de integración

```
timestep        1.0
rigidBonds      all
nonbondedFreq   1
fullElectFrequency 2
stepspercycle   20
```

Datos de salida

```
outputName      $outputname
dcdfreq         1000
outputEnergies  100
```

Condiciones de frontera

```
sphericalBC     on
sphericalBCcenter 7.77798271179, 5.73942756653, 3.54496049881
sphericalBCr1    22.2698405881
sphericalBCk1    10
sphericalBCexp1  2
```

Restricciones

```
fixedAtoms      on
fixedAtomsCol    B
fixedAtomsForces on
```

Protocolo de enfriamiento

```
temperature      1000
reassignFreq     25
reassignIncr     -0.35
reassignHold     300
```

Enfriamiento

```
run              50000
```

Anexo H: Protocolo para las simulaciones de Dinámica Molecular Dirigida

#Datos de entrada

```
structure      ../Estructuras/6pti.psf
coordinates    ../muestra1/6pti_100_min.pdb
set outputname 6pti_TMD1
```

Parámetros del Campo de Fuerza

```
paraTypeCharmm on
parameters     ../Estructuras/par_all22_prot.inp
exclude        scaled1-4
1-4scaling     1.0
cutoff         12.
switching      on
switchdist     8.
pairlistdist   13.5
```

Parámetros de integración

```
timestep       1.0
rigidBonds     all
nonbondedFreq  1
fullElectFrequency 2
stepspercycle  20
```

#Control de Temperatura constante

```
langevin       on
langevinDamping 5
langevinTemp    300
langevinHydrogen off
```

Datos de salida

```
outputName     $outputname
dcdfreq        1000
outputEnergies 100
```

#Protocolo de simulación dirigida

```
TMD on
TMDk 200
TMDFirstStep 0
TMDLastStep 100000
TMDfile ../Muestreo_300/6pti_min_300K.pdb
TMDFinalRMSD 0
```

Ejecución

```
firsttimestep      0  
numsteps 100000
```

Anexo I: Interpolación cúbica basada en triangulaciones

X: Vector que contiene la primera coordenada principal de las estructuras
Y: Vector que contiene la segunda coordenada principal de las estructuras
Z: Vector que contiene el valor de la energía de las estructuras

```
C=floor(min(X):0.1:max(X))  
D=floor(min(Y):0.1:max(Y))  
[E,F]=meshgrid(C,D)  
G=griddata(X,Y,Z,E,F,'cubic')  
surfc (E,F,G, 'FaceColor', 'interp',... 'Edgecolor','none')
```

Referencias

- [1] Dobson, C. 2001. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond.* **356**: 133-145.
- [2] Olivares, L., y García, L. 2004. Plegamiento de las proteínas: Un problema interdisciplinario, *Rev. Soc. Quim. Méx.* **48**: 95-105.
- [3] Pavia, S. 2000. Estudios sobre el plegamiento del Inhibidor Proteico de Carboxipeptidasa de Patata (PCI). Instituto de Biología Fundamental. Universidad Autónoma de Barcelona. Tesis para obtener el Doctorado en Ciencias Químicas.
- [4] Creighton, T. 1984. Pathways and mechanisms of protein folding. *Adv. Biophys.* **18**: 1-20.
- [5] Onuchic, N., Luthey-Schulten, Z., y Wolynes P. 1997. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**: 545-600.
- [6] Salamanca, S. 2003. Plegamiento y funcionalidad biológica del inhibidor de metalocarboxipeptidasa LCI (Leech Carboxypeptidase Inhibitor). Instituto de Biotecnología y Biomedicina. Universidad Autónoma de Barcelona. Tesis para obtener Doctorado en Ciencias.
- [7] Pain, H. 2000. Mechanisms of protein Folding. Oxford University Press. (Oxford).
- [8] Tinoco, J., Saner, K., Wang, J., y Poglisi, J. 2002. Physical Chemistry: Principles and Applications in Biological Sciences. Prentice Hall (New Jersey).
- [9] Rose, G., Fleming, P., Banavar, B. y Martin, A. 2006. A backbone-based theory of protein folding. *PNAS.* **45**: 16623-16633.
- [10] Levit, M., y Warshel, A. 1975. Computer simulation of protein folding. *Nature.* **253**: 694-698.

