



Análisis de Modos Normales en una Biomolécula

John Erick Cabrera Ramirez

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Física
Bogotá, Colombia
2016

Análisis de 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
2016

(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. Modelos Teóricos	2
1.1. Relevancia	2
1.2. Dinámica de una Biomolécula	2
1.3. Movimientos Locales	3
1.3.1. Análisis de Modos Normales Estándar (NMA estándar)	3
1.4. Movimientos Globales	4
1.4.1. Análisis de Modos Normales (NMA)	4
1.4.2. Modelos de Redes Elásticas (ENM)	5
1.4.3. Análisis por Componentes Principales (PCA)	5
1.5. Descripción Mecánica de NMA	5
1.6. Descripciones de los Movimientos Globales	13
1.6.1. ENMs	13
1.6.2. Análisis por Componentes Principales (PCA)	16
2. Estudios Actuales del Cotransportador vSGLT	18
2.1. Transportadores	18
2.2. Algunas Familias Proteicas y enrollamiento LeuT (LeuT fold)	18
2.3. Co-transportador vSGLT	18
3. Simulación Computacional	19
4. Resultados y Análisis	20
5. Conclusiones y recomendaciones	21
5.1. Conclusiones	21
5.2. Recomendaciones	21

A. Anexo 1: Deducción de las ecuaciones de movimiento	22
B. Anexo:	23
Bibliografía	24

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	$\frac{m_{\text{C}}}{m}$
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	Coefficiente de equilibrio	1	Ec...
L	Longitud	m	DF
L	Longitud del reactor	m	Figura...
m	Masa	kg	DF
\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	DF
P	Presión	Pa	$\frac{\vec{F} \cdot \vec{n}}{A}$
Q	Calor	kJ	1. LT
T	Temperatura	K	DF
t	Tiempo	s	DF
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	$\frac{\text{m}^2}{\text{g}}$	$(w_{\text{F,waf}})(A_{\text{BET}})$
β_i	Grado de formación del componente i	1	$\frac{m_i}{m_{\text{bm},0}}$
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_O	Ángulo superior de avalancha	1	Figura...
θ_U	Ángulo inferior de avalancha	1	Figura...
κ	Velocidad de calentamiento	$\frac{K}{s}$	$\frac{dT}{dt}$
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	Poros
p	Pirolizado
R	Reacción
t	Total
wf	Libre de agua
waf	Libre de agua y de ceniza
0	Estado de referencia

Superíndices

Superíndice	Término
n	Coefficiente x

Abreviaturas

Abreviatura	Término
1. <i>LT</i>	Primera ley de la termodinámica
<i>DF</i>	Dimensión fundamental
<i>RFF</i>	Racimos de fruta fresca

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

Modelos Teóricos

1.1. 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 este en un paso previo al metabolismo y la energética celular al interior de la célula.

1.2. 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, de ahí que

sea necesario realizar *dinámica molecular* (Molecular Dynamics que por sus siglas en inglés es MD) la cual estudia mediante simulaciones computacionales el movimiento de los átomos, 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 se escogen los modelos de potencial.

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 y tiempo, las simulaciones de biomoléculas se pueden hacer analizando los *movimientos locales* y los *movimientos globales*.

1.3. Movimientos Locales

Hace 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].

Como caso particular se pueden tomar 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.
- Interacciones entre pares: Lennard Jones, electrostático.

1.3.1. Análisis de Modos Normales Estándar (NMA estándar)

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 armónicos; ya que los potenciales tienen una forma complicada como la del modelo de Amber, 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 ponderadas.
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.

1.4. Movimientos Globales

Son aquéllas simulaciones en las que se desean conocer los *cambios globales* o el aspecto general que exhibe el movimiento de una biomolécula haciendo simplificaciones, ya sea en los potenciales presentes o en el número de átomos interconectados. Este tipo de simulaciones pueden ser realizadas a un orden de magnitud de los microsegundos, lo cual facilita su uso en computadores personales, al respecto ver [Gur et al., 2013].

