

Universidad Católica de Santa María

Facultad de Ciencias Farmacéuticas, Bioquímicas y

Biotechnológicas

Escuela Profesional de Ingeniería Biotechnológica



**ANÁLISIS BIOINFORMÁTICO DE MUTACIONES DE INSERCIÓN, DELECIÓN Y
SUSTITUCIÓN EN EL GEN B-GLOBINA HUMANA.**

Tesis presentada por el Bachiller:

Bustamante Cárdenas, Edgar Miguel

Para optar por el Título Profesional de:

Ingeniero Biotechnólogo

Asesor: Dr. Gómez Valdez, Badhin

Arequipa - Perú

2021

UCSM-ERP

UNIVERSIDAD CATÓLICA DE SANTA MARÍA
INGENIERIA BIOTECNOLOGICA
TITULACIÓN CON TESIS
DICTAMEN APROBACIÓN DE BORRADOR

Arequipa, 02 de Febrero del 2021

Dictamen: 002708-C--2021

Visto el borrador del expediente 002708, presentado por:

2007224171 - BUSTAMANTE CARDENAS EDGAR MIGUEL

Titulado:

**ANÁLISIS BIOINFORMÁTICO DE MUTACIONES DE INSERCIÓN, DELECIÓN Y SUSTITUCIÓN EN EL
GEN β -GLOBINA HUMANA**

Nuestro dictamen es:

APROBADO

**0737 - PAZ ALIAGA CARLOS EITEL IVAN
DICTAMINADOR**



**2781 - BARREDA DEL CARPIO JAIME ERNESTO
DICTAMINADOR**



**3060 - CARPIO CARPIO JOSE MIGUEL
DICTAMINADOR**



DEDICATORIA

Le agradezco a Dios por haberme acompañado y guiado a lo largo de mi carrera, por ser mi fortaleza en los momentos de debilidad y por brindarme una vida llena de aprendizajes, experiencias y felicidad.

Le doy gracias a mi familia en especial a mis padres por apoyarme en todo momento, por los valores que me han inculcado, y por haberme dado la oportunidad de tener una excelente educación en el transcurso de mi vida.

RESUMEN

La Beta-talasemia se describe como un grupo de trastornos sanguíneos hereditarios caracterizados por anomalías en la síntesis de cadenas beta de la hemoglobina. Estas anomalías dan lugar a diferentes fenotipos que van desde síntomas clínicos moderados a severo, hasta no tener ningún síntoma en absoluto. La mayoría de los defectos en la hemoglobina surgen directamente de las mutaciones en la estructura del gen β -globina (HBB).

Avances recientes en herramientas computacionales han permitido estudiar la relación entre el genotipo y el fenotipo en muchas enfermedades, incluida la β -talasemia. Debido a la alta prevalencia de β -talasemia, estos análisis han ayudado a comprender la base molecular de la enfermedad de una mejor manera. En esta dirección, una base de datos relacionada, denominada HbVar, fue desarrollada en 2001 por un esfuerzo colectivo académico para proporcionar información de calidad y actualizada sobre las variaciones genómicas que conducen a hemoglobinopatías y Talasemia. La base de datos registró detalles sobre cada variante, incluida la secuencia alterada, los defectos hematológicos, su patología y su fecha de aparición, junto con su lugar de procedencia.

En el presente estudio, el objetivo fue investigar mutaciones de inserción, sustitución y delección en el HBB recogido de HbVar y sus efectos utilizando el enfoque *in silico*.

En este análisis comparativo se estudió tres secuencias una wild-type y dos secuencias mutadas de HBB en la región de América del Sur. Se determinó que la secuencia mutada Brazilian nd-HPFH presenta la formación de una cadena beta en la estructura de la proteína no pudiendo alinearse o superponerse en la wild-type, esto da por resultado la alteración de la función de la proteína, y por consiguiente el desarrollo de la enfermedad Talasemia.

Palabras clave: Hemoglobina, Mutación, Talasemia.

ABSTRACT

Beta-thalassemia is described as a group of inherited blood disorders characterized by abnormalities in the synthesis of beta chains of hemoglobin. These abnormalities give rise to different phenotypes ranging from moderate to severe clinical symptoms, to having no symptoms at all. Most defects in hemoglobin arise directly from mutations in the structure of the β -globin gene (HBB).

Recent advances in computational tools have made it possible to study the relationship between genotype and phenotype in many diseases, including β -thalassemia. Due to the high prevalence of β -thalassemia, these analyzes have helped to understand the molecular basis of the disease in a better way. In this direction, a related database, called HbVar, was developed in 2001 by a collective academic effort to provide quality and up-to-date information on genomic variations that lead to hemoglobinopathies and Thalassemia. The database recorded details about each variant, including the altered sequence, the hematologic defects, its pathology, and its date of appearance, along with its place of origin.

In the present study, the aim was to investigate insertion, substitution, and deletion mutations in HBB collected from HbVar and their effects using the in-silico approach.

In this comparative analysis, three sequences, one wild-type and two mutated HBB sequences, were studied in the South American region. It was determined that the Brazilian nd-HPFH mutated sequence presents the formation of a beta chain in the protein structure, not being able to align or overlap in the wild-type, this results in the alteration of the function of the protein, and therefore the development of Thalassemia disease.

Key words: Hemoglobin, Mutation, Thalassemia.

INTRODUCCION

Las hemoglobinopatías son trastornos genéticos causados por variaciones de un gen en los grupos de genes de la globina humana tipo α y β . Estos son los trastornos hereditarios más comunes en humanos, con casi el 7% de la población humana actuando como portadores de las mutaciones en los genes de la globina. Las sustituciones de nucleótidos individuales en las regiones codificadoras o reguladoras de estos genes pueden conducir a diversos grados de defectos en su expresión^{1,2}. El gen HBB pertenece al grupo de genes de β -globina que codifica el polipéptido de β -globina. Se encuentra en el brazo corto del cromosoma 11 y contiene dos intrones y tres exones. Defectos moleculares en HBB humano puede provocar defectos estructurales que causan anormalidades en las hemoglobinas, como HbS, HbC y HbD, o puede provocar la ausencia o la síntesis reducida de cadenas de β -globina que causan β -talasemia³.

Las mutaciones en el HBB pueden implicar la sustitución, eliminación o inserción de uno o múltiples nucleótidos dentro del gen o sus regiones flanqueantes, lo que resulta en anemia y baja producción de glóbulos rojos^{4,5,6,7}. La β -talasemia se hereda de forma autosómico-recesiva y su manifestación clínica se puede dividir en talasemia mayor, intermedia y talasemia menor.^{8,9} Algunas mutaciones en el HBB conducen a la inactivación completa del gen que resulta en la ausencia de cadenas de β -globina (β^0), lo que a su vez conduce a la forma más grave de talasemia. Otras mutaciones permiten la producción de cadenas de β -globina en proporciones variables que conducen a la β -talasemia. Este caso se encuentra más comúnmente en el Medio Oriente, Asia Central, países mediterráneos, India y el sur de China y en algunas partes de África y América del Sur¹⁰. Es uno de los trastornos genéticos más comunes causados por mutaciones puntuales que causan efectos fenotípicos variables. Estas severidades fenotípicas pueden surgir de defectos en la transcripción, procesamiento de ARN o traducción del gen HBB¹¹.

Las mutaciones más comunes en la mayoría de los países poblados incluyen IVSI-110 (G> A), IVSI-1 (G> A), IVSI-6 (T> C), IVSII-1 (G> A), IVSI-5 (G> C), codón 5 (-CT) y codón 39 (C> T)^{12 - 17}. Debido a la alta prevalencia de fenotipos variables de talasemia y la notable heterogeneidad de sus defectos moleculares, se emplearon diversas estrategias para investigar los mecanismos moleculares de esta enfermedad. Recientes avances en las herramientas computacionales, se ha convertido los análisis in silico en uno de los métodos elegidos para investigar los vínculos entre las características fenotípicas, genómicas y resultantes en la talasemia.

HbVar (<http://globin.cse.psu.edu/globin/hbvar/>) es la base de datos más antigua y apreciada de variantes de hemoglobina y mutaciones de talasemia establecida en 2001^{18,19,20}. Es una base de datos específica de locus, que se desarrolló como un esfuerzo académico combinado para mantener un registro de variantes de hemoglobina, nuevas entradas de datos, actualizaciones y correcciones. Proporciona información actualizada y de alta calidad sobre las variaciones genómicas, los efectos fenotípicos y hematológicos asociados, la patología, la frecuencia de diferentes mutaciones, la prevalencia étnica y la procedencia^{21,22}. HbVar se ha convertido en un recurso primario para la comunidad de investigación que trabaja en las proteínas de globina y para los médicos que tratan con pacientes con hemoglobinopatías, para ayudarlos a hacer los diagnósticos adecuados²³⁻²⁸. El objetivo de este estudio es investigar los efectos sobre la estructura y función de la β -globina proteína utilizando el método in silico.

INDICE

DICTAMEN APROBATORIO

DEDICATORIA

RESUMEN

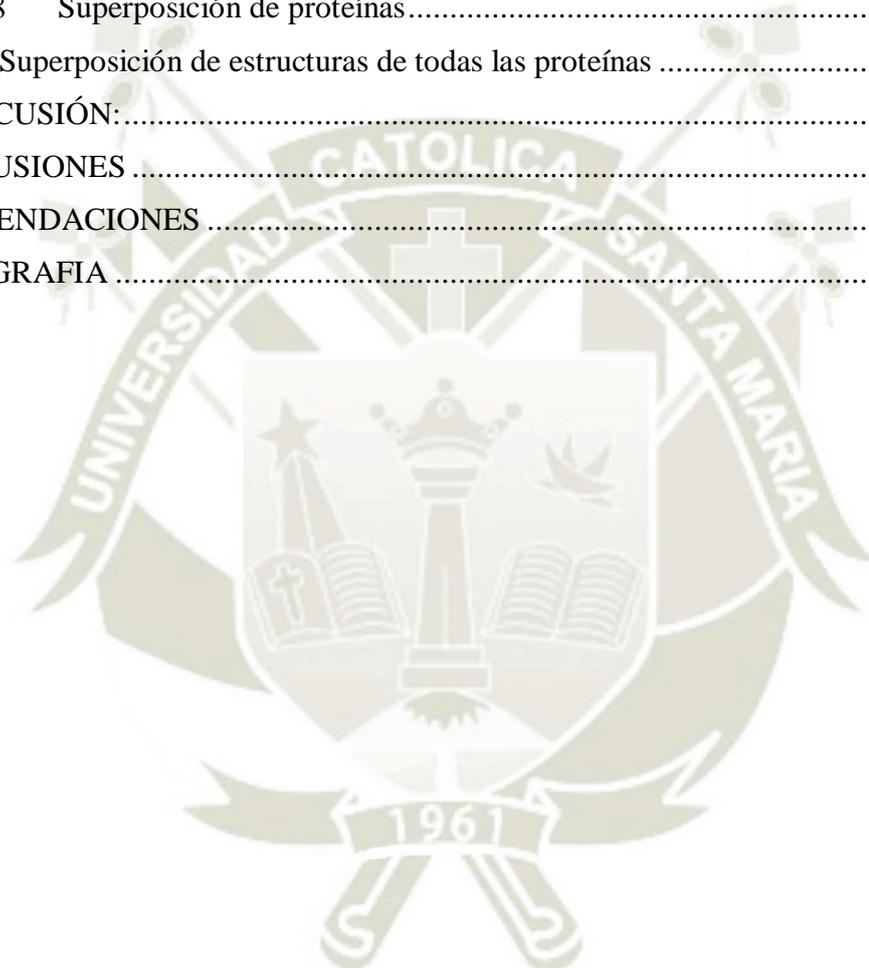
ABSTRACT

INTRODUCCION

CAPÍTULO I	1
HIPÓTESIS	1
OBJETIVOS	2
1 MARCO TEORICO.....	3
1.1 LA HEMOGLOBINA Y SU ESTRUCTURA.....	3
1.1.1 EL ERITROCITO.....	3
1.1.2 HEMOGLOBINA.....	3
1.1.3 TIPOS DE HEMOGLOBINAS	5
1.1.4 FUNCIÓN DE LA HEMOGLOBINA	6
1.1.5 ESTRUCTURA DE LOS GENES GLOBÍNICOS	8
1.2 HEMOGLOBINOPATÍAS	13
1.2.1 HEMOGLOBINOPATÍAS ESTRUCTURALES	14
1.2.2 SÍNDROMES TALASÉMICOS	15
1.3 MUTACIONES	16
1.3.1 TIPOS DE MUTACIONES	17
1.4 BIOINFORMÁTICA	20
1.5 BASE DE DATOS BIOLÓGICAS	22
1.5.1 HbVar.....	22
1.6 PROCESAMIENTO DE DATOS BIOLÓGICOS	23
1.6.1 Alineamiento de secuencias	23
1.6.2 Determinación de la estructura proteica.....	25
1.6.3 Simulación de dinámica molecular	32
CAPITULO II.....	36
2 MATERIALES Y METODOS.....	36
2.1 CAMPO DE APLICACIÓN	36
2.1.1 Área de conocimiento	36
2.1.2 Campo de verificación	36
2.2 MATERIALES	37

2.3	METODOLOGÍA	38
2.3.1	Recolección de secuencia:	38
2.3.2	Alineación de secuencias múltiples (MSA):	38
2.3.3	Traducción:.....	39
2.3.4	Modelado de homología:.....	39
2.3.5	Evaluación del modelo:.....	39
2.3.6	Simulación de dinámica molecular	40
CAPITULO III.....		42
3	RESULTADOS	42
3.1	Variantes mutacionales:	42
3.2	Exploración del MSA entre HB wild y proteínas mutadas:	43
3.3	Traducción:.....	45
3.4	Modelado por homología:	45
3.5	Evaluación:	47
3.6	Cambios conformacionales:	53
3.7	Simulaciones de dinámica molecular de -195(C->G) Agamma, the Brazilian nd-HPFH.	54
3.7.1	RMSD:	54
3.7.2	RMSF	55
3.7.3	ROG	56
3.7.4	Temperatura.....	57
3.7.5	Presión.....	58
3.7.6	Energía	59
3.7.7	Análisis de estructura secundaria (DSSP).....	60
3.7.8	Superposición de proteínas.....	61
3.8	Simulaciones de dinámica molecular de la proteína Hb de Wild-type	62
3.8.1	RMSD:	62
3.8.2	RMSF	63
3.8.3	ROG	64
3.8.4	Temperatura.....	65
3.8.5	Presión.....	66
3.8.6	Energía	67
3.8.7	Análisis de estructura secundaria (DSSP).....	68
3.8.8	Superposición de proteínas.....	69
3.9	Simulaciones de dinámica molecular de Hb Chile	70

3.9.1	RMSD:	70
3.9.2	RMSF	71
3.9.3	ROG	72
3.9.4	Temperatura.....	73
3.9.5	Presión.....	74
3.9.6	Energía	75
3.9.7	Análisis de estructura secundaria (DSSP).....	76
3.9.8	Superposición de proteínas.....	77
3.10	Superposición de estructuras de todas las proteínas	78
4	DISCUSIÓN:.....	79
CONCLUSIONES		82
RECOMENDACIONES		84
BIBLIOGRAFIA		85



INDICE DE TABLAS

Tabla 1 Variantes HBB	42
Tabla 2 Resultados de Molprobity,	49
Tabla 3 Cambios conformacionales en las 3 secuencias por análisis PSIPRED.	53



INDICE DE FIGURAS

Figura 1 De la estructura primaria de la hemoglobina a la cuaternaria.	4
Figura 2 Curva de disociación de la hemoglobina.	6
Figura 3 Esquema de la familia de genes globínicos α	9
Figura 4 Esquema de la familia de genes globínicos beta.....	10
Figura 5 Estructura interna de un gen globínico β	11
Figura 6 Diferencias bioquímicas entre las globinas γ^G y γ^A ..	12
Figura 7 Región del globínico β humano.	13
Figura 8 Tipos de mutaciones génicas. FUENTE: Fita et al, 2008	18
Figura 9: Diagrama de flujo del procesamiento computacional.....	21
Figura 10: Visualización de un alineamiento de secuencias múltiple	24
Figura 11: Visualización de dos estructuras de proteína.....	26
Figura 12: Visualización de la estructura de una proteína	27
Figura 13: Visualización de la entrada del Protein Data Bank.....	28
Figura 14: Esquema general del algoritmo de modelado por homología	30
Figura 15 Algoritmo básico de dinámica molecular.	34
Figura 16 Alineación de secuencia múltiple	44
Figura 17 Traducción de las secuencias.....	45
Figura 18 Modelo de wild-type HB generado por SWISS-MODEL.....	46
Figura 19 Modelo de -195 (C-> G) Agamma; the Brazilian nd-HPFH generado por SWISS-MODEL.....	46
Figura 20 Modelo Hb Chile generado por SWISS-MODEL	47
Figura 21 Gráfico de Wild type de Hb.....	48
Figura 22 Gráfico de -195 (CG) Agamma; the Brazilian nd-HPFH.....	48
Figura 23 El gráfico de Hb Chile.....	48
Figura 24 Estructura 3D de Hb Wild type. Obtenida de Chimera.....	51
Figura 25 Estructura 3D de -195 (C-> G) Agamma; Brazilian nd-HPFH. Obtenida de Chimera.	51
Figura 26 Estructura 3D de Hb Chile. Obtenida de Chimera.....	52
Figura 27 Estructura 3D de comparación de las 3 estructuras.	52
Figura 28 Gráfico de la desviación cuadrática media de la raíz (RMSD) del carbono alfa de la proteína. obtenida de GROMACS.	54
Figura 29 Gráfico de la fluctuación cuadrática media (RMSF) del carbono alfa de la proteína. obtenida de GROMACS	55

Figura 30 Gráfico del radio de giro (ROG) del carbono alfa de la proteína. obtenida de GROMACS.	56
Figura 31 Temperatura obtenida de GROMACS	57
Figura 32 Presión obtenida de GROMACS	58
Figura 33 Gráfico que representa la energía cinética, potencial y total del sistema. obtenida de GROMACS	59
Figura 34 Fluctuación entre la estructura secundaria de la proteína (cadena principal). obtenida de GROMACS	60
Figura 35 Superposición de proteínas en diferentes escalas de tiempo en nanosegundos. obtenida de Chimera	61
Figura 36 Gráfico de la desviación cuadrática media (RMSD) de los carbonos alfa de la proteína. obtenida de GROMACS	62
Figura 37 Gráfico de fluctuación cuadrática media (RMSF) que muestra los carbonos alfa de la proteína. obtenida de GROMACS	63
Figura 38 Gráfico del radio de giro (ROG) de los carbonos alfa de la proteína. obtenida de GROMACS	64
Figura 39 Temperatura obtenida de GROMACS	65
Figura 40 Presión obtenida de GROMACS	66
Figura 41 Gráfico que representa la energía cinética, potencial y total del sistema. obtenida de GROMACS	67
Figura 42 Fluctuación entre la estructura secundaria de la proteína (cadena principal). obtenida de GROMACS	68
Figura 43 Superposición de proteínas en diferentes nanosegundos. obtenida de Chimera....	69
Figura 44 Gráfica de la desviación cuadrática media (RMSD). obtenida de GROMACS	70
Figura 45 Gráfico de fluctuación cuadrática media (RMSF). obtenida de GROMACS	71
Figura 46 Gráfico del radio de giro (ROG) obtenida de GROMACS	72
Figura 47 Temperatura a 300 K obtenida de GROMACS	73
Figura 48 Presión a 0 bar obtenida de GROMACS	74
Figura 49 Gráfico que representa la energía cinética, potencial y total del sistema. obtenida de GROMACS	75
Figura 50 Fluctuación entre estructura secundaria de la proteína (cadena principal) obtenida de GROMACS	76
Figura 51 Superposición de proteínas en diferentes nanosegundos. Obtenida de Chimera...	77
Figura 52 Superposición de las tres proteínas usando PyMol.....	78

CAPÍTULO I

HIPÓTESIS

Al ser un estudio descriptivo no se precisa la formulación de hipótesis.

OBJETIVOS

General

Analizar mutaciones de inserción, sustitución y deleción en el gen β -globina humana utilizando herramientas computacionales.

Específicos

1. Determinar si el programa de análisis de homología permite identificar casos de mutación.
2. Analizar la estabilidad estructural de la proteína mediante simulaciones de DM.
3. Evaluar los efectos de las mutaciones sobre la estructura de la proteína β -globina.
4. Determinar sus efectos sobre la función de la proteína β -globina.
5. Analizar las alteraciones de las secuencias de nucleótidos y aminoácidos con el tipo wild-type / mutantes.

1 MARCO TEORICO

1.1 LA HEMOGLOBINA Y SU ESTRUCTURA

1.1.1 EL ERITROCITO

Los eritrocitos normales son discos bicóncavos con un diámetro medio de aproximadamente 7.8 μm y en el centro de 1 μm . El volumen de los hematíes es de 90 fl. El género y la altitud a la que vive la persona afecta al número de eritrocitos. En los varones normales, el número de eritrocitos por mm^3 es de 5'200,000 (+/- 300,000) y en las mujeres normales de 4'700,000 (+/- 300,000) ²⁹.

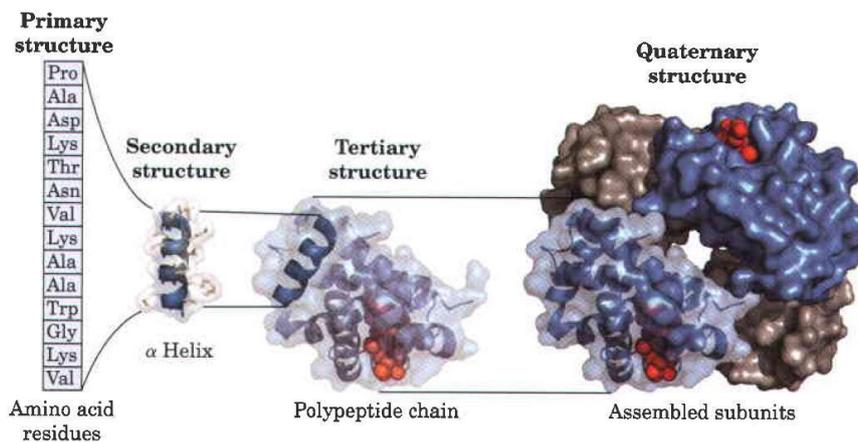
1.1.2 HEMOGLOBINA

La hemoglobina es una molécula casi esférica con un diámetro de cerca de 5,5nm. Es una proteína tetramérica con 4 grupos prostéticos hemo, uno asociado con cada cadena polipeptídica. La hemoglobina adulta (Hb A) contiene 2 tipos de globinas, 2 cadenas α (constituidas por 141 residuos cada una) y 2 cadenas β (146 residuos cada una) con dos ejes de simetría. En un término más amplio, la hemoglobina es un tetrámero de 2 homodímeros (2 cadenas α de globina y 2 cadenas no α) ³⁰.

La estructura de la hemoglobina presenta niveles de complejidad. La estructura primaria está formada por la secuencia de aminoácidos ensamblada mediante uniones covalentes (enlaces peptídicos y puentes disulfuro). Al contrario, la estructura secundaria hace semejanza a la disposición de los aminoácidos en patrones estructurales típicos como la hélice α o la lámina β que son especialmente estables. La estructura terciaria se define como el plegamiento de la

estructura secundaria para dar lugar a una estructura globular o subunidad proteica. Finalmente, la estructura cuaternaria está constituida por la asociación de las diferentes subunidades de la proteína (Figura 1) ³⁰.

Figura 1. De la estructura primaria de la hemoglobina a la cuaternaria.



Extraída de: Lehninger. Principles of biochemistry. 5ª Edición. Nelson DL, Cox MM. W.H. Freeman and company (2008), p 92.

La estructura primaria constituida por la secuencia de aminoácidos, la secundaria generada por la disposición en α -hélice de los aminoácidos. Estructura terciaria originada por plegamiento de la cadena peptídica y la cuaternaria producida por la agrupación de las 4 globinas para conformar la molécula de hemoglobina. Dos tercios de los aminoácidos se disponen de forma helicoidal, confiriendo rigidez a la estructura, que se estabiliza por puentes de hidrógeno, la cual adopta en el espacio una forma típica debido a varios plegamientos. La función principal de cada subunidad consiste en proporcionar un ambiente estable para el grupo hemo, en el cual pueda producirse la reacción reversible de oxigenación ³¹.

Este papel lo desempeña la estructura terciaria de cada subunidad, que está compuesta por 8 segmentos helicoidales, replegados sobre sí mismos, formando una estructura globular compacta en el centro de la cual está situado el grupo hemo, dentro de un bolsillo hidrofóbico,

que mantiene el hierro en estado ferroso. Este bolsillo de unión al hemo está conformado por las hélices E y F de cada subunidad proteica ³⁰⁻³².

Estas hélices no se distribuyen al azar, sino que se encuentran orientadas siguiendo unas reglas estándar de empaquetamiento de hélices, con el objetivo de formar una estructura globular ³³.

Los residuos de aminoácidos situados hacia el interior de la molécula son apolares (alanina, valina, leucina, etc), mientras que los situados en la superficie en contacto con el medio acuoso exterior tienen cadenas laterales hidrófilas (lisina, glutamina). La exclusión de los residuos hidrófobos del medio acuoso mantiene fija la estructura globular. Finalmente, la hemoglobina tiene una estructura cuaternaria, resultado de la relación entre las cuatro cadenas polipeptídicas ³⁰.

La estructura cuaternaria de la hemoglobina mantiene unas fuertes interacciones entre subunidades desiguales. Las interfases $\alpha 1\beta 1$ y $\alpha 2\beta 2$ involucran a más de 30 residuos cada una, mientras que en las interfases $\alpha 1\beta 2$ y $\alpha 2\beta 1$ participan 19 aminoácidos. Las interacciones hidrofóbicas son las que predominan en todas las interfases pero también hay algunos puentes de hidrógeno y pocas uniones iónicas o puentes salinos ³⁰.

1.1.3 TIPOS DE HEMOGLOBINAS

Existen 6 tipos de cadenas de globina en la persona humana, alfa (α), zeta (δ), beta (β), delta (δ), gamma (γ) y épsilon (ϵ). Las 2 primeras se les denominada cadenas tipo α y constan de 141 aminoácidos mientras que las restantes llamadas β -like o cadenas no α se forman levemente más largas, constituidas por 146 aminoácidos. De las combinaciones entre cadenas de globina tipo α y no α se van a instaurar las diferentes hemoglobinas, en los períodos germinal, fetal, neonatal y adulto.

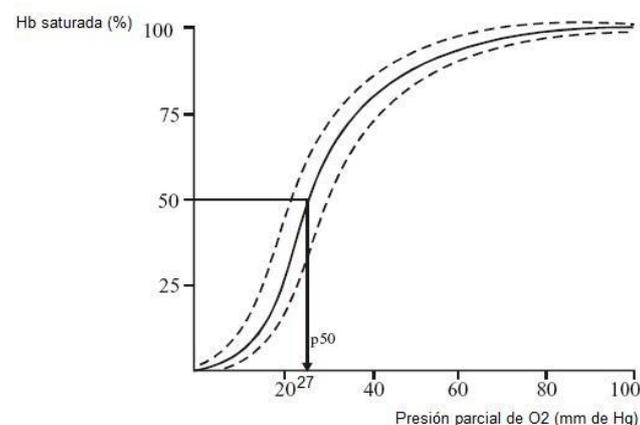
1.1.4 FUNCIÓN DE LA HEMOGLOBINA

La hemoglobina se localiza en el interior de los hematíes, siendo su función principal la de transportar oxígeno desde los pulmones hasta los tejidos, lo que se consigue gracias a la unión de la globina con el hemo, el cual se hace soluble, asegurándose un ambiente hidrófobo para captar de forma reversible el oxígeno. Para la realización de esta función es esencial el mantenimiento de la estructura y configuración espacial de la molécula de hemoglobina ^{31,32}.

La hemoglobina tiene dos formas cuaternarias distintas. La primera es la forma T (tensa) o desoxihemoglobina que es muy estable en ausencia de oxígeno a expensas de un gran número de pares iónicos en las interfases $\alpha 1\beta 2$ y $\alpha 2\beta 1$. La segunda, es la forma R (relajada) u oxihemoglobina que se estabiliza gracias al oxígeno. Además, el oxígeno al unirse a la hemoglobina induce el cambio de conformación de T a R, lo que provoca la ruptura de parte de los pares iónicos que estabilizan la forma T y la creación de otros nuevos ^{30,32}.

La curva de disociación del oxígeno es sigmoidea debido a la cooperatividad que existe para unir las moléculas de oxígeno (Figura 2) ³⁴, ya que estas favorecen un cambio conformacional (de la forma T a la forma R) en las subunidades restantes que facilita la unión de más moléculas de oxígeno en las otras subunidades. En consecuencia, la hemoglobina se comporta como una enzima alostérica ³⁰.

Figura 2 Curva de disociación de la hemoglobina.



