



UNIVERSIDAD
NACIONAL
DE COLOMBIA

Análisis computacional de la Hemaglutinina de los virus influenza A de linaje pandémico en Colombia

Jaime Andrés Úsuga Restrepo

Universidad Nacional de Colombia

Facultad de Minas

Medellín, Colombia

2023

Análisis computacional de la Hemaglutinina de los virus influenza A de linaje pandémico en Colombia

Jaime Andrés Úsuga Restrepo

Tesis presentada como requisito para optar al título de:
Magister en Ingeniería – Materiales y Procesos

Director (a):

Ph.D. Juan Pablo Hernández Ortiz

Codirector (a):

Ph.D. Karl Adolf Ciuoderis Aponte

Grupo de Investigación:

Center for Research and Surveillance of Tropical and Infectious Diseases (CRS-TID),
Colombia-Wisconsin One Health Consortium.

Universidad Nacional de Colombia
Facultad de Minas, Departamento Materiales y Minerales
Medellín, Colombia

2023

Declaración de obra original

Yo declaro lo siguiente:

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

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

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

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

Nombre Jaime Andrés Úsuga Restrepo

Fecha 01/08/2023

Resumen

Análisis computacional de la hemaglutinina de los virus de influenza A de linaje pandémico en Colombia

Aunque se ha demostrado la amplia circulación del virus influenza (VI) A en cerdos y humanos en el mundo y en Colombia, existen pocos estudios que hayan caracterizado los aspectos biológicos de estos virus. Por tanto, éste trabajo se basó en el análisis computacional de la proteína de superficie hemaglutinina (HA) de cepas pandémicas del VI circulantes en el país, y su relación con un cambio en las propiedades biológicas de estos. Este trabajo se desarrolló en tres capítulos, en el primero se caracterizó la composición genómica de la HA, seguido de la identificación de las mutaciones en esta proteína y su posible contribución a la diseminación de estos virus en la región. Los últimos dos capítulos, correspondieron al análisis estructural de la HA y su interacción con los receptores celulares. En general, se sugiere que las mutaciones I322V, P84S, S204T, E375K entre otras, pudieron haber contribuido al fitness y en el establecimiento de estos virus en las poblaciones actuales en Colombia. Las mutaciones más relevantes (E375K y S163N) estaban relacionadas con el aumento de la virulencia y la capacidad de evadir la respuesta inmune. Adicionalmente, se observó un clúster filogenéticamente relacionado que presentó dos mutaciones únicas (D223X y Q224X) no reportadas previamente. Por otra parte, para los modelos de los consensos tridimensionales de la HA, se observó que el modelo cBri18 (con mutaciones únicas R46G, P283A y I299V), representativo de su clúster, mostró comportamientos estructurales ligeramente distintos a los otros modelos y presentó cambios en el *loop*130 del sitio de unión al receptor. Adicionalmente, luego del análisis de los modelos de HA más relevantes, se encontró que estos mantienen preferencia la unión con los receptores celulares humanos (SA α 2,6) sobre los aviares (SA α 2,3). También que el modelo cCal09 tuvo mayor afinidad por SA α 2,6; y que el modelo

cBri18 fue el única interactuó con SA α 2,3, hecho que puede atribuirse a los cambios evidenciados en el *loop*130. En conclusión, la presencia de ciertas mutaciones en la proteína HA de VI H1N1pdm de Colombia están posiblemente influenciando las propiedades biológicas de estos virus, sin embargo, otros estudios son necesarios para poder confirmar estos hallazgos. No obstante, este conocimiento generado aporta en el fortalecimiento de las acciones en la vigilancia y control de estos virus en el país y la región.

Palabras clave: Virus de influenza, Colombia, HA, mutaciones, diversidad genética, docking, dinámica molecular.

Abstract

Computational analysis of the hemagglutinin of pandemic lineage influenza A viruses in Colombia

Despite circulation of influenza virus (IV) in pigs and humans globally, few studies have characterized its biological features in Colombia. Therefore, in this study a computational analysis of the hemagglutinin (HA) surface protein of IV pandemic strains circulating in the country, and its relationship with a change in their biological properties was carried out. This work was developed in three chapters, in the first one, the genomic composition of HA was characterized, followed by the identification of the mutations and their possible contribution to the dissemination of these viruses in the region. The last two chapters corresponded to the structural analysis of HA and its interaction with cell receptors. In general, the mutations I322V, P84S, S204T, E375K, among others, have contributed to the fitness and establishment of these viruses in current populations in Colombia. The most relevant changes (E375K and S163N) were related to increased virulence and the ability to evade the immune response. In addition, a phylogenetically related cluster was found that exhibited two unique changes (D223X and Q224X) not previously reported. On the other hand, for the three-dimensional consensus models of the HA, it was shown that the cBri18 model (with exclusive mutations R46G, P283A and I299V), representative of its cluster, showed structural changes in the loop130 of receptor binding domain behaviors slightly different from the other models. In addition, analyzing the most relevant HA models, it was found that they preferentially bind to human cell receptors (SA α 2,6) over avian (SA α 2,3). Also, that the cCal09 model had a higher affinity for SA α 2,6; and that the cBri18 model was the only one that interacted with SA α 2,3, a fact that can be attributed to the changes observed in loop130. In conclusion, the presence of certain mutations in the HA protein of

IV H1N1pdm from Colombia might be influencing the biological properties of these viruses, however, other studies are necessary to confirm these findings. However, this generated knowledge contributes to strengthening actions in the surveillance and control of these viruses in the country and the region.

Keywords: Influenza virus, Colombia, HA, mutations, genetic diversity, docking, molecular dynamics.

Contenido

Resumen.....	6
Abstract.....	8
Lista de figuras.....	11
Lista de tablas.....	15
Lista de abreviaturas.....	16
1. Introducción	17
1.1 Planteamiento del problema	17
1.2 Justificación	18
1.3 Objetivos.....	20
1.3.1 Objetivo General	20
1.3.2 Objetivos Específicos.....	20
2. Fundamentos teóricos	21
2.1 Biología del virus	21
2.1.1 El virus de influenza (VI)	21
2.1.2 Hemaglutinina y sitio de unión al receptor celular	22
2.1.3 Cambio y evolución viral	23
2.1.4 Diversidad genética	25
2.1.5 Transmisión viral.....	26
2.1.6 Infección viral.....	28

2.2	Técnicas computacionales para el estudio de estructuras proteicas y caracterización genética	29
2.2.1	Análisis filogenético	29
2.2.2	Modelamiento por homología de estructuras proteicas	30
2.2.3	Docking en la predicción de interacciones en complejos biológicos	31
2.2.4	Dinámica Molecular para el estudio de proteínas y complejos proteicos	32
2.2.5	Energía libre de unión para el estudio interacciones biomoleculares.....	34
3.	Antecedentes.....	35
4.	Capítulo 1: Análisis de mutaciones gen HA de VI de linaje H1N1pdm	36
4.1	Metodología.....	36
4.1.1	Preparación de la base de datos	36
4.1.2	Filogenia para la caracterización y análisis evolutivo	37
4.1.3	Detección de mutaciones	38
4.2	Resultados	39
4.2.1	Secuencias de H1 en Colombia	39
4.2.2	Identificación de mutaciones	39
4.2.3	Análisis evolutivo de los VI H1N1pdm	42
4.3	Discusión.....	47
5.	Capítulo 2: Modelación estructural consenso HA VI H1N1pdm	51
5.1	Metodología.....	51
5.1.1	Cálculo de secuencias consenso	52
5.1.2	Modelamiento por homología y validación de las estructuras	52
5.1.3	Dinámica molecular de las proteínas.....	53
5.2	Resultados	53
5.2.1	Cálculo de secuencias consenso representativas del VI H1N1pdm	53
5.2.2	Modelado de la estructura 3D.....	55
5.2.3	Dinámica molecular de las estructuras.....	60
5.3	Discusión.....	65
6.	Capítulo 3: Análisis unión HA y receptores celulares.....	67
6.1	Metodología.....	68
6.1.1	Preparación de ligandos de receptores celulares.....	68
6.1.2	Docking Ligando-Proteína	68
6.1.3	Dinámica molecular HA y receptor celular.....	69
6.1.4	Cálculo de energías de unión	70
6.2	Resultados	70
6.2.1	Selección de modelos.....	70
6.2.2	Docking molecular	71
6.2.3	Dinámica molecular ligando-proteína	73
6.2.4	Energías de unión.....	81
6.3	Discusión.....	85
7.	Conclusiones	88
8.	Anexos.....	90

Lista de figuras

- Figura 1. Esquema de la estructura y composición del Virus Influenza tipo A.** Dos proteínas de superficie, HA y NA, y la proteína M2 están incrustadas en la envoltura viral, que se deriva de la membrana plasmática del huésped. El complejo de ribonucleoproteína (RNP) comprende un segmento de ARN viral asociado con la NP y tres polimerasas (PA, PB1 y PB2). La proteína de la matriz (M1) está asociada tanto a la RNP como a la envoltura viral. (Horimoto & Kawaoka, 2005). 22
- Figura 2. Estructura de HA del VI H1N1 pandémico del 1918 resaltando regiones más importantes.** La cabeza contiene el sitio de unión al receptor de ácido siálico, que está rodeado por cinco sitios antigénicos (Sa, Sb, Ca1, Ca2 y Cb). El tallo comprende las hélices A y B y el péptido de fusión. (Bouvier & Palese, 2008) 23
- Figura 3. Fenómenos de cambio antigénico y re-arreglo antigénico en el VI.** Los dos mecanismos principales que determinan la evasión inmune. La primera imagen se refiere al cambio antigénico que se da por las mutaciones que se acumulan en las proteínas de superficie (HA y NA), generando que el virus evada la respuesta inmune. El segundo mecanismo es el re-arreglo antigénico, que ocurre cuando segmentos del material genético de diferentes cepas, dando como resultado un virion completamente nuevo. (Boyoglu-Barnum et al., 2021) 24
- Figura 4. Relación genética de los linajes de virus humanos y porcinos.** Las flechas verdes, naranjas, moradas y azules siguen la evolución de la hemaglutinina, mientras que las flechas negras siguen otros segmentos. La figura muestra la procedencia de distintos linajes que han afectado a humano y cerdos. (Rambo-Martin et al., 2020) 26
- Figura 5. El enlace glucosídico entre el ácido siálico y el penúltimo azúcar de las cadenas.** La estructura del ácido N -acetilneuramínico, un tipo de ácido siálico en sus conformaciones de enlace. En la conformación $\alpha 2,6$ es el receptor celular encontrado principalmente en el tracto humano, la conformación $\alpha 2,3$, es la forma de receptor que se encuentra principalmente en el tracto de las aves. Modificado de Byrd-Leotis y colaboradores (Byrd-Leotis et al., 2017) 27
- Figura 6. Proceso de infección de un virión del VI.** Para infectar una célula huésped, el virión debe unirse efectivamente a la forma del SA que esté recubriendo la pared celular, luego, ocurre la endocitosis, en donde el ARN viral se libera y la célula usa el material de replicación. El nuevo virión formado brota a la superficie unido nuevamente al SA. Es aquí donde la NA es de hidrolizar la unión, permitiendo la liberación del virus. Modificado de (Wennekes Tom, 2013) 29
- Figura 7. Diagrama de Ramachandran que muestra los valores de los ángulos ϕ y ψ .** En verde se muestran regiones de los valores permitidos de los ángulos ϕ y ψ de la

conformación de proteínas. Los gráficos de la izquierda son estructuras con sus ángulos en regiones permitidas y la figura de la derecha es desfavorecida debido a los choques estéricos. (Mark Berg Jeremy et al., 2006)..... 31

Figura 8. Gráfico del *docking* entre un ligando y un receptor. Este gráfico muestra la idea central de *docking*, fijar una región de estudio o caja en el sitio de unión, allí se posiciona el ligando para que este encuentre la conformación de mejor interacción. (Guedes et al., 2021). 32

Figura 9. Minimización de la energía de un sistema Una superficie de energía tridimensional que muestra los métodos de minimización, en donde los movimientos hacia abajo representa que el sistema se está moviendo en dirección del mínimo de energía más cercano.(Levitt, 2014) 33

Figura 10. Flujograma de la metodología capítulo 1. Resumen grafico de los pasos seguidos en la metodología para el análisis de mutaciones de HA VI H1N1pdm. 36

Figura 11. Árbol filogenético de máxima verosimilitud: Árbol de máxima verosimilitud de HA de los VI en Colombia con referencias de linajes Euroasiático (morado), Humano estacionario (naranja), porcino clásico (verde) y pandémico (azul) generado en IQtree con un un modelo sustitucional GTR+I+G y 100 *bootstraps*. Las ramas que pertenecen a secuencias del estudio se marcan de color rojo y las de referencia de color gris. El árbol se ilustra con el clado pandémico colapsado por su extensa longitud, pero su versión extendida se encuentra en el **anexo 1**. 40

Figura 12. Identificación y clasificación de mutaciones: En (A) se muestra en rectángulos el número de mutaciones que pertenece a cada categoría. En (B) se muestra frecuencia de las 50 mutaciones más frecuentes. Ambas gráficas están coloreadas de acuerdo a la categoría que pertenecen. 41

Figura 13. Árbol bayesiano en escala de tiempo: Filogenia del gen H1-VI de linaje H1N1pdm en Colombia en el tiempo, en donde se diferencian los grupos virales G1, G2 y G3. Adicionalmente se señala con una flecha una cepa viral lejana del clado de los grupos virales. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. los tiempos del ancestro común más reciente entre H1-IAV se muestran en los nodos de rama. El árbol completo se encuentra en el **anexo 3**. Tiempo en años hace referencia al periodo de aparición del ancestro común mas reciente. 43

Figura 14. Árbol bayesiano de G1 con mutaciones: Comparación de las relaciones filogenéticas de los virus y las sustituciones más frecuentes adquiridas en el tiempo para el grupo viral G1 formado por las HA de linaje H1N1pdm en Colombia. Los otros clados se muestran de forma colapsada. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. Las mutaciones se colorean de acuerdo su categoría descrita en la **tabla 2**. Las mutaciones que se introdujeron en un clado se resaltan en el árbol... 45

Figura 15. Árbol bayesiano de G2 y G3 con mutaciones: Comparación de las relaciones filogenéticas de los virus y las sustituciones más frecuentes adquiridas en el tiempo para los grupos virales G2 y G3 formado por las HA de linaje H1N1pdm en Colombia. El grupo G1 se muestra de forma colapsada. observadas de acuerdo con su nivel de significancia.

Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. Las mutaciones se colorean de acuerdo su categoría descrita en la **tabla 2**. 46

Figura 16. Flujograma de metodología capítulo 2. Resumen grafico de los pasos seguidos en la metodología de la modelación estructural consenso de HA VI H1N1pdm. 51

Figura 17. Gráficos de Ramachandran de las 5 estructuras consenso. Gráficos de Ramachandran con 4 regiones establecidas así: Regiones de color rojo (A, B y L) representan las más favorables. Regiones en amarillo brillante (a, b, l y p) que representan adicionales permitidas. Las regiones en amarillo opaco (~a, ~b, ~l y ~p) son generosamente permitidas y la región en blanco que corresponde a no permitidas. Cada gráfico se distribuye entre cada consenso de la siguiente manera: Cal09 (A), Mi15 (B), Bri18 (C), GM19 (D) y Syd22 (E). 57

Figura 18. Z-score de las proteínas consenso. El punto negro en las gráficas representa el Z-score de los modelos consenso: cCal09, cMi15, cBri18, cGM19 y cSyd22 contra el Z-score de las proteínas depositadas en el PDB. 58

Figura 19. Estructuras tridimensional de los consensos modelados y sus mutaciones. (A) Estructuras de los 5 consensos señalando la ubicación de las mutaciones presentes en el sitio de unión. (B) Lista de sustituciones en el sitio de unión respecto a la referencia (WT). 59

Figura 20. Medición de propiedades durante minimización y equilibrio: Propiedades termodinámicas del sistema en el tiempo. A) Energía de minimización del sistema, B) Densidad, C) Temperatura y D) Presión. 60

Figura 21. Métricas de dinámica molecular de las consensos modelados: Media) Desviación cuadrática media de la raíz (RMSD) y B) Radio de giro (Rg) del esqueleto de las proteínas cCal09, cMi15, cBri18, cGM19 y cSyd22 en el tiempo. 61

Figura 22. RMSF de los consensos modelados. raíz de la fluctuación media cuadrática (RMSF) de las proteínas cCal09, cMi15, cBri18, cGM19 y cSyd22. 62

Figura 23. Alineación del RBD de las estructuras modeladas. Comparación de las estructuras tridimensionales del sitio de unión al receptor de las proteínas de HA: cCal09 (verde), cMi15 (Morado), cBri18 (Azul), cGM19 (Amarillo) y cSyd22 (Rojo), después de 50ns de producción en dinámica molecular. 63

Figura 24. Distancia de las regiones del RBD. Distancias entre los aminoácidos T134, L192 y Q224 de las proteínas cCal09 (A), cMi15 (B), cBri18 (C), cGM19 (D) y cSyd22. Durante 50ns de producción de dinámica molecular. 64

Figura 25. Flujograma de metodología capítulo 3. Resumen grafico de los pasos seguidos en la metodología del Análisis unión HA y receptores celulares. 68

Figura 26. Diagrama de las interacciones de docking. Interacciones entre los receptores humano (A-C) y aviar (D-F) con las proteínas consenso cCal09, cBri18 y cSyd22. *Docking* realizado con Qvina2 y Diagrama visualizado con LigPlot. 73

- Figura 27. Medición de las propiedades en minimización y equilibrio:** Propiedades termodinámicas de los complejos ligando-proteína en el tiempo. A) Energía de minimización del sistema, B) Densidad, C) Temperatura y D) Presión..... 74
- Figura 28. Medición entre el sitio de unión de los consensos y los receptores celulares.** Medición de distancias entre el centro geométrico de los receptores celulares aviar y humano y el sitio de unión de las proteínas cCal09, cBri18 y cSyd22..... 75
- Figura 29. RMSD y poses del ligando a lo largo de la simulación:** RMSD de ligandos proteínas durante los 100ns de los complejos: A) cCal09/ α 2,6, C) cBri18/ α 2,6, E) cSyd22/ α 2,6 y G) cBri18/ α 2,3 y las poses en 0, 50 y 100ns de los mismos complejos respectivamente en: B), D), F) y H). 77
- Figura 30. Radio de giro:** Medición del radio de giro del esqueleto de las proteínas en complejo con A) el receptor α 2,6 y B) α 2,3 a lo largo de 100ns de producción en dinámica molecular. 78
- Figura 31. Gráficos de las interacciones.** Diagrama de interacciones bidimensionales receptores con las proteínas (A, D, G y J) realizado con LigPlot. Pose tridimensional de los ligandos en el sitio de unión de las proteínas (B, E, H y K). Distancia de 3 enlaces H de durante 100ns de producción (C, F, I y L). Para los complejos cCal09/ α 2,6(A-C), cBri18/ α 2,6 (D-F), cSyd22/ α 2,6 (G-I) y cBri18/ α 2,3 (J-L)..... 80
- Figura 32. Energías de unión de los complejos:** Medición de la energía de unión MM/GBSA: (A) durante los últimos 10ns (B) Promedio de los últimos 10ns. Entre los complejos cCal09/ α 2,6, cBri18/ α 2,6, cSyd22/ α 2,6, y cBri18/ α 2,3. Los cálculos se realizaron con el software *gmx_MMPBSA*. 81
- Figura 33. Descomposición de energía por residuos:** Energía de unión MMGBSA promedio por residuos pertenecientes al RBD de las proteínas cCal09, cBri18 y cSyd22 con el receptor humano α 2,6 en (A) y de cBri18 con el receptor aviar α 2,3 en (B). Los residuos que no presentaron valores de energía no se mostraron. Los cálculos se realizaron con el software *gmx_MMPBSA* durante los últimos 10ns de la producción en dinámica molecular. Las posiciones que presentan mutación en alguna de las proteínas se muestran con un símbolo (*) 83
- Figura 34. Energía de unión promedio a dos constantes dieléctricas.** MMGBSA de los últimos 10ns de las proteínas cCal09, cBri18 y cSyd22 en unión con los receptores humano y aviar con una constante dieléctrica interna de 1 en (A), de 2 en (B). 84

Lista de tablas

Tabla 1. Referencias usadas en los análisis filogenéticos	37
Tabla 2. Descripción de las categorías de mutaciones ofrecidas por Flusurver.	38
Tabla 3. Número de secuencias del estudio que se mapearon a las cepas vacunales recomendadas por el OMS.	54
Tabla 4. Mutaciones totales y únicas identificadas a partir de la referencia A/California/07/2009 de los consensos: cCal09, cMi15, cBri18, cGM19 y cSyd22 calculados en este estudio.....	55
Tabla 5. Aminoácidos con los que se forma interacción electrostática entre los receptores humano y aviar y las proteínas cCal09, cMi15, cBri18, cGM19 y cSyd22.....	72

Lista de abreviaturas

Abreviatura	Significado
VI	Virus de Influenza
SA α 2,6	Receptor Celular Humano
SA α 2,3	Receptor Celular Aviar
H1N1pdm	Linaje Pandémico
Rg	Radio de giro
RMSD	Root Mean Square Root
RMSF	Root Mean Square Fluctuation
MMGBSA	Molecular Mechanics Generalized Poisson Boltzmann Surface Area
HA	Hemaglutinina
NA	Neuramindasa
NP	Nucleoproteína
M	Proteína Matriz
NS1	Proteína No Estructural
NEP	Proteína estructural
PB1	Polimerasa 1
PB2	Polimerasa 2
HS	Humano estacionario
EA	Euro asiático
CS	Porcino Clásico

1. Introducción

1.1 Planteamiento del problema

El virus Influenza (VI) causa una infección altamente contagiosa que afecta el tracto respiratorio de humanos y distintas especies de animales. Esta enfermedad afecta alrededor del 5 % al 15 % de la población mundial cada año (Petrova & Russell, 2017). Se calcula que la infección por VI ha causado entre 50 y 100 millones de muertes a lo largo de la historia en distintas pandemias y epidemias, lo que la ha categorizado como una de las enfermedades más mortales (Fraser et al., 2011). La primera pandemia de influenza documentada fue en el año 1918, la cual cobró la vida de alrededor de 70 millones de personas y una de las últimas fue ocasionada por un nuevo virus en el 2009 (Michaelis et al., 2009). Actualmente causa de 3 a 5 millones de casos severos y desde 250,000 hasta 500,000 muertes anuales (Petrova & Russell, 2017), afectando principalmente a niños, adultos mayores de 65 años y personas inmunocomprometidas o que padecen enfermedades crónicas (Tafalla et al., 2016). Sin embargo, como se mencionó inicialmente, el virus infecta una alta variedad de especies huésped como las aves y los cerdos. En el caso del virus aviar, las estrategias de preparación y prevención incluyen el sacrificio de aves enfermas y la vacunación (Otte et al., 2008). Desde 2003, más de 400 millones de aves han sido sacrificadas en todo el mundo como resultado directo de brotes de influenza aviar (Boni et al., 2013). En los cerdos, la infección por VI es un problema común en las granjas, se estima que el 57% de las granjas en Estados Unidos son prevalentes a VI (Klement et al., 2017). La vacunación en cerdos, el cierre temporal, entre otras, son medidas requeridas para el control del VI en las granjas afectadas (Cador et al., 2017).

Los cerdos han jugado un papel epidemiológico fundamental en la adaptación y evolución de VI. Estos animales poseen características en su tracto respiratorio que permite la coinfección de diferentes cepas virales lo que favorece la aparición y emergencia de

nuevos virus. Todo esto representa un gran riesgo de salud y producción animal, generando grandes impactos a la salud a nivel mundial.

El VI se caracteriza por estar en constante evolución, la cual se encuentra mediada principalmente por mutaciones en su genoma. Estas mutaciones son responsables de modificar las características biológicas del virus y alterar mecanismos tales como: la transmisión entre especies, el escape al reconocimiento por anticuerpos neutralizantes o la resistencia a medicamentos antivirales (Petrova & Russell, 2017).

De forma general, hay disponible mucha información acerca del virus en el mundo, pero en el contexto latinoamericano y especialmente colombiano, los estudios de caracterización de estos virus en cerdos y humanos son muy escasos. En la década de los 70s se reportó por primera vez la existencia del VI en cerdos en Colombia (Hanssen et al., 1977). Luego se realizó detección del H1N1 pandémico (Consuelo Ramirez-Nieto et al., 2012; Karlsson et al., 2013), el cual ha circulado en población humana y de manera endémica en poblaciones porcinas. Sin embargo, a la actualidad no existen análisis extensos a nivel molecular y estructural que permitan entender como es el comportamiento y la evolución de los virus en el país, cuáles son las características epidemiológicas y biológicas particulares de estos virus y qué efectos podrían tener dichas características en la interacción de virus-hospedero, en la prevención y control de la enfermedad en el país y la región.

Por lo tanto, este trabajo busca investigar: ¿cómo la presencia de mutaciones en la proteína de superficie (HA) de los IV reportados en Colombia pueden alterar las propiedades estructurales y biológicas (afinidad por hospederos, transmisión, evasión inmune y patogenicidad) del virus?

1.2 Justificación

El VI se caracteriza por su naturaleza impredecible. Debido a su biología el VI se encuentra en continuo cambio. Una de las maneras en las que el virus evoluciona se conoce como cambio antigénico, la cual consiste en acumulación de mutaciones en las

proteínas virales superficiales que pueden modificar dos aspectos importantes: primero, la especificidad del virus a la hora de interactuar con la célula huésped, causando un aumento o disminución de la afinidad del virión y modificando su habilidad para infectar y transmitirse entre hospederos; y segundo, estas mutaciones pueden ocurrir en sitios clave involucrados en la unión del virus a fármacos o en la neutralización por los anticuerpos, afectando la eficacia de los antivirales, las vacunas o la respuesta inmune ante la infección (Webster & Govorkova, 2014). Otro mecanismo de evolución más drástico pero menos frecuente es llamado re-arreglo antigénico. Este se genera cuando diferentes cepas provenientes coinfectan un mismo hospedero, provocando un reordenamiento del material genético durante la replicación viral, dando como resultado la emergencia de un virus completamente nuevo o mutado (Sandbulte et al., 2015).

Dados los altos riesgos que representa la alta tasa de cambio del VI, se puede apreciar el gran reto que se tiene para la salud pública en Colombia y en el mundo, ya que el surgimiento de nuevas cepas con potencial pandémico es inesperado. Estas y otras barreras dificultan la prevención y control de la enfermedad tanto en humanos como en animales. Una de las maneras tradicionales en las que se enfrenta este problema es esperar hasta que ocurra el brote de enfermedad de una nueva cepa viral y una vez esté circulando entre la población, se toman las medidas para mitigar sus efectos, ya sea con el desarrollo de vacunas o de fármacos para su control. Sin embargo, todo esto requiere tiempos de hasta 12 meses o más (Rajapaksha et al., 2014), causando un gran desafío dado que en dicho periodo puede propagarse el virus y tener efectos devastadores.

El gran aumento en el uso de herramientas computacionales ha traído gran avance en la investigación y desarrollo de técnicas para entender el comportamiento de los virus y sus efectos. El VI no es una excepción, en los últimos años se han desarrollado innumerables estudios computacionales acerca de la transmisión, respuesta inmune, afinidad por hospederos, respuesta a fármacos, patogenicidad, entre otros (Bouvier & Palese, 2008). Es por esto que las herramientas bioinformáticas son de sumo beneficio, ya que con ayuda de estas se pueden realizar análisis en un menor tiempo y a menores costos, permitiendo así estudiar múltiples cepas que se distribuyen a lo largo de las localidades e incluso

analizar las nuevas que van emergiendo y así apoyar la toma de decisiones de forma oportuna.

En Colombia se carece de estudios computacionales que permitan investigar las características y propiedades biológicas y epidemiológicas de los VI circulantes reportados y por ende se hace evidente la necesidad de ampliar el conocimiento que se tiene acerca de estos virus, y así entender el impacto potencial que tienen para la salud pública.

1.3 Objetivos

1.3.1 Objetivo General

Determinar por análisis in silico si los cambios en el gen de la hemaglutinina afectan las propiedades biológicas de los virus influenza tipo A del linaje pandémico en Colombia.

1.3.2 Objetivos Específicos

1. Determinar el efecto potencial de las mutaciones de la hemaglutinina sobre las propiedades biológicas de los virus influenza A del linaje pandémico en Colombia.
2. Generar un modelo estructural 3D de la hemaglutinina que represente los virus influenza A del linaje pandémico en Colombia.
3. Evaluar si las mutaciones presentes en el dominio de unión al receptor de la hemaglutinina de los virus estudiados alteran la unión con las células del hospedero (ave y humano).

Cada objetivo se nombró con los títulos listados a continuación:

Objetivo 1: Análisis de mutaciones HA VI H1N1pdm

Objetivo 2: Modelación estructural consenso HA VI H1N1pdm

Objetivo 3: Análisis Unión HA y receptores celulares

2. Fundamentos teóricos

2.1 Biología del virus

2.1.1 El virus de influenza (VI)

En la biología es de suma importancia estudiar los virus, su biopatología, su forma, su transmisión, las consecuencias que este puede traer a la población y la manera de prevenirlos. Uno de los virus más relevantes actualmente es el VI. Este pertenece a la familia orthomyxovirus, y se han descrito 4 tipos: A, B, C y D (Bouvier & Palese, 2008). Los tipos A y C tienen la capacidad de afectar humanos y animales como los cerdos y las aves; mientras que el tipo B sólo afecta a humanos. El tipo D, recientemente descrito, solo afecta a animales (Javanian et al., 2021).

La estructura biológica de VI se caracteriza por ser de forma esférica o tubular con tamaños que van de 80 nm a 120 nm (Stanley, 1944) (ver **figura 1**). Regularmente, cuenta con 8 segmentos de RNA monocatenarios en sentido negativo, en donde se almacena su información genética. Estos segmentos codifican para 11 proteínas (HA, NA, NP, M1, M2, NS1, NEP, PA, PB1, PB1-F2, PB2) que están envueltas en una membrana lipídica y en una proteína M1 o proteína de la matriz, que se encarga de dar fuerza y rigidez a la envoltura lipídica.