Un conjunto de modelos que permite calcular los movimientos globales de una molécula son los *Modelos de Redes Elásticas* (Elastic Network Models o ENM 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) y el análisis por modos normales estándar (Normal Mode Analysis o NMA por sus siglas en inglés).

1.4.1. Análisis de Modos Normales (NMA)

Existe una versión 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. Por ejemplo, en una proteína las unidades monoméricas son los aminoácidos y las posiciones de equilibrio son las de los C- α en la estructura cristalina, las cuales se encuentran en las bases de datos existentes.

1.4.2. Modelos de Redes Elásticas (ENM)

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. Una diferencia con los NMA (estándar y global) es la forma del potencial, en este caso se toman los potenciales de un resorte, los cuales se calculan con la diferencia de distancias y no los desplazamientos q_i del potencial (1-5), ver figura PONER.

Al simplificar el potencial, la biomolécula original se convierte en una red cuyos nodos están sometidos al potencial elástico, ver figura 1-1. Los nodos se consideran como bloques constituyentes de la biomolécula. Aunque el bloques constituyentes suelen ser los residuos de la biomolécula, puede variar ligeramente para dar compatibilidad al modelo con los datos experimentales. Por ejemplo en las proteínas como el inhibidor de la tripsina pancreática bovina (BPTI) [Gur et al., 2013], los constituyentes son los C- α de los aminoácidos.

El modelo es bueno cuando hay una buena estabilidad de los enlaces con respecto a su posición de equilibrio; para distancias mucho mayores que lo que se denomina la distancia de corte R_c no hay potencial alguno. 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.

1.4.3. 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].

1.5. Descripción Mecánica de NMA

Consid

sectionDescripción 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

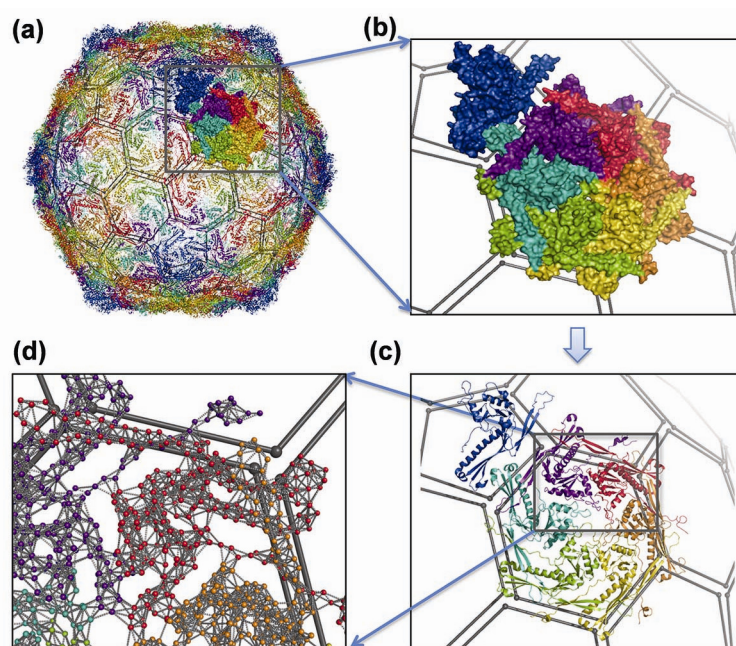


Figura 1-1.: (a) Vista exterior de un cápside vírico HK97 coloreado por cada cadena, todas las proteínas son idénticas. (b) Vista del arreglo proteínico en una cara del cápside. (c) Vista de la estructura secundaria de las proteínas (d) Esquema de cada proteína mostrando cada uno de sus átomos, las aristas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

$\mathbf{q}_0 = \mathbf{0}$ tal como describe la teoría de pequeñas oscilaciones, ver [Goldstein et al., 2001]:

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

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:

$$\left. \frac{\partial V}{\partial q_i} \right|_{\mathbf{q}=\mathbf{0}} = 0 \quad (1-2)$$

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 **1-2**). Ahora suponga que hay un 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]

’ere 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]:

$$V(\mathbf{q}) = V(\mathbf{0}) + \sum_{i=1}^n \left. \frac{\partial V}{\partial q_i} \right|_{\mathbf{q}=\mathbf{0}} q_i + \frac{1}{2} \sum_{ij} \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}=\mathbf{0}} q_i q_j + \dots \quad (1-3)$$

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:

$$\left. \frac{\partial V}{\partial q_i} \right|_{\mathbf{q}=\mathbf{0}} = 0 \quad (1-4)$$

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

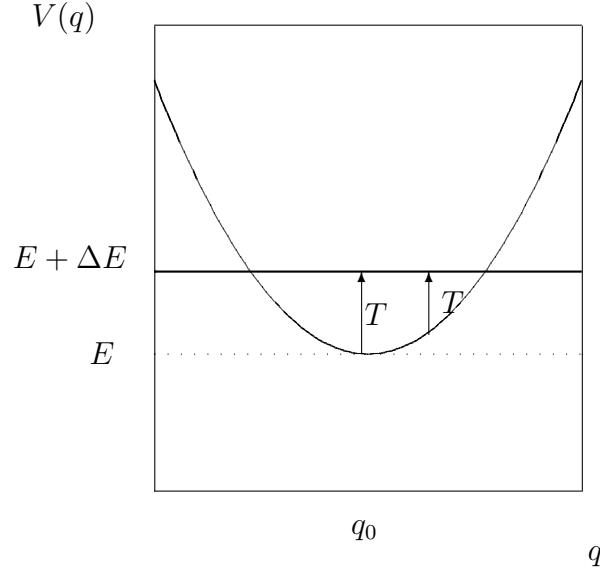


Figura 1-2.: Potencial en función de la posición y energía total E

punto de equilibrio donde los nodos no se mueven, es decir, $E = V_{min}$ (línea punteada de la figura 1-2). Ahora suponga que hay un 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]

Considerando la condición de equilibrio (1-4) y despreciando desplazamientos de orden superior se tiene que:

$$V(\mathbf{q}) = \frac{1}{2} \sum_{i,j=1}^n \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}=\mathbf{0}} q_i q_j \quad (1-5)$$

Donde se identifican las constantes elásticas como:

$$U_{ij} = \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}=\mathbf{0}} \quad (1-6)$$

En términos matriciales el potencial se puede escribir como:

$$V(\mathbf{q}) = \frac{1}{2} \mathbf{q}^t \mathbf{U} \mathbf{q} \quad (1-7)$$

En la ecuación (1-7) \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 enegía cinética de los constituyentes sólo dependerá de los cuadrados de las velocidades generalizadas:

$$T = \frac{1}{2} \mathbf{q}^t \mathbf{M} \mathbf{q} \quad (1-8)$$

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

$$M_{jk} = \sum_{i=1}^N m_i \frac{\partial \mathbf{r}_i}{\partial q_j} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \quad (1-9)$$

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:

$$q_i = x_i - x_{i0} \quad (1-10)$$

En (1-10) 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 (1-9) se convierte en:

$$\begin{aligned} M_{jk} &= \sum_{i=1}^N m_i \delta_{ij} \delta_{ik} \\ M_{jk} &= m_j \delta_{jk} \end{aligned} \quad (1-11)$$

La ecuación (1-11) dice que la matriz \mathbf{M} es una matriz diagonal cuyas componentes son las masas del sistema.

