

# Universidad Autónoma de Baja California

Facultad de Ingeniería, Arquitectura y Diseño

Campus Ensenada



Modelado *in silico* de proteína conformada de epítomos conservados de influenza A (H1N1 y H5N1) como potencial vacuna universal y construcción de casete de expresión

Para obtener el grado académico de Bioingeniero

Por

Molina Zamudio Kitzia Yashvelt

Director: José Luis Stephano Hornedo

CoDirector: Tatiana N. Olivares Bañuelos

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**  
FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO

MODELADO *IN SILICO* DE PROTEÍNA CONFORMADA DE EPÍTOPOS CONSERVADOS  
DE INFLUENZA A (H1N1 Y H5N1) COMO POTENCIAL VACUNA UNIVERSAL Y  
CONSTRUCCIÓN DE CASETE DE EXPRESIÓN

TESIS

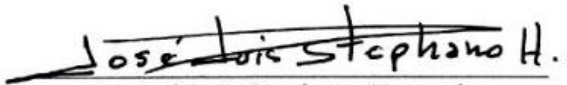
PARA CUBRIR LOS REQUISITOS NECESARIOS PARA OBTENER EL TÍTULO DE

**BIOINGENIERO**

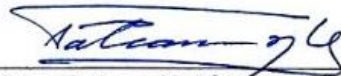
PRESENTA:

**KITZIA YASHVELT MOLINA ZAMUDIO**

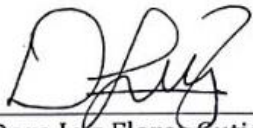
Aprobada por:



Dr. José Luis Stephano Hornedo  
Director  
(Presidente)



Dra. Tatiana N. Olivares Bañuelos  
Codirector  
(Secretario)



Dra. Dora Luz Flores Gutiérrez  
Sinodal  
(Vocal)



Dra. Claudia M. Gómez Gutiérrez  
Sinodal  
(Vocal)



Dr. David Cervantes Vásquez  
Sinodal  
(Vocal)

## **AGRADECIMIENTOS:**

A mis padres y a mi hermana que son mi apoyo y motivación para seguir adelante, que ha estado siempre para mí en los momentos más difíciles, gracias por ser un equipo y mis mejores amigos, sin ustedes no estaría donde estoy ahorita y en donde estaré mañana.

Le agradezco encarecidamente al Dr. José Luis Stephano Hornedo por compartirme un poco de su amplio conocimiento y su apoyo como persona, sobre todo por permitirme la entrada a su laboratorio y a su enorme familia conformada por todos los investigadores que ha forjado. Sus enseñanzas siempre las llevaré presentes; trabajo duro y humildad. Gracias por apoyarme aun cuando fallaba.

Gracias a todos mis maestros por sus enseñanzas, regaños y consejos, que me ayudaron a formarme como profesional, especialmente a mi comité evaluador;

A la Dra. Tatiana Olivares ya que sus consejos y correcciones me fueron de gran ayuda para realizar el presente trabajo, su paciencia a pesar de mis tiempos es algo que se lo agradeceré siempre, gracias por ser la profesora que es y motivar a todos los alumnos que tenemos el gusto de recibir clases de usted.

A la Dra. Claudia Gómez, gracias por ser estricta en sus clases, sus enseñanzas me fomentaron el hábito de la lectura y el gusto por las estructuras proteicas, deseo que más profesores sean como usted, que fomenten disciplina y dedicación en sus alumnos.

Al Dr. David Cervantes Vásquez por motivar mi curiosidad para comprender mejor cómo funciona la energía desde moléculas inorgánicas hasta las proteínas que hacen posible la vida.

A la Dra. Dora Luz Flores por transmitirme el gusto por las matemáticas, recuerdo muy bien sus clases de cálculo, lo que más admiro de usted es su perfeccionismo.

Al Dr. Rubén por hacer interesantes todas y cada una de sus clases, al hacerme captar su atención proyectando su pasión por la enseñanza.

## RESUMEN

Actualmente no existe una vacuna universal contra la Influenza tanto estacional como pandémica. Cada año las compañías farmacéuticas sintetizan vacunas nuevas contra esta enfermedad, ya que la deriva antigénica de estos virus es muy alta y, en caso de que se presentase un nuevo brote pandémico, tomaría más tiempo producir la vacuna que la enfermedad en matar a miles de personas. Por esto ha surgido la necesidad de generar una vacuna "Universal" que tenga un espectro de protección amplio contra diversas cepas de Influenza. En el presente trabajo se realizó el modelado 3D *in silico* de una proteína conformada por cuatro epítomos conservados de influenza A (Hemaglutinina HA, Neuroaminidasa NA, Nucleoproteína NP y Proteína de matriz 2 M2), a la cual se le llamó Vac-Quimera. La quimera resultante fue estable y reconocida por linfocitos B, linfocitos T y MHC-I y II. Adicional al diseño de la quimera, en el presente trabajo se sintetizó también el gen que codifica para la proteína Vac-Quimera, el cual se clonó en el vector pChlamy\_4 para su expresión en *Chlamydomonas reinhardtii* en trabajos posteriores al presente.

## ÍNDICE

Resumen.....	I
Agradecimientos.....	II
Índice de tablas.....	III
Índice de figuras.....	IV
1. Introducción.....	14
2. Antecedentes.....	16
2.1. La influenza.....	16
2.2. El virus de la influenza.....	23
2.3. Mecanismo de infección.....	31
2.4. Proteínas potenciales para la vacuna “Universal”.....	36
2.5. Diseño de proteínas <i>in silico</i> .....	40
2.6. Microalgas para la producción de vacunas.....	41
3. Hipótesis.....	44
4. Objetivo general.....	44
4.1. Objetivos específicos.....	45
5. Metodología.....	46

5.1. Modelado <i>in silico</i> de la proteína Vac-Quimera.....	46
5.1.1. Elección de epítomos conservados.....	46
5.1.2. Análisis fisicoquímico de la estructura primaria de Vac-Quimera.....	48
5.1.3. Reconocimiento de homología.....	48
5.1.4. Predicción de la estructura secundaria.....	49
5.1.5. Identificación de epítomos en células B y en MHC.....	49
5.1.6. Modelado tridimensional de la proteína.....	50
5.1.7. Análisis de la estructura final de Vac-Quimera.....	51
5.1.7.1. Análisis de estereología.....	51
5.1.7.2. Solubilidad.....	53
5.1.7.3. Superposición de la estructura tridimensional.....	54
5.2. Construcción de cassette de expresión.....	54
5.2.1. Síntesis del gen que codifica para Vac-Quimera.....	54
5.2.2. Amplificación por PCR.....	54
5.2.3. Digestión enzimática.....	55
5.2.4. Ligación del gen en el vector pChlamy_4.....	55
5.2.5. Transformación del plásmido en <i>E. coli</i> XL1-Blue.....	56
5.2.6. Confirmación de la transformación.....	56

5.2.6.1. Confirmación mediante digestión enzimática.....	56
5.2.6.2. Confirmación mediante PCR.....	57
5.2.6.3. Confirmación mediante secuenciación.....	57
6. Resultados.....	58
6.1. Diseño de la proteína .....	58
6.1.1. Epítopes conservados.....	58
6.1.2. Secuencia primaria.....	58
6.1.3. Análisis fisicoquímico de la estructura primaria de Vac-Quimera.....	59
6.1.4. Reconocimiento de homología.....	59
6.1.5. Estructura secundaria.....	60
6.1.6 Identificación de epítomos en células B y en MHC.....	62
6.1.7. Estructura tridimensional.....	68
6.1.8. Análisis estructural.....	68
6.1.8.1. Estereología.....	68
6.1.8.2. Solubilidad.....	76
6.1.8.3. Superposición de la estructura tridimensional.....	81
6.2. Construcción del cassette de expresión.....	82
6.2.1. Amplificación del gen por PCR.....	82

6.2.2. Digestión.....	84
6.2.3. Ligación.....	85
6.2.4. Confirmación.....	86
6.2.5. Secuenciación.....	88
7. Discusión.....	90
8. Conclusión.....	101
9. Referencias.....	102
10. Anexos.....	110

## **INDICE DE TABLAS**

TABLA I. Epidemias y pandemias por influenza.....	16
TABLA II. Número acumulado de casos humanos confirmados de gripe aviar A (H5N1) notificado a la OMS, 2003-2015.....	21
TABLA III. Nomenclatura comparativa entre el mismo virus con diferente hospedero.....	22
TABLA IV. Péptidos de la nucleoproteína del virus de la influenza A reconocidos por Linfocitos T Citotóxicos presentados por el complejo mayor de histocompatibilidad I.....	36

TABLA V. Comparación de las características de los sistemas de producción de proteínas recombinantes tradicionales respecto a las microalgas.....	39
TABLA VI. Proteínas recombinantes de uso médico producidas en el cloroplasto de <i>C. reinhardtii</i> .....	41
TABLA VII. Secuencias de diversos ectodominios de proteínas de matriz 2 en virus de la influenza A .....	44
TABLA VIII. Comparación entre algunos péptidos de fusión de diversas cepas conocidas del virus de la influenza .....	45
TABLA IX. Cebadores utilizados para amplificar el gen Vac-Quimera mediante Reacciones en Cadena de la Polimerasa (PCR).....	53
TABLA X. Datos del análisis de la proteína Vac-Quimera realizado por el programa Protparam.....	57
TABLA XI. Determinantes antigénicos lineares reconocidos por anticuerpos.....	60
TABLA XI. Epítopes reconocidos por el MHC-I de los serotipos HLA-A, HLA-B y HLA-C.....	61
TABLA XII. Unión de epítopes a MHC-I (HLA-A*0201) .....	63
TABLA XIII. Unión de epítopes a MHC-II (HLA-DRB1*0101,*0401*0701) .....	64
TABLA XIV. Péptidos antigénicos presentados por los complejos MHC-I y MHC-II con su respectivo valor antigénico.....	65
TABLA XV. Comparación entre el tipo de modelado de los servidores seleccionados.....	66

TABLA XVI. Comparación de la calidad de las proteínas modeladas con los cuatro diferentes servidores.....	68
TABLA XVII. Comparación de los diferentes valores de solubilidad obtenidos.....	73
TABLA XVIII. Concentraciones de DNA plasmídico obtenido por extracción.....	82

## ÍNDICE DE FIGURAS

Figura 1. Número de personas infectadas por diferentes subtipos de influenza alrededor del mundo a partir del año 2009 .....	18
Figura 2. Número de personas infectadas por diferentes subtipos de influenza en México a partir del año 2009 .....	19
Figura 3. Virus de la Influenza con sus proteínas estructurales y genes.....	23
Figura 4. Estructura cristalizada de hemaglutinina A/Hong Kong/1/1968 (H3N2).....	24
Figura 5. Estructura cristalizada de neuroaminidasa (NA) A/Dk/England/1/1956 (H11N6).....	26
Figura 6. Estructura cristalizada de nucleoproteína .....	28
Figura 7. Estructura cristalizada de la proteína de matriz 2 .....	29
Figura 8. Galactosa unida al ácido siálico en la forma 2,3 que reconoce el virus de influenza aviar.....	30
Figura 9. Hemaglutinina.....	32

Figura 10. Fagocitosis del virus de la influenza.....	33
Figura 11. Efecto del pH en la fagocitosis.....	33
Figura 12. Orden seleccionado para la síntesis del gen Quimera.....	56
Figura 13. Análisis de secuencia de la proteína Vac-Quimera por PSI-BLAST.....	58
Figura 14. Predicción de estructura secundaria basada en la secuencia de la proteína Vac-Quimera .....	59
Figura 15. Estructura predicha por el servidor Phyre2 de la proteína Vac-Quimera.....	59
Figura 16. Estructura secundaria de la proteína Vac-Quimera predicho por el servidor PROCHECK.....	59
Figura 17. Estructura secundaria predicha por NPSA .....	60
Figura 18. Estructura tridimensionales de la proteína Vac-Quimera .....	67
Figura 19. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera .....	69
Figura 20. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera .....	70
Figura 21. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera .....	71
Figura 22. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera modelada con el servidor I-TASSER.....	72
Figura 23. Estructura tridimensional de proteína Vac-Quimera modelada con el servido Phyre2.....	75

Figura 24. Estructura tridimensional de proteína Vac-Quimera modelada con el servidor Robetta.....	76
Figura 25. Estructura tridimensional de proteína Vac-Quimera modelada con el servidor QUARK.....	77
Figura 26. Empalme de las proteínas modeladas utilizando el servidor PhyreStorm.....	78
Figura 27. Sitios de hibridación de los cebadores diseñados para amplificar el gen Vac-Quimera y agregar en el 5' un sitio de corte de <i>EcoRI</i> .....	80
Figura 28. Amplificación del gen Quimera con diferentes concentraciones de cloruro de magnesio.....	81
Figura 29. Doble digestión enzimática para el vector pChlamy_4 con las enzimas <i>EcoRI</i> y <i>PstI</i> .....	81
Figura 30. Diagrama del cassette de expresión final.....	82
Figura 31. Clonas positivas de <i>E.coli</i> XL1-blue con el vector pChlamy_Quim4 después de que se hizo la digestión enzimática con <i>EcoRI</i> y <i>PstI</i> .....	84
Figura 32. Amplificación de colonias positivas de <i>E. coli</i> transformadas con la ligación del gen Vac-Quimera en el vector pChlamy_4 .....	85
Figura 33. Alineamiento de la secuencia Vac-Quimera con la muestra 1A, con 92 nucleótidos alineados de 264 del gen Vac-Quimera .....	86

## 1. INTRODUCCIÓN

Cada año, usualmente en invierno, el virus de la influenza estacionaria infecta a los humanos. Hay tres tipos de virus de influenza: A, B y C, de los cuales los virus A y B son los que causan epidemias de temporada.

Las infecciones causadas por virus de influenza tipo C generan enfermedades respiratorias leves y no se cree que puedan causar epidemias (OMS 2014; CDC 2015). La enfermedad generada por el virus de la influenza se transmite de persona en persona, al estornudar, toser o bien, al tocar superficies contaminadas. Para que un virus de influenza se considere como pandémico, éste no debe haber infectado humanos y, por lo tanto, la mayoría de la población al ser infectada por el virus no tendrá inmunidad ante la enfermedad, incrementando considerablemente el número de infectados, en comparación con el número de afectados por el virus de la influenza estacional. Un ejemplo de esto ocurrió en el año 2009, cuando el virus de la influenza A (H1N1), un virus que no había circulado anteriormente en humanos, rápidamente infectó a un número alto de habitantes y se convirtió en una pandemia mundial, para después una influenza estacional (OMS 2014).

Las aves son el hospedero natural del virus de la influenza A, aunque el virus puede infectar también humanos, puercos y caballos, entre otros animales. Este virus muta mayormente en dos de sus proteínas de membrana: HA y NA y, con base en esto se le ha asignado su nomenclatura. Cada una de las proteínas de membrana que muta tiene distintos serotipos; la proteína HA o hemaglutinina se divide en dieciocho serotipos que van del H1 al H18, mientras que la proteína NA o neuroaminidasa, se divide en once serotipos, denominados del N1 al N11. Todos los subtipos de

influenza A son capaces de infectar a las aves, pero no de provocar la enfermedad. Hasta la fecha todos los brotes de la forma más patógena conocida han sido causados por los subtipos H5 y H7 (dotados con las variantes 5 y 7 de la hemaglutinina). De los once subtipos de la proteína NA, solo la N1 y N2 han sido responsables de pandemias y epidemias anuales recurrentes en humanos (Tong et al. 2013). A diferencia del virus de la influenza A, el virus de la influenza B usualmente se encuentra en humanos y, al igual que el virus de influenza C, no se clasifica en subtipos ya que sus proteínas no mutan rápidamente en comparación con las proteínas del virus de la influenza A (Apisarnthanarak et al. 2008).

Considerando la variabilidad que tiene el virus de la influenza A y, el daño que puede causar en las poblaciones, el objetivo de los grupos de investigación es generar una vacuna universal que sirva para proporcionar inmunidad contra diferentes serotipos de influenza y, que a su vez, pueda proteger a la humanidad de una nueva cepa pandémica (Kaiser 2006). Debido a la alta tasa de mutación del virus de la influenza A, cada año las compañías farmacéuticas cambian las cepas que utilizan para la elaboración de vacunas. En la industria farmacéutica, el proceso de producción dura nueve meses, por lo tanto si una nueva pandemia surgiera, la producción de una vacuna tardaría más tiempo en realizarse de lo que el virus tardaría en matar a miles de personas. Las vacunas estacionales actuales que funcionan y que generan actividad inmunológica son vacunas elaboradas con el virus atenuado, el cual genera anticuerpos que reconocen a las glicoproteínas de la superficie del virus: la hemaglutinina HA y neuroaminidasa NA. Hoy en día, el proceso de producción de vacunas se lleva a cabo utilizando huevos de gallina embrionados. Éste método

presenta varias desventajas, entre las que se encuentra la limitación de la producción si la oferta del huevo es insuficiente, la baja inmunidad heteróloga generada por el virus y, las reacciones alérgicas a las proteínas del huevo. Debido a esto, ha sido necesario desarrollar nuevos procesos para la producción de vacunas hechas a partir de proteínas recombinantes que mimetizan a los epítomos del virus (Ren et al. 2015). Dentro de la producción de proteínas recombinantes se han realizado diseño *in silico* de estas para predecir estructura y funcionalidad de proteínas modificadas. En éste contexto, las microalgas han sido vistas en los últimos años como una nueva alternativa para la síntesis de proteínas recombinantes. Entre las ventajas que presenta utilizar estos organismos para la producción de vacunas, podemos mencionar que su cultivo es fácil y económico, que acumulan grandes cantidades de biomasa y, que son consideradas inocuas para los humanos (Siripornadulsil et al. 2007).

## **2. ANTECEDENTES**

### **2.1. LA INFLUENZA**

La influenza es una enfermedad muy contagiosa de las vías respiratorias, dicha enfermedad causa dolor de cabeza, cuerpo cortado, y malestar en general, en casos muy severos puede ocasionar la muerte, la influenza tiene una duración de al menos dos días hasta dos semanas pero la prolongación de esta dependerá de las complicaciones que lleguen a ocurrir. Las complicaciones de la influenza en su mayoría se deben a infecciones conjuntas con bacterias provocando neumonía

hasta meningitis (CDC n.d.). Después de la altamente patogénica influenza española de 1918, la cepa H1N1 fue considerada como una cepa dominante en la población hasta que surgieron nuevas cepas como la cepa H2N2, que causó la gripe asiática en 1957 y la cepa H3N2 causante de la gripe de Hong Kong en 1968. Se tienen registros que después de estos eventos resurgió la influenza H1N1 en 1976, por una cepa idéntica a la identificada en los años 50's en China y en la Unión Soviética. Actualmente, la influenza generada tanto por la cepa H1N1 como por la cepa H3N2, circula como influenza estacional en las poblaciones (Joshi et al. 2009). En la tabla I se presenta el orden cronológico de los registros históricos sobre epidemias y/o pandemias causadas por el virus de la influenza.

Es importante resaltar el impacto que han provocado estas gripes sobre la densidad poblacional mundial. La mayor mortandad se registró durante la gripe española, en la que enfermaron, aproximadamente, el 20%-40% de personas y murieron entre 20 y 50 millones de estos, lo que la convirtió en una de las gripes con más alta tasa de mortalidad en adultos de 20 a 50 años (FluGov 2010). La gripe asiática fue la segunda pandemia más fuerte del siglo XX después de la española, y la más alta en tasa de infección en niños (Caballero 2005); se presentaron 70 mil muertes en los Estados Unidos y 2 millones de muertes alrededor del mundo.

TABLA I. Epidemias y pandemias por influenza, datos históricos (Kilbourne 2006)

Año	Afección
1889-90	H2N8, serológico retrospectivo, pandemia intensa
1900-03	H3N8, serología retrospectiva, pandemia intensa
1918-19	H1N1, gripe española, 20-50 millones fallecidos
1933-35	H1N1, epidemia ligera, se aisló por primera vez la Influenza A
1946-47	H1N1, epidemia ligera
1957-58	H2N2, gripe asiática, pandemia intensa
1968-69	H3N2, gripe de Hong Kong, pandemia moderada
1976	H1N1, gripe porcina, no se diseminó
1977-78	H1N1, gripe rusa, pandemia ligera
1997	H5N1, gripe aviar aparecida en Hong Kong, casos esporádicos en humanos
2009	H1N1, gripe A iniciada en marzo declara pandemia por la OMS en junio, finalizó el 10 de agosto de 2010 dejando 19 mil víctimas, su nombre inicial fue gripe porcina

En años posteriores surgió la gripe de Hong Kong, la cual mató alrededor de 34 mil individuos en los Estados Unidos y 1 millón alrededor del mundo (Kilbourne 2006). En 2009 la influenza H1N1 reemergió como pandemia moderada, ya que los casos mencionados anteriormente, con excepción de la gripe española de 1918, no se presentó una mortandad tan alta, pero se estima que hubo un total de 151,700 a 575,400 muertes alrededor del mundo (Fig. 1). Actualmente, ésta gripa solo circula como gripe estacional.

Durante los primeros tres meses del 2016 se declararon en alerta epidemiológica por influenza los países de Panamá, Costa Rica, Perú y México, con 26 casos registrados para Panamá, 29 muertes en Costa Rica y 2000 casos y 123 muertes en Perú durante los últimos tres años (El comercio 2016). En México se han reportado 4121 casos de personas infectadas por influenza A de diversos subtipos, con un total de 149 muertes, mientras que por AH1N1 suman 1519 los casos y 111 muertes en lo que va del año 2016 (Fig. 2) (Excelsior 2016).

Número de especímenes positivos para influenza por subtipo

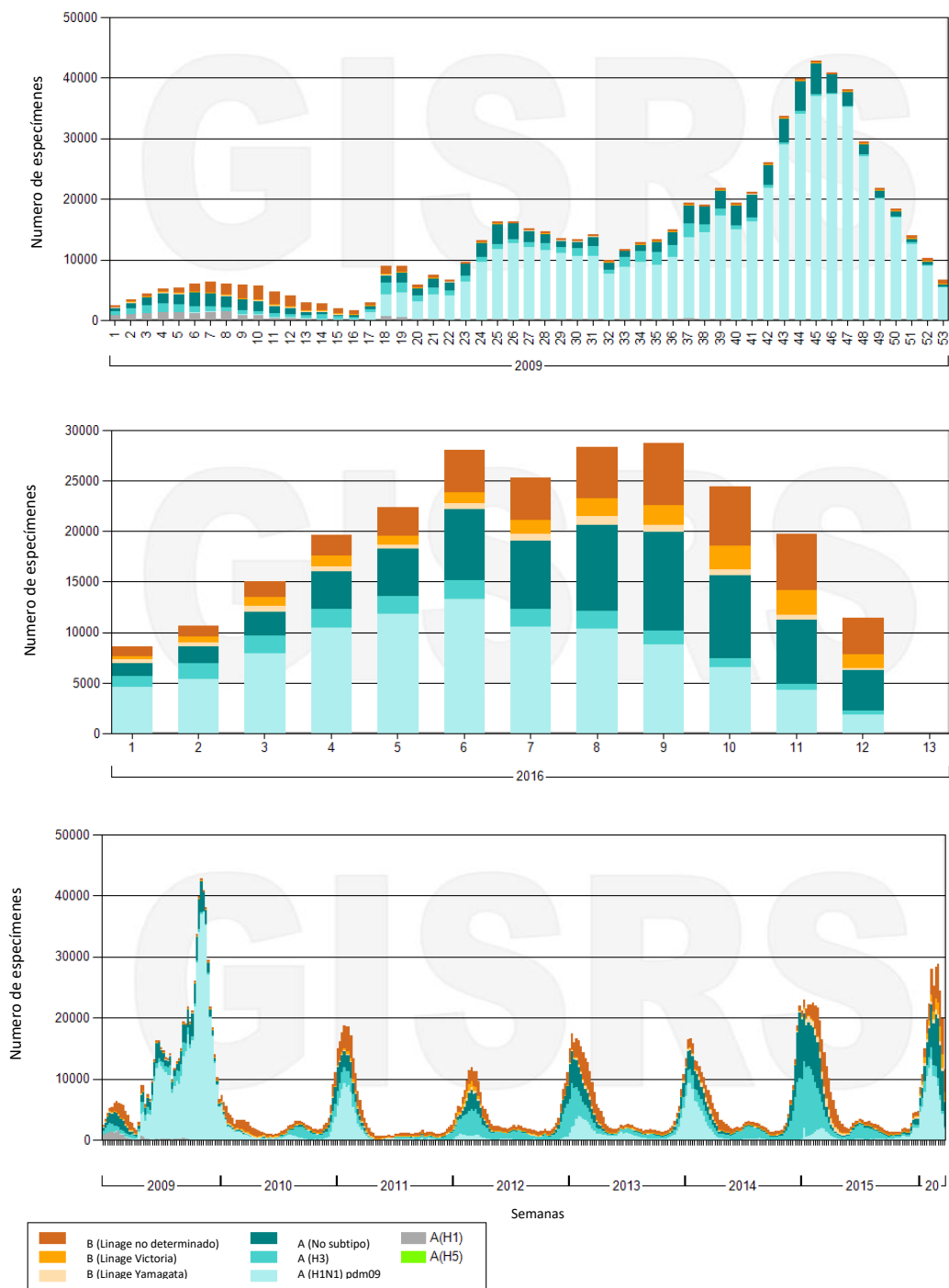


Figura 1. Número de personas infectadas por diferentes subtipos de influenza alrededor del mundo a partir del año 2009 (WHO 2016).

Número de especímenes positivos para influenza por subtipo

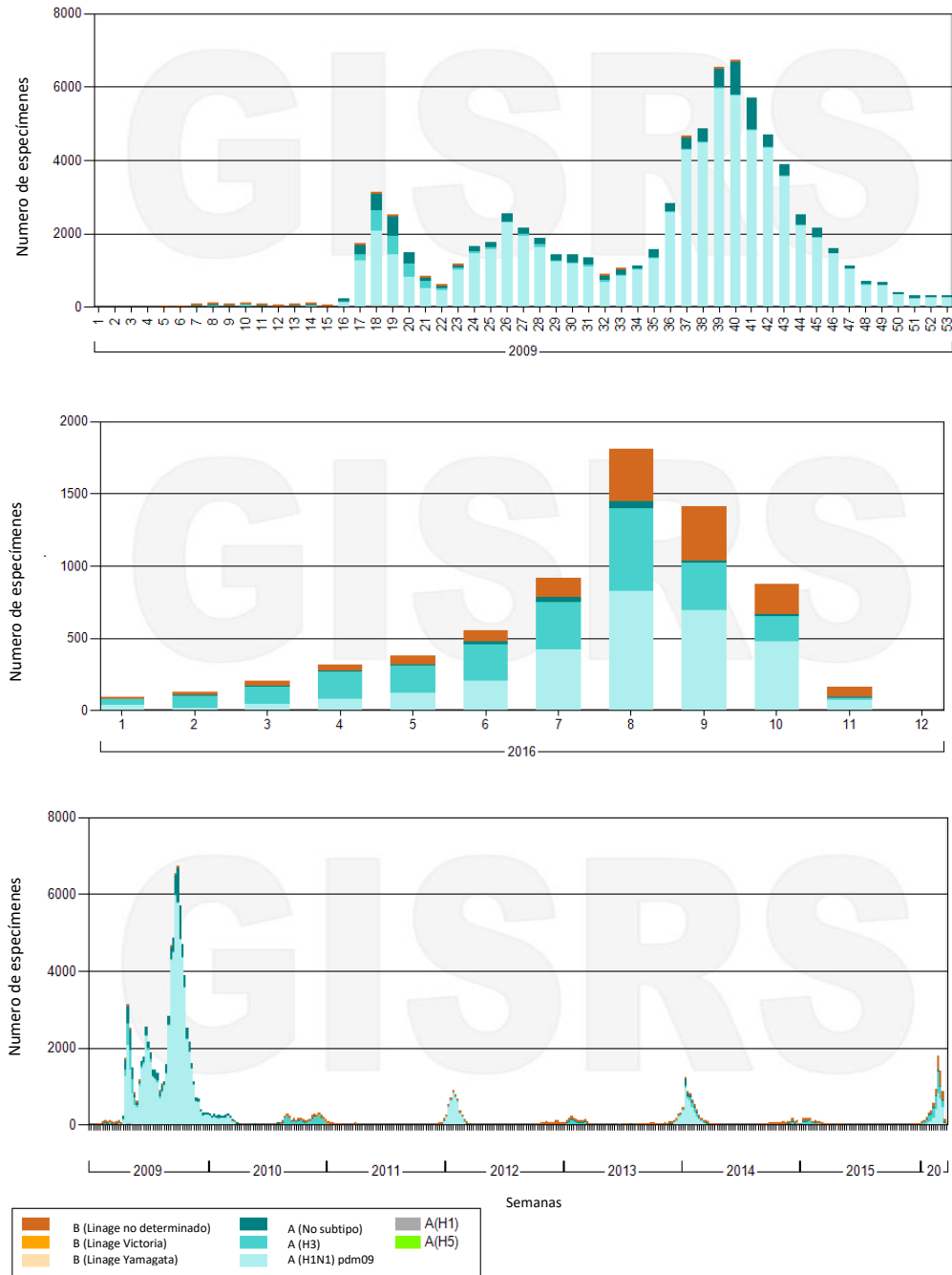


Figura 2. Número de personas infectadas por diferentes subtipos de influenza en México a partir del año 2009 (WHO 2016).