Fuente Smilg et al, 2012

Se indica la forma de calcular la P50. Cuando la afinidad por el oxígeno está aumentada, la curva de disociación de la hemoglobina está desplazada hacia la izquierda y por lo tanto, la P50 está disminuida.

Por el contrario, se encuentran valores de P50 más altos en situaciones que impliquen una disminución de la afinidad por el oxígeno. En esos casos la curva de disociación de la hemoglobina está desplazada hacia la derecha.

Se define P50 como la presión parcial de O₂ para la cual la hemoglobina se encuentra saturada al 50%. En condiciones fisiológicas la P50 es de 26,6 mmHg. La P50 se entiende como una medida inversa de la afinidad, es decir, a mayor afinidad de la hemoglobina por el oxígeno, menor será el valor de la P50 y viceversa.

Existen diversos factores fisiológicos que regulan la afinidad de la hemoglobina por el oxígeno tales como el pH, la temperatura, la pCO₂ (presión parcial de CO₂) y la presencia de 2,3-difosfoglicerato (2,3-DPG).

La influencia del pH se conoce como efecto Bohr y consiste en una disminución de la afinidad a pH ácido, lo que facilita la liberación de oxígeno en los tejidos donde el ión bicarbonato y el ácido láctico acidifican el medio. Lo mismo ocurre con el aumento de temperatura o de pCO₂.

La presencia de 2,3-DPG aumenta el valor de la P50, es decir, disminuye la afinidad de la hemoglobina por el oxígeno. Este metabolito se sitúa entre las cadenas β y forma cuatro puentes salinos con aminoácidos polares de carga positiva que estabilizan la forma T pero no altera la capacidad alostérica de la hemoglobina. El 2,3- DPG se une a la hemoglobina en un lugar apartado del sitio de unión a oxígeno y por lo tanto, el 2,3-DPG ejerce una modulación alostérica heterotrópica ³⁰.

Asimismo, existen otros factores exógenos y endógenos que pueden alterar a la P50. Entre los que producen una descenso de la semejanza se encuentran el desarrollo de acomodación a la altura, el hipertiroidismo, la obstrucción pulmonar grave y la anemia crónica. Todas estas situaciones generan una adición del 2,3-DPG que es el encargado de la restricción de la semejanza. Además, existen hemoglobinas patológicas con la semejanza disminuida.

Por el contrario, los estados de shock o la sangre almacenada hacen que el 2,3- DPG disminuya y en resultado la semejanza de la hemoglobina por el oxígeno aumenta. De similar modo, hay variantes de hemoglobina patológicas con la semejanza aumentada.

1.1.5 ESTRUCTURA DE LOS GENES GLOBÍNICOS

Los genes globínicos son unidades activas que dan origen a las diferentes Hb's, se localizan en dos cromosomas distintos de acuerdo a la familia a la cual pertenecen, dichos genes se hayan limitados en su expresión de acuerdo al orden espacio-temporal en el cual es requerida su actividad para el correcto desarrollo del ser humano.

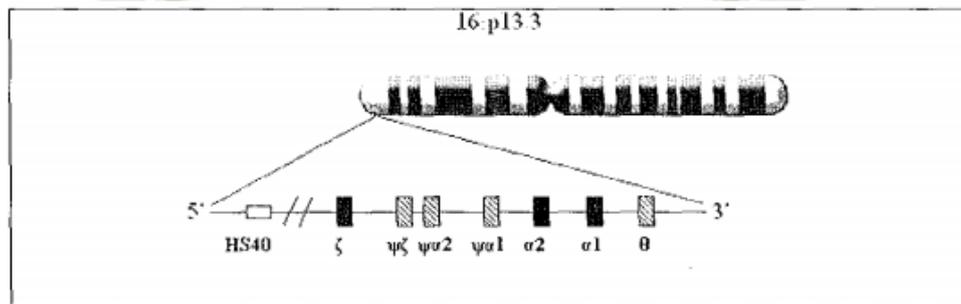
En total existen ocho genes globínicos funcionales y cinco genes inactivos o pseudogenes agrupados en dos familias³⁵: la Familia de genes globínicos alfa y la Familia de genes globínicos no alfa o semejantes a beta.

1.1.5.1 GENES GLOBÍNICOS α .

Esta familia se encuentra localizada en el cromosoma 16 p13.3, en un reglón de aproximadamente 40 Kb de longitud, está conformado por cuatro genes: ζ , $\alpha 2$, $\alpha 1$ y θ de los cuales solo los tres primeros son funcionales (Figura 3). El gen ζ únicamente se expresa en la etapa embrionaria y los genes $\alpha 2$ y $\alpha 1$ se expresan desde la etapa fetal hasta la fase adulta, mientras que el gen θ no se ha demostrado que forme una proteína estable. Además de los genes antes mencionados, en esta región se encuentran tres pseudogenes: $\psi\zeta$, $\psi\alpha 2$, y $\psi\alpha 1$. A 40 Kb

del gen embrionario ζ en el extremo 5' se encuentra una región denominada HS40 que está relacionada con la regulación de la expresión génica de esta familia, el arreglo de estos genes globínicos α es de manera ontogénica ³⁵.

Figura 3 Esquema de la familia de genes globínicos α . El recuadro en blanco representa la región HS40. Los recuadros en color negro indican los genes funcionales. Mientras que los recuadros sombreados representan a los pseudogenes o genes inactivos.



Fuente: Ibarra, 2001.

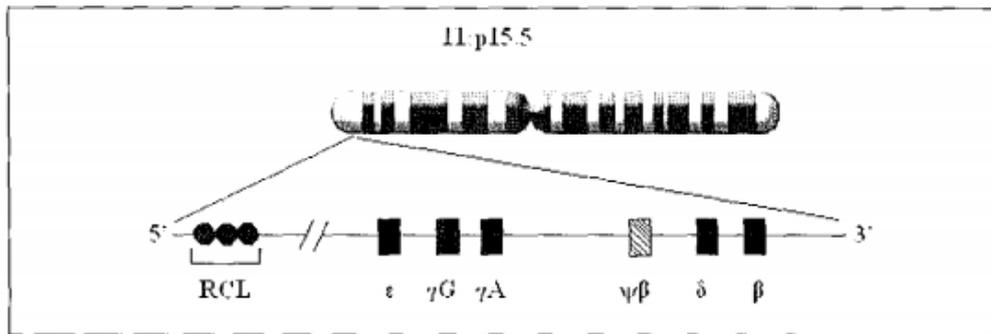
1.1.5.2 GENES GLOBÍNICOS β

La familia de los genes globínicos no α o semejantes a β se encuentra localizada en el cromosoma 11 en la región p15.5 y abarca un segmento de 70 Kb.

Dicha familia está constituida por cinco genes funcionales llamados ϵ , γ^G , γ^A , δ y β , además de un pseudogen denominado $\psi\beta$ que está localizado entre los genes γ^A y δ (Thein, 2005). Alrededor de 10 a 20 Kb hacia el extremo 5' del gen globínico ϵ existe una zona conformada por cuatro sitios hipersensibles a la enzima DNA_{asa} , conocida como región controladora del locus o RCL (Figura 4) que tiene tres funciones:

- 1) Organizar el dominio de cromatina activa en células eritroides para la expresión ontogénica de los genes globínicos;
- 2) Regula la actividad transcripcional de forma espacio temporal y,
- 3) Actuar como un amplificador de la transcripción de dichos genes ⁴².

Figura 4 Esquema de la familia de genes globínicos beta. Los círculos en negro representan la Región Controladora del Locus (RCL), los recuadros en negro indican los genes funcionales, el recuadro sombreado al gen inactivo ó pseudogen.



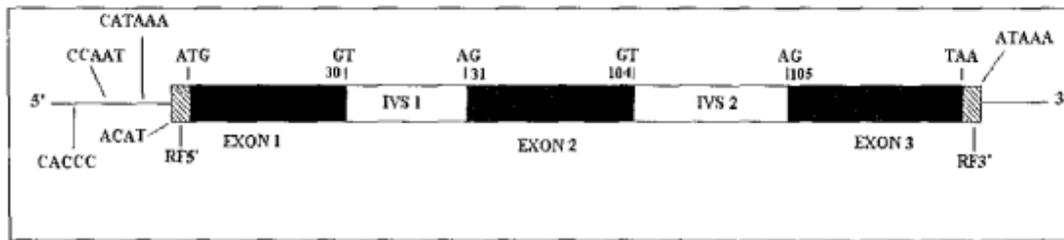
Fuente: Thein, 2005

1.1.5.3 ESTRUCTURA INTERNA DE LOS GENES GLOBÍNICOS

El genoma humano es una molécula compleja que depende de la interacción coordinada de diversas unidades como son la región promotora, las regiones flanqueadoras (RF) y secuencias consenso para que éste sea capaz de realizar procesos como la transcripción, de igual manera su organización establece las condiciones en las cuales será expresado.

La región promotora está localizada en el extremo 5' de los genes globínicos (genes alfa y beta), esta región no se transcribe y resulta indispensable para que la expresión de éstos sea normal, el promotor de un gen globínico está formado por las cajas consenso (Figura 5) que son elementos cis altamente conservados, estas cajas consenso son las secuencias 5'-CATAAA-AA-3' conocida como la caja TATA, la secuencia 5'-CCAAT-3' y la secuencia 5'-C(A ó T)CCC-3', estas cajas participan en procesos importantes como son la apertura de la doble hélice y unión del complejo de la transcripción así como para la unión del factor transcripcional específico de células eritroides conocido como EKLF³⁵. Se ha demostrado que el sitio para unión del EKLF resulta indispensable para que se lleve a cabo la transcripción del gen globínico β , y adicionalmente se ha demostrado que resulta necesario para la activación de los genes globínicos ϵ y γ ³⁶.

Figura 5 Estructura interna de un gen globínico β . En el extremo izquierdo, sobre la línea se señalan las cajas consenso en la región promotora de un gen globínico: 5'-CATAAA-3', 5'- CCAA T-3', 5'-CACCC-3'. En la RF5' se ubica la secuencia 5'-ACA T-3'. Los recuadros en color negro representan los exones mientras que los recuadros claros representan los intrones. En el límite de la RF5' y el exón 1 el codón de inicio de la traducción A TG, mientras que en el límite del exón 3 y fa RF3' se encuentra el codón de terminación TAA. Las secuencias GT y AG establecen fas fronteras entre los exones y Jos intrones. En la RF3' se encuentra la secuencia 5'-ATAAA-3'.



Fuente: Ibarra, 2001.

Como se muestra en la Figura 5 las regiones flanqueadoras, que también se conocen como secuencias adicionales, no se traducen y forman parte del RNA nuclear y mensajero. La RF5' contiene la secuencia consenso 5'-ACA(T ó C)- 3', mientras que en la RF3' se encuentra la secuencia consenso 5'-AATAAA-3' cuya función es señalar el sitio de unión de la enzima poliadenilato sintetasa ³⁵.

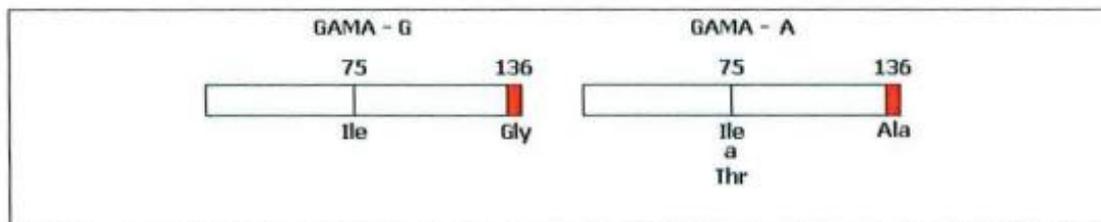
En todos los genes globínicos existen 3 exones los cuales tienen como función codificar la secuencia de aminoácidos (aa). De igual manera que los exones en los genes globínicos existen intrones (IVS) de éstos solo existen dos (Figura 5) a los que también se les conoce como secuencias intercaladas, que se encargan de separar a los exones en el DNA y RNAhn ³⁵.

1.1.5.4 GENES GLOBÍNICOS y

Los genes globínicos γ^G y γ^A han sido motivo de estudio debido a que dentro de la región promotora de estos genes se han encontrado variantes en la secuencia de nucleótidos que han sido denominadas polimorfismos, estas variantes resultan de interés para entender las variaciones en el fenotipo clínico de P-Talasemia y Anemia Drepanocítica (AD) ³⁷. Los genes globínicos y conforman la Hb F sintetizan dos tipos de globina los cuales se diferencian entre sí en la posición 136 de las cadenas γ^G está presente una glicina mientras que en las cadenas γ^A está presente alanina; además la cadena γ^A presenta una variante en la posición 75 que resulta

en el cambio de isoleucina por treonina como se muestra en la Figura 6 ³⁸. En humanos la expresión de los genes globínicos γ dan origen a aproximadamente al 1 % de Hb F, en pacientes con Persistencia Hereditaria de Hemoglobina Fetal (PHHF) existe un incremento en los niveles basales de globinas γ del 5 - 20% ³⁹.

Figura 6 Diferencias bioquímicas entre las globinas γ^G y γ^A . A la izquierda la globina γ^G se muestra el aa presente en las posiciones 75 y 136. En la globina γ^A se puede observar que en la posición 75 puede estar presente una Isoleucina o Treonina, mientras que en la posición 136 se aprecia una Alanina.

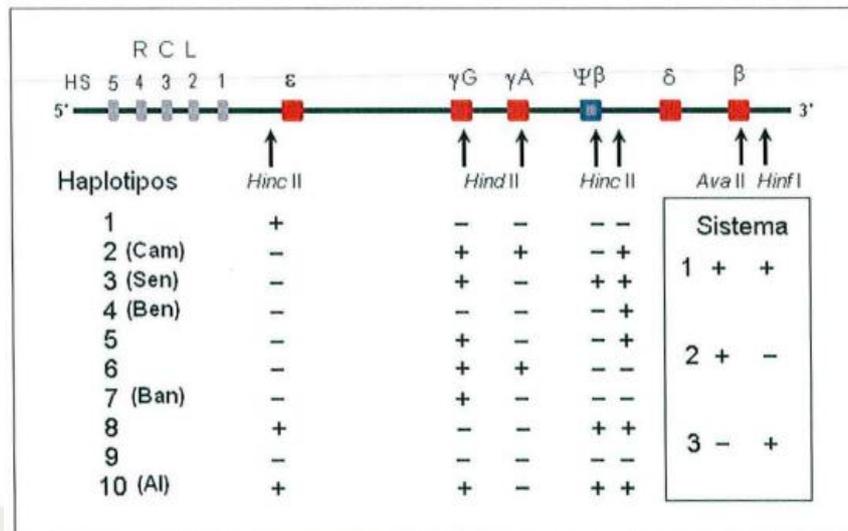


Fuente: Bunn, 1986

1.1.5.5 POLIMORFISMOS DE LOS GENES GLOBÍNICOS

Estudios poblacionales realizados en la secuencia de los genes globínicos ha mostrado variantes en la secuencia de nucleótidos detectadas por enzimas de restricción, que como se mencionó anteriormente, han sido denominadas polimorfismos al observarse con una frecuencia mayor al 1.0%. Se han descrito cerca de 17 sitios de restricción en la familia de los genes globínicos p de éstos se han seleccionado 7 para formar el denominado haplotipo (Hp), de los cuales 5 sitios forman en el subhaplotipo 5' y dos el subhaplotipo 3' (Figura 7). Un Hp es la combinación de los sitios polimórficos que conforman una región específica de un gen y son una herramienta útil que nos puede ayudar a determinar la relación genética entre individuos, reconstruir la historia de una mutación desde su origen y seguir las rutas de migración de poblaciones ³⁹.

Figura 7 Región del globínico β humano. Se muestran 10 Hp que son identificados por 7 enzimas de restricción diferentes.



Fuente: Flint, 1998.

1.2 HEMOGLOBINOPATÍAS

Las hemoglobinopatías son alteraciones cualitativas o cuantitativas de la Hb originada por mutaciones genéticas, cuyas consecuencias pueden ser:

- Una modificación estructural debida al reemplazo de un aminoácido por otro en cualquiera de los exones de los genes globínicos,
- Disminución de la síntesis de una cadena globínica estructuralmente normal (talasemias),
- Alteraciones que modifican el reemplazo de Hb F por Hb A en la etapa adulta ⁴⁰.

Se consideran hemoglobinopatías solo aquellas mutaciones que afectan regiones esenciales de la molécula y que, por tanto, poseen expresividad clínica ⁴¹. Se ha estimado que cerca del 7% de la población mundial es portadora de alguna hemoglobinopatía y que cada año nacen en el mundo aproximadamente 300,000 - 400,000 niños con formas severas de estas enfermedades⁴².

1.2.1 HEMOGLOBINOPATÍAS ESTRUCTURALES

Son el resultado de mutaciones al nivel de alguno de los genes que codifican la síntesis de una determinada cadena globínica: α , β , δ , ϵ y ζ . La alteración estructural puede ser explicada por la sustitución de una sola base en el triplete del codón correspondiente en el DNA del gen globínico. Unas pocas variantes tienen reemplazado un aminoácido en diferentes sitios dentro de la misma subunidad ³⁸.

Las variantes estructurales raras surgen a consecuencia de la delección de un nucleótido de un codón, lo que da lugar a un cambio en el marco de la lectura de los codones en sentido descendente, con la consiguiente delección de uno o más residuos de aminoácido, o a la inversa, con extensión de la cadena polipeptídica por alterarse el codón final ⁴³. Se reconocen tres tipos de alteraciones estructurales que serán descritas brevemente a continuación:

1.2.1.1 HEMOGLOBINAS INESTABLES

Son las variantes caracterizadas por una alteración estructural que causa una disminución de la estabilidad de la molécula, por lo que estas Hb pueden precipitar en los eritrocitos formando cuerpos de inclusión conocidos como cuerpos de Heinz ⁴⁴. Estas células poseen una menor capacidad de deformación y son eliminadas con mayor facilidad de la circulación, con el desarrollo de una anemia hemolítica ⁴³.

En la actualidad más de 100 diferentes variantes de Hb inestable han sido reportadas⁴⁵ de las cuales alrededor del 85 % corresponden a mutaciones en el gen β , hay cinco clases de mutaciones que conducen a la inestabilidad de la globina³⁵:

- Sustituciones en la cavidad del heme.
- Mutaciones que interfieren con la estructura secundaria.
- Inserción de residuos polares que interfieren con interacciones hidrófobas.
- Anormalidades estructurales.

- Interferencia con los contactos interglobinas.

1.2.1.2 ALTERACIÓN EN LA AFINIDAD POR EL O₂

La Hb tiene 2 estructuras cuaternarias: una para la desoxi-hemoglobina y otra para la oxihemoglobina. La sustitución de un residuo de aminoácido, situado en la zona de contacto $\alpha_1\beta_2$ en los alrededores del grupo hemo o en el lugar de unión del 2-3 DPG puede causar un desplazamiento del equilibrio alostérico, dando lugar a este tipo de variante de Hb. Cuando este equilibrio se desplaza hacia la forma "relajada" hay aumento de la afinidad; por el contrario, cuando la forma "tensa" es la favorecida hay disminución en la afinidad por el O₂⁴⁴.

1.2.1.3 METAHEMOGLOBINA

Es el derivado oxidado de la Hb en que el Fe del heme se encuentra en la forma férrica (Fe⁺³) y no en la forma ferrosa normal (Fe⁺²), en esta situación, el Fe es incapaz de unirse reversiblemente con el O₂ y, por lo tanto, de transportar el gas. La metahemoglobina es funcionalmente inútil⁴⁴. Pequeñas cantidades de metahemoglobina se forman de manera continua en los eritrocitos normales, debido a la interacción de la Hb con el ión superóxido⁴³.

1.2.2 SÍNDROMES TALASÉMICOS

Constituyen un grupo muy heterogéneo de trastornos caracterizados por alteraciones cuantitativas de la síntesis de cadenas polipeptídicas α o β de la Hb³⁸. La disminución de la síntesis de cadenas α se denomina α -talasemia, mientras que la disminución de cadenas β se denomina β -talasemia, características comunes a todos los síndromes talasémicos son las siguientes:

- Trastorno global de la síntesis de Hb, lo que da lugar a unos eritrocitos microcíticos e hipocrómicos.

- Producción desequilibrada de las cadenas de globina: ello conduce a la formación de tetrámeros de una sola cadena y niveles elevados de Hb A₂ , Hb F o de ambas en β -talasemia ⁴³.

1.3 MUTACIONES

el concepto de mutación se entiende como un cambio heredable repentino en el material genético que no puede justificarse a través de la segregación o la recombinación⁴⁶.

Las mutaciones son cambios repentinos en la estructura genética, siendo por ello hereditarios. Desde hace mucho tiempo se está tratando de inducir mutaciones bajo control experimental, para así lograr nuevas características hereditarias en una población⁴⁷.

Las mutaciones son la última fuente de variabilidad en los organismos; la variabilidad en el caso de las mutaciones inducidas no es necesariamente diferente a aquella variabilidad causada por las mutaciones espontáneas durante la evolución ⁴⁸.

Para que una mutación sea detectada, debe ocurrir algún cambio fenotípico en la planta. Un cambio visible en alguna característica morfológica como altura de planta, color de pericarpio, deficiencia de clorofila, entre otros que se identifican más fácilmente ⁴⁹.

Una mutación génica es todo cambio en la secuencia nucleotídica de un gen, que no proviene de una recombinación normal. Estos cambios dan origen a un gen mutante, cuya expresión será una proteína anormal; sin embargo, también existe una enorme variedad de cambios en la nucleotídica que no conllevan alteraciones en la expresión génica y que constituyen los polimorfismos o variaciones normales entre la población ⁵⁰.

Las mutaciones no impulsan la evolución, pero proporcionan la diversidad genética sobre la cual actúan la selección natural y otros agentes ⁵¹.

La mutación es un cambio heredable del material genético, los individuos que muestran estos cambios heredables reciben la denominación de mutantes⁵².

1.3.1 TIPOS DE MUTACIONES

1.3.1.1 POR SU ORIGEN

Las mutaciones se clasifican según su origen en espontáneas o inducidas.

la mutación espontánea es aquella que ocurre en la naturaleza, mientras que una mutación inducida es la que resulta de la acción de un agente mutagénico o mutágeno ⁴⁹.

las mutaciones espontáneas implican cambios permanentes en el material genético que ocurren sin influencia externa; en cambio, las inducidas ocurren cuando algún agente externo a la célula -un mutágeno- causa un cambio permanente en el ADN ⁵¹.

Las mutaciones se generan de manera espontánea en la naturaleza, pero en una frecuencia reducida (alrededor de 10^{-5} y 10^{-8}). Para poder incrementar la tasa de ocurrencia (10^{-6} y 10^{-4}) se pueden utilizar diversos agentes mutagénicos químicos y/o físicos ⁵³⁻⁵⁸.

La mutación inducida debe considerarse un complemento para incrementar la cantidad de variación natural existente, en particular cuando se haya reducido excesivamente la variación genética por una intensa selección o cuando no se encuentra lo que se desea ⁵⁵.

1.3.1.2 POR LA CANTIDAD DE DAÑO EN EL MATERIAL

Se clasifica las mutaciones en tres tipos: a) aquellas en las que ocurre cambio en la estructura química del gen (mutaciones genéticas o puntuales), b) las que implican alteraciones cromosómicas (traslocaciones, inversiones, duplicaciones) y c) las que implican cambios en el número de cromosomas (mutaciones genómicas)⁵⁹.

a) Mutaciones génicas o puntuales

Son aquellas mutaciones que afectan a una sola base o bien a un número relativamente pequeño de bases (hasta algunos miles) y que requieren técnicas de análisis genético molecular para su detección ⁶⁰. Dentro de este tipo de mutaciones se incluyen:

las adiciones, deleciones y las sustituciones. En estas últimas podemos distinguir las transiciones (cambios de una pirimidina a otra (C a T, o de T a C) o de una purina a otra (de A a G, o de G a A) y las transversiones (cambio de purina (A o G) a pirimidina (T o C) o viceversa).

Las sustituciones provocan la alteración de un solo triplete, no suelen ser perjudiciales a no ser que se forme un codón sin sentido o afecte a un aminoácido del centro activo de la proteína. Las deleciones y adiciones son mutaciones más graves que las sustituciones porque todos los tripletes de bases estarán cambiados a partir del punto en el que se ha producido la mutación y por tanto el mensaje codificado será totalmente distinto ⁶¹.

Figura 8 Tipos de mutaciones génicas.

	...-A-T-G-C-A-T-G-T-A-C-C-...	
	...-T-A-C-G-T-A-C-A-T-G-G-...	
SUSTITUCION	DELECIÓN	ADICIÓN
...-A-T-G-C-A-G-G-T-A-C-C-...	...-A-T-G-C-A-G-T-A-C-C-...	...-A-T-G-C-A-T-C-G-T-A-C-C-...
...-T-A-C-G-T-C-C-A-T-G-G-...	...-T-A-C-G-T-C-A-T-G-G-...	...-T-A-C-G-T-A-G-C-A-T-G-G-...

FUENTE: Fita et al, 2008

b) Mutaciones cromosómicas (Aberraciones estructurales)

Las mutaciones cromosómicas son aquellas que afectan una parte de uno o varios cromosomas o bien que implican un reordenamiento o traslocación⁶⁰. El mecanismo causante de las mutaciones cromosómicas es la recombinación no homóloga entre los cromosomas, ocurriendo con una frecuencia de 6×10^{-4} / división celular.

Tanto las aberraciones estructurales como numéricas de los cromosomas rara vez se perpetúan debido a que suelen ser incompatibles con la supervivencia y/o reproducción.

Las mutaciones en la estructura cromosómica resultan de la rotura de un cromosoma seguida de la reconstitución en una combinación anormal. Entre los cambios que afectan la estructura

de los cromosomas se pueden encontrar distintos tipos de reordenamientos: deleciones, inserciones, inversiones y translocaciones⁶².

- **Deleciones:** Son aquellos cambios en los que, tras la ruptura del cromosoma, se produce la reunión de los fragmentos con pérdida de un segmento cromosómico y consecuentemente pérdida de información. Algunas consisten en pérdida de tan solo uno o pocos nucleótidos, mientras que en otros casos se pueden perder miles de nucleótidos, afectando la secuencia de un gen o a un grupo de genes contiguos.
- **Inserciones:** Corresponde a la ganancia de material genético producto de la repetición de un segmento cromosómico, ya sea en el mismo cromosoma o en otro ⁵⁰.
- **Inversiones:** Se producen por dos roturas dentro de un mismo cromosoma y la unión invertida del segmento involucrado⁵⁰.
- **Translocaciones:** En ellas se produce el cambio de posición de un fragmento cromosómico. Si la translocación se produce dentro del mismo cromosoma se denomina translocación homóloga; cuando el fragmento cromosómico es trasladado y colocado en otro cromosoma se denomina translocación no homóloga o recíproca⁶².

c) Mutaciones genómicas (aberraciones numéricas)

Las mutaciones genómicas como aquellas que implican la ganancia o pérdida de todo un cromosoma (o de varios) entero. El mecanismo de las mutaciones genómicas es la no disyunción en la meiosis pero se desconoce la causa concreta. La frecuencia de las mutaciones genómicas es de 10^{-2} / división celular. Cuando el número de cromosomas es múltiplo exacto del número haploide se denominan con la terminación “-ploidía” (triploidía, tetraploidía). Cuando el número de cromosomas no es múltiplo exacto del número haploide se denominan aneuploidías (trisomía, monosomía) ⁶².

1.3.1.3 POR EL TIPO DE CÉLULA A LA QUE AFECTA

Las mutaciones pueden afectar a las células somáticas o a las células germinales (“reproductivas”). En el primer caso afectará solo al individuo que la padece, no se transmitirá a la descendencia y no habrá ninguna manifestación en la apariencia externa de la persona, salvo que tenga efecto dominante. Si afecta a las células germinales o reproductivas, puede transmitirse a la descendencia, que portara la mutación tanto en sus células germinales como en las somáticas⁶³.

1.4 BIOINFORMATICA

Existen discusiones alrededor de la definición y alcance de términos como bioinformática, biología computacional, química computacional, etc., de las cuales no participaremos en esta tesis. En un sentido amplio, puede entenderse el área de la bioinformática como el conjunto de técnicas computacionales para el almacenamiento, análisis y procesamiento de cualquier tipo de datos biológicos.

Si nos remitimos a los orígenes de la bioinformática como campo de estudio, podemos hablar de que los primeros algoritmos de procesamiento de información biológica estén probablemente asociados al tratamiento de secuencias de ADN, ya que el volumen de los mismos hizo impráctico su manejo manual a principios de los años setenta, época en la que otros datos que actualmente son procesados con técnicas bioinformáticas no estaban disponibles o eran más bien escasos⁶⁴.

