



UNIVERSIDAD MICHOACANA DE SAN NICOLÁS DE HIDALGO

FACULTAD DE CIENCIAS MÉDICAS Y BIOLÓGICAS
“DR. IGNACIO CHÁVEZ”
DIVISIÓN DE ESTUDIOS DE POSGRADO

Tesis:

Análisis bioinformático de la inmunogenicidad de la proteína
PE_PGRS18 de *Mycobacterium tuberculosis*

Que para obtener el grado de Maestría en Ciencias de la Salud

PRESENTA:

Q.F.B Eva Nelida Jiménez Ruiz

Directora de tesis:

D. en C. Ma. Soledad Vázquez Garcidueñas



Morelia, Michoacán, Agosto, 2020.

El comité tutorial designado por la División de Estudios de Posgrado de la Facultad de Ciencias Médicas y Biológicas “Dr. Ignacio Chávez” de la Universidad Michoacana de San Nicolás de Hidalgo Aprobó la tesis presentada por:

Química Farmacobióloga
EVA NELIDA JIMÉNEZ RUIZ

Doctora en Ciencias Biológicas
Ana Edith Higareda Mendoza

Doctora en Ciencias (Biología)
Ma. Soledad Vázquez Garcidueñas

Doctor en Biotecnología de Plantas
Gerardo Vázquez Marrufo

Doctora en Ciencias
Alejandra Ochoa Zarzosa

Doctora en Ciencias Biológicas
Ana Laura Guillén Nepita

La Maestría en Ciencias de la Salud de la
Universidad Michoacana de San Nicolás de Hidalgo
Pertenece al Padrón de Posgrado de Excelencia de CONACyT

La estudiante de la Facultad de Ciencias Médicas y Biológicas
“Dr. Ignacio Chávez” de la
Universidad Michoacana de San Nicolás de Hidalgo
agradece al CONACyT la beca que le fue otorgada
durante la realización de su tesis
Núm de becario CONACyT: 702755

Se agradece también el apoyo brindado por CONACyT
para la realización de una estancia de investigación en
el Institute for Integrative Systems Biology, Parc Científic of the
Universitat de València octubre – diciembre de 2019

El presente trabajo de Tesis se realizó en:

Universidad Michoacana de San Nicolás de Hidalgo
Facultad de Ciencias Médicas y Biológicas
“Dr. Ignacio Chávez”
División de Estudios de Posgrados
Laboratorio de Genética Molecular Microbiana

y

Institute for Integrative Systems Biology
of Parc Científic of the Universitat de València

Dirección de Tesis:

Doctora en Ciencias (Biología)

Ma. Soledad Vázquez Garcidueñas

Universidad Michoacana de San Nicolás de Hidalgo

Contribuciones:

Doctora en Ciencias Biológicas

Mireia Coscollá Devís

Parc Científic of the Universitat de València

ABREVIATURAS

BCG. Bacilo de Calmette y Guérin.

CENAPRECE. Centro Nacional de Programas Preventivos y Control de Enfermedades.

CPA. Célula presentadora de antígeno.

DCs. Células dendríticas.

HBHA. Adhesina hemaglutinina que se une a heparina.

HLA. Antígeno leucocitario humano.

iNOS. Especies reactivas de nitrógeno inducible.

LAM. Lipoarabinoamano.

LPS. Lipopolisacáridos.

MHC. Complejo Mayor de Histocompatibilidad.

OMS. Organización Mundial de la Salud.

PAMPS. Patrones moleculares asociados a patógenos.

PCR. Reacción en cadena de polimerasa.

RRPM. Receptores que reconocen patrones moleculares.

TB. Tuberculosis.

TBEP. Tuberculosis extra pulmonar.

TCR. Receptor de linfocitos T.

TNF. Factor de necrosis tumoral.

TRL. Receptor Toll-like.

VIH. Virus de inmunodeficiencia humana.

WHO. World Health Organization.

ÍNDICE

1. MARCO TEÓRICO	1
1.1 Agente causal de la tuberculosis	1
1.2 Tipos de tuberculosis	2
1.2.1 Tuberculosis pulmonar	2
1.2.2 Tuberculosis extrapulmonar	2
1.2.3 Tuberculosis diseminada	3
1.3 Epidemiología	4
1.4 Vía de transmisión	5
1.5 Mecanismo de invasión.....	5
1.6 Genes de virulencia	9
1.7 Familia de genes pe/ppe.....	10
1.7.1 Localización y función de las proteínas PE/PPE.....	12
1.7.2 Polimorfismos en las familias de proteínas PE/PPE.....	16
1.7.3 Regulación de la transcripción de genes pe/ppe.....	18
1.7.4 Antígeno Leucocitario Humano (HLA).....	19
1.7.5 Unión de las proteínas PE/PPE con el Complejo Mayor de Histocompatibilidad (MHC).....	21
1.8 Herramientas bioinformáticas en el análisis de la interacción patógeno hospedero	22
2. ANTECEDENTES	25
3. JUSTIFICACIÓN	27
4. HIPÓTESIS	28
5. OBJETIVOS	28
5.1 Objetivo general	28
5.2 Objetivos específicos	28
6. ESTRATEGIA EXPERIMENTAL	29
7. MATERIALES Y MÉTODOS	30
7.1 Material biológico	30
7.2 Caracterización in silico de la proteína PE_PGRS18	31
7.2.1 Propiedades	31
7.2.2 Modelado in silico de la proteína PE_PGRS18.....	32

7.2.3 Inmunoinformática	33
8. RESULTADOS	38
8.1 Caracterización de la proteína PE_PGRS18	38
8.2 Identificación de péptidos de la proteína PE_PGRS18 que interaccionan con moléculas de HLA clase I y II.....	40
8.2.1 Epítopes de unión a células T CD8+ en las secuencias de estudio.....	42
8.2.2 Epítopes de unión a células T CD4+ en las secuencias de estudio.....	46
8.3 Mutaciones que estimulan/evaden la respuesta inmune del hospedero	49
8.3.1 SNPs que estimulan y evaden la respuesta inmune	51
8.4 Modelado 3D de epítopes de la proteína PE_PGRS18 y modelado 3D de alelos de clase I y II del HLA	52
8.4.1 Epítopes	52
8.4.2 Alelos que reconocen un mayor número de epítopes en la proteína PE_PGRS18.....	52
8.5 Docking molecular entre los epítopes y los alelos de estudio.....	58
8.6 Modelado 3D por homología de la proteína PE_PGRS18.....	70
8.6.1 Comparaciones globales de la estructura tridimensional	72
8.7 Docking molecular entre la proteína PE_PGRS18 y los receptores TLRs 2 y 4.....	74
8.8 Predicción de la interacción de la proteína PE_PGRS18 H37Rv.	76
9. DISCUSIÓN	77
10. RESUMEN DE RESULTADOS	85
11. CONCLUSIÓN	86
12. BIBLIOGRAFÍA	87
ANEXOS	97

ÍNDICE DE TABLAS

Nº	Descripción	Pág.
1	Características clínicas, genotípicas y resistencia a antibióticos de las cepas de estudio	29
2	Alelos de clase I y II del HLA más frecuentes de distintas regiones del mundo	33
3	Epítopes de alta afinidad reconocidos por alelos de clase I del HLA de la región mexicana y distintas regiones	44
4	Probabilidad de que los alelos de clase I de la región mexicana y de distintas regiones, reconozcan el mismo número de epítopes en los dominios de la proteína	45
5	Epítopes reconocidos con alta afinidad por alelos de clase II del HLA de diferentes regiones geográficas	47
6	SNPs que estimulan y evaden la respuesta inmune mediada por células T CD4+ y CD8+	50
7	Información de los moldes representativos utilizados para el modelado 3D por homología de alelos de clase I y II del HLA	52
8	Calidad de los modelos 3D de los alelos de clase I y II del HLA	53
9	Términos del refinado de algunos alelos de clase I y II del HLA	54
10	Modelados 3D de los alelos de clase I y II del HLA	55
11	Epítopes reconocidos por los alelos de clase I y II del HLA	58
12	Docking molecular entre los epítopes y alelos de clase I y II del HLA	59
13	Características de los moldes representativos utilizados para el modelado 3D por homología de la proteína PE_PGRS18-H37Rv	69
14	Cepas modeladas 3D con mutaciones frecuentes en distintas cepas y criterios de validación de la calidad de cada modelo	71
15	Criterios de refinamiento y validación de la calidad de los TLRs 2 y 4	73

ÍNDICE DE FIGURAS

Nº	Descripción	Pág.
1	Mecanismo de invasión de <i>Mycobacterium tuberculosis</i>	6
2	Estructura de los genes <i>pe/ppa</i>	11
3	Regiones estructurales y no estructurales de la subfamilia PE_PGRS	15
4	Hélice transmembrana en la proteína PE_PGRS18 y regiones desordenadas y flexibles	37
5	Presencia, localización y péptido señal de la proteína PE_PGRS18	38
6	Regiones de unión proteína-proteína	39
7	Epítopes de células T CD4+ asociadas con los dominios de la proteína PE_PGRS18 H37Rv	40
8	Epítopes de células T CD8+ asociadas con los dominios de la proteína PE_PGRS18 H37Rv	42
9	Epítopes de células T CD8+ asociadas con los dominios de las proteínas PE_PGRS18 de las cepas de estudio	43
10	Epítopes de células T CD4+ asociadas con los dominios de las proteínas PE_PGRS18 de las cepas de estudio	46
11	Mutaciones que estimulan o evaden el reconocimiento de péptidos por las células T CD4+ Y CD8+	49
12	Modelado 3D de la proteína PE_PGRS18 H37Rv	70
13	Superposición de las estructuras de las cepas modeladas y la estructura de referencia PE_PGRS18 H37Rv	72
14	Modelados 3D de los receptores TLR2 y TLR4	74
15	Interacción de la proteína PE_PGRS18	75

RESUMEN

Las proteínas PE y PPE de *M. tuberculosis* han demostrado ser potentes antígenos que son reconocidos por células B y T. Nuestro grupo de trabajo encontró que el gen *pe_pgrs18* de cepas de *M. tuberculosis* de Michoacán era de los más polimórficos. Es importante buscar péptidos en la proteína PE_PGRS18 que se unan a moléculas del CMH clase I y II, para analizar si este gen es capaz de inducir una respuesta inmune celular (respuesta inmune adaptativa) y si la proteína de estudio puede interactuar con los TLRs 2 y 4 del hospedero (respuesta inmune innata). Estos péptidos podrían probarse a futuro para su inclusión en una vacuna contra la tuberculosis o para la identificación temprana de la enfermedad. Para ello se analizaron 46 secuencias del gen *pe_pgrs18* de cepas Michoacanas y de Queretaro de *M. tuberculosis*. Se utilizaron los servidores Mega, PredictProtein, Signal -IP, Computer pI/Mw, Pylor 2, Prism, NetMHCpan, NetMHCIIpan y Galaxyepdock. Resultados. La proteína PE_PGRS18 tiene loops (43.11%), laminas beta (40.04%), hélices alfa (16.85 %), una secuencia señal en la posición 36-56 aa, un punto isoeléctrico de 4.27, un peso molecular de 42074.79 Da y 17 regiones de unión a proteínas. La proteína es integral de membrana, un 78.56% es hidrofóbica y carece de enlaces disulfuro. Se identificaron 5,114 epítopes de alta afinidad a linfocitos T CD8+; además de 995 alelos de alta afinidad a linfocitos T CD4+. Se identificaron 332 mutaciones de las cuales 244 estimulan y 88 evaden la respuesta inmune mediada por los linfocitos T CD4+ y CD8+. No se observó una interacción directa de la proteína PE_PGRS18 con los receptores TLR2 y TLR4.

Palabras clave: Proteína PE_PGRS18, inmunogenicidad, bioinformática, *M. tuberculosis*, mutación.