Dentro de la diversidad de cepas de influenza existen algunas que se transmiten de humano a humano, mientras que otras se transmiten de animales a humanos o viceversa. Se sabe que el puerco es el principal organismo que sirve para la recombinación de virus de la influenza, a pesar de que su hospedero natural son las aves. El mecanismo de recombinación que tienen los virus de influenza es uno de los principales factores que hace impredecible saber cuándo habrá una nueva cepa que genere una pandemia. Una de las cepas menos comunes, pero sin embargo, de las más patogénicas, es la de la influenza aviar H5N1, la cual tiene un intervalo de mortalidad del 24% al 100%. Esta cepa se transmite de animales enfermos a humanos (actualmente no hay casos de transmisión de humano a humano), por lo tanto, tiene un gran potencial para convertirse en una nueva cepa pandémica (Joshi et al. 2009). Un ejemplo de lo que puede generar la recombinación de éstos virus se presentó en el año 2005, cuando por primera vez que se produjo una infección humana por una recombinación triple (tres virus involucrados) en Wisconsin, E.U.A. La primera que se tiene reportada de una recombinación cuádruple de virus fue en 2009, con el virus H1N1 que contenía los genes NP, NS y el H1 de la cepa de 1918 de gripe porcina, los genes PB2 y PA de la gripe de 1998 procedente de aves de América del Norte transferida a cerdos, el gen PB1 proviene de una transferencia de gripe de aves a seres humanos en 1968, que luego pasó a cerdos en 1998 y, el gen de matriz que tiene una mutación que le permite al virus resistencia a la amantadina y a la rimantadina (Perez n.d.). En la tabla II, se muestra el número acumulado de casos humanos confirmados de gripe aviar A (H5N1) notificado a la OMS, en el periodo comprendido entre los años 2003 a 2015.

TABLA II. Número acumulado de casos humanos confirmados de gripe aviar A (H5N1) notificado a la OMS, 2003-2015 [modificado de [www.who.int/influenza/human\\_animal\\_interface](http://www.who.int/influenza/human_animal_interface)].

País	2003-2009		2010		2011		2012	
	casos	muertes	Casos	Muertes	casos	muertes	Casos	muertes
Azerbaiyán	8	5	0	0	0	0	0	0
Bangladesh	1	0	0	0	2	0	3	0
Cambodia	9	7	1	1	8	8	3	3
Canada	0	0	0	0	0	0		
China	38	25	2	1	1	1	2	1
Djibouti	1	0	0	0	0	0		
Egipto	90	27	29	13	39	15	11	5
Indonesia	162	134	9	7	12	10	9	9
Iraq	3	2	0	0				
Lao People's Demo. Republic	2	2	0	0				
Myanmar	1	0	0	0				
Nigeria	1	1	0	0				
Pakistán	3	1	0	0				
Tailandia	25	17	0	0				
Turkey	12	4	0	0				
Viet Nam	112	57	7	2			4	2
<b>Total</b>	<b>468</b>	<b>282</b>	<b>48</b>	<b>24</b>	<b>62</b>	<b>34</b>	<b>32</b>	<b>20</b>
País	2013		2014		2015		Total	
	casos	muertes	casos	muertes	casos	muertes	Casos	muertes
Azerbaiyán							8	5
Bangladesh	1	1					7	1
Cambodia	26	14	9	4			56	37
Canada	1	1					1	1
China	2	2	2				47	30
Djibouti							1	0
Egipto	4	3	14	7	16	2	203	72
Indonesia	3	3	2	2			197	165
Iraq							3	2
Lao People's Demo. Republic							2	2
Myanmar							1	
Nigeria							1	1
Pakistán							3	1
Tailandia							25	17
Turkey							12	4
Viet Nam	2	1	2	2			127	64
<b>Total</b>	<b>39</b>	<b>25</b>	<b>29</b>	<b>15</b>	<b>16</b>	<b>2</b>	<b>694</b>	<b>402</b>

## 2.2. EL VIRUS DE LA INFLUENZA

Los virus de la influenza pertenecen a la familia Orthomyxoviridae. El genoma de éstos virus está compuesto de ocho segmentos de RNA de cadena sencilla de sentido negativo, que codifica para 11 genes virales: Hemaglutinina (HA); Neuroaminidasa (NA), matriz 1 (M1), matriz 2 (M2), Nucleoproteína (NP), proteína no estructural 1 (NSP1), proteína no estructural 2 (NS2: también conocida como proteína de exportación nuclear NEP), polimerasa ácida (PA), polimerasa básica 1 (PB1), polimerasa básica 2 (PB2) y, otra polimerasa básica 1-F2 (PB1-F2) (Ping et al. 2015; Samji 2009) . Los virus de la influenza se dividen en tipos A, B y C, de acuerdo a las diferencias antigénicas de sus nucleoproteínas (NP) y su proteína de matriz (M) (Fig. 2). La influenza A se subdivide a su vez en subtipos, clasificándolos con una nomenclatura en donde se incluye: hospedero de origen, origen geográfico, número de cepa y año en que fue aislado. En la tabla III se presenta un ejemplo de la nomenclatura comparativa entre el mismo virus con diferente hospedero (Fields 1985).

TABLA III. Nomenclatura comparativa entre el mismo virus con diferente hospedero (Fields 1985).

Subtipo	Humano	Cerdo	Caballo	Ave
H3	HK/1/68	Sw/Taiwan/70	Eq/Miami/1/63	Dk/Ukr/1/63

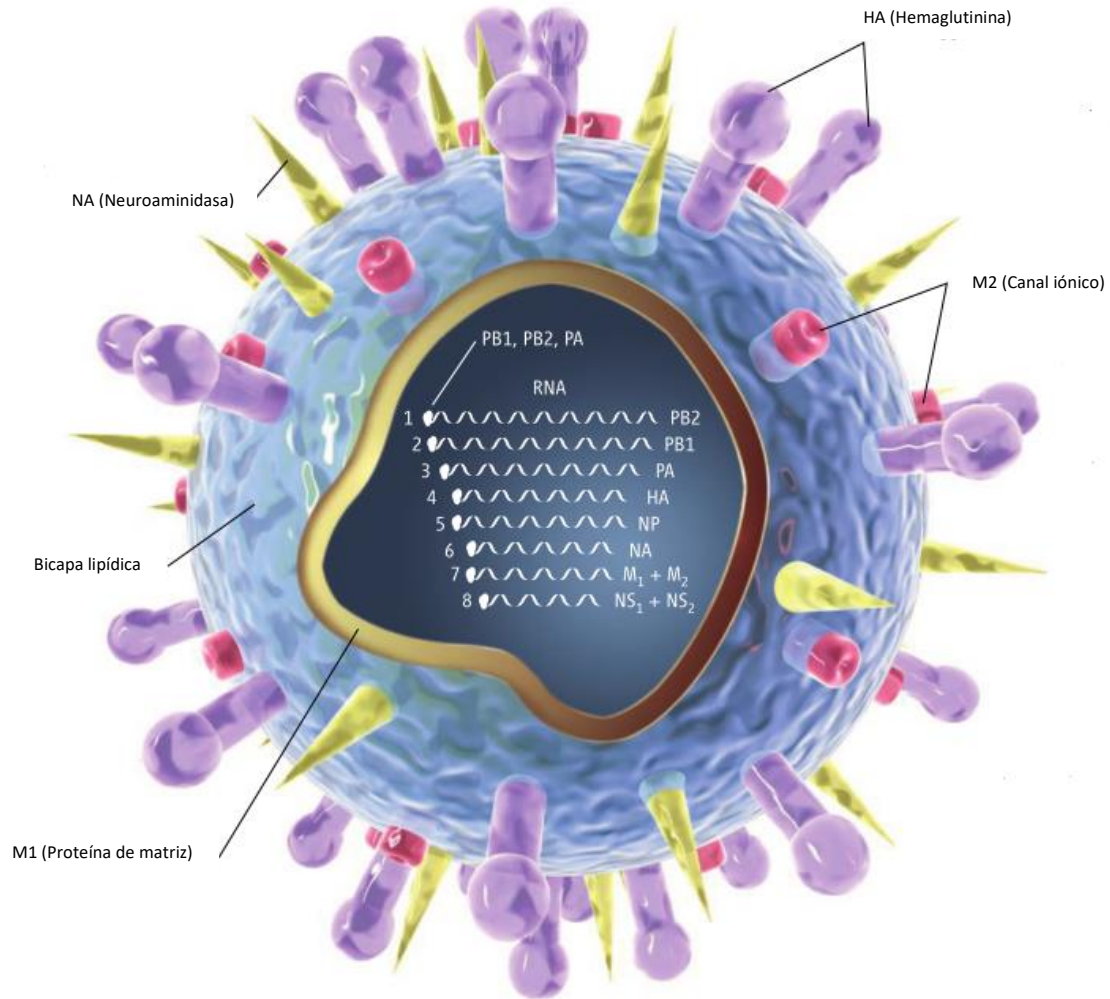


Figura 3. Virus de la Influenza con sus proteínas estructurales y genes. Ilustración tomada de Kaiser 2006.

El virus de la influenza A tienen dos antígenos principales que son la hemaglutinina y la neuroaminidasa. La hemaglutinina (HA) constituye el 25 % de las proteínas virales (~500 moléculas por virus) y es la responsable de la penetración del virus a las células durante las fases iniciales de la infección. La HA es un monómero que se encuentra codificado en el cuarto segmento del RNA y está formado por una sola cadena polipeptídica que es cortada post-traduccionalmente en tres secciones. De

estas tres secciones la parte del N-terminal tiene una secuencia señal que al ser removida, dará como resultado dos cadenas llamadas HA1 y HA2, las cuales tendrán pesos moleculares de 36,000 y 27,000 daltones, respectivamente. Los segmentos HA1 y HA2 están unidos por un solo enlace disulfuro y, en la parte del C-terminal de la HA2 hay una secuencia de 25-32 aminoácidos hidrofóbicos que le sirven a la proteína para anclarse a la membrana del virus (Fields 1985; Ping et al. 2015). La HA está construida por un trímero de dos regiones estructurales distintas llamadas tallo y cabeza. El tallo está compuesto de una triple hebra enrollada de alfa hélices que tienen una longitud de 76Å desde la membrana del virus hasta la zona de la cabeza (Fig. 3a), la cual es una región globular compuesta de una hoja beta que contiene el receptor al sitio de unión. Por su parte, los determinantes antigénicos se encuentran en dicho dominio globular (Fig. 3b) (Fields 1985; Ekiert et al. 2009).

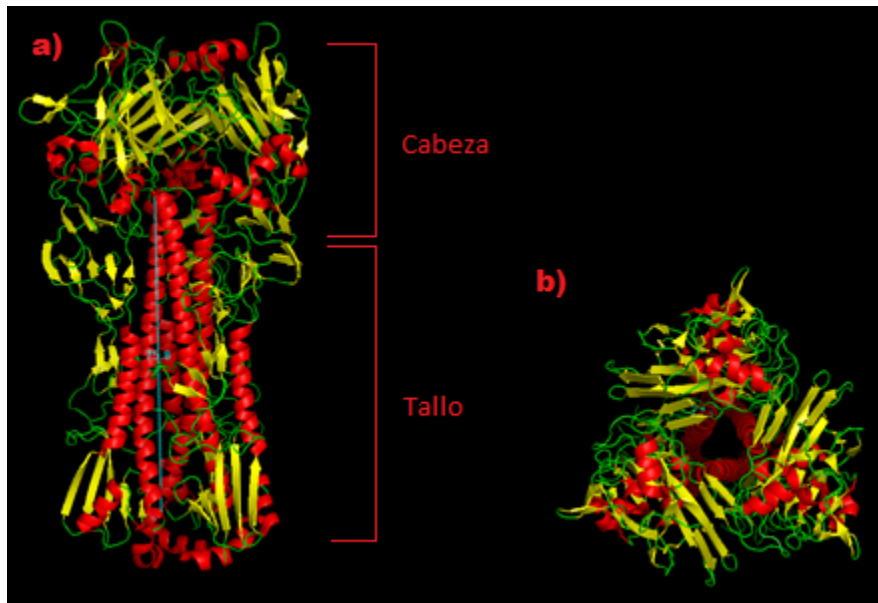


Figura 4. Estructura cristalizada de hemaglutinina A/Hong Kong/1/1968 (H3N2) a) Se puede observar las secciones de tallo 76Å (color azul) y cabeza, b) vista desde arriba la región globular, donde se pueden apreciar las tres secciones de la hemaglutinina [Estructura analizada en Pymol, 4FNK].

El segundo antígeno del virus de la influenza A, la Neuroaminidasa (NA), es la segunda glicoproteína subtipo-específica del virus y constituye el 5% de las proteínas virales (~100 moléculas por virus). La proteína NA es una proteína tetramérica con simetría circular cuádruple codificada en el quinto segmento de RNA (Fig. 5c, d). La cadena polipeptídica que contiene se pliega en seis secciones idénticas de hojas beta antiparalelas unidas por giros que cambian con la deriva antigénica (Fig. 5a, b), dejando como zonas conservadas los residuos cercanos al sitio de unión al ácido siálico. Estructuralmente, la proteína NA tiene forma de hongo con una cabeza en forma de caja hecha de cuatro subunidades semi esféricas y coplanares, unidas centralmente a un tallo con región hidrofóbica con el que se incrusta a la membrana viral (Fields 1985; Sun et al. 2014).

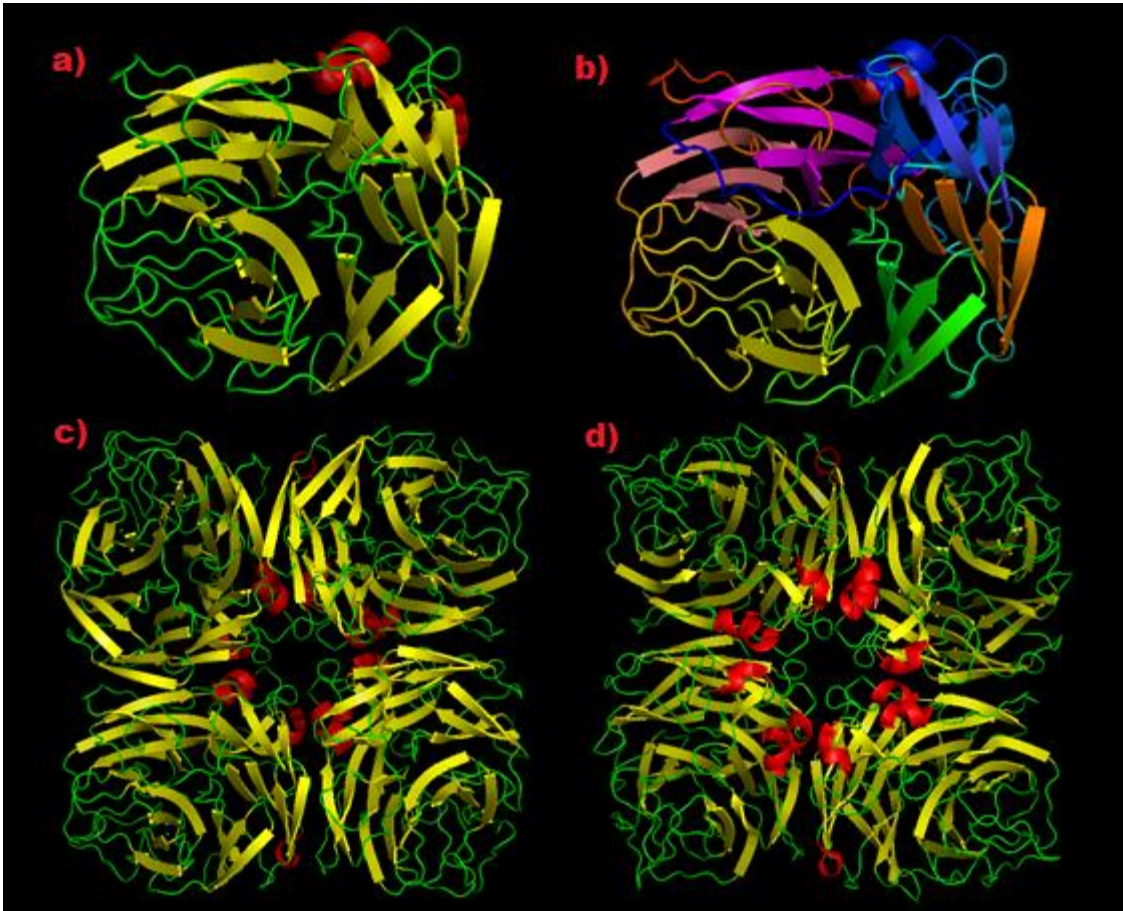


Figura 5. Estructura cristalizada de neuroaminidasa (NA) A/Dk/England/1/1956 (H11N6), a) cadena polipeptídica, b) cadena polipeptídica en donde se identifican los seis grupos de cuatro hojas betas antiparalelas en colores azul, naranja, verde, amarillo, salmón y morado, c) vista anterior del tetrámero, d) vista posterior mostrando las regiones hidrofóbicas [Analizado en Pymol, 1W20].

Otra proteína de suma importancia para el virus de la influenza A es la Nucleoproteína (NP), la cual es una de las proteínas distintivas de cada tipo específico de influenza (ya sea A, B o C). La NP está codificada en el segmento número cinco del genoma viral y se conforma con aproximadamente 500 aminoácidos, lo cual constituye el 17% de las proteínas de la envoltura viral. Su principal función es unirse a los segmentos del RNA viral (vRNA) y al RNA complementario (cRNA), en donde hace la función de ribonucleoproteína (RNP) ya

que se une a las polimerasas PB2, PB1 y PA formando un heterotrímero en el cual la polimerasa se une al final del vRNA mientras la NP encapsula el resto del vRNA con una periodicidad de 24 bases de vRNA por NP. Esta unión se lleva a cabo mediante una interacción ribosa-fosfato dejando las bases expuestas para permitirle a la polimerasa transcribirlas (Cianci et al. 2013; Xu et al. 2011). Estructuralmente, la NP es un trímero en el cual cada monómero tiene forma de media luna. La proteína NP está formada principalmente por alfa hélices; la estructura completa se divide en cabeza y cuerpo, que se unen entre sí por la cadena polipeptídica en tres regiones. Los dominios de cabeza y cuerpo están formados por regiones de polipéptidos no contiguos, en los cuales el dominio de la cabeza está formada por los residuos 150-272 y 438-452 y, el dominio corporal se compone de los tres segmentos de polipéptidos 21-149, 273-396 y 453-489 (Ye et al. 2006). Debido a que esta proteína es distinta en los diversos tipos de influenza, se han hecho estudios filogenéticos para determinar ancestros en común y dilucidar como es que se recombina para infectar distintas especies. Gracias a estos estudios se ha descubierto que la influenza H1N1 en humanos y en cerdos, tienen un ancestro en común (Xu et al. 2011).

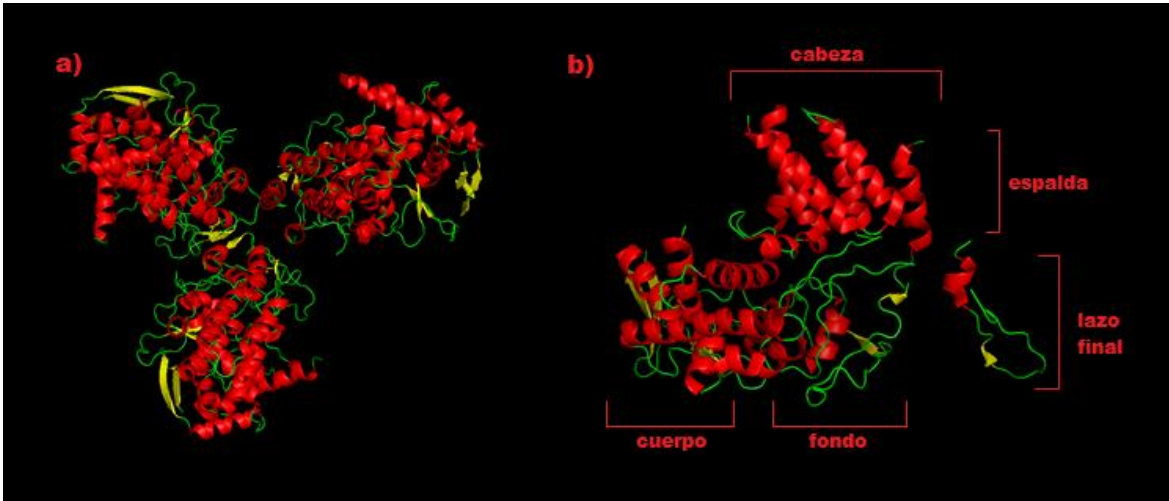


Figura 6. Estructura cristalizada de nucleoproteína a) Vista anterior del trímero completo, b) monómero de la proteína con sus partes especificadas [Estructura modificada y visualizada en Pymol].

La proteína de Matriz II (M2) del virus de la influenza A es otra molécula importante de éste virus. La proteína es codificada en el segmento número siete del RNA, al igual que la proteína de Matriz I (M1). M1 es una proteína de la cápside, mientras que M2 forma un canal iónico. Ambas proteínas son altamente conservadas, con aproximadamente 16 a 20 moléculas por cada virión (Gabbard et al. 2009). Estructuralmente, M2 es un tetrámero de 97 aminoácidos formado por unidades idénticas unidas por dos puentes disulfuro, cuyo monómero es una hélice conformada de tres segmentos: 1) un dominio extracelular amino terminal que comprende los aminoácidos del 1 al 23, 2) un dominio transmembranal con los residuos del 24 al 46 y 3) la región intracelular del carboxilo terminal compuesta del residuo 47 al 97 (Fig. 7). Los residuos del 46 al 62 pertenecen a una zona altamente conservada de la hélice anfipática de la proteína M2, la cual tiene la función de alterar la curvatura de la membrana. La principal función de esta proteína es mantener el pH bajo durante la penetración del virus a la célula, de manera tal que

una vez dentro de la célula el pH bajo desestabiliza la unión de la ribonucleoproteína RNP para empezar a transcribir el vRNA y traducir las proteínas del virus (entre ellas HA, NA y M2). Las proteínas generadas por éste mecanismo se insertan en el retículo endoplasmático para ser transportadas posteriormente a la superficie de la célula hospedera mediante la red *trans*-Golgi (TGN). En este momento del ciclo viral M2 lleva a cabo su función secundaria, la cual es proteger a las proteínas formadas del pH del lumen del aparato de Golgi, para que la HA inmadura no pierda su conformación antes de ser transportada a la membrana (Ye et al. 2006; Pielak & Chou 2011).

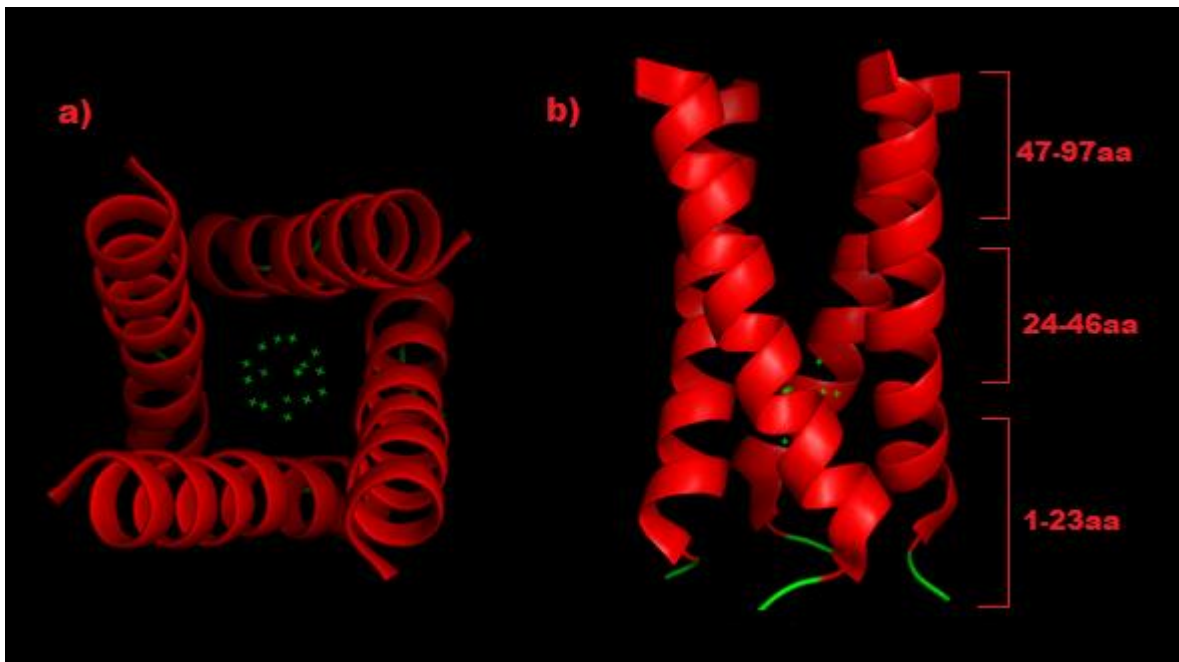


Figura 7. Estructura cristalizada de la proteína de matriz 2 a) Vista del canal iónico, b) monómero de la proteína con sus partes especificadas [Estructura modificada y visualizada en Pymol].

### **2.3. MECANISMO DE INFECCIÓN**

Los virus animales solamente pueden infectar células que poseen receptores específicos para el virus en particular. Por lo tanto, se requieren diferentes receptores para diferentes tipos de virus. En el caso de la influenza, el virus infecta las células epiteliales, especialmente las del tracto respiratorio, y su ingreso a la célula huésped dependerá de que el virus pueda unirse al ácido siálico unido a la galactosa (polisacárido de la membrana celular). Este reconocimiento es efectuado por la proteína HA, la cual puede adherirse a diferentes tipos de uniones entre el ácido siálico y la galactosa entre ellos están la unión 2,3 que se realiza entre el carbón 2 del ácido siálico y el carbón 3 de la galactosa (Fig. 8a) que se encuentra en aves y es reconocida por la influenza H5N1. La galactosa también se puede encontrar en el epitelio profundo de los pulmones de humanos, sin embargo, en humanos la unión más común de las vías respiratorias altas es la 2, 6 (Fig. 8b). Variaciones en la estructura de la HA del virus permiten que esta se asocie a distintos tipos de uniones galactosa-ácido siálico, lo que le brinda un espectro de infección más amplio (Montiel & G. 2010; Slonczewski & Foster 1923).

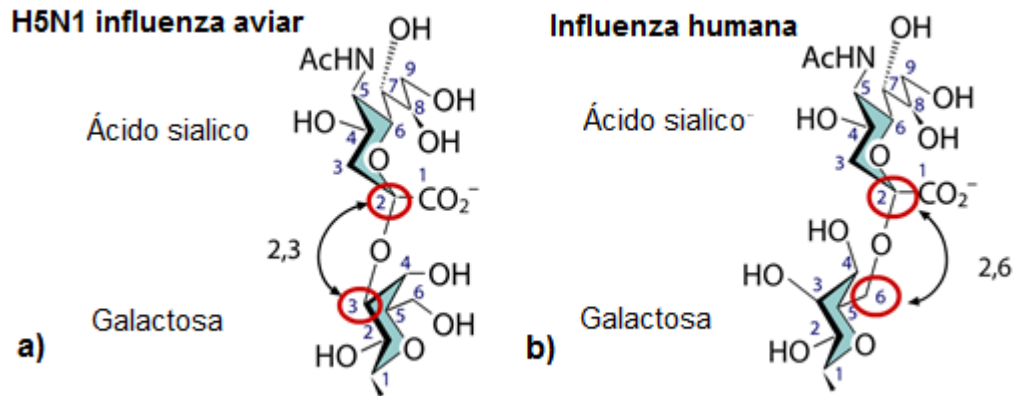


Figura 8. Galactosa unida al ácido siálico en la forma 2,3 que reconoce el influenza aviar, b) galactosa unida al ácido siálico en la forma 2,6 típica en vías respiratorias superiores, reconocida por la influenza en humanos (Slonczewski & Foster 2008).

En el trímero de la hemaglutinina cada monómero se une a un ácido siálico (Fig. 9a) dándole a la HA la capacidad de hacer ingresar al virus a la célula. Para que esto ocurra es necesario que la proteína HA se rompa en HA1 y HA2, mediante proteasas (tripsinas, urocinasas, trombinas, calicreína, factor Xa, acrosina, entre otras) del tracto respiratorio que el mismo hospedero libera, cuyo sitio de corte es un giro que puede contener uno o varios residuos de Arg o Lys, con la forma R-X-K/R-R (Fig. 9b) (Zhirnov et al. 2002; Chaipan et al. 2009)(Chaipan et al. 2009). Se sabe que otros organismos como *Staphylococcus aureus* y *Pseudomonas aeruginosa*, también pueden liberar éste tipo de proteasas.

Una vez ligada la HA del virus al ácido siálico, la reacción de la célula es fagocitar al virus para después poder degradarlo mediante vesículas endosomales con un pH bajo (Fig. 10). Cuando el pH del endosoma celular es aproximadamente 5.0, la HA sufre un cambio conformacional que genera se rompa la unión con el ácido siálico y lo libera, permitiendo que la zona hidrofóbica que había sido cortada con la

enzima, se una a la membrana de la célula hospedera (Fig. 11a). Es así que el pH bajo hace que las regiones hidrofóbicas se introduzcan más en la membrana del endosoma (Fig. 11b) y, al hacer más pequeño el espacio entre las dos membranas, estas finalmente se fusionan (Fig. 11c). Antes de que esto pase la proteína M2 se activa con el pH bajo y comienza a bombear protones al interior del virión para que el vRNA se pueda desprender de la RNP (Sanjuan et al. 2009).

