



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Modelación de sistemas biológicos metaestables en la mesoescala**

**Daniel Oswaldo Maldonado Pérez**

Universidad Nacional de Colombia  
Facultad de Minas, Departamento de Procesos y Energía  
Medellín, Colombia  
2022



# **Modelación de sistemas biológicos metaestables en la mesoescala**

**Daniel Oswaldo Maldonado Pérez**

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

**Magister en Ingeniería Química**

Director (a):

Ph.D. Juan Pablo Hernández Ortiz

Línea de Investigación:

Biofísica

Biología computacional

Grupo de Investigación:

Center for Research and Surveillance of Tropical and Infectious Diseases (CRS-TID)

Universidad Nacional de Colombia

Facultad de Minas, Departamento Procesos y Energía

Medellín, Colombia

2022



*Do not quench your inspiration and your imagination; do not become the slave of your model.*

*Vincent Van Gogh,*



# **Declaración de obra original**

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.

**Daniel Oswaldo Maldonado Perez**

Nombre

Fecha 28/05/2022

## **Agradecimientos**

En primera instancia a Minciencias por la financiación parcial del trabajo mediante la convocatoria 807-2018, a mi director el Profesor Juan Pablo Hernández Ortiz, a los profesores John William Brach y Jorge Osorio, a mis compañeros de grupo; Alberto Mario Ceballos, Daniel Alejandro Olaya, Karl Ciuderis, Laura Perez, Juan David Torrenegra, Maria Stella López Carvajal y Jarol Molina por sus constantes retroalimentaciones, enseñanzas y apoyo en las discusiones académicas. De la misma manera, al apoyo de los demás compañeros e integrantes del grupo CRS-TID. Por otra parte, agradecer a los desarrolladores de todos aquellos softwares libres que han permitido la contribución de investigaciones en campos de la ciencia.



# Resumen

## **Modelación de sistemas biológicos metaestables en la mesoescala**

Este trabajo presenta diferentes métodos y plataformas de entendimientos de procesos biológicos como proteínas, complejos supramoleculares y enfermedades transmitidas por vectores como dengue en diferentes escalas. A nivel de proteínas, se emplea métodos atomísticos de dinámica molecular para analizar la evolución de proteínas en agua, para luego medir propiedades fisicoquímicas en el tiempo y así extraer las superficies en un estado metaestable de cada una de ellas. La formación de estructuras supramoleculares a través de agregaciones o segregaciones en sistema biológicos entre diferentes moléculas se da principalmente por interacciones electrostáticas en la escala mesoscópica, por consiguiente, se estudia modelos como Monte Carlo y dinámica molecular permiten entender el comportamiento energético y de interacción de las configuraciones del sistema como las formaciones de hélices. Por último, análisis retrospectivo de rezagos entre diferentes variables climáticas, fenómeno del niño, índices asociados al atlántico permite entender las incidencias de estas en casos de dengue

El cálculo de la constante dieléctrica se puede ver afectada debido a las interacciones entre el agua y las proteínas como posibles efectos de polarización entre ellas. La energía libre de los sistemas helicoidales del sistema específico del disminuye a medida que aumenta la longitud de debye en modelo de Monte Carlo. Las principales variables que inciden o afectan en la aparición de casos del dengue son; el fenómeno del niño, índice del Caribe (CAR) y Noratlántico Tropical (NTA) debido a las fluctuaciones de temperaturas.

**Palabras clave: Proteínas, Superficies, Molecular, Agregación.**

## Abstract

### **Modeling metastable biological systems at the mesoscale**

This work shows different methods and platforms for understanding biological processes such as proteins, supramolecular complexes, and vector-borne diseases at different scales. At the level of proteins, atomistic molecular dynamics methods are used to analyze the evolution of proteins in water, measure physicochemical properties over time and thus extract the surfaces in a metastable state of each of them. The formation of supramolecular structures through aggregations or segregations in biological systems between different molecules is mainly due to electrostatic interactions on the mesoscopic scale; therefore, models such as Monte Carlo and Molecular Dynamics are studied, allowing us to understand the energy and interaction behavior of molecules. System configurations such as helix formations. Finally, retrospective analysis of laps between different climatic variables, El Niño phenomenon, indices associated with the Atlantic allows us to understand the incidences of these in dengue cases.

The calculation of the dielectric constant can be affected due to the interactions between water and proteins as possible polarization effects between them. The free energy of the helical systems of the specific system decreases as the Debye length increases in the Monte Carlo model. The main variables that influence or affect the appearance of dengue cases are the El Niño phenomenon, the Caribbean index (CAR), and the North Atlantic Tropical index (NTA) due to temperature fluctuations.

**Keywords: Proteins, Surfaces, Molecular, Aggregation.**

# Contenido

Resumen.....	9
Lista de figuras.....	13
Lista de tablas.....	15
1. Introducción.....	16
2. Objetivos.....	18
2.1 Objetivo General.....	18
2.2 Objetivos Específicos.....	18
3. Fundamentación de Métodos y Modelos.....	19
3.1 Escalas.....	19
3.2 Modelamiento por homología.....	21
3.3 Dinámica Molecular.....	22
3.3.1 Campo de fuerza.....	24
3.3.2 Modelo SPC.....	25
3.3.3 Termóstato.....	26
3.3.4 Baróstato.....	28
3.4 Dinámica Browniana.....	29
3.5 Interacciones Electroestáticas.....	31
3.5.1 Ecuaciones de Maxwell.....	31
3.5.2 Debye-Huckel.....	32
4. Proteínas de vías del complemento.....	34
4.1 Vías de complemento.....	34
4.2 Factor B.....	36
4.3 C3.....	37
4.4 Metodología.....	39
4.4.1 Proteínas.....	39
4.4.2 Protocolo de Simulación.....	39
4.5 Resultados y análisis.....	40
4.6 Conclusiones.....	48
5. Agregación y segregación de partículas biológicas.....	49
5.1 RNPs.....	49
5.2 Metodología.....	52
5.3 Resultados y análisis.....	53
5.4 Conclusiones.....	65
6. Análisis Retrospectivo Dengue.....	67
6.1 Áreas de Estudio.....	68
6.2 Datos de Dengue.....	69
6.3 Datos climáticos.....	69
6.4 Datos del Niño.....	69
6.5 Metodología.....	70

6.6	Resultados y análisis .....	70
6.7	Conclusiones .....	78
7.	Conclusiones y recomendaciones.....	79
7.1	Conclusiones .....	79
	Bibliografía.....	80

## Lista de figuras

FIGURA 1. Escalas de espaciales y temporales en sistemas biológicos. ....	20
FIGURA 2. Organización de autoensamblaje en diferentes escalas espaciales. ....	21
FIGURA 3. Metodología de simulación de dinámica molecular a través de Gromacs .....	23
FIGURA 4. Representación esquemática de las tres vías del complemento. ....	34
FIGURA 5. Representación tridimensional de la proteína Factor B en cartoon.....	37
FIGURA 6. Enlace tioéster de la proteína C3 formando en los enlaces con los residuos de la cisteína y la glutamina. ....	38
FIGURA. 7. Representación de la proteína C3 con la cadena A, y la cadena B. ....	39
FIGURA 8. Propiedades termodinámicas del sistema en el tiempo. a) Densidad, b) Presión, c) Temperatura y d) Energía de minimización del sistema. ....	40
FIGURA 9. a) Desviación cuadrática media de la raíz (RMSD) de backbone de la proteína del factor b en el tiempo. b) Evolución temporal del Radio de giro (Rg).....	41
FIGURA 10. a) Momento total dipolar en el tiempo; momento en x (línea azul), momento dipolar en y (línea roja), z y momento dipolar total (línea negra). b) Evolución temporal de la constante dieléctrica en el tiempo.....	41
FIGURA 11. Representación de la Superficie de Van der Waals; a) Factor B cristalizada (color rojo) y Factor B obtenido de la dinámica Molecular (color azul). b) Representación de la superficie de Van der Waals de la proteína Factor B obtenida desde la dinámica molecular con la configuración cartoon. ....	42
FIGURA 12. Potencial Electroestático: Factor Ba (línea roja), Factor Bb (línea negra) y Factor B.....	43
FIGURA 13. Propiedades termodinámicas del sistema en el tiempo. a) Densidad, b) Presión, c) Temperatura y d) Energía de minimización del sistema.....	44
FIGURA 14. a) Desviación cuadrática media de la raíz (RMSD) de backbone de la proteína del factor b en el tiempo. b) Evolución temporal del Radio de giro (Rg).....	45
FIGURA 15. a) Momento total dipolar en el tiempo; momento en x (línea azul), momento dipolar en y (línea roja), z y momento dipolar total (línea negra). b) Evolución temporal de la constante dieléctrica en el tiempo.....	45
FIGURA 16. Representación de la Superficie de Van der Waals; a) C3 cristalizada (color rojo) y C3 obtenido de la dinámica Molecular (color azul). B) Representación de la superficie de Van der Waals de la proteína C3 obtenida desde la dinámica molecular con la configuración cartoon. ....	46

FIGURA 17. Potencial Electroestático: C3 (línea roja) y C3b (línea negra) .....	47
FIGURA. 18. Visualización de RNPs a través de fluorescencia en microscopía de luz ...	50
FIGURA 19. Imagen experimental de RNPs germinales de Zebrafish formando supra-estructuras helicoidales .....	51
FIGURA 20. Representación esquemática de los sitios de hibridación entre las secuencias de RNA (Dazl y Bruno) y la energía libre de interacción.....	54
FIGURA 21. Caracterización de regiones de residuos desordenados y ordenados de la proteína BUC. Los aminoácidos de color negro son configuraciones ordenadas. Los aminoácidos de color rojos tienen configuración desordenada. ....	56
FIGURA 22. Caracterización de regiones de residuos desordenados y ordenados de la proteína Buc2L. Los aminoácidos de color negro son configuraciones ordenadas. Los aminoácidos de color rojo tienen configuración desordenada .....	57
FIGURA 23. Modelo tridimensional de las proteínas: a) Buc y b) Buc2 .....	57
FIGURA 24. Representación esquemática de una partícula de RNP.....	59
FIGURA 25. Energía libre en keV con cargas positivas y negativas distribuidas en proporciones diferentes 50:50, 60:40, 70:30, 80:20 y 90:10 respectivamente, variando consigo la longitud de Debye de en 1, 10, 50, 100, 250, 500, 1000 nm. ....	60
FIGURA 26. Representación esquemática de la formación helicoidal entre RNPs acorde a la distribución de carga positiva y negativa; 50:50, 60:40, 70:30, 80:20 y 90:10 respectivamente.....	61
62	
FIGURA 27. El diagrama izquierdo (a) representa el mapa de calor asociado a la energía de acuerdo a longitud de Debye (eje y) respecto a la proporción en cargas (eje x). En el diagrama de la derecha (b) representa el mapa de calor asociado las distancia promedios de las configuraciones de hélices respecto a cada bead; En el eje x está la proporción de cargas y en el eje y la longitud de Debye. ....	62
FIGURA 28. El diagrama izquierdo (a) representa el mapa de calor asociado a la energía de acuerdo a longitud de Debye (eje y) respecto a la proporción en cargas (eje x). En el diagrama de la derecha (b) representa el mapa de calor asociado las distancia promedios de las configuraciones de hélices respecto a cada bead; En el eje x está la proporción de cargas y en el eje y la longitud de Debye. ....	62
FIGURA 29. Función de distribución radial (rdf) y desplazamiento cuadrático medio (msd) para para un sistema de 8 RNPs con diferentes cargas.....	64
FIGURA 30. Ciclo de vida del mosquito <i>Aedes aegypti</i> .....	68
FIGURA 31. Esquema de las cuatro regiones del niño en océano pacifico[55]. ....	70
FIGURA 32. Casos de dengue semanales a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés: Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio. ....	71

FIGURA 33. Temperatura media, rango de temperatura máxima y mínima semanal a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio. ....72

FIGURA 34. Humedad relativa, rango de humedad relativa máxima y mínima semanal a través de los años desde el 2008 hasta comienzo del 2020 en cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio. ....73

FIGURA 35. Precipitaciones semanales a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio.....74

FIGURA 36. Análisis de correlación entre las variables climáticas y los casos de dengue para tres ciudades: Villavicencio, Cucúta y Medellín. ....75

## Lista de tablas

Tabla 1. Principales campos de fuerzas implementado en sistemas biológico, químicos y físicos en simulaciones de dinámica molecular. ....24

Tabla 2. Propiedades y parámetros físicos de los modelos más implementados para el agua. ....26

Tabla 3. Distribución de las pares de secuencias analizadas entre ARN de germoplasma animal y vegetal. ....53

Tabla 4. Intervalos de hibridación entre las secuencias de ARN germoplasma animal y vegetal con su correspondiente energía de interacción en Kcal/mol.....55

Tabla 5. Rezagos entre las variables climáticas, índices del NIÑO, NTA y CAR en los casos de dengue en las diferentes ciudades. ....77

# 1. Introducción

En los últimos años, el menester y la prioridad de entender los procesos biológicos, como sus sistemas dinámicos, han producido diferentes investigaciones a nivel *in silico*, en ámbitos como; proteómica, metabolómica, genómica, la inmunología, división y diferenciación celular, entre otros [1]. Debido a la complejidad de los fenómenos biológicos y a la dificultad de comprender el comportamiento de las interacciones fisicoquímicas entre biomoléculas, complejos supramoleculares o las mismas células. De tal manera que se han desarrollado técnicas de simulación computacional para estudiar dichos sistemas,

Los modelos fisicoquímicos han permitido entender y explicar a través de conectividades entre diferentes escalas el funcionamiento de algunos sistemas complejos. Sin embargo, existen fenómenos de auto-ensamble o agregación que no se han conseguido dilucidar y al ser estados metaestables se hace difícil entender los mecanismos por el cual presentan configuraciones estructurales o energéticas específicas. A menudo, se realizan experimentos *in vivo* para esclarecer formaciones de complejos supramoleculares, activación de reacciones químicas, agregaciones moleculares o incluso dispersiones de vectores con el objetivo de determinar y delinear mecanismos funcionalidades o de acción. Desafortunadamente, los sistemas biológicos son altamente complejos y se componen de múltiples redes funcionales que actúan en diversas escalas temporales y espaciales[1]–[4]. La formación de estructuras supramoleculares, agregaciones o segregaciones ocurre por el auto-ensamble entre diferentes moléculas, un caso particular entres proteínas y ARN mediante interacciones moleculares desde la escala atomística hasta la escala mesoscópica en equilibrio termodinámico en fases fluidas. Sin embargo, se desconocen los mecanismos fisicoquímicos que permiten dicha formación, motivando el modelamiento multi-escala. No obstante, dichos procesos, modelos y técnicas computacionales han generado el interés de la ingeniera molecular para la creación de metodologías aplicadas a creaciones de nuevos materiales, farmacéutica, control de vectores, órganos sintéticos, bioconductores, entre otros [1], [5]–[7].

La falta de una metodología teórica para explicar fenómenos de agregación como los de RNPs a través de modelos físico-químicos y la necesidad de hallar mecanismos que permitan entender la formación de estructuras en estados metaestables a través de



modelamiento y simulación multi-escala, asociados incluso a interacciones electrostáticas y/o hidrodinámicas, evidencian la estrategia de crear diseños racionales de modelos mínimos que permitan explicar diferentes comportamientos de interés en sistemas biológicos y a su vez que puedan ser usados en otras áreas como modelos fluctuantes fuera del equilibrio [1][8].

Actualmente, algunos estudios proponen modelos con electrostática para el plegamiento de proteínas o de complejos de ligandos. Para el cálculo de energía libre y el análisis de posibles efectos en el docking molecular, como la interacción de pequeños medicamentos con proteínas para el uso en diseño de biocatálisis y proteínas terapéuticas [9], [10].

La implementación de modelos mesoscópicos permite estudiar las interacciones entre proteínas, la formación de auto-organizaciones de estructuras en solución, como las condiciones o factores que permiten las agregaciones entre ellas a través de del cálculo de la energía libre [11]. Por otro lado, las interacciones electrostáticas son estudiadas y calculadas a través de técnicas computacionales como las simulaciones de Monte Carlo. Un ejemplo de estas, son proteínas en soluciones iónicas cuyo objetivo es entender el comportamiento de dichos sistemas y analizar regulación de las interacciones con el medio.

Estudios recientes, postulan modelos basados en dinámica browniana para estudiar interacciones en sistemas iónicos y poliméricos a través de cargas electrostáticas. Por otro lado, estudios de suspensiones activas con hidrodinámica permiten modelar sistemas autopropulsados con efectos quimiotaxis en la movilidad de interacciones entre partículas. Dichos modelos pueden generar modelos dinámicos asociados a transmisiones espaciales de vectores e incluso en sistemas proteicos o supramoleculares [12], [13].

## 2. Objetivos

### 2.1 Objetivo General

Modelar sistemas biológicos meta-estables en la escala mesoscópica para medir las dinámicas de auto-ensamblaje molecular, estados de agregación e interacciones superficiales, a través de simulaciones moleculares, estocásticas y brownianas.

### 2.2 Objetivos Específicos

1. Establecer superficies de Van der Waals de proteínas de interés a través de la implementación de Dinámica Molecular, para medir interacciones superficiales que permitan establecer el diseño de plataformas de simulaciones multi-protéicas.
2. Modelar la agregación de partículas de ribonucleopartículas (RNP) mediante el uso de dinámica browniana e interacciones electrostáticas, para caracterizar conformaciones estructurales y entender las dinámicas de agregación.
3. Construir un modelo de análisis estadístico multivariado para análisis de retrospectivos de casos de dengue o enfermedades virales.

El abordaje de los objetivos se realizó a través de capítulos o secciones que permiten evidenciar el desarrollo y los resultados obtenidos. Por consiguiente; a cada objetivo le corresponde un capítulo de la siguiente manera:

Objetivo 1: Proteínas de la vía del complemento

Objetivo 2: Agregación y segregación de partículas biológicas

Objetivo 3: Análisis retrospectivo de dengue.

## **3. Fundamentación de Métodos y Modelos**

En esta sección se hace una pequeña explicación de manera general de algunos métodos y modelos implementados para el cumplimiento de los objetivos, el cual permiten entender conceptualmente los abordajes teóricos que se implementan en las metodologías y las herramientas utilizada en los capítulos para el desarrollo de las respectivas metodologías acorde a objetivos.

### **3.1 Escalas**

Los sistemas biológicos son inherentemente complejos en naturaleza; se componen de múltiples redes funcionales que operan a través de diversos dominios espacios-temporales. Existen diferentes tipos de escala en los sistemas biológicos, clasificadas según su nivel de organización. La cual, inicia desde diferentes precursores (aminoácidos, ácidos nucleicos azucres y ácidos grasos), formación de biomoléculas (proteínas, ADN, RNA, lípidos, polisacárido), supramoleculas, organelas, células, tejidos, órganos hasta la formación del mismo organismo [1]–[4], [6].

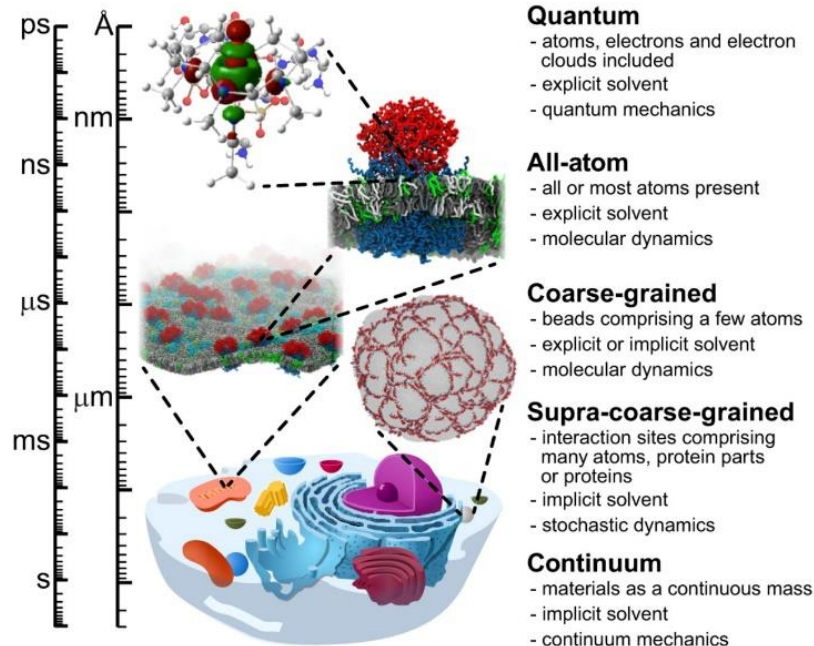


FIGURA 1. Escalas de espaciales y temporales en sistemas biológicos[1], [6].

La dinámica ordenada en sistemas biológicos emerge a partir de una combinación de complejas interacciones estereo-específicas (auto-ensamblaje), electrostáticas, puentes de hidrogeno, hidrofílicas e hidrofóbicas que se producen en el equilibrio termodinámico, y/o en estados metaestables. Mediante interacciones dinámicas entre moléculas que disipan energía para su auto-organización [1]–[4], [6].

El auto-ensamblaje molecular es el ensamblaje de moléculas de estructuras tridimensionales mediado por enlaces débiles y no covalentes; enlaces de hidrógeno, interacciones electrostáticas, interacciones hidrofóbicas e hidrofílicas. Aunque, estas fuerzas son consideradas débiles, sus interacciones colectivas pueden producir estructuras químicamente estables. Un ejemplo de esto son las moléculas biológicas como péptidos, proteínas, ácidos nucleicos, lípidos y otros componentes, que se autoensamblan de manera versátil y funcional. El entendimiento y la comprensión de este fenómeno ha inspirado el desarrollo de materiales nanoestructurados, productos farmacéuticos, biotecnológicos, entre otros.

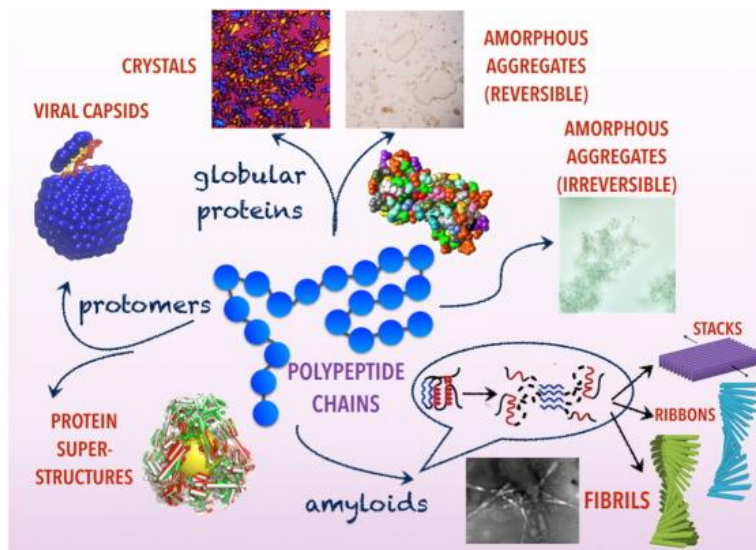


FIGURA 2. Organización de autoensamblaje en diferentes escalas espaciales [6].

### 3.2 Modelamiento por homología

El modelado de homología predice la estructura tridimensional de una proteína respecto a una de referencia, debido a su alta similitud e identidad de las secuencias permite obtener una estructura similar entre ellas. La metodología consiste en cuatro pasos: identificación del target, alineamiento de secuencias, construcción del modelo y optimización del modelo [14], [15].

Las herramientas más utilizadas para el modelamiento o predicción de proteínas tridimensionales son: MODELLER, I-TASSER, SWISS-MODEL y PRIMO [15], [16]. No obstante, existen otros softwares especializados o servidores que permiten cumplir la misma función. MODELLER es un programa que permite modelar las estructuras acordes a las restricciones espaciales de ensamblaje cómo distancias, ángulos, ángulos diedros, pares de ángulos diedros y algunas otras características espaciales definidas por átomos o pseudo-átomos[17][18]. Inicialmente, se proporciona el archivo de alineamiento entre las secuencias, las restricciones comparativas de la estructura 3D de referencia, posteriormente MODELLER calcula automáticamente un modelo que satisface las

restricciones basadas en experimentos de resonancias magnéticas nucleares RMN, reglas de empaquetado de estructuras secundarias, experimentos de reticulación, espectroscopia de fluorescencia, reconstrucción de imágenes en microscopía electrónica, mutagénesis dirigida al sitio, potenciales de fuerza media de residuo y átomo-átomo, entre otras[19], [20]. Finalmente, el modelo es optimizados a través de una función de densidad de probabilidad molecular, métodos gradientes conjugados y dinámica molecular[18]–[20].

