



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA QUÍMICA Y BIOTECNOLOGÍA

MEJORAMIENTO DE LA TERMOESTABILIDAD DE ENZIMAS MEDIANTE DINÁMICA MOLECULAR Y ANÁLISIS DE COMPONENTES PRINCIPALES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL CON
MENCION EN BIOTECNOLOGÍA.

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIAS DE LA
INGENIERÍA, MENCION QUÍMICA.

JAVIER ANDRÉS JOFRÉ ESCOBAR

PROFESOR GUÍA:

ORIANA SALAZAR AGUIRRE, Ph. D

PROFESOR CO-GUÍA:

J. CRISTIAN SALGADO HERRERA, Ph. D

MIEMBROS DE LA COMISIÓN:

MARÍA ELENA LIENQUEO CONTRERAS, Ph. D;

GERALD ZAPATA TORRES, Ph. D

SANTIAGO DE CHILE

DICIEMBRE, 2012

RESUMEN

La industria biotecnológica precisa del uso de enzimas optimizadas, en particular, de termoestabilidad elevada ya que procesos industriales como la elaboración de biocombustibles de segunda generación requieren temperaturas de operación por sobre las temperaturas óptimas para la mayoría de las enzimas. El diseño de nuevas enzimas modificando enzimas ya existentes puede ser un proceso extensivo en términos experimentales y analíticos y no asegura necesariamente un grado determinado de mejora. En el presente trabajo se desarrolló una nueva herramienta para el mejoramiento de la termoestabilidad de enzimas con énfasis en dinámica molecular.

Se diseñó un procedimiento para simular el comportamiento de la estructura de la proteína estudiada a tres diferentes temperaturas, 300, 350 y 400K. Se elaboró un nuevo índice de flexibilidad usando análisis de componentes principales, el valor *if*, que a partir de la simulación, permite determinar las regiones más flexibles y candidatas a ser rigidizadas para mejorar la termoestabilidad. Las estructuras de las variantes diseñadas a partir de lo anterior se simularon para evaluar su grado de mejora en términos de la flexibilidad y compactación. El procedimiento se validó mediante su aplicación a las estructuras de las enzimas Cel7A de *Talaromyces emersonii* y Cel7B de *Melanocarpus albomyces*. Se recuperaron regiones identificadas en otro estudio como flexibles y se encontraron nuevas regiones para ser rigidizadas. De la aplicación del procedimiento a la enzima Cel72 se obtuvieron 3 nuevas enzimas de las cuales dos mostraron reducir la compactación respecto de la nativa y una mostró reducir la flexibilidad de la estructura.

Se diseñó una estrategia para mejorar la termoestabilidad de enzimas, fue posible identificar regiones flexibles y se vieron cambios en la flexibilidad y compactación de variantes respecto de sus enzimas nativas. El procedimiento aquí descrito tiene el potencial de ser una herramienta rápida y de bajo costo, sin embargo, se requerirá de ensayos de termoestabilidad de las enzimas aquí propuestas para validar experimentalmente el procedimiento.

Palabras Claves: Dinámica molecular, Cel7, Flexibilidad, Termoestabilidad, Análisis de Componentes Principales.

ABSTRACT

The biotechnology industry requires the use of optimized enzymes, particularly, of high thermostability because industrial processes like elaboration of second generation biofuels requires operation temperatures above the optimal temperatures of most enzymes. The design of new enzymes modifying existing ones can be an extensive process in terms of experimental and analytical and it doesn't necessarily ensures a level of improvement. In this work a new thermostability improvement tool was developed, with an emphasis in molecular dynamics.

A procedure to simulate the behavior of protein structure at three different temperatures was designed. A new flexibility index was created using principal component analysis, the *if* value, from the simulation data and it allows identify flexible regions to be stiffened to improve the thermostability. Variants structures designed from the above were re-simulate to assess the degree of improvement in terms of the flexibility and compactness. The procedure was validated by applying it to the structures of the enzymes *Talaromyces emersonii* Cel7A and *Melanocarpus albomyces* Cel7B. Regions identified by another study as flexible were recovered and other new regions were found to be stiffened. From the application of the procedure to the enzyme Cel72 three new enzymes were obtained two of which showed a reduced compactness from the native structure and one of which showed to reduce the structure flexibility. Thermostability assays are required to validate experimentally the procedure.

A strategy to improve the enzyme thermostability were designed. Flexible regions were identified and changes in flexibility and compactness were seen between variants and native enzyme. The procedure here described has the potential to be a fast and cheap tool, however, thermostability assays are required to validate the procedure.

Keywords: Molecular Dynamic, Cel7, Flexibility, Thermostability, Principal Component Analysis.

Índice

1. INTRODUCCIÓN	1
1.1. Definición y descripción del problema	1
1.2. Motivación	2
2. ANTECEDENTES	4
2.1. Industria Biotecnológica	4
2.1.1. Etanol como biocombustible	5
2.2. Enzimas	9
2.2.1. General	9
2.2.2. Plegamiento de las enzimas	10
2.2.3. Flexibilidad de las proteínas	13
2.2.4. Flexibilidad y Termoestabilidad	15
2.2.5. Celulasas de hongos	17
2.2.6. Celulasas de hongos termófilos	18
2.3. Ingeniería de Proteínas	19
2.3.1. Métodos de Mejoramiento Enzimático	19
2.4. Determinación de la estructura tridimensional de una enzima	22
2.4.1. Determinación experimental	22
2.4.2. Predicción de estructuras	24
2.5. Dinámica de una Enzima: simulación <i>in silico</i>	27
2.5.1. Generalidades	27
2.5.2. Dinámica Molecular	28
2.5.3. Simulación	56

3. MODELOS DE ESTUDIO	70
3.1. Celobiohidrolasa Cel7A de <i>Talaromyces emersonii</i>	70
3.2. Celobiohidrolasa Cel7B de <i>Melanocarpus albomyces</i>	72
4. HIPÓTESIS DE TRABAJO	74
5. OBJETIVOS	75
5.1. Objetivo General	75
5.2. Objetivos Específicos	75
6. MATERIALES Y MÉTODOS	76
6.1. Materiales	76
6.1.1. Hardware	76
6.1.2. Software	76
6.2. Métodos	77
6.2.1. Cálculo de Trayectorias	77
6.2.2. Cálculo de valores f y if	77
6.2.3. Determinación de la Raíz Cuadrática Media de las Desviaciones	79
6.2.4. Compactación	80
7. RESULTADOS	81
7.1. Caracterización de la enzima Cel72	83
7.2. Procedimiento para la simulación y identificación de mutantes	85
7.2.1. Conversión de la estructura y generación de la topología	86
7.2.2. Minimización de energía de la estructura en el vacío	87
7.2.3. Condiciones de Borde Periódicas	88
7.2.4. Adición de Solvente e Iones	89
7.2.5. Minimización de energía del sistema solvatado	89

7.2.6. Relajación del solvente y átomos de hidrógeno y aumento de la temperatura	89
7.2.7. Control de la Temperatura	90
7.2.8. Control de la Presión	91
7.2.9. Producción y Equilibrio	91
7.3. Validación	92
7.3.1. Simulación de las enzimas nativas T y M	92
7.3.2. Cálculo de Flexibilidad	110
7.3.3. Simulación del efecto de la temperatura en la trayectoria de residuos de las enzimas variantes de T y M.	118
7.3.4. Resultados destacados de la etapa de validación	127
7.4. Aplicación de la metodología a Cel72	129
7.4.1. Modelamiento	129
7.4.2. Simulación de Cel72	135
7.4.3. Cálculo de la flexibilidad	146
7.4.4. Simulación de las variantes	150
8. DISCUSIÓN	155
9. CONCLUSIONES	164
10. TEMAS DE INVESTIGACIÓN FUTURA	167
BIBLIOGRAFÍA	169
A. ANEXOS	189
A.1. Cálculo de trayectorias con GROMACS	189
A.2. Análisis de componentes principales	193
A.3. Alineamiento múltiple de Cel72	196

A.4. Figuras resultado de la evaluación con PROCHECK	198
A.5. Registros de salida	208
A.6. Tabla t-student	208

Índice de figuras

2.1.	Movilidad y Deformabilidad	15
2.2.	Proceso para determinar la estructura mediante rayos X.	23
2.3.	Múltiples estructuras obtenidas por NMR.	24
2.4.	Representación gráfica del grado de separación entre átomos enlazados en una molécula	47
2.5.	Condiciones de borde periódicas en dos dimensiones.	57
2.6.	Dodecaedro rómbico y octaedro truncado	59
3.1.	Estructura tridimensional se Cel7A de <i>T. emersonii</i>	71
3.2.	Estructura tridimensional se Cel7B de <i>M. albomyces</i>	73
7.1.	Esquema general del procedimiento de mejoramiento de termoestabilidad	82
7.2.	Esquema de los dominios conservados encontrados para Cel72	84
7.3.	Estructura tridimensional de Cel72	85
7.4.	Proteína en una caja cúbica en dónde el solvente es agua	88
7.5.	Temperatura y presión del sistema en función del tiempo de relajación	93
7.6.	Energía del sistema durante la etapa de relajación	94
7.7.	Temperatura y presión del sistema en la etapa NVT	95
7.8.	Energía medida en la etapa NVT	96
7.9.	Temperatura y presión del sistema en la etapa NPT	98
7.10.	Energía medida en la etapa NPT	99
7.11.	Temperatura y presión del sistema en la etapa de producción	101

7.12. Energía del sistema en la etapa de producción	102
7.13. $C\alpha$ -RMSD y distribución de $C\alpha$ -RMSD de la enzima T y la enzima M en la etapa de producción	104
7.14. Rg de la cadena principal para la enzima T y la enzima M en la etapa de producción.	105
7.15. RMSf de cada residuo para las enzimas T y la enzima M.	107
7.16. Visión general de la trayectoria de T	108
7.17. Visión general de la trayectoria de la enzima M	109
7.18. Flexibilidad local de las enzimas T y M por residuo	110
7.19. Regiones definidas como flexibles en base al análisis de valor <i>if</i> para la enzima T.	113
7.20. Regiones definidas como flexibles en base al análisis de valor <i>if</i> para la enzima M.	114
7.21. Mutantes de T.	116
7.22. Mutantes de la enzima M.	117
7.23. $C\alpha$ -RMSD de las enzimas mutantes de T la etapa de producción. . .	119
7.24. $C\alpha$ -RMSD de las enzimas mutantes de la enzima M la etapa de pro- ducción.	119
7.25. RMSD promedio para cada residuo de las mutantes de T	120
7.26. RMSD promedio para cada residuo de las mutantes de la enzima M .	120
7.27. Rg calculado para las mutantes de T en función del tiempo	122
7.28. Rg calculado para las mutantes de la enzima M en función del tiempo . . .	122
7.29. Valor <i>f</i> de la enzima T y sus variantes.	125
7.30. Valor <i>f</i> de la enzima M y sus variantes.	126
7.31. Modelo de Cel72	130

7.32. Modelos de las enzimas Cel72, T y M superpuestos	131
7.33. QMEAN score del modelo	133
7.34. Estimación de la calidad local del modelo	134
7.35. Temperatura y presión del sistema en función del tiempo de relajación	137
7.36. Temperatura, presión y energía del sistema en la etapa NVT	137
7.37. Temperatura, presión y energía del sistema en la etapa NPT	139
7.38. Temperatura, presión y energía del sistema en la etapa de producción	141
7.39. $C\alpha$ -RMSD en la etapa de producción	143
7.40. R_g de la enzima en función del tiempo	144
7.41. Visión general de la trayectoria de Cel72	145
7.42. Flexibilidad local de C por residuo	146
7.43. Regiones definidas como flexibles en base al análisis de valor if para Cel72. En color azul se destacaron los residuos pertenecientes a la región A, en verde la B, rojo para C, amarillo para D, celeste para E, rosado para F y naranja para G.	148
7.44. Mutantes de Cel72	149
7.45. $C\alpha$ -RMSD de las enzimas mutantes de Cel72 en la etapa de producción.	151
7.46. RMSD promedio para cada residuo de las variantes de Cel72	152
7.47. R_g calculado para las mutantes de Cel72 en función del tiempo	153
7.48. Valor f de la enzima Cel72 y sus variantes.	154
A.1. Gráfico de Ramachandran de Cel72.	198
A.2. Gráfico de Ramachandran para todos los tipos de residuos	199
A.3. Gráfico Chi1-Chi2 para todos los tipos de residuos	200
A.4. Propiedades por residuo de Cel72	201

A.5. Propiedades por residuo de Cel72	202
A.6. Propiedades por residuo de Cel72	203
A.7. Propiedades por residuo de Cel72	204
A.8. Largo de enlace de la cadena principal de Cel72	205
A.9. Ángulos de enlace de la cadena principal de Cel72	206
A.10.RMS de la distancia respecto a la planaridad	207

Índice de tablas

2.1. Campos de fuerza.	48
2.2. Caja cúbica, dodecaedro rómbico y el octaedro truncado.	60
2.3. Tipos de átomos en GROMACS.	61
7.1. Resultado del alineamiento de la secuencia de Cel72 utilizando BLAST	83
7.2. Temperatura, presión y energía del sistema para las enzimas T y M en la etapa NVT	97
7.3. Temperatura, presión y energía del sistema de T y M en la etapa NPT	100
7.4. Propiedades del sistema medidas en la etapa de producción	103
7.5. Medias y desviaciones del Rg.	106
7.6. Resultados del test t	106
7.7. Ranking de valores <i>if</i>	111
7.8. Media y desviación estándar del Rg a cada temperatura para las mutantes de la enzima T	121
7.9. Media y desviación estándar del Rg a cada temperatura para las mutantes de la enzima M	123
7.10. Análisis t-student entre las enzimas mutantes de T y la enzima nativa. . .	123
7.11. Análisis t-student entre las enzimas mutantes de la enzima M y la enzima nativa.	124
7.12. Pseudo energías para el cálculo de QMEAN.	132
7.13. Temperatura, presión y energía del sistema para la enzima T en la etapa NVT	138
7.14. Temperatura, presión y energía del sistema en la etapa NPT	140

7.15. Propiedades medidas en la etapa de producción a las temperaturas de estudio	142
7.16. Medias y desviaciones del Rg.	144
7.17. Resultados del test t	144
7.18. Ranking de valores <i>if</i>	147
7.19. Media y desviación estándar del Rg a cada temperatura para las mutantes de Cel72	152
7.20. Análisis t-student entre las enzimas mutantes de Cel72 y la enzima nativa.	153
A.1. Parámetros para la minimización de energía.	190
A.2. Tabla de t-student	208

Abreviaciones y nomenclatura

CBM	<i>Cellulose Binding Module</i> , en español, Módulo de Unión a Celulosa
RACHITT	<i>Random Chimeragenesis on Transient Templates</i> , en español, Generación de Quimeras al Azar en Templados Transitorios
ISM	<i>Iterative Saturation Mutagenesis</i> , en español, Mutagénesis por Saturación Iterativa
RMSD	<i>Root mean square deviation</i> o Raíz cuadrática media de las desviaciones
C α -RMSD	<i>Root-mean-square deviation</i> o Raíz cuadrática media de las desviaciones de los carbonos alpha entre dos estructuras
Rg	Radio de giro
\hat{A}	Operador A
$[\hat{A}, \hat{B}]$	Conmutador de los operadores \hat{A} y \hat{B}
MM	Mecánica Molecular
NVT	Acople con número constante de partículas, volumen y temperatura
GROMACS	<i>Groningen Machine for Chemical Simulations</i>
PDB	<i>Protein Data Bank</i>
PCA	Análisis de Componentes Principales
CBH	Celobiohidrolasa
PBC	Condiciones de borde periódicas
SPC	<i>Simple point charge</i> o Carga puntual simple
PME	<i>Particle-Mesh Ewald</i>
Enzima T	Celobiohidrolasa Cel7A de <i>Talaromyces emersonii</i> , código PDB: 1Q9H
Enzima M	Celobiohidrolasa Cel7B de <i>Melanocarpus albomyces</i> , código PDB: 2RFW
Cel72	Celobiohidrolasa de la familia 7 de <i>Phanerochaete chrysosporium</i> .

1 . INTRODUCCIÓN

1.1. Definición y descripción del problema

Las industrias a nivel global se encuentran en continua carrera por obtener procesos menos costosos y más eficientes haciendo los proyectos más rentables con el principal objetivo de ser más competitivos. En busca por la eficiencia, uno de los factores importantes para la industria química y biotecnológica es el rendimiento de cada una de las etapas del proceso. Se busca mejorar el rendimiento de las reacciones para así obtener mayores cantidades de producto a menor costo. Para lograr esto existen diversas alternativas enmarcadas en el contexto de las condiciones de operación en las que se realizan las reacciones químicas como presión y temperatura entre otros. Por otro lado, es importante también la utilización de reactivos y catalizadores que ofrezcan procesos eficientes en costo y tiempo, para poder lograr rendimientos altos a tiempos de residencia por proceso unitario bajos.

Por otro lado, el desarrollo de nuevas tecnologías para la obtención de procesos más eficientes es un desafío en sí. No sólo los procesos propiamente tales son un cuello de botella para que una industria en particular sea altamente competitiva sino también lo es el desarrollo de nuevas herramientas para mejorar tales procesos debido a su alto costo y mediana eficiencia. Es por esto que también se busca mejorar en lo posible e innovar en técnicas y herramientas de desarrollo aplicadas a la obtención de procesos productivos más eficientes.

Una de las industrias en notorio crecimiento en los últimos años es la de los biocombustibles. Enmarcada en el contexto de la búsqueda de energías alternativas al petróleo, la producción de combustibles a base de etanol obtenido a partir de biomasa

ha pasado a ser una necesidad país para poder lograr un mayor grado de independencia energética. Para esto se necesitan procesos altamente competitivos en donde los costos y eficiencias de producción deben ser lo más cercanos posibles al óptimo. A diferencia de la mayoría de los procesos productivos químicos, la producción de biocombustibles, por ser un proceso biotecnológico, presenta una dificultad mayor al momento de intentar mejorar los niveles y eficiencia de producción. Esta dificultad se debe a que precisa de la utilización de enzimas en algunas de las etapas de la línea productiva. Las enzimas típicamente necesitan de condiciones de operación suaves respecto de las condiciones óptimas requeridas en los procesos químicos comunes. Esto implica que si se trabaja con enzimas es necesario alejarse de las condiciones óptimas de un proceso químico. Una de estas condiciones es la temperatura, por esto, se necesita de enzimas que sean más resistentes y que mantengan su actividad a temperaturas altas para poder trabajar en condiciones más cercanas a las que necesita un proceso químico para que sea de alto rendimiento.

1.2. Motivación

Para el mejoramiento de la termoestabilidad de enzimas existe un amplia lista de técnicas que permiten modificar las proteínas para hacerlas más resistentes al efecto de la temperatura. La mayoría de los enfoques implican un gran gasto en tiempo e insumos de laboratorio que obstaculizan el proceso de desarrollo de nuevas enzimas que ofrezcan alta termoestabilidad. Así, con esta motivación se ha buscado una metodología para la obtención de enzimas con termoestabilidad mejorada a partir de un nuevo enfoque, basado en el modelamiento computacional por homología y la simulación mediante dinámica molecular, para poder contar con una herramienta de bajo costo y relativamente rápida respecto a los métodos tradicionales.

Está descrito que un índice importante de termo estabilidad es la flexibilidad de la proteína (Sección 2.2.4). Es por esto que se ha decidido enfocar el procedimiento en la búsqueda de sitios de alta flexibilidad en la enzima y rigidizarlos para así mejorar la resistencia a altas temperaturas.

2 . ANTECEDENTES

2.1. Industria Biotecnológica

La aplicación de principios científicos e ingenieriles al procesamiento de materiales mediante el uso de agentes biológicos ha emergido con gran fuerza debido a las presiones hacia la industria química de mejorar los procesos y hacerlos más amigables con el medio ambiente [58, 70]. Así mismo, existe un reciente interés por productos de tipo alimentarios que además de sus cualidades nutricionales primarias como satisfacer el hambre o aportar con contenido calórico a la dieta tienen otro tipo de propiedades nutricionales y beneficios adicionales para la salud del consumidor como contenido de Omega 3, bajo contenido de colesterol y ácidos grasos trans, productos con alto contenido de probióticos, etc [23, 68, 96]. El creciente uso de procesos biotecnológicos [32] ha generado alta competitividad y la urgencia por proveer de procedimientos eficientes y de bajo costo.

Las exigencias del mercado moderno, la economía y política mundial están haciendo de la industria de servicios globales o offshoring una opción cada vez más atractiva para los bionegocios [24]. El fortalecimiento de las redes de negocios y los nuevos mercados han creado un espacio fértil para la investigación y desarrollo aplicados. Las políticas de control y restricción de precios en países industrializados junto a la creciente preocupación por los costos de la salud han provocado que las industrias farmacéuticas ofrezcan productos no solamente seguros y efectivos, si no también que poseen ventajas económicas [74], lo que ha generado la necesidad de desarrollar nuevas tecnologías que permitan agregar valor a los productos nuevos o existentes, para poder permanecer insertos en el mercado. Por otro lado la bio-

tecnología en procesos industriales requiere también del desarrollo de herramientas que permitan hacer más eficientes los procesos investigativos así como los procesos industriales.

En particular, la industria de la biotecnología utiliza microorganismos completos o algunos de sus componentes como parte fundamental de la línea productiva. Debido a las diferencias en los requerimientos de operación entre los procesos químicos y biológicos es necesario crear herramientas que acerquen las condiciones de operación óptimas de ambos y así hacer viable los procesos biotecnológicos. Para esto se precisa que, tanto los microorganismos como enzimas, presenten características que se adecuen a los requerimientos del proceso tomando en cuenta que el rendimiento se encuentra sujeto a las condiciones de operación de cada uno de los elementos interactuantes en la reacción. Las condiciones medioambientales en las que se pueden encontrar los agentes biológicos que se utilizan en los procesos son generalmente distintas y significativamente lejanas a las que se necesitan para que el proceso sea eficiente. Es por esto que es necesario el diseño y la ingeniería de microorganismos o partes de éstos, según se requiera, para lograr acercar en lo posible, los óptimos de operación y así maximizar la eficiencia del proceso.

2.1.1. Etanol como biocombustible

Una de las industrias en notorio crecimiento es la de los biocombustibles. Enmarcada en el contexto de la búsqueda de energías alternativas al petróleo, la producción de combustibles a base de etanol obtenido a partir de biomasa ha pasado a ser una necesidad país para poder lograr un mayor grado de independencia energética.

El incremento de uso de combustibles fósiles ha causado un incremento en las emisiones de gases de efecto invernadero y generado externalidades negativas y daño

al medio ambiente. Esto junto a la inestabilidad de los suministros de petróleo y la continua fluctuación de los precios ha provocado un gran interés en el uso y búsqueda de energías alternativas [121]. Por esto que una rama de la biotecnología, llamada biotecnología blanca [120], se ha interesado en la bioproducción de combustibles a partir de fuentes renovables. Estas tecnologías utilizan células y enzimas para sintetizar productos de fácil biodegradación, con requerimientos energéticos bajos y menor producción de desechos tóxicos y contaminantes durante la producción que los resultantes de la producción de combustibles a partir de fuentes fósiles. Dada la complejidad de la biomasa en términos de composición, un proceso de producción de biocombustibles típicamente conlleva varios procesos tales como la colección de la biomasa, la separación de los constituyentes de la materia prima para obtener monosacáridos ácidos grasos, entre otros, y finalmente la posterior conversión al biocombustible propiamente tal.

Entre los diferentes biocombustibles existentes el bioetanol destaca por ser uno de los biocombustibles líquidos más utilizados. En el año 2011, la producción mundial de bioetanol alcanzó a los 109 mil millones de litros. Los mayores productores en el mundo son América del Norte y Central (50%), América del Sur (20%) y Brasil (19%)¹. El año 2010 se esperaba un aumento del 6% en la producción de energías renovables debido al esfuerzo por reducir las emisiones de gases de efecto invernadero [148].

El etanol como combustible puede ser utilizado en numerosas maneras. Es utilizado como aditivo en gasolinas para reducir las emisiones de monóxido de carbono, óxidos nitrosos e hidrocarburos [145]. El etanol posee un mayor octanaje con respecto a los combustibles provenientes del petróleo, permitiendo a los motores a combustión

¹RFA: Renewable Fuels Association, F.O. Lichts: 2011 World Fuel Ethanol Production

funcionar a tasas de compresión mayores y así entregar un mejor desempeño [145]. Además, el etanol posee una mayor presión de vapor y calor de evaporación de la bencina por lo que se ha observado mayor potencia si se usa etanol [148].

El bioetanol es producido mediante la conversión directa de azúcares obtenidos de cultivos como caña de azúcar o remolacha, e indirectamente desde almidón de maíz, trigo, papas, o desde celulosa obtenida de biomasa, a etanol mediante la fermentación seguida de destilación. Mejorar el rendimiento de la producción de etanol y reducir los costos asociados a la producción es crítico para hacer del proceso viable. Durante la fermentación de caña de azúcar, cerca del 10 % de azúcares disponibles se desvían de la ruta de producción de etanol y son convertidos en glicerol y en ácido succínico. La síntesis de glicerol ocurre en respuesta al estrés osmótico y por tanto posee un rol esencial en la viabilidad de las células.

Los biocombustibles pueden ser producidos a partir de una gran variedad de materias primas como cultivos alimenticios y sustratos celulósicos. Los cultivos alimenticios contribuyen al 60 % de la producción mundial de bioetanol [148]. La expansión de la producción de biocombustibles junto con el incremento en la demanda mundial de cereales y el aumento de los costos de la energía a llevado al aumento drástico de los precios de los cereales [28].

La cosecha total mundial de granos se estimó en el año 2006 en 1.79 mil millones de toneladas, 4 % menos que el consumo estimado [80] siendo la sexta vez en los últimos 10 años que la producción de grano es menor que el consumo mundial. En los Estados Unidos, la competencia por tierras y otros recursos para la producción de maíz para biocombustibles ha llevado a la restricción en el uso de tierras y consecuentemente a la reducción de los niveles de producción. La búsqueda de menores costos en la producción de biocombustibles y la competencia con los cultivos

alimentarios tradicionales ha llevado a considerar los sustratos celulósicos como una potencial materia prima.

La celulosa es el biopolímero más abundante en la tierra con una cantidad estimada de 7.5×10^{10} toneladas sintetizadas anualmente a través de procesos fotosintéticos [102]. Se encuentra principalmente en la pared celular de células vegetales y está dentro de una matriz heterogénea compuesta también por xylano, hemi-celulosa y lignina. La celulosa es un biopolímero lineal, insoluble y compuesto por residuos repetidos de β -D-glucopiranososa ligados por enlaces glicosídicos β -1,4. En contraste con otros polímeros de glucano, como el almidón, la unidad repetida en la celulosa no es glucosa sino celobiosa, un disacárido. La celulosa presenta un alto grado de polimerización puesto que una cadena individual puede alcanzar largos superiores a 25.000 residuos de glucosa [44].

La celulosa producida por plantas está compuesta tanto de regiones amorfas con grandes huecos y otras irregularidades como regiones compactas. Debido a resistencia a la mayoría de las formas de degradación, la celulosa se acumula en la naturaleza. Se ha estimado que cerca de la mitad del carbón fijado anualmente en ecosistemas es almacenado como celulosa. La síntesis de celulosa se asocia principalmente a las plantas, sin embargo algunos animales, bacterias y algas también pueden producirla [21].

Enzimas empleadas para la degradación de celulosa

La degradación enzimática de la celulosa es realizada por enzimas endo-1,4- β -glucanasas (EC 3.2.1.4) y exo-1,4- β -glucanasas (celobiohidrolasas, EC 3.2.1.91). Las endoglucanasas son capaces de hidrolizar celulosa amorfa, carboximetilcelulosa y celulosa impregnada con ácido fosfórico, produciendo oligosacáridos solubles que son

luego degradados en celobiosa y glucosa mediante la acción de una β -glucosidasa (EC 3.2.1.21). Las celobiohidrolasas degradan la celulosa cortando unidades de celobiosa del extremo no reducido de una fibra de celulosa [57]. Las celulasas son enzimas inducibles sintetizadas por una gran diversidad de microorganismos, entre ellos bacterias y hongos [82, 119]. Estos microorganismos pueden ser aeróbicos, anaeróbicos, mesofílicos o termofílicos. Entre los más estudiados productores de celulasas se encuentran las bacterias *Clostridium*, *Cellulomonas*, *Thermomonospora*, y los hongos de los géneros *Trichoderma*, y *Aspergillus* [83, 84, 129, 130].

Las celulasas provenientes de hongos termofílicos han emergido recientemente como una alternativa para aplicaciones biotecnológicas, debido a que, por las condiciones ambientales naturales en las que se desenvuelven, producen enzimas más estables. Sin embargo, se han caracterizado en detalle una pequeña cantidad de celulasas termofílicas provenientes de hongos. Herramientas como mutación sitio-dirigida y evolución dirigida (Sección 2.3.1) han sido utilizadas de manera preferente para la obtención de enzimas más termoestables. El mejoramiento de celulasas termofílicas permitirá el desarrollo de celulasas mejores y más versátiles para su aplicación en biotecnología y entregar nuevas oportunidades para los esfuerzos en ingeniería de proteínas [89].

2.2. Enzimas

2.2.1. General

Las proteínas, y por lo tanto las enzimas, están compuestas por una cadena de aminoácidos que se pliegan de acuerdo a la química característica de cada residuo y a su interacción con los residuos vecinos, en una estructura tridimensional que le entrega funcionalidad [38]. Dependiendo de la estructura tridimensional de una en-

zima, se establecen estructuras más complejas de modo que uno o más aminoácidos conformarán un centro catalítico denominado sitio activo. Además, la conformación espacial entrega características estéricas que pueden ser importantes en la actividad de una enzima, por ejemplo, para el ingreso del sustrato en la enzima y su aproximación al sitio catalítico [38]. Es importante destacar que el plegamiento de una enzima será estable según las condiciones ambientales que conforman el contexto en el que se desenvuelve la enzima en cuestión.

Los aminoácidos poseen características químicas diferentes entre ellos, por ejemplo, algunos aminoácidos poseen carga eléctrica, la que dependerá del pH del medio en el que se encuentra. Algunos aminoácidos son polares y otros apolares, por lo que poseen diferente afinidad al agua [98]. Los aminoácidos polares, más afines al agua, tenderán a exponerse en una enzima y los apolares a ocultarse hacia el interior de la proteína, si es que el medio externo es polar. La interacción que tiene un aminoácido con otros residuos vecinos también dependerá del ambiente puesto que, como se mencionó, algunos aminoácidos poseen carga y pueden formar puentes iónicos mediante interacciones de tipo coulombicas con otros residuos cargados. Además existen residuos que pueden formar puentes de hidrógeno. Estas interacciones pueden variar, también, dependiendo de la temperatura puesto que cada interacción como las mencionadas es débil, es decir, son reversibles y requieren de poca energía para ser disueltas.

2.2.2. Plegamiento de las enzimas

La secuencia de aminoácidos contiene toda la información necesaria para el plegamiento de una cadena polipeptídica en su estructura tridimensional. Esto quiere decir que bajo determinadas condiciones la cadena de aminoácidos establecerá la

forma de la enzima, o sea, se plegará según los residuos por los que esté constituida y su orden, plegándose primero estructuras locales y luego estructuras más complejas para formar la estructura final

Se postuló que la estructura nativa de una proteína es una estructura termodinámicamente estable, depende solamente de la secuencia de aminoácidos y en las condiciones de la solución en la que se encuentra [8,63], no depende de cómo la enzima fue sintetizada. Hay algunas excepciones como la insulina, proteasa α -lítica [124], y serpinas [144].

Hasta antes de los años 80 se propuso que el mecanismo de plegamiento es una suma de diferentes interacciones como puentes de hidrógeno, puentes iónicos, interacciones de van der Waals e interacciones hidrofóbicas. De esta manera la secuencia primaria determina las estructuras secundarias de la enzima y éstas a su vez determinan la estructura terciaria [9]. Sin embargo, mediante modelamiento estadístico nació una nueva propuesta: el componente dominante son las interacciones del tipo hidrofóbicas, siendo la estructura secundaria consecuencia del plegamiento terciario [46,47]; sin embargo, ningún tipo de interacción intermolecular debe ser descartada [147].

Los puentes de hidrógeno son un tipo de interacción importante debido a que la mayoría de los enlaces de hidrógeno posibles en una estructura nativa están presentes. Los puentes de hidrógeno entre amidas del esqueleto de una proteína y grupos carbonilos son una clave importante de la estructura secundaria [11, 27, 43, 55]. El estrecho empaque en las proteínas implica que las fuerzas de van der Waals también son interacciones importantes [31].

A pesar de que existe una gran cantidad de estructuras posibles para una enzima dada su estructura primaria, las proteínas pueden converger rápidamente a su estado nativo [134], algunas veces en microsegundos, lo que podría tardar en un

computador hasta años. La manera de cómo se pliega una enzima está explicada de diferentes maneras según distintos enfoques, los que finalmente se complementan para dilucidar el mecanismo de plegamiento. En el modelo de difusión-colisión, microdominios se forman primero luego difunden y colisionan para formar estructuras más grandes [76,77]. El mecanismo de nucleación-condensación propone que estados de transición difusos junto a algunas estructuras secundarias conforman núcleos que forman posteriormente la estructura terciaria. Algunas proteínas parecieran seguir un mecanismo de difusión-colisión jerárquico, en donde las estructuras secundarias se forman y ensamblan siguiendo un orden jerárquico [95,104].

Se han construido modelos para estudiar la naturaleza del espacio conformacional² [25,45], es decir la forma de la superficie de energía de las posibles conformaciones por las que tiene que pasar una proteína para alcanzar su estructura nativa, expresado mediante una función que describe la energía libre de las interacciones intramoleculares y de la solvatación de una proteína dada en función de los grados de libertad. Se determinó que las proteínas poseen un espacio conformacional con forma de embudo, es decir, muchos estado de alta energía y pocos estados más estables de baja energía [25,47–49,88]. De esta forma, las proteínas se denaturan a altas temperaturas debido a que al entregar energía al sistema se está elevando el estado energético y se hacen posibles conformaciones que antes no eran estables.

Muchas enzimas poseen puentes disulfuro en su estructura. Estos puentes son enlaces covalentes entre dos grupos tiol y aportan en la rigidez de las enzimas estabilizando su estructura terciaria haciendo de ellos una herramienta atractiva para mejorar la estabilidad de enzimas a través de mutaciones sitio dirigidas.

²El espacio de estructuras posibles de una enzima tanto finales como las intermedias necesarias para llegar a la conformación final

2.2.3. Flexibilidad de las proteínas

Existen diferentes tipos de interacciones interatómicas que definen campos de fuerzas que a su vez, dado un determinado estado energético del sistema, establecen restricciones de desplazamiento para los átomos. Estas interacciones pueden ser enlazantes, coulómbicas y de van der Waals.

El conjunto de interacciones establece restricciones para el desplazamiento de los átomos que componen una proteína. Esto hace que los átomos se organicen de una manera particular buscando siempre un estado de menor energía.

Existe una relativa libertad de desplazamiento de los átomos conformantes de una proteína producto de que el estado basal de éstos posee una energía cinética distinta de cero aportada principalmente por la temperatura del sistema. Esta libertad no dependerá tan solo de cada átomo sino también de los átomos vecinos que interactúen con éste. Así, en una enzima, los átomos interactúan en diferente grado con todos los otros átomos presentes en la macromolécula y con el solvente que los rodea produciéndose restricciones de desplazamiento según el estado energético. Dado que todo sistema busca el menor estado energético posible, los átomos presentarán la configuración espacial más cómoda en términos energéticos según se respete el potencial definido por la configuración global del sistema. Esto produce efectos macro-moleculares como el plegamiento particular a cada proteína ya que si dos enzimas presentan características idénticas en su composición, así también será el potencial energético que definen sus estructuras si se encuentran bajo las mismas condiciones, por lo tanto éstas serán idénticas en estructura. Lo opuesto sucederá si ambas enzimas son diferentes. Este plegamiento será necesario para que una enzima sea activa. Se deben configurar espacialmente los aminoácidos de una manera particular para conformar el sitio activo y si se requiere, sitios de unión a sustrato.

Dado que los átomos de un sistema pueden desplazarse de acuerdo a las fuerzas a las que están sometidos, las mismas registrarán el comportamiento de regiones más amplias en una macromolécula. Será posible identificar dos tipos principales de comportamientos en términos del movimiento a nivel macro-atómico en una enzima: (1) la movilidad, la cual corresponde al desplazamiento que una región de la enzima realiza a determinadas condiciones de temperatura, y (2) la deformabilidad, que corresponde al cambio estructural de una región a determinadas condiciones de temperatura sin involucrar necesariamente grandes desplazamientos [81] (Figura 2.1). Ambos comportamientos son variables en el tiempo, es decir, si el sistema es estable se apreciarán deformaciones y movimientos en una región determinada a lo largo del tiempo. Así, se define la flexibilidad de una enzima o de una región de ésta como la combinación de los comportamientos mencionados anteriormente. Una región será flexible si ésta presenta movimientos de sus átomos componentes tal que se presenten deformaciones y desplazamientos, a una temperatura determinada, modificando la estructura de la enzima en el tiempo.

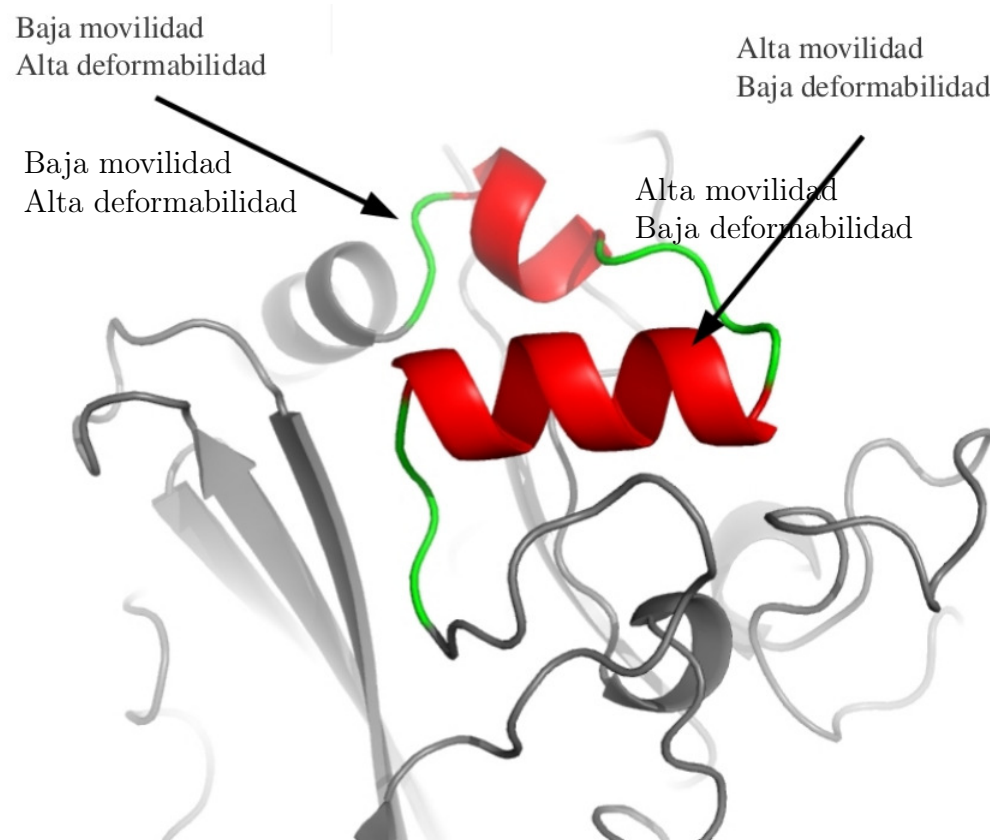


Figura 2.1: Movilidad y Deformabilidad. En la figura se señala una zona de alta deformabilidad y baja movilidad, Esta región actúa como una bisagra con amplitudes de desplazamiento bajas, sin embargo, alta deformabilidad. También se señala una región de alta movilidad en donde ocurre un desplazamiento amplio de toda una región pero sin grandes cambios conformacionales de esta misma. Adaptado de Julio A. Kovacs y cols. [81]

2.2.4. Flexibilidad y Termoestabilidad

En una proteína, su actividad se ve afectada con el aumento de la temperatura. De hecho es posible reducir considerablemente la actividad de una enzima con solo calentar el sistema a una temperatura por sobre su óptimo. Dependiendo de qué tipo de enzima se trate, será como cambia la actividad frente la variación de temperatura

respecto de la temperatura óptima de la enzima. Al comparar enzimas que en la naturaleza se presentan como termoestables con enzimas de temperatura óptima menores es posible notar ciertas características comunes en enzimas termofílicas [138]. Así, una enzima termofílica tendrá mayor actividad a mayores temperaturas en comparación con enzimas mesófila o criofílicas. Ha sido posible ver que proteínas termofílicas son menos flexibles que enzimas mesofílicas a temperaturas mesofílicas [39], dando señales de la relación entre flexibilidad y termoestabilidad.

La identificación de los factores que contribuyen a la termoestabilidad de enzimas termófilas ha sido un desafío durante varios años [42,86,100,107]. Como se mencionó, la temperatura tiene efectos en las interacciones intramoleculares en una enzima, es por esto que al aumentar la temperatura del sistema por sobre la temperatura óptima se producen cambios conformacionales que pueden afectar la actividad de una enzima. Por otro lado, algunas restricciones locales en la flexibilidad de una proteína afecta su desempeño catalítico. Se ha visto como el diseño de enzimas termoestables logran incrementar la estabilidad a altas temperaturas al reducir la flexibilidad local [142]. La flexibilidad puede afectar no solamente la función si no también la estabilidad [17]. La identificación de sitios altamente flexibles puede ser utilizado para mejorar las propiedades catalíticas entre ellas la termoestabilidad [116]. Por el contrario, para enzimas que trabajan a bajas temperaturas un aumento en la flexibilidad puede mejorar las características catalíticas [85,92].

Se ha propuesto que enzimas termófilas no solo son termoestables si no que también son más estables frente a la mayoría de los denaturantes comunes respecto de enzimas mesofílicas. Debido a su destacada flexibilidad, enzimas aisladas desde microorganismos termófilos pueden tener aplicaciones tecnológicas [137].

Se ha visto que la energía de activación de enzimas en reacciones oxido-reducción

decrece con el aumento de la temperatura y que este comportamiento puede ser correlacionado con los cambios en la actividad dependientes de la temperatura indicando una reorganización estructural de la enzima dependiente del tiempo. Comparaciones entre enzimas mesofílicas y termófilas indican que la flexibilidad conformacional de ambos tipos de enzimas es similar a sus respectivas temperaturas fisiológicas [146].

Se han realizado simulaciones mediante dinámica molecular para entender la estructura, dinámica y flexibilidad de enzimas y como éstas pueden ser afectadas por la temperatura. De esta forma, aumentando la temperatura en una simulación la vía de denaturación no se ve modificada sino su velocidad, siendo más rápida a mayores temperaturas [41, 75]. Así, un mejor entendimiento de estas propiedades puede proveer una guía para el mejoramiento de la estabilidad a través de la identificación de regiones de alta flexibilidad críticas para la termoestabilidad, y así ser útil para el diseño de biocatalizadores más estables para el uso industrial. Hay ejemplos de mejoramiento de la termoestabilidad mediante random mutagénesis [123], la introducción de secuencias de consenso [109] y la introducción de puentes disulfuro [91, 112], sin embargo, poder utilizar herramientas computacionales para tales efectos podría disminuir el tiempo y costos económicos para la obtención de enzimas mejoradas.

2.2.5. Celulasas de hongos

Las celulasas están presentes en las familias de glicosil hidrolasas 1, 3, 5, 6, 7, 8, 9, 10, 12, 16, 44, 45, 48, 51, y 61. Las variantes termófilas están presentes en las familias 1, 3, 5, 6, 7, 12 y 45. La producción de celulasas de hongos es comúnmente inducida bajo la presencia de celulosa y controlada mediante un sistema de inducción-represión [131]. La celulosa o oligosacáridos obtenidos producto de la degradación de la celulosa participan como inductores mientras que moléculas de glucosa actúan

como inhibidores [16, 35, 60, 103, 113].

Una característica común en las celulasas es su estructura molecular. Endoglucanasas y celobiohidrolasas están compuestas por cuatro dominios, un péptido señal para exportación extracelular, un módulo de unión a carbohidratos (CBM, por cellulose binding module), un dominio catalítico, y un puente o *linker* de unión entre el CBM y el dominio catalítico, que es rico en Ser, Thr, y Pro. Las proteínas maduras están O-glicosiladas y N-glicosiladas en el linker y en el dominio catalítico, respectivamente. Existen variantes entre las diferentes clases de celulasas, que combinan el orden o la presencia de dominios, con la restricción de que siempre están presentes el dominio catalítico y la señal de exportación. La sustracción del CBM de Cel7A de *T. reesei* y de CBH1 de *H. grisea* provocó una reducción en la actividad enzimática sugiriendo que se requiere de una unión firme entre la enzima y el sustrato mediado por el CBM [132].

2.2.6. Celulasas de hongos termófilos

Los hongos termófilos son especie que crecen a temperaturas máximas por sobre los $50^{\circ}C$ y mínimas por sobre los $20^{\circ}C$ [93], por lo que son fuente de enzimas con el potencial de se utilizadas para degradación de biomasa como una aplicación biotecnológica. Para la degradación de la celulosa se requiere que las enzimas estén en contacto directo con el sustrato, sin embargo, el empaque intrincado que posee requiere de pretratamientos para mejorar la eficiencia de la degradación del sustrato. Entre estos pretratamientos se encuentra el calentamiento que hace que la celulosa se hinche y se fraccione con mayor facilidad. De ahí la importancia de trabajar con celulasas termoestables.

2.3. Ingeniería de Proteínas

Los requerimientos específicos de operación de industrias que utilizan enzimas en sus procesos no siempre coinciden con las condiciones óptimas de operación de la enzimas, por lo que se hace necesario buscar enzimas que tengan tanto la actividad catalítica que se necesita y además condiciones de operación cercanas al óptimo del proceso. En la búsqueda de enzimas con características que respondan a las necesidades, la modificación de enzimas ya existentes es una de las alternativas, de esta forma, es posible obtener enzimas mejor adaptadas a las condiciones de operación requeridas.

2.3.1. Métodos de Mejoramiento Enzimático

Uno de los principales objetivos en el diseño de enzimas es la generación de enzimas con propiedades mejoradas. Existen dos grandes enfoques para enfrentar el problema de modificar una enzima. Estos enfoques son de carácter experimental y requieren de etapas posteriores al proceso de mejoramiento mismo que pueden ser altamente intensivas en análisis. Un enfoque es el diseño racional, en donde las proteínas son modificadas mediante mutaciones puntuales definidas a base del entendimiento que se tiene de su estructura, mecanismos de reacción y de las consecuencias de un cambio puntual, o varios cambios, en las propiedades de la enzima.

Por otro lado, la evolución dirigida de proteínas abarca un conjunto de técnicas para la generación de variantes y la selección de estas a base de una característica de interés. Una de las estrategias más efectivas es acumular gradualmente mutaciones a medida que se aplica una presión selectiva. Esto significa generar múltiples variantes de una enzima, luego seleccionar según la característica que se está buscando mejorar y finalmente repetir el proceso sobre la o las mejores variantes para obtener como

producto una o más enzimas con varias mutaciones adquiridas de forma aditiva con cualidades resaltadas. Esto se puede lograr mediante la generación de grandes librerías de mutantes seguido de una *screening* eficiente para la propiedad de interés.

El éxito de la evolución dirigida recae en la relación que existe entre la calidad de la diversidad en la librería de variantes, el tamaño de ésta y de la habilidad de detectar el grado de mejoras en las variantes. La diversidad de las librerías es generada mediante mutagénesis o recombinación. Típicamente, las librerías han sido creadas mediante mutaciones al azar como *error-prone* PCR, o mutaciones sitio dirigidas. La frecuencia de mutaciones que mejora una enzima es relativamente baja, primando principalmente las mutaciones deletéreas, por lo que en cada ciclo sólo se agregan mutaciones favorables³ únicas.

Una técnica llamada DNA shuffling supera esta limitación permitiendo la recombinación directa de mutaciones favorables provenientes de múltiples genes. Una población de secuencias de DNA son fragmentadas al azar y luego reensambladas mediante PCR [125], en donde se introducen las mutaciones controlando la tasa de error de la polimerasa. Luego de seleccionar las variantes, éstas son utilizadas para generar de manera iterativa secuencias con múltiples mutaciones favorables mezclando las secuencias. Existen variantes del método DNA shuffling que buscan mejorar la diversidad biológica de las librerías y así obtener un espacio de variantes más amplio para detectar las mejoras de manera más temprana [37, 105].

Debido que para algunas aplicaciones industriales como la conversión de materia prima o la producción de proteínas para uso farmacéutico, se usan células completas en lugar de sólo enzimas, muchas veces la mejora de una sola enzima no es suficiente para adaptarse a las condiciones de operación del proceso. En su lugar, se puede

³Las mutaciones favorables corresponden a mutaciones en una enzima que generar un cambio benéfico según la o las características que se desean mejorar

hacer un shuffling del genoma completo de la célula [150] combinando las ventajas del DNA shuffling con los beneficios de la cruce convencional.

La recombinación de genes grandes como operones es difícil de lograr *in vitro*. Una alternativa es la recombinación *in vivo-in vitro* en un plásmido heteroduplex parental en un proceso de reparación *in vivo* [140]. Un heteroduplex fabricado *in vitro* es transformado dentro de células de bacteria en donde las regiones no homólogas son reparadas formando una librería de nuevas secuencias recombinadas formadas con elementos del heteroduplex parental. La desventaja de este método es que se requiere genes parentales con alta homología y está limitada a dos genes parentales por evento.

Otro ejemplo de evolución dirigida es la generación de librerías mediante la generación de quimeras al azar en templados o plantillas transitorias (RACHITT) [34]. Este método genera un templado transitorio de hebra simple con urácilo y fragmentos de hebra simple como donadores parciales. Como uno o más de los fragmentos donadores pueden acoplarse simultáneamente al templado, este enfoque genera crossovers de alta frecuencia.

También es posible generar mutaciones mediante el uso de un conjunto de oligonucleótidos sintetizados artificialmente con regiones homólogas a la zona a modificar y luego se completa la doble hebra mediante PCR. De esta manera se permite el *crossing over* de hebras de baja homología [149].

Además de las estrategias de evolución dirigidas mencionadas aquí y muchas otras, es posible reducir el espacio de variantes si es que existe información suficiente de la enzima como para enfocar el trabajo en regiones acotadas de la secuencia y así reducir el tamaño del *screening*. Un ejemplo de esto es la mutagénesis por saturación iterativa (ISM) [115] en donde se mejora la termoestabilidad de una enzima

2.4. DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE UNA ENZIMA

utilizando mutagénesis por saturación en sitios de alta flexibilidad determinados por los valores B obtenidas a partir de espectroscopia por resonancia magnética nuclear (NMR). De esta manera, se seleccionan regiones flexibles como sitios candidatos a mutar y se aplican, en un primer paso, mutagénesis por saturación en esos sitios, generando mutantes con mutaciones puntuales. Luego se evalúa la termoestabilidad y se seleccionan las mejores mutaciones. Sobre la más termoestable se realiza nuevamente una mutagénesis por saturación en el sitio donde se había mutado la segunda mejor enzima. Esto se itera otras tres veces hasta mejorar duplicar la termoestabilidad de la enzima en cuestión.

2.4. Determinación de la estructura tridimensional de una enzima

2.4.1. Determinación experimental

La determinación de la estructura de una enzima se puede realizar, en términos generales, de dos maneras: experimental y computacional. Hay dos principales alternativas para determinar experimentalmente la estructura tridimensional de una enzima. La primera consiste en obtener la configuración espacial de los átomos de una enzima a través de la difracción de rayos X al ser difractado por los átomos de la enzima cristalizada. Para una proteína el patrón de difracción debe ser analizado por computador. La luz emitida desde una fuente puntual se enfoca sobre un cristal de proteínas. Las ondas son dispersadas y captadas por sensores que registran el patrón de dispersión (Figura 2.2). La primera etapa consiste en cristalizar la proteína lo más condensada posible para obtener información detallada de la estructura. Una vez hecho esto se coloca el cristal entre un haz de rayos X y un detector. Del patrón de señales difractadas se genera un mapa de densidad electrónica de la proteína

2.4. DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE UNA ENZIMA

utilizando la transformada de Fourier de las señales.

Cabe destacar que el entorno físico de una enzima en un cristal no es idéntico al de la enzima en solución ya sea *in vitro* o *in vivo*. Debido a que se trabaja con un cristal de proteína, los estudios de difracción de rayos X aportan poca información de los movimientos intermoleculares de la enzima, inexactitudes que se generan debido a la interacción proteína-proteína dentro del cristal.

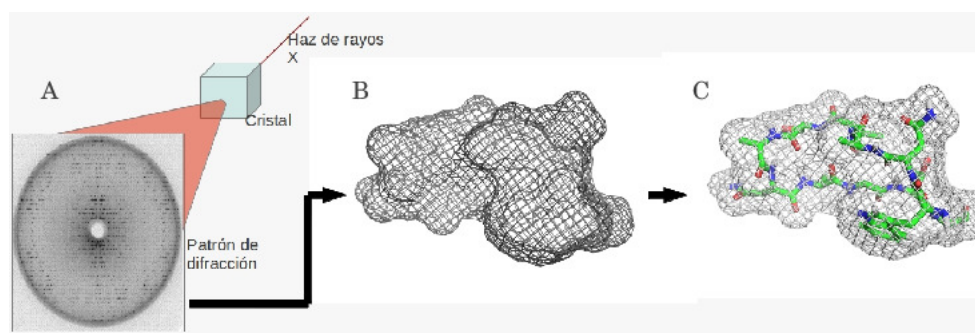


Figura 2.2: Representación del proceso para determinar la estructura de una enzima mediante rayos X. (A) La primera etapa consiste en obtener el patrón de difracción. (B) Luego, los datos obtenidos se procesan para determinar la densidad electrónica de la estructura. (C) Posteriormente es posible inferir la posición de los átomos y determinar la estructura final.

El segundo método implica la utilización de NMR. Con respecto a la difracción de rayos X, posee la ventaja de poder trabajar con proteínas en solución entregando además información de la dinámica de la molécula. La NMR es una manifestación del momento angular debido al spin del núcleo de algunos átomos, entre los que están ^1H , ^{13}C , ^{15}N , ^{19}F y ^{31}P . Debido al spin nuclear se genera un dipolo magnético que se alinea al ser sometido a un campo magnético estático. Al aplicar un pulso electromagnético a la frecuencia de resonancia, parte de la energía es absorbida por el núcleo dejando un espectro resultante con información que permite determinar la identidad del núcleo y de su entorno inmediato.

2.4. DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE UNA ENZIMA

Se puede determinar el espectro de NMR para ^1H . Mediante la medición del acoplamiento del spin de átomos vecinos y de átomos enlazados covalentemente permite determinar ciertas restricciones espaciales para los átomos. La información de restricciones de distancia obtenidos, junto a restricciones como radios de van der Waals, ángulos y distancias de enlace se utilizan para obtener, de forma computacional, familias de estructuras (Figura 2.3) que implican una gama de estructuras válidas de acuerdo con las restricciones impuestas. Este conjunto de estructuras reflejará las vibraciones intermoleculares de la proteína.

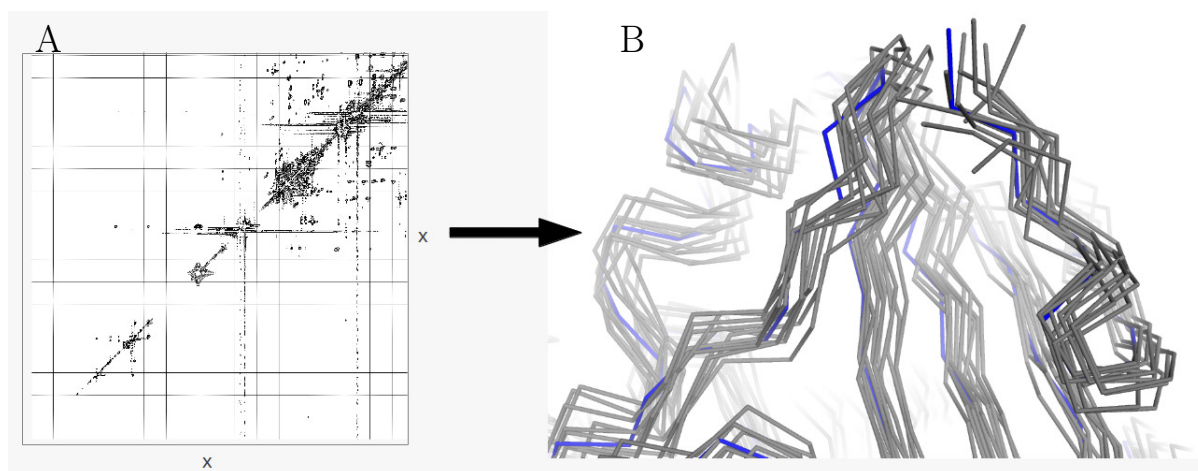


Figura 2.3: Representación de las múltiples estructuras de una enzima obtenidas a través de NMR. En A se muestra una representación del gráfico de correlación el espectro obtenido en un experimento de NMR y permite determinar la identidad del núcleo y de su entorno inmediato. De esta manera es posible deducir la estructura de una proteína. En B se muestra una representación de las múltiples estructuras obtenidas del NMR producto de las vibraciones intermoleculares.