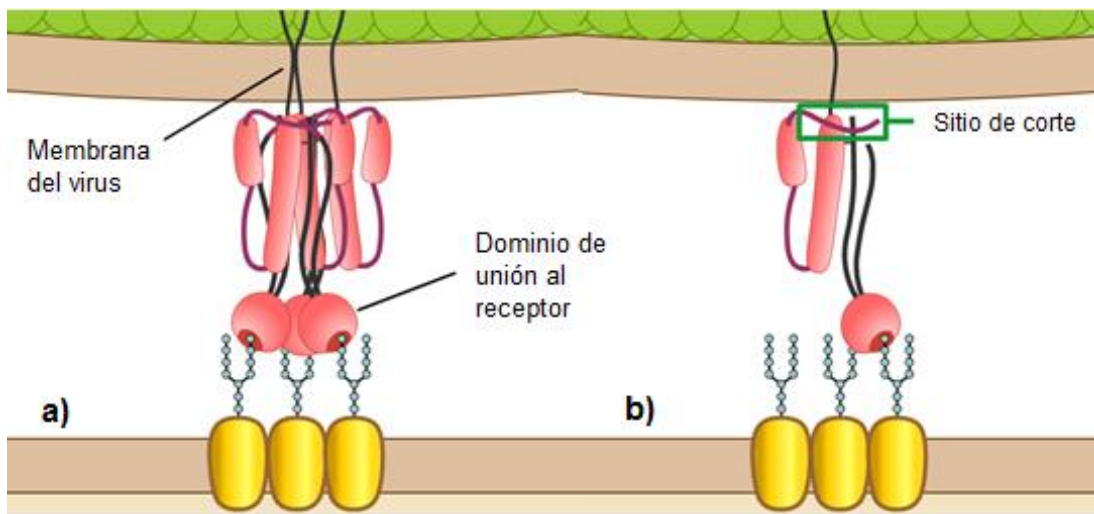


Figura. 9. Hemaglutinina. a) unión de la hemaglutinina al receptor galactosa-ácido siálico, b) monómero de hemaglutinina después del corte peptídico dando como resultado HA1 y HA2 (Slonczewski & Foster 2008).

Posteriormente, el vRNA se introduce en el núcleo de la célula hospedera junto con NP, PA, PB1 y PB2, proteínas que tienen señales de localización del núcleo para poder ingresar, y se lleva a cabo la transcripción y la replicación del genoma viral (Samji 2009). El ensamblaje del virus requiere varios pasos, entre ellos todas las enzimas sintetizadas deben ensamblarse en la membrana de células no polarizadas o en el plasma de la membrana apical de células epiteliales polarizadas; Esto se

logra mediante la red de transportación *trans*-Golgi que se mencionó anteriormente. En este punto la neuroaminidasa juega un papel importante en la replicación del virus, ya que esta proteína libera al virus de la superficie del hospedero removiendo el ácido siálico de los glicoconjugados celulares y las glicoproteínas virales mediante una hidrolisis (2,3), (2,6) y (2,8) (Tadanobu et al. 2012).

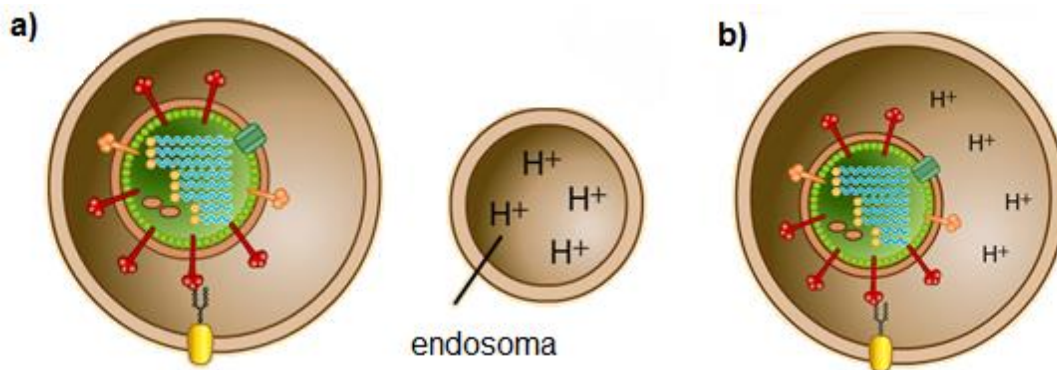


Figura 10. Fagocitosis del virus de la influenza. a) Virus fagocitado y vesícula con pH ácido en su interior, b) endosoma con pH bajo (Slonczewski & Foster 2008).

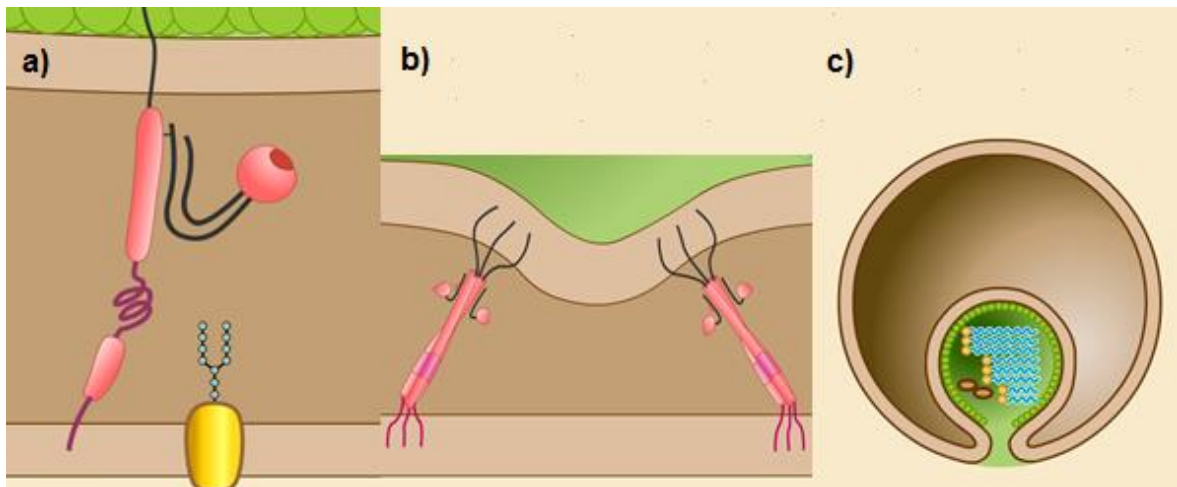


Figura 11. Efecto del pH en la fagocitosis. a) el pH bajo hace que la zona hidrofóbica que ha sido cortada de la hemaglutinina se introduzca a la membrana del endosoma, b) entre más bajo el pH se hace más pequeña la distancia entre las membranas ya que las zonas hidrofóbicas se introducen más, c) membrana viral fusionándose con la membrana del endosoma (Slonczewski & Foster 2008).

## 2.4. PROTEÍNAS POTENCIALES PARA UNA VACUNA “UNIVERSAL”

La deriva antigénica de los virus de la influenza es lo que hace que estos muten rápidamente, razón por la cual es necesario hacer vacunas anuales o bien, de crear una vacuna universal. Para el diseño de dicha vacuna, en años recientes se ha estudiado la capacidad neutralizante de tienen algunos anticuerpos hacia el virus. Se sabe que dichos anticuerpos reconocen las regiones conservadas de la hemaglutinina (zonas del tallo), por lo cual estas regiones se han tomado como puntos de partida para elaborar potenciales vacunas universales en las que el prototipo consiste en remover la cabeza y dejar expuesto el tallo. Cabe mencionar que esta vacuna ya se ha usado en ratones (Kilbourne 2006; Li et al. 2010).

El problema de usar la hemaglutinina completa como vacuna es la característica inmunodominante de la cabeza, lo que genera una respuesta inmunitaria sesgada hacia esta zona y una limitada reacción al área del tallo. En respuesta a esta problemática se han realizado experimentos en donde se expresa HA sin cabeza. En los ensayos inmunológicos *in vitro* se ha podido observar que los anticuerpos neutralizantes generados por los ratones inmunizados con esta vacuna experimental reconocen la HA de influenza H1N1, H2N2, H5N1 y H6N1. En experimentos *in vivo* con esta vacuna, ratones inoculados con la vacuna, posteriormente, han sido infectados con las cepas H1N1, H6N1 y H5N1. Los resultados mostraron que los ratones presentaban una protección total contra H1N1 y H6N1 y, una protección parcial contra H5N1, en ambos casos resistieron la infección con una pérdida mínima de peso (Krammer 2015).

En otro experimento se aisló un anticuerpo llamado CR6261 (IgG1), de un repertorio de individuos sanos vacunados. El anticuerpo fue seleccionado mediante HAH5 recombinante expresada en fagos. Se sabe que CR6261 también es capaz de neutralizar H1,H2,H5,H6,H8 y H9 y, además, fue capaz de proteger a ratones de H1N1 y H5N1 aplicándolo 5 días después de la infección (Ekiert et al. 2009). La caracterización del epítipo de HA y el mecanismo de neutralización se realizó mediante cristalización del complejo antígeno-anticuerpo utilizando los HA de las cepas "A/SouthCarolina/1/1918; SC1918H1" (responsable de la denominada gripe española de 1918) y la cepa "A/Vietnam/1203/2004; Viet04/H5" (causante de una gripe altamente patogénica) (Igarashi et al. 2010; Wang et al. 2009). Durante la maduración viral cada polipéptido HA es separado, permitiendo que el anticuerpo CR6261 se una a la zona proximal de la membrana de cada uno de los polipéptidos e interactúa principalmente con la alfa hélice de la HA2, y con los residuos de la región del tallo de HA1. La interacción Fab (fracción de unión al antígeno) es mediada exclusivamente por la cadena pesada y los epítopes a los que se une, son altamente conservados por su importante funcionalidad para la hemaglutinina. La región conservada de la HA1 donde se une el anticuerpo CR6261 consiste en Val<sup>40</sup>, Leu<sup>42</sup>, y Leu<sup>292</sup> (V,L,L) junto con los residuos de HA2 Thr<sup>49</sup>, Val<sup>52</sup> y Ile<sup>56</sup> (T,V,I) (Ekiert et al. 2009; Sui et al. 2009; Greenbaum et al. 2009; Li et al. 2010).

En el sistema inmunológico, los linfocitos T citotóxicos (LTC) responden a la NP del virus de la influenza A, pero el reconocimiento no depende de una estructura definida lo que sugiere que la NP puede ser reconocida después de la degradación de la célula blanco. Se ha demostrado con experimentos que la proteína completa

NP es ineficiente para sensibilizar a los LTC que realizarán la lisis de las células blanco, en comparación con la sensibilización de los LTC por péptidos de la NP destrozada (Carragher et al. 2008). En la tabla IV se muestran los péptidos de la nucleoproteína de ratón que se usaron para sensibilizar a los LTC y su análogo en humanos. El resultado de este experimento fortalece la hipótesis de que las proteínas virales son degradadas y presentadas por medio del complejo mayor de histocompatibilidad clase I a los LTC (Varich et al. 2009). El amino terminal de la NP contiene 87 aminoácidos, el epítipo conservado de la NP, que es bien reconocido por los LTC, consta de los aminoácidos del 50-63 que contiene la secuencia Arg-Leu-Ile-Gln (R,L, I, Q) (Ketha & Atreya 2008).

TABLA IV. Péptidos de la nucleoproteína del virus de la influenza A reconocidos por Linfocitos T Citotóxicos presentados por el complejo mayor de histocompatibilidad I.

Secuencia	Especificidad	Elemento de restricción
<b>365-380 (IASNENMETMESSTLE)</b>	1934-1945	H-2D
<b>365-379(IASNENMDAMESSTL)</b>	Todas las A	H-2D
<b>50-63 (SDYEGRLIQNSLTI)</b>	Todas las A	H-2K
<b>335-349(SAAFEDLRVLSFIRG)</b>	Todas las A	HLA-B37

Las proteínas arriba mencionadas son muy importantes tanto para la propagación del virus como para la activación de la respuesta inmune del hospedero. Estudios recientes demostraron que la inoculación en ratones con las proteínas recombinantes HA, NA y M1, expresadas en baculovirus pero basadas del virus

A/meerkat/Shanghai]/SH-1/2012, brindaban protección efectiva contra diferentes cepas de influenza aviar altamente patogénica como la (HPAI) H5. El mecanismo de la vacuna consiste en activar la respuesta humoral y celular y, se demostró que la inoculación con las proteínas por separado brindaban un 100% de sobrevivientes en comparación a los organismos inoculados con el virus completo (Ren et al. 2015; Fields 1985).

Derivado de las investigaciones anteriores, actualmente se están realizando pruebas médicas de fase II con 443 adultos a los cuales se les inocula la vacuna experimental Multimeric-001 (M-001). Ésta vacuna se diseñó con base en los epítomos conservados que son reconocidos por los linfocitos T y B que reconocen a la influenza. Estos epítomos pertenecen a las proteínas HA, NP y proteínas M de las cepas de influenza A y B. Los nueve epítomos están combinados y triplicados en una sola proteína recombinante expresada en *E. coli* y, el hecho de expresar los péptidos en una sola proteína en vez de varias minimiza los costos de producción (Atsmon et al. 2014). Con base en los resultados anteriores, podemos inferir que para crear una vacuna universal contra las diferentes cepas de la influenza lo ideal es expresar una proteína recombinante con epítomos conservados de las diversas proteínas del virus, tanto de superficie como internas, de manera tal que sea posible activar la respuesta celular y humoral del sistema inmune del hospedero.

## 2.5. DISEÑO DE PROTEINAS *IN SILICO*

En los últimos 20 años se ha recopilado nueva información sobre los patrones de péptidos que se unen al complejo mayor de histocompatibilidad (MHC) y que son presentados para la estimulación de los receptores de las células T del sistema inmune. Los primeros en describir estos patrones de unión a la superficie del MHC fueron Rötzschke y Falk, un ejemplo de estos patrones sería “\_ L \_ \_ \_ \_ V”, donde los espacios pueden ser cualquier aminoácido, sin embargo, si se eligiese un antígeno para un MHC individual, la diferencia alélica sería un problema. Esta problemática se minimizó con el descubrimiento de Sette donde describe los supertipos de familias alélicas de MHC, por lo cual si se encuentra un ligando que se una a un miembro de una superfamilia, por ejemplo A3, este también se unirá a A11. Este método de diseñar vacunas de regiones patogénicas que son presentadas por moléculas del complejo mayor de histocompatibilidad se ha determinado como inmunogenética reversa (De Groot et al. 2009). La tecnología ya se ha empleado en el diseño de proteínas que son utilizadas como vacunas, tal es el caso de la proteína Der f 25, la cual se realizó mediante una predicción *in silico* de los epítomos que causan alergia al polvo que son reconocidos por células T y B. Como producto de éste diseño *in silico*, actualmente esto se utiliza como vacuna contra la alergia al polvo (Prabhavathy et al. 2011). También se han realizado vacunas contra virus utilizando esta misma metodología, tal es el caso de la vacuna contra el virus Lassa, el cual atacan principalmente a los habitantes del oeste de África y tiene una mortalidad del 15-20% (Verma et al. 2015).

## 2.6. MICROALGAS PARA LA PRODUCCIÓN DE VACUNAS

Las microalgas han sido utilizadas como una alternativa para la síntesis de proteínas recombinantes para la producción de vacunas dado que tienen la ventaja de que su cultivo es fácil y económico, de que es relativamente sencillo producir grandes cantidades de biomasa y, sobretodo, porque son consideradas inocuas para los humanos. En la tabla V se hace una comparación de las características de los sistemas de producción de proteínas recombinantes tradicionales respecto a la producción en microalgas. Entre las microalgas que se han utilizado para dicho fin, se encuentra *Chlamydomonas reinhardtii*.

La microalga verde *Chlamydomonas reinhardtii* ha sido estudiada por casi sesenta años, a esta microalga se le apoda la levadura verde, ya que se multiplica velozmente y, crecer de forma fotótrofa o heterótrofa en medio nutritivo, dependiendo de la disponibilidad de los compuestos del medio y de la intensidad lumínica. Se han producido diversas proteínas recombinantes de interés médico en cloroplastos, en los cuales se ha obtenido hasta un 5% de rendimiento de proteína soluble total, mientras que la producción proteica en núcleo, como la de la eritropoyetina humana, ha generado rendimientos aproximados de 100 µg/L de cultivo. En la Tabla VI se presentan las proteínas recombinantes de uso médico producidas en el cloroplasto de *C. reinhardtii*. Cabe mencionar que la diferencia entre la inserción del transgen en núcleo y en cloroplasto, es que en núcleo la inserción es azarosa, y puede haber depleción o silenciamiento de algún gen, mientras que en cloroplasto la recombinación homóloga se hace de manera dirigida mediante la transformación.

TABLA V. Comparación de las características de los sistemas de producción de proteínas recombinantes tradicionales respecto a las microalgas.

Organismo	Costo de producción	Tiempo de producción	Capacidad de escalamiento	Calidad de producto	Capacidad de glicosilación	Riesgo de contaminación	Costo de almacenamiento
<b>Bacterias</b>	Bajo	Corto	Alto	Bajo	No	Endotoxinas	Moderado
<b>Levaduras</b>	Medio	Medio	Alto	Medio	Si	Bajo	Moderado
<b>Células de insecto</b>	Alto	Medio	Medio	Alto	Si	Alto	Costoso
<b>Células de mamíferos</b>	Alto	Largo	Muy bajo	Muy alto	Si	Virus-priones	Costoso
<b>Animales</b>	Alto	Muy largo	Bajo	Muy alto	Si	Virus-priones	Costoso
<b>Células de plantas</b>	Bajo	Corto	Alto	Alto		Bajo	Poco costoso
<b>Plantas</b>	Muy bajo	Largo	Muy alto	Alto	Núcleo: Si Cloroplasto: no	Bajo	Poco costoso
<b>Microalgas</b>	Muy bajo	corto	Muy alto	Alto	Núcleo: Si Cloroplasto: no	Bajo	Poco costoso

TABLA VI. Proteínas recombinantes de uso médico producidas en el cloroplasto de *C. reinhardtii*

Gen expresado	Tipo de proteína	Función	Promotor + 5'UTR/3' UTR	Rendimiento (proteína soluble total, PST)	Aplicación	Referencia
<b>HSV8-Isc</b>	Anticuerpo	Actividad contra la glicoproteína D del virus herpes simple	atpA/rbcL	0.5% PST	Anticuerpos monoclonales	Mayfield et al. 2003
<b>83K7C</b>	Anticuerpo derivado de IgG1	Vacuna oral contra <i>S. aureus</i>	LC: psbA/rbcL HC: psbA/psbA	0.1% de biomasa seca	Vacunal oral	Tran et al, 2009
<b>CTB-D2</b>	Toxina B de cólera fusionada con el dominio D2 de la proteína de unión a fibronectina	Tratamiento del cáncer	rbcl/rbcL	0.7% PST	Terapéutico	Dressen et al, 2010
<b>hTRAIL</b>	Factor de necrosis tumoral	Estimulación de la producción de mucina en el intestino para inmunizar a recién nacidos	atpA/rbcL	0.43%-0.67%	Terapéutico	Yang et al, 2006
<b>M-SAA</b>	Suero bovino amiloide asociado a las glándulas mamarias	Diagnóstico temprano de la diabetes tipo I	psbA/psbA	5.0% PST	Terapéutico	Manuell et al, 2007
<b>hGAD65</b>	Descarboxilasa de ácido glutámico humana 65	Simil de anticuerpo	rbcl/rbcL	0.25-0.3% PST	Terapéutico	Wang et al, 2008
<b>M-SAA-10NF3</b>	10° dominio de la fibronectina III humana	Simil de anticuerpo	psbA/psbA	~3.0% PST	Terapéutico	Rasala et al, 2010
<b>14NF3</b>	14° dominio de la III humana	Tratamiento del enfisema pulmonar y de la disfunción eréctil; antidepresivo	psbA/psbA	3% PST	Terapéutico	Rasala et al, 2010
<b>VEGF</b>	Factor de crecimiento endotelial vascular humano	Curación de heridas mediante la activación de células endoteliales	psbA/psbA	2% PST	Terapéutico	Rasala et al, 2010

### **3. HIPÓTESIS**

La proteína Vac-Quimera diseñada a base de epítomos conservados altamente inmunogénicos del virus de la influenza A genera en el hospedero una respuesta inmune humoral y celular que puede ser utilizada como fuente potencial de una vacuna universal contra éste virus.

### **4. OBJETIVOS**

#### **4.1. OBJETIVO GENERAL**

Modelar *in silico* una proteína conformada de epítomos conservados de influenza A (H1N1 y H5N1), determinar si activa al sistema inmune humoral y celular, y construir su casete de expresión.

#### **4.2. OBJETIVOS ESPECÍFICOS**

- 1.- Realizar un análisis fisicoquímico de la secuencia de la proteína mediante la media de las características intrínsecas de cada residuo utilizando el servidor ProtParam.
- 2.- Predicción de la estructura secundaria de la proteína utilizando el servidor NSP@.
- 3.- Modelado de la estructura tridimensional mediante los algoritmos Phyre<sup>2</sup>, Robetta, QUARK y I-TASER.

4.- Análisis de la calidad estructural de la proteína modelada mediante el servidor ProTSAV

5.- Determinación de la solubilidad de la proteína modelada mediante un algoritmo realizado por el grupo Warwicker de Bioinformática de la universidad de Manchester Inglaterra.

6.- Determinar si la proteína es capaz de activar el sistema inmune humoral y celular con base en sus epítopes utilizando los servidores ProPred1, MHC-Pred, VaxiJen y ElliPro.

7.- Clonación del gen de la proteína Vac-Quimera en el vector pChlamy\_4 mediante la ligación de sitios pegajosos utilizando T4 ligasa.

## **5. METODOLOGÍA**

### **5.1. MODELADO *IN SILICO* DE LA PROTEÍNA VAC-QUIMERA**

#### **5.1.1. ELECCIÓN DE EPÍTOPOS CONSERVADOS**

Se seleccionaron 4 proteínas de alta importancia para el ingreso del virus a la célula hospedera, así como para la replicación y propagación de este. Las proteínas seleccionadas fueron 1) hemaglutinina, 2) neuroaminidasa, 3) proteína de matriz 2 (las tres de la membrana) y 4) nucleoproteína (proteína del interior de la cápside).

El epítipo de la proteína de matriz 2 se seleccionó del ectodominio de la cepa H5N1 A/HongKong/156/97, ya que esta zona es conservada y dicha cepa es altamente patogénica (Pichyangkul et al. 2009; Gabbard et al. 2009; Tompkins et al. 2007; Greenbaum et al. 2009). En la tabla VII se muestran las secuencias de diversos ectodominios de proteínas de matriz 2 en virus de la influenza A, en donde los sitios conservados se indican con amarillo.

TABLA VII. Secuencias de diversos ectodominios de proteínas de matriz 2 en virus de la influenza A (Tompkins et al. 2007). Los sitios conservados se indican en color amarillo en las secuencias.

Cepa	Abreviación	Subtipo	Secuencia
<b>A/Tailandia/SP-83/04</b>	M2e-H5(SP-83)	H5N1	MSLLTEVETPT <b>TR</b> NEW <b>E</b> CRC <b>S</b> DSSD
<b>A/HongKong/156/97</b>	M2e-H5(HK)	H5N1	MSLLTEVET <b>L</b> TRNGWGCR <b>C</b> S <b>D</b> SSD
<b>A/FM/1/47-MA</b>	M2e-FM	H1N1	MSLLTEVETPT <b>K</b> NEW <b>C</b> ERCNDSSD
<b>A/PR/6</b>	M2e-PR6	H1N1	MSLLTEVETPIRNEWGCR <b>C</b> N <b>G</b> SSD
<b>Consenso</b>	M2e-Con		MSLLTEVETPIRNEWGCR <b>C</b> N <b>D</b> SSD

De la nucleoproteína se eligió la zona conservada de la región S2 de la cepa H1N1 A/PuertoRico/8/34, la cual tiene 11 aminoácidos idénticos a las cepas (A/California/VRDL395/2009), (A/Sendai/TU60/2009(H1N1)) (Greenbaum & Kгаа 2008; Xing & Cardona 2009; Liu & Zhao 2010) entre otras cepas con las que se alineó para hacer la comparación de secuencias con la herramienta Blast del National Center for Biotechnology Information (<http://blast.ncbi.nlm.nih.gov/>) (S. F. Altschul et al. 1997).

En el caso de la hemaglutinina se tomó la única secuencia universalmente conservada de todos los subtipos de influenza A y dos linajes de influenza B que se tienen registrados, esta secuencia corresponde a la región del péptido de fusión (Hu

et al. 2013; Chun et al. 2008). En la tabla VII se muestran las secuencias de péptidos de fusión para diversas cepas del virus de la influenza.

TABLA VIII. Comparación entre algunos péptidos de fusión de diversas cepas conocidas del virus de la influenza (Chun et al. 2008).

Cepa	NCBI No. acceso	Subtipo	Secuencia
<b>A/NewCaledonia/20/99</b>	AAP34324	H1N1	GLFGAIAGFIEGGW
<b>A/HongKong/1073/99</b>	EF154959.1	H9N2	GIFGAIAGFIEGGW
<b>A/Netherlands/219/03</b>	AB307735.1	H7N7	GLFGAIAGFIENGW
<b>A/Wisconsin/67/2005</b>	EU100670.1	H3N2	GIFGAIAGFIENGW

De la neuroaminidasa se eligió un epítipo que se sabe genera respuesta inmune de los linfocitos colaboradores CD4<sup>+</sup>T. La secuencia se alineo también con el programa Blast y como resultado se observó que la región estaba altamente conservada en más de 100 cepas de influenza (Greenbaum et al. 2009; Roti et al. 2013; Homan & Bremel 2011) .

### 5.1.2. ANÁLISIS FÍSICOQUÍMICO DE LA ESTRUCTURA PRIMARIA DE VAC-QUIMERA

El análisis físicoquímico se llevó a cabo con la herramienta ProtParam (<http://web.expasy.org/protparam/>), la cual permite determinar *in silico* varios parámetros físicos y químicos de una proteína determinada en tres organismos distintos: células de mamífero, levaduras y *E. coli*; el programa permite además extrapolar los resultados obtenidos a un organismo particular (Dos Santos & Soares de Souza 2014; Gasteiger et al. 2005 ). En el análisis se incluyó la determinación de los siguientes parámetros: peso molecular, punto isoelectrico (PI) teórico, composición de aminoácidos, índice de inestabilidad (el cual debe ser menor a 40

para considerarse estable), índice alifático (volumen relativo ocupado por las cadenas laterales alifáticas: alanina, isoleucina, valina y leucina, donde valores positivos para proteínas globulares termoestables), coeficiente de extinción y, promedio de hidropaticidad (GRAVY) (correspondiente a la media de los valores de hidropaticidad de los aminoácidos). Cuando se presentaron zonas con puntos de fosforilación (con un 80% de probabilidad de aparición) éstas fueron analizadas con NETPhos v2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>).

### 5.1.3. RECONOCIMIENTO DE HOMOLOGÍA

Se buscó homología en la secuencia de la proteína Vac-Quimera para tener como template esas estructuras en el modelado posterior mediante el servidor PSI-BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Stephen F. Altschul et al. 1997).

### 5.1.4. PREDICCIÓN DE LA ESTRUCTURA SECUNDARIA

La identificación de los elementos que conforman la estructura secundaria, con base en la secuencia de aminoácidos de la proteína Vac-Quimera, alfa hélices, hojas beta, giros y flexibilidad, fueron predichos con DNASTar protean system, Phyre<sup>2</sup>, y el servidor NSP@ (<https://npsa-prabi.ibcp.fr/>). Dicho programa hace un consenso de 7 algoritmos diferentes, divididos en dos fundamentos distintos: los basados en las frecuencias relativas de las apariciones de cada residuo (SOPM, PREDATOR) y, los basados en la probabilidad de inferencia bayesiana (GOR4, DPM DSC) (Díaz de los Ríos 2013).

### 5.1.5. IDENTIFICACIÓN DE EPITOPOS EN CÉLULAS B Y EN MHC

Los epítopes que son reconocidos por células B (anticuerpos), se identificaron utilizando cinco programas computacionales diferentes. El primer análisis fue realizado con bioinformatics predicted antigenic peptides (BPAP) system (<http://imed.med.ucm.es/Tools/antigenic.pl>), el cual se basa en el método de Kolaskar y Tongaonkar (Kolaskar & Tongaonkar 1990), al igual que el segundo programa llamado Antibody Epitope Prediction ([http://tools.immuneepitope.org/tools/bcell/iedb\\_input](http://tools.immuneepitope.org/tools/bcell/iedb_input)). El tercer programa utilizado fue DNASTar, el cual basa su reconocimiento en la identificación de patrones en la estructura secundaria, tales como flexibilidad, hidropaticidad y antigenicidad sugestiva. El cuarto programa de análisis fue ElliPro: antibody epitope prediction (<http://tools.iedb.org/ellipro/>) (Ponomarenko et al. 2008), cuyo fundamento es el modelo oculto de Markov (Eddy 2004), al igual que la herramienta BepiPred que también fue utilizada para determinar epítopes (<http://tools.immuneepitope.org/tools/bcell/BepipredDisplayServlet>) (Erik et al. 2006).

Para determinar que alelos que expresan el complejo MCH-I reconocen a la proteína Vac-Quimera se utilizó la herramienta ProPred1 (<http://www.imtech.res.in/cgibin/propred1/propred1.pl>) (Singh & Raghava 2003) de igual manera se realizó para el complejo MHC-II utilizando la herramienta ProPred (<http://www.imtech.res.in/raghava/propred/>) (Singh & Raghava 2002). Posteriormente, la secuencia se analizó en el programa MHCPred (<http://www.ddg-pharmfac.net/mhcpred/MHCPred/>), el cual hace un análisis de la afinidad con la que

se une el epítipo a los complejos MHC-I y II. La afinidad se midió en IC<sub>50</sub>, siendo los menores a 5000 más afines ( Guan et al. 2003a; Guan et al. 2003b; Hattotuwagama et al. 2004). Para la identificación específica del MHC-I (Lundegaard et al. 2008) y MHC-II (Wang et al. 2010) por separado, se utilizó Immune Epitope Database and Analysis Resource (<http://www.iedb.org/>), en la cual se tiene una opción independiente para cada tipo de complejo.

#### 5.1.6. MODELADO TRIDIMENSIONAL DE LA PROTEÍNA

El modelado final de la proteína se realizó en Phyre<sup>2</sup>, Robetta, QUARK y I-TASER, tomando como templete principal la estructura tridimensional de la neuroaminidasa. Una vez obtenido los modelos tridimensionales en la base de datos de proteínas (PDB, [www.uniprot.org](http://www.uniprot.org)), estos se visualizaron en PyMOL.