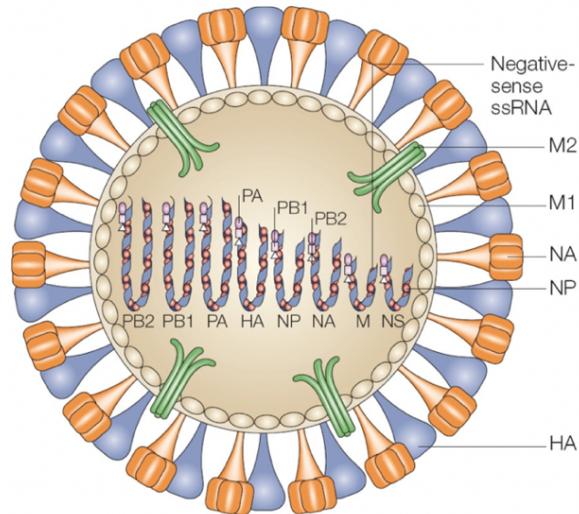


Figura 1. Esquema de la estructura y composición del Virus Influenza tipo A. Dos proteínas de superficie, HA y NA, y la proteína M2 están incrustadas en la envoltura viral, que se deriva de la membrana plasmática del huésped. El complejo de ribonucleoproteína (RNP) comprende un segmento de ARN viral asociado con la NP y tres polimerasas (PA, PB1 y PB2). La proteína de la matriz (M1) está asociada tanto a la RNP como a la envoltura viral. (Horimoto & Kawaoka, 2005).

Dentro del tipo de influenza A, existen muchos subtipos agrupados según las diferencias que poseen en las proteínas de superficie: hemaglutinina (HA) y neuraminidasa (NA). Existen 18 subtipos diferentes de HA (desde H1 hasta H18) y 11 subtipos de NA (desde N1 hasta N11) identificados a la actualidad (Petrova & Russell, 2017).

2.1.2 Hemaglutinina y sitio de unión al receptor celular

La HA es una glicoproteína que se encuentra en la superficie de la membrana del virus. Esta se ensambla en homotrimeros compuesta de monómeros que se divide en dos dominios HA1 y HA2. La región HA1 cuenta con el sitio de unión al receptor del ácido siálico del hospedero, que a su vez está rodeado por 5 sitios epítomos o sitios antigénicos (Sa, Sb, Ca1, Ca2 y Cb), los cuales son reconocidos por anticuerpos que se utilizan para inhibir la unión con el receptor, neutralizando la infectividad del virus (ver **figura 2**). Mientras que la región HA2 es la encargada de mediar la fusión de la envoltura del virus para ingresar a la membrana celular.

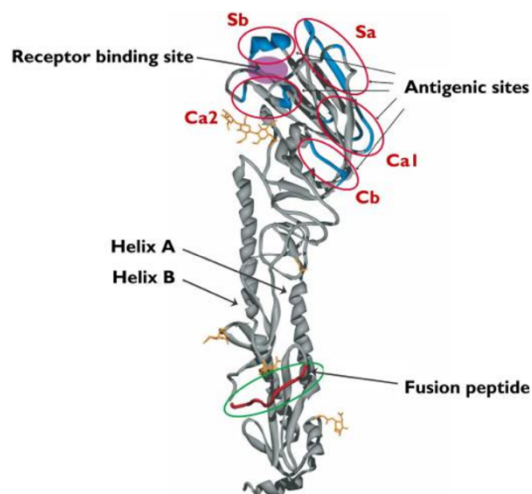


Figura 2. Estructura de HA del VI H1N1 pandémico del 1918 resaltando regiones más importantes. La cabeza contiene el sitio de unión al receptor de ácido siálico, que está rodeado por cinco sitios antigénicos (Sa, Sb, Ca1, Ca2 y Cb). El tallo comprende las hélices A y B y el péptido de fusión. (Bouvier & Palese, 2008)

La HA en su dominio HA1 cuenta con el sitio de unión al receptor (RBD, por sus siglas en inglés) (ver **figura 2**). Esta contiene 3 estructuras secundarias conservadas: dos loops en 130 y 220 y una hélice alfa en 190 (Byrd-Leotis et al., 2017; Tzarum et al., 2015). El RBS tiene la función de reconocer el receptor celular ubicado en la membrana de las células epiteliales del hospedero y que al entrar en interacción, permite tanto la entrada del virus a la célula huésped como la fusión de las membranas, en donde, con la ayuda de todo el material replicativo de la célula se logra la formación de nuevos viriones (Bouvier & Palese, 2008). En un estudio en donde se realizó la construcción de la HA pandémica H1N1 del 2009, se identificó que el RBS se encuentra en los residuos del 63 a 286 (residuos 55 a 271 en la numeración H3) (DuBois et al., 2011).

2.1.3 Cambio y evolución viral

Distintos factores ambientales como variaciones en la temperatura y PH, o errores en la replicación inducen cambios o mutaciones en el VI (Y. Wang et al., 2022), muchos de los cuales alteran las propiedades de las glicoproteínas HA y NA. Estos cambios se pueden dar a nivel de nucleótidos o a nivel de aminoácidos. La acumulación de estas mutaciones

en la superficie de las proteínas se denomina cambio antigénico y da como resultado una alteración parcial del estado antigénico del virus o en el caso de HA o cambios en la afinidad por los hospederos) (Koel et al., 2013), lo que genera que el virus evada las respuestas inmunitarias de los huéspedes. Esto ha generado epidemias de influenza estacional, provocando actualizaciones anuales o semestrales en las vacunas para que consideren eficazmente este cambio antigénico (Kim et al., 2018). También se puede presentar otro fenómeno denominado re-arreglo antigénico, que es un proceso se ve favorecido dada la naturaleza segmentada del material genético y que consiste en la combinación de dos cepas de distinto origen que provocan una coinfección una misma célula huésped provocando un intercambio en los segmentos de los virus, lo que da como resultado un nuevo virión que deja completamente desprotegido al huésped, variando sus características de transmisión y patogenicidad o incluso pudiendo dar como resultado la aparición de una cepa pandémica (Kim et al., 2018; Y. Wang et al., 2022).

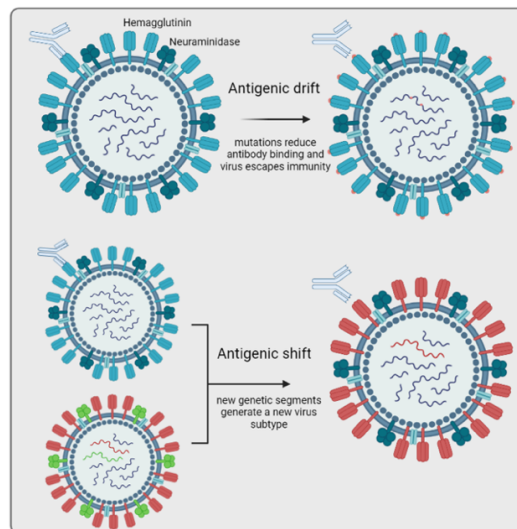


Figura 3. Fenómenos de cambio antigénico y re-arreglo antigénico en el VI. Los dos mecanismos principales que determinan la evasión inmune. La primera imagen se refiere al cambio antigénico que se da por las mutaciones que se acumulan en las proteínas de superficie (HA y NA), generando que el virus evada la respuesta inmune. El segundo mecanismo es el re-arreglo antigénico, que ocurre cuando segmentos del material genético

de diferentes cepas, dando como resultado un virion completamente nuevo. (Boyoglu-Barnum et al., 2021)

2.1.4 Diversidad genética

Todos estos Cambios que experimenta el VI, lo han llevado no solo a diversificarse en distintos subtipos como se mencionó anteriormente, sino también a caracterizarse dentro de los subtipos en distintos linajes. En particular, el subtipo H1 que es de especial interés en este estudio, es un subtipo dentro del cual los cerdos han desempeñado un papel fundamental en su diversificación dada la alta frecuencia de transmisión de otros huéspedes cepas VI humano a porcinos, la repetida introducción de estas cepas virales humanas ha determinado la evolución que se ha establecido, jugando un papel fundamental en la diversificación genética de los principales linajes que se han identificado (Anderson et al., 2016).

Existen distintos tipos de clasificaciones según la ruta evolutiva en cada continente como se muestra en la **figura 4**, dentro de las cuales se encuentran la clasificación europea, norteamericana y la global (Anderson et al., 2016). Por ejemplo la clasificación Europea definió 4 clados principales de H1 basado en el huésped o la introducción regional (Watson et al., 2015). El primero en datar fue ocasionado por una transmisión bidireccional entre los cerdos y humanos en la pandemia del 1918, lo que dio como resultado al linaje denominado “porcino clásico” según la clasificación europea (o 1A en la clasificación global) (Vincent et al., 2008). En la década de los 90’s surgió el segundo linaje llamado “triple reordenamiento” que ocurrió debido al reordenamiento que tuvieron los genes del porcino clásico con virus de H3N2 humano (Anderson et al., 2016), aumentando la diversificación de los genes de la HA del porcino a tres clados distintos llamados H1- α , H1- β y H1- γ , según la notación estadounidense, o 1A.1, 1A.2 y 1A.3.3.3, según la notación global (Rajao et al., 2018). Luego, en la década de los 2000, virus provenientes de H1N1 estacional humana se introdujeron a cerdos, dando origen al tercer linaje “*human seasonal-like H1*” (1B en notación global) del cual surgieron dos clados filogenéticos H1- δ 1 (1B.2.2) y H1- δ 2 (1B.2.1) (Chepkwony et al., 2021). Finalmente, en el 2009 la pandemia de H1N1 formo el cuarto linaje (Chepkwony et al., 2021), A pesar de que este virus es producto del reordenamiento de genes derivados de los linajes porcinos clásico, triple reordenado, entre

otros, su rápida propagación y establecimiento en las poblaciones generaron el clado denominado “H1N1pdm09” o “pdm” como se muestra en la **figura 4** (Gao et al., 2017).

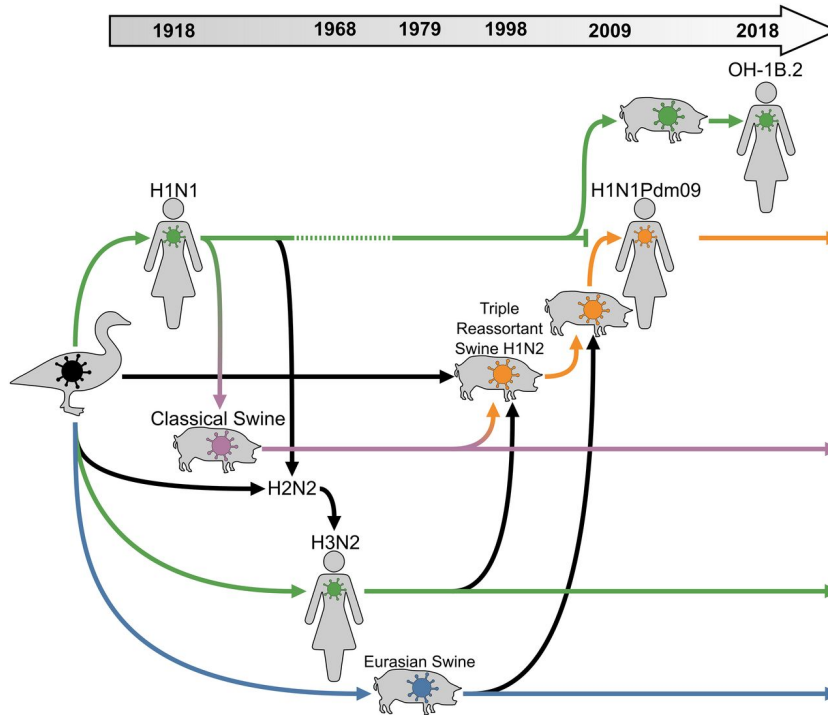


Figura 4. Relación genética de los linajes de virus humanos y porcinos. Las flechas verdes, naranjas, moradas y azules siguen la evolución de la hemaglutinina, mientras que las flechas negras siguen otros segmentos. La figura muestra la procedencia de distintos linajes que han afectado a humano y cerdos. (Rambo-Martin et al., 2020)

2.1.5 Transmisión viral

El virus de influenza se transmite principalmente a través de la diseminación de partículas respiratorias expulsadas durante la tos o los estornudos de huésped infectado. Cuando otro huésped susceptible inhala las gotas, el virus invade sus vías respiratorias infectando las células epiteliales en el tracto respiratorio (Taubenberger & Morens, 2008). La pared celular del tracto respiratorio se encuentra recubierta de polisacáridos que terminan en ácido siálico. El ácido siálico (SA) que se encuentra enlazado a una galactosa provee un sitio de reconocimiento para el virus. El SA puede unirse a la galactosa en distintas posiciones:

- Si esta unión se da entre el carbono 2 (C2 alfa) del SA y el carbono 3 (C3 alfa) de la galactosa (ver **figura 5**), se le denominará como SA α 2,3. Las cepas de influenza de ave reconocen principalmente esta versión del carbohidrato, el cual se encuentra principalmente en las vías respiratorias superiores de las aves. En humanos, estos carbohidratos se encuentran en las vías respiratorias inferiores. Esta es una de las razones por las cuales la cepa H5N1 a pesar de ser muy letal, no se transmite entre humanos.
- La unión entre el carbono 2 (C2 alfa) del SA y el carbono 6 (C6 alfa) de la galactosa (ver **figura 5**), genera un ácido siálico α 2,6, esta forma de receptor celular es reconocida por las cepas de influenza humana (H1, H2 Y H3). Además, esta forma de carbohidrato se encuentra principalmente en el tracto respiratorio de humanos. (Byrd-Leotis et al., 2017; Graaf Miranda & Fouchier Ron A, 2014; Rogers' And & D'souz~, 1989).

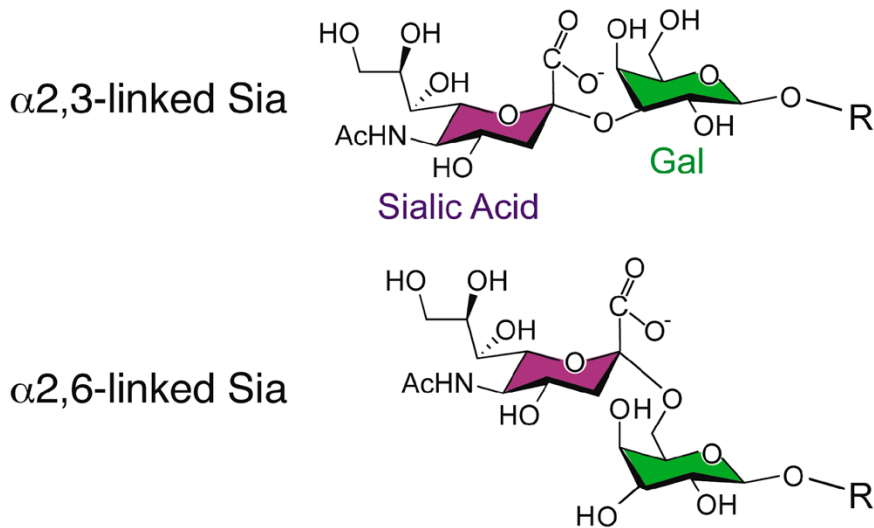


Figura 5. El enlace glucosídico entre el ácido siálico y el penúltimo azúcar de las cadenas. La estructura del ácido N -acetilneuramínico, un tipo de ácido siálico en sus conformaciones de enlace. En la conformación α 2,6 es el receptor celular encontrado principalmente en el tracto humano, la conformación α 2,3, es la forma de receptor que se encuentra principalmente en el tracto de las aves. Modificado de Byrd-Leotis y colaboradores (Byrd-Leotis et al., 2017)

Existe una barrera de infección que impide que una cepa de origen aviar se propague entre humanos, principalmente porque, como se mencionó anteriormente, las cepas de virus de influenza aviar utilizan preferiblemente receptores expresados en células de aves y los virus humanos utilizan preferiblemente receptores expresados en el tracto respiratorio humano. Sin embargo, las cepas de origen porcino son capaces de reconocer tanto el carbohidrato con la forma de 2,3-SA como 2,6-SA, provocando que los cerdos sean susceptibles tanto a cepas de origen humano o de ave, lo que convierte a los cerdos en un recipiente de mezcla para el reordenamiento de distintas cepas, siendo así un agente capaz de causar distintas pandemias humanas (Graaf Miranda & Fouchier Ron A, 2014).

2.1.6 Infección viral

Para que el VI genere todo el ciclo de infección es necesario que el virión que proviene de las gotas infectadas se acople a las moléculas de SA descritas anteriormente. Esta unión ocurre cuando el RBS de la HA reconoce el receptor celular y se acopla, generando una interacción efectiva. Luego la célula huésped absorbe el virión por endocitosis, allí se liberan los segmentos del material genético al citoplasma en una fusión mediada por HA (Dou et al., 2018); todo esto permite observar que tanto la unión del virus al receptor celular, como el ingreso al interior de la célula es mediado por la HA. Una vez se encuentre todo el RNA segmentado en el citoplasma, se inicia una serie de procesos (Dou et al., 2018) en donde todo el material genético del virus usa la maquinaria de traducción de la célula huésped para replicar, transcribir, sintetizar y ensamblar un nuevo virión que brota finalmente a la superficie. Allí nuevamente la HA se unirá al SA, manteniendo el virión unido a la superficie celular. Es en este punto del proceso donde NA cliva la unión del SA con las moléculas de azúcar subyacente, permitiendo la liberación de los viriones y la propagación a nuevas células (**figura 6**). Por lo que la NA ha sido objeto de varios estudios para el diseño de fármacos que inhiban su acción (McAuley et al., 2019) y por lo tanto mitiguen la propagación.

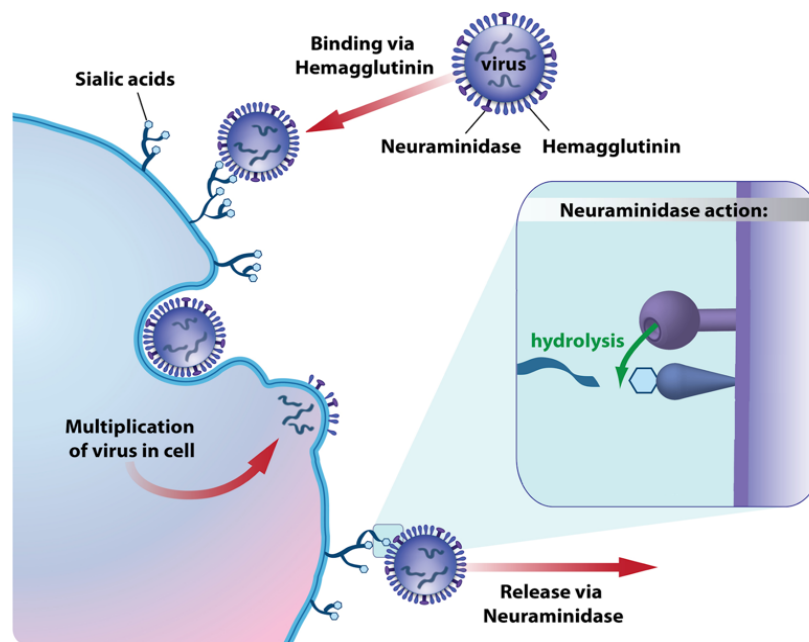


Figura 6. Proceso de infección de un virión del VI. Para infectar una célula huésped, el virión debe unirse efectivamente a la forma del SA que esté recubriendo la pared celular, luego, ocurre la endocitosis, en donde el ARN viral se libera y la célula usa el material de replicación. El nuevo virión formado brota a la superficie unido nuevamente al SA. Es aquí donde la NA es de hidrolizar la unión, permitiendo la liberación del virus. Modificado de (Wennekes Tom, 2013)

2.2 Técnicas computacionales para el estudio de estructuras proteicas y caracterización genética

2.2.1 Análisis filogenético

El análisis filogenético es una técnica ampliamente usada en la biología para estudiar la historia evolutiva de las entidades biológicas, la similitudes morfológicas o características moleculares que existen entre ellos mismos, esto da como resultado la ascendencia común en algún momento de la historia de estos organismos (Bawono & Heringa, 2014). Estos análisis se representan a través de árboles filogenéticos, que muestran las relaciones de parentesco y divergencia de especies a lo largo del tiempo. La construcción de estos árboles requiere de la información del organismo a estudiar a través de sus secuencias de

ADN, ARN o proteínas. Esta información es analizada con el objetivo de comparar las características entre sí y buscar patrones o de similitud o de diferenciación (Kapli et al., 2020). Esta técnica usa distintos métodos estadísticos para la construcción de árboles, dentro de las que se encuentra métodos de: distancias, máxima parsimonia, máxima verosimilitud, inferencia bayesiana, entre otros. Los cuales varían en eficiencia y limitaciones dependiendo de los datos proporcionados y el objetivo de estudio (Munjal et al., 2019).

En el ámbito de la virología esta herramienta ayuda a obtener un seguimiento de la diversidad y propagación de los virus lo que ha dado como resultado: identificación de nuevas variantes, monitoreo de pandemias y epidemias, evaluación de efectividad de vacunas y diseño de control y prevención. Adicionalmente puede aportar información acerca de propiedades biológicas de los virus como: transmisibilidad, virulencia, entre otros (Gorbalenya & Lauber, 2017).

En este estudio se hará uso de esta técnica para identificar las cepas colombianas que se caracterizan en el linaje H1N1pdm y para analizar la evolución de este linaje en el país.

2.2.2 Modelamiento por homología de estructuras proteicas

Existen métodos experimentales que permiten obtener la estructura tridimensional de una proteína como lo es la cristalografía de rayos X o la espectroscopia de RMN, entre otros (Vyas et al., 2012). Sin embargo, el cálculo de la estructura de proteínas a través de métodos computacionales representa un reto debido a que su esqueleto peptídico está conformado por ángulos de rotación ϕ y ψ en el enlace, estos tienen a su vez cierto rango de valores que puede tomar (ver **Figura 6**). A medida que se tiene un mayor número de aminoácidos en la cadena, el número de posibles conformaciones empieza a ser mucho mayor, la búsqueda de la estructura de una proteína basada únicamente en su secuencia se ve obstaculizada por el muy alto costo computacional asociado a la evaluación de todas las posibles combinaciones de ángulos, generando así una limitación considerable.

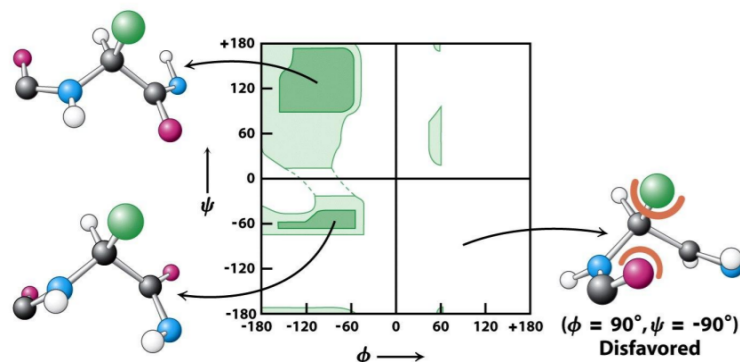


Figura 7. Diagrama de Ramachandran que muestra los valores de los ángulos ϕ y ψ . En verde se muestran regiones de los valores permitidos de los ángulos ϕ y ψ de la conformación de proteínas. Los gráficos de la izquierda son estructuras con sus ángulos en regiones permitidas y la figura de la derecha es desfavorecida debido a los choques estéricos. (Mark Berg Jeremy et al., 2006)

Por lo tanto, para llegar a una solución aproximada de la estructura, se utiliza una alternativa llamada modelamiento por homología, la cual consiste en usar una secuencia a la cual ya se le haya calculado su estructura tridimensional, usualmente a través de métodos experimentales y, si esta estructura es lo suficientemente parecida a la proteína de interés, se toma como molde y se extrapola la solución de la proteína de interés basado en la estructura conocida. Esto bajo la suposición de que las secuencias con relaciones evolutivas muy similares (y por lo tanto secuencias muy parecidas), tienen estructuras muy similares (Tramontano & Morea, 2003; Vyas et al., 2012). Este modelo es uno de los más utilizados ya que ha mostrado ser muy fiable. (Vyas et al., 2012).

En este estudio, se hará uso del modelamiento por homología para crear las estructuras tridimensionales de diversas cepas representativas de HA H1N1pmd en Colombia.

2.2.3 Docking en la predicción de interacciones en complejos biológicos

El acoplamiento molecular o *Docking* en inglés, es un método en el cual se estudia la interacción de un complejo biológico, prediciendo la conformación preferida de una molécula al estar unida con otra. A su vez, este permite la medición de fuerzas de unión entre las moléculas. Una de sus aplicaciones es el modelado del comportamiento de una molécula pequeña como un ligando en el sitio activo de una macromolécula (por lo general

una proteína, ver **figura 8**) que permite estudiar procesos bioquímicos fundamentales (Kitchen et al., 2004) evaluando las distintas direcciones, lugares, posiciones y conformación preferenciales del ligando con respecto a la macromolécula a través de funciones scoring. Este método compara el número de posiciones o disposiciones particulares que evalúa termodinámicamente el número de uniones favorables con el objetivo de buscar idealmente, el mínimo global de energía. El *docking* tuvo inicio en los años 80 (Kuntz et al., 1982), hoy en día ha llegado a convertirse en una de las técnicas más utilizadas para el diseño y descubrimiento de nuevos fármacos.

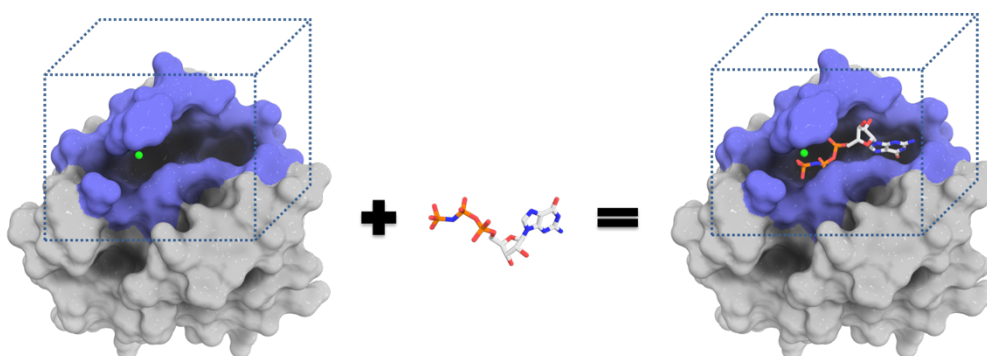


Figura 8. Gráfico del *docking* entre un ligando y un receptor. Este gráfico muestra la idea central de *docking*, fijar una región de estudio o caja en el sitio de unión, allí se posiciona el ligando para que este encuentre la conformación de mejor interacción. (Guedes et al., 2021).

El método de *docking* ha demostrado ser de bajo costo y computacionalmente muy eficiente, permitiendo evaluar una gran cantidad de moléculas en distintas posiciones en un bajo tiempo. Sin embargo, los cálculos de energías libres pueden carecer de precisión (Torres et al., 2019), por lo que se recomienda evaluar una mayor cantidad de posiciones y someter el resultado a otros estudios o análisis posteriores que permitan tener una mayor confiabilidad en el resultado.

El *docking* será usado en este estudio para predecir la conformación y las interacciones iniciales que forman las HA de VI H1N1pdm con los receptores celulares humano y aviar.

2.2.4 Dinámica Molecular para el estudio de proteínas y complejos proteicos

La dinámica molecular es un método que permite estudiar un sistema a través del tiempo para determinar su evolución, comportamiento y características a partir de las

interacciones o fuerzas que se dan entre los átomos. Una de las bondades principales de esta herramienta es que permite estudiar una gran cantidad de moléculas en movimiento, las cuales serían imposibles de calcular mediante expresiones analíticas. Además, ha roto con la barrera del estudio del comportamiento de sistemas rígidos, atributo por el cuál este método es ampliamente utilizado en sistemas biológicos para analizar cómo los cambios a nivel molecular, como las mutaciones, que generan cambios conformacionales y de plegamiento que por consiguiente, afecta sus procesos biológicos. (Hollingsworth & Dror, 2018)

En este método los sistemas de estudio se pueden describir por ecuaciones físicas como la segunda ley de Newton, la cual acelera sus átomos en dirección de la fuerza para llevar los átomos a un estado de menor energía.

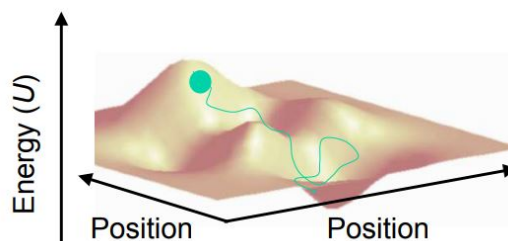


Figura 9. Minimización de la energía de un sistema Una superficie de energía tridimensional que muestra los métodos de minimización, los desplazamientos hacia abajo indican que el sistema se está moviendo en dirección al mínimo de energía más cercano. (Levitt, 2014)

Este método requiere el uso de campos de fuerzas, los cuales son funciones de energía potencial que dependen de funciones de forma y parámetros empíricos que determinan qué tan fuerte es la interacción entre los átomos. Dentro de estas interacciones se incluyen interacciones enlazadas y no enlazadas como una atracción o repulsión electrostática. En su forma más simple se puede representar como:

$$E_{total} = E_{bonded} + E_{non-bonded}$$

La precisión de los estudios realizados por dinámica molecular son dependientes de los campos de fuerza utilizados. Existe una amplia diversidad de campos de fuerza, en donde

las familias más utilizadas son: OPLS, AMBER, CHARMM Y GROMOS para la simulación de sistemas biológicos (Guvench & MacKerell, 2008).

Una simulación típica de dinámica molecular consta de una preparación del sistema, en donde se establecen las condiciones iniciales para realizar una minimización, seguida de una etapa de equilibrio, fijando variables termodinámicas como: energía E, temperatura T, presión P, volumen V, etc. A través del ensamble de sistemas: microcanónico (NVE), canónico (NVT), isotérmico-isobárico (NPT) o gran canónico (VT). Después, se realiza un paso de producción, fase en la cual se procede a formar la trayectoria del sistema evaluando el comportamiento en el tiempo (Schneider et al., 2008).

2.2.5 Energía libre de unión para el estudio interacciones biomoleculares

La energía libre es la cantidad de energía interna utilizable de un sistema para realizar un trabajo, esta constituye el impulso de todos los procesos moleculares como uniones o reacciones (E. Wang et al., 2019), motivo por el cual se utiliza para predecir cuantitativamente la afinidad de complejos biológicos como ligando-proteína (Brice & Dominy, 2011) y así determinar si la una unión es efectiva y estable. A partir de simulaciones de dinámica molecular, los cálculos de energía libre se consideran el método más preciso para predecir la energía de afinidad de un ligando a su diana (Hou et al., 2011a).