### 3.3 Dinámica Molecular

La simulación de dinámica molecular es una técnica computacional para calcular las propiedades de equilibrio y transporte de un sistema clásico de  $N$ -partículas, resolviendo las ecuaciones de movimiento de Newton hasta que las propiedades del sistema ya no cambian con el tiempo. El algoritmo de Dinámica Molecular consiste: Inicialmente en la inicialización del sistema (posiciones iniciales y velocidades), procediendo con el cálculo de la fuerza en cada partícula, y resolviendo las ecuaciones de movimiento de Newton a través integración numérica. A continuación, se describen las ecuaciones a resolver:

$$\frac{dr_i(t)}{dt} = v_i(t)$$

$$\frac{dv_i(t)}{dt} = \frac{F_i(t)}{m_i}$$

Donde  $i = 1, 2, 3 \dots N$  es número de átomos,  $F_i$  es la fuerza ejercida sobre el átomo,  $r_i$  es la posición,  $v_i$  la velocidad,  $m_i$  la masa y  $t$  el tiempo. La fuerza neta que ejerce las partículas entre ellas misma se calcula a través de la derivada de la energía potencial  $F_i = -\nabla_{r_i} V$  por lo tanto:

$$F_i(t) = \frac{\partial V(r_1(t), r_2(t), \dots, r_N(t))}{\partial r_i(t)}$$

Donde  $F_i$  es la fuerza ejercida sobre el átomo,  $r_i$  es la posición,  $V$  la energía potencial,  $m_x$  la masa y  $t$  el tiempo. La integración numérica de las ecuaciones de Newton es calculada a través de diferentes métodos de integración. Entre los más usados son los algoritmos de Verlet y de leap-frog[21], este último, se describe a continuación:

$$v_i \left( t + \frac{\Delta t}{2} \right) = v_i \left( t - \frac{\Delta t}{2} \right) + \frac{F_i(t)}{m_i} \Delta t$$

$$r_i(t + \Delta t) = r_i(t) + v_i \left( t + \frac{\Delta t}{2} \right) \Delta t$$

En donde las variables y parámetros son:  $i = 1, 2, 3 \dots N$  el número de átomos,  $F_i$  la fuerza ejercida sobre el átomo,  $m_i$  la masa,  $v_i$  la velocidad,  $t$  el tiempo,  $r_i$  posición y el  $\Delta t$  paso de tiempo.

Para este trabajo, vamos a usar el software de simulación Gromacs [22] para realizar la dinámica molecular y la minimización de la energía de las proteínas de interés, cuyo objetivo es comprender y predecir macroscópicamente los sistemas biofísicos a través de un modelo atomístico. Por consiguiente, implementaremos el siguiente esquema de simulación:

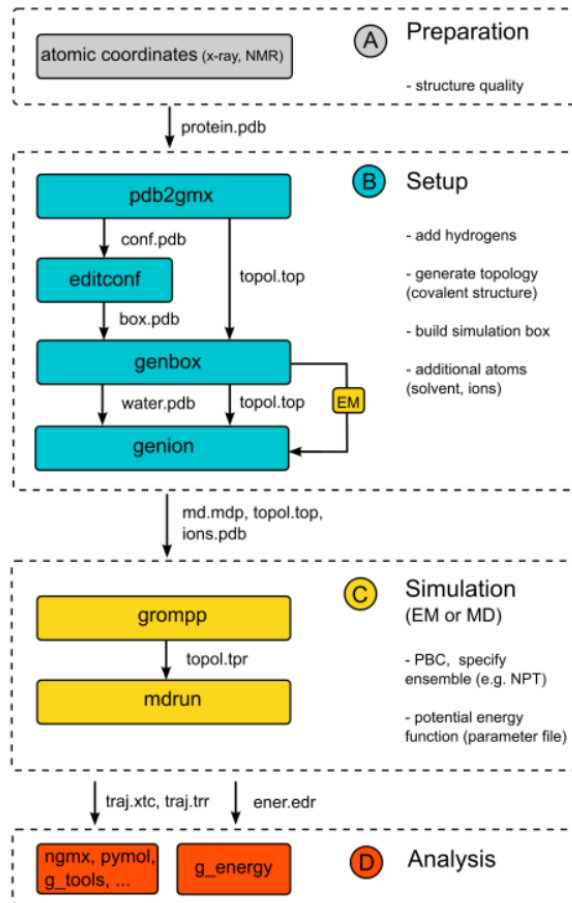


FIGURA 3. Metodología de simulación de dinámica molecular a través de Gromacs [23].

### 3.3.1 Campo de fuerza

El campo de fuerza describe la energía potencial intra e intermolecular de interacciones entre partículas, átomos y moléculas de manera ideal cuya derivación proviene de cálculos precisos de la mecánica cuántica y trabajos experimentales de acuerdo a las propiedades del sistema, obteniendo consigo parámetros que permiten describir la energía mediante el uso de simulaciones por computadora a diferentes escalas de tiempo y espacio[24] . Existen diferentes tipos de campos de fuerza reportados en la literatura para sistemas biológicos y bioquímicos como; proteínas, coloides, lípidos, entre otros en la siguiente tabla se reporta alguno de ellos [24].

Campo de fuerza	Aplicación
AMBER 96	Péptidos, pequeñas moléculas orgánicas
AMBER 99	Péptidos, pequeñas moléculas orgánicas
OPLS	Péptidos, pequeñas moléculas orgánicas
GROMOS 53a6	Proteínas, nucleótidos, azúcares
GROMOS 53a5	Proteínas, nucleótidos, pequeñas moléculas
CHARMM 22	Proteína
CHARMM 27	DNA, RNA, y lípidos
CHARMM 36	Proteínas, DNA, RNA, lípidos
COMPASS	Pequeñas moléculas, polímeros
MARTINI	Coarse-grained, proteínas, lípidos, polímeros

Tabla 1. Principales campos de fuerzas implementado en sistemas biológico, químicos y físicos en simulaciones de dinámica molecular.

Para nuestro sistema particular de proteínas el campo de fuerza implementado es CHARMM36 representado como la suma de contribuciones de las fuerzas de interacción entre átomos:

$$\begin{aligned}
 U = & \sum_{bonds} k_b(b - b_o)^2 + \sum_{angles} k_\theta(\theta - \theta_o)^2 + \sum_{dihedrales} k_\phi(1 + \cos(n\phi - \delta))^2 \\
 & + \sum_{impropers} k_\omega(\omega - \omega_o)^2 + \sum_{Urey-Brandley} k_\mu(\mu - \mu_o)^2
 \end{aligned}$$



$$+ \sum_{nonbondend} \left( \varepsilon \left[ \left( \frac{r}{r_{ij}} \right)^{12} - \left( \frac{r}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\varepsilon r_{ij}} \right)$$

El primer término de la función de energía describe los enlaces moleculares donde  $k_b$  es la constante de fuerza de enlace y  $b - b_0$  es la distancia desde el equilibrio a la que se ha movido el átomo. El segundo término de la ecuación describe los ángulos de enlace, donde  $k_\theta$  es la constante de fuerza del ángulo y  $\theta_0$  es el ángulo de equilibrio entre 3 átomos enlazados. El tercer término es para los diedros (ángulos de torsión) donde  $k_\phi$  es la constante de fuerza diedro,  $n$  es la multiplicidad de la función,  $\phi$  es el ángulo diedro y  $\delta$  es la fase cambio. El cuarto término explica los impropios, la flexión fuera del plano, donde  $k_\omega$  es la constante de fuerza y  $\omega - \omega_0$  es el ángulo fuera del plano. El quinto término de ecuación comprende el componente de Urey-Bradley (contabilidad de términos cruzados para la flexión de ángulos usando interacciones 1,3 no unidas), donde  $k_\mu$  es la constante de fuerza respectiva y  $\mu$  es la distancia entre los átomos 1,3 en el potencial armónico. Las interacciones no enlazadas entre pares de átomos ( $i, j$ ) están representadas por los dos últimos términos. La energía de van Der Waals (VDW) se calcula con un potencial de Lennard-Jones estándar de 12-6 y la energía electrostática con un potencial el Coulomb.

### 3.3.2 Modelo SPC

La capacidad disolvente o solvente del agua permite mediar funciones importantes en reacciones químicas y biológicas. La molécula de agua se caracteriza por sus parámetros geométricos como longitudes de enlace y ángulos que podría mantenerse rígido o flexible durante la simulación presentando posibles efectos de polarización. Existen diferentes modelos parametrizados con cargas atómicas parciales de oxígeno e hidrógeno asociados a interacciones de dispersión y repulsión semejantes al potencial de Lennard-Jones[25], [26]. Entre los modelos más usados son los de tres y cuatro sitios o puntos para sistemas biológicos[26].

Los modelos de agua de tres sitios como SCP o TIP3P se definen con un átomo de oxígeno y dos de hidrogeno con cargas parciales para cada átomo, pero el oxígeno tiene

interacciones de Lennar-Jones con los otros átomos, presentado un ángulo rígido tetrahedral de  $104.5^\circ$  excepto para SCP que es de  $109.47^\circ$ [25]. A su vez, hay configuraciones actualizadas o mejoradas como SCP/E con corrección de polarización promedio a la función de energía potencial, obteniendo una mejor densidad y constante de difusión que el modelo SPC. Por otro lado, los modelos de cuatro sitios como TIP4P es un potencial de interacción plano rígido de cuatro sitios para el agua donde la carga negativa se desplaza del átomo de oxígeno a un punto 0.15A a lo largo de la bisectriz entre los átomos de hidrógeno[25], [26].

	$\sigma(\text{A})$	$\varepsilon(\text{kJ/mol})$	$l_1(\text{A})$	$l_2$	$q_h(e)$	$\frac{q_o}{q_1}(e)$	$\theta$	$\phi$	Moment o Dipolar	Constant e Dieléctric a
SPC	3.166	0.650	1		0.41	-0.82	109.47		2.27	65
SPC/E	3.166	0.650	1		0.423	-0.847	109.47		2.35	71
TIP3P	3.150	0.636	0.957		0.417	-0.834	104.52		2.35	82
TIP4P	3.153	0.648	0.957	0.15	0.520	-1.04	104.52	52.26	2.18	53
TIP4P-Ew	3.164	0.680	0.957	0.125	0.524	-1.04	104.52	52.26	2.32	78.4
TIP4P-FQ	3.153	0.648	0.957	0.15	0.630	-1.26	104.52	52.26	2.64	79

Tabla 2. Propiedades y parámetros físicos de los modelos más implementados para el agua.

### 3.3.3 Termóstato

En dinámica molecular las fluctuaciones de energía total con ensambles canónicos son controladas a través de algoritmos conocidos como termostato que permiten tener la temperatura constante del sistema. Los cuales otorgan restricciones a ecuaciones de movimiento mediante grados de libertad (parámetros) clasificándolos de acuerdo a su accionar de todo el conjunto o subconjunto de partículas del sistema.

Los termostatos más implementados en las simulaciones de MD son: Andersen, Nosé Hoover y Berendsen. No obstante, existen otros tipos de termostatos modificados como V-resacle provenientes de los nombrando anteriormente[24]. Para este trabajo el termostato utilizado fue Berendsen, por lo tanto, lo describiremos a continuación.

## Termostato de Berendsen

El termostato Berendsen escala la velocidad de las partículas mediante un baño termostático externo que está a una temperatura de referencia, removiendo o aplicando calor al sistema en cada paso de integración del sistema. Donde la temperatura y la energía cinética se relaciona de la siguiente manera:

$$\frac{3}{2}Nk_B T = \frac{1}{2} \sum_i m_i v_i^2$$

Donde  $k_B$  es la constante de Boltzmann,  $T$  la temperatura,  $v_i$  la velocidad,  $m_i$  la masa y  $t$  el tiempo. Por consiguiente, el cambio de la temperatura del sistema se da a través de un factor de escalamiento  $\lambda$  en la velocidad:

$$\Delta T = \frac{2}{3Nk_B} \sum_{i=1}^N \frac{m_i (\lambda v_i)^2}{2} - \frac{2}{3Nk_B} \sum_{i=1}^N \frac{m_i (v_i)^2}{2}$$

$$\Delta T = (\lambda^2 - 1)T$$

Obteniendo

$$\lambda = \sqrt{1 + \frac{\Delta T}{T}}$$

Donde  $k_B$  es la constante de Boltzmann,  $T$  la temperatura,  $v_i$  la velocidad,  $m_i$  la masa y  $t$  el tiempo,  $\Delta T$  el cambio de la temperatura. No obstante, al escalarse la velocidad en cada paso de integración y al ser la diferencia de temperatura proporcional al tiempo, el factor de escalamiento queda definido de la siguiente manera:

$$\lambda = \left(1 + \frac{\Delta t}{\tau} \left(\frac{T}{T_0} - 1\right)\right)^{\frac{1}{2}}$$

La constante de tiempo  $\tau$  del termostato que indica el acoplamiento al sistema y el reservorio de calor,  $\Delta t$  el paso de tiempo de integración,  $T$  la temperatura y  $T_0$  la temperatura inicial.

### 3.3.4 Baróstato

En algunas simulaciones de dinámica molecular es necesario mantener la presión constante, por consiguiente, hay diferentes algoritmos que permiten tener un valor deseado concertando el volumen del sistema. Entre los baróstatos más utilizados están: Andersen, Berendsen, Nosé Hoover, MMTK y Parinello-Rahman [22]. A continuación, describieron el Parinello-Rahman, ya que fue utilizado como baróstato en la simulación de Gromacs.

#### Baróstato Parinello-Rahman

En esta sección describiremos el termostato Parinello-Rahman, el cual es un algoritmo basado el baróstato de Andersen. No obstante, los vectores de la caja siguen la siguiente ecuación de movimiento.

$$\frac{db^2}{dt^2} = VW^{-1}b'^{-1}(P - P_{ref})$$

Donde  $b$  es la matriz que contiene los vectores de la caja,  $V$  es el volumen de la caja,  $W$  es la matriz de parámetros que determina la fuerza de acoplamiento,  $P$  es la presión actual y  $P_{ref}$  es la presión de referencia del sistema. Dado que la ecuación de movimiento de las partículas cambia en el sistema, la función de Hamiltoniano está dada por:

$$E_{pot} + E_{kin} + \sum_i P_{ii}V + \sum_{i,j} \frac{1}{2} W_{ij} \left( \frac{db_{ij}}{dt} \right)^2$$

Donde  $E_{pot}$  es la energía potencial y  $E_{kin}$  es la energía cinética. Por lo tanto, la ecuación de movimiento obtenida es:

$$\frac{d^2 r_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - M \frac{dr_i}{dt}$$

$$M = b^{-1} \left[ b \frac{db'}{dt} + \frac{db}{dt} b' \right] b'^{-1}$$

Es importante mencionar que la matriz inversa de parámetros  $W^{-1}$  determina la fuerza del acoplamiento y cómo se puede deformar la caja, relacionando las compresibilidades isotérmicas aproximadas  $\beta$  y la constante de tiempo de presión  $\tau_p$  en las condiciones de entrada donde  $L$  es el elemento de matriz de caja más grande.

$$(W^{-1})_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2L}$$

### 3.4 Dinámica Browniana

La dinámica browniana es un método utilizado en la escala mesoscópica en las que moléculas explícitas del solvente son reemplazadas por una fuerza estocástica. Es una técnica implementada en sistemas complejos; poliméricos, coloides, proteínas, entre otros en estados de equilibrio y no equilibrio[27], [28]. La ecuación browniana proviene de una solución particular de la ecuación de Fokker-Planck

Considerando un sistema no inercial de partículas en la meso-escala con un solvente viscoso, el balance de fuerza está dado por:

$$F_{net} = m_i a_i = F_{d,i} + F_{e,i} + F_{f,i} \simeq 0 \quad \text{Ec.1}$$

Donde cada partícula es denota por  $i$ , la fuerza neta  $F_{net}$ , la masa por  $m_i$ , la aceleración  $a_i$ , las fuerzas de arrastre  $F_{d,i}$ , la fuerza browniana  $F_{f,i}$  y las fuerzas externas  $F_{e,i}$  asociadas a resortes, fuerzas electrostáticas, volumen excluido y otras fuerzas.

La fuerza de arrastre es dada por ley de Stoke's en una esfera:

$$F_d = -\zeta \left( \frac{dr_i}{dt} - u^\infty(r_i) \right) \quad \text{Ec. 2}$$

Donde  $\zeta$  es el coeficiente de arrastre, Por lo tanto, considerando el flujo no perturbado  $u^\infty(r_i)$ , la ecuación que rige el sistema está denotada por:

$$\frac{dr_i}{dt} = u^\infty(r_i) + \frac{1}{\zeta} (F_{e,i}(r_j) + F_{f,i}(t)) \quad \text{Ec. 3}$$

Comúnmente llamada ecuación de Langevin. La fuerza browniana al presentar una distribución aleatoria, y al satisfacer el teorema de fluctuación-disipación, su distribución media es igual a cero:

$$\langle F_{f,i}(t) \rangle = 0 \quad \text{Ec. 4}$$

$$\langle F_{f,i}(t)F_{f,j}(t') \rangle = 2k_B T \zeta \delta_{ij} \delta(t - t') \delta \quad \text{Ec.5}$$

Donde  $\delta_{ij}$  es el delta de Kronecker,  $k_B$  es la constante de Boltzmann,  $T$  es la temperatura,  $\delta(t - t')$  es la función de delta de Dirac y  $\delta$  es el tensor unitario de segundo orden. Por lo tanto, a la integral y operar se obtiene la siguiente ecuación:

$$\frac{dr_i}{dt} \simeq u^\infty(r_i) + \frac{1}{\zeta} F_{e,i}(r_i(t)) + \sqrt{\frac{2k_B T}{\zeta \delta}} dW_i \quad \text{Ec.6}$$

$W_i$  es un proceso Wiener independiente, que cumple los siguientes procesos

$$\langle dW_i \rangle = 0 \quad \text{Ec. 7}$$

$$\langle dW_i dW_j \rangle = \delta_{ij} \delta \quad \text{Ec.8}$$

Por consiguiente, al implementar el método de primer orden en el tiempo como el de Euler obtenemos

$$r_i(t + \delta t) \simeq r_i(t) + \left( \frac{dr_i}{dt} \right) (t) \delta t \quad \text{Ec.9}$$

### 3.5 Interacciones Electroestáticas

La electrostática es el estudio de las cargas eléctricas estacionarias. Por tanto, al considerarse un sistema Coulombico, es necesario obtener potencial electrostático a partir de las ecuaciones de Maxwell [29]–[31].

#### 3.5.1 Ecuaciones de Maxwell

Describen los fenómenos electromagnéticos, cómo se genera los campos magnéticos y eléctricos por las cargas y corrientes.

$$\nabla \cdot E = \frac{1}{\epsilon_0} \rho \quad (\text{Ley de Gauss}) \quad \text{Ec.10}$$

$$\nabla \cdot B = 0 \quad (\text{Ley no asociada}) \quad \text{Ec. 11}$$

$$\nabla \times E = -\frac{\partial B}{\partial t} \quad (\text{Ley de Faraday}) \quad \text{Ec.12}$$

$$\nabla \times B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t} \quad (\text{Ley de Amperé}) \quad \text{Ec. 13}$$

El sistema a estudiar considera que el campo eléctrico irrotacional, es decir  $\nabla \times E = 0$ . Por lo cual, obtenemos que el potencial electrostático;

$$E = -\nabla\phi \quad \text{Ec. 14}$$

El potencial electrostático considera cargas puntuales y cuya fuerza es ejercida por la presencia de otras cargas. Por consiguiente, al no haber presencia de campos magnéticos variables en el tiempo e implementando la ley de Gauss obtenemos la ecuación de Poisson:

$$\nabla^2 \phi = -\frac{1}{\epsilon_0} \rho \quad \text{Ec. 15}$$

### 3.5.2 Debye-Huckel

La ecuación de Poisson bajo las siguientes consideraciones tiene una solución particular; La desviación de la idealidad es función de la energía potencial debido a las interacciones electrostáticas: Dichas interacciones están asociadas a los iones y a la nube electrónica circundante. Los iones son tratados como cargas puntuales esféricas. Las interacciones entre iones obedecen a ley de Coulomb. El disolvente es un medio dieléctrico continuo con constante dieléctrica  $\epsilon_0$ . Por lo tanto,

$$\nabla^2 \phi_i(r) = -\frac{1}{\epsilon_0} \rho_i(r) \quad \text{Ec. 16}$$

Donde  $\nabla^2 \phi(r)$  es potencial dado a una distancia  $r$ , y  $\rho(r)$  es la densidad de carga promedio. Es relevante aclarar que la nube circundante tiene simetría esférica en función de la distancia alrededor del ion central. A partir de la distribución de Boltzmann calculamos la densidad de carga;

$$n'_i(r) = n_i \exp\left(-\frac{z_i e \phi(r)}{k_B T}\right) \quad \text{Ec. 17}$$

Al operar las dos ecuaciones anteriores obtenemos la ecuación Poisson-Boltzmann

$$\nabla^2 \phi(r) = -\frac{1}{\epsilon_0} \sum_i \left[ n_i e z_i \exp\left(-\frac{z_i e \phi(r)}{k_B T}\right) \right] \quad \text{Ec. 18}$$

Al expandir el término exponencial por serie de Taylor hasta el primer orden. El primer término (orden cero) se evade, ya que el sistema presenta la condición de ser eléctricamente neutro  $\sum_i n_i z_i = 0$ . Por ende, se obtiene la ecuación de Helmholtz.

$$\nabla^2 \phi_i(r) = \kappa^2 \phi_i(r) \quad \text{Ec. 19}$$

$$\kappa^2 = \frac{e}{\epsilon_0 \epsilon_r k_B T} \sum_i n_i z_i^2 \quad \text{Ec. 20}$$

La solución general dada la condición cuando  $r \rightarrow \infty$ , y para electrolitos simétricos, operando e integrando se obtiene.



$$\phi_i(r) = -\frac{z_j e}{4\pi\epsilon_0\epsilon_r} \frac{e^{\kappa a_0}}{1+\kappa a_0} \frac{e^{-\kappa r}}{r} \quad \text{Ec. 21}$$

Las interacciones electrostáticas están asociadas a cargas puntuales localizadas entre cationes y aniones. Dichas interacciones pueden ser atractivas o repulsivas según el signo de la carga. Si las cargas que interactúan son del mismo signo, entonces habrá repulsión entre ellas, lo que se significa que es necesario trabajo para poder mantenerlas juntas. Sin embargo, si las cargas que interactúan tienen diferente signo, habrá atracción entre ellas. Por lo tanto, es necesario aplicar trabajo para poder separarlas [10], [29]–[32].

## 4. Proteínas de vías del complemento

### 4.1 Vías de complemento

El sistema del complemento es un sistema funcional de reacciones enzimáticas medidas por diversas proteínas que reaccionan entre sí induciendo una serie de respuestas inflamatorias para combatir infecciones ocasionadas por patógenos. Existen tres vías principales de activación; clásica, lectinas, alternativa y terminal, las cuales están entrelazadas a su vez con vías de coagulación[33]. En figura 4 se puede observar las diferentes vías de activación como sus respectivas funciones [33].

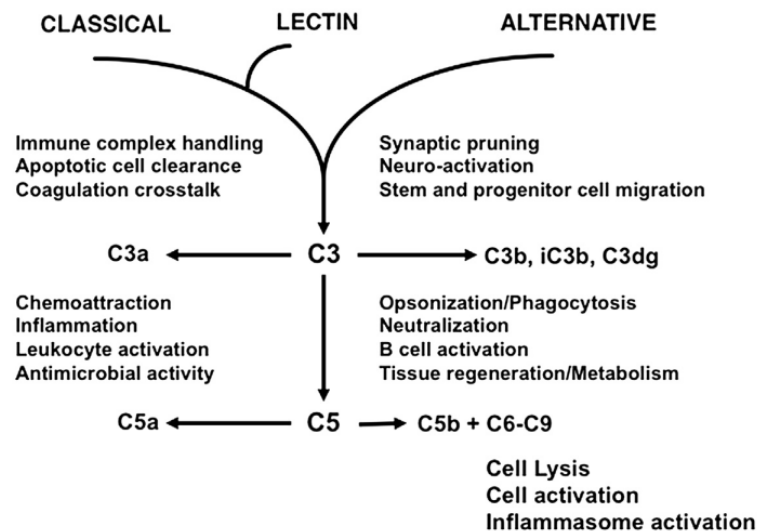


FIGURA 4. Representación esquemática de las tres vías del complemento.