- [11] Anfinsen, C. 1973. Principles that govern the folding of protein chains. *Science*. **181**: 223-230.
- [12] Morra, G. 2005. Role of electrostatics explored with molecular dynamics simulations for protein stability and folding. Universidad de Berlin. Tesis para obtener diplomado en Física.
- [13] Karplus, M. 1997. The Levinthal paradox: yesterday and today. *Folding & Design*. **2**: S69-S75.
- [14] Zwanzig, R., Szabo, A. y Bagchi, B. 1992. Levinthal's paradox. *PNAS*. **89**: 20-22
- [15] Levinthal, C. 1968. Are there pathways for protein folding?. *J. Chim. Physique*. **65**: 44-45.
- [16] Cárdenas, M. y Vázquez, E. 2003. ¿Es el glóbulo fundido un intermediario general en la vía de plegamiento de las proteína globulares?. *Rev. Soc. Quim. Méx.* **47**: 320-327.
- [17] Onuchic, N., Luthey-Schulten, Z., y Wolynes P. 1997. Theory of Protein Folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **48**: 545-600.
- [18] Ma, B., Kumar, S., Tsai, C. y Nussinov, R. 1999. Folding funnels and binding mechanism. *Protein Engineering*. **9**: 713-720.
- [19] Bryngelson, J., Onuchi, J. y Wolynes P. 1995. Funnels, Pathways and the Energy Landscape of Protein Holding: A Synthesis. *Proteins*. **21**:167-195
- [20] Bogatyreva, N. y Finkelstein, A. 2001. Cunning simplicity of protein holding landscapes. *Protein Engineering*. **8**: 521 - 523
- [21] Lois, G., Blawdziewicz, J., y O'Hern, C. 2008. Reliable protein holding on complex energy landscapes: The free energy reaction path. *Biophysical Journal*. **95**: 2692 - 2701
- [22] Anfinsen, C., Haber, E., Sela, M. y White, Jr. 1961. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *PNAS*. **47**: 1309 - 1314.
- [23] Wedemeyer, WJ., Welker, E., Narayan, M. y Scheraga, HA.2000. Disulfide bonds and protein folding. *Biochemistry*. **39**: 4207-4216.

- [24] Dittrich, M. y Kanchanawarin, C. Case Study: BPTI. Theoretical and Computational Biophysics Group. Descargado de: <http://www.ks.uiuc.edu/Training/CaseStudies/>. Septiembre 2008
- [25] Ascenzi, P., Bocedi, A., Bolognesi, M., Spallarossa, A., Coletta, M., De Cristofaro, R., Menegatti, E. 2003. The bovine basic pancreatic trypsin inhibitor (Kunitz inhibitor): a milestone protein., *Curr. Protein Peptide Sci.* **4**: 231-251.
- [26] Landais, R., Li, R. y Woodward, C. 2001. The antithrombotic and antiinflammatory mechanisms of action of aprotinin. *Ann. Thorac. Surg.* **72**: 2169-2175
- [27] Karplus, M. y McCammon, A. 2002. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**: 646-652
- [28] Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C. 1987. Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J.Mol.Biol.* **198**: 469-480
- [29] Kress, L. y Laskowsky, M. 1967. The Basic Trypsin Inhibitor of Bovine Pancreas. *J. Biol Chem.* **242**: 4925-4929
- [30] Rader, A & otros. 2002. Protein unfolding: Rigidity lost. *PNAS.* **99**: 3540-3545
- [31] Chelvanayagam, G., y Argos, P. 1993. Prediction of protein folding pathways: Bovine Pancreatic Trypsin Inhibitor. *Cytotechnology.* **11**: 67-71.
- [32] Creighton, T. 1990. Protein Folding. *Biochem. J.* **270**: 1-16.
- [33] Weissman, J. y Kim, P. 1991. Reexamination of the folding of BPTI: predominance of native intermediates. *Science.* **253**: 1386-1393.
- [34] Staley, J. y Kim, P. 1992. Complete folding of bovine pancreatic trypsin inhibitor with only a single disulfide bond. *PNAS.* **89**: 1519-1523.
- [35] Zhang, J. y Goldenberg D. 1997. Mutational analysis of BPTI folding pathway: I. Effects of aromatic-leucine substitutions on the distribution of folding intermediates. *Protein Science.* **6**: 1549-1562.
- [36] Zhang, J. y Goldenberg D. 1997. Mutational analysis of BPTI folding pathway: II. Effects of aromatic-leucine substitutions on the distribution of folding intermediates. *Protein Science.* **6**: 1563-1576.

- [37] Wlodawer A, Thirumalai, D. 1995. Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin inhibitor. *PNAS*. **92**: 1277-1281.
- [38] Troyer, J. y Cohen, F. 1995. Protein Conformational Landscapes: Energy minimization and clustering of a long molecular trajectory. *Proteins: Structure, Function and Genetics*. **23**: 97-110.
- [39] Caves, L., Evanseck, J y Karplus, M. 1998. Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science*. **7**: 649-666.
- [40] Nienhaus, U. 1999. Conformational motions in the rugged energy landscape of proteins. *AIP Conference Proceedings*. **487**: 99-116.
- [41] Ercolessi, F. 1997. A molecular dynamics primer. International School for Advanced Studies (SISSA-ISAS), Trieste, Italia. 4-10.
- [42] Bhadarkar, M. ., Brunner, R., Chipot, A., Dalke, A., Dixit, S., Grayson, P., Gullingsrud, J., Gursoy, A., Hardy, D., Hénin, J, Humphrey, W., Hurwitz, D., Krawetz, N., Kumar, S., Nelson, M., Phillips,A., Shinozaki, A., Zheng, G. y Zhu, F. 2008. NAMD User's Guide. Version 2.6. Theoretical Biophysics Group, University of Illinois and Beckman Institute.
- [43] Frenkel, D. y Smit, B. 2002. Understanding Molecular Simulation. Londres. Academic Press. Computational Science Series.
- [44] Phillips, J., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R., Kale, L. y Schulten K. 2005. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. **26**: 1781-1802
- [45] Zhigilel, L. 2000. Modelling in Materials Science: Introduction to interatomic potentials. University of Virginia, MSE 524. 1-23.
- [46] Karplus, M. Y Kuriyan, J. 2005. Molecular dynamics and protein function. *PNAS*. **19**: 6679-6685.
- [47] Humphrey, W., Dalke, A. y Schulten, K. 1996. "VMD - Visual Molecular Dynamics". *J. Molec. Graphics*. **14**: 33-38.
- [48] Schwieters, C., Kuszewski, J., Tjandra, N. y Clore, M. 2003. "The Xplor-NIH NMR Molecular Structure Determination Package". *J. Magn. Res.* **160**: 66-74.