2.4.2. Predicción de estructuras

No todas las enzimas descritas ni tampoco las no descritas poseen registros experimentales de su estructura tridimensional. Para estudios de la dinámica estructural

2.4. DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE UNA ENZIMA

de una enzima se requiere conocer su conformación y cómo los átomos se configuran espacialmente. Para una enzima cuya estructura es desconocida, existen métodos que permiten determinarla de manera computacional sin tener que recurrir a ensayos experimentales. Estos métodos pueden ser dependientes o no de información recopilada de estructuras ya conocidas, generalmente, de manera experimental. Los métodos no dependientes de información de otras enzimas, denominados métodos *ab initio*, se basan en la química propia de la enzima e intentan predecir el plegamiento de ésta mediante la optimización del potencial sujeto a restricciones impuestas por la física del problema, o bien, mediante la utilización de química cuántica para determinar el estado de los átomos y su configuración para así dilucidar la estructura de las enzimas. Los métodos *ab initio* pueden ser muy costosos en términos de la cantidad y complejidad de las ecuaciones que se deben resolver, sobre todo los que utilizan química cuántica. Por otro lado, los métodos dependientes de otras estructuras se basan principalmente en la similitud de la secuencia primaria para luego determinar de una u otra forma la estructura terciaria por lo que se utiliza la información de la estructura de las otras estructuras para aproximar la configuración de la estructura de la enzima que se está modelando para luego optimizar el potencial.

El modelamiento por comparación surge de la necesidad de conocer estructuras no determinadas experimentalmente. Un enfoque para elaborar un modelo sin requerir de un desarrollo experimental es realizarlo a partir de estructuras ya conocidas. Debido a la enorme cantidad de estructuras depositadas y a la homología estructural que poseen algunas enzimas debido a la conservación de sus secuencias en diferentes organismos, es posible modelar una enzima comparando su secuencia a la de estructuras ya conocidas.

El objetivo de la comparación es la obtención de un modelo válido en términos

2.4. DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE UNA ENZIMA

de la calidad del modelo que se necesita de acuerdo al uso que se le va a dar. La calidad del modelo dependerá principalmente del alineamiento con la o las secuencias que se utilicen como plantilla para generarlo. Una identidad de secuencia sobre el 30 % puede generar estructuras de buena calidad, con un $C\alpha$ -RMSD⁴ entre el modelo generado y el de la plantilla menor a 2Å.

Las principales etapas del modelamiento por homología consisten primero en determinar cuál o cuales son las secuencias similares a la secuencia de la enzima que se desea modelar. Se debe determinar a qué familia pertenece la enzima y buscar mediante alineamiento con múltiples secuencias cuál es la que posee mejor nivel de identidad. Una vez hecho el alineamiento se pueden conocer las regiones de mayor similitud y así establecer una aproximación a la estructura terciaria mediante la identificación de estructuras secundarias posibles a lo largo de la secuencia. Posteriormente se genera el modelo propiamente tal adecuando la estructura a la de la plantilla a modo de minimizar las diferencias entre ambas estructuras, mientras se minimizan los contactos erróneos entre átomos que pudieran estar generando interacciones desfavorables en términos energéticos. Esto se realiza mediante una minimización de energía de la estructura de la enzima, a la vez que se adecua la estructura a la plantilla. Con esta minimización de energía se intenta evitar superposiciones entre átomos, interacciones moleculares desfavorables, de modo de acomodar los residuos en una configuración lo menos tensionada posible.

Existen diversos programas computacionales que permiten llevar a cabo un modelamiento por homología. Entre estos están: Swissmodel [10, 78, 111], 3D-JIGSAW [12, 13, 36] y Modeller [52, 56, 94, 118]

Una vez realizado el modelo es necesario evaluarlo para determinar si éste posee

⁴Root-mean-square deviation de los carbonos alpha entre dos estructuras (Sección 6.2.3).

la calidad que se requiere para el uso que se le dará. Entre las muchas herramientas de evaluación de estructuras se encuentran: PROCHECK [87] que revisa la calidad estereoquímica de la proteína y ANOLEA [99] evalúa la calidad del empaque del modelo.

2.5. Dinámica de una Enzima: simulación *in silico*

2.5.1. Generalidades

Una manera de profundizar el entendimiento del comportamiento de sistemas biológicos es mediante el uso de recursos matemáticos; actualmente esto es imprescindible para complementar la información experimental que se tiene de tales sistemas y así desarrollar estrategias para su optimización y uso en tecnologías aplicadas a diversos áreas de la industria biotecnológica.

Una de las herramientas matemáticas es la simulación computacional. Esta permite evaluar distintos escenarios para un sistema y extraer información que de manera experimental no se podría. Dentro de las aplicaciones de simulaciones computacionales aplicadas a sistemas biológicos se encuentran las de redes neuronales [114,127], de redes metabólicas [29, 30, 53, 73], interacciones ligando-receptor [2], mecanismos de reacción e interacciones atómicas, entre muchas otras.

Un método de simulación computacional puede ser aplicado a la resolución de diversos problemas y la confianza de sus resultados dependerá del modelo matemático que se esté aplicando y de la precisión de los cálculos. Consiste básicamente en realizar experimentos ficticios a base de las reglas que impone un modelo. Por ejemplo, en el caso de una red metabólica, es necesario tener un modelo para la cinética de cada componente y algunos datos de entrada para poder obtener una dinámica del comportamiento a lo largo del período de simulación. Para el caso de la dinámica

molecular es necesario conocer el modelo que interprete las leyes físicas bajo las cuales está sometido el sistema y las condiciones iniciales para conocer el comportamiento del sistema en el tiempo.

2.5.2. Dinámica Molecular

La simulación del comportamiento de una macromolécula frente al efecto de la temperatura y del solvente, se le conoce como Dinámica Molecular y consiste en conocer la posición de cada uno de los componentes del sistema en todo tiempo. A partir de esto se elaboran los análisis que se requieran según el tipo de estudio que se realice. La manera de lograr esto dependerá de los recursos computacionales de los que se disponga. En este sentido, se deberá intentar lograr un equilibrio entre la fidelidad de la representación del sistema respecto de la realidad frente al costo de implementar y resolver el problema. Generalmente es preferible considerar algunas aproximaciones que faciliten el cálculo. Esto no necesariamente resulta en una falta de exactitud en los resultados si es que se adoptan las suposiciones adecuadas y si el modelo y la simulación se encuentran correctamente calibrados.

A pesar de la aproximaciones que se hagan, es posible extraer de la resolución del problema un análisis cualitativo que entregue direcciones de acción. Si bien es difícil representar el problema como una fiel copia de la realidad, será posible aplicar las conclusiones de análisis a la resolución de un problema real según sea la pregunta que se haya planteado desde un comienzo.

La teoría detrás de la dinámica molecular, algunas aproximaciones matemáticas comúnmente utilizadas y las herramientas computacionales para su resolución se presentan y explicarán a continuación. En esta Tesis se utilizó GROMACS para realizar las simulaciones [19, 67, 90, 136], por lo que se presentan definiciones y conceptos

necesarios para entender el uso y procedimientos llevados a cabo con este *software*.

Medida simultánea de varias propiedades

Se considerará a continuación el criterio a utilizar para decir a qué propiedades del sistema se pueden asignar valores definidos simultáneamente.

Si una función de estado Ψ es función propia del operador \hat{A} con valor propio s , entonces la medida de la propiedad física A proporcionará con seguridad el valor s . Si Ψ es simultáneamente función propia de dos operadores \hat{A} y \hat{B} , (es decir, si $\hat{A}\Psi = s\Psi$ y $\hat{B}\Psi = t\Psi$), entonces podemos asignar, también simultáneamente, valores definidos a las cantidades físicas A y B . Es necesario poder conocer cuándo es posible que Ψ sea función propia de dos operadores a la vez. Primero, la condición necesaria para que exista un conjunto completo⁵ de funciones propias simultáneas de dos operadores es que los operadores conmuten entre sí. Segundo, si \hat{A} y \hat{B} son dos operadores que corresponden a magnitudes físicas y que conmutan, entonces existe un conjunto completo de funciones que son funciones propias de los citados operadores \hat{A} y \hat{B} . Por tanto, si $[\hat{A}, \hat{B}] = 0$, entonces Ψ puede ser una función propia de \hat{A} y \hat{B} .

Operadores

Los operadores que intervienen en mecánica cuántica son lineales. \hat{A} es un operador lineal si y sólo si cumple las dos propiedades siguientes:

$$\hat{A}[f(x) + g(x)] = \hat{A}f(x) + \hat{A}g(x) \quad (2.1)$$

$$\hat{A}[cf(x)] = c\hat{A}f(x) \quad (2.2)$$

⁵Un conjunto de funciones $g_1, g_2, \dots, g_i, \dots$ se dice que es un conjunto completo si cualquier función f que se comporte bien y que obedezca las mismas condiciones de límite que las funciones g_i , puede desarrollarse como una combinación lineal de estas funciones o sea, $f = \sum_i a_i g_i$ donde los coeficientes a_i son constantes.

donde f y g son funciones arbitrarias y c es una constante arbitraria (no necesariamente real)

El cuadrado de un operador se define como el producto de un operador por si mismo: $\hat{A} = \hat{A}\hat{A}$. La potencia n -ésima de un operador ($n = 1, 2, 3, \dots$) se define como la aplicación n veces sucesivas del operador.

Identidad de conmutadores

Definimos el conmutador $[\hat{A}, \hat{B}]$ de los operadores \hat{A} y \hat{B} como el operador $\hat{A}\hat{B} - \hat{B}\hat{A}$:

$$[\hat{A}, \hat{B}] \equiv \hat{A}\hat{B} - \hat{B}\hat{A} \quad (2.3)$$

$$[\hat{A}, \hat{B}] f = \hat{A}\hat{B}f - \hat{B}\hat{A}f \quad (2.4)$$

donde f es una función derivable. Si $\hat{A}\hat{B} = \hat{B}\hat{A}$, entonces $[\hat{A}, \hat{B}] = 0$ y en ese caso decimos que \hat{A} y \hat{B} conmutan y si $\hat{A}\hat{B} \neq \hat{B}\hat{A}$, entonces \hat{A} y \hat{B} no conmutan.

Se muestran a continuación las siguiente identidades, que pueden ser demostradas desarrollando los conmutadores:

$$[\hat{A}, \hat{B}] = -[\hat{B}, \hat{A}] \quad (2.5)$$

$$[\hat{A}, \hat{A}^n] = 0, n = 1, 2, 3, \dots \quad (2.6)$$

$$[k\hat{A}, \hat{B}] = [\hat{A}, k\hat{B}] = k[\hat{A}, \hat{B}] \quad (2.7)$$

$$[\hat{A}, \hat{B} + \hat{C}] = [\hat{A}, \hat{B}] + [\hat{A}, \hat{C}], [\hat{A} + \hat{B}, \hat{C}] = [\hat{A}, \hat{C}] + [\hat{B}, \hat{C}] \quad (2.8)$$

$$[\hat{A}, \hat{B}\hat{C}] = [\hat{A}, \hat{B}]\hat{C} + \hat{B}[\hat{A}, \hat{C}], [\hat{A}\hat{B}, \hat{C}] = [\hat{A}, \hat{C}]\hat{B} + \hat{A}[\hat{B}, \hat{C}] \quad (2.9)$$

donde k es una constante y se considera que los operadores son lineales.

Funciones propias y valores propios de operadores

Suponiendo que el efecto que produce el operador \hat{A} al actuar sobre una función $g(x)$ es simplemente multiplicar la función $g(x)$ por cierta constante C . Decimos entonces que $g(x)$ es una función propia de \hat{A} con valor propio C . Exigimos en esta definición que la función propia $g(x)$ no sea idénticamente cero en todos sus puntos. Tenemos

$$\hat{A}g(x) = Cg(x) \quad (2.10)$$

Operadores momento angular

En mecánica cuántica hay dos tipos de momento angular, el momento angular orbital, que proviene del movimiento de la partícula a través del espacio, y que es el análogo de la magnitud mecano-clásica \vec{L} :

$$\vec{L} \equiv \vec{r} \times \vec{p} \quad (2.11)$$

$$L_x = yp_z - zp_y, L_y = zp_x - xp_z, L_z = xp_y - yp_x \quad (2.12)$$

donde \vec{p} es el momento lineal de una partícula ($\vec{p} = m\vec{v}$, m es la masa y \vec{v} es la velocidad) y \vec{r} es el vector de posición, y el momento angular de espín, que es una propiedad intrínseca de muchas partículas microscópicas y que no tiene análogo mecano-clásico. Se considerará en esta ocasión solamente el momento angular orbital. Los operadores mecano-cuánticos para las componentes del momento angular orbital de una partícula se construye reemplazando las coordenadas y los momentos en las ecuaciones clásicas 2.12 por sus correspondientes operadores. De este modo se obtiene:

$$\hat{L}_x = -i\hbar \left(y \frac{\partial}{\partial z} - z \frac{\partial}{\partial y} \right) \quad (2.13)$$

$$\hat{L}_y = -i\hbar \left(z \frac{\partial}{\partial x} - x \frac{\partial}{\partial z} \right) \quad (2.14)$$

$$\hat{L}_z = -i\hbar \left(x \frac{\partial}{\partial y} - y \frac{\partial}{\partial x} \right) \quad (2.15)$$

Utilizando

$$\hat{L} = \left| \hat{\vec{L}} \right| = \hat{\vec{L}} \cdot \hat{\vec{L}} = \hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2 \quad (2.16)$$

es posible construir el operador para el cuadrado del módulo del momento angular a partir de los operadores dados por las Ecuaciones 2.13 a 2.15.

Puesto que las relaciones de conmutación determinan a qué magnitudes físicas se pueden asignar simultáneamente valores definidos, se investigan esas relaciones para el momento angular. Haciendo actuar el operador \hat{L}_y sobre una función dada $f(x, y, z)$, obtenemos

$$\hat{L}_y f = -i\hbar \left(z \frac{\partial f}{\partial x} - x \frac{\partial f}{\partial z} \right) \quad (2.17)$$

y operando sobre esta ecuación con \hat{L}_x , queda

$$\hat{L}_x \hat{L}_y f = -\hbar \left(y \frac{\partial f}{\partial x} + yz \frac{\partial f}{\partial z \partial x} - yx \frac{\partial f}{\partial z} - z \frac{\partial f}{\partial y \partial x} + zx \frac{\partial f}{\partial y \partial z} \right) \quad (2.18)$$

De modo análogo,

$$\hat{L}_y \hat{L}_x f = -\hbar \left(zy \frac{\partial f}{\partial x \partial z} - z \frac{\partial f}{\partial x \partial y} - xy \frac{\partial f}{\partial z} + x \frac{\partial f}{\partial y} + xz \frac{\partial f}{\partial z \partial y} \right) \quad (2.19)$$

Restando la Ecuación 2.18 de la 2.19, se obtiene

$$\hat{L}_x \hat{L}_y f - \hat{L}_y \hat{L}_x f = -\hbar \left(y \frac{\partial f}{\partial x} - x \frac{\partial f}{\partial y} \right) \quad (2.20)$$

$$\left[\hat{L}_x, \hat{L}_y \right] = i\hbar \hat{L}_z \quad (2.21)$$

Se entiende por permutación cíclica de x , y y z el cambio de x por y , y por z , y z por x . Si se realiza una permutación cíclica en \hat{L}_x obtenemos \hat{L}_y , si se realiza en \hat{L}_y se obtiene \hat{L}_z y si se realiza en \hat{L}_z se obtienen \hat{L}_x . Así, realizando dos permutaciones cíclicas sucesivas en la Ecuación 2.21, se obtiene

$$\left[\hat{L}_y, \hat{L}_z \right] = i\hbar \hat{L}_x, \quad \left[\hat{L}_z, \hat{L}_x \right] = i\hbar \hat{L}_y \quad (2.22)$$

Evaluando los conmutadores de \hat{L} con cada uno de sus componentes usando las identidades de las Ecuaciones 2.5 a 2.9 se tiene:

$$\begin{aligned}
 \left[\hat{L}^2, \hat{L}_x \right] &= \left[\hat{L}_x^2 + \hat{L}_y^2 + \hat{L}_z^2, \hat{L}_x \right] \\
 &= \left[\hat{L}_x^2, \hat{L}_x \right] + \left[\hat{L}_y^2, \hat{L}_x \right] + \left[\hat{L}_z^2, \hat{L}_x \right] \\
 &= \left[\hat{L}_y^2, \hat{L}_x \right] + \left[\hat{L}_z^2, \hat{L}_x \right] \\
 &= \left[\hat{L}_y, \hat{L}_x \right] \hat{L}_y + \hat{L}_y \left[\hat{L}_y, \hat{L}_x \right] + \left[\hat{L}_z, \hat{L}_x \right] \hat{L}_z + \hat{L}_z \left[\hat{L}_z, \hat{L}_x \right] \\
 &= -i\hbar \hat{L}_z \hat{L}_y - i\hbar \hat{L}_y \hat{L}_z + i\hbar \hat{L}_y \hat{L}_z + i\hbar \hat{L}_z \hat{L}_y \\
 \left[\hat{L}, \hat{L}_x \right] &= 0
 \end{aligned} \tag{2.23}$$

Puesto que una permutación cíclica de x , y y z deja la expresión $\hat{L} = \hat{L}_x + \hat{L}_y + \hat{L}_z$ inalterada, si realizamos dos de dichas permutaciones en la Ecuación 2.23, se obtiene:

$$\left[\hat{L}, \hat{L}_y \right] = 0, \quad \left[\hat{L}, \hat{L}_z \right] = 0 \tag{2.24}$$

Puesto que \hat{L} conmuta con cada uno de sus componentes, podemos especificar valores definidos para L y para una cualquiera de sus componentes. Sin embargo, ninguna pareja de componentes de \hat{L} conmutan entre si, de modo que no podemos especificar más de una componente simultáneamente. Nótese que al fijar $L = \left| \vec{L} \right|$ no especificamos el vector \vec{L} si no que solamente su módulo. La especificación completa de \vec{L} implica fijar simultáneamente cada una de sus tres componentes, lo que normalmente no puede hacerse. En mecánica clásica, cuando el momento angular se conserva, cada una de sus tres componentes tiene valor definido. En mecánica cuántica, cuando el operador momento angular se conserva, sólo es posible especificar su magnitud y una de sus componentes.

Podríamos intentar determinar los valores propios y las funciones propias comunes de los operadores \hat{L} y \hat{L}_z usando las expresiones para estos operadores en coordenadas cartesianas. Sin embargo, encontraríamos que las ecuaciones en derivadas parciales que se obtienen no son separables. Por esto que se realiza una transformación a

coordenadas polares esféricas. La coordenada r es la distancia desde el origen al punto (x, y, z) ; el ángulo θ es el que forma el vector \vec{r} con la parte positiva del eje z ; y el ángulo ϕ es el que forma la proyección del vector \vec{r} en el plano xy con la parte positiva del eje x . Usando las correspondientes relaciones trigonométricas se obtiene:

$$x = r \sin \theta \cos \phi, y = r \sin \theta \sin \phi, z = r \cos \theta \quad (2.25)$$

$$r = \sqrt{x^2 + y^2 + z^2}, \cos \theta = \frac{z}{\sqrt{x^2 + y^2 + z^2}}, \tan \phi = \frac{y}{x} \quad (2.26)$$

Expresando la Ecuación 2.13 en coordenadas esféricas queda:

$$\hat{L}_x = i\hbar \left(\sin \phi \frac{\partial}{\partial \theta} + \cot \theta \cos \phi \frac{\partial}{\partial \phi} \right) \quad (2.27)$$

$$\hat{L}_y = -i\hbar \left(\cos \theta \frac{\partial}{\partial \theta} + \cot \theta \sin \phi \frac{\partial}{\partial \phi} \right) \quad (2.28)$$

$$\hat{L}_z = -i\hbar \frac{\partial}{\partial \phi} \quad (2.29)$$

Elevando al cuadrado cada uno de los operadores \hat{L}_x , \hat{L}_y y \hat{L}_z , y sumándolos, se construye el operador \hat{L} (Ecuación 2.16)⁶:

$$\hat{L} = -\hbar^2 \left(\frac{\partial}{\partial \theta} + \cot \theta \frac{\partial}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial}{\partial \phi} \right) \quad (2.30)$$

Problema de fuerzas centrales en una partícula

Considerar una partícula simple que se mueve sometida a una fuerza central. Una fuerza central es una fuerza que proviene de una función de energía potencial de simetría esférica; es decir, que depende únicamente de la distancia de la partícula al origen: $V = V(r)$. La relación entre la fuerza y la energía potencial es:

$$\vec{F} = -\vec{\nabla}V(x, y, z) = -\hat{i}x \left(\frac{\partial V}{\partial x} \right) - \hat{j} \left(\frac{\partial V}{\partial y} \right) - \hat{k} \left(\frac{\partial V}{\partial z} \right) \quad (2.31)$$

⁶Notar que aunque los operadores momento angular dependen de las tres coordenadas cartesianas x , y y z , sólo son función de las dos coordenadas esféricas θ y ϕ .

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

donde \hat{i} , \hat{j} y \hat{k} corresponden a los vectores unitarios a lo largo de los ejes x , y y z , respectivamente. Como V solamente es función de r , tenemos que $(\partial V/\partial\theta)_{r,\phi} = 0$, y $(\partial V/\partial\phi)_{r,\theta} = 0$. Así,

$$\left(\frac{\partial V}{\partial x}\right)_{y,z} = \frac{dV}{dr} \left(\frac{\partial r}{\partial x}\right)_{y,z} = \frac{x}{r} \frac{dV}{dr} \quad (2.32)$$

$$\left(\frac{\partial V}{\partial y}\right)_{x,z} = \frac{y}{r} \frac{dV}{dr}, \quad \left(\frac{\partial V}{\partial z}\right)_{x,y} = \frac{z}{r} \frac{dV}{dr} \quad (2.33)$$

La ecuación 2.31 queda de la siguiente forma:

$$\vec{F} = -\frac{1}{r} \frac{dV}{dr} (x\hat{i} + y\hat{j} + z\hat{k}) = -\frac{dV(r)}{dr} \frac{\vec{r}}{r} \quad (2.34)$$

La cantidad \vec{r}/r en 2.34 es el vector unitario en la dirección radial. Una fuerza central solo tiene componente radial. Para la mecánica cuántica de una partícula simple sometida a una fuerza central, el operador Hamiltoniano es:

$$\hat{H} = \hat{T} + \hat{V} - \left(\frac{\hbar^2}{2m}\right) \nabla^2 + V(r) \quad (2.35)$$

donde \hat{T} corresponde al operador para la energía cinética T

$$\hat{T} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$$

y ∇ está definido por la siguiente expresión:

$$\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Puesto que V tiene simetría esférica, utilizamos coordenadas polares esféricas (Ecuaciones 2.25 y 2.26). Se debe expresar por tanto el operador laplaciano en estas coordenadas. Así,

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{1}{r^2} \cotg\theta \frac{\partial}{\partial \theta} + \frac{1}{r^2 \text{sen}^2\theta} \frac{\partial^2}{\partial \phi^2} \quad (2.36)$$

Comparando la expresión anterior con la 2.30, se obtiene:

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} - \frac{1}{r^2 \hbar^2} \hat{L}^2 \quad (2.37)$$

y el Hamiltoniano como:

$$\hat{H} = -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) + \frac{1}{2mr^2} \hat{L}^2 + V(r) \quad (2.38)$$

En mecánica clásica, una partícula sometida a una fuerza central conserva su momento angular. En mecánica cuántica, podemos preguntarnos si es posible que haya estados con valores definidos simultáneamente para la energía y el momento angular. Para que el conjunto de funciones propias de \hat{H} lo sea también de \hat{L}^2 , el conmutador $[\hat{H}, \hat{L}^2]$ debe anularse. Tenemos

$$\begin{aligned} [\hat{H}, \hat{L}^2] &= [\hat{T}, \hat{L}^2] + [\hat{V}, \hat{L}^2] \\ [\hat{T}, \hat{L}^2] &= \left[-\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} \right) + \frac{1}{2mr^2} \hat{L}^2, \hat{L}^2 \right] \\ [\hat{T}, \hat{L}^2] &= -\frac{\hbar^2}{2m} \left[\frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r}, \hat{L}^2 \right] + \frac{1}{2m} \left[\frac{1}{r^2} \hat{L}^2, \hat{L}^2 \right] \end{aligned} \quad (2.39)$$

Como \hat{L}^2 depende solamente de θ y de ϕ , y no de r (Ecuación 2.30), conmuta con cualquier operador que sea sólo función de r . Así, el primer conmutador de la ecuación 2.39 es cero. Más aún, puesto que cualquier operador conmuta consigo mismo, el segundo conmutador de 2.39 también es cero. Por tanto, $[\hat{T}, \hat{L}^2] = 0$. Además, puesto que \hat{L}^2 no depende de r , y V es sólo función de r , tenemos $[\hat{V}, \hat{L}^2] = 0$. Por lo tanto,

$$[\hat{H}, \hat{L}^2] = 0 \text{ si } V = V(r) \quad (2.40)$$

el Hamiltoniano \hat{H} conmuta con \hat{L}^2 cuando la función de energía potencial es independiente de θ y de ϕ .

Teniendo en cuenta ahora que el operador $\hat{L}_z = -i\hbar\partial/\partial\phi$ (Ecuación 2.29), como \hat{L}_z no depende de r , y puesto que conmuta con \hat{L}^2 (Ecuación 2.24), vemos que \hat{L}_z conmuta con el Hamiltoniano 2.38:

$$\left[\hat{H}, \hat{L}_z\right] = 0 \text{ si } V = V(r) \quad (2.41)$$

Es posible, por tanto, tener un conjunto de funciones propias simultaneas de \hat{H} , \hat{L}^2 y \hat{L}_z para el problema de fuerzas centrales. Llamando ψ a estas funciones propias comunes, tenemos

$$\hat{H}\psi = E\psi \quad (2.42)$$

$$\hat{L}^2\psi = l(l+1)\hbar^2\psi, \quad l = 0, 1, 2, \dots \quad (2.43)$$

$$\hat{L}_z\psi = m\hbar\psi, \quad m = -l, -l+1, \dots, l \quad (2.44)$$

Las expresiones para las ecuaciones 2.42, 2.43 y 2.44, provienen del desarrollo mostrado y de las ecuaciones para las funciones propias y valores propios del momento angular orbital de una partícula:

$$\hat{L}^2 Y_l^m(\theta, \phi) = l(l+1)\hbar^2 Y_l^m(\theta, \phi), \quad l = 0, 1, 2, \dots \quad (2.45)$$

$$\hat{L}_z Y_l^m(\theta, \phi) = m\hbar Y_l^m(\theta, \phi), \quad m = -l, -l+1, \dots, l-1, l \quad (2.46)$$

en donde las funciones propias vienen dadas por la siguiente ecuación:

$$Y_l^m(\theta, \phi) = \left[\frac{2l+1}{4\pi} \frac{(l-|m|)!}{(l+|m|)!} \right]^{\frac{1}{2}} P_l^{|m|}(\cos\theta) e^{im\theta} \quad (2.47)$$

$P_l^{|m|}$ corresponden a las funciones asociadas de Legendre definidas del siguiente modo:

$$P_l^{|m|}(w) \equiv \frac{1}{2^l l!} (1-w^2)^{\frac{|m|}{2}} \frac{d^{l+|m|}}{dw^{l+|m|}} (w^2-1)^l, \quad l = 0, 1, 2, \dots \quad (2.48)$$

donde $w = \cos\theta$.

Usando las ecuaciones 2.38 y 2.43 se obtiene para la ecuación de Schrödinger 2.42:

$$\begin{aligned}
 & -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \frac{1}{2mr^2} \hat{L}^2 \psi + V(r) \psi = E \psi \\
 & -\frac{\hbar^2}{2m} \left(\frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} \right) + \frac{l(l+1)\hbar^2}{2mr^2} \psi + V(r) \psi = E \psi
 \end{aligned} \tag{2.49}$$

Las funciones propias de \hat{L}^2 son armónicos esféricos $Y_l^m(\theta, \phi)$, y, puesto que \hat{L}^2 es independiente de r , se puede multiplicar $Y_l^m(\theta, \phi)$ por una función arbitraria de r y seguiremos teniendo una función propia de \hat{L}^2 y de \hat{L}_z . Por tanto,

$$\psi = R(r) Y_l^m(\theta, \phi) \tag{2.50}$$

Haciendo uso de la Ecuación 2.50 en 2.49 y dividiendo por Y_l^m , obtenemos una ecuación diferencial ordinaria para la función incógnita $R(r)$:

$$-\frac{\hbar^2}{2m} \left(R'' + \frac{2}{r} R' \right) + \frac{l(l+1)\hbar^2}{2mr^2} R + V(r) R = ER(r) \tag{2.51}$$

Así, se ha demostrado que para cualquier problema de una partícula con una función de energía potencial de simetría esférica $V(r)$, las funciones de onda estacionarias son $\psi = R(r) Y_l^m(\theta, \phi)$, donde el factor radial $R(r)$ satisface la Ecuación 2.51. Empleando una forma específica para $V(r)$ en la ecuación 2.51, es posible resolverla para el correspondiente problema particular.

Movimiento Nuclear en Moléculas Diatómicas

La ecuación de Schrödinger para el movimiento nuclear en un estado electrónico ligado de una molécula diatómica es

$$\left[-\frac{\hbar^2}{2m_\alpha} \nabla_\alpha^2 - \frac{\hbar^2}{2m_\beta} \nabla_\beta^2 + U(R) \right] \psi_N = E \psi_N \tag{2.52}$$

donde α y β son los núcleos, y la función de onda para el movimiento nuclear ψ_N es función de la coordenada nuclear $x_\alpha, y_\alpha, z_\alpha, x_\beta, y_\beta, z_\beta$. La energía potencial $U(R)$ es

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

función, solamente, de las coordenadas relativas de los dos núcleos, y la ecuación de Schrödinger de dos partículas (2.52) se puede reducir a dos ecuaciones de Schrödinger de una partícula separadas, una para la energía traslacional de la molécula como un todo y otra para el movimiento relativo interno de los núcleos, cada uno con respecto al otro. Tenemos

$$\psi_N = \psi_{N,tr}\psi_{N,int} \text{ y } E = E_{tr} + E_{int} \quad (2.53)$$

La ecuación de Schrödinger para $\psi_{N,int}$ es

$$\left[-\frac{\hbar^2}{2\mu}\nabla^2 + U(R) \right] \psi_{N,int} = E_{int}\psi_{N,int}, \text{ donde } \mu \equiv m_\alpha m_\beta (m_\alpha + m_\beta) \quad (2.54)$$

donde $\psi_{N,int}$ es función de las coordenadas relativas de un núcleo con respecto a otro. Las mejores coordenadas que se emplean son las coordenadas esféricas de un núcleo relativas al otro. El radio r , en coordenadas esféricas relativas es la distancia internuclear R , y denotamos las coordenadas angulares mediante θ_N y ϕ_N . La energía potencial en la ecuación 2.54 depende solamente de R , por lo que se trata de un problema de fuerzas centrales (Sección 2.5.2),

$$\psi_{N,int} = P(R) Y_J^M(\theta_N, \phi_N), \quad J = 0, 1, 2, \dots, \quad M = -J, \dots, J \quad (2.55)$$

donde las funciones Y_J^M son las funciones armónicas esféricas para los números cuánticos J y M .

A partir de la Ecuación 2.51, la función radial $P(R)$ se obtiene resolviendo

$$-\frac{\hbar^2}{2\mu} \left[P''(R) + \frac{2}{R}P'(R) \right] + \frac{J(J+1)\hbar^2}{2\mu R^2} P(R) + U(R)P(R) = E_{int}P(R) \quad (2.56)$$

Definiendo

$$F(R) \equiv RP(R) \quad (2.57)$$

y sustituyendo $P = F/R$ en la Ecuación 2.56, se obtiene:

$$-\frac{\hbar^2}{2\mu}F''(R) + \left[U(R) + \frac{J(J+1)\hbar^2}{2\mu R^2} \right] F(R) = E_{int}F(R) \quad (2.58)$$

que se puede considerar como una ecuación de Schrödinger unidimensional con energía potencial efectiva $U(R) + J(J+1)\hbar^2/(2\mu R^2)$.

La resolución más simple de la Ecuación 2.58 consiste en desarrollar $U(R)$ por medio de una serie de Taylor en torno a R_e :

$$U(R) = U(R_e) + U'(R_e)(R - R_e) + \frac{1}{2}U''(R_e)(R - R_e)^2 + \frac{1}{6}U'''(R_e)(R - R_e)^3 + \dots \quad (2.59)$$

En la separación internuclear de equilibrio R_e , la pendiente de la curva $U(R)$ es cero, de forma que $U'(R_e) = 0$. Para R próximas a R_e , $(R - R_e)^3$, y las potencias superiores serán pequeñas, con lo que podemos despreciar estos términos. Definiendo la constante de fuerza de equilibrio k_e como $k_e \equiv U''(R_e)$, tenemos:

$$U(R) \approx U(R_e) + \frac{1}{2}k_e(R - R_e)^2 = U(R_e) + \frac{1}{2}k_e x^2 \quad (2.60)$$

$$k_e \equiv U''(R_e) \text{ y } x \equiv R - R_e$$

De esta manera la Ecuación 2.58 se transforma en:

$$-\frac{\hbar^2}{2\mu}S''(x) + \left[U(R_e) + \frac{1}{2}k_e x^2 + \frac{J(J+1)\hbar^2}{2\mu(x + R_e)^2} \right] S(x) \approx E_{int}S(x) \quad (2.61)$$

$$\text{donde } S(x) \equiv F(R) \quad (2.62)$$

Desarrollando $1/(x + R_e)^2$ mediante una serie de Taylor se obtiene que

$$1/(x + R_e)^2 \approx 1/R_e^2 \quad (2.63)$$

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

La sustitución de la expresión anterior en 2.61 y reagrupando términos da

$$-\frac{\hbar^2}{2\mu}S''(x) + \frac{1}{2}k_e x^2 S(x) \approx \left[E_{int} - U(R_e) - \frac{J(J+1)\hbar^2}{2\mu(R_e)^2} \right] S(x) \quad (2.64)$$

La Ecuaciones 2.64 es la misma que la ecuación de Schrödinger para un oscilador armónico unidimensional con coordenadas x , masa μ , energía potencial $\frac{1}{2}k_e x^2$, y autovalores de energía $E_{int} - U(R_e) - J(J+1)\hbar^2/2\mu R_e^2$. Por lo tanto, es posible igualar el término entre corchetes en 2.64 a los valores propios del oscilador armónico, y se tiene:

$$E_{int} - U(R_e) - \frac{J(J+1)\hbar^2}{2\mu R_e^2} \approx \left(v + \frac{1}{2} \right) h\nu_e$$

$$E_{int} \approx U(R_e) + \left(v + \frac{1}{2} \right) h\nu_e + \frac{J(J+1)\hbar^2}{2\mu R_e^2} \quad (2.65)$$

$$\nu_e = \frac{1}{2\pi} \left(\frac{k_e}{\mu} \right)^{1/2}, \quad v = 0, 1, 2, \dots \quad (2.66)$$

donde ν_e es la frecuencia vibracional de equilibrio. La energía molecular interna E_{int} es aproximadamente la suma de la energía electrónica $U(R_e) \equiv E_{elec}$ (que difiere para los diferentes estados electrónicos de la misma molécula), la energía vibracional $(v + \frac{1}{2}h\nu_e)$, y la energía rotacional $J(J+1)\hbar^2/2\mu R_e^2$. Las aproximaciones 2.60 y 2.63 corresponden al oscilador armónico y al rotor rígido. De 2.65 y 2.53, la energía molecular, $E = E_{tr} + E_{int}$, es aproximadamente la suma de las energías traslacional, rotacional, vibracional y electrónica:

$$E \approx E_{tr} + E_{rot} + E_{vib} + E_{elec} \quad (2.67)$$

A partir de 2.53, 2.55, 2.57 y 2.62, la función de onda para el movimiento nuclear es

$$\psi_N \approx \psi_{N,tr} S_v(R - R_e) R^{-1} Y_J^M(\theta_n, \phi_N) \quad (2.68)$$

donde $S_v(R - R_e)$ son las funciones propias del oscilador armónico con número cuántico v .

Mecánica Molecular

En mecánica clásica el movimiento de una partícula está gobernado por la segunda ley de Newton:

$$F = m \cdot a = m \cdot \frac{d^2x}{dt^2} \quad (2.69)$$

donde F es la fuerza que actúa sobre la partícula, m es su masa, t es el tiempo y a es la aceleración, que viene dada por $a = dv/dt = (d/dt)(dx/dt) = d^2x/dt^2$, donde v es la velocidad. La Ecuación 2.69 contiene la segunda derivada de la coordenada x con respecto al tiempo. Para resolverla hemos de realizar dos integraciones, lo que introduce dos constantes arbitrarias c_1 y c_2 en la solución, de modo que

$$x = g(t, c_1, c_2) \quad (2.70)$$

donde g es alguna función del tiempo. Preguntémonos ahora: ¿qué información debemos poseer en un tiempo dado t_0 para poder predecir el movimiento futuro de la partícula. Si sabemos que en el instante t_0 la partícula está en el punto x_0 , tenemos que

$$x_0 = g(t_0, c_1, c_2) \quad (2.71)$$

Puesto que hemos de determinar dos constantes necesitamos sin embargo, más información. Derivando la Ecuación 2.70 obtenemos

$$\frac{dx}{dt} = v = \frac{d}{dt}g(t, c_1, c_2) \quad (2.72)$$

de modo que, si sabemos también que la partícula tiene una velocidad v_0 en el instante de tiempo t_0 , entonces disponemos de la relación adicional

$$v_0 = \left. \frac{d}{dt}g(t, c_1, c_2) \right|_{t=t_0} \quad (2.73)$$

Podemos usar pues las Ecuaciones 2.71 y 2.73 para determinar c_1 y c_2 en función de x_0 y de v_0 . Conociendo c_1 y c_2 , podemos emplear la Ecuación 2.70 para predecir exactamente el movimiento futuro de la partícula.

$$\frac{\partial V(x, t)}{\partial x} = -F(x, t) \quad (2.74)$$

De acuerdo con la segunda ley de Newton, dado el estado de un sistema en cualquier instante de tiempo, su estado y movimientos futuros quedan completamente determinados, como muestran las Ecuaciones 2.70 a 2.73.

Conociendo de forma exacta el estado presente de un sistema mecano-clásico, podemos predecir su estado futuro. Sin embargo, el principio de incertidumbre de Heisenberg pone de manifiesto que no podemos determinar simultáneamente la posición y la velocidad exactas de una partícula microscópica, de modo que no podemos disponer de la información que requiere la mecánica clásica para predecir el movimiento futuro del sistema. En mecánica cuántica debemos contentarnos con algo menos que la predicción completa del movimiento futuro exacto del sistema.

Nuestra aproximación a la mecánica cuántica va a consistir en postular los principios básicos y luego usar esos postulados para deducir consecuencias que puedan probarse experimentalmente, como los niveles de energía de los átomos. Para describir el *estado* de un sistema, lo hacemos mediante una *función de onda* o *función de estado* Ψ . Puesto que el estado cambia, en general, con el tiempo, Ψ es también función del tiempo, $\Psi = \Psi(x, t)$. La función de onda contiene toda la información que es posible conocer acerca del sistema, de manera que en lugar de hablar de “estado descrito por la función de onda Ψ ”, simplemente se hablará de “estado Ψ ”. La segunda ley de Newton nos dice cómo encontrar el estado futuro de un sistema mecano-clásico conociendo el estado presente. Para encontrar el estado futuro de un sistema mecano-

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

cuántico conociendo el estado presente necesitamos una ecuación que nos diga cómo cambia la función de onda con el tiempo. Para un sistema unidimensional de una sola partícula se postula que esta ecuación es

$$-\frac{\hbar}{i} \frac{\partial \Psi(x, t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x, t)}{\partial x^2} + V(x, t) \Psi(x, t) \quad (2.75)$$

donde la constante \hbar se define como

$$\hbar \equiv \frac{h}{2\pi} \quad (2.76)$$

Esta ecuación es conocida como *Ecuación de Schrödinger dependiente del tiempo*, $i = \sqrt{-1}$, m es la masa de la partícula, y $V(x, t)$ es la función de energía potencial del sistema.

La ecuación de Schrödinger dependiente del tiempo contiene la primera derivada de la función de onda con respecto al tiempo y permite calcular la función de onda futura en cualquier tiempo, si conocemos la función de onda en el instante de tiempo t_0 .

La función de onda contiene toda la información que es posible conocer sobre el sistema. No podemos esperar que Ψ proporcione una especificación concreta de la posición, como hace el estado mecano-clásico del sistema. La respuesta correcta a esta pregunta la dio Max Born poco después de que Schrödinger descubriese su ecuación. Born postuló que la cantidad

$$|\Psi(x, t)|^2 dx \quad (2.77)$$

da la probabilidad de encontrar a la partícula en el tiempo t en la región del eje x comprendida entre x y $x + dx$. En la Ecuación 2.77, las barras indican valor absoluto y dx es una longitud infinitesimal sobre el eje x . La función $|\Psi(x, t)|^2$ es la **densidad**

de probabilidad de encontrar a la partícula en cualquier lugar infinitesimal del eje x .

La mecánica cuántica proporciona las leyes de movimiento de las partículas macroscópicas. La experiencia muestra que los objetos macroscópicos obedecen a la mecánica clásica. Por tanto, para que la mecánica cuántica sea una teoría válida, debe reducirse a la mecánica clásica conforme pasemos de partículas microscópicas a macroscópicas. Los efectos cuánticos van asociados a la longitud de onda de de Broglie $\lambda = h/mv$. Puesto que h es muy pequeña, a la longitud de onda de de Broglie para los objetos macroscópicos es prácticamente cero. Así, cabe esperar que en el límite $\lambda \rightarrow 0$, la ecuación de Schrödinger dependiente del tiempo se reduzca a la segunda ley de Newton.

Para el tratamiento de moléculas existen métodos semi-empíricos y de mecánica molecular (MM). El método de mecánica molecular es muy diferente a los métodos semi empíricos ya que no es un método mecanocuántico, puesto que no trata con un Hamiltoniano, con una función de onda o con densidad electrónica. En lugar de ello usa un modelo de una molécula compuesta por átomos que se mantienen unidos por enlaces. Usando parámetros tales como constantes de fuerza de tensión de enlace y de flexión de enlace, y permitiendo interacciones entre los átomos no enlazados, el método construye una expresión de la energía potencial que es unión de las posiciones atómicas. Minimizando esta expresión para varios conformeros moleculares, el método MM predice geometrías de equilibrio y energías relativas.⁷ Este método es aplicable a estados electrónicos fundamentales y en su mayoría a compuestos orgánicos. Sin embargo, están creciendo las aplicaciones a compuestos organometálicos y compuestos de coordinación de metales de transición. Debido a que los cálculos MM

⁷El método fue desarrollado por Westheimer, Hendrickson, Wiberg, Allinger, Warshel y otros

son mucho más rápidos que los cálculos mecanocuántico, se pueden tratar sistemas con miles de átomos.

La mecánica molecular se ocupa de los cambios en la energía electrónica de la molécula debido al estiramiento del enlace (V_{ten}), flexión del ángulo de enlace (V_{flex}), flexión fuera del plan (V_{ffp}), rotación interna (torsión) en torno a los enlaces (V_{tor}), interacciones entre estos movimientos (que producen un término cruzado, V_{cruz}), atracciones y repulsiones de van der Waals entre átomos no ligados (V_{vdW}) e interacciones electrostáticas entre átomos (V_{el}). La suma de estas contribuciones da la energía potencial de la mecánica molecular V (llamada energía estérica) para el movimiento de los átomos en la molécula o moléculas:

$$V = V_{ten} + V_{flex} + V_{ffp} + V_{tor} + V_{cruz} + V_{vdW} + V_{el} \quad (2.78)$$

Las expresiones explícitas usadas para cada uno de los términos en (2.78) definen a qué se llama campo de fuerza de la mecánica molecular, ya que las derivadas de la función de energía potencial determinan las fuerzas sobre los átomos. Un campo de fuerzas contiene fórmulas analíticas para los términos de (2.78), y los valores para todos los parámetros que intervienen en esas fórmulas. Los campos de fuerza empíricos no solo se emplean para cálculos de mecánica molecular si no también para dinámica molecular de líquidos y disoluciones, donde se integra la segunda ley de Newton para seguir los movimientos de los átomos con el tiempo, en sistemas que contienen cientos de moléculas.

Un campo de fuerza MM asigna a cada átomo de una molécula un número de posibles tipos de átomo, dependiendo del número atómico de átomo y del entorno molecular. Por ejemplo, algunos tipos de átomos comúnmente usados en campos de fuerza para compuestos orgánicos son carbono sp^3 (saturado), carbono sp^2 (do-

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

blemente enlazado), carbono *sp* (triplemente enlazado), carbono carbonilo, carbono aromático, y así sucesivamente, H enlazado a C, H enlazado a O, H enlazado a N, y así sucesivamente. Los diferentes campos de fuerza contienen, a veces, diferente número de clases de tipos de átomos, basados en las decisiones tomadas por sus constructores. Un campo de fuerza para compuestos orgánicos contiene de 50 a 75 tipos de átomos.

En los cálculos electrónicos *ab initio* y semi empíricos de punto simple o de optimización de geometría, se introducen los números atómicos de los átomos y un conjunto de coordenadas (cartesianas o internas) para cada átomo, y no se especifica qué átomos están enlazados a qué átomos. En un cálculo de MM, se debe especificar no solamente las coordenadas atómicas iniciales, sino también qué átomos están ligados a qué átomos, de forma de que se pueda construir apropiadamente la expresión de V .

Para una posterior descripción de los diferentes términos matemáticos que conforman en potencial se establece la siguiente terminología: los átomos 1, 2 son átomos ligados entre sí; los átomos 1, 3 están separados por dos enlaces; los átomos 1, 4 están separados por tres enlaces; y así sucesivamente, como se puede ver en la figura 2.4.

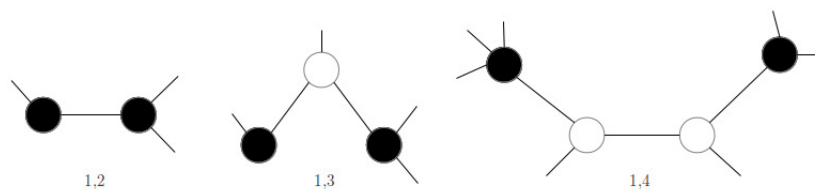


Figura 2.4: Representación gráfica del grado de separación entre átomos enlazados en una molécula. Los átomos sombreados son los átomos 1, 2, 1, 3 y 1, 4.

Algunos campos de fuerzas y usos se listan en la Tabla 2.1.

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

Tabla 2.1: Campos de fuerza.

Campo de fuerza	Uso	Autor
MM2	Compuestos orgánicos de tamaño pequeño y medio	Burkert y Allinger
MM3	Compuestos orgánicos pequeños, polipéptidos y proteínas	N.L. Alinger y L. Yan, <i>J. Am. Chem. Soc.</i> , 115, 11918 (1993) y referencias citadas en él.
MM4	Hidrocarburos	N.L Alinger et al., <i>J. Comput. Chem.</i> , 17, 642, 747 (1996); N. Nevind et al., <i>J. Comput. Chem.</i> , 17, 669, 695, 730.
AMBER	polipéptidos, proteínas y ácidos nucleicos	S.J. Weiner et al., <i>J. Comput. Chem.</i> , 7, 230 (1986); S.J. Weiner et al., <i>J. Am. Chem. Soc.</i> , 106, 765 (1984); versión AMBER94- W.D. Ornell et al., <i>J. Am. Chem. Soc.</i> , 117, 5179 (1995); www.amber.ucsf.edu
CHARMM	polipéptidos, proteínas y ácidos nucleicos	B.R. Brooks et al. <i>J. Comput. Chem.</i> 4, 187 (1983); A.D. Mackerell et al., <i>J. Am. Chem. Soc.</i> , 117, 11946 (1995); <i>J. Phys. Chem.</i> , B, 102, 3596 (1998); yuri.harvard.edu
OPLS	polipéptidos, proteínas y ácidos nucleicos	W.L. Jorgensen y J. Tirado-Rives, <i>J. Am. Chem. Soc.</i> , 110, 1657 (1988); J. Pranata et al., <i>J. Am. Chem. Soc.</i> , 113, 2810 (1991); versión OPLS-AA: W.L Jorgensen et al., <i>J. Am. Chem. Soc.</i> , 118, 11225 (1996)
GROMOS	polipéptidos, proteínas y ácidos nucleicos	X. Daura, A.E. Mark y W.F. van Gunsteren, <i>J. Comput. Chem.</i> , 19, 535 (1998)
CHARMm	versión comercial de CHARMM aplicable también a compuestos orgánicos más pequeños	F.A. Momany y R. Rone, <i>J. Comput. Chem.</i> , 13, 888 (1992)
MMFF94	versión 1994 del campo de Merck: compuestos orgánicos pequeños, proteínas y ácidos nucleicos	T.A. Halgren, <i>J. Comput. Chem.</i> , 17, 490, 520, 553, 616 (1996)
CFF93 y CFF95	versiones 1993-1995: compuestos orgánicos pequeños, proteínas y ácidos nucleicos	M.J. Hwang, T.P. Stockfish y A.T. Hagler, <i>J. Am. Chem. Soc.</i> , 116, 2515 (1994); J.R. Maple et al., <i>J. Comput. Chem.</i> , 19, 430 (1998)
DREIDING	Campo de Fuerza general aplicable a compuestos orgánicos e inorgánicos de todos los elementos de los grupos principales.	S.L. Mayo, B.D. Olafson y W.A. Goddard, <i>J. Phys. Chem.</i> , 94, 8897 (1990)
UFF	compuestos de elementos de la tabla periódica	A.K. Rappé et al., <i>J. Am. Chem. Soc.</i> , 114 10024 (1992)
Tripes	moléculas orgánicas pequeñas, diseño de drogas	M. Clark et al., <i>J. Comput. Chem.</i> , 10, 1989

Se debe distinguir entre un campo de fuerzas, que está definido por la expresión para V en (2.78), y un programa de mecánica molecular, que es un programa computacional que usa un campo de fuerzas para llevar a cabo cálculos de mecánica molecular. Consideremos ahora los diferentes términos de (2.78).

Tensión.

La energía potencial V_{ten} de la tensión del enlace se toma como la suma de las energías potenciales $V_{ten,ij}$ para la tensión de cada enlace de la molécula: $V_{ten} = \sum_{1,2} V_{ten,ij}$, donde la suma se extiende a todos los pares de átomos enlazados entre sí (esto es, sobre todos los pares de átomos 1,2). La elección más simple para $V_{ten,ij}$ es usar la aproximación del oscilador armónico y tomar $V_{ten,ij}$ como una función cuadrática del desplazamiento de la longitud de enlace, l_{ij} , con respecto a su longitud de referencia (o natural), l_{IJ}^0 , esto es,

$$V_{ten,ij} = \frac{1}{2}k_{IJ} (l_{ij} - l_{IJ}^0)^2 \quad (2.79)$$

Aquí, las letras mayúsculas I y J denotan los tipos de átomos para los átomos i y j de la molécula. La constante de fuerza k_{IJ} y la longitud de referencia l_{IJ}^0 , dependen de los tipos de átomos de los dos átomos que forman en enlace.

Nótese que la expresión 2.79 para $V_{ten,ij}$ es similar a la expresión del potencial U que se desprende de la ecuación de Schrödinger para el movimiento nuclear en un estado electrónico ligado de una molécula diatómica (Ecuación 2.60. Aquí las letras mayúsculas I y J denotan los tipos de átomos de los átomos i y j de la molécula. La constante de fuerza f_{IJ} y la longitud de referencia l_{IJ}^0 , dependen de los tipos de átomos de los dos átomos que forman el enlace. Notar que, debido a que muchos términos en la energía estérica 2.78 implican que los átomos forman un enlace particular, el valor referencial l_{IJ}^0 no es necesariamente igual a la longitud de

enlace de equilibrio del enlace entre los átomos i y j en cualquier molécula particular. Sin embargo, estará muy próximo a las longitudes de equilibrio de la mayor parte de los enlaces formados por átomos de tipo I y J .

Flexión

La energía potencial V_{flex} de flexión del enlace se toma como la suma de las energías potenciales, $V_{flex,ijk}$, para la flexión de cada ángulo de enlace de la molécula: $V_{flex} = \sum V_{flex,ijk}$, donde la suma se extiende a todos los ángulos de enlace de la molécula. La opción más simple es tomar $V_{flex,ijk}$ como una función cuadrática:

$$V_{flex,ijk} = \frac{1}{2} k_{IJK} (\theta_{ijk} - \theta_{IJK}^0)^2 \quad (2.80)$$

donde θ_{IJK}^0 es el valor de referencia para el ángulo de enlace tipo IJK .

Torsión

El término V_{tor} se toma como suma de términos $V_{tor,ijkl}$ para todos los pares de átomos 1, 4: $V_{tor} = \sum_{1,4} V_{tor,ijkl}$. Por ejemplo, para el etano, H_3CCH_3 , cada hidrógeno de la izquierda del carbono tiene una relación 1,4 con cada uno de los otros tres hidrógenos de la derecha del carbono, dando un total de nueve términos en la suma V_{tor} . Para $H_2C=CH_2$, hay cuatro términos de torsión en la suma. Una forma comúnmente usada (MMFF94, CFF93, MM2, MM3, ver tabla 2.1) para $V_{tor,ijkl}$ es

$$V_{tor,ijkl} = \frac{1}{2} [V_1 (1 + \cos \phi) + V_2 (1 - \cos 2\phi) + V_3 (1 + \cos 3\phi)] \quad (2.81)$$

donde ϕ es el ángulo diedro $D(ijkl)$ y V_1 , V_2 y V_3 son parámetros cuyos valores dependen de los tipos de átomos de i , j , k y l . Tripos, CHARMMm y DREIDING usan las siguientes formas más simples, que contienen solamente un parámetro: $\frac{1}{2} V_n [1 + \cos (n\phi - \phi_0)]$, donde n da el número de mínimos contenidos en 360° del

potencial de torsión, y ϕ_0 determina las posiciones de los mínimos. Por ejemplo, para $n = 3$ y ϕ_0 , tenemos un potencial con mínimos a 60° , 180° y 300° , como ocurre en el C_2H_6 .

El cambio de energía potencial cuando un grupo metilo se rota en relación al otro en el etano, se puede determinar experimentalmente por observación de transiciones entre los niveles de energía vibracional producidas por este potencial. En un campo de fuerzas MM, V_{tor} del etano la suma de las contribuciones $V_{tor,ijkl}$ de nueve ángulos diedros, y un $V_{tor,ijkl}$ no es determinable experimentalmente. Tomando las segundas derivadas de la energía electrónica del etano apropiadas e integrando el resultado, se pueden usar los cálculos *ab initio* para estimar $V_{tor,ijkl}$.

Flexión fuera del plano

La molécula de ciclohexanona, $C_6H_{10}O$, contiene un oxígeno doblemente enlazado a uno de los carbonos de un anillo de cuatro miembros. El carbono del carbonilo y los tres átomos enlazados a él están ligados todos en el mismo plano. El ángulo de enlace CCC del grupo carbonilo está próximo a los 90° , y los ángulos de enlace CCO están próximos a $\frac{1}{2}(360^\circ - 90^\circ) = 135^\circ$. Debido a que esos ángulos de 135° se desvían considerablemente de los ángulos de referencia de aproximadamente 120° en el carbono carbonilo, los términos de flexión de enlace $V_{flex,ijk}$ en el carbono del carbonilo harán que en la posición de equilibrio del átomo de oxígeno esté ligado por encima del plano de los carbonos del anillo, de forma que se producirán ángulos CCO de casi 120° . Para asegurar la planaridad de los tres átomos ligados al carbono del carbonilo, (lo cual está favorecido por los enlaces π entre C y O), se introduce un término de flexión fuera del plano en cada carbono del carbonilo. Este término puede tener la forma $\frac{1}{2}k_{ffp}\chi_{ffp}^2$ donde χ_{ffp} es el ángulo entre el enlace CO y el plano del

carbono del carbonilo y los dos carbonos enlazados a él. A menudo se usa un término de flexión fuera del plano para un carbono doblemente enlazado a otro carbono, ya que esto permite que las frecuencias vibracionales se reproduzcan mejor.

En lugar de términos de flexión fuera del plano, algunos campos de fuerzas usan los impropriamente llamados términos de torsión, para llegar al mismo resultado.

Términos cruzados

Los términos cruzados V_{cruz} en 2.78 permiten interacciones entre los movimientos de tensión, flexión y torsión. Por ejemplo, si los enlaces C-O y O-H de un ángulo de enlace COH se estiran, aumenta la distancia entre los átomos finales del ángulo de enlace COH, lo cual hace más fácil doblar el ángulo COH. Del mismo modo, reduciendo el ángulo COH tienden a aumentar las longitudes O-H y C-H. Para permitir esta interacción, se puede añadir un término cruzado tensión-flexión con la forma $\frac{1}{2}k_{12}(\Delta l_1 + \Delta l_2)\Delta\theta$, donde Δl_1 , Δl_2 y $\Delta\theta$ son las desviaciones de las longitudes de enlace y el ángulo de enlace con respecto a los valores de referencia.

Los términos cruzados usados más comúnmente son: tensión-flexión y tensión-tensión para dos enlaces al mismo átomo, y tensión-torsión, flexión-torsión y flexión-flexión para dos ángulos con un átomo central común. MM2 y MMFF94 solamente incluyen interacciones tensión-flexión. Los campos de fuerzas TRIPOS, AMBER, CHARMM, DREIDING y UFF no tienen términos cruzados. MM3 y MM4 incluyen términos tensión-flexión, flexión-flexión y tensión-torsión.