ABSTRACT

The PE and PPE proteins of *M. tuberculosis* have proven to be potent antigens that are recognized by B and T cells. Our working group found that the *pe_pgrs18* gene of *M. tuberculosis* strains of Michoacán was one of the most polymorphic. It is important to look for peptides in the PE_PGRS18 protein that bind to CMH class I and II molecules, to analyze whether this gene is capable of inducing a cellular immune response (adaptive immune response) and whether the study protein can interact with the host's TLRs 2 and 4 (innate immune response). These peptides could be tested in the future for inclusion in a tuberculosis vaccine or for early identification of the disease. For this, 46 sequences of the *pe_pgrs18* gene from *M. tuberculosis* strains from Michoacan and Queretaro of were analyzed. Mega, PredictProtein, Signal -IP, Computer pl / Mw, Ppyre 2, Prism, NetMHCpan, NetMHCIIpan and Galaxypepdock servers were used. Results. The PE_PGRS18 protein has loops (43.11%), beta sheets (40.04%), alpha helices (16.85%), a signal sequence at position 36-56 aa, an isoelectric point of 4.27, a molecular weight of 42074.79 Da and 17 protein-binding regions. The protein is membrane integral; 78.56% is hydrophobic and lacks disulfide bonds. 5,114 epitopes of high affinity to CD8 + T lymphocytes were identified, the A_6802 allele binds to a greater number of epitopes, in addition to 995 high affinity alleles to CD4 + T lymphocytes, the DRB_0407 allele binds to a greater number of epitopes. 332 mutations were identified, of which 244 stimulate and 88 evade the immune response mediated by CD4 + and CD8 + T cells. No direct interaction of the PE_PGRS18 protein was observed with the TLR2 and TLR4 receptors.

Key words. PE_PGRS18 protein, immunogenic, bioinformatics, *M. tuberculosis*, mutation.

1. MARCO TEÓRICO

La tuberculosis es una de las diez principales causas de mortalidad en el mundo, la cual afecta principalmente a los adultos en los años más productivos de su vida, aunque otros grupos de edad no están exentos de riesgo. Las personas infectadas por el virus de inmunodeficiencia humana (VIH) tienen entre 20 y 30 veces más probabilidades de desarrollar tuberculosis activa. El riesgo también es más elevado en las personas que padecen otros trastornos que comprometen el sistema inmunitario, como desnutrición o diabetes, así como en los consumidores de tabaco (WHO, 2018). La tuberculosis puede ser mortal si el paciente no recibe el tratamiento adecuado. Actualmente, se considera un severo problema de salud pública la presencia de cepas fármaco resistentes (WHO, 2017).

1.1 Agente causal de la tuberculosis

La tuberculosis (TB) es una enfermedad infecciosa crónica causada por un grupo de bacterias denominado “complejo *Mycobacterium tuberculosis*”, que incluye a las especies *Mycobacterium tuberculosis*, *M. bovis* (incluida la cepa BCG) y *Mycobacterium africanum*, siendo la primera especie la que se aísla con mayor frecuencia de pacientes humanos. Se incluye también *M. microti*, productor de tuberculosis en rata (Ramírez *et al.*, 2005)

La especie *M. tuberculosis* se caracteriza por ser un microorganismo intracelular obligado, aerobio, inmóvil, que se replica dentro de los fagosomas de los macrófagos. Su tiempo de duplicación es de 12 horas o más, por lo que su crecimiento en medios de cultivo *in vitro* es muy lento. Es sensible al calor, los rayos ultravioleta y la radiación solar directa; presenta resistencia a ácidos, alcoholes, álcalis, desinfectantes y a la desecación. Además, es naturalmente resistente a distintos antibióticos debido principalmente a la envoltura celular altamente hidrofóbica que actúa como una barrera permeable, lo que hace difícil su tratamiento (Ramírez *et al.*, 2005).

1.2 Tipos de tuberculosis

Aunque la TB se trata principalmente de una enfermedad pulmonar, que presentan el 85% de los casos, también afecta a otros órganos y tejidos. Alrededor del 10% de los individuos infectados desarrollan la enfermedad activa, la mayor parte en el transcurso de los dos primeros años posteriores a la adquisición de la infección (Martínez *et al.*, 2015).

1.2.1 Tuberculosis pulmonar

Como se mencionó anteriormente, la localización pulmonar de la tuberculosis (TBP) es la más frecuente, pero puede observarse en cualquier órgano. El cuadro clínico de la enfermedad pulmonar es de comienzo insidioso y naturaleza crónica. El síndrome de impregnación bacilar está constituido por síntomas generales como astenia, adinamia, hiperoxia, pérdida de peso, febrícula vespertina y sudoración nocturna asociada a sintomatología respiratoria que incluye tos, expectoración y disnea. En aproximadamente 20% de los casos, los síntomas característicos como astenia, anorexia, febrícula y pérdida de peso, son el único hallazgo de la enfermedad. La tos es por mucho el síntoma pulmonar más frecuente; al inicio puede ser seca, y a medida que avanza la enfermedad se transforma en productiva, con expectoración mucosa, mucopurulenta o hemoptoica. Las manifestaciones clínicas de la TB dependen de la edad y del sistema inmune del paciente (Aidar *et al.*, 2013).

1.2.2 Tuberculosis extrapulmonar

En el año 2013, el Programa Nacional de Control de Tuberculosis reportó 0.8 millones de casos de tuberculosis extrapulmonar (TBEP) en todo el mundo. La TBEP alcanza alrededor del 15-20% de los casos de tuberculosis; sin embargo, en los pacientes con VIH la TBEP puede constituir hasta el 50% de los casos (Raj *et al.*, 2016). La TBEP se define como enfermedad de tuberculosis fuera del parénquima celular. Al igual que la tuberculosis pulmonar, la TBEP puede ocurrir como una infección primaria en la cual los bacilos ingresan al cuerpo a través de los

pulmones y luego se diseminan a otras partes del cuerpo ya sea a través del sistema linfático o por vía hematológica. Un foco latente de bacilos puede reactivarse y luego difundir de la misma manera (a través del sistema linfático o por vía hematológica) (Karakousis *et al.*, 2017). La TBEP se puede encontrar en distintas formas, incluyendo tuberculosis pleural, linfadenitis tuberculosa, tuberculosis del sistema nervioso central, tuberculosis ocular, tuberculosis pericardial, tuberculosis esquelética, tuberculosis genitourinaria, tuberculosis gastrointestinal y tuberculosis cutánea; no obstante, todos los órganos pueden verse afectados (Bouhout *et al.*, 2017).

1.2.3 Tuberculosis diseminada

Esta forma del padecimiento compromete más de dos órganos. Hay gran deterioro del estado general, presentándose hipertermia, disnea y astenia, e incluye además signos clínicos de aquellos órganos que pueden estar afectados como derrame pleural, trastornos digestivos, hematológicos (anemia, linfopenia, trombocitopenia), hepatoesplenomegalia, y a veces signos meníngeos. En las evoluciones crónicas se observa compromiso del estado general y fiebre de origen desconocido. En las formas agudas puede presentarse como síndrome del distrés respiratorio del adulto. La naturaleza no específica de la presentación es probablemente la causa más frecuente del retraso en el diagnóstico (Aidar *et al.*, 2013).

El potencial de infectividad de un paciente con tuberculosis depende, fundamentalmente de cuatro factores: 1) severidad y frecuencia de la tos, 2) carácter y volumen de las secreciones, 3) número de bacilos de la fuente de infección (los pacientes con baciloscopia positiva son los más infectantes) y 4) uso de la quimioterapia, ya que se ha determinado que después de dos semanas de tratamiento, se produce una reducción en el número de bacilos cercana al 99% (Aidar *et al.*, 2013).

1.3 Epidemiología

En 2016 se reportaron 10.4 millones de casos de tuberculosis y 1.7 millones de muertes por esta enfermedad, entre los cuales, 0.4 millones de personas presentaron VIH. Más del 95% de las muertes por tuberculosis se producen en países de ingresos bajos y medianos. Siete países concentran el 64% de la mortalidad total, incluyendo a la India, seguida de Indonesia, China, Filipinas, Pakistán, Nigeria y Sudáfrica (OMS, 2017). En el caso de América, la OMS reportó en el 2017 un estimado de 273 574 casos con una tasa de incidencia de 28 casos por 100 000 habitantes. Por lo que, los datos epidemiológicos muestran que la incidencia anual ha disminuido constantemente desde 1990, aunque esta disminución parece ser más lenta en el último año (WHO, 2015). La tasa de incidencia ha estado disminuyendo en todas las subregiones del continente, excepto México y América Central, donde se ha mantenido constante durante los últimos siete años.

Según el registro y notificación de la Organización Mundial de la Salud en el año 2016, se estimó un total de 28 000 casos nuevos de tuberculosis de todas las formas en México, correspondiendo a una incidencia de 16.8 casos por cada 100 mil habitantes. La tasa más elevada se ubicó en Veracruz con 27.4 caso por cada 100 mil habitantes y Baja California Norte con 57.9 casos por cada 100 mil habitantes (Cenaprece, 2016). En la zona norte del país, que es la más desarrollada, se ha registrado que los estados de Baja California (57.9 por 100,000 hab.), Sonora (37 por 100,000 hab.), Chihuahua (19 por 100,000 hab.), Coahuila (17.4 por 100,000 hab.), Nuevo León (28.9 por 100,000 hab.) y Tamaulipas (31.4 por 100,000 hab.) presentan la mayor incidencia de tuberculosis de todas las formas a nivel nacional (Infanzó-Ruiz, 2018). Por otro lado, Michoacán es uno de los estados con menor incidencia de casos de tuberculosis, reportándose en 2016 en la entidad 7.4 casos por cada 100 mil habitantes de nuevos casos de tuberculosis de todas las formas, incluyendo tuberculosis pulmonar y meníngea, entre otras formas (Cenaprece, 2016).

1.4 Vía de transmisión

El ser humano es el reservorio natural preponderante de *M. tuberculosis*, siendo la vía aérea la principal ruta de transmisión. Un paciente con tuberculosis pulmonar puede transmitir este bacilo por medio de la tos, el estornudo y otros movimientos respiratorios similares, dispersándose a su alrededor por medio de pequeñas gotas llamadas “gotitas de *flugge*” (Taraco, 2007).

1.5 Mecanismo de invasión

La infección inicia con la entrada del bacilo de la tuberculosis en el cuerpo, a través del tracto respiratorio. Cuando el bacilo pasa a vías respiratorias bajas, se cree que es fagocitado por los macrófagos alveolares, neutrófilos y células dendríticas (DCs), los cuales constituyen la primera línea de defensa del hospedero (O’Garra *et al.*, 2013). El bacilo es reconocido a través de receptores que reconocen patrones moleculares (RRPM) como receptores de mantenimiento, receptor de manosa, CD14, CD44, DC-SIGN (*Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin*), receptores para opsoninas, receptores para proteínas surfactantes, FcR (Receptor del Factor cristalizante), y receptor de proteínas del complemento (Taraco, 2007). El reconocimiento de LAM (lipoarabinoamano) y la proteína secretora de 19 kDa de *M. tuberculosis* es mediada por receptores tipo Toll (TLRs) 2 y 4 (TLR2 y TLR4) expresados en la membrana de macrófagos alveolares (Taraco, 2007).

Después de la fagocitosis del bacilo por macrófagos alveolares pulmonares, neutrófilos y DCs del pulmón, estas células producen péptidos antimicrobianos, citocinas y quimiocinas que activan la respuesta inmune del hospedero (Keane *et al.*, 2001). Dentro del macrófago ya infectado, los mediadores lipídicos, prostaglandina E2 (proapoptótico) o lipoxina LX (pronecrótica), juegan un papel importante en la determinación de la vía de muerte de dicha célula que conducen a la apoptosis o necrosis. Las células apoptóticas infectadas pueden ser fagocitadas por las células dendríticas pulmonares o esferocitadas (fagocitosis de células apoptóticas) por los macrófagos pulmonares no infectados. Las DCs infectadas por

M. tuberculosis migran a los ganglios linfáticos locales de los pulmones entre los 8 y los 12 días después de la infección bajo la influencia de IL-12(p40)2 e IL-12p70 y de las quimiocinas CCL19 y CCL2, para que las células T inmaduras se diferencien en Th1. Enseguida, estas células Th1 ya diferenciadas migran de regreso a los pulmones de una manera dependiente de quimiocinas entre los 14 y 17 días después de la infección inicial y producen IFN- γ , que conduce a la activación de macrófagos, además de la producción de citocinas, la inducción de factores microbicidas incluyendo iNOS (especies reactivas de nitrógeno inducible), y al control bacteriano (O'Garra *et al.*, 2013). Este mecanismo se muestra en la figura 1.