#### 5.1.7. ANÁLISIS DE LA ESTRUCTURA FINAL DE VAC-QUIMERA

##### 5.1.7.1. Análisis de Estereología

Para la validación del modelado tridimensional de la proteína se utilizó el programa ProTSAV (<http://www.scfbio-iitd.res.in/>), el cual se basa en los siguientes diez métodos bien reportados y probados:

Naces: Calcula la disposición con la que los átomos y los residuos que se encuentran en la superficie de la proteína pueden disolverse, considerando óptimas

aquellas proteínas que mejor mantienen su estructura nativa al contacto con el agua.

Verify 3D: Es un algoritmo que compara la compatibilidad de la estructura tridimensional de la proteína con su estructura primaria, la cual está dada por las propiedades intrínsecas de cada aminoácido que la conforma.

Errat: Esta herramienta distingue si una región determinada es correcta o incorrecta de acuerdo a las características de las interacciones atómicas de la proteína; es considerada una muy buena herramienta ya que tiene un 95% de resolución.

Procheck: Evalúa la calidad estereoquímica midiendo longitud de uniones, ángulos de uniones, cadena principal y cadena lateral, contacto entre aminoácidos, geometría y distribución de los ángulos de torsión  $\Psi$  y  $\Phi$  plasmados en una gráfica de Ramachandran.

MolProbity: Utiliza una variedad de algoritmos basados en física y en los conocimientos de las estructuras; esta herramienta es la mejor para validar contacto entre todos los átomos, paridad entre cadenas laterales y, distribución de los ángulos de torsión en la gráfica de Ramachandran.

ProSA: Es una herramienta que diagnostica la estructura terciaria de acuerdo a los análisis estadísticos derivados de estructuras de proteínas experimentales, genera gráficas en valores de Z-score, las energías de los aminoácidos de la proteína problema y las proteínas de las bases de datos, tomando como mejores las que son más similares a las nativas.

dDFIRE: Determina la cantidad de energía libre de cada conformación de la proteína y considera como de mejor calidad aquellas estructuras con valores de energía libre bajos.

D2N: Predice la calidad de una estructura con base en seis diferentes características fisicoquímicas, energía de todos los átomos no unidos, accesibilidad total del área de superficie, restricción geométrica C $\beta$ , similitud a la estructura secundaria, superficie accesible a solventes área de aminoácidos polares y aminoácidos cargados, y delibera en puntajes rmsd TM y GDT para estructuras sin tener que usar información de la proteína nativa.

ProQ: Evalúa la calidad de un aminoácido empleando una base de datos donde integra los contactos entre átomos y aminoácidos, superficies accesibles a solventes y, estadística de estructura secundaria. Predice la calidad estructural de los modelos en base a las características medidas en puntaje LG y puntaje MaxSub Score. Si el puntaje LG es mayor 3 y el puntaje MaxSub mayor 0.5 representa que la estructura es de buena calidad.

PSN-QA: Esta herramienta utiliza una base de datos (networks) de estructuras de proteínas nativas y modeladas en combinación con máquinas de soporte de vector, y con base a esto estima la calidad estructural de las proteínas, dicha base de datos fue construida a partir de interacciones no covalentes entre cadenas de polipéptidos. Para poder determinar la calidad de la proteína compara la similitud entre la proteína problema y 16 representantes de conformaciones nativas y 10 representantes de conformaciones no nativas.

El puntaje que muestra el programa es el promedio de todos los valores obtenidos por los diez algoritmos, todos representados en valores Z (Z-score) (Singh et al. 2016).

#### 5.1.7.2. Solubilidad

Se analizó la solubilidad de la proteína tridimensional en colaboración del grupo Warwicker de Bioinformática estructural en comunicación personal de Alejandro Carballo, utilizando el programa Protein-Sol basado en la clasificación de ciertas áreas superficiales de la proteína, las cuales el algoritmo las denomina como parches, marcando en color azul las zonas con carga positiva, en rojo las áreas con carga negativa y con blanco las regiones sin carga (hidrofóbicas), ya que las áreas con carga positiva son las que determinan la solubilidad, el algoritmo solo se enfoca en estas para determinar si la proteína es soluble, los valores que en el algoritmo representan solubilidad son los que la relación tamaño de parche / threshold size distance sea menor a 1, dicho algoritmo aún no está disponible al público, sin embargo tiene una versión basada en la secuencia de aminoácidos (<http://protein-sol.manchester.ac.uk/>) (Chan et al. 2013).

#### 5.1.7.3. Superposición de la Estructura Tridimensional

Para determinar que la estructura tridimensional modelada tuviera homología con otras proteínas existentes se utilizó PhyreStorm, programa que hace una alineación tridimensional con proteínas cristalizadas de la base de datos de proteínas (PDB).

## 5.2. CONSTRUCCIÓN DE CASSETTE DE EXPRESIÓN

### 5.2.1. SÍNTESIS DEL GEN QUE CODIFICA PARA VAC-QUIMERA

El gen que codifica para la proteína Vac-Quimera se mandó sintetizar en el plásmido pUC57 con codones optimizados para cloroplasto de *Dunaliella salina* en la compañía GeneScript.

### 5.2.2. AMPLIFICACIÓN POR PCR

Se realizó una amplificación del gen Quimera por medio de la reacción en cadena de la polimerasa (PCR), utilizando la enzima HotStar Taq Plus QIAGEN®, con los cebadores de la Tabla IX a una temperatura de alineación de 63°C por 0.50 min. El producto de PCR obtenido se precipitó con acetato de sodio.

TABLA IX. Cebadores utilizados para amplificar el gen Vac-Quimera mediante Reacciones en Cadena de la Polimerasa (PCR).

Nombre	Secuencia 5´-3´
<b>EcoRI_F</b>	CGCAG <b>AATT</b> CCCATGGCACCACCACCA
<b>PstI_R</b>	GATC <b>CTGCAG</b> TTAACCTGAAGTCCAATAG

### 5.2.3. DIGESTIÓN ENZIMÁTICA

Se realizó una doble digestión enzimática con las enzimas EcoRI HF RE-Mix y PstI HF de NewEngland BioLabs®, tanto al producto de PCR (gen Quimera), como al vector (pChlamy\_4), utilizando un tiempo de incubación de 1hr a 37°C. Se digirieron 1.5 µg de producto de PCR (gen Quimera) previamente precipitado, y directamente de la digestión se purificó, tratándolo como producto de PCR con el kit ilustra GFX

PCR DNA and Gel Band Purification de GE Healthcare®. Posterior a la digestión del vector pChlamy se realizó una electroforesis y se purificó el fragmento a partir de gel con el kit illustra GFX PCR DNA and Gel Band Purification de GE Healthcare®.

#### 5.2.4. LIGACIÓN DEL GEN EN EL VECTOR PCHLAMY\_4

Se realizó la ligación del gen Quimera y el vector pChlamy\_4 con el kit T4 DNA ligase de Fermentas, utilizando 20µL (10.8ng/µL) del vector y 20 µL (9.4ng/µL) del inserto con 30 µL del Buffer de ligación 2x más 4 µL de T4 ligasa, con un volumen final de 74 µL. Se dejó incubando el producto de ligación toda la noche a temperatura ambiente y, posteriormente se precipitó con acetato de sodio.

#### 5.2.5. TRANSFORMACIÓN DEL PLÁSMIDO EN *E.COLI* XL1-BLUE

Una vez precipitada la ligación se transformaron 50 µL de células quimiocompetentes de *E.coli* XL1-Blue, utilizando el método de shock térmico. Se sembraron cuatro cajas de LB-agar-80 µg/ml de ampicilina de acuerdo a lo siguiente: la caja 1 fue control negativo con células sin transformar, las cajas 2, 3 y 4 se sembraron con 50, 75 y 125 µL de células, respectivamente, transformadas en LB sin antibiótico. Se dejaron crecer 24 hrs. a 37°C, se eligieron 3 colonias de cada caja para crecer en medio LB líquido con 80 µg/ml de ampicilina, durante 14 hrs. a 37°C.

## 5.2.6. CONFIRMACIÓN DE LA TRANSFORMACIÓN

### 5.2.6.1. Confirmación mediante digestión enzimática

Se tomaron tres colonias al azar de cada caja para crecerlas en medio LB, las cuales se denominaron 1A, 2A, 3A, 1B, 2B, 3B, 1C, 2C y 3C (la letra corresponde a la caja y el número a la clona). A cada colonia se le realizó una extracción de DNA plasmídico y con él se hizo una doble digestión enzimática con EcoRI HF RE-Mix y PstI HF de NewEngland BioLabs®, dicha digestión se incubó a 37°C por 1h. Los productos de la digestión se identificaron por electroforesis en geles de agarosa al 1.5%, usando buffer TAE, se corrió el gel durante 40min a 90V-1A.

### 5.2.6.2. Confirmación mediante PCR

Para confirmar la transformación bacteriana se hicieron ensayos de PCR con los cebadores previamente descritos. De las 9 clonas aisladas, se amplificó la clona 1A y 2A, utilizando los cebadores de la Tabla VIII y la enzima HotStar Taq Plus QIAGEN®, con una temperatura de alineación de 63°C por 0.50min.

### 5.2.6.3. Confirmación mediante secuenciación

Además de la confirmación por PCR, se mandó a secuenciar la clona 1A con los cebadores de la tabla VIII a la empresa EtonBioscience Inc., para confirmar que el inserto correspondía a la secuencia de interés.

## 6. RESULTADOS

### 6.1. DISEÑO DE LA PROTEÍNA

#### 6.1.1. EPITOPES CONSERVADOS

Los cuatro epítopes conservados analizados con el servidor blast presentaron un 100% de similitud con más de 200 proteínas (Ver anexo I).

#### 6.1.2. SECUENCIA PRIMARIA

Para el diseño de la proteína se decidió poner al inicio la secuencia de la proteína M2 (de los 4 epítopes conservados: Matriz 2, Nucleoproteína, Hemaglutinina y Neuroaminidasa) (Fig. 12), ya que este epítope comienza con los aminoácidos metionina (Met) y serina (Ser). Cabe recordar que ATG es el codón que codifica para la metionina y da inicio a la transcripción. Además, la metionina, al igual que la serina, le confiere mayor tiempo de vida media a la proteína final.



Figura 12. Orden seleccionado para la síntesis del gen Quimera.

### 6.1.3. ANÁLISIS FISICOQUÍMICO DE LA ESTRUCTURA PRIMARIA DE VAC-QUIMERA

El análisis fisicoquímico realizado por la herramienta Protparam, mostró que la proteína Vac-Quimera tiene un peso molecular de 7.8 kDa, un punto isoeléctrico de 5.22, un índice de inestabilidad menor a 40 (mostrándola como estable), cuyo aminoácido con mayor incidencia es la glicina (Gly) con un 13.9%, y los de menor son glutamina (Gln), metionina (Met) y prolina (Pro) con un 1.4% (Ver Anexo 2), y un promedio de hidropaticidad de -0.153 (dicho valor representa la energía libre que se libera en un proceso de transferencia de un aminoácido de un medio no polar a uno polar, (TABLA X).

TABLA X. Datos del análisis de la proteína Vac-Quimera realizado por el programa Protparam.

Parámetros	Unidades
Número de aminoácidos	72
Peso molecular	7.8 kDa
pl teórico	5.22
Número total de residuos cargados negativamente (Asp + Glu)	8
Número total de residuos cargados positivamente (Arg + Lys)	7
Vida media en eritrocitos mamíferos in vitro	30hrs
Vida media en E. coli in vivo	>10 hrs
Índice de inestabilidad II	27.52 Estable
Índice alifático	75.83
Promedio de hidropaticidad (GRAVY)	-0.153
Fórmula	C <sub>345</sub> H <sub>537</sub> N <sub>97</sub> O <sub>107</sub> S <sub>4</sub>
Número total de átomos	1090

#### 6.1.4. RECONOCIMIENTO DE HOMOLOGIA

Se analizó la secuencia de Vac-Quimera en el servidor PSI-BLAST y se reconocieron los dominios conservados de la superfamilia de Neuroaminidasas y de proteína de matriz 2 (Fig. 13). Esto nos indicó que el templete que mejor se ajusta para el diseño de la proteína es cualquier proteína perteneciente a la superfamilia de la neuroaminidasa, ya que la secuencia de la proteína tiene un 86.1% de aminoácidos que se alinean a dicha familia.

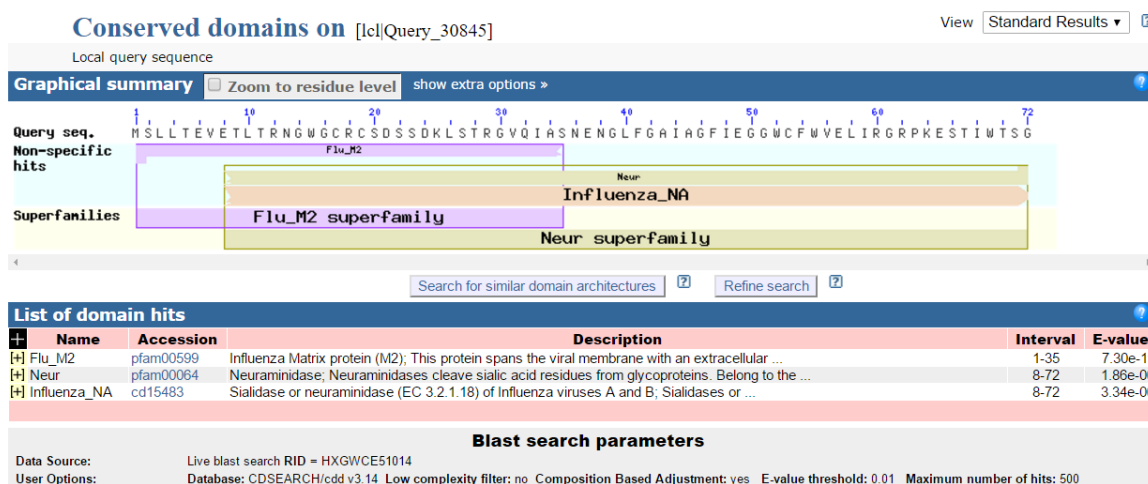


Figura 13. Análisis de secuencia de la proteína Vac-Quimera por PSI-BLAST, donde se observan los dominios conservados de las superfamilias Flu\_M2 (pfam00599) y Neur (pfam00064).

#### 6.1.5. ESTRUCTURA SECUNDARIA

La estructura secundaria predicha por el programa DNASTar proteam system (Fig. 14) consta de dos regiones de alfa hélice (1-10; 40-48), tres de hojas beta (7-10; 31-35, 53-61) y el resto de la secuencia son giros libres. En el caso del servidor Phyre<sup>2</sup> (Fig. 15), las regiones de las alfa hélices están comprendidas por los aminoácidos 2-12 y las hojas betas por los siguientes: 25-35, 40-49, 52-60, 67-70. La estructura

secundaria determinada por el servidor PROCHECK (Fig. 16) tiene una hoja que comprende de los residuos 2-11 y hojas beta de 41-46, 50-56, 58-60,69-71.

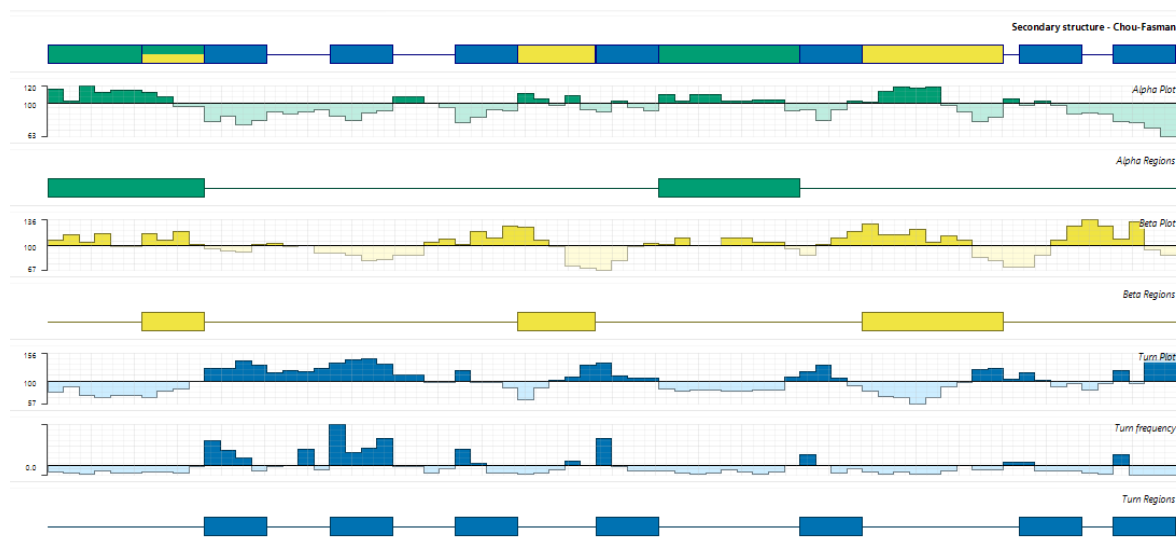


Figura 14. Predicción de estructura secundaria basada en la secuencia de la proteína Vac-Quimera empleando DNASTar protean system como herramienta.

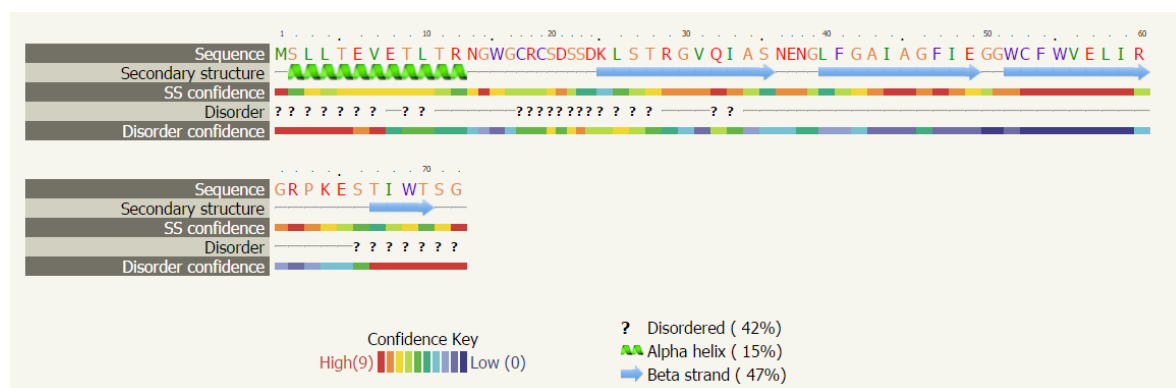


Figura 15. Estructura predicha por el servidor Phyre2 de la proteína Vac-Quimera.

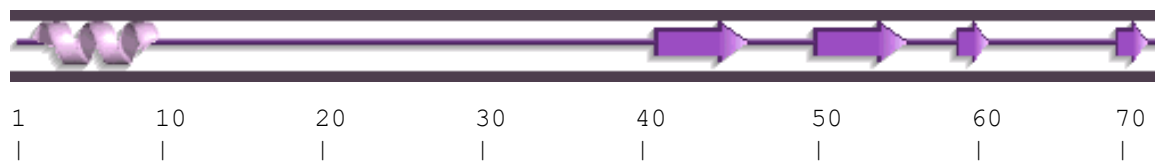


Figura 16. Estructura secundaria de la proteína Vac-Quimera predicho por el servidor PROCHECK.

La última predicción se realizó con el servidor NPSA, el cual utiliza siete algoritmos y, posteriormente, muestra un consenso de cada uno de los resultados arrojados (Fig. 17), conformado por una alfa hélice 3-10 y cuatro regiones de hojas betas (30-35, 40-47, 52-58, 67-70), sin embargo, la región conformada por los aminoácidos del 40-47 muestran un empalme entre hojas beta y alfa hélices.



Figura 17. Estructura secundaria predicha por NPSA, el consenso se encuentra en la parte inferior, alfa hélices = h, hojas beta = e, libres giros = c.

### 6.1.6. IDENTIFICACIÓN DE EPÍTOPOS EN CÉLULAS B Y EN MHC

Los epítopes que son capaces de inducir la respuesta inmunológica en células B y células T son considerados buenos candidatos para la síntesis de vacunas recombinantes. En el caso de los epítopes para anticuerpos de linfocitos B se utilizó el Immune Epitope Database and Analysis Resource IEDB, en donde se

identificaron los determinantes mediante secuencia con el método de antigenicidad de Kolaskar y Tongaonkar (Kolaskar & Tongaonkar 1990) y mediante estructura con el método ElliPro (Ponomarenko et al. 2008) (Ver Anexo 7).

TABLA XI. Determinantes antigénicos lineares reconocidos por anticuerpos.

Antigenicidad			ElliPro		
Longitud de péptido	Péptido	Proteína	Longitud de péptido	Péptido	Proteína
6	LTEVET	M2	14	IRGRPKESTIWTSG	NA
7	GCRCSDS	M2	4	IEGG	HA
7	TRGVQIA	NP	20	TLTRNGWGCRCS DSSDKLST	M2
7	FGAIAGF	HA		TLTRNGWGCRCS DSSDKLST	NP
7	CFWVELI	NA			

Con la herramienta BepiPred1 se determinó cuáles de los 47 alelos que expresan el complejo mayor de histocompatibilidad que se encuentran en la base de datos de esta herramienta pueden reconocer de manera individual a los cuatro epítopes de la proteína Vac-Quimera, en la tabla XI se observa que péptidos correspondientes a cada una de los cuatro fragmentos de las proteínas de la influenza seleccionadas son procesadas y presentadas por los 47 diferentes complejos MHC-I, además cabe resaltar que el grupo A2 donde se encuentran los alelos más frecuentes de la población mundial reconoce todos los epítopes de la proteína, sin embargo, en los epítopes correspondientes a las proteínas HA y NA comparten los aminoácidos IEGGWCFWVEL, posiblemente *in vivo*, la respuesta inmune celular específica a estas proteínas sea deficiente.

TABLA XI. Epítopes reconocidos por el MHC-I de los serotipos HLA-A, HLA-B y HLA-C

Alelo	M2 MSLLTEVETLTRNGWGCRCSDDSSD	NP KLSTRGVQIASNEN	HA GLFGAIAGFIEGGW	NA CFWVELIRGRPKESTIWTSG
HLA-A1	LTEVETLTR EVETLTRNG	SNENGLFGA	SNENGLFGA	
HLA-A2	SLLTEVETL	KLSTRGVQI QIASNENGL	IEGGWCFWV	IEGGWCFWV
HLA-A*0201	SLLTEVETL LLTEVETLT TLTRNGWGC		IEGGWCFWV	
HLA-A*0205	SLLTEVETL LLTEVETLT	QIASNENGL	GGWCFWVEL	GGWCFWVEL
HLA-A3	SLLTEVETL	KLSTRGVQI	GLFGAIAGF	
HLA-B14	SLLTEVETL	KLSTRGVQI DKLSTRGVQ	GGWCFWVEL	
HLA-C*0301	SLLTEVETL RCSDSSDKL	NENGLFGAI	GFIEGGWCF	

Se analizó la afinidad de unión de los epítopes al complejo MHC-I correspondientes al alelo HLA-A\*0201 ya que era el único disponible en la herramienta MHC-Pred, la afinidad se obtuvo en IC<sub>50</sub>. Valores menores de 5000 indican buena unión de los péptidos al complejo (Ver anexo 3). Posteriormente, se analizó cada uno de los péptidos que resultaron del análisis con MHC-Pred con la herramienta VaxiJen, en la cual se determinó cuáles de estos son considerados como antígenos a aquellos péptidos con un puntaje mayor a 0.4. Como resultado se obtuvieron 19 péptidos antigénicos de 34 que son reconocidos por el complejo mayor de histocompatibilidad tipo I, y de los 19 que son antigénicos, 7 son fragmentos de proteínas sin traslaparse (TABLA XII).

TABLA XII. Unión de epítopes a MHC-I (HLA-A\*0201), expresado en IC<sub>50</sub>, así como el puntaje obtenido por VaxiJen.

Aminoácidos	IC50 (nM)	Confianza (Max=1)	VaxiJen	Proteína
<b>TLTRNGWGC</b>	1967.89	1.00	1.2725	M2
<b>KLSTRGVQI</b>	464.52	1.00	1.4886	NP
<b>GGWCFWVEL</b>	2238.72	1.00	0.9809	HA/NA
<b>IEGGWCFWV</b>	101.62	0.89	1.1322	HA/NA
<b>ELIRGRPKE</b>	236.59	0.78	1.2557	NA
<b>ASNENGLFG</b>	283.14	0.78	0.8031	NP/HA
<b>CFIEGGWCF</b>	508.16	0.89	1.3042	HA/NA
<b>SDKLSTRGV</b>	731.14	0.89	1.2033	M2/NP
<b>ENGLFGAIA</b>	756.83	0.78	0.5928	NP/HA
<b>TRGVQIASN</b>	948.42	0.78	0.9074	NP
<b>SNENGLFGA</b>	1196.74	0.89	0.8930	NP/HA
<b>NENGLFGAI</b>	1221.80	0.89	0.6938	NP/HA
<b>FIEGGWCFW</b>	1355.19	0.89	1.3237	HA/NA
<b>TRNGWGCRC</b>	2177.71	0.89	1.7009	M2
<b>IRGRPKEST</b>	2280.34	0.89	0.6674	NA
<b>SSDKLSTRG</b>	2301.44	0.78	1.0901	M2/NP
<b>DSSDKLSTR</b>	2754.23	0.89	0.4865	M2/NP
<b>GWCFWVELI</b>	2978.52	0.89	0.9220	HA/NA
<b>VELIRGRPKE</b>	4226.69	0.78	0.7650	NA

En el caso del análisis de los péptidos que se unen al complejo MHC-II solo se analizaron los péptidos procesados por el alelo HLA-DRB1\*01,04,07 debido a que eran los disponibles en la herramienta MCH-II, además de, pertenecer al grupo DRB1, el cual está presente en todos los individuos (Ver anexos 4, 5 y 6).

Se analizó la afinidad de unión de los epítopes al complejo MHC-II correspondientes a los alelos DRB1\*0101, DRB1\*0401 y DRB1\*0701 ya que son los que se encuentran disponibles en la herramienta MHC-Pred (TABLA XIV), la afinidad se obtuvo en IC<sub>50</sub>. Valores menores de 5000 indican buena unión de los péptidos al complejo (Ver anexo 4). Al igual que para el MHC-I se analizaron cada uno de los péptidos

TABLA XIII. Unión de epítopes a MHC-II (HLA-DRB1\*0101,\*0401\*0701), expresados en IC<sub>50</sub>

Aminoácidos	IC 50 (nM)			Confianza (Max=1)			VaxiJen	Proteína
	*0701	*0401	*0101	*0701	*0401	*0101		
<b>DSSDKLSTR</b>	4395.42	814.70	16.87	1.00	0.78	1.00	0.4865	M2/NP
<b>NENGLFGAI</b>	1109.17	638.26	53.95	1.00	0.78	1.00	0.6938	NP/HA
<b>SDKLSTRGV</b>	1253.14	2978.52	63.68	1.00	0.89	1.00	1.2033	NP
<b>LSTRGVQIA</b>	626.61	1409.29	93.76	0.89	0.78	1.00	1.5966	NP
<b>VELIRGRP</b>	4008.67	273.53	3061.96	0.89	1.00	0.78	0.7650	NA
<b>WVELIRGRP</b>	594.29	845.28	10.86	1.00	1.00	0.89	0.6822	NA
<b>RGVQIASNE</b>	60.12	763.84	6.75	1.00	0.67	0.89	0.4830	NP
<b>ELIRGRPKE</b>	145.21	691.83	61.94	1.00	0.67	0.89	1.2557	NA
<b>DKLSTRGVQ</b>	173.38	331.89	111.94	1.00	0.89	0.89	1.5104	NP
<b>ENGLFGAIA</b>	338.06	301.30	251.19	1.00	0.78	0.89	0.5928	HA
<b>EGGWCFWVE</b>	402.72	1006.93	220.29	1.00	0.33	0.56	0.9908	HA/NA
<b>IAGFIEGGW</b>	501.19	2460.37	331.89	1.00	0.89	0.89	0.6004	HA
<b>FIGGWCFW</b>	545.76	398.11	15.74	1.00	0.67	0.78	1.3237	HA/NA
<b>GGWCFWVEL</b>	717.79	625.17	100.69	1.00	0.44	0.89	0.9809	HA/NA
<b>GFIGGWCF</b>	879.02	2079.70	613.76	1.00	0.67	0.67	1.3042	HA
<b>ASNENGLFG</b>	941.89	254.68	1931.97	1.00	0.89	0.89	0.8031	NP/HA
<b>STRGVQIAS</b>	941.89	709.58	126.18	1.00	0.67	0.89	1.1199	NP
<b>RGRPKESTI</b>	950.60	1227.44	74.64	1.00	0.67	0.78	0.4336	NA
<b>GWCFWVELI</b>	1191.24	171.00	208.45	1.00	0.78	0.89	0.9220	NA
<b>TRNGWCRC</b>	1879.32	712.85	277.33	1.00	0.78	0.89	1.7009	M2
<b>WGCRCSDSS</b>	4325.14	141.58	36.06	1.00	0.67	0.67	0.6489	M2
<b>NGWCRCSD</b>	4436.09	1088.93	111.69	1.00	0.67	0.89	1.8346	M2

TABLA XIV. Péptidos antigénicos presentados por los complejos MHC-I y MHC-II con su respectivo valor antigénico

Complejo	Epítotope	VaxiJen >0.4=Antigénico	Resultado	Proteína
<b>MHC-I</b>	TLTRNGWGC	1.2725	Antígeno	M2
	TRNGWGCRC	1.7009	Antígeno	M2
	DSSDKLSTR	0.4865	Antígeno	M2/NP
	SSDKLSTRG	1.0901	Antígeno	M2/NP
	SDKLSTRGV	1.2033	Antígeno	M2/NP
	KLSTRGVQI	1.4886	Antígeno	NP
	TRGVQIASN	0.9074	Antígeno	NP
	ASNENGLFG	0.8031	Antígeno	NP/HA
	SNENGLFGA	0.8930	Antígeno	NP/HA
	NENGLFGAI	0.6938	Antígeno	NP/HA
	ENGLFGAIA	0.5928	Antígeno	NP/HA
	GFIEGGWCF	1.3042	Antígeno	HA/NA
	FIEGGWCFW	1.3237	Antígeno	HA/NA
	IEGGWCFWV	1.1322	Antígeno	HA/NA
	GGWCFWVEL	0.8909	Antígeno	HA/NA
	GWCFWVELI	0.9220	Antígeno	HA/NA
	VELIRGRPKE	0.7650	Antígeno	NA
	ELIRGRPKE	1.2557	Antígeno	NA
	IRGRPKEST	0.6674	Antígeno	NA
	<b>MHC-II</b>	TRNGWGCRC	1.7009	Antígeno
WGCRCSDSS		0.6489	Antígeno	M2
NGWGCRCSD		1.8346	Antígeno	M2
DSSDKLSTR		0.4865	Antígeno	M2/NP
SDKLSTRGV		1.2033	Antígeno	NP
STRGVQIAS		1.1199	Antígeno	NP
DKLSTRGVQ		1.5104	Antígeno	NP
LSTRGVQIA		1.5966	Antígeno	NP
RGVQIASNE		0.4830	Antígeno	NP
ASNENGLFG		0.8031	Antígeno	NP/HA
NENGLFGAI		0.6938	Antígeno	NP/HA
ENGLFGAIA		0.5928	Antígeno	NP/HA
IAGFIEGGW		0.6004	Antígeno	HA
GFIEGGWCF		1.3042	Antígeno	HA/NA
FIEGGWCFW		1.3237	Antígeno	HA/NA
EGGWCFWVE		0.9908	Antígeno	HA/NA
GGWCFWVEL		0.9809	Antígeno	HA/NA
GWCFWVELI		0.9220	Antígeno	HA/NA
WVELIRGRP		0.6822	Antígeno	NA
VELIRGRPKE		0.7650	Antígeno	NA
ELIRGRPKE	1.2557	Antígeno	NA	
RGRPKESTI	0.4336	Antígeno	NA	

### 6.1.7. ESTRUCTURA TRIDIMENSIONAL

La proteína Vac-Quimera se modeló en cuatro diferentes servidores, Phyre<sup>2</sup>, Robetta, I-TASSER y QUARK (Fig. 18). Los tres primeros programas realizan la predicción mediante homología en secuencia, en la cual alinean la secuencia problema con otras de las bases de datos y las más similares las toma como template. En la figura 18 se muestran la estructura 3D que modeló cada uno de los servidores. En el caso de QUARK, este servidor hace la predicción de *novo*, esto significa que no utiliza proteínas template, solo se basa en las características intrínsecas de los aminoácidos. En la siguiente tabla (tabla XV) se comparan las características del modelado de cada uno de los servidores utilizados, en donde se especifica la estructura template y la secuencia secundaria generada.