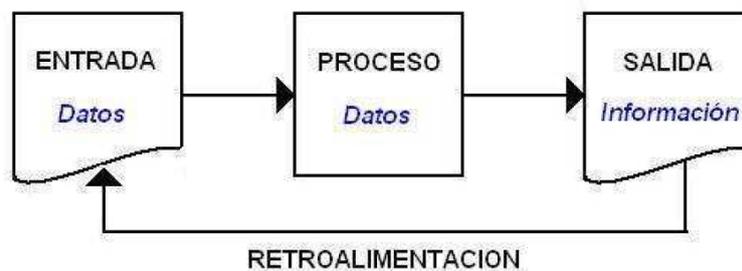
En la actualidad, el principal problema al que se enfrenta la bioinformática como campo, es la integración y extracción de información combinando de diferentes fuentes de datos de origen heterogéneo⁶⁵.

A lo largo del presente trabajo, manejaremos diferentes tipos de datos (que servirán como insumo de cada uno de los desarrollos realizados) los cuales comparten las siguientes características:

- Son almacenables
- Tienen una sintaxis precisa
- Son legibles con una semántica no ambigua
- Describen objetos o entidades biológicas mediante alguna o algunas de sus características fisicoquímicas.

Los primeros tres elementos se refieren todos aquellos datos que pueden ser procesados por una computadora mediante técnicas informáticas. El último agrega el componente biológico. Esos datos sufrirán combinaciones y transformaciones que darán lugar a nuevo conocimiento, plausible de ser utilizado tanto para validar teorías, como para alimentar modelos predictivos. En desarrollos bioinformáticos correctos y consistentes, los datos de salida deben poseer características recién enumeradas, para servir como insumo de futuros procesos, y que los métodos utilizados sean replicables, como muestra la figura 9.

Figura 9: Diagrama de flujo del procesamiento computacional de la información como sistema replicable, que en esta tesis tendrá como datos e información la descripción fisicoquímica y contextual de entidades biológicas.



Fuente: Radusky, 2017

1.5 BASE DE DATOS BIOLÓGICAS

Una base de datos, como concepto general, es un banco de información que contiene registros desglosados en propiedades y vinculados de una manera que permite relacionarlos como conjunto, ejercer sobre dicho conjunto búsquedas en función del valor de las propiedades, clasificarlos en función de filtros a las mismas, etc. En la actualidad, la mayoría de las bases de datos se encuentran almacenadas de una manera digital porque esto permite su rápido procesamiento debido a la capacidad de cómputo de los dispositivos en la actualidad.

Los programas que permiten interactuar de manera computacional con bases de datos digitales se denominan "gestores de bases de datos".

Podemos definir como base de datos biológica cualquier colección de información cuyos registros consistan en datos acerca de entidades biológicas. La misma puede provenir de experimentos científicos, literatura publicada, tecnología de experimentación de alto rendimiento, análisis computacional, etc.

Una base de datos biológica puede contener información de muy variadas áreas de investigación incluyendo genómica, proteómica, metabolómica, expresión génica mediante microarrays, filogenética y un largo etcétera. La información contenida en bases de datos biológicas incluye funciones, estructura y localización (tanto celular como cromosómica), efectos clínicos de mutaciones, así como similitudes de secuencias y estructuras tridimensionales de moléculas⁶⁶.

1.5.1 HbVar

HbVar (<http://globin.bx.psu.edu/hbvar>) es una de las bases de datos específicas de locus más antiguas y apreciadas, lanzada en 2001 por un esfuerzo académico multicéntrico para proporcionar información oportuna sobre las alteraciones genómicas que conducen a variantes de hemoglobina y todo tipo de talasemias y hemoglobinopatías. Los registros de la base de

datos incluyen descripciones fenotípicas extensas, efectos bioquímicos y hematológicos, patología asociada y ocurrencia étnica. Informa actualizaciones de > 600 entradas de HbVar, inclusión de datos específicos de la población para 28 poblaciones y 27 grupos étnicos para α - y β -talasemias y opciones de consulta adicionales en la página de consulta de HbVar. El contenido de HbVar también está interconectado con otras dos bases de datos genéticas establecidas, a saber, FINDbase (<http://www.findbase.org>) y la base de datos Leiden Open-Access Variation (<http://www.lovd.nl>), que permite datos comparativos consulta y análisis⁶⁷.

1.6 PROCESAMIENTO DE DATOS BIOLÓGICOS

1.6.1 Alineamiento de secuencias

Uno de los principales procesos en el estudio de la biología, es el de la comparación. Las secuencias (de ADN, ARN y proteínas), por supuesto, no escapan a este proceso. La forma más básica en la que pueden compararse secuencias es de a pares.

Un alineamiento de a pares es una forma de representar y comparar dos secuencias de proteínas (o de ADN, o de ARN) para resaltar sus zonas de similitud. Las mismas, podrían indicar relaciones funcionales y evolutivas entre las proteínas involucradas. Cuando dos proteínas comparten un ancestro común, las posiciones del alineamiento que no coinciden pueden interpretarse como mutaciones puntuales (sustituciones), y los huecos como indels (mutaciones de inserción o deleción) introducidas en el proceso en el que se produjo la divergencia evolutiva. En el alineamiento de secuencias proteicas, el nivel de semejanza entre los aminoácidos que ocupan un lugar concreto en la cadena puede interpretarse como una medición aproximada de conserva en una zona específica, o cadena motivo, entre linajes.

La extensión natural de la comparación (alineamiento) de a pares, es la que se realiza para conjuntos de secuencias, lo que da lugar a alineamientos múltiples. El problema de alinear

múltiples secuencias (MSA, por sus siglas en inglés) despertó tempranamente, en la década de 1970, el interés de la comunidad por las inferencias que era posible extraer, en la medida que la cantidad de secuencias disponibles fue siendo más voluminosa. La misma involucra el alineamiento de tres o más secuencias y busca encontrar posiciones equivalentes a lo largo de cada una de las secuencias alineadas (figura 10).

Figura 10: Visualización de un alineamiento de secuencias múltiple en el software Clustal 19, de secuencias relacionadas evolutivamente. Las posiciones están coloreadas por características fisicoquímicas similares, permitiendo interpretar amigablemente sustituciones que provocan pequeños o grandes cambios por pertenecer a grupos de aminoácidos de distintas características.

```

RLA0_METVA --MIDAKSEHKIAPWKIEEVNALKLLKSANVIALIDMMEVPAVQLQEIIRDK
RLA0_METJA ---METKVKAHVAPWKIEEVKTLKGLIKSKPVAIVDMDVPAVQLQEIIRDK
RLA0_PYRAB -----MAHVAEWKKKEVEELANLIKSYPPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRHO -----MAHVAEWKKKEVEELAKLIKSYPPVIALVDVSSMPAYPLSQMRRL
RLA0_PYRFU -----MAHVAEWKKKEVEELANLIKSYPPVALVDVSSMPAYPLSQMRRL
RLA0_PYRKO -----MAHVAEWKKKEVEELANLIKSYPPVIALVDVAGVPAYPLSKMRDK
RLA0_HALMA MSAESERKTETIPEWKQEEVDATVEMIESYESVGVVNIAGIPSRLOQDMRRD
RLA0_HALVO MSESEVRQTEVIPQWKREEVDELVDIESYESVGVVAGIPSRLOQSMRRE
RLA0_HALSA MSAEEQRTTEEVPEWKRQEVVELVDLLETYDSVGVVNVGTIPSKLOQDMRRG
RLA0_THEAC -----MKEVSQKKELVNEITRIKASRSVAIVDTAGIRTRQIQDIRGK
RLA0_THEVO -----MRKINPKKKEIVSELAODITKSKAVAVDIKGVTRROMQDIRAK
RLA0_PICTO -----MTEPAQWKIDFVKNLENEINSRKVAIVSIKGLRNNFQKIRNS
    
```

Fuente: Radusky, 2017

Un alineamiento múltiple puede proveer información de peso acerca de la relación estructural y/o funcional dentro de un conjunto de secuencias de proteínas (por ejemplo: la conservación evolutiva de aminoácidos importantes estructural o funcionalmente se corresponde con patrones en determinadas regiones de las secuencias). Pueden también ser útiles, por ejemplo, siendo almacenados como modelos ocultos de Markov2 0 para la asignación de pertenencia de una nueva proteína a una familia de proteínas, permitiendo en muchos casos inferir, por ejemplo, su función o su dominio de plegado.

Los alineamientos múltiples son un intento de representar secuencias relacionadas evolutivamente en una forma consistente. Encontrar el alineamiento óptimo en base a un modelo evolutivo dado es equivalente a maximizar la probabilidad de que las secuencias hayan evolucionado de la forma en la que el alineamiento indica ⁶⁸.

Existen rasgos críticos que definen la composición y actividad de la proteína. La zona activa de un catalizador, por ejemplo, requiere que determinados residuos de aminoácidos tengan una disposición tridimensional precisa. Además, una interfaz de unión proteína-proteína puede constituir de una amplia área con restricciones en la hidrofobicidad o polaridad de los residuos de aminoácidos. Las regiones funcionalmente restringidas de las proteínas evolucionan más lento que las regiones sin restricción, como bucles superficiales, dando parte a bloques discernibles de secuencias conservadas cuando se compara las secuencias de una estirpe de proteínas. Esos bloques son designados como "motivos" y proteínas que pertenecen a la misma estirpe muestran un alto conservación en estos motivos, presentando habitualmente una estructura tridimensional (plegado) análogo, y un mismo desempeño.

1.6.2 Determinación de la estructura proteica

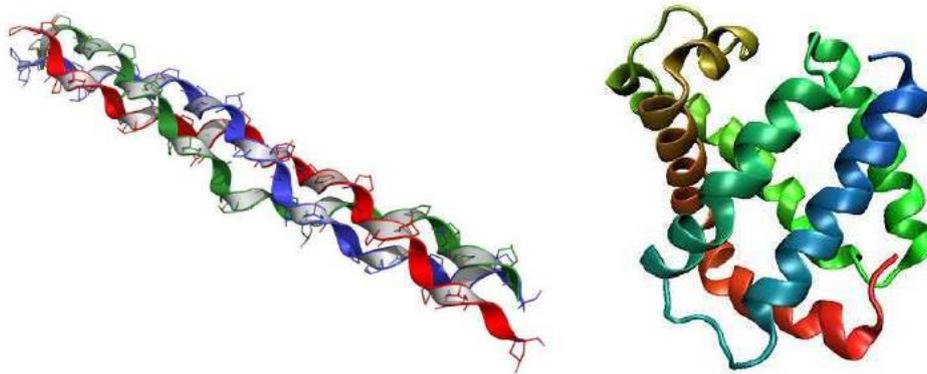
La estructura tridimensional de una proteína es la disposición en el espacio de los átomos que la componen, provenientes de su secuencia de aminoácidos. Existen diferentes niveles de estructuración de una proteína, influyendo los niveles inferiores en la disposición de los niveles superiores.

La estructura primaria está definida por la secuencia de aminoácidos que componen a la proteína, unidos covalentemente mediante el enlace peptídico. El orden de los aminoácidos es consecuencia del material genético: cuando se traduce el RNA se obtiene el orden que va a dar lugar a la secuencia de la proteína.

La estructura secundaria es la disposición espacial local que ocupa la "columna vertebral" (backbone) de la proteína, y se determina mediante la conformación de enlaces tipo puente de hidrógeno. Existen estructuras secundarias bien definidas y ordenadas, como alfa-hélices, hojas-beta y regiones desestructuradas en cuanto a su estructura secundaria.

La estructura terciaria se define como la estructura que adopta la cadena polipeptídica en el espacio. El modo en que la secuencia de aminoácidos se pliega en el espacio (de forma globular, como fibra, etc) determinará el o los dominios de plegado que pueden ser asignados a la proteína (figura 11).

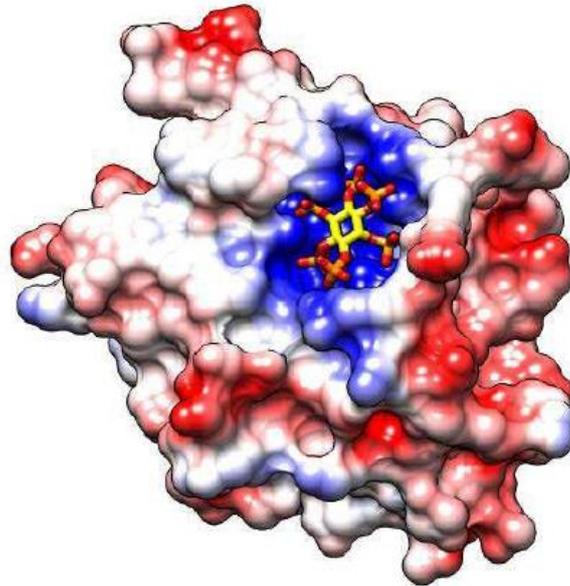
Figura 11: Visualización de dos estructuras de proteína. A la izquierda una proteína cuyo plegamiento es fibrilar y la derecha una cuyo plegamiento es globular.



Fuente: Radusky, 2017

El plegado suele realizarse de modo tal que los aminoácidos apolares que componen la proteína estudiada se sitúan hacia el interno y los polares hacia el externo, en medios acuosos. Esto provoca un equilibrio por interacciones hidrofóbicas, de fuerzas de van der Waals y de puentes disulfuro (covalentes, entre aminoácidos de cisteína convenientemente orientados) y mediante enlaces iónicos (figura 12).

Figura 12: Visualización de la estructura de una proteína coloreando cada región en base a su densidad de carga, siendo las regiones más polares las de un rojo más fuerte y las regiones más apolares las de un azul más fuerte. Como puede observarse, los aminoácidos se organizan de manera tal que las zonas expuestas hacia el exterior son las más polares. Regiones apolares pueden constituir una zona expuesta al solvente con la consecuencia de constituir zonas de probable acoplamiento a compuestos que estén disponibles en el medio en el cual está inserta la proteína.

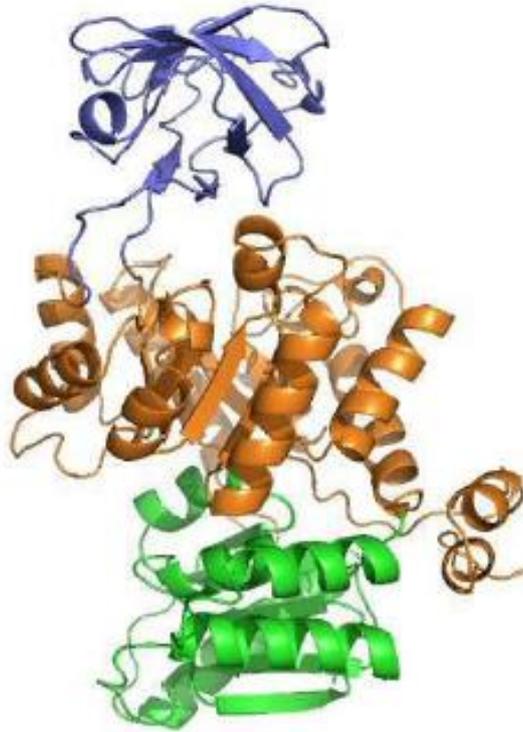


Fuente: Radusky, 2017

Se denomina dominio estructural a una parte constitutiva de la organización de las proteínas que estabiliza su plegado de modo autónomo. Los dominios, habitualmente, son seleccionados evolutivamente puesto que poseen un desempeño característico de la fisiología de la proteína que pertenecen.

La estructura cuaternaria se encuentra definida por diferentes cadenas peptídicas que se pliegan formando dominios que interactúan de una manera particular formando multímeros (figura 13), el cual posee propiedades distintas a la de los monómeros que la conforman.

Figura 13: Visualización de la entrada del Protein Data Bank (PDB)[CITA] 1pkm, la cual está compuesto por tres dominios estructurales que han sido remarcados con distintos colores.



Fuente: Radusky, 2017

Alrededor del 90% de las estructuras de las proteínas conocidas actualmente han sido determinadas mediante cristalografía de rayos X. Este método permite medir la densidad de distribución de los electrones de la proteína en las tres dimensiones del espacio, determinando de esta forma las coordenadas de los átomos relativas a las demás posiciones con certeza.

El Protein Data Bank (PDB) nace en el año 1971 en Brookhaven National Laboratory conteniendo únicamente 7 estructuras cristalográficas. Es en la actualidad el repositorio internacionalmente admitido de estructuras tridimensionales de macromoléculas como proteínas, ácidos nucleicos y sus complejos: con lípidos, azúcares, como asimismo además con diversos ligandos. En su última lectura cuenta con ~115.000 estructuras depositadas.

La obtención de cristales y la resolución de estructuras a partir de experimentos de difracción de rayos X enfrentan una limitación práctica; en la medida en que la cantidad de datos de secuencias generados crecen de manera exponencial, gracias a nuevas técnicas de secuenciación, la cantidad de estructuras crecen de manera lineal. Esta limitación, produce en la práctica una brecha que se amplía día a día, entre la cantidad de secuencias conocidas y aquellas para las que se conoce (de manera experimental) su estructura tridimensional.

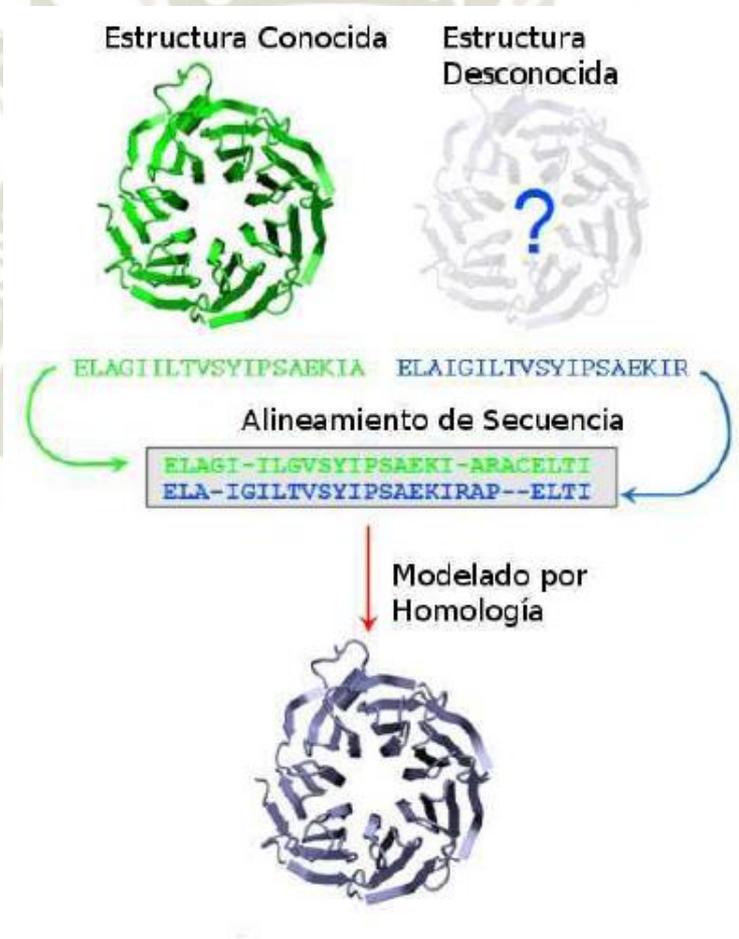
El uso de algoritmos para la determinación mediante métodos computacionales de estructuras de proteínas de la que sólo tenemos su información de secuencia es denominado modelado. Existen en la actualidad dos estrategias usadas mayoritariamente para modelar estructuras de proteína: modelado ab-initio y modelado por homología.

Los métodos ab-initio, que también pueden encontrarse bajo el nombre de métodos de-novo, comparten la estrategia de intentar generar la estructura de la proteína de interés basándose únicamente en principios fisicoquímicos y teniendo como dato de entrada únicamente la secuencia de aquella molécula que se intenta modelar. Fundamentalmente lo que computa el algoritmo es, mediante el uso de algún potencial definido por el programa que estemos usando, las transiciones que ocurren del estado desplegado de la proteína hasta su estado plegado final mediante dinámica molecular o muestreo del tipo Monte Carlo. Estos métodos poseen la desventaja de requerir grandes tiempos de cómputo⁶⁹.

Por otro lado, los métodos de modelado por homología (o modelado comparativo) son computacionalmente mucho menos costosos. El aspecto fundamental de estos algoritmos es que incorporan información conocida sobre proteínas que ya tienen resuelta su estructura tridimensional. El modelado comparativo utiliza estructuras completas de proteínas conocidas como molde (usualmente denominadas templados) para construir a partir de éste el modelo de interés. Un esquema general se muestra en la figura 14 y en los métodos computacionales de

esta tesis se desarrollará el concepto más en detalle. Servirán como templatado aquellas proteínas que puedan ser consideradas homólogas (que posean una identidad de secuencia tal, que en base a los criterios que cada algoritmo defina, la relación de homología pueda ser establecida). A medida que los datos aumentan, también aumenta el cubrimiento del espacio conformacional de estructuras y por ende el poder descriptivo de esta técnica⁷⁰.

Figura 14: Esquema general del algoritmo de modelado por homología, Teniendo dos proteínas homólogas en secuencia, se usa como molde la estructura desconocida, intentando predecir los cambios estructurales generados por las diferencias de aminoácidos.



Fuente: Radusky, 2017

Cabe destacar que hay proteínas de las cuáles no se obtiene información alguna buscando en bases de datos, principalmente porque ninguna proteína que posea resuelta su estructura cumple con la condición de tener una identidad suficiente que permita que se la use como templatado.

La ventaja de estos métodos es que, a diferencia del modelado por homología, los modelos generados no se encuentran sesgados por las estructuras de proteínas actualmente resueltas, lo que nos da el potencial de generar modelos de calidad sobre moléculas para las cuales su plegado no se encuentra resuelto experimentalmente.

SWISS-MODEL <https://swissmodel.expasy.org> fue el primer servidor de modelado de homología de proteínas totalmente automatizado y se ha mejorado continuamente durante los últimos 25 años ⁷¹ Su funcionalidad de modelado se ha ampliado recientemente para incluir el modelado de complejos homo y heteroméricos. Otras características introducidas recientemente incluyen el desarrollo de un nuevo motor de modelado, ProMod3, con mayor precisión de los modelos producidos, y un método de estimación de calidad de modelo local mejorado basado en una versión novedosa de QMEAN ⁷².

SWISS-MODEL genera actualmente ~3000 modelos por día (~2 modelos por minuto), frente a ~1500 modelos por día en 2014, lo que lo convierte en uno de los servidores de modelado de estructuras más utilizados en todo el mundo ⁷³.

ERRAT es un servidor en línea para analizar las estadísticas de interacciones no unidas entre diferentes tipos de átomos y traza el valor de la función de error frente a la posición de una ventana deslizante de 9 residuos, calculada en comparación con las estadísticas de estructuras altamente refinadas (<https://servicesn.mbi.ucla.edu/ERRAT/>). Es un método, de fácil disponibilidad, que utiliza información estructural para predecir la calidad de cada residuo. Errat analiza las estadísticas de interacciones no enlazadas entre átomos de nitrógeno (N), carbono (C) y oxígeno (O) ⁷⁶.

MOLPROBITY es un sistema ampliamente utilizado de validación de modelos para estructuras de proteínas y ácidos nucleicos, al que se accede en <http://molprobity.biochem.duke.edu>. Se basa en el trabajo de sistemas anteriores como ProCheck, WhatIf, y Oops, que introdujeron el uso de la validación mediante los criterios de rotamer de cadena lateral y diagrama de Ramachandran. Complementa los sistemas para validar datos y la coincidencia de modelo a datos como Rfree o residual de espacio real. MolProbity tiene algunas características específicamente diseñadas para cristalografía de rayos X, y también es adecuada y se utiliza para crioEM, neutrones, RMN y modelos computacionales ⁷⁷.

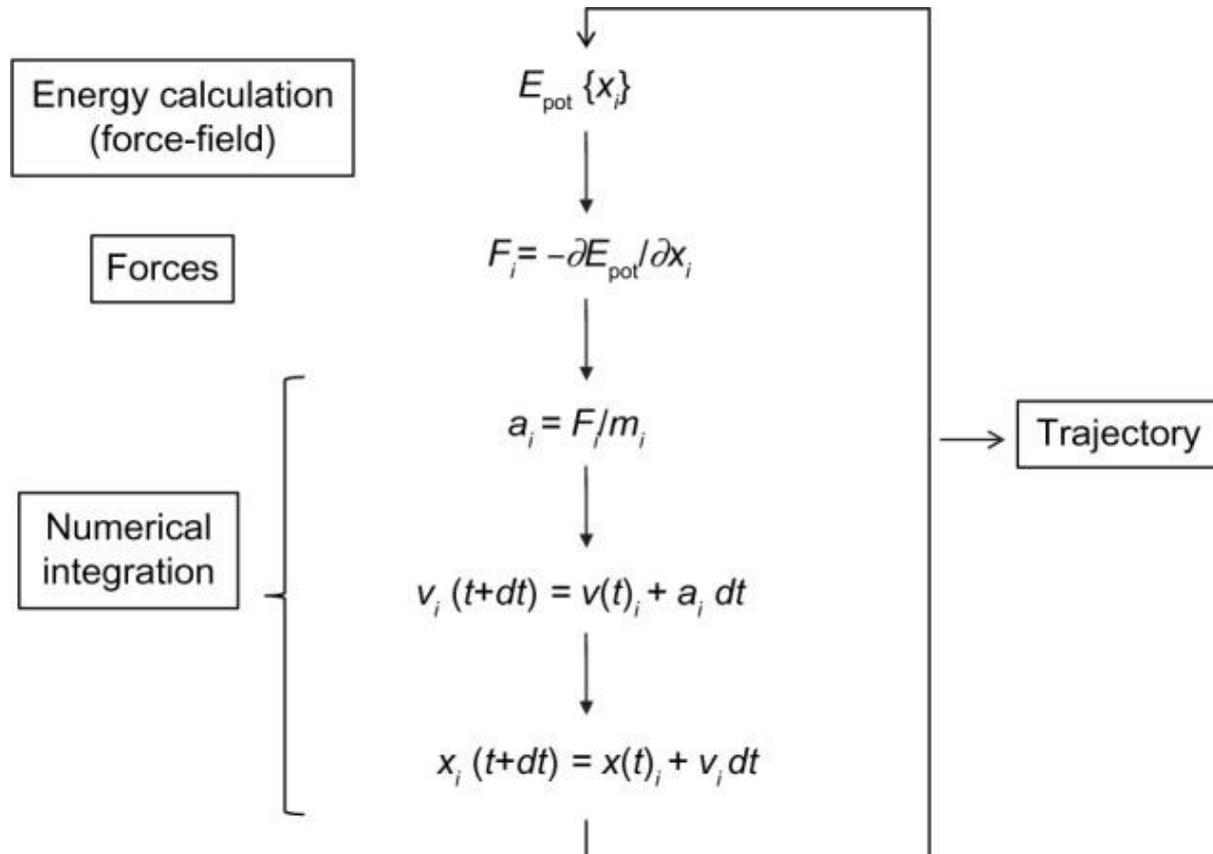
1.6.3 Simulación de dinámica molecular

La simulación de dinámica molecular (MD), desarrollada por primera vez a finales de los años 70, ha evolucionado a partir la simulación de varios cientos de átomos a sistemas con relevancia biológica, proteínas incrustadas en la membrana o grandes complejos macromoleculares como nucleosomas o ribosomas. La simulación de sistemas que tienen ~ 50 000 a 100 000 átomos es actualmente una rutina, y las simulaciones de alrededor de 500 000 átomos son comunes cuando se dispone de las infraestructuras informáticas adecuadas. Este importante adelanto es en gran detalle un resultado del uso de informática de alto rendimiento (HPC) y la sencillez del algoritmo MD básico (Figura 3). Se obtiene un modelo inicial del sistema a partir de estructuras experimentales o de datos de modelos comparativos. El sistema simulado se podría caracterizar con diferentes niveles de detalle. La representación atomística es la que conduce a la mejor reproducción de los sistemas actuales. Las representaciones se están volviendo muy populares cuando se requieren grandes sistemas o simulaciones largas ⁷⁸.

Una vez construido el sistema, las fuerzas que actúan sobre cada átomo se obtienen derivando ecuaciones, donde la energía potencial se deduce de la estructura molecular. Los campos de fuerza son ecuaciones complejas, pero son fáciles de calcular. La simplicidad de la representación del campo de fuerza de las características moleculares: resortes para la longitud y los ángulos del enlace, funciones periódicas para las rotaciones del enlace y los potenciales de Lennard-Jones, y la ley de Coulomb para las interacciones de van der Waals y electrostáticas, respectivamente, asegura que los cálculos de energía y fuerza son extremadamente rápidos incluso para sistemas grandes⁷⁹.

Los campos de fuerza que se utilizan actualmente en las simulaciones moleculares atómicas difieren en la forma en que se parametrizan. Una vez que se obtienen las fuerzas que actúan sobre los átomos individuales, la ley clásica de movimiento de Newton se utiliza para calcular aceleraciones y velocidades y para actualizar las posiciones de los átomos. Como la integración del movimiento se realiza numéricamente, para evitar la inestabilidad, se debe usar un intervalo de tiempo más corto que los movimientos más rápidos en la molécula. Esto se ubica normalmente entre 1 y 2 fs para simulaciones atómicas. Las simulaciones de microsegundos, que apenas alcanzan las escalas de tiempo de los procesos biológicos, requieren iterar este ciclo de cálculo 10⁹ veces. Éste es uno de los puntos fuertes de las estrategias generales. A medida que se utiliza una representación más simplificada del sistema, son posibles pasos de tiempo mucho mayores y, por lo tanto, la duración efectiva de las simulaciones se amplía drásticamente. Por supuesto, esto se puede obtener a expensas de la precisión del conjunto de simulación. Los avances algorítmicos, que incluyen el ajuste fino de los cálculos de energía, la paralelización o el uso de unidades de procesamiento gráfico (GPU), han mejorado en gran medida el rendimiento de las simulaciones MD⁸⁰.

