



Análisis de Modos Normales en una Biomolécula

John Erick Cabrera Ramirez

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2017

Análisis de Modos Normales en una Biomolécula

John Erick Cabrera Ramirez

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Físico

Director(a):
PhD, ,Yuly Edith Sánchez Mendoza

Línea de Investigación:
Biofísica Molecular
Grupo de Investigación:
Biofísica Molecular

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2017

(Dedicatoria o un lema)

Su uso es opcional y cada autor podrá determinar la distribución del texto en la página, se sugiere esta presentación. En ella el autor dedica su trabajo en forma especial a personas y/o entidades.

Por ejemplo:

A mis padres

o

La preocupación por el hombre y su destino siempre debe ser el interés primordial de todo esfuerzo técnico. Nunca olvides esto entre tus diagramas y ecuaciones.

Albert Einstein

Agradecimientos

Esta sección es opcional, en ella el autor agradece a las personas o instituciones que colaboraron en la realización de la tesis o trabajo de investigación. Si se incluye esta sección, deben aparecer los nombres completos, los cargos y su aporte al documento.

Resumen

El resumen es una presentación abreviada y precisa (la NTC 1486 de 2008 recomienda revisar la norma ISO 214 de 1976). Se debe usar una extensión máxima de 12 renglones. Se recomienda que este resumen sea analítico, es decir, que sea completo, con información cuantitativa y cualitativa, generalmente incluyendo los siguientes aspectos: objetivos, diseño, lugar y circunstancias, pacientes (u objetivo del estudio), intervención, mediciones y principales resultados, y conclusiones. Al final del resumen se deben usar palabras claves tomadas del texto (mínimo 3 y máximo 7 palabras), las cuales permiten la recuperación de la información.

Palabras clave: (máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado).

A continuación se presentan algunos ejemplos de tesauros que se pueden consultar para asignar las palabras clave, según el área temática:

Artes: AAT: Art y Architecture Thesaurus.

Ciencias agropecuarias: 1) Agrovoc: Multilingual Agricultural Thesaurus - F.A.O. y 2) GEMET: General Multilingual Environmental Thesaurus.

Ciencias sociales y humanas: 1) Tesauro de la UNESCO y 2) Population Multilingual Thesaurus.

Ciencia y tecnología: 1) Astronomy Thesaurus Index. 2) Life Sciences Thesaurus, 3) Subject Vocabulary, Chemical Abstracts Service y 4) InterWATER: Tesauro de IRC - Centro Internacional de Agua Potable y Saneamiento.

Tecnologías y ciencias médicas: 1) MeSH: Medical Subject Headings (National Library of Medicine's USA) y 2) DECS: Descriptores en ciencias de la Salud (Biblioteca Regional de Medicina BIREME-OPS).

Multidisciplinarias: 1) LEMB - Listas de Encabezamientos de Materia y 2) LCSH- Library of Congress Subject Headings.

También se pueden encontrar listas de temas y palabras claves, consultando las distintas bases de datos disponibles a través del Portal del Sistema Nacional de Bibliotecas¹, en la sección Recursos bibliográficos.^opción "Bases de datos".

Abstract

Es el mismo resumen pero traducido al inglés. Se debe usar una extensión máxima de 12

¹ver: www.sinab.unal.edu.co

renglones. Al final del Abstract se deben traducir las anteriores palabras claves tomadas del texto (mínimo 3 y máximo 7 palabras), llamadas keywords. Es posible incluir el resumen en otro idioma diferente al español o al inglés, si se considera como importante dentro del tema tratado en la investigación, por ejemplo: un trabajo dedicado a problemas lingüísticos del mandarín seguramente estaría mejor con un resumen en mandarín.

Keywords: palabras clave en inglés(máximo 10 palabras, preferiblemente seleccionadas de las listas internacionales que permitan el indizado cruzado)

Contenido

Agradecimientos	VII
Resumen	IX
Lista de símbolos	XIII
1. Estudios del Cotransportador vSGLT	2
1.1. Introducción	2
1.2. Formación de Péptidos y Proteínas	3
1.2.1. Estructura de las Proteínas	4
1.3. Proteínas de Membrana	11
1.4. Transportadores	11
1.4.1. Cotransportadores	12
1.5. Algunas Familias Proteicas y el plegamiento LeuT (LeuT fold)	12
1.6. Co-transportador vSGLT	13
1.7. Estudios actuales del Co-transportador vSGLT	13
2. Modelos Teóricos	14
2.1. Relevancia	14
2.2. Dinámica de una Biomolécula	14
2.3. Movimientos Locales	15
2.3.1. MD usando todos los átomos	15
2.3.2. Análisis de Modos Normales Estándar (NMA estándar)	16
2.4. Movimientos Globales	17
2.4.1. Análisis de Modos Normales (NMA)	17
2.4.2. Análisis por Componentes Principales (PCA)	19
2.5. Descripción Mecánica de NMA	19
2.6. Descripciones de los Movimientos Globales	25
2.6.1. ENMs	25
2.6.2. Análisis por Componentes Principales (PCA)	31
2.6.3. Modelos de Bloque Rígido (BNM o RTB)	32

2.7.	Comparación Entre distintos modelos	32
2.7.1.	Diferencias entre GNM y ANM	32
2.8.	Clasificación Modelos Teóricos	32
3.	Simulación Computacional	34
3.1.	Introducción	34
3.2.	NMA previo de vSGLT	34
3.3.	ANM para vSGLT	36
3.3.1.	Preparación del PDB	36
3.3.2.	Resultados	36
3.4.	Modelo GNM de vSGLT	36
3.5.	Mutante K294A de vSGLT con C- α	36
A.	Anexo 1: Matriz de Masa del sistema	38
B.	Anexo:	40
	Bibliografía	41

Lista de símbolos

Esta sección es opcional, dado que existen disciplinas que no manejan símbolos y/o abreviaturas.

Se incluyen símbolos generales (con letras latinas y griegas), subíndices, superíndices y abreviaturas (incluir sólo las clases de símbolos que se utilicen). Cada una de estas listas debe estar ubicada en orden alfabético de acuerdo con la primera letra del símbolo.

Símbolos con letras latinas

Símbolo	Término	Unidad SI	Definición
A	Área	m^2	$\int \int dx dy$
A_{BET}	Área interna del sólido	$\frac{\text{m}^2}{\text{g}}$	ver DIN ISO 9277
A_g	Área transversal de la fase gaseosa	m^2	Ec...
A_s	Área transversal de la carga a granel	m^2	Ec...
a	Coeficiente	1	Ec...
a	Contenido de ceniza	1	$\frac{m_{\text{ceniza}}}{m_{\text{bm},0}}$
c	Contenido de carbono	1	$\frac{m_C}{m}$
c	Longitud de la cuerda	m	Figura...
c	Concentración de la cantidad de materia	$\frac{\text{mol}}{\text{m}^3}$	$\frac{n}{V}$
D	Diámetro	m	
E_A	Energía de activación	$\frac{\text{kJ}}{\text{mol}}$	Ec....
F	Fracción de materia volátil	1	ver DIN 51720
Fr	Número de Froude	1	$\frac{\omega^2 R}{g_0}$
\vec{g}	Aceleración de la gravedad	$\frac{\text{m}}{\text{s}^2}$	$\frac{d^2 \vec{r}}{dt^2}$
H	Entalpía	J	$U + PV$

Símbolo	Término	Unidad SI	Definición
H_o	Poder calorífico superior	$\frac{\text{MJ}}{\text{kg}}$	ver DIN 51857
h	Contenido de hidrógeno	1	$\frac{m_H}{m}$
K	Coeficiente de equilibrio	1	Ec...
L	Longitud	m	<i>DF</i>
L	Longitud del reactor	m	Figura...
m	Masa	kg	<i>DF</i>
\dot{m}	Flujo de masa	$\frac{\text{kg}}{\text{s}}$	$\frac{m}{t}$
n	Velocidad de rotación	$\frac{1}{\text{s}}$	$\frac{\omega}{2\pi}$
n	Cantidad de materia	mol	<i>DF</i>
P	Presión	Pa	$\frac{\vec{F} \cdot \vec{n}}{A}$
Q	Calor	kJ	1. <i>LT</i>
T	Temperatura	K	<i>DF</i>
t	Tiempo	s	<i>DF</i>
x_i	Fracción de la cantidad de materia	1	$\frac{n_i}{n}$
V	Volumen	m^3	$\int dr^3$
\vec{u}	Velocidad	$\frac{\text{m}}{\text{s}}$	$(\frac{dr}{dt}, r \frac{dv}{dt}, \frac{dz}{dt})$
w_i	Fracción en masa del componente i	1	$\frac{m_i}{m_0}$
$w_{w,i}$	Contenido de humedad de la sustancia i	1	$\frac{m_{H_2O}}{m_{i,0}}$
Z	Factor de gases reales	1	$\frac{pv}{RT}$

Símbolos con letras griegas

Símbolo	Término	Unidad SI	Definición
α_{BET}	Factor de superficie	$\frac{\text{m}^2}{\text{g}}$	$(w_{F,waf})(A_{\text{BET}})$
β_i	Grado de formación del componente i	1	$\frac{m_i}{m_{bm,0}}$
γ	Wandhaftreibwinkel (Stahlblech)	1	Sección...
ϵ	Porosidad de la partícula	1	$1 - \frac{\rho_s}{\rho_w}$
η	mittlere Bettneigungswinkel (Stürzen)	1	Figura...

Símbolo	Término	Unidad SI	Definición
θ	Ángulo de inclinación de la cama	1	Figura...
θ_0	Ángulo superior de avalancha	1	Figura...
θ_U	Ángulo inferior de avalancha	1	Figura...
κ	Velocidad de calentamientoe	$\frac{K}{s}$	$\frac{dT}{dt}$
ν	Coeficiente estequiométrico	1	ver DIN 13345
ρ_b	Densidad a granel	$\frac{kg}{m^3}$	$\frac{m_S}{V_S}$ (Sección...)
ρ_s	Densidad aparente	$\frac{kg}{m^3}$	$\frac{m_F}{V_P}$ (Sección...)
ρ_w	Densidad verdadera	$\frac{kg}{m^3}$	$\frac{m_F}{V_F}$ (Sección...)
τ	Tiempo adimensional	1	Ec....
Φ_V	Flujo volumétrico	$\frac{m^3}{s}$	$\frac{\Delta V}{\Delta t}$
ω	Velocidad angular	$\frac{1}{s}$	$\frac{dv}{dt}$

Subíndices

Subíndice	Término
bm	materia orgánica
DR	Dubinin-Radushkevich
E	Experimental
g	Fase gaseosa
k	Condensado
Ma	Macroporos
P	Partícula
p	Poro
p	Piroizado
R	Reacción
t	Total
wf	Libre de agua
waf	Libre de agua y de ceniza
0	Estado de referencia

Abreviaturas

20 Aminoácidos Proteinogénicos

Aminoácido	Abreviatura	
	3 letras	1 letra
Alanina	Ala	A
Arginina	Arg	R
Asparagina	Asn	N
Aspartato	Asp	D
Cisteína	Cys	C
Glutamato	Glu	E
Glutamina	Gln	Q
Glicina	Gly	G
Histidina	His	H
Isoleucina	Ile	I
Leucina	Leu	L
Lisina	Lys	K
Metionina	Met	M
Fenilalanina	Phe	F
Prolina	Pro	P
Serina	Ser	S
Treonina	Thr	T
Triptófano	Trp	W
Tirosina	Tyr	Y
Valina	Val	V

Abreviatura	Definición
PDB	Protein Data Bank
ANM	Anisotropic Network Model
GNM	Gaussian Network Model

Abreviatura	Definición
vSLGT	Cotransportador de Sodio/Galactosa de la bacteria
SSS//Familia de simportadores de Sodio/Soluto	

Introducción

EDITAR El estudio de los cotransportadores de azúcar, en particular, transportadores de glucosa SGLT dependientes de sodio son esenciales en la producción de metabolismo y la energía celular. Los cotransportadores SGLT son miembros de la familia de portadores de soluto (SLC5) y algunos de estos transportadores de interés tienen una secuencia y estructura similar tridimensional similar. En este caso se examinó el co-transportador dependiente sodio galactosa del Vibrio parahaemolyticus (vSGLT), que media el transporte de galactosa en el citoplasma de las bacterias Vibrio parahaemolyticus. Según la literatura, la cinética del co-transportador tiene entre 5 y 6 estados o conformaciones, pero en este caso de que la desvinculación de los sustratos se estudia la conformación, también conocido como modelo de liberación estado de co-transportador que mira hacia dentro. se realizó un estudio computacional para analizar los movimientos globales de un transportador vSGLT, y comparamos nuestros resultados computacionales con los que se encuentran en los anteriores informes experimentales. análisis de modos normales con un modelo elástico de red (ENM) fue utilizado para explorar los cambios en los movimientos globales entre vSGLT en la presencia o ausencia de los iones que transportan (Na^+ , galactosa). ENM se ha demostrado que es un cálculo útil herramienta para predecir la dinámica de las proteínas de membrana en muchas aplicaciones. los modos normales más bajas generadas por la ENM proporcionar información valiosa sobre la dinámica global de las biomoléculas que son relevantes para su función.

- 2 a 4 páginas

Capítulo 1

Estudios del Cotransportador vSGLT

1.1. Introducción

El presente estudio se ocupa de una proteína encontrada en la membrana de la bacteria *Vibrio Parahaemolyticus*. Pero antes de realizar dicho estudio es necesario responder las siguientes preguntas fundamentales: ¿qué es una proteína? ¿de qué está conformada una proteína? ¿dónde se encuentran las proteínas? ¿cuál es el papel que desempeñan las proteínas en los seres vivos? ¿Cuál es la forma de las proteínas?.