En tres dimensiones, como es nuestro caso real, la transformación 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:

$$\begin{aligned} q_1 &= x_1 - x_{10} & q_4 &= x_2 - x_{20} & \cdots & q_{3i-2} &= x_i - x_{i0} & \cdots & q_{3N-2} &= x_N - x_{N0} \\ q_2 &= y_1 - y_{10} & q_5 &= y_2 - y_{20} & \cdots & q_{3i-1} &= y_i - y_{i0} & \cdots & q_{3N-1} &= y_N - y_{N0} \\ q_3 &= z_1 - z_{10} & q_6 &= z_2 - z_{20} & \cdots & q_{3i} &= z_i - z_{i0} & \cdots & q_{3N} &= z_N - z_{N0} \end{aligned} \quad (1-12)$$

Para esta transformación, (1-9) se convierte en:

$$\begin{aligned}
 M_{jk} &= \sum_{i=1}^N m_i (\delta_{i,3j-2} \delta_{jk} + \delta_{i,3j-1} \delta_{jk} + \delta_{i,3j} \delta_{jk}) \\
 M_{jk} &= m_{3j-2} \delta_{jk} + m_{3j-1} \delta_{jk} + m_{3j} \delta_{jk} \\
 M_{jk} &= (m_{3j-2} + m_{3j-1} + m_{3j}) \delta_{jk}
 \end{aligned} \tag{1-13}$$

En (1-13) 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:

$$m_{N+1}, m_{N+2}, \dots, m_{3N} = 0 \tag{1-14}$$

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:

$$j = \left\lfloor \frac{N + 3}{3} \right\rfloor \tag{1-15}$$

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] o puede ser revisado en el anexo A:

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{U}\mathbf{q} = \mathbf{0} \tag{1-16}$$

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:

$$\mathbf{r} = \mathbf{M}^{1/2} \mathbf{q} \tag{1-17}$$

$$\mathbf{K} = \mathbf{M}^{-1/2} \mathbf{U} \mathbf{M}^{-1/2} \tag{1-18}$$

Reemplazando dichas transformaciones en la ecuación de movimiento se obtiene:

$$\ddot{\mathbf{r}} + \mathbf{K}\mathbf{r} = \mathbf{0} \tag{1-19}$$

El efecto que tienen las coordenadas de masa ponderada sobre la energías 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:

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

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

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 :

$$\mathbf{r} = \mathbf{a} e^{i\omega t} \tag{1-22}$$

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 (1-19) se convierte en un problema de autovalores y autovectores:

$$\begin{aligned}
 \ddot{\mathbf{r}} &= -\mathbf{K} \mathbf{r} \\
 \omega^2 \mathbf{a} &= \mathbf{K} \mathbf{a} \\
 \lambda_k \mathbf{a}_k &= \mathbf{K} \mathbf{a}_k
 \end{aligned} \tag{1-23}$$

En (1-23) 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 (1-23).

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:

$$\mathbf{r} = \sum_k \mathbf{a}_k e^{i\omega_k t} \tag{1-24}$$

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

Ensamble Estadístico

Cuando el sistema (biomolécula) se encuentre a temperaturas fisiológicas, el sistema sólo intercambiará calor con los alrededores (baño térmico). Sin embargo, se ha demostrado que a esta temperatura, la aproximación armónica falla, presentando múltiples mínimos y no uno solo [Hayward and de Groot, 2008], [Elber and Karplus, 1987]; por otro lado, a temperatura fisiológica el ensamble es clásico [Sethna, 2006]. 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:

$$p = \frac{\exp(-E_s/k_B T)}{Z} \quad (1-25)$$

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:

$$Z = \frac{1}{h^n} \int d^n p d^n q \exp[-H(\mathbf{p}, \mathbf{q})/k_B T] \quad (1-26)$$