TABLA XV. Comparación entre el tipo de modelado de los servidores seleccionados

	Phyre <sup>2</sup>	Robetta	QUARK	I-TASSER
<b>De novo</b>	-----	-----	✓	-----
<b>Template</b>	C3salB	3cye	-----	2htvA
<b>Empalme</b>	0.70	0.79	No empalma	0.67
	NA 1SALA	NA 1V0ZC		NA 3CML
<b>Alfa hélice</b>	2-6	44-49	2-11, 24-34, 37-47	14-16, 31-33
<b>Hojas beta</b>	41-46, 49-55	7-11, 23-29, 36-40, 53-58	53-58, 65-71	35-37, 56-58

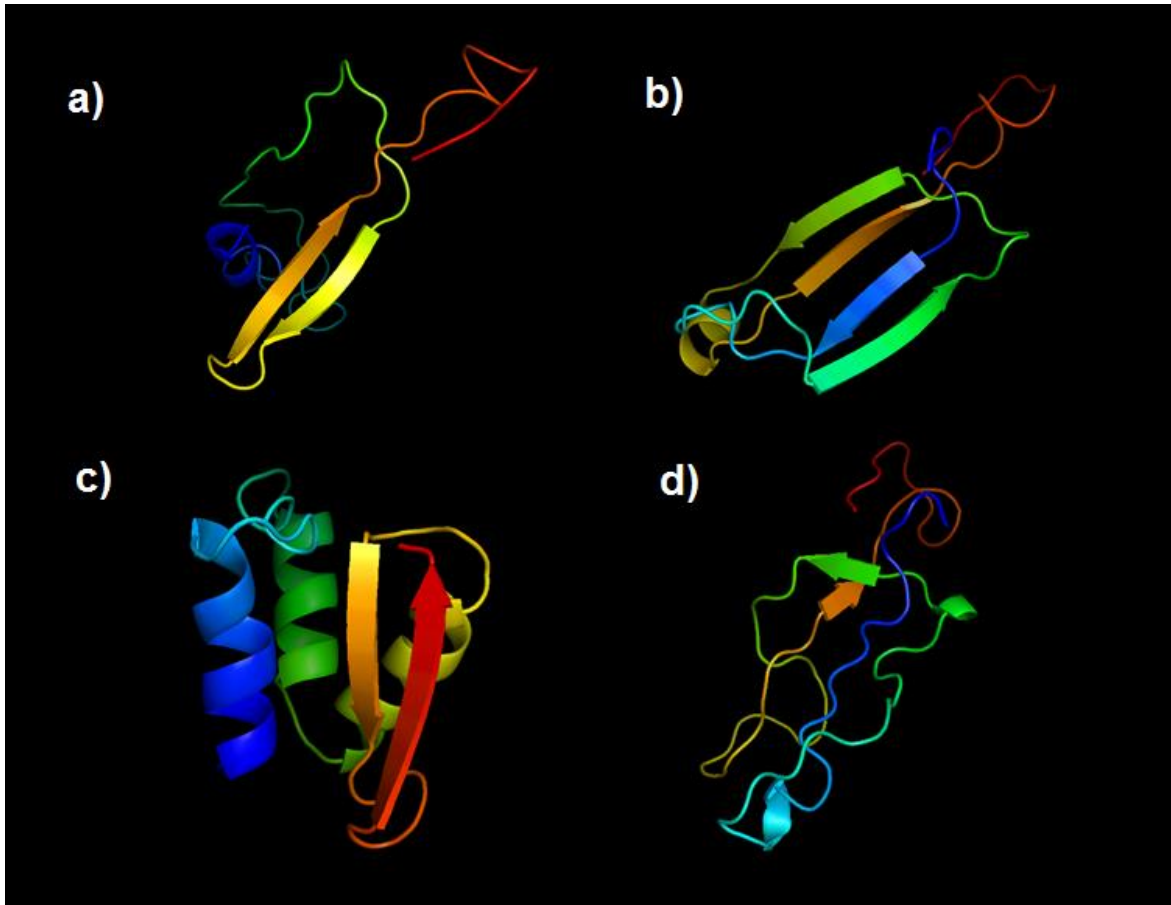


Figura 18. Estructuras tridimensionales de la proteína Vac-Quimera modelada por a) Phyre<sup>2</sup>, b) Rosetta, c) QUARK, d) I-TASSER.

## 6.1.8. ANÁLISIS ESTRUCTURAL

### 6.1.8.1. Estereología

La estructura tridimensional de la secuencia de la proteína Vac-Quimera obtenida por los cuatro diferentes servidores, se analizó con el algoritmo ProTSAV, las proteínas que obtengan en su análisis valores menores a 0.62 o que en la gráfica se encuentren en los colores amarillo y verde, son consideradas proteínas con un modelado de buena calidad. La estructura de la proteína Vac-Quimera modelada

con el servidor QUARK, obtuvo un puntaje de 0.55, considerándola como una proteína con modelado de buena calidad. En la tabla XVI se pueden comparar los puntajes obtenidos en cada uno de los modelados. En las figuras 19, 20, 21 y 22 se observa, de manera gráfica, la media de los valores obtenidos por cada uno de los 10 algoritmos que conforman a ProTSAV.

TABLA XVI. Comparación de la calidad de las proteínas modeladas con los cuatro diferentes servidores

Algoritmo	Phyre <sup>2</sup>	Robetta	QUARK	I-TASSER
<b>Naccess</b>	0.79	0.76	0.38	0.72
<b>Verify3D</b>	1.00	0.81	0.13	1.00
<b>Errat</b>	1.00	0.00	0.41	0.44
<b>Procheck</b>	0.70	0.52	1.00	0.82
<b>MolProbity</b>	1.00	1.00	1.00	1.00
<b>ProSA</b>	0.95	0.98	0.37	1.00
<b>dDIFIRE</b>	0.87	0.82	0.57	0.93
<b>D2N</b>	0.80	0.80	0.43	0.79
<b>ProQ</b>	0.00	0.00	0.61	0.00
<b>PSN-QA</b>	0.70	0.70	0.79	0.70
<b>Promedio</b>	0.79	0.63	0.55	0.74

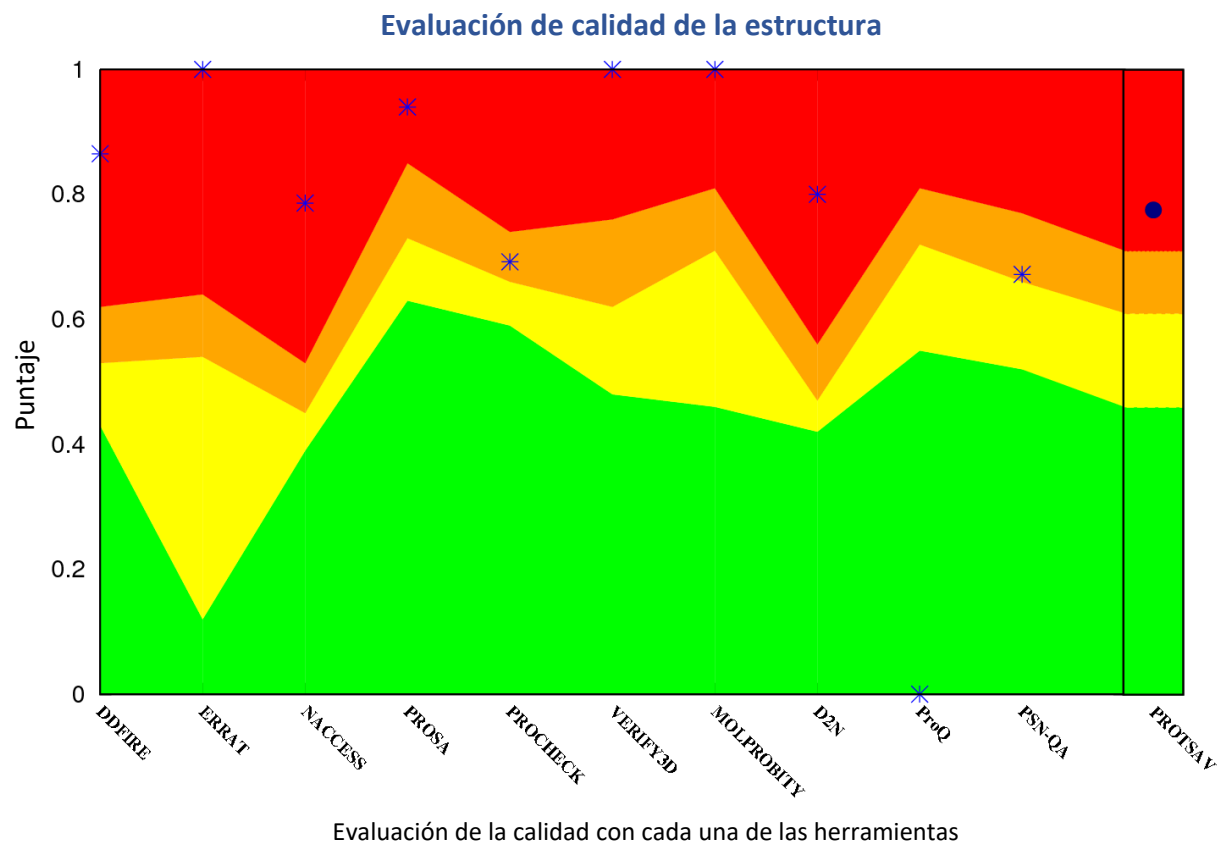


Figura 19. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera modelada con el servidor Phyre<sup>2</sup>

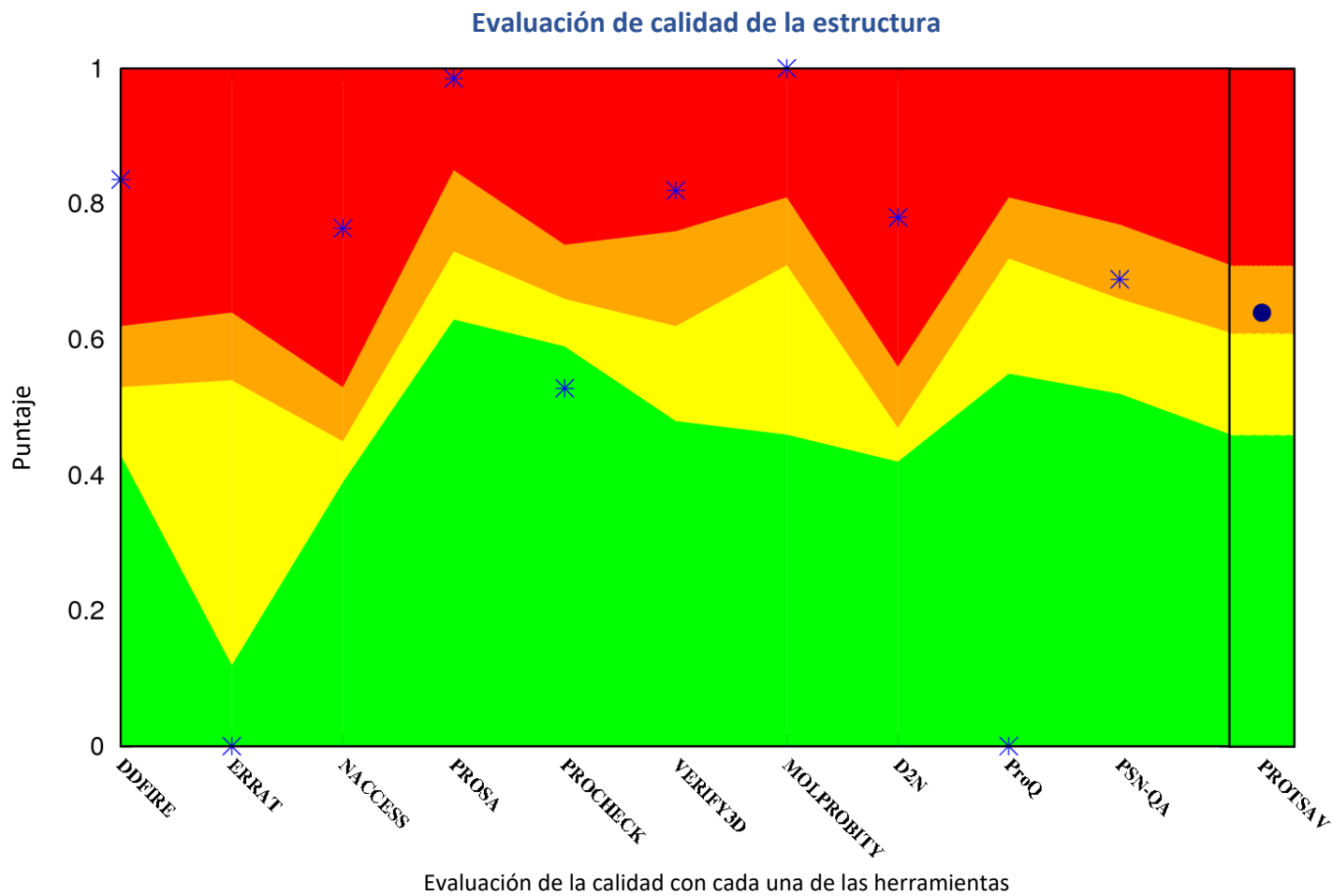


Figura 20. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera modelada con el servidor Robetta.

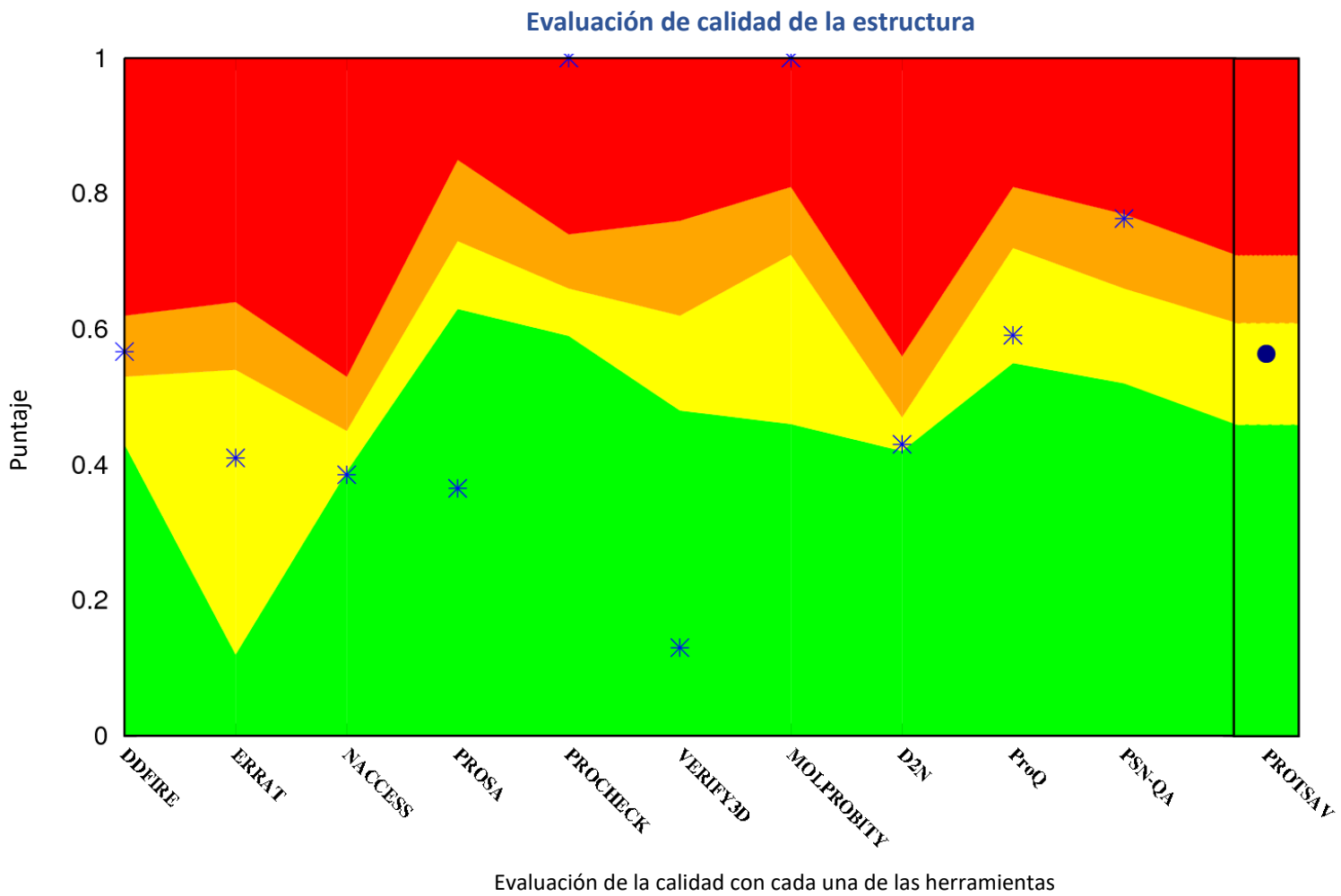


Figura 21. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera modelada con el servidor QUARK.

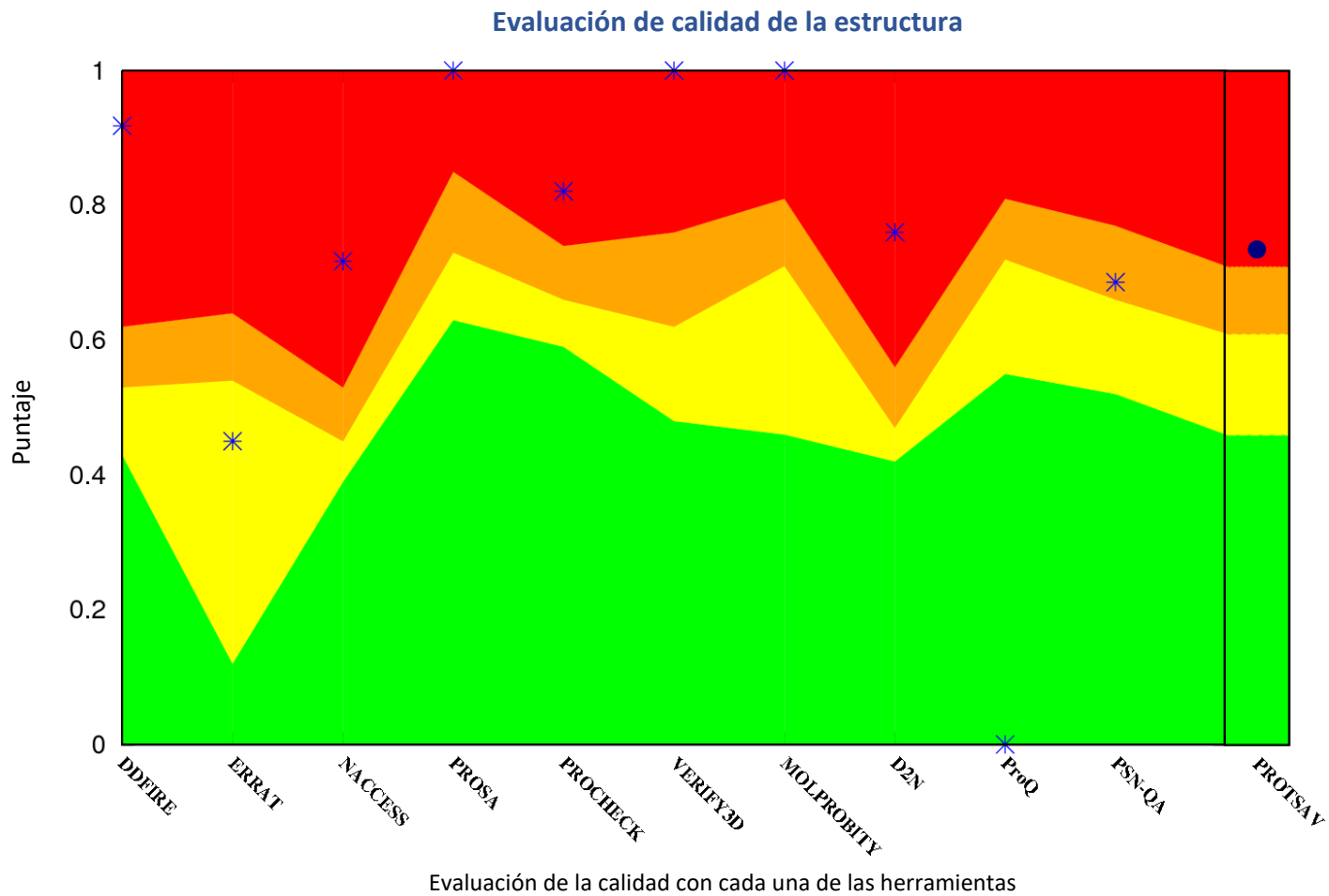


Figura 22. Gráfica del valor obtenido por el servidor ProTSAV para la proteína Vac-Quimera modelada con el servidor I-TASSER.

### 6.1.8.2. Solubilidad

La solubilidad de la estructura tridimensional las proteínas modeladas se determinó utilizando un programa aun no disponible para el público, en el cual se analiza la solubilidad de una proteína basado en su estructura. Estos valores fueron realizados por el Dr. Alejandro Carballo de la universidad de Manchester. También, existe una versión en donde se analiza la solubilidad basandose en la secuencia de aminoácidos ([protein-sol.manchester.ac.uk/](http://protein-sol.manchester.ac.uk/)). El algoritmo solo se enfoca en los parches positivos, los cuales están representados por el color azul, los parches negativos en color rojo y los parches sin carga (hidrofóbicos) en color blanco. De acuerdo a la predicción, para que una proteína se considere positiva, la relación tamaño de parche / distancia de tamaño umbral  $\geq (\geq)$  1.0. El valor obtenido con el algoritmo para el modelado más estable (QUARK) de la proteína Vac-Quimera fue de 0.23 (Ver anexo 8), indicándonos que hipotéticamente la proteína es soluble. En la tabla XVII se observan los valores de solubilidad obtenidos con el algoritmo. Todas las proteínas modeladas son solubles ya que el puntaje de relación es menor a 1. Dicho algoritmo también permite visualizar los parches con carga de la superficie de las proteínas modeladas (Fig. 23).

TABLA XVII. Comparación de los diferentes valores de solubilidad obtenidos

Proteína	Radio = Tamaño del parche / tamaño de distancia
Phyre2	0.25
<b><u>QUARK</u></b>	<b><u>0.23</u></b>
I-TASSER	0.46

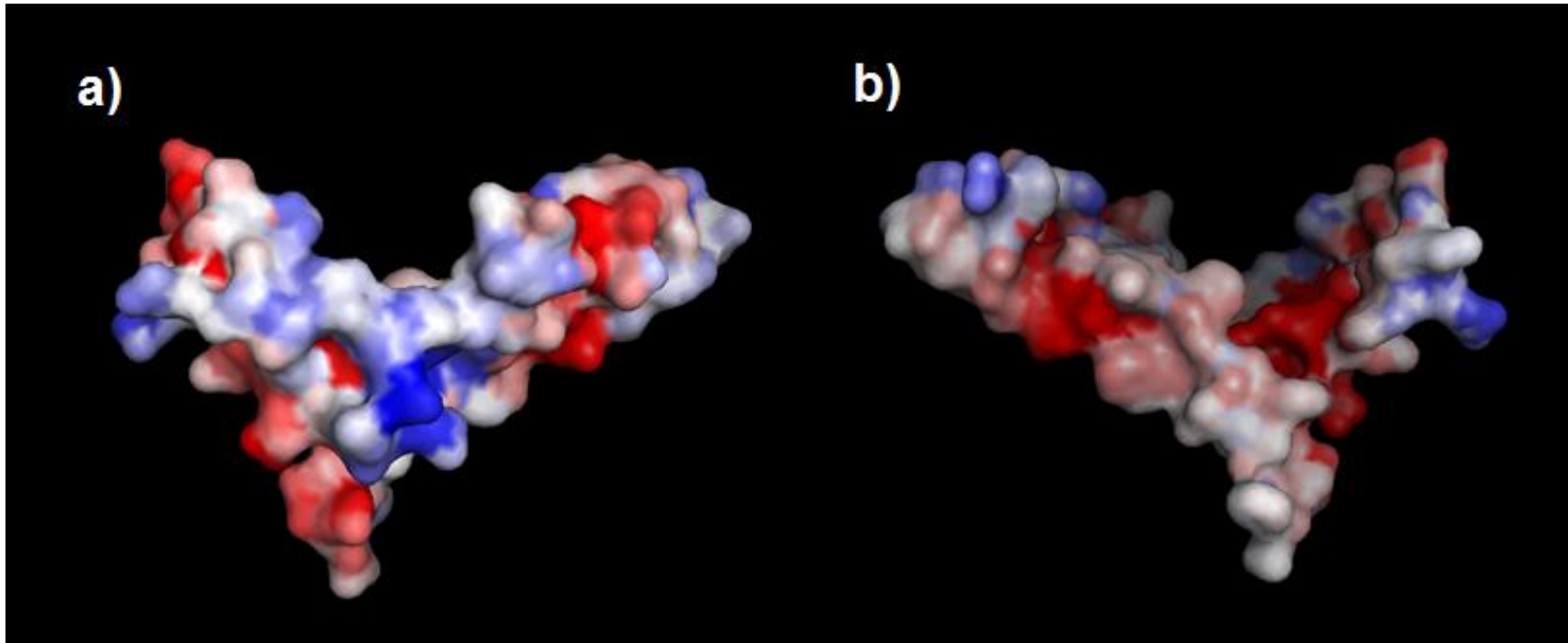


Figura 23. Estructura tridimensional de proteína Vac-Quimera modelada con el servidor Phyre<sup>2</sup>. Se observan en color rojo las regiones negativas, en azul las positivas y en blanco las neutras o hidrofóbicas, a) vista anterior de la proteína, b) vista posterior de la proteína [Visualizada en Pymol].