Figura 15 Algoritmo básico de dinámica molecular. Notas: La salida de la simulación, la trayectoria, es una lista ordenada de coordenadas de átomos de $3N$ para cada tiempo de simulación. Abreviaturas: E_{pot} , energía potencial; t , tiempo de simulación; dt , tiempo de iteración; Para cada coordenada espacial de los N átomos simulados (i): x , coordenada del átomo; F , componente de fuerzas; a , aceleración; m , masa atómica; v , velocidad.



Fuente: Gelpí et al, 2015.

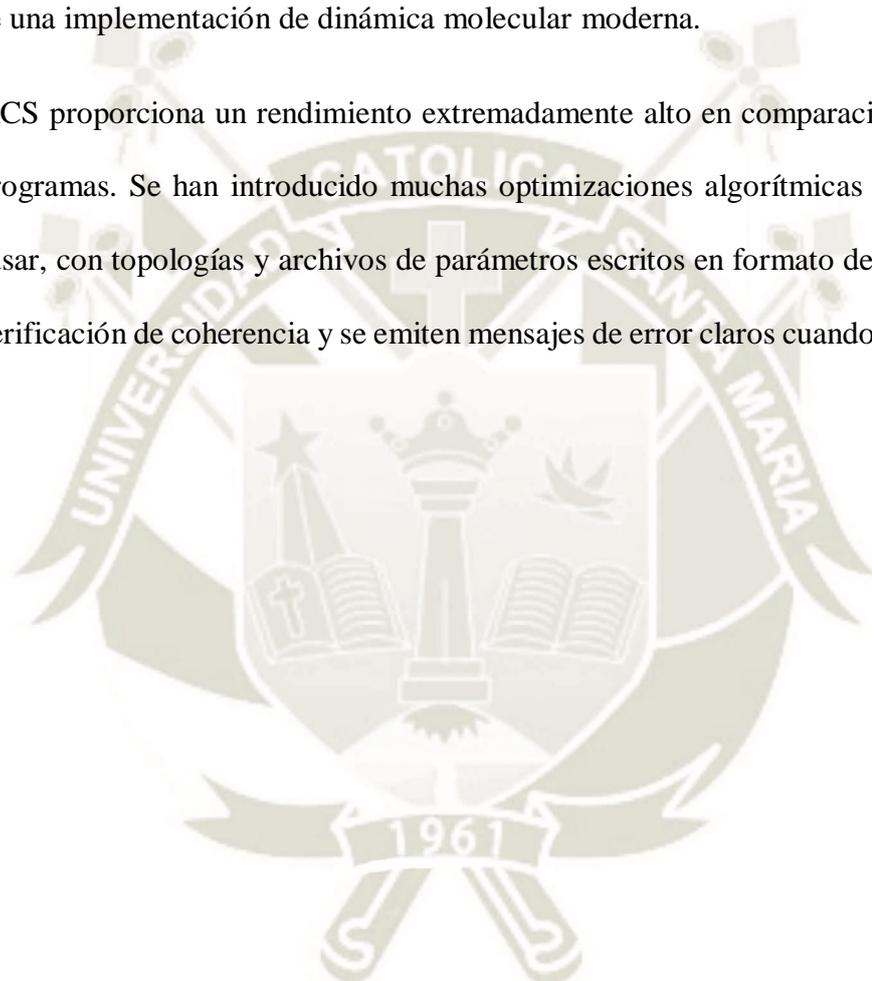
1.6.3.1 GROMACS

Es un nombre propio. No es un acrónimo. Es un programa de alto rendimiento, de código abierto, multiplataforma utilizado para realizar la simulación de la dinámica molecular de sistemas con cientos de millones de partículas. Gromacs fue originalmente desarrollado en la Universidad de Groningen, pero por ser de código abierto ha sido reimplementado por muchos otros desarrolladores.

Es un paquete versátil para realizar dinámica molecular, es decir, simular las ecuaciones de movimiento newtonianas para sistemas con cientos o millones de partículas. Está diseñado

principalmente para moléculas bioquímicas como proteínas, lípidos y ácidos nucleicos que tienen muchas interacciones enlazadas complicadas, pero dado que GROMACS es extremadamente rápido para calcular las interacciones no enlazadas (que generalmente dominan las simulaciones), muchos grupos también lo utilizan para la investigación de sistemas biológicos, p. ej. polímeros. Es compatible con todos los algoritmos habituales que se espera de una implementación de dinámica molecular moderna.

GROMACS proporciona un rendimiento extremadamente alto en comparación con todos los demás programas. Se han introducido muchas optimizaciones algorítmicas en el código. Es fácil de usar, con topologías y archivos de parámetros escritos en formato de texto claro. Hay mucha verificación de coherencia y se emiten mensajes de error claros cuando algo anda mal⁸¹.



CAPITULO II

2 MATERIALES Y METODOS

2.1 CAMPO DE APLICACIÓN

2.1.1 Área de conocimiento

- **Área:** Ciencias de la salud
- **Campo:** Biotecnología
- **Línea:** Bioinformática

2.1.2 Campo de verificación

2.1.2.1 UBICACIÓN ESPACIAL

No se precisó una infraestructura específica dado que el estudio es de tipo descriptivo y como técnica se aplicó la observación documental.

2.1.2.2 UBICACIÓN TEMPORAL

El estudio se llevó a cabo en el año 2020 durante los meses de agosto a diciembre.

2.1.2.3 UNIDADES DE ESTUDIO

Las unidades de estudio estuvieron conformadas por los datos informáticos de las secuencias de nucleótidos obtenidas del NCBI que presentan mutaciones en el HBB en la región Sudamérica.

2.2 MATERIALES

Materiales

1. EQUIPOS:

Computadora t7910. Dell server, Dual processors, Gtx 1080ti GPU

2. PROGRAMA INFORMÁTICO:

GROMACS 2020.4

2.3 METODOLOGÍA

2.3.1 Recolección de secuencia:

Las secuencias del gen HBB no caracterizados que tienen mutaciones de sustitución, delección y de inserción se recogen de HbVar. HbVar es una base de datos de hemoglobina humana y Talasemias (<http://globin.bx.psu.edu/hbvar/>). En HbVar hay un total de 1.823 variantes de HBB. Se recogieron variantes de HBB de la región de América del Sur a partir de 1823 secuencias. Sólo hubo 4 de las 1823 variantes, que eran de la región de América del Sur, con título de Venezuelan nd-HPFH, Hb Ecuador, -195 (C-> G) Agamma; the Brazilian nd-HPFH⁸² y Hb Chile (Hojas-Bernal et al., 1999). Venezuelan nd-HPFH, Hb Ecuador no estaban disponibles las secuencias de ADN. De -195 (C-> G) Agamma; the Brazilian nd-HPFH (Costa et al., 1990) y Hb Chile⁸³ fue recolectado de NCBI (<https://www.ncbi.nlm.nih.gov/>). La secuencia de wild-type del gen HBB (gen ID 3043) también se recolectó de NCBI (<https://www.ncbi.nlm.nih.gov/gene/?term=3043>).

2.3.2 Alineación de secuencias múltiples (MSA):

Para alinear las 3 secuencias, -195 (C-> G) Agamma; Brazilian nd-HPFH, HBB de Wild-type y Hb Chile se realizó a partir de Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Clustal Omega es una herramienta en línea para la alineación de secuencias múltiples desde el servidor web EMBL-EBI (<https://www.ebi.ac.uk/>). La alineación de múltiples secuencias se realizó utilizando parámetros de formato de salida: ClustalW con recuento de caracteres.

2.3.3 Traducción:

Para la traducción a proteínas de las secuencias alineadas -195 (C-> G) Agamma; the Brazilian nd-HPFH, HBB de Wild-type y Hb Chile, fue usado ExPASy translate, es un servidor de traducción en línea. Se introduce secuencias de ADN y proporciona una secuencia de proteínas como salida. Se utilizó parámetros de formato de salida: Compacto: M, -,sin espacios; Hebras de DNA: forward. (<https://web.expasy.org/translate/>).

2.3.4 Modelado de homología:

El modelo de homología de HBB de wild-type y de los tipos mutantes (Hb Chile y -195 (C-> G) Agamma; the Brazilian nd-HPFH) se realizó para comparar la estructura 3D de las proteínas. El modelado de homología se realizó utilizando SWISS-MODELER ⁸⁸.

SWISS-MODEL es un servidor de modelado de homología de estructuras de proteínas totalmente automatizado al que se puede acceder a través del servidor web Expasy o desde el programa DeepView. El propósito de este servidor es hacer que el modelado de proteínas sea accesible para todos los investigadores de ciencias de la vida en todo el mundo. (<https://swissmodel.expasy.org/>). Se buscó la plantilla para cada secuencia antes de construir el modelo. La mejor plantilla se eligió sobre la base de la similitud de secuencia.

2.3.5 Evaluación del modelo:

Los 3 modelos construidos con SWISS-MODELER fueron evaluados más a detalle por ERRAT para verificar los factores de calidad de las estructuras ⁸⁶.

ERRAT es un servidor en línea para analizar las estadísticas de interacciones no enlazadas entre diferentes tipos de átomos y traza el valor de la función de error frente a la posición de una ventana deslizante de 9 residuos, calculada mediante una comparación con estadísticas de

estructuras altamente refinadas (<https://servicesn.mbi.ucla.edu/ERRAT/>). Además, estas estructuras se evaluaron a partir de molprobity (<http://molprobity.biochem.duke.edu/>)⁹⁰. Molprobity es un servidor web de uso general que ofrece validación de calidad para estructuras 3D de proteínas, ácidos nucleicos y complejos ⁸⁷.

2.3.6 Simulación de dinámica molecular

GROMACS Se utilizó el paquete de simulación (GROMACS 2020.4) para realizar simulaciones de dinámica molecular. Se realizaron simulaciones de wild-type y proteínas mutantes (HPFH y Hb Chile) durante 30 ns en agua utilizando el campo de fuerza GROMOS96 53A6; Los archivos de trayectoria y energía se escribieron cada 2 ps.

Estos sistemas se solvataron en una caja octaédrica truncada periódicamente, que contenía 8805, 4611, 3502 SPC moléculas de agua para la de wild-type, HPFH y Hb Chile, respectivamente. La proteína se centró en la caja de simulación a una distancia mínima del borde de la caja de 1 nm para satisfacer de manera eficiente la convención de imagen mínima. Los estados de protonación de los grupos protonables se establecieron en pH neutro. Se agregaron iones de cloruro (2 en wild-type y 2 en hb chile) para neutralizar el sistema general. Sin embargo, no se agregaron iones en el sistema de proteína HPFH ya que ya estaba neutralizado.

La minimización de los tres sistemas se llevó a cabo durante 5000 pasos utilizando el método de descenso más pronunciado y la convergencia se logró dentro de la fuerza máxima <1000 ($\text{KJ mol}^{-1} \text{ nm}^{-1}$), para eliminar cualquier choque estérico seguido de equilibrio del sistema en el conjunto NPT durante 1000 ps a 300 K.

Las corridas de producción para todas las simulaciones se llevaron a cabo a una temperatura constante de 300 K y una presión de 1 atm (NPT) utilizando un escalado de velocidad de acoplamiento débil (termostato Berendsen modificado) y algoritmos Parrinello-Rahman, respectivamente. Los tiempos de relajación se fijaron en $\tau_T = 0,1$ ps y $\tau_P = 1,0$ ps y se utilizó una compresibilidad isotérmica estimada del agua de $4,5e - 5$ ($\text{kJ mol}^{-1} \text{ nm}^{-3}$)⁻¹. Todas las longitudes de enlace que involucran átomos de hidrógeno se mantuvieron rígidas en longitudes de enlace ideales utilizando el algoritmo Linear Constraint Solver (lincs), lo que permite un intervalo de tiempo de 2 fs. Se utilizó el esquema Verlet para el cálculo de interacciones no vinculadas. Se utilizaron condiciones de contorno periódicas (PBC) en todas las direcciones x, y, z. Se calcularon las interacciones dentro de un límite de corto alcance de 1,0 nm en cada paso de tiempo.

CAPITULO III

3 RESULTADOS

3.1 Variantes mutacionales:

4 variantes de la región HBB de América del Sur tienen diferentes mutaciones en diferentes ubicaciones (Tabla 1). La información de diferentes mutaciones de HBB en la región de América del Sur es la siguiente:

Tabla 1 Variantes HBB

Nombre	ID de HbVar	Nombre de HGVS	Categoría	Ubicación de la mutación	Procedencia
Venezuelan Nd-HPFH	2513	HBB1: c.-264C> T	Talasemias	No dentro de la unidad de transcripción conocida hg38: chr11 5.250.068	Origen étnico Negro
Hb Ecuador	2586	HBA1: c.416C> G	Variantes de Hb	Exón hg38: chr16 177,398	Encontrado en una mujer de Ecuador

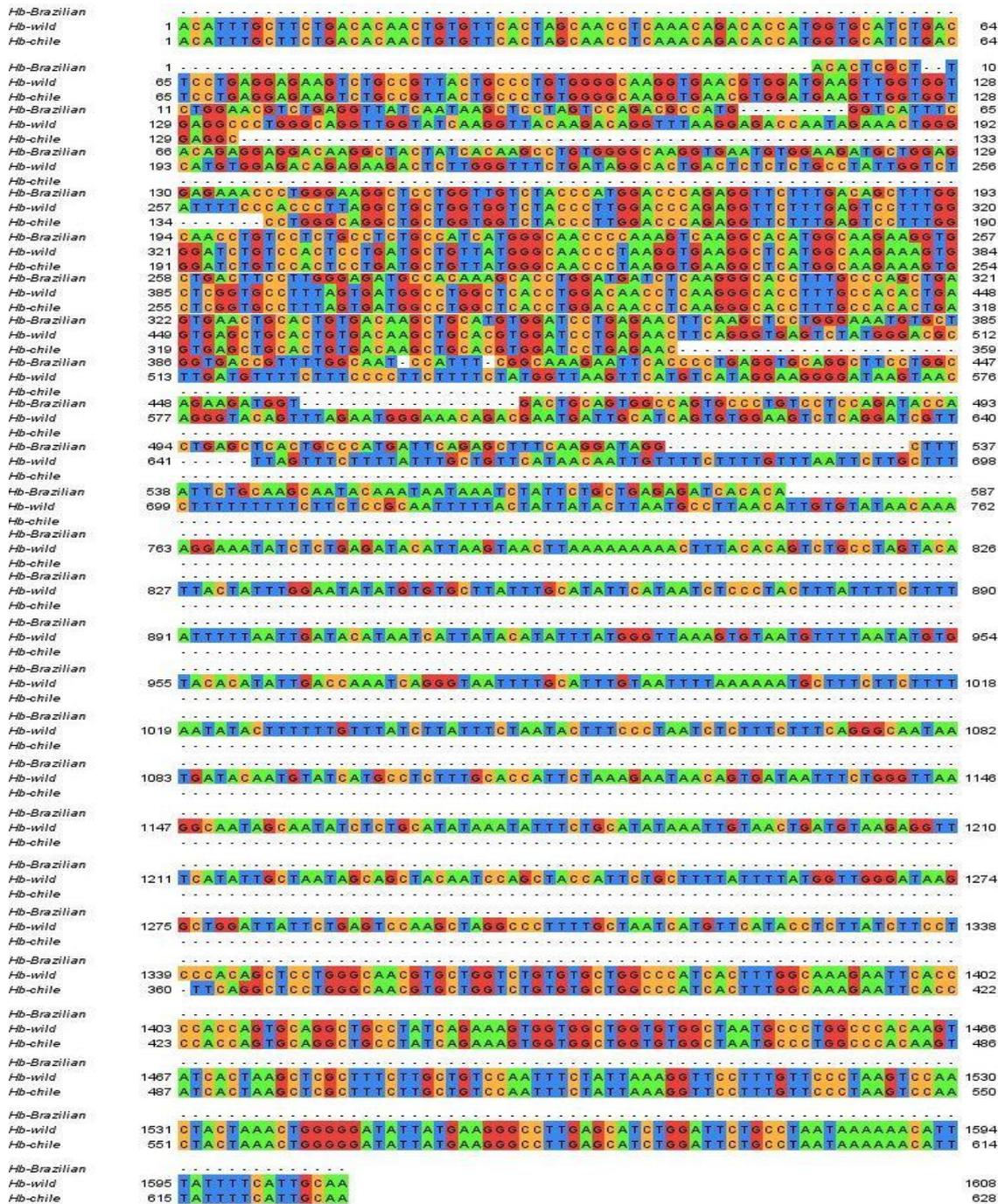
<p>-195 (C-> G) Agamma; The Brazilian nd- HPFH</p>	<p>1011</p>	<p>HBG1: c.-248C> G</p>	<p>Talasemias</p>	<p>No dentro de la unidad de transcripción conocida hg38: chr11 5,250,052</p>	<p>Origen étnico brasileño</p>
<p>Hb Chile</p>	<p>287</p>	<p>HBB: c.85C> A</p>	<p>Variantes de Hb</p>	<p>exón hg38: chr11 5,226,937</p>	<p>Origen étnico indio americano</p>

Hb Ecuador y Hb Chile tienen mutaciones en la parte del exón que alteran la secuencia de proteínas. La variante de Hb Ecuador cambia la serina en cisteína mientras que la Hb Chile variante transforma leucina en metionina.

3.2 Exploración del MSA entre HB wild y proteínas mutadas:

La alineación de secuencias múltiples se realizó utilizando Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). La alineación mostró que -195 (C-> G) Agamma; the brazilian nd-HPFH tiene 587 nucleótidos, el Wild-type contiene 1608 y el Hb Chile contiene 628 nucleótidos (Figura16). Los resultados mostraron que las 3 secuencias son diferentes entre sí. Hb Wild-type y Hb Chile tienen regiones más conservadas, mientras que Hb wild-type y -195 (C-> G) Agamma; the Brazilian nd-HPFH tienen regiones menos conservadas.

Figura 16 Alineación de secuencia múltiple. Mostró que la Hb wild-type tiene regiones más conservadas con Hb Chile.



Fuente: Generado por Clustal Omega, 2020.

3.3 Traducción:

La secuencia de proteínas se obtuvo del servidor ExPASy Translate (<https://web.expasy.org/translate/>).

La secuencia de proteínas muestra que -195 (C->G) Agamma; the Brazilian nd-HPFH contiene 195 residuos de aminoácidos, la proteína de Wild-type contiene 536 residuos de aminoácidos mientras que la proteína Hb Chile contiene 209 residuos de aminoácidos. La traducción de las secuencias muestra que las 3 variantes tienen diferencias debido a la mutación (Figura 17).

Figura 17 Traducción de las secuencias. Mostró que las 3 secuencias son diferentes entre sí debido a la mutación.

-195 (C->G) Agamma; the Brazilian nd-HPFH

TLASGTSEVINKLLVQTPWVISQRRTRLLSQACGAR-MNKMLEEKPEWEGSWLSTHGPRGSLTALATCPLPLPSWATPKSRHMARRC-LPWEMPOSTWIMISRAPLPS-VNCTVTSCMWILRTSSSWEMCW-PFWQSI SAKNSPLRCRLPGRRW-LQWPVPCPPDTELTATHDSELSRIGFILQAIQIINLFC-EIT

Hb Chile

TFASDITVFTSNLQTPWCI-LLRRSLPLLPCGAR-TWMKLVVRPWAGCWWSTLGPGRSLSPLGICPLLMLLWATLR-RIMARKCSVPLVMWLTWTTSRAPLPH-VSCTVTSCWILRTSGSWATCWSVCWPITLAKNSPHQCRLPIRKWVLWLMFWPTSITKLAFLLSNFY-RFLCSSLPTTKLGDIMKGLEHLSA--KTFIFIA

Wild Type

TFASDITVFTSNLQTPWCI-LLRRSLPLLPCGAR-TWMKLVVRPWAGWYQGYKTGLRRPIETGHVETEKTLGFLIGTDSLCLLVYFPTLRLLLVVYPWTQRFESF GDLSTPDVAVMGNPKVKAHGKRVLGAFS DGLAHLNLDLNGKTFATLSELHCDKLVDPENFRVSLWDA-CFLSPSFLWLSSCHRKGISNRVQFRMGNRRMIASVWKSQD RFSFFYLLFITIVFPCLILAFPPFLLRNFYYT-CLNIVYKRYL-DTLSNLKKNFTQSA-YITWNICVLICIFIISLLYFLLFLIDT-SLYIFMG-SVMF-YV YTY-PNQCNFAPVILKNAFFP-YTFLPILPLISLISFPQGNNDTMYHASLHHSKE-Q--FLG-GNSNISAYKYFCI-IVTDVRFILLIAATIQLPFCFYFMVGI RLDYSESKLGPFANHVHTSYLPPTAPGQRAGLCAGPSLWQRIHPTSAGCLSESGWCG-CPGPQVLSLSLSCCPISIRKGSFVP-VQLLNWGL-RALSIWILPNKK HLFSLQ

Fuente: Generado por el servidor ExPASy Translate, 2020.

3.4 Modelado por homología:

El modelado por homología se realizó utilizando SWISS-MODELER ⁸⁶. La búsqueda de plantillas para cada secuencia se realizó antes de construir el modelo. El modelo se construyó contra la mejor plantilla que se eligió sobre la base de la similitud de secuencia. Para generar estructura 3D de secuencias, plantilla con ID 1dxu.1 ⁸⁷ se eligió para la secuencia de wild-type de HBB que tiene una identidad de secuencia del 73,64% (Figura 18), plantilla con ID 2lp6.1 ⁸⁶ fue elegido para -195 (C->G) Agamma; el nd-HPFH Brazilian tiene una identidad de

secuencia 28,57% (Figura19) y plantilla con ID 2w7k.1⁸⁷ se eligió para la Hb Chile que tiene una identidad de secuencia del 6,25% (Figura20).

Figura 18 Modelo de wild-type HB

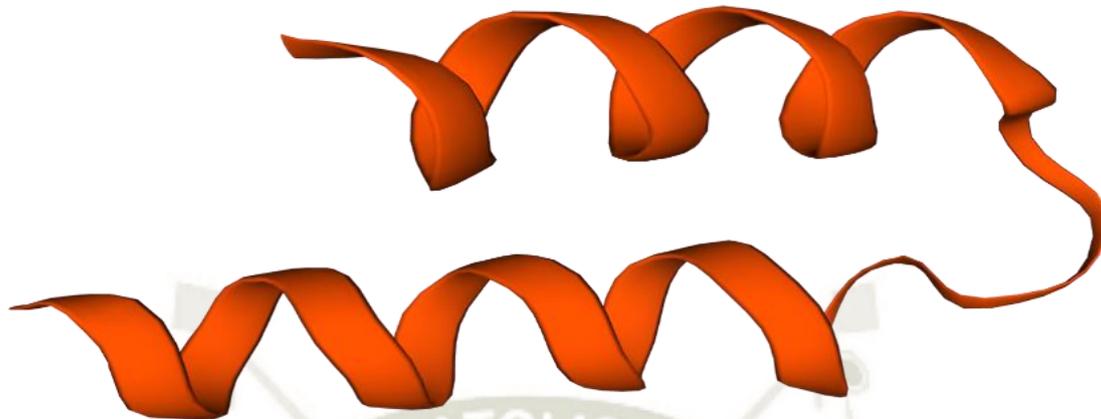


Fuente: Generado por SWISS-MODEL, 2020.

Figura 19 Modelo de -195 (C-> G) Agamma; the Brazilian nd-HPFH



Fuente: Generado por SWISS-MODEL, 2020.

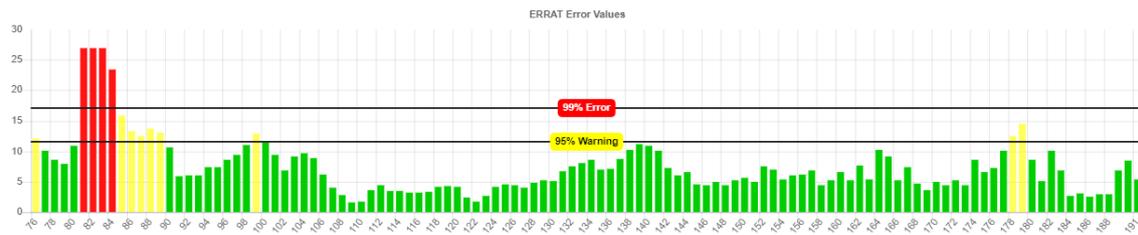
Figura 20 Modelo Hb Chile

Fuente: Generado por SWISS-MODEL,2020.

3.5 Evaluación:

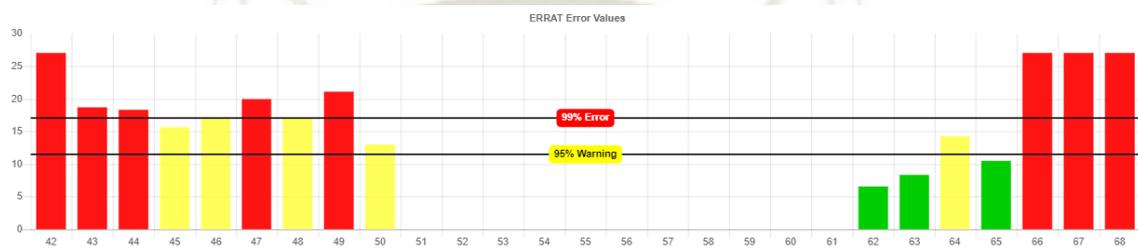
Los 3 modelos construidos con SWISS-MODELER fueron evaluados más a fondo por ERRAT para verificar los factores de calidad de las estructuras ⁸⁴. Los resultados muestran que la Hb de Wild-type (Figura21) tiene una estructura que tiene un factor de calidad 88,7931, -195 (C > G) Agamma; the Brazilian nd-HPFH (Figura22) tiene una estructura que tiene un factor de calidad de 18,75 y Hb Chile (Figura23) tiene una estructura que tiene un factor de calidad 100. Una estructura que tiene un factor de calidad superior a 95 significa una estructura de alta resolución. Entonces, Hb Chile tiene una estructura de alta resolución mientras que otros 2 no tienen una estructura de alta resolución. El gráfico muestra los factores de calidad de la estructura.

Figura 21 Gráfico de Wild type de Hb. Las líneas rojas muestran que hay algunos errores en la estructura.



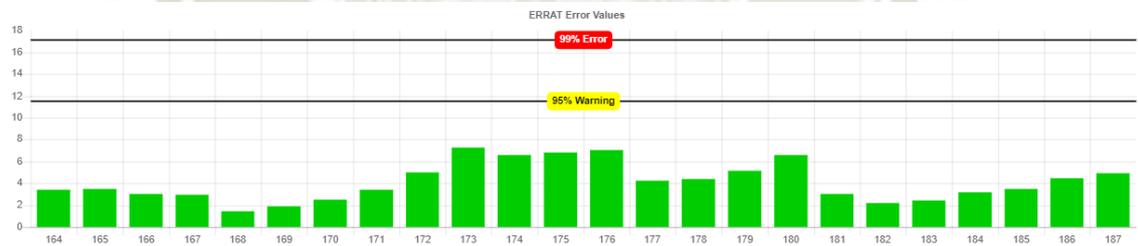
Fuente: Generado por ERRAT, 2020.

Figura 22 Gráfico de -195 (CG) Agamma; the Brazilian nd-HPFH. Las líneas rojas muestran que hay errores en la estructura.



Fuente: Generado por ERRAT, 2020.

Figura 23 El gráfico de Hb Chile. Muestra que no hay errores en la estructura.



Fuente: Generado por ERRAT, 2020.

Las estructuras también se analizaron por molprobrity. Analiza la geometría de las proteínas, como los rotámeros pobres, los rotámeros favorecidos, los valores atípicos de Ramachandran, los favorecidos de Ramachandran, Puntaje Z de distribución de Rama, desviaciones $C\beta > 0.25\text{Å}$, malos enlaces y malos ángulos.

También analiza péptidos de omegas como cis-Prolines y péptidos retorcidos, los criterios de baja resolución incluyen valores atípicos de CaBLAM y valores atípicos de CA Geometry y algunas otras validaciones adicionales como valores atípicos de volumen quiral. Para cada análisis se establece un objetivo. Según el objetivo, los resultados se dividen en 3 tipos: Bueno, Precaución y Advertencia, que se muestran en color verde, amarillo y rojo, respectivamente. Los resultados muestran que la Hb Chile tiene una buena estructura que la de tipo salvaje y -195 (C> G) Agamma; la nd-HPFH brasileña (Tabla2).