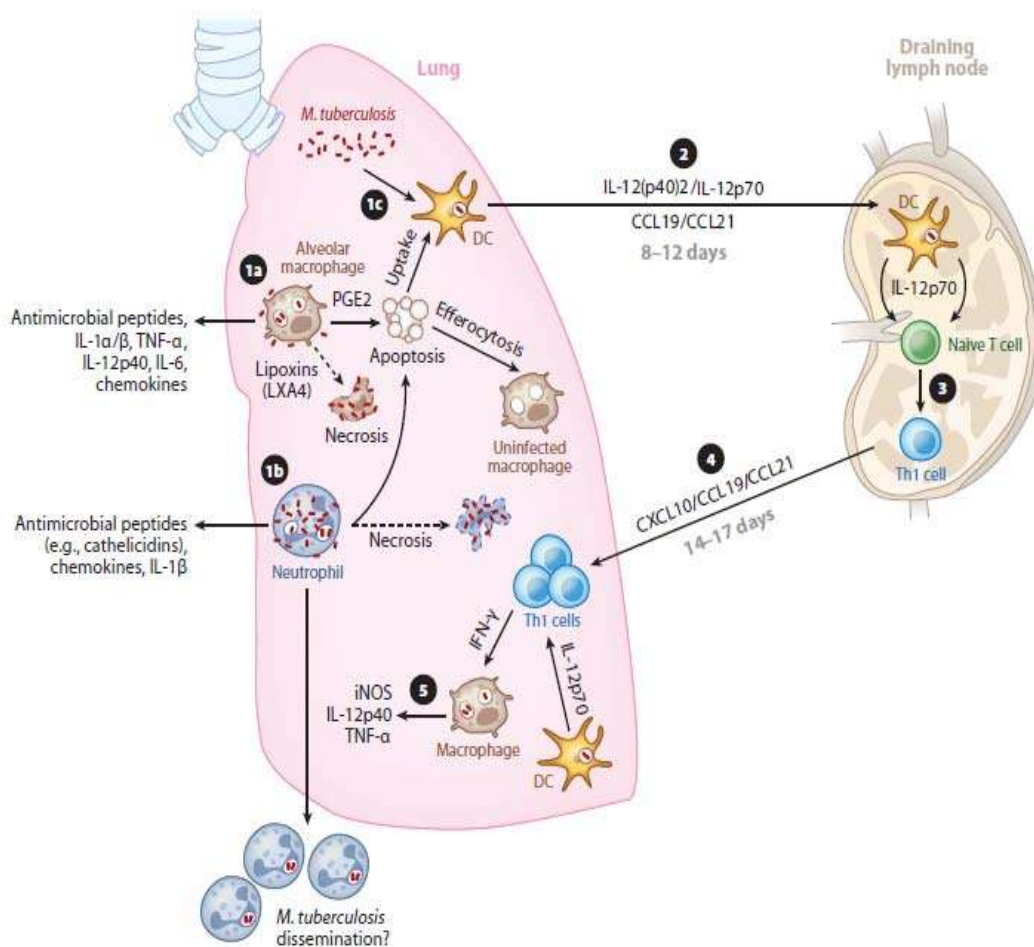


Figura 1. Mecanismo de invasión de *Mycobacterium tuberculosis*. 1a, b, c) El bacilo es fagocitado por macrófagos, neutrófilos y células dendríticas (DCs), 2) Las DCs infectadas viajan al ganglio linfático entre los 8 y 12 días después de la infección bajo la influencia de interleucinas (IL) y quimiocinas, 3) las células T inmaduras se diferencian en Th1, 4) Las células Th1 ya diferenciadas

migran de regreso a los pulmones de una manera dependiente de quimiocinas entre los 14 y 17 días después de la infección inicial, 5) producen IFN- γ , que conduce a la activación de macrófagos, producción de citosinas, inducción de factores microbicidas incluyendo iNOS (especies reactivas de nitrógeno inducible)(O'Garra *et al.*, 2013).

Sin embargo, *M. tuberculosis* es capaz de evitar la fusión del fagosoma y el lisosoma, así como de evadir la acción del fagolisosoma. Además, cuando el bacilo de tuberculosis es transportado a los ganglios linfáticos, éste puede diseminarse para migrar del sitio primario de infección hacia el sistema linfático y el torrente sanguíneo. La diseminación hematológica es un paso clave en el establecimiento de la enfermedad extrapulmonar y la colonización de lóbulos pulmonares previamente no infectados.

Las bacterias pueden cruzar la barrera epitelial alveolar por invasión directa y lisis de células epiteliales o viajando dentro de fagocitos profesionales (macrófagos). Además, los factores bacterianos como la adhesina hemaglutinina que se une a heparina (HBHA) y los productos de la vía de secreción de ESX-1 (ESAT-6 / CFP-10) ayudan en la translocación de bacterias a través del epitelio alveolar. Estudios sugieren que ESAT-6 es un factor citotóxico que debido a la lisis celular permite a las bacterias invadir y diseminarse dentro de las células que recubren los alveolos. La HBHA permite que las micobacterias se unan a glicoconjugados sulfatados en células epiteliales (Locht *et al.*, 2006). Otra forma de diseminación se presenta cuando el bacilo es capturado por los macrófagos de los alveolos, a los que se les denomina “macrófagos humanos infectados con *M. tuberculosis*”. Estos invaden la capa del epitelio subyacente y secretan varias citocinas y quimiocinas que permiten el reclutamiento y la activación de células inflamatorias (Russell *et al.*, 2010). Esta respuesta del hospedero a la bacteria da como resultado la formación del granuloma, el cual es una estructura que consiste en el recubrimiento de células infectadas por células epiteliales y, células gigantes multinucleadas rodeadas por linfocitos T activados (Polena *et al.*, 2016).

En la etapa inicial, el granuloma está altamente vascularizado. A medida que madura, los vasos sanguíneos se vuelven menos abundantes, aparece un núcleo central necrótico como resultado de la muerte de los macrófagos infectados. Sin

embargo, los granulomas se observan tanto en formas activas como latentes de TB. Por lo tanto, la formación de un granuloma no nos indica una respuesta inmune efectiva. Sin embargo, aunque las bacterias no se eliminan, generalmente se considera que los granulomas son estructuras protectoras del hospedero, que contienen la infección primaria (Seiscento *et al.*, 2010). Este dogma ha sido cuestionado en los últimos años con estudios en embriones de pez cebra infectados con *M. marinum*, lo que demuestra que la diseminación de las micobacterias se ve facilitado durante la formación temprana de granulomas (Volkman *et al.*, 2004).

En un estudio reciente se analizó el proceso de diseminación de tuberculosis mediante la investigación de macrófagos infectados de micobacterias y se demostró que el factor angiogénico VEGF, (que aumenta la formación de vasos sanguíneos y la propagación bacteriana), es secretado por macrófagos infectados con *M. tuberculosis* de una manera dependiente de RD1 (Region of difference 1 por sus siglas en inglés) (Polena *et al.*, 2016). De acuerdo con los resultados *in vitro* e *in vivo*, se observó una mayor angiogénesis en pacientes con TB. La concentración de VEGF en suero de pacientes con TB fue elevada, lo que confirma estudios previos (Matsuyama *et al.*, 2000); el nivel de células progenitoras endoteliales circulantes en sangre también se incrementó en comparación con los controles sanos. En general, este trabajo apoya la idea de que las micobacterias explotan a los macrófagos infectados con *M. tuberculosis* para la diseminación al inducir la formación de nuevos vasos sanguíneos (Polena *et al.*, 2016).

Como se mencionó anteriormente, la respuesta inmune innata es la primera línea de defensa del hospedero contra la infección. Este sistema inmunitario reconoce estructuras de los microorganismos que son esenciales para su supervivencia e infectividad, además de moléculas que liberan las células dañadas o necrosadas. Un tipo de receptores que reconocen dichas estructuras de los microorganismos son los receptores TLRs. Una vez que los TLRs se unen a sus ligandos se generan señales que activan factores de transcripción, los cuales estimulan la expresión de genes que codifican citocinas, enzimas y proteínas implicadas en las funciones antimicrobianas (Abbas, 2014).

A diferencia de la respuesta inmune innata, la respuesta inmune adaptativa se desarrolla de forma más lenta y proporciona una defensa más especializada y eficaz contra las infecciones. Esta se divide en dos tipos, inmunidad humoral y celular. La inmunidad humoral está mediada por los linfocitos B que producen anticuerpos. Estos anticuerpos se secretan en la circulación y los líquidos de las mucosas, se encargan de neutralizar y eliminar microbios y toxinas microbianas presentes fuera de las células del hospedero. Sin embargo, los anticuerpos no pueden acceder a los microbios que viven y se dividen dentro de las células infectadas (Abbas, 2014).

La inmunidad celular se encarga de eliminar a los microorganismos intracelulares mediante los linfocitos T. Para su activación requiere que las células presentadoras de antígeno muestren un antígeno peptídico unido a moléculas del CMH de clase I y II (Complejo Mayor de Histocompatibilidad) a los linfocitos T CD4+ o CD8+ vírgenes, los cuales reconocen el antígeno peptídico mediante sus receptores TCR. Los péptidos reconocidos por moléculas del CMH de clase II se unen a los linfocitos T CD4+, que a su vez secretan citocinas que activan a los macrófagos para inducir la destrucción del antígeno. Los péptidos reconocidos por moléculas del CMH de clase I se unen a los linfocitos T CD8+, y estos secretan citocinas las cuales activan la muerte de la célula diana que presenta el antígeno (Abbas, 2014).

1.6 Genes de virulencia

La habilidad de un patógeno bacteriano para sobrevivir dentro del hospedero requiere de la expresión de una serie de determinantes genéticos involucrados en la interacción patógeno-hospedero, situación que le permite resistir el estrés fisiológico y ambiental (Maulén, 2011). *Mycobacterium tuberculosis* ha demostrado tener una gran capacidad para sobrevivir en diversas condiciones como la tensión bactericida dentro del macrófago, el ambiente anaeróbico y alterado nutricionalmente del granuloma, así como también el estado latente en el cual el bacilo se somete a condiciones limitantes de nutrientes, oxígeno y moléculas

microbicidas. La comprensión de este comportamiento adaptativo radica en la identificación de genes como determinantes de virulencia, los cuales se expresan en estas condiciones variadas.

La pared celular de las micobacterias está constituida por un complejo conjunto de glicolípidos propios y ácidos micólicos, los cuales son los responsables de la forma y rigidez de la bacteria (Ramírez *et al.*, 2005). Además, dicha estructura celular ha sido asociada a la presencia de determinantes de virulencia como las proteínas tuberculina vieja (OT), Derivado Proteico Purificado (PPD), 65 kDa, lípidos como manósidos de fosfatidil inositol (PIM), antígeno 60 (Ag 60) y complejo proteico-lipopolisacárido procedente del citoplasma y de la membrana celular, común en *M. tuberculosis*, *M. bovis* y otras micobacterias (Ramírez *et al.*, 2005). Las bacterias son capaces de responder a un entorno complejo, mediante la integración de una gran variedad de señales para detectar su ubicación y expresar genes apropiados (Mehrotra, 2001). En el caso de *M. tuberculosis* se ha observado la expresión de genes de virulencia asociados al ambiente del hospedero, como *acr1* cuando la bacteria se encuentra expuesta a hipoxia u óxido nítrico en estado de latencia y durante el curso de la infección en macrófagos, o el gen *acr2* que se activa por un golpe de calor y estrés oxidativo, principalmente. De manera similar, *ahpC* es un regulador central del ambiente peroxidativo y respuesta al estrés nitrosante en *M. tuberculosis* (Forrellad *et al.*, 2013). Sin embargo, existen otros genes expresados en dicha interacción cuya función no es conocida. Por ejemplo, no se conoce la función exacta del gen *Rv2136c*, pero por su similitud con una enzima de *Escherichia coli* involucrada en la síntesis de peptidoglicano sugiere una función equivalente en *M. tuberculosis* (Mehrotra, 2001). Este grupo de genes relacionados con virulencia, pero de función desconocida, pertenece la familia *pe/ppe*.