Interacción electrostática

El término electrostático V_{el} se toma comúnmente como la suma de interacciones electrostáticas incluyendo todos los pares de átomos excepto los pares 1,2 y 1,3: $V_{el} = \sum_{1, \geq 4} V_{el,ij}$, donde los átomos i y j tienen una relación 1,4 o mayor. $V_{el,ij}$ se

calcula asignando cargas atómicas parciales Q_i a cada átomo, y usando la expresión de la energía potencial coulombicas

$$V_{el,ij} = \frac{Q_i Q_j}{\epsilon_r R_{ij}} \quad (2.82)$$

donde ϵ_r es una constante dieléctrica. Para cálculos que modelan moléculas en fase gaseosa, ϵ_r tiene un valor típicamente incluido en el intervalo de 1 a 1.5. Una ϵ_r mayor a 1 permite tener en cuenta interacciones de apantallamiento debido a polarización de partes de una molécula entre los átomos i y j .

Se usan varios métodos para asignar las cargas atómicas parciales MMFF94 supone que la polaridad de un tipo de enlace dado sea independiente de su entorno, y asigna cargas de acuerdo con $Q_i = Q_{i,formal} + \sum \omega_{KI}$, donde $Q_{i,formal}$ es la carga formal sobre el átomo i (obtenido a partir de la estructura de punto de Lewis, dividiendo los electrones en cada enlace por igual entre los átomos enlazados), ω_{KI} es un parámetro que representa la contribución de la carga al átomo i procedente del enlace entre los átomos i y k (de los tipos I y K), y la suma se extiende a todos los enlaces del átomo i . Nótese que $\omega_{KI} = -\omega_{KI}$.

Interacciones de van der Waals

El término de van der Waals en 2.78 se toma usualmente como la suma de interacciones que incluyen todas las interacciones posibles de pares de átomos 1, 4, 1, 5, 1, 6, ..., $V_{vdW} = \sum_{1, \geq 4} V_{vdW,ij}$, donde los átomos i y j están en relación 1, 4 o mayor. Las interacciones de van der Waals 1, 2 y 1, 3 y las interacciones electrostáticas se consideran incluidas implícitamente en los parámetros de la tensión de enlace u de flexión de enlace. Cada término de van der Waals de un par $V_{vdW,ij}$ es la suma de una atracción debida a fuerzas de dispersión de London y un término de repulsión debido, principalmente, a la repulsión de Pauli. Los campos de fuer-

za de AMBER, CHARMM, DREIDING, UFF y TRIPOS, toman $V_{vdW,ij}$ como un potencial de Lennard-Jones, que se puede escribir de dos formas equivalentes:

$$V_{vdW,ij} = \varepsilon_{IJ} \left[\left(\frac{R^*_{IJ}}{R_{ij}} \right)^{12} - 2 \left(\frac{R^*_{IJ}}{R_{ij}} \right)^6 \right] = 4\varepsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{R_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{IJ}}{R_{ij}} \right)^6 \right] \quad (2.83)$$

donde R_{ij} es la distancia entre los átomos i y j , el parámetro de profundidad del pozo ε_{IJ} es el valor de $V_{vdW,ij}$ en el mínimo de la curva de interacción, el parámetro R^*_{IJ} da el valor de R_{ij} en el mínimo de $V_{vdW,ij}$, y el parámetro alternativo σ_{IJ} es la distancia a la que $V_{vdW,ij}$ es cero.

Para las interacciones 1, 4, algunos campos de fuerzas multiplican $V_{vdW,ij}$ por un factor de escala que reduce su valor.

Cortes o Cutoff

Las interacciones electrostáticas y de van der Waals se denominan interacciones no enlazantes, y consumen la mayor parte del tiempo necesario para calcular $V_{estérico}$ de una molécula muy grande. Para una molécula de 3000 átomos, V_{el} y V_{vdW} son cada uno de ellos la suma de entorno a $\frac{1}{2}(3000)(2990) \approx 4 \times 10^6$ términos. (el último factor se reduce admitiendo que se omitan las interacciones 1, 2 y 1, 3.) Para acelerar los cálculos MM para moléculas grandes (y los cálculos de dinámica molecular en sistemas que contienen muchas moléculas), muchos programas usan un corte, significando que los términos $V_{el,ij}$ y $V_{vdW,ij}$ se omiten para parejas de átomos que están más separados que una distancia dada.

Las interacciones de van der Waals son proporcionales a $1/R_{ij}^6$ para grandes distancias, y son de corto alcance. Habitualmente se usaba una distancia de corte de van der Waals de 8 que se creía que producía poco error. Sin embargo, las simulaciones de dinámica molecular de alcanos líquidos usando el campo de fuerzas GROMOS demostraron que las entalpías de vaporización y las presiones de vapor cambiaban

substantialmente cuando la distancia de corte de van der Waals se variaba en el intervalo de 8 a 14, y, de esta forma, el radio de corte de van der Waals de 16 es mucho más justificable que el tradicional de 8 [40].

A menudo, se usa una distancia de corte de entre 10 y 15. Sin embargo, las interacciones electrostáticas son de largo alcance; así, para ellos el uso de una distancia de corte no es justificable, y produce errores serios.

Si un corte se aplica abruptamente a una distancia interatómica particular, esta discontinuidad puede causar problemas en la minimización de la energía y en los cálculos de dinámica molecular. Para evitar esto, se puede usar un corte que haga que las interacciones no enlazantes se hagan gradualmente cero para una distancia, digamos de 1. Esto se hace usando lo que se llama función interruptor.

Enlace de hidrógeno

Para parejas de átomos que poseen un enlace de hidrógeno entre ellos, cualquier campo de fuerza modifica la interacción de van der Waals a una forma como $A/R^{12} - C/R^{10}$, sin embargo, muchos campos de fuerzas no contienen términos especiales para el enlace de hidrógeno, y recaen en los términos electrostáticos y de van der Waals para producir el enlace de hidrógeno.

Parametrización

El procedimiento general para obtener los valores de los parámetros en un campo de fuerzas es usar la información experimental o teórica *ab initio* para elegir un conjunto inicial de valores de los parámetros y, entonces, variar esos parámetros de forma que minimicen la desviaciones de las propiedades moleculares predichas por el campo de fuerzas respecto a las propiedades experimentales o calculadas *ab initio* para un conjunto de moléculas seleccionadas, llamado conjunto de entrenamiento. Las

propiedades moleculares usadas en las parametrizaciones de los campos de fuerzas incluyen estructuras moleculares, diferencias de energía conformacional, frecuencias vibracionales, barreras de rotación interna, momentos di-polares e interacciones intermoleculares.

2.5.3. Simulación

Condiciones de Borde Periódicas

Una manera clásica de minimizar los efectos de borde en un sistema infinito es aplicar condiciones de borde periódicas. Los átomos del sistema a ser simulado son puestos en una caja, la cual está rodeada de copias de sí misma (Figura 2.5). De esta manera no hay límites en el sistema; artefactos causados por bordes en un conjunto aislado de moléculas es remplazado por artefactos producidos por las condiciones de borde periódicas. Si se desea simular sistemas no periódicos como líquidos o soluciones, la periodicidad por si sola puede causar errores. Los errores pueden ser evaluados comparando varios tamaños de sistemas; se espera que estos errores son menos graves que los errores resultantes de simular un límite poco natural como el vacío.

Existen varias formas posibles de celdas unitarias utilizadas para condiciones de borde periódicas. Algunas como el dodecaedro rómbico y el octaedro truncado [3] son formas cercanas a una esfera más que a un cubo y por lo tanto más económicas en términos computacionales, para el estudio de una macromolécula en solución, dado que se requieren menos moléculas de solvente para completar el espacio dentro de la caja, dada una distancia mínima entre imágenes macromoleculares. Sin embargo, un sistema periódico basado en un dodecaedro rómbico u octaedro truncado es equivalente a un sistema periódico basado en una celda triclinica. Esta última forma es

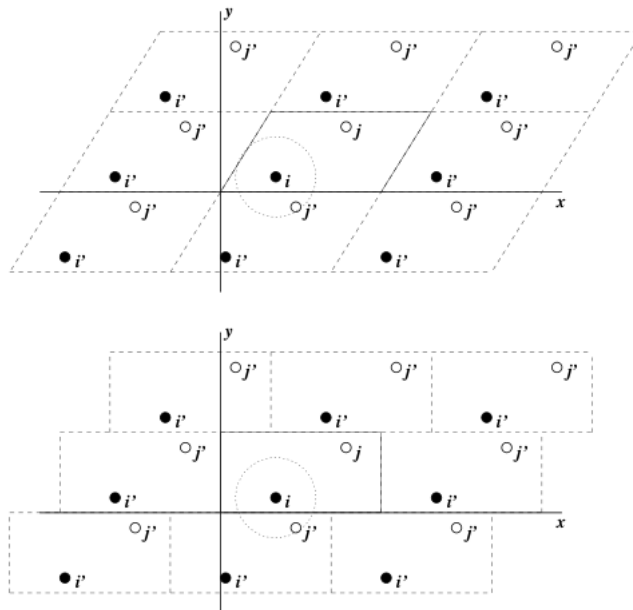


Figura 2.5: KAKAKAKSJSJSKACondiciones de borde periódicas en dos dimensiones.

la forma de celda más general; comprende todas las posibles formas de caja [14]. Es por esto que GROMACS está basado en la celda unitaria triclinica.

El programa computacional GROMACS utiliza condiciones de borde periódica, combinado a la convención de imágenes mínimas: solo una imagen, la más cercana, de cada partícula es considerada para términos de interacciones no enlazantes de corto alcance. para interacciones electrostáticas de largo alcance esto no siempre lo suficientemente preciso y GROMACS utiliza en ese caso incorpora métodos de suma en malla como la suma de Ewald y PME [51].

GROMACS soporta cajas triclinicas de cualquier forma. La caja está definida por 3 vectores a , b , y c . Los vectores deben satisfacer las siguientes condiciones:

$$a_y = a_z = b_z = 0 \tag{2.84}$$

$$a_x > 0, b_y > 0, c_z > 0 \quad (2.85)$$

$$|b_x| \leq \frac{1}{2}a_x, |c_x| \leq \frac{1}{2}a_x, |c_y| \leq \frac{1}{2}b_y \quad (2.86)$$

Las ecuaciones 2.84 siempre pueden ser satisfechas mediante rotar la caja. Las inecuaciones (2.85) y (2.86) pueden ser siempre satisfechas mediante la adición y sustracción de vectores de la caja.

Incluso cuando en la simulación se utiliza una caja triclinica, GROMACS siempre coloca las partículas en un volumen con forma de ladrillo, por razones de eficiencia. Esto se ilustra en la Figura 2.5 para un sistema bi-dimensional. Es por esto que la trayectoria de salida podría verse como si la simulación fue realizada en una caja rectangular.

Es también posible simular sin condiciones de borde periódicas, pero es más eficiente simular un grupo aislado de moléculas en una gran caja periódica, dado que la búsqueda rápida en una malla puede ser usada solamente si se utiliza un sistema periódico.

Tres de las más útiles tipos de cajas para simulaciones de sistemas solvatados se describen en la Tabla 2.2. El dodecaedro rómbico (Figura 2.6) es la más pequeña y más regular celda unitaria *space-filling*. Cada una de las 12 imágenes están a la misma distancia. El volumen es 71 % del volumen de un cubo con la misma distancia entre imágenes. Esto ahora cerca del 29 % de tiempo de CPU cuando se simula una molécula esférica o flexible en solvente. Hay dos orientaciones diferentes de un dodecaedro rómbico que satisfacen las ecuaciones 2.84, 2.85 y 2.86. En una orientación los primeros dos vectores de la caja coinciden con los ejes x e y . La otra orientación, que puede ser útil para simulaciones de proteínas de membrana. La sección transversal

con el plano xy es un hexágono, el cual es 14% más pequeña que un cuadrado con la misma distancia entre imágenes. La altura de la caja (c_z) puede ser modificada para obtener un espaciado más óptimo. Esta caja no solo ahorra tiempo de CPU si no que también resulta en un arreglo más uniforme de las proteínas.

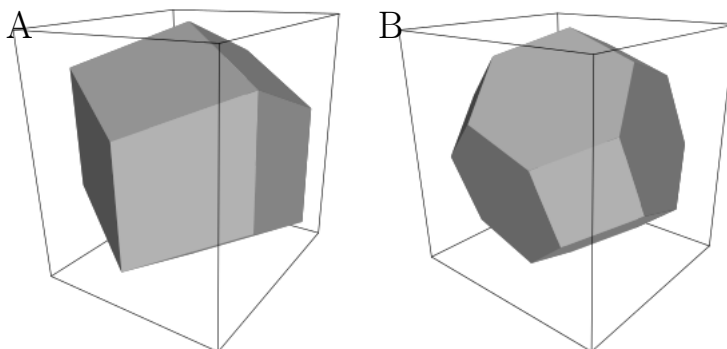


Figura 2.6: A) Dodecaedro rómbico y B) octaedro truncado

Restricciones para el Cutoff

La convención de imagen mínima implica que el radio de cutoff usado para interacciones no enlazantes truncadas no debe exceder la mitad del vector de la caja más pequeño:

$$R_c < \frac{1}{2} \min(\|a\|, \|b\|, \|c\|) \quad (2.87)$$

de otro modo más de una imagen podría estar dentro de la distancia de cutoff de fuerza. Cuando una macromolécula, como una proteína, es estudiada en solución, esta restricción no es suficiente. En principio una molécula de un solvente único no debería ser capaz de ver ambos lados de la macromolécula. Esto significa que el largo de cada vector de la caja debe exceder el largo de la molécula más dos veces el radio de cutoff.

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

Tabla 2.2: Caja cúbica, dodecaedro rómbico y el octaedro truncado.

Tipo de caja	Distancia entre imágenes	Volumen de la caja	Vectores de la caja			Ángulos de los vectores		
			a	b	c	∠bc	∠ac	∠ab
cúbica	d	d^3	$\begin{pmatrix} d \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ d \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 0 \\ d \end{pmatrix}$	90°	90°	90°
dodecaedro rómbico (cubo-xy)	d	$\frac{1}{2}\sqrt{2}d^3$	$\begin{pmatrix} d \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0 \\ d \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2}d \\ \frac{1}{2}d \\ \frac{1}{2}\sqrt{2}d \end{pmatrix}$	60°	60°	60°
dodecaedro rómbico (hexágono-xy)	d	$\frac{1}{2}\sqrt{2}d^3$	$\begin{pmatrix} d \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2}\sqrt{3}d \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{2}d \\ \frac{1}{6}\sqrt{3}d \\ \frac{1}{3}\sqrt{6}d \end{pmatrix}$	60°	60°	60°
octaedro truncado	d	$\frac{4}{9}\sqrt{3}d^3$	$\begin{pmatrix} d \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} \frac{1}{3}d \\ \frac{2}{3}\sqrt{2}d \\ 0 \end{pmatrix}$	$\begin{pmatrix} -\frac{1}{3}d \\ \frac{1}{3}\sqrt{2}d \\ \frac{1}{3}\sqrt{6}d \end{pmatrix}$	71.53°	109.47°	71.53°

Topología

Para realizar una simulación se requiere de la especificación de los tipos de átomos y combinaciones de átomos en donde las funciones de potencial actuarán. Esta información se encuentra en la topología en donde se definen todos los atributos de los componentes del sistema además de las interacciones de tipo enlazantes. Además, se especifican las fuerzas, posiciones y velocidades de los átomos. Un ejemplo de tipos de átomos en GROMACS se puede ver en la tabla 2.3.

Tabla 2.3: Tipos de átomos en GROMACS.

Tipo	Nombre	
O	Oxígeno carbonilo	(C=O)
OM	Oxígeno carboxilo	(CO-)
OA	Oxígeno hidroxilo	(OH)
OW	Oxígeno del agua	
N	Nitrógeno peptídico	(N or NH)
NT	Nitrógeno Terminal	(NH ₂)
NL	Nitrógeno Terminal	(NH ₃)
NR5	N aromático	(anillo de 5 lados, doble enlace)
NR5*	N aromático	(anillo de 5 lados, triple enlace)
NP	Nitrógeno porfirínico	
C	Carbono simple	(péptido, C=O, C-N)
CH1	Grupo CH Alifático	
CH2	Grupo CH ₂ Alifático	
CH3	Grupo CH ₃ alifático	

Campos de fuerza

Un campo de fuerza es la distribución espacial de la fuerza sobre una partícula resultante de todas las fuerzas que se ejercen sobre ella. Existen diferentes modelos de campo de fuerza según su naturaleza y las aproximaciones que se asuman. Estos modelos se a base del potencial que define tales fuerzas, luego, para calcular el comportamiento de una partícula sometida a un campo de fuerzas es necesario especificar las ecuaciones del modelo que lo describe considerando las aproximaciones asumidas y los parámetros del modelo.

GROMACS incluye diferentes tipos de campos de fuerzas pero recomienda GROMOS-

96 para el trabajo de macromoléculas como proteínas. El campo de fuerzas GROMOS-96 está parametrizado con un cut-off de Lennard-Jones de $1.4nm$, por lo que se debe utilizar un cutoff de Lennard-Jones para las interacciones de van der Waals de al menos $1.4nm$.

Coordenadas y velocidades

En comienzo de la simulación, en un tiempo $t = t_0$, todas las coordenadas de los átomos del sistema deben conocerse. El algoritmo leap-frog (Sección 2.5.3) requiere además de las velocidades en el instante $t = t_0 + \Delta t$. Si no se dispone de tales velocidades, se escogen de una distribución Maxwell-Boltzmann a una temperatura dada T :

$$p(v_i) = \sqrt{\frac{m_i}{2\pi kT}} \exp\left(-\frac{m_i v_i^2}{2kT}\right) \quad (2.88)$$

donde k es la constante de Boltzmann, m es la masa y v es la velocidad. Para esto, se generan números al azar normalmente distribuidos mediante la suma de doce números R_k con $0 \leq R_k < 1$ y restándole 6. El resultado se multiplica por la desviación estándar de la distribución de velocidad $\sqrt{kT/m_i}$. Dado que la energía total no corresponde exactamente a la que tendría a una temperatura T , se debe realizar una corrección eliminando el movimiento del centro de masas y escalando las velocidad de manera de que la energía total sea coherente con la temperatura T (Ecuación 2.90)

Energía cinética y temperatura

La temperatura está dada por la energía cinética total del sistema:

$$E_{kin} = \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \quad (2.89)$$

donde N corresponde a la cantidad de partículas del sistema. La temperatura absoluta del sistema se obtiene de la siguiente ecuación:

$$\frac{1}{2}N_{df}kT = E_{kin} \quad (2.90)$$

donde N_{df} corresponde a los grados de libertad del sistema:

$$N_{df} = 3N - N_c - N_{com} \quad (2.91)$$

N_c es el número de restricciones impuesta. N_{com} corresponde al numero de restricciones aportado por la restricción del movimiento del centro de masas que quita tres grados de libertad. Cuando la simulación se realiza sin solvente (en vacío), la rotación del centro de masas también puede ser removida, en ese caso N_{com} es seis.

Presión y virial El tensor de presión \mathbf{P} se calcula de la diferencia entre la energía cinética E_{kin} y el virial Ξ :

$$\mathbf{P} = \frac{2}{V} (\mathbf{E}_{kin} - \Xi) \quad (2.92)$$

donde V es el volumen de la caja, para el caso de sistemas isotrópicos la presión como escalar se calcula de la siguiente manera:

$$P = \frac{\text{trace}(\mathbf{P})}{3} \quad (2.93)$$

El virial está definido como:

$$\Xi = -\frac{1}{2} \sum_{i < j} r_{ij} \otimes \mathbf{F}_{ij} \quad (2.94)$$

Integrador leap frog

Para la integración de las ecuaciones de movimiento GOMACS utiliza, entre otros, el algoritmo leap-frog [69]. Para esto se necesita conocer las posiciones de los

átomos $\vec{r}(t)$ y la velocidad $\vec{v}(t - \frac{1}{2}\Delta t)$, así, va actualizando la posición y velocidad calculando la fuerza $\vec{F}(t)$ determinada por la posición a en el tiempo t :

$$\vec{v}\left(t + \frac{1}{2}\delta t\right) = \vec{v}\left(t - \frac{1}{2}\delta t\right) + \frac{\delta t}{m}\vec{F}(t) \quad (2.95)$$

$$\vec{r}(t + \delta t) = \vec{r}(t) + \delta t\vec{v}\left(t + \frac{1}{2}\delta t\right) \quad (2.96)$$

Temperature coupling

Consiste en control la temperatura y de esta forma realizar simulaciones con acople NVT (número constante de partículas, volumen y temperatura). Para simular temperaturas constantes GROMACS trabaja con tres enfoques, Berendsen, Nosé-Hoover, y reescalamiento de velocidad.

Berendsen temperature coupling El algoritmo de Berendsen imita un baño de calor a una temperatura dada T_0 con un acoplamiento débil con cinética de primer orden. De esta manera, las variaciones de temperatura respecto de T_0 son lentamente corregidas según:

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (2.97)$$

Esto quiere decir que la temperatura decae exponencialmente con la constante de tiempo τ . Para efectos de equilibrio del sistema este enfoque permite simulaciones cortas, de $0.01ps$ pero para un equilibrio más confiable se pueden hacer en más tiempo por ejemplo $0.5ps$

El termostato de Berendsen suprime las fluctuaciones de energía cinética. Esto implica en que no se genera el apropiado acoplamiento. Este error decrece según $1/N$ así que para sistemas grandes no se verán afectados significativamente salvo la

distribución de la energía cinética y propiedades de fluctuación como la capacidad calorífica. Un termostato similar que sí produce un adecuado acoplamiento es el de reescalamiento de velocidad.

El flujo de calor desde o hacia el sistema es afectado por el escalamiento de las velocidades de cada partícula en cada n_{TV} pasos, con un factor dependiente del tiempo dado por:

$$\lambda = \left[1 + \frac{n_{TC}\Delta t}{\tau_T} \left\{ \frac{T_0}{T(t - \frac{1}{2}\Delta t)} - 1 \right\} \right]^{\frac{1}{2}} \quad (2.98)$$

El parámetro τ_T es cercano al τ del temperature coupling (Ecuación 2.97):

$$\tau = \frac{2C_V\tau_T}{N_{df}k} \quad (2.99)$$

donde C_V es la capacidad calorífica total del sistema, k es la constante de Boltzmann y N_{df} es el número total de grados de libertad.

Velocity rescaling thermostat El termostato de reescalamiento de velocidad es esencialmente un termostato de Berendsen con un término estocástico adicional que asegura una distribución de la energía cinética correcta [26]:

$$dK = (K_0 - K) \frac{dt}{\tau_T} + 2\sqrt{\frac{KK_0}{N_f}} \frac{dW}{\sqrt{\tau_T}} \quad (2.100)$$

donde K es la energía cinética N_f es el numero de grados de libertad y dW es un proceso de Wiener⁸.

Nosé-Hoover El hamiltoniano del sistema se modifica introduciendo un reservorio termal y un termino de fricción en las ecuaciones de movimiento. La fuerza de fricción

⁸Un proceso de Wiener W_t es un proceso estocástico tiempo-continuo donde $W_0 = 0$, la función $t \rightarrow W_t$ es casi siempre continua y W_t tiene incrementos independiente con $W_t - W_s$ se distribuye normalmente con media 0 y varianza $t - s$.

es proporcional al producto de la velocidad de cada partícula y a un parámetro de fricción ξ . Este parámetro es una cantidad con momentum propio y ecuaciones de movimiento.

En este enfoque las ecuaciones de movimiento para la partícula son reemplazadas por:

$$\frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} - \frac{p_\xi}{Q} \frac{dr_i}{dt} \quad (2.101)$$

La ecuación de movimiento del parámetro ξ es:

$$\frac{dp_\xi}{dt} = (T - T_0) \quad (2.102)$$

La temperatura de referencia es T_0 y T es la temperatura instantánea del sistema. La cantidad conservada para las ecuaciones de movimiento Nosé-Hoover no es la energía total si no:

$$H = \sum_{i=1}^N \frac{p_i^2}{2m_i} + U(r_1, r_2, \dots, r_N) + \frac{p_\xi^2}{2Q} + N_f kT\xi \quad (2.103)$$

N_f es el número de grados de libertad. Q es el parámetro de masa:

$$Q = \frac{\tau_T^2 T_0}{4\pi^2} \quad (2.104)$$

Pressure Coupling

El acoplamiento de presión busca controlar la presión del sistema de manera análoga al acoplamiento de la temperatura.

Berendsen pressure coupling El algoritmo de Berendsen recalcula las coordenadas y vectores de la caja cada n_{PC} pasos con una matriz μ que tiene el efecto de una relajación de primer orden de la presión dada una presión de referencia P_0 :

$$\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} \quad (2.105)$$

La matriz μ viene dada por:

$$\mu_{ij} = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_p} \beta_{ij} \{P_{0ij} - P_{ij}(t)\} \quad (2.106)$$

β es la compresibilidad isothermal del sistema. En muchos casos será una matriz diagonal con elementos iguales cuyos valores se desconocen. Es suficiente tener una estimación de ellos puesto que el valor de β solo influencia la constante de tiempo no crítica de la relajación de la presión sin afectar el promedio de la presión misma.

Para agua a $1atm$ a $300K$ $\beta = 4.6 \times 10^{-10} Pa^{-1} = 4.6 \times 10^{-5} bar^{-1}$

Algoritmo LINCS

El algoritmo LINCS restablece el largo de los enlaces a su largo correcto [66].

Se considera un sistema de N partículas cuyas posiciones están dadas por el vector de largo $3N$ $r(t)$. Las ecuaciones de movimiento dadas por la ley de Newton es:

$$\frac{d^2r}{dt^2} = M^{-1}\vec{F} \quad (2.107)$$

donde \vec{F} es el vector de Fuerzas de largo $3N$ y M es una matriz diagonal de largo $3N \times 3N$ con las masas de las partículas. El algoritmo se realiza en dos etapas, en la primera las proyecciones de los nuevos enlaces en los antiguos se fijan en cero. En el segundo paso, se aplica una corrección para el alargamiento de los enlaces debido a la rotación. El sistema está posee K restricciones independientes del tiempo:

$$g_i(r) = |r_{i1} - r_{i2}| - d_i = 0 \quad i = 1, \dots, K \quad (2.108)$$

Se define B_{hi} como la matriz de gradiente de las restricciones⁹:

$$B_{hi} = \frac{\partial g_h}{\partial r_i} \quad (2.109)$$

⁹ B es una matriz de tamaño $K \times 3N$. Contiene las direcciones d de las restricciones

2.5. DINÁMICA DE UNA ENZIMA: SIMULACIÓN *IN SILICO*

La relación entre las nuevas coordenadas restringidas r_{n+1} están relacionadas a las coordenadas no restringidas r_{n+1}^{unc} por:

$$\begin{aligned} \vec{r}_{n+1} &= (I - T_n B_n) \vec{r}_{n+1}^{unc} + T_n \vec{d} = \\ r_{n+1}^{unc} - M^{-1} B_n (B_n M^{-1} B_n^T)^{-1} (B_n r_{n+1}^{unc} - \vec{d}) \end{aligned} \quad (2.110)$$

donde $T = M^{-1} B^T (B M^{-1} B^T)^{-1}$.

En este primer paso no se establece el largo real del enlace sino que su proyección en el antiguo. Para corregir por la rotación del enlace i , La proyección de enlace se establece como:

$$p_i = \sqrt{2d_i^2 - l_i^2} \quad (2.111)$$

donde l_i es el largo del enlace luego de la primera proyección. Las posiciones corregidas son:

$$r_{n+1}^* = (I - T_n B_n) r_{n+1} + T_n p \quad (2.112)$$

La mitad del tiempo que el CPU utiliza en cálculos lo ocupa invirtiendo la matriz de tamaño $K \times K$ $B_n M^{-1} B_n^T$ en cada paso. La matriz T es invertida con una expansión de potencias.

$$\begin{aligned} (B_n M^{-1} B_n^T)^{-1} &= S S^{-1} (B_n M^{-1} B_n^T)^{-1} S^{-1} S \\ S (S B_n M^{-1} B_n^T S)^{-1} S &= S (I - A_n)^{-1} S \end{aligned} \quad (2.113)$$

S es la raíz cuadrática inversa de la diagonal de $B_n M^{-1} B_n^T$. La inversa de $I - A_n$ se calcula de la siguiente manera:

$$(I - A_n)^{-1} = I + A_n + A_n^2 + A_n^3 + \dots \quad (2.114)$$

Minimización de energía con *steepest descent*

En GROMACS la minimización de energía puede ser realizada mediante el método *steepest descent*.

Se define el vector r como el vector de las $3N$ coordenadas y se define un h_0 dato, como el desplazamiento máximo inicial. Se calculan las fuerzas F y el potencial para determinar las nuevas posiciones como:

$$\vec{r}_{n+1} = \vec{r}_n + \frac{\vec{F}_n}{\max(|\vec{F}_n|)} h_n \quad (2.115)$$

donde h_n es el desplazamiento máximo y \vec{F}_n es la fuerza. Se calculan las fuerzas y potencial con las nuevas coordenadas y se compara el potencial del paso anterior con el nuevo. Si el nuevo potencial es menor que el anterior, entonces las nuevas posiciones son aceptadas y el paso $h_{n+1} = 1.3h_n$. Si el potencial es mayor, las nuevas coordenadas son rechazadas y se achica el paso, $h_n = 0.2h_n$. Esto se repite hasta que el algoritmo alcance un número máximo de iteraciones o bien cuando la fuerza sea mayor que un determinado valor ϵ

3 . MODELOS DE ESTUDIO

En el presente trabajo se utilizaron enzimas celobiohidrolasas de la familia de las Cel7 de *Talaromyces emersonii* y de *Melanocarpus albomyces* como modelos para la validación del procedimiento de mejoramiento propuesto. En esta sección se describen las principales características de tales enzimas en términos de estructura y actividad. La enzima de *Phanerochaete chrysosporium* estudiada en este trabajo se denominó Cel72, y se buscó mutaciones que reduzcan la flexibilidad de la estructura para mejorar su termoestabilidad. Se realizó una caracterización de la enzima Cel72 y se muestran estos resultados en la sección 7. Estas enzimas son útiles para ser utilizadas en la validación debido a su similitud estructural con Cel72 (sección 7.4.1) y a que poseen antecedentes de mutaciones que mejoran su termoestabilidad [141,142].

3.1. Celobiohidrolasa Cel7A de *Talaromyces emersonii*

La celobiohidrolasa Cel7A de *T. emersonii* es una enzima termoestable perteneciente a la familia 7 de las glicosil hidrolasas [60]. Posee un óptimo de temperatura de 68°C a pH 5 y una vida media de 68 minutos a 80°C y pH 5 [60].

Su estructura posee un plegamiento de β -sándwich (ver Figura 3.1), en donde 6 loops se prolongan desde él cerrándose y formando un túnel de alrededor de 50 [50, 108] para el ingreso y posterior hidrólisis del sustrato. El sitio catalítico está compuesto por tres ácidos carboxílicos y ubicados cerca de un extremo del túnel para permitir la liberación del producto (celobiosa) del extremo reducido de la cadena de celulosa. En el sitio activo, los residuos Glu209 y Glu214 están ubicados en forma opuesta al sitio de clivaje del enlace glicosídico y actúan como nucleófilo y donador de

3.1. CELOBIOHIDROLASA CEL7A DE *TALAROMYCES EMERSONII*

protones, respectivamente, y el residuo Asp211 actúa controlando la posición correcta y protonación de Glu209. Posee una secuencia señal para exportación extracelular y carece de módulo de unión a carbohidratos (CBM). Posee nueve puentes disulfuro. En la Figura 3.1 se muestra el modelo de la estructura tridimensional del cristal de la enzima Celobiohidrolasa Cel7A de *Talaromyces emersonii*.

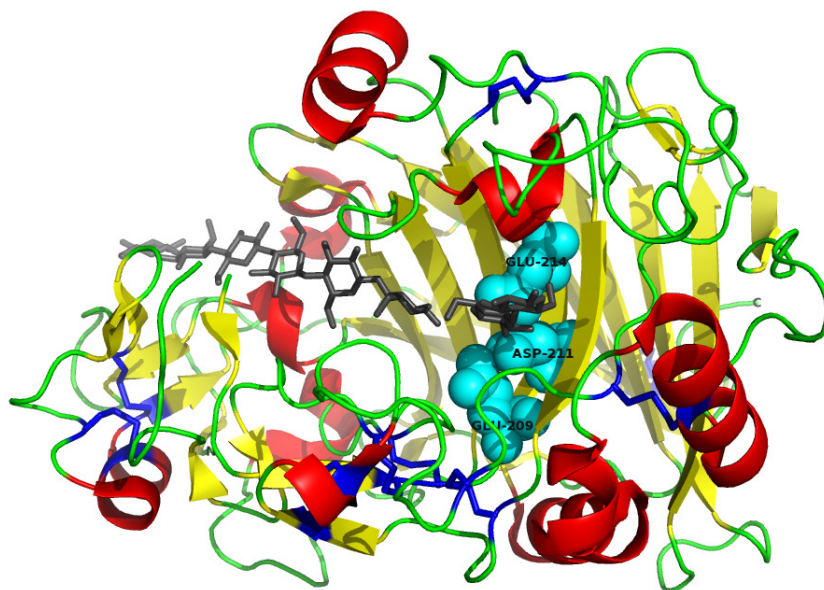


Figura 3.1: Estructura tridimensional de Cel7A de *T. emersonii* (1Q9H) que muestra el túnel por donde entra el sustrato. En celeste se muestran los residuos Glu209, Glu214 y Asp211 del sitio activo y azul los enlaces disulfuro. Las hojas beta están representadas en color amarillo y las hélices alfa en rojo. La estructura está superpuesta con la de Cel7A de *T. reesei* (7CEL) para mostrar una cadena de celulosa (gris) unida al sitio activo. Se utilizó el *software* PyMol para visualizar la estructura

El gen *cel7* que codifica para la Cel7A de *T. emersonii* consiste en un ORF (*open reading frame*) de 1365bp que codifica una secuencia de 455 aminoácidos, interrumpido por dos intrones. Posee ocho regiones de unión a Cre río arriba del gen y es probable que la represión por glucosa de la transcripción de la celulasa sea mediada por Cre [60].

3.2. CELOBIOHIDROLASA CEL7B DE *MELANOCARPUS ALBOMYCES*

La degradación de la celulosa ocurre de forma procesiva, en donde la cadena de celulosa avanza hacia el interior del túnel a medida que va saliendo el producto. El ataque hidrolítico sucede en el extremo de la cadena donde se produce glucosa o bien celotetrosa al comienzo y luego el producto principal que es celobiosa.

3.2. Celobiohidrolasa Cel7B de *Melanocarpus albomyces*

Cel7B de *M. albomyces* es una enzima perteneciente al grupo 7 de las glicosil hidrolasas (GH-7) con una termoestabilidad de 68°C a pH 6.0. Está compuesta por 430 aminoácidos y posee un peso molecular de 47,5 kDa. El producto principal de la hidrólisis de celulosa es celobiosa; posee el plegamiento característico de las celobiohidrolasas del grupo GH-7 y que consiste en hojas beta que comprenden un sándwich con forma cóncava de donde salen loops para formar un túnel. El túnel es un espacio en donde penetra la celulosa para ser hidrolizada de forma progresiva por el sitio activo conformado por tres residuos, Glu212, Asp214 y Glu217. De igual manera que con la Cel7A de *T. emersonii*, los residuos Glu212 y Glu217 actúan como nucleófilo y dador de protones mientras que Asp214 actúa controlando la posición y protonación del residuo nucleófilo [108]. Esta enzima no posee un CBM. En la Figura 3.2 se muestra el modelo de la estructura tridimensional del cristal de la enzima.

3.2. CELOBIOHIDROLASA CEL7B DE *MELANOCARPUS ALBOMYCES*

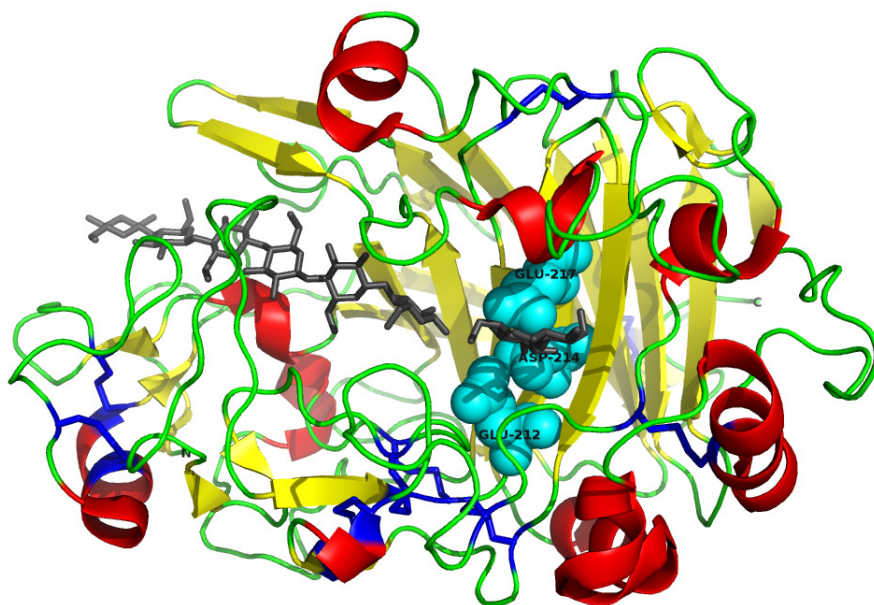


Figura 3.2: Estructura tridimensional de Cel7B de *M. albomyces* (2RFW) que muestra el túnel por donde entra el sustrato. En celeste se muestran los residuos Glu212, Asp214 y Glu217 del sitio activo y azul los enlaces disulfuro. Las hojas beta están representadas en color amarillo y las hélices alfa en rojo. La estructura está superpuesta con la de Cel7A de *T. reesei* (7CEL) para mostrar una cadena de celulosa (gris) unida al sitio activo. Se utilizó el *software* PyMol para visualizar la estructura

4 . HIPÓTESIS DE TRABAJO

La simulación del comportamiento térmico de proteínas mediante herramientas de dinámica molecular permite, junto al análisis de componentes principales, la identificación de zonas que difieren en el grado de flexibilidad de los residuos para, de esta forma, buscar mutaciones que puedan mejorar la termoestabilidad de proteínas mediante el cambio de residuos por otros que rigidicen las regiones más flexibles.

5 . OBJETIVOS

5.1. Objetivo General

Diseñar un procedimiento que permita proponer mutaciones para el mejoramiento de la termoestabilidad de enzimas.

5.2. Objetivos Específicos

- Diseñar una estrategia basada en dinámica molecular y Análisis de Componentes Principales para la identificación de zonas de alta flexibilidad en proteínas.
- Validar la estrategia planteada, usando proteínas modelo de estructura y termoestabilidad conocidas, las enzimas Cel7A de *T. emersonii* y Cel7B de *M. albomyces*
- Obtener un modelo de la estructura tridimensional de la enzima Cel72 y evaluar la flexibilidad de sus residuos usando Análisis de Componentes Principales.
- Proponer cambios de aminoácidos en la estructura de Cel72 que aumenten la termoestabilidad de la enzima, usando criterios de flexibilidad identificados por medio de la estrategia planteada en esta tesis.
- Evaluar comparativamente *in silico* el grado de mejoramiento de la termoestabilidad de las variantes propuestas, mediante el análisis de flexibilidad y compactabilidad de las enzimas

6 . MATERIALES Y MÉTODOS

6.1. Materiales

6.1.1. Hardware

Los cálculos de la simulación se realizaron en un cluster de procesadores de 80 núcleos, CPU Quad Core AMD Opterontm, 16 x 10 GBi RAM, Espacio en disco de 1 TBi en RAID 1, conectado mediante un switch de red de 1 Gbs. Para cada simulación se utilizaron 8 núcleos del cluster. El análisis de los resultados fue realizado en un computador laptop Dell Vostro 3500 con procesador Intel Core i5 M520 de 2.4GHz, doble núcleo y tecnología Hiper Threating y sistema operativo Ubuntu 11.10. Para determinar los valores F se programó el cálculo utilizando el lenguaje Python.

6.1.2. Software

GROMACS

Para todas las simulaciones computacionales realizadas en este trabajo se utilizó GROMACS [19,67,90,136]. GROMACS es un paquete de herramientas computacionales cuyo principal uso es la simulación de las ecuaciones de movimiento de Newton para sistemas de millones de partículas. GROMACS está diseñado para su uso con moléculas bioquímicas como proteínas.

Swiss-Modell

Para la fabricación de los modelos de enzimas de las que no se poseía registro de su estructura en el PDB (Protein Data Bank), se utilizó el servidor para modelamiento de estructuras de proteínas por homología Swiss-Model [10,78,111].

6.2. Métodos

6.2.1. Cálculo de Trayectorias

Para el cálculo de las trayectorias se utilizó el programa GROMACS. El procedimiento de cálculo consistió en varias etapas de preparación de la estructura para realizar finalmente la etapa de producción. Para esto se utilizaron varios módulos de GROMACS y la descripción de su uso se detalla en la Sección de Anexos A.1.

El protocolo general para el cálculo de la trayectoria de una macromolécula (Figura 7.1), en particular una enzima, consiste en tres etapas. La primera parte consiste en preparar el sistema de modo de alcanzar las condiciones a las cuales se desea realizar las mediciones. En un comienzo, se establece el contexto en el que la estructura de la enzima está inmersa. De esta forma, primero se agregan el solvente e iones y posteriormente se debe aumentar la temperatura y presión hasta alcanzar las condiciones de simulación. Posteriormente, una vez establecidas las condiciones de operación, se corre una simulación que considera dos etapas. La primera es de equilibrio, en donde el sistema se estabiliza y los átomos se acomodan para dar paso a la etapa de producción propiamente de la que los datos son recogidos. Las etapas de equilibrio y producción corresponden a una simulación de 11 *ns*, de los cuales los últimos 10 *ns* corresponden a la producción.

6.2.2. Cálculo de valores f y if

Para escoger las regiones a mutar en una enzima con el fin de mejorar su termoestabilidad, se planteó estudiar los desplazamientos de los átomos y se propuso como herramienta el cálculo de la varianza de la trayectoria de cada átomo en el tiempo utilizando análisis de componentes principales (PCA) [133]. A partir de la

información contenida en la trayectoria determinar ejes de máxima varianza para cada uno de los átomos de la enzima y así su varianza en este eje.

Valor f

La trayectoria de un átomo i se puede escribir de la siguiente forma:

$$M_i = \begin{bmatrix} x_{i1} & y_{i1} & z_{i1} \\ x_{i2} & y_{i2} & z_{i2} \\ \vdots & \vdots & \vdots \\ x_{ij} & y_{ij} & z_{ij} \\ \vdots & \vdots & \vdots \\ x_{in} & y_{in} & z_{in} \end{bmatrix} \quad (6.1)$$

donde (x_{ij}, y_{ij}, z_{ij}) es la posición del átomo i en el tiempo $t = t_j$. La matriz de covarianza C_i para la posición del átomo i es

$$C_i = \begin{bmatrix} var(x_i) & cov(x_i, y_i) & cov(x_i, z_i) \\ cov(y_i, x_i) & var(y_i) & cov(y_i, z_i) \\ cov(z_i, x_i) & cov(z_i, y_i) & var(z_i) \end{bmatrix} \quad (6.2)$$

donde $x_i = (x_{i1}, \dots, x_{in})$. De la misma manera se define y_i y z_i . Si se descompone C_i en sus valores principales (sección A.2) se obtiene

$$C_i = U_i S_i V_i^t \quad (6.3)$$

donde $U_i = (u_{i1}, u_{i2}, u_{i3})$ es la matriz con los vectores singulares u_{ik} en cada columna, tal que $U_i^t U_i = I$ y $S_i = (s_{i1}, s_{i2}, s_{i3})$ es una matriz con los valores singulares s_{i1}, s_{i2}, s_{i3} en la diagonal y ceros en todo el resto de la matriz. En este caso, como la matriz que se descompone es la matriz de covarianza, los vectores singulares corresponden a los ejes principales y los valores singulares corresponden a la varianza a lo largo de esos ejes. El primer vector singular es el eje de máxima varianza y la varianza asociada a este eje comprende la porción mayor de varianza. Así, la flexibilidad se

definió como la desviación estándar de la trayectoria a lo largo del eje de máxima varianza [17]:

$$f_i = \sqrt{s_{i1}} \quad (6.4)$$

Para determinar la flexibilidad de un residuo se utilizó el promedio de los valores f de los átomos que lo constituyen. De esa flexibilidad determinada para cada residuo en cada temperatura se obtuvo el valor if (Figura 7.18 en la Sección 7.3.2). Este valor se obtiene de la suma entre el f a $300K$ y el promedio de la norma de la diferencia entre los valores f a la primera temperatura ($300K$) y las otras dos (350 y $400K$):

$$if_i = f_{300,i} + \frac{|f_{300,i} - f_{350,i}| + |f_{300,i} - f_{400,i}|}{2} \quad (6.5)$$

Con este nuevo valor de flexibilidad para cada residuo se escogieron las regiones más flexibles candidatas a mutación mediante un ranking. El ranking se hizo ordenando los residuos de mayor a menor if para escoger los sitios más flexibles y luego proponer mutaciones.

6.2.3. Determinación de la Raíz Cuadrática Media de las Desviaciones

La raíz cuadrática media de las desviaciones (RMSD) fue calculada por el modulo de GROMACS `g_rms`. El cálculo se realiza a base de la siguiente expresión:

$$RMSD(t_1, t_2) = \left[\frac{1}{M} \sum_{i=1}^N m_i \|\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (6.6)$$

donde $M = \sum_{i=1}^N m_i$ y $\mathbf{r}_i(t)$ es la posición del átomo i en el tiempo t .

6.2.4. Compactación

Para visualizar el efecto de la temperatura en la estructura, en términos de la pérdida de la configuración espacial a lo largo de la simulación, se calcula la compactación de la enzima en función del tiempo. Como medida tosca de la compactación de una estructura se calcula el radio de giro (Radius of gyration, R_g) con el programa de GROMACS `g_gyrate`, el que lo calcula de la siguiente manera:

$$R_g = \left(\frac{\sum_{i=1}^N \|\mathbf{r}_i\|^2 m_i}{\sum_{i=1}^N m_i} \right)^{\frac{1}{2}} \quad (6.7)$$

donde m_i es la masa del átomo i y \mathbf{r}_i es su posición con respecto al centro de masas de la molécula.

7 . RESULTADOS

Se desarrolló un procedimiento para el mejoramiento de la termoestabilidad de enzimas, en particular Cel72, que consiste, en términos generales, en 3 etapas: (1) simular el comportamiento de la estructura bajo 3 temperaturas, 300, 350 y 400K; (2) evaluar la flexibilidad de la estructura según los datos obtenidos de las simulaciones y proponer mutaciones que rigidicen las regiones más flexibles de la estructura, y (3) repetir las etapas anteriores para las estructuras de las variantes de la enzima con el fin de determinar el grado de mejoramiento (Figura 7.1). La elección de las temperaturas se hizo en base a representar las vibraciones moleculares de la enzima y cómo la estructura terciaria de ésta cambia al aumentar la temperatura del sistema, y no busca, necesariamente, representar las condiciones ideales o de laboratorio ni las condiciones de operación de la enzima en un proceso productivo.

Como herramienta adicional de comparación entre las estructuras se calculó el Rg para observar los cambios en la compactación de las estructuras frente al aumento de la temperatura y de las enzimas variantes respecto de la enzima nativa. Una etapa previa (0) será necesaria en caso de que la enzima no posea una estructura conocida con anterioridad. Esta etapa se denomina modelamiento y consiste en la determinación de la estructura de la enzima ya sea por métodos experimentales, o bien, por métodos computacionales. Debido a que la estructura de Cel72 no se ha determinado experimentalmente, se realizó una etapa de modelamiento por homología previa.

El procedimiento antes mencionado fue aplicado a las estructuras de las enzimas Cel7A de *T. emersonii* (denominada enzima T) y Cel7B de *M. albomyces* (denominada enzima M) con el objetivo de validar el procedimiento para luego utilizarlo con la estructura de la enzima Cel72. A continuación se detallan los resultados obtenidos

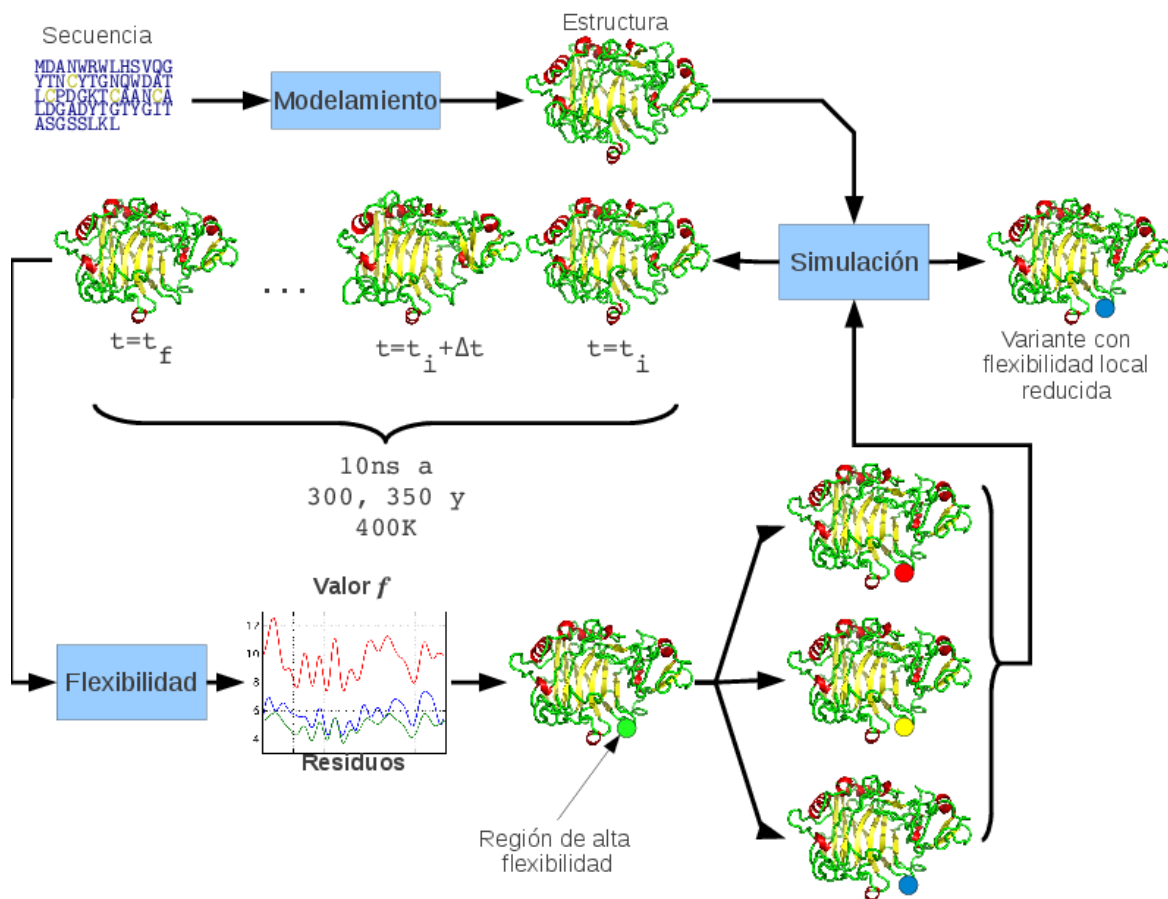


Figura 7.1: Esquema general del procedimiento de mejoramiento de termoestabilidad. Con la estructura de la enzima, se simula el efecto de la temperatura sobre el sistema para luego determinar la flexibilidad y en base a eso generar variantes. Las variantes serán simuladas nuevamente para ser evaluadas y así determinar cuál o cuáles de las variantes poseen características de compactabilidad y flexibilidad mejores que la enzima nativa.

en la validación y en la aplicación del procedimiento con la estructura de Cel72.

7.1. Caracterización de la enzima Cel72

Debido a que la estructura de la enzima Cel72 aún no está reportada, se muestran a continuación el resultado del análisis mediante la herramienta *on line* BLAST [5]) y en particular el programa BLASTP 2.2.26+ [6,7]. Al realizar la búsqueda se encontraron 100 estructuras con puntajes superiores a 500. En la Tabla 7.1 se muestran los 15 mejores puntajes obtenidos. Se detectaron dominios conservados putativos (Figura 7.2) determinando que la enzima corresponde a una CBH de la familia GH-7. El alineamiento múltiple para la determinación de los dominios conservados se muestra en la Sección A.3. Se encontró que posee residuos conservados respecto a GH-7. 16 de los 16 residuos que componen esta fracción conservada están presentes en Cel72.

Tabla 7.1: Resultado del alineamiento de la secuencia de Cel72 utilizando BLAST. Se muestran los 15 primeros resultados.

Descripción	Puntaje	E. value	Identidad	Positivos	Gaps
exo-cellobiohidrolase I precursor [Phanerochaete chrysosporium]	803	0,0	388/396 (98 %)	390/396 (98 %)	0/396 (0 %)
RecName: Full=Exoglucanase 1; AltName: Full=1,4-beta-cellobiohidrolase	802	0,0	388/396 (98 %)	390/396 (98 %)	0/396 (0 %)
cellulase [Phanerochaete chrysosporium]	800	0,0	387/396 (98 %)	389/396 (98 %)	0/396 (0 %)
Chain A, Cellobiohidrolase Cel7d From Phanerochaete Chrysosporium.	697	0,0	321/397 (81 %)	363/397 (91 %)	0/397 (0 %)
exocellulase [Irpex lacteus]	625	0,0	295/403 (73 %)	345/403 (86 %)	6/403 (1 %)
cellobiohidrolase [Irpex lacteus]	619	0,0	292/403 (72 %)	343/403 (85 %)	6/403 (1 %)
cellulase [Coniophora puteana]	602	0,0	283/403 (70 %)	331/403 (82 %)	6/403 (1 %)
cellobiohidrolase I [Neolentinus lepideus]	585	0,0	281/402 (70 %)	325/402 (81 %)	12/402 (3 %)
glycoside hydrolase family 7 cellobiohidrolase [Schizophyllum commune]	575	0,0	268/404 (66 %)	321/404 (79 %)	9/404 (2 %)
cellulose 1,4-beta-cellobiosidase precursor [Pleurotus sp. 'Florida']	574	0,0	272/406 (67 %)	322/406 (79 %)	11/406 (3 %)

Continúa en la página siguiente

7.1. CARACTERIZACIÓN DE LA ENZIMA CEL72

Tabla 7.1 : Resultado de alineamiento. (Continuación de la página anterior).

Descripción	Puntaje	E. value	Identidad	Positivos	Gaps
Chain A, 3-Dimensional Structure Of Native Cel7a From <i>Talaromyces Emersonii</i>	570	0,0	266/405 (66 %)	324/405 (80 %)	9/405 (2 %)
glycoside hydrolase family 7 protein [<i>Botryotinia fuckeliana</i>]	567	0,0	274/407 (67 %)	321/407 (79 %)	11/407 (3 %)
glycoside hydrolase family 7 protein [<i>Myceliophthora thermophila</i> ATCC 42464]	562	0,0	272/405 (67 %)	309/405 (76 %)	11/405 (3 %)
Structure Of Cellobiohydrolase 1 (Cel7a) From <i>Heterobasidion Annosum</i>	560	0,0	269/406 (66 %)	322/406 (79 %)	11/406 (3 %)
cellobiohydrolase I-I [<i>Volvariella volvacea</i>]	561	0,0	274/406 (67 %)	320/406 (79 %)	10/406 (2 %)

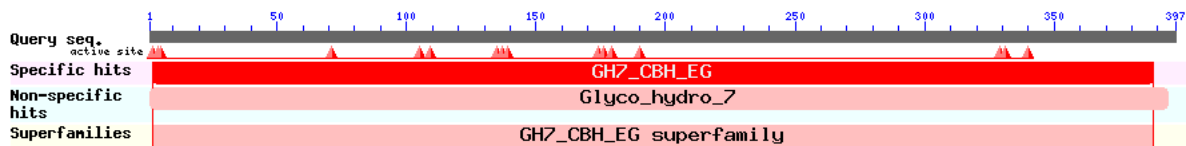


Figura 7.2: Esquema de los dominios conservados encontrados para Cel72 y de los aciertos específicos obtenidos. Se detectaron dominios conservados putativos determinando que la enzima corresponde a una CBH de la familia GH-7.

La estructura tridimensional de la enzima (Figura 7.3) se modeló mediante modelamiento por homología y se detalla en la sección 7.4.1, y comprende un túnel formado por hojas beta dispuestas de manera tal que forman un sándwich cóncavo con loops que cierran el túnel. De la estructura tridimensional y del alineamiento realizado es posible deducir que el sitio activo lo conforman los residuos Glu174, Asp176 y Glu179.