En (1-26), $n = 3N$ 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 (1-26) 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 (1-26) teniendo en cuenta que \mathbf{M} es diagonal:

$$\begin{aligned} Z &= \frac{1}{h^n} \int d^n p d^n q \exp[-(\dot{\mathbf{q}}^t \mathbf{M} \dot{\mathbf{q}}/2 + \mathbf{q}^t \mathbf{U} \mathbf{q}/2)/k_B T] \\ Z &= \frac{1}{h^n} \int d^n p \exp\left[-\sum_i \frac{p_i^2}{2m_i}\right] \int d^n r \left(\prod_i \sqrt{m_i}\right) \exp\left[-\frac{1}{2} \mathbf{r}^t \mathbf{K} \mathbf{r}/k_B T\right] \\ Z &= \frac{(2\pi)^{n/2}}{h^n} \int d^n r \exp\left[-\frac{1}{2} \mathbf{r}^t \mathbf{K} \mathbf{r}/k_B T\right] \end{aligned} \quad (1-27)$$

El resultado (1-27) es el mostrado en [Lezon et al., 2009] o en [Sethna, 2006]:

$$Z = \frac{(k_B T)^{n/2}}{\hbar^n} \frac{1}{|\mathbf{K}|^{n/2}} \quad (1-28)$$

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:

$$Z = (k_B T)^{n/2} \prod_{i=1}^n \frac{1}{\omega_i} \quad (1-29)$$

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

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

1.6. Descripciones de los Movimientos Globales

1.6.1. ENMs

Modelos de Redes Anisotrópicas (ANM)

Como se ha dicho, la aproximación a segundo orden del potencial, minimización, (1-7) 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:

$$V = \frac{1}{2} \sum_{i,j>i} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \quad (1-31)$$

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 **1-3**, el término $R_{ij} - R_{ij}^0$ es la elongación del resorte, lo cual es diferente a la resta de los desplazamientos, ecuación (1-5); entonces no se está aproximando el potencial ni se requiere una minimización de la energía, como se ha recalado, 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:

$$V_{ANM} = \frac{1}{2} \sum_{i \neq j=1} \gamma_{ij} (R_{ij} - R_{ij}^0)^2 \quad (1-32)$$

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

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

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

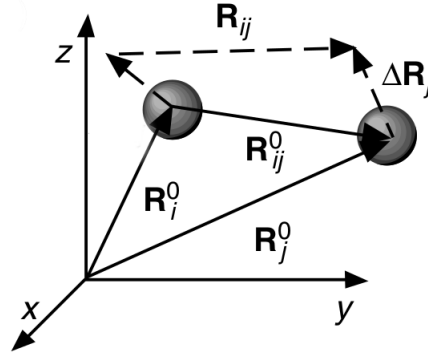


Figura 1-3.: Desplazamientos de los nodos $i - j$ de su posición de equilibrio. Tomado de [Rader et al., 2006].

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 (1-32), 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} :

$$\mathcal{H} = \{H_{ij}\} \text{ con } i, j = 1, \dots, N \quad (1-33)$$

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

$$\mathbf{H}_{ij} = \begin{pmatrix} \partial_{xi}\partial_{xj}V & \partial_{xi}\partial_{yj}V & \partial_{xi}\partial_{zj}V \\ \partial_{yi}\partial_{xj}V & \partial_{yi}\partial_{yj}V & \partial_{yi}\partial_{zj}V \\ \partial_{zi}\partial_{xj}V & \partial_{zi}\partial_{yj}V & \partial_{zi}\partial_{zj}V \end{pmatrix} \quad (1-34)$$

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 $i \neq j$):

$$\frac{\partial V}{\partial y_k} = \sum_j \frac{\gamma_{jk} (R_{kj} - R_{kj}^0) y_{kj}}{R_{kj}} \quad (1-35)$$

$$\left. \frac{\partial^2 V}{\partial x_i \partial y_j} \right|_{R_{ij}=R_{ij}^0} = -\frac{\gamma_{ij} x_{ij} y_{ij}}{R_{ij}^2} \quad (1-36)$$

$$\left. \frac{\partial^2 V}{\partial x_i \partial x_j} \right|_{R_{ij}=R_{ij}^0} = -\frac{\gamma_{ij} x_{ij}^2}{R_{ij}^2} \quad (1-37)$$