1.7 Familia de genes *pe/ppe*.

Los genes *pe* (prolina- glutámico) / *ppe* (prolina-prolina-glutámico) representan uno de los aspectos más intrigantes del genoma de *Mycobacterium tuberculosis*. Estos genes son especialmente abundantes en micobacterias patógenas, con más de 160

miembros en el genoma de *M. tuberculosis*. Fueron descubiertos hace más de 15 años y su función sigue siendo incierta (Fishbein *et al.*, 2015). Esta familia de genes comprende el 10% del genoma de *M. tuberculosis* identificado en la cepa de estudio H37Rv. Además, codifican para dos familias de proteínas PE/PPE, PE con 100 genes y PPE con 70 genes, las cuales se caracterizan por tener dominios N-terminal conservados de aproximadamente 110 y 180 aminoácidos; las proteínas PE/PPE podrían presentarse como la principal fuente de variación antigénica por medio del polimorfismo entre cepas (Mukhopadhyay y Balaji, 2011). Las proteínas PE/PPE están presentes tanto en micobacterias patógenas como las saprófitas y probablemente estén más asociadas al metabolismo micobacteriano de lo que se creía. En general, las micobacterias no patógenas tienden a tener menos genes *pe/ppe* (Fishbein *et al.*, 2015).

El nombre PE deriva del motivo distintivo prolina (P) – glutámico (E) y el PPE motivo prolina (P)- prolina (P)- glutámico (E), los cuales se encuentran presentes dentro de los 10 primeros aminoácidos de esta proteína (Fig. 2) (Mukhopadhyay y Balaji, 2011). La familia PE puede dividirse en PE_PGRS que incluye secuencias polimórficas ricas en GC y PE que no presentan características distintivas. La familia PPE se divide en PPE_MPTR (repeticiones tándem polimórficas mayores), PPE_SVP (con un motivo GxxSVPxxW), PPE_PPW (con un motivo PxxPxxW) y genes *ppe* sin características distintivas. Esta variedad de subclases proporciona la primer indicación que los miembros de las familias pueden jugar distintos roles funcionales (Fishbein *et al.*, 2015).

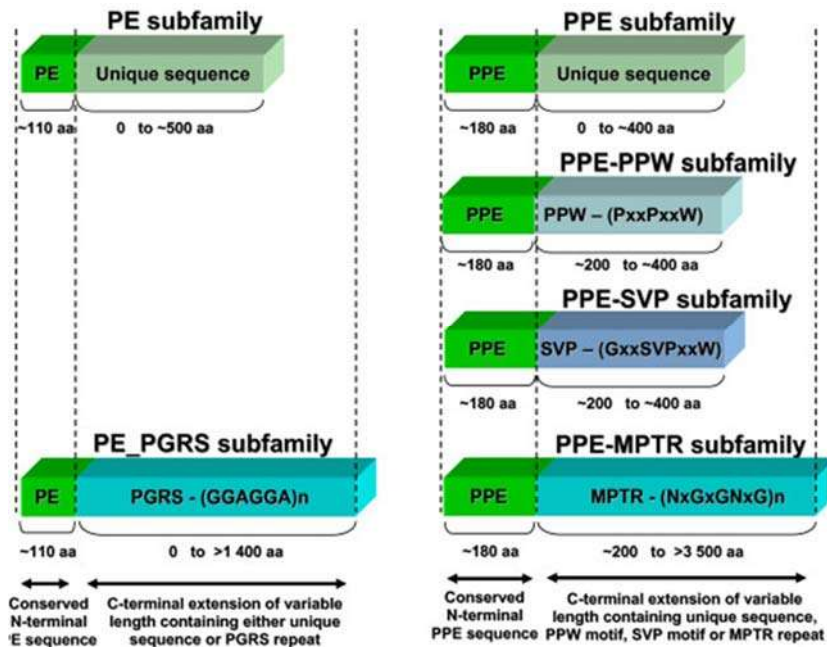


Figura 2. Estructura de los genes *pe*/*ppe*. Representación de la estructura de los miembros de las familias de genes *pe*/*ppe*, se muestra el dominio N-terminal conservado (verde), las posiciones de los motivos y las diferencias entre las subfamilias encontradas en ambos genes (gris y azul). (van Pittius *et al.*, 2006).

El impacto de la variación genética sobre la estructura y consecuentemente la función de las proteínas PE/PPE sigue siendo en gran parte desconocido. La primera estructura para un par de proteínas PE/PPE (PE25_{III}/ PPE41_{III}) se determinó en 2006 (Strong *et al.*, 2006). La estructura cristalina de un dominio enzimático putativo de PE_PGRS16v fue reportada en 2013 (Barathy and Suguna 2013), y dos estructuras completas de PE25_{III}/ PPE41_{III} (en complejos con EspG) se publicaron en 2014 (Ekiert and Cox 2014; Korotkova *et al.*, 2014). Los datos limitados estructurales confirmados experimentalmente dificultan la capacidad de predecir las consecuencias funcionales de la variabilidad PE/PPE (Fishbein *et al.*, 2015), además de que no se cuenta con la función precisa de los dominios de la proteína.

1.7.1 Localización y función de las proteínas PE/PPE

Existe evidencia experimental que apoya la asociación de PE/PPE con la micomembrana, o la exposición superficial huésped-patógeno (Sampson, 2001). Se han encontrado más de 35 proteínas PE/PPE en o sobre la membrana micobacteriana y/o la pared celular. Aunque las funciones PE/PPE pueden haber

variado por cambios en la secuencia de los genes que las codifican, su función específica depende con frecuencia de la localización en la pared celular o extracelular (proteínas secretadas). Recientemente se reportó que el dominio PE N-terminal de PE_PGRS33V es necesario para la localización de la proteína en la pared celular en *M. marinum* y *M. tuberculosis* (Cascioferro *et al.*, 2007, 2011; Zumbo *et al.*, 2013). Por otro lado, se encontró que la secuencia N-terminal de las proteínas Lip_{IV} (PE_PGRS63V) y PE_PGRS30V, permite la secreción y que el C-terminal tiene la actividad de lipasa (Daleke *et al.*, 2011; Iantomasi *et al.*, 2012). Trabajos recientes sobre PPE17_{IV} sugieren que los dominios de PE y PPE pueden cumplir funciones similares de orientación en la pared celular (Donà *et al.*, 2013). Además, se ha observado una asociación de varias proteínas PE/PPE con la pared celular micobacteriana, lo que ha llevado a sugerir que algunas proteínas PE/PPE están implicadas en la integridad estructural y en la morfología de colonias de *M. tuberculosis* (Chatrath *et al.*, 2011).

Por su exposición en la superficie, como la micomembrana, o a la secreción en el medio extracelular, se implica a las proteínas PE/PPE en la interacción directa con el hospedero, y existe evidencia experimental en ese sentido (Sampson, 2001). Se han encontrado más de 35 proteínas PE/PPE en o sobre la membrana micobacteriana y/o la pared celular. Aunque las funciones PE/PPE pueden haber variado por cambios en la secuencia de los genes que las codifican, su función específica depende con frecuencia de la localización en la pared celular o extracelular (proteínas secretadas). Se ha propuesto que las proteínas PE/PPE pueden contribuir en la patogénesis de *M. tuberculosis* influyendo en la interacción con el receptor Toll-like 2 (TLR2) de los macrófagos del hospedero. Estos receptores TLR son proteínas transmembranales caracterizados por un dominio extracelular rico en leucina y una cola citoplasmática. Reconocen diversos patrones moleculares asociados a patógenos (PAMPs); por ejemplo, los TLR2 reconocen lipoglucanos bacterianos, mientras que los TLR4 son específicos frente a los lipopolisacáridos (LPS) bacterianos (endotoxina) (Abbas, 2014).

Se cuenta con evidencia de que miembros de la familia PE/PPE pueden interactuar con las moléculas del hospedero, como el receptor TLR2 (Bansal *et al.*, 2010; Basu *et al.*, 2007; Nair *et al.*, 2009; Tiwari *et al.*, 2012; Zumbo *et al.*, 2013). Además, se demostró que la maduración y la activación de las células dendríticas por PE_PGRS11 y PE_PGRS17 dependen de la señalización de TLR2 (Bansal *et al.*, 2010). Sin embargo, aún no se ha establecido si la interacción con TLR2 es una propiedad exclusiva de las proteínas PE/PPE. Por el contrario, también se ha demostrado un papel importante de TLR2 y TLR4 en la inducción de la apoptosis de macrófagos sobre la infección por *M. tuberculosis* (Sánchez *et al.*, 2010). Adicionalmente, se demostró la interacción del receptor TLR4 de los macrófagos del hospedero con el dominio PGRS de la proteína PE_PGRS de *M. tuberculosis* Rv0297, el cual está implicado en la apoptosis mediada por el estrés del retículo endoplásmico (Grover *et al.*, 2018).

Como se comentó anteriormente, antes de que se presente la respuesta inmune adaptativa, la infección por *M. tuberculosis* se establece mediante la entrada a los macrófagos del hospedero (Thi *et al.*, 2012). Varias proteínas de PE/PPE se han ligado a la inhibición y/o activación del macrófago. Se sugiere que el sistema de secreción ESX-5 y las proteínas PE/PPE son responsables de la virulencia de *M. tuberculosis* y contribuyen a su capacidad para crecer en un macrófago (Bottai *et al.*, 2012). Por ejemplo, PE_PGRS30V se ha asociado con la virulencia micobacteriana mediante la inhibición de la fusión fagolisosomal (Lantomasi *et al.*, 2012). Además, PE_PGRS62, PE5, PPE15 y PPE2 inhiben la producción de especies reactivas de oxígeno; mientras que PE5, PE15 y PE4 ayudan a incrementar la supervivencia de la micobacteria dentro del macrófago (Tiwari *et al.*, 2012). Otro estudio *in vivo* reveló la notable abundancia relativa de proteínas PE/PPE detectadas durante la infección en cobayos, lo que subraya su posible papel en la virulencia (Kruh *et al.*, 2010); además, un 60% de las proteínas PE_PGRS y PPE_MPTR se expresan cuando el patógeno se encuentra en pulmones de cobayo a los 90 días después de la infección.

Basu y colaboradores (2007) demostraron que PE_PGRS33 es capaz de mejorar la expresión del factor de necrosis tumoral alfa (TNF- α) de una manera dependiente de TLR2 que conduce a la apoptosis del macrófago. Sin embargo, se observó que las eliminaciones dentro del dominio PGRS (en aislados clínicos) inhibieron esta capacidad. Existen estudios que sugieren que las proteínas PE_PGRS pueden ser capaces de inhibir el procesamiento y/o presentación de antígenos, además se ha propuesto que las repeticiones Gly-Ala en los dominios PGRS C-terminal son capaces de inhibir la degradación proteosomal del dominio N-terminal PE, lo que inhibe el procesamiento del antígeno por las células T CD8+ (McEvoy *et al.*, 2012).

1.7.1. a Regiones estructurales y no estructurales de las proteínas PE/PPE

La estructura terciaria o tridimensional de las proteínas puede estar constituida por regiones estructurales, es decir regiones con estructuras definidas o dominios (son combinaciones concretas de hélices α y láminas β y giros β los cuales forman estructuras compactas y estables con una función específica) (Karp, 2014). Además, dentro de las estructuras terciarias de las proteínas se pueden encontrar regiones desordenadas, loops, bucles o coils. Anteriormente se les conocía como conectores entre estructuras definidas o dominios funcionales (Cuevas-Velázquez y Covarrubias-Robles, 2011).

Sin embargo, en los últimos años se ha descubierto y caracterizado a las proteínas conocidas como proteínas no estructuradas o intrínsecamente desordenadas (IDPs). Estas proteínas poseen una elevada flexibilidad estructural la cual les permite adoptar estructuras diferentes y, por tanto, reconocer ligandos diversos conservando la especificidad en el reconocimiento de los mismos (Cuevas-Velázquez y Covarrubias-Robles, 2011).

El proteoma de *M. tuberculosis* tiene un alto contenido de IDPs (Malen *et al.*, 2007, Ahmad *et al.*, 2019) principalmente debido a las proteínas de la familia PE/PPE/PGRS, que a menudo contienen largas regiones de desorden estructural (Meszaron

et al., 2011, Ahmad *et al.*, 2019). En la figura 3 se observan las regiones estructurales de la subfamilia PE_PGRS y la región no estructural.



Figura 3. Regiones estructurales y no estructurales de la subfamilia PE_PGRS. Modificado de Meena LS, 2014.