Una proteína es un polímero (polí- Muchas -mero: Partes) que está formado por una gran cantidad de unidades del mismo tipo, estas unidades se conocen como aminoácidos. Específicamente se dice que las proteínas son polipéptidos¹ que tienen más de 50 aminoácidos, haciendo que su peso molecular sea mayor a 5000 Da [Kuchel et al., 2009].

Un aminoácido es una molécula orgánica formada por un carbono llamado α alrededor del cual se encuentran los grupos funcionales carboxilo y amino, además de un hidrógeno y un radical que le da la identidad a cada aminoácido, ver figura 1-1.

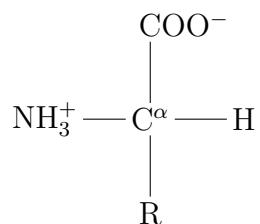


Figura 1-1.: Forma general de un L-aminoácido a pH 7. El radical R cambia para cada aminoácido.

¹Péptido: Molécula formada por una cadena de varios aminoácidos mediante un enlace llamado peptídico; normalmente se le dice peptido a una cadena con menos de 20 ~ 30 aminoácidos

A continuación se muestran los 20 aminoácidos proteinogénicos comunes² clasificados de acuerdo a la carga, la polaridad y la formación de grupos aromáticos. Las abreviaciones de los aminoácidos se encuentran en la sección Lista de Símbolos.

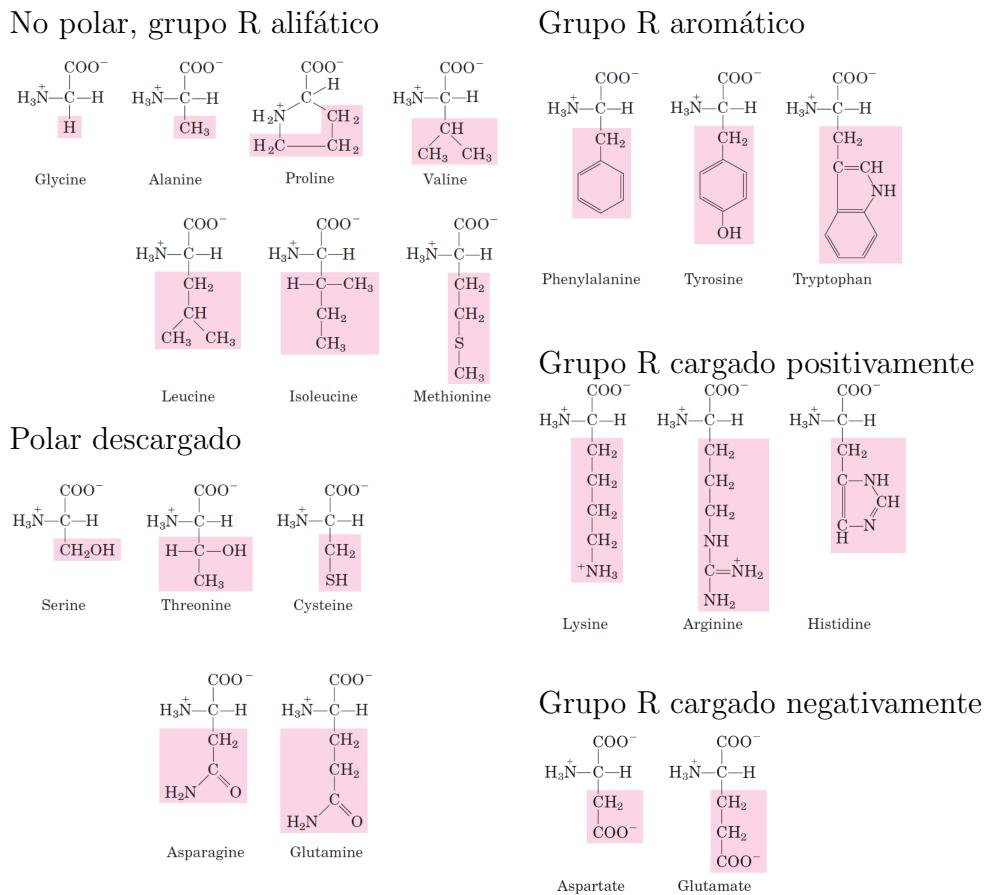


Figura 1-2.: Fórmulas estructurales de los 20 aminoácidos proteinogénicos a pH 7 clasificados según su radical de color rosado. Tomado de [Nelson and Cox, 2011].

Exceptuando la glicina, todos los aminoácidos presentan la propiedad de la quiralidad, existiendo dos formas posibles para cada aminoácido: L-aminoácidos o D-aminoácidos. La distinción va según la dirección en la que desvía la luz con respecto al centro quiral que es el C- α del aminoácido. Prácticamente todos los aminoácidos encontrados en proteínas tienen la forma L.

1.2. Formación de Péptidos y Proteínas

Dos aminoácidos reaccionan formando un enlace llamado peptídico, esto ocurre cuando el carbono del grupo carboxilo se enlaza covalentemente con el nitrógeno del grupo amino

²Son los incorporados en la síntesis de proteínas durante la traducción en el ribosoma

produciendo una deshidratación, es decir, liberando agua. En la figura 1-3 se muestran los reactantes y los productos de la reacción.

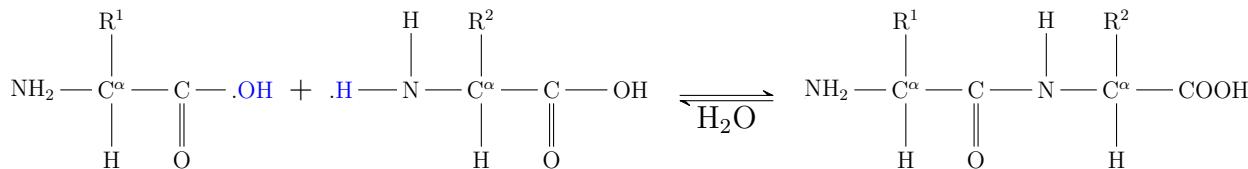


Figura 1-3.: Formación de un dipéptido. Se muestran los reactantes sin ionizar para exemplificar, en sus formas poliiónicas se ioniza el amino-terminal y el carboxi-terminal

El enlace peptídico es un enlace amida de la forma $R^1C(O)NHR^2$. Debido a que las amidas forman una estructura de resonancia tal como se ilustra en la figura 1-4, el enlace peptídico forma un enlace parcial doble. Este enlace parcial doble, como los característicos de las estructuras de resonancia, es un híbrido entre el enlace simple y el enlace doble. De hecho, Linus Pauling y Robert Corey [Nelson and Cox, 2011], encontraron que la longitud del enlace peptídico era de $1,32\text{\AA}$ la cual es menor a la de un enlace C–N simple ($1,49\text{\AA}$) y mayor a la de un enlace doble covalente C=N ($1,27\text{\AA}$).

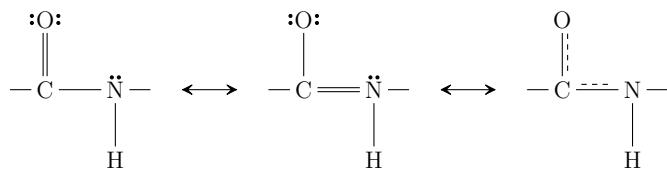


Figura 1-4.: Posibles estructuras de Lewis de la anión amida $[C(O)NH]^{-2}$ que al mezclarlas producen el estado resonante.

Otro apunte que hay resaltar es el hecho de que la distancia entre los carbonos α es de $3,8\text{\AA}$, [Smith et al., 1996], la cual se pone en evidencia cuando se involucra la estructura de los péptidos y de las proteínas en el espacio tridimensional.

1.2.1. Estructura de las Proteínas

De acuerdo a Kai Linderstrøm-Lang, la estructura de las proteínas puede considerarse en varios niveles, conocidos como estructuras *primaria, secundaria, terciaria y cuaternaria*. En lo que sigue se detallan cada uno de estos niveles.

Estructura Primaria y Secundaria

Por *estructura primaria* se entiende la secuencia u orden en que se encuentran los residuos unidos por enlaces peptídicos dentro de un péptido o proteína. Generalmente se conoce

primero por métodos experimentales de clivaje, la secuencia de los residuos. Conocer la secuencia es fundamental ya que esta determina la estructura tridimensional.

En la figura 1-3 se observa un residuo contiguo al otro en forma horizontal, es decir en su estructura primaria. Sin embargo, al observar al péptido tridimensionalmente la cadena que se forma no es lineal. Esto se debe a interacciones no covalentes entre los átomos que no necesariamente están conectados por el enlace peptídico. Estas interacciones hacen que los grupos funcionales dentro del péptido se encuentren a ángulos diferentes. Para estudiar la estructura tridimensional se describirán inicialmente, los ángulos de rotación entre los residuos que componen al péptido o proteína.

Linus Pauling y Robert Corey encontraron que el enlace peptídico forma una estructura planar (rígida) que contiene 4 átomos enlazados alrededor del C–N, la cual se presenta debido al enlace parcial doble, de tal manera que fija el ángulo de rotación del grupo C=O con respecto al grupo N–H. El ángulo ω entre los dos grupos es, excepto cuando hay prolinas, de 180° (conformación trans). En la figura 1-5 se pueden observar los *planos amida* de los residuos.

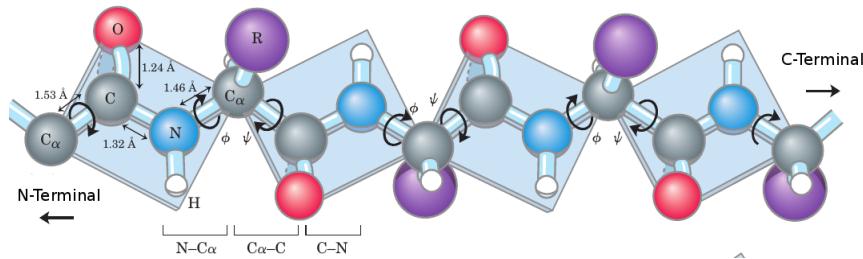


Figura 1-5.: Ángulos dihedros ϕ , ψ y ω (No mostrado). Tomado de [Nelson and Cox, 2011].

El ángulo ϕ es 0° cuando el grupo N–H está en la posición trans respecto al enlace C α –C mientras que el ángulo ψ es 0° cuando el enlace C–N está en la posición trans respecto al enlace C=O, ver [Kuchel et al., 2009].

La *estructura secundaria* se define como la conformación local que genera estructuras repetitivas. Al definir la estructura secundaria se simplifica el entendimiento de la estructura tridimensional de la proteína ya que permite identificar patrones que se repiten en diferentes proteínas.

Existen tres tipos comunes de estructura secundaria, ver [Kuchel et al., 2009]:

1. α Hélice: Ocurre si la cadena principal se enrolla formando una hélice, de tal manera que el oxígeno del grupo carbonilo en un residuo forma un *puente de hidrógeno* con el

hidrógeno de la amina secundaria encontrado 4 residuos más adelante. En la figura **1-6** se observan mediante las líneas punteadas, átomos de oxígeno (rojo) interactuando con los protones (gris claro) en otro residuo. La cadena principal o esqueleto se representa con una cinta enrollada, figura de la izquierda cumpliendo la propiedad de tener 3.6 residuos por cada vuelta. Se observa que la secuencia en la hélice avanza según la regla de la mano derecha, esto es característico de todas las hélices formadas por L-aminoácidos, estas se les denomina α_R hélices (Dextróginas).

2. Hoja plegada β o lámina β : Las hojas plegadas beta aparecen casi formando láminas, contrario a la α hélice. La hoja es formada por dos o más segmentos polipeptídicos, contrario a la α hélice, donde sólo se requiere una cadena polipeptídica para formar la hélice. Estos segmentos polipeptídicos están unidos por puentes de hidrógeno en los grupos carbonilo y amina de sus residuos. Hay dos tipos de láminas β , las *paralelas* que tienen sus cadenas polipeptídicas las mismas direccións, es decir del amino terminal al carboxi terminal o *antiparalelas*, es decir, con direcciones opuestas entre sí. En la figura
3. Vueltas (Turns): Son aquéllas conformaciones que no tienen una estructura secundaria repetitiva (pierden su estructura secundaria) e invierten la dirección de la cadena principal. Usualmente aparecen involucrados de 1 a 5 residuos, teniendo a estas cadenas un tamaño menor a 7 Å. Existe una variedad de tipos de vuelta según los ángulos dihedros que formen: [Pavone et al., 1996]
 - a) α Vuelta (En inglés α -turn): Es común que sea la unidad repetitiva de una α_R hélice, [Pavone et al., 1996]. Se forman puentes de hidrógeno entre los grupos amino y carbonilo del primer y el quinto residuo.
 - b) β Vuelta (En inglés β -turn): Es la más común y se presenta entre las conexiones de las hojas β antiparalelas. Ocurre cuando hay puentes de hidrógeno entre los grupos amino y carbonilo del primer y el cuarto residuo. Existen 4 tipos principales de β vueltas, las cuales, a pesar de ser estructuras no repetitivas, se distinguen por tener ángulos ϕ y ψ determinados. Frecuentemente se encuentran con residuos de glicina y prolina.
 - c) γ Vuelta (En inglés γ -turn): Forman sus puentes de hidrógeno entre el residuo i y el $i + 2$. Tienen una región más bien definida en el gráfico de Ramachandran, encontrándose de dos tipos: Clásicas $(\phi, \psi) = (75, -65)$ e invertidas $(\phi, \psi) = (-75, 65)$. En la figura se muestran estas regiones.
 - d) Lazo ω (En inglés ω -loop): Se definen como elementos que aunque tienen una longitud del enlace fijo, las direcciones o ángulos no están correlacionados, es decir, se distribuyen de manera aleatoria. [Smith et al., 1996].