Tabla 2 Resultados de Molprobit, los colores se muestran según el objetivo, el color rojo muestra una advertencia y el color amarillo muestra precaución mientras que el color verde muestra bien. El recuento sin procesar y su porcentaje se muestra en cada columna.

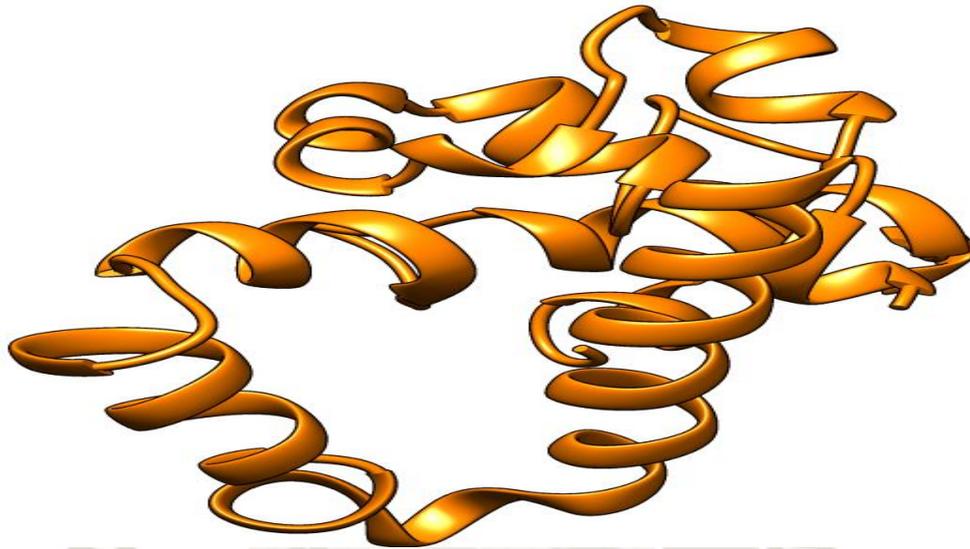
Parámetros		Hb Wild type	-195 (C-> G) Agamma; the brazilian nd- HPFH	Hb Chile	Objetivo
Proteína Geometría	Rotámeros pobres	5 (4,55%)	0 (0,00%)	0 (0,00%)	Objetivo: <0,3%
	Rotámeros preferidos	101 (91,82%)	29 (96,67%)	29 (100,0%)	Objetivo:> 98%
	Valores atípicos de Ramachandran	4 (3,28%)	4 (12,12%)	0 (0,00%)	Objetivo: <0,05%
	Ramachandran favorecido	108 (88,52%)	25 (75,76%)	28 (93,33%)	Objetivo:> 98%

	Puntaje Z de distribución de Rama	-1,11 ± 0,77	-3,72 ± 1,44	1,83 ± 1,58	Objetivo: abs (puntuación Z) <2
	Desviaciones de $C\beta > 0.25\text{Å}$	4 (3,45%)	1 (3,12%)	0 (0,00%)	Objetivo: 0
	Malos enlaces:	7/1020 (0,69%)	0/284 (0,00%)	0/261 (0,00%)	Objetivo: 0%
	Malos ángulos:	29/1383 (2,10%)	12/390 (3,08%)	2/350 (0,57%)	Objetivo: <0,1%
Omegas de péptidos	Cis Prolines:	0/6 (0,00%)	0/5 (0,00%)	0/1 (0,00%)	Esperado: ≤ 1 por cadena, o $\leq 5\%$
	Péptidos retorcidos:	2/123 (1,63%)	-	-	Objetivo: 0
Criterios de baja resolución	Valores atípicos de CABLAM	7 (5,8%)	2 (6,5%)	1 (3,6%)	Objetivo: <1.0%
	Valores atípicos de geometría de CA	4 (3,33%)	0 (0,00%)	0 (0,00%)	Objetivo: <0,5%
Validaciones adicionales	Valores atípicos de volumen quiral	0/152	0/41	0/41	

Estructuras modeladas en 3D:

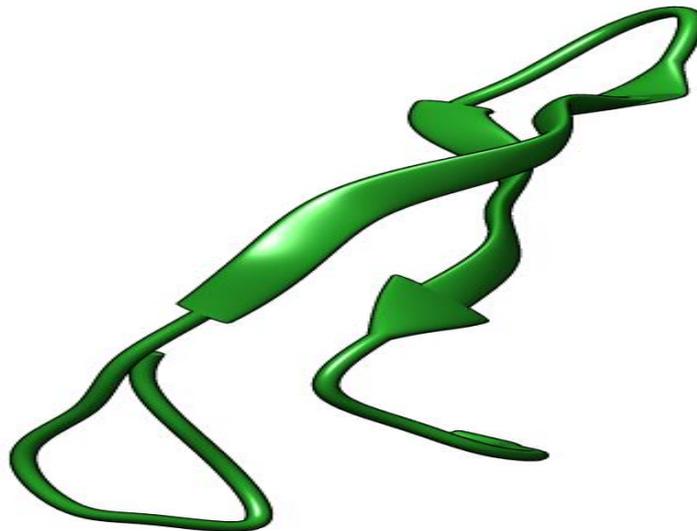
Se crearon estructuras a partir de Chimera para comparar los cambios estructurales debidos a mutaciones.(Figura 24-27)

Figura 24 Estructura 3D de Hb Wild type.



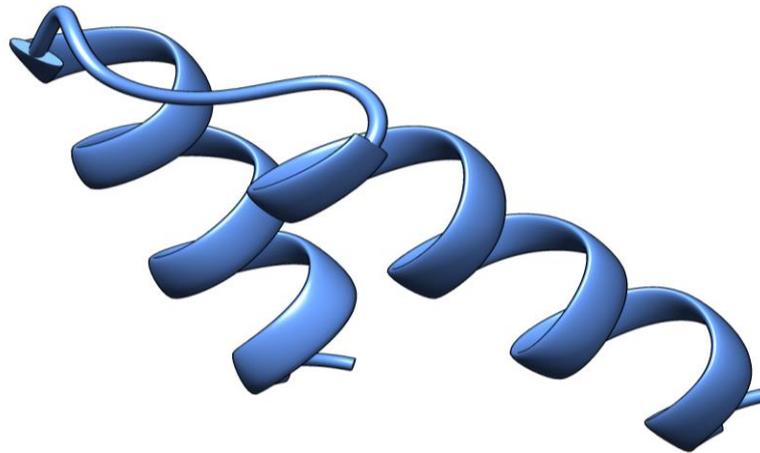
Fuente: Obtenida de Chimera, 2020.

Figura 25 Estructura 3D de -195 (C-> G) Agamma; Brazilian nd-HPFH.



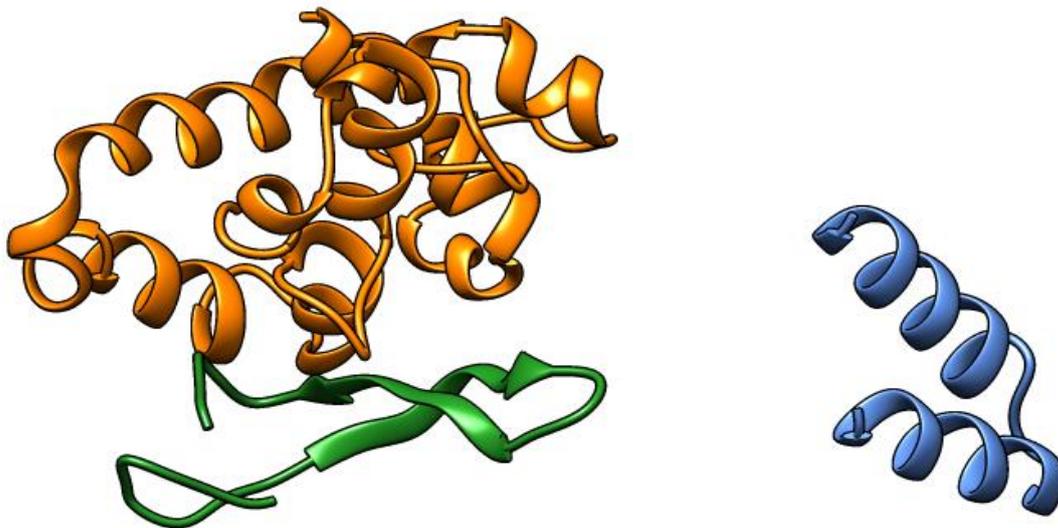
Fuente: Obtenida de Chimera, 2020.

Figura 26 Estructura 3D de Hb Chile.



Fuente: Obtenida de Chimera, 2020.

Figura 27 Estructura 3D de comparación de las 3 estructuras. El naranja muestra el Wild-type, el verde muestra -195 (C->G) Agamma; Brazilian nd-HPFH y azul muestra Hb Chile.



Fuente: Obtenida de Chimera, 2020.

3.6 Cambios conformacionales:

Hay un cambio significativo en la conformación o estructura de wild-type, -195 (C-> G) Agamma; the Brazilian nd-HPFH y Hb Chile (Tabla 3). Los cambios se analizaron a partir del análisis PSIPRED.

PSIPRED es un servidor en línea para la predicción de la estructura secundaria de proteínas (<http://bioinf.cs.ucl.ac.uk/>). Los cambios conformacionales son los siguientes.

Tabla 3 Cambios conformacionales en las 3 secuencias por análisis PSIPRED, Generado por PSIPRED.

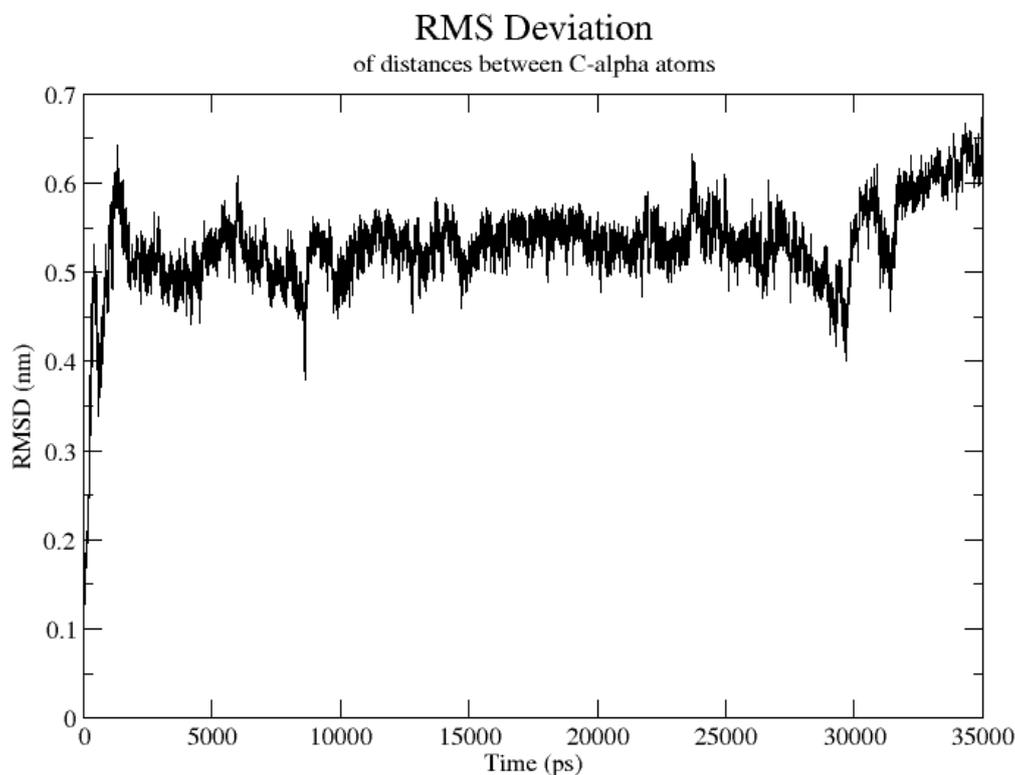
Estructuras secundarias	Wild-type	-195 (C-> G)	
		Agamma; Brazilian nd-HPFH	The Hb Chile
Hebras	14	7	8
Hélice	21	6	6
Coil	31	13	14

3.7 Simulaciones de dinámica molecular de -195(C->G) Agamma, the Brazilian nd-HPFH.

3.7.1 RMSD:

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps). El RMSD medio es de 0,54 nm con una desviación estándar de 0,04 nm. El gráfico RMSD indica que en el primer 2000 PS el sistema muestra una desviación hasta ~ 0.5 nm que permanece estable hasta 27 ns. Luego nuevamente se observa una fuerte fluctuación de 30 a 35 ns (Figura 28)

Figura 28 Gráfico de la desviación cuadrática media de la raíz (RMSD) del carbono alfa de la proteína.

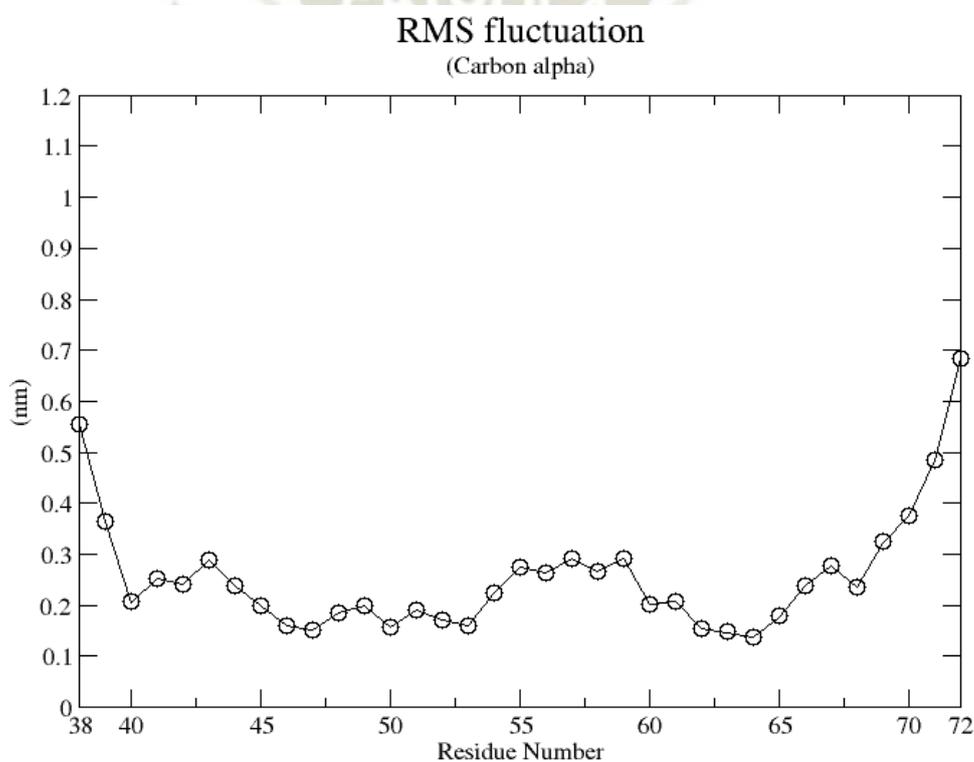


Fuente: Obtenida de GROMACS, 2020.

3.7.2 RMSF

Se realizó un total de simulación de 35 ns y para el análisis RMSF, se descartó la primera trayectoria de 5 ns (5000 ps). El gráfico RMSF muestra que la fluctuación de los terminales N y C alcanza hasta 0,6 nm y 0,7 nm, respectivamente. Sin embargo, el resto de residuos están por debajo de 0,3 nm (Figura 29).

Figura 29 Gráfico de la fluctuación cuadrática media (RMSF) del carbono alfa de la proteína.



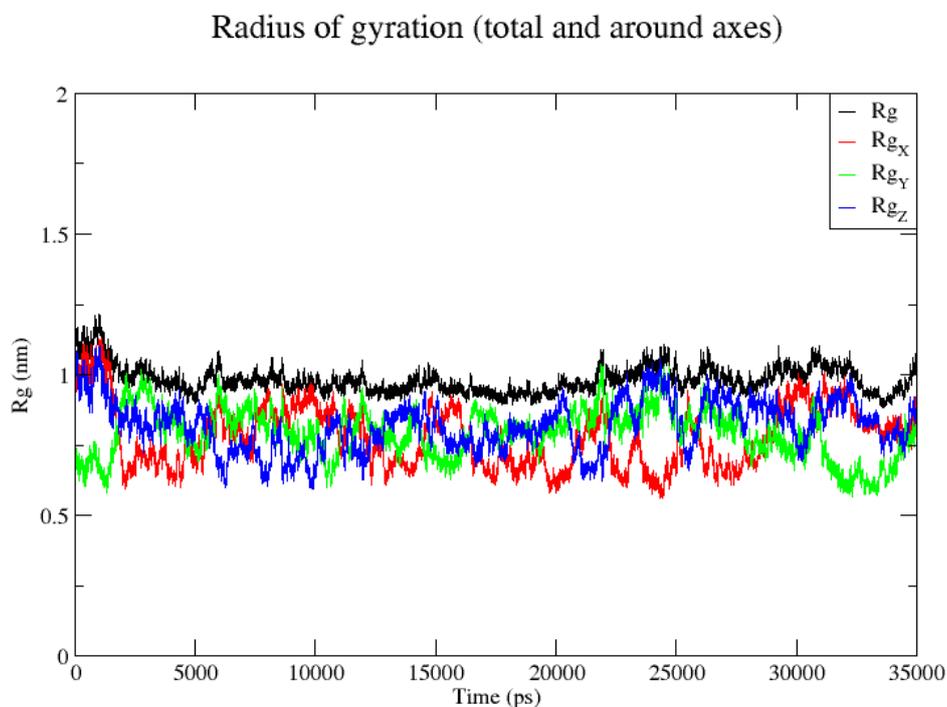
Fuete: Obtenida de GROMACS, 2020.

3.7.3 ROG

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps).

Se han observado ligeras fluctuaciones hasta 30 ns. Sin embargo, se observaron fluctuaciones pronunciadas entre 30 y 35 ns. El radio medio de giro es de 0,98 nm con una desviación estándar de 0,04 nm (Figura 30).

Figura30 Gráfico del radio de giro (ROG) del carbono alfa de la proteína.

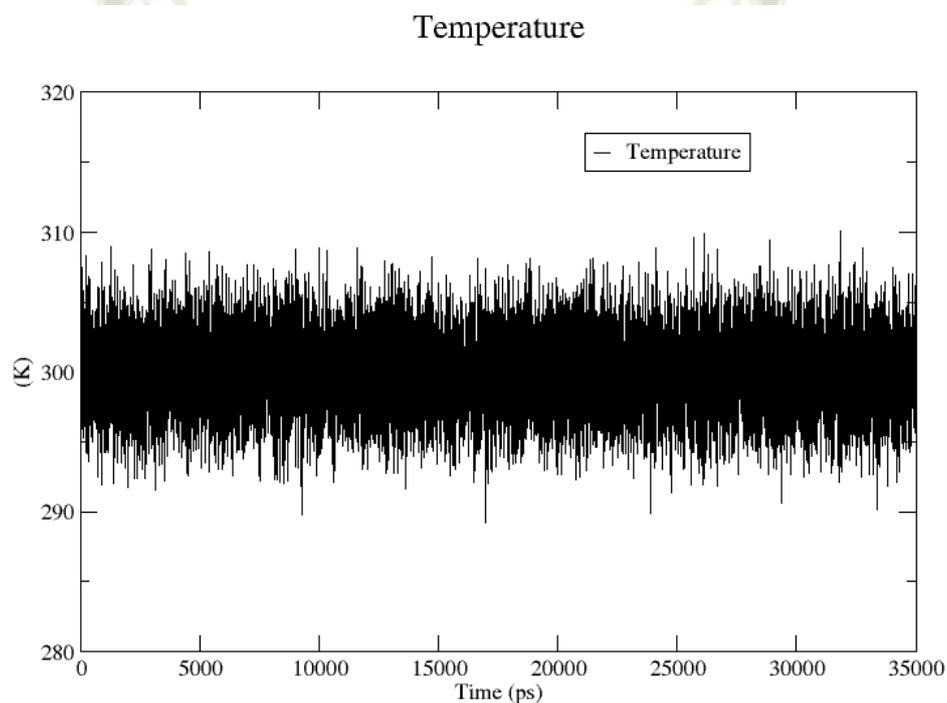


Fuente: obtenida de GROMACS, 2020

3.7.4 Temperatura

El gráfico de temperatura representa que, a lo largo de las simulaciones, la temperatura se mantuvo constante a 300 K. También muestra la estabilidad del sistema con ligeras fluctuaciones (Figura 31).

Figura 31 Temperatura

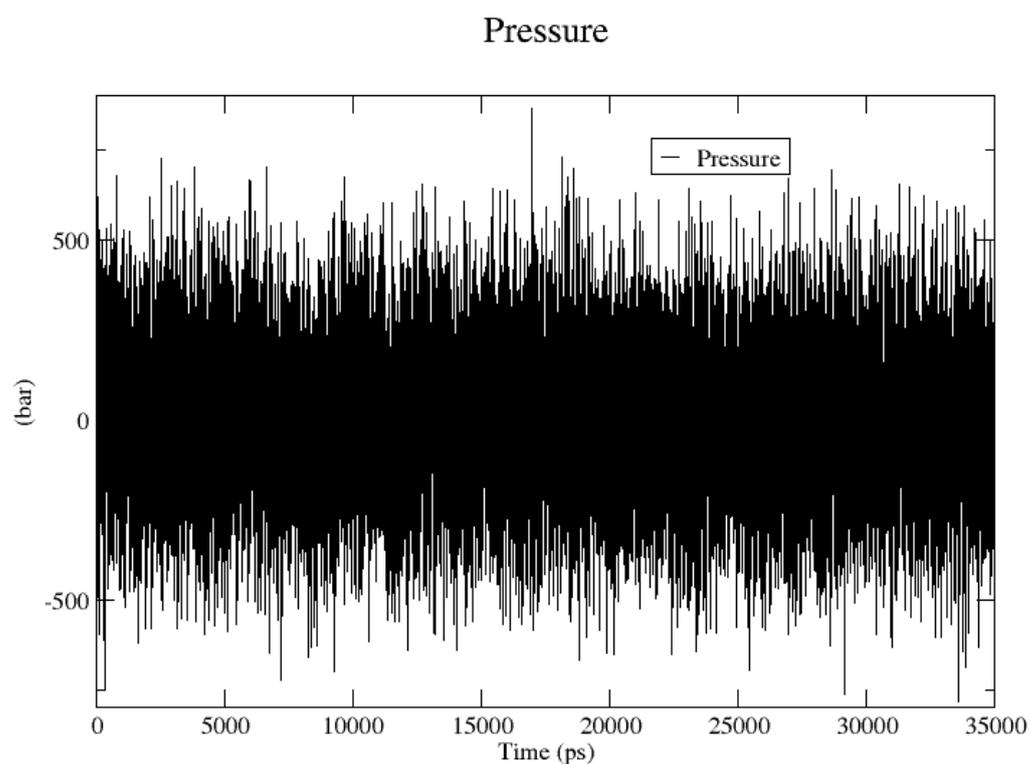


Fuente: Obtenida de GROMACS, 2020.

3.7.5 Presión

La gráfica de presión ilustra que la presión permaneció constante a 0 bar durante toda la simulación que representa la estabilidad del sistema (Figura 32).

Figura 32. Presión.

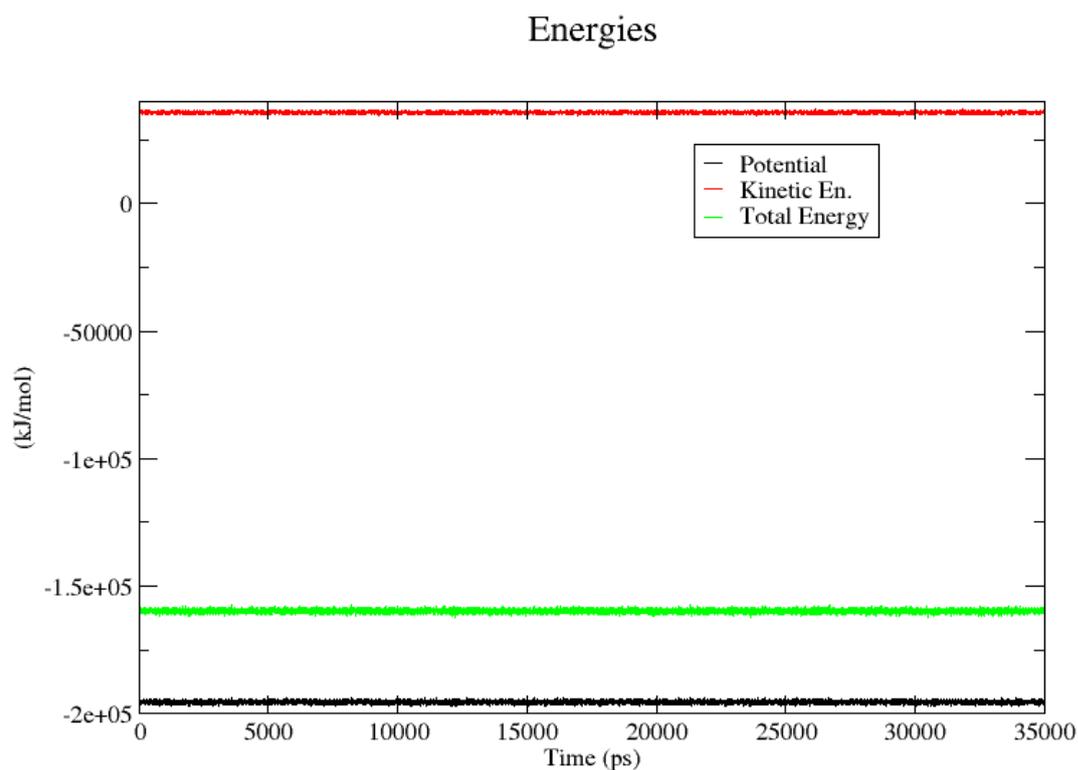


Fuente: Obtenida de GROMACS, 2020.

3.7.6 Energía

La energía total del sistema alcanza aproximadamente $-1,5e + 05$ KJ / mol, lo que ilustra la estabilidad del sistema (Figura 33).

Figura 33. Gráfico que representa la energía cinética, potencial y total del sistema.

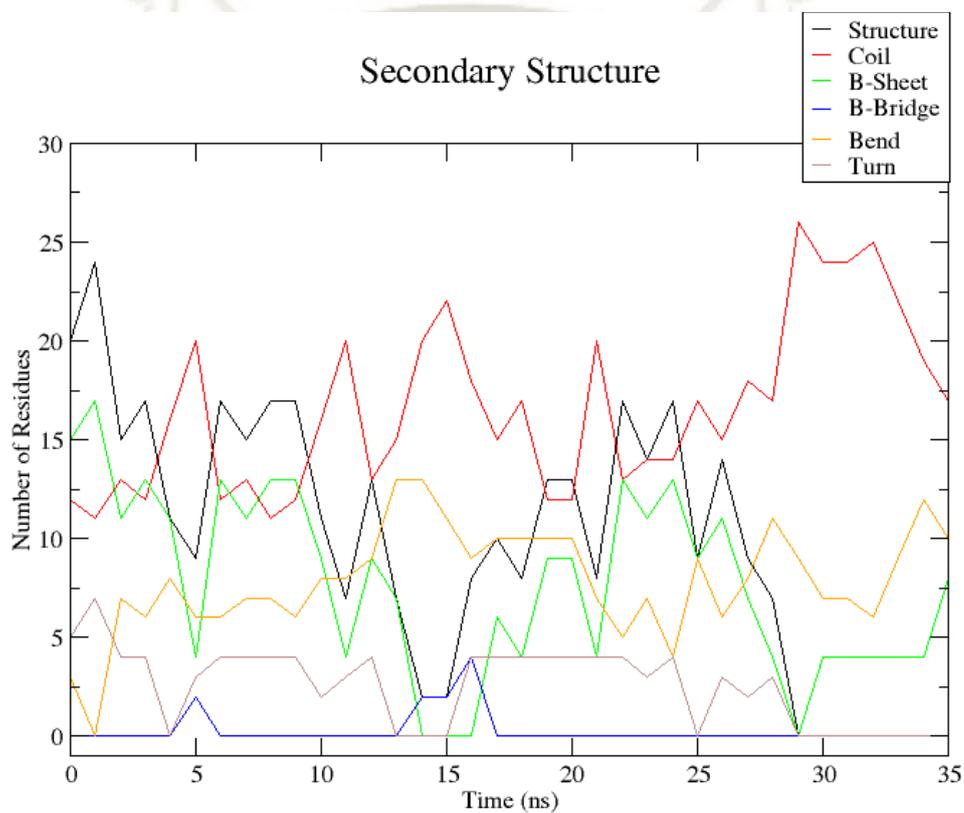


Fuente: Obtenida de GROMACS, 2020

3.7.7 Análisis de estructura secundaria (DSSP)

El gráfico DSSP muestra las fluctuaciones entre la cadena principal de proteínas a lo largo de la escala de tiempo de simulación (Figura 34).

Figura 34 Fluctuación entre la estructura secundaria de la proteína (cadena principal).