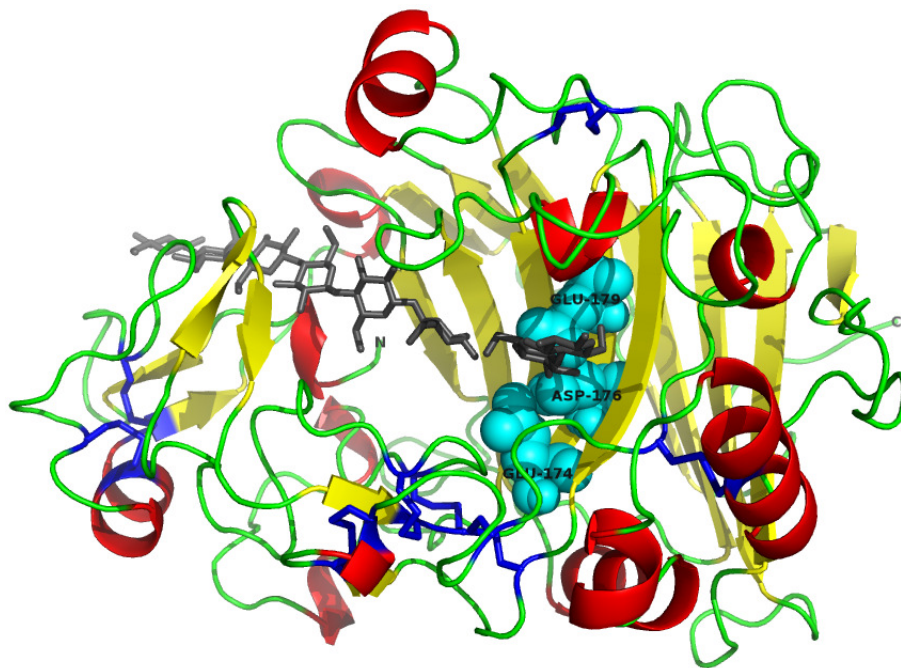


Figura 7.3: Modelo de la estructura tridimensional de Cel72 que muestra el túnel por donde entra el sustrato. En celeste se muestran los residuos Glu174, Asp176 y Glu179 del sitio activo y azul los puentes disulfuro. Las hojas beta están representadas en color amarillo y las hélices alfa en rojo. La estructura está superpuesta con la de Cel7A de *T. reesei* (código PDB: 7CEL) para mostrar una cadena de celulosa (gris) unida al sitio activo.

7.2. Procedimiento para la simulación del comportamiento de la estructura enzimática, a tres diferentes temperaturas e identificación de regiones de flexibilidad

Se elaboró un procedimiento para proponer cambios en la secuencia aminoacídica de que generen una mejora en la termoestabilidad de enzimas. El procedimiento consiste principalmente en dos etapas, primero se simula el comportamiento de la estructura de la enzima a tres temperaturas para posteriormente evaluar su flexibilidad

7.2. PROCEDIMIENTO PARA LA SIMULACIÓN Y IDENTIFICACIÓN DE MUTANTES

estructural en términos del valor f y proponer mutantes que rigidicen la estructuras en una o más regiones identificadas como flexibles, y finalmente se repite el protocolo de simulación para las variantes propuestas para poder determinar la compactación de las enzimas y valores f con el fin de identificar la o las mutaciones que resulten en posibles mejoras en la termoestabilidad (Figura 7.1).

El protocolo de dinámica molecular en términos generales consisten en tres etapas: inicialización, equilibrio y producción [54, 65, 135, 143]. En la etapa de inicialización se prepara la estructura y el sistema en términos de temperatura, presión y energía y se definen el solvente e iones. En la etapa de equilibrio se busca la minimización de artefactos estructurales y dinámicos realizando una simulación previa a la producción con las mismas condiciones del sistema [22, 143]. La etapa de producción es la simulación de la cual se extraerán los datos para el análisis posterior.

A continuación se detalla el procedimiento completo utilizado tanto para la validación como para la enzima Cel72. Todas las primeras etapas antes de la etapa de equilibrio comprenden la inicialización. Los registros de salida de cada etapa se encuentran en la sección de Anexos.

7.2.1. Conversión de la estructura y generación de la topología

Para la simulación se utilizó el software GROMACS. Dado que la estructura obtenida del PDB sólo contiene coordenadas, es necesario primero construir la topología de la molécula, la cual describe el sistema en términos de tipos de átomos, cargas, enlaces y otros tipos de interacciones no enlazantes como interacciones de Van der Waals. Esta topología es específica para un determinado campo de fuerza. El campo de fuerza empleado fue GROMOS 53a6 [106].

7.2. PROCEDIMIENTO PARA LA SIMULACIÓN Y IDENTIFICACIÓN DE MUTANTES

Es importante que la topología esté emparejada con la estructura, es decir que no sólo se genera la topología del sistema si no que también se convierte la estructura para adecuarse al campo de fuerza. Para convertir la estructura y construir la topología se utiliza el programa `pdb2gmx`. Este programa está diseñado para construir topologías para moléculas que se basan en *building blocks* como son los aminoácidos. El programa utiliza un biblioteca de *building blocks* para la conversión y entregará un error al intentar reconocer moléculas o residuos que no están presentes en la biblioteca. Se eliminaron los átomos de hidrógeno presentes en la estructura PDB para ser restituidos de acuerdo a la descripción en el campo de fuerza. Los archivos de registro de obtenido de la conversión de estructura se muestran en A.5.

7.2.2. Minimización de energía de la estructura en el vacío

La conversión de la estructura involucra quitar y adicionar átomos de hidrógeno lo que puede introducir tensión al sistema debido a átomos posicionados muy cerca entre ellos. Por lo tanto es necesario realizar una minimización de energía en la estructura y dejar que se relaje levemente. Esto corresponde a una corta simulación y requiere de dos procesos. Primero la estructura y la topología son combinadas en una sola descripción del sistema junto a una serie de parámetros de control. Esto produce un archivo de entrada para el programa que realiza el segundo proceso, la minimización. Los parámetros de control son listados en la Tabla A.1, en la sección 6.2.1.

Una vez minimizada la energía de la estructura se procede a establecer las condiciones de borde periódicas y agregar el solvente.

7.2.3. Condiciones de Borde Periódicas

Antes de agregar el solvente se debe escoger la disposición espacial de sistema. Se utilizó condiciones de borde periódicas (PBC), es decir, se define una sola celda la cual se repite de manera de completar totalmente el espacio. De este modo es posible simular un sistema infinito y periódico evitando problemas producto de simular los bordes de un sistema finito. Se escogió una celda de forma de un dodecaedro rómbico dado que esta forma corresponde al empaque óptimo de una esfera y así es la mejor alternativa para poder rotar libremente la molécula que se desea simular. Para evitar interacciones directas entre imágenes periódicas se establece una distancia mínima entre la molécula y el borde de la celda de 1 [nm]. En la Figura 7.4 se ejemplifica mostrando la celda para la enzima T.

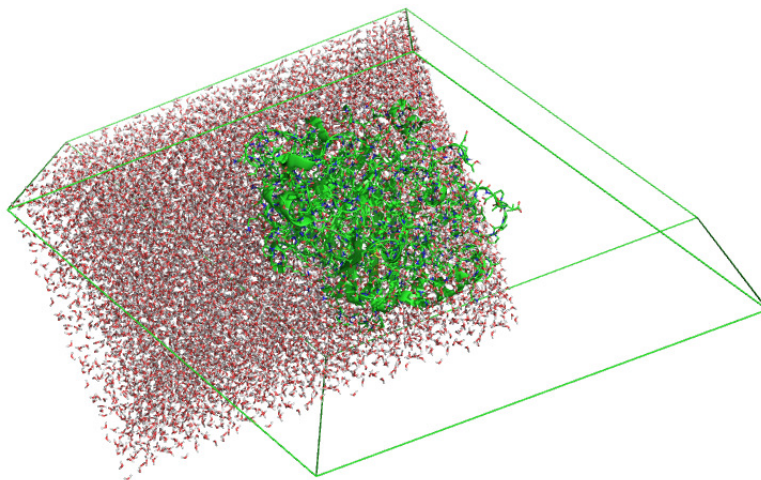


Figura 7.4: Proteína en una caja cúbica en dónde el solvente es agua. Se ha graficado los bordes de una celda triclinica. Esta celda triclinica corresponde a una parte de la celda dodecaédrica mencionada. Se muestra un bloque rectangular de agua que no está contenido completamente en la celda y la enzima parcialmente fuera del solvente.

7.2.4. Adición de Solvente e Iones

Una vez que se ha determinado el tipo de celda unitaria que se utilizará es posible agregar el solvente al sistema. Hay diferentes modelos de solvente y pueden ser explícitos o implícitos. En este caso se utilizó agua explícita como solvente de tipo SPC (Simple point charge). De este modo se obtienen una topología actualizada del sistema y una estructura que ahora incluyen a las moléculas de agua. Además, se agregan iones para neutralizar el sistema. Con estos propósitos se utilizó una concentración de NaCl de $0.15M$ [59].

7.2.5. Minimización de energía del sistema solvatado

Antes de comenzar con las etapas de simulación es necesario realizar un último paso de minimización de la energía del sistema. Al igual que luego de la etapa de conversión de la estructura y generación de la topología, la adición del solvente e iones pueden provocar interacciones no favorables en el sistema como superposición de átomos o cargas del mismo signo muy cerca entre ellas. Por lo tanto, se realiza una minimización de energía nuevamente utilizando los mismos parámetros antes usados (Sección 7.2.2) pero ahora utilizando condiciones de borde periódicas.

7.2.6. Relajación del solvente y átomos de hidrógeno y aumento de la temperatura

Esta etapa corresponde a la primera etapa de simulación en donde se deja que el solvente se acomode al soluto mientras se aumenta la temperatura gradualmente. Con gran parte de la tensión del sistema disipada mediante la minimización de energía anterior, es necesario dejar que el solvente se adapte a la proteína, es decir, permitir que el solvente se mueva libremente mientras los átomos de la proteína que no sean

7.2. PROCEDIMIENTO PARA LA SIMULACIÓN Y IDENTIFICACIÓN DE MUTANTES

hidrógenos quedan con movimiento restringido.

Para integrar las ecuaciones de movimiento de Newton se utilizó el algoritmo leap-frog con un paso de $2fs$ y un total de 27500 pasos para completar $55ps$ de simulación [4,20,75]. Para el tratamiento de las interacciones electrostáticas se utilizó PME (Particle-Mesh Ewald) luego de un radio $1nm$ con un espaciado de $0.12nm$ y un orden de interpolación de 4. Para las interacciones de tipo van der Waals se consideraron sólo dentro de un radio de $1nm$. El aumento de la temperatura se realizó gradualmente utilizando simulated annealing simple aumentando la temperatura en $30^{\circ}K$ cada $5ps$, salvo los últimos $5ps$ en donde se mantuvo la temperatura constante para estabilizar el sistema. Se generaron las velocidades para cada átomo según una distribución de Maxwell. Los enlaces fueron tratados como limitados, de largo fijo. Para las temperaturas 350 y $400K$ se partió de 300 y $350K$ respectivamente y se aumentó la temperatura gradualmente $5K$ cada $5ps$ y los últimos $5ps$ se mantuvo la temperatura final constante.

7.2.7. Control de la Temperatura

El sistema está ahora relativamente libre de tensión. En esta etapa se pone al sistema bajo un baño térmico. Se utiliza un *ensemble* NVT, es decir, se mantiene el número de partículas, volumen y temperatura constantes, así, se corre una corta simulación en donde el *temperature coupling* será encendido y se le permitirá al sistema adaptarse a estas nuevas condiciones.

Esta etapa consistió en una simulación de 27500 pasos de $2fs$ y algoritmo leap-frog para la integración de las ecuaciones de Newton. Las interacciones electrostáticas y de van der Waals se trataron de igual manera que en la Sección 7.3.1. La temperatura se controló mediante un reescalamiento de velocidades, con una temperatura

7.2. PROCEDIMIENTO PARA LA SIMULACIÓN Y IDENTIFICACIÓN DE MUTANTES

de referencia de $300K$, 350 y $400K$ según cada simulación y un τ_t igual a $0.1ps$. Los parámetros utilizados en esta etapa se encuentran detallados en la sección A.5.

7.2.8. Control de la Presión

En esta etapa se realiza una simulación utilizando *pressure coupling*. Se reasignan nuevas velocidades a las partículas. La simulación consta de 500000 pasos de $2fs$ y los parámetros para las interacciones de van der Waals, electrostáticas y el control de temperatura son idénticas a las de la sección anterior. Para el manejo de la presión se utilizó un barostato de Berendsen isotrópico con τ_p igual a $1ps$, presión de referencia de $1bar$ y compresibilidad del agua de $4.5 \times 10^{-5}bar^{-1}$. El detalle de los parámetros utilizados en esta etapa se encuentran en la sección A.5

7.2.9. Producción y Equilibrio

Una vez realizados los pasos de preparación para la simulación se lleva a cabo la producción que consiste en una simulación con parámetros similares a los de la etapa de equilibrio sin embargo alguno de ellos son diferentes con el objetivo de que la simulación sea lo más fiel posible a lo que sucede en la realidad, dentro de las capacidades informáticas de las que se dispone.

La etapa de producción fue una simulación de $11ns$ usando el algoritmo leap-frog para resolver las ecuaciones de movimiento de Newton y paso de $2fs$ [62,65,139,143]. Las condiciones de borde periódicas, interacciones de tipo electrostáticas y de Van der Waals, temperatura y presión se consideraron de igual manera que en la sección anterior. Las restricciones para los enlaces se trataron mediante el algoritmo LINCS de orden 4. Los parámetros utilizados en la etapa de producción se detallan en la sección A.5.

El equilibrio se determinó calculando el RMSD en la producción para identificar desde qué tiempo es posible considerar el sistema como equilibrado. Finalmente, para todas las simulaciones se consideró como etapa de equilibrio el primer ns de la etapa de producción.

7.3. Validación

Para la validación del procedimiento se trabajó con dos enzimas, celobiohidrolasa Cel7A de *Talaromyces emersonii* (denominada enzima T) y celobiohidrolasa Cel7B de *Melanocarpus albomyces* (denominada enzima M). Esta validación consistió en simular el comportamiento de ambas estructuras en un sistema líquido utilizando agua como solvente a tres diferentes temperaturas 300, 350 y 400°K. Como se mencionó en la sección 3, se utilizan estas enzimas en la validación debido a su similitud estructural con Cel72 (Sección 7.4.1) y a que poseen antecedentes de mutaciones que mejoran su termoestabilidad [141,142]. Con el objetivo de ver la evolución del sistema a medida que se prepara para la simulación final, se muestran en cada etapa los datos de temperatura, presión y energía del sistema.

7.3.1. Simulación de las enzimas nativas T y M

Adición de Solvente e Iones

Para la inicialización del sistema se agregaron a la celda 13430 moléculas de agua para la enzima T y 15388 moléculas para la enzima M. La concentración de iones utilizada fue de 0.15M de NaCl. Así como se debe neutralizar el sistema una de las especies se encuentra en exceso respecto de la otra. Se agregaron 73 átomos de Na⁺ y 42 átomos de Cl⁻ para la enzima T y 76 átomos de Na⁺ y 47 de Cl⁻ para la enzima

M.

Minimización de energía del sistema solvado

El registro de salida para la minimización de energía del sistema solvado para las enzimas T y M se encuentra en la sección A.5. Las estructuras obtenidas de las simulaciones a $300K$ y $350K$ fueron utilizadas para las simulaciones a 350 y $400K$, respectivamente.

Relajación del solvente y átomos de hidrógeno y aumento de la temperatura

Se midió la temperatura, presión, energía cinética, potencial y total para ambas enzimas (Figuras 7.5 y 7.6) mediante el módulo de GROMACS `g_energy`. Es posible notar como el aumento gradual de la temperatura entrega energía al sistema.

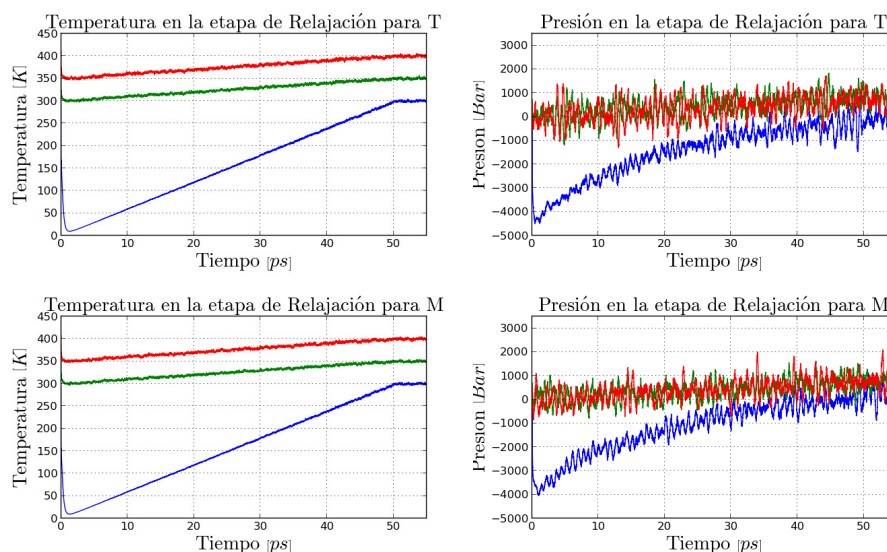


Figura 7.5: Temperatura y presión del sistema en función del tiempo de relajación. Se aumentó la temperatura de 0 a $300K$ (*azul*), de 300 a $350K$ (*verde*), y de 350 a $400K$ (*rojo*). En los últimos $5ps$ de cada relajación la temperatura de referencia se dejó constante.

Se puede ver que los últimos $5ps$ el sistema para ambas enzimas se estabiliza, es

decir, las presiones, temperaturas y energías registradas se mantienen constantes en torno a los parámetros de referencia. Es esperable que el sistema posea temperatura nula en el instante que comienza la simulación, para la primera relajación. Para las dos posteriores relajaciones el fenómeno observado es el mismo, la temperatura al comienzo de la etapa es distinta de la temperatura de referencia al instante inicial pero rápidamente comienza a ajustarse al proceso de aumento de temperatura establecido. Sin embargo se observa que el sistema comienza la simulación con temperatura distinta de cero, para ambas enzimas. Esto sucede dado que el programa automáticamente asigna velocidades a los átomos pero dado que se utiliza un calentamiento controlado del sistema el problema se corrige rápidamente de modo de lograr que el sistema alcance el estado que se desea.

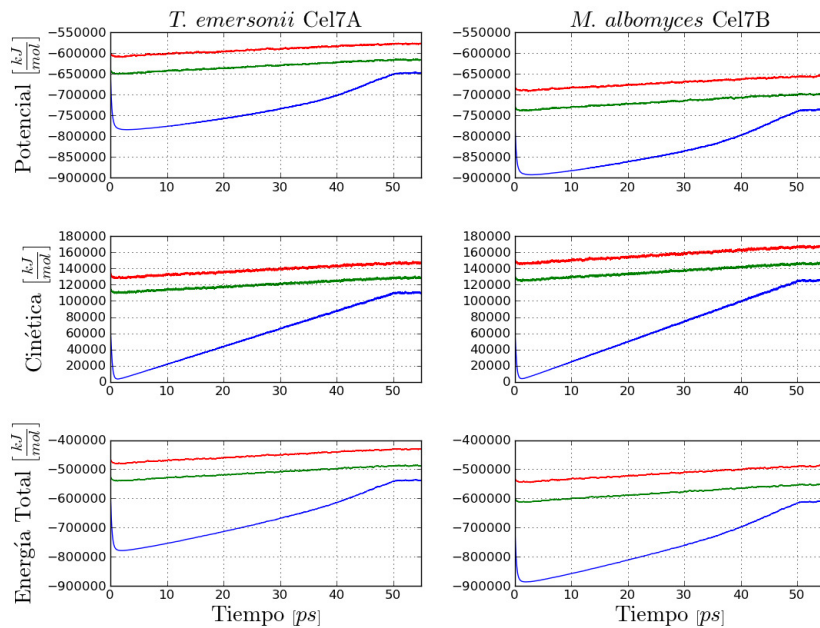


Figura 7.6: Energía del sistema durante la etapa de relajación para las enzimas T y M. Se muestra la energía potencial, cinética y total a las tres temperaturas estudiadas (azul: $300K$, verde: $350K$ y rojo: $400K$) en función del tiempo de relajación.

Lo observado anteriormente con la temperatura se observa también con las otras propiedades medidas en esta etapa (energía total, potencial y cinética, Fig. 7.6).

Control de la Temperatura

Se midieron las propiedades temperatura y presión con el módulo `g_energy` para las enzimas T y M; los resultados se muestran en la Figura 7.7. Se obtuvo que la temperatura y presión del sistema con temperatura de referencia $300K$, $350K$ y $400K$ fue de $299.856 \pm 1.423 [K]$, $349.824 \pm 1.696 [K]$ y $399.803 \pm 1.930 [K]$ para la enzima T, respectivamente; y de $299.876 \pm 1.341 [K]$, $349.837 \pm 1.595 [K]$ y $399.817 \pm 1.743 [K]$ para la enzima M, respectivamente.

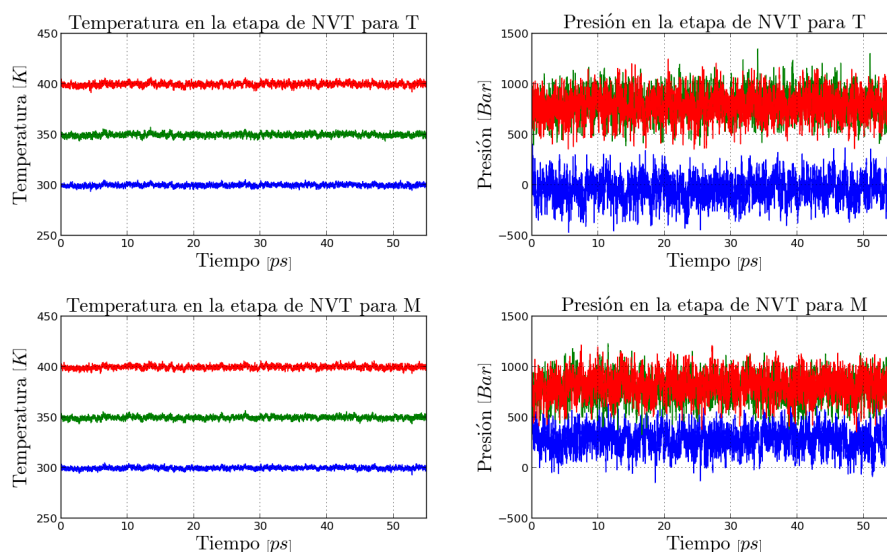


Figura 7.7: Temperatura y presión del sistema en la etapa NVT para las enzimas T y M (azul: $300K$, verde: $350K$ y rojo: $400K$).

Lo graficado en la Figura 7.7 muestra la constancia de la temperatura y se puede notar una diferencia de su comportamiento respecto de la temperatura en la etapa de relajación (Figura 7.5) producto de la aplicación de un termostato. Se registró

además, la energía del sistema durante la etapa NVT. En la Figura 7.8 se muestran los gráficos obtenidos del cálculo de la energía del sistema. Se midió la energía potencial, cinética y total.

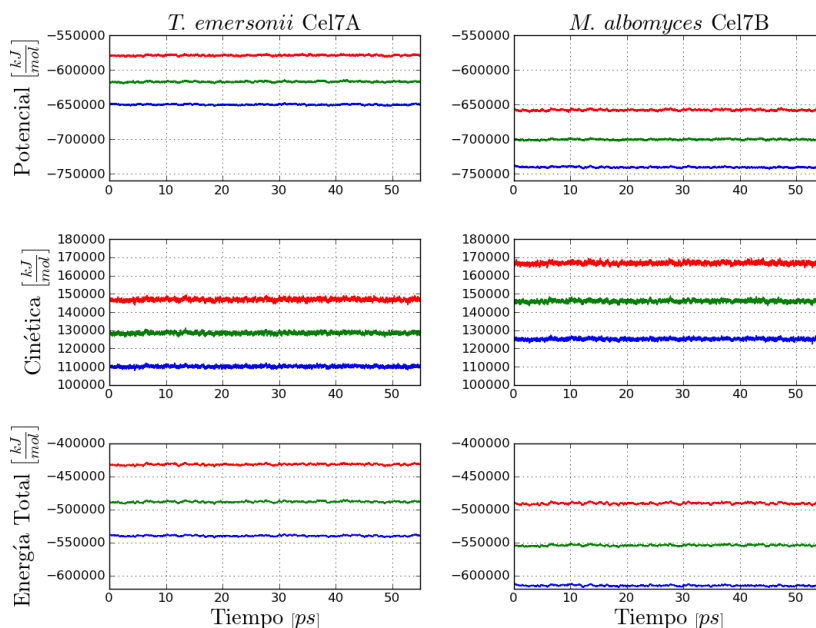


Figura 7.8: Energía medida en la etapa NVT para las enzimas T y M (azul: 300K, verde: 350K y rojo: 400K).

En la Tabla 7.2 se resumen los cálculos de las propiedades del sistema en función del tiempo de simulación para las enzimas T y M, respectivamente. Se muestra el Promedio, error de estimación y RMSD. El error de estimación del promedio lo calcula `q_energy` en base al promedio de bloques de 5 bloques.

En esta etapa, debido a la introducción del control de la temperatura es posible notar la constancia de la temperatura en torno a las temperaturas de referencia. Esto es posible verlo para ambas enzimas. El registro de salida la etapa de Temperature coupling se encuentra en la sección A.5.

Tabla 7.2: Temperatura, presión y energía del sistema para las enzimas T y M en la etapa NVT

Enzima	Temp. de referencia		Promedio	Err. Est.	RMSD	Unidades
T	300K	Temperatura	299,86	0,088	1,42335	[K]
		Presión	-39,3077	5,5	118,006	[bar]
		Potencial	-650030	67	790,455	[kJ/mol]
		Cinética	110080	32	522,526	[kJ/mol]
		Energía Total	-539949	67	927,593	[kJ/mol]
	350K	Temperatura	349,82	0,11	1,69616	[K]
		Presión	805,49	8,6	127,477	[bar]
		Potencial	-617067	130	821,261	[kJ/mol]
		Cinética	128459	40	622,847	[kJ/mol]
		Energía Total	-488608	170	1071,9	[kJ/mol]
	400K	Temperatura	399,80	0,12	1,92992	[K]
		Presión	783,986	2,9	125,354	[bar]
		Potencial	-579000	100	854,244	[kJ/mol]
		Cinética	146812	46	708,686	[kJ/mol]
		Energía Total	-432188	150	1104,34	[kJ/mol]
M	300K	Temperatura	299,88	0,076	1,34089	[K]
		Presión	282,58	5,1	112,386	[bar]
		Potencial	-740294	140	785,543	[kJ/mol]
		Cinética	125146	32	559,59	[kJ/mol]
		Energía Total	-615148	140	931,868	[kJ/mol]
	350K	Temperatura	349,84	0,095	1,5946	[K]
		Presión	783,248	4,6	118,219	[bar]
		Potencial	-700317	96	797,23	[kJ/mol]
		Cinética	145996	40	665,47	[kJ/mol]
		Energía Total	-554321	130	1029,94	[kJ/mol]
	400K	Temperatura	399,82	0,11	1,7433	[K]
		Presión	816,196	3,4	118,035	[bar]
		Potencial	-657800	110	1007,53	[kJ/mol]
		Cinética	166854	46	727,524	[kJ/mol]
		Energía Total	-490946	150	1279,76	[kJ/mol]

Control de la Presión

Se registró la temperatura y presión en función del tiempo de simulación durante la etapa NPT a las tres temperaturas para las enzimas T y M. Los datos obtenidos se muestran en la Figura 7.9.

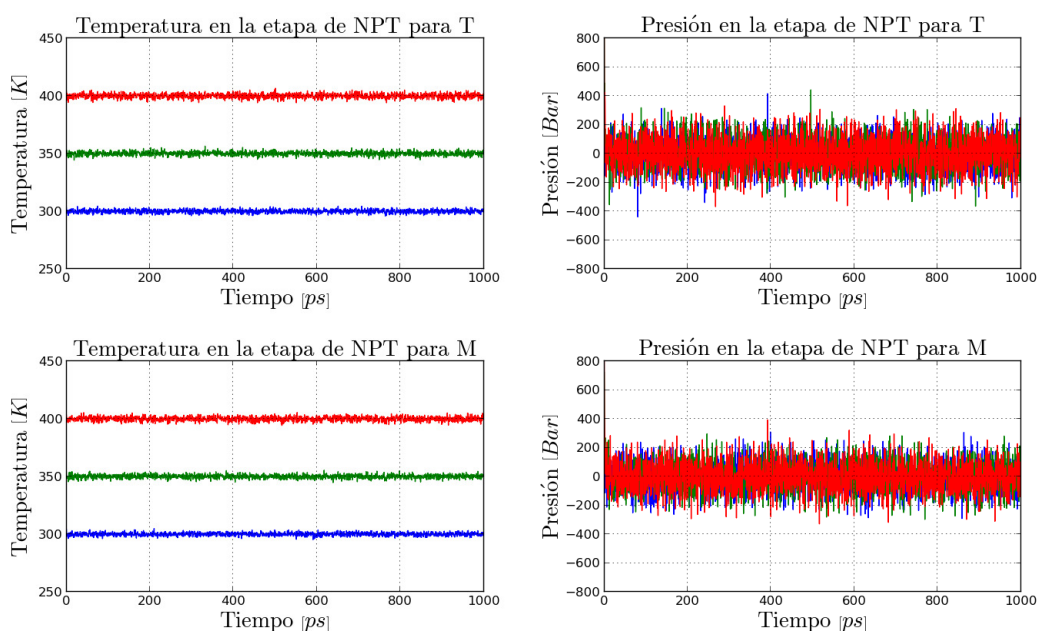


Figura 7.9: Temperatura y presión del sistema en la etapa NPT para las enzimas T y M (azul: 300K, verde: 350K y rojo: 400K).

Se registró, además, la energía del sistema durante a lo largo de la simulación a las tres temperaturas estudiadas. En la Figura 7.10 se grafican los datos de energía cinética, potencial y total calculados.

Se puede ver que la energía del sistema tiene una tendencia constante debido a que la temperatura, que es lo que aporta introduciendo energía al sistema, se mantiene constante a partir de esta etapa. Es por esto que se espera que la energía permanezca con la misma tendencia en los pasos siguientes, posiblemente variando su amplitud

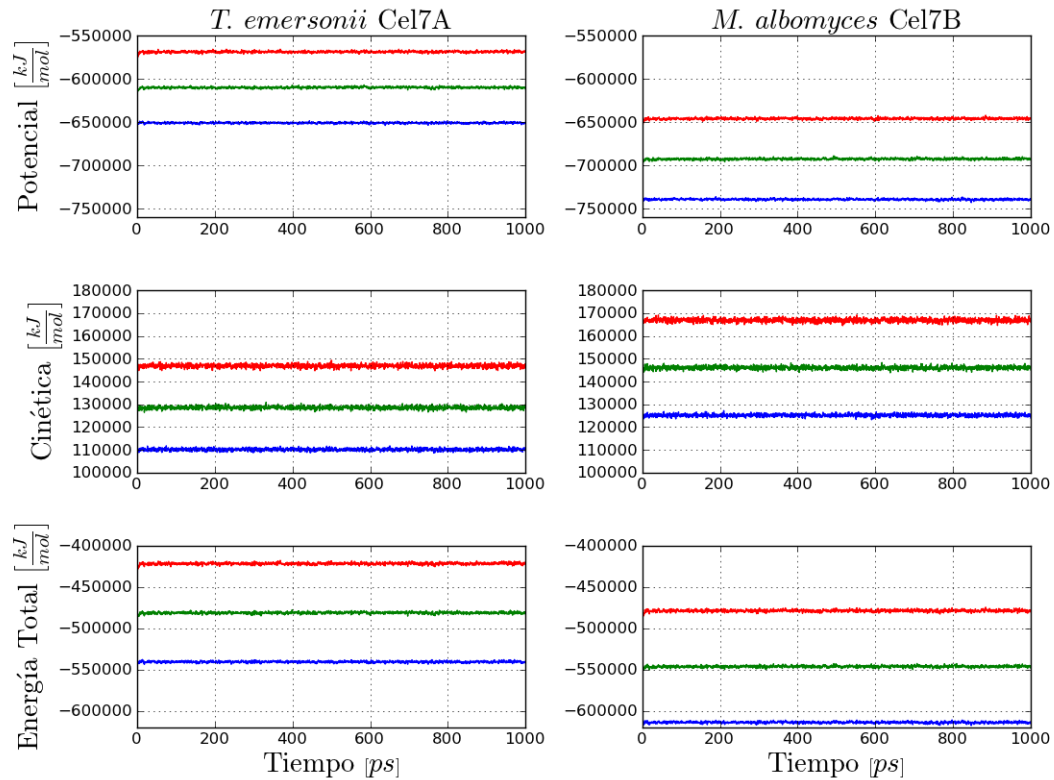


Figura 7.10: Energía medida en la etapa NPT para las enzimas T y M (azul: 300K, verde: 350K y rojo: 400K).

debido al control de la presión en las etapas NPT y de producción. En la Tabla 7.3 se resumen los datos obtenidos para la energía, presión y temperatura a las diferentes temperaturas para las dos enzimas, T y M.

Al introducir control de presión en esta etapa, la presión del sistema, tanto para la enzima T como para la enzima M, permanece estable en torno a al valor de referencia, sin embargo, es posible notar una gran fluctuación de valores de presión. Esto es normal y está informado por los desarrolladores de GROMACS [61].

Tabla 7.3: Temperatura, presión y energía del sistema de T y M en la etapa NPT

Enzima	Temp. de referencia		Promedio	Err. Est.	RMSD	Unidades
T	300K	Temperatura	299,98	0,013	1,41542	[K]
		Presión	0,9802	0,09	103,037	[bar]
		Potencial	-650816	17	763,399	[kJ/mol]
		Cinética	110125	4,9	519,616	[kJ/mol]
		Energía Total	-540691	20	927,319	[kJ/mol]
	350K	Temperatura	349,97	0,017	1,63982	[K]
		Presión	2,038	0,99	107,567	[bar]
		Potencial	-609940	79	895,223	[kJ/mol]
		Cinética	128514	6,1	602,159	[kJ/mol]
		Energía Total	-481426	81	1092,77	[kJ/mol]
	400K	Temperatura	399,97	0,02	1,88285	[K]
		Presión	2,299	1,2	110,366	[bar]
		Potencial	-568851	73	1002,64	[kJ/mol]
		Cinética	146873	7,2	691,403	[kJ/mol]
		Energía Total	-421978	77	1228,68	[kJ/mol]
M	300K	Temperatura	299,98	0,013	1,32502	[K]
		Presión	1,335	0,3	95,8929	[bar]
		Potencial	-738996	29	800,87	[kJ/mol]
		Cinética	125190	5,2	552,965	[kJ/mol]
		Energía Total	-613806	27	975,316	[kJ/mol]
	350K	Temperatura	349,98	0,015	1,54201	[K]
		Presión	1,951	0,86	100,565	[bar]
		Potencial	-692389	78	931,843	[kJ/mol]
		Cinética	146054	6,3	643,52	[kJ/mol]
		Energía Total	-546334	83	1141,33	[kJ/mol]
	400K	Temperatura	399,97	0,019	1,76165	[K]
		Presión	2,264	1,3	102,227	[bar]
		Potencial	-645872	56	1073,9	[kJ/mol]
		Cinética	166919	7,7	735,184	[kJ/mol]
		Energía Total	-478953	62	1312,82	[kJ/mol]

Producción

Se midieron, para ambas enzimas, la temperatura, presión y energía con el módulo `g_energy` de GROMACS, con el fin de ver la evolución del sistema a lo largo de la producción. En las Figuras 7.11 y 7.12 se grafican los resultados obtenidos de los cálculos mencionados. Es posible notar que tanto la temperatura como la presión del sistema se mantienen en torno a los valores de referencia, y que la presión fluctúa con gran amplitud, tal como se mencionó anteriormente para la etapa de control de presión.

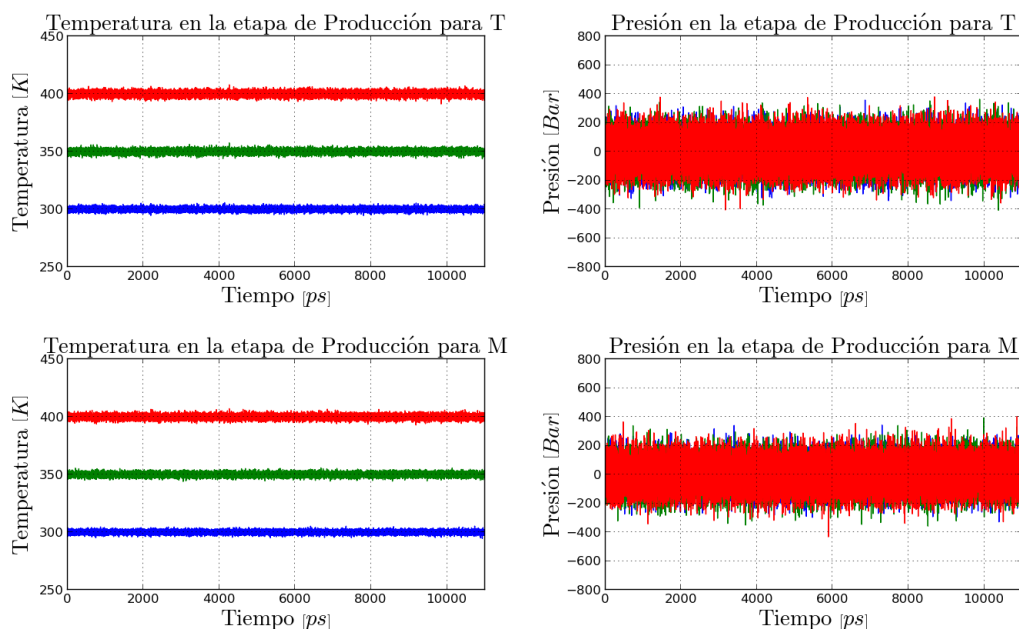


Figura 7.11: Temperatura y presión del sistema en la etapa de producción para las enzimas T y M (azul: $300K$, verde: $350K$ y rojo: $400K$).

La energía se muestra constante y coherente con las diferencias de temperatura para cada simulación, es decir, a mayor temperatura, mayor es la energía cinética del sistema. El archivo de registro de salida luego de realizar esta etapa se muestra

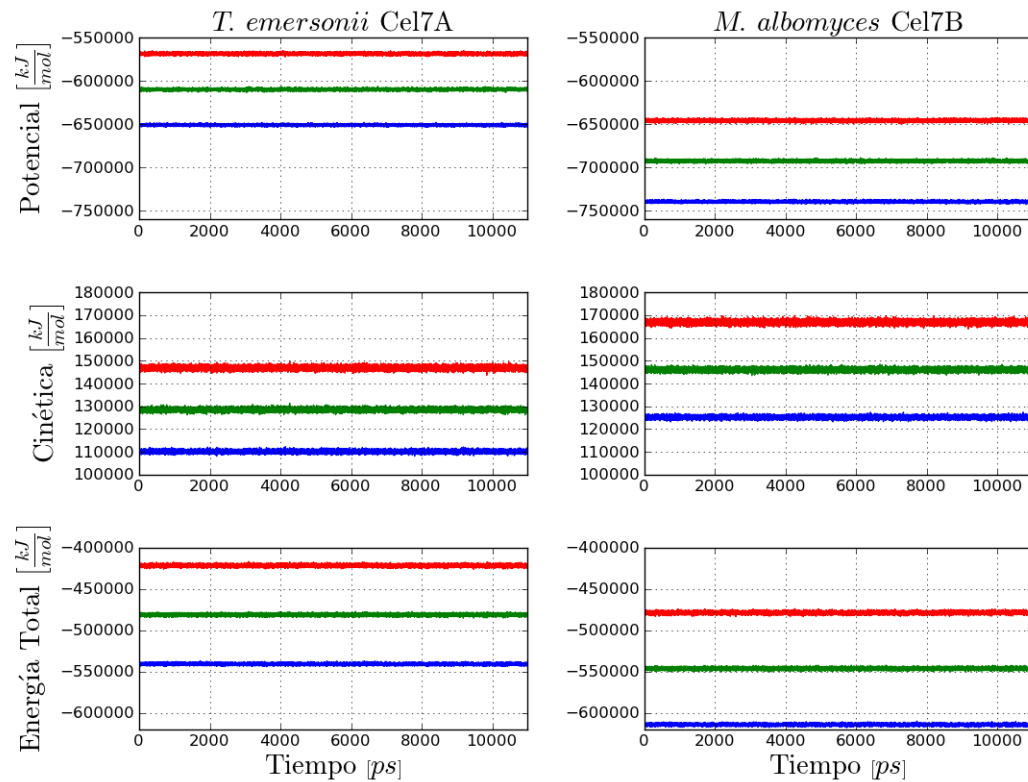


Figura 7.12: Energía del sistema en la etapa de producción (azul: 300K, verde: 350K y rojo: 400K).

en la sección A.5.

En la Tabla 7.4 se resumen los resultados de estos cálculos para ambas enzimas a las tres temperaturas estudiadas.

Tabla 7.4: Propiedades del sistema medidas en la etapa de producción para las enzimas T y M.

Enzima	Temp. de referencia		Promedio	Err. Est.	RMSD	Unidades
T	300K	Temperatura	299,977	0,0059	1,40926	[K]
		Presión	1,01471	0,0083	93,6245	[bar]
		Potencial	-650929	32	773,517	[kJ/mol]
		Cinética	110125	2,2	517,355	[kJ/mol]
		Energía Total	-540805	32	940,404	[kJ/mol]
	350K	Temperatura	349,973	0,0069	1,64315	[K]
		Presión	1,01244	0,0043	97,981	[bar]
		Potencial	-609899	24	866,78	[kJ/mol]
		Cinética	128514	2,5	603,382	[kJ/mol]
		Energía Total	-481386	27	1071,4	[kJ/mol]
	400K	Temperatura	399,969	0,0079	1,87882	[K]
		Presión	1,00317	0,011	100,238	[bar]
		Potencial	-568729	30	958,15	[kJ/mol]
		Cinética	146873	2,9	689,922	[kJ/mol]
		Energía Total	-421856	31	1197,7	[kJ/mol]
M	300K	Temperatura	299,98	0,0055	1,32119	[K]
		Presión	1,00496	0,0042	86,4807	[bar]
		Potencial	-739287	28	819,501	[kJ/mol]
		Cinética	125189	2,3	551,369	[kJ/mol]
		Energía Total	-614098	27	996,921	[kJ/mol]
	350K	Temperatura	349,976	0,0064	1,54099	[K]
		Presión	1,00411	0,0035	90,7725	[bar]
		Potencial	-692396	22	915,183	[kJ/mol]
		Cinética	146054	2,7	643,097	[kJ/mol]
		Energía Total	-546342	24	1131,3	[kJ/mol]
	400K	Temperatura	399,972	0,0073	1,76028	[K]
		Presión	1,00764	0,0035	92,3656	[bar]
		Potencial	-645765	50	1021,34	[kJ/mol]
		Cinética	166919	3,1	734,613	[kJ/mol]
		Energía Total	-478846	52	1276,18	[kJ/mol]

Con el fin de observar la estabilidad del sistema se determinó el RMSD de la cadena principal (C-alpha) de la enzima a lo largo de la simulación utilizando como referencia de la estructura minimizada respectiva para ambas enzimas y la distribución de los RMSD para cada enzima a cada temperatura simulada (Figura 7.13). Se considera el primer nanosegundo como etapa de equilibrio. Así, el RMSD obtenido para cada enzima a temperatura de referencia $300K$, $350K$ y $400K$ fue de $0.218 \pm 0.0368 [nm]$, $0.244 \pm 0.0269 [nm]$ y $0.410 \pm 0.0587 [nm]$ para la enzima T, respectivamente; y de $0.196 \pm 0.0158 [nm]$, $0.242 \pm 0.0365 [nm]$ y $0.315 \pm 0.0398 [nm]$ para la enzima M, respectivamente.

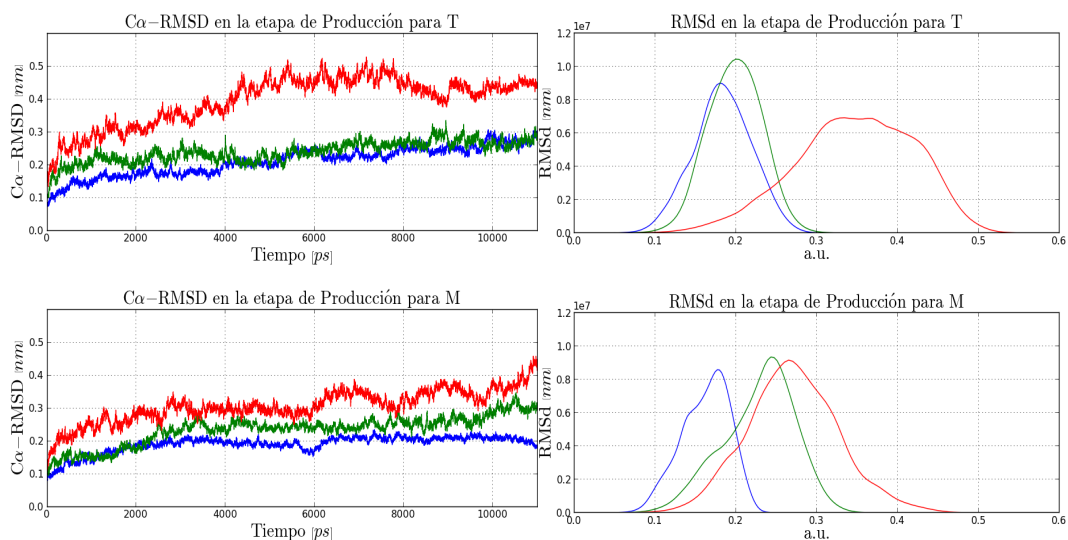


Figura 7.13: $C\alpha$ -RMSD y distribución de $C\alpha$ -RMSD de la enzima T y la enzima M con respecto a la estructura inicial en función del tiempo en la etapa de producción (azul: $300K$, verde: $350K$ y rojo: $400K$).

Como medida de compactación se calculó el Rg (Radius of gyration) en función del tiempo (Figura 7.14).

Para ver si los Rg a cada temperatura son significativamente diferentes se realizó un test de hipótesis t-Student. En la Tabla 7.5 se muestran los valores de los promedios y desviación estándar de los datos de Rg a cada temperatura para cada enzima.

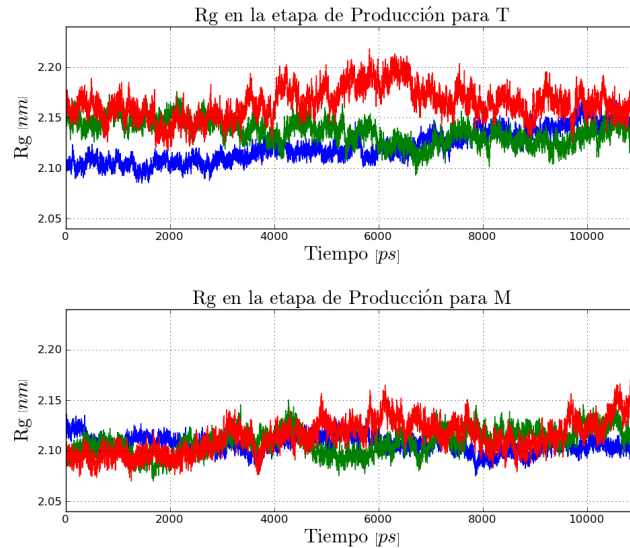


Figura 7.14: Rg de la cadena principal para la enzima T y la enzima M en la etapa de producción (azul: 300K, verde: 350K y rojo: 400K).

A base de esos valores se determinaron los t-value para cada caso. Se comparó, para cada enzima, el Rg a 300K versus 350 y 400K el Rg a 350K versus 400K. Como hipótesis nula se consideró que las medias del Rg en cada comparación son iguales. El t-value se calculó de la siguiente manera:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad (7.1)$$

donde n_1 y n_2 son el tamaño de la muestra para cada uno de los conjuntos de datos que se está analizando. \bar{x} y s son las medias y desviación estándar respectivas a cada conjunto de datos. Para 44002 grados de libertad y una confianza del 99,9% el valor de t tal que $p(|T| > t) = 0,001$ es 3,291, luego, en base a esto y los valores de t calculados (Tabla 7.6) se obtiene que se rechaza la hipótesis nula en todos los casos. De esta forma, se puede decir que la diferencia entre las medias comparadas es significativa (P-valor < 0,0001).

Tabla 7.5: Medias y desviaciones del Rg.

Enzima	300K		350K		400K	
	media	desv. est.	media	desv. est.	media	desv. est.
M	2,10659	0,00818	2,10898	0,01243	2,11388	0,01663
T	2,12196	0,01650	2,13624	0,01221	2,16389	0,01568

Tabla 7.6: Resultados del test t. Se muestra en la Tabla el t-value, intervalos de confianza de del promedio de la diferencia.

	t-value	Intervalos de confianza (95 %)			
		$\bar{x}_2 - \bar{x}_1$	Límite inferior	Límite superior	
T	300K-350K	103,1745	0,014279	0,014007	0,014551
	350K-400K	206,3378	0,027650	0,027387	0,027913
	300K-400K	273,1655	0,041929	0,041627	0,042230
M	300K-350K	23,8858	0,002396	0,002199	0,002594
	350K-400K	34,9941	0,004898	0,004623	0,005173
	300K-400K	58,3875	0,007294	0,007049	0,007540

Se determinó el RMS de las fluctuaciones de la posición de los átomos en la trayectoria para las enzimas T y la enzima M lo que se puede ver en la Figura 7.15.

Una visión general del comportamiento de las enzimas en la simulación cada 1.0ns de simulación se puede ver en las Figuras 7.16 y 7.17. Se utilizó el programa PyMol [1] para la visualización de la enzima. La elección de mostrar las estructuras cada 1.0ns se decidió de manera arbitraria.

Los archivos de registro de salida para las enzimas T y M se muestran en la sección A.5.

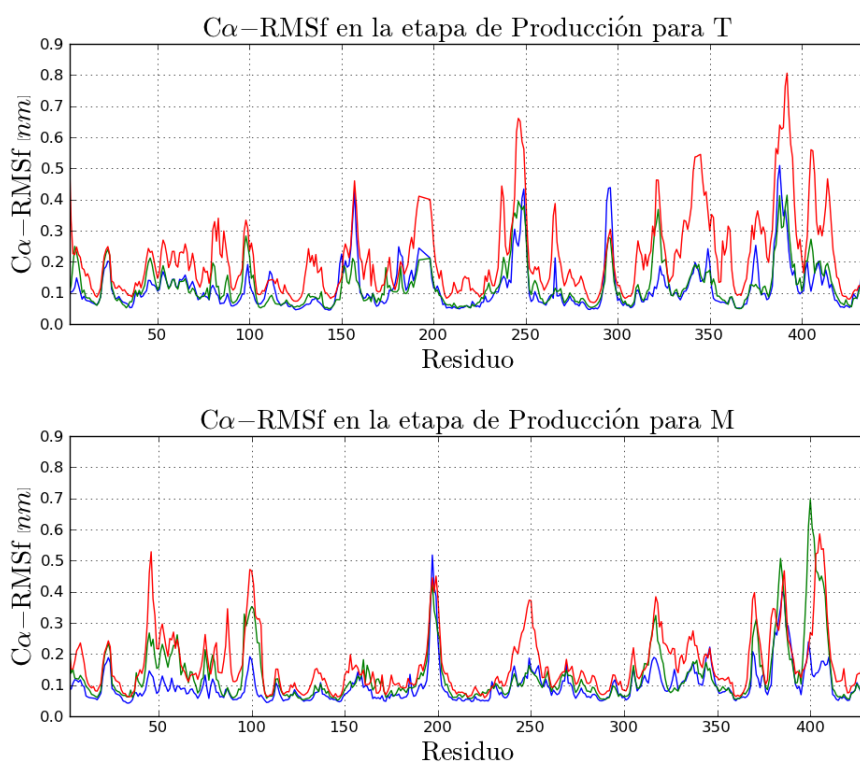


Figura 7.15: RMSf de cada residuo para las enzimas T y la enzima M (azul: 300K, verde: 350K y rojo: 400K).

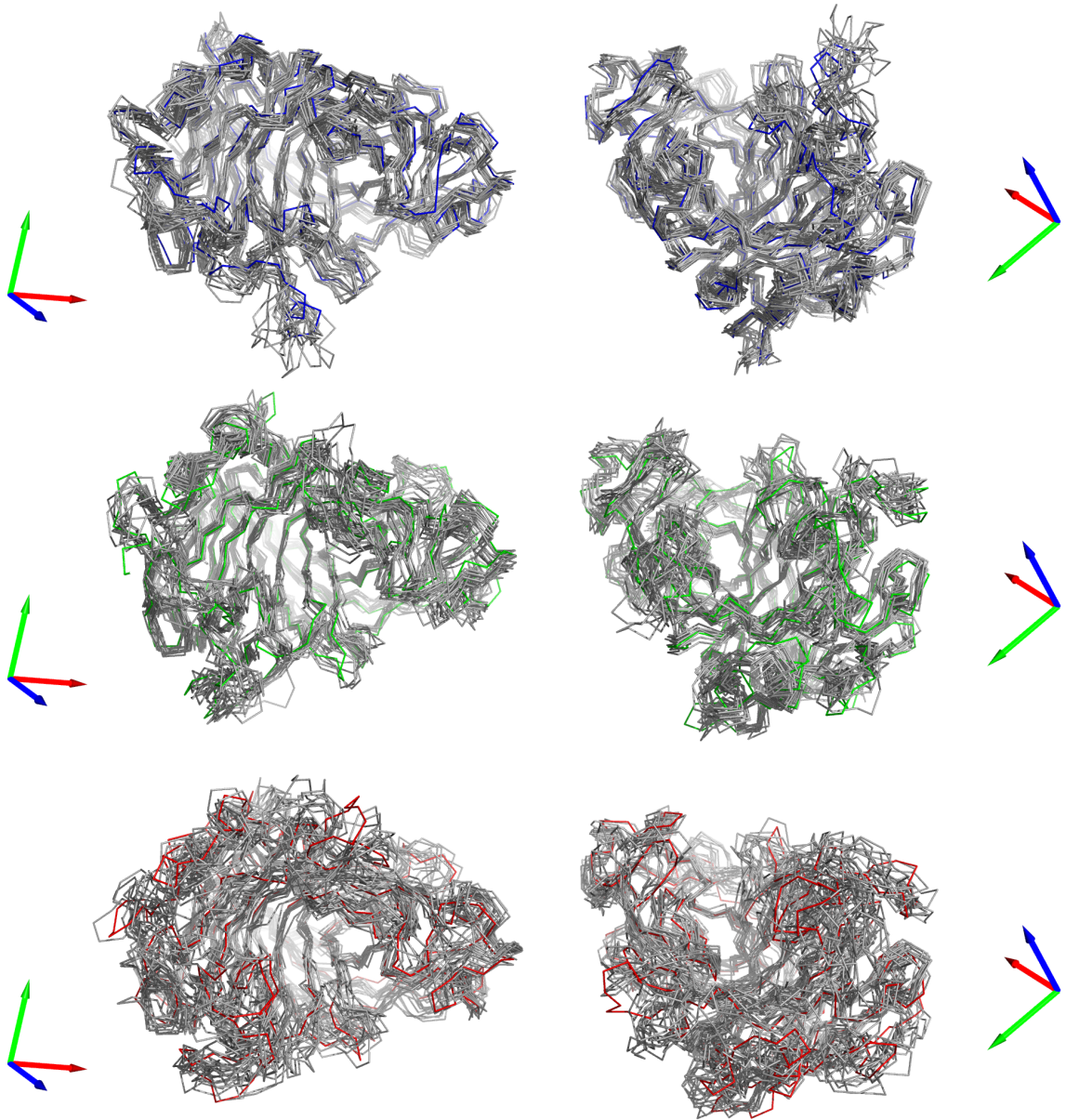


Figura 7.16: Visión general de la trayectoria de T a 300K (azul), 350K (verde) y 400K (rojo). Se muestran el esqueleto de las estructuras cada $1ns$ en 2 perspectivas diferentes. La elección de mostrar las estructuras cada $1.0ns$ se decidió de manera arbitraria. En color azul, verde o rojo se muestra la estructura inicial y en gris las estructuras a lo largo de la simulación.