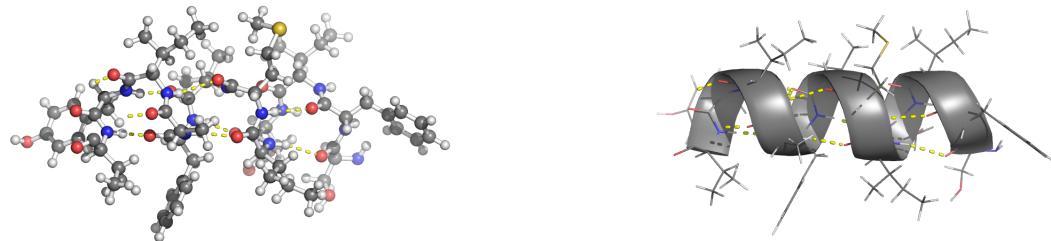


Figura 1-6.: Porción de una α hélice perteneciente a la proteína vSGLT (Código pdb 3DH4) formada por los enlaces de hidrógeno mostrados en color amarillo.



Figura 1-7.: Porción de una hoja β antiparalela perteneciente a la proteína hRBP2 (Código pdb 2RCQ) formada por los enlaces de hidrógeno mostrados en color amarillo.

También hay otros tipos de estructura secundaria como las hélices de colágeno, las π hélices, las 3_{10} hélices, entre otras, que no son de interés para estudiar nuestra proteína.

Los patrones estructurales se forman cuando los ángulos entre residuos son fijos. Esto diferencia un tipo de estructura secundaria de la otra, los ángulos ϕ y ψ entre residuos son diferentes. Aunque no hay únicos ángulos que determinen si conformación es una α hélice o una hoja β , al graficar ψ contra ϕ , se han encontrado regiones permitidas y prohibidas para la formación de estas conformaciones locales. El gráfico de ϕ contra ψ se conoce como gráfico de Ramachandran. En la figura 1-8 se muestran las ángulos permitidos en color oscuro para una α_R hélice, una hoja β y una α_L hélice.

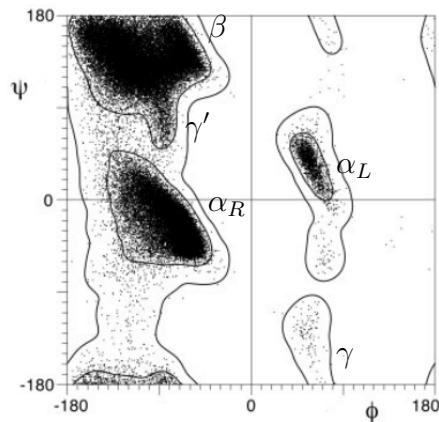


Figura 1-8.: Gráfico de Ramachandran para 97368 residuos tomados de 500 estructuras diferentes mostrando las regiones permitidas para una α_R hélice, una α_L hélice, una hoja β y las vueltas γ y γ' . Imagen tomada de [Lovell et al., 2003]

Estructura Terciaria y cuaternaria

La *estructura terciaria* es la forma en que las cadenas polipeptídicas como un todo (lo global en oposición a lo local) se pliegan en su estructura tridimensional [Kuchel et al., 2009]. Las péptidos y proteínas se pliegan en el entorno biológico, mientras que los péptidos sintéticos en solución generalmente muestran una conformación aleatoria, ver figura 1-9 (a). El plegamiento de las proteínas no es un proceso favorable termodinámicamente, sin embargo, las proteínas adquieren su estructura en condiciones biológicas debido a las interacciones no covalentes y covalentes (puente de disulfuro) presentes dentro de ella misma y con el medio acuoso.

En la *estructura cuaternaria* diferentes cadenas polipeptídicas, denominadas subunidades, se arreglan formando un complejo multiproteico. Estas subunidades están unidas por interacciones proteína-protína, que no son más que interacciones no covalentes. Un complejo multiproteico puede estar formado por dos o más proteínas del mismo tipo, multímeros homotípicos³, o diferentes, multímeros heterotípicos. En la figura 1-9 (b) se muestra un dímero.

³Cuando hay pocas subunidades monoméricas se denomina un homooligómero, en particular, si hay dos monómeros se le dice un dímero, tres monómeros trímero, cuatro monómeros tetrámero, etc.

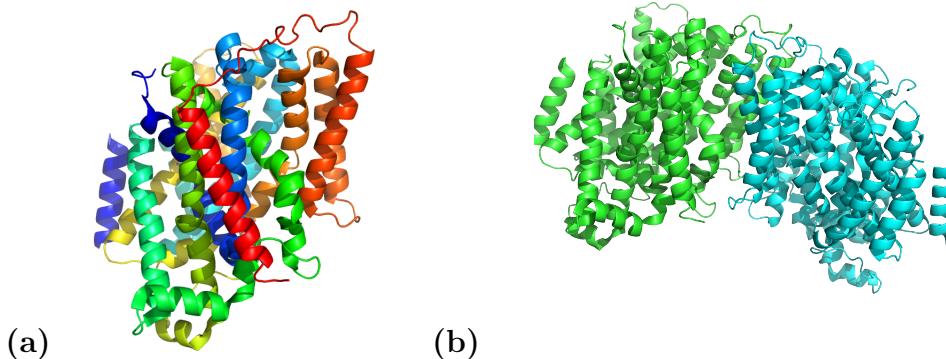


Figura 1-9.: Monómero del cotransportador vSGLT(Código pdb 3dh4); (b)vSGLT mostrado como dímero, como aparece en su forma biológica

Las interacciones no covalentes que pliegan la proteína en la estructura terciaria y unen las subunidades proteicas son: [Kuchel et al., 2009]

- *Interacción electrostática:* Es la interacción de Coulomb entre grupos cargados en un medio dieléctrico. A partir de la fuerza de interacción de Coulomb se puede obtener la energía potencial, que de forma equivalente, es el trabajo requerido para separar un sistema de cargas, en este caso, para separar los grupos que están interactuando y que también viene siendo la energía almacenada por el sistema. La energía potencial electrostática para N grupos funcionales que están interactuando es:

$$V_{elec}(r) = \sum_{j=1}^{N-1} \sum_{i=j+1}^N \frac{Z_i Z_j e^2}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (1-1)$$

Donde N es el número de grupos interactuantes, $i = 1, 2, \dots, N - 1$ y $j = j + 1, \dots, N$ son etiquetas que se les coloca a cada grupo con el fin de sumar todas las interacciones posibles. Z_i y Z_j son las cargas electrónicas de los grupos i y j respectivamente. ϵ_r es la constante dieléctrica o permitividad relativa del medio el cual suele ser acuoso, cuyo valor es $\epsilon_r = 80,10$ a $20^\circ C$ aunque en proteínas de membrana también se puede considerar la bicapa lipídica como otro dieléctrico.

Ha de notarse que la sumatoria interna corre desde $j + 1$ para no repetir la interacción entre el grupo i y el grupo j con la del grupo j y el i , que es la misma.

- *Puente de Hidrógeno:* Es una interacción electrostática entre grupos polares, debida a la asimetría de las cargas. Uno de estos grupos polares necesariamente posee un átomo de hidrógeno unido covalentemente a un átomo electronegativo, lo cual convierte al hidrógeno en un átomo electropositivo. El otro grupo posee un átomo con pares de

electrones sin enlazar.

El grupo al cual está unido el hidrógeno se le conoce como donor de puentes de hidrógeno y al otro grupo como acceptor de puentes de hidrógeno. Se dice también que es un enlace de polarización.

- *Puente salino:* Ocurre cuando entre dos grupos están presentes la interacción electrostática y el puente de hidrógeno.

- *Interacción de Van der Waals:* Es también un enlace de polarización, pero en este caso debida a la presencia de dipolos momentáneos en ciertos grupos con altas diferencias en sus electronegatividades. El potencial de Van der Waals, también conocido como Potencial de Lennard-Jones 6,12 es:

$$V_{len}(r) = \sum_{j=1}^{N+1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] \right\} \quad (1-2)$$

Donde El potencial de Van der Waals para dos se encuentra graficado en la figura 1-10. Las sumatorias expresan el total de interacciones presentes en los grupos con dipolos transitorios. Hay dos términos presentes, uno con potencia r_{ij}^{-6} y el otro con potencia r_{ij}^{-12} . El de r_{ij}^{-12} se conoce como término repulsivo y se puede por analogía con la interacción de Coulomb, cuando las cargas son positivas el potencial electrostático es positivo. Por oposición, el término r_{ij}^{-6} el término atractivo.

- *Interacciones Hidrofóbicas:* Es la forma de organizarse las moléculas hidrofóbicas en un medio acuoso. Cuando se agregan moléculas apolares en el agua, como por ejemplo, los hidrocarburos, las moléculas de agua se organizan de manera energéticamente desfavorable, entonces las moléculas hidrofóbicas se organizan de tal manera que decrezca la entropía de la solución.

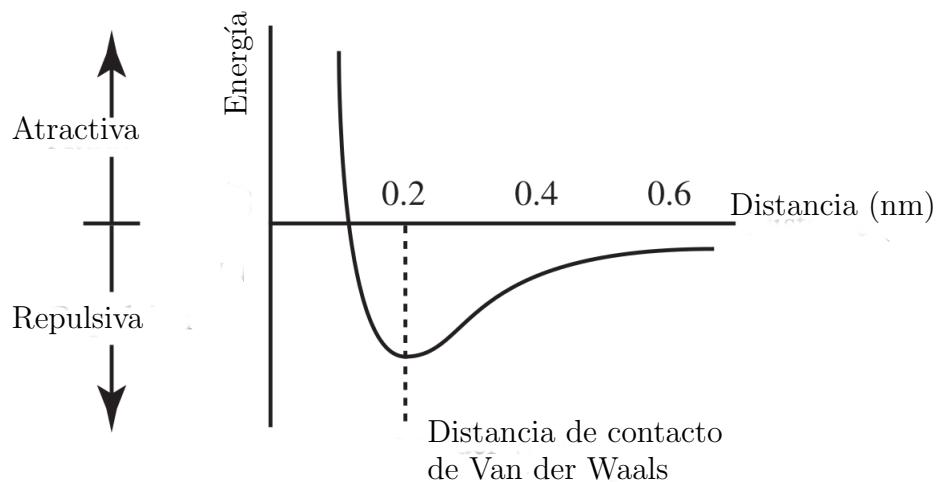


Figura 1-10.: Potencial de Van der Waals entre átomos. Imagen tomada de [Kuchel et al., 2009]

Mientras que en la estructura cuaternaria no aparecen interacciones covalentes, en la estructura terciaria normalmente se considera una sola interacción covalente: *El puente disulfuro*. Esta interacción ocurre cuando los tioles de dos cisteína, que pueden estar lejanas, se oxidan formando el enlace entre los dos azufres de las cisteína.

En la tabla 1-1 se muestran las energías de enlace y los rangos en los que actúan algunas de estas interacciones para ciertos casos particulares o en otros casos una estimación de su valor.

Interacción	Rango (\AA)	Energía de Enlace(kJ/mol)	Descripción
Enlace covalente	1.5	356	Para un enlace $-\text{C}-\text{C}-$
Enlace peptídico	1.32	-	La energía de enlace varía por cada proteína
Electrostática	3	12-20	Residuos en el interior de la proteína
Puente de Hidrógeno	1.9-2.0	5	Entre un carbonilo y un amino
Van der Waals	0.3-0.5	0.4-2.0	$-\text{CH}_2 \cdots \text{H}_2\text{C}-$
Hidrofóbicas	-	~ 3	Transferencia de un $-\text{CH}_2$ de una cadena lateral al agua

Tabla 1-1.: Rango de interacción y energía de enlace para las interacciones presentes en una proteína

1.3. Proteínas de Membrana

1.4. Transportadores

La membrana celular al ser hidrofóbica permite protegerse de la región extracelular, sin embargo, ella necesita ingresar y expulsar todos los compuestos necesarios para realizar su fisiología OJO. La membrana celular tiene proteínas que permiten el ingreso y la expulsión

de estos compuestos. Entre los tipos de proteínas se encuentran los poros, El potencial de Van der Waals para dos se encuentra graficado en la figura **1-10**.

Los transportadores se clasifican, de acuerdo al sistema de clasificación de transportadores [Nelson and Cox, 2011], en dos categorías principales de las cuales se desprenden otras subcategorías :

1. Portadores:

1. Transportadores activos primarios
 - 2.a Transportadores activos secundarios
 - a) Simportadores
 - b) Antiportadores
 - 2.b Uniportadores
2. **Canales iónicos:** Los canales se diferencian de los portadores en la razón a la que transportan los iones, que es de 10^6 iones/segundo (muy alta) así como en que los canales no necesitan energía metabólica para transportar los iones.

Para visualizar las proteínas se puede usar el microscopio de fuerza atómico

1.4.1. Cotransportadores

Hacia la década de los años sesenta Robert Crane, ver [Hamilton, 2013], estableció una relación acoplada o de *cotransporte* entre el ion de sodio y la glucosa los cuales son absorbidos por el intestino delgado. El conocimiento de este mecanismo ha permitido realizar el tratamiento de la diarrea y del cólera mediante la rehidratación oral. La hipótesis del cotransporte que ha sido numerosamente validada y ha sido una piedra angular para el entendimiento de el metabolismo de los carbohidratos, claves en la energética celular. La hipótesis del cotransporte también se ha extendido a otros organismos vivos, con la diferencia de que el acoplamiento del sodio se puede dar con cualquier otro soluto orgánico [Faham et al., 2008].