**Vía clásica:** Es activada cuando C1q, la porción de reconocimiento de patrones del complejo C1, interactúa a uno de diversos tipos de activadores, incluidos los complejos antígeno-anticuerpo (Ag-Ab), proteína C reactiva, células apoptóticas y otras estructuras. La activación por complejos Ag-Ab que contienen IgM e IgG se da al unirse a la región Fc de los anticuerpos. Un complejo de IgM pentamérico con antígeno activará la vía, requiriendo múltiples moléculas de IgG presencia de  $Ca^{2+}$  e inhibida por cationes quelantes[33].

**Vía de las Lectinas:** La activación de esta se debe a proteínas macromoleculares como; MBL, ficolina y la colectina las cuales se asemeja al C1q con tallos de colágenos y dominios de unión a carbohidratos. Las ficolinas se asocian con las proteínas MASP para activar la vía de la lectina. A sí mismo, el heterodímero de las colectinas CL-L1 y CL-K1 (denominadas CL-LK) en complejo con MASP-1/3 y MASP-2 activa la vía de la lectina del mismo modo. Una vez que MBL, ficolina o colectina se unen a un activador, un cambio conformacional conduce a la auto activación de MASP-1, que luego activa MASP-2. La activación de MASP-1 de MASP-2 puede ocurrir en cis, si ambas proteínas están en el mismo complejo, o en trans si están ubicadas en complejos adyacentes. MASP-2 escinde tanto C4 como C2 de forma análoga a C1 en la vía clásica. MASP-1 también puede escindir C2, mejorando la activación de la vía[33].

**Vía alternativas:** Esta vía es única entre las vías del complemento porque siempre está activa a un nivel muy bajo y, como tal, está preparada para responder rápidamente a la infección. La activación constitutiva de bajo nivel de la vía alternativa se conoce como "tickover". En este proceso, la proteína nativa C3 sufre un cambio conformacional reversible que exhibe el dominio que contiene tioéster (TED). El tioéster se hidroliza y la molécula C3(H<sub>2</sub>O) puede interactuar con el sitio de unión para el factor B, formando consigo el complejo de la convertasa C3(H<sub>2</sub>O)Bb [33].

En este trabajo nos enfocaremos en las proteínas Factor B y C3, las cuales son importante en la activación de la vía alternativas y sus reacciones en cadena para contribuir a las respuestas inflamatorias para así desarrollar y cumplir con el objetivo 1 propuesto anteriormente. Por consiguiente, es necesario entender las interacciones superficiales entre ellas y sus configuraciones estructurales. No obstante, hasta el momento las proteínas han sido caracterizadas de manera bioquímica, explicando las funciones biológicas en vías del complemento. A través de este estudio se desarrollará una metodología que permite realizar simulaciones multi-proteicas a través de las superficies de las proteínas. Sin embargo, inicialmente es necesario hallar la mejor configuración estructural mediante la dinámica molecular para obtener así las superficies de Van der Waals de cada una de ellas, permitiendo así proponer un método de interacciones de superficies y propiedades fisicoquímicas.

## 4.2 Factor B

El factor B es un polipéptido monocatenario de aproximadamente 93 kD que contiene cinco dominios distintos y está compuesto por un total de 739 aminoácidos. Tras la activación de la vía alternativa el factor B se segmenta debido al factor D en un fragmento Ba (~ 33 kD y residuos 1-234) y un fragmento Bb (~ 60 kD y residuos 235-739). La región Ba de la proteína abarca tres dominios de la proteína de control del complemento (CCP), que son importantes para la unión inicial del factor B a C3b. El fragmento Ba está unido a con el fragmento Bb a través del residuo 45 y el enlace scissile es cortado por el factor D en la posición (Ag259 – Lys260). El enlace escindible del factor B no unido está protegido de la escisión no específica y se mantiene en una conformación "bloqueada" por dos puentes salinos, que unen Arg259 a Glu471 y Glu232. La unión del factor B a C3b interrumpe el puente salino Arg259-Glu471. Esto induce un cambio conformacional que conduce a la activación de un sitio de adhesión dependiente de iones metálicos Mg<sup>2+</sup> (MIDAS) y la estabilización del complejo C3bB-Mg<sup>2+</sup>. El cambio conformacional también expone el enlace escindible y permite la escisión de las proteínas por el factor D.

El fragmento Bb contiene un dominio de tipo A, un factor de von Willebrand y un dominio de peptidasa S1 (serina proteasa). El dominio SP del fragmento Bb es el componente catalíticamente activo de la ruta alternativa C3 y C5 convertasas. Una vez que el factor B se ha unido a C3 (H<sub>2</sub>O) o C3b, sufre un cambio conformacional permitiendo que el enlace covalente sea accesible para el clivaje por el factor D. El fragmento Ba es soluble, pero el fragmento Bb permanece unido a C3b[34].

Las convertasas C3 escinden C3 en un enlace Arginina-Serina en la cadena  $\alpha$  del complemento C3 (residuos 726–727), produciendo C3a y C3b. La convertasa C5 escinde C5 en un enlace arginina-leucina en la cadena  $\alpha$  de C5 (residuos 74–75), que produce C5a y C5b. La inactivación de las convertasas C3 y C5 se consigue mediante la disociación de Bb de C3b. La vida media de C3(H<sub>2</sub>O)Bb y C3bBb es de aproximadamente 77 y 90s, respectivamente, pero el factor acelerador del decaimiento (DAF o CD55) y el factor H aceleran la disociación de estas moléculas las cuales no se pueden volver a unir. La actividad de la convertasa es regulada a través del factor D cuando el factor B está unido con C3(H<sub>2</sub>O) o C3b.

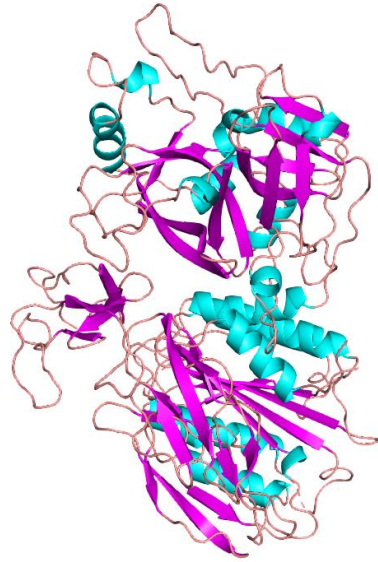


FIGURA 5. Representación tridimensional de la proteína Factor B en cartoon.

### 4.3 C3

C3 está compuesto por dos cadenas polipeptídicas (cadenas  $\alpha$  y  $\beta$ ) que están unidas por disulfuro con un enlace disulfuro intra-cadena adicional en la cadena  $\alpha$ . La proteína C3 está conformada por 1663 aminoácidos, con un punto isoeléctrico PI aproximado de 5.9, además de tener un tamaño de 187 kDa. La cadena  $\alpha$  tiene un enlace tioéster altamente reactivo formado en la posición del residuo de la cisteína 988 y el residuo de la glutamina 991. La proteína C3 tiene 13 dominios principales; ocho de los dominios tienen un pliegue de núcleo de fibronectina de tipo 3 y se denominan dominios de macroglobulina (MG). La cadena  $\beta$  está compuesta por seis dominios MG en tándem: MG1 a MG6 $\beta$  seguido de una región enlazadora denominada LNK. A su vez, la cadena  $\alpha$  continúa con el dominio MG7, seguido de un dominio denominado CUB, luego el dominio tioéster (TED)[35]. La reactividad alterada del C3 se debe a una transformación del tioéster en un anión tiolato libre, en Cys 988, y a la formación de un intermediario acil-imidazol34,35 por Gln 991 e His 1104 que posiblemente esté estabilizado por Glu 1106[36]. La orientación del dominio TED con el tioéster apuntando hacia adentro parece ser esencial para mantener el TED protector. Interacciones de MG8 en C3 nativo. TED está respaldado en esta posición por interacciones con MG2, MG8 y CUB. La escisión de C3 en Ser 726–Arg 727 elimina el

dominio ANA (que se convierte en la anafilatoxina C3a) y produce un C3b activado con un tioéster expuesto y reactivo[36].

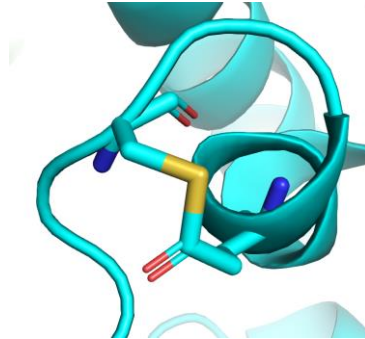


FIGURA 6. Enlace tioéster de la proteína C3 formando en los enlaces con los residuos de la cisteína y la glutamina.

La proteína C3 tiene como principal función la activación y regulación del sistema del complemento, interactuando con otras proteínas como; C3aR, CR1, CR2, CR3, CR4, factor I, factor H, CHFRs, MCP, DAF, CSMD1, C2 y factor B para mediar la respuesta inmunológica del complemento las cuales las más importantes son descritas a continuación: La C3a es mediador proinflamatorio ante infecciones, traumas, autoinmunidad entre otros. La interacción entre la C5a y el C3 al unirse al receptor de las células mieloides contribuyen a la activación celular, desgranulación y quimio atracción. Además de atribuirle funciones antimicrobianas bactericida, propiedades de supresión de citocinas inflamatorias y shock inducido por lipopolisacáridos. C3b/iC3b cumple con las funciones de eliminación de patógenos a través de la fagocitosis opsonizada a través de neutrófilos, macrófagos y células dendríticas. A su vez, dirigir patógenos intracelulares al proteosoma a través de la señalización y activación por el citosol la producción de citoquinas proinflamatorios. No obstante, existen otros tipos de funciones asociados a los complejos C3f/C3c y C3d/CEd<sub>g</sub> las cuales no son descritas en este documento.

La deficiencia de la proteína C3 o bajos niveles de está puede inducir enfermedades ocasionadas por infecciones como; neumonía, otitis, meningitis, osteomielitis, entre otras. Del mismo modo, pueden presentar enfermedades por inmunocomplejos o polimorfismos.

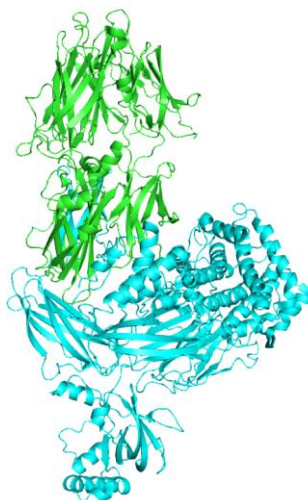


FIGURA. 7. Representación de la proteína C3 con la cadena A, y la cadena B.

## 4.4 Metodología

### 4.4.1 Proteínas

Los archivos de las proteínas tridimensionales fueron obtenidos de la base de datos Protein Data Bank; para el Factor B la proteína de referencia fue 2oK5 y para el C3 fue la 2a73. Los cuales son los únicos archivos presente en el repositorio para las proteínas de interés hasta el momento de la publicación de esta tesis. Posteriormente, a través del modelamiento por homología utilizando el software Modeller se completó los aminoácidos faltantes en las estructuras tridimensionales a partir de la secuencia completa de cada una de ellas, implementando el algoritmo de AutoModel, eligiendo la proteína con menor valor de Energía proteica optimizada discreta (DOPE) score.

### 4.4.2 Protocolo de Simulación

El software Gromacs (v2020.3) fue utilizado para las simulaciones de dinámica molecular para las proteínas completadas del C3 y Factor B. El potencial o el campo de fuerza utilizado fue CHARMM36, la caja de simulación fue cubica acorde al tamaño de cada proteína, el solvente explícito efectuado fue el modelo de agua SCP, se adiciono los iones de NA o CL al solvente para equilibrar la carga total del sistema (Electro-neutro). Posteriormente, se minimiza el sistema con el algoritmo steepest descent [22] con una tolerancia menor a 1000.0 kJ/mol/nm. Seguidamente, el sistema es equilibrado mediante

dos fases, iniciando por el ensamble canónico y posteriormente por ensamble isotérmico-isobárico a condiciones de 310 K, presión constante de 1 bar (isotrópica), utilizando el termóstato, V-rescale (Berendsen modificado) y el baróstato de Parinello-Rahmann respectivamente, cada fase fue ejecutada durante 10 ns. El tiempo de producción se realiza durante 100 ns con condiciones periódicas de contorno, interacciones de largo alcance Van der Waals y electrostáticas con tratamiento de Particle Mesh Ewald. El integrador empleado es Leap-frog, cuyo paso de integración es de 0.002.

## 4.5 Resultados y análisis

La minimización del factor B, tiene una energía potencial de  $-5 \times 10^6$  kJ/mol a 2000 pasos de minimización. Posteriormente a esta relajación, el sistema se equilibra obteniendo consigo valores promedios para las siguientes propiedades; la densidad del sistema es de  $1000 \text{ Kg/m}^3$ , la temperatura de 310.02 K, y la presión de 1.01 bar, como se puede observar en la figura 8.

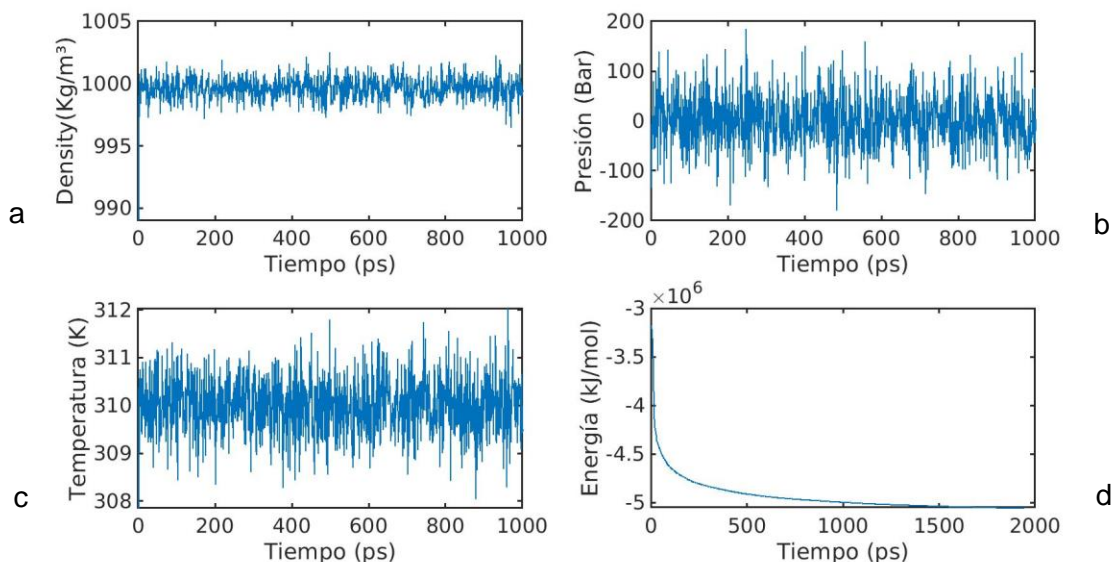


FIGURA 8. Propiedades termodinámicas del sistema en el tiempo. a) Densidad, b) Presión, c) Temperatura y d) Energía de minimización del sistema.

Al finalizar la producción se calculó la raíz de la desviación media cuadrática del backbone del factor B en el tiempo durante 100 ns. Inicialmente el RMSD aumento hasta alcanzar un valor promedio de 10 nm a partir de los 10 ns. A su vez, el radio de giro de la proteína disminuyo a través del tiempo obteniendo un valor promedio en 2.82 nm al final de la simulación. Lo anteriormente descrito se puede observar en la figura 9.



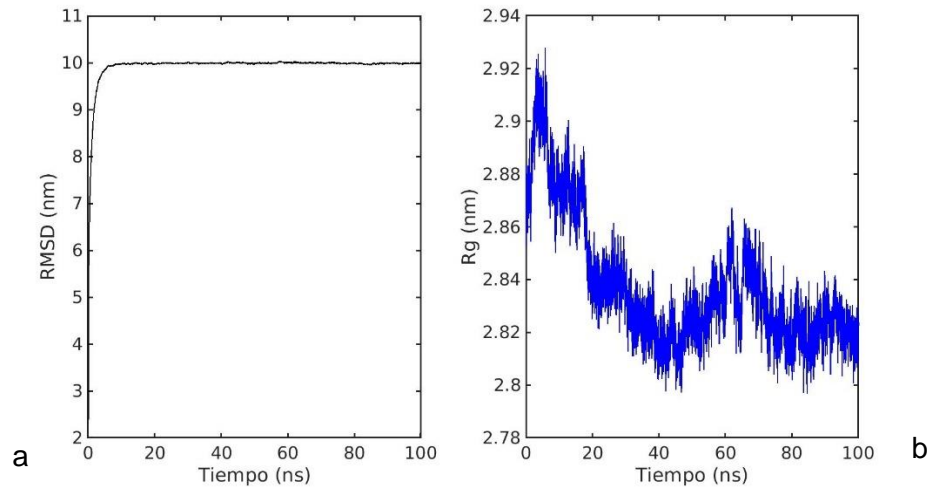


FIGURA 9. a) Desviación cuadrática media de la raíz (RMSD) de backbone de la proteína del factor b en el tiempo. b) Evolución temporal del Radio de giro (Rg).

Los momentos dipolares son calculados a través del tiempo en coordenadas x, y, z como su total con valores promedios de -1000, 200, 1200 y 1700 debye respectivamente. De igual manera, la constante eléctrica es calculada obteniendo consigo un promedio de 7 +/- 1.6 durante 100 ns de producción como se puede observar en la figura 10.

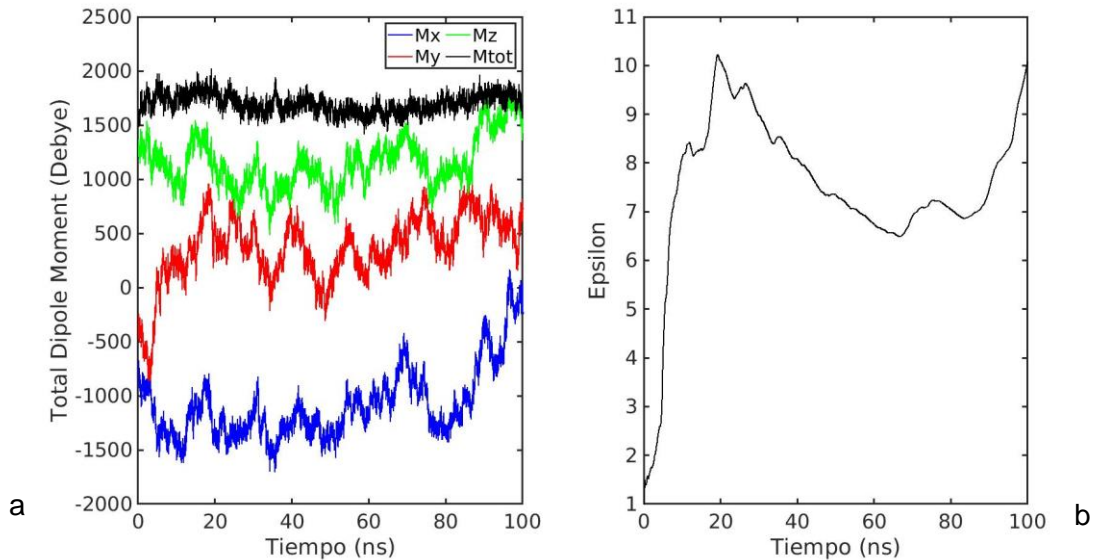


FIGURA 10. a) Momento total dipolar en el tiempo; momento en x (línea azul), momento dipolar en y (línea roja), z y momento dipolar total (línea negra). b) Evolución temporal de la constante dieléctrica en el tiempo.

Al finalizar el tiempo de producción, se extrajo la proteína a los 100 ns, para posteriormente obtener la superficie de esta y ser comparada con la estructura reportada en Protein Data Bank, como se puede observar en la figura 11. En donde se puede notar diferencias estructurales a nivel de superficie; de color rojo la superficie del Factor B de la proteína reportada y de color azul la superficie al terminar la producción de la dinámica molecular.

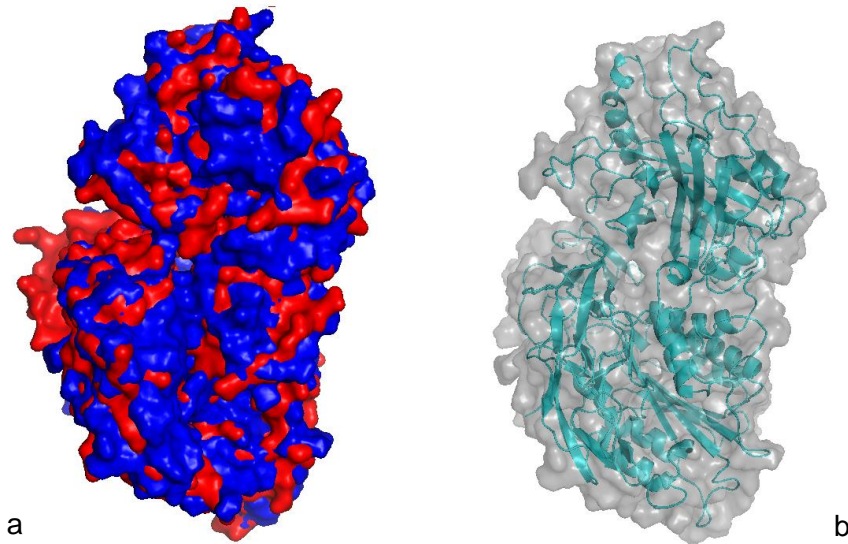


FIGURA 11. Representación de la Superficie de Van der Waals; a) Factor B cristalizada (color rojo) y Factor B obtenido de la dinámica Molecular (color azul). b) Representación de la superficie de Van der Waals de la proteína Factor B obtenida desde la dinámica molecular con la configuración cartoon.

A partir de la malla de la proteína Factor B, se realizó una caracterización electroestática mediante BEM (Boundary Element Method) basado en la formulación de ecuaciones integrales. Es importante resaltar que este método nos permite tener en cuenta la polarización de la proteína al estar inmersa en un medio. Por consiguiente, se hizo dos cortes asociados al fragmento de Factor Ba y el fragmento de Factor Bb, como se puede observar en la figura 12. Por lo tanto, al hacer la distribución de frecuencia del potencial en cada corte de la proteína, se evidenciando que el factor B y el fragmento Bb tiene una distribución más negativa que el factor Ba (Figura 12). Esto se debe principalmente a que el factor Ba tiene más aminoácidos cargados positivamente en dicha región a diferencia del fragmento FBb que tiene mayores aminoácidos cargados negativamente. No obstante,

el fragmento FBb está polarizado en regiones positivas y negativas las cuales pueden tener un mejor interacción y reacción con otras proteínas en este caso con la proteína C3.