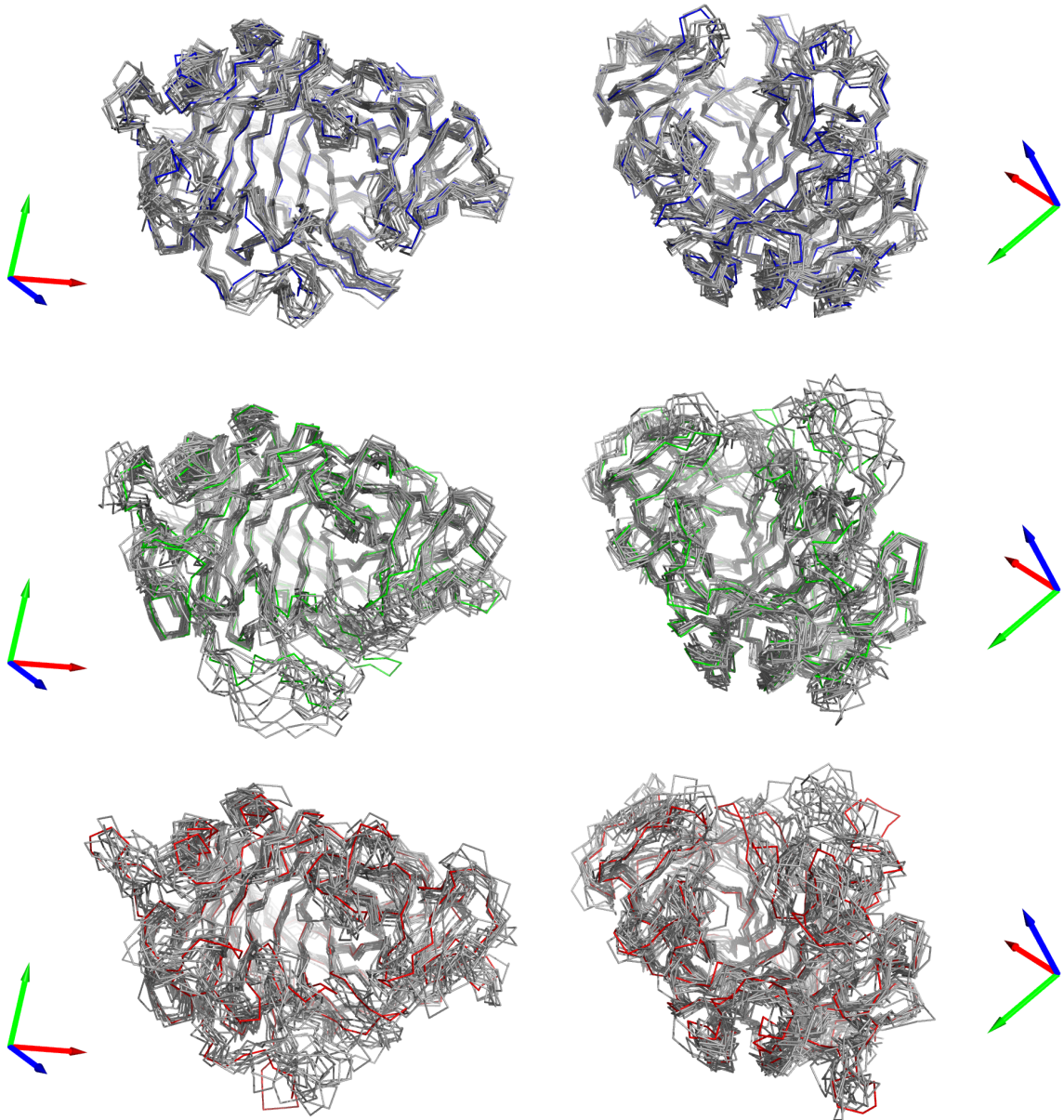


Figura 7.17: Visión general de la trayectoria de la enzima M a 300K (azul), 350K (verde) y 400K (rojo). Se muestran el esqueleto de las estructuras cada $1ns$ en 2 perspectivas diferentes. La elección de mostrar las estructuras cada $1.0ns$ se decidió de manera arbitraria. En color azul, verde o rojo se muestra la estructura inicial y en gris las estructuras a lo largo de la simulación.

7.3.2. Cálculo de Flexibilidad

Con el fin de identificar regiones de mayor flexibilidad en la enzimas modelo se procedió a determinar el valor f y if para residuo de cada enzima según los datos de trayectoria obtenidos de la etapa de producción considerando los últimos $10ns$ de simulación (Figura 7.18).

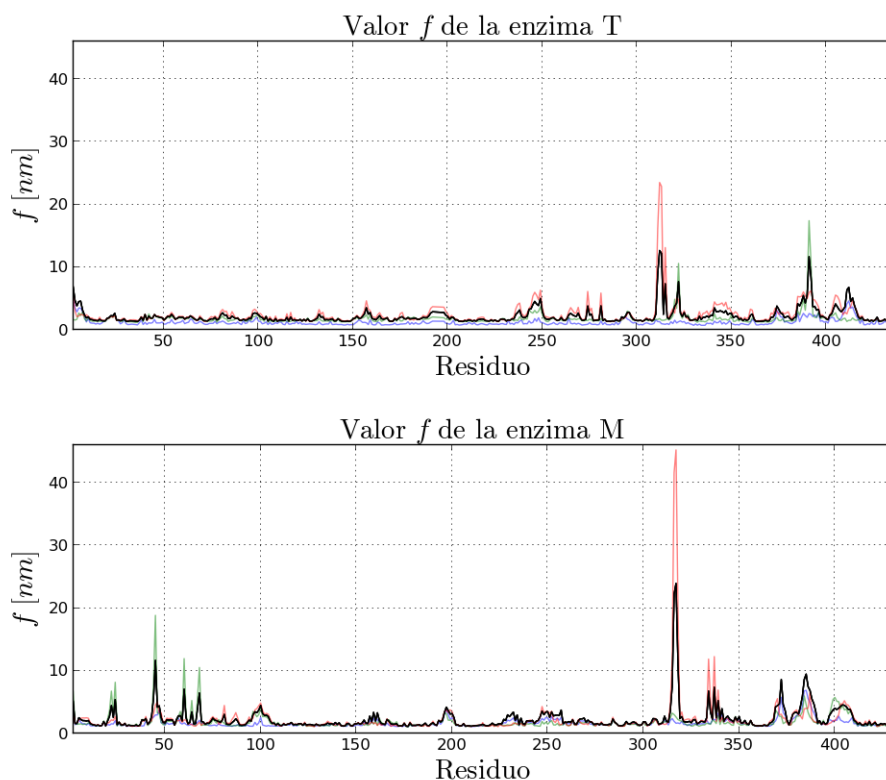


Figura 7.18: Flexibilidad local de las enzimas T y M por residuo. Se muestra en color negro la suma de entre el f a $300K$ y el promedio del módulo de la diferencia entre las otras dos temperaturas y la temperatura inicial (if). En color azul, verde y rojo se muestran los valores f determinados para cada enzima a las temperaturas 300 , 350 y $400K$ respectivamente.

A partir del ranking de if obtenido para cada enzima se definen las regiones mutables en base a la cercanía entre residuos destacados por su flexibilidad. De este

modo, se generan tres mutantes por cada enzima que permitan reducir el if en las regiones más flexibles de la enzima mediante puentes disulfuro. En la Tabla 7.7 se muestra el ranking de los aminoácidos de las enzimas T y M ordenados de acuerdo al valor de if determinado para las enzimas T y M. Pertenecen al ranking los residuos cuyos valores if son mayores que la media por enzima más su correspondiente desviación estándar¹.

Tabla 7.7: Ranking de valores if para las enzimas T y M. Se indican además las regiones a las cuales pertenecen los residuos.

N°	T			M		
	Aminoácido ^a	if [nm]	Región	Aminoácido ^a	if [nm]	Región
1	312	12,651	A	317	23,942	A
2	313	12,201	A	316	22,299	A
3	391	11,674	B	45	11,672	B
4	311	8,877	A	318	11,301	A
5	392	8,506	B	385	9,473	C
6	322	7,702	E	384	8,640	C
7	315	7,349	A	372	8,618	D
8	2	6,855	C	315	7,507	A
9	412	6,784	D	386	7,481	C
10	411	6,358	D	337	7,383	E
11	390	5,552	B	60	7,067	
12	388	5,473	B	387	6,873	C
13	414	5,103	D	334	6,785	E
14	249	5,005	F	68	6,452	
15	3	4,684	C	383	5,898	C
16	6	4,573	C	44	5,565	B
17	246	4,541	F	388	5,511	C
18	413	4,521	D	24	5,424	F
19	5	4,507	C	339	5,260	E
20	387	4,437	B	2	5,149	

Continúa en la página siguiente

¹Para cada enzima se calcula la media y la desviación estándar del valor if , luego estos dos valores se suman y su resultado se toma como límite inferior del rango de valores que pueden estar en el ranking, es decir, se excluyen del ranking todos los residuos con valores if menores que este límite.

^aÍndice relativo a la secuencia aminoacídica de la enzima madura.

Tabla 7.7 : Ranking de *if*. (Cont. de la página anterior).

N°	T			M		
	Aminoácido ^a	<i>if</i> [nm]	Región	Aminoácido ^a	<i>if</i> [nm]	Región
21	389	4,285	B	389	4,767	C
22	321	4,258	E	371	4,721	D
23	385	4,176	B	373	4,643	D
24	405	4,135	D	404	4,572	G
25	247	4,108	F	405	4,495	G
26	245	4,020	F	100	4,492	
27	435	3,899		22	4,468	F
28	415	3,884	D	370	4,437	D
29	248	3,878	F	47	4,423	B
30	4	3,831	C	430	4,412	
31	281	3,825		403	4,347	G
32	274	3,803		382	4,336	C
33	386	3,802	B			
34	374	3,790				
35	320	3,770	E			
36	394	3,749	B			
37	393	3,691	B			
38	410	3,663	D			
39	406	3,653	D			

Se definieron las regiones mutables en base a la cercanía entre residuos destacados por su flexibilidad. Los residuos que se encontraban dentro del ranking pero que no tenían residuos vecinos en la secuencia pertenecientes al ranking no se consideraron como parte de ninguna región. Así, se definieron 6 regiones mutables (regiones A a F de la Tabla 7.7) para la enzima T y 7 (regiones A a G de la Tabla 7.7) para la enzima M.

Se han descrito mutaciones que mejoran la termoestabilidad tanto para la enzima T [142] como para la enzima M [141]. Estas mutaciones son G4C-A72C, N54C-P191C, Q190C-I200C, T243C-A375C y G266C-D320C para la enzima T y corresponden todas a puentes disulfuros nuevos, y G4C-M70C, que es un nuevo puente disulfuro y

^aÍndice relativo a la secuencia aminoacídica de la enzima madura.

S290T para la enzima M. Para T, tres de estas mutaciones fueron recuperadas por el procedimiento aquí propuesto (regiones C, E y F recuperan las mutaciones G4C-A72C, G266C-D320C y T243C-A375C, respectivamente). La región F no recupera exactamente la mutación T243C-A375C sin embargo su cercanía con la mutación propuesta se consideró como suficiente para poder rigidizar esa zona. En las Figuras 7.19 y 7.20 se ilustran las regiones encontradas como de alta flexibilidad.

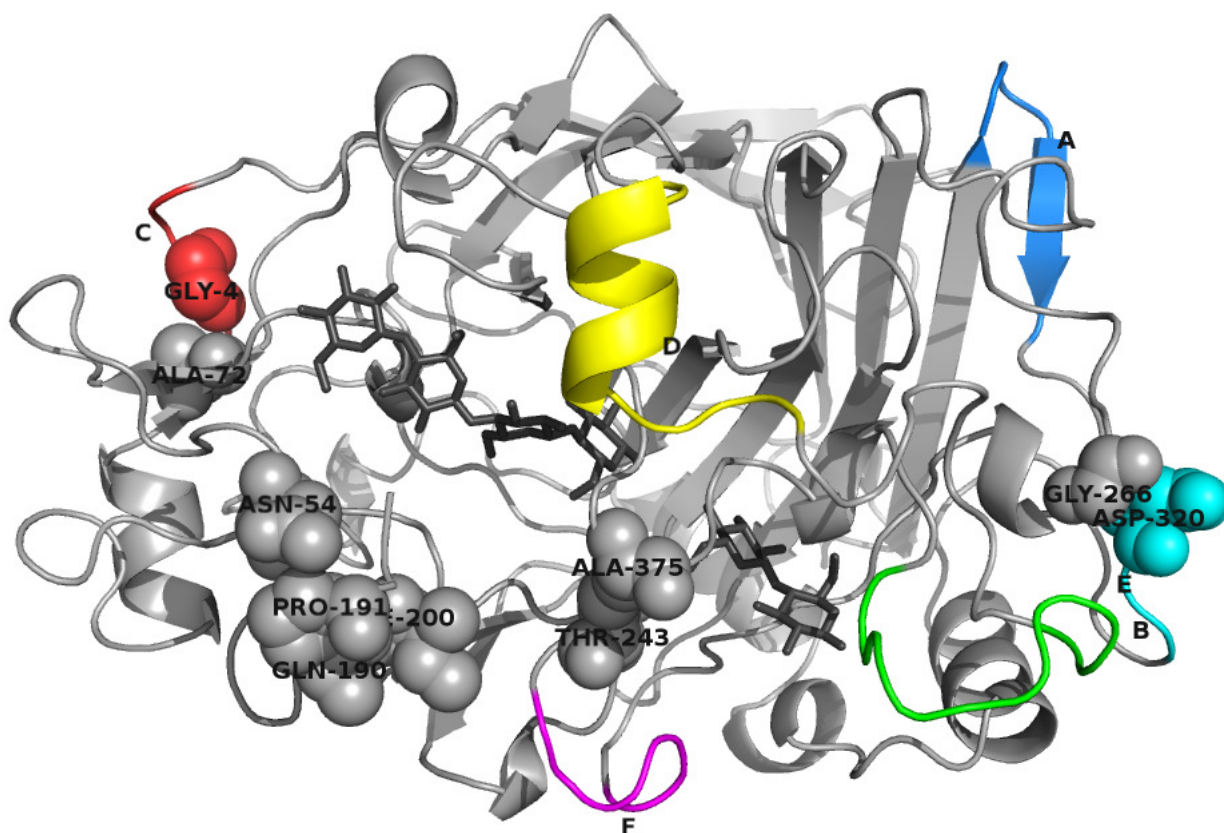


Figura 7.19: Regiones definidas como flexibles en base al análisis de valor *if* para la enzima T. En color azul se destacaron los residuos pertenecientes a la región A, en verde la B, rojo para C, amarillo para D, celeste para E y rosado para F. Los aminoácidos que aparece señalan son los mutados en la literatura [142].

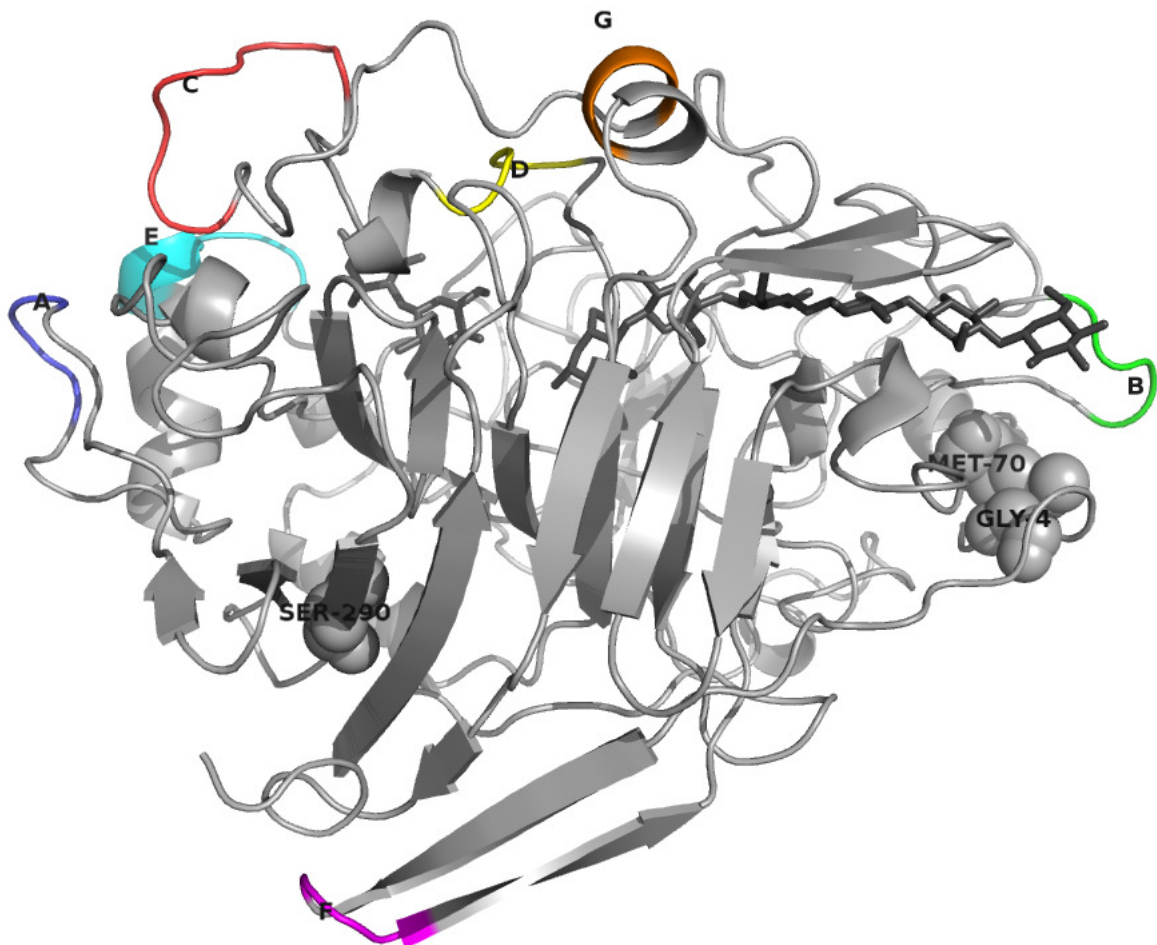


Figura 7.20: Regiones definidas como flexibles en base al análisis de valor *if* para la enzima M. En color azul se destacaron los residuos pertenecientes a la región A, en verde la B, rojo para C, amarillo para D, celeste para E, rosado para F y naranja para G. Los aminoácidos que aparece señalas son los mutados en la literatura [141].

A partir del análisis de flexibilidad realizado, se de lograron establecer regiones flexibles. Estas regiones fueron consideradas como regiones candidatas a rigidizar mediante puentes disulfuro. Para cada enzima se diseñaron 3 mutantes según los datos de flexibilidad que se obtuvieron. Además, se generaron las estructuras de las variantes reportadas de cada enzima, T y M. De esta forma se obtuvieron 8 enzimas para la T y 5 para la enzima M. En las Figuras 7.21 y 7.22 se muestran en diagrama de estructura tridimensional las mutantes generadas.

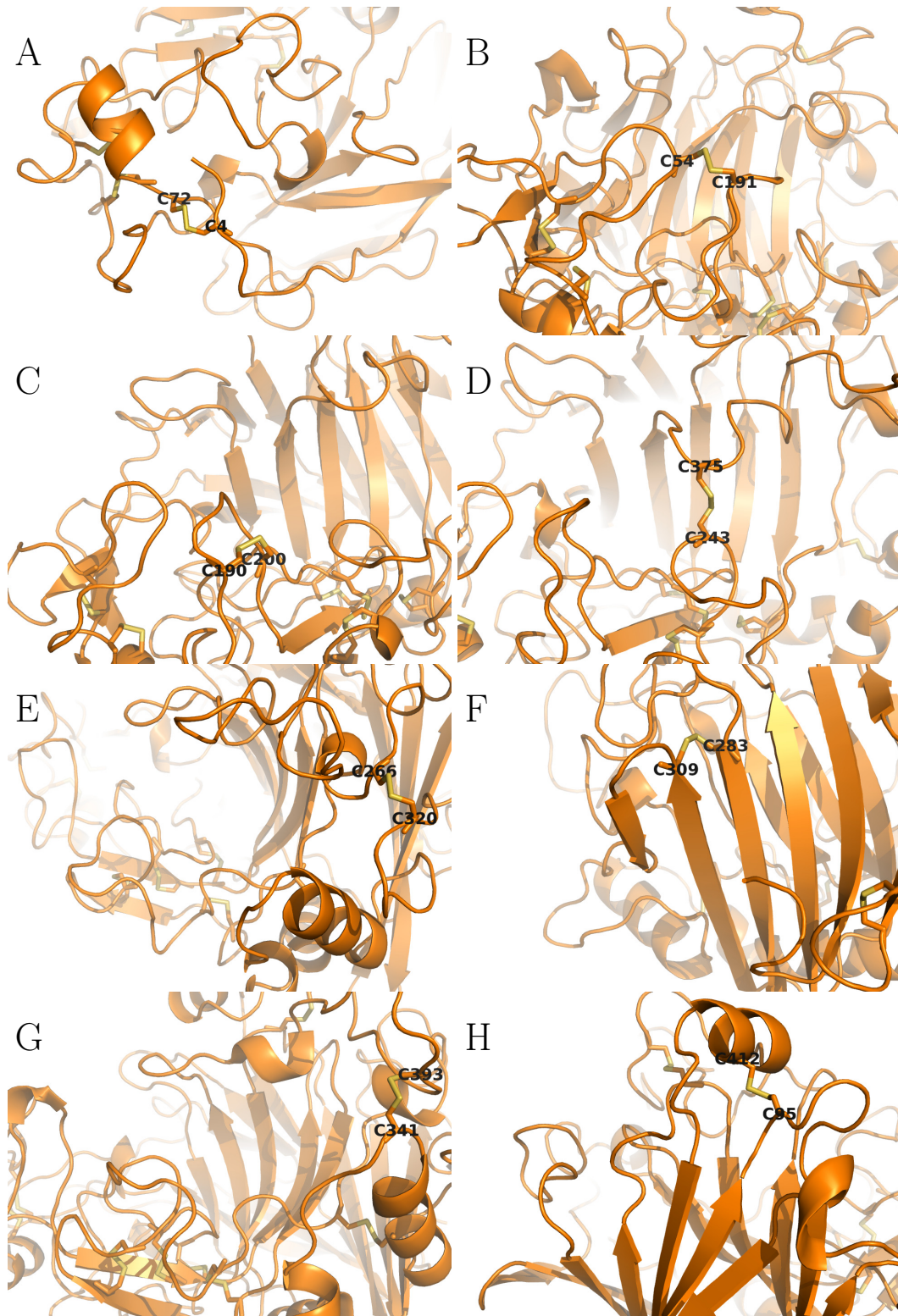


Figura 7.21: Mutaciones termoestabilizantes propuestas para la enzima T.

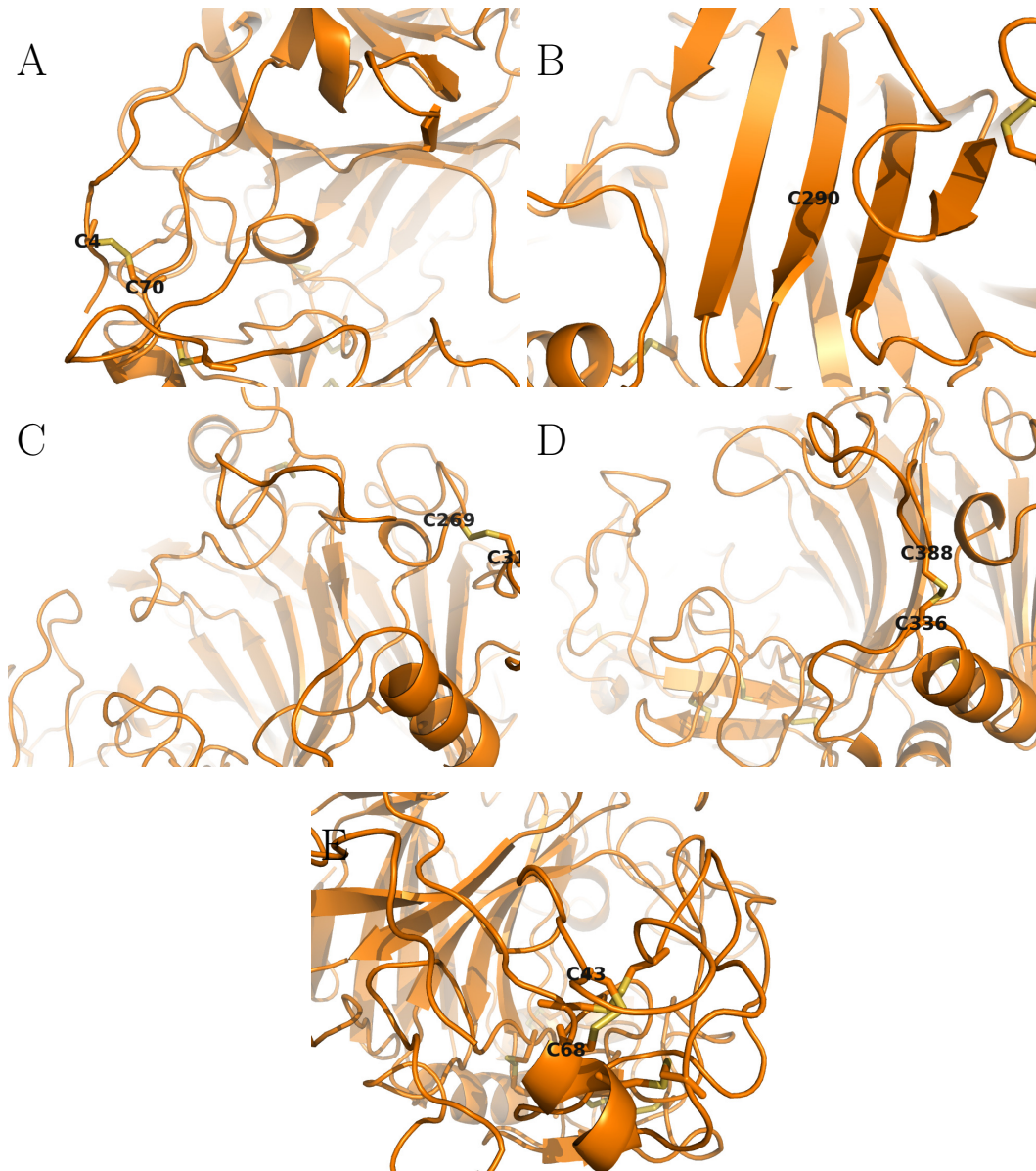
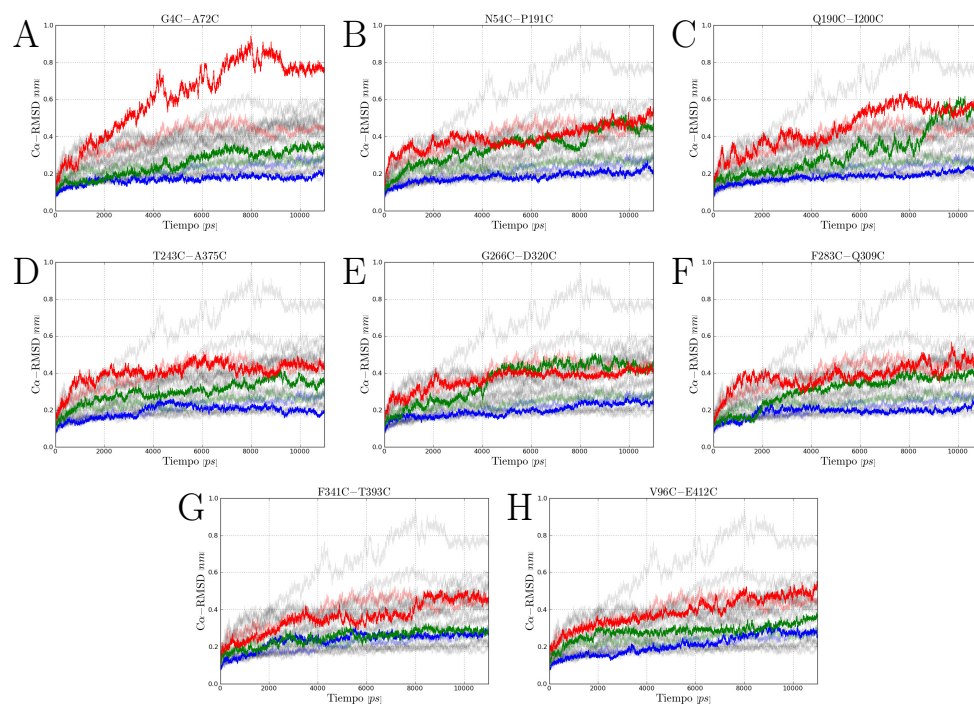
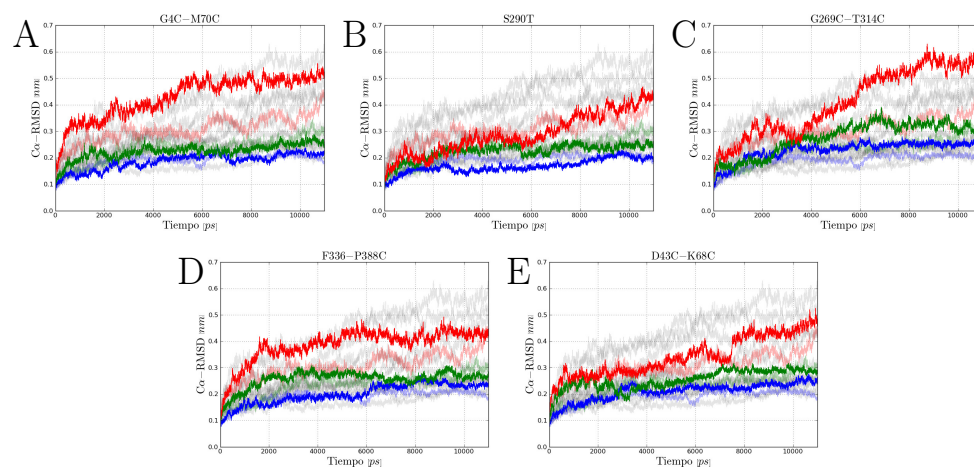


Figura 7.22: Mutaciones termoestabilizantes propuestas para la enzima M.

7.3.3. Simulación del efecto de la temperatura en la trayectoria de residuos de las enzimas variantes de T y M.

Para validar la estrategia es necesario simular el comportamiento de las estructuras de las enzimas variantes de T y M para, de esta forma, encontrar regiones flexibles en sus estructuras y comparar esto con la información que se tiene en la literatura de estas enzimas [141, 142]. Las estructuras de las mutadas se generaron mediante la herramienta para mutagénesis de PyMol. Los puentes disulfuro se establecieron mediante la opción `-ss` de `pdb2gmx` y se corrió una minimización de energía previa idéntica a la primera minimización en el procedimiento de simulación antes de llevar a cabo las siguientes etapas preparativas a la simulación, debido a que las variaciones se realizaron reemplazando los aminoácidos y haciendo modificaciones locales a al sistema teniendo que forzar la formación de los puentes para luego dejar que el sistema se acomode.

Se determinó el $C\alpha$ -RMSD en función del tiempo de producción para cada una de las enzimas variantes tanto de la enzima T como de la enzima M y se muestra en las Figuras 7.23 y 7.24. También se determinó el promedio del RMSD a lo largo de la simulación en cada residuo para las variantes de T (Figura 7.25) y para la enzima M (Figura 7.26).

Figura 7.23: C_{α} -RMSD de las enzimas mutantes de T la etapa de producción.Figura 7.24: C_{α} -RMSD de las enzimas mutantes de la enzima M la etapa de producción.

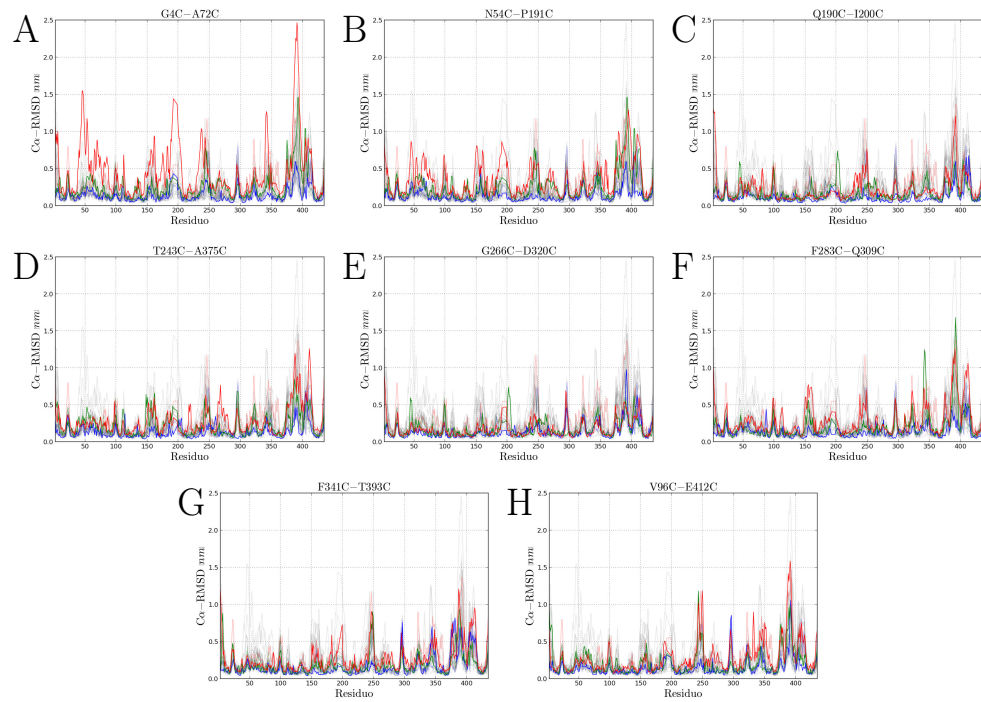


Figura 7.25: RMSD promedio para cada residuo de las mutantes de T

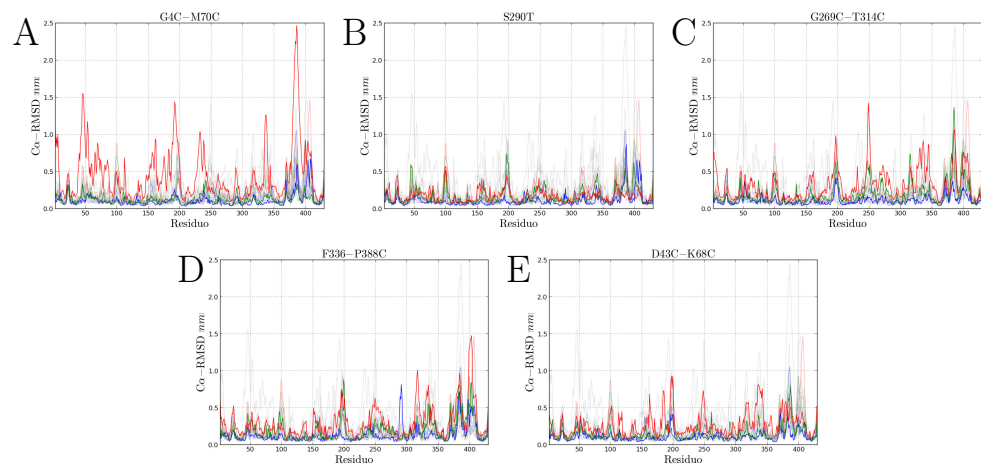


Figura 7.26: RMSD promedio para cada residuo de las mutantes de la enzima M

La compactación de las enzimas variantes de las enzimas T y M en la etapa de producción a las tres temperaturas simuladas se determinó en base al cálculo de Rg en función del tiempo de simulación y se muestra en las Figuras 7.27 y 7.28. En las Tablas 7.8 y 7.9 se muestran la media y la desviación estándar del Rg calculado para cada temperatura y para cada enzima. En base a estos datos se realizó un test t para evaluar la diferencia entre Rg medio de cada mutante respecto de la enzima nativa a una misma temperatura. Como hipótesis nula se consideró que las medias del Rg en cada comparación son iguales. Se estimó el promedio de la diferencia de Rg en cada comparación con un respectivo intervalo de confianza al 95 %. Si la comparación entrega que los Rg son significativamente diferentes, los valores negativos de la estimación del promedio de la diferencia indican que el valor comparado para la mutante es menor que el respectivo valor para la enzima nativa. Este análisis se resume en las Tablas 7.10 y 7.11.

Tabla 7.8: Media y desviación estándar del Rg a cada temperatura para las mutantes de la enzima T

Variante	300K		350K		400K	
	Rg medio [nm]	desv. est [nm]	Rg medio [nm]	desv. est [nm]	Rg medio [nm]	desv. est [nm]
A(G4C-A72C)	2,10249	0,0108191	2,13208	0,0157253	2,15512	0,0166725
B(N54C-P191C)	2,10811	0,0096405	2,14928	0,0205429	2,16140	0,0149647
C(Q190C-I200C)	2,11423	0,0087318	2,15412	0,0279318	2,20404	0,0189324
D(T243C-A375C)	2,10863	0,0076456	2,12843	0,0162320	2,13783	0,0193240
E(G266C-D320C)	2,12034	0,0082413	2,14021	0,0166655	2,16565	0,0162694
F(F383C-Q309C)	2,11725	0,0065979	2,12648	0,0138023	2,16240	0,0184387
G(F341C-T393)	2,11954	0,0096080	2,14591	0,0129451	2,17823	0,0148078
H(V96C-E412)	2,13022	0,0127630	2,15737	0,0091185	2,15272	0,0135960

7.3. VALIDACIÓN

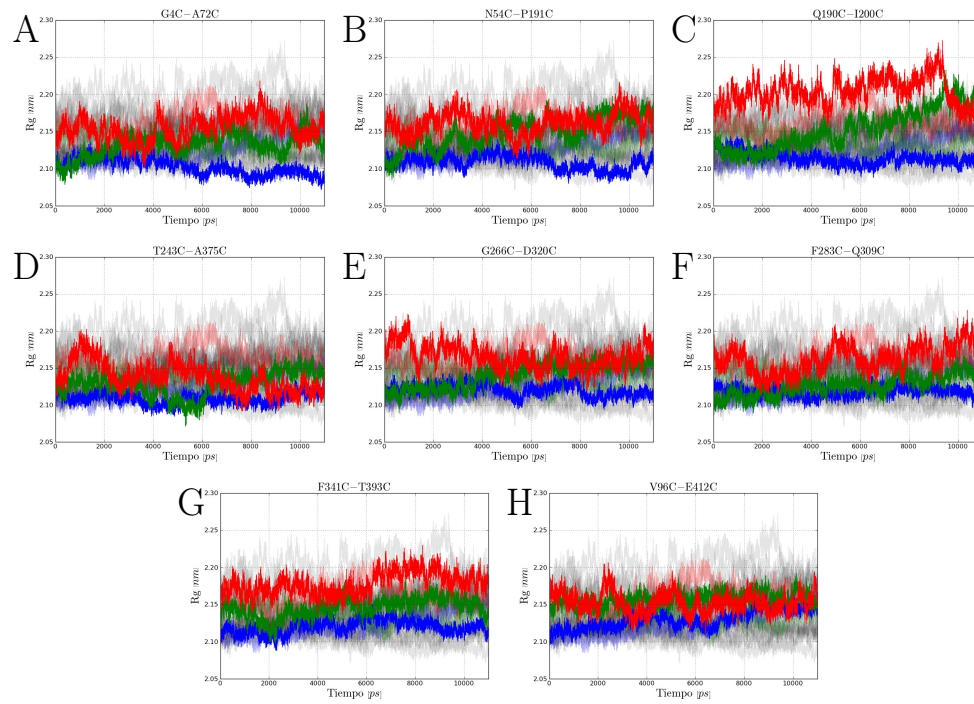


Figura 7.27: R_g calculado para las mutantes de T en función del tiempo

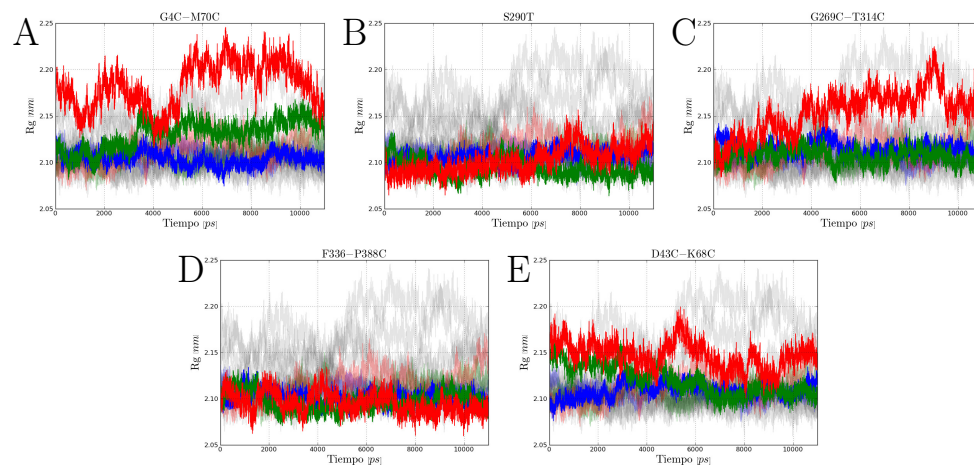


Figura 7.28: R_g calculado para las mutantes de la enzima M en función del tiempo

7.3. VALIDACIÓN

Tabla 7.9: Media y desviación estándar del Rg a cada temperatura para las mutantes de la enzima M

Variante	300K		350K		400K	
	Rg medio [nm]	desv. est [nm]	Rg medio [nm]	desv. est [nm]	Rg medio [nm]	desv. est [nm]
A(G4C-M70C)	2,10373	0,0074799	2,13046	0,0152561	2,18341	0,0241051
B(S290T)	2,10596	0,0070548	2,09379	0,0097930	2,10207	0,0140155
C(G269C-T314C)	2,11557	0,0073443	2,10671	0,0083961	2,15461	0,0241392
D(F336C-P388C)	2,10107	0,0069967	2,09661	0,0100731	2,09553	0,0113598
E(D43C-K68C)	2,10900	0,0084344	2,12009	0,0162001	2,14704	0,0148381

Tabla 7.10: Análisis t-student entre las enzimas mutantes de T y la enzima nativa.

		Intervalos de confianza (95 %)			
		t-value	$\bar{x}_2 - \bar{x}_1$	Límite inferior	Límite superior
A	300K	146,35	-0,01947	-0,01973	-0,01921
	350K	30,98	-0,004159	-0,004422	-0,003895
	400K	56,82	-0,008769	-0,009072	-0,008465
B	300K	107,49	-0,01385	-0,01410	-0,01360
	350K	80,95	0,01304	0,01272	0,01336
	400K	17,03	-0,002489	-0,002776	-0,002201
C	300K	61,41	-0,007730	-0,007977	-0,007482
	350K	87,01	0,01788	0,01748	0,01829
	400K	242,24	0,04015	0,03983	0,04048
D	300K	108,71	-0,01333	-0,01357	-0,01309
	350K	57,02	-0,007809	-0,008078	-0,007540
	400K	155,30	-0,02606	-0,02639	-0,02573
E	300K	13,03	-0,001620	-0,001864	-0,001376
	350K	28,51	0,003971	0,003698	0,004245
	400K	11,56	0,001761	0,001462	0,002061
F	300K	39,31	-0,004710	-0,004945	-0,004474
	350K	78,55	-0,009759	-0,01000	-0,009514
	400K	9,12	-0,001489	-0,001809	-0,001168
G	300K	18,80	-0,002420	-0,002673	-0,002167
	350K	80,62	0,009671	0,009436	0,009907
	400K	98,62	0,01434	0,01406	0,01463
H	300K	58,73	0,008260	0,007984	0,008536
	350K	205,69	0,02113	0,02093	0,02133
	400K	79,81	-0,01117	-0,01144	-0,01089

Tabla 7.11: Análisis t-student entre las enzimas mutantes de la enzima M y la enzima nativa.

		Intervalos de confianza (95 %)			
		t-value	$\bar{x}_2 - \bar{x}_1$	Límite inferior	Límite superior
	300K	38,21	-0,002856	-0,003002	-0,002709
A	350K	161,88	0,02148	0,02122	0,02174
	400K	352,18	0,06953	0,06914	0,06992
	300K	8,59	-0,0006256	-0,0007687	-0,0004825
B	350K	142,39	-0,01519	-0,01540	-0,01498
	400K	80,55	-0,01181	-0,01210	-0,01152
	300K	121,22	0,008984	0,008839	0,009130
C	350K	22,46	-0,002272	-0,002471	-0,002073
	400K	206,11	0,04073	0,04034	0,04112
	300K	76,00	-0,005516	-0,005658	-0,005373
D	350K	114,69	-0,01237	-0,01258	-0,01216
	400K	135,16	-0,01835	-0,01862	-0,01808
	300K	30,48	0,002414	0,002259	0,002570
E	350K	80,69	0,01111	0,01084	0,01138
	400K	220,71	0,03316	0,03286	0,03346

7.3. VALIDACIÓN

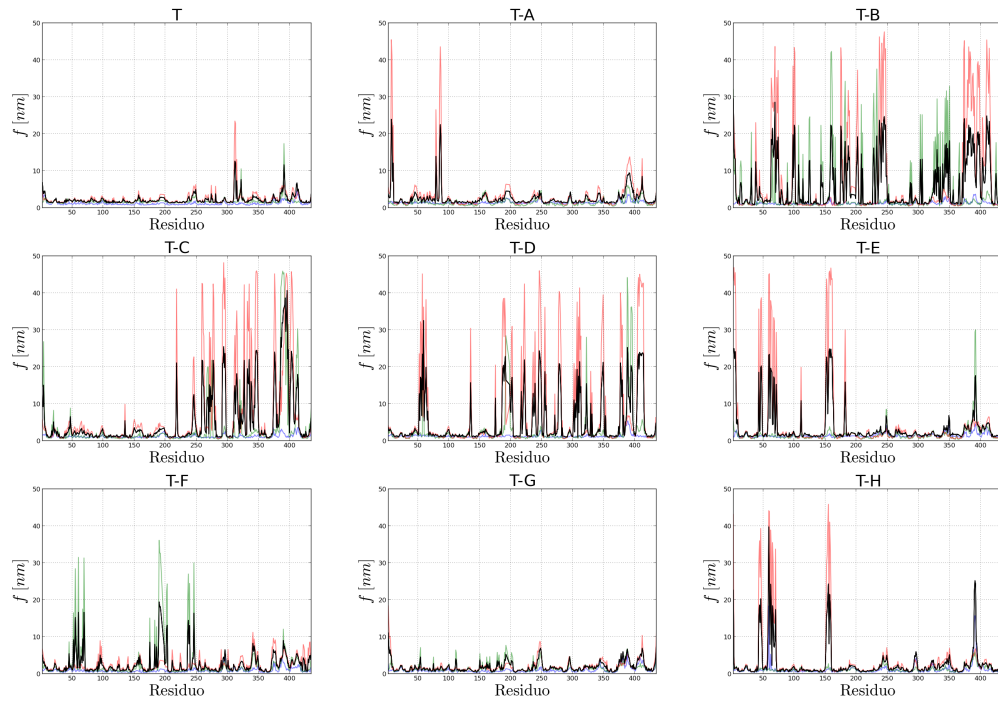


Figura 7.29: Valor f de la enzima T y sus variantes. En color azul, verde y rojo se muestran los valores f a 300, 350 y 400K, respectivamente. En color negro se muestra el valor if de cada enzima.

En las Figuras 7.29 y 7.30 se muestran los gráficos de valor f de cada residuo de las enzimas mutantes de las enzimas T y M.

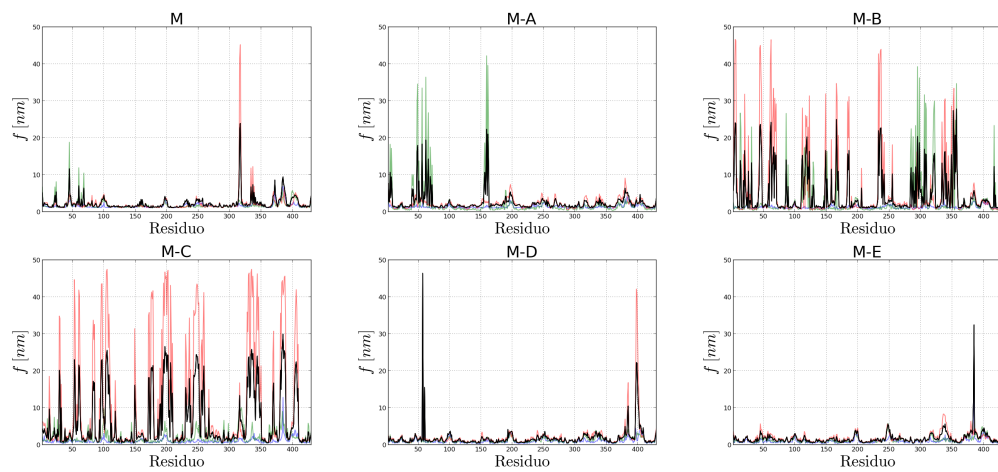


Figura 7.30: Valor f de la enzima M y sus variantes. En color azul, verde y rojo se muestran los valores f a 300, 350 y 400K, respectivamente. En color negro se muestra el valor if de cada enzima.

7.3.4. Resultados destacados de la etapa de validación

En la etapa de validación se simuló el comportamiento de las estructuras de las enzimas T y M con tres objetivos principales: (1) observar la evolución del sistema a lo largo de las diferentes etapas del procedimiento de simulación para de esta manera verificar que el procedimiento esté diseñado correctamente, (2) para determinar la flexibilidad de las enzimas previamente reportadas y establecer su existe coherencia entre ambos resultados y (3) identificar las regiones más flexibles de las enzimas T y M mediante el cálculo de los valores if y de esta manera proponer variantes en base a mutaciones que rigidicen las regiones más flexibles. Así, se observó un comportamiento en las etapas preparativas de la simulación, acorde con lo esperado, es decir, concordantes con los parámetros utilizados, en particular, los valores de referencia empleados en cada etapa (temperatura y presión específicamente). Se determinaron las regiones de alta flexibilidad para las enzimas T y M (Tabla 7.7). En base a esto se diseñaron 3 nuevas variante de cada enzima.

A partir de la simulación de comportamiento de las estructuras de las variantes de las enzimas T y M (tanto las propuestas en base a la metodología aquí desarrollada como las variantes propuestas en trabajos anteriores [141, 142]) se comparó la compactación de las enzimas no mutadas T y M, versus sus respectivas variantes, para así, establecer los cambios en la compactación de las enzimas en las simulaciones a 300, 350 y 400K, al introducir las mutaciones.

Con la información de las trayectorias de las estructuras de las enzimas variantes se determinó los valores if por residuo y se comparó estos resultados para determinar cambios en la flexibilidad al introducir mutaciones en regiones flexibles. Se pudo identificar que tres de las variantes de T y 2 de M mostraron, en promedio, una compactación mayor que su respectiva enzima nativa.

Cabe destacar que , para la enzima T, tres de las cinco variantes reportadas previamente como más termoestables poseen una mutación en o cerca una región identificada como flexible a partir del protocolo aquí desarrollado.

De la comparación de flexibilidad entre las enzimas nativas y sus respectivas variantes se obtuvo que una de las enzimas propuestas en este trabajo mostró disminuir su flexibilidad en términos generales, respecto de la enzima nativa. Así también sucede con una de las variantes aquí propuestas para M.

Para T, todas las enzimas variantes aquí propuestas mostraron valores if menores en la región mutada. Para M, dos de las tres variantes lo hicieron.

Así, se consideró el procedimiento validado y se procedió a continuar con su aplicación a la determinación de las regiones flexibles presentes en la estructura de la enzima Cel72.

7.4. Aplicación de la metodología a Cel72

7.4.1. Modelamiento

Modelo

Debido a que de la enzima Cel72 no se tiene modelo, si no que sólo la secuencia, se procede a modelar su estructura. El modelo que se obtuvo se muestra en la Figura 7.31. Para la generación del modelo se utilizó el programa Swiss-Model [10] y como molde la estructura reportada del módulo catalítico celobiohidrolasa Cel7D de *Phanerochaete chrysosporium* a resolución 1.32.

Al realizar un alineamiento del modelo obtenido con las estructuras de las enzimas T y M obtenidas del Protein Data Bank se obtuvo un RMS de 0.475 y 0.767, respectivamente. En la Figura 7.32 se muestran las estructuras de las enzimas Cel72, T y M superpuestas para mostrar la similitud en sus estructuras de manera gráfica.

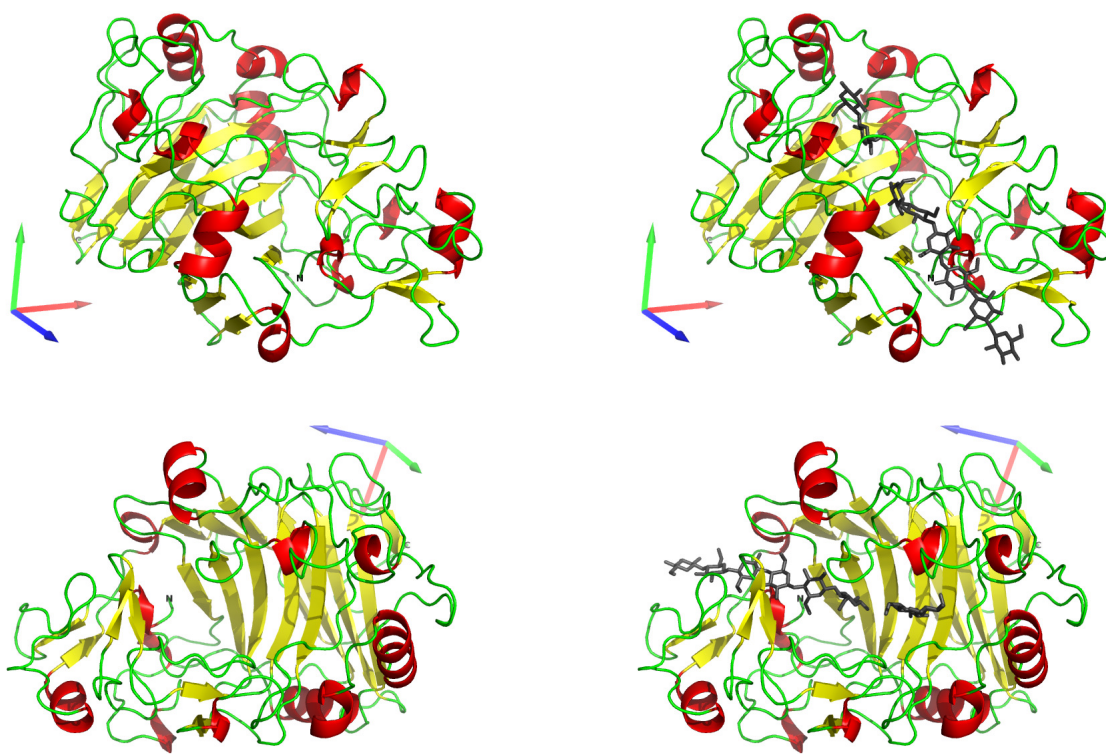


Figura 7.31: Modelo de Cel72. Se muestran en color rojo las hélices alfa y en color amarillo las hojas beta. La primera columna de estructuras corresponden a dos vistas diferentes de la enzima. La segunda columna corresponde a los mismos ángulos de observación de Cel72 superpuestas con la estructura de Cel7A de *T. reesei* (código PDB: 7CEL) para mostrar una cadena de celulosa (gris).

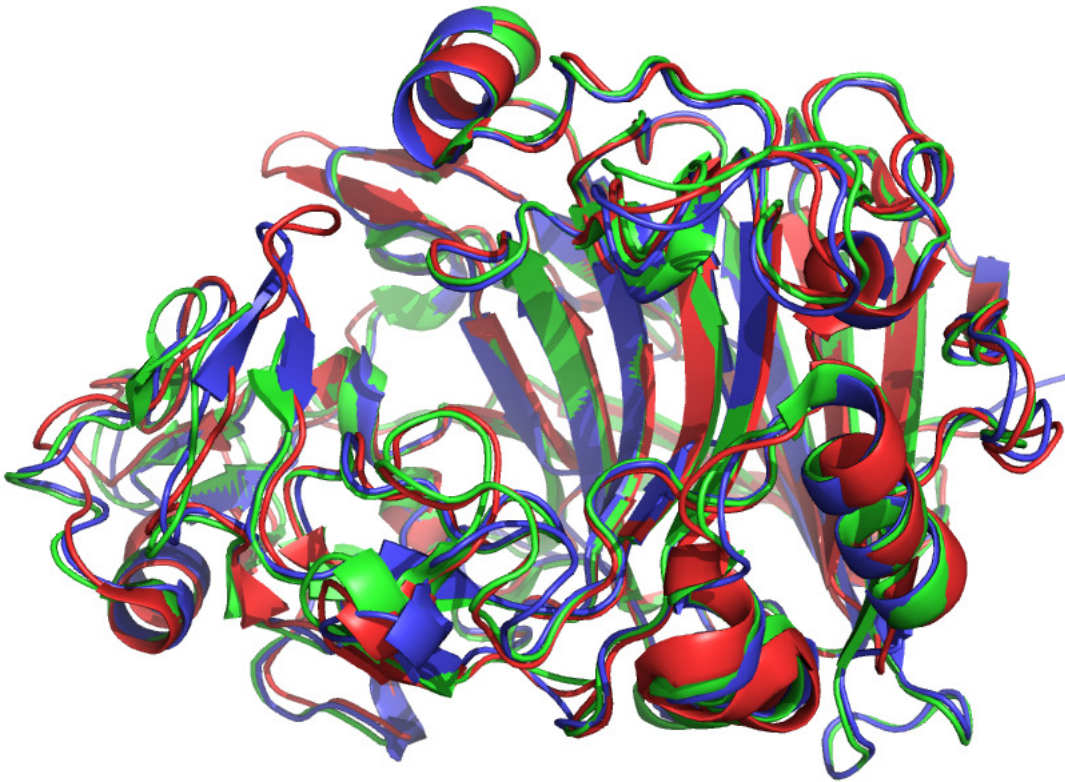


Figura 7.32: Modelos de las enzimas Cel72 (azul), T (verde) y M (rojo) superpuestos.

Evaluación del modelo

Se realizaron diferentes evaluaciones para poder validar el modelo. Se utilizaron los programas que ofrece Swiss-Model workspace y PROCHECK en el Structural Analysis and Validation Server de la UCLA. Los resultados de la validación se listan a continuación.

El puntaje global del modelo completo, QMEAN4 score [15], que refleja el ranking de confiabilidad del modelo predicho y que va de 0 a 1, fue de 0,79. La medida de la calidad absoluta del modelo, QMEAN Z-score, fue de 0,41, y se grafica en la Figura 7.33.

Los términos de pseudo energía, que contribuyen al cálculo juntos con sus Z-score correspondientes, respecto a los puntajes obtenidos de estructuras de alta resolución, obtenidas de manera experimental, de similar tamaño mediante cristalografía de rayos X, se muestran en la Tabla 7.12.