1.5. Algunas Familias Proteicas y el plegamiento LeuT (LeuT fold)

Las proteínas globulares tienden a formar ciertas arreglos estables de estructura secundaria. Estos arreglos de estructura secundaria se conocen como *estructuras supersecundarias*, *motivos estructurales* o *plegamientos*.

1.6. Co-transportador vSGLT

El cotransportador de Na⁺/Galactosa presente en la proteobacteria *Vibrio Parahaemolyticus* denotado como vSGLT, es un simportador perteneciente a la familia de simportadores de Na⁺-sóluto SSS [?].

La caracterización molecular del compuesto se realizó por primera vez hacia el año 2000 ver [Turk et al., 2000], mientras que la determinación de su estructura fue posible hacia el año 2008 [Faham et al., 2008]. En dichos estudios se establece que vSGLT es similar en la estructura primaria y terciaria a otros transportadores de la familia SLC5 como por ejemplo NIS, SGLT1, SGLT2.

1.7. Estudios actuales del Co-transportador vSGLT

Capítulo 2

Modelos Teóricos

2.1. Relevancia

Es importante estudiar los movimientos de una biomolécula ya que como es señalado en [Lezon et al., 2009] y en [Rader et al., 2006], la dinámica de la molécula vincula la estructura con la función de la biomolécula. La función es el papel que desempeña la biomolécula y que está intimamente relacionado con las interacciones de la biomolécula a un ligando. La estructura de una biomolécula tiene 4 niveles de organización, denominadas *estructuras primaria, secundaria, terciaria y cuaternaria*.

Por ejemplo, en una proteína la estructura primaria es la secuencia u orden de los monómeros (aminoácidos) que la constituyen. La estructura secundaria está determinada por los puentes de hidrógeno existentes entre los grupos amino y carboxilo de dos aminoácidos diferentes; predice la estructura tridimensional en forma local. Mientras que la estructura terciaria da el arreglo tridimensional de cada uno de sus átomos, específicamente se determina con las coordenadas de cada uno de los átomos. La estructura cuaternaria es el arreglo de unidades proteicas o cadenas peptídicas formando lo que se conoce como complejo multiproteico.

Algunas proteínas están en la clase de proteínas de transporte que se encuentran en la membrana celular, la función que cumplen es actuar como mediadoras para el transporte de iones y sustratos, convirtiéndose en un paso previo al metabolismo y la energética celular al interior de la célula.

2.2. Dinámica de una Biomolécula

La dinámica de una biomolécula se determina por las ecuaciones de movimiento para cada uno de los átomos que la constituyen. Usualmente en una biomolécula el número de monómeros es mayor a 20, que al multiplicarlo por el número de átomos en cada monómero incrementa considerablemente el número de ecuaciones de movimiento a resolver. En el

presente caso, se ha determinado que en cada cadena del cotransortador vSGLT hay 530 aminoácidos, conformando un total de 3854 átomos (ver [Faham et al., 2008] y el archivo PDB). De ahí que sea necesario realizar simulaciones computacionales por *dinámica molecular* (Molecular Dynamics que por sus siglas en inglés es MD) en las cuales se estudia el movimiento como función del tiempo de los átomos (Full atomic MD) de acuerdo a las interacciones que presenten.

Las ecuaciones de movimiento se pueden conocer a partir de los formalismos lagrangiano o hamiltoniano [Goldstein et al., 2001], en los cuales es necesario conocer los potenciales con los que interactúan los átomos. Las soluciones a las ecuaciones de movimiento se encuentran mediante los métodos de la dinámica molecular, el análisis de modos normales (Normal Mode Analysis que por sus siglas en inglés es NMA) o el análisis por componentes principales (Principal Component Analysis, PCA por sus siglas en inglés) los cuales difieren en los modelos de potencial y en los resultados relevantes. Mientras que en MD se obtiene la posición como función del tiempo, en los NMA y en los PCA es más importante analizar el espectro, es decir, los autovalores y autovectores obtenidos de una diagonalización.

Los diversos modelos de potencial pueden ser tomados según la naturaleza del polímero a analizar, ver [Lezon et al., 2009]. Sin embargo, al escoger el potencial para hacer un análisis *in silico* de la dinámica de una biomolécula, debe tenerse en cuenta el costo computacional requerido, esto es, el tiempo de simulación de la molécula y la exactitud requerida en el movimiento de cada uno de los constituyentes de la molécula.

De acuerdo a los parámetros de costo, tiempo, escala y detalle en el sistema, las simulaciones de biomoléculas se pueden hacer analizando los *movimientos locales* y los *movimientos globales*.

2.3. Movimientos Locales

Hacen referencia a las simulaciones en las que se incluyen todos los átomos junto con las interacciones presentes, es decir, en las que se analizan los *cambios locales*. Estas se pueden simular a un orden de magnitud de los nanosegundos en una máquina usual, al respecto ver [Gur et al., 2013].

2.3.1. MD usando todos los átomos

Existen simulaciones MD usando todos los átomos (Full atomic MD) o usando los movimientos globales (Coarse-grained models). Como caso particular se pueden considerar los potenciales usados en [Amber, 2016], que siguen el modelo de Amber. El modelo de Amber tiene en cuenta las contribuciones debidas a:

- Interacciones intermoleculares: Son las producidas por los enlaces covalentes entre grupos de átomos, las de valencia y las torsiones.

$$V_{cov}(r) = \sum_{enlaces} k_r (r - r_0)^2 \quad (2-1)$$

$$V_{val}(r) = \sum_{val} k_\theta (\theta - \theta_0)^2 \quad (2-2)$$

$$V_{tor}(r) = \sum_{torsiones} \sum_n \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \quad (2-3)$$

- Interacciones entre pares: Lennard Jones, electrostático.

$$V_{len}(r) = \sum_{j=1}^{N+1} \sum_{i=j+1}^N f_{ij} \left\{ \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^6 \right] \right\} \quad (2-4)$$

$$V_{elec}(r) = \sum_{j=1}^{N-1} \sum_{i=j+1}^N \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2-5)$$

2.3.2. Análisis de Modos Normales Estándar (NMA estándar)

El análisis de modos normales es el análisis armónico que en este caso es aplicado a las proteínas. Los modos normales ocurren cuando todos los constituyentes del sistema oscilan sinusoidalmente y con la misma frecuencia. Los potenciales que cumplen con estas características son los potenciales tipo Hooke; ya que los potenciales tienen una forma complicada como la del modelo de Amber, entonces se aproxima el potencial alrededor de un punto de equilibrio a una cuadrática.

En el análisis *estándar* los constituyentes son *todos los átomos* de la biomolécula. Al utilizar todos los átomos, los costos computacionales de las simulaciones pueden ser mayores que los de otros modelos en los que no se usan todos los átomos. A pesar de esto, el NMA estándar es útil para estudiar los movimientos funcionales de la biomolécula.

De acuerdo a [Hayward and de Groot, 2008], los pasos requeridos para realizar un modelo de NMA estándar son:

1. Una minimización de la energía potencial conformacional en función de las coordenadas cartesianas atómicas. Se usan las coordenadas atómicas cristalográficas.
2. El cálculo de la matriz “Hessiana”, que es la matriz de las segundas derivadas de la energía potencial evaluada en el mínimo.
3. La diagonalización de la matriz de Hessiana.

2.4. Movimientos Globales

Son aquéllas simulaciones en las que se desean conocer los *cambios globales* o el aspecto general que excibe el movimiento de una biomolécula haciendo simplificaciones, ya sea en los potenciales presentes o en el número de átomos interconectados. Algunas de estas simulaciones pueden ser realizadas a un orden de magnitud de los microsegundos y en algortas pueden obtenerse resultados en tan solo unos minutos, lo cual facilita su uso en computadores personales, al respecto ver [Gur et al., 2013]. Los modelos en que se reemplaza la descripción atomística por una de más baja resolución se conocen como modelos de grano grueso, en inglés *coarse-grained models*.

De forma semejante a la red cristalina en el estado sólido, los bloques constituyentes de la biomolécula se reemplazan por un punto representativo llamdo *nodo*. Aunque los bloques constituyentes son los residuos (monómeros) de la biomolécula.

En una proteína las unidades monoméricas son los aminoácidos y se suelen tomar como nodos las posiciones de equilibrio de los C- α en la estructura cristalina. Ver por ejemplo el tratamiento realizado al inhibidor de la tripsina pancreática bovina (BPTI) y al transportador de aspartato en arqueas GltPh, ver [Gur et al., 2013]. Aunque se suelen tomar los C- α esto puede variar ligeramente para dar compatibilidad al modelo con los datos experimentales. Existen programas como MAVEN [Zimmermann et al., 2011] que permiten escoger el centro de gravedad de cada monómero como posición de equilibrio del nodo, usar modelos de grano grueso esférico o usar modelos de resolución mixtas.

Un conjunto de modelos que permite calcular los movimientos globales de una molécula son los análisis por modos normales (Normal Mode Analysis o NMA por sus siglas en inglés). Otros modelos que describen los movimientos globales son los análisis por componentes principales (Principal Component Analysis o PCA por sus siglas en inglés).

2.4.1. Análisis de Modos Normales (NMA)

Existe una versión equivalente del análisis de modos normales en el cual se usa el mismo potencial pero no se consideran todos los átomos de la red sino las unidades monoméricas de la biomolécula.

Modelos de Redes Elásticas (ENM)

Propuesto por primera vez en [Tirion, 1996], los ENM, como la palabra *elástico* lo indica, se basan en una simplificación de la energía potencial a una energía potencial elástica, es decir de tipo Hooke, ver figura 2-1.

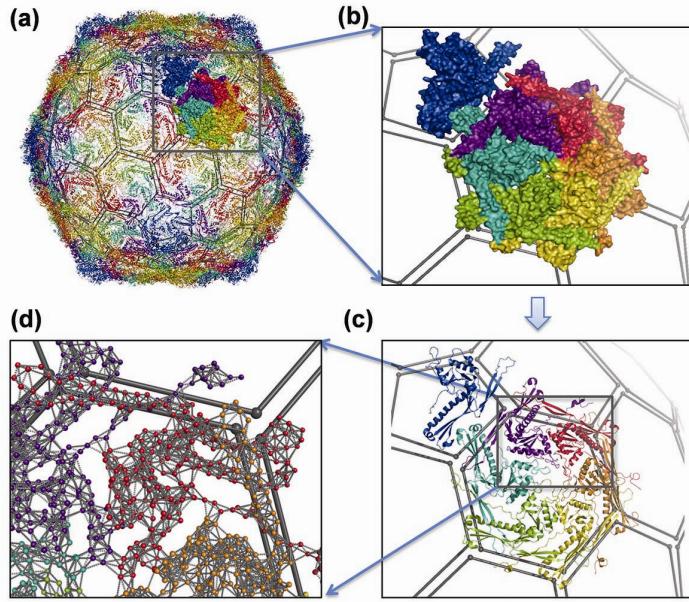


Figura 2-1.: (a) Vista exterior de un cápside vírico HK97 coloreado por cada cadena, todas las proteínas son idénticas. (b) Vista del arreglo proteínico en una cara del cápside. (c) Vista de la estructura secundaria de las proteínas (d) Esquema de cada proteína mostrando cada uno de sus átomos, las aristas de cada cara son carbonos α unidos por lados (ligaduras elásticas). Tomado de [Lezon et al., 2009].

El modelo es bueno cuando la estabilidad de los enlaces con respecto a su posición de equilibrio es menor a una estimación de 13\AA ; para distancias mucho mayores el potencial se corta; la distancia a partir de la que se corta el potencial se denomina distancia de corte R_c . La distancia de corte es uno de los parámetros requeridos en el modelo, el otro es la constante elástica que puede ser la misma para todas las unidades o puede variar.

La primera vez que se usaron los ENM en la investigación de proteínas fue gracias a Monique Tirion, ver [Tirion, 1996]. Entre las proteínas investigadas se encuentran la actina G (Código PDB: 1atn) y la miosina H1 (Código PDB: 1my). La G-actina es modelada bien por los ENM.

Existen dos tipos de ENM: Modelos de redes anisotrópicas (ANM por sus siglas en inglés) y Modelos de Redes Gaussianas (GNM por sus siglas en inglés) los cuales se diferencian en el modelo de potencial usado. El modelo de redes anisotrópicas inicialmente no usa el mismo potencial que el análisis de modos normales estándar sino que usa el potencial presente entre resortes tridimensionales, en la figura [?] se muestra que el potencial entre los resortes depende de la diferencia de distancias.