Dentro del cálculo de las energías libres se utilizan distintos métodos, muchos de estos requieren un alto costo computacional como lo es el método de energía libre de perturbación (FEP) o el método de integración termodinámica (TI) (Brice & Dominy, 2011; Patel & Kukol, 2021). Por lo tanto, con el objetivo de mitigar este problema, se han implementado métodos basados en la energía libre del estado final como lo es el método de Mecánica Molecular de Superficie de Área de Poisson-Boltzmann (Ivan et al., 2020), el cual reduce el costo computacional disminuyendo los estados intermedios, lo que resulta en un buen balance entre la eficiencia y precisión computacional, este método ha sido utilizado para evaluar las poses acopladas, determinar la estabilidad de las estructuras obtenidas y predecir las energías de afinidad, también se ha utilizados para analizar las contribuciones energéticas que cada residuo aporta (E. Wang et al., 2019).

3. Antecedentes

En términos del virus de influenza de origen pandémico en Colombia, se ha confirmado que este circula en el país desde el año 2009 (Karlsson et al., 2013). Además, un estudio publicado recientemente acerca de una caracterización genética de 10 virus circulando en cerdos en el periodo comprendido entre los años 2008 a 2015 (Osorio-Zambrano et al., 2022), reveló que dos cepas que datan en el año 2008 fueron identificadas en el linaje de porcino clásico (1A.1), mientras que los virus colectados después de la pandemia del 2009, fueron caracterizados en el linaje H1N1pdm09. Esto sugiere que desde la introducción del virus pandémico del 2009, el clado pandémico ha mantenido su dominancia en los ocho virus que circularon en los años posteriores. A pesar de esta caracterización, hasta la fecha, no se han reportado detecciones de las mutaciones que caracterizan a estos virus.

El virus pandémico del 2009 fue resuelto estructuralmente por primera vez durante ese mismo año (Soundararajan et al., 2009). Y, a su vez, este virus ha demostrado en repetidas ocasiones mostrar una preferencia por la unión con receptores humano SA α 2,6 sobre el receptor aviar SA α 2,3 (Xu et al., 2012; Bradley et al., 2011; Carbone et al., 2015; L. M. Chen et al., 2011; Childs et al., 2009; Maines et al., 2009; Yang et al., 2010). Además, estudios alrededor del mundo se han realizado para comprender el virus pandémico y las distintas variantes que van surgiendo (Chua & Chai, 2012; Chutinimitkul et al., 2010; Lee et al., 2011; Tse et al., 2011). Sin embargo, hasta la fecha no se han reportado estudios que caractericen estructuralmente los virus de influenza en Colombia o que permitan evaluar su preferencia por los hospederos.

La ausencia de este tipo de investigaciones dificulta el seguimiento de las condiciones del virus y la determinación de posibles factores locales que puedan estar influyendo en la dominancia del clado pandémico observado, llevándolo a destacarse sobre otros linajes virales, lo que destaca la importancia de continuar con la vigilancia y el estudio de su evolución para una mejor comprensión de su comportamiento y su posible impacto en la salud pública.

4. Capítulo 1: Análisis de mutaciones gen HA de VI de linaje H1N1pdm

En este capítulo, se hizo una investigación enfocada en la identificación y estudio de las principales mutaciones presentes en los genes HA de los virus de influenza del linaje H1N1pdm en Colombia. El objetivo es determinar el posible efecto de estas mutaciones en las propiedades biológicas del virus, buscando obtener un panorama más amplio acerca de cuáles de estas variantes han prevalecido y se han perpetuado en los virus circulantes.

4.1 Metodología

En la **figura 10** se muestra un resumen gráfico del paso a paso seguido en la metodología de este capítulo.

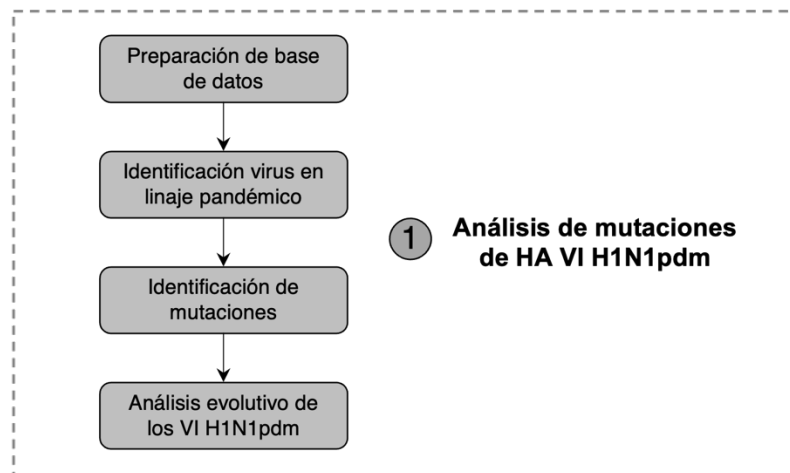


Figura 10. Flujograma de la metodología capítulo 1. Resumen gráfico de los pasos seguidos en la metodología para el análisis de mutaciones de HA VI H1N1pdm.

4.1.1 Preparación de la base de datos

Con el objetivo de crear una base de datos con todas las secuencias de HA de Colombia, se solicitaron las secuencias de HA disponibles en el laboratorio One Health de la

Universidad Nacional de Colombia y de la base de datos *Global Initiative on Sharing All Influenza Data* (GISAID, <https://www.gisaid.org/>), en donde se buscó con filtros como: “human”, “swine”, “H1” y “Colombia” de donde se descargaron todas las secuencias depositadas hasta el 23 de diciembre del 2022. Para depurar las secuencias redundantes, incompletas o de calidad insuficiente se hace uso del algoritmo *Sequence cleaner* (https://biopython.org/wiki/Sequence_Cleaner), en donde se establecieron límites del 80% para la longitud mínima de la secuencia y un máximo del 10% de indeterminaciones (N) admitidas en cada una de las secuencias, las secuencias que pasaron exitosamente este filtro las cuales se alinearon en MAFFT (Kato et al., 2019) y los extremos 5' y 3' fueron removidos en MEGA Software V.11 (Tamura et al., 2021).

4.1.2 Filogenia para la caracterización y análisis evolutivo

Posteriormente, para caracterizar las secuencias de HA obtenidas, se realizó un análisis filogenético de máxima verosimilitud (MLE) usando IQtree v2.2.0.3 (Minh et al., 2020). El mejor modelo sustitucional fue estimado con uso de la función “ModelFinder” de IQtree a través del Criterio de Información Bayesiano (BIC) y 1000 *bootstraps*, los cuales se etiquetan en cada rama. Finalmente, la edición y visualización del árbol se llevó a cabo en iTOL v5 (Letunic & Bork, 2021). Las referencias de los linajes incluidos en los estudios filogenéticos se muestran en la **tabla 1**, cada referencia se ingresó a BLAST (McGinnis & Madden, 2004) para encontrar secuencias de alta homología que permitieran darle soporte de rama a su filogenia.

Tabla 1. Referencias usadas en los análisis filogenéticos

Linaje/clado	ID NCBI	
Classical swine (CS)	Alpha (1A.1)	CY099119
	Beta (1A.2)	CY157999
	Gamma (1A.3)	CY158217
	Pandemic (1A.3.3.2)	NC_026433
Human seasonal (HS)	HS	AHG96683

Eurasian swine-avian like (EA)	EA	AJ176067
-----------------------------------	----	----------

Además, se realizó un análisis bayesiano en escala de tiempo para estimar los eventos de transmisión asociados a los virus de linaje pandémicos detectados. El método MCMC (*Markov Chain Monte Carlo*) se usó en Beast (Drummond et al., 2012) y Beagle (Ayres et al., 2012) con 100 millones de generaciones. Se seleccionó el modelo de sustitución según el BIC (*Bayesian Information Criterium*) encontrado en el apartado anterior. Se calcularon los parámetros de convergencia usando tracer v1.7 (Rambaut et al., 2018) de manera que los que recibieron valores de tamaños de muestra efectivos superiores a 200 se consideraron suficientemente muestreados (Kuhner, 2009). Con la ayuda de TreeAnnotator v1.8.4 se generó un árbol de máxima credibilidad (MCC). Los árboles resultantes se visualizaron y editaron utilizando FigTree V 1.4.3.

4.1.3 Detección de mutaciones

Luego, cada una de las secuencias identificadas en el linaje pandémico fue traducida a su cadena de aminoácidos con ayuda del software MEGA (Tamura et al., 2021) en el marco de lectura estándar. Adicionalmente, se usó FluSurver (<https://flusurver.bii.a-star.edu.sg>) para determinar las mutaciones y el nivel de riesgo de las secuencias y de cada consenso. Dicha significancia viene representada por un color distinto y, a su vez, determina por los números del 0 al 3, donde 3 es la más y 0 la menos significativa. Con el objetivo de identificar dicha representación de las mutaciones a lo largo del estudio, se nombrará cada tipo de mutación con categorías nombradas de la A a la F. La correspondencia de cada significancia se resume en la **tabla 2**.

Tabla 2. Descripción de las categorías de mutaciones ofrecidas por Flusurver.

Categoría	Significancia	Observaciones
A	3	Conocida por alterar la virulencia, causar resistencia a los medicamentos.
B	2	Crea o elimina un sitio potencial de glicosilación.
C	2	Ocurre en sitio que se sabe que está implicado en la unión de un fármaco, altera la especificidad del huésped.
D	1	Que ocurre en algún sitio de interacción.

E	1	Común
E	0	Sin efectos conocidos

4.2 Resultados

4.2.1 Secuencias de H1 en Colombia

Se obtuvieron 298 secuencias de H1, de las cuales 160 provienen del laboratorio One Health de la Universidad Nacional de Colombia y 138 se descargaron de la base de datos pública GISAID depositadas hasta el 23 de diciembre del 2022 en Colombia. Estas fueron posteriormente depuradas, dando como resultado 257 secuencias que pasaron exitosamente el filtro de calidad

4.2.2 Identificación de mutaciones

Dado que no existe una identificación clara del linaje de estos virus, se llevó a cabo un análisis de las relaciones filogenéticas de la HA de los virus objeto de estudio. Para ello, se construyó un árbol de máxima verosimilitud con 257 secuencias de referencia que permitieron identificar los linajes de estos virus. Las cepas de referencia utilizadas en el análisis incluyeron: EA, HU y CS, en este último se incluyeron los clados 1A.1, 1A.2, 1A.3 y el 1A.3.3.2 (clado pandémico). Se encontró que, al comparar las secuencias del estudio con respecto a las referencias, una de ellas no se agrupó en ninguno de los clústeres, cuatro de estas fueron agrupadas en el linaje HU, y las secuencias restantes (n=252) tuvieron una relación cercana al clado pandémico (1A.3.3.2), como se muestra en la **figura 11**.

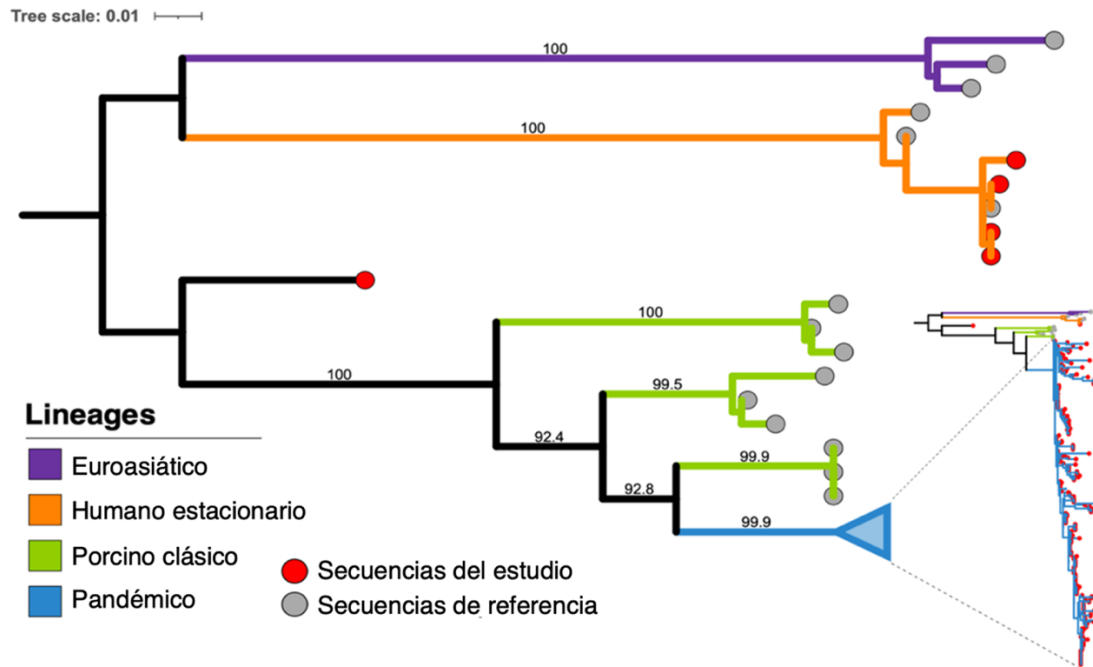


Figura 11. Árbol filogenético de máxima verosimilitud: Árbol de máxima verosimilitud de HA de los VI en Colombia con referencias de linajes Euroasiático (morado), Humano estacionario (naranja), porcino clásico (verde) y pandémico (azul) generado en IQtree con un modelo sustitucional GTR+I+G y 100 *bootstraps*. Las ramas que pertenecen a secuencias del estudio se marcan de color rojo y las de referencia de color gris. El árbol se ilustra con el clado pandémico colapsado por su extensa longitud, pero su versión extendida se encuentra en el **anexo 1**.

Una vez clasificados los virus se identificaron los que pertenecen al linaje pandémico, y se detectaron las mutaciones de dichos virus tomando como referencia la cepa pandémica A/California/07/2009(H1N1). Esta detección y categorización se realizó en el servidor en línea Flusurver, el cual sigue la clasificación dada en la **tabla 2**.

Inicialmente, la clasificación y frecuencia de las mutaciones observadas en los virus analizados en este trabajo se resumen en la **figura 12 (A)**. Se identificaron 6 categorías de acuerdo con nivel de significancia de las mutaciones. La mayoría (63.2%, 2158/3415) de las mutaciones identificadas fueron clasificadas en la categoría D, seguida de la categoría E (17.2%, 588/3415), y la categoría C (13.0%, 445/3415). Sin embargo, se observó que las categorías A (0.2%, 8/3415), F (2.9%, 102/3415) y B (3.3%, 114/3415) respectivamente, tuvieron la menor cantidad de mutaciones observadas.

Se observó que las tres mutaciones más frecuentes (P84S, I322V y S204T) pertenecieron a la categoría E, seguido de la mutación E375K perteneciente a la categoría C. También, para la categoría D, ocho mutaciones fueron las más frecuentes, siendo la mutación D98N la más frecuente de estas. Para la categoría B la mutación S163N fue la más frecuente, mientras que para la categoría A fue D223G. En la **figura 12 (B)** se resume la frecuencia de las primeras 50 mutaciones observadas para cada categoría. Adicionalmente, en el material suplementario (**anexo 2**) se provee una figura con la frecuencia de todas las mutaciones observadas.

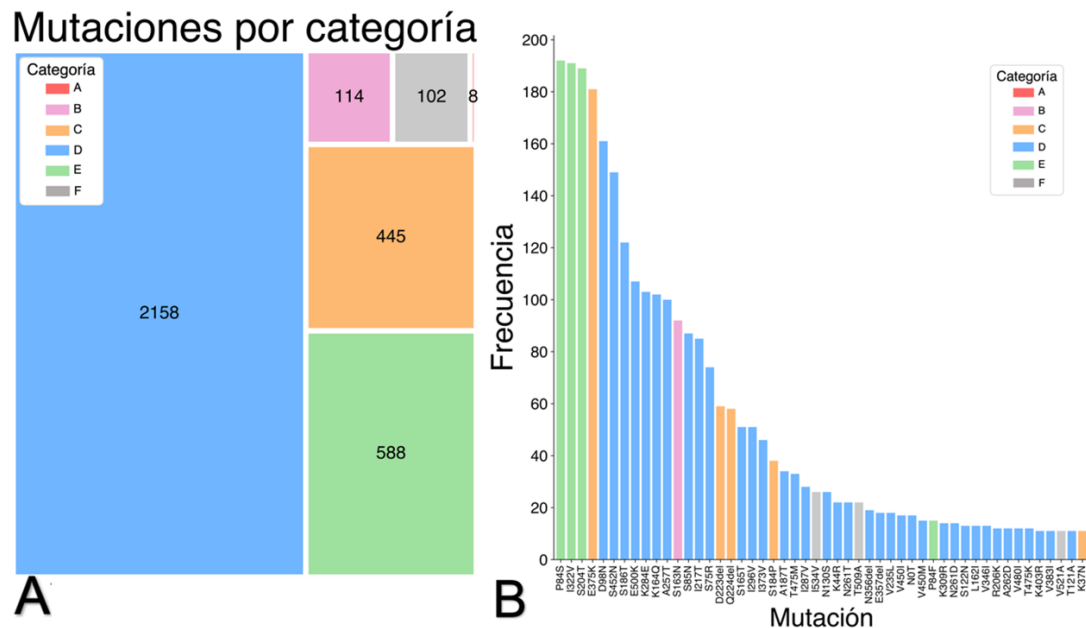


Figura 12. Identificación y clasificación de mutaciones: En (A) se muestra en rectángulos el número de mutaciones que pertenece a cada categoría. En (B) se muestra frecuencia de las 50 mutaciones más frecuentes. Ambas gráficas están coloreadas de acuerdo a la categoría que pertenecen.

4.2.3 Análisis evolutivo de los VI H1N1pdm

Luego de haber identificado las mutaciones más frecuentes en el estudio y su clasificación por categorías, se procedió a hacer un análisis bayesiano de las relaciones filogenéticas de estos virus como se muestra en la **figura 13**, y luego se comparó este patrón de agrupamiento con respecto a las mutaciones observadas. El análisis realizado permitió identificar un patrón de agrupamiento en los 252 virus analizados, como sigue: 167 virus datados desde el 2009 hasta el 2022, se agruparon creando el clado denominado G1; 62 virus datados entre el 2013 y 2017 conformaron el grupo viral G2; y 12 virus datados entre los años 2015 y 2017, se ubicaron en el grupo viral G3. Además, se identificó que el virus M8110_2016 (destacado con una flecha en la **figura 13**) estaba filogenéticamente más distante de los otros clados que conformaron el árbol.

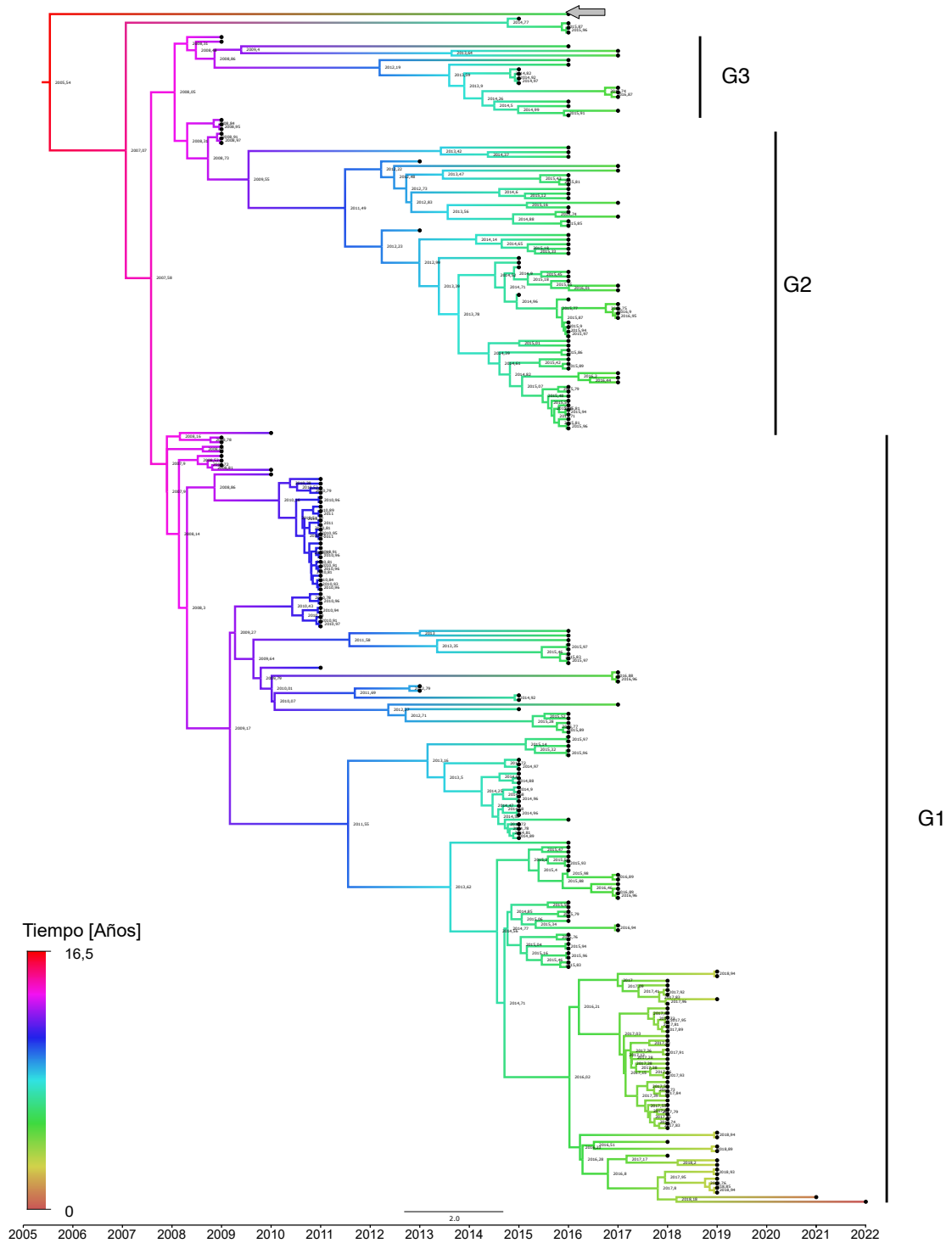


Figura 13. Árbol bayesiano en escala de tiempo: Filogenia del gen H1-VI de linaje H1N1pdm en Colombia en el tiempo, en donde se diferencian los grupos virales G1, G2 y

G3. Adicionalmente se señala con una flecha una cepa viral lejana del clado de los grupos virales. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. los tiempos del ancestro común más reciente entre H1-IAV se muestran en los nodos de rama. El árbol completo se encuentra en el **anexo 3**. Tiempo en años hace referencia al periodo de aparición del ancestro común más reciente.

Luego, las sustituciones de cada virus en cada uno de los grupos filogenéticos identificados fueron visualmente graficadas (**figura 14**), para explorar si existía alguna tendencia de agrupamiento de los VI H1N1pdm en Colombia. Se observó que al parecer hay una relación entre cada grupo filogenético (G1, G2, G3) con la presencia de ciertas mutaciones que no se observaron en los otros grupos. Por ejemplo, como se muestra en la **figura 13**, las mutaciones I322V, P84S, S204T y E375K estuvieron presentes en casi todos los virus pertenecientes al grupo viral G1, y que del análisis bayesiano se demostró que estas mutaciones tuvieron su origen alrededor del año 2008 y que aún se encuentran presentes en cepas que circularon recientemente.

De esta comparación se encontró también que a partir del año 2008 hubo una divergencia de estos virus creando dos clústeres, uno de ellos posiblemente asociado a la presencia de ciertas mutaciones (K44R, N130S y A187T), sin embargo, muchos de estos virus solamente circularon en el año 2011 pero no fueron reportados en años posteriores. Por otra parte, el otro clúster mostró la presencia de las sustituciones S186T y S452N, que se mantuvieron hasta recientemente. Adicionalmente, en el año 2011, se produjo la aparición de las sustituciones K164Q, K284E, A356T y E500K que nuevamente aportaron al parecer a la divergencia de estos virus. Posteriormente en el año 2013, aparecieron las sustituciones S85N, S163N y I217T, y en el año 2016 las sustituciones S75R, S165T e I296V.

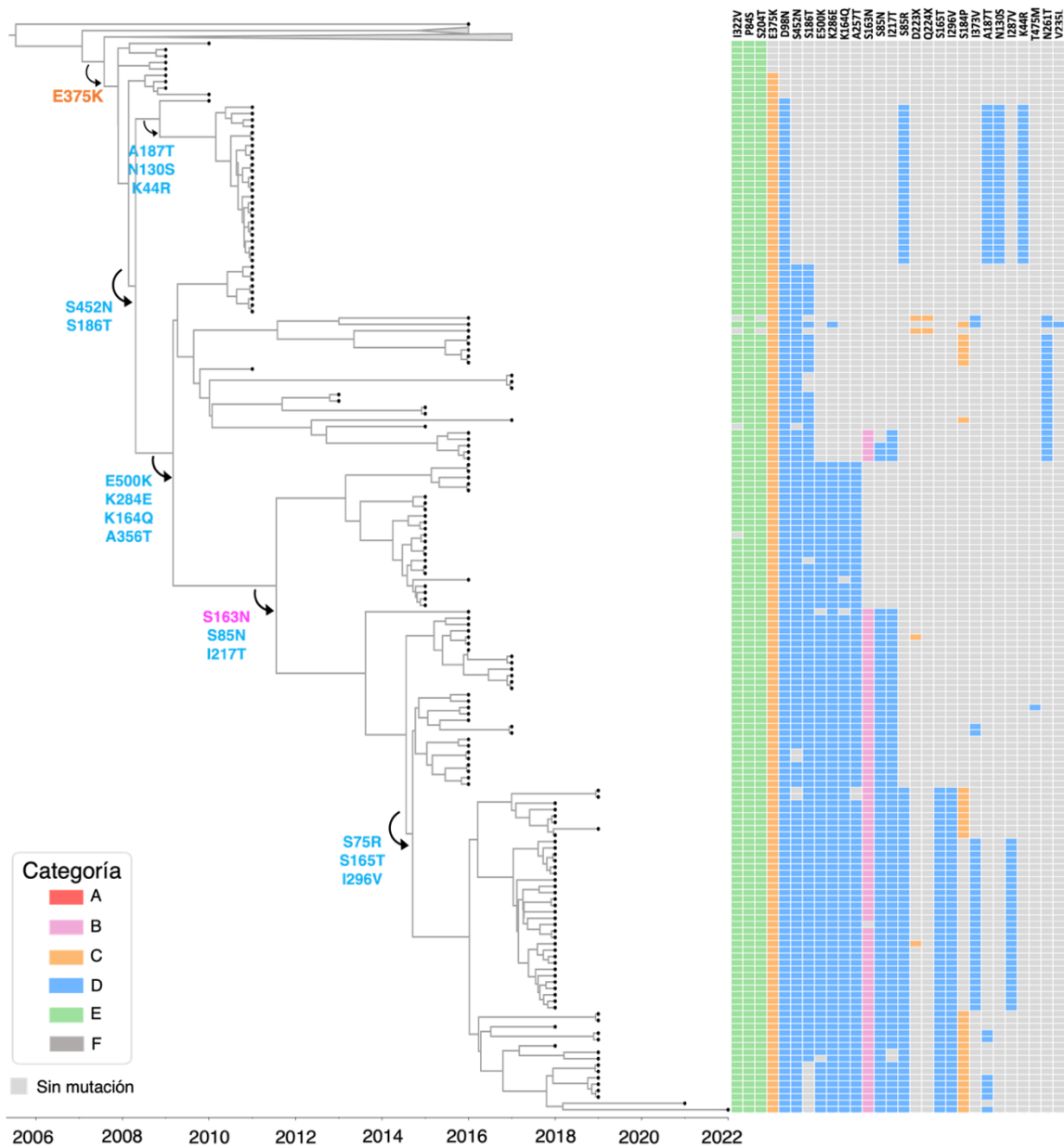


Figura 14. Árbol bayesiano de G1 con mutaciones: Comparación de las relaciones filogenéticas de los virus y las sustituciones más frecuentes adquiridas en el tiempo para el grupo viral G1 formado por las HA de linaje H1N1pdm en Colombia. Los otros clados se muestran de forma colapsada. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. Las mutaciones se colorean de acuerdo su categoría descrita en la **tabla 2**. Las mutaciones que se introdujeron en un clado se resaltan en el árbol.

El mismo análisis realizado para el grupo viral G1 ilustrado en la **figura 14**, se realizó para los grupos virales G2 y G3 como se muestra en la **figura 15**. Se identificó que las sustituciones D223X y Q224X, estuvieron presentes en casi todos los HA del grupo viral G2. Mientras que las sustituciones P84F, T375M, A262D y K309R se presentaron en los virus del grupo G3. En el material suplementario se provee la ilustración de la comparación entre el árbol bayesiano con todas las mutaciones identificadas en este estudio (**anexo 4**).