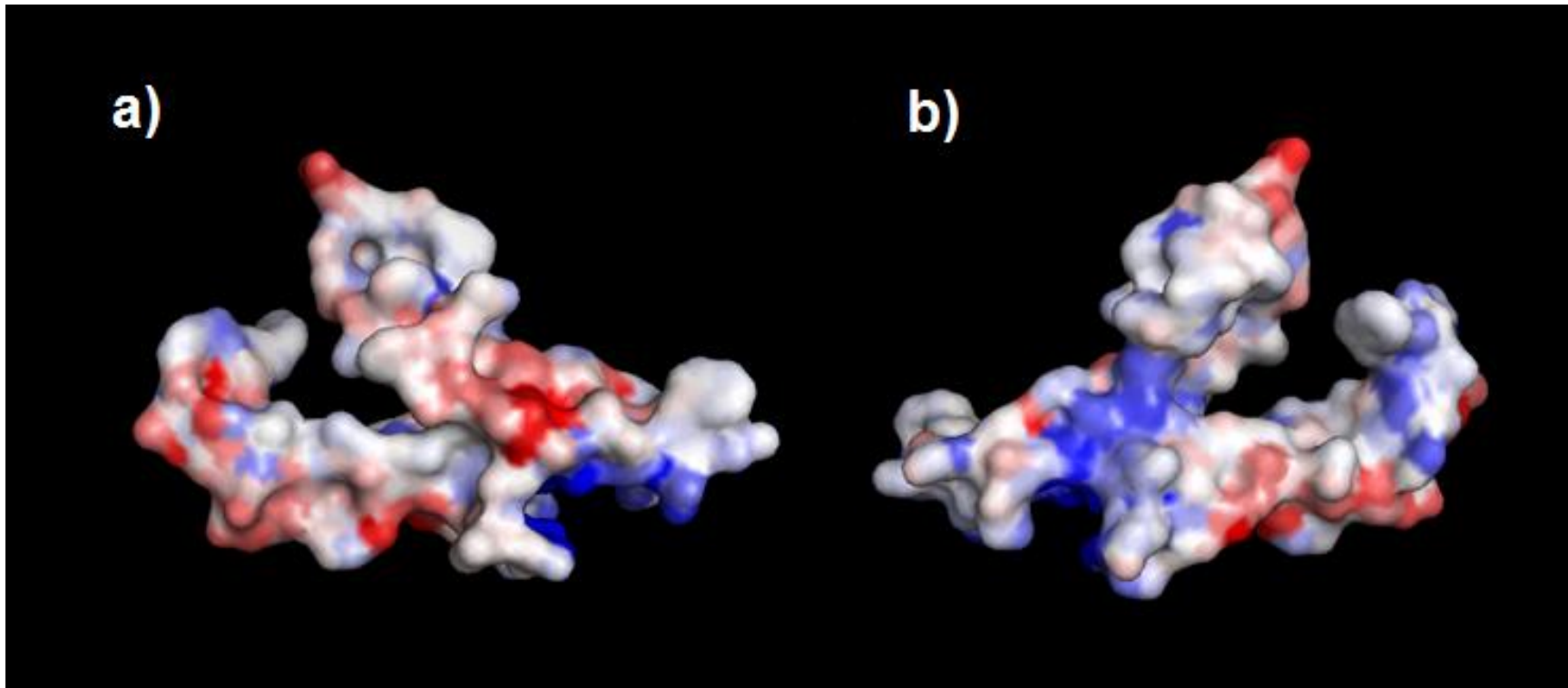


Figura 24. Estructura tridimensional de proteína Vac-Quimera modelada con el servidor Robetta. Se observan en color rojo las regiones negativas, en azul las positivas y en blanco las neutras o hidrofóbicas, a) vista anterior de la proteína, b) vista posterior de la proteína [Visualizada en Pymol].

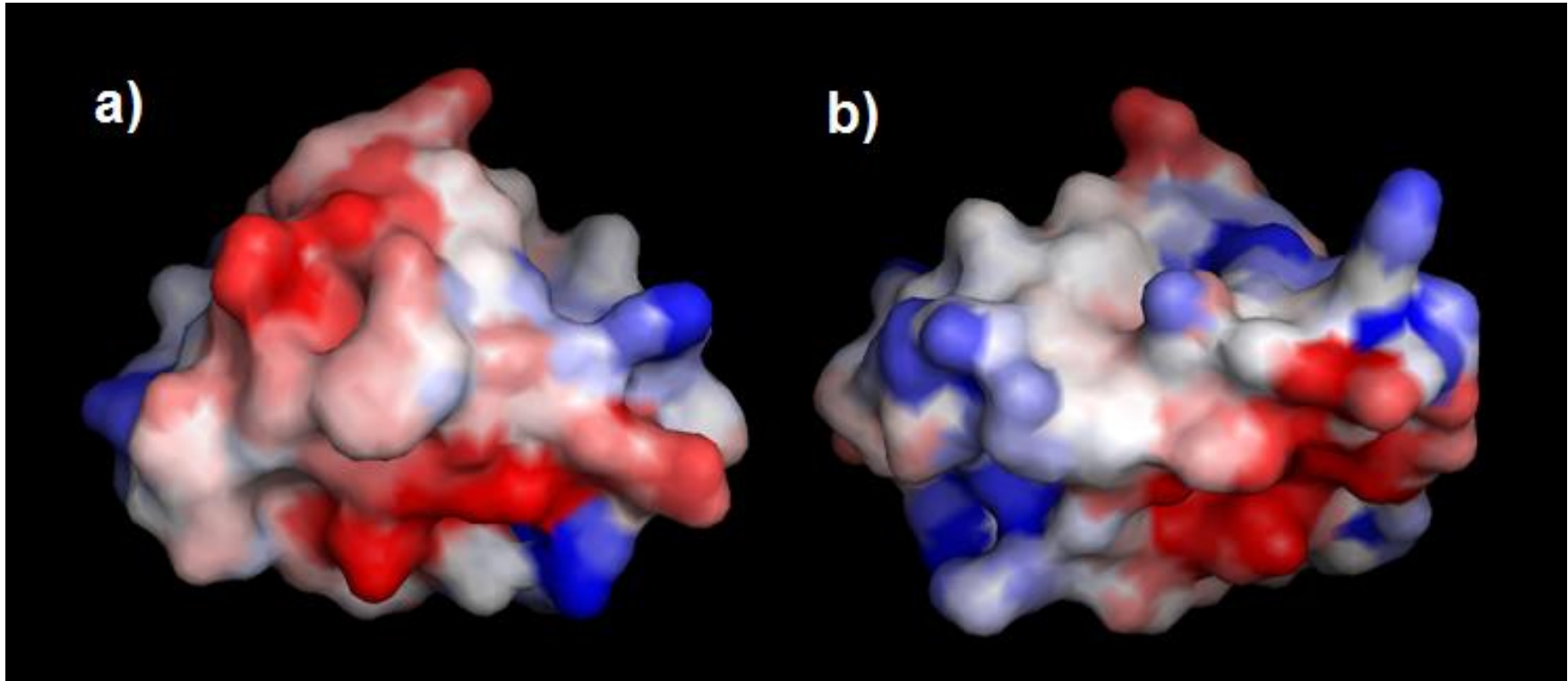
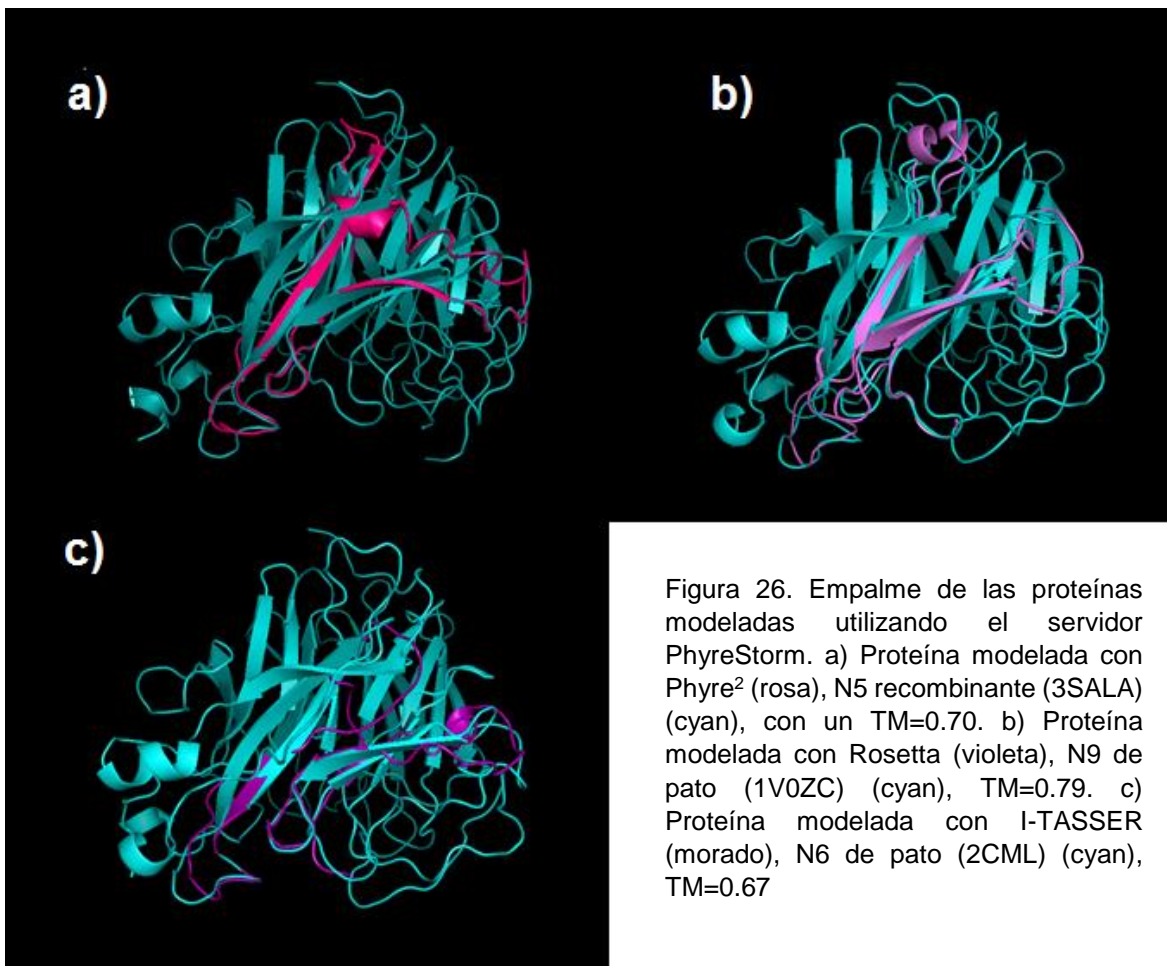


Figura 25. Estructura tridimensional de proteína Vac-Quimera modelada con el servidor QUARK. Se observan en color rojo las regiones negativas, en azul las positivas y en blanco las neutras o hidrofóbicas, a) vista anterior de la proteína, b) vista posterior de la proteína [Visualizada en Pymol].

### 6.1.8.3. Superposición

Se realizó el empalme estructural con el servidor PhyreStorm, el cual muestra sus resultados en TM, que es la similitud estructural global entre la proteína problema y la alineada, donde 1.0 es cien por ciento de empalme. La proteína que mejor empalmó fue la diseñada con el servidor Robetta, ya que obtuvo un TM de 0.79 de alineación con la proteína Neuroaminidasa de Pato NA 1V0ZC (Fig. 26).



## 6.2. CONSTRUCCIÓN DEL CASSETTE DE EXPRESIÓN

Se mandó sintetizar el gen que codifica para la proteína Vac-Quimera a GeneScript clonado en el vector pUC57 con codones optimizados para el cloroplasto de *D.salina*. Esta secuencia cuenta con un sitio para su recuperación de seis histidinas y una zona de corte para enterocinasas. Dicha secuencia se encuentra flanqueada por los sitios de corte para NcoI y PstI (Ver anexo 4).

### 6.2.1. AMPLIFICACIÓN DEL GEN POR PCR

Para su clonación en el vector pChlamy\_4 de *C. reinhardtii*, se amplificó dicho gen para cambiarle el sitio de corte NcoI por EcoRI, ya que el sitio de clonación múltiple del vector no cuenta con esa secuencia. Se diseñaron cebadores para dicho cambio (Fig. 19) y, posteriormente, se corrió una electroforesis en gel de agarosa para verificar cuál de las tres concentraciones de cloruro de magnesio presentaba una mejor amplificación (Fig. 20). Con base en los resultados se determinó que la concentración 1 mM era la más adecuada para las amplificaciones. De la purificación del producto de PCR se obtuvieron 936.6 ng/μL de ADN del gen amplificado.



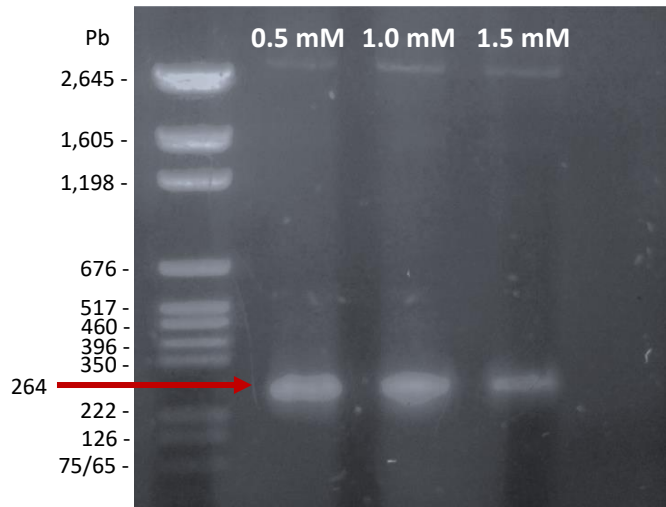


Figura 28. Amplificación del gen Quimera con diferentes concentraciones de cloruro de magnesio. Carril 1: estándar, carril 2: MgCl 0.5mM, carril 3: MgCl 1mM, carril 4: MgCl 1.5 mM.

### 6.2.2. DIGESTIÓN

Se digirió el vector pChlamy\_4 con las enzimas *EcoRI* y *PstI*, al igual que el producto de la amplificación del gen Vac-Quimera, para dejar ambos sitios pegajosos expuestos. Los resultados de las digestiones se muestran en la figura 29.

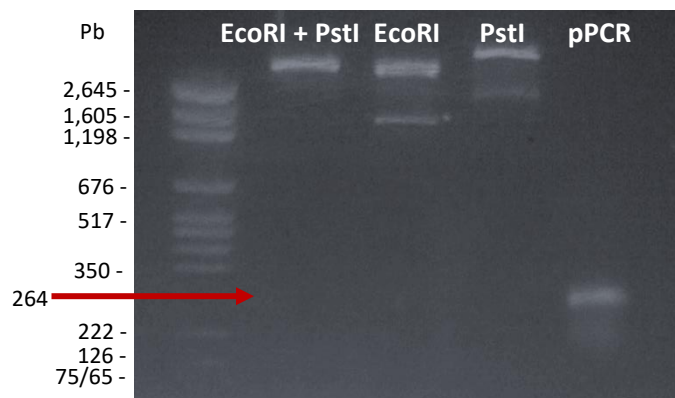


Figura 29. Doble digestión enzimática para el vector pChlamy\_4 con las enzimas *EcoRI* y *PstI*. Carril 1: estándar, carril 2: pChlamy\_4+EcoRIHIRE-Mix+PstIHF, carril 3: pChlamy\_4+EcoRIHFRE-Mix, carril 4: pChlamy+PstIHF, carril 5: 1µL de producto de PCR digerido y purificado directo sin electroforesis previa.



#### 6.2.4. CONFIRMACIÓN

De las células *E.coli* XL1-Blue transformadas con la ligación del gen Vac-Quimera en el vector pChlamy\_4 se seleccionaron nueve colonias: 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B y 3C, donde los números pertenecen a la caja seleccionada y la letra a la colonia seleccionada. Las concentraciones de ADN plasmídico obtenido se muestran en la tabla XVIII.

Tabla XVIII. Concentraciones de ADN plasmídico obtenido por extracción

Clona	DNA ng/ $\mu$ L
<b>1A</b>	144.6
<b>1B</b>	72.4
<b>1C</b>	70.4
<b>2A</b>	226.2
<b>2B</b>	222.6
<b>2C</b>	35.9
<b>3A</b>	68.3
<b>3B</b>	56.6
<b>3C</b>	60.7

De cada uno de los plásmidos obtenidos se realizó una doble digestión enzimática con *EcoRI* y *PstI*; para poder visualizar la banda liberada de 264 bp, se corrieron las digestiones en un gel de agarosa 1%. Los resultados se muestran en la figura 31.

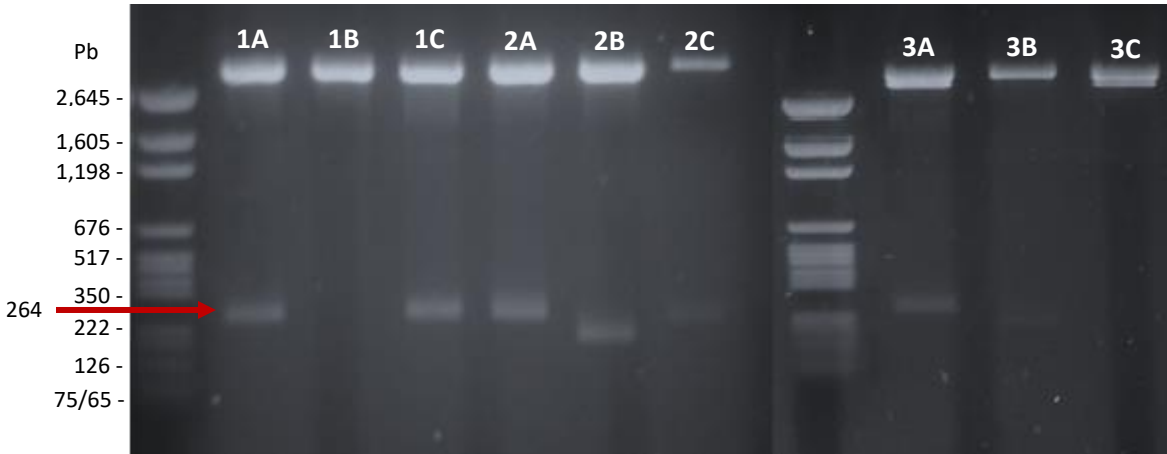


Figura 31. Clonas positivas de *E.coli* XL1-blue con el vector pChlamy\_Quim4 después de que se hizo la digestión enzimática con *EcoRI* y *PstI*.

Para corroborar que el gen Vac-Quimera se encontraba realmente ligado en el vector pChlamy\_4 se realizó una amplificación del gen con los cebadores previamente sintetizados (Tabla VIII). Se utilizó como control positivo el gen Vac-Quimera clonado en pUC57, que fue el sintetizado por la empresa GeneScrip (Fig. 24); se observó amplificación de una banda de aproximadamente 264bp en las clonas 2A y 2B.

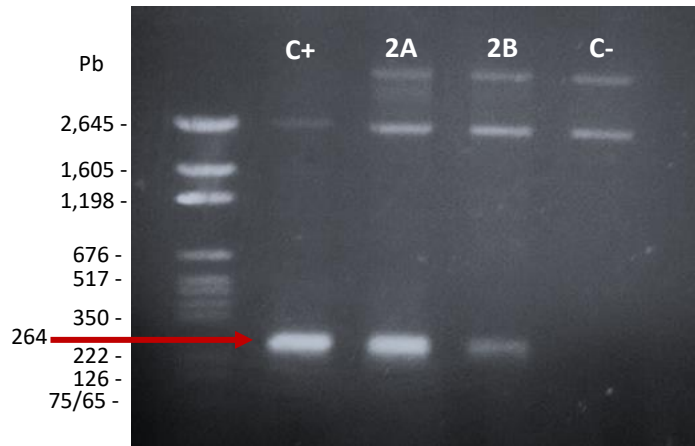


Figura 32. Amplificación de colonias positivas de *E. coli* transformadas con la ligación del gen Vac-Quimera en el vector pChlamy\_4 utilizando la técnica de shock térmico. Carril 1: estándar, Carril 2: pUC57-Quimera (control positivo), Carril 3: pChlamy\_Quim4 2A, Carril 4: pChlamy\_Quim4 2B, Carril 5: pChlamy\_Quim4 2A sin polimerasa, Carril 6: cebadores y polimerasa sin DNA.

### 6.2.5. SECUENCIACIÓN

Se mandó plásmido de la clona 2A a la empresa EtonBioscience para ser secuenciado y verificar que el inserto realmente fuese el del gen Vac-Quimera. En la figura 33 se presenta la secuencia obtenida, en la cual se observa que solo 92 nucleótidos de 264 del gen, ya que estos fueron los únicos que se secuenciaron sin impurezas. De los 92 nucleótidos puros se alinearon 67, teniendo un 72% de similitud entre la zona secuenciada y el gen original.



## 7. DISCUSIÓN

Las vacunas de la influenza convencionales son realizadas con virus atenuados, en los últimos años ha surgido una generación de vacunas realizadas solo con ciertas proteínas recombinantes de la influenza tales como la hemaglutinina y la neuroaminidasa (Kaiser 2006). Una proteína antigénica utilizada como vacuna genera una respuesta inmunológica más eficaz y específica ante la enfermedad ya que tiene más superficie antigénica expuesta que el virus atenuado, sin embargo, los epítopes que son responsables de la respuesta inmune solo comprenden una pequeña porción de la proteína total (Prabhavathy et al. 2011), por lo tanto, las nuevas vacunas basadas en solo péptidos antigénicos han demostrado ser las mejores vacunas para activar la respuesta inmune celular, ya que las células presentadoras de antígeno (APC), al procesar únicamente péptidos aseguran que solo se presentarán las regiones más inmunodominantes en lugar de otros péptidos no antigénicos que también se encuentran en la proteína (Atsmon et al. 2014; Groot et al. 2001; Prabhavathy et al. 2011).

Considerando lo anterior, en este trabajo se seleccionaron los péptidos más conservados e inmunodominantes de las proteínas NA, HA, M2 y NP de la influenza y se acomodaron consecutivamente como una proteína completa. Se analizaron las propiedades fisicoquímicas de la proteína diseñada utilizando la herramienta PROTPARAM mediante la cual se determinó que el punto isoeléctrico de la proteína es de 5.22, lo cual indica que a este pH las cargas de la proteína son neutras (Cox et al. 2009). De manera paralela se estimó que el tiempo de inestabilidad de la

proteína es de 30 horas y, de acuerdo a la literatura, cualquier valor menor a 40 horas demuestra que la proteína es estable y los valores mayores a 40 h presumen de una proteína inestable (Jaimes-Melo 2012). Cabe mencionar que este valor se debe a las propiedades intrínsecas del amino terminal de la proteína, ya que al tener metionina en el N-terminal eleva el tiempo de vida media (Berg et al. 2007). Mediante análisis se determinó que el promedio de hidropaticidad es de  $-0.153$ , promedio GRAVY que representa el cambio de la energía libre de un aminoácido al pasar de un medio no polar a uno polar. De acuerdo a éste promedio sabemos que valores mayores a cero representan aminoácidos hidrofóbicos y, valores menores a cero a aminoácidos hidrofílicos. Por lo tanto, la proteína Vac-Quimera diseñada en el presente trabajo demostró ser hidrofílica, de acuerdo a su estructura primaria (Kyte et al. 1982).

La proteína Vac-Quimera generada presentó un valor alifático de 75.83, el cual se define como el volumen ocupado por los aminoácidos alanina, valina, isoleucina y leucina, en una proteína. Los valores positivos indican termoestabilidad en proteínas globulares, lo cual demuestra que la proteína Vac-quimera es termoestable, de acuerdo a lo reportado por Gasteiger y colaboradores (Gasteiger et al. 2005). El análisis de la secuencia de aminoácidos de la proteína Vac-Quimera para determinar si existía o no homología con las proteínas del virus de la influenza, se hizo utilizando la herramienta BLAST de NCBI. Con ésta herramienta se determinó que la proteína Vac-Quimera tuvo un alineamiento del 86.1% con los aminoácidos de la superfamilia de la neuroaminidasa. De acuerdo a la literatura, las secuencias de aminoácidos que tengan una identidad mayor al 80% con otras secuencias de

la base de datos, son consideradas evolutivamente cercanas (Pearson 2014). Este dato nos permite suponer que a pesar de que la proteína Vac-Quimera está conformada por tan solo 72 aminoácidos, es una proteína reconocida como neuroaminidasa.

Los antecedentes anteriores nos indican que los epítopes que conforman a la proteína Vac-Quimera son capaces de inducir la respuesta inmune celular de los linfocitos T citotóxicos (CD8<sup>+</sup>) y de los T colaboradores (CD4<sup>+</sup>). Los linfocitos T citotóxicos reconocen los péptidos que son presentados por el complejo mayor de histocompatibilidad MHC-I, el cual es expresado en la membrana citoplasmática de todas las células nucleadas de vertebrados. Sin embargo, ésta presentación solo se lleva a cabo por medio de reacción cruzada y, la eficiencia de dicha presentación está determinada por la afinidad del péptido antigénico al sitio de unión del MHC-I (Joffre et al. 2012). Para determinar si los cuatro epítopes conservados que conforman a la proteína Vac-Quimera son reconocidos individualmente por distintos alelos del MHC-I, se utilizó la herramienta en línea ProPred1. Ésta herramienta se basa en un método de arreglos de matrices que permite predecir que péptidos lineales de una proteína se pueden unir a 47 alelos distintos del MHC-I (Singh & Raghava 2003). Al utilizar este método es posible determinar que alelos generarán respuesta inmunológica adaptativa a las cuatro proteínas de la influenza seleccionada (HA, NA, NP y M2) y por ende, que poblaciones se encontrarán mejor protegidas. Los resultados de ProPred1 demostraron que los alelos del grupo HLA-A2 son capaces de procesar y presentar a los cuatro epítopes (ver tabla XIII), sin embargo, hay un traslape de los aminoácidos IEGGWC y FWVEL pertenecientes a

la hemaglutinina y a la neuroaminidasa, respectivamente. Éstos resultados nos indican que *in vivo* podría generarse una identificación débil de dichas proteínas ante una infección por el virus de la influenza. No obstante, los alelos que obtuvieron mejor cobertura son los pertenecientes al grupo A2, el cual comprende todos los serotipos designados con el prefijo \*02. Cabe destacar que éste alelo es el más frecuente en la población mundial (Bardi et al. 2012; Gras et al. 2013), lo que permitiría generar una respuesta de los linfocitos T citotóxicos en la mayoría de los humanos que utilizarán ésta vacuna.

Se utilizó la herramienta MHC-Pred para determinar cuáles serían las combinaciones posibles de péptidos que presentaría el complejo mayor de histocompatibilidad del grupo alélico A2 y se obtuvieron 34 péptidos que hipotéticamente son procesados y presentados por el MHC-I, además que su unión es estable y fuerte ya que presentaron una afinidad  $IC_{50}$  menor a 5000 nM. Los 34 péptidos se analizaron individualmente en la herramienta VaxiJen en la que se determina la antigenicidad de un péptido o proteína de acuerdo a su secuencia (Doytchinova & Flower 2007). En el análisis se obtuvo que 19 de los péptidos serán reconocidos por los linfocitos T como antigénicos, por lo tanto el 55.8% de los péptidos procesados y presentados por el MHC-I serán efectivos para generar una respuesta inmunológica. Una de las limitantes es que 12 de los 19 péptidos antigénicos están constituidos por secuencias de dos proteínas y, en el caso de la hemaglutinina, es la única que siempre se presenta junto con péptidos de las proteínas aledañas, por lo que, no generaría una respuesta específica a una infección *in vivo*. No obstante, todos los fragmentos de proteínas tienen segmentos

de péptidos individuales representativos, dos para la proteína de matriz 2, dos para las nucleoproteínas y, tres para la neuroaminidasa. Los efectos que se obtendrían al utilizar la vacuna hipotética en un organismo vivo sería que la vacuna al no provocar una infección, solo podría ser fagocitada y presentada por el complejo MHC-I mediante mecanismo de reacción cruzada, sin embargo, al no poder procesar y presentar los péptidos pertenecientes a la proteína hemaglutinina, en caso de una infección por el virus, los linfocitos T citotóxicos no estarían previamente entrenados para mandar a apoptosis a las células que presentaran péptidos de dicha proteína. Sin embargo, solo con ser péptidos distintos a los propios, estos linfocitos provocarían la apoptosis de la célula, la única diferencia sería la eficiencia con la que lo realizarían.

Para determinar los péptidos que se presentarían por el MHC-II no fue necesario hacer un análisis de los alelos que mejor reconocen a la proteína, ya que dicho complejo se expresa por los alelos HLA-DRA y HLA-DRB correspondientes a las cadenas alfa y beta de la molécula MHC-II, respectivamente. Sabemos que la cadena beta es altamente polimórfica, sin embargo, el alelo DRB1 es expresado cinco veces más que los serotipos DRB3, DRB4 y DRB5, además que el alelo DRB1 está presente en todos los individuos, por lo que fue el alelo seleccionado para el análisis de los péptidos que se unen al MHC-II (Bardi et al. 2012; Fernandez Vina et al. 2012; Patronov & Doytchinova 2013). La herramienta MHC-Pred obtuvo 63 péptidos que son hipotéticamente procesados y reconocidos, todos se analizaron en la herramienta VaxiJen y se obtuvieron 22 péptidos reconocidos como antigénicos, de los cuales 16 corresponden a las cuatro proteínas de la proteína

Vac-Quimera sin traslaparse. De estos péptidos tres corresponden a la proteína de matriz 2, cinco a la nucleoproteína, tres para la hemaglutinina y cinco para la neuroaminidasa. A diferencia del MHC-I, en el MHC-II si se obtendría una respuesta inmune celular completa para las cuatro proteínas. El efecto que se obtendría al utilizar la vacuna hipotética sería que los linfocitos T colaboradores estarían previamente entrenados para actuar ante las cuatro proteínas de la influenza, ya que las células presentadoras de antígeno fagocitarían la vacuna y presentarían los péptidos antigénicos conservados. Sin embargo, la respuesta real a una infección solo se podría determinar con pruebas *in vivo*.

Además de la respuesta inmune celular mediada por los linfocitos T se presenta la respuesta inmune humoral realizada por los linfocitos B, los cuales producen anticuerpos en la superficie de su membrana. Dichos anticuerpos reconocen dos tipos de epítopes, continuos y discontinuos, los primeros son péptidos lineares y los segundos son péptidos estructurales. Aproximadamente, el 10% de los epítopes discontinuos tienen reacción cruzada con péptidos lineares de la misma proteína antígeno. A pesar de que por sí solos los epítopes lineares no generan una respuesta inmune importante son necesarios ya que indican los determinantes antigénicos de una proteína (Erik et al. 2006).