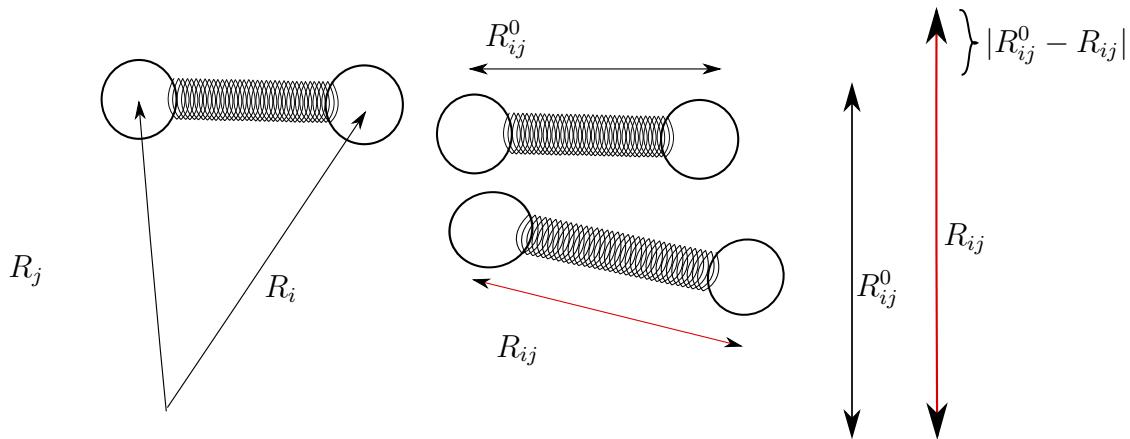


Figura 2-2.: Modelo representativo de las elongaciones de un resorte para ANM

Modelos de Bloques rígidos (BNM)

El modelo de bloques rígidos (BNM por sus siglas en inglés) o también conocido como rotaciones y traslaciones de bloques (RTB por sus siglas en inglés) es un modelo que surge cuando se desean estudiar los modos de un sistema grande como rotaciones y traslaciones de sus partes. Para esto se escogen n_b bloques rígidos con $n_b < N$, para N el número de constituyentes.

2.4.2. Análisis por Componentes Principales (PCA)

En el análisis por componentes principales (PCA por sus siglas en inglés) se buscan las contribuciones principales a las fluctuaciones, entiéndase por esto a las covarianzas entre las componentes de la posición de los átomos. Contrario a los ENM y NMA, los modos que más contribuyen a las fluctuaciones son los de más altas frecuencias [Amadei et al., 1993].

2.5. Descripción Mecánica de NMA

Considérese una biomolécula con N partículas constituyentes, la energía potencial V que representa las interacciones entre los constituyentes de la biomolécula (válido tanto para NMA estándar como para NMA global), se puede expresar alrededor de las posiciones de equilibrio $\mathbf{q}_0 = \mathbf{0}$ tal como describe la teoría de pequeñas oscilaciones, ver [Goldstein et al., 2001]:

$$V(\mathbf{q}) = V(\mathbf{0}) + \sum_{i=1}^n \frac{\partial V}{\partial q_i} \Big|_{\mathbf{q}=\mathbf{0}} q_i + \frac{1}{2} \sum_{ij}^n \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{\mathbf{q}=\mathbf{0}} q_i q_j + \dots \quad (2-6)$$

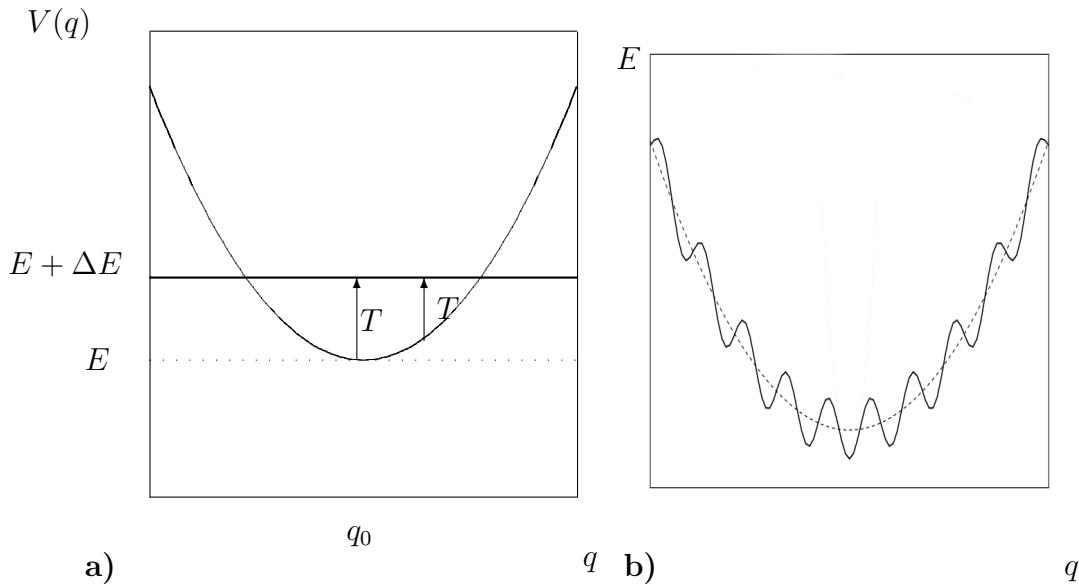


Figura 2-3.: Potencial en función de la posición, a) mostrando la energía total E y los cambios en la energía cinética y b) mostrando el modelo de múltiples mínimos.

Donde q_i son los desplazamientos con respecto a las posiciones de equilibrio, n es el número de posibles desplazamientos en la biomolécula. $V(\mathbf{0})$ es el potencial en equilibrio que por conveniencia puede ser calibrado a cero: $V(\mathbf{0}) = 0$.

El sistema se encuentra alrededor del equilibrio cuando las fuerzas generalizadas se anulan, esto es:

$$\left. \frac{\partial V}{\partial q_i} \right|_{\mathbf{q}=\mathbf{0}} = 0 \quad (2-7)$$

En este tipo de casos como la energía se minimiza, se dice que hay un equilibrio estable. Para entender esto, supóngase que la energía total, que en este caso es $E = T + V$ corresponde al punto de equilibrio donde los nodos no se mueven, es decir, $E = V_{min}$ (línea punteada de la figura 2-3). Ahora suponga que hay una incremento de la energía total en una cantidad ΔE , este incremento generará un aumento en la energía potencial, línea continua. Si el sistema se desvía de la posición de equilibrio, por conservación de la energía y como $E = T + V$, la energía cinética disminuye. En un equilibrio inestable la energía cinética aumenta, alejando al sistema del equilibrio. [Goldstein et al., 2001]

En la figura 2-3 también debe considerarse que la escala de las distancias está entre 0.1 y 10nm, pero que al acercarnos a distancias menores la aproximación armónica falla, presentando múltiples mínimos relativos, [Elber and Karplus, 1987].

Considerando la condición de equilibrio (2-7) y despreciando desplazamientos de orden su-

terior se tiene que:

$$V(\mathbf{q}) = \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{\mathbf{q}=\mathbf{0}} q_i q_j \quad (2-8)$$

Donde se identifican las constantes elásticas como:

$$U_{ij} = \frac{\partial^2 V}{\partial q_i \partial q_j} \Big|_{\mathbf{q}=\mathbf{0}} \quad (2-9)$$

En términos matriciales el potencial se puede escribir como:

$$V(\mathbf{q}) = \frac{1}{2} \mathbf{q}^t \mathbf{U} \mathbf{q} \quad (2-10)$$

En la ecuación (2-10) \mathbf{q} es el vector columna formado por los desplazamientos de las posiciones de equilibrio para cada constituyente, i. e., $\mathbf{q}^t = (q_1, q_2, \dots, q_n)$. Por otro lado la matriz de constantes elásticas \mathbf{H} , es una matriz simétrica debido a que la fuerza generalizada se considera conservativa, lo cual permite intercambiar el orden de las derivadas parciales.

Para las pequeñas oscilaciones, no existen ligaduras dependientes explícitamente de el tiempo (holonómicas) luego, la energía cinética de los constituyentes sólo dependerá de los cuadrados de las velocidades generalizadas:

$$T = \frac{1}{2} \mathbf{q}^t \mathbf{M} \mathbf{q} \quad (2-11)$$

Donde \mathbf{M} es la masa generalizada, la cual se expresa en términos de los factores de escala entre sistemas coordenados:

$$M_{jk} = \sum_{i=1}^N m_i \frac{\partial \mathbf{r}_i}{\partial q_j} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \quad (2-12)$$

La forma en que se escribirán los elementos de la masa del sistema depende de la transformación aplicada. Por ejemplo, en una dimensión las componentes de \mathbf{q} se definen como:

$$q_i = x_i - x_{i0} \quad (2-13)$$

En (2-13) x_i es la posición variable de la partícula i con respecto a un sistema fijo y x_{i0} su posición de equilibrio. Para este caso particular (2-12) se convierte en:

$$\begin{aligned} M_{jk} &= \sum_{i=1}^N m_i \delta_{ij} \delta_{ik} \\ M_{jk} &= m_j \delta_{jk} \end{aligned} \quad (2-14)$$

La ecuación (2-14) dice que la matriz \mathbf{M} es una matriz diagonal cuyas componentes son las masas del sistema.

Para el caso tridimensional, la matriz de masas del sistema es (Los detalles del cálculo se encuentran en el anexo A):

$$M_{jk} = (m_{3j-2} + m_{3j-1} + m_{3j}) \delta_{jk} \quad (2-15)$$

Ecuación de Movimiento El sistema satisface la ecuación de un oscilador armónico, lo cual se muestra para las pequeñas oscilaciones en [Goldstein et al., 2001, Chapter 6]:

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{U}\mathbf{q} = \mathbf{0} \quad (2-16)$$

Para convertir la ecuación a la forma estándar se usan las coordenadas de masa ponderada (en inglés mass-weighted coordinates) las cuales cambian la escala en la que se mide la posición:

$$\mathbf{r} = \mathbf{M}^{1/2}\mathbf{q} \quad (2-17)$$

$$\mathbf{K} = \mathbf{M}^{-1/2}\mathbf{U}\mathbf{M}^{-1/2} \quad (2-18)$$

Reemplazando dichas transformaciones en la ecuación de movimiento se obtiene:

$$\ddot{\mathbf{r}} + \mathbf{Kr} = \mathbf{0} \quad (2-19)$$

El efecto que tienen las coordenadas de masa ponderada sobre la energía cinética es el de desaparecer la dependencia con la matriz de masas del sistema, que como se vió puede resultar en una expresión tediosa. Las energías cinética y potencial en las nuevas coordenadas son:

$$\begin{aligned} T &= \frac{1}{2}\mathbf{q}^t\mathbf{M}\mathbf{q} \\ T &= \frac{1}{2}\dot{\mathbf{r}}^t\mathbf{M}^{-1/2}\mathbf{U}\mathbf{M}^{1/2}\dot{\mathbf{r}} \\ T &= \frac{1}{2}\dot{\mathbf{r}}^t\dot{\mathbf{r}} \end{aligned} \quad (2-20)$$

$$\begin{aligned} V(\mathbf{q}) &= \frac{1}{2}\mathbf{q}^t\mathbf{U}\mathbf{q} \\ V(\mathbf{q}) &= \frac{1}{2}\mathbf{r}^t\mathbf{M}^{-1/2}\mathbf{U}\mathbf{M}^{1/2}\mathbf{r} \\ V(\mathbf{q}) &= \frac{1}{2}\mathbf{r}^t\mathbf{Kr} \end{aligned} \quad (2-21)$$

Una solución (de prueba) a la ecuación de movimiento es la oscilatoria, en la cual los constituyentes de la biomolécula tienen la misma frecuencia ω :

$$\mathbf{r} = \mathbf{a}e^{i\omega t} \quad (2-22)$$

El vector \mathbf{a} es denominado vector de amplitudes debido a que cada una de sus componentes corresponde a las amplitudes de cada monómero. Sus valores son fijos en el tiempo.

Al reemplazar la solución de prueba en la ecuación (2-19) se convierte en un problema de autovalores y autovectores:

$$\begin{aligned}\ddot{\mathbf{r}} &= -\mathbf{K}\mathbf{r} \\ \omega^2 \mathbf{a} &= \mathbf{K}\mathbf{a} \\ \lambda_k \mathbf{a}_k &= \mathbf{K}\mathbf{a}_k\end{aligned}\tag{2-23}$$

En (2-23) se ha definido $\lambda = \omega^2$ y posteriormente se ha agregado el subíndice k para distinguir cada uno de los autovalores y autovectores que se encuentren en el problema (2-23). Cuando cada uno de los monómeros tiene la misma frecuencia, a cada frecuencia le corresponde un vector de amplitudes \mathbf{a} , cada uno de estos posibles vectores se les conoce como un *modo normal de oscilación*.

Ha de notarse que la solución total no está compuesta por un único modo, sino por la superposición de estos:

$$\mathbf{r} = \sum_k \mathbf{a}_k e^{i\omega_k t}\tag{2-24}$$

Cuando se tiene esta solución y se desean conocer los modos normales, pueden aplicarse dos métodos: El de la transformada de Fourier para conocer las frecuencias y la transformación a las coordenadas normales; las coordenadas normales son un espacio en el cual se separan los movimientos compuestos.

Las coordenadas normales ζ se definen como:

$$\mathbf{r} = \mathbf{T}\zeta\tag{2-25}$$

Donde \mathbf{T} es una transformación ortogonal $\mathbf{T}\mathbf{T}^T = \mathbf{I}$ formada por los modos normales \mathbf{a}_k pero modificados para que se cumpla la ortonormalidad. Cada componente de ζ es:

$$\zeta_k = C_k e^{i\omega_k t}\tag{2-26}$$

Es decir, las coordenadas normales ζ_k representan a un único modo normal de oscilación, una única frecuencia. Mientras que las columnas de \mathbf{T} proporcionan las direcciones en las que se da cada modo normal.

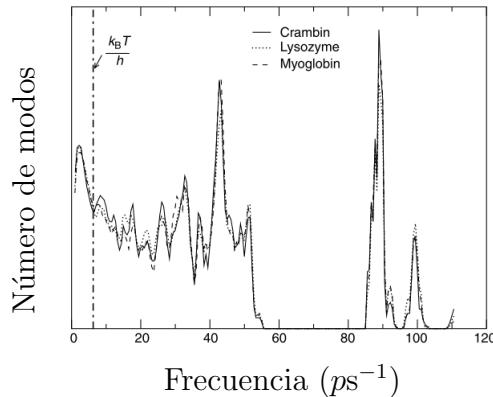


Figura 2-4.: Número de modos normales en función de la frecuencia para tres proteínas diferentes. $k_B T / \hbar$ es la frecuencia correspondiente a la más baja energía promedio a temperatura T . Tomado de [Cui and Bahar, 2006].