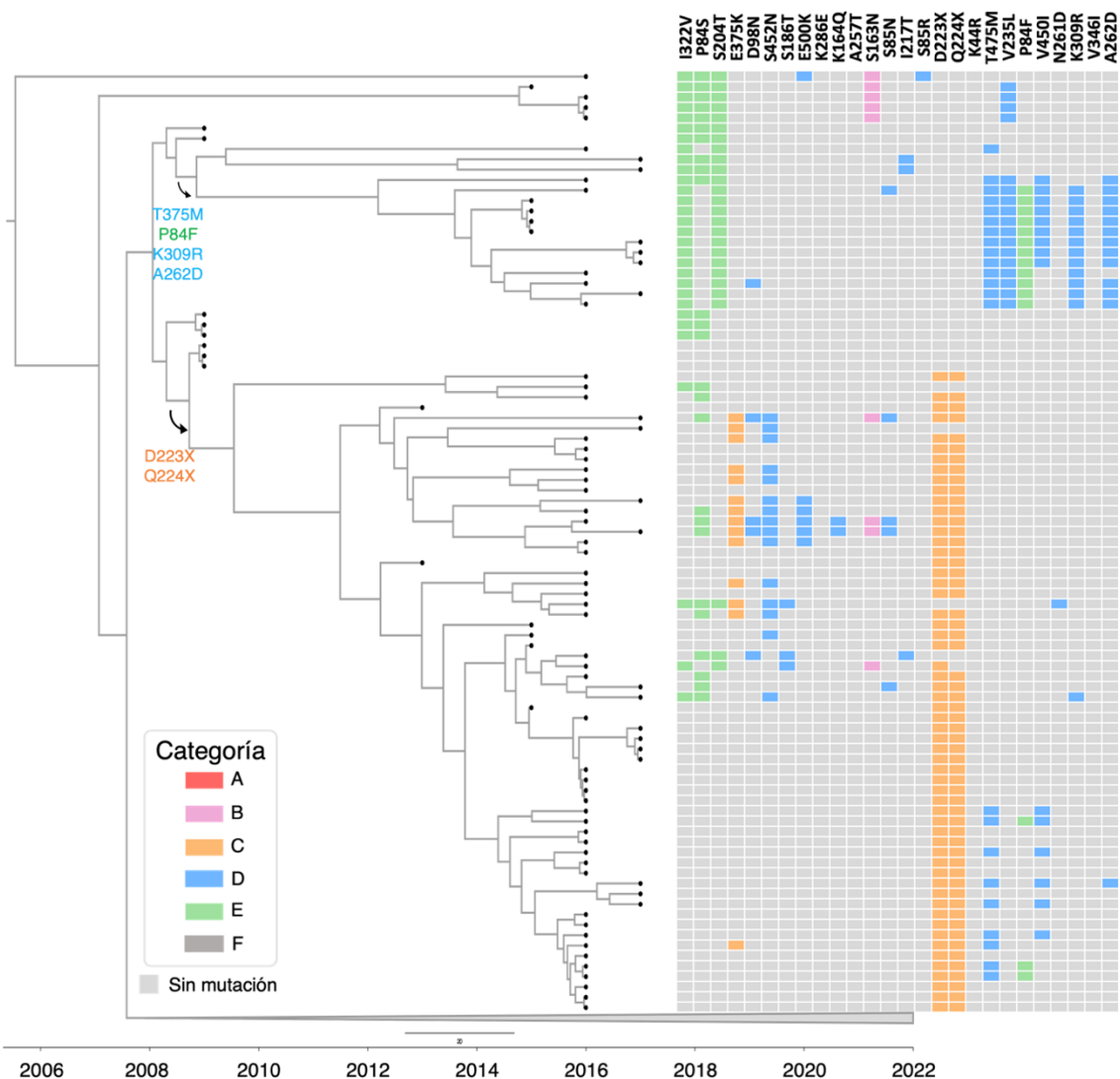


Figura 15. Árbol bayesiano de G2 y G3 con mutaciones: Comparación de las relaciones filogenéticas de los virus y las sustituciones más frecuentes adquiridas en el tiempo para los grupos virales G2 y G3 formado por las HA de linaje H1N1pdm en Colombia. El grupo G1 se muestra de forma colapsada. observadas de acuerdo con su nivel de significancia. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el

método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. Las mutaciones se colorean de acuerdo su categoría descrita en la **tabla 2**.

Finalmente, para el virus que no quedó agrupado en ninguno de los tres grupos virales establecidos en la **figura 13** (identificada con una flecha) se observó que el ancestro más cercano de esta cepa se originó en el año 2005 y que este virus presentó mutaciones únicas que no se encontraron en ninguna de los otros virus estudiados. Entre estas mutaciones se identificaron las siguientes: Q194I (categoría A), K212N (categoría B), K495N (categoría B), K164I (categoría D), I287M (categoría D) y D502E (categoría D).

4.3 Discusión

Por años el VI ha representado un problema de salud pública para el mundo, y se sabe que este virus tiene una alta capacidad de mutar, y, por tanto, esto puede dificultar las acciones de prevención y control. Por ende, el estudio y caracterización de las principales mutaciones es esencial. Dado que dicha información es completamente desconocida en Colombia, se realizó un análisis de las mutaciones en los VI de linaje H1N1pdm reportados desde 2008 hasta 2022.

Este estudio reveló que la mayoría de las mutaciones fueron identificadas en una categoría D. De esta categoría, la mutación D98N fue la que más se encontró, en el 59% de los virus estudiados, similar a lo notificado en flusurver, donde esta mutación se encuentra en el 51% de los virus reportados en este servidor. En la literatura no se han reportado efectos asociados a esta mutación. Solo se reporta su aparición en altas proporciones en diferentes estudios (Al Khatib et al., 2018c; Han et al., 2016), hecho que es coherente con la alta frecuencia de ocurrencia en este estudio.

La siguiente categoría de mutaciones más frecuente fue la categoría E (588/3415) (ver **figura 12 (A)**), en esta categoría se encontraron 3 mutaciones (P84S, I322V, y S204T), que son las mutaciones más frecuentes de todo el estudio (192/3415, 191/3415 y 189/3415) como se observa en la **figura 12 (B)**. Al hacer una búsqueda en la literatura de estas mutaciones, se encontró que estas mutaciones se encuentran en más del 90% de

las secuencias en la base de datos de flusurver y no se encontraron efectos relevantes reportados, aparte de ser mutaciones dominantes en las poblaciones estudiadas (Han et al., 2016; Koul et al., 2011; Ramos et al., 2013). Lo que tiene coherencia con la alta frecuencia encontrada en el estudio.

Luego, 445 de 3415 mutaciones fueron identificadas en la categoría C, en donde E375K es la mutación más frecuente de este grupo (181/588). Esta mutación está posicionada en la subunidad HA2, en una hélice cercana al péptido de fusión. Otras sustituciones en esta posición han determinado que E375K puede alterar procesos relacionados con la membrana de fusión, como lo es la reducción en el límite del pH necesario para la fusión, factor que afecta la infectividad y transmisión viral. Los virus de influenza humano tienen un límite de pH para la fusión más bajo que el aviar. Razón por la que los virus con esta mutación se adaptaron muy bien en humanos (Al Khatib et al., 2018a; Cotter et al., 2014).

La categoría B presentó 114 mutaciones, en donde la más frecuente fue S163N (92/114) y se encuentra en el 33% de las cepas en flusurver. Estas mutaciones se encuentran localizadas en la región HA1, más específicamente en el sitio antigénico Sa (Al Khatib et al., 2018c). Además, se ha reportado que esta mutación genera un sitio de glicosilación que produce características propias de evasión de respuesta inmune, lo cual ayuda a proteger el epítipo Sa y evitar la inmunidad preexistente (Al Khatib et al., 2018b; Guldemir et al., 2019).

La categoría F corresponde a una categoría de significancia nula, por lo que no se analizará. Y finalmente, en la categoría A, que corresponde a la de mayor significancia. Solo se identificaron 8 mutaciones distribuidas en las sustituciones en D223G, Q224R y Q194R. En donde D223G y Q224R apenas se encuentran en el 1.1 y 0,7% de las secuencias. Estas mutaciones se encuentran ubicadas en el loop 220 del dominio de unión al receptor. Y son reconocidas por causar cambios en la especificidad del receptor, provocando que cepas que únicamente se unían a SA α 2,6, pasen a tener especificidad cruzada α 2,3/ α 2,6 (Z. Chen et al., 2010; Matrosovich et al., 2000). Adicionalmente, estas mutaciones están relacionadas con baja tasa de transmisión humano a humano (Yasugi et al., 2012) siendo consecuente con el número de veces que se identificó en las secuencias de este estudio. Por otro lado, Q194R es una mutación que solo ha reportado 1 vez en la literatura según flusurver, este reporte corresponde a un estudio que se le hizo

a la cepa aviar H5N1, en este estudio estas variantes se encontraron exclusivamente en aislados provenientes de humanos y se ha observado que Q194R en combinación con las mutaciones Q226L y G228S generó pérdida importante en la unión a $\alpha 2,3$, pero un fuerte incremento en la unión a $\alpha 2,6$ en H5N1 (L. M. Chen et al., 2012).

Por otro lado, el análisis bayesiano mostró que las cepas se agruparon en tres grupos virales (G1, G2 y G3) como se muestra en la **figura 13**, G1 y G2 son los grupos que representan la mayoría de cepas, con 67% y 25% de los virus correspondientemente.

En la **figura 15** se observó que el grupo viral G2 se encuentra conformado por cepas que comparten en su mayoría dos mutaciones (D223X y G224X) de nivel de significancia medio (ya que están ubicadas localizadas en el *loop* 220 del *RDB*), lo que sugiere que estas mutaciones contribuyeron al establecimiento de este grupo viral. Al realizar una búsqueda en la literatura acerca de estas mutaciones, se encuentra que no se han reportado efectos asociados y, según flusurver, estas no han sido reportadas en otros lugares del mundo, lo que sugiere que estas mutaciones son propias de Colombia en el grupo G2. Adicionalmente, el análisis bayesiano encontró que estas mutaciones pudieron haber tenido surgimiento en Colombia a mediados del año 2008, con la mayoría de las cepas registradas en el 2016 y la última cepa que lo reporta data de 2017, lo que podría sugerir que estas mutaciones no aportaron al fitness.

Por otro lado, como se observa en la **figura 14**, el grupo viral G1 se encuentra conformado por cepas que comparten las mutaciones I322V, P84S y S204T, y por mutaciones como: E375K, K44R, N130S, A187T, S186T, S452N, K164Q, K284E, A356T, E500K, S85N, S163N, I217T S75R, S165T y I296V que se fueron introduciendo a lo largo de los años hasta las cepas que circulan en fechas actuales. Casi todas estas mutaciones pertenecen a las categorías D y E, que corresponde a un nivel de significancia bajo, a excepción de E375K y S163N, que pertenecen a las categorías C y B, correspondientemente, que son categorías de nivel de significancia medio.

En síntesis, en este estudio se logró clasificar las mutaciones acordes a su significancia y acorde a la frecuencia y prevalencia que ha tenido en el país. Esto permitió dividir las de la siguiente manera:

Primero, mutaciones de alta frecuencia y prevalencia, pero de baja significancia. En este grupo se encuentran: 322V, P84S, S204T, N130S, A187T, S186T, S452N, K164Q, K284E, A356T, E500K, S85N, I217T S75R, S165T y I296V. Estas mutaciones se consolidaron en las poblaciones con el transcurso del tiempo, permaneciendo presentes en prácticamente todas las cepas desde su introducción. Este patrón sugiere que dichas mutaciones han conferido una ventaja en el fitness y han desempeñado un papel crucial en el establecimiento de estos virus en las poblaciones actuales. Y acorde a sus categorías, se clasifican como de baja significancia, lo que está relacionado con mutaciones comunes o que ocurren en algún sitio de interacción, pero sin ningún efecto de alto riesgo, como se encontró en la literatura para algunas de ellas.

Segundo, mutaciones de alta significancia, pero baja frecuencia. Dentro de estas se encuentran: D223G, Q224R y Q194R, las cuales según lo reportado en la literatura tienen capacidad de alterar la virulencia o el tropismo, sin embargo, estas mutaciones no le han aportado fitness al virus para lograr establecerse en la población.

Finalmente, mutaciones de alta frecuencia y nivel de significancia medio. En este grupo se encuentran: E375K, S163N, D223X y G224X. Sin embargo, D223X y G224X solo estuvieron presentes hasta 2017, mientras que E375K y S163N son mutaciones persistentes que se introdujeron en 2008 y 2013, respectivamente, y continúan en las cepas actuales. Además, acorde a encontrado en la literatura, E375K aumenta la virulencia, lo que puede incrementar la eficiencia de la infección en hospederos, mientras que S163N ayuda a evadir la respuesta inmune, lo que puede llevar a cargas virales más altas y el aumento de tasas de transmisión. Estas características podrían explicar cómo estas mutaciones han contribuido al fitness del virus para establecerse en la población.

5. Capítulo 2: Modelación estructural consenso HA VI H1N1pdm

En este capítulo, se pretende obtener secuencias que puedan representar la diversidad de la mayoría de las HA de los virus de linaje pandémico presentes en Colombia, que tengan consigo las mutaciones más predominantes. El objetivo es modelar la estructura estas proteínas y determinar si la presencia de dichas mutaciones puede generar cambios significativos en su estructura o conformación, y en consecuencia, impactar en las propiedades biológicas de los virus.

5.1 Metodología

Este capítulo toma los virus que fueron caracterizados en el linaje pandémico del capítulo anterior (**capítulo 1, sección 3.2.2**) y se incluyen en el paso a paso desarrollado en la metodología, en la **figura 16** se muestra un resumen gráfico.

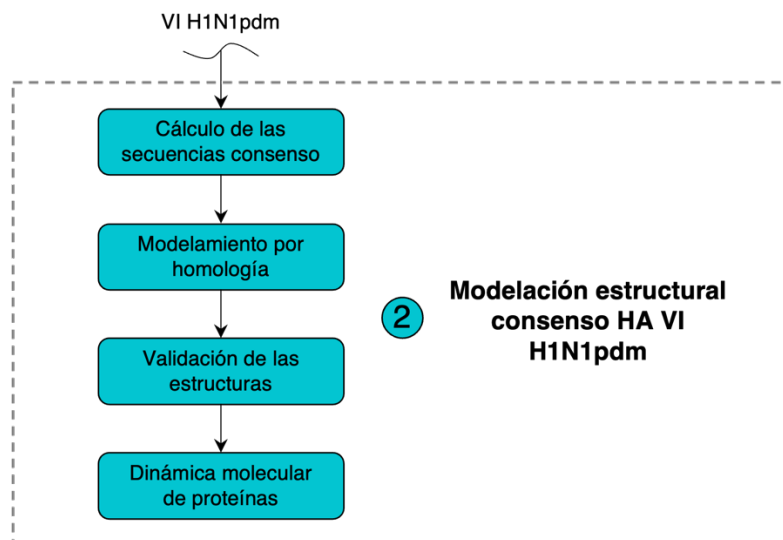


Figura 16. Flujograma de metodología capítulo 2. Resumen gráfico de los pasos seguidos en la metodología de la modelación estructural consenso de HA VI H1N1pdm.

5.1.1 Cálculo de secuencias consenso

Con el objetivo de formar *clusters* de secuencias de virus similares entre sí, se tomaron las secuencias del linaje pandémico estudiadas y se compararon contra las cepas vacunales recomendadas por la Organización Mundial de la Salud (<https://www.who.int/teams/global-influenza-programme>) usando el servidor online FluServer (<https://flusurver.bii.a-star.edu.sg>). Este alinea y determina la referencia más cercana haciendo un alineamiento de BLAST (McGinnis & Madden, 2004) contra cada referencia. Las secuencias analizadas se agruparon en *clusters* acorde a sus cepas de referencia y cada agrupación se analizó de manera independiente para determinar el consenso con uso de la herramienta Jalview (A. M. Waterhouse et al., 2009).

5.1.2 Modelamiento por homología y validación de las estructuras

Para realizar la modelación estructural de los consensos encontrados a través de homología, se realizó una búsqueda proteínas homólogas con el servidor web SWISS-MODEL (A. Waterhouse et al., 2018) en la opción “*search for templates*”. Para esto, seleccionó el molde de código PDB 6N41 y resolución de con 2.50 Å, obtenido a partir de difracción de rayos X. Los criterios usados para la selección del molde fueron mayor porcentaje de identidad y mayor resolución. Una vez seleccionada la estructura molde, se utilizó el software Modeller v10.4 (Šali & Blundell, 1993), el cual toma por entrada el molde elegido anteriormente y cada una de las secuencias objetivo a modelar y da por resultado una estructura homóloga. Se calcularon 30 modelos para cada consenso, los cuales fueron comparados a través de la función de puntuación DOPE (Discrete Optimized Protein Energy) (Shen & Sali, 2006), el cual evalúa la energía de cada modelo generado, de manera que los modelos que presentaron mejor energía fueron seleccionados para analizar posteriormente en los servidores en línea PROCHECK (Laskowski et al., 1993) y ProSA-Web(Wiederstein & Sippl, 2007). El modelo seleccionado a partir de los criterios expuestos se selecciona como el modelo representativo de los consensos generados anteriormente.

5.1.3 Dinámica molecular de las proteínas

Para el estudio de la estructura de los modelos en el tiempo, se realizó análisis de dinámica molecular en Gromacs version 2021.3 (Abraham et al., 2015) durante 50 ns. Cada sistema fue estudiado en una caja cúbica a un rango de 1.0 nm entre las paredes del componente y la proteína con condiciones periódicas en todas las direcciones. Éste se solató con un modelo de agua TIP3P (Jorgensen et al., 1983) y con un campo de fuerzas Amber99SB (Hornak et al., 2006). Para neutralizar los sistemas se usaron iones de sodio (Na⁺) y cloro (Cl⁻). Las interacciones de largo alcance se calcularon con el método PME (particle-mesh Ewald summation method) (Darden et al., 1993). Luego, se minimizó usando el algoritmo *steepest descent*, el cual se realizó hasta alcanzar un valor de fuerza inferior a 0.01 kJ/mol/nm con el objetivo de eliminar contactos incongruentes entre los átomos. Una vez se obtuvo la estructura minimizada, se procedió a equilibrarla con un número de partículas, una temperatura y un volumen constante en un sistema NVT por 100ps controlando una temperatura a 300K, utilizando un re-escalado de velocidad V-rescale (Bussi et al., 2007) y un tiempo constante T de 0.1. Subsecuentemente, se equilibró con una temperatura, una presión y un número de partículas constante (NPT) por 100 ps con uso del baróstato Parrinello-Rahman (Parrinello et al., 1981) controlando la presión a 1 bar utilizando una constante de acoplamiento de 2 ps y una compresibilidad isotérmica de $4.5 \times 10^{-5} \text{ bar}^{-1}$. Finalmente, 50 ns de producción se corrieron con un paso de tiempo de 2 fs.

5.2 Resultados

5.2.1 Cálculo de secuencias consenso representativas del VI H1N1pdm

Los 252 virus encontrados en el linaje pandémico se comparan con las cepas vacunales recomendadas por la Organización Mundial de la Salud (OMS) depositadas en Flusurver. Los resultados sugieren que los virus poseen cercanía a las cepas que se listan en la **tabla 3**, allí se nombra cada cepa y el periodo en el cual fue válida como recomendación.

Tabla 3. Número de secuencias del estudio que se mapearon a las cepas vacunales recomendadas por el OMS.

Referencia vacunal H1N1	Periodo	N° de secuencias
A/California/7/2009	2010 - 2016	142
A/Michigan/45/2015	2017 - 2019	86
A/Brisbane/02/2018	2020	13
A/Guangdong-Maonan/SWL1536/2019	2021	10
A/Sydney/5/2021	2021 - Actual	1

De los 252 virus, 142 se encontraron más cercanos a A/California/7/2009, la cual corresponde a la cepa pandémica del 2009 que estuvo vigente desde el año 2010 hasta el 2016 como recomendación de la OMS. A esta le siguen las cepas A/Michigan/45/2015, A/Brisbane/02/2018, A/Guangdong-Maonan/SWL1536/2019 y finalmente A/Sydney/5/2021 las cuales tuvieron cercanías con 86, 13, 10 y 1 virus respectivamente como se lista en la **tabla 3**.

Cada grupo de secuencias se alinea con el fin de identificar el aminoácido que ocurre con mayor frecuencia en cada sitio de la alineación, de manera que el conjunto de los aminoácidos más frecuentes forma lo que se conoce como secuencia consenso, generando así una secuencia representativa que recoge las características esenciales y permite comprender mejor las similitudes entre grupos de proteínas relacionadas. Este proceso da como resultado 5 secuencias de proteínas consenso que recogen las características más representativas de los 5 *clusters* formados e identificados. Estas secuencias consenso se nombran de acuerdo con el nombre de la cepa vacunal de referencia de cada grupo que representa, así: cCal09, cMi15, cBri18, cGM19 y cSyd22.

Con el fin de identificar las diferencias y/o similitudes mutacionales entre cada secuencia consenso frente a la referencia A/California/7/2009 -cepa que define el linaje pandémico- se realiza el alineamiento de estos 5 consensos, tal como se muestra en el **anexo 5**.

En la tabla 5 se muestran el total de mutaciones identificadas, destacando las mutaciones únicas con el fin de reconocer aquellas proteínas que no contienen mutaciones compartidas. En este caso, se puede apreciar que cBri18 y cSyd22 contienen mutaciones

que no son compartidas por ninguno de los consensos. Esta información tiene alta relevancia para el desarrollo de los siguientes apartados.

Tabla 4. Mutaciones totales y únicas identificadas a partir de la referencia A/California/07/2009 de los consensos: cCal09, cMi15, cBri18, cGM19 y cSyd22 calculados en este estudio.

Proteína	Total de mutaciones	Mutación única
cCal09	P84S, S204T, I233V, E375K	
cMi15	P84S, S85N, D98N, S163N, K164Q, S186T, S204T, I217T, A257T, K284E, I233V, E375K, S452N	
cBri18	R46G, S91R, P84S, S85N, D98N, S163N, K164Q, S165T, S184P, S186T, S204T, I217T, A257T, P283A, K284E, I296V, I299V, I233V, E375K, S452N	R46G, P283A, I299V
cGM19	S91R, P84S, S85N, D98N, N130D, S163N, K164Q, S165T, S184P, S186I, S204T, I217T, A257T, N261D, K284E, I296V, I233V, E375K, S452N	
cSyd22	K55Q, S91R, P84S, S85N, D98N, N130D, K131N, N157K, L162I, S163N, K164Q, S165T, S184P, S186I, A186T, Q190E, S204T, I217T, E225A, V151A, A257T, R260K, N261D, K284E, I296V, I233V, K309R, E375K, I419V, S452N	K55Q, A186T, Q190E, E225A, V151A, R260K, K309R, I419V

5.2.2 Modelado de la estructura 3D

Con base en la secuencia de aminoácidos de los 5 consensos encontrados, se genera una búsqueda de proteínas homologas que tengan previamente su estructura modelada en la base de datos PDB (*Protein Data Bank* <https://www.rcsb.org/>). En primer lugar, se realizó la búsqueda de un monómero de hemaglutinina que tenga las cadenas HA1 y HA2 con los siguientes parámetros de selección: un porcentaje de identidad mayor a 70% y una resolución inferior a 3Å con el fin de garantizar una alta homología del molde con la secuencia de estudio. El modelo seleccionado tiene por código de PDB 6N41, el cual fue resuelto a través de difracción de rayos X y cuenta con una resolución de 2.50 Å

garantizando así los parámetros de selección establecidos. Luego, al realizar un alineamiento local con la herramienta BLAST entre las secuencias del molde y las secuencias de estudio se encontró un porcentaje de identidad de igual a 80.49, 80.08, 80.01, 79.88 y 79.27% y una similaridad de 90, 89, 89, 88 y 88% para las secuencias cCal09, cMi15, cBri18, cGM19 y cSyd22 respectivamente.

Una vez seleccionado el molde, se procedió a modelar la estructura de las secuencias objetivo con ayuda de MODELLER v10.4, en esta plataforma se generaron 30 modelos para cada secuencia, los cuales fueron evaluados posteriormente a partir de la función DOPE. La función DOPE evalúa los 30 modelos generados y les asigna un puntaje de acuerdo con su calidad energética de manera que el modelo con el valor más bajo corresponde al más estable (Fiser & Šali, 2003; Shen & Sali, 2006). Los modelos cCal09, cMi15, cBri18, cGM19 y cSyd22 obtuvieron un puntaje de DOPE igual a -53339, -53793, -53577, -53669 y -53438 correspondientemente, esto indica que los modelos son energéticamente favorables y están bien optimizados en términos de la energía potencial de la estructura de la proteína, generando así modelos altamente plausibles y de buena calidad, por lo que estos modelos fueron seleccionados para ser sometidos a análisis posteriores.

Con el objetivo de evaluar los modelos seleccionados se hace uso del servidor en línea PROCHECK, el cual evalúa la calidad estereoquímica a través del gráfico de Ramachandran, este determina las regiones energéticamente permitidas de los ángulos ϕ y ψ de los aminoácidos para establecer si la proteína se encuentra en un estado conformacional razonable. Según el criterio que ofrecen los resultados proveídos por PROCHECK, un modelo de buena calidad debe tener mínimo el 90% de sus aminoácidos en las regiones más favorecidas. La **figura 17** muestra el gráfico de Ramachandran de cada uno de los modelos generados para las secuencias consenso.

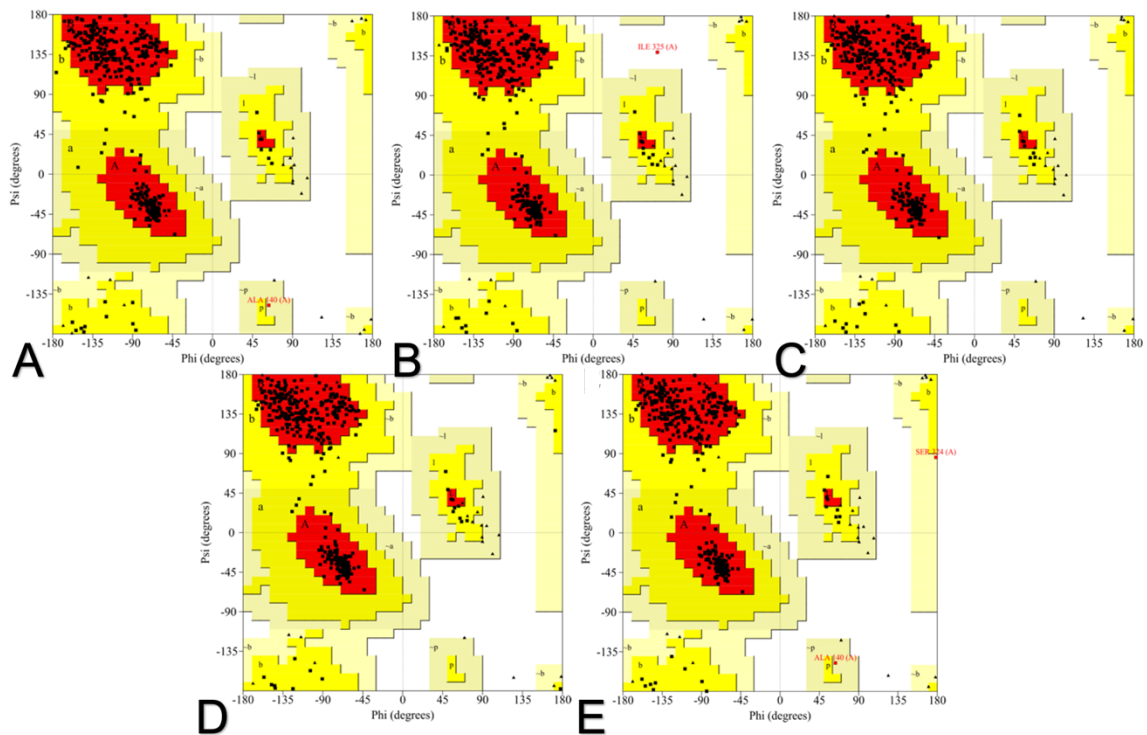


Figura 17. Gráficos de Ramachandran de las 5 estructuras consenso. Gráficos de Ramachandran con 4 regiones establecidas así: Regiones de color rojo (A, B y L) representan las más favorables. Regiones en amarillo brillante (a, b, l y p) que representan adicionales permitidas. Las regiones en amarillo opaco (\sim a, \sim b, \sim l y \sim p) son generosamente permitidas y la región en blanco que corresponde a no permitidas. Cada gráfico se distribuye entre cada consenso de la siguiente manera: Cal09 (A), Mi15 (B), Bri18 (C), GM19 (D) y Syd22 (E).

Las estadísticas extraídas de los diagramas de Ramachandran en PROCHECK muestran que el consenso cCal09 posee el 91.7% de los residuos en las regiones más favorables (A,B,L), seguida del 8.1% en regiones adicionales permitidas (a, b, l, p), 0.02% en regiones generosamente permitidas (\sim a, \sim b, \sim l, \sim p) y 0% de residuos en regiones no permitidas. El consenso cMi15 posee el 91.9% en regiones (A,B,L), 7.9% en (a, b, l, p), 0% en (\sim a, \sim b, \sim l, \sim p) y 1 residuo en regiones no permitidas. Para el caso de cBri19 el 91.4% en (A,B,L), 8,6% en (a, b, l, p) y 0% para las otras dos regiones. cGm19 se distribuye así: 91.9% en (A,B,L), 8.1% en (a, b, l, p) y sin residuos en las otras dos regiones. Finalmente, el consenso cSyd22 posee el 91.4% de residuos en (A,B,L), 8,1% en las regiones (a, b, l, p), 2 residuos en (\sim a, \sim b, \sim l, \sim p) y 0 en regiones no permitidas.

Como se muestra en la **figura 17 (B)** solo se encontró un residuo en región no permitida, correspondiente a una Isoleucina en la posición 325 del modelo cMi15, por el contrario, la **figura 17 (A, B, C, D y E)** muestra que los modelos de los consensos evaluados tiene un porcentaje de aminoácidos en las regiones más favorables mayor al 90%, lo que implica que todos los modelos se encuentran en un estado conformacional razonable.

Adicionalmente, la calidad general del modelo se evaluó utilizando el puntaje Z calculado por el servidor en línea ProSA-web. Este servidor proporciona las puntuaciones Z de todas las proteínas que han sido determinadas experimentalmente por rayos X y resonancia magnética nuclear (NMR) en el PDB, lo que permite verificar si la puntuación Z de las proteínas ingresadas se encuentra dentro del rango típico. El puntaje Z es una medida estadística utilizada para evaluar la calidad general de un modelo de estructura de proteína. Un z-score más bajo generalmente sugiere una estructura de proteína más confiable y plausible, mientras que un z-score más alto puede indicar la necesidad de una mayor refinación o validación del modelo (Sippl, 1993).

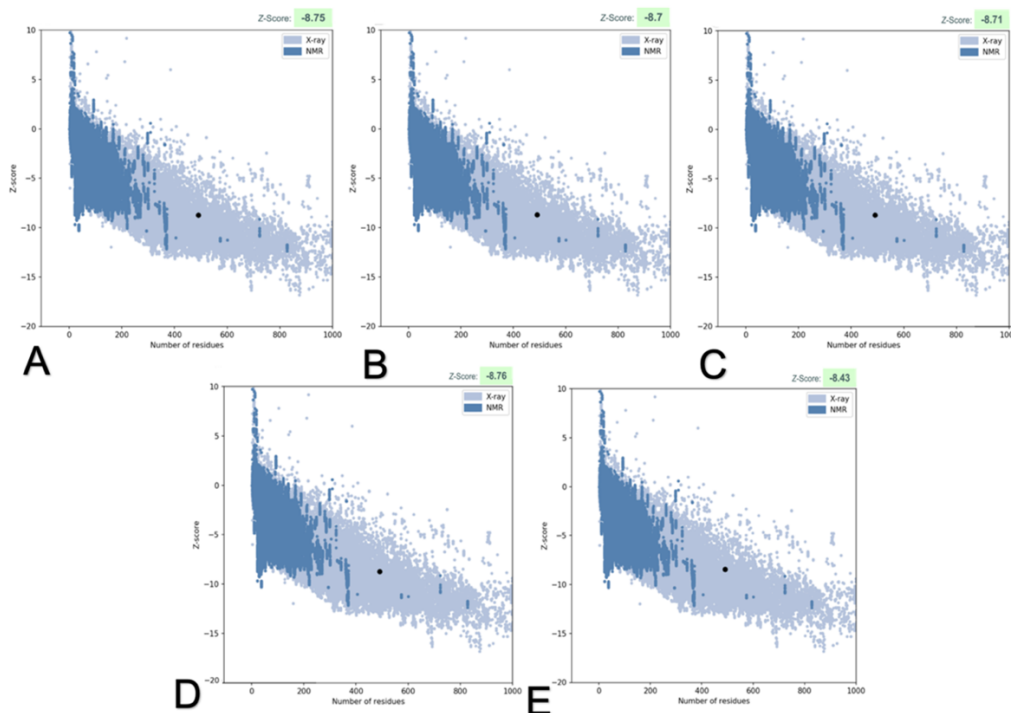


Figura 18. Z-score de las proteínas consenso. El punto negro en las gráficas representa el Z-score de los modelos consenso: cCal09, cMi15, cBri18, cGM19 y cSyd22 contra el Z-score de las proteínas depositadas en el PDB.