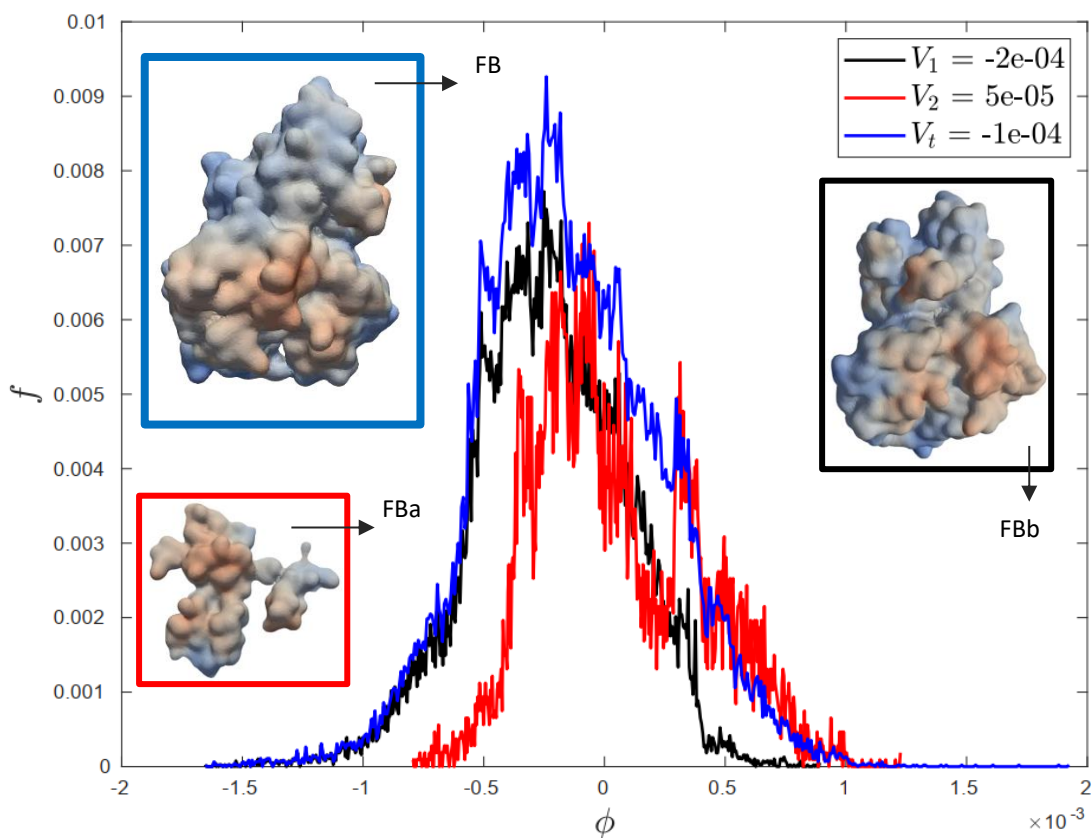


FIGURA 12. Potencial Electroestático: Factor Ba (línea roja), Factor Bb (línea negra) y Factor B

Por otro lado, los resultados obtenidos de la simulación de la proteína C3 es la siguiente: La minimización de la proteína de C3, obtuvo una energía potencial de  $-1 \times 10^7$  kJ/mol en 5000 pasos de minimización. Consecutivamente a la minimización, al equilibrarse él se mide a través del tiempo las siguientes propiedades termodinámicas promedios; la

densidad del sistema cuyo valor es de  $995.5 \text{ Kg/m}^3$ , la temperatura de  $310 \text{ K}$ , y la presión de  $0.99 \text{ bar}$ , como se puede observar en la figura 13.

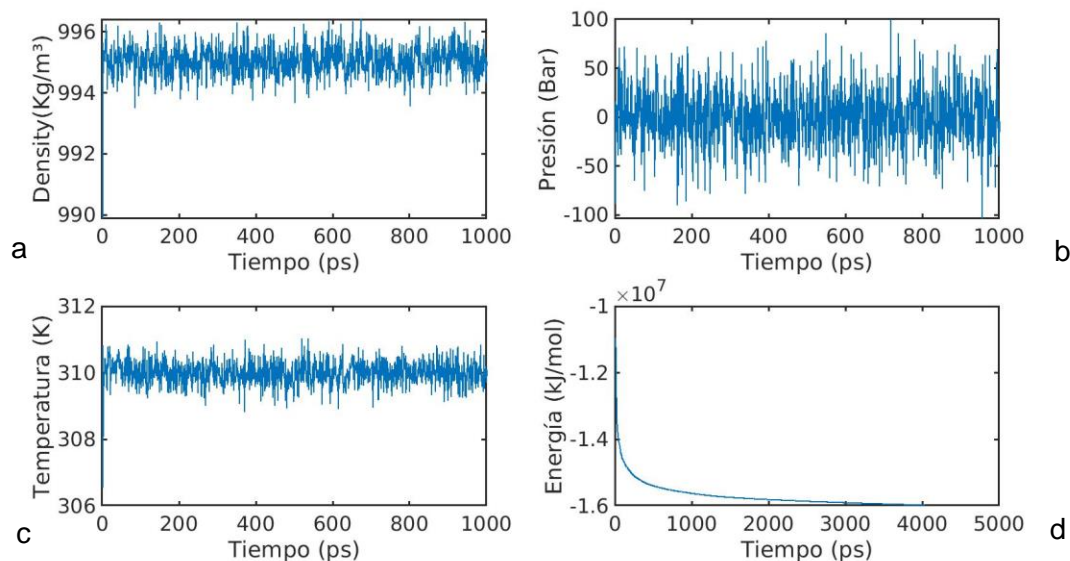


FIGURA 13. Propiedades termodinámicas del sistema en el tiempo. a) Densidad, b) Presión, c) Temperatura y d) Energía de minimización del sistema.

Al terminar la producción se calculó la desviación cuadrática media de la raíz (RMSD) de la backbone del factor B a través del tiempo. Inicialmente el RMSD aumento hasta alcanzar un valor promedio de  $0.45$  entre  $25$  a  $80 \text{ ns}$ . Sin embargo, a partir de los  $80 \text{ ns}$  a  $100 \text{ ns}$  la tendencia fue en aumento hasta alcanzar un valor  $0.68 \text{ nm}$ . Seguidamente, el valor se calculó el valor de radio de giro; En donde inicialmente disminuye de  $4.37 \text{ nm}$  hasta  $4.25$  en los primeros  $20 \text{ ns}$ , manteniéndose constante hasta alcanzar los  $78 \text{ ns}$ , para luego disminuir a  $4.2 \text{ nm}$  en  $100 \text{ ns}$  como se puede evidenciar en la figura 14.

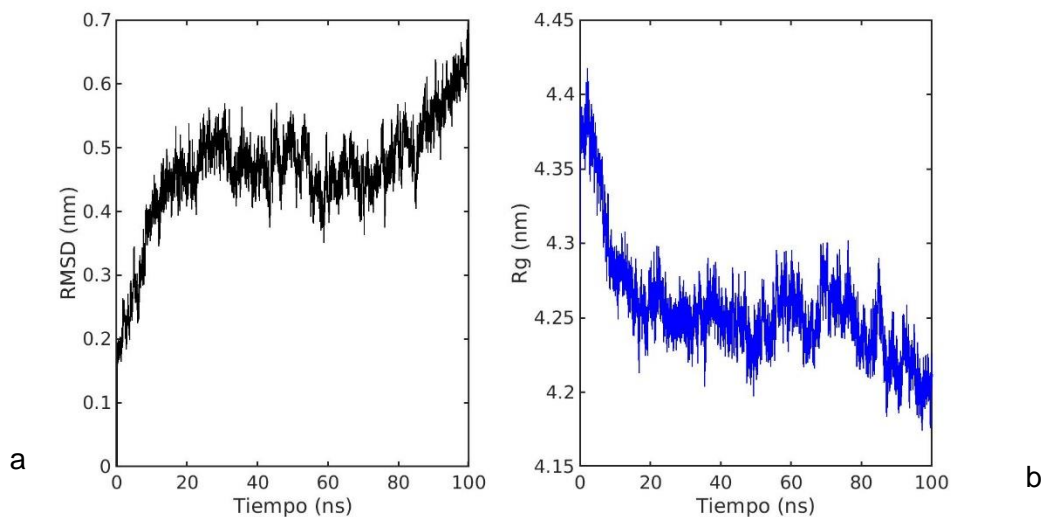


FIGURA 14. a) Desviación cuadrática media de la raíz (RMSD) de backbone de la proteína del factor b en el tiempo. b) Evolución temporal del Radio de giro (Rg).

A su vez, los momentos dipolares son calculados a través del tiempo en coordenadas x, y, z como su total con valores promedios de -700, -700, -3500 y 4000 Debye respectivamente. Seguidamente, la constante dieléctrica presento un incremento durante los 70 ns de producción. Sin embargo, a medida disminuyo hasta 18 en 100 ns como se puede observar en la figura 15. El promedio de la constante dieléctrica es del C3 es de 11+/- 2.0.

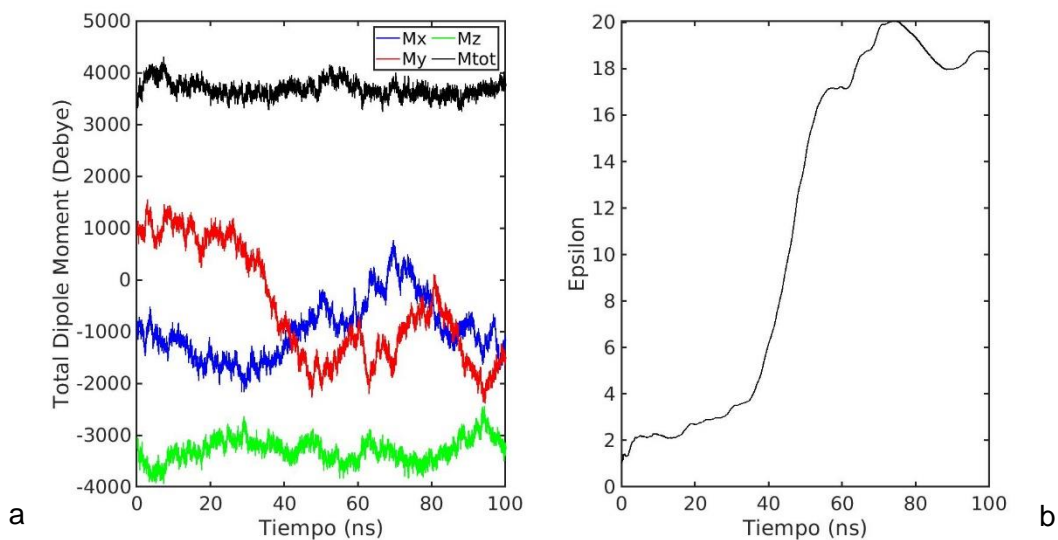


FIGURA 15. a) Momento total dipolar en el tiempo; momento en x (línea azul), momento dipolar en y (línea roja), z y momento dipolar total (línea negra). b) Evolución temporal de la constante dieléctrica en el tiempo.



Al finalizar el tiempo de producción, se extrajo la proteína a los 100 ns, para posteriormente obtener la superficie de esta y ser comparada con la estructura reportada en Protein Data Bank como se puede observar en la figura 16 a través del software PYMOL[37] o Chimera [38]. En donde se puede notar diferencias estructurales a nivel de superficie; de color roja la superficie C3 de la proteína reportada y de color azul la superficie al terminar la producción de la dinámica molecular.

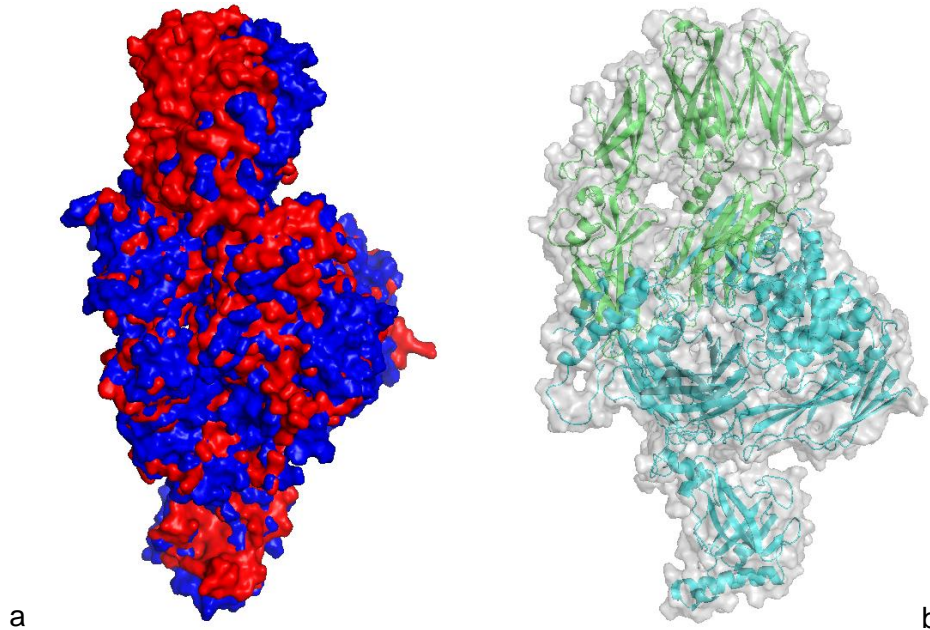


FIGURA 16. Representación de la Superficie de Van der Waals; a) C3 cristalizada (color rojo) y C3 obtenido de la dinámica Molecular (color azul). B) Representación de la superficie de Van der Waals de la proteína C3 obtenida desde la dinámica molecular con la configuración cartoon.

A partir de la malla de la proteína C3, se realizó la caracterización electroestática mediante BEM (Boundary Element Method) basado en la formulación de ecuaciones integrales. Por consiguiente, se hizo un corte en el sitio activo de la proteína C3, específicamente en el fragmento C3b, el cual reacciona con el factor B para generar la convertasa C3, dichos fragmento se puede observar en la figura 17, como la proteína C3 completa. Por lo tanto, al hacer la distribución de frecuencia del potencial el corte del sitio activo y al compararla con la proteína completa, se evidenciando que el fragmento del sitio activo C3b tiene una distribución más negativa que la proteína C3 (Figura 17). Dada la distribución de aminoácidos cargados como aminoácidos polares en dicha región se evidencia mayor

potencial negativo permitiendo la posible interacción o reacción mencionada anteriormente.

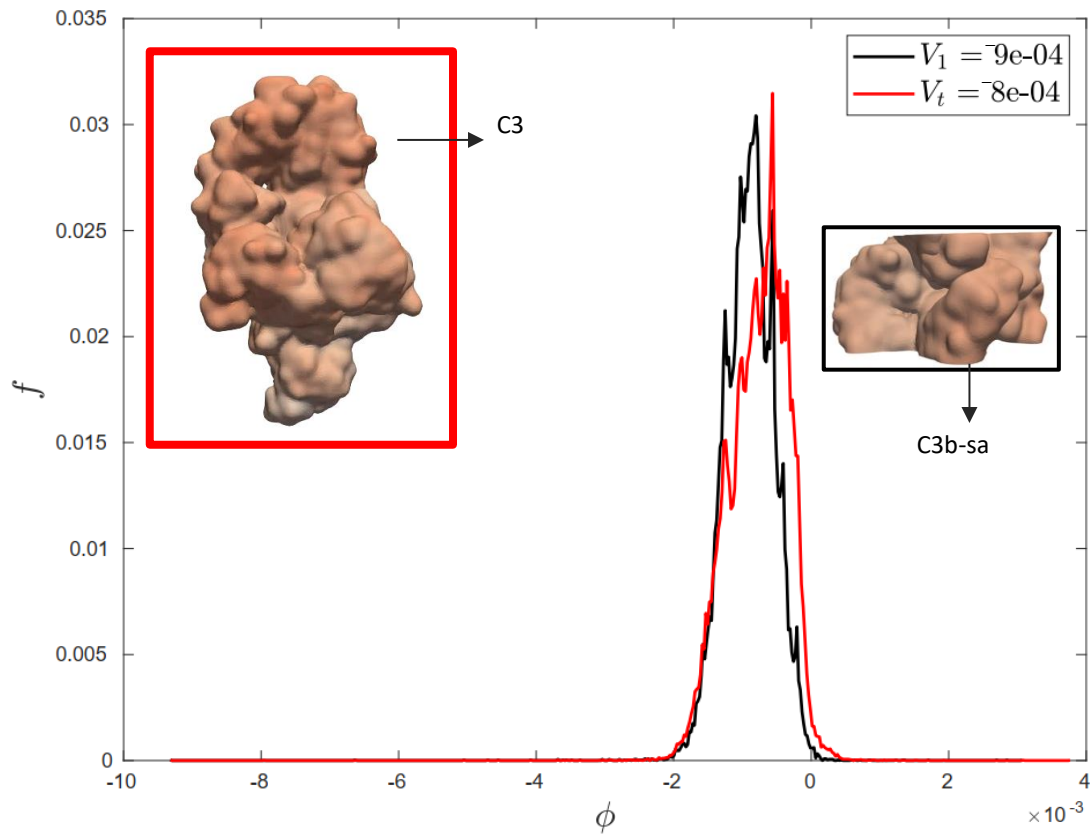


FIGURA 17. Potencial Electrostático: C3 (línea roja) y C3b (línea negra)

## 4.6 Conclusiones

La metodología aplicada permite obtener superficies de proteínas a través de la implementación de modelamiento por homología, y dinámica molecular en donde a través de una estructura de referencia, es completada para poder llevar a cabo simulaciones atómicas en medio de solvente explícitos, para así medir propiedades termodinámicas y físicas de las proteínas, analizando y explorando el comportamiento del sistema en el tiempo. A su vez, la metodología desarrollada permitió la obtención de las distribuciones de las cargas como sus respectivas coordenadas, las cuales pueden ser ubicadas en las superficies, y así ser un modelo inicial de posibles interacciones multi-proteínas entre sus respectivas superficies. Permitiendo consigo entender mecanismos de interacción entre proteínas de interés como Factor B y C3.

El cálculo de la constante dieléctrica a través del software GROMACS mostro variaciones de la constante a través que el tiempo de simulación, sin alcanzar un valor estable. Esto probablemente a los efectos de polarización que pueden tener el solvente del agua con las proteínas y/o interacciones electroestáticas locales entre ellas. Del mismo modo, es importante estudiar la evolución de los sistemas a mayores tiempos de producción dada a los posibles estados metaestables que puede presentar las proteínas a través de sus cambios conformaciones en el tiempo. El radio de giro presenta diferentes fluctuaciones con intervalos no mayores a 0.2 nm en tiempos específicos los cuales pueden estar relaciones al tipo de interacción y potencial del sistema. A su vez, propiedades como los momentos dipolares presentan fluctuaciones en valores concretos, indicando las posibles estabildades en valores promedios.



## 5. Agregación y segregación de partículas biológicas

La formación de estructuras supramoleculares se da a través del auto-ensamble entre proteínas y ARN mediante interacciones electromagnéticas y estereo-específicas en equilibrio termodinámico en fases fluidas. Sin embargo, se desconoce los mecanismos fisicoquímicos que permiten dicha formación. De tal modo, es prioridad entender los procesos biológicos como sistemas dinámicos, por medio de investigaciones *in-silico*, en ámbitos como: ingeniería molecular, proteómica, metabolómica y la genómica, [5][6],[7]. Por consiguiente, este estudio tiene la finalidad de construir modelos moleculares a partir de ARN y proteínas para la formación de estructuras supramoleculares como los ribonucleopartículas (RNPs) de germoplasma vegetales y animales mediante herramientas computacionales que expliquen los plegamientos, auto-ensamble e interacciones presentes en los mecanismos de formación, para entender y controlar las funciones de diferenciación y división celular.

### 5.1 RNPs

La formación de RNPs han sido estudiados experimentalmente como interacciones entre proteínas y RNA. Los RNPs presentan diversos compartimientos; gránulos de estrés, cuerpos de procesamiento y exosomas. Los cuerpos cromatoides se encuentran en las células germinales, a diferencia de las de transporte que se hallan presentes en ovocitos y neuronas. En muchos organismos como en el pez cebra, la especificación de las células germinales primordiales (PGC) requiere la transmisión del plasma germinal derivado de la madre. Los RNP de germoplasma se heredan típicamente como partículas dentro del huevo, que durante los primeros ciclos celulares o embriogénesis se agregan en grandes masas. Estos, se segregan en (PGC), donde se cree que promueven el inicio el proceso transcripcional de células germinales. Se han realizado experimentos para determinar la relación espacial entre los componentes proteicos de los granos a partir de la microscopia de luz, mediante el uso de fluorescencia e inmuno-tinción en las proteínas [40]–[42]. A su vez, el método de detección de ARN smFISH con imágenes de alta resolución, permitió

determinar la posición de los ARNm al interior de los gránulos para establecer límites entre ARN y proteínas, confirmando consigo agregados conjuntos heterogéneos [40], [43], [44].

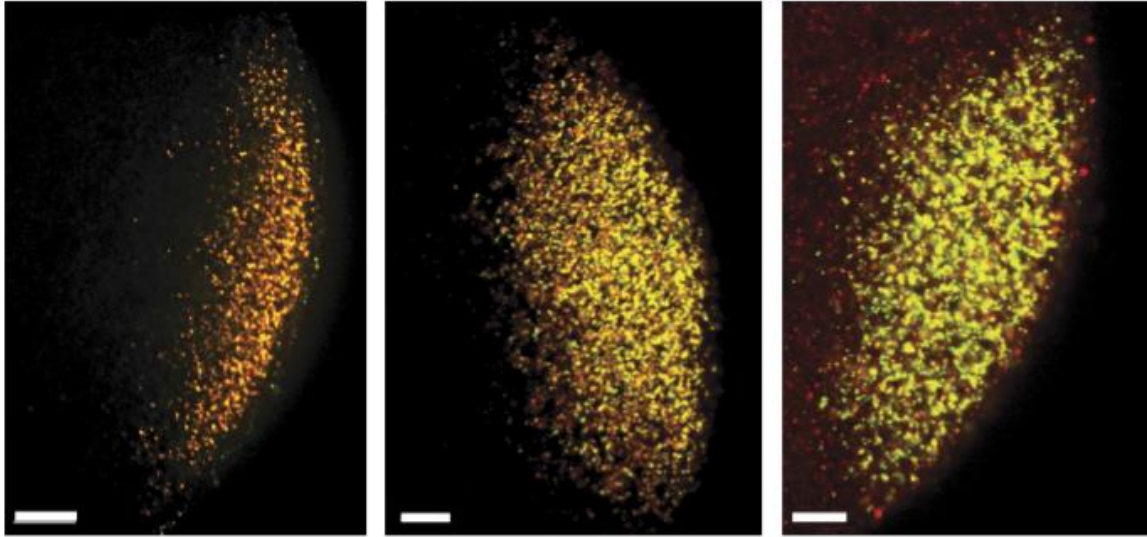


FIGURA. 18. Visualización de RNPs a través de fluorescencia en microscopía de luz.[42]

Estudios anteriores en peces cebras [41], [45], concluyen que el plasma germinal del pez cebra consiste en ribonucleopartículas (RNP) con ARN que se originan en los dos polos (animal y vegetal) del huevo. Se han evidenciado agregaciones y segregaciones de RNPs en peces cebra mediante inmunofluorescencia combinada e hibridación in situ [40]–[43]. Visualizando consigo formas esféricas de las partículas RNPs en el cual contiene un solo ARN. Sin embargo, otros estudios postulan que las interacciones entre proteínas intrínsecamente desordenadas con ARNs permiten la formación de RNPs, como estructuras micelares que posteriormente dan una apariencia circular [46]–[48]. No obstante, recientes estudios han permitido evidenciar agregaciones entre varios RNPs formando estructuras similares a una hélice, conservando consigo el aspecto esférico de los RNPs [40], [44], [47].

Existen actualmente 11 ARN de germoplasma animal reportados en la literatura como: *Askopos*, *carbonic anhydrase*, *dead end (dnd)*, *granullit (gran)*, *hook microtubule-tethering protein 2 (hook2)*, *microRNA 202-5p (miR-202-5p)*, *vasa/DEAD box polypeptide 4 (vasa/ddx4)*, entre otros. Del mismo modo, existen 2 ARN germoplasma vegetal RNPs: *celf-1/bruno-like (celf1)* y *deleted in azoospermia-like (dazl)*. No obstante, muchas

proteínas son parte de los agregados del germoplasma entre los cuales destacan las proteínas: Bucky ball (Buc), Celf1/Bruno-like y Tudor domain protein 6a (Tdrd6a).

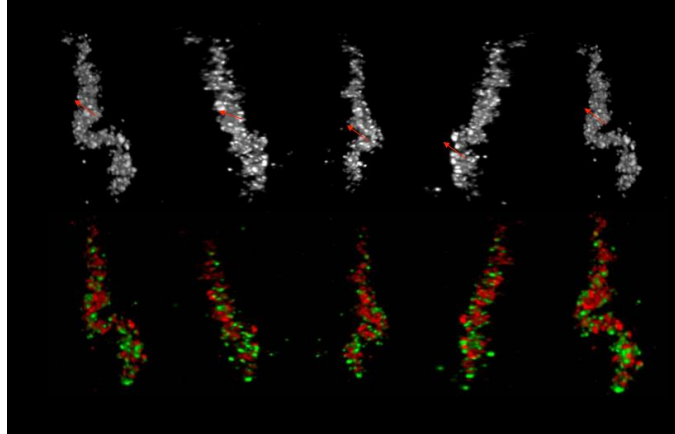


FIGURA 19. Imagen experimental de RNPs germinales de Zebrafish formando supraestructuras helicoidales [40].