Ensamble Estadístico

Cuando el sistema (biomolécula) se encuentre a temperaturas fisiológicas, sólo intercambiará calor con los alrededores (baño térmico). Sin embargo, se ha demostrado que a esta temperatura, la aproximación armónica falla , [Hayward and de Groot, 2008]; por otro lado, a temperatura fisiológica el ensamble es clásico dependiendo del rango de frecuencias a analizar[Sethna, 2006], [Cui and Bahar, 2006]. En la figura 2-4, se observa el número de nodos en función de la frecuencia para diferentes proteínas, para frecuencias menores que la frecuencia de vibración del oscilador armónico a temperatura ambiente, se puede trabajar con la mecánica clásica.

Como únicamente hay intercambio de energía y se trabaja a temperatura fisiológica, la estadística apropiada es la del ensamble canónico. En el ensamble canónico se sigue la distribución de probabilidad de Boltzmann:

$$p = \frac{\exp(-E_s/k_B T)}{Z} \quad (2-27)$$

Donde E_s es la energía del sistema y Z la función de partición. La función de partición es el factor de normalización de la densidad de probabilidad y cuenta todos los posibles microestados a temperatura constante:

$$Z = \frac{1}{\hbar^n} \int d^n p d^n q \exp [-H(\mathbf{p}, \mathbf{q})/k_B T] \quad (2-28)$$

En (2-28), $n = 3N$ cuando el número de coordenadas q_i escogido es el mismo que el número de coordenadas cartesianas.

Ha de notarse que la función de partición en (2-28) desacopla los momentos y las posiciones, ya que el hamiltoniano está dado por la fórmula $H = T + V$, donde la energía cinética depende sólo de los momentos generalizados y el potencial de las coordenadas generalizadas.

Se desarrolla (2-28) teniendo en cuenta que \mathbf{M} es diagonal:

$$\begin{aligned} Z &= \frac{1}{\hbar^n} \int d^n p d^n q \exp \left[-(\dot{\mathbf{q}}^t \mathbf{M} \dot{\mathbf{q}} / 2 + \mathbf{q}^t \mathbf{U} \mathbf{q} / 2) / k_B T \right] \\ Z &= \frac{1}{\hbar^n} \int d^n p \exp \left[-\sum_i \frac{p_i^2}{2m_i} \right] \int d^n r \left(\prod_i \sqrt{m_i} \right) \exp \left[-\frac{1}{2} \mathbf{r}^t \mathbf{K} \mathbf{r} / k_B T \right] \\ Z &= \frac{(2\pi)^{n/2}}{\hbar^n} \int d^n r \exp \left[-\frac{1}{2} \mathbf{r}^t \mathbf{K} \mathbf{r} / k_B T \right] \end{aligned} \quad (2-29)$$

El resultado (2-29) es el mostrado en [Lezon et al., 2009] o en [Sethna, 2006]:

$$Z = \frac{(k_B T)^{n/2}}{\hbar^n} \frac{1}{|\mathbf{K}|^{n/2}} \quad (2-30)$$

Tomando $\hbar = 1$ y teniendo en cuenta que el determinante de \mathbf{K} es el producto de los autovalores no nulos, es decir, de las frecuencias al cuadrado:

$$Z = (k_B T)^{n/2} \prod_{i=1}^n \frac{1}{\omega_i} \quad (2-31)$$

$$Z = (k_B T)^{n/2} |\mathbf{K}^{-1}|^{1/2} \quad (2-32)$$

Se observa que los modos de más bajas frecuencias contribuyen más a la función de partición.

2.6. Descripciones de los Movimientos Globales

2.6.1. ENMs

Modelos de Redes Anisotrópicas (ANM)

Como se ha dicho, la aproximación a segundo orden del potencial, minimización, (2-10) es cambiada por un potencial de Hooke en el cual se tienen en cuenta las distancias en lugar de los desplazamientos alrededor del equilibrio y también se pueden tener en cuenta diferentes constantes elásticas entre diferentes nodos:

$$V = \frac{1}{2} \sum_{i,j>i} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \quad (2-33)$$

Con $R_{ij}, R_{ij}^0 < R_c$

Donde R_{ij}^0 es la distancia de equilibrio entre los nodos $i - j$ de la red, R_{ij} su distancia variable y γ_{ij} las constantes elásticas entre los nodos $i - j$. Como se ilustra en la figura 2-5, el término $R_{ij} - R_{ij}^0$ es la elongación del resorte, lo cual es diferente a la resta de los desplazamientos, ecuación (2-8); entonces no se está approximando el potencial ni se requiere una minimización

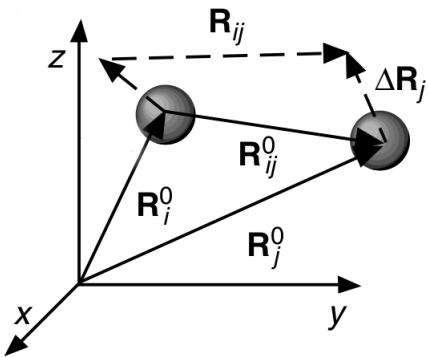


Figura 2-5.: Desplazamientos de los nodos $i - j$ de su posición de equilibrio. Tomado de [Rader et al., 2006].

de la energía, como se ha recalcado, sólo se cambia la forma del potencial.

Se requiere que $j > i$ para que no aparezca repetido el término $(R_{ij} - R_{ij}^0)^2$ con $(R_{ji} - R_{ji}^0)^2$ ya que representan la misma interacción, de ser así, se vuelven obsoletas las constantes elásticas γ_{ij} con $j < i$. Sin embargo, el potencial también puede escogerse dejando los términos repetidos y redefiniendo la matriz de constantes elásticas como una matriz simétrica, esto es: $\gamma_{ij} = \gamma_{ji}$. Escogiendo el potencial de esta manera, como es usual en la literatura [Rader et al., 2006], queda de la forma:

$$V_{ANM} = \frac{1}{2} \sum_{i \neq j=1z} \gamma (R_{ij} - R_{ij}^0)^2 \quad (2-34)$$

Con $R_{ij}, R_{ij}^0 < R_c$

$$\text{y } \gamma_{ij} = \gamma_{ji}$$

Existe un servidor web [Eyal et al., 2015] basado en ANM que permite ver los movimientos globales de una proteína o de un ácido nucleico y permite obtener información como los autovalores, los factores-B, las correlaciones entre otras.¹

Matriz Hessiana El cálculo de la matriz de segundas derivadas \mathbf{U} para el análisis de modos normales, si parte de una minimización de la energía. Las derivadas se calculan vía el potencial anisotrópico (2-34), para el presente caso la matriz \mathbf{U} es llamada \mathcal{H} . La matriz \mathcal{H} es de tamaño $3N \times 3N$ con N el número de residuos y está constituida por $N \times N$ submatrices de tamaño 3×3 llamadas \mathbf{H}_{ij} :

$$\mathcal{H} = \{H_{ij}\} \text{ con } i, j = 1, \dots, N \quad (2-35)$$

¹ Recurso disponible en <http://anm.csb.pitt.edu/cgi-bin/anm2/anm2.cgi>

Los elementos de \mathbf{H}_{ij} son:

$$\mathbf{H}_{ij} = \begin{pmatrix} \partial_{xi}\partial_{xj}V & \partial_{xi}\partial_{yj}V & \partial_{xi}\partial_{zj}V \\ \partial_{yi}\partial_{xj}V & \partial_{yi}\partial_{yj}V & \partial_{yi}\partial_{zj}V \\ \partial_{zi}\partial_{xj}V & \partial_{zi}\partial_{yj}V & \partial_{zi}\partial_{zj}V \end{pmatrix} \quad (2-36)$$

Calculando algunas de estas derivadas y definiendo $x_{ik} = x_k - x_i$, $y_{ik} = y_k - y_i$ y $z_{ik} = z_k - z_i$ se obtiene (Caso $i \neq j$):

$$\frac{\partial V}{\partial y_k} = \sum_j \frac{\gamma_{jk} (R_{kj} - R_{kj}^0) y_{kj}}{R_{kj}} \quad (2-37)$$

$$\left. \frac{\partial^2 V}{\partial x_i \partial y_j} \right|_{R_{ij}=R_{ij}^0} = -\frac{\gamma_{ij} x_{ij} y_{ij}}{R_{ij}^2} \quad (2-38)$$

$$\left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_{R_{ij}=R_{ij}^0} = -\frac{\gamma_{ij} x_{ij}^2}{R_{ij}^2} \quad (2-39)$$

De forma análoga se encuentran los otros elementos de \mathbf{H}_{ij} con lo cual los superelementos no diagonales de \mathcal{H} quedan escritos como [Lezon et al., 2009]:

$$\mathbf{H}_{ij} = -\frac{\gamma_{ij}}{R_{ij}^2} \begin{pmatrix} x_{ij}^2 & x_{ij} y_{ij} & x_{ij} z_{ij} \\ y_{ij} x_{ij} & y_{ij}^2 & y_{ij} z_{ij} \\ z_{ij} x_{ij} & z_{ij} y_{ij} & z_{ij}^2 \end{pmatrix} \quad (2-40)$$

Para los superelementos \mathbf{H}_{ii} que se encuentran en la diagonal, algunos de sus elementos son:

$$\begin{aligned} \left. \frac{\partial^2 V}{\partial x_i^2} \right|_{R_{ij}=R_{ij}^0} &= -\sum_{j \neq i} \gamma_{ij} \left(\frac{x_{ij}^2}{R_{ij}^2} \right) \\ \left. \frac{\partial^2 V}{\partial x_i^2} \right|_{R_{ij}=R_{ij}^0} &= -\sum_{j \neq i} H_{ij,xx} \end{aligned} \quad (2-41)$$

En (2-49), $H_{ij,xx}$ es el primer elemento de la submatriz \mathbf{H}_{ij} . El resultado (2-49) puede generalizarse para todos los superelementos de la diagonal, ver [Lezon et al., 2009]:

$$\mathbf{H}_{ii} = \sum_{j \neq i} \mathbf{H}_{ij} \quad (2-42)$$

Una vez obtenida la matriz Hessiana, es necesario calcular sus autovalores y autovactores de la misma forma que en NMA, ya que esto describe los movimientos de la biomolécula. La diagonalización de \mathcal{H} lleva a lo sumo $3N - 6$ modos normales no nulos, como se demuestra

en [Tirion, 1996]. Los 6 modos normales restantes, corresponden a las rotaciones y traslaciones de un cuerpo rígido; se puede escoger un sistema fijo en la molécula y que rote con ella.

El término de la distancia de corte R_c puede ser incluido en las constantes elásticas, es decir, $\gamma_{ij} = 0$ si $R_{ij} > R_c$ sino $\gamma_{ij} \neq 0$. Adicional a esto, las constantes elásticas pueden ser iguales (isotrópicas): $\gamma_{ij} = \gamma$ si $R_{ij} < R_c$ de lo contrario $\gamma_{ij} = 0$.

Modelo de Redes Gaussianas (GNM)

El modelo se denomina gaussiano debido a que las desviaciones del equilibrio o fluctuaciones son isotrópicas y siguen una distribución gaussiana [Rader et al., 2006]. Se requieren dos parámetros para describir el modelo: Una misma constante elástica γ para todas las unidades de la biomolécula y una distancia de corte a partir de la cual no se ejerce el potencial.

Contrario al ANM, en el modelo de redes gaussianas se consideran las desviaciones del equilibrio como desplazamientos:

$$V_{GNM} = \frac{\gamma}{2} \sum_{i,j=1}^N \Gamma_{ij} (\mathbf{R}_{ij} - \mathbf{R}_{ij}^0)^2 \quad (2-43)$$

Al ser γ_{ij} una matriz de constantes elásticas, se incluyen en la sumatoria los elementos Γ_{ij} donde $\gamma_{ij} = \gamma \Gamma_{ij}$. Γ_{ij} representa la simetría de la matriz γ y como se puede ver en [Rader et al., 2006], la topología de la red.

Ya que $\mathbf{R}_{ij} = \mathbf{R}_j - \mathbf{R}_i$ y definiendo $\Delta\mathbf{R}_i = \mathbf{R}_i - \mathbf{R}_i^0$ se tiene que $\mathbf{R}_{ij} - \mathbf{R}_{ij}^0 = \Delta\mathbf{R}_j - \Delta\mathbf{R}_i$, el potencial gaussiano se puede escribir en términos de los desplazamientos individuales:

$$\begin{aligned} V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \\ V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} [(\Delta x_j - \Delta x_i)^2 + (\Delta y_j - \Delta y_i)^2 + (\Delta z_j - \Delta z_i)^2] \\ V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} [(\Delta x_{ij})^2 + (\Delta y_{ij})^2 + (\Delta z_{ij})^2] \end{aligned} \quad (2-44)$$

Escribiendo (2-44) en forma matricial se tiene:

$$V_{GNM} = \frac{\gamma}{2} (\Delta\mathbf{X}^T \boldsymbol{\Gamma} \Delta\mathbf{X} + \Delta\mathbf{Y}^T \boldsymbol{\Gamma} \Delta\mathbf{Y} + \Delta\mathbf{Z}^T \boldsymbol{\Gamma} \Delta\mathbf{Z}) \quad (2-45)$$

Con $\Delta\mathbf{X}^T$ un vector N dimensional con componentes: $\Delta\mathbf{X}^T = (\Delta x_1, \dots, \Delta x_N)$ y $\boldsymbol{\Gamma}$ la matriz de Kirchhoff que se discutirá en o que sigue con el fin de entender la expresión (2-45).