En la **figura 18 (A, B, C, D y E)** se observa dónde se posicionan cada uno de los modelos de los consenso con respecto a otras estructuras obtenidas a través de cristalografía de rayos X. Con valores de Z-score igual a: -8.75, -8.7, -8.71, -8.76 y -8.43 para los cCal09, cMi15, cBri18, cGm19 y cSyd22, se puede apreciar que todos los modelos se encuentran en los rangos característicos de proteínas de tamaño similar, indicando que todos los modelos tiene una estructura confiable.

Una vez se modeló y validó las proteínas de estudio, se procedió a graficar su estructura tridimensional (**figura 19**) en donde se presentan las mutaciones y su ubicación en la estructura que posee cada variante solamente en su dominio de unión al receptor en comparación con la cepa de referencia A/California/7/2009 (WT). A partir de esta gráfica, se puede apreciar que la cepa cCal09 no presentó mutaciones en el dominio de unión al receptor, mientras que las cepas posteriores van acumulando más mutaciones. Las posiciones no mutadas no se muestran en la figura.

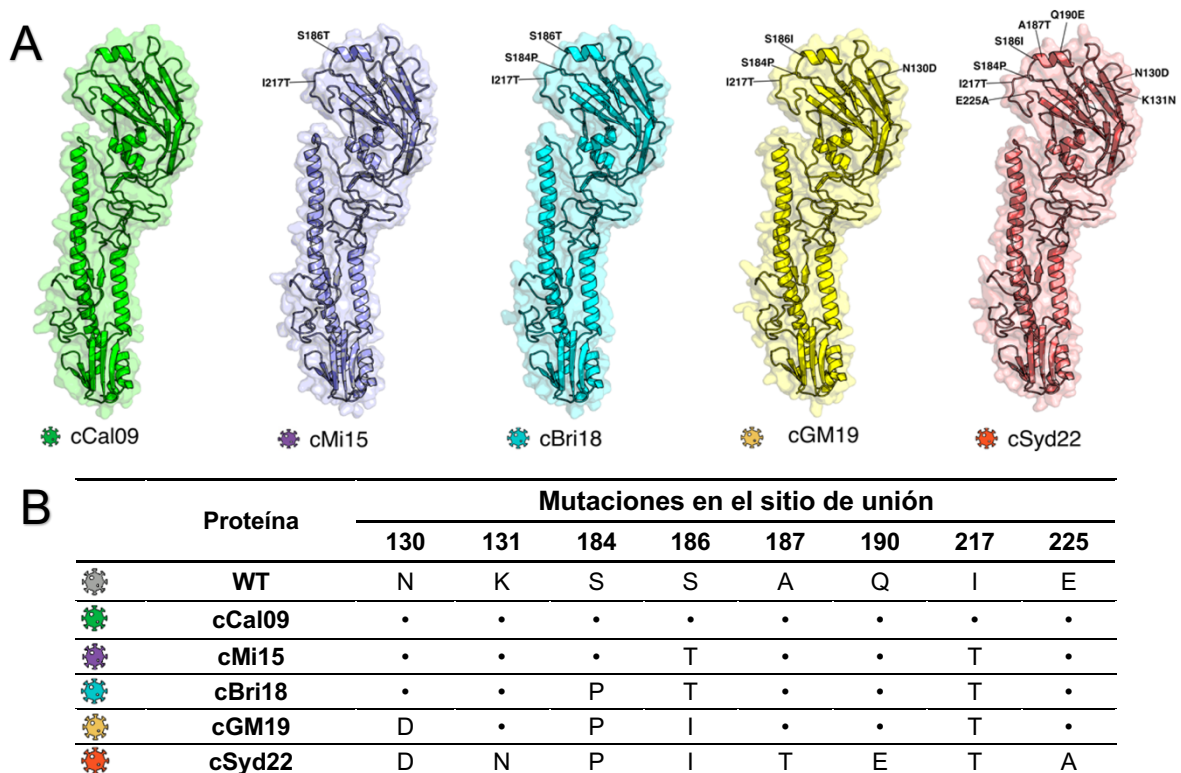


Figura 19. Estructuras tridimensional de los consensos modelados y sus mutaciones. (A) Estructuras de los 5 consensos señalando la ubicación de las mutaciones

presentes en el sitio de unión. (B) Lista de sustituciones en el sitio de unión respecto a la referencia (WT).

5.2.3 Dinámica molecular de las estructuras.

Minimización y equilibrio

Una vez obtenidas las estructuras se realiza una minimización de cada una de las estructuras, este proceso relaja la estructura al generar un re-arreglo espacial de los átomos a favor del gradiente de energía hasta una fuerza inferior a $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. La energía potencial en este proceso se grafica en la **figura 20 (A)** donde se observa su comportamiento hasta alcanzar valores de: $-5.61 \times 10^6 \text{ kJ/mol}$ en el paso 2769, $-5.36 \times 10^6 \text{ kJ/mol}$ en el paso 3015, $-5.57 \times 10^6 \text{ kJ/mol}$ en el paso 1940, $-5.47 \times 10^6 \text{ kJ/mol}$ en el paso 3168 y $-5.78 \times 10^6 \text{ kJ/mol}$ en el paso 2011 para cCal09, cMi15, cBri18, cGM19 y cSyd22, respectivamente. Luego, el sistema se equilibra en valores promedios aproximados de: densidad de 993 Kg/m^3 , temperatura de 300 K y presión de 1 bar , como se muestra en las **figuras 20 (B-D)**.

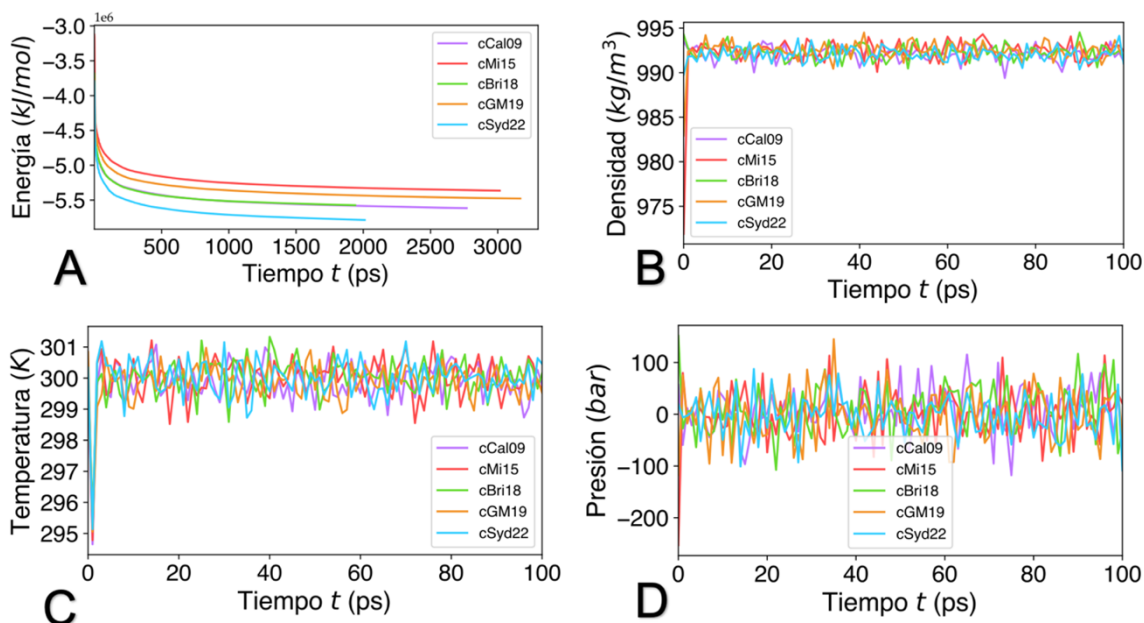


Figura 20. Medición de propiedades durante minimización y equilibrio: Propiedades termodinámicas del sistema en el tiempo. A) Energía de minimización del sistema, B) Densidad, C) Temperatura y D) Presión.

Al finalizar el equilibrio se simula durante 50 ns y se calculó la raíz de la desviación cuadrática media (RMSD) del esqueleto de los modelos como función del tiempo con el objetivo de monitorear cambios estructurales. En la **figura 21 (A)** se observa que el RMSD del modelo de cBri18 presenta un aumento a partir del nanosegundo 25, el cual vuelve a estabilizarse y mantener valores cercanos a los iniciales cerca al nanosegundo 30. En general, el valor RMSD de todas las proteínas presentan la misma tendencia con valores iguales a: 3.05 ± 0.54 , 2.54 ± 0.38 , 3.72 ± 0.58 , 2.89 ± 0.45 y 2.62 ± 0.38 Å para los modelos: cCal09, cMI15, cBri18, cGM19 y cSyd22 respectivamente, exhibiendo una variación estable. Adicionalmente, se calculó de radio de giro (R_g) del esqueleto de las estructuras a lo largo de la simulación, éste representa una medida característica de la compacidad de la estructura, en donde un valor de R_g bajo implica una proteína más compacta y menos flexible. La **figura 21 (B)** muestra que el radio de giro varió en valores alrededor de 3.8 y 3.7 nm para todas las estructuras, excepto para cGM19, que experimentó una ligera disminución hasta 3.65 nm a partir del nanosegundo 40, seguido de un incremento posterior que la llevó a recuperar su valor inicial.

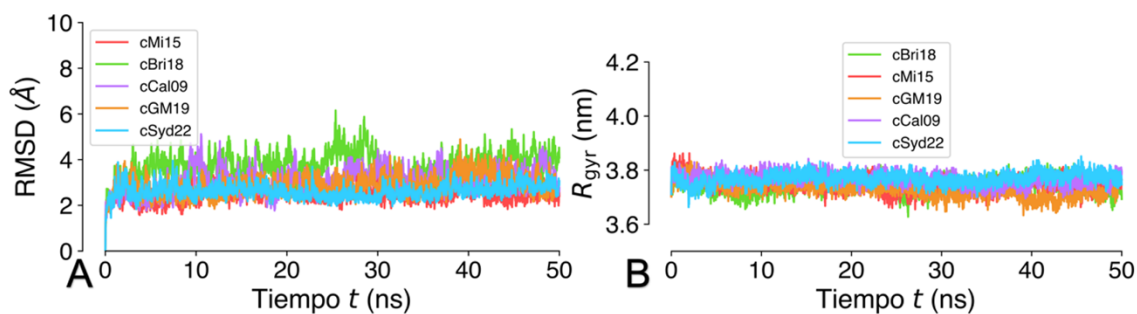


Figura 21. Métricas de dinámica molecular de las consensos modelados: A) Desviación cuadrática media de la raíz (RMSD) y B) Radio de giro (R_g) del esqueleto de las proteínas cCal09, cMI15, cBri18, cGM19 y cSyd22 en el tiempo.

Luego, se calculó la raíz de la fluctuación media cuadrática (RMSF), esta es una medida de la flexibilidad o movilidad de cada átomo o residuo en una estructura de proteína, de manera que un RMSF más alto indica que un residuo en particular es más flexible, mientras que un valor más bajo indica que es más rígido. La **figura 22** indica que la mayoría de los residuos fluctuaron con una intensidad menor a 3 Å incluyendo los residuos del dominio de unión al receptor, los cuales son resaltados en naranja, rojo y azul. Asimismo, se

observaron valores superiores de RMSF en otras regiones, dentro de las cuales se destaca la región cercana al residuo 331, que corresponde a la región de unión de los dominios HA1 y HA2, lo cual explica la alta tendencia del RMSF para todas las proteínas. Otra región que se destaca es la región posterior a los residuos del 391, resaltada en amarillo. Esta región mostró una alta fluctuación especialmente para cBr18. Al considerar las mutaciones únicas presentes en cBr18 mostradas en la tabla 5, se encontró que dos de ellas (P283A y I299V), resaltadas en verde en la **figura 22**, se encuentran cerca de esa región fluctuante, lo que sugiere que estas mutaciones podrían estar generando una desestabilización que se traduce en una mayor fluctuación.

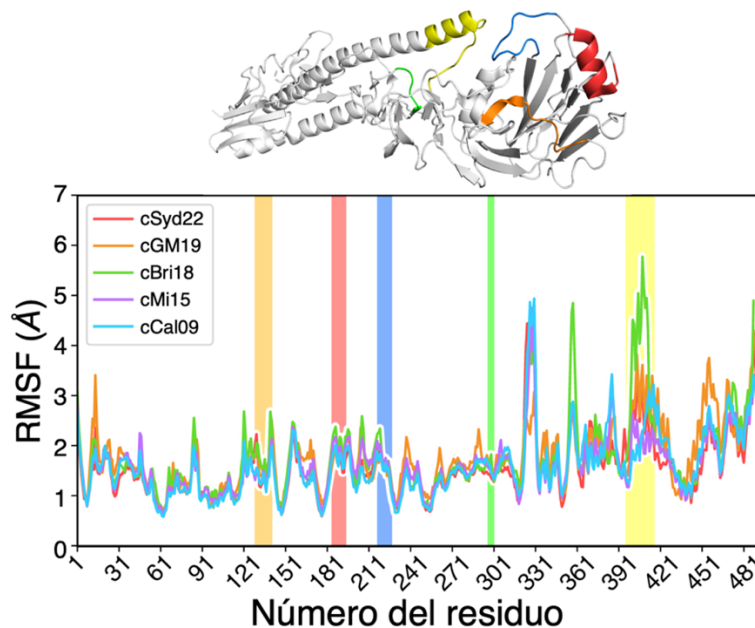


Figura 22. RMSF de los consensos modelados. raíz de la fluctuación media cuadrática (RMSF) de las proteínas cCal09, cMi15, cBri18, cGM19 y cSyd22.

Con el objetivo de realizar una caracterización estructural del sitio de unión al receptor, en la figura 25 se muestra la alineación de las estructuras tridimensionales del *loop* 130, la hélice 190 y el *loop* 220 de las 5 proteínas estudiadas después de los 50ns de producción de dinámica. Una inspección de esta figura permite observar que cBri18 experimentó un cambio conformacional significativo en el *loop* 130 que llevó a este dominio a alejarse de la región de fluctuación encontrada en las otras proteínas.

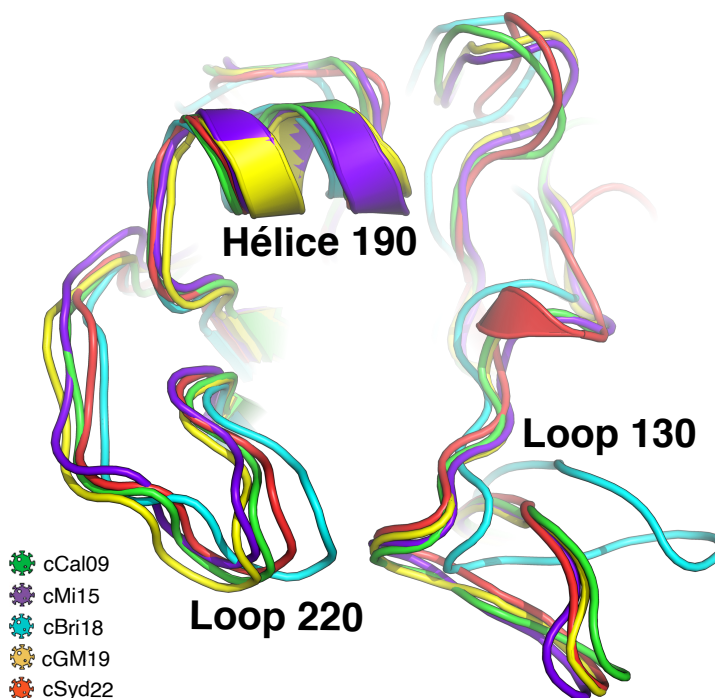


Figura 23. Alineación del RBD de las estructuras modeladas. Comparación de las estructuras tridimensionales del sitio de unión al receptor de las proteínas de HA: cCal09 (verde), cMi15 (Morado), cBri18 (Azul), cGM19 (Amarillo) y cSyd22 (Rojo), después de 50ns de producción en dinámica molecular.

Con el objetivo de verificar el cambio conformacional de la estructura observado en el apartado anterior y evaluar la estabilidad del “*pocket*” formado en el sitio de unión, se evaluará una métrica adicional. Para ello, se ejecuta un cálculo de distancias entre las 3 regiones que lo conforman (*loop* 130, hélice 190 y *loop* 220) como se observa en la figura 24 (A). Este cálculo se realiza a lo largo de los 50ns en cada una de las proteínas de estudio, para el cual se seleccionaron 3 aminoácidos conservados en las 5 proteínas: Thr134, Leu192 y Gln224. Las distancias fueron calculadas entre los carbonos α de cada uno de estos aminoácidos. Como se aprecia en la figura 24 (B-F), la distancia promedio d_1 (entre Thr134 y Leu192), d_2 (Thr134 y Gln224) y d_3 (Leu192 y Gln224) fue alrededor de 13.3 nm, 9.1 nm, y 14.6 nm para los modelos cCal09, cMi15, cGM19 y Syd22 y esta se mantuvo cerca de estos valores a lo largo del tiempo. Por otra parte, para el modelo cBri18, las distancias d_1 y d_2 se mantuvieron en 13.3 nm y 9.1 nm, pero aproximadamente en el nanosegundo 35, d_1 se aumentó hasta 14.6 nm y d_2 se redujo

hasta 7.9 nm, fluctuando alrededor de estos valores hasta llegar a los 50 ns. La distancia d_3 sí conservó un valor promedio de 14.6 nm como los otros modelos. Por lo que se puede apreciar que los aumentos encontrados en las distancias d_1 y d_3 corresponden al cambio estructural observado en el *loop* 130 que se detalla en la **figura 23**.

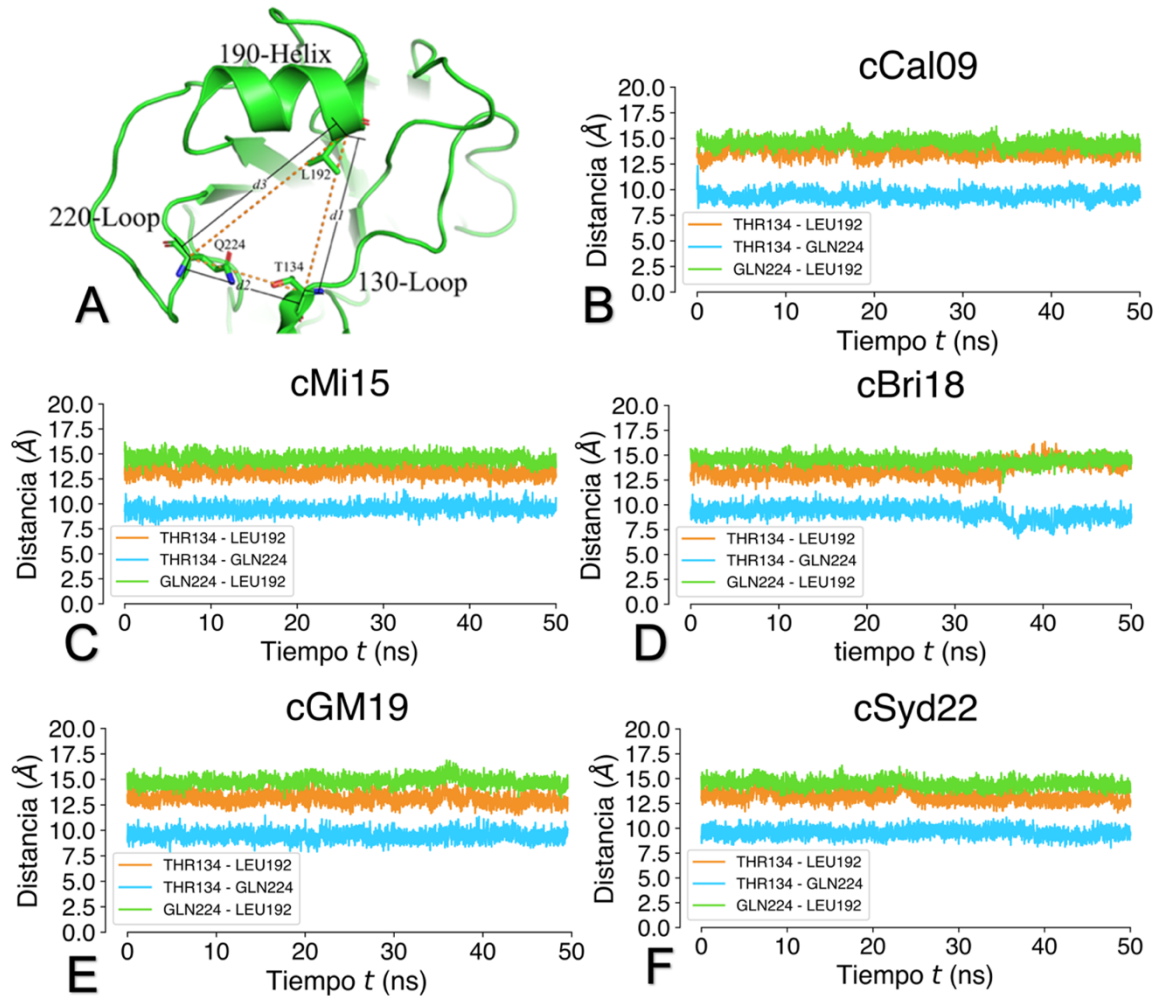


Figura 24. Distancia de las regiones del RBD. Distancias entre los aminoácidos T134, L192 y Q224 de las proteínas cCal09 (A), cMi15 (B), cBri18 (C), cGM19 (D) y cSyd22. Durante 50ns de producción de dinámica molecular.

5.3 Discusión

La funcionalidad de una proteína depende fuertemente de su estructura, por lo que en este estudio se realiza un modelamiento estructural de las proteínas H1 consenso que represente las cepas reportadas en de H1N1pdm en Colombia y que aporte a la brecha de conocimiento que se tiene acerca de las características con las que circula el VI H1N1pdm en el país.

En este estudio, se obtuvieron cinco secuencias consenso (cCal09, cMi15, cBri18, cGM19 y cSyd22) que representan la diversidad de los virus de la influenza H1N1pdm circulantes en Colombia. Cada secuencia consenso refleja las características principales y, por lo tanto, las mutaciones más predominantes del estudio (ver **tabla 5**). Estos clústers se crearon con base en la homología encontrada con cepas vacunales establecidas por la OMS, las cuales tienen en cuenta los principales clados que circulan dentro del linaje pandémico.

La **tabla 3** muestra el número de cepas virales representadas por cada consenso, la gran mayoría de los virus de linaje pandémico identificados en Colombia muestran una homología más alta con la cepa pandémica vacunal del 2009 A/California/07/2009, con un 57% de los virus analizados. Estos hallazgos indican que la proteína cCal09 es el consenso que recoge las características más predominantes de la mayoría de los virus circulantes en Colombia entre 2009 y 2022.

Luego, se obtuvieron las estructuras de los 5 consensos, los cuales se analizaron a través de dinámica molecular para evaluarlas en el tiempo. Estas simulaciones mostraron que las proteínas cCal09, cMi15, cGM19 y cSyd22 mostraron fluctuaciones similares como se observó en las **figuras 21 y 22**, mientras que las fluctuaciones para cBri18 presentó diferencias a las encontradas en las otras proteínas, lo que indica que dicha proteína experimentó cambios conformacionales, distintos a los encontrados en las otras proteínas a lo largo del tiempo.

Adicionalmente, un alineamiento estructural observado en la **figura 23** entre las regiones principales del RBD (*loop* 130, hélice 190 y *loop* 220) de las proteínas después de la

dinámica molecular reveló que el *loop* 130 de la proteína cBri18 se encontraba en una posición más alejada a la encontrada en los *loops* 130 de las otras proteínas, hecho que fue confirmado al medir las distancias de estas regiones del RBD a lo largo de los 50ns en la **figura 24**, ya que se observó cómo alrededor del nanosegundo 35 el *loop* 130 del cBri18 se alejó de la hélice 190 y el *loop* 220. Estos cambios conformacionales en el sitio de unión pueden ser factores claves para determinar la especificidad del receptor de los virus, como se afirma en un estudio realizado en el 2018 en el que Ni y colaboradores reportaron que las variaciones en las longitudes del *loop* 130 son responsables del cambio preferencia del receptor (Ni et al., 2018). Esto sugiere que cBri18 puede presentar cambios en la afinidad por los receptores, alterando factores como la eficiencia de infección y transmisibilidad en huéspedes como cerdos, humanos y aves.

A la hora de buscar factores que hayan repercutido en los cambios en el comportamiento de cBri18, se encontró que dicha proteína posee 3 mutaciones (R46G, P283A y I299V) que no se encuentran en ninguna de las otras proteínas estudiadas como se muestra en la **tabla 4**. Si bien las mutaciones en cuestión no están directamente ubicadas en las regiones principales del Receptor Binding Domain (RBD), su influencia podría manifestarse en los cambios conformacionales que resultaron en la modificación del *loop* 130. Es importante destacar que dos de estas mutaciones sí se encuentran en proximidad conformacional a la región que exhibió una alta fluctuación en el cálculo del Root Mean Square Fluctuation (RMSF). en la **figura 22**.

Al hacer una revisión en la literatura con ayuda de la base de datos de flusurver acerca de la relación que puedan tener estas mutaciones con los cambios conformacionales observados, se encontró que estas mutaciones se encuentran en 2,64% (R46G), (P283A) 2,37% y (I299V) 2.5% de los virus reportados en flusurver y no se encontró ningún estudio que relacione dichas mutaciones con cambios conformacionales.

En resumen, en este estudio se aplicó una metodología para crear 5 estructuras tridimensionales (cCal09, cMi15, cBri18, cGM19 y cSyd22) que representan la diversidad de los virus en Colombia, permitiendo evaluar a través de dinámica molecular los efectos que las mutaciones más predominantes generan sobre las estructuras de estas proteínas en el tiempo. Esto permitió encontrar que las mutaciones contenidas en las proteínas cCal09, cMi15, cGM19 y cSyd22 no tuvieron efecto significativo que se manifestara en la

fluctuaciones evaluadas en dinámica, ni en la conformación final de su estructura, revelando que no han generado cambios estructurales significativos. Por otro lado, se encontró que el consenso cBri18 sí manifestó fluctuaciones más elevadas y cambios estructurales en comparación a las otras proteínas, en especial en el *loop* 130 del RBD, hecho que podría estar asociado a las mutaciones únicas (R46G, P283A y I299V) identificadas en esta proteína.

6. Capítulo 3: Análisis unión HA y receptores celulares

El objetivo de este capítulo es llevar a cabo un estudio enfocado en analizar la afinidad entre los virus de influenza en Colombia, específicamente del linaje H1N1pdm, y los principales hospederos entre los cuales circula que incluyen al cerdo, humano y ave. Para ello, nos centraremos en los receptores $\alpha 2,3$ y $\alpha 2,6$. Estos receptores fueron seleccionados debido a que, como se detalla en la sección 2.13, son las formas predominantes expresadas en dichos hospederos. Analizaremos las interacciones entre estos receptores celulares y las proteínas consenso de interés en Colombia, todas pertenecientes al linaje H1N1pdm. El propósito fundamental de este análisis es evaluar si las sustituciones presentes en estas proteínas afectan significativamente la capacidad del virus para infectar distintos tipos celulares y si hay una preferencia evidente por un tipo específico de receptor celular.

6.1 Metodología

Este capítulo toma los modelos de los consensos de HA después de los 50ns de dinámica calculados en el capítulo anterior (**capítulo 2, sección 4.2.1**) y se incluyen en el paso a paso desarrollado en la metodología, en la **figura 25** se muestra un resumen gráfico.

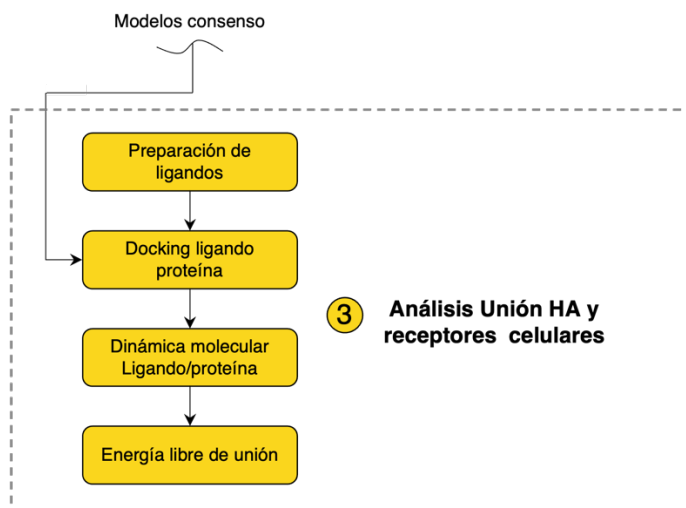


Figura 25. Flujograma de metodología capítulo 3. Resumen gráfico de los pasos seguidos en la metodología del Análisis unión HA y receptores celulares.

6.1.1 Preparación de ligandos de receptores celulares

Para obtener la estructura de los receptores celulares, se descargaron del banco de datos de proteínas (PDB, <https://www.rcsb.org/>) de las estructuras que tienen por código 4N5K para $\alpha 2,3$ y 4N60 para $\alpha 2,6$. Luego, se les realizó un proceso de optimización geométrica y molecular, que consta de una minimización en el programa Avogadro-1.2.0 (J. Wang et al., 2004) haciendo uso del campo de fuerzas generalizado de Amber (GAFF).