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]:

$$\mathbf{H}_{ij} = -\frac{\gamma_{ij}}{R_{ij}^2} \begin{pmatrix} x_{ij}^2 & x_{ij}y_{ij} & x_{ij}z_{ij} \\ y_{ij}x_{ij} & y_{ij}^2 & y_{ij}z_{ij} \\ z_{ij}x_{ij} & z_{ij}y_{ij} & z_{ij}^2 \end{pmatrix} \quad (1-38)$$

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

$$\begin{aligned} \left. \frac{\partial^2 V}{\partial x_i^2} \right|_{R_{ij}=R_{ij}^0} &= -\sum_{j \neq i} \gamma_{ij} \left(\frac{x_{ij}^2}{R_{ij}^2} \right) \\ \left. \frac{\partial^2 V}{\partial x_i^2} \right|_{R_{ij}=R_{ij}^0} &= -\sum_{j \neq i} H_{ij,xx} \end{aligned} \quad (1-39)$$

En (1-39), $H_{ij,xx}$ es el primer elemento de la submatriz \mathbf{H}_{ij} . El resultado (1-39) puede generalizarse para todos los superelementos de la diagonal, ver [Lezon et al., 2009]:

$$\mathbf{H}_{ii} = \sum_{j \neq i} \mathbf{H}_{ij} \quad (1-40)$$

Una vez obtenida la matriz Hessiana, es necesario calcular sus autovalores y autovectores, 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 con respecto al equilibrio son gaussianas e isotrópicas [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:

$$V_{GNM} = \frac{\gamma}{2} \sum_{i,j=1}^N \Gamma_{ij} (\mathbf{R}_{ij} - \mathbf{R}_{ij}^0)^2 \quad (1-41)$$

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:

$$\begin{aligned}
 V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta\mathbf{R}_j - \Delta\mathbf{R}_i)^2 \\
 V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} [(\Delta x_j - \Delta x_i)^2 + (\Delta y_j - \Delta y_i)^2 + (\Delta z_j - \Delta z_i)^2] \\
 V_{GNM} &= \frac{\gamma}{2} \sum_{i \neq j=1}^N \Gamma_{ij} [(\Delta x_{ij})^2 + (\Delta y_{ij})^2 + (\Delta z_{ij})^2]
 \end{aligned} \tag{1-42}$$

Escribiendo (1-42) en forma matricial se tiene:

$$V_{GNM} = \frac{\gamma}{2} (\Delta\mathbf{X}^T \mathbf{\Gamma} \Delta\mathbf{X} + \Delta\mathbf{Y}^T \mathbf{\Gamma} \Delta\mathbf{Y} + \Delta\mathbf{Z}^T \mathbf{\Gamma} \Delta\mathbf{Z}) \tag{1-43}$$

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:

$$\begin{aligned}
 \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta x_{ij})^2 &= \sum_{i \neq j=1}^N \Gamma_{ij} (\Delta x_i^2 + \Delta x_j^2 - 2\Delta x_i \Delta x_j) \\
 &= 2 \sum_{i \neq j=1}^N \Gamma_{ij} \Delta x_i^2 - 2 \sum_{i \neq j=1}^N \Delta x_i \Gamma_{ij} \Delta x_j
 \end{aligned} \tag{1-44}$$

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

1.6.2. 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 aquella que mide la dependencia entre cada una de las variables ($3N$ componentes) y cuyos

² Recurso disponible en http://web.archive.org/web/20070615210550/http://ignm.ccbb.pitt.edu/GNM_Online_Calculation.htm

elementos de la diagonal son las varianzas de dichas variables.

Matriz de covarianza:

$$C = \langle (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)(\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \rangle \quad (1-45)$$

En (1-46), 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 $i < 3N + 1$.