Para garantizar que Γ_{ij} sea adimensional, se escogen (no de forma única) los elementos fuera de la diagonal de acuerdo al procedimiento descrito a continuación.

Nótese que los términos cuadráticos por cada componente pueden expandirse:

$$\begin{aligned} \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta x_{ij})^2 &= \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta x_i^2 + \Delta x_j^2 - 2\Delta x_i \Delta x_j) \\ &= 2 \sum_{i \neq j=1}^N \Gamma_{ij} \Delta x_i^2 - 2 \sum_{i \neq j=1}^N \Delta x_i \Gamma_{ij} \Delta x_j \end{aligned} \quad (2-46)$$

La idea es que los términos de (2-46) y (2-45) coincidan. Escribiendo explícitamente el término de la componente x en (2-45):

$$\text{editar} \quad (2-47)$$

De lo anterior se concluye que la matriz de Kirchhoff Γ es:

$$\Gamma_{ij} = \begin{cases} -1 & \text{si } i \neq j \text{ y } R_{ij} \leq r_c \\ 0 & \text{si } i \neq j \text{ y } R_{ij} \geq r_c \\ -\sum_{j,j \neq i} \Gamma_{ij} & \text{si } i = j \end{cases} \quad (2-48)$$

Fluctuaciones, Correlaciones .

El objetivo primario es calcular las fluctuaciones cuadráticas medias (ms por sus siglas en inglés) para un residuo y las correlaciones entre residuos. Las fluctuaciones ms proporcionan los movimientos de la biomolécula y las correlaciones establecen la dependencia entre distintas partes de la biomolécula, es decir, los movimientos colectivos.

Las fluctuaciones ms para el residuo i se calculan como:

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_i \rangle = \langle \Delta X_i^2 \rangle + \langle \Delta Y_i^2 \rangle + \langle \Delta Z_i^2 \rangle \quad (2-49)$$

Y las correlaciones entre el residuo i y el j son:

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \langle \Delta X_i \Delta X_j \rangle + \langle \Delta Y_i \Delta Y_j \rangle + \langle \Delta Z_i \Delta Z_j \rangle \quad (2-50)$$

Los promedios calculados en (2-49) y en (2-50) son calculados con respecto al espacio de coordenadas.

De manera similar al resultado , donde la matriz \mathbf{K} es cambiada por la matriz $\gamma \Gamma$ y $n = 3N$ se llega a la función de partición gaussiana:

$$Z_{GNM} = (2\pi)^{3N/2} \left| \frac{k_B T}{\gamma} \Gamma^{-1} \right|^{1/2} \quad (2-51)$$

En [Rader et al., 2006] se encuentra que estas medidas están relacionadas con la inversa de la matriz de Kirchhoff:

$$\langle \Delta \mathbf{R}_i^2 \rangle = \frac{3k_B T}{\gamma} (\boldsymbol{\Gamma}^{-1})_{ii} \quad (2-52)$$

$$\langle \Delta \mathbf{R}_i \cdot \Delta \mathbf{R}_j \rangle = \frac{3k_B T}{\gamma} (\boldsymbol{\Gamma}^{-1})_{ij} \quad (2-53)$$

Factor-B .

El factor-B, factor de Debye-Waller o factor de temperatura es un factor que está directamente relacionado con las amplitudes de los haces dispersados cuando se hace difracción de rayos-x y por lo tanto describe la atenuación de la dispersión.

Para un modelo de armónico isotrópico, donde cada constituyente fluctúa de su posición de equilibrio como: $\Delta \mathbf{R}_i = R_i - R_i^0$, se encuentra que la intensidad de onda dispersada es:

$$I = I_0 \exp \left(-\frac{1}{3} \langle \Delta \mathbf{R}_i \rangle^2 \mathbf{q}^2 \right) \quad (2-54)$$

Con \mathbf{q} el vector de la red recíproca, I_0 la intensidad dispersada para la estructura rígida. El factor-B se define como el término dentro de la exponencial en (2-55). Como está directamente relacionado con la fluctuación ms, entonces da una medida de la rigidez o de la flexibilidad de la biomolécula, además tiene unidades de \AA^2 .

Como en la dispersión de Bragg, el vector de onda de la red recíproca es $4\pi \sin \theta / \lambda$, entonces:

$$B = \frac{8\pi}{3} \langle \Delta \mathbf{R}_i \rangle^2 \quad (2-55)$$

Para el caso GNM, de acuerdo a (2-52) el factor de estructura se convierte en:

$$B = \frac{8\pi^2 k_B T}{\gamma} (\boldsymbol{\Gamma}^{-1})_{ii} \quad (2-56)$$

Existe un servidor web basado en GNM que permite ver los movimientos globales de una proteína o de un ácido nucleico y permite obtener información como los autovalores, los factores-B, las correlaciones entre otras.²

² Recurso disponible en <http://gnm.csb.pitt.edu/index.php>

2.6.2. Análisis por Componentes Principales (PCA)

Sea $\mathbf{x}(\mathbf{t})$ la trayectoria en forma de vector columna de $3N$ componentes que tiene las posiciones a cada instante de tiempo de todos los átomos. La matriz de covarianza, en este caso, es aquélla que mide la dependencia entre cada una de las variables ($3N$ componentes) y cuyos elementos de la diagonal son las varianzas de dichas variables.

Matriz de covarianza:

$$C = \langle (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)(\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \rangle \quad (2-57)$$

En (2-58), la media es calculada para la trayectoria en todos los tiempos.

La matriz de covarianza puede ser expresada en términos de sus autovalores λ_i y sus autovectores \mathbf{t}_i , con $i < 3N + 1$.

$$\mathbf{C}\mathbf{t}_i = \lambda_i \mathbf{t}_i$$

Tomando $\Lambda = \lambda_i$, la matriz diagonal formada por los autovalores y $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{3N})$ como la matriz cuyas columnas son los autovectores, escritas de tal manera que los autovalores estén ordenados de mayor a menor y los autovectores ordenados de acuerdo a los autovalores entonces:

$$\Lambda = \mathbf{T}^T \mathbf{C} \mathbf{T} \quad (2-58)$$

Donde se ha tenido en cuenta que $\mathbf{T}^{-1} = \mathbf{T}^T$. Expresado de otra forma:

$$\mathbf{C} = \mathbf{T} \Lambda \mathbf{T}^T \quad (2-59)$$

La esencia de un PCA es descartar los modos menos relevantes que corresponden a los autovalores más pequeños y generar un nuevo conjunto de datos de menor tamaño a partir de estos (componentes principales). En [Amadei et al., 1993] se demuestra que las desviaciones más grandes corresponden a los autovalores más grandes, por tanto, se trabaja sólamente con los autovalores más grandes.

Dejando sólamente en \mathbf{T} los M autovectores o columnas más relevantes, se define la matriz de las componentes principales $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_M)$, donde $M < 3N$ como:

$$\begin{aligned} \mathbf{P} &= \mathbf{T}^T (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \\ \mathbf{P} &= (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle) \mathbf{T} \end{aligned} \quad (2-60)$$

O en términos de cada componente principal:

$$\mathbf{p}_i = \mathbf{t}_i^T (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \quad (2-61)$$

Si se desea volver a los datos originales hay que realizar la transformación inversa:

$$\mathbf{x}(\mathbf{t})^T = \mathbf{T} \mathbf{P} + \langle \mathbf{x} \rangle^T \quad (2-62)$$

2.6.3. Modelos de Bloque Rígido (BNM o RTB)

Los modos normales de los bloques rígidos tienen una dimensión de tamaño n_b con respecto a los de un NMA, $3N$. Los modos normales de bloque rígido se definen a partir de una transformación denominada \mathbf{P} del espacio de las q_i al de los bloques rígidos r'_i .

La ecuación de movimiento es:

2.7. Comparación Entre distintos modelos

2.7.1. Diferencias entre GNM y ANM

2.8. Clasificación Modelos Teóricos

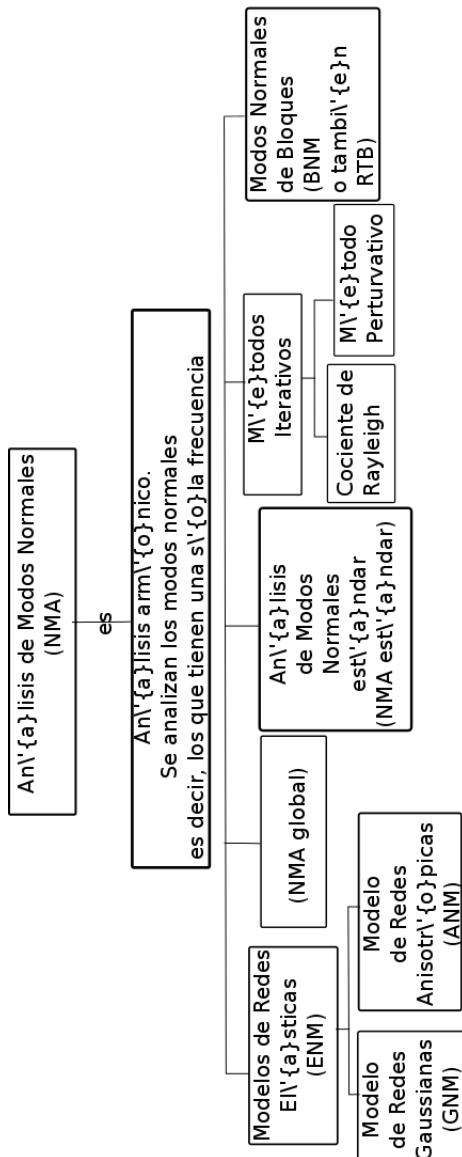


Figura 2-6.: Mapa conceptual en el que se dividen los movimientos locales de los movimientos globales, mostrando los tipos de movimientos globales utilizados

Capítulo 3

Simulación Computacional

3.1. Introducción

Se estudia el cotransportador vSGLT mediante el servidor ANM (En inglés Anisotropic Network Model server) de la universidad de Pittsburg [Eyal et al., 2015]¹. El servidor ANM, como su nombre lo indica, es un servidor web el cual procesa la información de entrada del lado del cliente, en este caso, la información de entrada será el pdb y los parámetros de entrada para realizar los cálculos del ANM, mientras que del lado del servidor se procesa la información suministrada vía un código escrito en C, el cual permite calcular los n primeros modos normales, los factores b, las constantes elásticas, las correlaciones entre distintas partes de la molécula y permite visualizar los modos de vibración mediante java o pymol. En la FIGURA se encuentra un esquema de la entrada y salida de datos

3.2. NMA previo de vSGLT

Mediante el servidor ANM, en aparece un estudio previo del importador vSGLT, en el cual se utiliza un archivo de entrada que es el pdb de 3DH4 encontrado en la base de datos del Protein Data Bank. En este archivo no se determinan los residuos de la primera hélice en cada subunidad. Aunque aparecen las posiciones de los residuos, el ANM server los ignora, haciendo el cálculo únicamente con los residuos conocidos.

Los cálculos de salida mostrados son los factores b por cada número de residuo y para cada distancia de corte r_c , variable entre 7 Å y 14 Å. Ver figura ??.

¹Recurso disponible en <http://anm.csb.pitt.edu/cgi-bin/anm2/anm2.cgi>

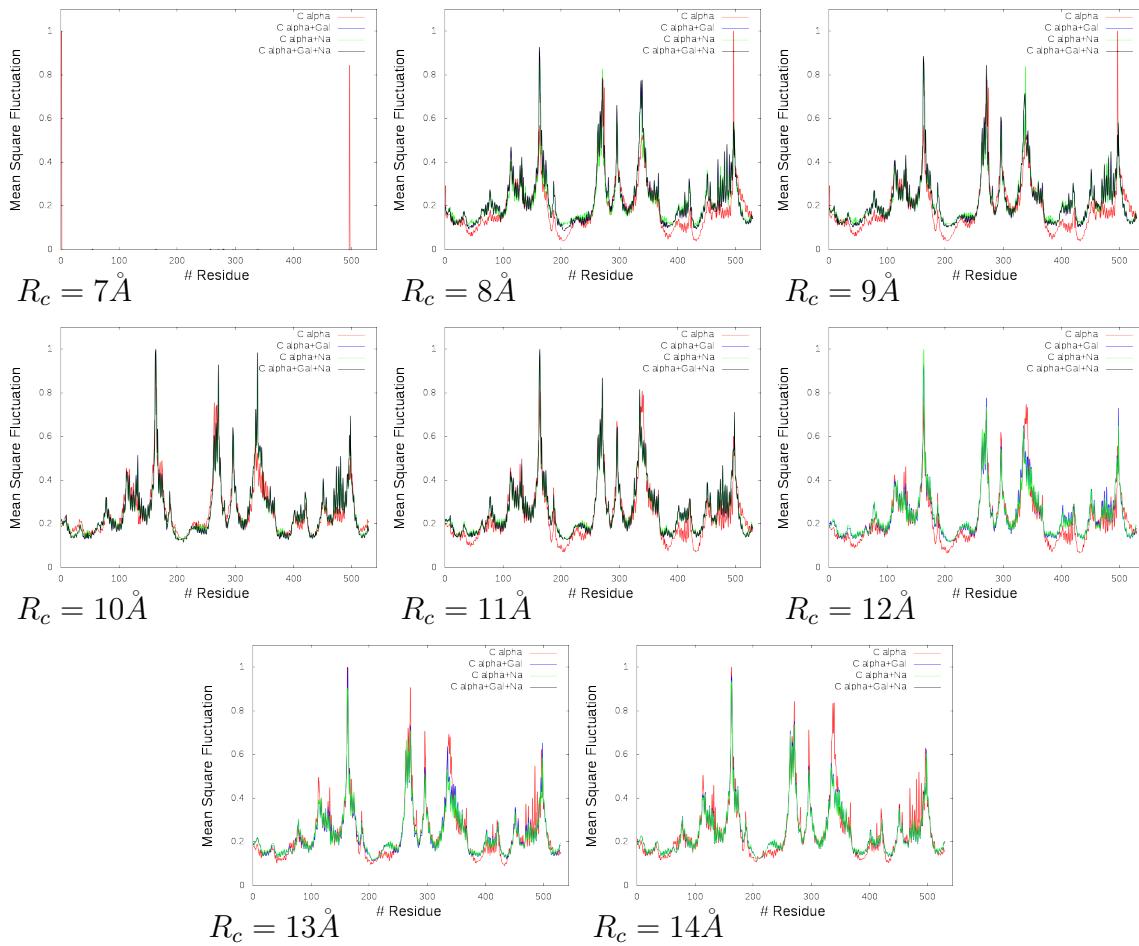


Figura 3-1.: Fluctuaciones ms normalizadas en función del número de residuo entre $7\text{\AA} \leq R_c \leq 14\text{\AA}$ usando los primeros 100 modos. Los diferentes colores indican si la simulación fue realizada sin el ion, el sustrato, con alguno de ellos o ambos.