6.1.2 Docking Ligando-Proteína

Con el objetivo de obtener las conformaciones de los ligandos descritos en la **sección 5.1.1** y las proteínas modeladas, se realizó un *docking* molecular con uso del *software* Qvina2 (Alhossary et al., 2015) en donde se ubicó la caja de estudio en RBD de las proteínas, región donde se obtuvieron los complejos unidos y se analizaron las interacciones. A cada una de las proteínas se les evaluó la afinidad respecto a cada uno de los ligandos. Cada corrida de *docking* se realizó 200 veces de manera independiente,

de cada corrida se extrajeron los 3 modos más estables obtenidos. Adicionalmente, se hace uso del parámetro “*Exhaustiveness*” para modificar el esfuerzo computacional durante el *docking*, este se fijó en 24 para obtener resultados más consistentes (Forli et al., 2016). La herramienta MGLtools V 1.5.7 (<https://ccsb.scripps.edu/mgltools/>) se utilizó para posicionar y centrar la caja en el sitio de unión con el receptor celular, cada tamaño fue elegido independientemente para garantizar la inclusión del sitio de unión permitiendo rotaciones libres del ligando. Cada corrida se realizó de manera semi-rígida, lo que significa que la proteína se mantiene rígida y los ligandos del receptor flexibles. Los complejos extraídos de este procedimiento fueron visualizados en Pymol (<http://www.pymol.org/pymol>) y las interacciones del ligando con la proteína de manera bidimensional fueron visualizadas Ligplot (Laskowski & Swindells, 2011).

6.1.3 Dinámica molecular HA y receptor celular

Para el estudio de los complejos formados ligando-proteína (H1pdm/ α 2,3 y H1pdm/ α 2,6) en la **sección 5.1.2**, se realizó análisis de dinámica molecular usando la versión 2021.3 de Gromacs (Abraham et al., 2015). Se usó el campo de fuerzas Amber99SB(Hornak et al., 2006) para las proteínas de H1. Para los ligandos (α 2,3 y α 2,6) se usó el campo de fuerzas generalizado de Amber (GAFF) (J. Wang et al., 2004), en el cual el programa Acpype (Sousa Da Silva & Vranken, 2012) fue usado para generar la topología compatible con Gromacs. El modelo TIP3P (Jorgensen et al., 1983) se usó como solvente. Para la minimización se usó el algoritmo *steepest descent* hasta llegar a una fuerza máxima que no excediera 1000 kJ mol⁻¹ nm⁻¹. Luego, se equilibró en un sistema NVT a una temperatura de 300K utilizando un re-escalado de velocidad V-rescale (Bussi et al., 2007) y un tiempo constante T de 0.1 por 100ps. Posteriormente, un equilibró en un sistema (NPT) se llevó a cabo por 1000 ps con uso del barostato Parrinello-Rahman (Parrinello et al., 1981) a 1 bar de presión con una compresibilidad isotérmica de 4.5x10⁻⁵ bar⁻¹ y constante de acoplamiento de 2ps. Por último, se corrieron 100ns de producción con un paso de tiempo de 2 fs.

6.1.4 Cálculo de energías de unión

Finalmente, para el cálculo de las energías libres de unión de los complejos proteína-ligando se hizo uso del método de MMGBSA (Molecular Mechanics Generalized Born Surface Area) (Hou et al., 2011b) por sus siglas en inglés. Este cálculo es realizado por medio de la siguiente ecuación:

$$\Delta G_{\text{unión}} = G_{\text{complejo}} - (G_{\text{proteína}} + G_{\text{ligando}})$$

En donde el total de la energía libre del complejo está representada por $\Delta G_{\text{unión}}$, mientras que la energía de la proteína y del ligando de manera independiente se representan por G_{complejo} y G_{ligando} respectivamente;

$$\Delta G_{\text{unión}} = \Delta G_{\text{gas}} + \Delta G_{\text{sol}} = \Delta E_{\text{vdw}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdw}} + \Delta G_{\text{polar}} + \Delta G_{\text{no-polar}}$$

En donde ΔE_{gas} es la energía potencial promedio que incluye dos términos, las interacciones de Van del Waals y las interacciones electrostáticas ($\Delta G_{\text{gas}} = \Delta E_{\text{vdw}} + \Delta E_{\text{ele}}$). ΔG_{sol} es la contribución de la energía libre de solvatación compuesta de la energía libre de solvatación (ΔG_{polar}) y la energía de solvatación no polar ($\Delta G_{\text{no-polar}}$), las contribuciones polares calculadas por el Born generalizado (GB) (*electrostatic solvation energy*) y los no polares al SASA (*Solvent-Accessible Surface Area*) por sus siglas en inglés (*non-electrostatic solvation component*):

$$\Delta G_{\text{sol}} = \Delta G_{\text{GB}} + \Delta G_{\text{SASA}}$$

Este cálculo se estimó usando el paquete gmx_MMPBSA (Valdés-Tresanco et al., 2021), el cual se realizó con un total del 1000 frames que se tomaron en los últimos 10 ns de la simulación. Las contribuciones energéticas de cada residuo se realizaron con los aminoácidos pertenecientes al RBD y aquellos aminoácidos que se encontraban a una distancia inferior a 4 Å.

6.2 Resultados

6.2.1 Selección de modelos

Para evaluar el efecto que las mutaciones tienen en la unión del ligando receptor y disminuir el tiempo de cómputo asociado a los análisis de dinámica, se procedió a

seleccionar aquellas proteínas que presentaban características relevantes. Teniendo como proteínas candidatas a cCal09, cBri18 y cSyd22. cCal09 porque dicha variante es altamente homóloga a la cepa pandémica de California del 2009, proporcionando una cepa de referencia para comparaciones. Y cBri18 y cSyd22 porque, como se muestra en la **tabla 4**, presentan mutaciones exclusivas que no se encuentran en ninguna otra variante. En adición, cBri18 ha evidenciado cambios estructurales que son de interés para la evaluación con respecto a los ligandos y cSyd22 representa las cepas virales que están circulante durante los últimos años.

6.2.2 Docking molecular

Una vez obtenidas y validadas las estructuras de los consensos y los ligandos, se procede a realizar un acoplamiento entre los dos receptores: aviar ($\alpha 2,3$) y humano ($\alpha 2,6$) contra cada una de las proteínas en su dominio de unión al receptor. En las **figuras 26 (A-F)** se muestra el valor de la energía de afinidad de los complejos, el cual es un puntaje asignado a las interacciones generadas en la unión, en donde a valores más negativos de la afinidad, mejor es la unión entre el ligando y la proteína. Dicho lo anterior, los resultados obtenidos en las **figuras 26 (A-F)** corresponden a las mejores conformaciones encontradas de 200 experimentos realizados para cada complejo, de **(A-C)** para las variantes en unión con el receptor humano y de **(D-F)** en unión con el receptor aviar. En estas gráficas se ilustran las interacciones generadas, donde se detallan tanto los puentes de hidrógeno con sus respectivas distancias como los residuos con los cuales se logra tener una interacción hidrofóbica.

Los aminoácidos con los que se formaron puentes de hidrógeno se sintetizaron en la **tabla 5**. Los residuos con los que se formaron enlaces son consistentes con las interacciones reportadas en otras estructuras cristalográficas (Xu et al., 2012). Adicionalmente, se puede apreciar que los residuos que pertenecen al *loop* 200 (224, 225 y 226) y Try92 estuvieron presentes en casi todas las conformaciones, en especial las que se encuentran unidas al receptor humano.

Tabla 5. Aminoácidos con los que se forma interacción electrostática entre los receptores humano y aviar y las proteínas cCal09, cMi15, cBri18, cGM19 y cSyd22.

Sistema	Posición															
	92	131	133	134	135	136	143	181	184	185	188	191	220	224	225	226
cCal09	Y	K	-	T	-	-	-	H	-	-	-	-	-	-	E	-
α 2,6	cBri18	Y	K	-	-	-	-	K	-	-	-	-	-	-	E	G
	cSyd22	Y	N	-	-	A	A	-	-	-	D	-	-	Q	A	G
	cCal09	-	K	-	T	-	-	K	-	-	-	-	K	-	-	-
α 2,3	cBri18	Y	-	-	-	-	-	-	-	-	-	-	-	-	E	G
	cSyd22	Y	N	V	T	A	-	K	-	-	-	-	-	Q	A	G

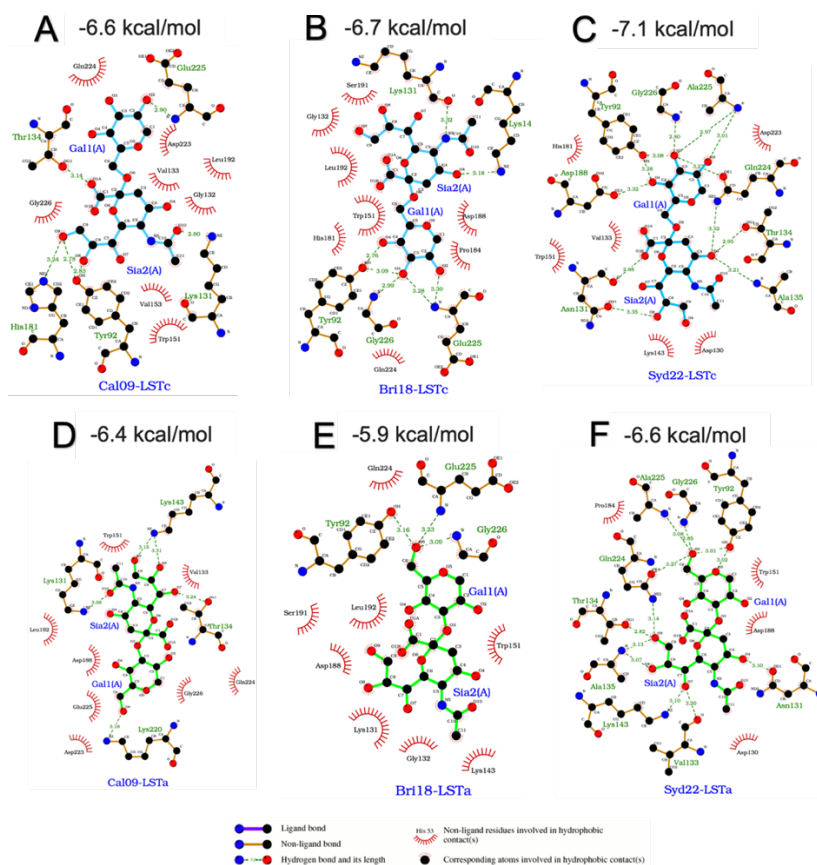


Figura 26. Diagrama de las interacciones de *docking*. Interacciones entre los receptores humano (A-C) y aviar (D-F) con las proteínas consenso cCal09, cBri18 y cSyd22. *Docking* realizado con Qvina2 y Diagrama visualizado con LigPlot.

Los cálculos de energía de unión realizados por la función de *scoring* de *docking* poseen muchas limitaciones dado que el experimento se realiza en condiciones ideales como: una unión en el vacío -dado que no tiene solvente- o una proteína rígida, lo que implica que el ligando no altera la conformación de esta. Es por esto que para realizar un cálculo más realista acerca de cómo afectan las mutaciones la unión entre el ligando y la proteína, se procederá a realizar un análisis de dinámica molecular que mejore estas condiciones que no se pudieron evaluar.

6.2.3 Dinámica molecular ligando-proteína

Minimización y equilibrio

Con los complejos seleccionados se realizó nuevamente un proceso de dinámica molecular, la cual inicia con una minimización hasta una fuerza inferior a $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ en cada átomo, que genera una energía potencial de: $-4.14 \times 10^6 \text{ kJ/mol}$ en el paso 2952, y $-4.14 \times 10^6 \text{ kJ/mol}$ en el paso 2491 para los complejos de cCal09. $-3.8 \times 10^6 \text{ kJ/mol}$ en el paso 3282 y $-3.79 \times 10^6 \text{ kJ/mol}$ en el paso 2448 para los complejos de cBri18. Y $-4.32 \times 10^6 \text{ kJ/mol}$ en el paso 2918 y $-4.31 \times 10^6 \text{ kJ/mol}$ en el paso 2376 para los complejos de cSyd22 como se muestra en la **figura 27 (A)**. Los sistemas fueron equilibrados a una densidad de densidad de 993 Kg/m^3 , temperatura de 300 K y presión de 1 bar , como se muestra en las **figuras 27 (B-D)**.

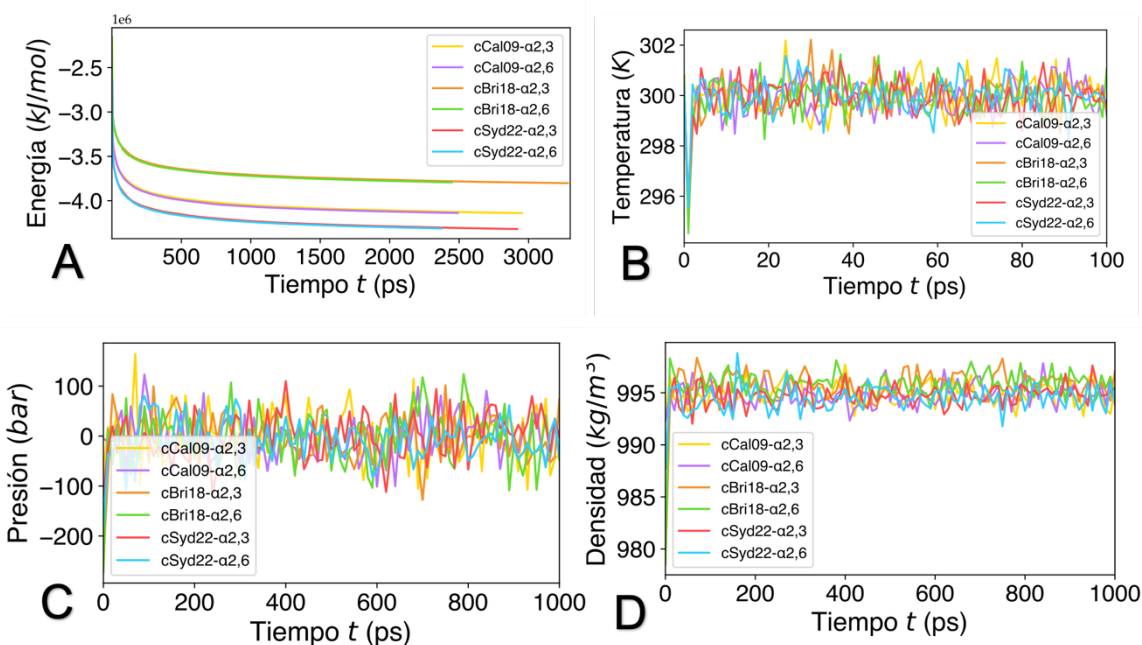


Figura 27. Medición de las propiedades en minimización y equilibrio: Propiedades termodinámicas de los complejos ligando-proteína en el tiempo. A) Energía de minimización del sistema, B) Densidad, C) Temperatura y D) Presión.

Luego de que los sistemas se encontraran en equilibrio se realizó una producción durante 100 ns y se procedió a medir las distancias entre el centro geométrico del RBD de las proteínas y los receptores celulares humano y aviar, como muestra el esquema realizado en la **figura 28 (C)**, para estudiar si los ligandos de cada uno de los complejos permanecían en RBD. En la **figura 28 (A)**, se observan las distancias entre el receptor humano $\alpha 2,6$ y

el sitio de unión de las proteínas. Se puede apreciar que en el caso de cBri18 y cSyd22, la distancia se mantuvo estable con un promedio inferior a 5 Å. Sin embargo, cCal09 experimentó fluctuaciones desde el inicio de la simulación hasta antes del nanosegundo 30 y, a partir de ese momento, se mantuvo estable junto con las otras proteínas. **La figura 28 (B)** muestra las distancias de las proteínas con respecto al receptor aviar. Aquí se puede apreciar que el ligando abandonó el sitio de unión de cCal09 antes de iniciar la producción, durante las etapas de equilibrio. En el caso de cSyd22, $\alpha_{2,3}$ abandonó el sitio antes del nanosegundo 20, mientras que el ligando de cBri18 aumentó su distancia, permaneciendo por encima de 5 Å, pero sin abandonar completamente el sitio de unión. El receptor $\alpha_{2,3}$ abandonó el sitio de unión de las proteínas cCal09 y cSyd22 debido a que las interacciones formadas no eran lo suficientemente fuertes como para mantener el receptor unido al ligando de manera estable a lo largo del tiempo. Los complejos que mantuvieron su ligando en el sitio de unión se tomaron para análisis posteriores.

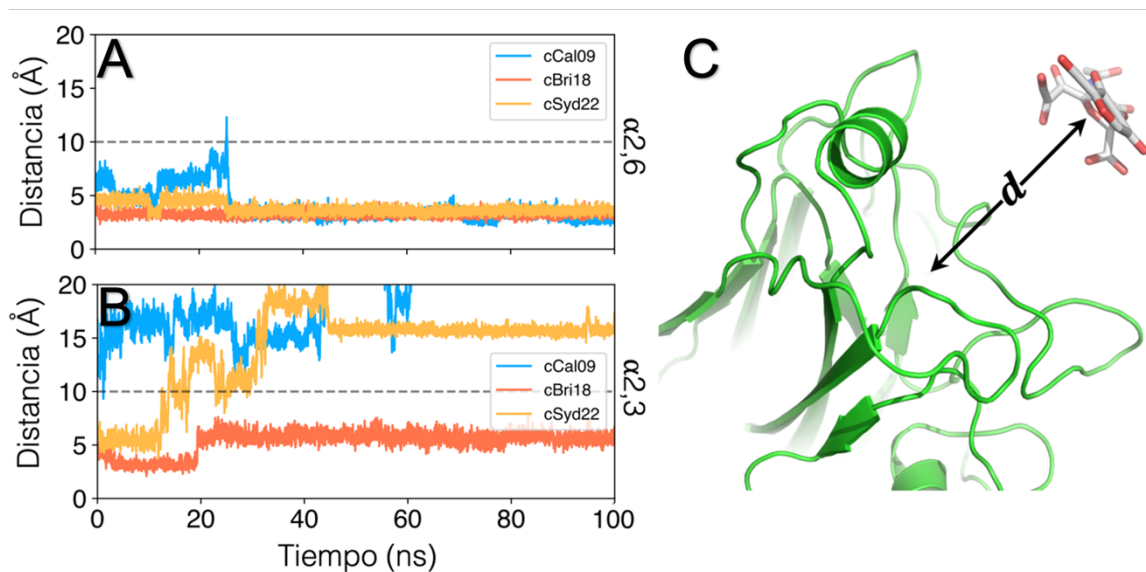


Figura 28. Medición entre el sitio de unión de los consensos y los receptores celulares. Medición de distancias entre el centro geométrico de los receptores celulares aviar y humano y el sitio de unión de las proteínas cCal09, cBri18 y cSyd22.

RMSD

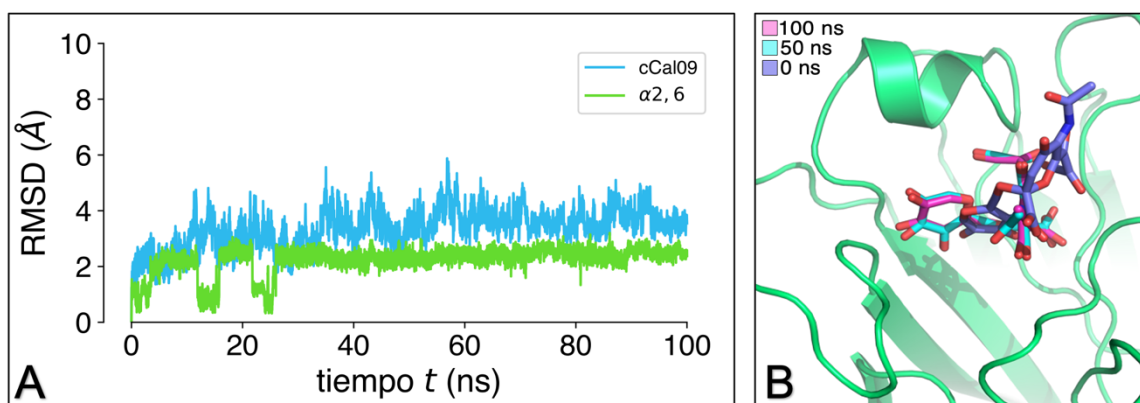
A lo largo de los 100ns también se midió el RMSD para el receptor humano y sus respectivas proteínas como se ilustra en la **figura 29 (A, C y E)**. Además, se muestran

imágenes superpuestas de la posición del ligando en el RBD en los tiempos 0, 50 y 100ns en las **figuras 29 (B, D y F)**. El RMSD del α 2,6 unido a cCal09 en la **figura 29 (A)** muestra fluctuaciones con promedios entre 0,7 y 2,1 hasta llegar al nanosegundo 26, equilibrándose alrededor de 2.1, esto también se ve reflejado en la **figura 29 (B)** ya que se ilustra cómo la posición del ligando en el nanosegundo 0 difiere de las posiciones en 50 y 100ns, las cuales no muestran cambios significativos.

Para el caso de α 2,6 unido a cBri18, el RMSD en el nanosegundo 0 crece hasta estabilizarse en 2 Å, de igual manera esto se ve reflejado en las posiciones del ligando superpuestas en la **figura 29 (D)**, en donde el ligando en el nanosegundo 0 tiene una posición ligeramente distinta a la obtenida en las posiciones 50 y 100ns.

Las figuras 29 (E y F) muestran el comportamiento para α 2,6 unido a cSyd22, el ligando en este caso tuvo el valor más pequeño de todos, con un promedio de 0.7, reflejando una alta similitud de la posición del ligando a lo largo de la simulación.

Para el caso del receptor α 2,3 unido a la proteína cBri18, la **figura 29 (G)** muestra el RMSD, el cual presenta fluctuaciones a lo largo de la simulación que se mantuvieron por debajo de 2Å. Además, la proteína cBri18 también muestra valores estables inferiores a 4 Å.



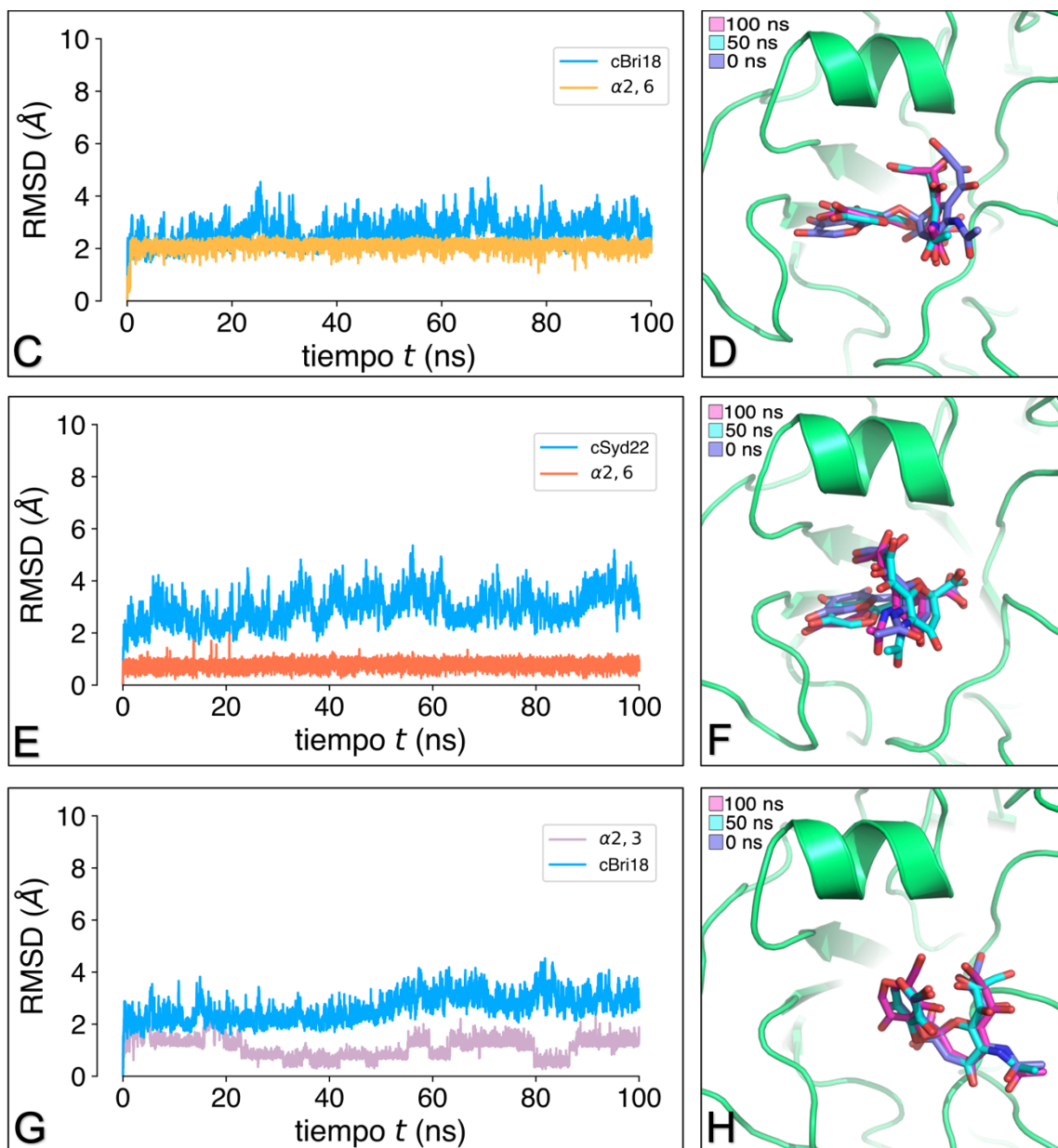


Figura 29. RMSD y poses del ligando a lo largo de la simulación: RMSD de ligandos proteínas durante los 100ns de los complejos: A) cCal09/ $\alpha 2, 6$, C) cBri18/ $\alpha 2, 6$, E) cSyd22/ $\alpha 2, 6$ y G) cBri18/ $\alpha 2, 3$ y las poses en 0, 50 y 100ns de los mismos complejos respectivamente en: B), D), F) y H).

Además, se realizó la medición del radio de giro de las proteínas unidas a cada uno de los receptores, como se muestra en la **figura 30 (A y B)**. En la **figura 30 (A)**, se observa que el radio de giro de las proteínas unidas al receptor humano se mantiene alrededor de 3.8 y 3.7 nm, con ligeras fluctuaciones. Por otro lado, en la **figura 30 (B)**, se muestra el radio de giro de la proteína cBri18 unida al receptor aviar, el cual presenta fluctuaciones entre 3.7 y 3.6 nm. Estos resultados evidencian que la unión de las proteínas a los receptores celulares no tuvo un impacto significativo en su compacidad.

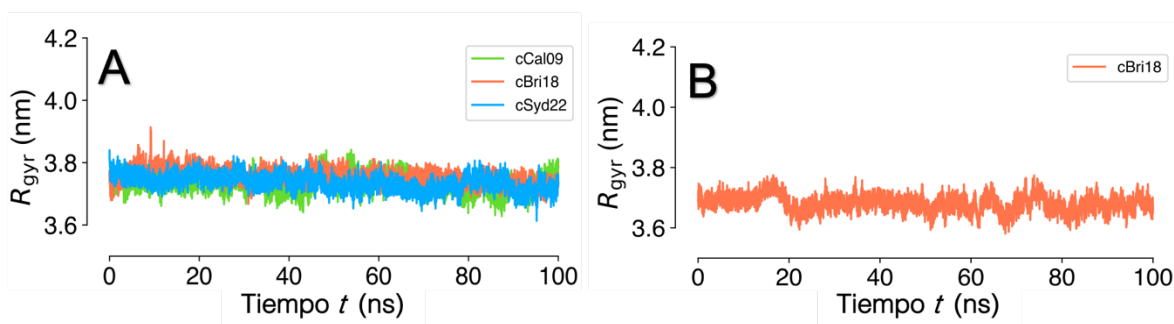


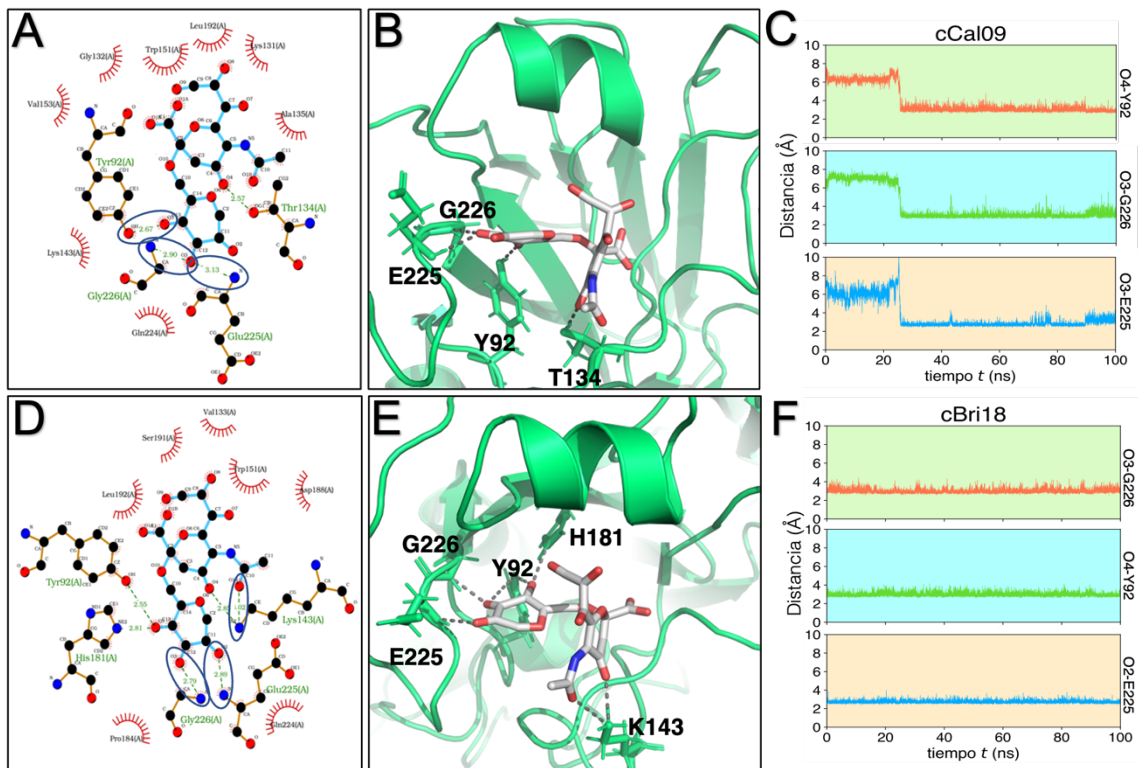
Figura 30. Radio de giro: Medición del radio de giro del esqueleto de las proteínas en complejo con A) el receptor $\alpha 2,6$ y B) $\alpha 2,3$ a lo largo de 100ns de producción en dinámica molecular.