1.7.2 Polimorfismos en las familias de proteínas PE/PPE.

En los microorganismos patógenos en general se encontró que las proteínas que se encuentran directamente expuestas a la vigilancia inmune muestran mayores niveles de polimorfismos (McEvoy *et al.*, 2012). Como se ha mencionado anteriormente, muchas proteínas PE/PPE se encuentran situadas en la superficie celular y otras son probablemente secretadas. Esto en conjunto con su inmunogenicidad y la naturaleza polimórfica bien establecida de sus repeticiones C-terminal, ha llevado a la sugerencia de que dichas proteínas pueden estar implicadas en la variación antigénica y la evasión inmune. Estudios posteriores han reportado un gran número de polimorfismos en los genes *pe_pgrs33*, *pe_pgrs16* y *pe_pgrs26* (McEvoy *et al.*, 2012), además algunos genes PE/PPE han mostrado reportes de niveles elevados de la integración de IS6110 y recombinación homóloga.

En un estudio epidemiológico realizado en 19 cepas clínicas de *M. tuberculosis* se detectaron hasta 25 regiones cromosómicas eliminadas, y se observó una correlación inversa entre el porcentaje del genoma con eliminaciones y el porcentaje de pacientes infectados que presentaban cavitación pulmonar. Los

autores de este estudio han interpretado estos datos considerando que la acumulación de eliminaciones en *M. tuberculosis* conduce a una disminución de la virulencia. Sin embargo, parece ser que estas eliminaciones ocurren como consecuencia de la recombinación homóloga entre copias de IS6110. En otro trabajo se observó que las eliminaciones identificadas no eran esenciales para causar la enfermedad, sin embargo su espectro de frecuencia sugiere que la mayoría no tienen efectos deletéreos para el patógeno (Tsolaki *et al.*, 2004).

Otro estudio relacionado con la seguridad y la potencia vacunal de algunas mutantes de *M. tuberculosis* para su evaluación como candidatos vacunales contra la tuberculosis demostró que la eliminación de varios genes involucrados en la patogenicidad y virulencia confirió a los mutantes la capacidad de generar respuestas inmunes protectoras frente al reto con *M. tuberculosis* en modelos animales (Borrero *et al.*, 2011). Como se mencionó anteriormente, los genes *pe_pgrs* se caracterizan por su alto contenido de GC y extensas secuencias repetitivas, haciéndolos propensos a eventos de recombinación y la variabilidad genética. Se realizó un análisis comparativo de secuencias de los genes *pe_pgrs17* (Rv0978c) y *pe_pgrs18* (Rv0980c) de *M. tuberculosis* el cual reveló una variación genética sorprendente en comparación con la cepa de referencia H37Rv. La variación, denominada polimorfismo 12/40, consiste en una inserción de 12 pb acompañada por un conjunto de 40 polimorfismos de nucleótidos únicos (SNPs) que se producen en uno o en ambos genes (Karboul *et al.*, 2006). Estos estudios proporcionaron evidencia directa y clara de la ocurrencia natural de la conversión génica en micobacterias, que parece estar restringida a cepas modernas de *M. tuberculosis*; además, proporciona una nueva perspectiva a explorar los eventos moleculares que acompañaron la evolución, expansión clonal y reciente diversificación del bacilo tuberculoso (Karboul *et al.*, 2006).

La generación de polimorfismos entre distintas cepas puede contribuir a la variabilidad antigénica, pero la presencia de un gran número de proteínas PE_PGRS dentro de un solo genoma podría potencialmente conducir a una variación antigénica debido a su expresión durante procesos de infección o estrés.

La expresión de muchos de los genes PE_PGRS está alterada en *M. tuberculosis*, portando mutaciones en los genes *sigD*, *sigL*, *sigM*, *sigB*, *sigE* y *sigF*, lo cual sugiere un papel de estos factores en el control de la expresión de dichos genes durante la patogénesis (Mukhopadhyay y Balaji, 2011).

En conclusión, la relación entre polimorfismos de los genes *pe/ppe* y otras familias de *M. tuberculosis* se encuentra poco explorada.

1.7.3 Regulación de la transcripción de genes *pe/ppe*.

La regulación de la familia de genes *pe/ppe* está controlada por un grupo heterogéneo de reguladores de la transcripción (Voskuil, 2004), entre los que se encuentran los factores sigma. Los factores sigma micobacterianos facilitan la adaptación fisiológica a estímulos ambientales y virulencia en las micobacterias patógenas (Sachdeva *et al.*, 2010). En particular, se ha documentado que el factor SigF regula todos los sublinajes de la familia de genes *pe* (Fishbein *et al.*, 2015). SigF se presenta en micobacterias patógenas y saprófitas, y se le ha relacionado con la adaptación bacteriana al estrés ambiental, incluyendo variaciones de pH, estrés oxidativo y choque térmico (Williams *et al.*, 2007). Además de ser un factor relevante para la supervivencia de las micobacterias ambientales, la capacidad adaptativa proporcionada por SigF también resulta importante para la supervivencia de las micobacterias patógenas dentro del macrófago (Russell, 2011). Otro factor sigma estudiado es SigB, un factor ubicuo que además de participar en el control de la respuesta al estrés de la pared celular micobacteriana (Manganelli *et al.*, 1999), regula la expresión de diversos genes *pe_pgrs* y algunos genes *ppe* posiblemente implicados en la interacción huésped-bacteria mediada por la pared celular (Lee *et al.*, 2008; Fontán *et al.*, 2009). SigD es otro factor sigma que también regula genes *pe_pgrs*, aunque mayoritariamente ejerce su efecto regulatorio sobre genes *pe*, con la excepción de *ppe19IV*. Este regulador responde al medio ambiente pulmonar humano, la modulación de la expresión en respuesta a la inanición y la hipoxia, y suprime la expresión de múltiples proteínas PE_PGRS (Raman *et al.*, 2004; Calamita *et al.*, 2005). Tanto SigB como SigD están relacionados con el control de

la expresión del más reciente y evolucionado sublinaje de genes *pe/ppe*, encontrados solo en micobacterias patógenas y asociados al control de la expresión de genes esenciales para la respuesta a estrés y patogenicidad de la bacteria *in vivo* (Calamita *et al.*, 2005; Lee *et al.*, 2008). Esto indica que las proteínas PE/PPE bajo su control también pueden jugar un rol en la virulencia bacteriana (Fishbein *et al.*, 2015).

Otros reguladores identificados son los sistemas reguladores de dos componentes (TCSs, por sus siglas en inglés), los cuales son sensores de células bacterianas que en *Mycobacterium tuberculosis* contribuyen a la adaptación de la bacteria a su estilo de vida intracelular (Bretl *et al.*, 2011). Varios TCSs afectan la expresión de los genes *pe/ppe*. El TCS más influyente en la expresión génica *pe/ppe* es *phoPR*, controlando la expresión de por lo menos 13 genes *pe/ppe* diferentes (Walters *et al.*, 2006).

1.7.4 Antígeno Leucocitario Humano (HLA)

Los alelos del MHC son proteínas de membrana situadas en las células presentadoras de antígeno, las cuales tienen la función de mostrar los antígenos peptídicos para el reconocimiento por las células T. Las células que capturan antígenos microbianos y los muestran para que los reconozcan los linfocitos T se llaman células presentadoras de antígenos (CPA) como las células dendríticas, macrófagos y linfocitos B (Abbas, 2014). Las células dendríticas son las CPAs más potentes para activar los linfocitos T vírgenes. El macrófago, que abunda en todos los tejidos y en las reacciones inmunitarias celulares, fagocita microorganismos y muestra los antígenos de estos patógenos a los linfocitos T efectores (CD4+), que activan a los macrófagos para que maten a los microorganismos. Los linfocitos B ingieren antígenos proteícos y los muestran a los linfocitos T efectores (CD4+) dentro de los tejidos linfáticos (Abbas, 2014).

Los genes del MHC se ubican en el brazo corto del cromosoma 6 y ocupan un segmento de 3.5 millones de pares de bases (Vielma, 2010). El grupo de genes que componen este *locus* del MHC se encuentran en todos los mamíferos. A las

proteínas del MHC humano se les denomina “antígenos leucocitarios humanos” (HLA, por sus siglas en inglés), porque se descubrieron como antígenos de los leucocitos que podrían ser identificados con anticuerpos específicos (Abbas, 2014). El sistema HLA se divide en regiones de clase I, II y III. La región de genes de clase I se ubica en el *locus* 11 (Vielma, 2010), que es la región más telomérica, y los genes de esta clase se ubican en el *locus* A, B y C que codifican para los alelos clásicos HLA A, B y C, así como los genes que codifican para los alelos no clásicos HLA E, F, G, H, Y y J. La región de genes de clase II se ubica en el *locus* 5 (Vielma, 2010), estos genes se localizan más cerca del centrómero y se ubican en tres principales loci (DP, DQ Y DR), que codifican para los alelos HLA DP, DR y DQ. En esta región se encuentran también los genes que codifican para los alelos no clásicos (TAP, LMP y DM). La región de genes de la clase III se ubica entre las regiones genéticas de clase I y clase II, donde se encuentra un grupo heterogéneo de genes que codifica para varias proteínas secretadas con funciones inmunes o inflamatorias, como aquellas que son componentes del complemento (C2, C4 y el factor B), las proteínas relacionadas con la inflamación (citocinas como TNF- α , LTA, LTB) o las de choque térmico (Trujillo *et al.*, 2017).

Cada una de los alelos de la clase I y II del HLA presenta una hendidura de unión al péptido en el extremo amino terminal. Los alelos de clase I constan de una cadena α asociada de forma no covalente a una proteína llamada β 2-microglobulina, que es codificada por un gen situado fuera del HLA. En los dominios aminoterminales α 1 y α 2 se encuentra el surco o hendidura de unión al péptido, la cual posee aminoácidos altamente polimórficos, y es lo suficientemente grande para acomodar péptidos entre 8 y 11 aminoácidos. El dominio α 3 es invariable y contiene la zona de unión para el correceptor CD8 del linfocito T (Abbas, 2014).

Los alelos de clase II del HLA consta de dos cadenas α y β . Las regiones aminoterminales de ambas, llamadas dominios α 1 y β 1, contienen aminoácidos polimórficos y forman una hendidura que es lo suficientemente grande para acomodar péptidos de entre 10 y 30 aminoácidos. El dominio β 2 no polimórfico contiene la zona de unión para el correceptor CD4 del linfocito T (Abbas, 2014).

Las proteínas extracelulares que son interiorizadas por las CPAs especializadas, como células dendríticas, macrófagos y linfocitos B, son procesadas en vesículas endocíticas, mostradas a las moléculas de la clase II del HLA y reconocidas por los linfocitos T CD4+ cooperadores. Por su parte, las proteínas del citosol de cualquier célula nucleada son procesadas en orgánulos citoplásmicos, mostradas en moléculas de la clase I del HLA y reconocidas por los linfocitos T CD8+ citotóxicos (Abbas, 2014).

Las regiones de las proteínas del patógeno que interactúan directamente con el HLA son secuencias cortas denominadas epítopes, siendo definidos como péptidos de las proteínas bacterianas que interactúan de manera específica con las células hospederas, generando la respuesta celular o humoral. Los epítopes T, que tienen una longitud generalmente ente 9 a 15 aminoácidos, generan una respuesta celular al activar los linfocitos T. Esta respuesta es generada por el complejo formado entre el epítope, el HLA y el receptor de linfocitos T (TCR) (Janeway y Golstein, 2001). Los epítopes B, que generalmente tiene una longitud de 20 aminoácidos, son los encargados de generar la respuesta humoral inmune, es decir la producción de anticuerpos. La identificación de estos epítopes es importante en el desarrollo de vacunas y pruebas inmunodiagnósticas (Manzalawy y Hinavar, 2010).

1.7.5 Unión de las proteínas PE/PPE con el Complejo Mayor de Histocompatibilidad (MHC).

Algunas proteínas PE y PPE de *M. tuberculosis* han demostrado ser potentes antígenos que son reconocidos por células B y T. Las proteínas RRv1759c y Rv3367 de la subfamilia PE_PGRS son expresadas durante la infección por *M. tuberculosis* y muestran respuesta a anticuerpos en humanos y conejos (Abou-Zeid *et al.*, 1991; Singh *et al.*, 2001; Chaitra *et al.*, 2005). Otro estudio ha demostrado que el dominio PE de la proteína PE_PGRS Rv1818c, tras la inmunización en ratones, induce una buena respuesta inmune celular, mientras que el dominio PGRS es responsable de inducir una buena respuesta inmune humoral (Delogu y Brennan, 2001). Hay

suficiente evidencia experimental que apoya la importancia de los linfocitos T CD4+ en el control de la infección por *M. tuberculosis*. Además, hay evidencia que sugiere un rol importante de los linfocitos T CD8+ en el control de la infección por *M. tuberculosis* en el hospedero. Sin embargo, pocos son los epítopes de antígenos micobacterianos que han sido identificados por células T CD8+ de humanos (Chaitra *et al.*, 2005).