Los RNPs son homotípicos y cada partícula tiene una forma esférica que contiene un solo tipo de ARN, tiene un radio estimado entre 0.231  $\mu\text{m}$  y 1.274  $\mu\text{m}$ , con un promedio de 0.576  $\mu\text{m}$ . Por consiguiente, presenta un volumen promedio de 0.800  $\mu\text{m}^3$ , alcanzado límites inferiores de 0.42  $\mu\text{m}^3$  y superiores de 1.86  $\mu\text{m}^3$

Métodos de simulación computacional como la dinámica molecular, Monte Carlo, la dinámica browniana, guían y contribuyen al diseño de experimentos *in vivo* e *in vitro* de los sistemas biológicos [4], [5], [49]. Un ejemplo de estos, son los estudios a través de interacción de electrostáticas entre proteínas en solución para entender las implicaciones de la energía libre de Gibbs a través de simulaciones de Monte Carlos, de tal manera que permita regular y entender posibles agregaciones. Otro de los métodos que se están implementado para interacciones entre proteínas, interacciones de plegamiento y proceso de auto-organización es la dinámica browniana con el fin de ver como se afecta las agregaciones el cambio concentraciones de pH en solución [21], [50]–[52].

Los modelos computacionales permiten capturar conectividades entre diferentes escalas de las funciones biológicas, ya que pueden cerrar la brecha en la comprensión entre los experimentos aislados *in vitro* y los modelos *in vivo*. Es decir, el acoplamiento de las escalas de espacio y tiempo en el que ocurre los fenómenos biológicos que permiten

evidenciar la organización de materia. A medida que las interacciones entre procesos atomísticos se transmiten, otorga información a precursores o moléculas para formación de biomoléculas hasta alcanzar sistemas ordenados y complejos de auto-organización que permiten realizar múltiples funciones biológicas, químicas y físicas [1]–[5], [49], [53]. Por este motivo, el modelamiento multi-escala de los RNPs a nivel molecular y mesoscópico conllevará a comprender los mecanismos y propiedades en cada etapa de formación. La conformación de las estructuras supra-moleculares (Figura 19) como los RNPs por sus interacciones, plegamiento y propiedades de auto-ensamble, son de gran interés en la ingeniería por presentar posibles aplicaciones a nivel de bio-conductores eléctricos, biosensores, diseño de órganos sintéticos, creación de nuevos bio-materiales y diferentes áreas de la biotecnología.

## 5.2 Metodología

En la base de datos de NCBI se obtuvieron las secuencias de las de proteínas de referencia Bruno y Buc. A su vez, las secuencias de RNA; Bruno, Dazl, Dnd y nano. Posteriormente, se hace un análisis a cada una de ellas de la siguiente manera; para las proteínas se realiza un modelamiento por homología mediante el software MODELLER, se identifica las distribuciones y cuantificación de aminoácidos, las cargas locales y totales a través cálculos matemáticos en Python. A su vez, para las secuencias de RNA se predicen las estructuras secundarias y las posibles hibridaciones en entre ellas a través de modelos de mínima energía libre (MFE). Posteriormente, se propone un modelo mesoscópico acorde a los resultados anteriores en el cual los RNPs son representados como una esfera o icosaedro tridimensional dadas las observaciones experimentales entre las proteínas y RNAs. Para entender el mecanismo de formación de hélices dadas las interacciones es implementada la dinámica browniana en un sistema de partículas de RNPs electroneutras con diferentes distribuciones de carga puntuales en la superficie, cuyas interacciones electrostáticas son debido al potencial eléctrico de Debye-Huckel, además de tener un potencial de Lennard-Jones (LJ) truncado para describir la interacción del volumen de exclusión entre las entre ellas. Es importante mencionar que el desarrollo de la representación del icosaedro de la estructura de los RNPs, la ejecución de la dinámica browniana y Monte Carlo se desarrolló a través de programación en el lenguaje de Fortran 90.

### 5.3 Resultados y análisis

La predicción de hibridación de moléculas de ARN se basa en modelos de mínima energía libre (MFE). Las interacciones se predicen minimizando la puntuación de energía de interacción, mediante la programación dinámica de los aportes energéticos de los nucleótidos. La puntuación de energía de una interacción predicha de ARN-ARN es la suma de las siguientes contribuciones: la energía libre de hibridación de las subsecuencias interactivas y las energías libres necesarias para mostrar los sitios de interacción en ambas moléculas de ARN. Por consiguiente, el análisis de hibridación se realizó entre pares de las secuencias de ARN del germoplasma vegetal y animal como se muestra en la siguiente tabla:

Sequence 1	sequence 2
Bruno	dazl
Bruno	dnd
Bruno	nanos
Bruno	vasa
Bruno	Bruno
dazlv	dnd
dazlv	nanos
dazlv	vasa
dazlv	dazlv
dnd	nanos
dnd	vasa
dnd	dnd
nanos	vasa
nanos	nanos
vasa	vasa

Tabla 3. Distribución de las pares de secuencias analizadas entre ARN de germoplasma animal y vegetal.

En la figura 20 se pueden ver las secuencias de ARN con una flecha roja indicando la posición 5' a 3' y una flecha verde indicando la posición 3' a 5' en la alineación. El mapeo muestra la interacción de las cadenas que están en el eje x y el eje Y; la intersección con el cuadro rojo indica las mayores interacciones entre las secuencias de ARN. El cálculo de la energía de interacción en la hibridación está asociado a la energía libre mínima a una temperatura de 310 Kelvin. De acuerdo con la energía, habrá hibridaciones que no es tan probable que interactúen pero que aún se contemplan. Por lo tanto, en la figura podemos

ver una representación esquemática de los intervalos y sitios hibridados entre las secuencias. Asimismo, la tabla 4 resume los intervalos de las secuencias donde es más probable que ocurran interacciones o hibridaciones con su respectiva energía en calorías y kBT.

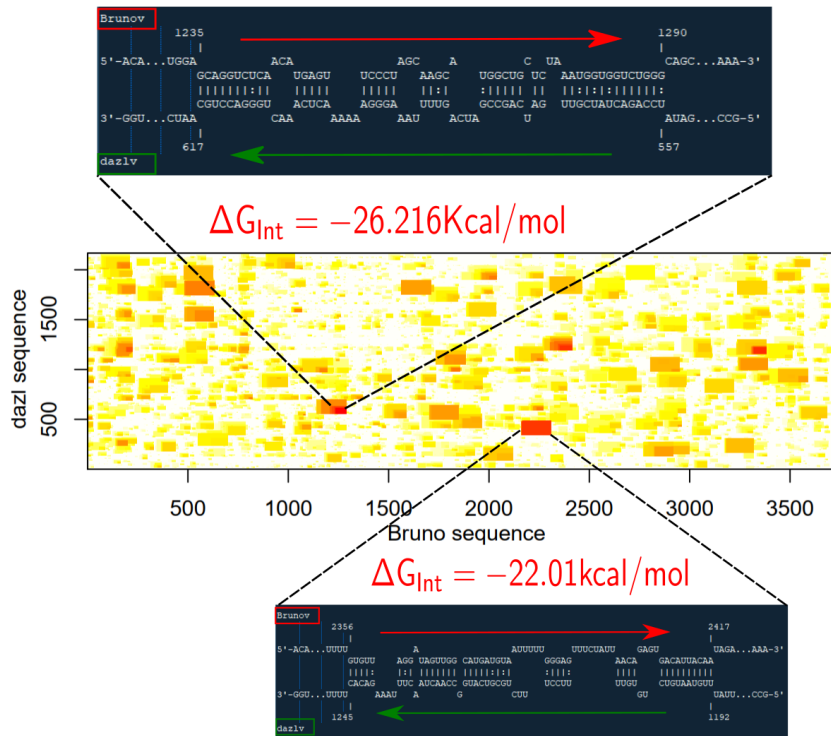


FIGURA 20. Representación esquemática de los sitios de hibridación entre las secuencias de RNA (Dazl y Bruno) y la energía libre de interacción.

Entre las secuencias de RNA del germoplasma vegetal y animal se identificaron interacciones que alcanzan valores promedios entre 20.000 y 23.000 kcal/mol. No obstante, se evidencio que entre las mismas especies de RNA existen mayor afinidad de hibridación. Sin embargo, interacción Nano-Nano fue la que presento menor energía de hibridación que las demás. A diferencia de las hibridaciones entre Bruno y Vasa que obtuvieron mayor energía de interacción alcanzando valores de 40.000 kcal/mol. Indicando consigo que las interacciones entre dos cadenas de RNA de los germoplasmas como animal o vegetal se debe a hibridaciones de las secuencias, el cual permitir la formación de complejos con proteínas.

Sec 1	Sec 2	Intervalo Secuencia 1	Intervalo Secuencia 2	Energía	Unidad	Energía	Unidad
Bruno	Dazl	1235 -- 1290	557 -- 617	-26,2162	kcal/mol	-4,25E+01	kBT
		3320 -- 3385	1156 -- 1226	-22,7185	kcal/mol	-3,69E+01	kBT
		2356 -- 2417	1192 -- 1235	-22,0099	kcal/mol	-3,57E+01	kBT
Bruno	Dnd	3034 -- 3113	1641 -- 1716	-23,140	kcal/mol	-3,75E+01	kBT
		3280 -- 3362	1650 -- 1730	-22,435	kcal/mol	-3,64E+01	kBT
		1269 -- 1373	811 -- 888	-22,195	kcal/mol	-3,60E+01	kBT
Bruno	Nanos	2112 -- 2253	429 -- 546	-22,106	kcal/mol	-3,59E+01	kBT
		772 -- 807	635 -- 675	-21,696	kcal/mol	-3,52E+01	kBT
		2913 -- 2979	222 -- 288	-21,005	kcal/mol	-3,41E+01	kBT
		2326 -- 2467	758 -- 905	-20,288	kcal/mol	-3,29E+01	kBT
Bruno	Vasa	1070 -- 1194	510 -- 655	-41,202	kcal/mol	-6,69E+01	kBT
		2067 -- 2208	253 -- 392	-40,358	kcal/mol	-6,55E+01	kBT
		1825 -- 1968	249 -- 390	-39,673	kcal/mol	-6,44E+01	kBT
Bruno	Bruno	104 -- 235	1830 -- 1977	-29,317	kcal/mol	-4,76E+01	kBT
		1830 -- 1911	151 -- 235	-24,280	kcal/mol	-3,94E+01	kBT
Dazl	Dnd	1149 -- 1228	1389 -- 1457	-22,286	kcal/mol	-3,62E+01	kBT
		362 -- 506	1217 -- 1365	-18,975	kcal/mol	-3,08E+01	kBT
		1158 -- 1228	1389 -- 1457	-18,779	kcal/mol	-3,05E+01	kBT
Dazl	Nanos	1161 -- 1212	1093 -- 1141	-19,626	kcal/mol	-3,18E+01	kBT
Dazl	Vasa	1489 -- 1595	315 -- 461	-34,444	kcal/mol	-5,59E+01	kBT
		1930 -- 2078	464 -- 612	-31,365	kcal/mol	-5,09E+01	kBT
Dazl	Dazl	1819 -- 1964	1808 -- 1955	-19,935	kcal/mol	-3,23E+01	kBT
		1126 -- 1163	875 -- 913	-19,529	kcal/mol	-3,17E+01	kBT
		506 -- 648	495 -- 640	-19,439	kcal/mol	-3,15E+01	kBT
Dnd	Nanos	1604 -- 1710	267 -- 380	-26,700	kcal/mol	-4,33E+01	kBT
Dnd	Dnd	434 -- 568	295 -- 402	-21,150	kcal/mol	-3,43E+01	kBT
		295 -- 361	497 -- 568	-19,456	kcal/mol	-3,16E+01	kBT
		65 -- 111	835 -- 877	-18,828	kcal/mol	-3,05E+01	kBT
Nano	Vasa	13 -- 68	561 -- 623	-27,702	kcal/mol	-4,49E+01	kBT
		532 -- 669	167 -- 309	-24,092	kcal/mol	-3,91E+01	kBT
		900 -- 1048	414 -- 535	-23,689	kcal/mol	-3,84E+01	kBT
Nano	Nanos	495 -- 639	742 -- 869	-18,859	kcal/mol	-3,06E+01	kBT
		127 -- 190	590 -- 650	-18,564	kcal/mol	-3,01E+01	kBT
		1007 -- 1041	105 -- 135	-17,650	kcal/mol	-2,86E+01	kBT
		105 -- 143	998 -- 1041	-17,485	kcal/mol	-2,84E+01	kBT
Vasa	Vasa	520 -- 643	2155 -- 2302	-37,076	kcal/mol	-6,02E+01	kBT
		2139 -- 2278	537 -- 663	-35,637	kcal/mol	-5,78E+01	kBT

Tabla 4. Intervalos de hibridación entre las secuencias de ARN germoplasma animal y vegetal con su correspondiente energía de interacción en Kcal/mol.

El análisis de las proteínas se da partir Buc y Buc2l. Las cuales a través de cada secuencia y del uso de algoritmos predictivos como PONDR, se encontró qué las regiones o conjuntos de aminoácidos que tienen mayor probabilidad de presentar una estructura desordenada. Para los análisis desarrollados se utilizaron los siguientes predictores; PONDR® VL-XT las cuales tienen en cuenta regiones desordenadas largas caracterizadas de diversas formas y regiones desordenadas terminales caracterizadas por rayos X. Al hacer predicciones, las salidas de NNP están entre 0 y 1. Si el valor de un residuo supera o coincide con un umbral de 0,5, el residuo se considera desordenado. Por lo tanto, se puede observar en las figuras 21 y 22 la representación esquemática de los aminoácidos desordenados y ordenados correspondiente a las proteínas Buc y Buc2l

1	MEGINNNSQP	MGVGQPHHPV	NHTRPFFVYQ	PFSQPYFMYQ	WPMNPYGHYG	80
51	FPGPALHFGR	PYMAFYQFMQ	YPGYVIPHAP	MQPIDYRRIN	PHYPSVASVD	100
101	LRVRHHFQNA	GMHRETACSE	VQTDPSDSVN	KLIDKIESLK	ACELGSDKGP	150
151	NNVVSSTFDV	VQGEKLTRLN	EDSNLEVATK	ECKEDPVTRP	TTYSDSAYDA	200
201	ESSQGRLEDC	VFSDVLPDLS	SSVHEEEEE	EKDVNEDEP	QTVADEICSQ	250
251	NEMSASTTSN	VFCSGVQZIA	DPTECHDLEK	LGDEQKQDIP	SADAAAVIEP	300
301	LISLSEDFDL	PYQILRLPCN	KTTTGLSLER	EIDPLVYFDS	PSTLLPPQNY	350
351	LSSIGSAYS	SYYPQVTQER	QSVLSPSIDE	LSSRDEMFST	DVEDLEVVP	400
401	HVYVGGGRLA	EASDMFVRSR	KELPSVDKTC	SVCQKTCACC	GSTLQDEVGM	450
451	CKMAEHSHPF	RDEMSDQDCD	YDLEAEVRSN	CESPRVSKRK	CCSRHALPSC	500
501	GHHCAKRRHR	KLLCEGQESC	DLREQARVHP	KGECCEEYGA	LAKADKRIQK	550
551	GALCRPCIEQ	QWREGVVSQD	ENWASCGAKF	RSWRQVTGFP	DQGRTPLRSS	600
601	TCKSIHQRRP	RSEYNDYDET	EFTYCQRGRG	SMKRGSR		650

FIGURA 21. Caracterización de regiones de residuos desordenados y ordenados de la proteína BUC. Los aminoácidos de color negro son configuraciones ordenadas. Los aminoácidos de color rojos tienen configuración desordenada.

De acuerdo con los resultados hallados los aminoácidos de color negro son aquellos que tiene una configuración ordenada. Sin embargo, los de color rojo son aquellos que tiene mayor probabilidad de presentar una configuración desordenada. Sin embargo, la proteína Buc presenta un conjunto mayor de aminoácidos que pueden presentar una estructura tridimensional desordenada en comparación con BucL2.



1	<b>MAAISTHEPI</b>	<b>GGLSGVLSEF</b>	<b>LHAQAPLCAG</b>	<b>FLPSRILHFD</b>	<b>EQLQLNQQFY</b>	50
51	VFTVQPFMPY	QFPTSSLCVP	YSGFHGLGYS	VMPLQLPSCV	EVPGFITPQL	100
101	QMLDYRRITP	HVAPATAHRT	<b>RHSNFQNFVP</b>	<b>LGPVMVNSEV</b>	<b>QTEPICHSIG</b>	150
151	<b>SHGPAARSDT</b>	<b>CSESGRGTGC</b>	<b>DSPVSTTPRS</b>	<b>TENKSIACPE</b>	<b>GVSVLTQNCN</b>	200
201	<b>TSATETTDLS</b>	TVPKSEIIQA	EQVQSKCGVL	LSKHEIIQNK	<b>NNKMASYNKN</b>	250
251	DIM <b>QCSLGSV</b>	<b>QSSQEDVVVCS</b>	YRSLAFREDK	<b>QRVEAGMLSY</b>	<b>SRKHHLTSTCT</b>	300
301	ELNVLRTFSC	TLKKISKPSK	SVIA <b>QCVKVK</b>	<b>NNSETQFQTK</b>	<b>DKDDGNSFNN</b>	350
351	NFKILRLPFD	VQNHKLPCQL	DASIWSVESL	LPYVPSSEWL	AENGLLTPQK	400
401	TLSSSLTKPSN	VISKPYCVSV	GKNL <b>QSGPSA</b>	KYCVSPHQLE	ASIWSIDSLM	450
451	PYVPSNEWMV	ENG <b>FSTPQKG</b>	<b>VSPQIKSSDG</b>	<b>HSSTDGIPLR</b>	<b>KNLQACGSTK</b>	500
501	YHGVSQLEA	SIWSVDSLIP	YVFPNELMVD	<b>NGFSTPQKSV</b>	<b>SPQIKSSDGD</b>	550
551	<b>FSTDGIPLRK</b>	NLQTCGSTKY	HRVSHQLEAS	IWSVDSLMPY	VPSNELMVDN	600
601	<b>GFSTPQKGIS</b>	<b>PQIKSSDGYS</b>	<b>STDGIPLKKN</b>	LQTCGSAKYH	GVSHQLEASI	650
651	WSVESLKPYP	PSNGYVDNG	YLTSH <b>QALSP</b>	<b>FFKSSNVVSE</b>	<b>FNQTSDDHFQ</b>	700
701	<b>VGERLQTCAS</b>	<b>AKLKGSIKSL</b>	<b>EKRSPYRPSS</b>	SWLADFGNVY	YYSKLPFVQQ	750
751	QSEILERWQP	KVSLTTHLEP	SFINL <b>KKTRN</b>	<b>QTPATGISDQ</b>	KHCVSNKCKC	800
801	KVKRSPRLSC	VTLESKLEPLC	QSCRY <b>DTKGK</b>	<b>DSKKHSDCPG</b>	<b>CKKEDYVEVK</b>	850
851	TFDKISNEGV	LDAKNMALGN	LLPECCGAAR	VELRQQTALT	<b>SPAQDKCGDL</b>	900
901	EWRLWEKKKD	<b>FMNQNKSIW</b>	<b>SLQKENAAKE</b>	<b>SMQASISVRR</b>	<b>IENKESMLTR</b>	950
951	<b>NRTRGEHVTG</b>	<b>RE</b>				1000

FIGURA 22. Caracterización de regiones de residuos desordenados y ordenados de la proteína Buc2L. Los aminoácidos de color negro son configuraciones ordenadas. Los aminoácidos de color rojo tienen configuración desordenada

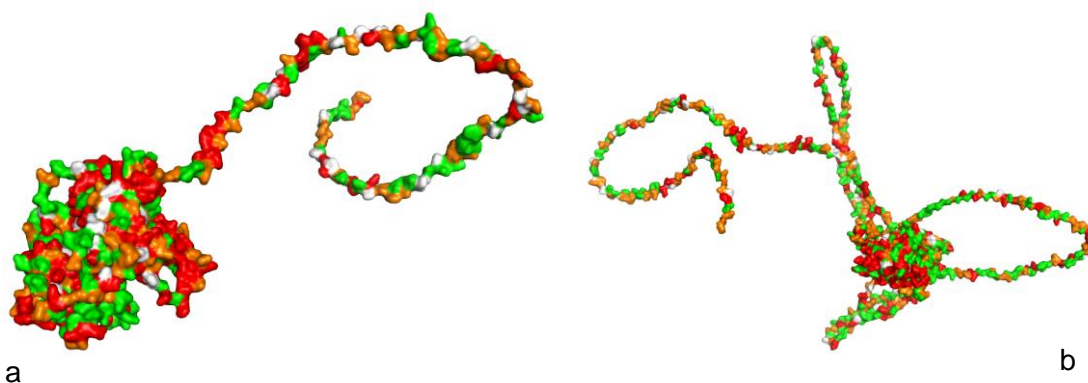


FIGURA 23. Modelo tridimensional de las proteínas: a) Buc y b) Buc2

Posteriormente, se hizo un modelamiento por homología de cada proteína para obtener un modelo tridimensional de cada una de estas y así identificar las regiones antes mencionadas como el tipo de aminoácidos presente en cada estructura como hidrofóbicos (naranja), cargados (rojos) y polares (verde). Identificado consigo la carga total de cada una de ellas. No obstante, de acuerdo con los resultados obtenidos del RNA, asumimos que la formación de los RNPs se debe a las interacciones electrostáticas entre las proteínas en cada cadena de RNA. Por consiguiente, de acuerdo con las estructuras reportadas a nivel experimental se propone un modelo icosaédrico con una forma similar a un ovoide. A continuación, se construyó la agregación de RNPs de manera helicoidal

entre ellas, basada en una función analítica, parametrizando la ubicación del centro de masa de cada partícula RNP en un cuadro periódico, para así controlar todos los parámetros geométricos de la hélice a través de esta función, como el paso, el radio y la altura, permitiendo consigo modificar la cantidad de RNP que construyen la formación helicoidal. En nuestro modelo, asumimos que cada RNP (radio 100 nm) está formada por 43 beads cargadas, garantizando la electroneutralidad a pesar de la evidente segregación de carga en la superficie de las partículas. A sí mismo, el modelo es lo suficientemente general como para incluir una segunda hélice, que se entrelaza con la otra, para imitar el comportamiento observado para las hélices RNPs del germoplasma animal y vegetal.

Para determinar la conformación RNP con una distribución de carga superficial particular que permite la agregación helicoidal, desarrollamos una simulación de Monte Carlo que encuentra la configuración de energía del estado fundamental permitiendo la segregación de carga superficial y así modificar la agregación local de las RNPs. Por lo tanto, el funcional de energía libre, expresado como:

$$H_T = H_E + H_v$$

Donde  $H_E$  es la energía electrostática en  $K_B T$  asociada al potencial Debye-Huckel anteriormente mencionado y la energía de volumen excluido está dada por  $H_v$  de la siguiente manera:

$$H_v = \begin{cases} \frac{1}{K_{BT}} \left\{ 4\varepsilon \left[ \left( \frac{\sigma}{r} \right)^{12} - \left( \frac{\sigma}{r} \right)^6 \right] + \varepsilon \right\}, & r \leq 2\frac{1}{6}\sigma \\ 0, & r > 2\frac{1}{6}\sigma \end{cases}$$

Donde  $\varepsilon$  y  $\sigma$  son la energía y tamaño característicos para la energía repulsiva de Lennard-Jones, y  $r$  es la distancia entre centros de masa de las partículas RNP. Una vez calculada la energía total para la configuración inicial, desarrollamos una estrategia de muestreo en el espacio de configuración para obtener la carga y configuración espacial que minimiza la energía libre. El procedimiento propuesto comienza con la selección aleatoria de una RNP y una selección aleatoria de un método de muestreo. Hay tres movimientos posibles en el espacio de configuración para lograr la energía de estado mínima: traslación, rotación y redistribución de carga. Si se selecciona el muestreo de traslación, proponemos un

desplazamiento aleatorio de toda la partícula RNP en un intervalo dado, verificando las condiciones de superposición. Si se elige el muestreo de rotación, proponemos una rotación aleatoria de la RNP a través de un jacobiano en el espacio euclidiano para cada bead en la partícula, y si se selecciona la redistribución de carga, modificamos la carga de un par de beads aleatorias, sumando y restando una cantidad de carga aleatoria para garantizar la electroneutralidad. La configuración propuesta genera un cambio local en la energía libre. Si la diferencia entre la energía anterior de la configuración propuesta y la energía anterior es menor o igual a cero, aceptamos el cambio:

$$\Delta H_T \leq 0$$

Si el cambio de energía es mayor que cero, aceptamos la nueva configuración con una probabilidad dada por el Criterios de Metrópolis, es decir:

$$\Delta H_T > 0; (old \rightarrow new) = \min \left[ 1, \exp \left( -\frac{\Delta H_T}{k_B T} \right) \right]$$

Luego, repetimos el procedimiento un número N de pasos de Monte Carlo hasta que el sistema alcance una configuración de energía mínima.