En el presente proyecto se determinaron los epítopes lineares de la proteína Vac-Quimera utilizando los métodos ElliPro y de antigenicidad de Kolaskar y Tongaonkar. El primero está considerado como el mejor método basado en estructura para la predicción de epítopes de unión a anticuerpo (Ponomarenko et al. 2008), pero en ambos métodos se reconocieron epítopes lineares. Sin embargo,

solo el método de antigenicidad de Kolaskar y Tongaonkar obtuvo determinantes antigénicos para las cuatro proteínas sin traslaparse, lo cual indica que el uso de la vacuna hipotética Vac-Quimera sería capaz de inmunizar individuos contra las cuatro proteínas del virus de la influenza seleccionadas. No obstante, al igual que en el caso de la respuesta inmune celular, es necesario realizar pruebas *in vivo* para determinar su efectividad.

La vacuna hipotética es una proteína nombrada Vac-Quimera la cual como su nombre lo dice es una quimera de cuatro epítopes conservados de las proteínas de la influenza: matriz II, nucleoproteína, neuroaminidasa y hemaglutinina. Por lo tanto, al ser solo epítopes conservados, el tamaño de la proteína es pequeño, conformada de solo 72 aminoácidos. Su estructura tridimensional es desconocida por ser una proteína putativa, dicho lo anterior se requirió determinar la estructura tridimensional de la proteína para corroborar que esta fuese globular e hidrofílica evitando su precipitación en el torrente sanguíneo del paciente. La estructura secundaria de la proteína Vac-Quimera se predijo con el servidor NPSA y se observó que está conformada por una alfa hélice y cuatro regiones de hojas betas. Esta conformación es la que tendría la proteína con base a las propiedades intrínsecas de los aminoácidos. El saber esto fue pieza clave para determinar la estructura tridimensional correcta, ya que la mayoría de los programas que realizan predicción estructural de proteínas se basan en proteínas existentes, las cuales son tomadas como templete o molde para realizar el plegamiento tridimensional de la proteína problema.

En este trabajo se utilizaron cuatro programas distintos Phyre2, Robetta, QUARK y I-TASSER. De estos cuatro solo QUARK tiene un sistema de modelado *de novo*, es decir, no utiliza proteína templete, los otros tres realizaron primeramente un alineamiento de secuencia de la proteína Vac-Quimera con las secuencias de proteínas de bases de datos, siendo la neuroaminidasa la proteína que mejor coincidió en el empalme. Para validar el modelado se analizaron cada una de las proteínas en el algoritmo ProTSAV, el cual determina un modelado de calidad a aquellas estructuras que tengan un puntaje menor de 0.62. La proteína modelada con el servidor QUARK obtuvo un puntaje de 0.55, siendo esta la estructura tridimensional hipotética que se obtendría al ser expresada la proteína Vac-Quimera. Aunado a esto se realizó un análisis de empalme estructural para determinar si las proteínas modeladas tenían similitud con proteínas de las bases de datos, por lo que, se utilizó el servidor PhyreStorm en donde un puntaje de 1.0 significa 100% de similitud con la proteína empalmada. Las tres proteínas que fueron modeladas con base a una proteína templete se alinearon con un promedio de 0.72 a neuroaminidasas, N5 recombinante, N9 de pato y N6 de pato. Analizando este resultado con lo obtenido por el algoritmo ProTSAV es evidente que la metodología de alineamiento con proteína templete no es el ideal para proteínas putativas de pocos aminoácidos. Esto es congruente con los resultados, ya que la estructura tridimensional de las neuroaminidasas templete está formada por diferentes fuerzas y nubes de cargas que son resultado del tamaño de la proteína y del acomodo discontinuo de sus aminoácidos. Es decir, que aminoácidos que están a distancia en la estructura primaria pueden encontrarse en la estructura terciaria y generar una carga parcial que afecta el acomodo de otros parches de aminoácidos,

lo cual, es lo que le confiere su estructura tridimensional única a cada proteína. En cambio, la proteína Vac-Quimera al ser pequeña en comparación con una neuroaminidasa, no se ve afectada por cargas parciales de parches de aminoácidos, por lo que, querer hacer que se parezca en estructura a una neuroaminidasa es agregarle fuerza de torsión a aminoácidos que no la tienen, y por ende es una estructura tridimensional forzada e irreal. En cambio la estructura tridimensional creada *de novo* del servidor QUARK solo se basa en las características intrínsecas de los aminoácidos que la conforman. Se comparó la estructura secundaria de la proteína tridimensional con la obtenida en el servidor NPSA y la similitud fue del 88%. Se realizó el análisis de solubilidad a la proteína validada en colaboración con el Dr. Alejandro Carballo de la Universidad de Manchester. El programa donde se analizó la proteína considera solubles a las proteínas con puntajes menores o iguales a 1.0. La proteína modelada con el servidor QUARK obtuvo un puntaje de 0.23, lo que indica que es soluble, por lo tanto, si se expresara la proteína Vac-Quimera esta no tendría similitud estructural con ninguna proteína de la influenza por lo que la respuesta inmune humoral sería débil ya que solo dependería de los epítopes lineales reconocidos por reacción cruzada. Sin embargo, la estructura tridimensional nos indica que la proteína es soluble y podría ser buena candidata para su uso como vacuna.

Para poder llevar a cabo esta investigación, de lo teórico a lo práctico, se sintetizó la secuencia de aminoácidos que expresan a la proteína Vac-Quimera en el vector pUC-57 en la empresa GenScript, y a esta secuencia de ADN se le denominó gen *Vac-Quimera*. Originalmente, se pensaba expresar la proteína en cloroplastos de

*Dunaliella salina*, por lo que se optimizaron los codones del gen para dicha expresión, sin embargo, el proyecto cambió a expresión en núcleo en *Chlamydomonas reinhardtii* debido a que la transformación de *Dunaliella salina* wild type representaba un reto, por sí solo. Debido a este cambio, se sabía que aunque se realizara la correcta inserción del gen en el genoma de *C. reinhardtii* su expresión protéica sería baja ya que no tendría codones optimizados (Badalà et al. 2008).

Se realizó un cassette de expresión mediante la clonación del gen *Vac-Quimera* en el vector pChlamy\_4. Para llevar a cabo la clonación fue necesario cambiar uno de los sitios de corte que flanqueaban el gen Quimera. En un principio el gen clonado en el vector pUC57 estaba flanqueado por los sitios NcoI y PstI por lo que se cambió NcoI por EcoRI, para que de esta manera pudiera ligarse en el sitio de clonación múltiple del vector pChlamy\_4. Se utilizó la técnica de amplificación en cadena de la polimerasa en donde el cebador forward tenía un sitio de corte **GAATTC** para la enzima de restricción EcoRI. En el primer ciclo de reacción, éste cebador solo se unía al extremo del gen, dejando sin unión a la secuencia del sitio de corte, pero en los ciclos siguientes la secuencia del sitio de corte se sintetiza adjunta al gen. La ligación se realizó por medio de la unión de los sitios pegajosos complementarios del gen y del vector, a los cuales primero se les realizó una doble digestión enzimática para dejar expuestos los nucleótidos. Cabe mencionar que se utilizaron sitios pegajosos diferentes para asegurar la correcta clonación en sentido 5´3´ del gen en el vector y evitar falsos positivos en donde la secuencia estuviese clonada en sentido contrario. No se utilizó ningún indicador selectivo de la clonación para diferenciar las colonias transformadas con el vector, por lo tanto, todas las colonias

se eligieron al azar y, posteriormente, fueron analizadas mediante doble digestión enzimática con las enzimas correspondientes a los sitios de restricción con los que se ligó. De las nueve colonias seleccionadas, cinco liberaron un fragmento del tamaño esperado, pero para confirmar que realmente se hubiese clonado el gen en el vector se hizo una amplificación del plásmido presente en dos de las cinco colonias que presentaron un fragmento del tamaño esperado. La amplificación fue positiva, lo cual corroboró la correcta clonación del gen en el vector. Posteriormente, se mandó a secuenciar el plásmido de una de las dos colonias que amplificaron, llamada cepa 2A, pero el análisis de la secuencia demostró que ésta tenía ciertas mutaciones, ya que presentó bases que no correspondían a lo que se había clonado. Este resultado puede ser consecuencia de que al realizar varias amplificaciones del gen, la misma polimerasa insertó algunas bases que no correspondían a la secuencia original, o bien, que los cebadores utilizados flanqueaban sitios distintos a los esperados pero generaban un fragmento del tamaño que se esperaba.

## 8. CONCLUSIONES

El análisis fisicoquímico de la secuencia de la proteína Vac-Quimera demostró que los aminoácidos que la conforman son estables y le confieren una vida media de 30 h.

Fue posible predecir los componentes de la estructura secundaria de Vac-Quimera, gracias a las propiedades intrínsecas de los aminoácidos de la proteína.

Los linfocitos T citotóxicos solo reconocerán anticipadamente a las proteínas de matriz II, nucleoproteína y neuroaminidasa.

La hemaglutinina es la única proteína de las cuatro que conforman a la proteína Vac-Quimera que no podrá ser reconocida específicamente por los linfocitos T citotóxicos.

Los linfocitos T colaboradores brindarán una protección completa a una infección causada por el virus de la influenza, ya que podrán reconocer anticipadamente a las proteínas de matriz II, nucleoproteína, hemaglutinina y neuroaminidasa.

Los alelos pertenecientes al grupo A2 reconocen mejor los epítopes que conforman la proteína Vac-Quimera que el resto de los serotipos.

La mayoría de las personas que usarán la vacuna podrán tener buena respuesta inmune debido a que los alelos del grupo A2 son los más frecuentes en la población mundial.

Una de las características de la vacuna hipotética Vac-Quimera, de acuerdo a su estructura secundaria indica que es soluble en el torrente sanguíneo.

Se hizo la correcta clonación del gen Vac-Quimera en el vector pChlamy\_4 para su posterior expresión en *Chlamydomonas reinhardtii*.

La expresión de la proteína Vac-Quimera en *Chlamydomonas reinhardtii* será débil ya que la frecuencia de codones del gen pertenece a cloroplasto de *Dunaliella salina* y la transformación sería en núcleo de *Chlamydomonas reinhardtii*.

En la secuenciación del gen Vac-Quimera se mostró mutaciones en algunos nucleótidos por lo que no es apta para expresarse en *Chlamydomonas reinhardtii*.

## 9. REFERENCIAS

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25, pp.3389–3402.

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17), pp.3389–3402.

Apisarnthanarak, A. et al., 2008. Avian influenza research progress. *Nova science publishers*, p.165.

Atsmon, J. et al., 2014. Priming by a novel universal influenza vaccine ( Multimeric-001 )— A gateway for improving immune response in the elderly population. *Vaccine*, 32(44), pp.5816–5823. Available at: <http://dx.doi.org/10.1016/j.vaccine.2014.08.031>.

Badalà, F., Nouri-mahdavi, K. & Raoof, D.A., 2008. A critical analysis of codon optimization in human therapeutics. *Computer*, 144(5), pp.724–732.

Bardi, M.S. et al., 2012. HLA-A, B and DRB1 allele and haplotype frequencies in volunteer bone marrow donors from the north of Parana State. *Revista brasileira de hematologia e hemoterapia*, 34(1), pp.25–30. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3459602&tool=pmc-entrez&rendertype=abstract>.

Berg, J.M., Stryer, L. & Tymoczko, J.L., 2007. *Bioquímica*, Reverte.

Caballero, J., 2005. Las gripes mas mortales de la historia. *Elmundo.es*. Available at: <http://www.elmundo.es/suplementos/magazine/2005/318/1130522632.html> [Accessed November 3, 2015].

Carragher, D.M. et al., 2008. A novel role for non-neutralizing antibodies against nucleoprotein in facilitating resistance to influenza virus. *Journal of immunology (Baltimore, Md. : 1950)*, 181(6), pp.4168–4176.

CDC, Flu symptoms. Available at: [www.cdc.gov/flu/about/disease](http://www.cdc.gov/flu/about/disease).

CDC, 2015. Tipos de virus de la influenza. *Centro para el control y la Prevención de enfermedades*.

Chaipan, C. et al., 2009. Proteolytic Activation of the 1918 Influenza Virus Hemagglutinin. *Journal of Virology*, 83(7), pp.3200–3211.

Chan, P., Curtis, R.A. & Warwicker, J., 2013. with a lack of positively-charged surface. , pp.41–45.

Chun, S. et al., 2008. Universal antibodies and their applications to the quantitative determination of virtually all subtypes of the influenza A viral hemagglutinins. *Vaccine*, 26(48), pp.6068–6076.

Cianci, C. et al., 2013. Review Influenza nucleoprotein : promising target for antiviral chemotherapy. , 91, pp.77–91.

El comercio, 2016. OMS alerta de posible rebrote de influenza en America Latina. Available at: <http://elcomercio.pe/ciencias/medicina/oms-alerta-posible-rebote-influenza-america-latina-noticia-1874355> [Accessed January 27, 2016].

Cox, M., Nelson, D.L. & Lehninger, A.L., 2009. *Lehninger Principles of Biochemistry*,

Díaz de los Ríos, D.A., 2013. *Predicción estructural y funcional de la proteína hipotética pa2481 de Pseudomonas Aeruginosa PAO1*. Pontificia Universidad Javeriana, Bogotá D.C.- Colombia.

Doytchinova, I.A. & Flower, D.R., 2007. VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC bioinformatics*, 8(1), p.4. Available at: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-4>.

Eddy, S.R., 2004. What is a hidden Markov model ? , 22(10), pp.1315–1316.

Ekiert, D.C. et al., 2009. Antibody recognition of a highly conserved influenza virus

- epitope. *Science (New York, N. Y.)*, 324(5924), pp.246–251.
- Erik, J. et al., 2006. Improved method for predicting linear B-cell epitopes. , 7, pp.1–7.
- Excelsior, 2016. Van 149 muertes por influenza. Available at: <http://www.excelsior.com.mx/nacional/2016/03/07/1079320> [Accessed February 3, 2016].
- Fernandez Vina, M.A. et al., 2012. Tracking human migrations by the analysis of the distribution of HLA alleles, lineages and haplotypes in closed and open populations. *Philosophical Transactions of The Royal Society*, 367, pp.820–829.
- Fields, B., 1985. *Virology*,
- FluGov, 2010. historia de la gripe pandemica. Available at: <http://espanol.flu.gov/pandemia/historia/u3s/%C3%ADndice.html> [Accessed November 3, 2015].
- Gabbard, J. et al., 2009. A humanized anti-M2 scFv shows protective in vitro activity against influenza. *Protein Engineering, Design and Selection*, 22(3), pp.189–198.
- Gasteiger, E. et al., 2005. Protein Identification and Analysis Tools on the ExPASy Server. *the proteomics protocols handbook*, pp.571–608.
- Gras, S. et al., 2013. Acute emergence and reversion of influenza A virus quasispecies within CD8. , (May), pp.1–10.
- Greenbaum, J. a et al., 2009. Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population. *Proceedings of the National Academy of Sciences of the United States of America*, 106(48), pp.20365–20370.

- Greenbaum, W.V. & Kga, 2008. Supporting Information. *Science And Technology*, 3(c), pp.1–9. Available at: <http://www.springerlink.com/index/f25411752643g829.pdf>.
- Groot, A.S. De et al., 2001. From genome to vaccine : in silico predictions , ex vivo verification. , 19, pp.4385–4395.
- De Groot, A.S. et al., 2009. Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008-2009 conventional influenza vaccine. *Vaccine*, 27(42), pp.5740–5747.
- Guan, P. et al., 2003. MHCpred: bringing a quantitative dimension to the online prediction of MHC binding. , p.4.
- Hattotuwigama, C.K. et al., 2004. Quantitative online prediction of peptide binding to the major histocompatibility complex. *Journal of Molecular Graphics and Modelling*, 22(3), pp.195–207.
- Homan, E.J. & Bremel, R.D., 2011. Patterns of predicted T-cell epitopes associated with antigenic drift in influenza H3N2 hemagglutinin M. A. Barry, ed. *PLoS ONE*, 6(10), p.e26711.
- Hu, W. et al., 2013. Fully human broadly neutralizing monoclonal antibodies against influenza A viruses generated from the memory B cells of a 2009 pandemic H1N1 influenza vaccine recipient. *Virology*, 435(2), pp.320–328.
- Igarashi, M. et al., 2010. Predicting the antigenic structure of the pandemic (H1N1) 2009 influenza virus hemagglutinin. *PLoS ONE*, 5(1), pp.1–7.
- Jaimes-Melo, M., 2012. *Predicción IN-SILICO de la estructura y función de la proteína hipotética p284 de Trypanosoma cruzi*.
- Joffre, O.P. et al., 2012. Cross-presentation by dendritic cells. *Nature reviews. Immunology*, 12(8), pp.557–569. Available at: <http://www.nature.com/doi/10.1038/nri3254>.

- Joshi, S.R., Shaw, A.C. & Quagliarello, V.J., 2009. Pandemic influenza H1N1 2009, innate immunity, and the impact of immunosenescence on influenza vaccine. *The Yale journal of biology and medicine*, 82(4), pp.143–151.
- Kaiser, J., 2006. A One-Size-Fits-All Flu Vaccine ? *Science (New York, N.Y.)*, 312, pp.380–382. Available at: [www.sciencemag.org](http://www.sciencemag.org).
- Ketha, K.M. V & Atreya, C.D., 2008. Application of bioinformatics-coupled experimental analysis reveals a new transport-competent nuclear localization signal in the nucleoprotein of influenza A virus strain. *BMC cell biology*, 9, p.22.
- Kilbourne, E.D., 2006. Influenza Pandemics of the 20th Century. , 12(1), pp.9–14.
- Kolaskar, A.S. & Tongaonkar, P.C., 1990. A semi-empirical method for prediction of antigenic determinant proteins. , 276(1), pp.172–174.
- Krammer, F., 2015. In Translation The Quest for a Universal Flu Vaccine : Headless HA 2 . 0. *Cell Host and Microbe*, 18(4), pp.395–397. Available at: <http://dx.doi.org/10.1016/j.chom.2015.10.003>.
- Kyte, J. et al., 1982. A Simple Method for Displaying the Hydrophobic Character of a Protein. , pp.105–132.
- Li, C. et al., 2010. A simple slot blot for the detection of virtually all subtypes of the influenza A viral hemagglutinins using universal antibodies targeting the fusion peptide. *Nature protocols*, 5(1), pp.14–19. Available at: <http://dx.doi.org/10.1038/nprot.2009.200>.
- Liu, X. & Zhao, Y.P., 2010. Switch region for pathogenic structural change in conformational disease and its prediction. *PLoS ONE*, 5(1).
- Lundegaard, C. et al., 2008. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic acids research*, 36(Web Server issue), pp.509–512.

Montiel, J.L. & G., B.-R., 2010. Mecanismos de infección del virus de influenza. *Narraciones de la ciencia*, pp.51–55.

OMS, 2014. Virus de la influenza, infección en humanos. Available at: [http://www.who.int/influenza/human\\_animal\\_interface/virology\\_laboratories\\_and\\_vaccines/influenza\\_virus\\_infections\\_humans\\_feb14.pdf?ua=1](http://www.who.int/influenza/human_animal_interface/virology_laboratories_and_vaccines/influenza_virus_infections_humans_feb14.pdf?ua=1) [Accessed August 1, 2015].

Patronov, A. & Doytchinova, I., 2013. T-cell epitope vaccine design by immunoinformatics. *Open biology*, 3(1), p.120139. Available at: <http://rsob.royalsocietypublishing.org/content/3/1/120139>.

Pearson, W.R., 2014. An Introduction to Sequence and Series. *International Journal of Research*, 1(10), pp.1286 – 1292. Available at: <http://edupediapublications.org/journals/index.php/ijr/article/view/964>.

Perez, G., Recombinacion de virus H1N1. Available at: [http://www.virusah1n1.com/recombinacin\\_de\\_virus\\_h1n1](http://www.virusah1n1.com/recombinacin_de_virus_h1n1) [Accessed February 3, 2016].

Pichyangkul, S. et al., 2009. Cross-reactive Antibodies against Avian Influenza Virus A ( H5N1 ). , 15(9), pp.1537–1539.

Pielak, R.M. & Chou, J.J., 2011. Influenza M2 proton channels. *Biochimica et Biophysica Acta - Biomembranes*, 1808(2), pp.522–529.

Ping, J. et al., 2015. Development of high-yield influenza A virus vaccine viruses. *Nature Communications*, pp.1–15. Available at: <http://dx.doi.org/10.1038/ncomms9148>.

Ponomarenko, J. et al., 2008. ElliPro : a new structure-based tool for the prediction of antibody epitopes. , 8, pp.1–8.

Prabhavathy, K., Perumal, P. & Sundarabaalaji, N., 2011. In silico identification of B- and T-cell epitopes on OMPLA and LsrC from *Salmonella typhi* for peptide-

based subunit vaccine design. , 10(October), pp.440–451.

Ren, Z. et al., 2015. H5N1 influenza virus-like particle vaccine protects mice from heterologous virus challenge better than whole inactivated virus. *Virus Research*, 200, pp.9–18. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0168170215000131>.

Roti, M. et al., 2013. H5N1 Influenza Virus 1.

Samji, T., 2009. Influenza A : Understanding the Viral Life Cycle. , 82, pp.153–159.

Sanjuan, M. et al., 2009. Endocytosis of influenza viruses. *National institute of health*, 9(10), pp.699–711.

Dos Santos, R. & Soares de Souza, A., 2014. *Análisis de secuencias de aminoácidos*,

Singh, A. et al., 2016. Biochimica et Biophysica Acta ProTSAV : A protein tertiary structure analysis and validation server. *BBA - Proteins and Proteomics*, 1864(1), pp.11–19. Available at: <http://dx.doi.org/10.1016/j.bbapap.2015.10.004>.

Singh, H. & Raghava, G.P.S., 2002. ProPred: Prediction of HLA-DR binding sites. *Bioinformatics*, 17(12), pp.1236–1237.

Singh, H. & Raghava, G.P.S., 2003. ProPred1: Prediction of promiscuous MHC class-I binding sites. *Bioinformatics*, 19(8), pp.1009–1014.

Siripornadulsil, S., Dabrowski, K. & Sayre, R., 2007. Microalgal Vaccines. In *Transgenic Microalgae as Green Cell Factories*. New York, NY: Springer New York, pp. 122–128.

Slonczewski, J.L. & Foster, 2008. *Microbiology: An Evolving Science* First edit., W. W.Norton & Company. Available at: [www.sumanasinc.com/webcontent/animations](http://www.sumanasinc.com/webcontent/animations).

- Slonczewski, J.L. & Foster, J.W., 1923. *Microbiology An Evolving Science* First edit., W. W.Norton & Company.
- Sui, J. et al., 2009. Structural and Functional bases for broad spectrum neutralization of Avian and human influenza A viruses. *Nat Struct Mol Biol*, 16(3), pp.265–273.
- Sun, X. et al., 2014. Structure of influenza virus N7: the last piece of the neuraminidase “jigsaw” puzzle. *Journal of virology*, 88(16), pp.9197–9207.
- Tadanobu, T. et al., 2012. Amino acid determinants conferring stable sialidase activity at low pH for H5N1 influenza A virus neuraminidase. , pp.261–266.
- Tompkins, S.M. et al., 2007. Matrix protein 2 vaccination and protection against influenza viruses, including subtype H5N1. *Emerging Infectious Diseases*, 13(3), pp.426–435.
- Tong, S. et al., 2013. New World Bats Harbor Diverse Influenza A Viruses. *PLoS Pathog.*
- Varich, N.L. et al., 2009. Location of antigenic sites recognized by monoclonal antibodies in the influenza A virus nucleoprotein molecule. *The Journal of general virology*, 90(Pt 7), pp.1730–1733.
- Verma, S.K., Yadav, S. & Kumar, A., 2015. In silico prediction of B - and T - cell epitope on Lassa virus proteins for peptide based subunit vaccine design.
- Wang, C.-C. et al., 2009. Glycans on influenza hemagglutinin affect receptor binding and immune response. *Proceedings of the National Academy of Sciences of the United States of America*, 106(43), pp.18137–18142.
- Wang, P. et al., 2010. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11, p.568.
- WHO, 2016. FluNet data entry. Available at: <http://apps.who.int/flumart> [Accessed March 9, 2016].

- Xing, Z. & Cardona, C.J., 2009. Preexisting immunity to pandemic (H1N1) 2009. *Emerging Infectious Diseases*, 15(11), pp.1847–1849.
- Xu, J. et al., 2011. Evolutionary dynamics of influenza A nucleoprotein (NP) lineages revealed by large-scale sequence analyses. *Infection, Genetics and Evolution*, 11(8), pp.2125–2132. Available at: <http://dx.doi.org/10.1016/j.meegid.2011.07.002>.
- Ye, Q., Krug, R.M. & Tao, Y.J., 2006. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. , 444(December).
- Zhirnov, O.P., Ikizler, M.R. & Wright, P.F., 2002. Cleavage of Influenza A Virus Hemagglutinin in Human Respiratory Epithelium Is Cell Associated and Sensitive to Exogenous Antiproteases. , 76(17), pp.8682–8689.

# 10. ANEXOS

## Anexo 1

Description	Max score	Total score	Query cover	E value	Ident	Accession
<a href="#">neuraminidase [Influenza A virus (A/Cygnus olor/Germany/R1372/2007(H5N1))]</a>	71.0	71.0	100%	2e-13	100%	<a href="#">CAQ63504.1</a>
<a href="#">neuraminidase [Influenza A virus (A/duck/Met Nam/Ncvd1/2003(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABP52006.1</a>
<a href="#">neuraminidase [Influenza A virus (A/chicken/Met Nam/Ncvd8/2003(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABP52009.1</a>
<a href="#">neuraminidase [Influenza A virus (A/goose/Met Nam/113/2001(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABP52015.1</a>
<a href="#">neuraminidase [Influenza A virus (A/teal/Hong Kong/2978/03(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABO21698.1</a>
<a href="#">neuraminidase [Influenza A virus (A/swine/Manitoba/D0274/2013(H1N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ALZ43882.1</a>
<a href="#">neuraminidase [Influenza A virus (A/duck/Guangxi/1378/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66658.1</a>
<a href="#">neuraminidase [Influenza A virus (A/goose/Guangxi/914/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66654.1</a>
<a href="#">neuraminidase [Influenza A virus (A/Chinese pond heron/Hong Kong/18/2005(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66699.1</a>
<a href="#">neuraminidase [Influenza A virus (A/goose/Guangxi/1198/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66656.1</a>
<a href="#">neuraminidase [Influenza A virus (A/grey heron/Hong Kong/728/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66697.1</a>
<a href="#">neuraminidase [Influenza A virus (A/duck/Guangxi/1586/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66659.1</a>
<a href="#">neuraminidase [Influenza A virus (A/goose/Guangxi/2112/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66663.1</a>
<a href="#">neuraminidase [Influenza A virus (A/duck/Guangxi/1681/2004(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">ABC66660.1</a>
<a href="#">neuraminidase [Influenza A virus (A/chicken/Hong Kong/YU562/2001(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAO46823.1</a>
<a href="#">neuraminidase 1 [Influenza A virus (A/environment/Hong Kong/437-8/99 (H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAG01216.1</a>
<a href="#">neuraminidase 1 [Influenza A virus (A/environment/Hong Kong/437-4/99 (H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAG01196.1</a>
<a href="#">neuraminidase [Influenza A virus (A/environment/Hong Kong/437-6/99 (H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAG01206.1</a>
RecName: Full=Neuraminidase	71.0	71.0	100%	3e-13	100%	<a href="#">Q809U7.2</a>
RecName: Full=Neuraminidase	71.0	71.0	100%	3e-13	100%	<a href="#">Q809V2.1</a>
<a href="#">neuraminidase [Influenza A virus (A/goose/Guangdong/3/1997(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAK38299.1</a>
<a href="#">neuraminidase [Influenza A virus (A/Goose/Hong Kong/76.1/01 (H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AOS2956.1</a>
<a href="#">neuraminidase [Influenza A virus (A/Pheasant/HongKong/FY155/01(H5N1))]</a>	71.0	71.0	100%	3e-13	100%	<a href="#">AAO46820.1</a>

## Alineamiento de epitope de NA: CFWVELIRGRPKESTIWTSG

Description	Max score	Total score	Query cover	E value	Ident	Accession
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Tunisia/145/2012(H9N2))]</a>	81.7	81.7	100%	7e-18	100%	<a href="#">AJD77940.1</a>
<a href="#">membrane ion channel M2 [Influenza A virus (A/Hong Kong/485/1997(H5N1))]</a>	81.7	81.7	100%	8e-17	100%	<a href="#">AAK49254.1</a>
<a href="#">matrix protein M2 [Influenza A virus (A/quail/Dubai/301/2000(H9N2))]</a>	81.7	81.7	100%	8e-17	100%	<a href="#">ABM21861.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/partridge/Shantou/24/2000(H9N2))]</a>	81.7	81.7	100%	8e-17	100%	<a href="#">ABV46320.1</a>
<a href="#">matrix protein M2 [Influenza A virus (A/quail/Hong Kong/G1/97 (H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AAD49078.1</a>
<a href="#">membrane ion channel M2 [Influenza A virus (A/Hong Kong/491/97(H5N1))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AAK49260.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/Hong Kong/482/97(H5N1))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AAF74333.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Tunisia/12/2010(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ADX99479.2</a>
RecName: Full=Matrix protein 2; AltName: Full=Proton channel protein M2	81.7	81.7	100%	9e-17	100%	<a href="#">Q70632.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Israel/1638/2006(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACJ68620.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Israel/32/2009(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AEZ67275.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/turkey/Israel/311/2009(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AEZ67264.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Israel/1184/2010(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AEZ67231.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Israel/1293/2010(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">AEZ67220.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/AS120/2007(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ADQ74932.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/AS130/1999(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ADQ74930.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH84/2005(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56272.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH87/2008(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56270.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH186/2007(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56268.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH82/2003(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56264.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH80/2001(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56262.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH78/1999(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56260.1</a>
<a href="#">matrix protein 2 [Influenza A virus (A/chicken/Iran/TH77/1998(H9N2))]</a>	81.7	81.7	100%	9e-17	100%	<a href="#">ACR56258.1</a>