Ya que la infección por *M. tuberculosis* puede inducir respuesta de células T CD4+ y CD8+ a diversas proteínas PE/PPE, algunas de estas proteínas han sido evaluadas como posibles antígenos para formar parte de una nueva vacuna contra la TB (Sampson, 2011; Sayes, 2012; Mortier *et al.*, 2015).

1.8 Herramientas bioinformáticas en el análisis de la interacción patógeno hospedero

La Bioinformática es el uso de herramientas computacionales que permite analizar, depurar y agilizar el manejo de grandes cantidades de datos de la biología principalmente en términos fisicoquímicos-matemáticos y permite comprender y organizar la información asociada. Parte, de datos encontrados experimentalmente, los cuales son almacenados y sobre estos se aplican técnicas de consulta y de análisis (Bustos *et al.*, 2011).

Se han realizados un gran número de experimentos, en el que principalmente parten de un análisis *in silico* el cual es rápido, de bajo costo y brinda las bases, para enseguida realizar el experimento *in vitro* y/o *in vivo*.

Uno de los primeros ejemplos fue la identificación de potenciales candidatos vacúnales del patógeno de humanos *Neisseria meningitidis* serogrupo B (MenB). Esta bacteria es una de las principales causas, en todo el mundo, de meningitis bacteriana. A pesar de los esfuerzos realizados, no se había podido desarrollar una vacuna efectiva contra MenB debido a la variación de la secuencia de las proteínas de superficie y la reactividad cruzada del polisacárido capsular de este serogrupo con los tejidos humanos (Ferreira *et al.*, 2008).

Pizza y colaboradores (2000) analizaron la secuencia del genoma de la cepa virulenta MC58. Utilizaron los servidores BLAST, FASTA, MOTIFS, FINDPATTERNS y PSORT, además de ProDom, Pfam y Blocks; identificaron 570 marcos de lectura abiertos que codificaban para antígenos expuestos en la superficie o secretados, de los cuales 350 fueron expresados exitosamente en *E. coli* (Pizza *et al.*, 2000). Las proteínas recombinantes fueron purificadas y utilizadas para inmunizar ratones y el antisuero obtenido para cada una fue entonces analizado por ELISA y FACS para evaluar la localización celular de los antígenos.

Fueron seleccionadas siete proteínas representativas para estudios posteriores, las cuales se denominaron antígenos de Neisseria derivados del genoma (GNA por sus siglas en inglés *genome Neisseria antigen*). Cada una de estas proteínas presentó una respuesta inmune que inducía la actividad bactericida mediada por el complemento (Pizza *et al.*, 2000). Por tanto se demostró que los antígenos identificados por análisis *in silico* son buenos candidatos para el desarrollo clínico de una vacuna contra MenB (Ferreira *et al.*, 2008).

Otro ejemplo, es el estudio realizado por Grover y colaboradores (2018) en el cual descubren la importancia del dominio PGRS de la cepa Rv0297 de *Mycobacterium tuberculosis* partiendo de un análisis *in silico*. En la última década, ha habido un creciente interés en determinar el papel de las proteínas PE_PGRS en la fisiopatología de la TB debido a su presencia limitada en micobacterias no patógenas. Es por ello que fue de interés para los investigadores analizar si el dominio PGRS de la Rv0297 es capaz de interactuar con el TLR 2 y TLR 4 de los macrófagos del hospedero, ya que se sabe que pocas de las proteínas PE_PGRS de *M. tuberculosis* interactúan con los TLR, principalmente el tipo TLR2. Por lo tanto, llevaron a cabo *in silico* simulaciones para predecir la identidad del TLR involucrado en la interacción (Grover *et al.*, 2018).

Para ello, la secuencia de la proteína Rv0297 se obtuvo de la base de datos UniProtKB (identificador [ID] Q6MX50) y la estructura 3D se modeló usando I-TASSER. La estructura de la proteína TLR4 humana (27 a 627 aminoácidos) se obtuvo del banco de datos de proteínas RCSB (PDB ID 3FXI). Las interacciones

Rv0297 con TLR4 y TLR2 se compararon a través del servidor web HADDOCK 2.2. La validación de la estructura se realizó a través de PSVS y las simulaciones de dinámica molecular (MD) se realizaron con GROMACS v5.0.7. Este análisis reveló que el TLR4 tiene una mejor interacción con el PGRS Rv0297 que el TLR2 (Grover *et al.*, 2018).

Habiendo predicho la interacción del TLR4 con el PGRS de la Rv0297, realizaron un ensayo de interacción *in vitro*. Para este análisis utilizaron células de macrófagos murinos RAW 264.7 y RAW- Δ TLR4, estas células se incubaron con el PGRS de la proteína Rv0297 y se tiñeron con sueros anti- PGRSRv0297 para evaluar la interacción del TLR4 localizado en la superficie de los macrófagos con la proteína PGRS-Rv097. Las imágenes confocales revelaron la presencia del complejo que interactúa en la superficie de las células RAW 264.7 incubadas con la proteína PGRS-Rv097. A diferencia de las células RAW 264.7, las células RAW- Δ TLR4 no mostraban tales complejos en su superficie, incluso en presencia de la proteína PGR-Rv097. La presencia de esta interacción en la superficie de las células RAW 264.7 solamente, y no en las células deficientes en TLR4, indica que PGRS-Rv029 realmente interactúa con TLR4 de los macrófagos (Grover *et al.*, 2018).

2. ANTECEDENTES

En el laboratorio de Genética Molecular Microbiana se tiene una colección de 120 cepas de micobacterias, recolectadas a partir del año 2009. Se cuenta con los datos epidemiológicos de los pacientes incluyendo: la ubicación de la enfermedad, edad, género y jurisdicción de procedencia. De estos aislados, 19 pertenecen al estado de Querétaro. Además, se ha determinado la familia, linaje y genotipo por medio de las técnicas MIRU-VNTR (*Mycobacterial Interspersed Repetitive Units – Variable Number of Tandem Repeats*) y espoligotipificación (Guillén-Nepita, 2014; Vázquez-Heredia, 2017). Adicionalmente, se determinó la resistencia de estas cepas a uno o más de los antibióticos de primera línea usados en el tratamiento de TB: (Isoniacida (INH), Rifampicina (RIF), Etambutol (EMB), Estreptomina (SM) y Pirazinamida (PZA)) (Tabla 1). Así como a la presencia de mutaciones en genes asociados a dichas resistencias (Pérez -Reyes, 2014).

En un estudio previo realizado por Frutos-Hernández (2015) se encontraron polimorfismos en genes *pe* y *ppe* en cepas de *Mycobacterium tuberculosis*, aisladas de pacientes del estado de Michoacán, que se relacionan con características genotípicas y clínicas específicas. El gen *pe_pgrs18* resultó ser el más polimórfico ya que el 71.1% de las cepas analizadas presentaron algún polimorfismo en su secuencia. Además, se observó un patrón de asociación con base en las posiciones de los cambios de la secuencia del gen *pe_pgrs18* y las familias en las que se presentaban. Adicionalmente en un trabajo posterior se detectaron los polimorfismos en el gen *pe_pgrs18* de 59 cepas de *M. tuberculosis*, 40 del estado de Michoacán y 19 de Querétaro. De todas las cepas, 26 presentaron variaciones en las regiones conservadas del gen, de las cuales 16 son de Michoacán, con 77 SNPs, 1 inserción y 11 eliminaciones, y 10 son de Querétaro, con 224 SNPs, 2 inserciones y 14 eliminaciones. Las cepas más polimórficas fueron MYC 106 y MYC 107 de Querétaro y la cepa MYC 066 de Michoacán. El programa PROVEAN predijo que seis de las cepas poseen polimorfismos perjudiciales para la proteína, de las cuales 4 son de Michoacán y 2 de Querétaro. Se observó el polimorfismo 12/40 en 27 de las cepas de estudio, consistente en una inserción de 12 nucleótidos en la

posición 354 a 365 y 40 SNPs entre la posición 352 y 579. Cuatro de las cepas de Michoacán de las familias X, LAM y EAI presentaron asociación con un grupo de SNPs en las posiciones 83 y 300. No se encontró asociación entre la localización clínica de la enfermedad y el sexo de los pacientes con los polimorfismos en las cepas de Michoacán (Jiménez- Ruiz, 2017).

3. JUSTIFICACIÓN

El genoma de *Mycobacterium tuberculosis*, el organismo causal de la tuberculosis (TB), codifica para una familia de proteínas única conocida como la familia PE/PPE/PGRS, presente exclusivamente en el género *Mycobacterium* y en ningún otro género de organismos procariontes o eucariotes. Se ha reportado que la subfamilia PE_PGRS constituye aproximadamente el 10% del genoma de *M. tuberculosis*. Los polimorfismos en esta subfamilia de genes son una de las mayores fuentes de variación en el complejo *Mycobacterium tuberculosis*. Aunque la localización subcelular de estas proteínas sigue siendo desconocida, algunas de las proteínas PE_PGRS han sido consideradas como posibles factores de virulencia, y algunas son constituyentes de la superficie celular, por lo que están involucradas en la interacción de micobacterias y macrófagos.

Sin embargo, no existe suficiente evidencia epidemiológica y experimental que demuestre que la subfamilia del gen *pe_pgrs18* juega un papel importante en la virulencia e interacción de *M. tuberculosis* en los tejidos del hospedero, y su presencia limitada en micobacterias no patógenas genera un gran interés en determinar el papel de las proteínas PE_PGRS en la fisiopatología de la TB, es decir, es importante conocer el papel de las proteínas PE_PGRS presentes en el complejo *M. tuberculosis* para el desarrollo de la TB. Además, se ha demostrado que varios miembros de la familia PE_PGRS estimulan fuertes respuestas de células T.

Es por ello que resulta interesante buscar epítopes (péptidos) codificados por el gen *pe_pgrs18*, para seleccionar aquellos que se unan a moléculas del HLA de clase I y II con el objetivo de analizar si este gen, que ha mostrado ser muy polimórfico, es capaz de inducir una respuesta inmune celular que le brinde protección al hospedero generando linfocitos T de memoria, eliminando al microorganismo y por tanto eliminando la infección. Los epítopes predichos podrían entonces probarse experimentalmente para su inclusión en una vacuna potencial contra la tuberculosis y la identificación temprana de la enfermedad.

4. HIPÓTESIS

La proteína PE_PGRS18 de *Mycobacterium tuberculosis* posee epítopes que son reconocidos por proteínas del HLA y presenta polimorfismos que incrementan significativamente la probabilidad de interacción con el sistema inmune del hospedero.