Para verificar la estabilidad de los receptores celulares dentro de la cavidad del RBD de cada una de las proteínas, se midió la distancia promedio de los enlaces H entre los ligandos y residuos de la proteína con los que interactúan. Las **figuras 31 (A y B)** muestran los enlaces de hidrógeno que se formaron en el nanosegundo 100 entre el receptor humano y los residuos Y92, T134, E225 y G226 de la proteína cCal09. De estos, se eligieron tres enlaces para medir su distancia a lo largo de la simulación, los cuales se graficaron en la **figura 31 (C)**, allí se muestra que estos enlaces se generaron alrededor del nanosegundo 26, y a partir de ahí, permanecieron estables hasta llegar al nanosegundo 90, en donde los enlaces formados con los residuos E225 y G226 tuvieron un aumento en su fluctuación mientras que el enlace formado con Y92 tuvo una disminución de las fluctuaciones.

En las **figuras 31 (D y E)** se muestran los enlaces formados con el receptor humano y los residuos Y92, K143, H181, E225 y G226 de la proteína cBri18, de los cuales a 3 (resaltados

en la **figura 31 (D)** se les graficó la distancia en la **figura 31 (F)** durante los 100ns. Allí, se aprecia que esta distancia permaneció estable a desde el inicio de la simulación con una baja fluctuación.

Para los enlaces entre cSyd22 y el receptor humano, los residuos con los que se formó enlace son: Y92, H181, D223 y G226, de los cuales, los resaltados en la **figura 31 (G)** se les graficó la distancia en el tiempo en la **figura 31 (H)**, en donde se aprecia que las distancias son estables a lo largo del tiempo.



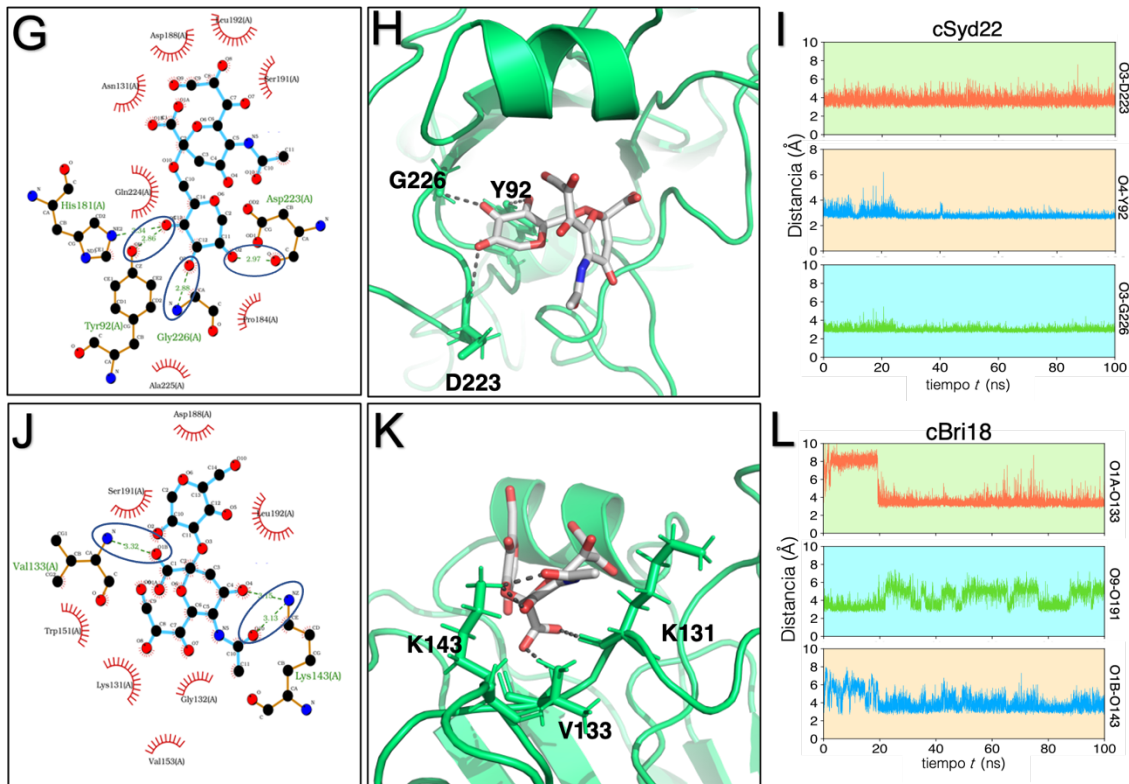


Figura 31. Gráficos de las interacciones. Diagrama de interacciones bidimensionales receptores con las proteínas (A, D, G y J) realizado con LigPlot. Pose tridimensional de los ligandos en el sitio de unión de las proteínas (B, E, H y K). Distancia de 3 enlaces H de durante 100ns de producción (C, F, I y L). Para los complejos cCal09/α2,6(A-C), cBri18/α2,6 (D-F), cSyd22/α2,6 (G-I) y cBri18/α2,3 (J-L).

Se hizo el mismo procedimiento para los enlaces formados entre el receptor aviar y los residuos: K131, V133 y K143 de la proteína cBri18 mostrados en las **figuras 31 (J y K)** y se grafican sus distancias en la **figura (L)**. Aquí se puede apreciar que a partir del nanosegundo 20 se formó un enlace con el residuo V133 a una distancia alrededor de los 3Å, mientras que las otras dos interacciones con 143 y 191, fueron menos estables a lo largo de los 100 nanosegundos. Adicionalmente, **La figura 31 (J)** muestra que el receptor aviar no formó interacciones con residuos pertenecientes al *loop* 220 como lo tuvo el receptor humano en los casos anteriores, sino que este formó interacciones principalmente con el *loop* 130 y con la hélice 190 (residuos del 188 al 190).

La figura 31 (E) muestra cómo el receptor humano está posicionado en la región central del sitio de unión de cBri18, en donde logra formar enlaces con los 3 dominios principales (*loop* 130, hélice 190 y *loop* 220), por el contrario, el receptor aviar no se encuentra

posicionado en el centro del sitio de unión de cBri18, sino que se encuentra desplazado hacia el *loop* 130 como se muestra en las **figuras 31 (J y K) y figura 29 (H)**, formando enlaces con los dominios *loop* 130 y hélice 190.

6.2.4 Energías de unión

La evaluación de la dinámica molecular de los compuestos como se ilustra en las **figuras 28 (A-B)**, sugirió que el receptor celular humano permaneció en el sitio de unión de cada una de las proteínas de manera estable, mientras que el receptor aviar permaneció en el RBS solo para el caso de la proteína cBri18, presentando fluctuaciones más significativas que el resto. Por lo tanto, la energía libre de unión MMGBSA fue calculado para estimar la unión de estos 4 complejos. La **figura 32 (A)** presenta los valores de energía durante los últimos 10ns, los cuales se muestran estables a lo largo del tiempo. Adicionalmente, se calcula el valor promedio de dicha energía y se grafica en la **figura 33 (B)**, la cual permite observar que el receptor aviar unido a la proteína cBri18 presentó un valor de energía de -14.58 kcal/mol, mientras que el receptor humano presentó valores de -30.06, -24.65 y -26.34 kcal/mol en unión con cCal09, cBri18 y cSyd22, respectivamente.

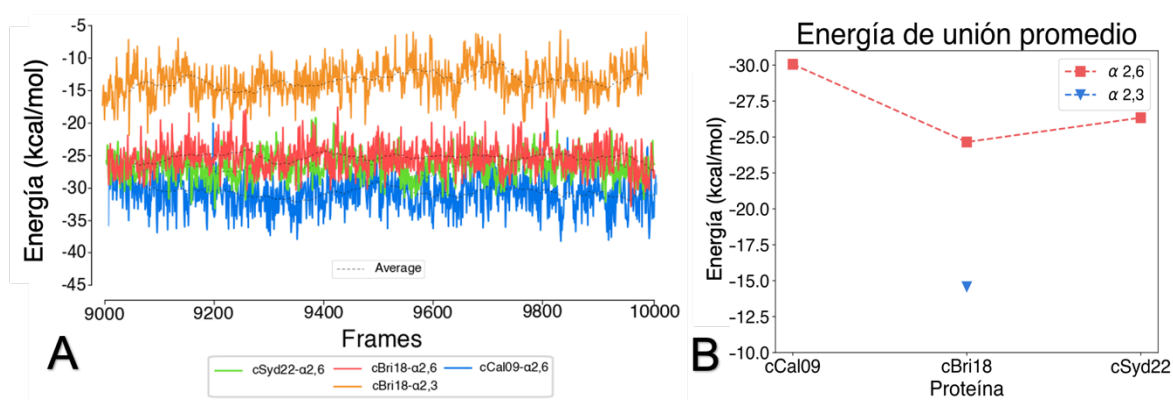


Figura 32. Energías de unión de los complejos: Medición de la energía de unión MM/GBSA: (A) durante los últimos 10ns (B) Promedio de los últimos 10ns. Entre los complejos cCal09/ α 2,6, cBri18/ α 2,6, cSyd22/ α 2,6, y cBri18/ α 2,3. Los cálculos se realizaron con el software *gmx_MMPBSA*.

En la **figura 33(A-B)** se muestra la descomposición por residuos de las energías de unión para cada uno de los aminoácidos con los que forma interacción, esta medición se hizo

para el RBS de los virus cCal09, cBri18 y cSyd22 en interacción con el receptor celular humano en la **figura 33 (A)** y el RBS de cBri18 en interacción con el receptor aviar en figura 34 (B). Un valor positivo indica que los residuos son desestabilizadores, mientras que el signo negativo indica que son estabilizadores. El análisis de los virus unidos al receptor humano muestra que los residuos: Y92, T134, W151, H181, L192, Q224, E225(A para cSyd22) y G226 presentaron la contribución más significativa (menor a -1 kcal/mol) en por lo menos uno de los virus.

Haciendo una comparación entre los residuos que tuvieron contribución con el receptor humano se observó que los residuos T134, W151 y Q224 presentaron una disminución más significativa de la energía de unión en los virus posteriores a cCal09 con valores de disminución alrededor de -2.5, 0.5 y 1.0 kcal/mol, respectivamente. Mientras que los residuos Y92 y L192 evidenciaron aumentos en la energía de unión.

En el caso del receptor aviar, el análisis mostró que no generó interacción significativa con residuos del *loop* 220, mientras que el receptor humano sí. Adicionalmente, se encontró que los residuos que generan contribución más significativa son K131, W151 y L192 con un valor de energía aproximado de 2.7, 1 y 2.7 kcal/mol, respectivamente, siendo K131 un residuo clave para la unión del virus cBri18 con el receptor aviar, mientras que con el receptor humano no.

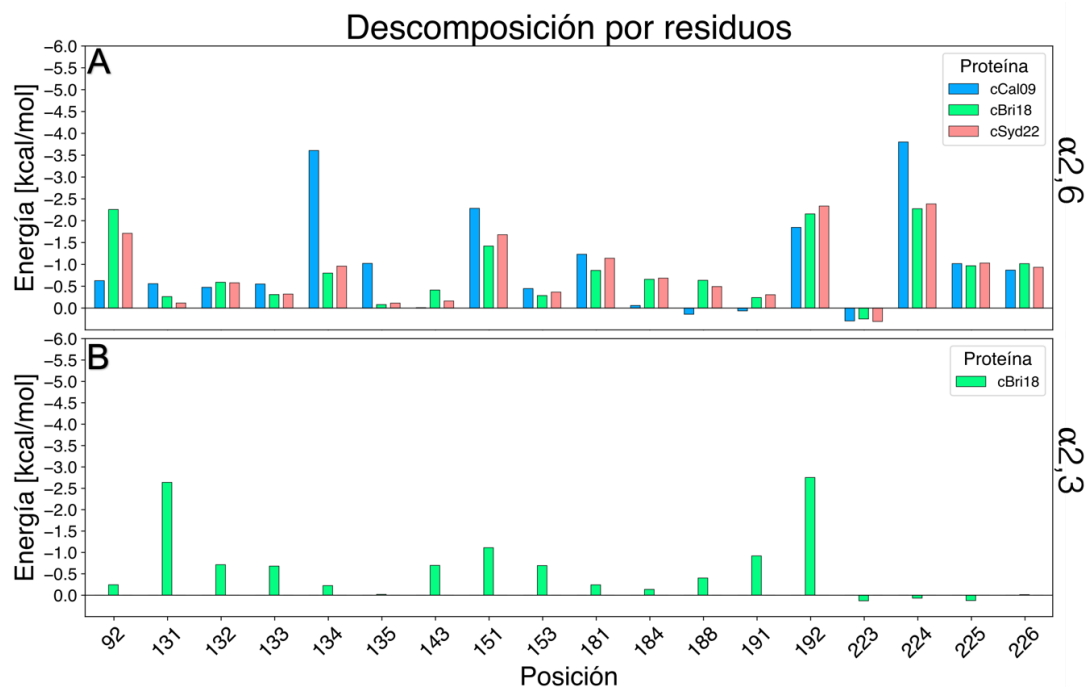


Figura 33. Descomposición de energía por residuos: Energía de unión MMGBSA promedio por residuos pertenecientes al RBD de las proteínas cCal09, cBri18 y cSyd22 con el receptor humano $\alpha 2,6$ en (A) y de cBri18 con el receptor aviar $\alpha 2,3$ en (B). Los residuos que no presentaron valores de energía no se mostraron. Los cálculos se realizaron con el software *gmx_MMPBSA* durante los últimos 10ns de la producción en dinámica molecular. Las posiciones que presentan mutación en alguna de las proteínas se muestran con un símbolo (*).

Estos análisis se realizan con un constante dieléctrica del soluto igual a 78.5, la cual corresponde al agua a temperatura ambiente, pero dado que se desconoce el valor de la constante dieléctrica de la hemaglutinina, los cálculos se realizaron con un valor de constante dieléctrica del soluto igual a 1, que es el valor que el software *gmx_MMPBSA* trae por defecto. Sin embargo, dada la alta sensibilidad que tiene la constante dieléctrica para el cálculo de las energías, se repitió el procedimiento con una constante de soluto igual a 2 para verificar que no existan cambios muy significativos que alteren los resultados encontrados anteriormente. Para esto, en la **figura 34** se muestran los valores promedios de la energía de unión para una constante dieléctrica de 1 y 2, permitiendo observar los cambios en los valores promedios de la energía de unión, un ejemplo de ello es la energía de unión entre cCal09 contra el receptor humano, con una constante dieléctrica de 1 el valor fue de -30, con una constante de 2 el valor fue de -36,54. Esto mismo sucedió en

todos los casos, donde se observó que los valores absolutos de la energía aumentaron, pero las tendencias encontradas inicialmente se mantuvieron para los valores de constante dieléctrica estudiados, como lo es el caso de las proteínas unidas al receptor humano. Por ejemplo, se evidenció que la proteína cCal09 siempre mantuvo el valor de energía de unión más grande de todos los complejos, seguido del complejo cSyd22 y finalmente cBri18. Otra tendencia que se mantuvo fue la de cBri18 en unión con los receptores aviar y humano, la cual fue superior en todos los casos en unión con el receptor humano.

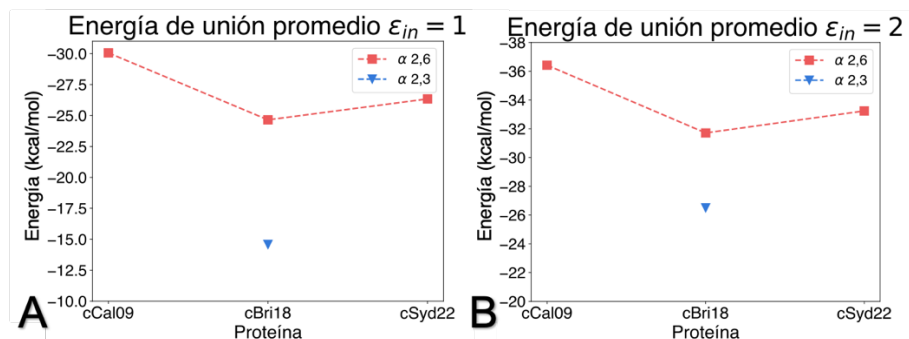


Figura 34. Energía de unión promedio a dos constantes dieléctricas. MMGBSA de los últimos 10ns de las proteínas cCal09, cBri18 y cSyd22 en unión con los receptores humano y aviar con una constante dieléctrica interna de 1 en (A), de 2 en (B).

Todos estos resultados mostrados anteriormente, confirman que los cambios conformacionales presentados en el *loop* 130 de la cepa viral cBri18 tuvieron efectos más notables en la interacción con el receptor aviar. Específicamente, se observó que el receptor aviar tuvo asentamiento más cerca del *loop* 130 que del *loop* 220, lo que otorgó mayor relevancia a los enlaces formados con los residuos del *loop* 130 y la *hélice* 190, así como la contribución energética de cada uno de ellos.

Adicionalmente, al comparar los resultados de las proteínas en unión con el receptor aviar, se hizo evidente que solo la cepa viral cBri18 logró establecer una unión con el receptor aviar. En este caso, el *loop* 130 desempeñó un papel clave para que dicha unión fuera efectiva. Por otro lado, las cepas virales cCal09 y cSyd22, las cuales no presentaron cambios conformacionales en el *loop* 130, no generaron interacciones lo suficientemente fuertes como para mantener los receptores en el sitio de unión.

6.3 Discusión

Evaluar las mutaciones en el dominio de unión al receptor de la hemaglutinina de los virus es crucial, ya que puede facilitar la transmisión entre aves, cerdos y humanos, impulsar la evolución viral hacia variantes más virulentas e influir en la infectividad. En Colombia no se han reportado estudios que revelen el estado actual de las cepas circulantes con respecto a los receptores. Por lo tanto, en esta investigación, se abordó el estudio de la interacción de los receptores humano SA α 2,6 y aviar SA α 2,3 con 3 proteínas de interés es Colombia (cCal09, cBri18 y cSyd22), estas proteínas hacen parte de un grupo de 5 proteínas que representan la diversidad de HA VI de linaje H1N1pdm circulantes en Colombia, pero fueron elegidas dadas características que las distinguieron de las otras.

Este estudio reveló que el receptor aviar SA α 2,3 no logró formar interacción efectiva con las proteínas cCal09 y cSyd22, abandonando RBD de estas proteínas como se observa en las **figuras 28 (A-B)**, siendo cBri18 la única proteína que mantuvo el receptor aviar en el sitio de unión a lo largo de toda la simulación, mientras que el receptor humano SA α 2,6 sí logró formar interacción efectiva en todos los casos. Adicionalmente, al analizar la energía de unión de cada una de estas proteínas con los receptores que permanecieron, como se ilustra en la **figura 33**, se pudo observar que el receptor humano SA α 2,6 presentó valores de afinidad superiores al receptor aviar, lo que sugiere que las proteínas evaluadas presentan preferencia por el receptor humano que el receptor aviar. Esta información es coherente con lo reportado en (Bradley et al., 2011; Carbone et al., 2015; L. M. Chen et al., 2011; Childs et al., 2009; Maines et al., 2009; Yang et al., 2010), en donde se encontró que el virus de origen pandémico A/California/07/09 presentaba preferencia por el receptor humano sobre el receptor aviar. Lo que permite observar que la tendencia por preferencia hacia el receptor humano sobre el aviar se sigue manteniendo en los virus de linaje pandémico evaluados en este estudio.

Adicionalmente, en este estudio se comparó la energía entre las proteínas consenso que se unieron al receptor humano SA α 2,6 y se encontró que la proteína cCal09, que como se mencionó en el **capítulo 2**, es el consenso que representa el 57% de las cepas virales H1N1pdm en Colombia, tuvo el valor de energía más alto como se mostró en la **figura 32**

(A y B), seguida de la cepa cSyd22 que corresponde a la cepa que circula en años actuales y finalmente cBri18. Esto podría implicar que las cepas que se representan con el consenso cCal09 pueden ser aquellas que tienen capacidad de infección y transmisión más elevada, mientras que cepas más actuales, representadas en cBri18 y cSyd22, pudieron haber disminuido la transmisibilidad o infectividad hacia receptores humano. Sin embargo, estudios posteriores son requeridos para evaluar esta hipótesis.

Las mutaciones que han acumulado las cepas cBri18 y cSyd22, con respecto a cCal09, las han llevado a disminuir su afinidad por el receptor humano como se mencionó en el párrafo anterior. A la hora de evaluar si las mutaciones presentes en el RBD (ver **figura 19**) tienen un efecto directo en la contribución de esta afinidad (ver **figura 33**), se encontró que estas mutaciones no tuvieron un impacto directo, sugiriendo que el impacto generado sobre la afinidad por SA α 2,6 se dio de manera indirecta, generando que posiciones no mutadas como: T134, W151, L192 y Q224, sean las más destacadas en la interacción entre las proteínas evaluadas y el receptor SA α 2,6, en especial para cCal09.

También se pudo observar que, a pesar de que la proteína cBri18 tuvo cambios conformacionales en el *loop* 130 del RBD (**capítulo 2**), estos no fueron muy significativo en términos de la unión con el receptor SA α 2,6, dado que, como se muestra en las **figuras 32 y 33 (A)** la energía total y la contribución energética fue muy similar a la observada en cSyd22.

Sin embargo, los análisis realizados entre cBri18 y el receptor SA α 2,3, en comparación con el receptor SA α 2,6, podrían dar pie a que estos cambios conformacionales sí tuvieron efectos en la unión con SA α 2,3.

Primero, pudo haber tenido efecto en el asentamiento del receptor SA α 2,3 en el RBD de cBri18 y las regiones con quién formó enlace. Como se ilustra en las **figuras 31 (D y E) y 29 (D)**, el receptor SA α 2,6 se mantuvo en el centro del RBD de cBri18, formando enlaces con las tres regiones, *loop* 130, hélice 190 y *loop* 220, mientras que el receptor SA α 2,3 se desplazó en dirección del *loop* 130, formando enlaces principalmente con los residuos del *loop* 130 y la hélice 190, pero no con el *loop* 220 como se muestra en la **figura 31 (K) y 29 (H)**. Esto se reforzó con lo encontrado en un estudio previo, en donde Según Ni y colaboradores, los cambios conformacionales que sufre el *loop* 130 determinan las

diferentes posiciones de asiento de los receptores unidos dentro RBD del receptor de los HA.

Segundo, pudo haber tenido efecto en la unión con el receptor SA α 2,3. **La figura 28 y 12 (B)** muestra que SA α 2,3 no formó interacción efectiva con cCal09 y cSyd22, pero con cBri18 sí, esto también es reforzado por Ni y colaboradores, dado que ellos encontraron en el mismo estudio que los cambios conformacionales del *loop* 130 puede alterar la uniones y la especificidad por los receptores. Dado que, como se mostró en el **capítulo 2**, estos cambios conformacionales no son presentados por cCal09 y cSyd22, se sugiere que esta pudo haber sido la razón por la que el receptor SA α 2,3 formó interacción con cBri18, pero no con cCal09 y cSyd22.

Al retomar la hipótesis planteada en el **capítulo 2**, en donde se propuso que las mutaciones únicas que presenta cBri18 (R46G, P283A y I299V, ver **tabla 4**) podrían tener relación con los cambios conformacionales que esta presentó, se sugiere que dichas mutaciones también podrían estar ocasionando un incremento en la afinidad por el receptor aviar SA α 2,3. Al hacer una búsqueda en la literatura en la base de datos de Flusurver acerca de la relación de estas mutaciones con la alteración de especificidad por el receptor, solo se encontró que según Huang y colaboradores la mutación R46G se encontró en un grupo de mutaciones que permitieron que virus de H7N9 se unieran con mayor afinidad al receptor SA α 2,6, mostrando que esta mutación está relacionada con la alteración de unión a receptores (Huang et al., 2017). Aun así, estudios adicionales son necesarios para evaluar la presencia de las mutaciones en relación al adaptación al receptor aviar SA α 2,3.

En síntesis, en este estudio se evaluó el efecto de las mutaciones en la energía de las proteínas consenso de interés del linaje H1N1pdm en Colombia denominadas cCal09, cBri18 y cSyd22 en complejo con los receptores celulares de ave SA α 2,3 y humano SA α 2,6. Los cálculos de la energía de unión calculada con MMGBSA muestra que cBri18 y cSyd22 disminuyeron la afinidad por SA α 2,6 en comparación con cCal09, una cepa altamente homologa con la cepa de referencia A/California/07/2009, pero no se pudo establecer qué mutaciones generaron dicha disminución. Por el contrario, sí se lograron identificar 3 mutaciones que pueden haber generado un incremento (R46G, P283A y

I299V) por la afinidad aviar a cBri18, a pesar de que estas no se encuentran en el RBD. Finalmente, todas las proteínas evaluadas presentaron mayor afinidad por SA α 2,6 que por SA α 2,3, sugiriendo que el linaje pandémico sigue teniendo preferencia por el receptor humano SA α 2,6 que por el aviar SA α 2,3.

7. Conclusiones

1. La caracterización de las mutaciones de la HA del VI pandémico reportados en Colombia, permitió su clasificación según su significancia, frecuencia y prevalencia en el país. Demostrando que las mutaciones de alta frecuencia con baja significancia (I322V, P84S, S204T, K44R, N130S, A187T, S186T, S452N, K164Q, K284E, A356T, E500K, S85N, I217T S75R, S165T y I296V) posiblemente han favorecido el establecimiento del virus en esta región, mientras que las mutaciones de alta significancia con baja frecuencia (D223G, Q224R y Q194R) al parecer no ha aportado al fitness de estos virus. Por otro lado, las mutaciones de alta frecuencia con nivel de significancia medio (E375K y S163N) pueden estar contribuyendo al éxito del virus, favoreciendo el aumento de la virulencia y la capacidad de evadir la respuesta inmune.
2. La metodología de modelamiento estructural del VI aplicada en este estudio demostró que permite hacer análisis de mutaciones de manera exitosa. Sirviendo de base para futuros estudios sobre la diversidad genética de las cepas circulantes del linaje H1N1pdm en Colombia y otras regiones del mundo. Estos modelos que incluyen las mutaciones más predominantes generaron la oportunidad de estudiar con mayor profundidad el efecto que dichas mutaciones tienen sobre la estructura viral a lo largo del tiempo. Destacando que los modelos cBri18 y cSyd22, que representaron las cepas que circulan actualmente, tuvieron una disminución en su afinidad por el receptor celular humano en comparación con la referencia cCal09,

lo que conlleva a la posibilidad de que estos virus puedan ampliar a futuro su preferencia de huésped hacia otras especies o de lo contrario perder esta habilidad.

3. Todas las proteínas evaluadas de los virus repostados para Colombia demostraron una mayor afinidad por el receptor humano, lo que sugiere que los virus pertenecientes al linaje pandémico aún mantienen su preferencia por el huésped humano.
4. El modelo cBri18 fue el único que reportó afinidad por el receptor aviar SA α 2,3. Y que además puede que este fenómeno estuviera fuertemente influenciado por el cambio conformacional en el *loop* 130, probablemente provocado por las mutaciones únicas observadas (R46G, P283A y I299V), lo que sugiere que estas mutaciones podrían estar influenciando la preferencia de unión por el receptor aviar.

8. Anexos

Los anexos se adjuntan en una carpeta de drive para facilitar el acceso por parte de lector: https://drive.google.com/drive/folders/1dZapWX0my5jrD3pW4v73Y_AdOrPAMLIR?usp=sharing

- **Anexo 1. Árbol filogenético de máxima verosimilitud:** Árbol de máxima verosimilitud de HA de VI en Colombia con referencias de linajes Euroasiático, Humano estacionario, porcino clásico y pandémico generado en IQtree con un modelo sustitucional GTR+I+G y 100 *bootstraps*. Las ramas que pertenecen a secuencias del estudio se marcan de color rojo y las de referencia de color gris.
- **Anexo 2. frecuencia de mutaciones:** La frecuencia de todas las mutaciones identificadas en la HA de los VI de linaje H1N1pdm en Colombia se muestran en frecuencias de mayor a menor. Adicionalmente estas mutaciones se colorean de acuerdo con las categorías mostradas en la **tabla 2**.
- **Anexo 3. Árbol bayesiano en escala de tiempo:** Filogenia del gen HA de los VI de linaje H1N1pdm en Colombia en el tiempo, este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. los tiempos del ancestro común más reciente entre H1-IAV se muestran en los nodos de rama.
- **Anexo 4. Árbol bayesiano con mutaciones:** Comparación de las relaciones filogenéticas de los virus y las todas las sustituciones adquiridas en el tiempo para las HA de los VI de linaje H1N1pdm en Colombia. Este árbol fue inferido usando análisis bayesiano y se construyó con Beast usando el método MCMC (*Markov Chain Monte Carlo*) con 100 millones de generaciones. Las mutaciones se colorean de acuerdo su categoría descrita en la **tabla 2**.
- **Anexo 5. Alineamiento de secuencias consenso:** Muestra el alineamiento de 5 secuencias consenso cCal09, cMi15, cBri18, cGM19 y cSyd22 calculados a partir de 5 *clústeres* formados por las HA de los VI de linaje H1N1pdm en Colombia.