## Alineamientos de epítoto de M2: MSLLETVETLTRNGWGCRCSDDSSD

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Nucleoprotein M9L Influenza A virus (A/Puerto Rico/8/1934(H1N1))	45.6	45.6	100%	6e-05	100%	<a href="#">CDH98260.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Ningjin/03/2014(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ANK78256.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Henan/PL01/2012(H1N2))	45.6	45.6	100%	9e-05	100%	<a href="#">ALT19014.1</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/swine/Korea/JNS06/2004(H3N2))	45.6	45.6	100%	9e-05	100%	<a href="#">ABY81789.2</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/swine/Yokohama/138/2011(H3N2))	45.6	45.6	100%	9e-05	100%	<a href="#">BAM78401.1</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/swine/Mie/R02/2012(H1N2))	45.6	45.6	100%	9e-05	100%	<a href="#">BAM78385.1</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/swine/Yamaqata/19/2010(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">BAM78369.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Binh Duong/02_16/2010(H1N2))	45.6	45.6	100%	9e-05	100%	<a href="#">BAM65747.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Norway/02_11342/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">AEX30601.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Binh Duong/03_13/2010(H3N2))	45.6	45.6	100%	9e-05	100%	<a href="#">BAK41897.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Binh Duong/03_06/2010(H3N2))	45.6	45.6	100%	9e-05	100%	<a href="#">BAK41893.1</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/turkey/Ontario/FAV114-17/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">AD52818.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-22/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACU51033.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-3/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACU51031.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-2/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACU51030.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-1/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACU51029.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-8/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACRO1026.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Alberta/OTH-33-14/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ACU51032.1</a>
<input type="checkbox"/> nucleoprotein Influenza A virus (A/Ontario/1252/2007(H3N2))	45.6	45.6	100%	9e-05	100%	<a href="#">ABY70960.1</a>
<input type="checkbox"/> Chain A_Crystal Structure Of Wsn/a Influenza Nucleoprotein With Three Mutations (e53d_Y289h_Y313v)	45.6	45.6	100%	9e-05	100%	<a href="#">ADYT_A</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/North Carolina/02023/2008(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ADM18080.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/swine/Missouri/02060/2008(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">ADM18081.1</a>
<input type="checkbox"/> nucleocapsid protein Influenza A virus (A/California/VRDL_135/2009(H1N1))	45.6	45.6	100%	9e-05	100%	<a href="#">AD77353.1</a>
<input type="checkbox"/> Chain A_Influenza A Virus Nucleoprotein No At 3.2a Resolution	45.6	45.6	100%	9e-05	100%	<a href="#">2IQH_A</a>

## Alineamiento de epítipo de NP: KLSTRGVQIASNEN

Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/turkey/Hartzlehn/R26/99(H6N2))	47.3	47.3	100%	4e-06	100%	<a href="#">CAP58166.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/chicken/Italy/22/98(H5N9))	47.3	47.3	100%	4e-06	100%	<a href="#">CAP58165.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Anas domestica/Germany/R2583/2007)	47.3	47.3	100%	5e-06	100%	<a href="#">CAP59544.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/quill/Stralsund/Wv1136-40/03(H13N6))	47.3	47.3	100%	5e-06	100%	<a href="#">CAP58176.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Anas platyrhynchos/Germany/R2709/2007(H2N9))	47.3	47.3	100%	5e-06	100%	<a href="#">CAP59547.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Anas platyrhynchos/Germany/R2219/2006(H11))	47.3	47.3	100%	5e-06	100%	<a href="#">CAP58174.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/anus ridibundus/Germany/R2064/2006(H13))	47.3	47.3	100%	6e-06	100%	<a href="#">CAP58177.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/black-headed quill/Sweden/5/99(H16N3))	47.3	47.3	100%	6e-06	100%	<a href="#">CAP58180.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Anas platyrhynchos/Germany/R1649/2007(H1N1))	47.3	47.3	100%	6e-06	100%	<a href="#">CAP59539.1</a>
<input type="checkbox"/> Chain A_The Hemaagglutinin Fusion Peptide (H1 Subtype) At Ph 7.4	47.3	47.3	100%	7e-06	100%	<a href="#">ZKXA_A</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Anas platyrhynchos/Germany/R2706/2007(H2N9))	47.3	47.3	100%	7e-06	100%	<a href="#">CAP59546.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/chicken/Emirates/R66/2002(H9N2))	47.3	47.3	100%	9e-06	100%	<a href="#">CAP58172.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Hong Kong/156/97(H5N1))	47.3	47.3	100%	1e-05	100%	<a href="#">CAP58163.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/chicken/Siemani/1/2015(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ALV82016.1</a>
<input type="checkbox"/> HA-M2e [synthetic construct]	47.3	47.3	100%	2e-05	100%	<a href="#">AJC97681.1</a>
<input type="checkbox"/> HA-ferritin nanoparticle [synthetic construct]	47.3	47.3	100%	2e-05	100%	<a href="#">AGR55396.1</a>
<input type="checkbox"/> alpha-mating factor secretion signal peptide/hemaagglutinin fusion protein [Expression vector pPICK9KH1N1HA]	47.3	47.3	100%	2e-05	100%	<a href="#">ADO95142.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/green peafowl/Thailand/VSMMU-3-CBI/2005(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ABQ11261.2</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/muscovy duck/Vietnam/NCVD02/2005(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ABQ09850.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Chicken/Tapu/BBPV1-576/2005(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ABU99019.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/swine/Nebraska/1/1992(H1N1))	47.3	47.3	100%	2e-05	100%	<a href="#">AAB29091.2</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/duck/Vietnam/367/2005(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ABQ09853.1</a>
<input type="checkbox"/> hemaagglutinin subtype H5 Influenza A virus (A/Hong Kong/481/97(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">AAC32098.1</a>
<input type="checkbox"/> hemaagglutinin Influenza A virus (A/Thailand/01/1997(H5N1))	47.3	47.3	100%	2e-05	100%	<a href="#">ABY25551.1</a>

## Alineamiento de epítipo de HA: GLFGAIAGFIEGGW

## Anexo 2

Numero de aminoácidos	72
Peso molecular	7883.8
pI teórico	5.22
Número total de residuos cargados negativamente (Asp + Glu)	8
Número total de residuos cargados positivamente (Arg + Lys)	7
Coeficiente de extinción molar	22125 M <sup>-1</sup> cm <sup>-1</sup>
Vida media en reticulocitos mamíferos in vitro	30hrs
Vida media en E. coli in vivo	>10 hrs
Índice de inestabilidad II	27.52 Estable
Índice alifático	75.83
Gran promedio de hidropaticidad (GRAVY)	-0.153
Formula	C <sub>345</sub> H <sub>537</sub> N <sub>97</sub> O <sub>107</sub> S <sub>4</sub>
Número total de átomos	1090

Ala (A)	3	4.2%
Arg (R)	5	6.9%
Asn (N)	3	4.2%
Asp (D)	2	2.8%
Cys (C)	3	4.2%
Gln (Q)	1	1.4%
Glu (E)	6	8.3%
Gly (G)	10	13.9%
His (H)	0	0.0%
Ile (I)	5	6.9%
Leu (L)	6	8.3%
Lys (K)	2	2.8%
Met (M)	1	1.4%
Phe (F)	3	4.2%
Pro (P)	1	1.4%
Ser (S)	8	11.1%
Thr (T)	6	8.3%
Trp (W)	4	5.6%
Tyr (Y)	0	0.0%
Val (V)	3	4.2%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

## Anexo 3 MHCPred Alelo HLA-A\*0201

The HLA allele used in the test is: A0201			
The query sequence			
MSLLTEVETLTRNGWGCRCSDDSKLSTRGVQIASNENGLFGAIAAGFIEGGWCFWV ELIRGRPKESTIWTS			
Amino acid groups	Predicted $-\log I_{C_{50}}$ (M)	Predicted $I_{C_{50}}$ Value (nM)	Confidence of prediction (Max = 1)
<u>AIAGFIEGG</u>	7.607	24.72	0.89
<u>GLFGAIAAGF</u>	7.557	27.73	1.00
<u>SLLTEVETL</u>	7.201	62.95	1.00
<u>IEGGWCFWV</u>	6.993	101.62	0.89
<u>FWVELIRGR</u>	6.771	169.43	0.89
<u>ELIRGRPKE</u>	6.626	236.59	0.78
<u>ASNENGLFG</u>	6.548	283.14	0.78
<u>LTEVETLT</u>	6.544	285.76	1.00
<u>CFWVELIRG</u>	6.495	319.89	0.78
<u>KLSTRGVQI</u>	6.333	464.52	1.00
<u>WCFWVELIR</u>	6.314	485.29	0.89
<u>GFIEGGWCF</u>	6.294	508.16	0.89
<u>SDKLSTRGV</u>	6.136	731.14	0.89
<u>QIASNENGL</u>	6.122	755.09	1.00
<u>ENGLFGAIA</u>	6.121	756.83	0.78
<u>KESTIWTS</u>	6.083	826.04	0.78
<u>MSLLTEVET</u>	6.049	893.31	0.89
<u>TRGVQIASN</u>	6.023	948.42	0.78
<u>SNENGLFGA</u>	5.922	1196.74	0.89
<u>NENGLFGAI</u>	5.913	1221.80	0.89
<u>GAIAGFIEG</u>	5.874	1336.60	0.89
<u>FIEGGWCFW</u>	5.868	1355.19	0.89
<u>TLTRNGWGC</u>	5.706	1967.89	1.00
<u>TRNGWGCRC</u>	5.662	2177.71	0.89
<u>CSDSSDKLS</u>	5.66	2187.76	0.89
<u>GGWCFWVEL</u>	5.65	2238.72	1.00
<u>IRGRPKEST</u>	5.642	2280.34	0.89
<u>SSDKLSTRG</u>	5.638	2301.44	0.78
<u>DSSDKLSTR</u>	5.56	2754.23	0.89
<u>GWCFWVELI</u>	5.526	2978.52	0.89
<u>SDSSDKLST</u>	5.401	3971.92	0.89
<u>EVETLTRNG</u>	5.385	4120.98	0.78

<u>VELIRGRPK</u>	5.374	4226.69	0.78
<u>TEVETLTRN</u>	5.324	4742.42	0.78

ANEXO 4 MHCpred Alelo HLA-DRB1\*0101

The HLA allele used in the test is: DRB0101			
The query sequence			
MSLLTEVETLTRNGWGCRCSDDSSDKLSTRGVQIASNENGLFGAIAAGFIEGGWCFWV ELIRGRPKESTIWTS			
Amino acid groups	Predicted -logIC <sub>50</sub> (M)	Predicted IC <sub>50</sub> Value (nM)	Confidence of prediction (Max = 1)
<u>FWVELIRGR</u>	8.358	4.39	0.89
<u>RGVQIASNE</u>	8.171	6.75	0.89
<u>WVELIRGRP</u>	7.964	10.86	0.89
<u>TEVETLTRN</u>	7.879	13.21	0.78
<u>FIEGGWCFW</u>	7.803	15.74	0.78
<u>DSSDKLSTR</u>	7.773	16.87	1.00
<u>RCSDSSDKL</u>	7.605	24.83	0.67
<u>LFGAIAAGFI</u>	7.598	25.23	1.00
<u>CSDSSDKLS</u>	7.582	26.18	0.78
<u>EVETLTRNG</u>	7.575	26.61	0.78
<u>KLSTRGVQI</u>	7.52	30.20	0.89
<u>WGCRCSDDSS</u>	7.443	36.06	0.67
<u>CFWVELIRG</u>	7.382	41.50	0.78
<u>SDSSDKLST</u>	7.361	43.55	0.89
<u>VETLTRNGW</u>	7.302	49.89	0.89
<u>LIRGRPKE</u>	7.298	50.35	0.67
<u>NENGLFGAI</u>	7.268	53.95	1.00
<u>CRCSDSSDK</u>	7.253	55.85	0.67
<u>ELIRGRPKE</u>	7.208	61.94	0.89
<u>SDKLSTRGV</u>	7.196	63.68	1.00
<u>QIASNENGL</u>	7.188	64.86	1.00
<u>MSLLTEVET</u>	7.166	68.23	0.78
<u>RGRPKESTI</u>	7.127	74.64	0.78
<u>SNENGLFGA</u>	7.116	76.56	0.78
<u>TLTRNGWGC</u>	7.107	78.16	0.78
<u>PKESTIWTS</u>	7.095	80.35	0.78
<u>KESTIWTS</u>	7.094	80.54	1.00
<u>GVQIASNEN</u>	7.03	93.33	0.89
<u>LSTRGVQIA</u>	7.028	93.76	1.00
<u>GGWCFWVEL</u>	6.997	100.69	0.89

<u>SSDKLSTRG</u>	6.972	106.66	1.00
<u>NGWGCRCS</u>	6.952	111.69	0.89
<u>DKLSTRGVQ</u>	6.951	111.94	0.89
<u>GWGCRCS</u>	6.909	123.31	0.67
<u>TRGVQIASN</u>	6.901	125.60	0.89
<u>STRGVQIAS</u>	6.899	126.18	0.89
<u>LLTEVETLT</u>	6.889	129.12	0.89
<u>IASNENGLF</u>	6.888	129.42	0.78
<u>GCRCS</u>	6.878	132.43	0.67
<u>GLFGAIAGF</u>	6.865	136.46	1.00
<u>IEGGWCFWV</u>	6.836	145.88	0.89
<u>LTEVETLTR</u>	6.835	146.22	0.78
<u>LLTEVETL</u>	6.811	154.53	0.78
<u>WCFWVELIR</u>	6.807	155.96	0.78
<u>RNGWGCRCS</u>	6.769	170.22	0.56
<u>NGLFGAIAG</u>	6.702	198.61	0.89
<u>RPKESTIWT</u>	6.683	207.49	0.78
<u>GWCFWVELI</u>	6.681	208.45	0.89
<u>EGGWCFWVE</u>	6.657	220.29	0.56
<u>LTRNGWGCR</u>	6.644	226.99	0.67
<u>ENGLFGAIA</u>	6.6	251.19	0.89
<u>TRNGWGCR</u>	6.557	277.33	0.89
<u>IAGFIEGGW</u>	6.479	331.89	0.89
<u>IRGRPKEST</u>	6.317	481.95	0.89
<u>GFIEGGWCF</u>	6.212	613.76	0.67
<u>ETLTRNGWG</u>	6.191	644.17	0.67
<u>GAIAGFIEG</u>	6.147	712.85	0.89
<u>GRPKESTIW</u>	6.018	959.40	0.78
<u>FGAIAGFIE</u>	5.967	1078.95	1.00
<u>ASNENGLFG</u>	5.714	1931.97	0.89
<u>AGFIEGGWC</u>	5.514	3061.96	0.89
<u>VELIRGRP</u>	5.514	3061.96	0.78
<u>AIAGFIEGG</u>	5.369	4275.63	1.00

## ANEXO 5 MHCpred alelo HLA-DRB1\*0401

The HLA allele used in the test is: DRB0401			
The query sequence			
MSLLTEVETLTRNGWGCRCS DSSDKLSTRGVQIASNENGLFGA IAGFIEGGWCFWV ELIRGRPKESTIWTS G			
Amino acid groups	Predicted $-\log IC_{50}$ (M)	Predicted $IC_{50}$ Value (nM)	Confidence of prediction (Max = 1)
<u>GVQIASNEN</u>	7.586	25.94	0.78
<u>WGCRCS DSS</u>	6.849	141.58	0.67
<u>GWCFWVELI</u>	6.767	171.00	0.78
<u>CRCS DSSDK</u>	6.621	239.33	0.78
<u>ASNENGLFG</u>	6.594	254.68	0.89
<u>VELIRGRP K</u>	6.563	273.53	1.00
<u>NGLFGA IAG</u>	6.521	301.30	0.78
<u>DKLSTRGVQ</u>	6.479	331.89	0.89
<u>MSLLTEVET</u>	6.409	389.94	0.89
<u>FIEGGWCFW</u>	6.4	398.11	0.67
<u>ETLTRNGWG</u>	6.376	420.73	0.56
<u>PKESTIWTS</u>	6.362	434.51	0.67
<u>TEVETLTRN</u>	6.211	615.18	0.89
<u>GAIAGFIEG</u>	6.204	625.17	0.56
<u>GGWCFWVEL</u>	6.204	625.17	0.44
<u>FGAIAGFIE</u>	6.198	633.87	0.89
<u>NENGLFGAI</u>	6.195	638.26	0.78
<u>TRGVQIASN</u>	6.191	644.17	0.89
<u>WCFWVELIR</u>	6.185	653.13	0.78
<u>GRPKESTIW</u>	6.185	653.13	0.67
<u>VQIASNENG</u>	6.173	671.43	0.78
<u>SNENGLFGA</u>	6.162	688.65	0.89
<u>ELIRGRPKE</u>	6.16	691.83	0.67
<u>TLTRNGWGC</u>	6.151	706.32	0.56
<u>STRGVQIAS</u>	6.149	709.58	0.67
<u>TRNGWGCR C</u>	6.147	712.85	0.78
<u>RGVQIASNE</u>	6.117	763.84	0.67
<u>SLLTEVETL</u>	6.11	776.25	0.78
<u>DSSDKLSTR</u>	6.089	814.70	0.78
<u>EVETLTRNG</u>	6.083	826.04	0.78
<u>WVELIRGRP</u>	6.073	845.28	1.00
<u>CSDSSDKLS</u>	6.06	870.96	0.89

<u>GLFGAIAGF</u>	6.047	897.43	0.89
<u>RNGWGCRC</u>	6.045	901.57	0.56
<u>EGGWCFWVE</u>	5.997	1006.93	0.33
<u>QIASNENGL</u>	5.993	1016.25	0.89
<u>RPKESTIWT</u>	5.989	1025.65	0.67
<u>AIAGFIEGG</u>	5.977	1054.39	0.78
<u>VETLTRNGW</u>	5.972	1066.60	0.78
<u>NGWGCRCSD</u>	5.963	1088.93	0.67
<u>GWGCRCSDS</u>	5.962	1091.44	0.67
<u>KLSTRGVQI</u>	5.935	1161.45	0.89
<u>RGRPKESTI</u>	5.911	1227.44	0.67
<u>IRGRPKEST</u>	5.911	1227.44	0.89
<u>AGFIEGGWC</u>	5.854	1399.59	0.78
<u>LSTRGVOIA</u>	5.851	1409.29	0.78
<u>RCSDSSDKL</u>	5.824	1499.68	0.78
<u>SDSSDKLST</u>	5.77	1698.24	0.89
<u>SSDKLSTRG</u>	5.72	1905.46	0.78
<u>LTEVETLTR</u>	5.698	2004.47	0.78
<u>GFIEGGWCF</u>	5.682	2079.70	0.67
<u>KESTIWTSG</u>	5.679	2094.11	0.78
<u>IEGGWCFWV</u>	5.617	2415.46	0.78
<u>LFGAIAGFI</u>	5.613	2437.81	1.00
<u>IAGFIEGGW</u>	5.609	2460.37	0.89
<u>GCRCDSSD</u>	5.601	2506.11	0.67
<u>CFWVELIRG</u>	5.599	2517.68	0.78
<u>LIRGRPKE</u>	5.531	2944.42	0.89
<u>SDKLSTRGV</u>	5.526	2978.52	0.89
<u>LTRNGWGCR</u>	5.506	3118.89	0.67
<u>FWVELIRGR</u>	5.48	3311.31	1.00
<u>IASNENGLF</u>	5.469	3396.25	0.89
<u>LLTEVETLT</u>	5.467	3411.93	0.89
<u>ENGLFGAIA</u>	5.352	4446.31	0.78

## ANEXO 6 MHCpred alelo HLA-DRB1\*0701

The HLA allele used in the test is: DRB0701			
The query sequence			
MSLLTEVETLTRNGWGCRCS DSSDKLSTRGVQIASNENGLFGA IAGFIEGGWCFWV ELIRGRPKESTIWTS G			
Amino acid groups	Predicted $-\log I C_{50}$ (M)	Predicted $I C_{50}$ Value (nM)	Confidence of prediction (Max = 1)
<u>RGVQIASNE</u>	7.221	60.12	1.00
<u>FWVELIRGR</u>	7.139	72.61	1.00
<u>LFGA IAGFI</u>	7.09	81.28	1.00
<u>KLSTRGVQI</u>	7.025	94.41	0.89
<u>MSLLTEVET</u>	7.004	99.08	0.89
<u>PKESTIWTS</u>	6.871	134.59	1.00
<u>VETLTRNGW</u>	6.858	138.68	0.89
<u>ELIRGRPKE</u>	6.838	145.21	1.00
<u>DKLSTRGVQ</u>	6.761	173.38	1.00
<u>TLTRNGWGC</u>	6.645	226.46	0.89
<u>QIASNENGL</u>	6.636	231.21	1.00
<u>EVETLTRNG</u>	6.621	239.33	1.00
<u>RPKESTIWT</u>	6.615	242.66	0.89
<u>LIRGRP KES</u>	6.569	269.77	0.89
<u>GAIAGFIEG</u>	6.511	308.32	1.00
<u>ENGLFGAIA</u>	6.471	338.06	1.00
<u>EGGWCFWVE</u>	6.395	402.72	1.00
<u>NGLFGA IAG</u>	6.336	461.32	1.00
<u>IAGFIEGGW</u>	6.3	501.19	1.00
<u>FIEGGWCFW</u>	6.263	545.76	1.00
<u>GWGCRCS DS</u>	6.228	591.56	0.78
<u>WVELIRGRP</u>	6.226	594.29	1.00
<u>RNGWGC RCS</u>	6.216	608.14	0.89
<u>SLLTEVETL</u>	6.21	616.60	1.00
<u>LSTRGVQIA</u>	6.203	626.61	0.89
<u>GGWCFWVEL</u>	6.144	717.79	1.00
<u>RCSDSSDKL</u>	6.141	722.77	0.89
<u>VQIASNENG</u>	6.134	734.51	1.00
<u>LTRNGWGCR</u>	6.105	785.24	1.00
<u>GFIEGGWCF</u>	6.056	879.02	1.00
<u>WCFWVELIR</u>	6.048	895.36	0.89
<u>ASNENGLFG</u>	6.026	941.89	1.00

<u>STRGVQIAS</u>	6.026	941.89	1.00
<u>RGRPKESTI</u>	6.022	950.60	1.00
<u>IEGGWCFWV</u>	6.019	957.19	0.78
<u>IRGRPKEST</u>	5.995	1011.58	1.00
<u>NENGLFGAI</u>	5.955	1109.17	1.00
<u>GWCFWVELI</u>	5.924	1191.24	1.00
<u>SDKLSTRGV</u>	5.902	1253.14	1.00
<u>IASNENGLF</u>	5.889	1291.22	1.00
<u>GRPKESTIW</u>	5.758	1745.82	1.00
<u>AIAGFIEGG</u>	5.757	1749.85	1.00
<u>LLTEVELTL</u>	5.749	1782.38	0.89
<u>ETLTRNGWG</u>	5.743	1807.17	0.78
<u>TRNGWGCRC</u>	5.726	1879.32	1.00
<u>FGAIAGFIE</u>	5.725	1883.65	1.00
<u>KESTIWTSG</u>	5.529	2958.01	1.00
<u>CRCSDSSDK</u>	5.519	3026.91	0.78
<u>TEVELTRN</u>	5.506	3118.89	1.00
<u>CFWVELIRG</u>	5.499	3169.57	0.89
<u>LTEVELLTR</u>	5.441	3622.43	1.00
<u>AGFIEGGWC</u>	5.432	3698.28	0.89
<u>VELIRGRPK</u>	5.397	4008.67	0.89
<u>CSDSSDKLS</u>	5.38	4168.69	0.89
<u>WGCRCSDSS</u>	5.364	4325.14	1.00
<u>DSSDKLSTR</u>	5.357	4395.42	1.00
<u>NGWGCRCSD</u>	5.353	4436.09	1.00

## ANEXO 7: ElliPro

### Predicted peptides:

No.	Start Position	End Position	Peptide	Peptide Length
1	4	9	LTEVET	6
2	16	22	GCRCSDS	7
3	28	34	TRGVQIA	7
4	41	47	FGAIAGF	7
5	53	59	CFWVELI	7

## IEDB Analysis Resource

[Home](#) [Help](#) [Example](#) [Reference](#) [Download](#) [Contact](#)

Input Sequences: fileXHpJbT.pdb

Chain: A  
1 MSLLTEVETL TRNGWGRCS DSSDKLSTRG VQIASNENGL FGAIAGFIEG GNCFWVELIR  
61 GRPKESTIWT SG

Predicted Linear Epitope(s):

No.	Chain	Start	End	Peptide	Number of residues	Score	3D structure
1	A	43	53	AIAGFIEGGWC	11	0.687	<a href="#">View</a>
2	A	11	28	TRNGWGRCS DSSDKL	18	0.657	<a href="#">View</a>
3	A	60	72	RGRPKESTIWTSG	13	0.622	<a href="#">View</a>
4	A	1	4	MSLL	4	0.549	<a href="#">View</a>

Predicted Discontinuous Epitope(s):

No.	Residues	Number of residues	Score	3D structure
1	A:G42, A:A43, A:I44, A:A45, A:G46, A:F47, A:I48, A:E49, A:G50, A:G51, A:W52	11	0.708	<a href="#">View</a>
2	A:R60, A:G61, A:R62, A:P63, A:H64, A:E65, A:S66, A:T67, A:I68, A:W69, A:T70, A:S71	12	0.639	<a href="#">View</a>
3	A:E8, A:L10, A:T11, A:R12, A:N13, A:G14, A:W15, A:G16, A:C17, A:R18, A:C19, A:S20, A:D21, A:S22, A:S23, A:D24, A:K25, A:L26	18	0.623	<a href="#">View</a>
4	A:S2, A:L3, A:L4	3	0.523	<a href="#">View</a>

## ANEXO 7

Following up ins/sol separation by posQmax (bigger = ins) list all CA whose closest patch is pos (P): neg (N) : neither (H) additionally for pos, give ratio = patch size / threshold size dist = distance of this CA to pos patch center (for modification) since we use CA for whole aa, can get odd results e.g. Lys in neg patch we could improve with more detailed atom-based analysis

residue	PNH	ratio	npoints	dist
CA MET -	1 POS	0.250	747	5.5
CA SER -	2 POS	0.250	747	6.9
CA LEU -	3 HYD		6860	14.5
CA LEU -	4 NEG		2218	12.4
CA THR -	5 NEG		2218	13.4
CA GLU -	6 NEG		852	5.3
CA VAL -	7 NEG		852	9.3
CA GLU -	8 NEG		2218	13.3
CA THR -	9 NEG		2218	14.5
CA LEU -	10 NEG		852	13.6
CA THR -	11 NEG		852	13.1
CA ARG -	12 NEG		852	14.1
CA ASN -	13 NEG		852	14.9
CA GLY -	14 HYD		6860	16.4
CA TRP -	15 HYD		6860	14.3
CA GLY -	16 HYD		6860	17.7

CA CYS - 17 HYD	6860	20.3
CA ARG - 18 HYD	6860	23.6
CA CYS - 19 HYD	6860	23.4
CA SER - 20 HYD	6860	18.0
CA ASP - 21 HYD	6860	19.1
CA SER - 22 NEG	337	2.9
CA SER - 23 HYD	6860	17.7
CA ASP - 24 HYD	6860	20.1
CA LYS - 25 HYD	6860	17.8
CA LEU - 26 HYD	6860	16.6
CA SER - 27 POS 0.005	16	1.1
CA THR - 28 NEG	2218	14.9
CA ARG - 29 POS 0.250	747	8.4
CA GLY - 30 HYD	6860	8.6
CA VAL - 31 POS 0.250	747	4.2
CA GLN - 32 HYD	6860	11.9
CA ILE - 33 HYD	6860	12.1
CA ALA - 34 HYD	6860	12.1
CA SER - 35 NEG	2218	13.9
CA ASN - 36 NEG	2218	10.7
CA GLU - 37 HYD	6860	12.3
CA ASN - 38 POS 0.035	106	3.8
CA GLY - 39 HYD	6860	9.5

CA LEU - 40 HYD		6860	8.4
CA PHE - 41 NEG		2218	6.5
CA GLY - 42 HYD		6860	8.4
CA ALA - 43 HYD		6860	8.5
CA ILE - 44 HYD		6860	12.7
CA ALA - 45 NEG		852	7.4
CA GLY - 46 HYD		6860	15.8
CA PHE - 47 HYD		6860	17.9
CA ILE - 48 HYD		6860	22.5
CA GLU - 49 HYD		6860	21.9
CA GLY - 50 HYD		6860	16.9
CA GLY - 51 NEG		852	3.1
CA TRP - 52 NEG		852	6.3
CA CYS - 53 HYD		6860	10.6
CA PHE - 54 POS	0.250	747	7.2
CA TRP - 55 POS	0.250	747	8.7
CA VAL - 56 POS	0.250	747	9.2
CA GLU - 57 NEG		2218	7.6
CA LEU - 58 NEG		2218	8.5
CA ILE - 59 NEG		2218	11.1
CA ARG - 60 NEG		2218	11.6
CA GLY - 61 NEG		2218	12.7
CA ARG - 62 HYD		6860	20.0

CA PRO - 63 HYD	6860	23.4
CA LYS - 64 HYD	6860	22.3
CA GLU - 65 NEG	2218	19.2
CA SER - 66 HYD	6860	26.3
CA THR - 67 HYD	6860	24.8
CA ILE - 68 HYD	6860	28.2
CA TRP - 69 HYD	6860	23.9
CA THR - 70 HYD	6860	21.9
CA SER - 71 NEG	2218	15.6
CA GLY - 72 NEG	2218	10.