5. OBJETIVOS

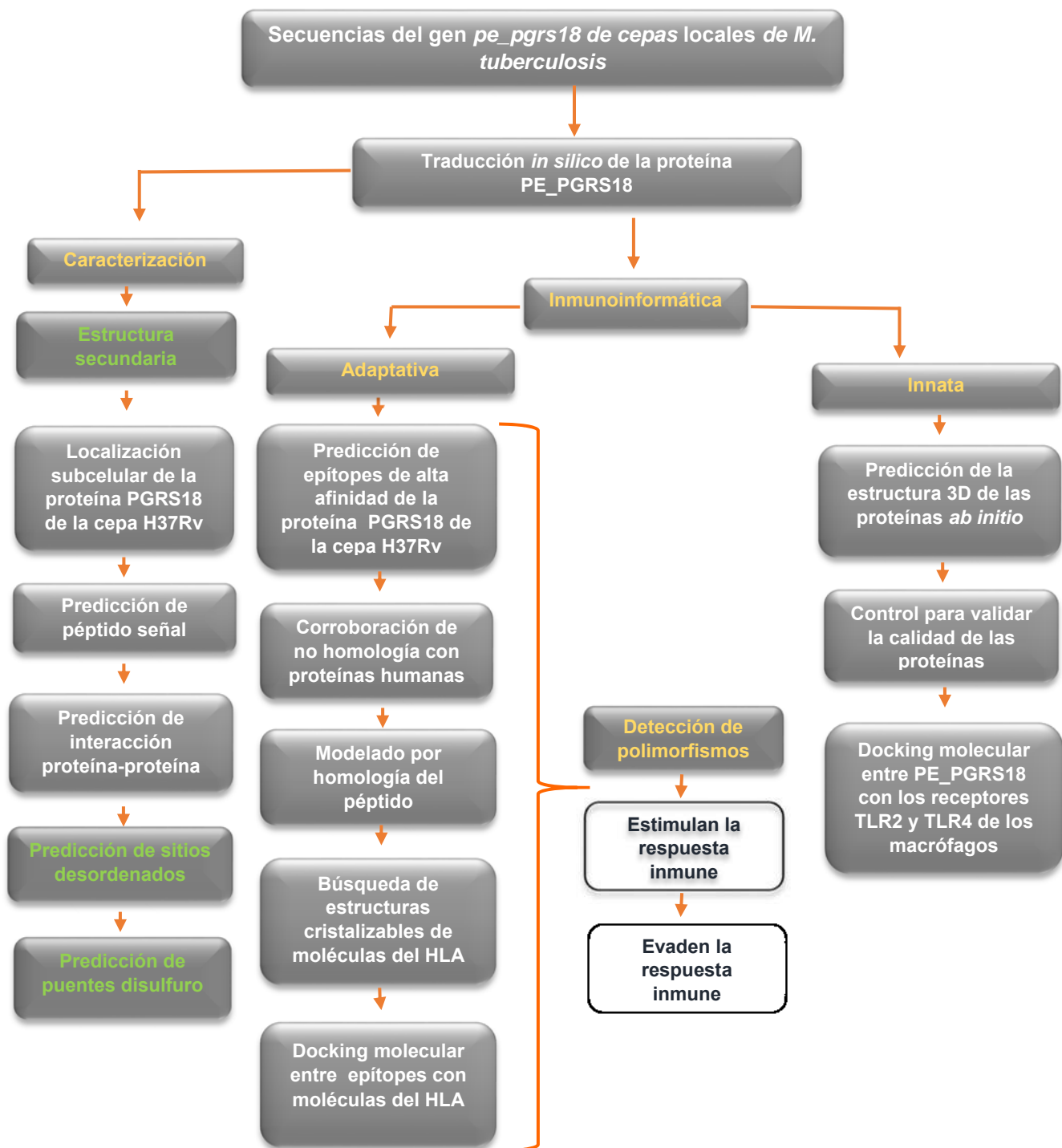
5.1 Objetivo general

Analizar *a dem* la relevancia inmunogénica de la proteína PE_PGRS18 de cepas de *Mycobacterium tuberculosis*.

5.2 Objetivos específicos

- 1) Determinar *in silico* la localización celular y posible función de la proteína PE_PGRS18 de *M. tuberculosis*.
- 2) Identificar epítopes de la proteína PE_PGRS18 de alta probabilidad de interacción con moléculas de HLA clase I y II.
- 3) Definir posibles polimorfismos de la proteína PE_PGRS18 que estimulen y/o permitan la evasión de la respuesta inmune del hospedero.
- 4) Determinar las secuencias conservadas de la proteína PE_PGRS18 que permiten la interacción con los receptores tipo TLR2 y TLR4 de los macrófagos del hospedero.

6. ESTRATEGIA EXPERIMENTAL



7. MATERIALES Y MÉTODOS

7.1 Material biológico

Para este estudio se utilizaron las secuencias del gen *pe_pgrs18* de 46 cepas de *M. tuberculosis*, de las cuales 29 fueron aisladas en el estado de Michoacán durante el periodo 2009-2014 de pacientes con tuberculosis pulmonar, meníngea, ganglionar y mixta (Tabla 1), donadas por el “Hospital General Dr. Miguel Silva” (HGMS), la clínica no. 75 del Instituto Mexicano del Seguro Social (IMSS) de Morelia Michoacán y el Laboratorio de Micobacterias del Laboratorio Estatal de Salud Pública de Michoacán (LESP). Las otras 17 cepas fueron aisladas en el estado de Querétaro durante un periodo 2015, de estas cepas no se cuenta con información clínica y epidemiológica. Como cepa de referencia se utilizará la cepa *M. tuberculosis* H37Rv donada por el Dr. Antonio Enciso Moreno, de la Unidad de Investigación del IMSS-Zacatecas.

Tabla 1. Características clínicas, genotípicas y resistencia a antibióticos de las cepas de estudio

Cepa (MYC)	Estado	Localización P/EP/M	Linaje	Familia	Genotipo AF	Resistencia S/ R
047	Mich.	NR	Euro-americano	New-1	NR	S
048	Mich.	NR	Euro-americano	New-1	NR	S
051	Mich.	Pulmonar	Euro-americano	Haarlem	NR	S
052	Mich.	Pulmonar	Euro-americano	Haarlem	NR	S
053	Mich.	NR	Euro-americano	X	NR	S
055	Mich.	Meníngea	Euro-americano	Lam	NR	R (SM)
058	Mich.	Pulmonar	Oeste de África	Ghana	NR	S
061	Mich.	Pulmonar	Euro-americano	La	NR	S
062	Mich.	Pulmonar	Indo-oceánico	EAI	NR	R (SM,EMB, PZA)
064	Mich.	Pulmonar	Euro-americano	X	NR	S
065	Mich.	Pulmonar	Euro-americano	LAM	NR	R (MDR-TB)
066	Mich.	Pulmonar	NR	NR	Haarlem	NR
068	Mich.	Pulmonar	Euro-americano	Haarlem	NR	NR
069	Mich.	NR	Indo-oceánico	EAI	Haarlem	R (RIF)
070	Mich.	Pulmonar	NR	NR	EAI	NR
071	Mich.	Pulmonar	NR	NR	EAI	NR
076	Mich.	NR	NR	NR	Haarlem	NR
078	Mich.	Meníngea	NR	NR	NR	NR
079	Mich.	Pulmonar	NR	NR	Haarlem	NR
080	Mich.	Pulmonar	NR	NR	LAM	NR
086	Mich.	NR	NR	NR	LAM	NR
087	Mich.	Renal	NR	NR	Haarlem	NR

089	Mich.	NR	NR	NR	Haarlem	NR
090	Mich.	NR	NR	NR	Haarlem	NR
091	Mich.	NR	NR	NR	Haarlem	NR
093	Mich.	Pulmonar	NR	NR	X	NR
096	Mich.	Pulmonar	NR	NR	Haarlem	NR
097	Mich.	Pulmonar	NR	NR	NR	NR
098	Mich.	Pulmonar	NR	NR	X	NR
100	Qro.	NR	NR	NR	LAM	NR
101	Qro.	NR	NR	NR	LAM	NR
102	Qro.	NR	NR	NR	Haarlem	NR
103	Qro.	NR	NR	NR	Haarlem	NR
104	Qro.	NR	NR	NR	Bovis	NR
106	Qro.	NR	NR	NR	Haarlem	NR
107	Qro.	NR	NR	NR	S	NR
108	Qro.	NR	NR	NR	Haarlem	NR
109	Qro.	NR	NR	NR	Haarlem	NR
110	Qro.	NR	NR	NR	Bovis	NR
112	Qro.	NR	NR	NR	Haarlem	NR
113	Qro.	NR	NR	NR	EAI	NR
114	Qro.	NR	NR	NR	Bovis	NR
115	Qro.	NR	NR	NR	LAM	NR
116	Qro.	NR	NR	NR	Haarlem	NR
117	Qro.	NR	NR	NR	S	NR
120	Qro.	NR	NR	NR	S	NR

LAM: Latino-América-Mediterráneo; EAI: Este África-Indio; NR: No registrado; S: sensible; R: resistente; MDR-TB: Multidrogresistente; XDR-TB: Extremadamente drogorresistente.

7.2 Caracterización *in silico* de la proteína PE_PGRS18

La secuencia FASTA de la proteína PE_PGRS18 de la cepa H37Rv, se descargó de la base de datos del NCBI (<http://www.ncbi.nlm.nih.gov/protein/2924431?report=fasta>), con el fin de realizar todos los procedimientos bioinformáticos de análisis (Sherry *et al.*, 2001). Se realizó la traducción *in silico* de las 49 secuencias del gen *pe_pgrs18* de las cepas de *M. tuberculosis*, con la ayuda del programa Mega versión 6.06.

7.2.1 Propiedades

Se utilizaron tres servidores para el análisis de las secuencias de aminoácidos de las proteínas PE_PGR18. PredictProtein, disponible en <https://www.predictprotein.org/home>, permite predecir la estructura secundaria de la proteína PE_PGRS18, su localización, los sitios desordenados, puentes de

disulfuro y la interacción proteína-proteína, lo cual es importante para conocer la organización estructural y funcional en la célula hospedero. SIGNAL-IP LOCTRE3 (<http://www.cbs.dtu.dk/services/SignalP/>) permite predecir la presencia y ubicación de los sitios de escisión (primer aminoácido en la proteína madura) del péptido señal en las secuencias de aminoácidos de la proteína PE_PGRS18. Compute pI/Mw permite calcular el punto isoeléctrico de la proteína, así como su peso molecular, este servidor se encuentra disponible en https://web.expasy.org/compute_pi/.

7.2.2 Modelado *in silico* de la proteína PE_PGRS18

Para el modelado de la proteína PE_PGRS18 por homología se utilizó el servidor Phyre2 (<http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index>), el cual construye la estructura 3D pronosticada para dicha proteína mediante plantillas obtenidas de distintas bases de datos y realiza modelado de *novoo* de aquellas regiones que no tienen similitud con las plantillas. Este mismo servidor permite predecir la función de la proteína a partir de sus secuencias de aminoácidos. Los servidores RAMPAGE y ProSA se utilizaron para evaluar la calidad de las proteínas predichas y el reconocimiento de errores de los modelos. Dichos servidores se encuentran disponibles en <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php> y <https://prosa.services.came.sbg.ac.at/prosa.php>, respectivamente.

a) Comparaciones globales de la estructura tridimensional de la proteína PE_PGRS18 H37Rv y de las cepas de estudio

Una vez obtenido el modelado de la proteína de referencia se realizó el modelado 3D de las cepas de estudio mediante el servidor Robetta (<http://rosetta.bakerlab.org/>). Este servidor permite agregar una plantilla de tu elección para realizar el modelado. Posteriormente se verificó la calidad de los modelos mediante los servidores Rampage y Prosa-web.

Enseguida, se realizó la superposición de los modelos obtenidos de las cepas de estudio con respecto al modelo de la cepa de referencia. Este análisis se realizó mediante el programa Chimera en la opción Structure Comparison > MatchMarker,

el cual indica el valor de RMSD (raíz de la desviación cuadrática media de las posiciones de pares de átomos de dos estructuras que queremos comparar) la cual mide la similitud entre diferentes estructuras, o entre las coordenadas de un modelado, y las estructuras obtenidas experimentalmente. Cuando el valor del RMSD es más cercano a 0 indica que hay una mayor similitud entre las estructuras de estudio y el molde (Meng *et al.*, 2006).

b) Docking molecular entre la proteína PE_PGRS18 y los receptores tipo Toll

Se realizó la búsqueda de estructuras cristalizables de los receptores Toll-like 2 (TLR2) y Toll-like 4 (TLR4) de los macrófagos mediante la base de datos Protein Data Bank (PDB). Además, se realizó el acoplamiento de la proteínas PE_PGRS18 con los receptores TLR2 y TLR4 obtenidos del servidor PDB mediante el programa Prism (<http://cosbi.ku.edu.tr/prism/>) con el objetivo de corroborar la interacción descrita por la literatura y observar si los polimorfismos encontrados en las secuencias de las proteínas PE_PGRS18 provocan algún cambio en la interacción de la proteína con su receptor específico.

c) Predicción de la Interacción de la proteína PE_PGRS18H37Rv

Para la predicción de la interacción de la proteína PE_PGRS18 se utilizó el servidor String (<http://string-db.org>). String es una base de datos de predicción y conocimiento de las interacciones proteína-proteína. La interacción incluye una asociación directa (física) e indirecta (funcional) (Szklarczyk *et al.*, 2019). Valores cercanos a 1 indican una mayor interacción.

7.2.3 Inmunoinformática

a) Predicción de epítopes

Para este análisis se eligieron los alelos del Antígeno Leucocitario Humano (HLA por sus siglas en ingles) más frecuentes de distintas partes del mundo (Tabla 2).