Tabla 7.12: Pseudo energías para el cálculo de QMEAN.

Scoring function term	Raw score	Z-score
C_beta interaction energy	-66.22	-1.54
All-atom pairwise energy	-6990.27	-1.59
Solvation energy	-26.74	-1.04
Torsion angle energy	-140.15	1.38
QMEAN4 score	0.788	0.41

Se realizó una estimación de la calidad del modelo de manera local utilizando ANOLEA, QMEAN y GROMOS (Figura 7.34). El potencial de fuerza promedio atómico empírico (atomic empirical mean force potential, ANOLEA [99]) se utiliza para evaluar la calidad del empaque de modelos. El programa realiza cálculos de energía en una cadena polipeptídica, evaluando el ambiente no local (non-local environment, NLE) de cada átomo pesado en la molécula. El puntaje local que calcula QMEAN

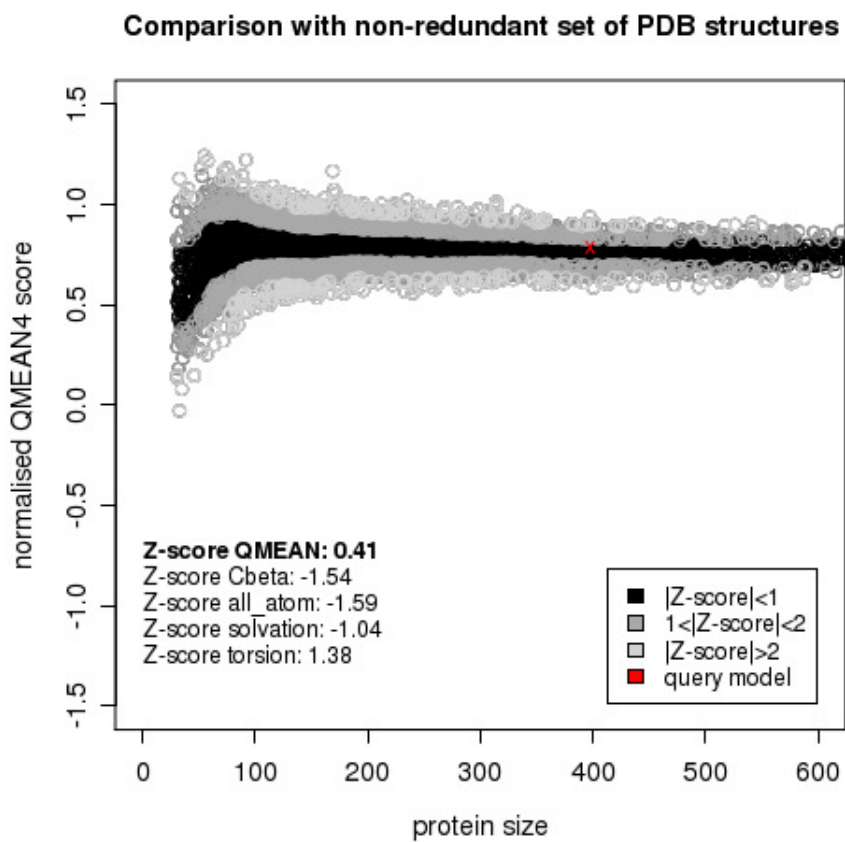


Figura 7.33: El QMEAN score del modelo es comparado a los puntajes de estructuras de alta resolución utilizados como referencia.

se realiza en una ventana de 9 residuos. Por otro lado, GROMOS [40], que es un paquete de simulación de dinámica molecular, realiza un análisis de conformaciones obtenidas de manera experimental o por simulación computacional.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

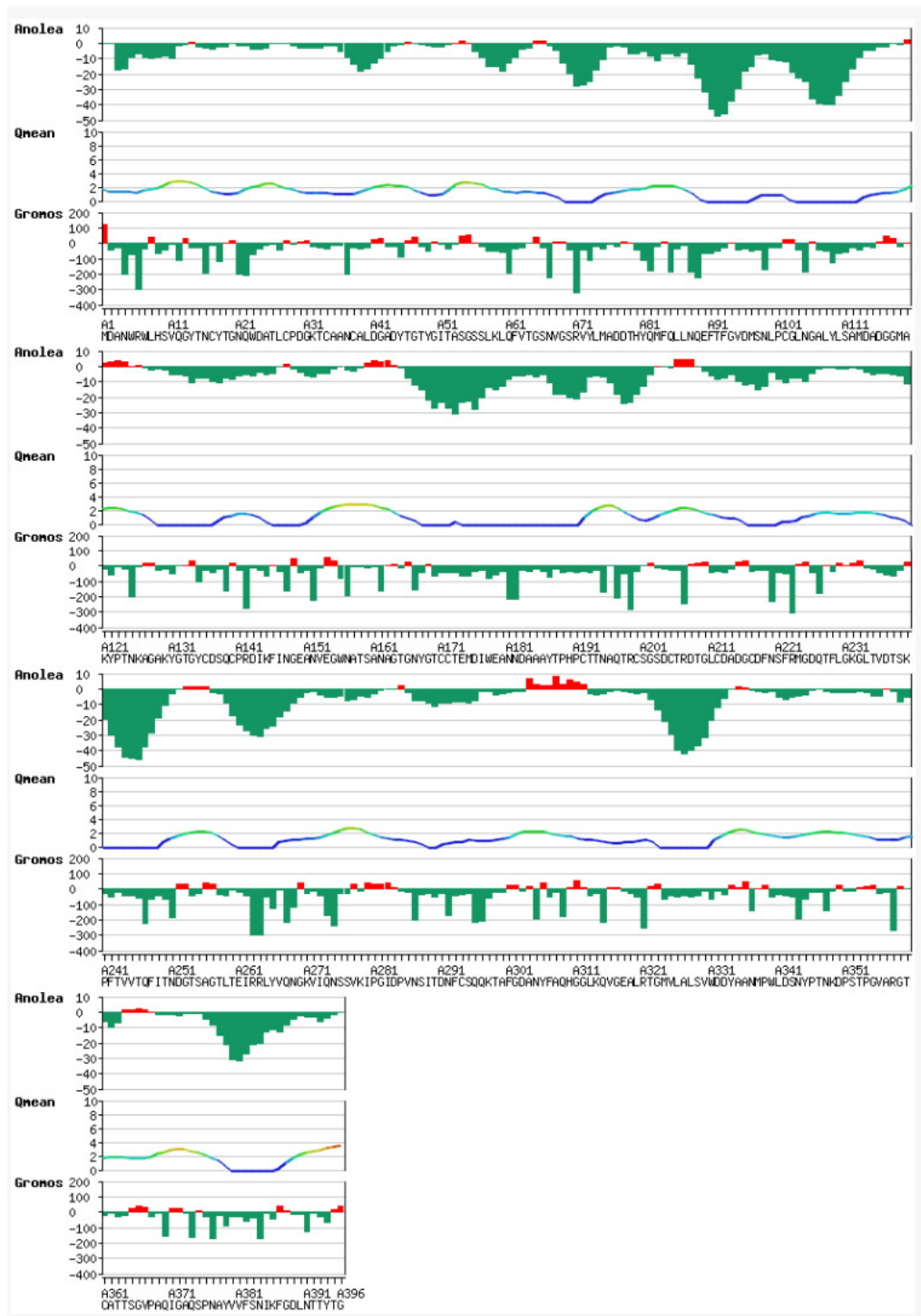


Figura 7.34: Estimación de la calidad local del modelo. Los valores negativos (verde), representa un ambiente energético favorable mientras que los valores positivos (rojo) representa un ambiente energético desfavorable.

Se realizó una evaluación del modelo utilizando PROCHECK. Al graficar los residuos de la enzima en un gráfico de Ramachandran se obtuvo que en 86.1 % de los aminoácidos se encuentran en el núcleo de la región favorable, el 13.7 % se encuentra en la zona menos favorable, y el 0.3 % está en la región de rechazo. Un modelo de buena calidad debería tener por sobre el 90 % de sus residuos en regiones más favorables. Para el gráfico de Ramachandran por tipo de residuo se encontraron 7 residuos con conformación no favorable. Al realizar un gráfico de los ángulos de torsión para residuos de cadena lateral lo suficientemente larga para tener estos ángulos en donde se obtuvo 3 residuos con conformación desfavorable. Se obtuvo que el 0.2 % de los enlaces en la cadena principal presentaron largo fuera del rango de referencia y el 1.5 % de sus ángulos están fuera del rango. Las Figuras del análisis de evaluación del modelo se adjuntan en la sección A.4.

Dichos valores entregados por los diferentes métodos de evaluación indican una aceptabilidad inferior a la ideal, sin embargo, debido a que la siguiente etapa es una simulación, el modelo será sometido a etapas de preparación a la simulación, y por esto, el modelo obtenido se acepta para ser utilizado en las etapas posteriores.

7.4.2. Simulación del efecto de la temperatura en la estructura de Cel72

El procedimiento seguido para la simulación del efecto de la temperatura para Cel72 fue idéntico al seguido para las enzimas T y M. Se realizó una simulación a 300K y luego la estructura final de la etapa de producción se utilizó como entrada para la simulación a 350K. Así mismo, la estructura obtenida de la etapa de simulación a 350K se utilizó como estructura de partida para la simulación a 400K.

Las etapas de conversión de la estructura, minimización de energía y condiciones

de borde periódicas no se detallan en esta sección pues se detallaron en la etapa de validación y corresponden a etapas realizadas bajo las mismas condiciones. Los parámetros y registros de salida de cada etapa se encuentran en la sección A.5. Con el objetivo de ver la evolución del sistema a medida que se prepara para la simulación final, se muestran en cada etapas los datos de temperatura, presión y energía del sistema.

Adición de solvente e Iones

De igual modo que en la etapa de validación, se utilizó como solvente agua explícita de tipo SPC. Se agregaron a la celda 14768 moléculas de agua. La densidad fue de $1067.84 [g/l]$ y el volumen $503.631 [nm^3]$. La concentración de iones utilizada fue de $0.15M$ de NaCl. Se reemplazaron 102 moléculas de soluto agregando 57 átomos de Na^+ y 45 átomos de Cl^- .

Relajación del solvente y átomos de hidrógeno y aumento de la temperatura

Las condiciones y parámetros utilizadas en la etapa de relajación en la validación se repitieron para la relajación del solvente y aumento de la temperatura con Cel72. Con el fin de observar la evolución del sistema se midió la temperatura, presión y energía del sistema en función del tiempo (Figura 7.35).

Control de la Temperatura

Se realizó la etapa de temperature coupling de igual modo que se realizó en la validación con las enzimas T y M. Se midieron las propiedades temperatura y presión, y se midió además la energía del sistema durante la etapa NVT (Figura 7.36). En la Tabla 7.13 se resume los valores de las propiedades del sistema en función del tiempo de simulación.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

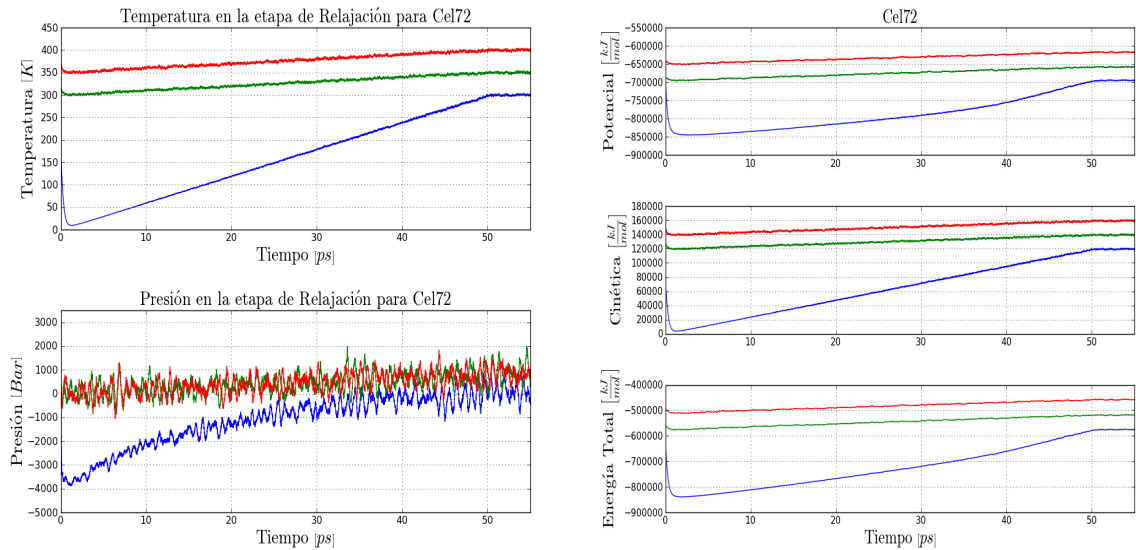


Figura 7.35: Temperatura, presión y energía del sistema en función del tiempo de relajación. Se aumentó la temperatura de 0 a 300K (azul), de 300 a 350K (verde), y de 350 a 400K (rojo). En los últimos 5ps de cada relajación la temperatura de referencia se dejó constante. Se muestra la energía potencial, cinética y total a las tres temperaturas estudiadas (azul: 300K, verde: 350K y rojo: 400K) en función del tiempo de relajación.

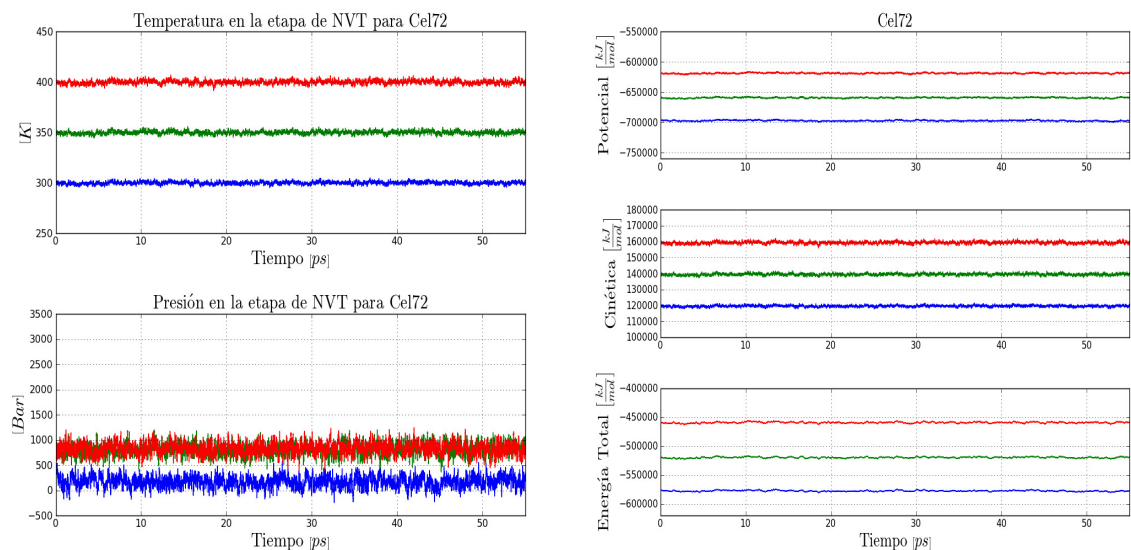


Figura 7.36: Temperatura, presión y energía del sistema en la etapa NVT para la enzima Cel72 (azul: 300K, verde: 350K y rojo: 400K).

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

Tabla 7.13: Temperatura, presión y energía del sistema para la enzima T en la etapa NVT

Temp. de Referencia		Promedio	Err. Est.	RMSD	Unidades
300K	Temperatura	299,86	0,082	1,35717	[K]
	Presión	169,779	4	112,291	[bar]
	Potencial	-697048	120	848,608	[kJ/mol]
	Cinética	119532	33	540,999	[kJ/mol]
	Total	-577516	120	1015,62	[kJ/mol]
350K	Temperatura	349,83	0,1	1,6041	[K]
	Presión	807,339	6,6	116,091	[bar]
	Potencial	-659216	65	850,867	[kJ/mol]
	Cinética	139451	40	639,43	[kJ/mol]
	Total	-519764	100	1091,67	[kJ/mol]
400K	Temperatura	399,81	0,12	1,85694	[K]
	Presión	830,2	4,2	117,169	[bar]
	Potencial	-618800	77	887,294	[kJ/mol]
	Cinética	159371	47	740,219	[kJ/mol]
	Total	-459429	100	1146,37	[kJ/mol]

Control de la Presión

En esta etapa se realiza una simulación utilizando *pressure coupling*. Se midió la temperatura y presión en función del tiempo de simulación durante la etapa NPT a las tres temperaturas, y se midió la energía del sistema a lo largo de la simulación, a las tres temperaturas estudiadas (Figura 7.37). En la Tabla 7.14 se resumen los datos obtenidos para la energía, presión y temperatura a las diferentes temperaturas.

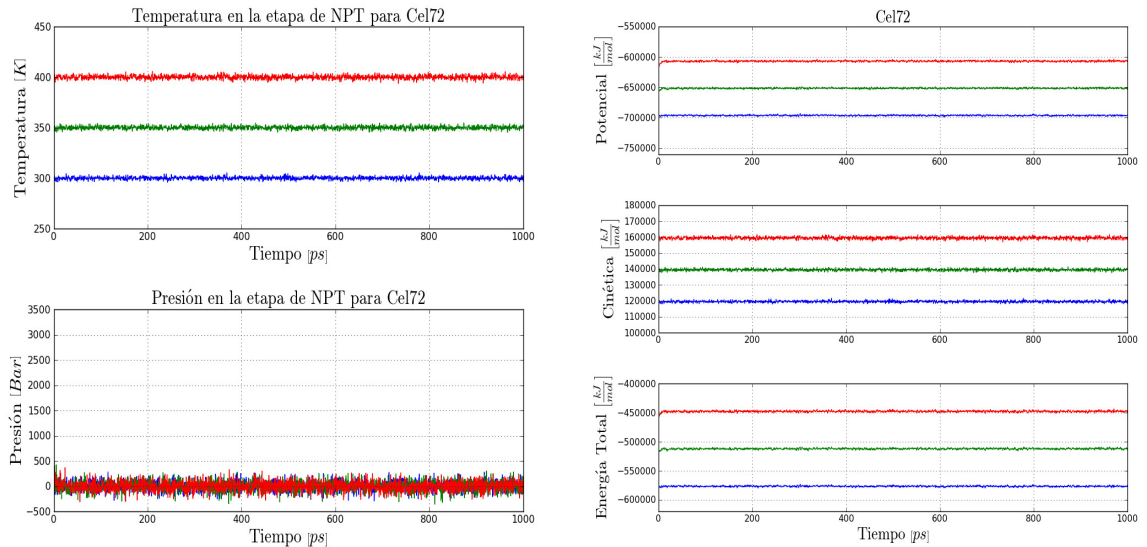


Figura 7.37: Temperatura, presión y energía del sistema en la etapa NPT (azul: 300K, verde: 350K y rojo: 400K).

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

Tabla 7.14: Temperatura, presión y energía del sistema en la etapa NPT

Temp. de Referencia		Promedio	Err. Est.	RMSD	Unidades
300K	Temperatura	299,98	0,013	1,35654	[K]
	Presión	1,24526	0,23	96,7327	[bar]
	Potencial	-696216	40	791,392	[kJ/mol]
	Cinética	119579	5,1	540,748	[kJ/mol]
	Total	-576637	43	955,67	[kJ/mol]
350K	Temperatura	349,98	0,016	1,57753	[K]
	Presión	1,99233	0,93	102,213	[bar]
	Potencial	-651504	57	922,525	[kJ/mol]
	Cinética	139509	6,2	628,839	[kJ/mol]
	Total	-511995	63	1128,71	[kJ/mol]
400K	Temperatura	399,97	0,019	1,80587	[K]
	Presión	2,36781	1,3	102,857	[bar]
	Potencial	-606907	57	1048,21	[kJ/mol]
	Cinética	159438	7,4	719,859	[kJ/mol]
	Total	-447469	63	1280,27	[kJ/mol]

Producción

La etapa de producción fue una simulación de $11ns$. Los parámetros usados son los que se utilizaron en la validación. Se midieron la temperatura, presión y energía con el módulo `g_energy` de GROMACS. En la Figura 7.38 se grafican los resultados obtenidos de los cálculos mencionados. En la Tabla 7.15 se resumen los resultados de estos cálculos a las tres temperaturas estudiadas.

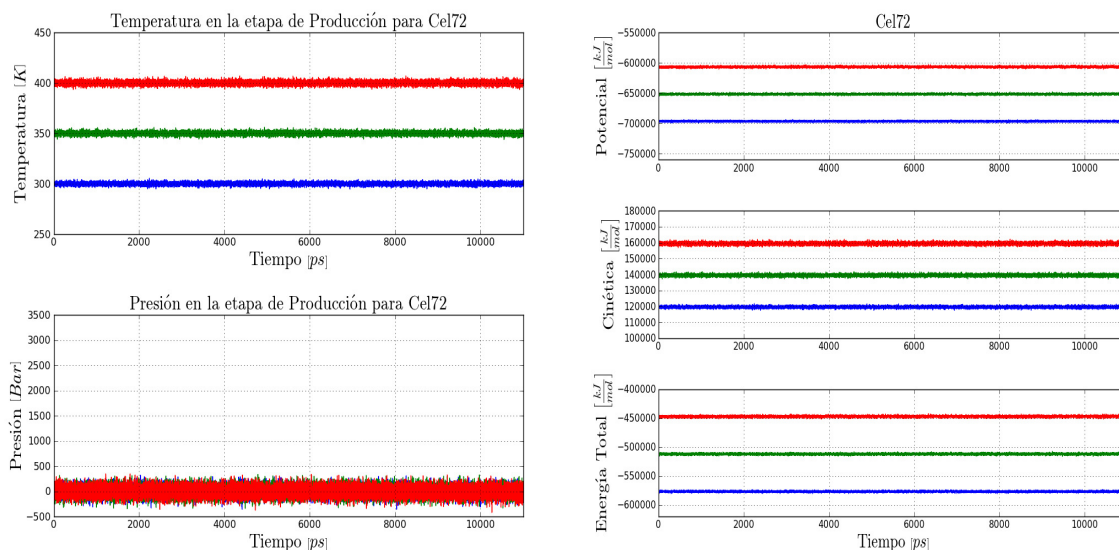


Figura 7.38: Temperatura, presión y energía del sistema en la etapa de producción (azul: $300K$, verde: $350K$ y rojo: $400K$).

Se determinó el RMSD de la cadena principal ($C\alpha$ -RMSD) de la enzima a lo largo de la simulación utilizando como referencia de la estructura inicial y se determinó también el RMSD promedio para cada residuo y la distribución de los RMSD para cada enzima a cada temperatura simulada (Figura 7.39). Se consideró como etapa de equilibrio el primer ns de la simulación. El RMSD de los últimos $10ns$ de simulación fue de $0.2584 \pm 0.06426[nm]$, $0.3066 \pm 0.06536[nm]$ y $0.4523 \pm 0.10265[nm]$ a cada temperatura respectivamente.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

Tabla 7.15: Propiedades medidas en la etapa de producción a las temperaturas de estudio.

Temp. de Referencia		Promedio	Err. Est.	RMSD	Unidades
300K	Temperatura	299,98	0,0056	1,35201	[K]
	Presión	1,01163	0,003	87,5073	[bar]
	Potencial	-696433	21	798,091	[kJ/mol]
	Cinética	119579	2,2	538,943	[kJ/mol]
	Total	-576854	22	972,263	[kJ/mol]
	Volumen	508,387	0,045	0,686471	[nm ³]
	Densidad	1010,98	0,089	1,36511	[kg/m ³]
350K	Temperatura	349,976	0,0065	1,57669	[K]
	Presión	1,00833	0,0045	91,9157	[bar]
	Potencial	-651577	20	897,061	[kJ/mol]
	Cinética	139508	2,6	628,504	[kJ/mol]
	Total	-512069	20	1108,27	[kJ/mol]
	Volumen	531,024	0,05	0,784628	[nm ³]
	Densidad	967,883	0,092	1,43001	[kg/m ³]
400K	Temperatura	399,973	0,0074	1,8001	[K]
	Presión	1,01019	0,011	93,7633	[bar]
	Potencial	-606695	41	997,188	[kJ/mol]
	Cinética	159438	3	717,56	[kJ/mol]
	Total	-447257	41	1244,21	[kJ/mol]
	Volumen	563,479	0,066	0,970176	[nm ³]
	Densidad	912,136	0,11	1,57028	[kg/m ³]

Como medida de compactación se calculó el Rg (Radius of gyration) en función del tiempo (Figura 7.40).

En la Tabla 7.16 se muestran los valores de los promedios y desviación estándar de los datos de Rg a cada temperatura. A base de esos valores se determinaron los t-value para realizar un test t-student. Se comparó el Rg a 300K versus 350 y 400K el Rg a 350K versus 400K. Como hipótesis nula se consideró que las medias del Rg en cada comparación son iguales. El t-value se calculó mediante la ecuación 7.1.

Los valores de t calculados (Tabla 7.17) se obtiene que se rechaza la hipótesis nula en todos los casos. De esta forma, se puede decir que la diferencia entre las medias comparadas es significativa (P-valor < 0,0001).

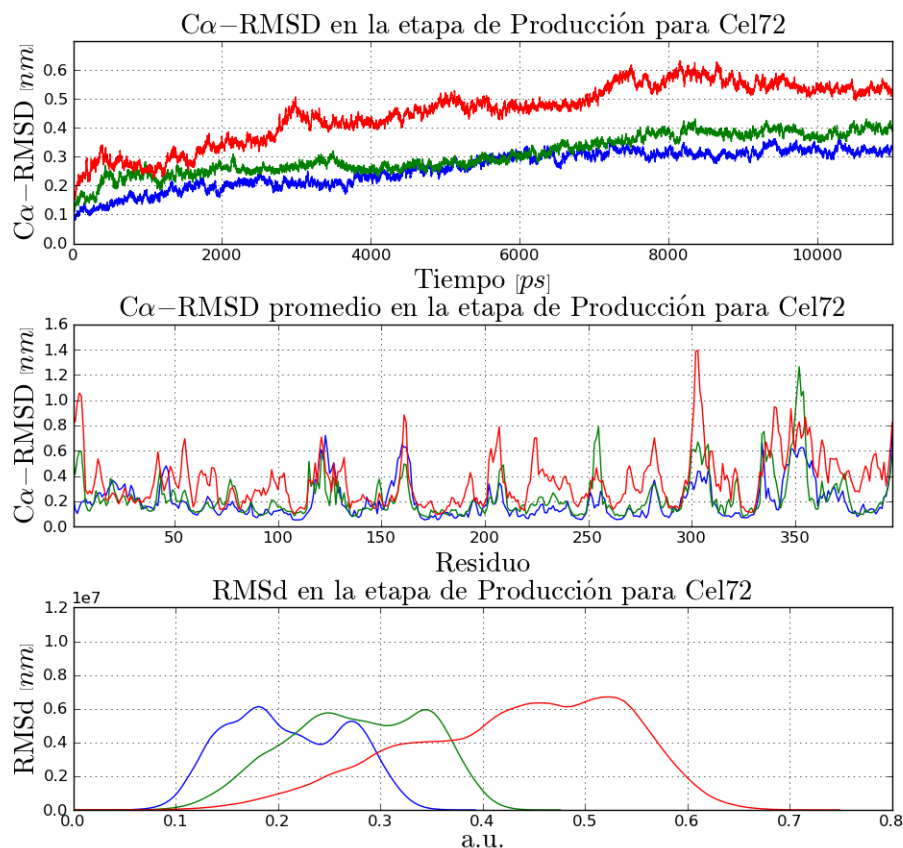


Figura 7.39: $C\alpha$ -RMSD en la etapa de producción. Se muestra el $C\alpha$ -RMSD a lo largo de la simulación, el $C\alpha$ -RMSD promedio para cada residuo y la distribución de $C\alpha$ -RMSD (azul: 300K, verde: 350K y rojo: 400K).

Una visión general del comportamiento de las enzimas en la simulación cada 1.0ns de simulación a 300, 350 y 400K se puede ver en la Figura 7.41. Se utilizó el programa PyMol para la visualización de la enzima. Se puede ver que a simple vista que la amplitud del desplazamiento de la cadena principal de Cel72 aumenta y la estructura se deforma al incrementar la temperatura.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

Tabla 7.16: Medias y desviaciones del Rg.

300K		350K		400K	
Media	desv. Est	Media	desv. Est	Media	desv. Est
2,1192	0,0170	2,1218	0,0230	2,1422	0,0145

Tabla 7.17: Resultados del test t. Valor t de cada comparación y los intervalos de confianza para el promedio de la diferencia.

	t-value	Intervalos de confianza (95 %)		
		$\bar{x}_2 - \bar{x}_1$	Límite inferior	Límite superior
300K-350K	13.46	0.002599	0.002220	0.002978
350K-400K	111.03	0.02037	0.02000	0.02073
300K-400K	152.27	0.02296	0.02267	0.02326

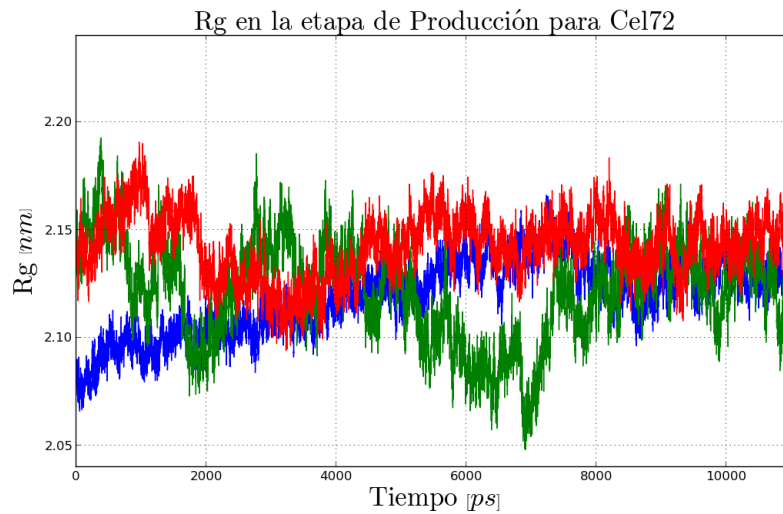


Figura 7.40: Rg de la enzima en función del tiempo a las tres temperaturas (azul: 300K, verde: 350K y rojo: 400K).

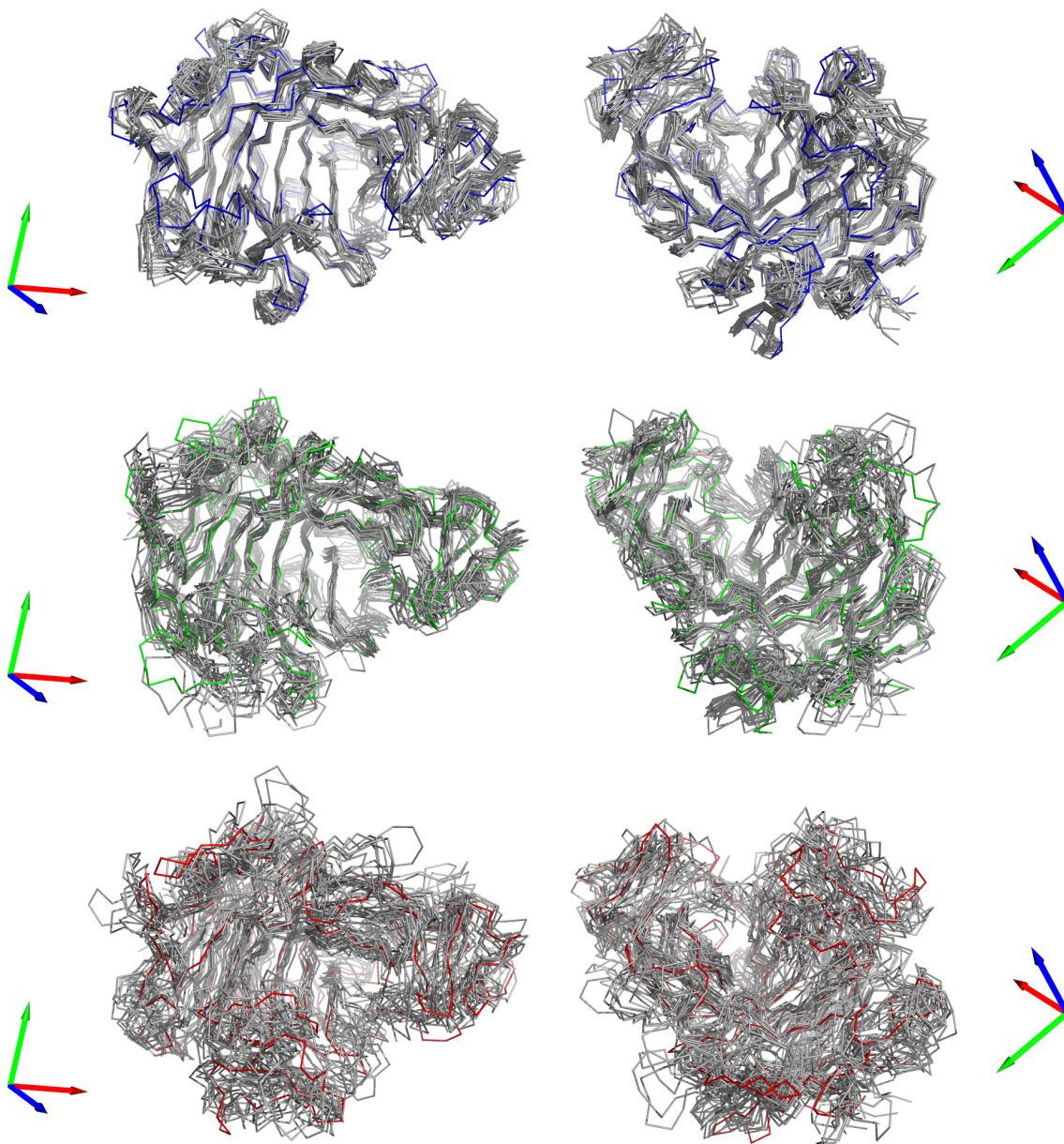


Figura 7.41: Visión general de la trayectoria de Cel72 a 300K (azul), 350K (verde) y 400K (rojo). Se muestran el esqueleto de las estructuras cada $1ns$ en 4 perspectivas diferentes. La elección de mostrar las estructuras cada $1.0ns$ se decidió de manera arbitraria. En color azul, verde o rojo se muestra la estructura inicial y en gris las estructuras a lo largo de la simulación.

7.4.3. Cálculo de la flexibilidad

Para determinar la flexibilidad local en cada residuo, se calculó el valor f y if de igual modo que en la validación. En la Figura 7.42 se muestra un gráfico con los valores f para cada temperatura, por residuo; y los valores if para cada residuo de la enzima Cel72.

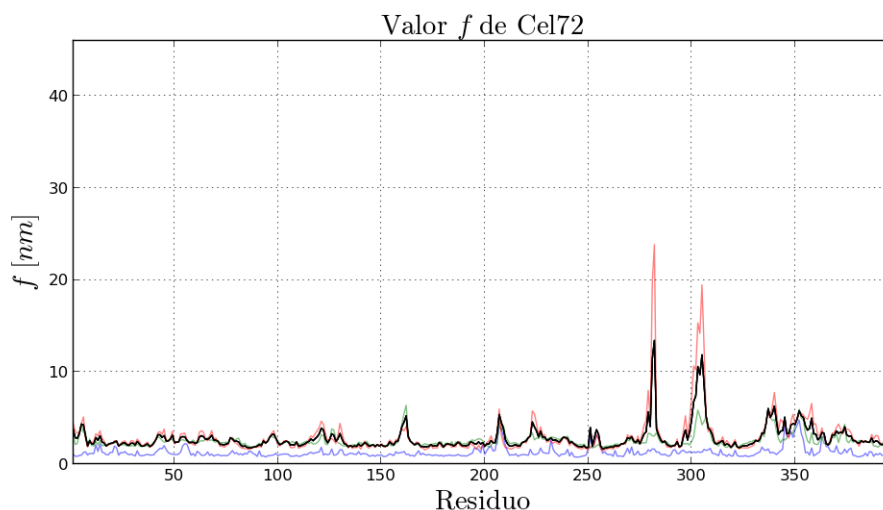


Figura 7.42: Flexibilidad local de Cel72 por residuo. Se muestra en color negro la suma de entre el f a 300K y el promedio del módulo de la diferencia entre las otras dos temperaturas y la temperatura inicial (if). En color azul, verde y rojo se muestran los valores f determinados para cada enzima a las temperaturas 300, 350 y 400K respectivamente.

Al partir de los valores if calculados, se realizó un ranking u ordenamiento con el objetivo de clasificar los residuos en base a su valor if ordenándolos de mayor a menor if . Se definió como región mutable los aminoácidos rankeados que sean vecinos en la secuencia. En la Tabla 7.18 se muestra el ranking realizado para Cel72. Pertenecen al ranking los residuos cuyos valores f son mayores que la media por enzima más su correspondiente desviación estándar.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

Tabla 7.18: Ranking de valores *if* para Cel72. Se indican además las regiones a las cuales pertenecen los residuos.

N°	Aminoácido ^a	<i>if</i>	Región	N°	Aminoácido ^a	<i>if</i>	Región
1	282	3,449	A	18	345	5,102	D
2	305	11,876	B	19	338	5,077	C
3	281	11,467	A	20	354	5,074	D
4	303	10,611	B	21	358	4,994	D
5	304	9,709	B	22	351	4,904	D
6	306	8,694	B	23	300	4,826	B
7	302	7,407	B	24	341	4,669	C
8	301	6,803	B	25	223	4,585	
9	340	6,318	C	26	161	4,575	F
10	337	6,030	C	27	208	4,514	E
11	352	5,800	D	28	336	4,433	C
12	279	5,680	A	29	349	4,397	D
13	339	5,480	C	30	5	4,351	G
14	207	5,399	E	31	350	4,313	D
15	353	5,257	D	32	6	4,279	G
16	162	5,250	F	33	348	4,236	D
17	307	5,205	B	34	355	4,169	D

Al igual que en la sección de validación (Sección 7.3.2), se definieron las regiones mutables en base a la cercanía entre residuos destacados por su flexibilidad dejando fuera los residuos que se encontraban dentro del ranking pero que no tenían residuos vecinos en la secuencia definiéndose de esta manera 7 regiones que se muestran en la Figura 7.43. En la Figura 7.44 se muestran la estructura del entorno de las mutaciones generadas a partir de la metodología propuesta.

^aÍndice relativo a la secuencia aminoacídica de la enzima madura.

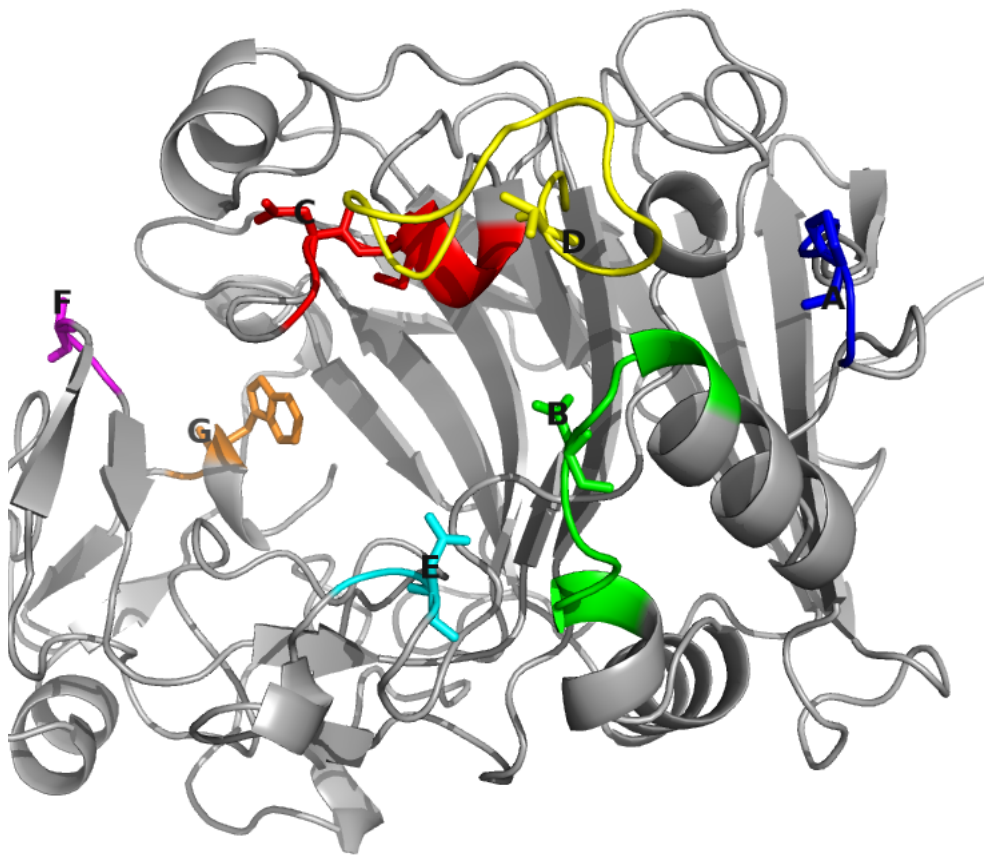


Figura 7.43: Regiones definidas como flexibles en base al análisis de valor if para Cel72. En color azul se destacaron los residuos pertenecientes a la región A, en verde la B, rojo para C, amarillo para D, celeste para E, rosado para F y naranja para G.

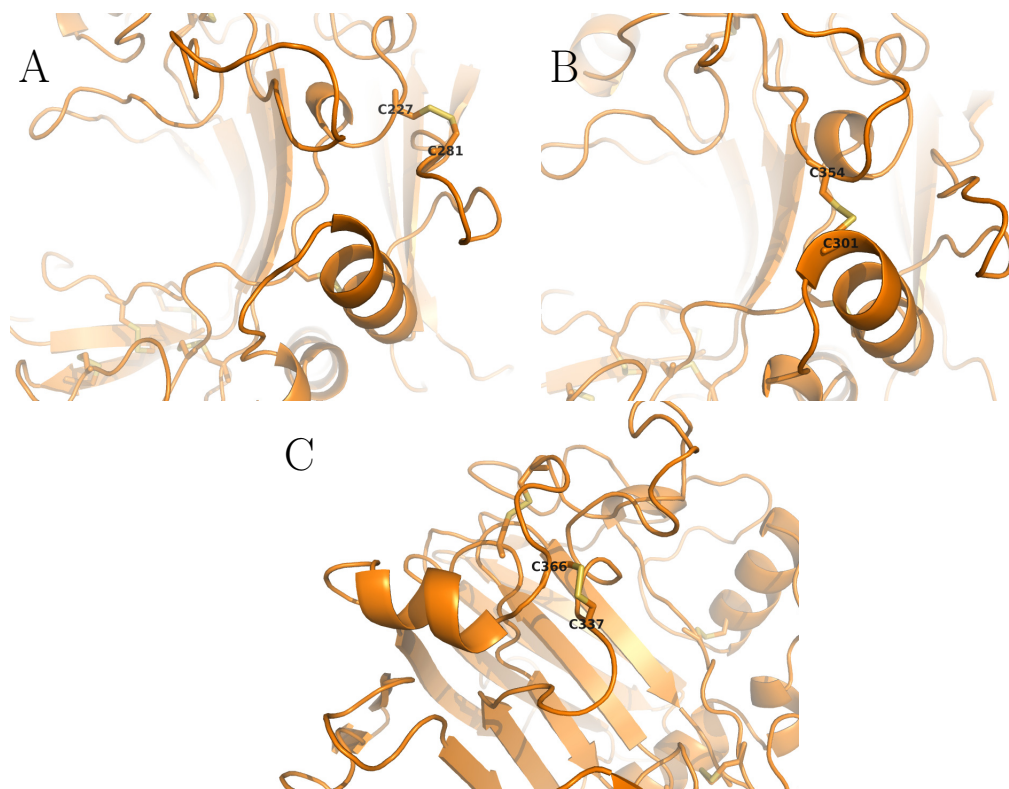


Figura 7.44: Estructura del entorno de las mutaciones termoestabilizantes propuestas en Cel72.

7.4.4. Simulación de las variantes

Las estructuras de las enzimas variantes de Cel72 propuestas para mejorar la termoestabilidad y señaladas en la sección anterior (Sección 7.4.3), fueron generadas mediante la herramienta de mutagénesis de Pymol y fueron tratadas posteriormente de igual manera que las variantes de las enzimas T y M.

Se determinó el $C\alpha$ -RMSD en función del tiempo de producción para cada una de las enzimas mutantes y se muestra en la Figura 7.45, en donde se puede comparar los RMSD de las variantes (azul: 300K, verde: 350K y rojo: 400K) con los valores calculados para la enzima nativa (azul claro: 300K, verde claro: 350K y rojo claro: 400K). También se determinó el promedio del RMSD a lo largo de la simulación en cada residuo para las mutantes de Cel72 (Figura 7.46).

La compactación de las enzimas variantes de Cel72 en la etapa de producción a las tres temperaturas simuladas se determinaron en base al cálculo de R_g en función del tiempo de simulación y se muestra en la Figura 7.47. En la Tabla 7.19 se muestran la media y la desviación estándar del R_g calculado para cada temperatura y para cada enzima. A base de estos datos se realizó un test t para evaluar la diferencia entre R_g medio de cada mutante respecto de la enzima nativa a una misma temperatura. Como hipótesis nula se consideró que las medias del R_g en cada comparación son iguales. Se estimó el promedio de la diferencia de R_g en cada comparación con un respectivo intervalo de confianza al 95%. Si la comparación entrega que los R_g son significativamente diferentes, los valores negativos de la estimación del promedio de la diferencia indican que el valor comparado para la mutante es menor que el respectivo valor para la enzima nativa. Los resultados de este análisis se resume en la Tabla 7.20.

En la Figura 7.48 se muestran los valores f para cada residuo de las variantes de

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

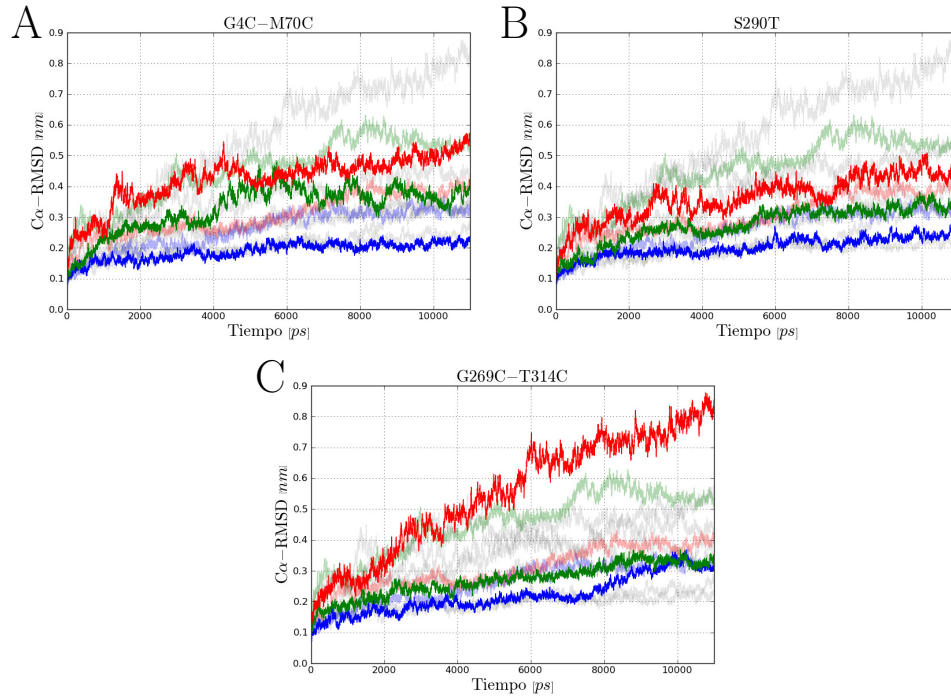


Figura 7.45: $C\alpha$ -RMSD de las enzimas mutantes de Cel72 en la etapa de producción (azul: 300K, verde: 350K y rojo: 400K). Se muestra también en colores claros los valores obtenidos para la enzima nativa.

Cel72.

De esta forma, fueron analizadas las tres variantes diseñadas a partir de los cálculos de *if*. Dos de estas enzimas mostraron mayor compactación, en promedio, a las tres temperaturas simuladas y la tercera variante mostró mayor compactación en las dos primeras temperaturas.

Respecto a la flexibilidad, se obtuvo que las tres enzimas mostraron disminuir el valor *if* en las regiones modificadas, y una en particular, la enzima Cel72-B, mostró una disminución global de la flexibilidad.

Será necesario evaluar la termoestabilidad de las variantes de manera experimental para poder establecer la real efectividad del método aquí propuesto.

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

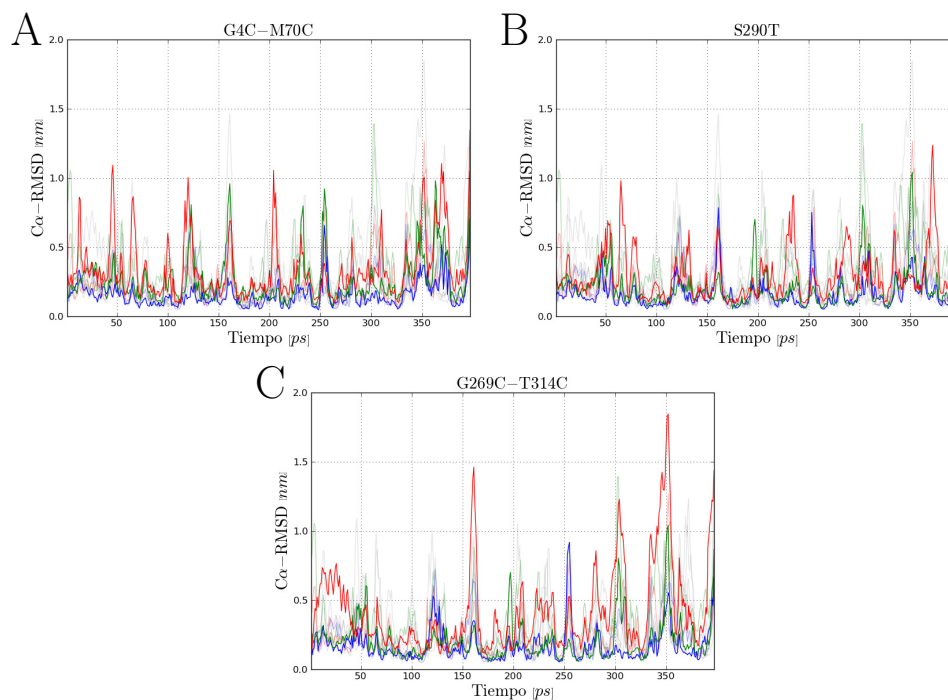


Figura 7.46: RMSD promedio para cada residuo de las variantes de Cel72

Tabla 7.19: Media y desviación estándar del Rg a cada temperatura para las mutantes de Cel72

Enzima	300K		350K		400K	
	media	desv. est.	media	desv. est.	media	desv. est.
A (Q227C-P281C)	2.0795	0.01589	2.1203	0.02012	2.1402	0.01945
B (F301C-G354C)	2.1114	0.01446	2.1210	0.01485	2.1398	0.01551
C (N337C-G366C)	2.0791	0.00785	2.0893	0.01366	2.1813	0.04610

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

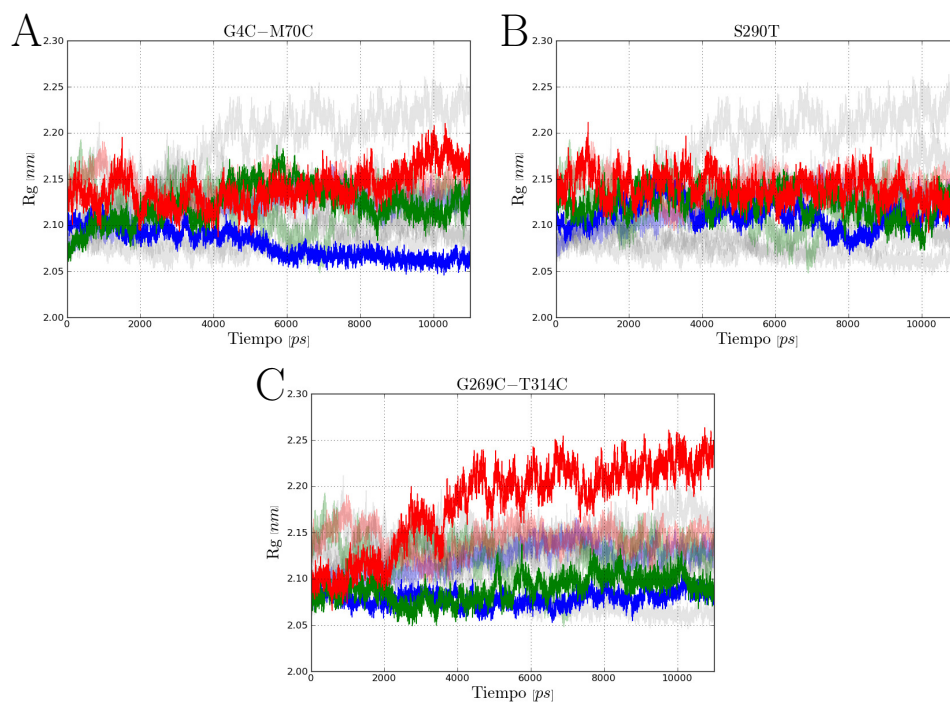


Figura 7.47: R_g calculado para las mutantes de Cel72 en función del tiempo

Tabla 7.20: Análisis t-student entre las enzimas mutantes de Cel72 y la enzima nativa.

		Intervalos de confianza (95 %)			
		t value	$\bar{x}_2 - \bar{x}_1$	Límite inferior	Límite superior
	300K	253.41	-0.03979	-0.04010	-0.03949
A	300K	7.61	-0.001569	-0.001974	-0.001164
	300K	12.45	-0.002036	-0.002358	-0.001715
	300K	52.26	-0.007871	-0.008167	-0.007575
B	300K	4.47	-0.0008256	-0.001188	-0.0004628
	300K	17.02	-0.002436	-0.002718	-0.002155
	300K	317.52	-0.04014	-0.04039	-0.03989
C	300K	180.43	-0.03256	-0.03291	-0.03220
	300K	119.86	0.03905	0.03841	0.03969

7.4. APLICACIÓN DE LA METODOLOGÍA A CEL72

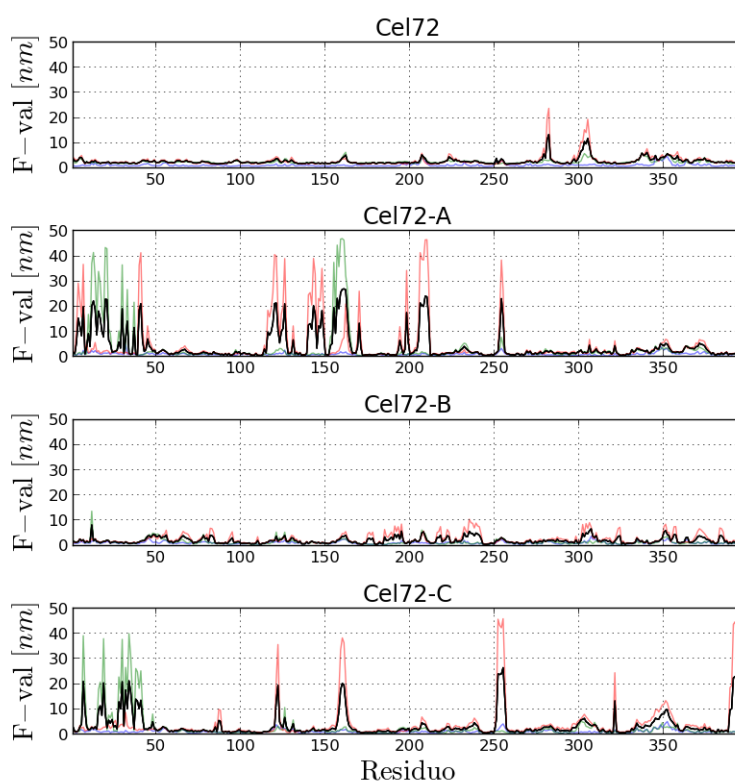


Figura 7.48: Valor f de la enzima Cel72 y sus variantes. En color azul, verde y rojo se muestran los valores f a 300, 350 y 400K, respectivamente. En color negro se muestra el valor if de cada enzima.

8 . DISCUSIÓN

El mejoramiento enzimático, en particular de la termoestabilidad, es necesario como desarrollo de tecnologías que permitan la aplicación de estrategias de biología molecular en procesos industriales que requieran de enzimas y en donde las condiciones de operación óptimas en términos de temperatura sean desfavorables para la estabilidad enzimática de las enzimas que se requiere utilizar. Por otro lado, las estrategias experimentales poseen desventajas sustantivas respecto de los enfoques computacionales dado que requieren de costos económicos asociados al volumen de ensayos, en cambio las herramientas computacionales requieren de elevados costos iniciales dependientes de la infraestructura informática que se implemente pero que el largo plazo estos pueden ser absorbidos por la obtención de resultados sin costos variables asociados a estos. De esta forma, se desarrollo una metodología que busca proponer mutaciones que mejoren la termoestabilidad de enzimas.

Como se mencionó en la Sección 2.2.4, una de las propiedades estructurales de las enzimas incidentes en la termoestabilidad es la flexibilidad. Enzimas más rígidas poseen una estabilidad mayor frente a los efectos desestabilizantes de la temperatura [39, 138].