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

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 esten ordenados de mayor a menor y los autovectores ordenados de acuerdo a los autovalores entonces:

$$\Lambda = \mathbf{T}^T C \mathbf{T} \quad (1-46)$$

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

$$C = \mathbf{T} \Lambda \mathbf{T}^T \quad (1-47)$$

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 sólomente con los autovalores más grandes.

Dejando sólomente 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:

$$\begin{aligned} \mathbf{P} &= \mathbf{T}^T (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \\ \mathbf{P} &= (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle) \mathbf{T} \end{aligned} \quad (1-48)$$

O en términos de cada componente principal:

$$\mathbf{p}_i = \mathbf{t}_i^T (\mathbf{x}(\mathbf{t}) - \langle \mathbf{x} \rangle)^T \quad (1-49)$$

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

$$\mathbf{x}(\mathbf{t})^T = \mathbf{T} \mathbf{P} + \langle \mathbf{x} \rangle^T \quad (1-50)$$

Estudios Actuales del Cotransportador vSGLT

El cotransportador de Na^+ /Galactosa presente en la proteobacteria *Vibrio Parparahaemolyticus* denotado como vSGLT, es un simportador presente en la familia de simportadores de sodio-soluto SSS.

2.1. Transportadores

2.2. Algunas Familias Proteicas y enrollamiento LeuT (LeuT fold)

2.3. Co-transportador vSGLT

El co-transportador es similar en la estructura primaria y terciaria a otros transportadores de la familia SLC5 como por ejemplo NIS, SGLT1, SGLT2.

Simulación Computacional

Capítulo 4

Resultados y Análisis

Conclusiones y recomendaciones

5.1. Conclusiones

Las conclusiones constituyen un capítulo independiente y presentan, en forma lógica, los resultados de la tesis o trabajo de investigación. Las conclusiones deben ser la respuesta a los objetivos o propósitos planteados. Se deben titular con la palabra conclusiones en el mismo formato de los títulos de los capítulos anteriores (Títulos primer nivel), precedida por el numeral correspondiente (según la presente plantilla).

5.2. Recomendaciones

Se presentan como una serie de aspectos que se podrían realizar en un futuro para emprender investigaciones similares o fortalecer la investigación realizada. Deben contemplar las perspectivas de la investigación, las cuales son sugerencias, proyecciones o alternativas que se presentan para modificar, cambiar o incidir sobre una situación específica o una problemática encontrada. Pueden presentarse como un texto con características argumentativas, resultado de una reflexión acerca de la tesis o trabajo de investigación.

Apéndice A

Anexo 1: Deducción de las ecuaciones de movimiento

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 Struct. Funct. Genet.*, 17(4):412–425.
- [Amber, 2016] Amber, C. (2016). Amber 2016 Reference Manual.
- [Elber and Karplus, 1987] Elber, R. and Karplus, M. (1987). Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science (80-.)*, 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.
- [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. *J. Chem. Phys.*, 139(12):121912.
- [Hayward and de Groot, 2008] Hayward, S. and de Groot, B. L. (2008). Normal modes and essential dynamics. *Methods Mol. Biol.*, 443:89–106.
- [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. *Handb. Biol. Networks*, pages 129–158.
- [Rader et al., 2006] Rader, A. J., Chennubhotla, C., Yang, L.-W., and Bahar, I. (2006). The Gaussian Network Model: theory and applications. *Norm. Mode Anal. - theory Appl. to Biol. Chem. Syst.*, 10(20):41–64.
- [Sethna, 2006] Sethna, J. P. (2006). Entropy, order parameters, and complexity. *Stat. Mech. Lab. At. Solid State Physics, Cornell Univ. Ithaca, NY*, pages 105–113.
- [Tirion, 1996] Tirion, M. M. (1996). Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. *Phys. Rev. Lett.*, 77(9):1905–1908.