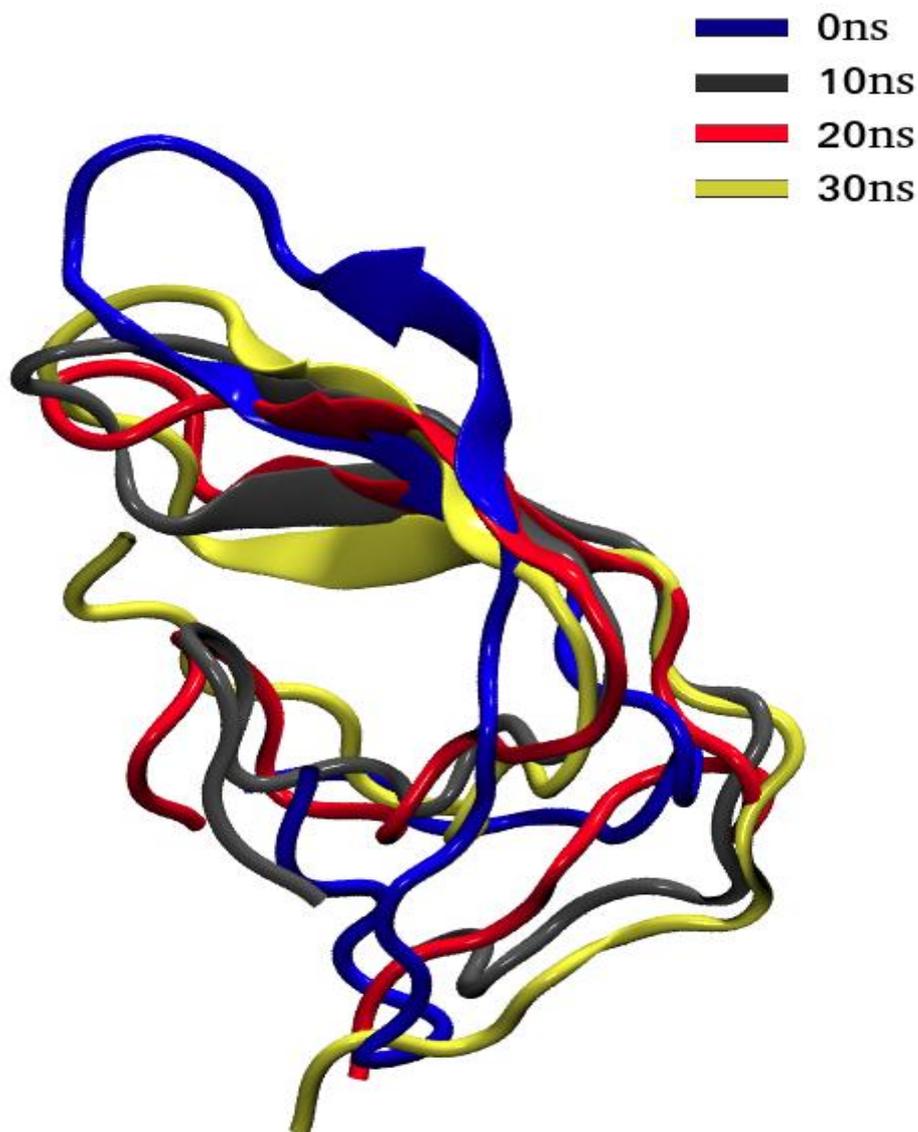


Fuente: obtenida de GROMACS, 2020

3.7.8 Superposición de proteínas

La proteína se superpuso en diferentes escalas de tiempo, es decir, 0 ns, 10ns, 20ns, 30ns para analizar los cambios en la estructura y posición de la proteína (Figura 35).

Figura 35 Superposición de proteínas en diferentes escalas de tiempo en nanosegundos



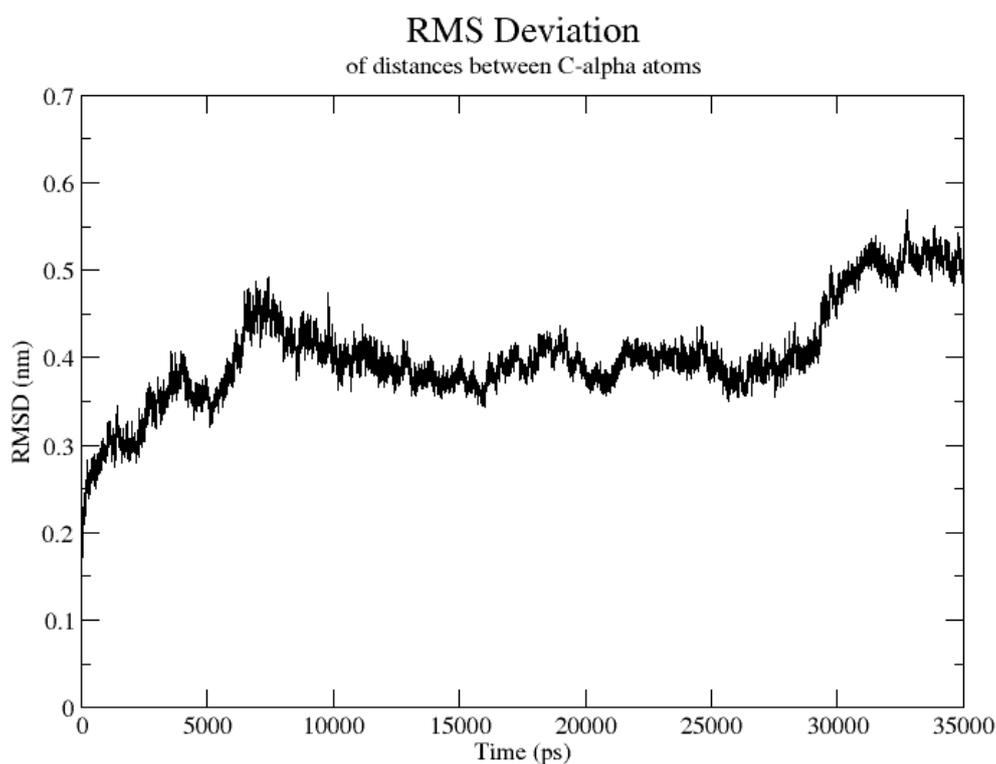
Fuente: Obtenida de Chimera,2020.

3.8 Simulaciones de dinámica molecular de la proteína Hb de Wild-type

3.8.1 RMSD:

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps). RMSD de aprox. 6 ns alcanza 0,4 nm, y luego alcanza una meseta con algunas fluctuaciones notables hasta 30 ns. Se observó un cambio estructural visible entre 30 y 35 ns donde la RMSD alcanza hasta 0,5 nm. Es probable que esta fuerte desviación esté asociada con el terminal C de la proteína (como lo indica su valor de RMSF más alto). La RMSD media es $0,42 \pm 0,05$ que representa la estabilidad general de la estructura (Figura 36).

Figura 36 Gráfico de la desviación cuadrática media (RMSD) de los carbonos alfa de la proteína.

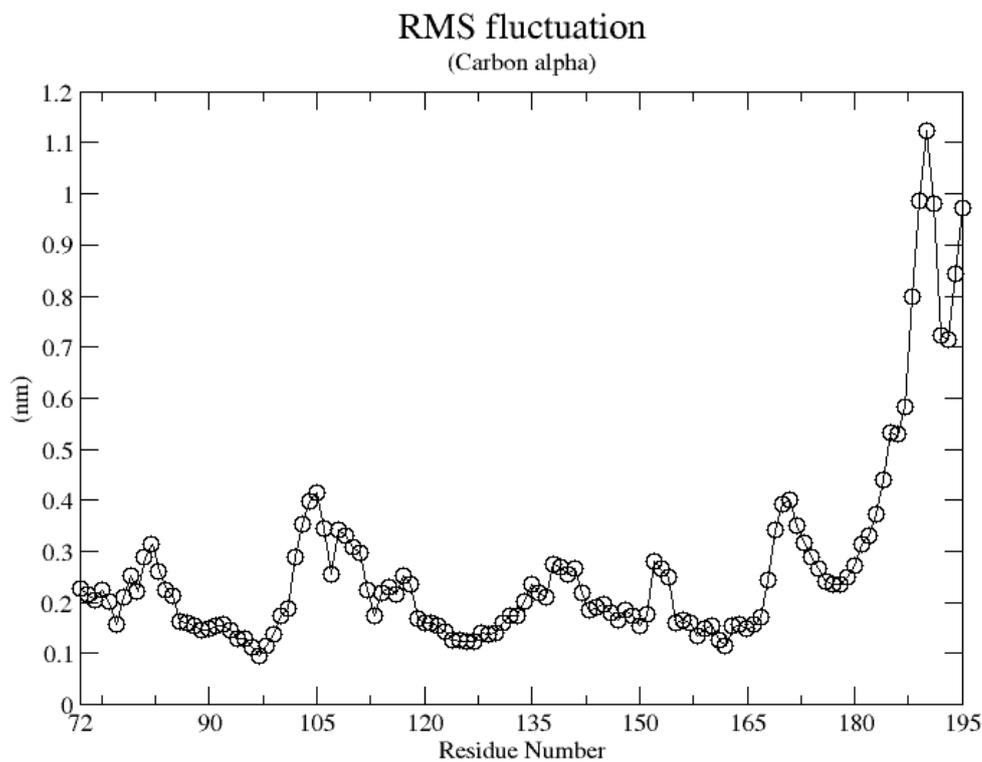


Fuente: Obtenida de GROMACS, 2020.

3.8.2 RMSF

Se realizó un total de simulación de 35 ns y para el análisis RMSF, se descartó la primera trayectoria de 5 ns (5000 ps). El gráfico RMSF muestra que el terminal N es bastante estable. Sin embargo, los residuos que van desde 79-83, 101-118 y 167-173 muestran una mayor fluctuación con un valor de RMSF superior a 0,3 nm. El terminal C es muy fluctuante y alcanza hasta 1,1 nm, lo que indica un cambio notable en la estructura (Figura 37).

Figura 37 Gráfico de fluctuación cuadrática media (RMSF) que muestra los carbonos alfa de la proteína

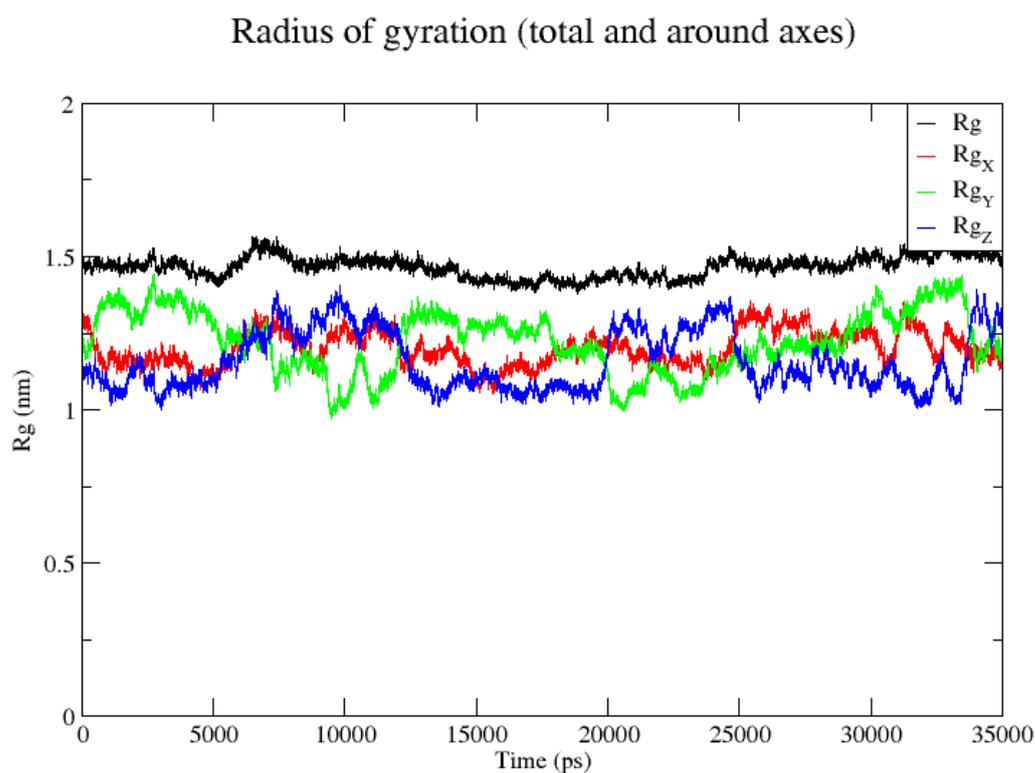


Fuente: Obtenida de GROMACS, 2020

3.8.3 ROG

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps). En general, el ROG es bastante estable con un valor medio de $1,47 \pm 0,04$ nm (Figura 38).

Figura 38 Gráfico del radio de giro (ROG) de los carbonos alfa de la proteína.

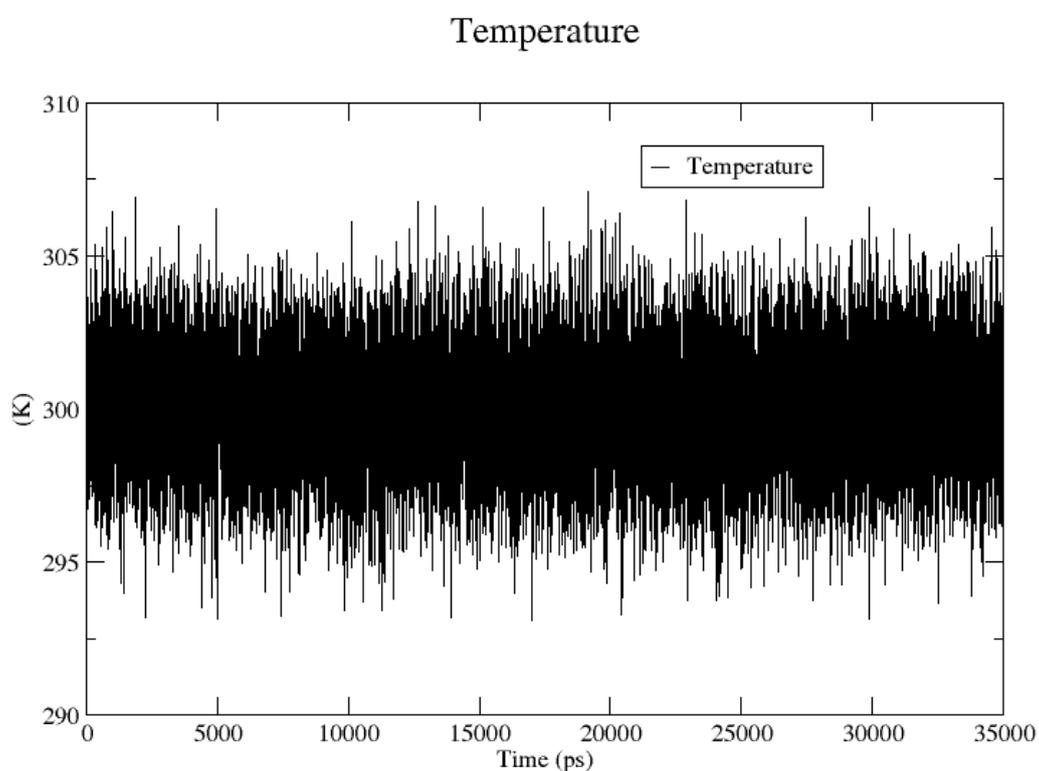


Fuente: Obtenida de GROMACS, 2020.

3.8.4 Temperatura

El gráfico de temperatura representa que, a lo largo de las simulaciones, la temperatura se mantuvo constante a 300 K. También muestra la estabilidad del sistema con ligeras fluctuaciones (Figura 39).

Figura 39 Temperatura

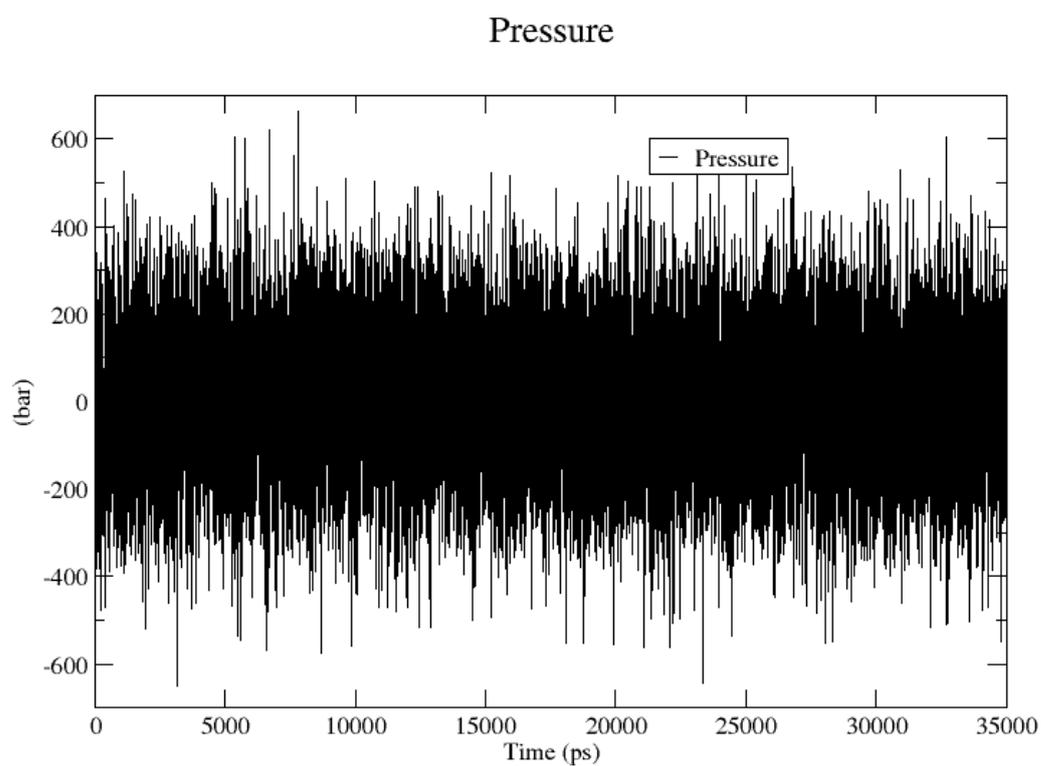


Fuente: Obtenida de GROMACS, 2020

3.8.5 Presión

El gráfico de presión ilustra que la presión se mantuvo constante a 0 bar durante toda la simulación que representa la estabilidad del sistema (Figura 40).

Figura 40 Presión

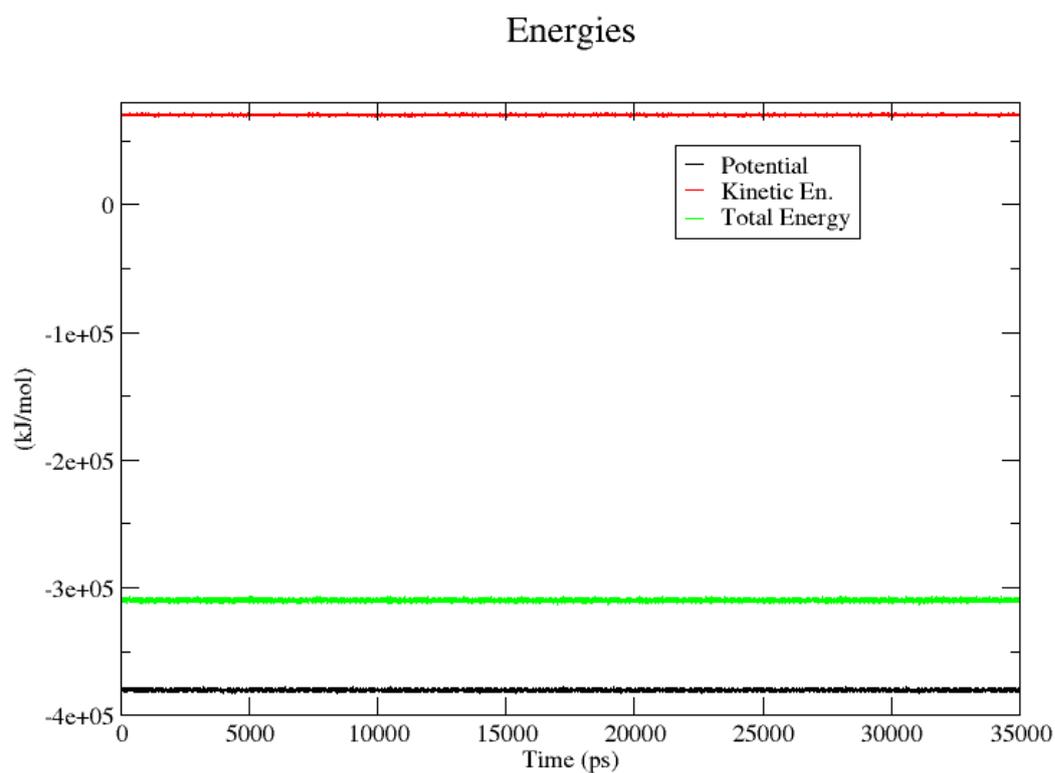


Fuente: Obtenida de GROMACS, 2020.

3.8.6 Energía

La energía total del sistema alcanza aproximadamente $-3e + 05$ KJ / mol, lo que ilustra la estabilidad del sistema (Figura 41).

Figura 41 Gráfico que representa la energía cinética, potencial y total del sistema.

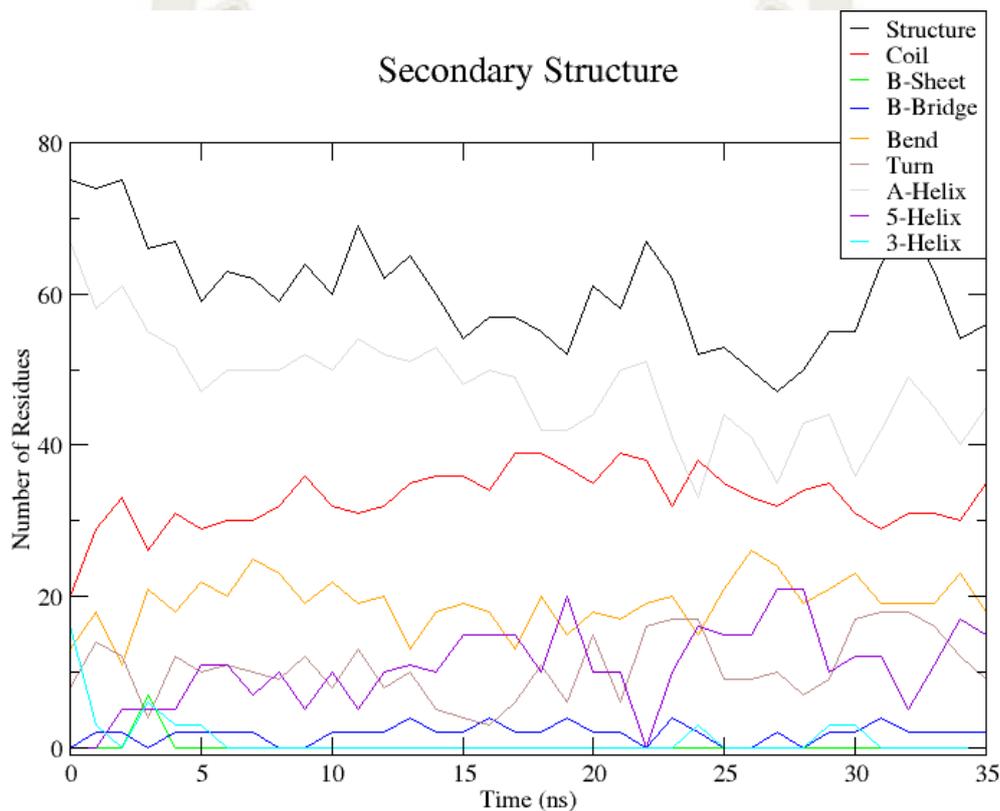


Fuente: Obtenida de GROMACS, 2020.

3.8.7 Análisis de estructura secundaria (DSSP)

El gráfico DSSP muestra las fluctuaciones entre la cadena principal de proteínas a lo largo de la escala de tiempo de simulación (Figura 42).

Figura 42 Fluctuación entre la estructura secundaria de la proteína (cadena principal).

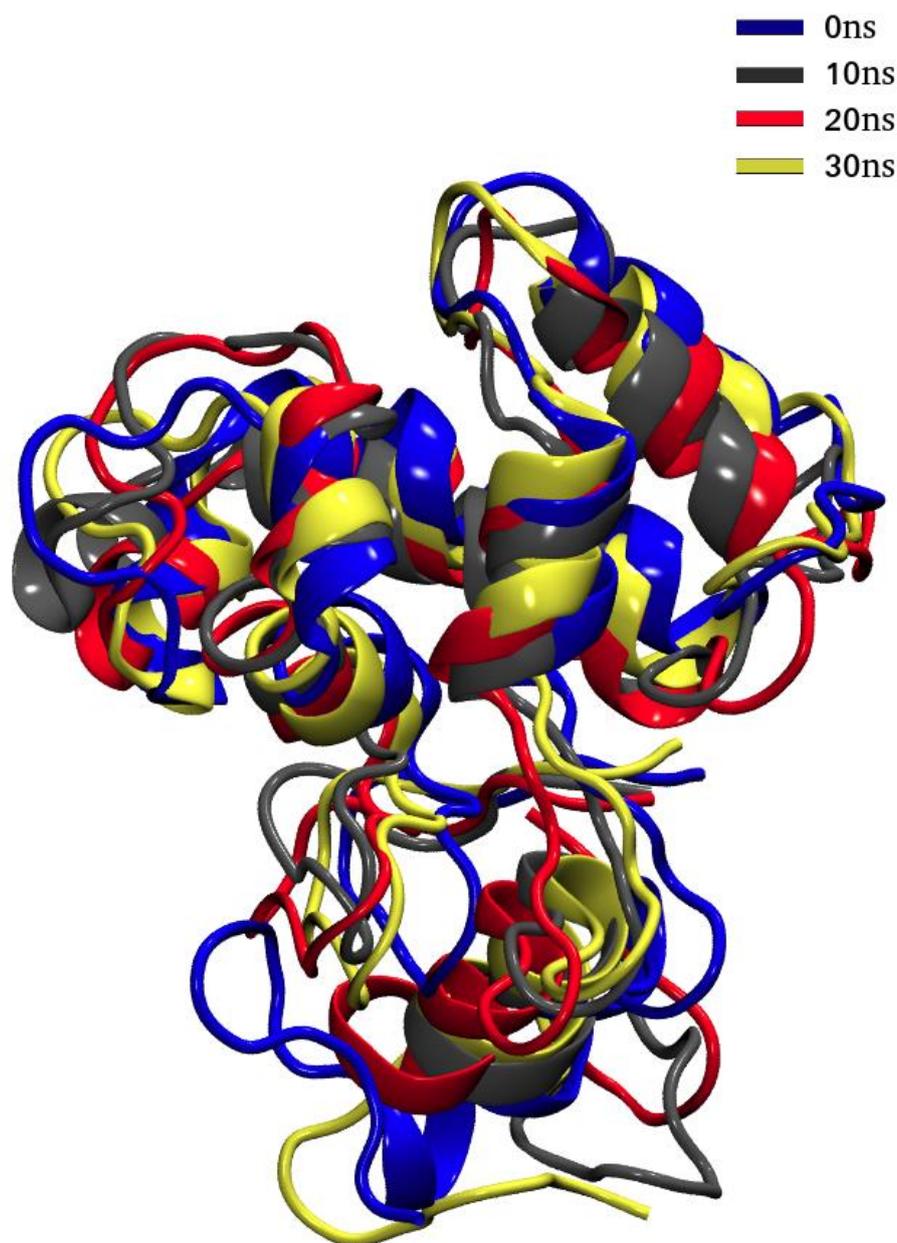


Fuente: obtenida de GROMACS, 2020.

3.8.8 Superposición de proteínas

La proteína se superpuso en diferentes escalas de tiempo, es decir, 0 ns, 10ns, 20ns, 30ns para analizar los cambios en la estructura y posición de la proteína (Figura 43).

Figura 43 Superposición de proteínas en diferentes nanosegundos.