Para llevar a cabo el mejoramiento enzimático se enfocó el trabajo en la relación entre estas dos propiedades, termoestabilidad y flexibilidad, de esta forma, el procedimiento desarrollado busca reducir la flexibilidad en la enzima y así disminuir el efecto desestabilizador que tiene la temperatura en su estructura y que finalmente puede llevar a la pérdida de su actividad. Sin embargo, es preciso destacar que no será posible rigidizar regiones de una proteína de manera irrestricta. Para eso, es

necesario poseer información básica de la enzima que permita evitar modificaciones que puedan terminar en pérdida de la actividad. Es requisito conocer en cierto grado el funcionamiento a nivel bioquímico de la enzima para protegerla de modificaciones en regiones que puedan alterar tanto estructuras como residuos necesarios para el mecanismo de reacción. En ese sentido, la enzima con la que se trabajó posee subdominios necesarios para su actividad que no fueron objetivo de modificación dado que poseían rigidez elevada, por sobre otras regiones más flexibles, pero por otro lado, las regiones propuestas para ser modificadas son vecinas en su mayoría, a las regiones subdominios que participan de la reacción. Esto puede significar alteraciones en la actividad por efectos distintos al que ejerce la temperatura sobre la estructura como dificultad de acceso del sustrato a túnel para enfrentar el sitio activo.

El diseño racional de enzimas está basado principalmente en la estructura enzimática más que en el comportamiento de esta frente a diferentes factores ambientales que podrían afectar a la estructura [33,141]. Los enfoques computacionales para la obtención de enzimas termoestables, en general, utilizan la información entregada por la dinámica molecular de la enzima para profundizar el conocimiento de la estructura y su comportamiento para, de esta manera enfrentar el problema de la termoestabilización de manera más eficiente y enfocada [97,126]. Otros métodos utilizan los datos de flexibilidad desprendidos de la espectroscopía para la determinación de la estructura de la enzima, NMR [79,115] El trabajo aquí desarrollado agrega un nivel de profundidad en el estudio de la termoestabilidad mayor respecto de los enfoques ya existentes puesto que se realiza un análisis de la dinámica molecular de una enzima a tres diferentes temperaturas además de utilizar la información que se desprende de la dinámica molecular para elaborar un índice de flexibilidad local que permite englobar el comportamiento de la estructura a medida que la temperatura

del sistema aumenta. Este índice de flexibilidad, denominado valor if , permite extraer nueva información de la estructura de una enzima con la utilidad de indicar explícitamente qué regiones son las más flexibles en una enzima y así poder proponer mutaciones que rigidicen estas regiones. De esta manera, la gran diferencia entre el presente trabajo con otros enfoques es la manera como se escogen las regiones a modificar, puesto que aquí son señaladas en base al valor if determinado a partir del análisis de componentes principales de las trayectorias de la enzima a tres diferentes temperaturas.

Se consideró el procedimiento descrito para la simulación como un resultado por sí sólo, debido a que existen diversas versiones de procedimientos de simulación [17, 62, 65, 72, 75, 143] con una estructura general del procedimiento relativamente estándar (inicialización, equilibrio y producción) y en el presente trabajo se diseñó un procedimiento de simulación específico que busca representar de la mejor forma posible los fenómenos que suceden en una enzima en solución acuosa para el propósito de mejorar la termoestabilidad de enzimas que no sólo involucra la simulación de la enzima sino que también la detección de regiones de alta flexibilidad. Para esto se utilizó GROMACS debido a que es una herramienta de libre acceso con uso reportado [4, 18, 62, 64, 101, 117, 122, 128, 139, 143], y en particular, en la caracterización del efecto de la temperatura en enzimas [54, 135].

Las temperaturas utilizadas para las simulaciones que se llevaron a cabo se escogieron para poder representar el efecto de la temperatura en la estructura de la enzima. El hecho de usar temperaturas alejadas de las condiciones fisiológicas reales de las enzimas estudiadas se debió a que el sistema estudiado se idealizó para facilitar el estudio y se aleja también de las condiciones reales en las que se encontrarían las enzimas en el laboratorio o en condiciones de uso industrial; aislando de mejor

manera el efecto de temperatura en la enzima.

Para las etapas en que se utilizó un ensamble NVT, fue posible notar el efecto del termostato en el control de la temperatura, así como también pudo ser observado el efecto del barostato cuando se controló la presión. Las mediciones realizadas para la presión arrojaron variaciones si bien para todas las simulaciones las presiones observadas se mantuvieron en promedio en torno a la presión de referencia, las amplitudes de la variación de la presión fueron, en todos los casos, muy elevadas con valores del RMS cercanos a los 100[bar]. Este comportamiento según los desarrolladores de GROMACS es normal [61].

Se determinó el Rg de las enzimas como medida de la compactación de su estructura. Para comparar el Rg calculado para una enzima a diferentes temperaturas y poder determinar si este valor varía al aumentar la temperatura, se realizó un test t-student comparando las medias del Rg a las diferentes temperaturas con hipótesis nula que las medias son iguales. Para todos los casos, tanto en la validación como con Cel72, se obtuvo que la compactación disminuye al aumentar la temperatura. El t-value obtenido en cada comparación indicó que los Rg, para todas las enzimas, entre una temperatura y otra son, en promedio, significativamente diferentes. Se estimó el promedio de la diferencia entre el Rg a una temperatura con respecto al Rg calculado a otra temperatura obteniendo valores que indicaron aumentos del Rg al aumentar la temperatura, como era de esperarse pues al aumentar la temperatura las vibraciones moleculares aumentan y la compactación de la enzima disminuye.

Se realizaron medidas del RMSf y del valor f de las enzimas simuladas en cada una de las temperaturas. Si bien ambos valores son medidas del desplazamiento de los átomos a lo largo de la producción, los residuos más flexibles poseen un componente principal que abarca la mayor porción de su varianza total, y por otro lado,

el componente principal de residuos menos flexibles cubren una menor porción de su varianza, de este modo, los residuos con valor f similar al su valor de RMSf son menos flexibles [17]. De este modo, mientras el RMSf mide la fluctuación total de los átomos respecto a una estructura promedio, el análisis mediante PCA aísla las principales características del movimiento de un átomo. Es por esto que la presente investigación enfoca la búsqueda de mutantes más termoestables en el cálculo del valor f como herramienta de detección de regiones de alta flexibilidad.

Basándose en el if , uno de los resultados importantes fue el recuperar, en el caso de la enzima T, regiones indicadas en una previa investigación como flexibles [141,142] a partir del ranking de valor if . Se encontró tres regiones identificadas previamente como regiones mutables para reducir la flexibilidad de la enzima con el fin de mejorar su termoestabilidad (regiones C, E y F). Esta disminución de la flexibilidad se debe a que en todos los casos se restringe el movimiento de regiones muy flexibles mediante una unión a regions que presentaron un menor desplazamiento.

Se comparó la media del Rg de las mutantes respecto de la media del Rg de sus respectivas enzimas nativas para evidenciar cambios en la compactación de las enzimas al introducir una mutación. Así, comparando, en este caso, los valores correspondientes a la simulación realizada a la misma temperatura. De este modo se obtuvo que, en todos los casos, la diferencias entre las medias eran significativas. En la etapa de validación se obtuvo que las variantes A, D y F de T aumentaron su compactación respecto de la enzima nativa a una misma temperatura, y para la enzima M, las variantes B y D también mostraron reducir el Rg con respecto a la nativa. El hecho de que las variantes B y C de T no mostraron reducir la compactación de la enzima puede deberse a que las regiones modificadas, como no son necesariamente las más flexibles (no aparecen en el ranking de if) no rigidizan lo

suficiente la estructura como para mantenerla compacta a medida que se aumenta la temperatura, de hecho, estas enzimas en ensayos experimentales no resultaron ser las más termoestables [142], sin embargo, las enzimas A y D si lo fueron y A tiene mutada una de las regiones más flexibles de la enzima, la región C, y es esta enzima la que en los ensayos experimentales mostró una mayor termoestabilidad. Ahora bien, de las tres enzimas propuestas por el procedimiento aquí desarrollado, las variantes F, G y H, sólo la variante F logró disminuir la compactación de la estructura frente al aumento de la temperatura. Esta enzima presenta una mutación en la región más flexible de la enzima. Las variantes G y H de T no presentaron mejorías en términos de la disminución del R_g frente al aumento de la temperatura. La flexibilidad de la variante G de T se disminuye respecto a la enzima nativa a lo largo de toda la enzima, lo que puede significar si bien la estructura se rigidiza la compactación de la enzima aumenta convirtiéndose en una enzima más estable pero menos compacta. Por otro lado, la variante H muestra disminuir su flexibilidad en la región mutada pero aumenta en otras regiones.

Para la enzima M, la variante D posee una mutación que rigidiza dos regiones de alta flexibilidad (C y E) lo que sugiere que por esto que su compactación aumenta respecto a la estructura nativa y su flexibilidad a lo largo de la enzima cambia en comparación con la enzima nativa pero con valores máximos menores que el if de la nativa. Si bien con la variante B de M el R_g disminuye respecto de la enzima nativa, su flexibilidad en términos generales es mayor que la de la enzima nativa. Esto puede deberse a que la mutación presente en la variante B altera la polaridad de la región que la rodea alterando de forma global la flexibilidad de la enzima. Las variantes A y C, si bien poseen mutaciones en las regiones A y B respectivamente, es posible que estas mutaciones fijen una determinada región y liberen los loops vecinos dándole

mayor libertad de movimiento aumentando la flexibilidad a lo largo de la enzima. La variante E presenta una mutación en la región B a diferencia de la enzima A en dónde la mutación es cercana a la región B pero no está en ella. La cercanía de la mutación de la variante E con los loops del canal podría favorecer el impedimento al desplazamiento de aquellos loops rigidizando la estructura, sin embargo, el R_g puede no haberse visto reducido respecto a la enzima nativa debido a que cuando la estructura se estabiliza en una configuración de compactación menor.

La aplicación del procedimiento a las estructuras de las enzimas T y M permitió validar el procedimiento debido a que se pudo recuperar, en el caso de la enzima T, sitios reportados como incidentes en la termoestabilidad de la enzima por su flexibilidad. Permitted también, comprobar la coherencia de los resultados obtenidos a partir del análisis de flexibilidad y así poder identificar patrones concordantes en las tres enzimas analizadas ya que las regiones destacadas como flexibles compartían cierta similitud estructural. Pese a lo anterior, será necesario evaluar la termoestabilidad de las variantes de forma experimental para determinar qué variantes son más termoestables y verificar si los datos experimentales con los que se cuenta son consistentes con el desarrollo computacional aquí propuesto.

De las variantes de Cel72, las enzimas A y B mostraron un menor R_g respecto de la enzima nativa a las temperaturas respectivas, sin embargo, tan solo la enzima B mostró reducir la flexibilidad de la estructura en las regiones que se buscó rigidizar. Esto puede deberse a que la mutación que posee esta enzima rigidiza dos regiones flexibles de la enzima, la región B y la D.

En las tres enzimas estudiadas, mutaciones en regiones flexibles asiladas y alejadas dentro del contexto flexible, es decir que no son vecinas de ninguna otra región flexible ni de loops menos flexibles, presentaron reducción de la flexibilidad en la

zona mutada pero aumento de flexibilidad en otras regiones lo que sugiere que estas mutaciones entregan una mayor libertad de desplazamiento de las otras regiones flexibles aumentando la flexibilidad global.

Desde el punto de vista computacional, la estrategia aquí planteada ofrece la posibilidad de analizar e identificar regiones de alta flexibilidad con costos operacionales bajos y costos computacionales moderados lo que hace posible que se pueda analizar enzimas de gran tamaño y evaluar numerosas cantidades de variantes sin aumentar costos de manera considerable, y, así, poder evaluar otros tipos de mutaciones, no solo puentes disulfuro, si no, evaluar todas las posibles modificaciones en paralelo y así, encontrar variantes óptimas, que rigidicen las regiones flexibles con resultados de termoestabilidad mejores.

Las enzimas que mostraron mayor compactación en relación a sus respectivas enzimas nativas, a las tres temperaturas simuladas, fueron las variantes A, D y F de la enzima T; las variantes B y D de la enzima M, y las variantes A y B de Cel72. Sin embargo, sólo la variante G de T y B de Cel72 mostraron menor flexibilidad a lo largo de toda la estructura. Las enzimas A, E, F y H, variantes de T, A, D y E, variantes de M, y las variantes A, B y C de Cel72, mostraron menor flexibilidad en la región que en las nativas respectivas habían sido identificadas como regiones máxima flexibilidad, sin embargo, mostraron incrementar la flexibilidad en otras regiones.

Esto sugiere que la relación entre la flexibilidad y la compactación de las enzimas no es necesariamente directa, lo que indica que al aumentar la temperatura, la estructura puede alcanzar configuraciones menos compactas y a la vez más estables en términos de flexibilidad, es decir, alcanza una configuración estructural menos compacta pero más rígida. Así también es posible que al introducir mutaciones rigidizantes, la tensión producto del aumento de la temperatura se libera en regiones

menos rígidas y menos estables, aumentando su flexibilidad y con posibles efectos negativos en la flexibilidad global de la estructura.

Lo anterior advierte la idea de diseñar enzima con propiedades de denaturarse por el efecto de la temperatura de manera programada, es decir, controlar la forma en que la estructura pierde su configuración ideal de manera de permitir a determinadas regiones ser más flexibles que otras y así, que la pérdida de su configuración tridimensional esté controlada para asegurar que la actividad de la enzima permanezca el mayor tiempo posible bajo condiciones de temperatura desfavorables. Para esto, el análisis de flexibilidad deberá ser más completo y abarcar mutaciones en diferentes regiones de la enzima al mismo tiempo.

De esta manera, será necesario para posteriores trabajos buscar indicadores que reflejen de mejor manera que el R_g la termoestabilidad de las enzimas, y que pueda ser obtenido a base de los datos obtenidos de la trayectoria de las estructuras simuladas. Por otro lado, será necesario calibrar estas mediciones contrastando los resultados de valores if y de termoestabilidad, con resultados de ensayos experimentales con el fin de comprobar el efecto de las mutaciones sugeridas por el procedimiento aquí desarrollado en la termoestabilidad de las enzimas.

9 . CONCLUSIONES

Se diseñó una estrategia basada en la dinámica de la estructura enzimática que permite la proposición de mutaciones rigidizantes de regiones flexibles en la estructura de una enzima. Para poder afirmar si estas mutaciones son termostabilizantes será necesario continuar el estudio en un campo experimental para poder validar la metodología. El procedimiento desarrollado consiste en una primera simulación del comportamiento de las enzimas a tres temperaturas diferentes escogidas como 300, 350, y 400K, luego se determina la flexibilidad a lo largo de la estructura, posteriormente se escogen las regiones a rigidizar y finalmente se repite el procedimiento con las mutantes diseñadas para, de esta forma, evaluar la flexibilidad de las nuevas variantes. Para determinar la flexibilidad de las estructuras se creó un nuevo índice de flexibilidad basado en la varianza asociada al componente principal de cada átomo. Esta varianza se determinó para cada residuo, a cada temperatura, para luego ser combinados en un solo valor *if* que entrega la flexibilidad para cada residuo.

El procedimiento se validó mediante la simulación de la estructura de dos enzimas, Cel7A de *T. emersonii* y Cel7B de *M. albomyces*. La flexibilidad de estas enzimas fue analizada mediante el cálculo de los valores *if*. Para la enzima Cel7A de *T. emersonii* había cinco enzimas variantes de ella en las que las mutaciones mejoraban la termoestabilidad de la enzima, tres de ellas fueron identificadas en las regiones más flexibles de la enzima. Para Cel7B de *M. albomyces* además de las enzimas mutantes propuestas por el procedimiento se analizaron dos más reportadas como mutantes que buscaban mejorar la termoestabilidad. De esas enzimas ninguna se encontraba cerca de alguna región identificada como flexible ni tampoco mostraron mejorar la flexibilidad de la enzima. De las seis enzimas variantes analizadas en la validación

cuatro mostraron reducir la flexibilidad (T-G, T-H, M-D y M-E).

La compactación de las enzimas se determinó a base del cálculo del Rg. Se obtuvo que en todas las simulaciones el Rg aumenta de manera significativa ($p < 0.0001$) al aumentar la temperatura. Se utilizó este valor para analizar las mutantes y compararlas con respecto a las enzimas nativas. No se observó una relación directa entre la compactación de las enzimas y su flexibilidad.

Del análisis de la enzima Cel72, se partió de una estructura desconocida para finalmente proponer tres mutaciones, las cuales solo una muestra una mejoría en la flexibilidad a lo largo de la enzima, la enzima B. Según el análisis de Rg de las mutaciones de las variantes se determinó que las variantes A y B lograron disminuir el Rg respecto de la enzima nativa.

Si bien fue posible notar el efecto de las variaciones de las enzimas en la flexibilidad de la estructura, serán necesarios estudios experimentales para poder establecer la relación entre los cálculos de flexibilidad, las modificaciones y la actividad de las enzimas para determinar el grado de mejora que se obtiene en base a la aplicación de la metodología aquí propuesta. Sin embargo, las proyecciones de la metodología es amplia dado que permitiría proponer modificaciones puntuales para la termoestabilización de enzimas de estructura desconocida, sin la necesidad de realizar ensayos experimentales explorativos, reduciendo de esta manera el *screening* presentes en las estrategias experimentales de mejoramiento enzimático.

Las enzimas de uso industrial son sometidas a condiciones de operación que extrapolan sus condiciones óptimas. La metodología aquí propuesta posee el potencial de mejorar la termoestabilidad de manera rápida y sin la necesidad de recurrir a ensayos experimentales para *screening*. Esto ofrece un nuevo enfoque que amplía las capacidades que tiene la perspectiva experimental debido a que será posible estu-

diar el efecto de múltiples variaciones puntuales de manera simultánea explorando diversas combinaciones de residuos. La metodología se apoya en el cálculo computacional lo que entrega la posibilidad de multiplicar el número de posibles variantes y de regiones a modificar sin necesariamente aumentar el costo de llevarlo a cabo, potenciando la aplicabilidad del procedimiento.

10 . TEMAS DE INVESTIGACIÓN FUTURA

Para un completo análisis el trabajo deberá realizar el estudio con un espectro de mutaciones más amplio que solo puentes disulfuro. Debido a como se estructura el trabajo y a que éste se sustenta de herramientas computacionales para la obtención de datos, será posible aumentar el espacio de variantes a todos los aminoácidos en paralelo sin tener que modificar el procedimiento. Debido a que el efecto de una mutación no es solo local si no que afecta a la estructura completa en mayor o menor grado, así, se podrá estudiar si hay otros tipos de mutaciones que producen efectos no esperados en la estructuras y que sean favorables para el objetivo que se persigue y que es mejorar la termoestabilidad.

Por otro lado, será necesario validar experimentalmente el procedimiento y evaluar las variantes propuestas por este a modo de determinar de una mejor manera el grado de mejoramiento. También puede ser posible buscar mediante el análisis experimental de la termoestabilidad un índice que pueda ser calculado a base de los datos arrojados por las simulaciones computacionales y así poder prescindir de los ensayos de laboratorio.

El hallazgo de nuevas regiones flexibles luego de la rigidización de las zonas más flexibles en una enzima sugiere la posibilidad de estudiar la denaturación por temperatura de enzimas de forma programada y así controlar la forma en que una enzima se denatura para mantener, en lo posible, la actividad enzimática, y por ende si termoestabilidad.

Del presente trabajo se obtuvo que las mutaciones propuestas a partir del procedimiento propuesto logran reducir la flexibilidad de la enzima, sin embargo el objetivo final de estas mutaciones es mejorar la termoestabilidad de las enzimas por lo que

se necesita del trabajo experimental para validar estos resultados en términos de la termoestabilidad.

Bibliografía

- [1] *The pymol molecular graphics system, version 1.5.0.1*, Schrödinger, LLC.
- [2] *112th annual meeting of the american association of colleges of pharmacy, san antonio, tx, july 9-13, 2011*, Am J Pharm Educ, 75 (2011), p. 105.
- [3] D. J. ADAMSA, E. M. ADAMSA, AND G. J. HILLSA, *The computer simulation of polar liquids*, Mol Phys, 38 (1979), pp. 387–400.
- [4] S. ADCOCK AND J. A. MCCAMMON, *Molecular dynamics: survey of methods for simulating the activity of proteins*, Chemical reviews, 106 (2006), pp. 1589–1615.
- [5] S. F. ALTSCHUL, W. GISH, W. MILLER, E. W. MYERS, AND D. J. LIPMAN, *Basic local alignment search tool*, J Mol Biol, 215 (1990), pp. 403–410.
- [6] S. F. ALTSCHUL, T. L. MADDEN, A. A. SCHÄFFER, J. ZHANG, Z. ZHANG, W. MILLER, AND D. J. LIPMAN, *Gapped blast and psi-blast: a new generation of protein database search programs*, Nucleic Acids Res, 25 (1997), pp. 3389–3402.
- [7] S. F. ALTSCHUL, J. C. WOOTTON, E. M. GERTZ, R. AGARWALA, A. MORGULIS, A. A. SCHÄFFER, AND Y.-K. YU, *Protein database searches using compositionally adjusted substitution matrices*, FEBS J, 272 (2005), pp. 5101–5109.
- [8] C. B. ANFINSEN, *Principles that govern the folding of protein chains*, Science, 181 (1973), pp. 223–230.

-
- [9] C. B. ANFINSEN AND H. A. SCHERAGA, *Experimental and theoretical aspects of protein folding*, Adv Protein Chem, 29 (1975), pp. 205–300.
- [10] K. ARNOLD, L. B. J. KOPP, AND T. SCHWEDE, *The swiss-model workspace: A web-based environment for protein structure homology modelling*, Bioinformatics, 22 (2006), pp. 195–201.
- [11] F. AVBELJ AND R. L. BALDWIN, *Role of backbone solvation in determining thermodynamic β propensities of the amino acids*, Proc Natl Acad Sci USA, 99 (2002), pp. 1309–1313.
- [12] P. A. BATES, L. A. KELLEY, R. M. MACCALLUM, AND M. J. E. STERNBERG, *Enhancement of protein modelling by human intervention in applying the automatic programs 3d-jigsaw and 3d-pssm*. proteins: Structure, function and genetics, Suppl, 5 (2001), pp. 39–46.
- [13] P. A. BATES AND M. J. E. STERNBERG, *Model building by comparison at casp3: Using expert knowledge and computer automation*. proteins: Structure, function and genetics, Suppl, 3 (1999), pp. 47–54.
- [14] H. BEKKER, E. J. DIJKSTRA, M. K. R. RENARDUS, AND H. J. C. BERENDSEN, *An efficient, box shape independent non-bonded force and virial algorithm for molecular dynamics*, Molecular Simulation, 14 (1995), pp. 137–151.
- [15] P. BENKERT, M. BIASINI, AND T. SCHWEDE, *Toward the estimation of the absolute quality of individual protein structure models*, Bioinformatics, 27 (2011), pp. 343–350.
- [16] Z. BENKO, E. DRAHOS, Z. SZENGYEL, T. PURANEN, J. VEHMAANPERA, AND K. RECZEY, *Thermoascus aurantiacus cbhi/cel7a production in tricho-*

-
- derma reesei* on alternative carbon sources, Applied Biochemistry and Biotechnology, 137–140 (2007), pp. 195–204.
- [17] N. C. BENSON AND V. DAGGETT, *Dynameomics: Large-scale assessment of native protein flexibility*, Protein Sci, 17 (2008), pp. 2038–2050.
- [18] R. W. BENZ, F. CASTRO-ROMÁN, D. J. TOBIAS, AND S. H. WHITE, *Experimental validation of molecular dynamics simulations of lipid bilayers: a new approach*, Biophysical journal, 88 (2005), pp. 805–815.
- [19] H. J. C. BERENDSEN, D. VAN DER SPOEL, AND R. VAN DRUNEN, *Gromacs: A message-passing parallel molecular dynamics implementation*, Comp Phys Comm, 91 (1995), pp. 43–56.
- [20] T. C. BISHOP, *Molecular dynamics simulations of a nucleosome and free dna*, Biomolecular Structure and Dynamics, 22 (2005), pp. 673–686.
- [21] C. BOISSET, H. CHANZY, B. HENRISSAT, R. LAMED, Y. SHOHAM, AND E. A. BAYER, *Digestion of crystalline cellulose substrates by the clostridium thermocellum cellulosome: structural and morphological aspects*, Biochem J, 340 (1999), pp. 829–835.
- [22] R. BOURGAULT, A. J. OAKLEY, J. D. BEWLEY, AND M. C. J. WILCE, *Three-dimensional structure of (1,4)- β -d-mannan mannanohydrolase from tomato fruit*, Protein Science, 14 (2005), pp. 1233–1241.
- [23] J. T. BRENNAN, N. S. JR, A. J. SINCLAIR, AND S. C. CUNNANE, *α -linolenic acid supplementation and conversion to n-3 long-chain polyunsaturated fatty acids in humans*, Prostaglandins Leukot Essent Fatty Acids, 80 (2009), pp. 85–91.

-
- [24] V. BROWER, *Going global in r & d* , EMBO Rep., 5 (2004), pp. 333–335.
- [25] J. D. BRYNGELSON AND P. G. WOLYNES, *Spin glasses and the statistical mechanics of protein folding*, Proc Natl Acad Sci USA, (1987), pp. 7524–7528.
- [26] G. BUSSI, D. DONADIO, AND M. PARRINELLO, *Canonical sampling through velocity rescaling*, Chem Phys, 126 (2007), p. 014101.
- [27] M. P. BYRNE, R. L. MANUEL, L. G. LOWE, AND W. E. STITES, *Energetic contribution of side chain hydrogen bonding to the stability of staphylococcal nuclease*, Biochemistry, 34 (1995), pp. 13949–13960.
- [28] C. R. CARERE, R. SPARLING, N. CICEK, AND D. B. LEVIN, *Third generation biofuels via direct cellulose fermentation*, Int. J. Mol. Sci., 9 (2008), pp. 1342–1360.
- [29] R. L. CHANG, L. GHAMSARI, A. MANICHAIKUL, E. F. Y. HOM, S. BALAJI, W. FU, Y. SHEN, T. HAO, B. Å. PALSSON, K. SALEHI-ASHTIANI, AND J. A. PAPIN, *Metabolic network reconstruction of chlamydomonas offers insight into light-driven algal metabolism*, Molecular Systems Biology, 7 (2011).
- [30] R. L. CHANG, L. XIE, L. XIE, P. E. BOURNE, AND B. Å. PALSSON, *Drug off-target effects predicted using structural analysis in the context of a metabolic network model*, PLoS Comput Biol, 6 (2010), p. e1000938.
- [31] J. CHEN AND W. E. STITES, *Packing is a key selection factor in the evolution of protein hydrophobic cores*, Biochemistry, 40 (2001), pp. 15280–15289.
- [32] Z. CHEN, H.-G. WANG, Z.-J. WEN, AND Y. WANG, *Life sciences and biotechnology in china*, Phil. Trans. R. Soc., 362 (2007), pp. 947–957.

- [33] D. CHOUDHURY, S. BISWAS, S. ROY, AND J. K. DATTA GUPTA, *Improving thermostability of papain through structure-based protein engineering*, Protein Eng Des Sel, 23 (2010), pp. 457–467.
- [34] W. M. COCO, W. E. LEVINSON, M. J. CRIST, H. J. HEKTOR, A. DARZINS, P. T. PIENKOS, C. H. SQUIRES, AND D. J. MONTICELLO, *Dna shuffling method for generating highly recombined genes and evolved enzymes*, Nat Biotechnol, 19 (2001), pp. 354–359.
- [35] C. M. COLLINS, P. G. MURRAY, S. DENMAN, J. P. MORRISSEY, L. BYRNES, T. T. TEERI, AND M. G. TUOHY, *Molecular cloning and expression analysis of two distinct β -glucosidase genes, *bg1* and *aven1*, with very different biological roles from the thermophilic, saprophytic fungus *Talaromyces emersonii**, Mycological Research, 111 (2007), pp. 840–849.
- [36] B. CONTRERAS-MOREIRA AND P. A. BATES, *Domain fishing: a first step in protein comparative modelling*, Bioinformatics, 18 (2002), pp. 1141–1142.
- [37] A. CRAMERI, S. A. RAILLARD, E. BERMUDEZ, AND W. P. STEMMER, *Dna shuffling of a family of genes from diverse species accelerates directed evolution*, Nature, 391 (1998), pp. 288–291.
- [38] T. E. CREIGHTON, *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company, New York, 2a. ed. ed., 1992.
- [39] M. J. DANSON, D. W. HOUGH, G. L. TAYLOR, AND L. PEARL, *Enzyme thermostability and thermoactivity*, Protein Engineering, 9 (1996), pp. 629–630.

-
- [40] X. DAURA, A. E. MARK, AND W. F. V. GUNSTEREN, *Parametrization of aliphatic ch_N united atoms of gromos96 force field*, Journal of Computational Chemistry, 19 (1998), pp. 535–547.
- [41] R. DAY, B. J. BENNION, S. HAM, AND V. DAGGETT, *Increasing temperature accelerates protein unfolding without changing the pathway of unfolding*, J Mol Biol, 322 (2002), pp. 189–203.
- [42] P. I. W. DE BAKKER, P. H. HUNENBERGER, AND J. A. MCCAMMON, *Molecular dynamics simulations of the hyperthermophilic protein sac7d from sulfobolus acidocaldarius: Contribution of salt bridges to thermostability*, J Mol Biol, 285 (1999), pp. 1811–1830.
- [43] S. DEECHONGKIT, P. E. DAWSON, AND J. W. KELLY, *Toward assessing the position-dependent contributions of backbone hydrogen bonding to β -sheet folding thermodynamics employing amide-to-ester perturbations*, J Am Chem Soc, 126 (2004), pp. 16762–16771.
- [44] M. DESVAUX, *Clostridium cellulolyticum: model organism of mesophilic cellulolytic clostridia*, FEMS Microbiology Reviews, 29 (2005), pp. 741–764.
- [45] K. A. DILL, *Theory for the folding and stability of globular proteins*, Biochemistry, 24 (1985), pp. 1501–1509.
- [46] —, *Dominant forces in protein folding*, Biochemistry, 29 (1990), pp. 7133–7155.
- [47] —, *Polymer principles and protein folding*, Protein Sci, 8 (1999), pp. 1166–1180.

-
- [48] K. A. DILL, D. O. V. ALONSO, AND K. HUTCHINSON, *Thermal stabilities of globular proteins*, *Biochemistry*, 28 (1989), pp. 5439–5449.
- [49] K. A. DILL AND H. S. CHAN, *From levinthal to pathways to funnels*, *Nat Struct Biol*, 4 (1997), pp. 10–19.
- [50] C. DIVNE, J. STAHLBERG, T. T. TEERI, AND T. A. JONES, *High-resolution crystal structures reveal how a cellulose chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei**, *J Mol Biol*, 275 (1998), pp. 309–325.
- [51] U. ESSMAN, L. PERELA, M. L. BERKOWITZ, T. DARDEN, H. LEE, AND L. G. PEDERSEN, *A smooth particle mesh ewald method*, *Chem Phys*, 103 (1995), pp. 8577–8592.
- [52] N. ESWAR, M. A. MARTI-RENOM, B. WEBB, M. S. MADHUSUDHAN, D. ERAMIAN, M. SHEN, U. PIEPER, AND A. SALI, *Comparative protein structure modeling with modeller*, *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., 15 (2006), pp. 5.6.1–5.6.30.
- [53] X. FANG, A. WALLQVIST, AND J. REIFMAN, *Development and analysis of an in vivo-compatible metabolic network of *Mycobacterium tuberculosis**, *BMC Systems Biology*, 4 (2010).
- [54] J. FERNER, A. VILLA, E. DUCHARDT, E. WIDJAJAKUSUMA, J. WOHNERT, G. STOCK, AND H. SCHWALBE, *Nmr and md studies of the temperature-dependent dynamics of rna ynmg-tetraloops*, *Nucleic Acids Res*, 36 (2008), pp. 1928–1940.

-
- [55] A. R. FERSHT, J. SHI, J. KNILL-JONES, D. M. LOWE, A. J. WILKINSON, D. M. BLOW, P. BRICK, P. CARTER, M. M. Y. WAYE, AND G. WINTER, *Hydrogen bonding and biological specificity analysed by protein engineering*, Nature, 314 (1985), pp. 235–238.
- [56] A. FISER, R. DO, AND A. SALI, *Modeling of loops in protein structures*, Protein Science, 9 (2000), pp. 1753–1773.
- [57] R. FLEIX AND L. G. LJUNGDAHL, *The cellulosome: the exocellular organelle of clostridium*, Annual Reviews of Microbiology, 47 (1993), pp. 791–819.
- [58] M. GAVRILESCU AND Y. CHISTI, *Biotechnology-a sustainable alternative for chemical industry*, Biotechnol Adv., 23 (2005), pp. 471–499.
- [59] T. G. IBRAGIMOVA AND R. C. WADE, *Importance of explicit salt ions for protein stability in molecular dynamics simulation*, The Biophysical Society, 74 (1998), pp. 2906–2911.
- [60] A. GRASSICK, P. G. MURRAY, R. THOMPSON, C. M. COLLINS, L. BYRNES, G. BIRrane, T. M. HIGGINS, AND M. G. TUOHY, *Three-dimensional structure of a thermostable native cellobiohydrolase, cbh ib, and molecular characterization of the cel7 gene from the filamentous fungus, Talaromyces emersonii*, European Journal of Biochemistry, 271 (2004), pp. 4495–4506.
- [61] GROMACS, *Pressure*. [en línea] <<http://www.gromacs.org/Documentation/Terminology/Pressure?highlight=pressure>>, 2011. [consulta: 3 de abril 2012].
- [62] A. GROTTESI, M. CERUSO, A. COLOSIMO, AND A. D. NOLA, *Molecular dynamics study of a hyperthermophilic and a mesophilic rubredoxin*, Journal of Chemical Physics, 294 (2002), pp. 287–294.

-
- [63] E. HABER AND C. B. ANFINSEN, *Side-chain interactions governing the pairing of half-cystine residues in ribonuclease*, J Biol Chem, 237 (1962), pp. 1839–1844.
- [64] T. H. HAMID, R. N. RAHMAN, A. B. SALLEH, AND M. BASRI, *The role of lid in protein-solvent interaction of the simulated solvent stable thermostable lipase from Bacillus strain 42 in water-solvent mixtures*, Biotechnol & Biotechnol Eq, 23 (2009), pp. 1524–1530.
- [65] S. HARRIS, Z. SANDS, AND C. LAUGHTON, *Molecular dynamics simulations of duplex stretching reveal the importance of entropy in determining the biomechanical properties of dna*, Biophys J, 88 (2005), pp. 1684–1691.
- [66] B. HESS, H. BEKKER, H. J. C. BERENDSEN, AND J. G. E. M. FRAAIJE, *Lincs: a linear constraint solver for molecular simulations*, J Comp Chem, 18 (1997), pp. 1463–1472.
- [67] B. HESS, C. KUTZNER, D. VAN DER SPOEL, AND E. LINDAHL, *Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation*, J Comp Chem, 4 (2008), pp. 435–447.
- [68] S. HIRAI, N. TAKAHASHI, T. GOTO, S. LIN, T. UEMURA, R. YU, AND T. KAWADA, *Functional food targeting the regulation of obesity-induced inflammatory responses and pathologies*, Mediator Inflamm, 367838 (2010).
- [69] R. W. HOCKNEY, S. GOEL, AND J. EASTWOOD, *Quiet high-resolution computer models of a plasma*, J Comp Phys, 14 (1973), pp. 148–158.
- [70] I. T. HORVÁTH AND P. T. ANASTAS, *Innovations and green chemistry*, Chem Rev., 107 (2007), pp. 2169–2173.

-
- [71] H. HOTELLING, *Analysis of a complex of statistical variables into principal components*, Journal of Educational Psychology, 24 (1933), pp. 417–441 and 498–520.
- [72] M. K. JONES, A. CATTE, J. C. PATTERSON, F. GU, J. CHEN, L. LI, AND J. P. SEGREST, *Thermal stability of apolipoprotein a-i in high-density lipoproteins by molecular dynamics*, Biophysical journal, 96 (2009), pp. 353–371.
- [73] T. A. A. KADIR, A. A. MANNAN, A. M. KIERZEK, J. MCFADDEN, AND K. SHIMIZU, *Modeling and simulation of the main metabolism in escherichia coli and its several single-gene knockout mutants with experimental verification*, Microb Cell Fact, 9 (2010).
- [74] K. KAITIN, *The landscape for pharmaceutical innovation: Drivers of cost-effective clinical research*, Pharm Outsourcing, (2010).
- [75] R. A. KARJIBAN, M. B. A. RAHMAN, M. BASRI, A. B. SALLEH, D. JACOBS, AND H. A. WAHAB, *Molecular dynamics study of the structure, flexibility and dynamics of thermostable l1 lipase at high temperatures*, The Protein Journal, 28 (2009), pp. 14–23.
- [76] M. KARPLUS AND D. L. WEAVER, *Diffusion-collision model for protein folding*, Biopolymers, 18 (1979), pp. 1421–1437.
- [77] —, *Protein folding dynamics: the diffusion-collision model and experimental data*, Protein Sci, 3 (1994), pp. 650–668.
- [78] F. KIEFER, K. ARNOLD, M. KÄIJNZLI, L. BORDOLI, AND T. SCHWEDE, *The swiss-model repository and associated resources*, Nucleic Acids Research, 37 (2009), pp. D387–D392.

-
- [79] S. J. KIM, J. A. LEE, J. C. JOO, Y. J. YOO, Y. H. KIM, AND B. K. SONG, *The development of a thermostable cip (coprinus cinereus peroxidase) through in silico design*, *Biotechnol Prog*, 26 (2010), pp. 1038–1046.
- [80] K. KLEIN AND D. G. LEROY, *The biofuels frenzy: What's in it for canadian agriculture?*, annual conference of alberta institute of agrologists, Alberta Institute of Agrologists, 2007.
- [81] J. A. KOVACS, P. CHACON, AND R. ABAGYAN, *Predictions of protein flexibility: First-order measures*, *Proteins*, 56 (2004), pp. 661–668.
- [82] C. P. KUBICEK, *From cellulose to cellulase inducers: facts and fiction*, in *Proceedings of the 2nd Symposium Trichoderma Reesei Cellulases and Other Hydrolases (TRICEL '93)*, vol. 8, Espoo, Finland, 1993, pp. 181–188.
- [83] R. C. KUHAD, R. GUPTA, AND Y. P. KHASA, *Bioethanol production from lignocellulosic biomass: an overview*, in *Wealth from Waste*, New Delhi, India, 2010, Teri Press.
- [84] R. C. KUHAD, M. MANCHANDA, AND A. SINGH, *Hydrolytic potential of extracellular enzymes from a mutant strain of fusarium oxysporum*, *Bioprocess Engineering*, 20 (1999), pp. 133–135.
- [85] L. KULAKOVA, A. GALKIN, T. NAKAYAMA, T. NISHINO, AND N. ESAKI, *Improvement of thermostability of cold-active serine alkaline protease from the psychrotrophic bacterium shewanella sp. strain ac10 by rational mutagenesis*, *Journal of Molecular Catalysis*, 22 (2003), pp. 113–117.
- [86] R. KUMAR AND R. NUSSINOV, *How do thermophilic proteins deal with heat?*, *Cell Mol Life Sci*, 58 (2001), pp. 1216–1233.

-
- [87] R. A. LASKOWSKI, M. W. MACARTHUR, D. S. MOSS, AND J. M. THORNTON, *Procheck: a program to check the stereochemical quality of protein structures*, J Appl Cryst, 26 (1993), pp. 283–291.
- [88] P. E. LEOPOLD, M. MONTAL, AND J. N. ONUCHIC, *Protein folding funnels: a kinetic approach to the sequence-structure relationship*, Proc Natl Acad Sci USA, 89 (1992), pp. 8721–8725.
- [89] D.-C. LI, A.-N. LI, AND A. C. PAPAGEORGIOU, *Cellulases from thermophilic fungi: Recent insights and biotechnological potential*, Enzyme Research, 2011 (2011).
- [90] E. LINDAHL, B. HESS, AND D. VAND DER SPOEL, *Gromacs 3.0: A package for molecular simulation and trajectory analysis*, J. Mol. Mod, 7 (2001), pp. 306–317.
- [91] H. L. LIU AND W. C. WANG, *Protein engineering to improve the thermostability of glucoamylase from aspergillus awamori based on molecular dynamics simulations*, Protein Engineering, 6 (2003), pp. 19–25.
- [92] T. LONHIENNE, C. GERDAY, AND G. FELLER, *Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility*, Biochim Biophys Acta, 1543 (2000), pp. 1–10.
- [93] R. MAHESHWARI, G. BHARADWAJ, AND M. K. BHAT, *Thermophilic fungi: their physiology and enzymes*, Microbiol Mol Biol Rev, 64 (2000), pp. 461–488.
- [94] M. A. MARTI-RENO, A. STUART, A. FISER, R. SÁNCHEZ, F. MELO, AND A. SALI, *Comparative protein structure modeling of genes and genomes*, Annu. Rev. Biophys. Biomol. Struct, 29 (2000), pp. 291–325.

-
- [95] J. A. MCCAMMON, B. R. GELIN, AND M. KARPLUS, *Dynamics of folded proteins*, Nature, 267 (1977), pp. 585–590.
- [96] M. MEDIĆ-ŠARIĆ, V. RASTIJA, M. BOJIĆ, AND ŽELJAN MALEĀA, *From functional food to medicinal product: Systematic approach in analysis of polyphenolics from propolis and wine*, Nutrition Journal, 8 (2009).
- [97] Y. T. MEHARENNA AND T. L. POULOS, *Using molecular dynamics to probe the structural basis for enhanced stability in thermal stable cytochromes p450*, Biochemistry, 49 (2010), pp. 6680–6686.
- [98] A. MEISTER, *Biochemistry of the Amino Acids*, vol. 1 y 2, Academic Press, Inc, New York, 2a. ed. ed., 1965.
- [99] F. MELO AND E. FEYTMANS, *Assessing protein structures with a non-local atomic interaction energy*, J Mol Biol, 277 (1998), pp. 1141–1152.
- [100] J. H. MISSIMER, M. STEINMETZ, R. BARON, F. WINKLER, R. KAMMERER, X. DAURA, AND W. VAN GUNSTEREN, *Configurational entropy elucidates the role of salt-bridge networks in protein thermostability*, Prot Sci, 16 (2007), pp. 1349–1359.
- [101] N. MOGHARRAB, H. GHOURCHIAN, AND M. AMININASAB, *Structural stabilization and functional improvement of horseradish peroxidase upon modification of accessible lysines: experiments and simulation*, Biophysical journal, 92 (2007), pp. 1192–1203.
- [102] E. MONSERRATE, S. B. LESCHINE, AND E. CANALE-PAROLA, *Clostridium hungatei sp. nov., a mesophilic, n₂-fixing cellulolytic bacterium isolated from*

- soil*, International Journal of Systematic and Evolutionary Microbiology, 51 (2001), pp. 123–132.
- [103] P. G. MURRAY, C. M. COLLINS, A. GRASSICK, AND M. G. TUOHY, *Molecular cloning, transcriptional, and expression analysis of the first cellulase gene (cbh2), encoding cellobiohydrolase ii, from the moderately thermophilic fungus Talaromyces emersonii and structure prediction of the gene product*, Biochem Biophys Res Commun, 301 (2003), pp. 280–286.
- [104] J. K. MYERS, T. G., AND OAS, *Preorganized secondary structure as an important determinant of fast protein folding*, Nat Struct Biol, 8 (2001), pp. 552–558.
- [105] J. E. NESS, S. KIM, A. GOTTMAN, R. PAK, A. KREBBER, T. V. BORCHERT, S. GOVINDARAJAN, E. C. MUNDORFF, AND J. MINSHULL, *Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently*, Nature Biotechnology, 20 (2002), pp. 1251–1255.
- [106] C. OOSTENBRINK, A. VILLA, A. E. MARK, AND W. F. V. GUNSTEREN, *A biomolecular force field based on the free enthalpy of hydration and solvation: The gromos force-field parameter sets 53a5 and 53a6*, Journal of Computational Chemistry, 25 (2004), pp. 1656–1676.
- [107] E. PAPALEO, L. RICCARDI, P. FANTUCCI, C. VILLA, AND L. D. GIOIA, *Flexibility and enzymatic cold-adaptation: a comparative molecular dynamics investigation of the elastase family*, Biochimica Biophys Acta, 1764 (2006), pp. 1397–1406.
- [108] T. PARKKINEN, A. KOIVULA, J. VEHEMAANPERA, AND J. ROUVINEN, *Crystal structures of Melanocarpus albomyces cellobiohydrolase cel7b in complex with*

-
- cello-oligomers show high flexibility in the substrate binding*, Protein Sci, 17 (2008), pp. 1383–1394.
- [109] S. PATKAR, J. VIND, E. KELSTRUP, M. W. CHRISTENSEN, A. SVENDSEN, K. BORCH, AND O. KIRK, *Effect of mutations in candida antarctica b lipase*, Chem Phys Lipids, 93 (1998), pp. 95–101.
- [110] K. PEARSON, *On lines and planes of closest fit to systems of points in space*, Philosophical Magazine, 2 (1901), pp. 559–572.
- [111] M. C. PEITSCH, *Protein modeling by e-mail*, Bio/Technology, 13 (1995), pp. 658–660.
- [112] M. G. PIKKEMAAT, A. B. LINSSEN, H. J. BERENDSEN, AND D. B. JANSSEN, *Molecular dynamics simulations as a tool for improving protein stability*, Protein Engineering, 15 (2002), pp. 185–192.
- [113] M. J. POCAS-FONSECA, I. SILVA-PEREIRA, B. B. ROCHA, AND M. D. O. AZEVEDO, *Substrate-dependent differential expression of humicola grisea var. thermoidea cellobiohydrolase genes*, Canadian Journal of Microbiology, 46 (2000), pp. 749–752.
- [114] A. B. PORTO-PAZOS, N. VEIGUELA, P. MESEJO, M. NAVARRETE, A. ALVARRELLOS, O. I. NEZ, A. PAZOS, AND A. ARAQUE, *Artificial astrocytes improve neural network performance*, PLoS ONE, 6 (2011), p. e19109.
- [115] M. REETZ AND J. CARBALLEIRA, *Iterative saturation mutagenesis (ism) for rapid directed evolution of functional enzymes*, Nature protocols, 2 (2007), pp. 891–903.

-
- [116] M. T. REETZ1 AND J. D. CARBALLEIRA, *Iterative saturation mutagenesis (ism) for rapid directed evolution of functional enzymes*, Nature Protocols, on line, 2 (2007), pp. 891–903.
- [117] A. G. ROCCO, L. MOLLIKA, P. RICCHIUTO, A. M. BAPTISTA, E. GIANAZZA, AND I. EBERINI, *Characterization of the protein unfolding processes induced by urea and temperature*, Biophysical journal, 94 (2008), pp. 2241–2251.
- [118] A. SALI AND T. BLUNDELL., *Comparative protein modelling by satisfaction of spatial restraints*, J Mol Biol, 234 (1993), pp. 779–815.
- [119] L. SANG-MOK AND Y. M. KOO, *Pilot-scale production of cellulase using trichoderma reesei rut c-30 in fed-batch mode*, Journal of Microbiology and Biotechnology, 11 (2001), pp. 229–233.
- [120] H. SCHEPENS, *White biotechnology: gateway to a more sustainable future*, EuropaBio, Lyon France, (2003).
- [121] C. SCHUBERT, *Can biofuels finally take center stage?*, Nat Biotechnol, 24 (2006), pp. 777–784.
- [122] C. M. SHEPHERD AND H. J. VOGEL, *A molecular dynamics study of ca(2+)-calmodulin: evidence of interdomain coupling and structural collapse on the nanosecond timescale*, Biophysical journal, 87 (2004), pp. 780–791.
- [123] A. SHINKAI, A. HIRANO, AND K. AISAKA, *Substitutions of ser for asn-163 and pro for leu-264 are important for stabilization of lipase from pseudomonas aeruginosa*, J Biochem, 120 (1996), pp. 915–921.

-
- [124] J. L. SOHL, S. S. JASWAL, AND D. A. AGARD, *Unfolded conformations of alpha-lytic protease are more stable than its native state*, *Nature*, 395 (1998), pp. 817–817.
- [125] W. P. STEMMER, *Dna shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution*, *Proc Natl Acad Sci U S A*, 91 (1994), pp. 10747–10751.
- [126] F. STERPONE AND S. MELCHIONNA, *Thermophilic proteins: insight and perspective from in silico experiments*, *Chem Soc Rev*, 41 (2012), pp. 1665–1676.
- [127] R. D. STEWART AND W. BAIR, *Spiking neural network simulation: numerical integration with the parker-sochacki method*, *J Comput Neurosci*, 27 (2009), pp. 115–133.
- [128] M. C. STUMPE AND H. GRUBMÄJLLER, *Urea impedes the hydrophobic collapse of partially unfolded proteins*, *Biophysical journal*, 9 (2009).
- [129] R. K. SUKUMARAN, R. R. SINGHANIA, AND A. PANDEY, *Microbial cellulases—production, applications and challenges*, *Journal of Scientific and Industrial Research*, 64 (2005), pp. 832–834.
- [130] Y. SUN AND J. CHENG, *Hydrolysis of lignocellulosic materials for ethanol production: a review*, *Bioresource Technology*, 82 (2002), pp. 1–11.
- [131] M. SUTO AND F. TOMITA, *Induction and catabolite repression mechanisms of cellulase in fungi*, *J Biosci Bioeng*, 92 (2001), pp. 305–311.
- [132] S. TAKASHIMA, H. IIKURA, A. NAKAMURA, M. HIDAKA, H. MASAKI, AND T. UOZUMI, *Isolation of the gene and characterization of the enzymatic pro-*

-
- perties of a major exoglucanase of humicola grisea without a cellulose-binding domain*, Journal of Biochemistry, 124 (1998), pp. 717–725.
- [133] M. L. TEODORO, G. N. PHILLIPS, AND L. E. KAVRAK, *Understanding protein flexibility through dimensionality reduction*, J Comp Biol, 10 (2003), pp. 617–634.
- [134] A. TORRIANI AND C. LEVINTHAL, *Ordered synthesis of proteins during outgrowth of spores of bacillus cereus*, J Bacteriol, 94 (1967), pp. 176–183.
- [135] Y. TSFADIA, R. FRIEDMAN, J. KADMON, A. SELZER, E. NACHLIEL, AND M. G. M., *Molecular dynamics simulations of palmitate entry into the hydrophobic pocket of the fatty acid binding protein*, FEBS Lett, 581 (2007), pp. 1243–1247.
- [136] D. VAN DER SPOEL, E. LINDAHL, B. HESS, G. GROENHOF, A. E. MARK, AND H. J. C. BERENDSEN, *Gromacs: Fast, flexible and free*, J Comp Chem, 26 (2005), pp. 1701–1719.
- [137] F. M. VERONESE, E. BOCCU, O. SCHIAVON, G. GRANDI, AND A. FONTANA, *General stability of thermophilic enzymes: studies on 6-phosphogluconate dehydrogenase from bacillus stearothermophilus and yeast*, J Appl Biochem, 6 (1984), pp. 39–47.
- [138] C. VIEILLE AND J. G. ZEIKUS, *Thermozymes: identifying molecular determinants of protein structural and functional stability*, Trends Biotechnol, 14 (1996), pp. 183–190.

-
- [139] A. VILLA, J. WÄŦHNERT, AND G. STOCK, *Molecular dynamics simulation study of the binding of purine bases to the aptamer domain of the guanine sensing riboswitch*, *Nucleic acids research*, 37 (2009), pp. 4774–4786.
- [140] A. A. VOLKOV, Z. SHAO, AND F. H. ARNOLD, *Recombination and chimeragenesis by in vitro heteroduplex formation and in vivo repair*, *Nucleic Acids Res*, 27 (1999), p. e18.
- [141] S. P. VOUTILAINEN, H. BOER, M. ALAPURANEN, J. JANIS, J. VEHMAANPERA, AND A. KOIVULA, *Improving the thermostability and activity of *Melanocarpus albomyces* cellobiohydrolase *cel7b**, *Applied microbiology and biotechnology*, 83 (2009), pp. 261–272.
- [142] S. P. VOUTILAINEN, P. G. MURRAY, M. G. TUOHY, AND A. KOIVULA, *Expression of *Talaromyces emersonii* cellobiohydrolase *cel7a* in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity*, *Protein engineering, design y selection : PEDS*, 23 (2010), pp. 69–79.
- [143] E. WALTON AND K. VANVLIET, *Equilibration of experimentally determined protein structures for molecular dynamics simulation*, *Physical Review E*, 74 (2006), pp. 1–8.
- [144] Z. WANG, J. MOTTONEN, AND E. J. GOLDSMITH, *Kinetically controlled folding of the serpin plasminogen activator inhibitor 1*, *Biochemistry*, 35 (1996), pp. 16443–16448.
- [145] A. E. WHEALS, L. C. BASSO, D. M. ALVES, AND H. V. AMORIM, *Fuel ethanol after 25 years*, *Trends Biotechnol*, 17 (1999), pp. 482–487.

- [146] A. WRBA, A. SCHWEIGER, V. SCHULTES, R. JAENICKE, AND P. ZAVODSKY, *Extremely thermostable d-glyceraldehyde-3-phosphate dehydrogenase from the eubacterium thermotoga maritima*, *Biochemistry*, 29 (1990), pp. 7584–7592.
- [147] J. S. YANG, W. W. CHEN, J. SKOLNICK, AND E. I. SHAKHNOVICH, *All-atom ab initio folding of a diverse set of proteins*, *Structure*, 15 (2006), pp. 53–63.
- [148] J. ZALDIVAR, J. NIELSEN, AND L. OLSSON, *Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration*, *Appl Microbiol Biotechnol*, 56 (2001), pp. 17–34.
- [149] D. ZHA, A. EIPPER, AND M. T. REETZ, *Assembly of designed oligonucleotides as an efficient method for gene recombination: a new tool in directed evolution*, *Chembiochem*, 4 (2003), pp. 34–39.
- [150] Y. X. ZHANG, K. PERRY, V. A. VINCI, K. POWELL, W. P. STEMMER, AND S. B. DEL CARDAYRÀL, *Genome shuffling leads to rapid phenotypic improvement in bacteria*, *Nature*, 415 (2002), pp. 644–646.

A . ANEXOS

A.1. Cálculo de trayectorias con GROMACS

Se describe aquí los programas y pasos para la simulación del movimiento de una enzima en solución acuosa, bajo el efecto de la temperatura; lo que consiste en el cálculo de la trayectoria de sus átomos en un determinado campo de fuerza. Para esto se utilizó el paquete de herramientas GROMACS [19, 67, 90, 136],

pdb2gmx

El primer paso corresponde a la conversión de la estructura y la creación de la topología del sistema. Para esto se utilizó el módulo `pdb2gmx` utilizando como entrada un archivo PDB y como salida la estructura convertida en el formato `.gro` y la topología como un archivo `.top`. En la conversión se definió el campo de fuerza **gromos53a6** como el campo bajo el que se registrará el sistema y se ignoraron los átomos de hidrógeno de la estructura de entrada mediante la opción `-ignh`, para ser incluidos posteriormente y así permitir su correcto posicionamiento.

En el caso de la simulación de la dinámica de las enzimas mutadas, en la creación de la topología se establecieron los puentes disulfuro, es decir, se especificó en la topología cuales son los pares de cisteínas que formarán enlace. Esto se realizó mediante la opción `-ss` para una posterior minimización de energía previa a la minimización realizada común a todas las enzimas simuladas.

grompp

Antes de cada simulación se deben generar los archivos de entrada ya sea para realizar una minimización de energía o para realizar una simulación. Esto se realiza

A.1. CÁLCULO DE TRAYECTORIAS CON GROMACS

mediante el módulo `grompp` y es aquí en donde se ingresan los parámetros para la simulación o minimización de energía en un archivo `.mdp`. Como entrada se utiliza un archivo de estructura (`.gro`) y de topología (`.top`) que se obtienen de la conversión de estructura o como salida de una minimización o simulación.

Para las etapas de minimización de energía se utilizó en los archivos de parámetros, la opción `integrator=steep`. Lo que se utiliza en esta etapa no corresponde a un integrador sino que indica como método minimización de energía el algoritmo `steepest descent`. Bajo esta opción, el parámetro `emstep` indica el paso máximo y `emtol` es la tolerancia. En la Tabla A.1 se listan los parámetros utilizados para las minimizaciones de energía. Luego de agregar el solvente, las etapas de minimización de energía se realizan bajo condiciones de borde periódicas.

Tabla A.1: Parámetros para la minimización de energía.

Parámetro	valor
Método	steepest descent
Número máximo de pasos	100
Tolerancia	1 $\left[\frac{kJ}{mol}\right]$
neighbor searching type ^b	simple
Interacciones de Coulomb y Van der Waals	Cut-off ^c
Radio de Coulomb	1[nm]
Radio de Van der Waals	1[nm]
Restricciones para los enlaces	ninguna ^d

Para la resolución de las ecuaciones de Newton en las etapas de simulación, se utilizó el algoritmo **leap frog** con un paso de $2fs$. Los parámetros restantes, de temperatura, manejo de interacciones enlazantes y no enlazantes, varían según cada

^bEste parámetro describe cómo encontrar los átomos vecinos a cada átomo.

^cSimplificación para las interacciones de tipo Coulomb y Van der Waals. Estas son consideradas nulas para radios mayores al radio de Coulomb y de Van der Waals definidos.