Bibliografía

- Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/J.SOFTX.2015.06.001>
- Al Khatib, H. A., Al Thani, A. A., & Yassine, H. M. (2018a). Evolution and dynamics of the pandemic H1N1 influenza hemagglutinin protein from 2009 to 2017. *Archives of Virology*, 163(11), 3035–3049. <https://doi.org/10.1007/S00705-018-3962-Z/FIGURES/5>
- Alhossary, A., Handoko, S. D., Mu, Y., & Kwoh, C. K. (2015). Fast, accurate, and reliable molecular docking with QuickVina 2. *Bioinformatics*, 31(13), 2214–2216. <https://doi.org/10.1093/BIOINFORMATICS/BTV082>
- Anderson, T. K., Macken, C. A., Lewis, N. S., Scheuermann, R. H., Reeth, K. Van, Brown, I. H., Swenson, S. L., Simon, G., Saito, T., Berhane, Y., Ciacci-Zanella, J., Pereda, A., Davis, C. T., Donis, R. O., Webby, R. J., & Vincent, A. L. (2016). A Phylogeny-Based Global Nomenclature System and Automated Annotation Tool for H1 Hemagglutinin Genes from Swine Influenza A Viruses. *mSphere*, 1(6). <https://doi.org/10.1128/MSPHERE.00275-16>
- Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O., Huelsenbeck, J. P., Ronquist, F., Swofford, D. L., Cummings, M. P., Rambaut, A., & Suchard, M. A. (2012). BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. *Systematic Biology*, 61(1), 170. <https://doi.org/10.1093/SYSBIO/SYR100>
- Bawono, P., & Heringa, J. (2014). Phylogenetic Analyses. *Comprehensive Biomedical Physics*, 6, 93–110. <https://doi.org/10.1016/B978-0-444-53632-7.01108-4>
- BII Flusurver - Frequently Asked Questions*. (s/f). Recuperado el 2 de julio de 2023, de <https://flusurver.bii.a-star.edu.sg/help/faq.html>
- Boni, M. F., Galvani, A. P., Wickelgren, A. L., & Malani, A. (2013). Economic epidemiology of avian influenza on smallholder poultry farms. *Theoretical Population Biology*, 90, 135–144. <https://doi.org/10.1016/j.tpb.2013.10.001>
- Bouvier, N. M., & Palese, P. (2008). THE BIOLOGY OF INFLUENZA VIRUSES. *Vaccine*.
- Boyoglu-Barnum, S., Ellis, D., Gillespie, R. A., Hutchinson, G. B., Park, Y. J., Moin, S. M., Acton, O. J., Ravichandran, R., Murphy, M., Pettie, D., Matheson, N., Carter, L.,

- Creanga, A., Watson, M. J., Kephart, S., Ataca, S., Vaile, J. R., Ueda, G., Crank, M. C., ... Kanekiyo, M. (2021). Quadrivalent influenza nanoparticle vaccines induce broad protection. *Nature*, *592*(7855), 623–628. <https://doi.org/10.1038/S41586-021-03365-X>
- Bradley, K. C., Jones, C. A., Tompkins, S. M., Tripp, R. A., Russell, R. J., Gramer, M. R., Heimbürg-Molinaro, J., Smith, D. F., Cummings, R. D., & Steinhauer, D. A. (2011). Comparison of the receptor binding properties of contemporary swine isolates and early human pandemic H1N1 isolates (Novel 2009 H1N1). *Virology*, *413*(2), 169–182. <https://doi.org/10.1016/J.VIROL.2011.01.027>
- Brice, A. R., & Dominy, B. N. (2011). Analyzing the robustness of the MM/PBSA free energy calculation method: Application to DNA conformational transitions. *Journal of Computational Chemistry*, *32*(7), 1431–1440. <https://doi.org/10.1002/JCC.21727>
- Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of chemical physics*, *126*(1). <https://doi.org/10.1063/1.2408420>
- Byrd-Leotis, L., Cummings, R. D., & Steinhauer, D. A. (2017). *Molecular Sciences The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase*. <https://doi.org/10.3390/ijms18071541>
- Cador, C., Andraud, M., Willem, L., & Rose, N. (2017). Control of endemic swine flu persistence in farrow-to-finish pig farms: A stochastic metapopulation modeling assessment. *Veterinary Research*, *48*(1), 1–14. <https://doi.org/10.1186/S13567-017-0462-1/FIGURES/7>
- Carbone, V., Schneider, E. K., Rockman, S., Baker, M., Huang, J. X., Ong, C., Cooper, M. A., Yuriev, E., Li, J., & Velkov, T. (2015). Molecular Characterisation of the Haemagglutinin Glycan-Binding Specificity of Egg-Adapted Vaccine Strains of the Pandemic 2009 H1N1 Swine Influenza A Virus. *Molecules*, *20*(6), 10415. <https://doi.org/10.3390/MOLECULES200610415>
- Chen, L. M., Blixt, O., Stevens, J., Lipatov, A. S., Davis, C. T., Collins, B. E., Cox, N. J., Paulson, J. C., & Donis, R. O. (2012). In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology*, *422*(1), 105. <https://doi.org/10.1016/J.VIROL.2011.10.006>
- Chen, L. M., Rivaller, P., Hossain, J., Carney, P., Balish, A., Perry, I., Davis, C. T., Garten, R., Shu, B., Xu, X., Klimov, A., Paulson, J. C., Cox, N. J., Swenson, S., Stevens, J., Vincent, A., Gramer, M., & Donis, R. O. (2011). Receptor specificity of subtype H1 influenza A viruses isolated from swine and humans in the United States. *Virology*, *412*(2), 401. <https://doi.org/10.1016/J.VIROL.2011.01.015>

- Chen, Z., Wang, W., Zhou, H., Amorsolo L. Suguitan, Jr., Shambaugh, C., Kim, L., Zhao, J., Kemble, G., & Jin, H. (2010). Generation of Live Attenuated Novel Influenza Virus A/California/7/09 (H1N1) Vaccines with High Yield in Embryonated Chicken Eggs. *Journal of Virology*, 84(1), 44. <https://doi.org/10.1128/JVI.02106-09>
- Chepkwony, S., Parys, A., Vandoorn, E., Stadejek, W., Xie, J., King, J., Graaf, A., Pohlmann, A., Beer, M., Harder, T., & Van Reeth, K. (2021). Genetic and antigenic evolution of H1 swine influenza A viruses isolated in Belgium and the Netherlands from 2014 through 2019. *Scientific Reports 2021 11:1*, 11(1), 1–12. <https://doi.org/10.1038/s41598-021-90512-z>
- Childs, R. A., Palma, A. S., Wharton, S., Matrosovich, T., Liu, Y., Chai, W., Campanero-Rhodes, M. A., Zhang, Y., Eickmann, M., Kiso, M., Hay, A., Matrosovich, M., & Feizi, T. (2009). Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. *Nature biotechnology*, 27(9), 797. <https://doi.org/10.1038/NBT0909-797>
- Chua, K., & Chai, H. (2012). Hemagglutinin protein of Asian strains of human influenza virus A H1N1 binds to sialic acid—a major component of human airway receptors. *Genetics and Molecular Research*, 11(1), 636–643. <https://doi.org/10.4238/2012.March.16.1>
- Chutinimitkul, S., Herfst, S., Steel, J., Lowen, A. C., Ye, J., van Riel, D., Schrauwen, E. J. A., Bestebroer, T. M., Koel, B., Burke, D. F., Sutherland-Cash, K. H., Whittleston, C. S., Russell, C. A., Wales, D. J., Smith, D. J., Jonges, M., Meijer, A., Koopmans, M., Rimmelzwaan, G. F., ... Fouchier, R. A. M. (2010). Virulence-Associated Substitution D222G in the Hemagglutinin of 2009 Pandemic Influenza A(H1N1) Virus Affects Receptor Binding. *Journal of Virology*, 84(22), 11802–11813. https://doi.org/10.1128/JVI.01136-10/SUPPL_FILE/27_8_10_REVISSED_SOM_JVI01136_10.DOC
- Consuelo Ramirez-Nieto, G., Augusto, C., Rojas, D., Julio, V., Alfonso, V., Correa, J., Dario, J., & Galvis, M. (2012). *First isolation and identification of H1N1 swine influenza viruses in Colombian pig farms*. 4, 983–990. <https://doi.org/10.4236/health.2012.430150>
- Cotter, C. R., Jin, H., & Chen, Z. (2014). A Single Amino Acid in the Stalk Region of the H1N1pdm Influenza Virus HA Protein Affects Viral Fusion, Stability and Infectivity. *PLoS Pathogens*, 10(1). <https://doi.org/10.1371/JOURNAL.PPAT.1003831>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089–10092. <https://doi.org/10.1063/1.464397>

- Dou, D., Revol, R., Östbye, H., Wang, H., & Daniels, R. (2018). Influenza A virus cell entry, replication, virion assembly and movement. En *Frontiers in Immunology* (Vol. 9, Número JUL). Frontiers Media S.A. <https://doi.org/10.3389/fimmu.2018.01581>
- Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, 29(8), 1969. <https://doi.org/10.1093/MOLBEV/MSS075>
- DuBois, R. M., Aguilar-Yañez, J. M., Mendoza-Ochoa, G. I., Oropeza-Almazán, Y., Schultz-Cherry, S., Alvarez, M. M., White, S. W., & Russell, C. J. (2011). The Receptor-Binding Domain of Influenza Virus Hemagglutinin Produced in *Escherichia coli* Folds into Its Native, Immunogenic Structure. *Journal of Virology*, 85(2), 865–872. <https://doi.org/10.1128/jvi.01412-10>
- Fiser, A., & Šali, A. (2003). Modeller: Generation and Refinement of Homology-Based Protein Structure Models. *Methods in Enzymology*, 374, 461–491. [https://doi.org/10.1016/S0076-6879\(03\)74020-8](https://doi.org/10.1016/S0076-6879(03)74020-8)
- Forli, S., Huey, R., Pique, M. E., Sanner, M. F., Goodsell, D. S., & Olson, A. J. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. *Nature Protocols* 2016 11:5, 11(5), 905–919. <https://doi.org/10.1038/nprot.2016.051>
- Fraser, C., Cummings, D. A. T., Klinkenberg, D., Burke, D. S., & Ferguson, N. M. (2011). *Special Article Influenza Transmission in Households During the 1918 Pandemic*. 174(5). <https://doi.org/10.1093/aje/kwr122>
- Gao, S., Anderson, T. K., Walia, R. R., Dorman, K. S., Janas-Martindale, A., & Vincent, A. L. (2017). The genomic evolution of H1 influenza A viruses from swine detected in the United States between 2009 and 2016. *The Journal of General Virology*, 98(8), 2001. <https://doi.org/10.1099/JGV.0.000885>
- Global Influenza Programme*. (s/f). Recuperado el 2 de julio de 2023, de <https://www.who.int/teams/global-influenza-programme/vaccines/who-recommendations>
- Gorbalenya, A. E., & Lauber, C. (2017). Phylogeny of Viruses. *Reference Module in Biomedical Sciences*. <https://doi.org/10.1016/B978-0-12-801238-3.95723-4>
- Graaf Miranda, & Fouchier Ron A. (2014). Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO*, 33(8), 823–841. <https://doi.org/https://doi.org/10.1002/emboj.201387442>
- Guedes, I. A., Costa, L. S. C., dos Santos, K. B., Karl, A. L. M., Rocha, G. K., Teixeira, I. M., Galheigo, M. M., Medeiros, V., Krempser, E., Custódio, F. L., Barbosa, H. J. C., Nicolás, M. F., & Dardenne, L. E. (2021). Drug design and repurposing with

- DockThor-VS web server focusing on SARS-CoV-2 therapeutic targets and their non-synonym variants. *Scientific Reports* 2021 11:1, 11(1), 1–20. <https://doi.org/10.1038/s41598-021-84700-0>
- Guldemir, D., Coskun-Ari, F. F., Altas, A. B., Bakkaloglu, Z., Unaldi, O., Bayraktar, F., Korukluoglu, G., Aktas, A. R., & Durmaz, R. (2019). Molecular characterization of the influenza A(H1N1)pdm09 isolates collected in the 2015-2016 season and comparison of HA mutations detected in Turkey since 2009. *Journal of Medical Virology*, 91(12), 2074–2082. <https://doi.org/10.1002/JMV.25565>
- Guvench, O., & MacKerell, A. D. (2008). Comparison of protein force fields for molecular dynamics simulations. *Methods in Molecular Biology*, 443, 63–88. https://doi.org/10.1007/978-1-59745-177-2_4/COVER
- Han, Y., Sun, N., Lv, Q. yue, Liu, D. hong, & Liu, D. peng. (2016). Molecular epidemiology and phylogenetic analysis of HA gene of influenza A(H1N1)pdm09 strain during 2010–2014 in Dalian, North China. *Virus Genes*, 52(5), 606–612. <https://doi.org/10.1007/S11262-016-1358-2/FIGURES/2>
- Hanssen, H., Hincapié, O., & López, J. H. (1977). *INFLUENZA EN PORCINOS DE ANTIOQUIA, COLOMBIA* .
- Hollingsworth, S. A., & Dror, R. O. (2018). Review Molecular Dynamics Simulation for All. *Neuron*, 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
- Horimoto, T., & Kawaoka, Y. (2005). *INFLUENZA: LESSONS FROM PAST PANDEMICS, WARNINGS FROM CURRENT INCIDENTS*. <https://doi.org/10.1038/nrmicro1208>
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple AMBER force fields and development of improved protein backbone parameters. *Proteins*, 65(3), 712. <https://doi.org/10.1002/PROT.21123>
- Hou, T., Wang, J., Li, Y., & Wang, W. (2011b). Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of chemical information and modeling*, 51(1), 69. <https://doi.org/10.1021/CI100275A>
- Huang, D. T. N., Lu, C. Y., Chi, Y. H., Li, W. L., Chang, L. Y., Lai, M. J., Chen, J. S., Hsu, W. M., & Huang, L. M. (2017). Adaptation of influenza A (H7N9) virus in primary human airway epithelial cells. *Scientific Reports*, 7(1). <https://doi.org/10.1038/S41598-017-10749-5>

- Ivan, F. X., Zhou, X., Lau, S. H., Rashid, S., Teo, J. S. M., Lee, H. K., Koay, E. S., Chan, K. P., Leo, Y. S., Chen, M. I. C., Kwoh, C. K., & Chow, V. T. (2020). Molecular insights into evolution, mutations and receptor-binding specificity of influenza A and B viruses from outpatients and hospitalized patients in Singapore. *International Journal of Infectious Diseases*, 90, 84–96. <https://doi.org/10.1016/J.IJID.2019.10.024>
- Javanian, M., Barary, M., Ghebrehewet, S., Koppolu, V., Vasigala, V. K. R., & Ebrahimpour, S. (2021). A brief review of influenza virus infection. En *Journal of Medical Virology* (Vol. 93, Número 8, pp. 4638–4646). John Wiley and Sons Inc. <https://doi.org/10.1002/jmv.26990>
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), 926–935. <https://doi.org/10.1063/1.445869>
- Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics* 2020 21:7, 21(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>
- Karlsson, E. A., Ciuderis, K., Freiden, P. J., Seufzer, B., Jones, J. C., Johnson, J., Parra, R., Gongora, A., Cardenas, D., Barajas, D., Osorio, J. E., & Schultz-Cherry, S. (2013). Prevalence and characterization of influenza viruses in diverse species in Los Llanos, Colombia. *Emerging Microbes and Infections*, 2. <https://doi.org/10.1038/emi.2013.20>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166. <https://doi.org/10.1093/BIB/BBX108>
- Kim, H., Webster, R. G., & Webby, R. J. (2018). Influenza Virus: Dealing with a Drifting and Shifting Pathogen. *Viral Immunology*, 31(2), 174–183. <https://doi.org/10.1089/vim.2017.0141>
- Kitchen, D. B., Decornez, H., Furr, J. R., & Bajorath, J. (2004). Docking and scoring in virtual screening for drug discovery: Methods and applications. En *Nature Reviews Drug Discovery* (Vol. 3, Número 11, pp. 935–949). <https://doi.org/10.1038/nrd1549>
- Klement, E., Weng, H.-Y., Poljak, Z., Orlando, F., Pardo, C., Alba-Casals, A., Nerem, J., Morrison, R. B., Puig, P., & Torremorell, M. (2017). *influenza herd-level Prevalence and seasonality in Breed-to-Wean Pig Farms in the Midwestern United states*. 4, 11. <https://doi.org/10.3389/fvets.2017.00167>
- Koel, B. F., Burke, D. F., Bestebroer, T. M., Van Der Vliet, S., Zondag, G. C. M., Vervaeet, G., Skepner, E., Lewis, N. S., Spronken, M. I. J., Russell, C. A., Eropkin, M. Y., Hurt, A. C., Barr, I. G., De Jong, J. C., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., Fouchier, R. A. M., & Smith, D. J. (2013). Substitutions near the receptor binding site

- determine major antigenic change during influenza virus evolution. *Science*, 342(6161), 976–979.
https://doi.org/10.1126/SCIENCE.1244730/SUPPL_FILE/KOEL.SM.PDF
- Koul, P. A., Mir, M. A., Bali, N. K., Chawla-Sarkar, M., Sarkar, M., Kaushik, S., Khan, U. H., Ahmad, F., Garten, R., Lal, R. B., & Broor, S. (2011). Pandemic and seasonal influenza viruses among patients with acute respiratory illness in Kashmir (India). *Influenza and Other Respiratory Viruses*, 5(6), e521. <https://doi.org/10.1111/J.1750-2659.2011.00261.X>
- Kuhner, M. K. (2009). Coalescent genealogy samplers: windows into population history. *Trends in ecology & evolution*, 24(2), 86.
<https://doi.org/10.1016/J.TREE.2008.09.007>
- Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A Geometric Approach to Macromolecule-Ligand Interactions. En *J. Mol. Bid.*
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, 26(2), 283–291.
<https://doi.org/10.1107/S0021889892009944>
- Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: Multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling*, 51(10), 2778–2786. https://doi.org/10.1021/CI200227U/ASSET/IMAGES/MEDIUM/CI-2011-00227U_0006.GIF
- Lee, A. N., Hartono, Y. D., Sun, T., Leow, M. L., Liu, X. W., Huang, X., & Zhang, D. (2011). Molecular dynamics studies of human receptor molecule in hemagglutinin of 1918 and 2009 H1N1 influenza viruses. *Journal of Molecular Modeling*, 17(7), 1635–1641. <https://doi.org/10.1007/S00894-010-0867-5/METRICS>
- Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. <https://doi.org/10.1093/NAR/GKAB301>
- Levitt, M. (2014). Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture). *Angewandte Chemie International Edition*, 53(38), 10006–10018.
<https://doi.org/10.1002/ANIE.201403691>
- Maines, T. R., Jayaraman, A., Belser, J. A., Wadford, D. A., Pappas, C., Zeng, H., Gustin, K. M., Pearce, M. B., Viswanathan, K., Shriver, Z. H., Raman, R., Cox, N. J., Sasisekharan, R., Katz, J. M., & Tumpey, T. M. (2009). Transmission and

- Pathogenesis of Swine-Origin 2009 A(H1N1) Influenza Viruses in Ferrets and Mice. *Science (New York, N. Y.)*, 325(5939), 484.
<https://doi.org/10.1126/SCIENCE.1177238>
- Mark Berg Jeremy, Stryer Lubert, & Tymoczko John. (2006). *Biochemistry* (6th ed.).
<https://books.google.com.cu/books?id=HRr4MNH2YssC&printsec=frontcover#v=onepage&q&f=false>
- Matrosovich, M., Tuzikov, A., Bovin, N., Gambaryan, A., Klimov, A., Castrucci, M. R., Donatelli, I., & Kawaoka, Y. (2000). Early Alterations of the Receptor-Binding Properties of H1, H2, and H3 Avian Influenza Virus Hemagglutinins after Their Introduction into Mammals. *Journal of Virology*, 74(18), 8502.
<https://doi.org/10.1128/JVI.74.18.8502-8512.2000>
- McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E., & McKimm-Breschkin, J. L. (2019). Influenza virus neuraminidase structure and functions. En *Frontiers in Microbiology* (Vol. 10, Número JAN). Frontiers Media S.A.
<https://doi.org/10.3389/fmicb.2019.00039>
- McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, 32(Web Server issue), W20.
<https://doi.org/10.1093/NAR/GKH435>
- Michaelis, M., Doerr, H. W., & Cinatl, J. (2009). An influenza A H1N1 virus revival - Pandemic H1N1/09 virus. En *Infection* (Vol. 37, Número 5, pp. 381–389).
<https://doi.org/10.1007/s15010-009-9181-5>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., Lanfear, R., & Teeling, E. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/MOLBEV/MSAA015>
- Munjal, G., Hanmandlu, M., & Srivastava, S. (2019). Phylogenetics Algorithms and Applications. *Ambient Communications and Computer Systems*, 904, 187.
https://doi.org/10.1007/978-981-13-5934-7_17
- Ni, F., Kondrashkina, E., & Wang, Q. (2018). Determinant of receptor-preference switch in influenza hemagglutinin. *Virology*, 513, 98.
<https://doi.org/10.1016/J.VIROL.2017.10.010>
- Osorio-Zambrano, W. F., Ospina-Jimenez, A. F., Alvarez-Munoz, S., Gomez, A. P., & Ramirez-Nieto, G. C. (2022). Zooming in on the molecular characteristics of swine influenza virus circulating in Colombia before and after the H1N1pdm09 virus. *Frontiers in Veterinary Science*, 9.
<https://doi.org/10.3389/FVETS.2022.983304/FULL>

- Otte, J., Hinrichs, J., Rushton, J., Roland-Holst, D., & Zilberman, D. (2008). Impacts of avian influenza virus on animal production in developing countries. En *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources* (Vol. 3). <https://doi.org/10.1079/PAVSNNR20083080>
- Parrinello, M., Rahman, A., Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *JAP*, *52*(12), 7182–7190. <https://doi.org/10.1063/1.328693>
- Patel, H., & Kukol, A. (2021). Integrating molecular modelling methods to advance influenza A virus drug discovery. *Drug Discovery Today*, *26*(2), 503–510. <https://doi.org/10.1016/J.DRUDIS.2020.11.014>
- Petrova, V. N., & Russell, C. A. (2017). The evolution of seasonal influenza viruses. *Nature Publishing Group*, *16*. <https://doi.org/10.1038/nrmicro.2017.118>
- Rajao, D. S., Anderson, T. K., Kitikoon, P., Stratton, J., Lewis, N. S., & Vincent, A. L. (2018). Antigenic and genetic evolution of contemporary swine H1 influenza viruses in the United States. *Virology*, *518*, 45–54. <https://doi.org/10.1016/J.VIROL.2018.02.006>
- Rajapaksha, H., Petrovsky, N., & Guan, Y. (2014). In Silico Structural Homology Modelling and Docking for Assessment of Pandemic Potential of a Novel H7N9 Influenza Virus and Its Ability to Be Neutralized by Existing Anti-Hemagglutinin Antibodies. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0102618>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901. <https://doi.org/10.1093/SYSBIO/SYY032>
- Rambo-Martin, B. L., Keller, M. W., Wilson, M. M., Nolting, J. M., Anderson, T. K., Vincent, A. L., Bagal, U. R., Jang, Y., Neuhaus, E. B., Davis, C. T., Bowman, A. S., Wentworth, D. E., & Barnes, J. R. (2020). Influenza A Virus Field Surveillance at a Swine-Human Interface. *mSphere*, *5*(1). <https://doi.org/10.1128/MSPHERE.00822-19/ASSET/43499F7B-FAE5-42D2-BBA4-4D2CB9F4741F/ASSETS/GRAPHIC/MSPHERE.00822-19-F0004.JPEG>
- Ramos, A. P., Herrera, B. A., Ramírez, O. V., García, A. A., Jiménez, M. M., Valdés, C. S., Fernández, A. G., González, G., Fernández, S. I. O., Báez, G. G., & Espinosa, B. H. (2013). Molecular and phylogenetic analysis of influenza A H1N1 pandemic viruses in Cuba, May 2009 to August 2010. *International Journal of Infectious Diseases*, *17*(7), e565–e567. <https://doi.org/10.1016/j.ijid.2013.01.028>

- Rogers' And, G. N., & D'souz~, B. L. (1989). 322 Receptor Binding Properties of Human and Animal HI Influenza Virus Isolates. En *VIROLOGY* (Vol. 173).
- Šali, A., & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3), 779–815.
<https://doi.org/10.1006/JMBI.1993.1626>
- Sandbulte, M. R., Spickler, A. R., Zaabel, P. K., & Roth, J. A. (2015). Optimal Use of Vaccines for Control of Influenza A Virus in Swine. *Vaccines 2015, Vol. 3, Pages 22-73*, 3(1), 22–73. <https://doi.org/10.3390/VACCINES3010022>
- Schneider, R., Sharma, A. R., & Rai, A. (2008). Introduction to molecular dynamics. *Lecture Notes in Physics*, 739, 3–40. https://doi.org/10.1007/978-3-540-74686-7_1/COVER
- Shen, M., & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Science : A Publication of the Protein Society*, 15(11), 2507.
<https://doi.org/10.1110/PS.062416606>
- Sippl, M. J. (1993). Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4), 355–362.
<https://doi.org/10.1002/PROT.340170404/ABSTRACT>
- Soundararajan, V., Tharakaraman, K., Raman, R., Raguram, S., Shriver, Z., Sasisekharan, V., & Sasisekharan, R. (2009). Extrapolating from sequence—the 2009 H1N1 “swine” influenza virus. *Nature Biotechnology 2009 27:6*, 27(6), 510–513. <https://doi.org/10.1038/nbt0609-510>
- Sousa Da Silva, A. W., & Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5, 367. <https://doi.org/10.1186/1756-0500-5-367>
- Stanley, W. M. (1944). THE SIZE OF INFLUENZA VIRUS. *Journal of Experimental Medicine*, 79(3), 267–283. <https://doi.org/10.1084/JEM.79.3.267>
- Tafalla, M., Buijssen, M., EGINE Geets, R., & Vonk Noordegraaf-Schouten, M. (2016). A comprehensive review of the epidemiology and disease burden of Influenza B in 9 European countries. *Human Vaccines & Immunotherapeutics*, 12.
<https://doi.org/10.1080/21645515.2015.1111494>
- Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Molecular Biology and Evolution*, 38(7), 3022–3027.
<https://doi.org/10.1093/MOLBEV/MSAB120>
- Taubenberger, J. K., & Morens, D. M. (2008). *The Pathology of Influenza Virus Infections*.
<https://doi.org/10.1146/annurev.pathmechdis.3.121806.154316>.

- Torres, P. H. M., Sodero, A. C. R., Jofily, P., & Silva-Jr, F. P. (2019). Key topics in molecular docking for drug design. En *International Journal of Molecular Sciences* (Vol. 20, Número 18). MDPI AG. <https://doi.org/10.3390/ijms20184574>
- Tramontano, A., & Morea, V. (2003). Assessment of Homology-Based Predictions in CASP5. *Proteins: Structure, Function and Genetics*, 53, 352–368. <https://doi.org/10.1002/prot.10543>
- Tse, H., Kao, R. Y. T., Wu, W. L., Lim, W. W. L., Chen, H., Yeung, M. Y., Woo, P. C. Y., Sze, K. H., & Yuen, K. Y. (2011). Structural basis and sequence co-evolution analysis of the hemagglutinin protein of pandemic influenza A/H1N1 (2009) virus, 236(8), 915–925. <https://doi.org/10.1258/EBM.2011.010264>
- Tzarum, N., De Vries, R. P., Paulson, J. C., & Wilson, I. A. (2015). Structure and Receptor Binding of the Hemagglutinin from a Human H6N1 Influenza Virus. *Cell Host & Microbe*. <https://doi.org/10.1016/j.chom.2015.02.005>
- Valdés-Tresanco, M. S., Valdés-Tresanco, M. E., Valiente, P. A., & Moreno, E. (2021). gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *Journal of chemical theory and computation*, 17(10), 6281–6291. <https://doi.org/10.1021/ACS.JCTC.1C00645>
- Vincent, A. L., Ma, W., Lager, K. M., Janke, B. H., & Richt, J. A. (2008). Chapter 3 Swine Influenza Viruses: A North American Perspective. *Advances in Virus Research*, 72, 127–154. [https://doi.org/10.1016/S0065-3527\(08\)00403-X](https://doi.org/10.1016/S0065-3527(08)00403-X)
- Vyas, V. K., Ukawala, R. D., Ghate, M., & Chintha, C. (2012). Homology modeling a fast tool for drug discovery: Current perspectives. *Indian Journal of Pharmaceutical Sciences*, 74(1), 1–17. <https://doi.org/10.4103/0250-474X.102537>
- Wang, E., Sun, H., Wang, J., Wang, Z., Liu, H., Zhang, J. Z. H., & Hou, T. (2019). End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chemical Reviews*, 119(16), 9478–9508. https://doi.org/10.1021/ACS.CHEMREV.9B00055/ASSET/IMAGES/MEDIUM/CR-2019-000558_0003.GIF
- Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and Testing of a General Amber Force Field. En *J Comput Chem* (Vol. 25).
- Wang, Y., Tang, C. Y., & Wan, X.-F. (2022). Antigenic characterization of influenza and SARS-CoV-2 viruses. *Analytical and Bioanalytical Chemistry*, 414, 3. <https://doi.org/10.1007/s00216-021-03806-6>

- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., De Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research*, 46(W1), W296–W303. <https://doi.org/10.1093/NAR/GKY427>
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9), 1189–1191. <https://doi.org/10.1093/BIOINFORMATICS/BTP033>
- Watson, S. J., Langat, P., Reid, S. M., Lam, T. T.-Y., Cotten, M., Kelly, M., Reeth, K. Van, Qiu, Y., Simon, G., Bonin, E., Foni, E., Chiapponi, C., Larsen, L., Hjulsager, C., Markowska-Daniel, I., Urbaniak, K., Dürrwald, R., Schlegel, M., Huovilainen, A., ... Kellam, P. (2015). Molecular Epidemiology and Evolution of Influenza Viruses Circulating within European Swine between 2009 and 2013. *Journal of Virology*, 89(19), 9920. <https://doi.org/10.1128/JVI.00840-15>
- Webster, R. G., & Govorkova, E. A. (2014). Continuing challenges in influenza. *Annals of the New York Academy of Sciences*, 1323(1), 115. <https://doi.org/10.1111/NYAS.12462>
- Wennekes Tom. (2013). *New Flu Vaccine? Drug Shuts Down Resistant Strains of Influenza Virus*. <https://www.medicaldaily.com/new-flu-vaccine-drug-shuts-down-resistant-strains-influenza-virus-244497>
- Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, 35(Web Server issue), W407. <https://doi.org/10.1093/NAR/GKM290>
- Xu, R., McBride, R., Nycholat, C. M., Paulson, J. C., & Wilson, I. A. (2012). Structural Characterization of the Hemagglutinin Receptor Specificity from the 2009 H1N1 Influenza Pandemic. *Journal of Virology*, 86(2), 982–990. <https://doi.org/10.1128/JVI.06322-11>
- Yang, H., Carney, P., & Stevens, J. (2010). Structure and Receptor binding properties of a pandemic H1N1 virus hemagglutinin. *PLoS Currents*, 2(MAR). <https://doi.org/10.1371/CURRENTS.RRN1152>
- Yasugi, M., Nakamura, S., Daidoji, T., Kawashita, N., Ramadhany, R., Yang, C. S., Yasunaga, T., Iida, T., Horii, T., Ikuta, K., Takahashi, K., & Nakaya, T. (2012). Frequency of D222G and Q223R Hemagglutinin Mutants of Pandemic (H1N1) 2009 Influenza Virus in Japan between 2009 and 2010. *PLoS ONE*, 7(2). <https://doi.org/10.1371/JOURNAL.PONE.0030946>