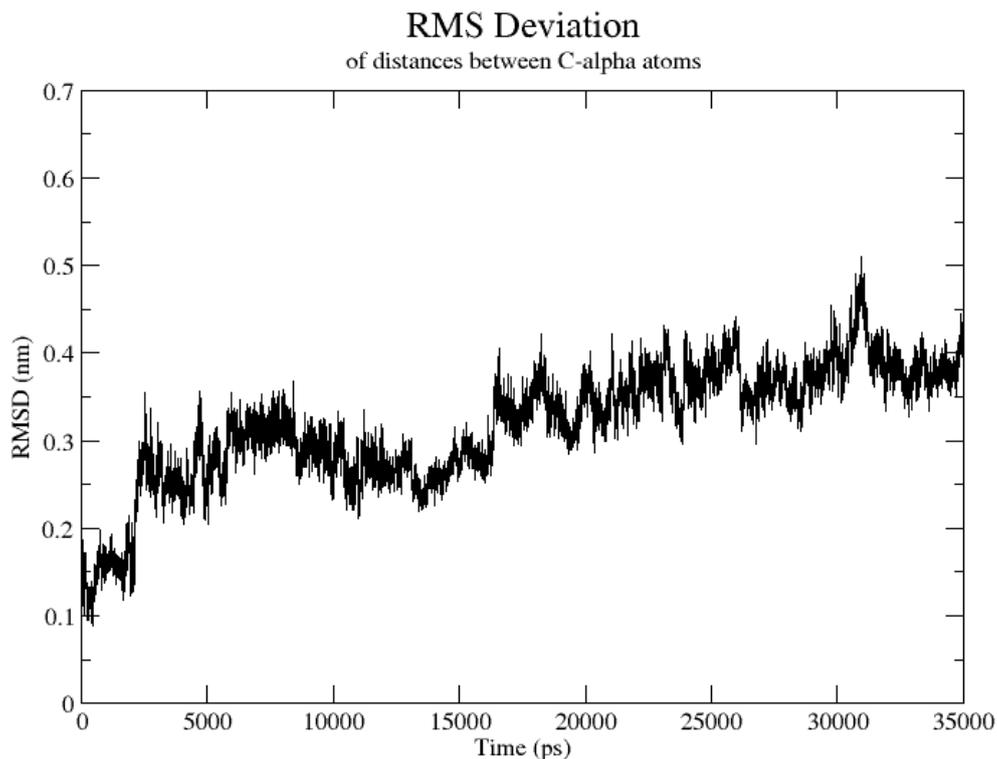
Fuente: Obtenida de Chimera, 2020.

3.9 Simulaciones de dinámica molecular de Hb Chile

3.9.1 RMSD:

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps). La desviación alcanza hasta 0,3 nm en los primeros 2,5 ns, luego, después de 17 ns, la RMSD aumenta aún más y se mantiene estable durante el resto de la simulación. En general, hay algunas fluctuaciones pronunciadas, que reflejan cambios estructurales. El RMSD medio es de 0,33 nm con una desviación estándar de 0,05 nm (Figura 44).

Figura 44 Gráfica de la desviación cuadrática media (RMSD).

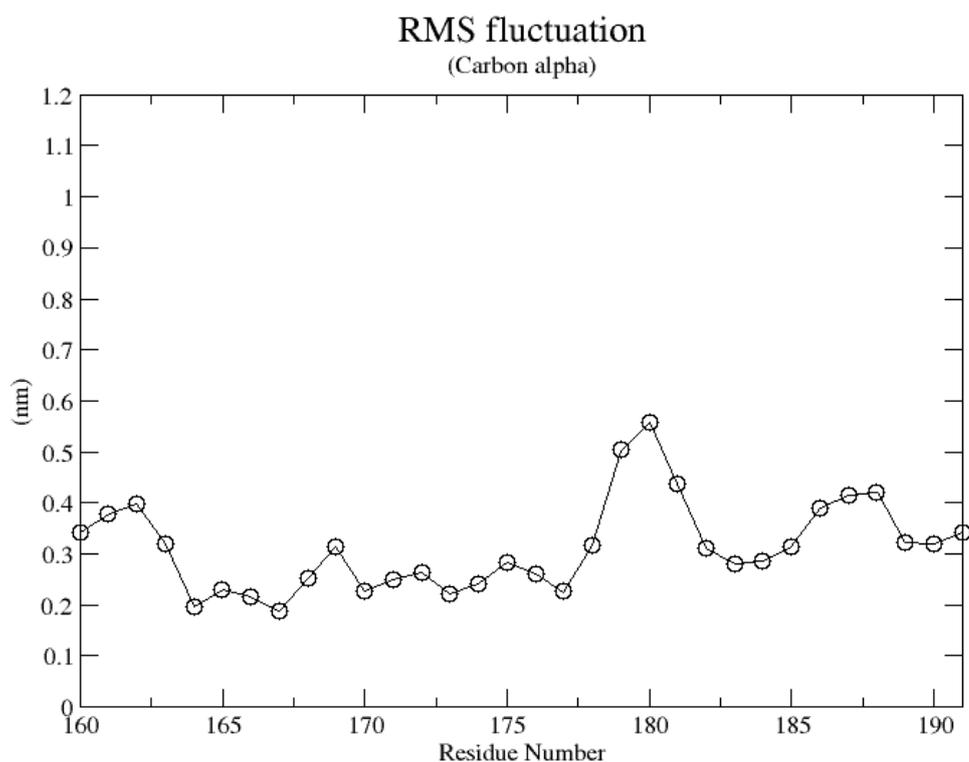


Fuente: Obtenida de GROMACS, 2020

3.9.2 RMSF

Se realizó un total de simulación de 35 ns y para el análisis RMSF, se descartó la primera trayectoria de 5 ns (5000 ps). Ambos terminales N y C son comparativamente menos fluctuantes y permanecieron por debajo de 0,4 nm. Sin embargo, los residuos 179-181 son muy fluctuantes y alcanzan más de 0,5 nm. (Figura 45).

Figura 45 Gráfico de fluctuación cuadrática media (RMSF).

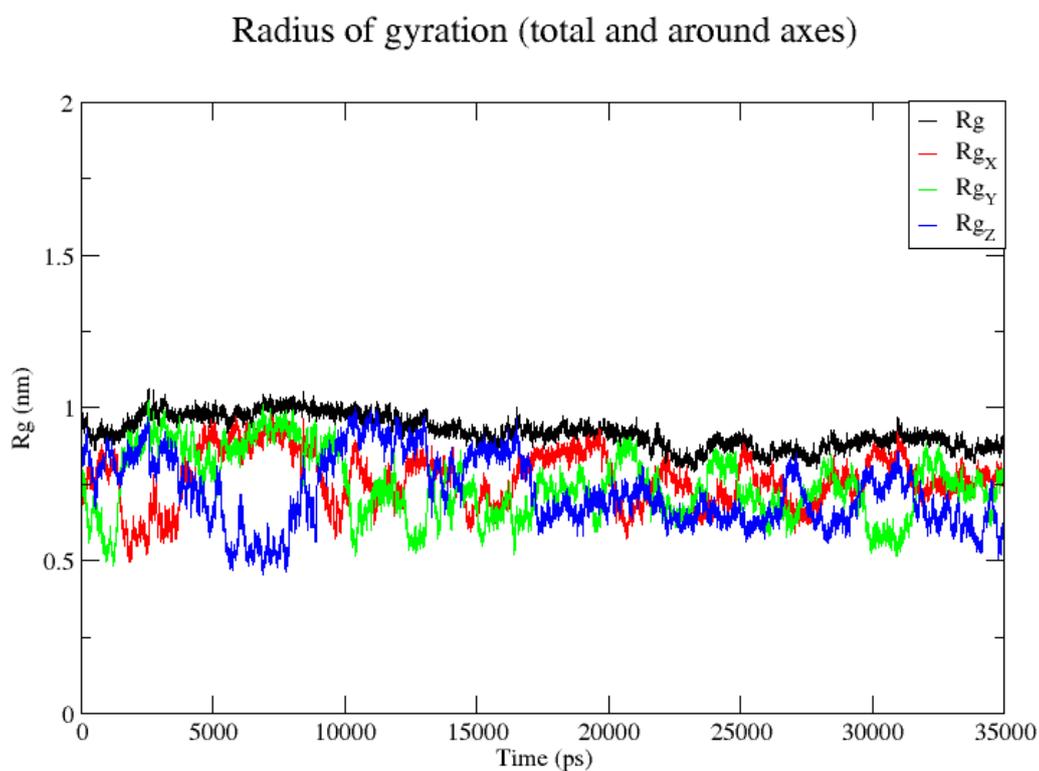


Fuente: Obtenida de GROMACS, 2020.

3.9.3 ROG

Se realizó un total de simulación de 35 ns y, para el análisis promedio, se descartó la trayectoria de los primeros 5ns (5000 ps). El gráfico muestra que ROG disminuyó en la simulación general. Sin embargo, después de 22 ns se observaron fluctuaciones prominentes, lo que indica algún cambio estructural. El ROG medio es de 0,92 nm con una desviación estándar de 0,05 nm (Figura 46).

Figura 46 Gráfico del radio de giro (ROG)

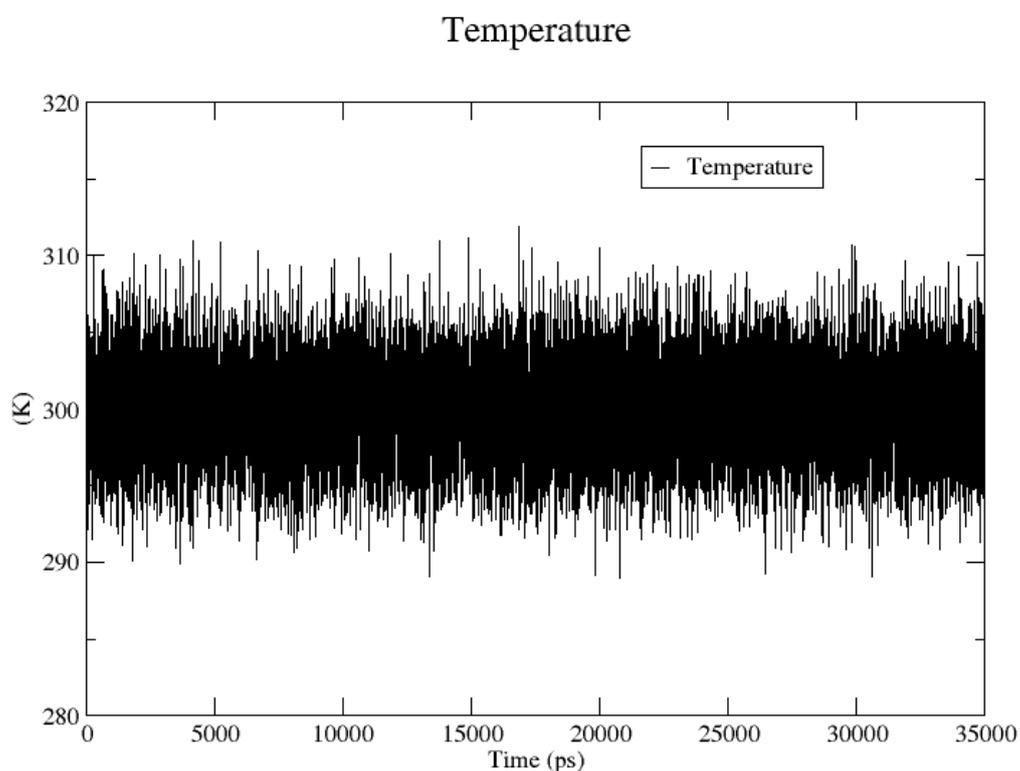


Fuente: Obtenida de GROMACS, 2020.

3.9.4 Temperatura

El gráfico de temperatura representa que a lo largo de las simulaciones, la temperatura se mantuvo constante a 300 K. También muestra la estabilidad del sistema con ligeras fluctuaciones (Figura 47).

Figura 47 Temperatura a 300 K

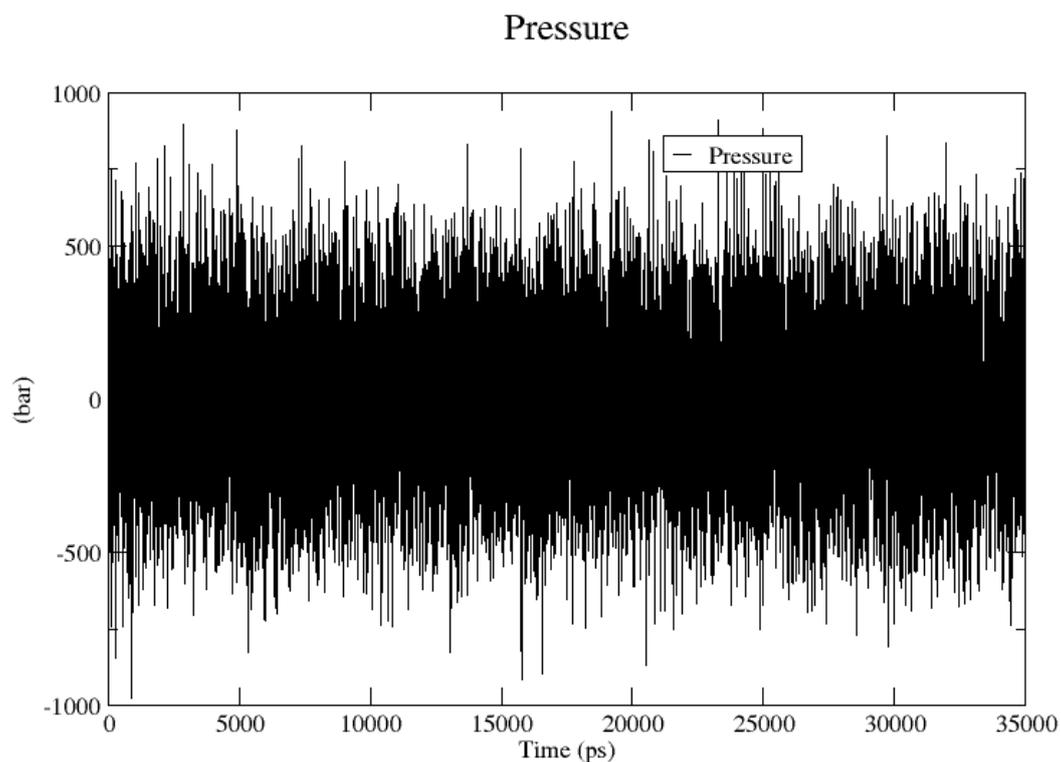


Fuente: obtenida de GROMACS, 2020.

3.9.5 Presión

El gráfico de presión ilustra que la presión se mantuvo constante a 0 bar durante toda la simulación que representa la estabilidad del sistema (Figura 48).

Figura 48. Presión a 0 bar.

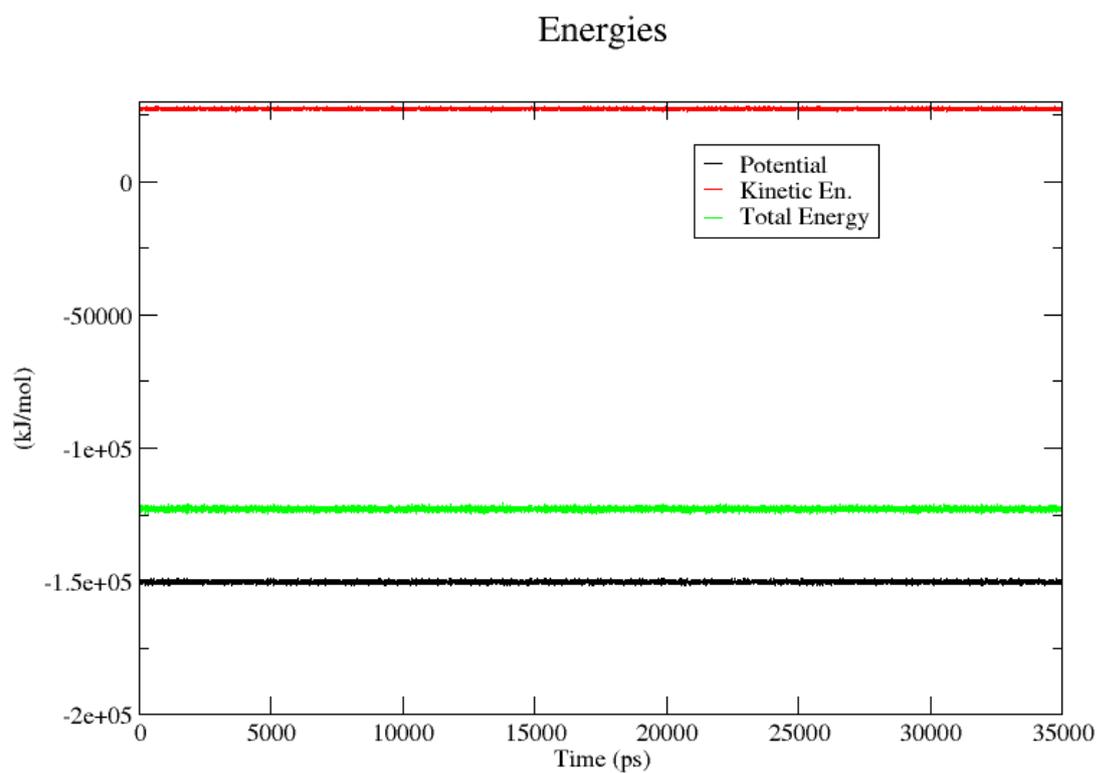


Fuente: Obtenida de GROMACS, 2020

3.9.6 Energía

La energía total del sistema alcanza aproximadamente $-1,3e + 05$ KJ / mol, lo que ilustra la estabilidad del sistema (Figura 49).

Figura 49 Gráfico que representa la energía cinética, potencial y total del sistema.

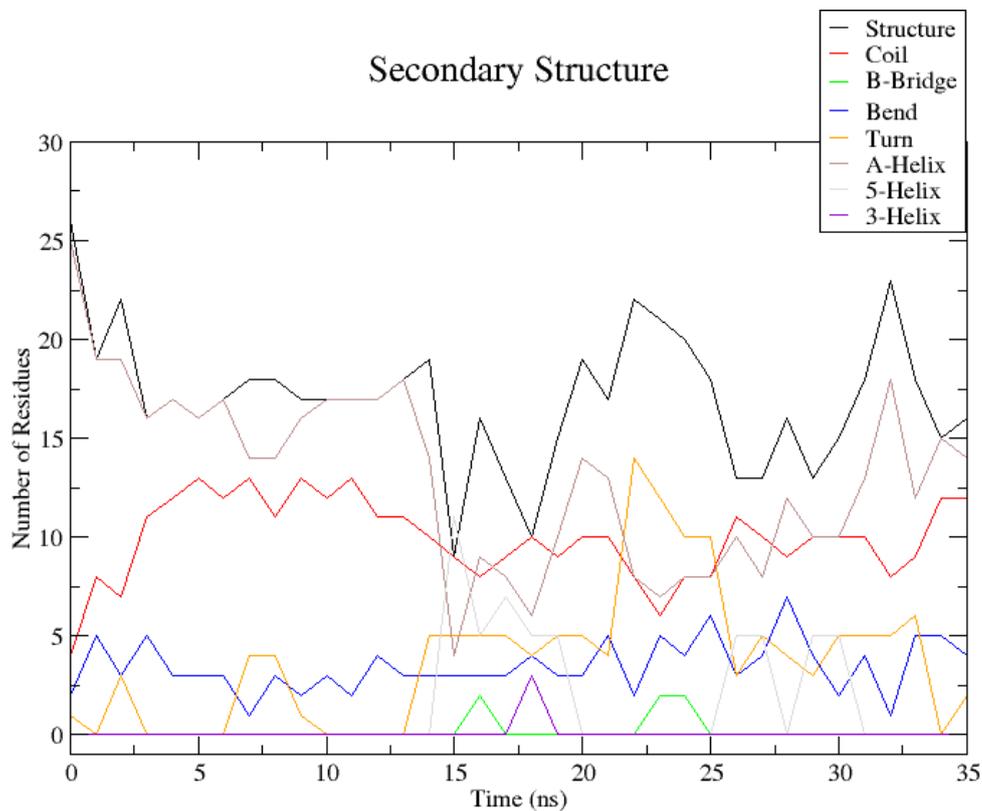


Fuente: Obtenida de GROMACS, 2020.

3.9.7 Análisis de estructura secundaria (DSSP)

El gráfico DSSP muestra las fluctuaciones entre la cadena principal de proteínas a lo largo de la escala de tiempo de simulación (Figura 50).

Figura 50 Fluctuación entre estructura secundaria de la proteína (cadena principal)

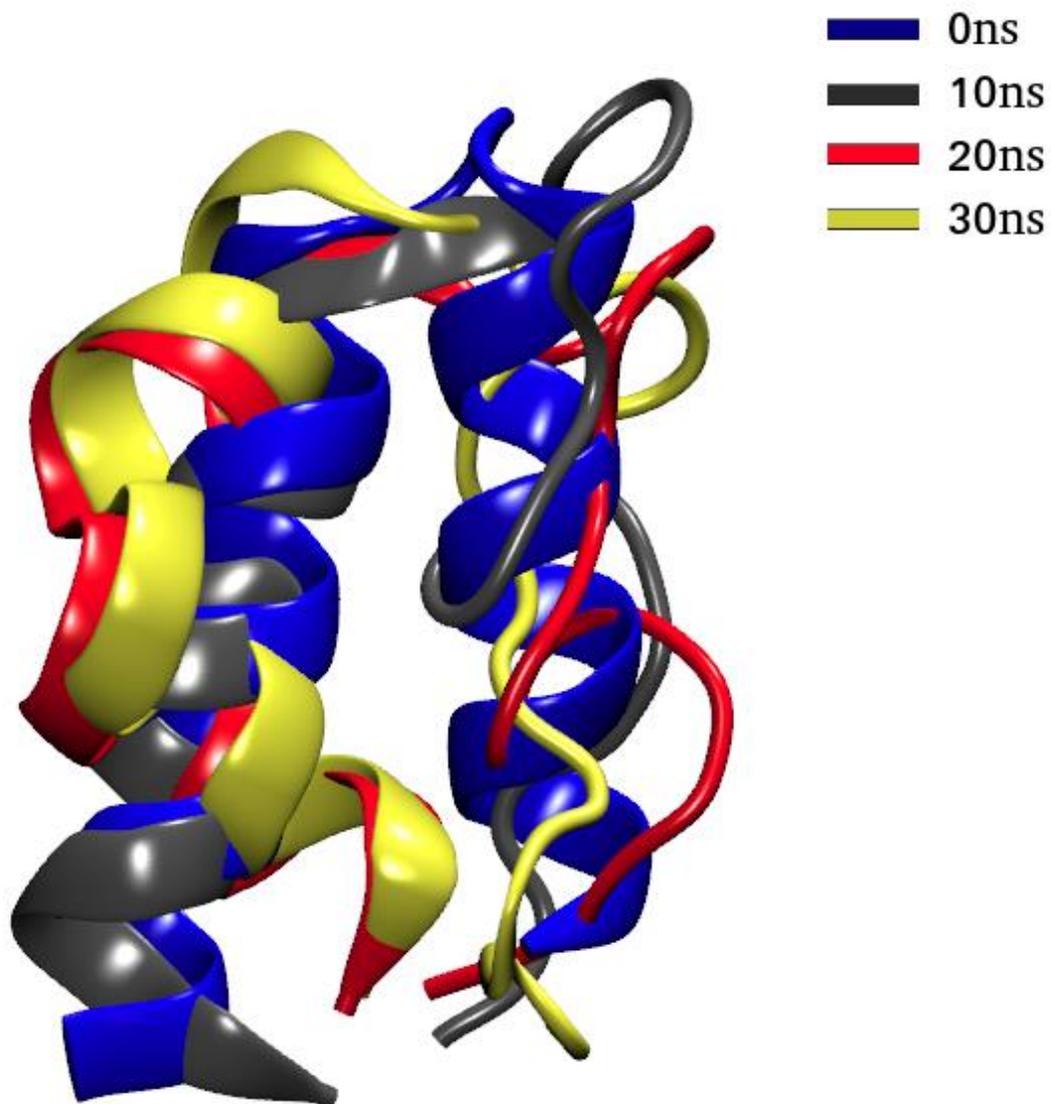


Fuente: Obtenida de GROMACS, 2020.

3.9.8 Superposición de proteínas

La proteína se superpuso en diferentes escalas de tiempo, es decir, 0 ns, 10ns, 20ns, 30ns para analizar los cambios en la estructura y posición de la proteína (Figura 51).

Figura 51 Superposición de proteínas en diferentes nanosegundos.

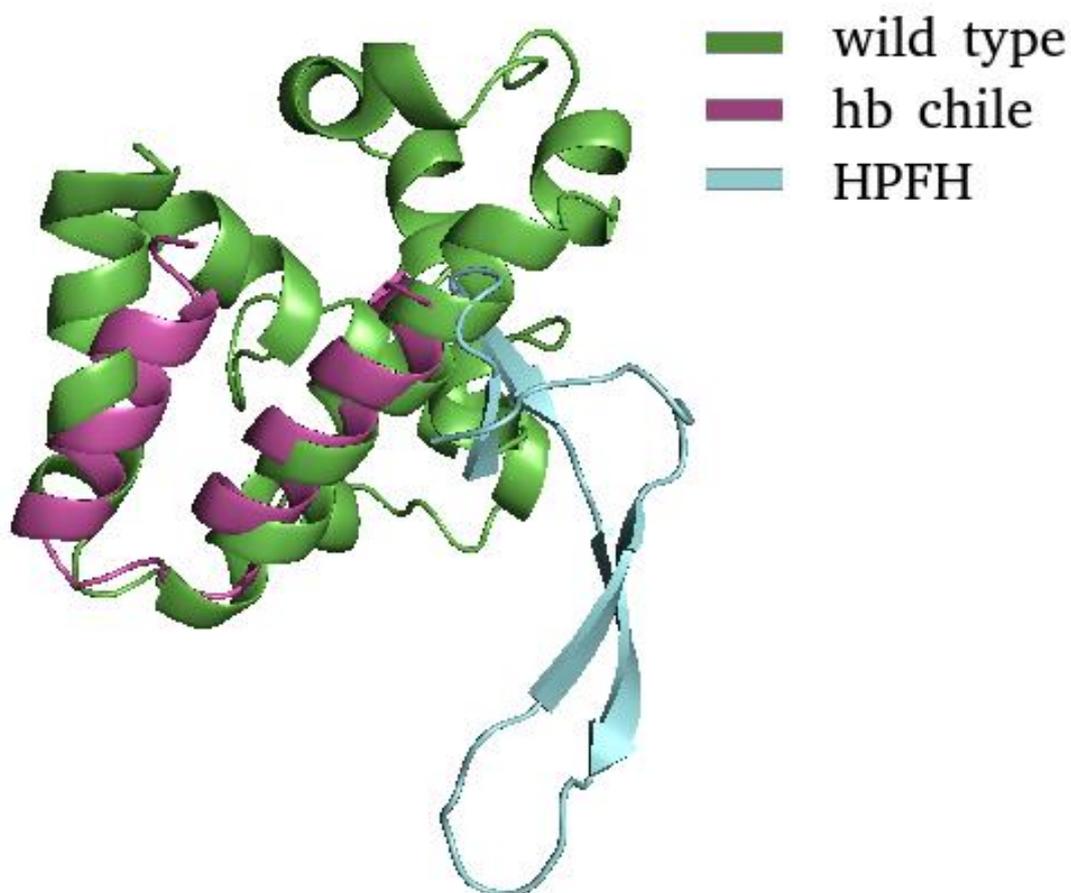


Fuente: Obtenida de Chimera, 2020.

3.10 Superposición de estructuras de todas las proteínas

Las tres estructuras de proteínas se superpusieron usando PyMol. Como la HPFH tiene cadena beta, no se superpone con otras proteínas (Figura 52).

Figura 52 Superposición de las tres proteínas



Fuente: Obtenido de PyMol, 2020.

4 DISCUSIÓN:

Construir la relación entre el fenotipo y el genotipo en el entorno clínico es uno de los principales objetivos de la investigación tradicional ⁸⁸. Sin embargo, los estudios sobre una gran cantidad de mutaciones son problemáticos, principalmente debido a los análisis experimentales. En contraste, el análisis in silico es más rápido y fácil de ejecutar, produce más resultados y cuesta menos, lo que lo hace más eficiente. Este tipo de análisis se basa en alteraciones en las secuencias de nucleótidos y / o aminoácidos, y su comparación con la secuencia nativa para correlacionar el efecto de estas alteraciones en el fenotipo del individuo ⁸⁹⁻⁹². Las mutaciones en el gen HBB, que se encuentra en el cromosoma 11 p15.5 ⁹³, son responsables de varias hemoglobinopatías graves, como anemia falciforme y β -talasemia. La anemia falciforme y la β -talasemia pueden provocar anemia grave y otras condiciones que amenazan la vida ⁹⁴.

Aunque el HBB está bien caracterizado, mutaciones en este gen registrados en la base de datos de HbVar son poco conocidos y no se han estudiado adecuadamente antes. En este estudio se recogió variantes de HBB de la región de América del Sur a partir de 1823 secuencias.

El análisis in silico mostró que la variante -195 (C-> G) Agamma; the Brazilian nd-HPFH adquirió una formación de bucle en la estructura 3D de la proteína (Figura 52). La variante Hb Chile a diferencia del anterior, no mostró formación de una cadena beta, pero en la simulación de dinámica molecular se presentó evidencia de efectos causados por la mutación en la flexibilidad de la proteína. El análisis RMSF mostró un alto grado de fluctuaciones en -195 (C-> G) Agamma; the Brazilian nd-HPFH, con la de wild-type en comparación con Hb Chile. (Figuras 29, 37 y 45 respectivamente). La sustitución de nuevos aminoácidos probablemente puede alterar el entorno de la proteína, lo que resulta en un nuevo conjunto de interacciones entre aminoácidos, que a su vez, podrían tener flexibilidad de las proteínas afectadas.