^dNo hay restricciones para los enlaces excepto aquellas definidas explícitamente en la topología, es decir, los enlaces son representados por un potencial armónico al igual que los ángulos, de acuerdo al campo de fuerza.

etapa y se especifican más adelante en la sección de resultados (Sección 7).

editconf

Al momento de definir la caja periódica con la que se trabajará se utiliza el módulo **editconf**. Con las opciones **-bt** y **-d** se definen la forma de la caja y la distancia entre la proteína y el borde de la caja. Como archivo de entrada se ingresa un archivo de estructura (**.gro**) proveniente de la minimización de energía de la enzima en el vacío.

genbox

Una vez que se ha definido el tamaño y forma de la caja se procede a definir el tipo de solvente. Esto se hace utilizando el módulo de GROMACS **genbox**. Mediante la opción **-cs** se define el tipo de solvente, que en este caso se usa un modelo de agua SPC, y se indica mediante el archivo **spc216.gro** que es el tipo de agua por defecto en GROMACS. Los archivos de entrada son un archivo de estructura, proveniente de la definición de las condiciones de borde periódicas, y la topología. La topología no se utiliza como requisito de la adición de solvente pero se necesita para ser actualizada ya que ahora hay nuevas moléculas en el sistema.

genion

Luego de agregar el agua, se ha solvatado a la proteína pero la carga neta del sistema es distinta de cero. Para neutralizar el sistema se debe agregar un número adecuado de iones de manera tal de neutralizar la carga. Para eso se utiliza el programa **genion**. Este programa utiliza como entrada tanto la estructura como la topología por lo que al igual que para la minimización de energía es necesario acoplar aquella información en un sólo archivo de salida que también es un archivo de

estructura (`.gro`). `genion` reemplaza moléculas de solvente por iones mono atómicos en la posición del átomo con el potencial electrostático más favorable. El potencial es calculado en todos los átomos y es recalculado luego de cada inserción. Es por esto que se deben definir parámetros tal como se realizó para la minimización de energía. Los parámetros utilizados son los que se muestran en la tabla A.1. También se requiere de la topología del sistema de igual manera que con `genbox`, para ser actualizada.

Con la opción `-conc` se define la concentración de sal $[M]$, con `-neutral` se indica que el sistema debe tener carga nula y con las opciones `-pname` y `-nname` se definen los tipos de iones, en este caso Na^+ y Cl^- .

mdrun

Cuando ya se han definido los parámetros comienza la simulación o bien la minimización de energía. GROMACS separa la definición de los parámetros del cálculo propiamente tal para facilitar la paralelización de los cálculos. El programa `mdrun` es la principal herramienta de todo el procedimiento pues es quien resuelve las ecuaciones de Newton, pero además realiza las minimizaciones de energía. Los archivos de entrada se preparan utilizando `grompp` que genera un archivo `.tpr` que sirve de entrada para `mdrun`. Este programa genera cuatro archivos de salida, un archivo de estructura con la conformación de salida luego de la simulación, un archivo de energía (`.edr`) en donde se registran la temperatura, presión, energías, densidad, volumen, entre otros, un registro de salida (`.log`) y un archivo de trayectorias (`.trr`) que contiene la información más importante en este trabajo pues contiene la información de la posición de los átomos en el tiempo.

A.2. Análisis de componentes principales

El análisis de componentes principales en una herramienta estadística que permite la transformación lineal de un set de variables a un nuevo set de variables no correlacionadas sustancialmente menor y que contiene casi toda la información de las variables no transformadas [71, 110].

Supongamos que tenemos una matriz de datos M con mediciones de $x = (x_1, x_2, \dots, x_I)$:

$$M = \begin{pmatrix} x_{11} & \dots & x_{1j} & \dots & x_{1I} \\ \vdots & \ddots & \vdots & & \vdots \\ x_{i1} & \dots & x_{ij} & \dots & x_{iI} \\ \vdots & & \vdots & \ddots & \vdots \\ x_{I1} & \dots & x_{Ij} & \dots & x_{II} \end{pmatrix} \quad (\text{A.1})$$

donde $j = 1, \dots, J$ es la j -ésima medición de la variable i -ésima. Se busca un conjunto de variables y_i combinación lineal de x_i no correlacionadas entre sí, cuyas varianzas vaya decreciendo progresivamente:

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \dots + a_{iI}x_I = \mathbf{a}_i^t \mathbf{x} \quad (\text{A.2})$$

$$\mathbf{a}_i = \begin{pmatrix} a_{i1} \\ \vdots \\ a_{iI} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_I \end{pmatrix}$$

Para mantener la ortogonalidad de la transformación se impone la restricción $\sum_{k=1}^I a_{ik}^2 = 1$. Así, se escogen los valores de \mathbf{a}_1 de modo que maximice la varianza de y_1 sujeto a la restricción antes impuesta,

$$Var(y_1) = Var(\mathbf{a}_1^t \mathbf{x}) = \mathbf{a}_1^t \Sigma \mathbf{a}_1 \quad (\text{A.3})$$

donde Σ es la matriz de covarianza de \mathbf{x} . Para resolver el problema de maximizar la función en la ecuación A.3 se utiliza el método de multiplicadores de Lagrange.

De esta forma:

$$L(\mathbf{a}_1) = \mathbf{a}_1^t \Sigma \mathbf{a}_1 - \lambda (a_1^t \mathbf{a}_1 - 1) \quad (\text{A.4})$$

$$\frac{\partial L}{\partial \mathbf{a}_1} = 2\Sigma \mathbf{a}_1 - 2\lambda \mathbf{I} \mathbf{a}_1 = 0$$

$$\Rightarrow (\Sigma - \lambda \mathbf{I}) \mathbf{a}_1 = 0 \quad (\text{A.5})$$

Para que el sistema de ecuaciones en A.5 tenga solución, según el teorema de Roché-Frobenius, la matriz $(\Sigma - \lambda \mathbf{T})$ debe ser singular, o sea,

$$|\Sigma - \lambda \mathbf{T}| = 0 \quad (\text{A.6})$$

luego, λ es un valor propio de Σ . La matriz de covarianza es de orden I y si además es definida positiva, tendrá I valores propios distintos, $\lambda_1, \lambda_2, \dots, \lambda_I$, tales que, por ejemplo, $\lambda_1 > \lambda_2 > \dots > \lambda_I$.

De la expresión A.6 se tiene que:

$$\Sigma \mathbf{a}_1 = \lambda \mathbf{I} \mathbf{a}_1 \quad (\text{A.7})$$

entonces,

$$\text{Var}(y_1) = \text{Var}(\mathbf{a}_1^t \mathbf{x})$$

$$= \mathbf{a}_1^t \Sigma \mathbf{a}_1$$

$$= \mathbf{a}_1^t \lambda \mathbf{I} \mathbf{a}_1$$

$$= \lambda \mathbf{a}_1^t \mathbf{a}_1$$

$$= \lambda \cdot 1$$

$$\text{Var}(y_1) = \lambda \quad (\text{A.8})$$

Para maximizar la varianza de y_1 , el primer componente principal, se debe tomar el mayor valor propio, por ejemplo, λ_1 y el correspondientes vector propio \mathbf{a}_1 . El segundo componente principal, y_2 , se obtiene mediante un desarrollo análogo.

$$L(\mathbf{a}_2) = \mathbf{a}_2^t \Sigma \mathbf{a}_2 - \lambda (\mathbf{a}_2^t \mathbf{a}_2 - 1) - \delta \mathbf{a}_2^t \mathbf{a}_1 \quad (\text{A.9})$$

$$\frac{\partial L(\mathbf{a}_2)}{\partial \mathbf{a}_2} = 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 - \delta \mathbf{a}_1 = 0 \quad (\text{A.10})$$

Como y_1 no está correlacionado con y_2 ,

$$\begin{aligned} \text{Cov}(y_2, y_1) &= \text{Cov}(\mathbf{a}_2^t \mathbf{x}, \mathbf{a}_1^t \mathbf{x}) \\ &= \mathbf{a}_2^t \cdot E[(\mathbf{x} - \mu)(\mathbf{x} - \mu)] \cdot \mathbf{a}_1 \\ &= \mathbf{a}_2^t \Sigma \mathbf{a}_1 = 0 \end{aligned} \quad (\text{A.11})$$

luego, $\mathbf{a}_2^t \Sigma \mathbf{a}_1 = 0$. Así, multiplicando por \mathbf{a}_1^t y como $\mathbf{a}_1^t \mathbf{a}_1 = 1$ se tiene que

$$\frac{\partial L(\mathbf{a}_2)}{\partial \mathbf{a}_2} = 2\Sigma \mathbf{a}_2 - 2\lambda \mathbf{a}_2 \quad (\text{A.12})$$

$$\Rightarrow (\Sigma - \lambda \mathbb{I}) \mathbf{a}_2 = 0 \quad (\text{A.13})$$

De manera análoga a y_1 se escoge λ como el segundo mayor valor propio de la matriz Σ con el vector propio asociado \mathbf{a}_2 . Extendiendo este razonamiento, para el i -ésimo componente le correspondería el i -ésimo valor propio asociado al vector propio \mathbf{a}_i . Así, de manera general se tiene que

$$\mathbf{y} = A\mathbf{x} \quad (\text{A.14})$$

donde

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_I \end{pmatrix}, A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1I} \\ a_{21} & a_{22} & \dots & a_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1} & a_{I2} & \dots & a_{II} \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_I \end{pmatrix} \quad (\text{A.15})$$

A.3. ALINEAMIENTO MÚLTIPLE DE CEL72

Ahora bien, como $Var(y_i) = \lambda_i$ y $Cov(y_k, y_l) = 0$ si $k \neq l$, la matriz de covarianza de y será:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_l \end{pmatrix} \quad (\text{A.16})$$

A.3. Alineamiento múltiple de Cel72 para determinar dominios conservados

A continuación se muestra el alineamiento obtenido de BLAST para determinar los dominios conservados. Se destacan con # los residuos conservados. Sólo se muestran en letras mayúsculas los residuos alineados, entre corchetes el número de residuos no alineados. la enzima **query** corresponde a la enzima Cel72. En color rojo se destacan los aminoácidos coincidentes en la mayoría de la enzimas.

Feature 1	#	##	#	#	#	#	#	#	#					
40Vw_A	7	[27]	.DAGIHGIRQKN	[26]	.ILSGM	[8]	.GITTS	[2]	.KLRLQQL	[5]	.VSPRVYLL	[6]	.YEMHLHT	123
query	1	[1]	.DANRWLHVSQ	[26]	.ALDGA	[6]	.GITAS	[2]	.SLKLQFV	[4]	.VGSRVYLM	[5]	.YQMFQLL	87
1Q9H_A	8	[27]	.DANRWVHVDN	[26]	.ALDGA	[6]	.GVTSS	[2]	.SLKLNQV	[4]	.VGSRLYLL	[5]	.YQIFKLL	120
20Vw_A	7	[27]	.DAGIHGIRQKN	[26]	.ILSGM	[8]	.GITTS	[2]	.KLRLQQL	[5]	.VSPRVYLL	[6]	.YEMHLHT	123
gi 121788	30	[27]	.DWNRYMHHDAN	[25]	.FIEGV	[5]	.GVVTS	[2]	.SLTMNQY	[10]	.VSPRVYLL	[5]	.YVMLKLN	147
gi 145666094	31	[27]	.DVMFRPMHVKD	[25]	.VLEGV	[5]	.GVATK	[2]	.ALKLNQY	[9]	.VAPRVYLL	[6]	.YELMKLV	147
gi 171682802	30	[27]	.DALSRHFHVS	[30]	.ALEGI	[5]	.GVSAV	[2]	.AVTLRQY	[9]	.VSPRVYLL	[6]	.YEPLTLL	151
1EG1_A	8	[27]	.DWNRYMHHDAN	[25]	.FIEGV	[5]	.GVVTS	[2]	.SLTMNQY	[10]	.VSPRVYLL	[5]	.YVMLKLN	124
1DYM_A	7	[27]	.DSLSPHPIHRAE	[29]	.IMEGI	[6]	.GVTTN	[2]	.SLRLQHI	[6]	.PSPRVYLL	[6]	.YEMHLHT	125
1H46_X	8	[27]	.DSNWRVHSTS	[26]	.ALDGA	[6]	.GITST	[2]	.ALTLKFV	[4]	.VGSRVYLM	[5]	.YQLLKLL	120
gi 66828465	27	[27]	.DANRWVHSAQ	[25]	.YLDGA	[6]	.GVVTS	[2]	.SVRLNFV	[7]	.IGSRFLM	[5]	.YQLFHVH	141
gi 121852	26	[27]	.DANRWLHSTS	[26]	.ALDGA	[6]	.GITAS	[2]	.SLKLQFV	[4]	.VGSRVYLM	[5]	.YQMFQLL	138
gi 15824271	20	[27]	.DSNWRVHSDS	[25]	.LIDGA	[6]	.GITSS	[2]	.SLKLVFV	[8]	.IGSRVYLL	[5]	.YQIFDLK	135
1CEL_A	8	[27]	.DANRWTHATN	[26]	.CLDGA	[6]	.GVVTS	[2]	.SLSIGFV	[7]	.VGARLYLM	[5]	.YQEFLL	123
1EGN_A	8	[27]	.DANRWTHATN	[26]	.CLDGA	[6]	.GVVTS	[2]	.SLSIDFV	[7]	.VGARLYLM	[5]	.YQEFLL	123
2RFW_A	8	[27]	.DANRWLHDDN	[24]	.MIEGA	[7]	.GASTS	[2]	.ALTLKFV	[8]	.VGSRVYLM	[5]	.YQMFNLM	123
gi 3132	27	[27]	.DVNWRVHSTT	[26]	.ALDGA	[6]	.GILPS	[2]	.SVTLRPV	[2]	.LGLRFLF	[5]	.YQMFQLL	137
gi 16117340	11	[26]	.DRHIGDWDRE	[10]	.ANVGV	[2]	.SADGK	[2]	.SQRLVSK	[7]	.VGSRVYIV	[6]	.YQLFQFV	106
2RG0_A	8	[27]	.DANRWLHDDN	[24]	.MIEGA	[7]	.GASTS	[2]	.ALTLKFV	[8]	.VGSRVYLM	[5]	.YQMFNLM	123
Feature 1	#	##	#	#	#	#	#	#	#	#	#	#	#	
40Vw_A	124	[1]	.TEFSEFDV	[1]	.EKLPCGMNGALYLSEMP	[12]	.AGAYYGAGYCDAAQC	[20]	.ELDIWEANS	[1]	.ATHIA	211		
query	88	[1]	.QEFTFDVD	[1]	.SNLPCGLNGALYLSAMD	[12]	.AGAKYGTGYCDSQC	[33]	.EMDIWEANN	[1]	.AAAYT	188		
1Q9H_A	121	[1]	.REFSEFDVD	[1]	.SNLPCGLNGALYFVAMD	[12]	.AGAKYGTGYCDSQC	[35]	.EMDVWEANS	[1]	.SNAVY	223		
20Vw_A	124	[1]	.TEFSEFDV	[1]	.EKLPCGMNGALYLSEMP	[12]	.AGAYYGAGYCDAAQC	[20]	.ELDIWEANS	[1]	.ATHIA	211		
gi 121788	147	[1]	.QELSEFDVD	[1]	.SALPCGNGSLYLSQMD	[10]	.AGANYGSGYCDAAQC	[20]	.EMDILEGNS	[1]	.ANALT	232		
gi 145666094	148	[1]	.MELAFDVD	[1]	.SQLCGMNGALYFVSM	[10]	.AGAAVGTGECDAQC	[1]	.QDQFTANR	[1]	.ATALT	214		
gi 171682802	152	[1]	.QELAVDVD	[1]	.SGLPCGMNSAVYLSEMD	[10]	.AGAYGTGYCDAAQC	[21]	.EMDIWEANS	[1]	.ANSFT	238		
1EG1_A	125	[1]	.QELSEFDVD	[1]	.SALPCGNGSLYLSQMD	[10]	.AGANYGSGYCDAAQC	[20]	.EMDILEGNS	[1]	.ANALT	210		
1DYM_A	126	[1]	.FEFTFDVD	[1]	.TKLPCGMNSALYLSM	[10]	.GGAYGTGYCDAAQC	[20]	.AMDIWEANS	[1]	.ASHVA	211		
1H46_X	121	[1]	.QEFTFDVD	[1]	.SNLPCGLNGALYLSAMD	[12]	.AGAKYGTGYCDSQC	[33]	.EMDIWEANN	[1]	.AAAFY	221		
gi 66828465	142	[1]	.QEFTFDVD	[1]	.SNLDCGLNGALYLVSM	[12]	.AGAKYGTGYCDAAQC	[35]	.EMDIWEANN	[1]	.ATAVT	244		
gi 121852	139	[1]	.QEFTFDVD	[1]	.SNLPCGLNGALYLSAMD	[12]	.AGAKYGTGYCDSQC	[33]	.EMDIWEANN	[1]	.AAAFY	239		
gi 15824271	136	[1]	.KEFTFDVD	[1]	.SNLDCGLNGALYFVSM	[12]	.AGAKYGTGYCDAAQC	[35]	.EMDIWEANK	[1]	.ATAYT	238		
1CEL_A	124	[1]	.NEFSEFDVD	[1]	.SQLPCGLNGALYFVSM	[12]	.AGAKYGTGYCDSQC	[35]	.EMDIWEANS	[1]	.SEALT	226		
1EGN_A	124	[1]	.NEFSEFDVD	[1]	.SQLPCGLNGALYFVSM	[12]	.AGAKYGTGYCDSQC	[35]	.EMDIWEANS	[1]	.SSHVA	226		
2RFW_A	124	[1]	.NELAFDVD	[1]	.STVECGINSALYFVAME	[12]	.AGAYGTGYCDAAQC	[35]	.EIDWESNA	[1]	.AFAYT	226		
gi 3132	138	[1]	.KEFTFDV	[1]	.PNMRCGSSGAIHLTAMD	[12]	.AGAKYGTGFCDAQC	[33]	.DIALWEAND	[1]	.SASFA	238		
gi 16117340	107	[1]	.KEFTYTDV	[1]	.SQIPCQVNAALYTVEMP	[6]	.GGVEYGYGCDAAQC	[7]	.EFDIQEASS	[1]	.AIVYT	175		
2RG0_A	124	[1]	.NELAFDVD	[1]	.STVECGINSALYFVAME	[12]	.AGAYGTGYCDAAQC	[35]	.EIDWESNA	[1]	.AFAYT	226		

A.4. Figuras resultado de la evaluación con PROCHECK

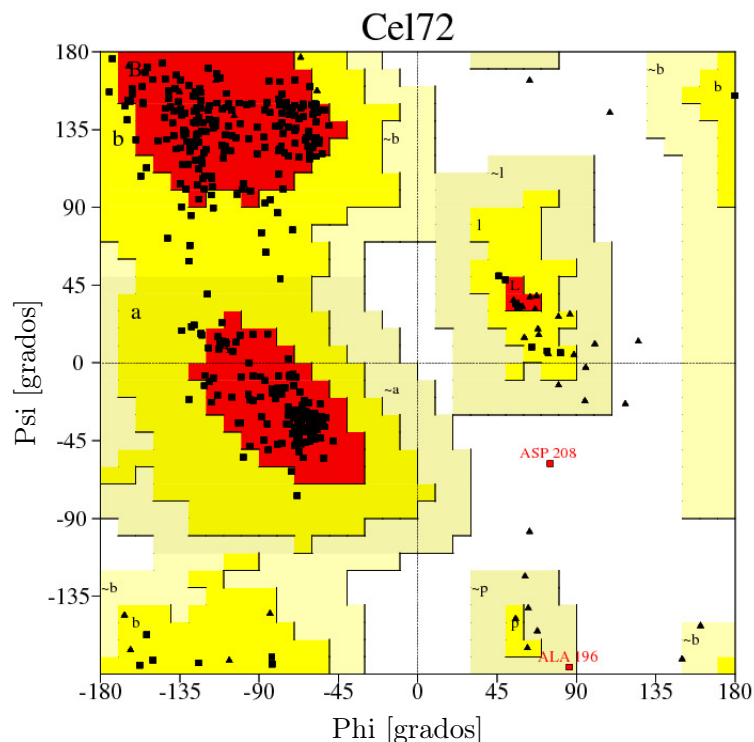


Figura A.1: Gráfico de Ramachandran de Cel72, basado en el análisis de 118 estructuras de resolución de al menos 2.0 y factor R no mayor al 20%. Las regiones de color rojo (A, B y L, denominadas así por α -hélice, hoja β y *loop*, respectivamente) son las regiones más favorables. Las regiones de color amarillo (a, b, l y p) son las regiones permitidas menos favorables. Las regiones de color beige (a, b, l, p) son las regiones adicionales permitidas. Se obtuvo 290 residuos (86.1%) en regiones A, B ó L; 45 residuos (13.4%) en regiones a, b, l ó p; 1 residuo (0.3%) en regiones permitidas adicionales, y 1 residuo (0.3%) en regiones desfavorables. 43 residuos son Gly (mostrado como triángulos) y 15 son Pro. El número total de residuos es 397.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

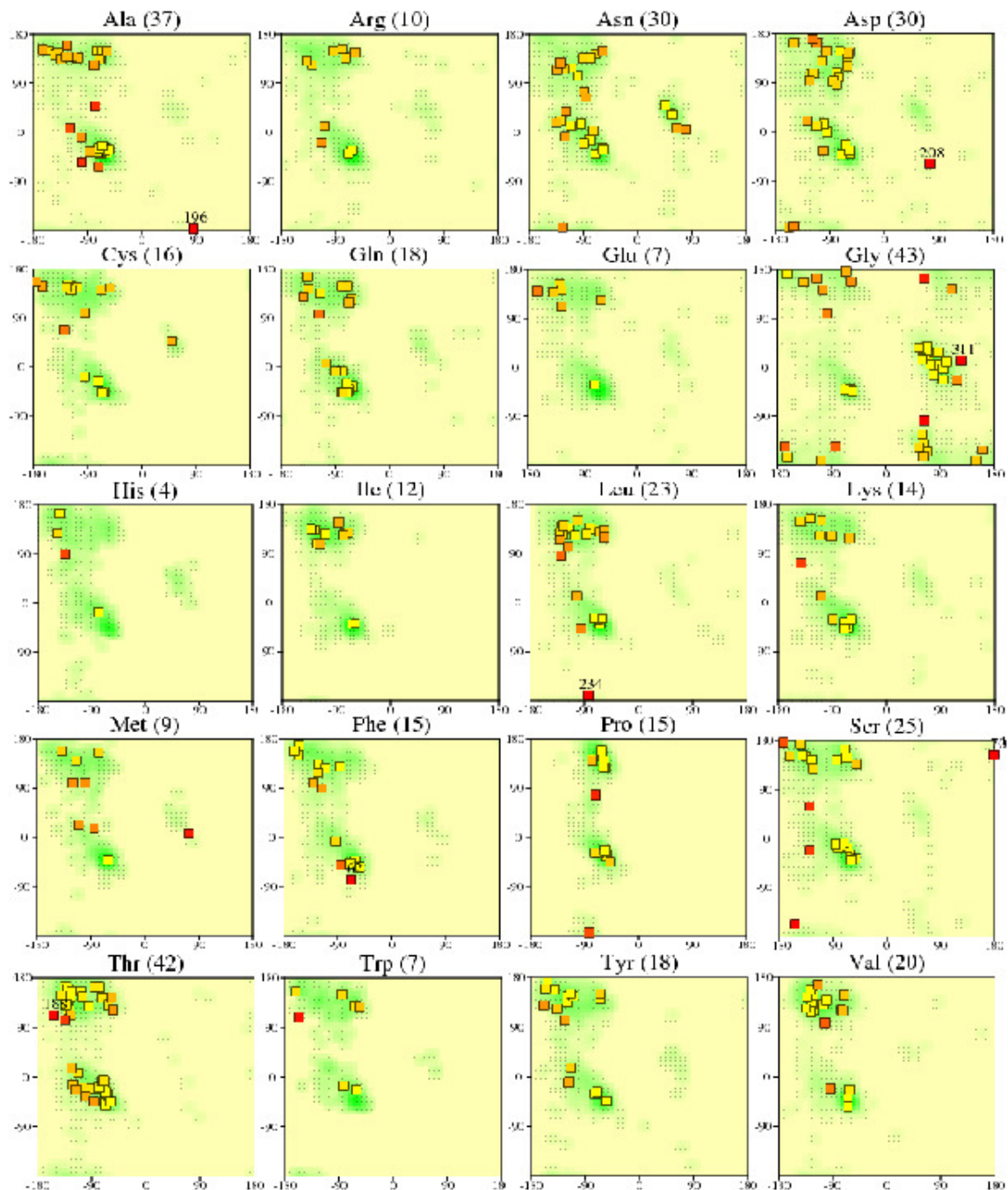


Figura A.2: Gráfico de Ramachandran para todos los tipos de residuos. La cantidad de cada residuo presente en la enzima se muestra entre paréntesis. Aquellos con conformación desfavorable se muestran en rojo y se muestra el número de residuo. El sombreado verde muestra las conformaciones favorables obtenidas del análisis de 163 estructuras a resolución de 2.0 o mejor.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

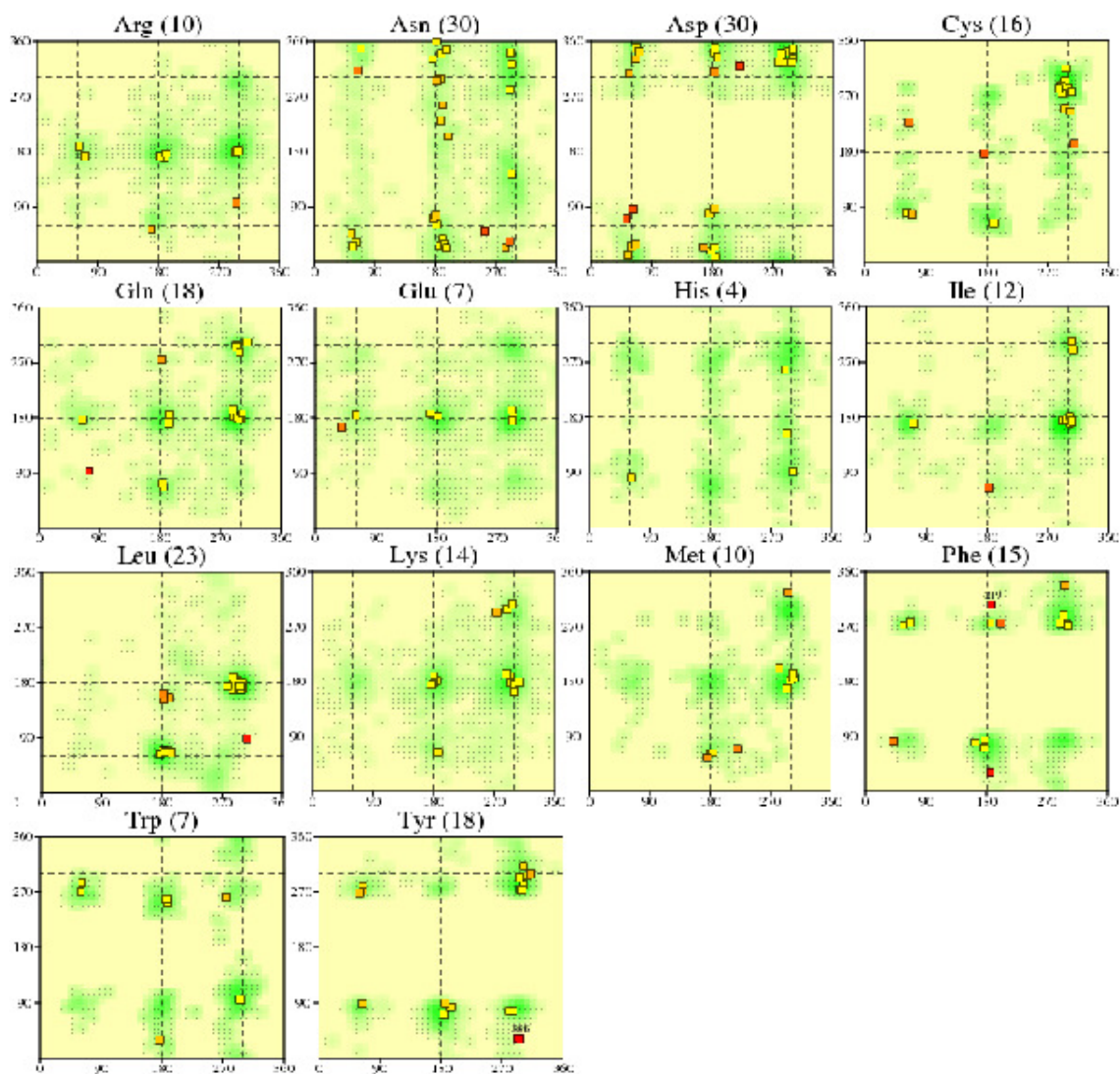


Figura A.3: Gráfico para los ángulos de torsión Chi1-Chi2, para todos los tipos de residuos. La cantidad de cada residuo presente en la enzima se muestra entre paréntesis. Aquellos con conformación desfavorable se muestran en rojo y se muestra el número de residuo. El sombreado verde muestra las conformaciones favorables obtenidas del análisis de 163 estructuras a resolución de 2.0 o mejor.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

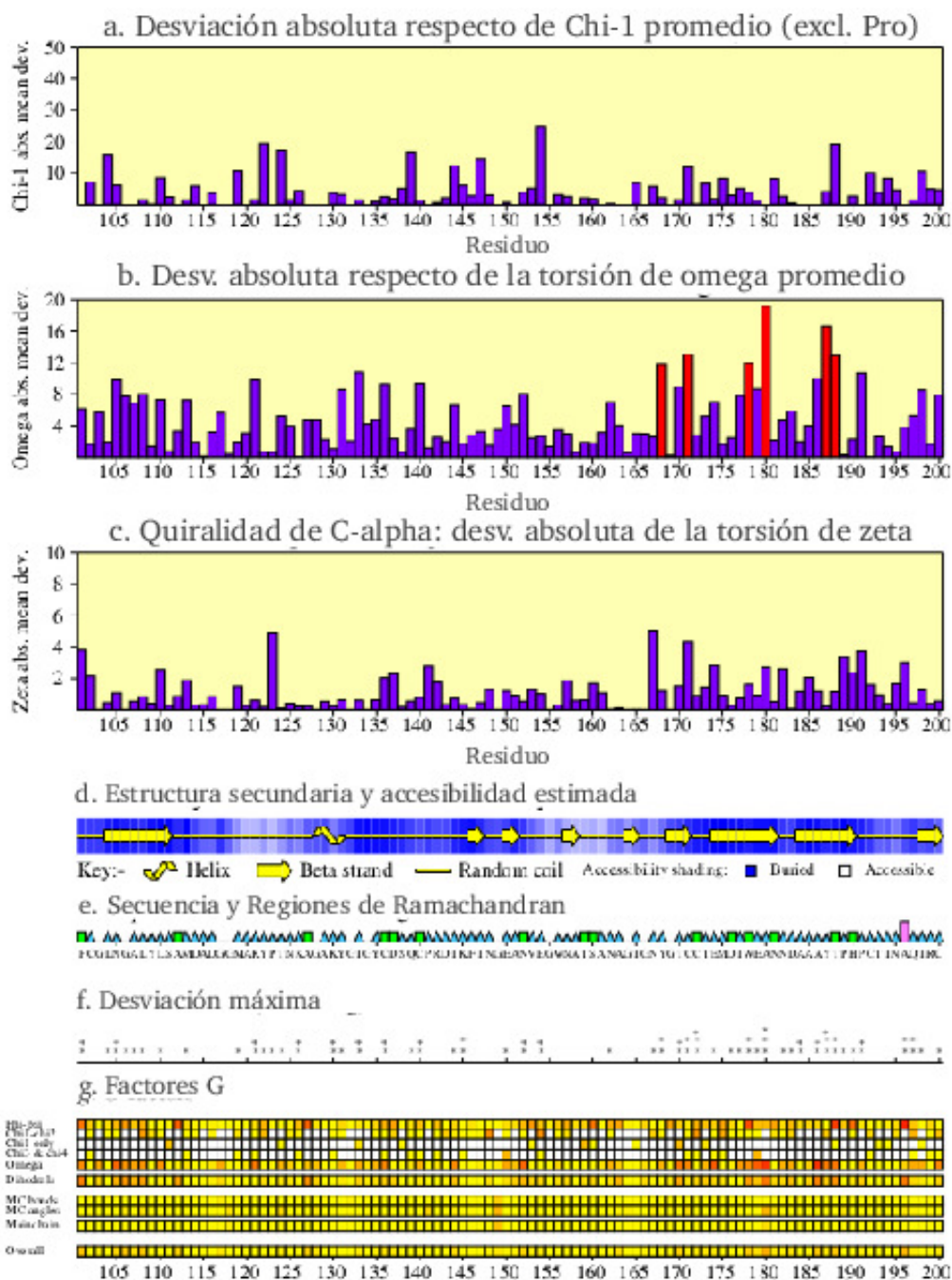


Figura A.5: Propiedades por residuo de Cel72. Los gráficos mostrados son para los residuos 101 al 200. Para los gráficos a, b y c, se destacan en color rojo los residuos cuya desviación es el doble o más de la desviación estándar ideal.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

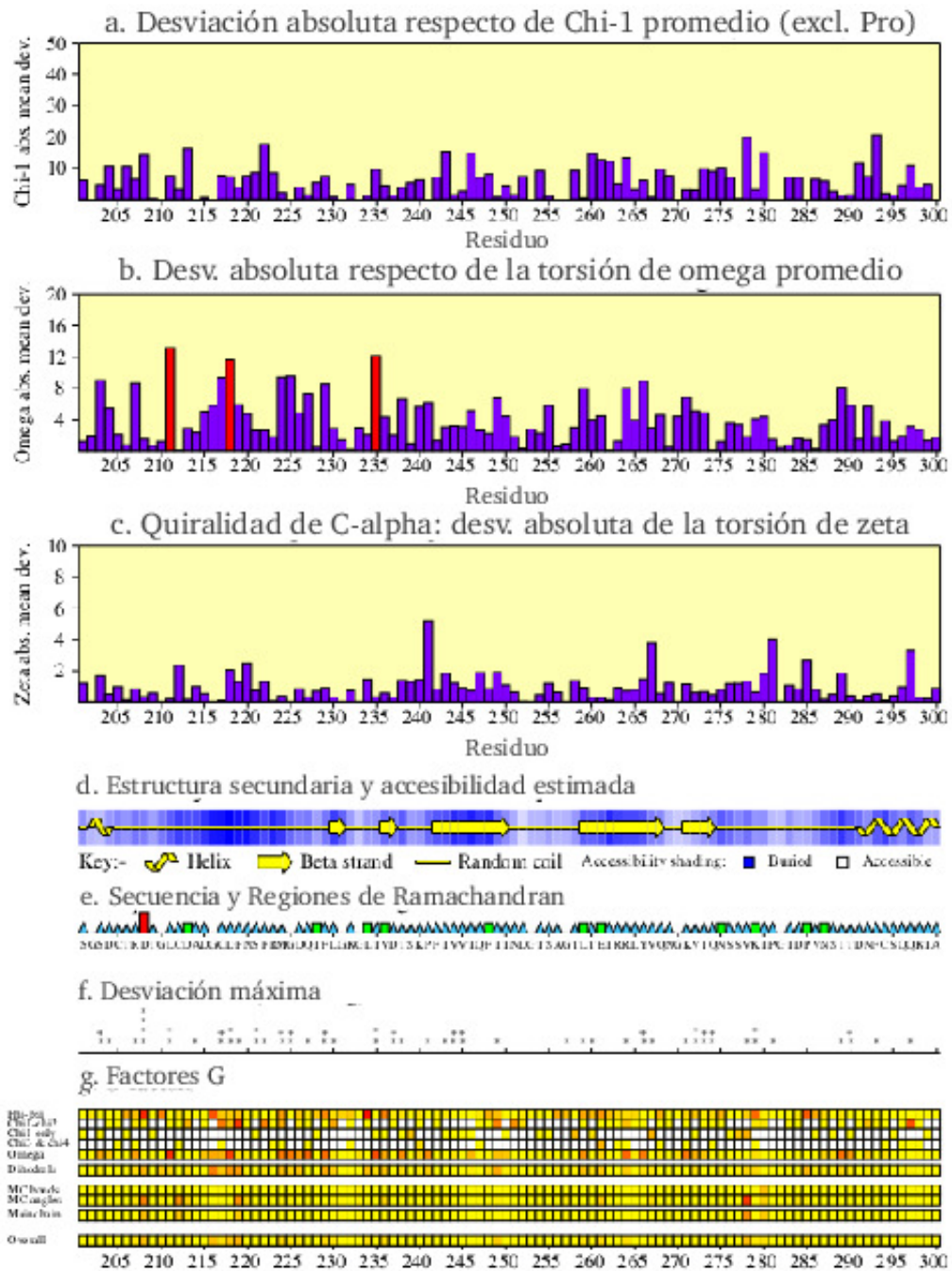


Figura A.6: Propiedades por residuo de Cel72. Los gráficos mostrados son para los residuos 201 al 300. Para los gráficos a, b y c, se destacan en color rojo los residuos cuya desviación es el doble o más de la desviación estándar ideal.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

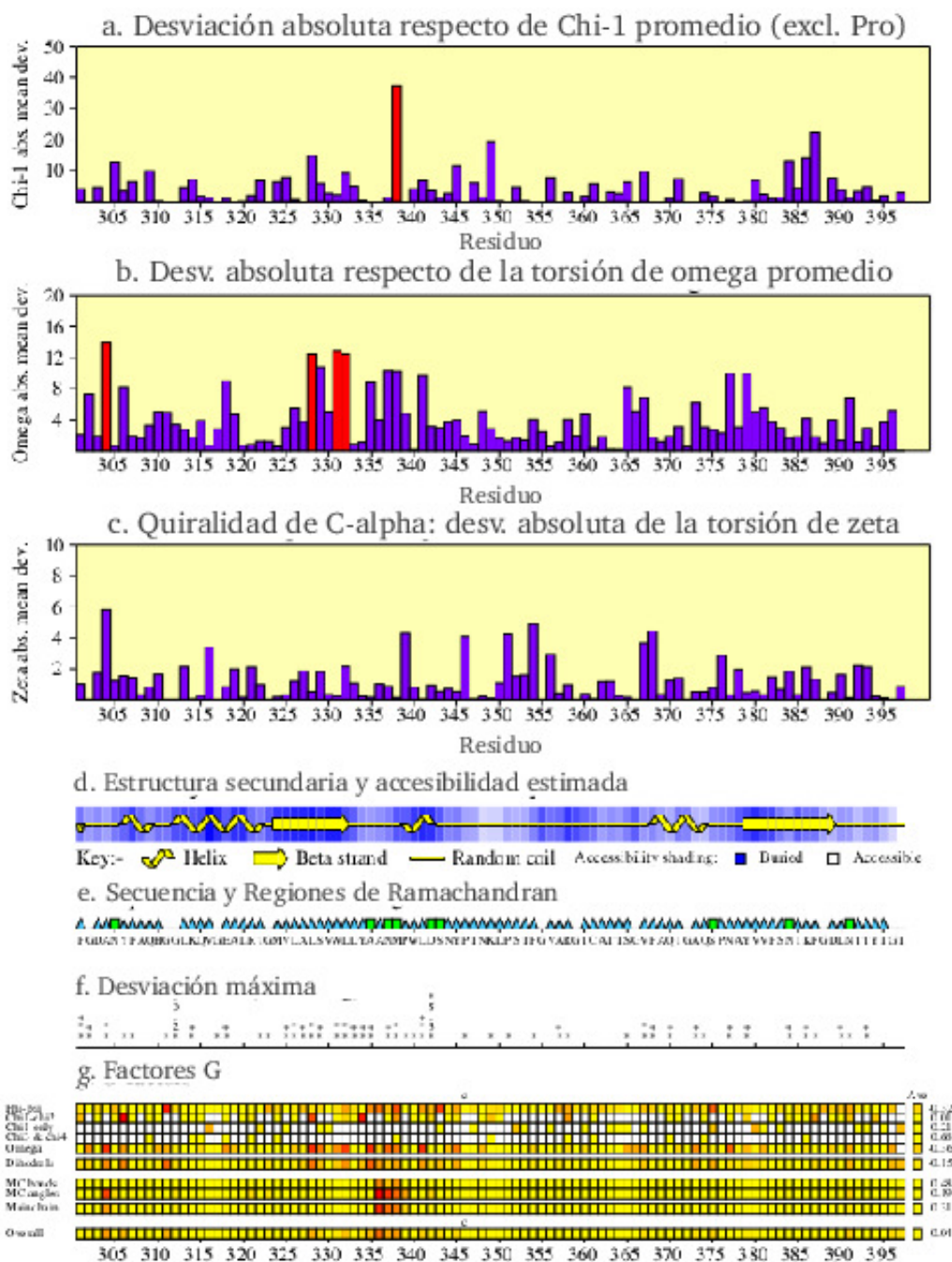


Figura A.7: Propiedades por residuo de Cel72. Los gráficos mostrados son para los residuos 301 al 397. Para los gráficos a, b y c, se destacan en color rojo los residuos cuya desviación es el doble o más de la desviación estándar ideal.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

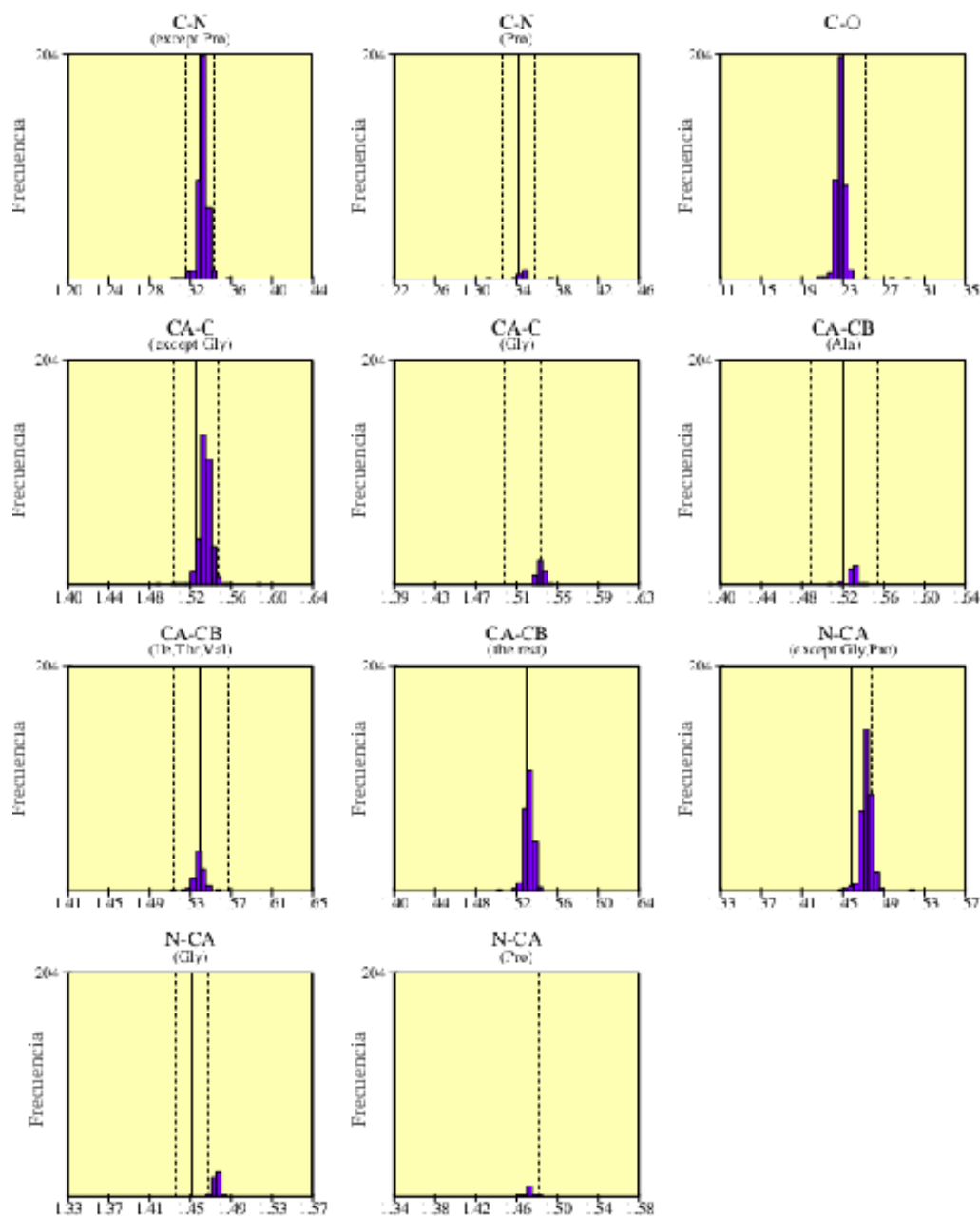


Figura A.8: Largo de enlace de la cadena principal de Cel72. Las líneas continúa y discontinua representan el promedio y desviación estándar, respectivamente, según los datos para pequeñas moléculas de Engh & Huber.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

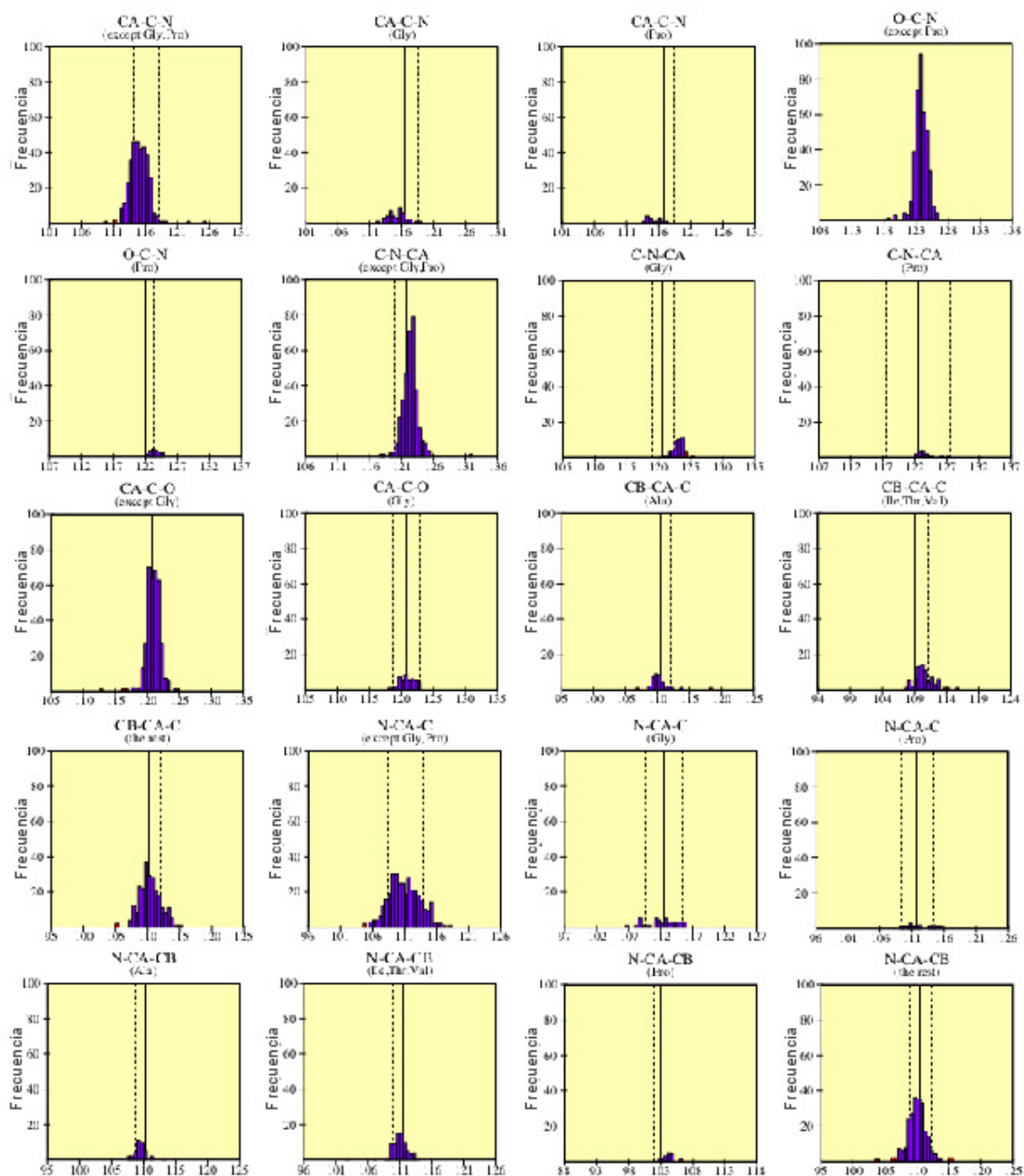


Figura A.9: Ángulos de enlace de la cadena principal de Cel72. Las líneas continúa y discontinua representan el promedio y desviación estándar, respectivamente, según los datos para pequeñas moléculas de Engh & Huber.

A.4. FIGURAS RESULTADO DE LA EVALUACIÓN CON PROCHECK

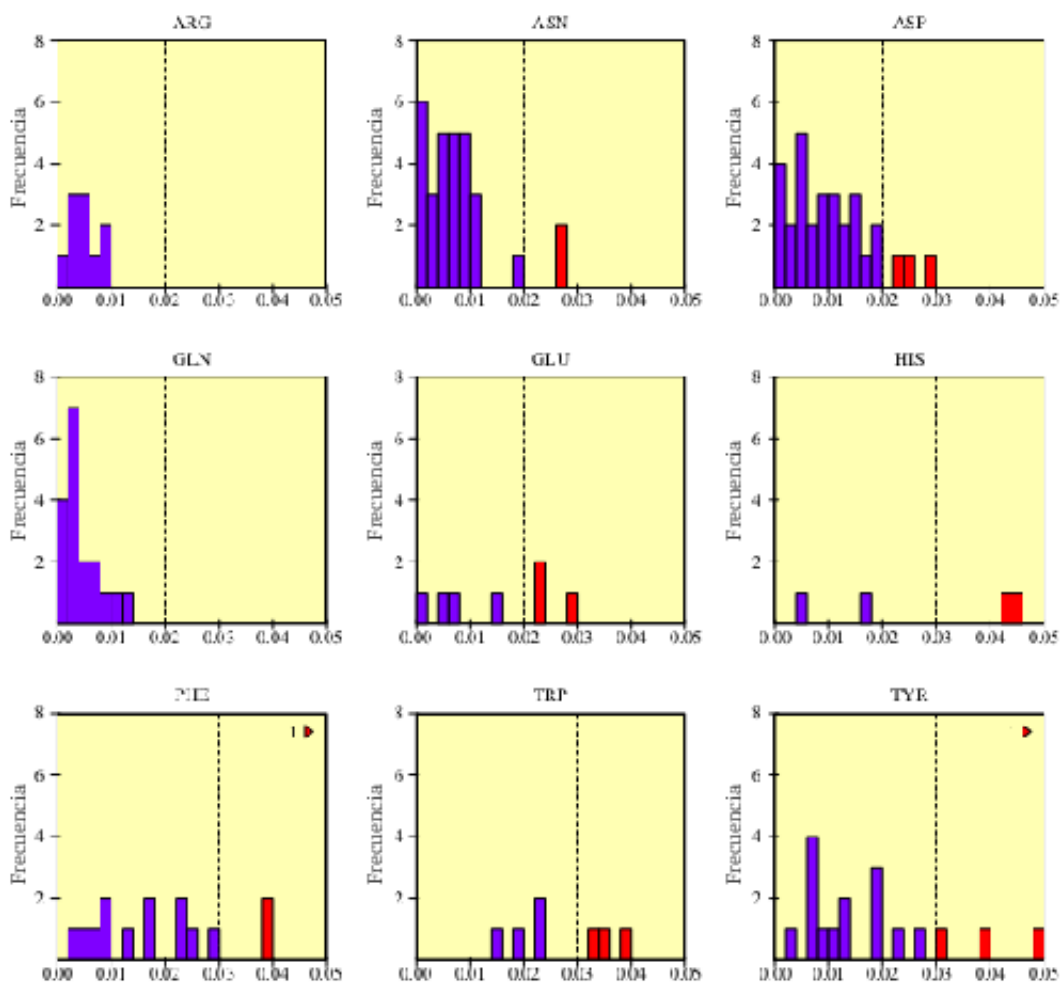


Figura A.10: Histogramas que muestran RMS de la distancia respecto a la planaridad. Las barras de color rojo indican grandes desviaciones respecto de la planaridad: RMS > 0.03 para anillos, y > 0.02 para residuos que no son anillos. Los triángulos de color rojo indican y el número a su lado, indican la cantidad de puntos fuera del gráfico.

A.5. Registros de salida

Los registros se adjuntan en formato digital, al final de este documento.

A.6. Tabla t-student

Tabla A.2: Tabla de t-student. Valores t tales que $p[|T| > t] = \alpha$. n son grados de libertad.

$n \setminus \alpha$	0,9	0,8	0,7	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,1584	0,3249	0,5095	1	1,9626	3,0777	6,3137	12,7062	31,821	63,6559	636,5776
2	0,1421	0,2887	0,4447	0,8165	1,3862	1,8856	2,92	4,3027	6,9645	9,925	31,5998
3	0,1366	0,2767	0,4242	0,7649	1,2498	1,6377	2,3534	3,1824	4,5407	5,8408	12,9244
4	0,1338	0,2707	0,4142	0,7407	1,1896	1,5332	2,1318	2,7765	3,7469	4,6041	8,6101
5	0,1322	0,2672	0,4082	0,7267	1,1558	1,4759	2,015	2,5706	3,3649	4,0321	6,8685
6	0,1311	0,2648	0,4043	0,7176	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074	5,9587
7	0,1303	0,2632	0,4015	0,7111	1,1192	1,4149	1,8946	2,3646	2,9979	3,4995	5,4081
8	0,1297	0,2619	0,3995	0,7064	1,1081	1,3968	1,8595	2,306	2,8965	3,3554	5,0414
9	0,1293	0,261	0,3979	0,7027	1,0997	1,383	1,8331	2,2622	2,8214	3,2498	4,7809
10	0,1289	0,2602	0,3966	0,6998	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693	4,5868
11	0,1286	0,2596	0,3956	0,6974	1,0877	1,3634	1,7959	2,201	2,7181	3,1058	4,4369
12	0,1283	0,259	0,3947	0,6955	1,0832	1,3562	1,7823	2,1788	2,681	3,0545	4,3178
13	0,1281	0,2586	0,394	0,6938	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123	4,2209
14	0,128	0,2582	0,3933	0,6924	1,0763	1,345	1,7613	2,1448	2,6245	2,9768	4,1403
15	0,1278	0,2579	0,3928	0,6912	1,0735	1,3406	1,7531	2,1315	2,6025	2,9467	4,0728
16	0,1277	0,2576	0,3923	0,6901	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208	4,0149
17	0,1276	0,2573	0,3919	0,6892	1,069	1,3334	1,7396	2,1098	2,5669	2,8982	3,9651
18	0,1274	0,2571	0,3915	0,6884	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784	3,9217
19	0,1274	0,2569	0,3912	0,6876	1,0655	1,3277	1,7291	2,093	2,5395	2,8609	3,8833
20	0,1273	0,2567	0,3909	0,687	1,064	1,3253	1,7247	2,086	2,528	2,8453	3,8496
21	0,1272	0,2566	0,3906	0,6864	1,0627	1,3232	1,7207	2,0796	2,5176	2,8314	3,8193
22	0,1271	0,2564	0,3904	0,6858	1,0614	1,3212	1,7171	2,0739	2,5083	2,8188	3,7922
23	0,1271	0,2563	0,3902	0,6853	1,0603	1,3195	1,7139	2,0687	2,4999	2,8073	3,7676
24	0,127	0,2562	0,39	0,6848	1,0593	1,3178	1,7109	2,0639	2,4922	2,797	3,7454
25	0,1269	0,2561	0,3898	0,6844	1,0584	1,3163	1,7081	2,0595	2,4851	2,7874	3,7251
26	0,1269	0,256	0,3896	0,684	1,0575	1,315	1,7056	2,0555	2,4786	2,7787	3,7067
27	0,1268	0,2559	0,3894	0,6837	1,0567	1,3137	1,7033	2,0518	2,4727	2,7707	3,6895
28	0,1268	0,2558	0,3893	0,6834	1,056	1,3125	1,7011	2,0484	2,4671	2,7633	3,6739

Continúa en la página siguiente

A.6. TABLA T-STUDENT

Tabla A.2 :Tabla de t-student. (Continuación de la página anterior).

$n \setminus \alpha$	0,9	0,8	0,7	0,5	0,3	0,2	0,1	0,05	0,02	0,01	0,001
29	0,1268	0,2557	0,3892	0,683	1,0553	1,3114	1,6991	2,0452	2,462	2,7564	3,6595
30	0,1267	0,2556	0,389	0,6828	1,0547	1,3104	1,6973	2,0423	2,4573	2,75	3,646
40	0,1265	0,255	0,3881	0,6807	1,05	1,3031	1,6839	2,0211	2,4233	2,7045	3,551
80	0,1261	0,2542	0,3867	0,6776	1,0432	1,2922	1,6641	1,9901	2,3739	2,6387	3,4164
120	0,1259	0,2539	0,3862	0,6765	1,0409	1,2886	1,6576	1,9799	2,3578	2,6174	3,3734
∞	0,126	0,253	0,385	0,674	1,036	1,282	1,645	1,96	2,326	2,576	3,291