- [49] Brooks, R., Bruccoleri, R., Olafson, B., States, D., Swaminathan, S. y Karplus, M. 1983. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**: 187-217.
- [50] Pearlman, D., Case, D., Caldwell, J., Ross, W., Cheatham, T., DeBolt, S., Ferguson, D., Seibel, G. y Kollman, P. 1995. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.* **91**: 1-41.
- [51] Phillips, J., Isgro, T., Sotomayor, M. y Villa, E. 2007. NAMD TUTORIAL. Windows Version. University of Illinois. NIH Resource for Macromolecular Modelling and Bioinformatics Beckman Institute.
- [52] Sagui, C. & Darden, T. (1999). Molecular Dynamics simulations of Biomolecules: Long-Range Electrostatics effects. *Annu. Rev. Biophys. Biomol. Struct.* **28**: 155-79.
- [53] Ellen, M. (2004). Introduction to molecular Dynamics Simulation. *NIC Series.* **23**:1-28.
- [54] Leach, A. Molecular Modelling. (2001). Gran Bretaña. Pearson Education Limited.
- [55] Oren, B., MacKerell, A., Roux, B. y Watanabe, M. 2001. Computational Biochemistry and Biophysics. USA. Marcel Dekker, Inc.
- [56] MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., y Karplus, M. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **102**: 3586-3616.
- [57] Rueda, M. 2005. Estudio teórico sobre la influencia del solvente en la estructura y dinámica del ADN. Universidad De Barcelona. Departamento De Bioquímica Y Biología Molecular. Tesis para obtener el Doctorado en Ciencias.
- [58] Bernstein, F., Koetzle, T., Williams, G., Meyer, E., Brice, M., Rodgers, J., Kennard, O., Shimanouchi, T. y Tasumi, M. 1977. The Protein Data Bank: A

- Computer-based Archival File For Macromolecular Structures. *J. of Mol. Biol.* **112** : 535.
- [59] Levy, Y. y Becker, O. 2001. Energy landscapes of conformational constrained peptides. *J. Chem. Phys.* **114**: 993-1009.
- [60] Dubey, A., Sharma, G., Mavroidis, C., Tomassone, M., Nickitzuk, K. y Yarmush, L. 2004. Computational Studies of Viral Protein Nano-Actuators. *Journal of Computational and Theoretical Nanoscience.* **1**: 18-28.
- [61] Pang-Ning, T., Steinbach, M. y Kumar, V. 2006. Introduction to Data Mining. Addison-Wesley / Prentice Hall.
- [62] Rencher, A. 2002. Methods of Multivariate Analysis. Wiley Series (Canadá).
- [63] Mojena, R.. 1977. Hierarchical grouping methods and stopping rules - an evaluation. *Computer J.* **20** : 359 -363
- [64] Becker, O. 1998. Principal Coordinate Maps of Molecular Potential Energy Surfaces. *Journal of Computational Chemistry.* **19**: 1255 - 1267.
- [65] Tous, J. y Ferrando, P. 1991. Aplicación de las EMD al análisis del diferencial semántico. *Psicothema.* **3**: 453-486.
- [66] Gower, J. 2004. The geometry of biplot scaling. *Biometrika.* **93**: 705 - 714.
- [67] Dagget, V. y Levitt, M. 1993. Protein Unfolding Pathways explored through molecular dynamics simulations. *J. Mol. Biol.* **232**: 600-619.
- [68] Perrier, X., Jacquemoud-Collet, J.P. 2006. DARwin software. <http://darwin.cirad.fr/darwin>
- [69] Elmaci, N. y Berry, E. 1999. Principal coordinate analysis on a protein model. *J. Chem. Phys.* **21 (1)** : 10607 - 10622.
- [70] Sandwell, D. 1987. Biharmonic Spline Interpolation of GEOS-3 and SEASAT Altimeter Data. *Geophysical Research Letters.* **2**: 139-142.