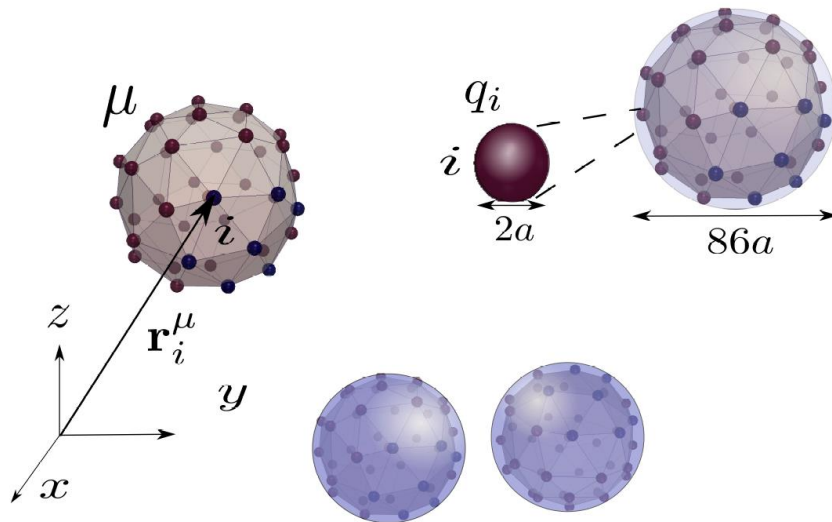


FIGURA 24. Representación esquemática de una partícula de RNP.

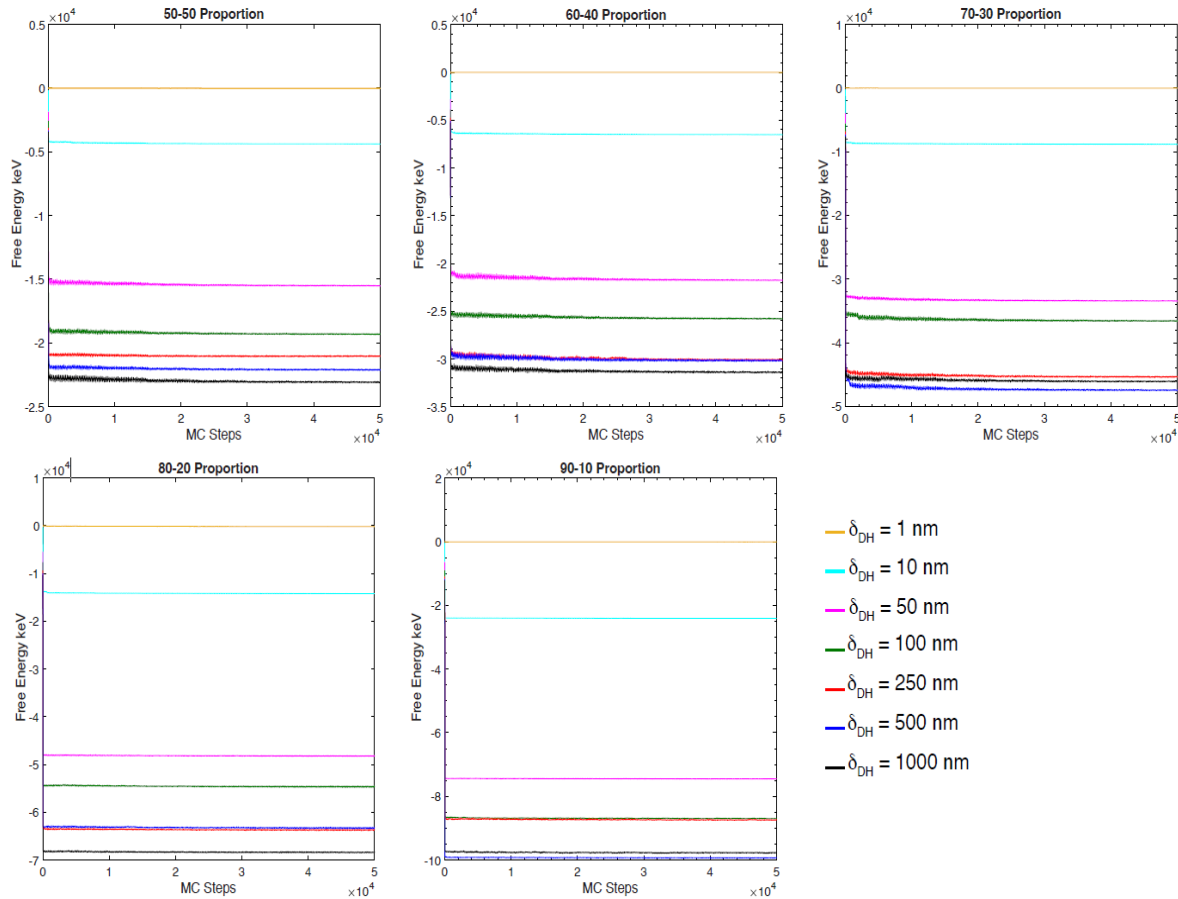


FIGURA 25. Energía libre en keV con cargas positivas y negativas distribuidas en proporciones diferentes 50:50, 60:40, 70:30, 80:20 y 90:10 respectivamente, variando consigo la longitud de Debye de en 1, 10, 50, 100, 250, 500, 1000 nm.

Las configuraciones de la distribución de carga de los RNPs son de manera positiva y negativa. En donde aumenta la proporción de la carga positivas mientras se reduce la proporción de la carga negativa, consiguiendo consigo distribuciones que permitan una focalización de interacción espacial entre ellas para la formación helicoidal de los RNPs. En la proporción 50:50 entre cargas positivas y negativas; a medida que la longitud de Debye variaba a mayor distancia, la energía disminuía inversamente proporcional a cada longitud, alcanzando un valor mayor a 1 nm y un valor menor a 1000 nm. No obstante, las energías presentaban un comportamiento constante a través de los pasos de Monte Carlo. En la proporción 60:40 tiene un comportamiento similar al de 50:50. Sin embargo, a longitudes de 250 y 500 nm alcanzan valores similares de energía libre próximos a -3 keV. Por otro lado, en la proporción 70:30, la energía libre alcanza valores similares para las longitudes de 250 y 1000 nm alrededor de -4.4 keV. A su vez, a 500 nm alcanza la menor

energía libre con un valor de  $-4.5$  keV. En la proporción 80:20, para longitudes de Debye de 250 y 500 nm alcanzan valores equivalentes en  $-6.5$  keV. Por otro lado, la proporción 90:10, la longitud 100 y 250 presentan valores análogos alrededor de  $-8.3$  keV. Las tendencias energéticas de cada distribución disminuían a medida que la longitud de Debye aumentaba. De igual manera, a medida que la distribución positiva era mayor a la negativa la energía de los sistemas alcanzaba valores menores respecto a su sucesor. Se evidenció que a longitud de Debye igual a 1 nm la energía libre fue igual a 0 para todas distribuciones de cargas.

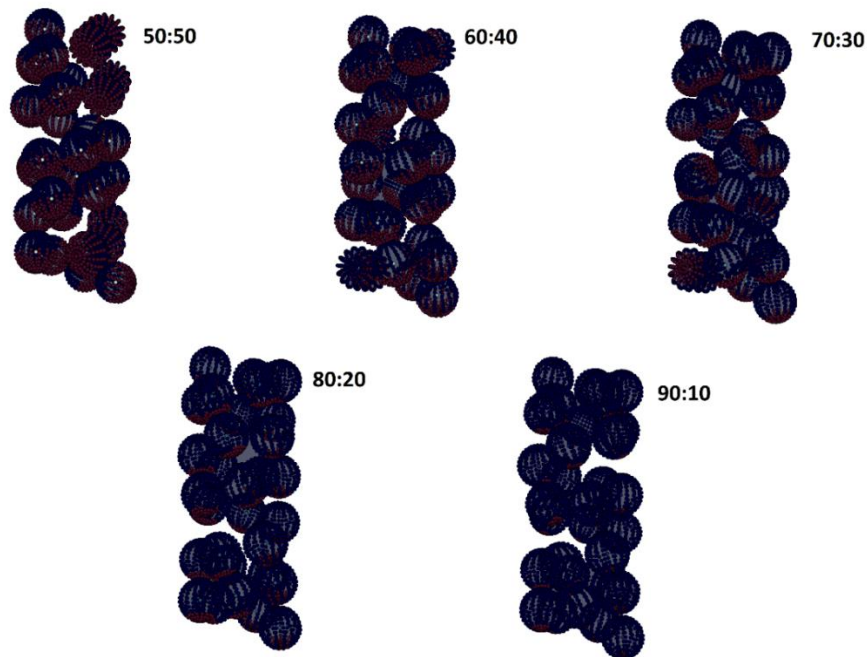


FIGURA 26. Representación esquemática de la formación helicoidal entre RNPs acorde a la distribución de carga positiva y negativa; 50:50, 60:40, 70:30, 80:20 y 90:10 respectivamente.

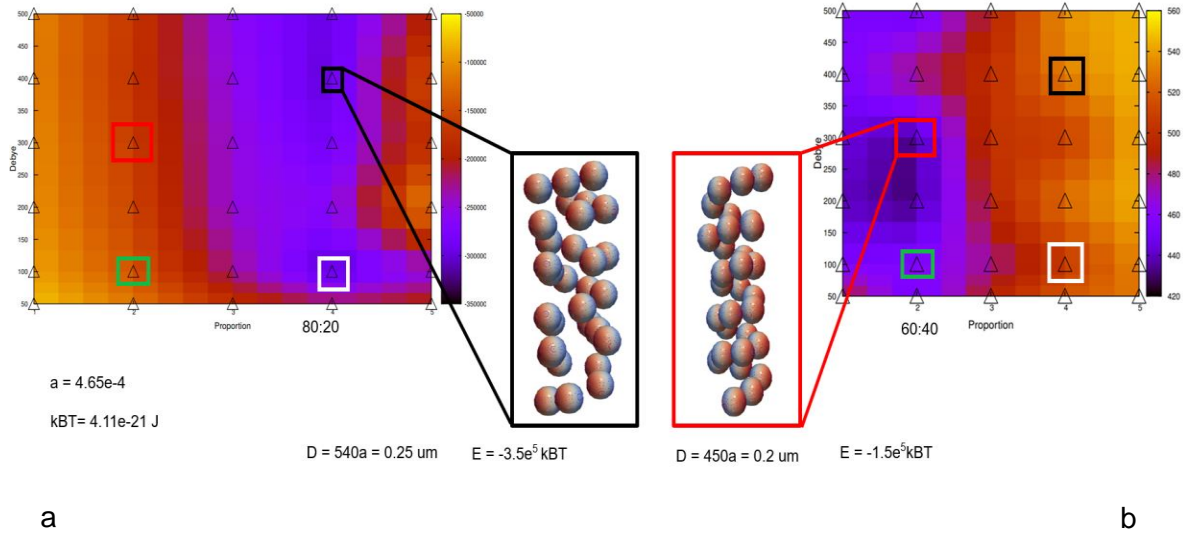


FIGURA 27. El diagrama izquierdo (a) representa el mapa de calor asociado a la energía de acuerdo a longitud de Debye (eje y) respecto a la proporción en cargas (eje x). En el diagrama de la derecha (b) representa el mapa de calor asociado las distancia promedios de las configuraciones de hélices respecto a cada bead; En el eje x está la proporción de cargas y en el eje y la longitud de Debye.

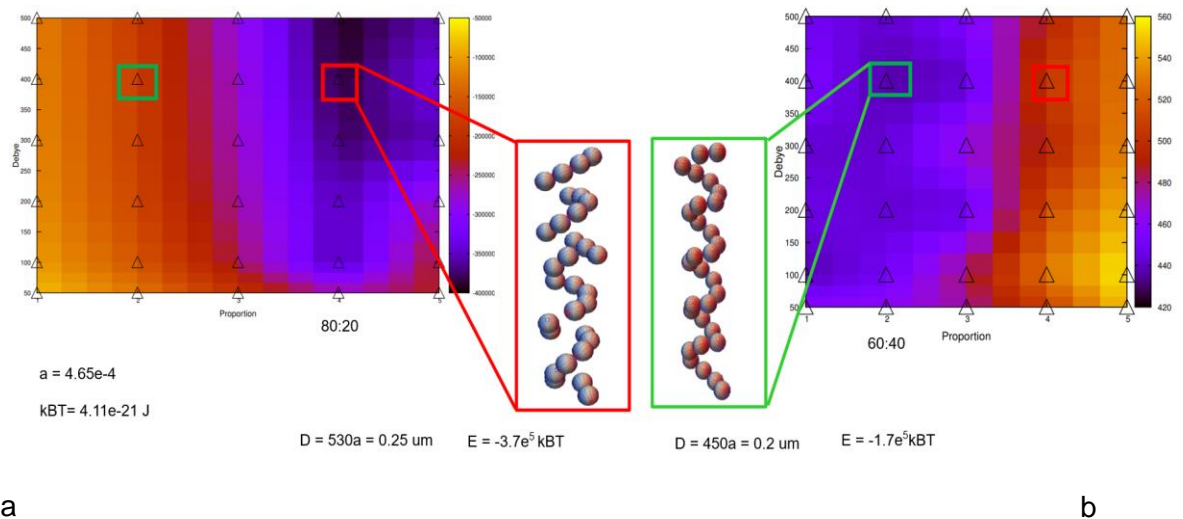


FIGURA 28. El diagrama izquierdo (a) representa el mapa de calor asociado a la energía de acuerdo a longitud de Debye (eje y) respecto a la proporción en cargas (eje x). En el diagrama de la derecha (b) representa el mapa de calor asociado las distancia promedios de las configuraciones de hélices respecto a cada bead; En el eje x está la proporción de cargas y en el eje y la longitud de Debye.

En la figura 27 y 28 se realizó un análisis a las simulaciones antes ejecutadas cambiando el pitch de la hélice de los RNPs; en donde se identificaba la mínima energía de cada conformación de acuerdo a la proporción de la distribución de las cargas (50:50, 60:40, 70:30, 80:20, 90:10) y la variación de la longitud de Debye (100, 200, 300, 400, 500). Del mismo modo, se realizó un procedimiento similar; identificando las distancias promedios de las configuraciones de cada bead a través de toda la simulación de acuerdo a la proporción de la distribución de las cargas y la variación de la longitud de Debye anteriormente expuestas.

El análisis anteriormente descrito tuvo la finalidad de encontrar una configuración de menor energía libre del sistema es decir encontrar un equilibrio global o local que permitiera optimizar la distribución de las cargas y la morfología helicoidal de los RNPs (estructuras más estables) para así implementar la dinámica browniana y dilucidar el mecanismo de agregación de las partículas en el sistema. Aunque los resultados obtenidos del Monte Carlo no permitieron obtener una configuración más estable dado variaciones de las variaciones de las cargas y sus respectivas distribuciones, la energía electrostática es tan grande que las partículas no conservan la forma helicoidal al final de los pasos de Monte Carlo. Por lo tanto, se propuso un modelo con 8 partículas de RNPs para con una configuración de hélice, para analizar la evolución del sistema mediante la simulación browniana; con una carga negativa máxima ( $q_{mx}$ ) y una proporción de cargas por RNPs de 75:25 cuya caja de simulación es 15 veces el tamaño del RNP, y una longitud de Debye adimensional aproximada de 0.96, mientras que el número de Bjerrum es del orden del tamaño del RNPs.



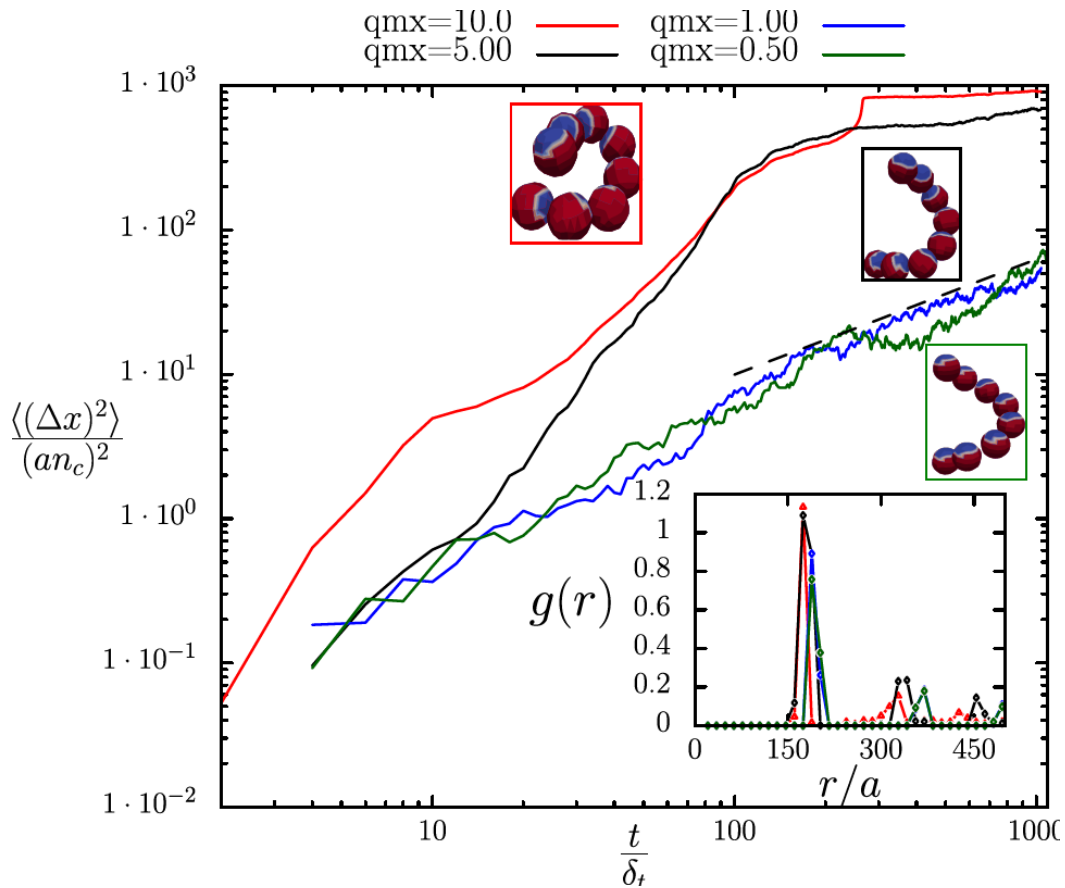


FIGURA 29. Función de distribución radial (rdf) y desplazamiento cuadrático medio (msd) para para un sistema de 8 RNPs con diferentes cargas.

En la figura 29 se observa el desplazamiento medio cuadrático (msd) y las funciones de distribución radial (interior de la figura) para una cadena de 8 RNPs cuyas distribuciones de cargas positivas y negativas (beads) es la misma 25-75%. Por otro lado, el valor  $q_{mx}$  es la carga negativa constante de cada bead. A Cargas mayor de un orden de magnitud (curva solida negra) el sistema tiende a aglomerarse por medio de las interacciones electrostáticas en cada bead de los rnp. Sin embargo, a partir de 200 pasos de tiempo adimensional el sistema alcanza la formación de una estructura similar, la cual se puede observar en el interior de imagen, enmarcada con bordes de color negro. La dinámica del msd da como resultado un comportamiento anómalo característico de sistemas agregados y que eventualmente lograría obtener un rango difusivo para tiempos muy largos.

Por otro lado, a medida que la carga disminuye y se acerca a la unidad (líneas negra y azul) la agregación de los RNPs disminuye y por tanto el comportamiento de las curvas



pasa de ser un comportamiento anómalo para aproximarse al de una curva para una suspensión difusiva. Los efectos estéricos entre los RNPs son mayores al efecto promedio de las interacciones electrostáticas entre los beads, esto hace que la tendencia de la curva del msd tienda a ser lineal  $x^n$ , donde  $n=0.85$ . Del mismo modo, Al interior de la figura, la en función de distribución radial puede apreciarse que para un sistema cuyas cargas negativas sean mayores a un orden de magnitud, los primeros vecinos están cerca de un valor a 150 el cual es equivalente al diámetro de un solo RNP ( $2nc$ ). Mientras que cuando disminuye la carga la distancia de los RNPs es mayor, sin embargo, la estructura de los primeros y segundos vecinos es similar a la de un sólido, es decir cuyas distancias promedio son equivalentes y constantes.

## 5.4 Conclusiones

Si bien el análisis de interacciones entre los diferentes RNAs evidencio posibles hibridaciones entre dos cadenas de ácido nucleico monocatenario, las condiciones de hibridación dependen de la solución, la temperatura, la formación de energía y la estabilidad del enlace de pares de bases complementarias. Sin embargo, debido a los sitios de interacción dados, no se puede concluir que solo la hibridación de las cadenas permita la generación de aglomerados estables. Por lo tanto, las interacciones electrostáticas y no covalentes permiten un autoensamblaje más estable de las RNPs, considerando como hipótesis que el proceso o mecanismo agregado está asociado a interacciones entre el ARN y las proteínas.

El método de Monte Carlo se fundamentó en encontrar la estructura más optima equilibrada del sistema helicoidal de los RNPs. Sin embargo, en las diferentes proporciones de cargas implementadas; se evidenció que los sistemas quedaban atrapados debido a las interacciones electroestáticas entre las partículas. Esto quiere decir, que las interacciones electrostáticas de Debye-Huckel era más fuertes que el volumen excluido de Lennard Jones, Por consiguiente, es necesario mejorar el modelo

propuesto para así encontrar un balance entre las fuerzas que rigen el sistema y así minimizar el sistema como por ejemplo: la longitud de Debye-Huckel y el volumen excluido.

El modelo discreto en forma de icosaedro planteado como representación de los RNPs, permitió tener cargas puntuales y con volumen excluidos, facilitando las interacciones entre las partículas a medida que avanza la evolución del sistema, mediante proporciones de cargas 75:25. Aunque se evidencia la conservación de la estructura helicoidal, el tiempo de simulación no es suficiente, dado que el tiempo de evolución por bead es muy lenta. Por consiguiente, no se ha alcanzado un régimen completamente difusivo del sistema, por ende, no sé puede caracterizar completamente la estructura. A su vez, es importante integrar la estructura minimizadas por Monte Carlo y realizar más ejecuciones *in Silico* de la dinámica browniana y así hallar una estructura difusiva que permita caracterizar y dilucidar completamente la formación de estructuras helicoidales de los RNPs.

## 6. Análisis Retrospectivo Dengue

El Dengue es una enfermedad viral la cual es transmitida por la picadura de mosquitos hembra infectadas (*Aedes Aegypti*) y el virus causante de dicha enfermedad es conocido como DENV, existen cuatro serotipos de dengue (DENV-1, DENV-2, DENV-3 y DENV-4) perteneciente al género *Flavivirus* de la familia *Flaviviridae*. Los principales síntomas asociados a la enfermedad son: fiebre, dolor en los huesos, dolor en las articulaciones, y dolor de cabeza. La Organización Mundial de la salud (OMS) clasifica la enfermedad según los síntomas de tres maneras; Dengue sin signos de alarma caracterizada por fiebre, náuseas, rash, mialgias, artralgias, leucopenia. Dengue con signos de alarma presenta dolor abdominal, vómitos, acumulación del líquido, sangrado, aumento de hematocrito con disminución de plaquetas. Dengue grave caracterizada por extravasación severa del plasma, hemorragias severas y daño grave en órganos.

Colombia es considerada una región endemo-epidémica para dengue debido a su ubicación en la zona del trópico y es considerada una enfermedad de gran importancia en la salud pública del país debido a sus ciclos de brotes tría-anual, y la circulación de los cuatro serotipos por medio dispersión del mosquito de *Aedes* han presentado gran cantidad de casos en las regiones nacional. Por consiguiente, el Instituto Nacional de Salud (INS) realiza vigilancia epidemiológica para prevención y control de la enfermedad mediante la notificación, registros individuales en el Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA) consolidando el seguimiento a nivel clínico los casos de dengue. La presencia del ciclo de vida Mosquito *Aedes*, está asociada a condiciones ambientales como; temperatura, humedad, precipitación, las cuales permite el desarrollo del vector en un periodo de 7 a 10 días mediante cuatro etapas como se puede evidenciar en la siguiente gráfica [54].