Esta pérdida de flexibilidad puede resultar en la pérdida de función de la proteína. Los estudios sugieren que un cambio en la función podría deberse debido al plegamiento de la secuencia de proteínas, esto particularmente ocurre debido al cambio en las propiedades fisicoquímicas, como hidrofobicidad, carga y geometría debido a la cadena lateral de los residuos de aminoácidos. Para probar aún más los efectos de estas mutaciones de sustitución sobre la estructura y función de las proteínas, es necesario realizar un análisis más profundo.

Los estudios sobre el radio de giro de C átomos de las proteínas de wild type y mutantes representadas en la estructura 3D, se ve más compacto y estable la wild type en comparación con las dos variantes. Estos datos sugieren que la sustitución en las mutaciones en la proteína HBB pueden estar afectando su estructura y función como se muestra en ambas variantes, pero se requieren más estudios intensivos para comprender completamente el alcance de estos efectos. Aún tenemos que determinar cómo, estas mutaciones afectan la flexibilidad de la proteína HBB y si esta pérdida afecta la función de las proteínas y en qué medida. Los estudios In vitro evaluarán aún más el comportamiento funcional de proteínas mutantes.

La OMS (Organización Mundial de la Salud) reporta que las hemoglobinopatías representan un importante problema sanitario cada año nacen más de 330 000 niños afectados (83% de casos de anemia de células falciformes y 17% de casos de talasemia). Las hemoglobinopatías causan aproximadamente un 3,4% de las defunciones entre los niños menores de 5 años.

En Perú que representa un complejo étnico importante en Latinoamérica el conocimiento sobre la situación actual de las hemoglobinopatías es limitado. El alto costo que significa realizar tamizaje de hemoglobinas anormales en recién nacidos, aunado a la importante mortalidad perinatal, contribuye a la escasez de diagnósticos de síndromes derivados de hemoglobinas anormales que tenemos, influye en la escasez de diagnósticos de síndromes derivados de hemoglobinas anormales. Las anormalidades de hemoglobina descritas en su forma homo u

heterocigotas, son todas producto de la inmigración, sólo podría ser demostrado si se realizan los estudios respectivos y se promueve el tamizaje neonatal de hemoglobinas anormales. Debido al entrecruzamiento genético podemos observar desde el punto de vista clínico tres tipos de talasemias: mayor u homocigota, intermedia y menor, que es la más frecuente anormalidad de hemoglobina que existe a nivel mundial, lamentablemente sub diagnosticada en nuestro medio por la costumbre que existe de dar hierro a todo paciente anémico sin realmente investigar la causa y porque esta alteración genética se confunde con la anemia por deficiencia de hierro. Del tipo mayor no hemos identificado un solo caso y aparentemente no se da todavía en el Perú. Dentro de las intermedias y las menores hemos encontrado varios pacientes, tanto en su forma alfa y beta. Solo se cuenta con información de un trabajo realizado en Lima por Castillo y colaboradores (1998) quienes estudiaron 5206 muestras de sangre, encontrando que un 7.2% eran hemoglobinas anormales, con 16 tipos de variantes, de los cuales el 3.2% correspondía a Hemoglobina S; 2.1% tenían Talasemia y 0.4% presentaban Hemoglobina S/ Talasemia ⁹⁵. Por otro lado, Roa y colaboradores analizaron 234 muestras de sangre de recién nacidos de tres ciudades andinas (Huancayo, Puno y Cerro de Pasco), ubicadas entre 3500 y 4400 metros sobre el nivel del mar. En dicha evaluación, no se detectó ninguna variante anormal de hemoglobina⁹⁶.

CONCLUSIONES

PRIMERA: Se estudió tres secuencias una wild-type y dos secuencias mutadas de HBB en la región de América del Sur. Se determinó que la secuencia mutada -195 (C-> G) Agamma; the Brazilian nd-HPFH presenta la formación de una cadena beta en la estructura de la proteína no pudiendo alinearse o superponerse en la Wild-type, ésto da por resultado la alteración de la función de la proteína, y por consiguiente el desarrollo de la enfermedad Talasemia.

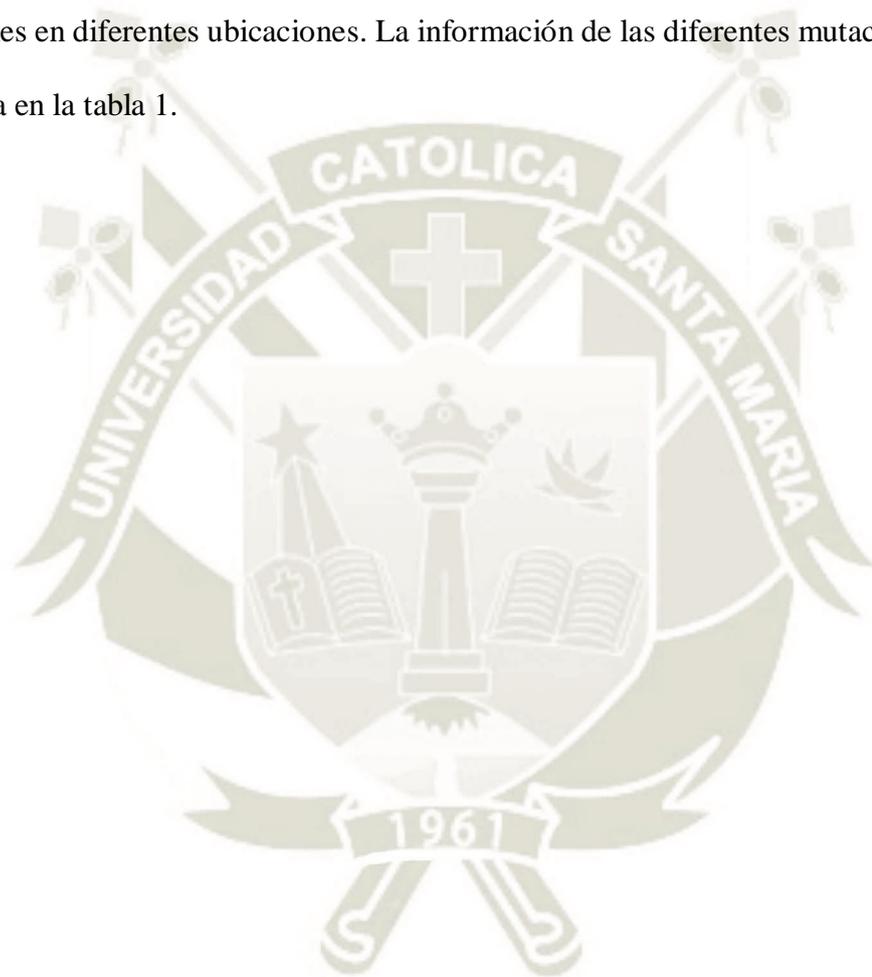
SEGUNDA: Se logró realizar el análisis de homología usando como plantilla con ID 1dxu.1 para la secuencia de Wild-type de HBB que tiene una identidad de secuencia del 73,64% , plantilla con ID 2lp6.1 fue elegido para -195 (C-> G) Agamma; the Brazilian nd-HPFH tiene una identidad de secuencia 28,57% , y la plantilla con ID 2w7k.1 se eligió para Hb Chile que tiene una identidad de secuencia del 6,25%

TERCERA: Se analizó la estabilidad estructural de la proteína mediante simulaciones de DM. -195 (C-> G) Agamma; the Brazilian nd-HPFH su gráfico RMSF muestra que la fluctuación de los terminales N y C alcanza hasta 0,6 nm y 0,7 nm, respectivamente. Sin embargo, el resto de los residuos están por debajo de 0,3 nm. Wild-type su gráfico RMSF muestra que el terminal N es bastante estable. Sin embargo, los residuos que van desde 79-83, 101-118 y 167-173 muestran una mayor fluctuación con un valor de RMSF superior a 0,3 nm. El terminal C es muy fluctuante y alcanza hasta 1,1 nm, lo que indica un cambio notable en la estructura. Hb Chile su gráfico RMSF, se descartó la primera trayectoria de 5 ns (5000 ps). Ambos terminales N y C son comparativamente menos fluctuantes y permanecieron por debajo de 0,4 nm. Sin embargo, los residuos 179-181 son muy fluctuantes y alcanzan más de 0,5 nm

CUARTA: Se evaluó los efectos de las mutaciones sobre la estructura de la proteína β -globina. Las tres estructuras de proteínas se superpusieron usando PyMol. Como la HPFH tiene cadena beta, no se superpone con otras proteínas.

QUINTA: Se determinó sus efectos sobre la función de la proteína β -globina. La formación de una cadena beta en la estructura de -195 (C-> G) Agamma; the Brazilian nd-HPFH explica la alteración de la función de la proteína, y por consiguiente el desarrollo de la enfermedad Talasemia.

SEXTA: Se analizó 4 variantes de la región HBB de América del Sur, tienen diferentes mutaciones en diferentes ubicaciones. La información de las diferentes mutaciones de HBB se encuentra en la tabla 1.



RECOMENDACIONES

PRIMERA: El presente trabajo de investigación y las conclusiones que se desprenden de él, es orientado a la obtención de información de las mutaciones que ocurren a nivel del gen beta hemoglobina en pacientes de procedencia sudamericana, se espera sea utilizada para posteriores estudios en busca de la terapéutica para las hemoglobinopatías.

SEGUNDA: Son escasas la documentación e investigaciones que analizan la frecuencia de hemoglobinopatías y su prevalencia en la población peruana, se recomienda la identificación temprana y correcta de ésta patología para asegurar el consejo genético e implementar los controles clínicos regulares con la correspondiente disponibilidad de tratamientos de apoyo (p. Ej., Transfusiones de glóbulos rojos, TIC orales), terapias génicas (p. Ej., Luspatercept) y opciones curativas (p. Ej., trasplante de médula ósea).

BIBLIOGRAFIA

1. D. J. Weatherall and J. B. Clegg. The Thalassaemia Syndromes, Blackwell Science, Oxford, UK, 2001.
2. Vives Corrons, J. L. Anemias por defectos congénitos de la hemoglobina. hemoglobinopatías estructurales y talasemias. Medicine - Programa de Formación Médica Continuada Acreditado, 2001, 8(51), 2684-2693.
3. Galanello, R., & Origa, R. Beta-thalassemia. Orphanet Journal of Rare Diseases, 2010, 5(1), 5-11.
4. Warsy, A. S., ElHazmi, M. A. F., Al Momin, A. K., AlHazmi, A., & Aleem, A. Extensive Polymorphisms in Saudi Beta Thalassaemia Patients. Biosciences Biotechnology Research Asia, 2013, 10(1), 127-132.
5. K. M. K.DeVooght, R. V. Wijk, and W. W. Van Solinge, "Management of gene promoter mutations in molecular diagnostics," Clinical Chemistry, 2009, vol. 55, no. 4, pp. 698–708,
6. Al-Ali, A. K., Al-Ateeq, S., Imamwerdi, B. W., Al-Sowayan, S., Al-Madan, M., Al-Muhanna, F., ... Qaw, F. Molecular Bases of β -Thalassemia in the Eastern Province of Saudi Arabia. Journal of Biomedicine and Biotechnology, 2005(4), 322-325.
7. AlMosawi, R. H. N., Al-Rashedi, N. A. M., & Ayoub, N. I. Clinical Laboratory Manifestation and Molecular Diagnosis of β -Thalassemia Patients in Iraq. Journal of Pediatric Hematology/Oncology, 2020, 42(1), 27-31.
8. De Sanctis, V. β -thalassemia distribution in the old world: a historical standpoint of an ancient disease. Mediterranean Journal of Hematology and Infectious Diseases, 2017, 9(1), e2017018.
9. Thein, S. L. Pathophysiology of β Thalassemia—A Guide to Molecular Therapies. Hematology, 2005(1), 31-37.

10. Qari, M. H., Wali, Y., Albagshi, M. H., Alshahrani, M., Alzahrani, A., Alhijji, I. A., ... Mousa, S. A. Regional consensus opinion for the management of Beta thalassemia major in the Arabian Gulf area. *Orphanet Journal of Rare Diseases*, 2013, vol.8 no.1, p. 143.
11. Giardine, B., van Baal, S., Kaimakis, P., Riemer, C., Miller, W., Samara, M., ... Patrinos, G. P. HbVar database of human hemoglobin variants and thalassemia mutations: 2007 update. *Human Mutation*, 2007, vol.28, no.2, p.206.
12. Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Bérout, G., Claustres, M., & Bérout, C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, 2009, 37(9), e67.
13. A map of human genome variation from population-scale sequencing. *Nature*, 2010, 467(7319), 1061-1073.
14. Medina, M. Á. Systems biology for molecular life sciences and its impact in biomedicine. *Cellular and Molecular Life Sciences*, 2012, vol.70, no 6, p. 1035-1053.
15. Schwede, T., Kopp, J., Guex, N. y Peitsch, MC. SWISS-MODEL: servidor automatizado de modelado de homología de proteínas. *Investigación de ácidos nucleicos*, 2003, vol. 31 no 13, p. 3381–3385.
16. Snyder, DA, Aramini, JM, Yu, B., Huang, YJ, Xiao, R., Cort, JR, Shastry, R., Ma, L.-C., Liu, J., Rost, B. y otros. Estructura de RMN en solución de la proteína ribosómica RP-L35Ae de *Pyrococcus furiosus*. *Proteínas: estructura, función y bioinformática*, 2012, vol. 80, no 7, p. 1901-1906.
17. Williams, CJ, Headd, JJ, Moriarty, NW, Prisant, MG, Videau, LL, Deis, LN, Verma, V., Keedy, DA, Hintze, BJ, Chen, VB y otros. MolProbity: más y mejores datos de referencia para una mejor validación de la estructura de todos los átomos. *Ciencia de las proteínas*, 2018, vol. 27, no 1, p. 293–315.

18. De Las Rivas, J., & Barber, J. Analysis of the Structure of the PsbO Protein and its Implications. *Photosynthesis Research*, 2004, vol. 81, no 3, p. 329-343.
19. Salwinski, L. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research*, 2004, vol. 32 no.90001, p. 449 - 451.
20. Millman, K. J., & Aivazis, M. Python for Scientists and Engineers. *Computing in Science & Engineering*, 2011, vol.13 no.2, p.9-12.
21. Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 2008 vol.4, vol.3, p. 435-447.
22. Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., ... Lopez, R. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Research*, 2015, vol. 43, no. W1, p.W580-W584.
23. Li, M., Suzuki, K., Qu, J., Saini, P., Dubova, I., Yi, F., ... Belmonte, J. C. I. Efficient correction of hemoglobinopathy-causing mutations by homologous recombination in integration-free patient iPSCs. *Cell Research*, 2011, vol. 21, no.12, p. 1740-1744.
24. Saraf, S. L., Molokie, R. E., Nouraie, M., Sable, C. A., Luchtman-Jones, L., Ensing, G. J., ... Gordeuk, V. R. (2014). Differences in the clinical and genotypic presentation of sickle cell disease around the world. *Paediatric Respiratory Reviews*, 15(1), 4-12.
25. Carlice-dos-Reis, T., Viana, J., Moreira, F. C., Cardoso, G. de L., Guerreiro, J., Santos, S., & Ribeiro-dos-Santos, Â. Investigation of mutations in the HBB gene using the 1,000 genomes database. *PLOS ONE*, 2017, vol.12, no.4, e0174637.
26. Miga, K. H., Newton, Y., Jain, M., Altemose, N., Willard, H. F., & Kent, W. J. Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Research*, 2014, vol. 24 no.4, p.697-707.

27. Riera, C., Lois, S., Domínguez, C., Fernandez-Cadenas, I., Montaner, J., Rodríguez-Sureda, V., & de la Cruz, X. Molecular damage in Fabry disease: Characterization and prediction of alpha-galactosidase A pathological mutations. *Proteins: Structure, Function, and Bioinformatics*, 2014, vol. 83 no.1, p.91-104.
28. Qadah, T., & Jamal, M. S. Computational Analysis of Protein Structure Changes as a Result of Nondeletion Insertion Mutations in Human β -Globin Gene Suggests Possible Cause of β -Thalassemia. *BioMed Research International*, 2019, p 1-8.
29. Guyton AC. Glóbulos sanguíneos. inmunidad y coagulación de la sangre. En: *Tratado de Fisiología Médica*. Interamericana: Edil: McGraw-Hill. México.1997, p. 465- 475.
30. Nelson DL, Cox MM. *Lehninger: Principles of Biochemistry*. Fifth ed. W. H. Freeman and Company. 2008, New York.
31. Villegas A, Roper P. Capítulo 10: Hemoglobinopatías y alteraciones de los eritrocitos. González JM, Medina JM. *Patología Molecular*. Mc Graw-Hill. Interamericana, Aravaca (Madrid). 2001, p. 171-192.
32. Safo MK, Ahmed MH, Ghatge MS, Boyiri T. Hemoglobin-ligand binding: understanding Hb function and allostery on atomic level. *Biochim Biophys Acta*, 2011. vol.1814, no. 6, p. 797- 809.
33. Sousounis K, Haney CE, Cao J, Sunchu B, Tsonis PA. Conservation of the threedimensional structure in non-homologous or unrelated proteins. *Hum Genomics*, 2012. 6: 10
34. Tárraga L, Smilg C. Tema 9: Transporte de gases. La hemoglobina. Volviendo a lo básico 2012 [cited; Available from: http://www.ffis.es/volviendoalobasico/tema_9_transporte_de_gases_la_hemoglobina.html.

35. Ibarra B, Perea FJ, Jaloma A, Romero F, González JR. Genética y Hematología. En Genética Clínica: Diagnostico de las Enfermedades Hereditarias. Guizar-Vázquez J (Ed) .. El Manual Moderno. México,2001, vol. 29 p.499 - 543.
36. Omori A., Tanabe O., Engel JO., Fukamizu A., Tanimoto K. Adult Stage gamma-Globin Silencing is Mediated by a Promoter Direct Repeat Element. Mol. Cell. Biol. 2005, vol 25, no.9, p. 3443 - 3451.
37. Patrinos GP., Kollia P., Papapanagiotou E., Loutradi-Anagnostou A., Loukopoulos D., Papadakis MN. Ay-Haplotypes: A New Group of Genetic Markers far Thalassemic Mutations Inside the 5' Regulatory Region of the Human Ay-Globin Gene. Am. J. Hemat. 2001, vol 66, p. 99 - 104.
38. Bunn F, Forget BG. Hemoglobin: Molecular, Genetic and Clinical Aspects: Edil: W.B. Saunders Company. USA. 1986.
39. Flint J., Harding RM., Boyce AJ., Clegg JB. The Population Genetics of the Haemoglobinopathies en Sickle Cell Disease and Thalassaemia. Rodgers G.P. Bailliere's Clínica Hematology. 1998 , vol 11, no 1, p.1 - 51.
40. Morales-Gonzáles KR. Caracterización Molecular de Hemoglobinopatías en Individuos de Tuxpan y Tamiahua en el Estado de Veracruz. Tesis de Licenciatura de QFB, Centro Universitario de Ciencias Exactas e Ingenierías. Universidad de Guadalajara. 2005.
41. Sans-Sabrafen J., Besses RC., Vives JL. Defectos Congénitos de la Hemoglobina: Hemoglobinopatías Estructurales. Hematología Clínica. Editorial Harcourt. España. 9: 2002, p. 183 - 222.
42. Weatherall DJ., Clegg JB. Inherited Haemoglobin Disorders: An Increasing Health Global Healt Problem. Bull WHO. 2001, vol 79, no. 8, p. 704- 712.
43. Rapaport SI. Introducción a la Hematología. Edil. Masson Doyma. España. 2002.

44. Colombo B., Svarch E., Martínez G. Introducción al Estudio de las Hemoglobinopatías. Edil: Científico-Técnica. La Habana Cuba. 1981.
45. Hardison RC., Chui DHK., Giardine B., Riemer C., Patrinos, GP., Anagnou, N., Miller W., and Wajcman H. Databases of Human Hemoglobin Variants and Other Resources at the Globin Gene Server. (Eds) Hemoglobin. 2001. (<http://globin.cse.psu.edu/hbvar/menu.html>), March 17 2006.
46. De Vries, H. (1901) Die Mutationstheorie.I. Leipzig:Veit & Co.Lieipzig, Germany. (English translation, 1910. The open court,Chicago).
47. Elliot, F. (1964) Citogenética y Manejo de Plantas. Cía Editorial Continental S.A. México. D.F
48. Sigurbjornsson, B. (1977) Manual on mutation breeding. 2° Edit. Technical reports series, No.119.International Atomic Energy Agency (IAEA).Viena.
49. Poehlman, J., Allen, D. Mejoramiento genético de las cosechas. Editorial Limusa. Segunda Edición.Mexico. 2003, p. 511.
50. Colombo C., M, Cornejo E., V., Raimann B., E. Errores innatos en el metabolismo del niño. Edit. Universitaria. Segunda Edición. Santiago de Chile. 2003.
51. Sadava, D., Purves, W. H. Vida: La Ciencia de la Biología. Ed. Médica Panamericana.Octava Edicion. Buenos Aires, Argentina. 2009.
52. Mba, C. Induced Mutations Unleash the Potentials of Plant Genetic Resources for Food and Agriculture. 2013, p. 200-231.
53. Brunner H. Radiation induced mutations for plant selection. Applied Radiation and Isotopes 1995, vol 46, no.6/7, p. 589-594.
54. Donini P., Sonnino A. (1998) Induced mutation in plant breeding: current status and future outlook. In Somaclonal Variation and Induced Mutation in Crop Improvement. S. Mohan Jain, D. S. Brar and B. S. Ahloowalia Editors. Kluwer Academic Publishers.1998, p. 255-291.

55. Predieri, S., R.H. Zimmerman. Pear mutagenesis: In vitro treatment with gamma rays and field selection for productivity and fruit traits. *Euphytica* 2001, vol 3, p.217–227.
56. Ravindra N. S., R. N. Kulkarn, M. C. Gayathri And S. Ramesh. Somaclonal variation for some morphological traits, herb yield, essential oil content and essential oil composition in an Indian cultivar of rose-scented geranium. *Plant breeding* 2004, vol 123, p. 84-86.
57. Able J., P. Langridge. Wild sex in the grasses. *Trends in Plant Science* 2006, vol 11, no. 6, p. 261-263.
58. Waugh R., D. J. Leader, N. McCallum And D. Caldwell. Harvesting the potential of induced biological diversity. *Trends in Plant Science* 2006, vol. 11, no. 2, p. 71-79.
59. Elliot, F. *Citogenética y Manejo de Plantas*. Cía Editorial Continental S.A. México. D.F, 1964.
60. Oliva V., R., Vidal T., J.M. *Genoma humano: nuevos avances en investigación, diagnóstico y tratamiento*. Ediciones Universitat Barcelona. Barcelona, España. 2006.
61. Van Harten, A. *Mutation breeding: Theory and practical applications*. Cambridge, USA: Cambridge Univ.Press. 1998.
62. Martin D., A. *Patología Quirúrgica*. Edit. Elsevier España. Tercera Edición. Madrid, España. (2004)
63. Monge-Nájera, J., Gomez F., P., Rivas R., M. *Biología General*. Editorial Universidad Estatal a Distancia EUNED. Primera edición. San José, Costa Rica. 2002.
64. Doolittle, R. F. Similar amino acid sequences: chance or common ancestry. *Science* , 1981; 214 (4517), 149-159.
65. Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., ... & Twigger, S.. Big data: The future of biocuration. *Nature* , 2008; 455 (7209), 47-50.

66. Radusky, Leandro Gabriel. Herramientas bioinformáticas para el análisis estructural de proteínas a escala genómica. 2017.
67. Belinda Giardine, Joseph Borg, et. al.. Updates of the HbVar database of human hemoglobin variants and thalassemia mutations. 2014 Jan 1; 42.
68. Feng, D. F., & Doolittle, R. F. Progressive sequence alignment as a prerequisite for correct phylogenetic trees. *Journal of molecular evolution*, 1987; 25 (4), 351-360.
69. Simons KT, Bonneau R, Ruczinski I, Baker Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins Suppl D* 1999; 3:171-6
70. Webb, B., & Sali, A. (2014). Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics*, 5-6.
71. Peitsch M.C. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* 1996; 24:274–279.
72. Benkert P., Biasini M., Schwede T.. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics.* 2011; 27:343–350.
73. Biasini M., Bienert S., Waterhouse A., Arnold K., Studer G., Schmidt T., Kiefer F., Gallo Cassarino T., Bertoni M., Bordoli L. et al. . SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* 2014; 42: W252–W258.
74. Haas J., Barbato A., Behringer D., Studer G., Roth S., Bertoni M., Mostaguir K., Gumienny R., Schwede T.. Continuous automated model evaluation (CAMEO) complementing the critical assessment of structure prediction in CASP12. *Proteins.* 2018; 86(Suppl. 1):387–398
75. Berman H.M., Battistuz T., Bhat T.N., Bluhm W.F., Bourne P.E., Burkhardt K., Feng Z., Gilliland G.L., Iype L., Jain S. et al. . The Protein Data Bank. *Acta Crystallogr. D, Biol. Crystallogr.* 2002; 58:899–907.

76. Björn Wallner and Arne Elofsson. Identification of correct regions in protein models using structural, alignment, and consensus information. 2006; 15(4): 900–913.
77. Christopher J. Williams, 1 Jeffrey J. Headd, 1 , 5 Nigel W. Moriart et al.. MolProbity: More and better reference data for improved all-atom structure validation. 2018 Jan; 27(1): 293–315
78. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature. 1977;267(5612):585–590.
79. Hermans J, Berendsen HJC, Vangunsteren WF, Postma JPM. A consistent empirical potential for water-protein interactions. Biopolymers. 1984;23(8):1513–1518.
80. Rueda M, Ferrer-Costa C, Meyer T, et al. A consensus view of protein dynamics. Proc Natl Acad Sci U S A. 2007;104(3):796–801.
81. About GROMACS. (2018).https://www.gromacs.org/About_Gromacs
82. Backofen R., Gilbert D. “Bioinformatics and Constraints”. Constraints, 2001, vol 6, p. 141-156.
83. Stephanopoulos G. “Bioinformatics and Metabolic Engineering” (Foreword). Metabolic Engineering 2000 vol. 2, p. 157-158.
84. Sackman H. Biomedical Information Technology. Global Social Responsibilities for the Democratic Age. Academic Press, San Diego, CA (USA), 1997.
85. Van Bemmelen J.H., Musen M.A.(eds.) Handbook of Medical Informatics. SpringerVerlag, Heidelberg, 1997.
86. Jones R.L. "The Internet and Healthcare Information Systems: How Safe Will Patient Data Be?". IS Audit & Control Journal, I; 1998, p. 25-30.
87. Roberts R. “Bioinformatics Analysis of Gene Banks Provides a Treasure Trove for the Functional Genomist”. J. Mol. Cell. Cardiol.; 2000, vol 32, p. 1917-1919.

88. Kadian Singh, P., & Mistry, K. N. A Computational Approach to Determine Susceptibility to Cancer by Evaluating the Deleterious Effect of nsSNP in XRCC1 Gene on Binding Interaction of XRCC1 Protein With Ligase III. *elsevier*, 2016, p. 141-149.
89. Lettre, G. The Search for Genetic Modifiers of Disease Severity in the Hemoglobinopathies. *Cold Spring Harbor Perspectives in Medicine*, 2012, 2(10), a015032.
90. Steinberg, M. H., & Sebastiani, P. Genetic modifiers of sickle cell disease. *American Journal of Hematology*, 2012, 87(8): 795-803.
91. Steinberg, M. H. Genetic Etiologies for Phenotypic Diversity in Sickle Cell Anemia. *The Scientific World JOURNAL*, 2009, 9, 46-67.
92. Borgio, J. F., AbdulAzeez, S., Al-Muslami, A. M., Naserullah, Z. A., Al-Jarrash, S., Al-Suliman, A. M., ... Al-Ali, A. K. KLF1 gene and borderline hemoglobin A2 in Saudi population. *Archives of Medical Science*, 2018, 1, 230-236.
93. Onda, M., Akaishi, J., Asaka, S., Okamoto, J., Miyamoto, S., Mizutani, K., ... Emi, M. Decreased expression of haemoglobin beta (HBB) gene in anaplastic thyroid cancer and recovery of its expression inhibits cell growth. *British Journal of Cancer*, 2005, 92(12), 2216-2224.
94. D. J. Weatherall, "Hemoglobinopathies worldwide: present and future," *Current Molecular Medicine*, 2008, vol. 8, no. 7, pp. 592-599.
95. Castillo J, Hazán E, Márquez M. Las Hemoglobinopatías en el Perú. *Rev Med Inst Peru Segur Soc* 1998; 7:7-17.
96. Roa D, Aguinaga M, Ruiz W, Ulloa V, Turner E. Búsqueda de Hemoglobinas anormales en los recién nacidos en las grandes alturas. *Rev Med Hered* 1997; 8:87-91.