??

3.3. ANM para vSGLT

3.3.1. Preparación del PDB

Ya que en el mutante de vSGLT en la posición A294A, llamado 2XQ2 aparece resuelta la estructura cristalina del TM1 perteneciente a la cadena A, se usa el programa PyMol, y como archivos de entrada los pdbs 3DH4 y el de su mutante 2XQ2, para usar un archivo que sea cercano a la estructura real objeto de estudio, esto es, al cotransportador vSGLT.

3.3.2. Resultados

3.4. Modelo GNM de vSGLT

3.5. Mutante K294A de vSGLT con C- α

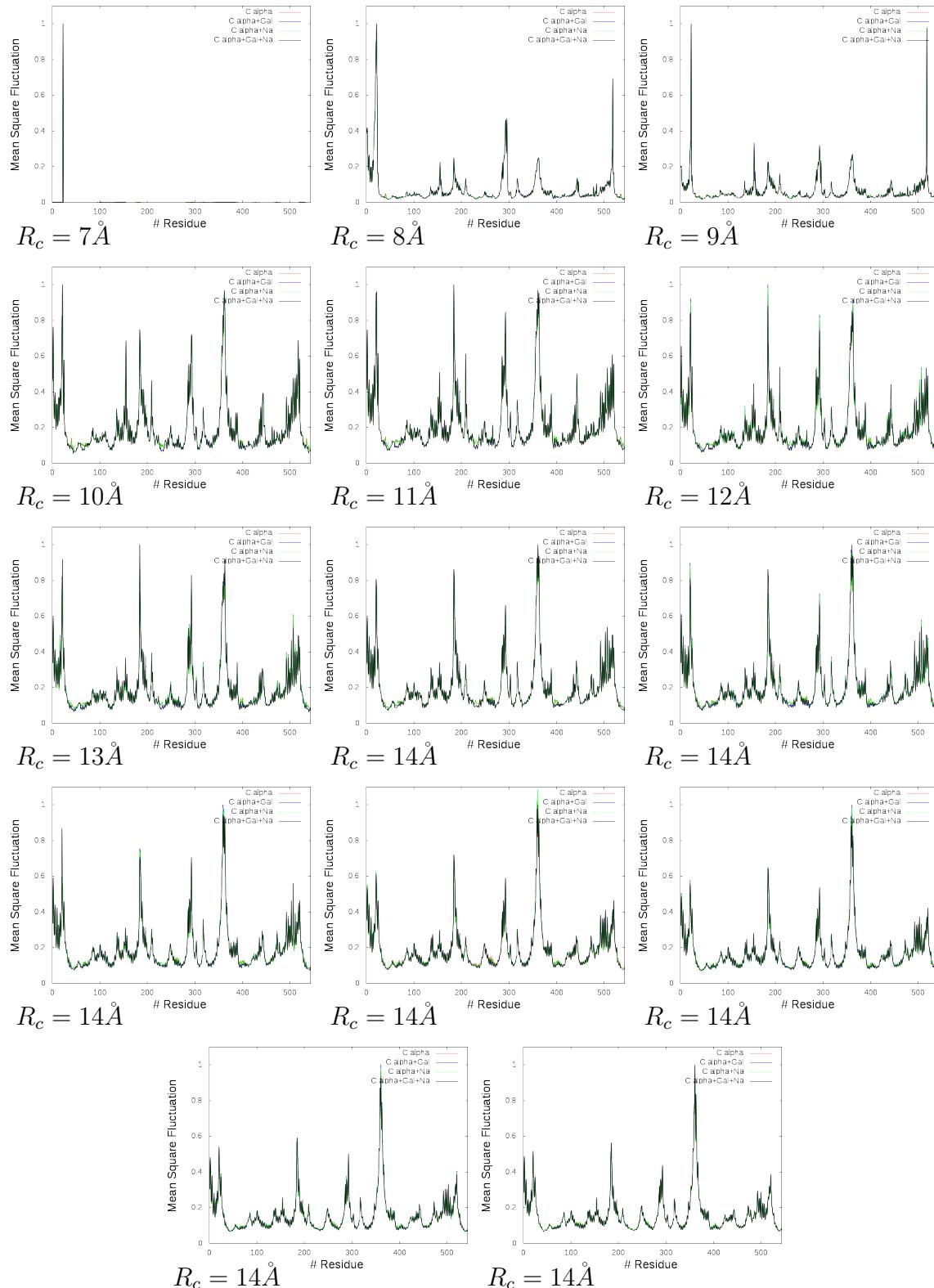


Figura 3-2.: Fluctuaciones ms normalizadas en función del número de residuo entre $7\text{\AA} \leq R_c \leq 20\text{\AA}$ usando los primeros 100 modos. Los diferentes colores indican si la simulación fue realizada sin el ion, el sustrato, con alguno de ellos o ambos.

Apéndice A

Anexo 1: Matriz de Masa del sistema

En tres dimensiones, como es nuestro caso real, la transformación (2-12) dependerá del modelo escogido. Exceptuando el modelo de redes gaussianas (Gaussian Network Model que por sus siglas en inglés es GNM) la transformación de coordenadas va de \mathbf{r}_i posiciones con $i = 1, 2, \dots, N$ ($3N$ coordenadas) a q_j coordenadas $j = 1, 2, \dots, 3N$:

$$\mathbf{r}_i \longrightarrow q_j$$

Con

$$i = 1, 2, \dots, N \quad j = 1, 2, \dots, 3N$$

Por cada componente en cartesianas:

$$\begin{aligned} q_1 &= x_1 - x_{10} & q_4 &= x_2 - x_{20} & \cdots & q_{3i-2} &= x_i - x_{i0} & \cdots & q_{3N-2} &= x_N - x_{N0} \\ q_2 &= y_1 - y_{10} & q_5 &= y_2 - y_{20} & \cdots & q_{3i-1} &= y_i - y_{i0} & \cdots & q_{3N-1} &= y_N - y_{N0} \\ q_3 &= z_1 - z_{10} & q_6 &= z_2 - z_{20} & \cdots & q_{3i} &= z_i - z_{i0} & \cdots & q_{3N} &= z_N - z_{N0} \end{aligned} \tag{A-1}$$

Para esta transformación, (2-12) se convierte en:

$$\begin{aligned} M_{jk} &= \sum_{i=1}^N m_i (\delta_{i,3j-2}\delta_{jk} + \delta_{i,3j-1}\delta_{jk} + \delta_{i,3j}\delta_{jk}) \\ M_{jk} &= m_{3j-2}\delta_{jk} + m_{3j-1}\delta_{jk} + m_{3j}\delta_{jk} \\ M_{jk} &= (m_{3j-2} + m_{3j-1} + m_{3j})\delta_{jk} \end{aligned} \tag{A-2}$$

En (A-2) debe resaltarse que para $j = k = 3N$, el elemento de matriz $M_{3N,3N}$ requiere las masas $m_{3(3N)-2} = m_{9N-2}$, $m_{3(3N)-1} = m_{9N-1}$ y $m_{3(3N)} = m_{9N}$, sin embargo ! no hay $9N$ masas!, el número de masas es el mismo número de nodos: N , entonces, para poder calcular la matriz \mathbf{M} es necesario definir lo siguiente:

$$m_{N+1}, m_{N+2}, \dots, m_{3N} = 0 \tag{A-3}$$

Como a partir de $N + 1$ las masas son nulas, la matriz de masa (que es diagonal) tiene elementos nulos si

$$\begin{aligned}3j - 2 &= N + 1 \\ j &= \frac{N + 3}{3}\end{aligned}$$

Como no siempre N es múltiplo de 3, se escoge el entero menor que este más cerca al valor:

$$j = \left\lfloor \frac{N + 3}{3} \right\rfloor \quad (\text{A-4})$$

Donde $\lfloor \cdot \rfloor$ representa la función piso.

Apéndice B

Anexo:

Bibliografía

- [Amadei et al., 1993] Amadei, A., Linssen, A. B. M., and Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Genetics*, 17(4):412–425.
- [Amber, 2016] Amber, C. (2016). Amber 2016 Reference Manual.
- [Cui and Bahar, 2006] Cui, Q. and Bahar, I. (2006). *Normal Mode Analysis Theory and Applications*.
- [Elber and Karplus, 1987] Elber, R. and Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science*, 235(4786):318–321.
- [Eyal et al., 2015] Eyal, E., Lum, G., and Bahar, I. (2015). The anisotropic network model web server at 2015 (ANM 2.0). *Bioinformatics*, 31(9):1487–1489.
- [Faham et al., 2008] Faham, S., Watanabe, A., Besserer, G. M., Cascio, D., Specht, A., Hirayama, B. A., Wright, E. M., and Abramson, J. (2008). The crystal structure of a sodium galactose transporter reveals mechanistic insights into Na⁺/sugar symport. *Science (New York, N.Y.)*, 321(5890):810–814.
- [Goldstein et al., 2001] Goldstein, H., Poole, C. P., and Safko, J. (2001). *Classical Mechanics*, volume 1. Adison Wesley, Pearson, 3 edition.
- [Gur et al., 2013] Gur, M., Zomot, E., and Bahar, I. (2013). Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions. *The Journal of chemical physics*, 139(12):121912.
- [Hamilton, 2013] Hamilton, K. L. (2013). Robert K. Crane-Na⁺-glucose cotransporter to cure? *Frontiers in Physiology*, 4 MAR(March):1–5.
- [Hayward and de Groot, 2008] Hayward, S. and de Groot, B. L. (2008). Normal modes and essential dynamics. *Methods in Molecular Biology*, 443:89–106.
- [Kuchel et al., 2009] Kuchel, P. W., Easterbrook-Smith, S. B., Gysbers, V., Guss, J. M., and Hancock, D. P. (2009). *Biochemistry*.

- [Lezon et al., 2009] Lezon, T. R., Shrivastava, I. H., Yang, Z., and Bahar, I. (2009). Elastic Network Models For Biomolecular Dynamics: Theory and Application to Membrane Proteins and Viruses. *Handbook on Biological Networks*, pages 129–158.
- [Lovell et al., 2003] Lovell, S. C., Davis, I. W., Adrendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003). Structure validation by C alpha geometry: phi,psi and C beta deviation. *Proteins-Structure Function and Genetics*, 50(August 2002):437–450.
- [Nelson and Cox, 2011] Nelson, D. L. and Cox, M. M. (2011). *Lehninger Principles of Biochemistry*. 4 edition.
- [Pavone et al., 1996] Pavone, V., Gaeta, G., Lombardi, A., Nastri, F., Maglio, O., Isernia, C., and Saviano, M. (1996). Discovering protein secondary structures: Classification and description of isolated α -turns. *Biopolymers*, 38(6):705–721.
- [Rader et al., 2006] Rader, A. J., Chennubhotla, C., Yang, L.-W., and Bahar, I. (2006). The Gaussian Network Model: theory and applications. *Normal Mode Analysis - theory and applications to biological and chemical systems*, 10(20):41–64.
- [Sethna, 2006] Sethna, J. P. (2006). Entropy, order parameters, and complexity. *Statistical Mechanics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY*, pages 105–113.
- [Smith et al., 1996] Smith, L. J., Fiebig, K. M., Schwalbe, H., and Dobson, C. M. (1996). The concept of a random coil. Residual structure in peptides and denatured proteins. *Folding & design*, 1(5):R95–R106.
- [Tirion, 1996] Tirion, M. M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Physical review letters*, 77(9):1905–1908.
- [Turk et al., 2000] Turk, E., Kim, O., Le Coutre, J., Whitelegge, J. P., Eskandari, S., Lam, J. T., Kreman, M., Zampighi, G., Faull, K. F., and Wright, E. M. (2000). Molecular characterization of *Vibrio parahaemolyticus* vSGLT. A model for sodium-coupled sugar cotransporters. *Journal of Biological Chemistry*, 275(33):25711–25716.
- [Zimmermann et al., 2011] Zimmermann, M. T., Kloczkowski, A., and Jernigan, R. L. (2011). MAVENs: motion analysis and visualization of elastic networks and structural ensembles. *BMC bioinformatics*, 12(1):264.