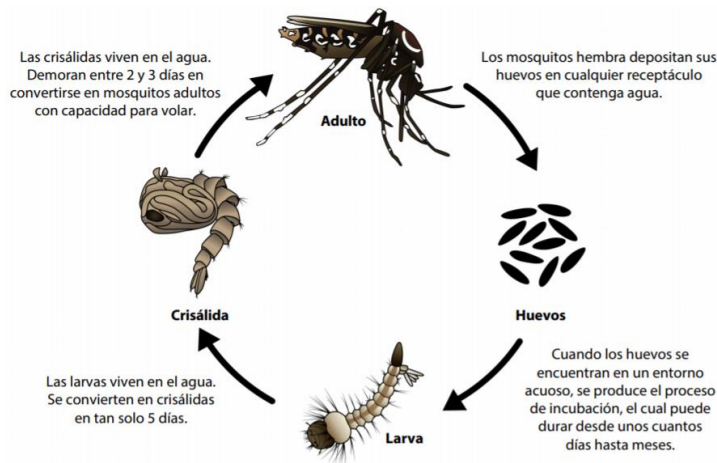


FIGURA 30. Ciclo de vida del mosquito *Aedes aegypti*[54].

Por lo anteriormente expuesto, es necesario crear herramientas de análisis y de predicción de los brotes de dengue en la región en Colombia que permitan a las entidades de salud estar preparadas, planificar o crear planes de contingencias y de intervención a posibles infecciones ocasionadas por la transmisión del virus por la picadura del mosquito. Por consiguiente, diferentes modelos matemáticos, estadísticos y de inteligencia artificial a través del uso de las variables anteriormente descritas, analizan los factores y los fenómenos que ocasionan los casos de dengue, cuyos resultados obtenidos son implementados en promover el entendimiento de los ciclos anuales de la aparición de la enfermedad.

## 6.1 Áreas de Estudio

Las ciudades para este estudio fueron elegidas acorde a la prevalencia de los casos de dengue en el tiempo, las condiciones climáticas y geográficas de las ciudades, cómo su ubicación a nivel del mar menor o igual a 2000 metros de altura. Por consiguiente, Cali, Medellín, Santa Marta, Cúcuta, Barranquilla y Villavicencio fueron las ciudades escogidas para realizar el análisis retrospectivo. Para analizar así los casos de dengue en el tiempo de cada una de ellas y como se ven afectados según su ubicación geográfica de acuerdo a las variables climáticas como índices oceánicos.

## 6.2 Datos de Dengue

Los casos de dengue fueron extraídos del portal del Sistema Nacional de Vigilancia en Salud Pública (SIVIGILA) reportados de manera semanal clasificados de la siguiente manera: Caso probable, confirmado por laboratorio y confirmado por nexo epidemiológico. Los casos probables son aquellos cuyos pacientes proceden de áreas endémicas que cumple con la definición de dengue con o sin signo de alarma. Los casos confirmados por laboratorio son aquellos pacientes los cuales son confirmados por algún criterio de diagnóstico por laboratorio para dengue: PCR para pacientes con menos de cinco días de inicio de fiebre, prueba de IgM con pacientes con 5 o más días de inicio de fiebre. Los casos confirmados por nexo epidemiológico consisten en confirmar los casos probables de dengue de acuerdo con los casos confirmados por laboratorio utilizando la asociación temporal y espacial de la ubicación de la persona. Todos los casos están reportados teniendo en cuenta la semana epidemiológica de la ocurrencia, grupo de edad, sexo, etnia, sistema de salud, lugar de procedencia, estado de hospitalización y clasificación del caso en una ventana de tiempo del año 2008 hasta el año 2020.

## 6.3 Datos climáticos

La información climática es obtenida de la plataforma web del Instituto de Hidrología, Meteorología Y Estudios Ambientales (IDEAM) durante el año 2008 hasta el año 2020 de las ciudades de interés. Los datos históricos del clima están reportados en tiempo de minutos de acuerdo con las estaciones de vigilancia de las ciudades. Entre los datos más relevantes para este análisis retrospectivo son: temperatura media, temperatura máxima, temperatura mínima, humedad mínima, humedad media, humedad máxima y precipitación. Por consiguiente, la base de datos se le realizó un preprocesamiento de estos; removiendo consigo mediciones anómalas, agregaciones en el tiempo de manera diaria, semanales y mensuales.

## 6.4 Datos del Niño

Los datos asociada a los fenómenos del Niño y la Niña se obtuvieron de la plataforma web The National Oceanic and Atmospheric Administration (NOAA) desde el año 2008 hasta el año 2020 [55]. Las variables asociadas a dichos fenómenos son: La temperatura superficial

del mar de las cuatro regiones de franja tropical de pacífico; Occidental (Región Niño 4), central (Región Niño 3), centro-occidental (Región Niño 3-4) y la oriental (Región Niño 1+2) con el fin de hacer seguimiento y vigilancia de los ciclos ENOS. Además, de los índices oceánicos del niño, los índices multivariado del ENSO y el índice transatlántico del niño. Los datos, son reportados de manera mensual, reorganizados de manera semanal.

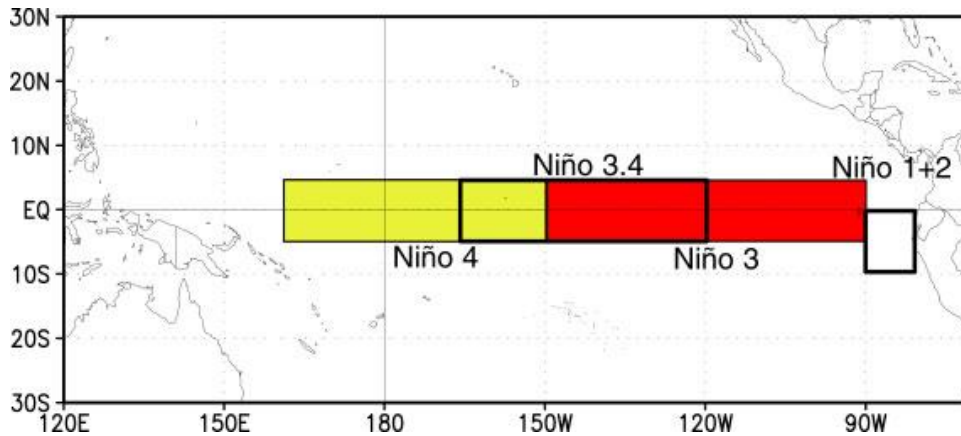


FIGURA 31. Esquema de las cuatro regiones del niño en océano pacífico[55].

## 6.5 Metodología

A partir de la información obtenida de la base de datos SIVIGILA, IDEAM, NOA se hizo un análisis de preprocesamiento y de exploración; identificando anomalías, continuando con la limpieza, selección, tratamiento de valores perdidos y transformación de los datos. Posteriormente, los datos climáticos son agregados de manera diaria, semanal y mensual, los datos provenientes del fenómeno del niño son reorganizados de mensual a semanal, y finalmente los casos de dengue son adheridos de acuerdo con su reporte semanal. Para cada conjunto de datos de la ciudad se realiza un análisis de correlaciones entre las diferentes variables mencionadas y cómo éstas afectan la incidencia de los casos de dengue en el tiempo.

## 6.6 Resultados y análisis

A continuación, se presenta los resultados asociados a la cantidad de casos semanas por cada ciudad durante los años del 2008 hasta comienzos del 2020.

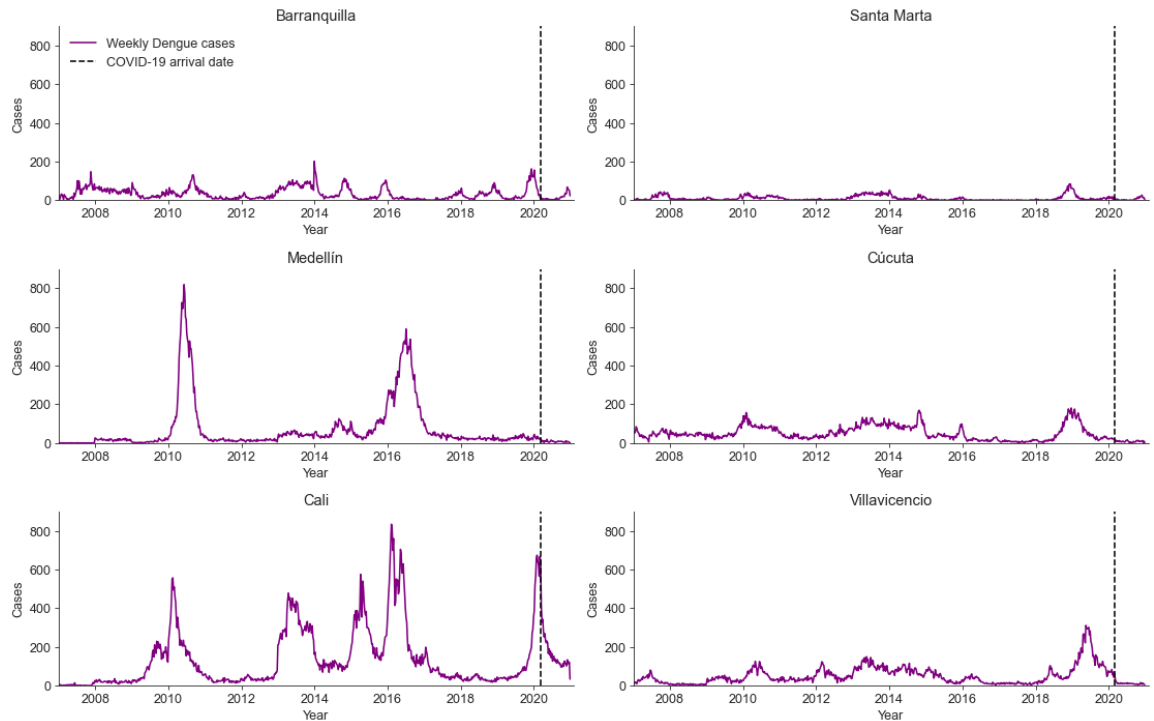


FIGURA 32. Casos de dengue semanales a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés: Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio.

Medellín durante el año 2010 y 2016 presento dos grandes brotes de dengue mayores a 300 casos reportados, alcanzando 800 y 600 casos semanales respectivamente durante eso dos años. Asimismo, la ciudad de Cali durante el intervalo de tiempo entre 2008 a 2020 presento cinco brotes de dengue epidémicos mayores a 400 casos por semana durante los años, 2010, 2013, 2015, 2016 y comienzos del año 2020. Alcanzando 800 casos en el 2016 y 2020. En cambio, ciudades como Barranquilla, Cúcuta, Villavicencio y Santa Marta tuvieron casos reportando menores a 200 por semana en el tiempo. No obstante, Santa Marta fue la ciudad con menor casos en el tiempo. Sin embargo, Villavicencio a mitad del 2019 reporto 200 casos semanales dengue.

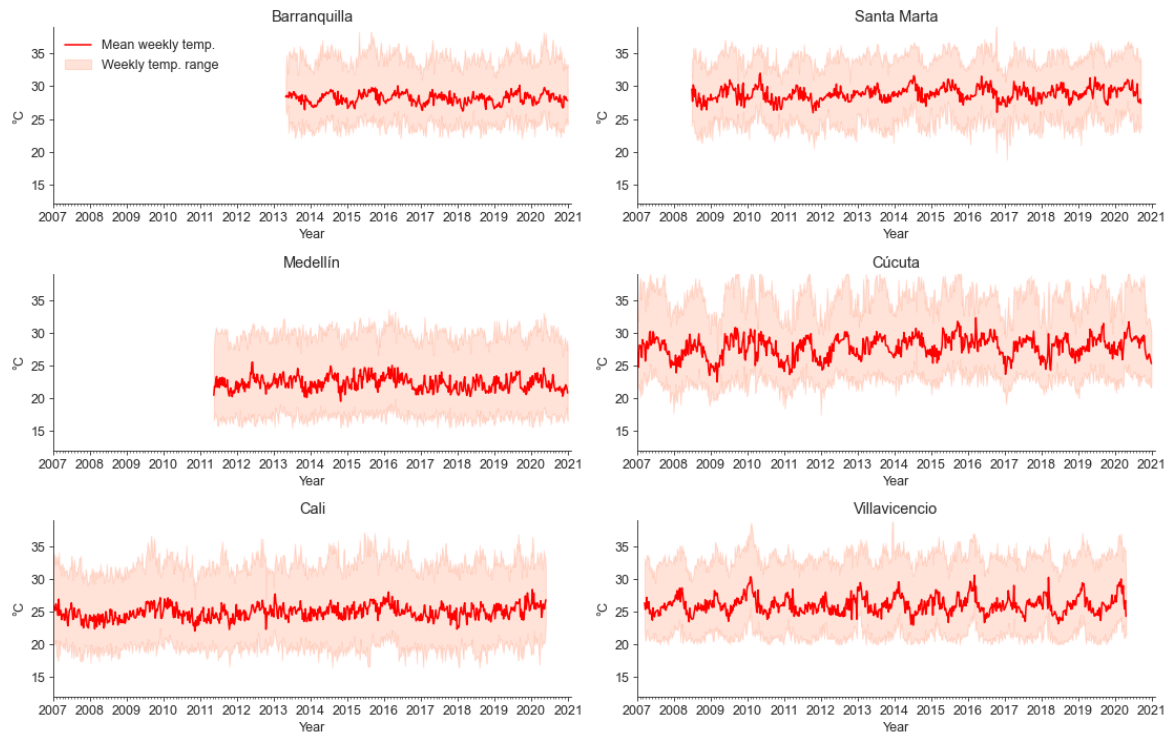


FIGURA 33. Temperatura media, rango de temperatura máxima y mínima semanal a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio.

En la figura 33 podemos evidenciar temperaturas medias para Barraquilla alrededor de 28 °C con temperatura máximas de 35 grados centígrados y temperatura mínima de 22 °C en el periodo de tiempo del 2014 y 2021. Sin embargo, no se pudo obtener las temperaturas antes del 2013 hacia atrás dado que en su momento la plataforma del IDEAM los datos no estaban disponibles. A su vez, para la ciudad de Medellín la temperatura media fue alrededor de 23°C con temperatura máximas de 30°C y temperaturas mínimas del 17°C entre los años 2011 y 2021. Es importante mencionar, que datos anteriores al año 2011 no se pudieron recolectar debido a que había muchos datos no reportados. La ciudad de Santa Marta, presento temperaturas medias alrededor de 27°C con temperaturas máximas de 35 °C y mínimas de 24 °C. La ciudad de Cali, reporto temperaturas medias en torno a 25 °C, alcanzando temperaturas máximas de 35°C y temperaturas mínimas de 18 °C. La ciudad de Cúcuta reporto temperaturas medias de 27°C con temperaturas máximas de 38 °C y mínimas de 20°C. Por último, la ciudad de Villavicencio presento temperaturas medias de 27°C con temperaturas máximas de 38 °C y mínimas de 20°C.



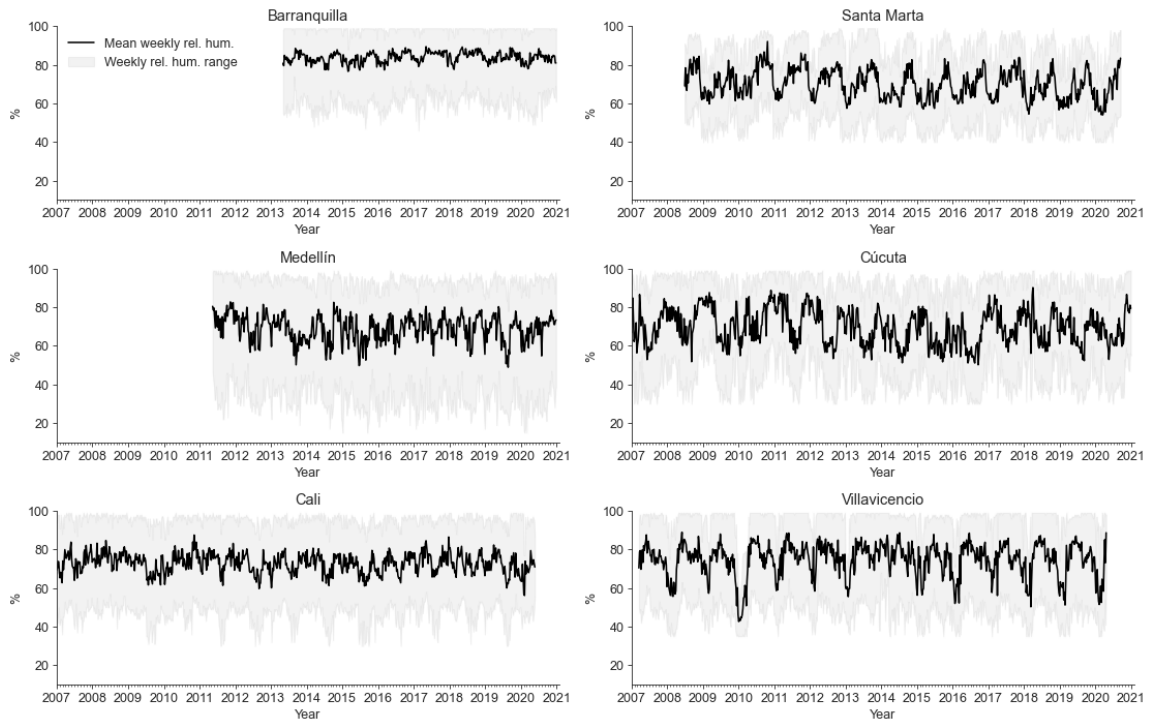


FIGURA 34. Humedad relativa, rango de humedad relativa máxima y mínima semanal a través de los años desde el 2008 hasta comienzo del 2020 en cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio.

Los datos no graficados de la humedad relativa para las ciudades de Barranquilla, Medellín y Santa Marta se debió principalmente a datos faltantes o no reportados en la base de datos. Barranquilla reportó una humedad relativa en torno al 85% con rangos máximos y mínimos de 100% y 35% respectivamente. Para la ciudad de Santa Marta la humedad relativa media varía en el 78% con rangos entre el 100% y 40% para la humedad relativa máxima y mínima respectivamente. La ciudad de Medellín tuvo humedad relativa media alrededor del 70% alcanzando valores mínimos y máximos del 20% a 100%. La ciudad de Cali reportó humedades relativas del 77% con rango mínimos y máximos del 35% al 100%. A su vez, la ciudad de Cúcuta reportó humedad relativa media del 77% con rangos mínimos y mayores de 38% a 100%. Por último, la ciudad de Villavicencio presentó humedad relativa del 78% con rangos entre 40% a 100%.

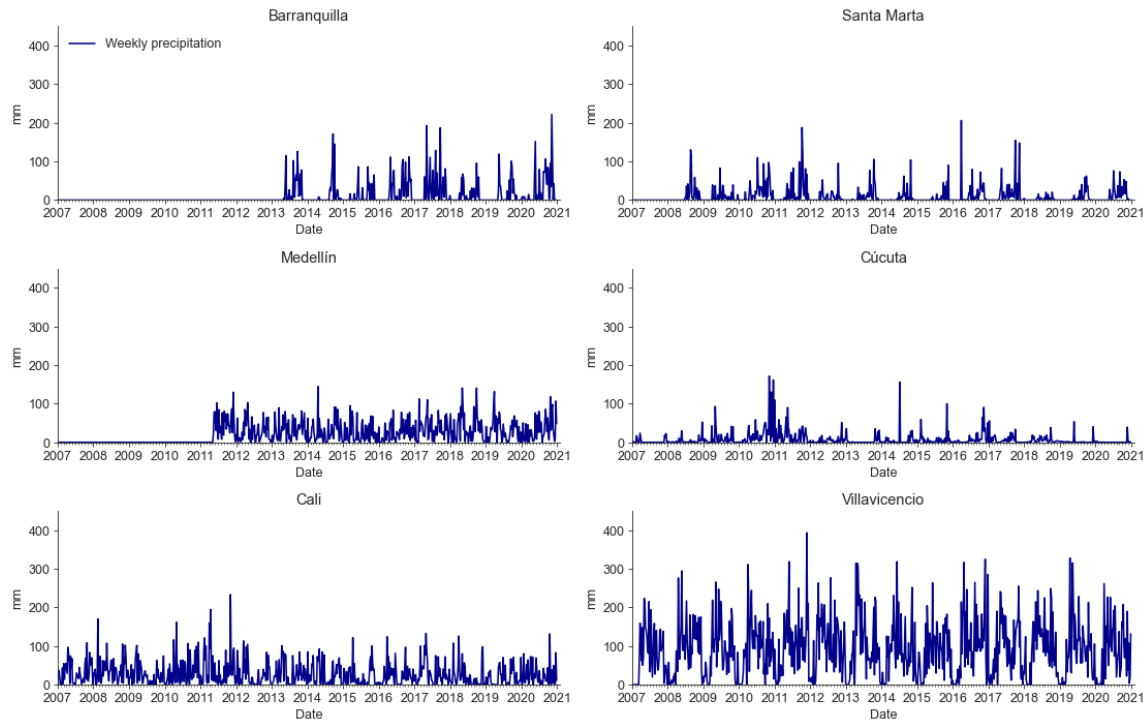


FIGURA 35. Precipitaciones semanales a través de los años desde el 2008 hasta comienzo del 2020 de cada ciudad de interés; Barranquilla, Medellín, Cali, Santa Marta, Cali, Cúcuta y Villavicencio.

Barranquilla, Santa Marta, Medellín, Cali, y Cúcuta presentaron precipitaciones menores a 150 mm. Donde Cúcuta es la ciudad con menor precipitación el tiempo, seguida de la ciudad de Santa Marta. Medellín y Cali presentaron precipitaciones constantes no mayores 100 mm y la ciudad de Villavicencio sostuvo lluvias mayores a 100 mm alcanzando semanas con precipitaciones mayores a 250 mm.

Para analizar cómo afecta las variables climáticas y el fenómeno niño a los casos dengue se generó un conjunto de datos con retraso de 4 a 25 semanas, normalizando las series temporales para las variables mencionadas, para luego proceder a calcular la correlación entre cada variable dependiente y los casos de dengue en cada rezago.

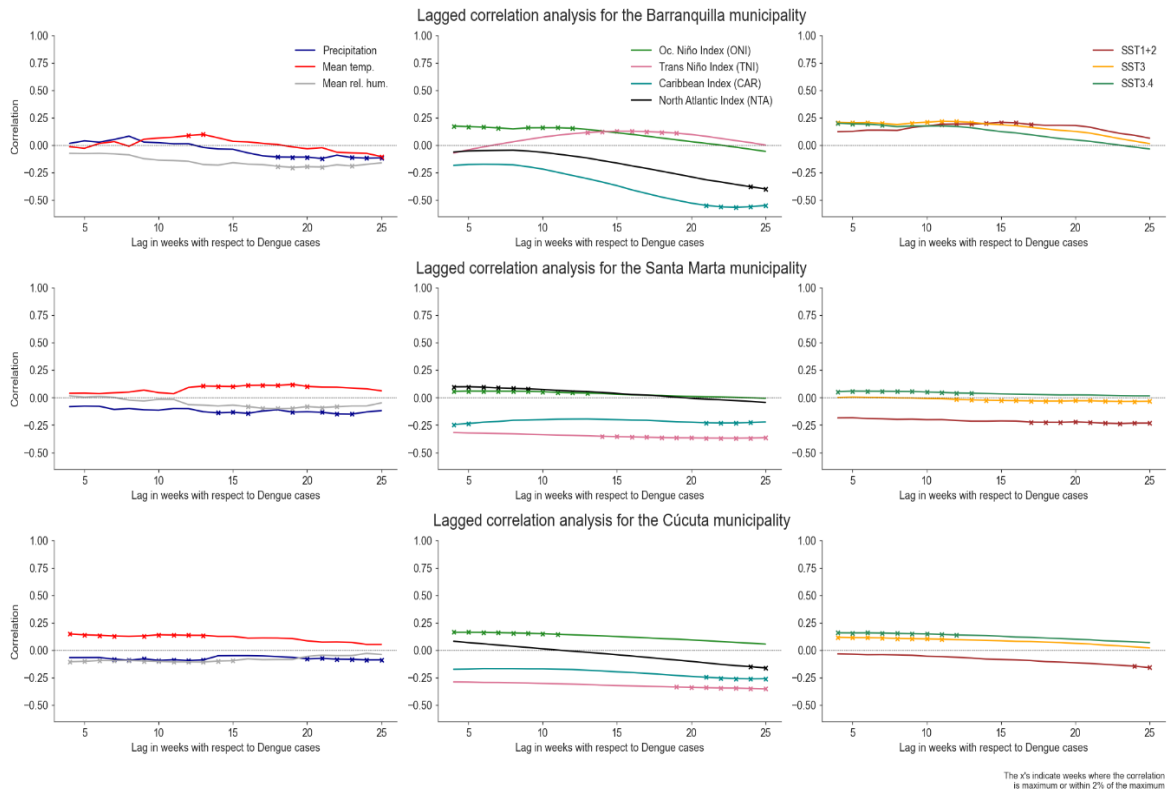


FIGURA 36. Análisis de correlación entre las variables climáticas y los casos de dengue para tres ciudades: Villavicencio, Cucúta y Medellín.

Las correlaciones lineales halladas entre -1 a 1 de las variables dependientes con los casos dengue, nos permite inferir que los valores mayores o menores a cero tienen un efecto o impacto de retardado en la incidencia de los casos de dengue. Por consiguiente, los brotes de dengue podrían estar relacionados con dichas variables acorde a ciclos de vida del mosquito en esos instantes de tiempo. Por consiguiente, en la figura 31, se muestran los resultados obtenidos en tres ciudades diferentes; Barranquilla, Santa Marta y Cúcuta, cuyo análisis es similar para las demás ciudades restantes seleccionadas. En la ciudad de Barranquilla la humedad relativa, la temperatura media y la precipitación tienen un efecto en los casos de dengue de manera inversa de acuerdo con la correlación negativa hallada a través de los rezagos. No obstante, la precipitación al inicio de los rezagos tiene un efecto directo significativo en los casos de dengue, pero después del rezago 9 tiene una incidencia inversa debido a la disminución de los valores negativos de correlación hasta el final de los rezagos. Del mismo modo, la temperatura media presentó una relación directa

representativa con los casos de dengue hasta el rezago 12, posteriormente a este la correlación disminuye a valores negativos. El análisis de correlación asociada a las variables de los índices oceánicos del Niño, del caribe, y del Atlántico respecto a los casos de dengue; evidencio que los índices del niño tienen una correlación directa con los casos de dengue a diferencia del caribe y del atlántico. Estos últimos a través de los rezagos tienen un efecto inverso al tener valores correlaciones menores a cero y cercanos a -1. Por otro lado, las temperaturas superficiales del mar tienen una relación directa con la presencia de casos de dengue manteniéndose relativamente constante hasta el rezago 15 a 17 semanas. Posteriormente, disminuye hasta alcanzar un valor aproximadamente a cero, por lo cual no podemos inferir o asumir alguna relación de covariación entre ellas.

En la ciudad de Santa Marta, el análisis de correlaciones para las variables climáticas como precipitación y humedad relativa presentan una relación inversa con los casos dengue. A diferencia, de la temperatura media que conserva una correlación positiva evidenciando un posible efecto directo entre los casos de dengue. Para el análisis de correlación entre los índices del fenómeno del niño y el atlántico los valores fueron cercanos a cero con valores positivos hasta el rezago 17, y posteriormente con valores negativos hasta el 25, indicando no inferencia entre los casos de dengue. A su vez, Los índices del Trans niño y el caribe evidencio relaciones inversas entre los casos del dengue a través de los rezagos con correlaciones negativas significativas. De manera similar, las temperaturas superficiales del mar SST3 y SST3.4. obtuvieron valores cercanos a cero, demostrando la no inferencia entre las variables con los casos de dengue. No obstante, la SST1+2 indico un efecto inverso con los casos de Dengue de manera constante en los rezagos propuestos.

El análisis de correlaciones para la ciudad de Cúcuta para las variables climáticas de humedad relativa y precipitación presentaron un efecto inverso en los casos de dengue. No obstante, las correlaciones negativas estuvieron cercanas a cero y la temperatura media fue positiva disminuyendo hasta valores cercanos a cero a medida que avanzaba los rezagos. Las correlaciones asociadas a los índices trans del niño, el caribe y atlántico, evidencio que las correlaciones eran negativas, implicando consigo la relación inversa en la incidencia con los casos de dengue. En cambio, el índice oceánico del niño tuvo valores positivos, pero a medida que avanzaba los rezagos disminuía hasta alcanzar valores próximos a cero. Por último, las temperaturas superficiales SST3.4 y SST3 tuvieron

correlaciones positivas pero cercanas a cero a diferencia de las SST1+2 que inicio valores cercanos a cero, pero disminuyeron a traves de los rezagos.

El análisis de correlaciones se hizo a cada ciudad y se analizó como afectaba cada variable a los casos de dengue identificando consigo sus rezagos. Estos son reportados a continuación;

City	Precip	Mean temp	Mean r. hum	ONI	TNI	CAR	NTA	SST1+2	SST3	SST34	Average lag
Barranquilla	21	13	20	8	16	23	24	14	9	5	15
Santa Marta	19	16	19	8	19	21	6	21	18	8	15
Medellín	22	21	23	24	18	7	8	24	24	24	19
Cúcuta	13	9	9	7	22	23	24	24	7	8	14
Cali	22	16	20	18	9	13	5	24	21	19	14
Villavicencio	17	17	20	19	10	17	23	5	22	20	17

Tabla 5. Rezagos entre las variables climáticas, índices del NIÑO, NTA y CAR en los casos de dengue en las diferentes ciudades.

De acuerdo con los resultados obtenidos; Barranquilla, tiene un impacto temprano en los casos de dengue en el rezago 5 debido a la temperatura superficial SST3.4. En cambio, los índices de CAR tiene un efecto en los casos dengue a los 22 rezagos. En la ciudad de Santa Marta, el índice de NTA tiene un efecto a las 8 semanas de rezago a diferencia del índice CAR que índice a las 21 semanas de rezago. La ciudad de Medellín, el índice de CAR tiene una incidencia temprana en los casos de dengue a las 8 semanas de rezago, la temperatura SST3.4 en contraste con la variable ONI que llega hasta 23 semanas de rezago para tener un efecto en los casos. La ciudad de Cúcuta la variable ONI presento efecto a las 7 semanas de rezago en comparación con los índices NTA, SST1+2 que llegaron hasta 23 semanas para tener una incidencia de los casos. La ciudad de Cali, la variable NTA tuvo un efecto a las 5 semanas en oposición a la precipitación que fue a las 22 semanas. Finalmente, la ciudad de Villavicencio, la variable SST1+2 tiene un efecto a las 5 semanas de rezago en oposición al índice Noratlántico Tropical (NTA) que fue a las 22 semanas.

La importancia de los rezagos es la implementación a futuros modelos de predicción de casos de dengue, en donde además de tener en cuentas las variables climáticas, fenómenos del niño y los casos de dengue, se pueda entender cómo afecta los impactos a través de esos intervalos de tiempo en la aparición de nuevos casos y los ciclos de vida del mosquito como la propagación de la enfermedad.

## 6.7 Conclusiones

El análisis realizado en este trabajo permite entender como las variables climáticas, fenómeno del niño, temperaturas superficies, así como los diferentes índices en el atlántico y caribe están relacionadas con los casos de dengue según la ubicación geográfica de cada ciudad. No obstante, el análisis de cada variable con los casos de dengue a través rezagos permite entender la afectación de estas a mayor o menor tiempo en la incidencia de brotes. Aunque las condiciones climáticas pueden ser similares en algunas regiones sus correlaciones con los casos de dengue son diferentes debido a mecanismos de incidencias principalmente a fenómenos como el Niño, índice Noratlántico Tropical (NTA) y el índice del Caribe (CAR). Eso se debe principalmente a las fluctuaciones de temperaturas en las regiones del mar, las cuales puede provisionar mejores condiciones para ciclo vida del mosquito. El análisis obtenido se evidencia este tipo de relaciones en diferentes ciudades: La ciudad de Barranquilla el rezago que tiene un efecto a menor plazo con la temperatura superficial SST3.4. En la ciudad de Santa Marta, el índice Noratlántico Tropical (NTA) tiene un efecto a las 8 semanas de rezago debido a la cercanía con el océano atlántico. La ciudad de Cúcuta la variable ONI presentó efecto de menor plazo a las 7 semanas de rezago, en la ciudad de Cali el índice que tuvo una afectación fue el índice NTA a las 5 semanas y en la ciudad de Villavicencio, la variable SST1+2 tiene un efecto a las 5 semanas de rezago. Sin embargo, en este análisis no se tuvieron en cuenta otros tipos de variables como las condiciones geográficas de cada región y principalmente con los lugares de mayores brotes de pacientes con Dengue, los comportamientos sociales y de salubridad de la población que pueden ser dicientes en las correlaciones conjuntas en la aparición de los casos de dengue.

# 7. Conclusiones y recomendaciones

## 7.1 Conclusiones

A través de la metodología utilizada para obtención de la superficie de las proteínas de complemento Factor B y C3, es posible obtener las coordenadas, la distribución de cargas en las superficies para así realizar futuras simulaciones múltiple proteínas. Sin embargo, es importante desarrollar otro método de cálculo basado en las fluctuaciones de los momentos dipolares del sistema debido al constante cambio en la evolución del sistema, o como alternativa implementar más tiempo de simulación mayor a 100 ns para analizar las tendencias en el tiempo.

La propuesta de métodos Monte Carlo y de dinámica browniana para la agregación y segregación de partículas como RNPs permiten entender la configuraciones y formaciones de hélices a través de la minimización del sistema entre las interacciones electrostática de Debye Huckel y volumen excluido de Lennard Jones. Aunque el sistema puede representar la dinámica de agregación, se evidencio la necesidad de mejorar los métodos propuestos como el muestreo y el balance entre la energía del sistema de acuerdo a las proporciones de carga de las partículas. Se evidenció, la necesidad de optimizar los cómputos de simulaciones de la dinámica browniana. Por otra parte, el análisis de bioinformático evidencio que interacciones entre las cadenas de ARN a través de la hibridación no es suficiente para generar una estructura estable. Por consiguiente, se expone en el modelo mínimo la formación de de RNPs a través de interacciones electrostáticas.

En análisis de rezagos entre las variables climáticas, índices como fenómeno del niño, índice del Noratlántico Tropical (NTA) y el índice del Caribe (CAR) con los casos de dengue reportados en cada ciudad, presentan una correlación significativa con los índices mayor a las variables climáticas. Esto se debe principalmente a las fluctuaciones de la temperatura de mar y a su vez, encontrando consigo que cada ciudad tiene un rezago diferente de acuerdo con la ubicación geográfica y su cercanía con las costas a pesar de tener en algunas de ellas condiciones climáticas similares.

## Bibliografía

- [1] J. Walpole, J. A. Papin, and S. M. Peirce, "Multiscale Computational Models of Complex Biological Systems," *Annu. Rev. Biomed. Eng.*, vol. 15, no. 1, pp. 137–154, 2013, doi: 10.1146/annurev-bioeng-071811-150104.
- [2] M. Tawhai, J. Bischoff, D. Einstein, A. Erdemir, T. Guess, and J. Reinbolt, "Multiscale modeling in computational biomechanics.," *IEEE Eng. Med. Biol. Mag.*, vol. 28, no. 3, pp. 41–9, 2009, doi: 10.1109/MEMB.2009.932489.
- [3] J. S. Yu and N. Bagheri, "Multi-class and multi-scale models of complex biological phenomena," *Curr. Opin. Biotechnol.*, vol. 39, pp. 167–173, 2016, doi: 10.1016/j.copbio.2016.04.002.
- [4] J. O. Dada and P. Mendes, "Multi-scale modelling and simulation in systems biology," *Integr. Biol.*, vol. 3, no. 2, p. 86, 2011, doi: 10.1039/c0ib00075b.
- [5] G. A. Vásquez-Montoya, J. S. Danobeitia, L. A. Fernández, and J. P. Hernández-Ortiz, "Computational immuno-biology for organ transplantation and regenerative medicine," *Transplant. Rev.*, vol. 30, no. 4, pp. 235–246, 2016, doi: 10.1016/j.trre.2016.05.002.
- [6] M. L. Martins, S. C. Ferreira, and M. J. Vilela, "Multiscale models for biological systems," *Curr. Opin. Colloid Interface Sci.*, vol. 15, no. 1–2, pp. 18–23, 2010, doi: 10.1016/j.cocis.2009.04.004.
- [7] C. T. S. Epistemus, "Métodos de simulación computacional en biología," pp. 84–92.
- [8] Y. L. Wang, Y. L. Zhu, Z. Y. Lu, and A. Laaksonen, "Electrostatic interactions in soft particle systems: Mesoscale simulations of ionic liquids," *Soft Matter*, vol. 14, no. 21, pp. 4252–4267, 2018, doi: 10.1039/c8sm00387d.
- [9] H. X. Zhou and X. Pang, "Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation," *Chem. Rev.*, vol. 118, no. 4, pp. 1691–1741, 2018, doi: 10.1021/acs.chemrev.7b00305.
- [10] P. Kuki and J. E. Nielsen, "Electrostatics in proteins and protein-ligand complexes," *Future Med. Chem.*, vol. 2, no. 4, pp. 647–666, 2010, doi: 10.4155/fmc.10.6.
- [11] M. Lund and B. Jönsson, "A Mesoscopic Model for Protein-Protein Interactions in Solution," *Biophys. J.*, vol. 85, no. 5, pp. 2940–2947, 2003, doi: 10.1016/S0006-3495(03)74714-6.
- [12] C. E. Dykstra, "Electrostatic Interaction Potentials in Molecular Force Fields," *Chem. Rev.*, vol. 93, no. 7, pp. 2339–2353, 1993, doi: 10.1021/cr00023a001.
- [13] J. P. Hernández-Ortiz, J. J. de Pablo, and M. D. Graham, "N log N method for hydrodynamic interactions of confined polymer systems: Brownian dynamics.," *J. Chem. Phys.*, vol. 125, no. 16, p. 164906, 2006, doi: 10.1063/1.2358344.
- [14] M. M. Gromiha, R. Nagarajan, and S. Selvaraj, "Protein Structural Bioinformatics: An Overview," *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp.



- 445–459, Jan. 2019, doi: 10.1016/B978-0-12-809633-8.20278-1.
- [15] S. Skariyachan and S. Garka, “Exploring the binding potential of carbon nanotubes and fullerene towards major drug targets of multidrug resistant bacterial pathogens and their utility as novel therapeutic agents,” *Fullerenes, Graphenes Nanotub. A Pharm. Approach*, pp. 1–29, Jan. 2018, doi: 10.1016/B978-0-12-813691-1.00001-4.
- [16] O. Sensoy, J. G. Almeida, J. Shabbir, I. S. Moreira, and G. Morra, “Computational studies of G protein-coupled receptor complexes: Structure and dynamics,” *Methods Cell Biol.*, vol. 142, pp. 205–245, Jan. 2017, doi: 10.1016/BS.MCB.2017.07.011.
- [17] S. Abeln, K. A. Feenstra, and J. Heringa, “Protein Three-Dimensional Structure Prediction,” *Encycl. Bioinforma. Comput. Biol. ABC Bioinforma.*, vol. 1–3, pp. 497–511, Jan. 2019, doi: 10.1016/B978-0-12-809633-8.20505-0.
- [18] A. Šali and T. L. Blundell, “Comparative Protein Modelling by Satisfaction of Spatial Restraints,” *J. Mol. Biol.*, vol. 234, no. 3, pp. 779–815, Dec. 1993, doi: 10.1006/JMBI.1993.1626.
- [19] Y. S. Watanabe, Y. Fukunishi, and H. Nakamura, “1P047 Modeling of Loops in Protein Structures,” *Seibutsu Butsuri*, vol. 44, no. supplement, p. S41, 2004, doi: 10.2142/biophys.44.s41\_3.
- [20] Poeran, “乳鼠心肌提取 HHS Public Access,” *Physiol. Behav.*, vol. 176, no. 12, pp. 139–148, 2017, doi: 10.1002/cpbi.3.Comparative.
- [21] J. J. Wendoloski and J. B. Matthew, “Molecular dynamics effects on protein electrostatics,” *Proteins Struct. Funct. Bioinforma.*, vol. 5, no. 4, pp. 313–321, 1989, doi: 10.1002/prot.340050407.
- [22] Lindahl, Abraham, Hess, and van der Spoel, “GROMACS Documentation,” *GROMACS 2021.3 Man.*, pp. 1–623, 2021.
- [23] Max Planck Institute, “Atomistic Simulation of Biomolecular function,” *Leonard Heinz*. [http://www2.mpibpc.mpg.de/groups/grubmueller/Lugano\\_Tutorial/part1/](http://www2.mpibpc.mpg.de/groups/grubmueller/Lugano_Tutorial/part1/).
- [24] M. P. Allen and D. J. Tildesley, “Computer simulation of liquids: Second edition,” *Comput. Simul. Liq. Second Ed.*, pp. 1–626, 2017, doi: 10.1093/oso/9780198803195.001.0001.
- [25] T. T. Nguyen, M. H. Viet, and M. S. Li, “Effects of water models on binding affinity: Evidence from all-atom simulation of binding of tamiflu to A/H5N1 neuraminidase,” *Sci. World J.*, vol. 2014, 2014, doi: 10.1155/2014/536084.
- [26] A. Emperador, R. Crehuet, and E. Guàrdia, “Effect of the water model in simulations of protein–protein recognition and association,” *Polymers (Basel)*, vol. 13, no. 2, pp. 1–9, 2021, doi: 10.3390/polym13020176.
- [27] I. Snook, “Brownian Dynamics,” *Langevin Gen. Langevin Approach to Dyn. At.*

- Polym. Colloid. Syst.*, pp. 133–156, 2007, doi: 10.1016/b978-044452129-3/50008-0.
- [28] R. R. Gabdoulline and R. C. Wade, “METHODS: A Companion to Methods in Brownian Dynamics Simulation of Protein–Protein Diffusional Encounter,” *Enzymology*, vol. 14, pp. 329–341, 1998, [Online]. Available: <https://dasher.wustl.edu/chem478/labs/lab-11/methods-14-329-98.pdf>.
- [29] S. Palacios, V. Romero-Rochin, and K. Volke-Sepulveda, “Brownian motion in typical microparticle systems,” pp. 1–12, 2011, [Online]. Available: <http://arxiv.org/abs/1108.3316>.
- [30] P. Debye and E. Huckel, “The theory of electrolytes I. The lowering of the freezing point and related occurrences,” *Phys. Zeitschrift*, vol. 24, no. 1923, pp. 185–206, 1923.
- [31] A. Vasquez Echeverri, “Solución de la ecuación de Fokker-Plank para simular ADN confinado fuera del equilibrio considerando condiciones electrostaticas,” p. 94, 2016, [Online]. Available: <http://www.bdigital.unal.edu.co/55094/>.
- [32] E. Hückel, “Zur Theorie der Elektrolyte,” *Ergebnisse der exakten naturwissenschaften*, pp. 199–276, 2007, doi: 10.1007/bfb0111753.
- [33] S. R. Barnum and T. N. Schein, *The Complement System*, Second Edi. Elsevier Ltd, 2018.
- [34] J. Laskowski and J. M. Thurman, *Factor B*, Second Edi. Elsevier Ltd, 2018.
- [35] S. R. Barnum, *C3*, Second Edi. 2018.
- [36] B. J. C. Janssen *et al.*, “Structures of complement component C3 provide insights into the function and evolution of immunity,” *Nature*, vol. 437, no. 7058, pp. 505–511, 2005, doi: 10.1038/nature04005.
- [37] W. DeLano and S. Bromberg, “PyMOL User’s Guide (Original),” *DeLano Sci. LLC*, pp. 1–66, 2004, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:PyMOL+User’s+Guide#2>.
- [38] C. User, “Chimera1.10\_UsersGuide.”
- [39] D. M. Hinckley, G. S. Freeman, J. K. Whitmer, and J. J. De Pablo, “An experimentally-informed coarse-grained 3-site-per-nucleotide model of DNA: Structure, thermodynamics, and dynamics of hybridization,” *J. Chem. Phys.*, vol. 139, no. 14, 2013, doi: 10.1063/1.4822042.
- [40] C. Eno, C. L. Hansen, and F. Pelegri, “Aggregation, segregation, and dispersal of homotypic germ plasm RNPs in the early zebrafish embryo,” *Dev. Dyn.*, vol. 248, no. 4, pp. 306–318, 2019, doi: 10.1002/dvdy.18.
- [41] Y. Hashimoto *et al.*, “Localized maternal factors are required for zebrafish germ cell formation,” *Dev. Biol.*, vol. 268, no. 1, pp. 152–161, 2004, doi: 10.1016/j.ydbio.2003.12.013.

- [42] T. Trcek, M. Grosch, A. York, H. Shroff, T. Lionnet, and R. Lehmann, "Drosophila germ granules are structured and contain homotypic mRNA clusters," *Nat. Commun.*, vol. 6, 2015, doi: 10.1038/ncomms8962.
- [43] C. Eno and F. Pelegri, "Gradual recruitment and selective clearing generate germ plasm aggregates in the zebrafish embryo," *Bioarchitecture*, vol. 3, no. 4, pp. 125–132, 2013, doi: 10.4161/bioa.26538.
- [44] S. Nijjar and H. R. Woodland, "Protein interactions in *Xenopus* germ plasm RNP particles," *PLoS One*, vol. 8, no. 11, 2013, doi: 10.1371/journal.pone.0080077.
- [45] C. Yoon, K. Kawakami, and N. Hopkins, "Zebrafish vasa homologue RNA is localized to the cleavage planes of 2- and 4-cell-stage embryos and is expressed in the primordial germ cells.," *Development*, vol. 124, no. 16, pp. 3157–65, 1997, [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9272956>.
- [46] J. A. Toretsky and P. E. Wright, "Assemblages: Functional units formed by cellular phase separation," *J. Cell Biol.*, vol. 206, no. 5, pp. 579–588, 2014, doi: 10.1083/jcb.201404124.
- [47] C. P. Brangwynne, P. Tompa, and R. V. Pappu, "Polymer physics of intracellular phase transitions," *Nat. Phys.*, vol. 11, no. 11, pp. 899–904, 2015, doi: 10.1038/nphys3532.
- [48] Y. Lin, D. S. W. Protter, M. K. Rosen, and R. Parker, "Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins," *Mol. Cell*, vol. 60, no. 2, pp. 208–219, 2015, doi: 10.1016/j.molcel.2015.08.018.
- [49] "Two segregation pathways of germ plasm RNPs in teleosts."
- [50] M. Rouhani, F. Khodabakhsh, D. Norouziyan, R. A. Cohan, and V. Valizadeh, "Molecular dynamics simulation for rational protein engineering: Present and future prospectus," *J. Mol. Graph. Model.*, vol. 84, pp. 43–53, 2018, doi: 10.1016/j.jmgm.2018.06.009.
- [51] A. Sasse, K. U. Laverty, T. R. Hughes, and Q. D. Morris, "Motif models for RNA-binding proteins," *Curr. Opin. Struct. Biol.*, vol. 53, no. August, pp. 115–123, 2018, doi: 10.1016/j.sbi.2018.08.001.
- [52] C. Cragnell, E. Rieloff, and M. Skepö, "Utilizing Coarse-Grained Modeling and Monte Carlo Simulations to Evaluate the Conformational Ensemble of Intrinsically Disordered Proteins and Regions," *J. Mol. Biol.*, vol. 430, no. 16, pp. 2478–2492, 2018, doi: 10.1016/j.jmb.2018.03.006.
- [53] C. Rodriguez. and M. Aluztisa., "Medicina molecular de FIBAO," *Medicina (B. Aires)*, vol. 1, pp. 7–9, 2007, [Online]. Available: [www.medmol.es](http://www.medmol.es).
- [54] CDC, "Centro Nacional para Enfermedades Infecciosas Emergentes y Zoonóticas División de Enfermedades Transmitidas por Vectores," *J. Infect. Dis.*, vol. 41, no. 2, p. 1, 2020, [Online]. Available:

- [55] NOAA, "El Niño Regions," 2005, [Online]. Available: [https://www.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/nino\\_regions.shtml](https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/nino_regions.shtml).