Tabla 2. Alelos de clase I y II del HLA más frecuentes de distintas regiones del mundo			
	HLA Clase I		HLA Clase II
	HLA-A	HLA-B	HLA-DRB1
Australia	A*34:01	B*13:01	
Oceanía	A*24:02	B*40:01	DRB1*15:02
Europa		B*07:02	
América del Norte	A*24:02	B*40:02	DRB1*14:02
México	A*02:01		DRB1*08:02
	A*24:02		DRB1*04:07
	A*02:06	B*39:05	DRB1*14:06
	A*68:01	B*39:06	DRB1*07:01
	A*11:01	B*35:01	DRB1*04:04
		B*40:02	DRB1*16:02
América del Sur	A*3101	B*15:04	DRB1*16:02
Noreste de Asia	A*11:01	B*51:01	
Sureste de Asia	A*24:02	B*40:01	DRB1*12:02
Suroeste de Asia	A*11:01	B*51:01	DRB1*03:01
África del Norte	A*02:01	B*44:03	DRB1*03:01
África subsahariana		B*53:01	DRB1*15:03

Para la predicción de péptidos de la proteína PE_PGRS18 de *M. tuberculosis* que pueden ser reconocidos por moléculas de clase I y II del HLA, se utilizaron los siguientes servidores:

- **HLA Clase I**

En este caso se utilizó el servidor NetMHCpan (<http://www.cbs.dtu.dk/services/NetMHCpan/>), el cual predice la unión de péptidos (epítopes) a moléculas de clase I del HLA utilizando redes neuronales artificiales (Jurtz *et al.*, 2017). Los péptidos predichos tienen un tamaño de nueve aminoácidos.

Para ello, al igual que en el caso anterior, se utilizaron las 46 secuencias de aminoácidos del gen *pe_pgrs18* de *M. tuberculosis*.

Se utilizó además la herramienta Blastn (búsqueda de alineación local básica) que permite encontrar regiones de similitud entre secuencias homólogas (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). Este servidor nos permite saber si los epítopes predichos presentan alguna homología con secuencias de proteínas humanas. Esto es importante ya que si los epítopes presentan homología con secuencias de proteínas humanas entonces no se activaría la respuesta inmune del hospedero.

- **HLA Clase II**

Se utilizó el servidor NetMHCIIpan (<http://www.cbs.dtu.dk/services/NetMHCIIpan/>) para predecir las regiones de unión (epítopes) a los alelos HLA de clase II. Este servidor utiliza redes neuronales artificiales (Karosiene *et al.*, 2013) para realizar las predicciones. Para ello se utilizaron las 46 secuencias de aminoácidos traducidas del gen *pe_pgrs18* de *M. tuberculosis*, con la finalidad de obtener sólo los péptidos codificados por este gen que pueden interaccionar con moléculas de clase II del HLA. Los péptidos (epítopes) predichos tienen un tamaño de nueve aminoácidos.

a) Docking molecular entre los epítopes y alelos del HLA

El servidor PEP-FOLD (<http://mobyli.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD>) permite predecir el modelado 3D de *novo* a partir de la secuencia aminoácidos, de péptidos de 5 a 50 aminoácidos. Este método se basa en las letras del alfabeto estructuras (SA), derivado de los modelos ocultos de Márkov o HMM (por sus siglas en inglés por, Hidden Markov Mode). Los mejores modelos se eligieron de acuerdo a dos parámetros principalmente:

- 1) sOPED. Indica el valor de la energía, donde una menor energía representa una mayor estabilidad del modelo.
- 2) Tm (modelado de plantilla). Permite la evaluación totalmente automatizada de las predicciones de la estructura de la proteína. Tiene el valor entre 0-1, donde 1 indica una buena predicción del modelo.

Por lo tanto, los mejores modelos tienen la energía sOPEP más baja y el valor más alto de Tm. En seguida se realizó el modelado 3D por homología de los alelos de clase I y II del MHC mediante el servidor Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>) en modo intensivo. Este servidor utiliza métodos avanzados de detección remota de homología para construir modelos 3D mediante el algoritmo de Poing.

En seguida se verificó la calidad de dichos modelos mediante los servidores Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) y Prosa web (<https://prosa.services.came.sbg.ac.at/prosa.php>).

1) Rampage. - Se basa en los criterios de Ramachandra, que permite distinguir conformaciones raras o erróneas de la estructura central (columna vertebral) de la estructura secundaria de una proteína, mediante los valores de los ángulos phi (Φ) y psi (Ψ) de cada aa (Lovell *et al.*, 2003). El servidor nos indica el número y porcentaje de residuos que se encuentran en regiones favorables, regiones permitidas y en regiones atípicas. Cuando el porcentaje de residuo en regiones favorables es >90 el servidor indica que el modelo 3D de la proteína es de buena calidad (Lovell *et al.*, 2003).

2) Prosa web. - Realiza la validación de la estructura de la proteína, mediante las coordenadas atómicas utilizando solo los C α de la estructura de entrada. Calcula el z-score el cual indica la calidad general del modelo y mide la desviación de la energía total de la estructura con respecto a una distribución de energía derivada de conformaciones aleatorias. El servidor hace una comparación con los z-score de todas las cadenas de proteínas determinadas experimentalmente por rayos X y RMN (resonancia magnética nuclear). Valores positivos del z-score indican partes problemáticas o erróneas de un modelo (Wiederstein and Sippl, 2007).

En seguida se realiza el refinamiento de cada uno de los modelos mediante el servidor Galaxirefine para mejorar la calidad de cada modelo. Este servidor realiza un reordenamiento de la cadena lateral de la proteína (las cadenas R de los aa que constituyen a la proteína), el ordenamiento de la cadena lateral y posteriormente la relajación general de la estructura (estructura con mayor afinidad

al oxígeno) mediante simulación de dinámica molecular (Heo *et al.*, 2013). El servidor muestra 5 modelos que genera a partir del modelo inicial por lo que nos permite elegir el modelado que mejor nos convenga de acuerdo al puntaje de GDT-HA. El GDT-HA indica la distancia global de modelos basados en plantados de alta precisión. Esta prueba identifica un grupo de residuos en la predicción que se desvían del objetivo, pero no más que la distancia de corte (Kopp *et al.*, 2007).

A continuación, se realiza el docking molecular entre los modelados 3D obtenidos de los epítopes y los alelos de clase I y II del HLA mediante el servidor GalaxyPepdock con el objetivo de corroborar si los epítopes predichos coinciden con los alelos indicados por los servidores mencionados anteriormente. Este servidor arroja el puntaje de “precisión estimada” que se refiere a la fracción de aa pronosticados del sitio de unión y la similitud con el molde obtenido de distintas bases de datos. Dicho puntaje va de 1 a 0 donde los valores más cercanos a 1 indican una mayor precisión en la interacción péptido- alelo (Lee *et al.*, 2015).

8. RESULTADOS

8.1 Caracterización de la proteína PE_PGRS18

Debido a que la estructura tridimensional de las proteínas y su actividad biológica dependen de las propiedades fisicoquímicas de sus aminoácidos constituyentes, este análisis básico es esencial. Los resultados mostraron que la proteína PE_PGRS18 tiene una longitud de 457 aminoácidos (aa) y que dicha proteína está constituida por hélices (hélice α , pi- y 310) en un 16.85%, láminas β en un 40.04 % y bucles (*loops*) en un 43.5%. Estos resultados indican que la mayor parte de la estructura de la proteína PE_PGRS18 está constituida de *loops* y láminas β .

La accesibilidad del solvente por parte de la proteína PE_PGRS18 mostró que el 78.56 % de su estructura es hidrofóbica, por lo que esta parte de la proteína se encuentra oculta, mientras que un 14.88% permanece expuesta (hidrofílica) y un 6.56% es intermedia. En la predicción de hélices transmembrana de la estructura de la proteína PE_PGR18 se observó una porción de 19 aa localizada en la posición 38-56. Sin embargo, el servidor no asegura que se trate de una hélice transmembrana o del sitio de escisión de un péptido señal (Fig. 4). Para corroborar este resultado se utilizó el servidor SignalP, estos resultados son mostrados más adelante. También se predijeron las regiones flexibles y desordenadas de la estructura secundaria de la proteína PE_PGRS18 (Fig. 4). En relación a la predicción de enlaces disulfuro en la proteína PE_PGRS18 no se encontraron.

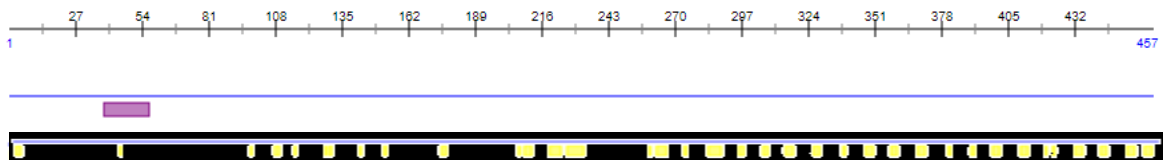


Figura 4. Hélice transmembrana en la proteína PE_PGRS18 y regiones desordenadas y flexibles. Banda morada= posible hélice transmembrana de 19 a.a de longitud en posición 38-56. Bandas amarillas= regiones desordenadas y flexibles.

Con respecto a la localización subcelular de la proteína PE_PGRS18, el servidor TBpred indicó con un puntaje 0.48979843 (es el más alto con respecto a las localizaciones restantes) sugiriendo que dicha proteína es transmembranal. Este servidor indica la localización de proteínas específicamente pertenecientes a *M. tuberculosis*. De acuerdo a la composición de aminoácidos de la proteína, genera un puntaje para cada una de las 4 localizaciones. El puntaje mayor indica la localización de la proteína.

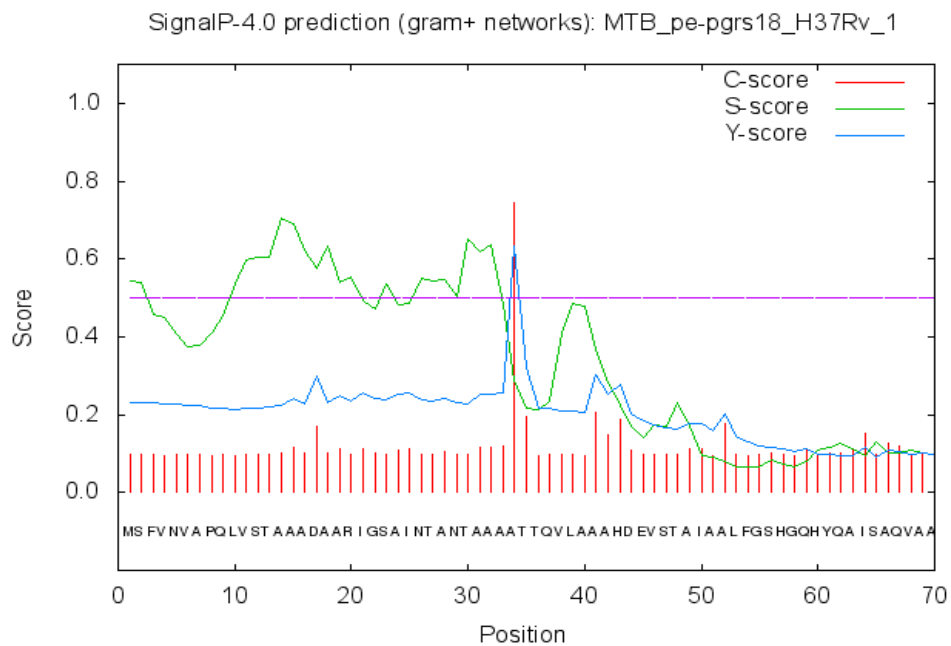


Figura 5. Presencia, localización y péptido señal de la proteína PE_PGRS18. C-score= puntaje de escisión del péptido señal en bruto, S-score= posición del péptido señal en la parte madura de la proteína, Y-score= puntuación del sitio de escisión combinada de C-score y S-score.

Se identificó el sitio de escisión de un péptido señal entre la posición 38-56 aa (Fig. 5), así como 17 regiones de unión a proteína en las posiciones 127, 129, 133-136, 180, 217, 220, 223, 225, 235, 238, 241-244, 246-247, 249, 251-252, 255-256, 266, 268 (Fig. 6).

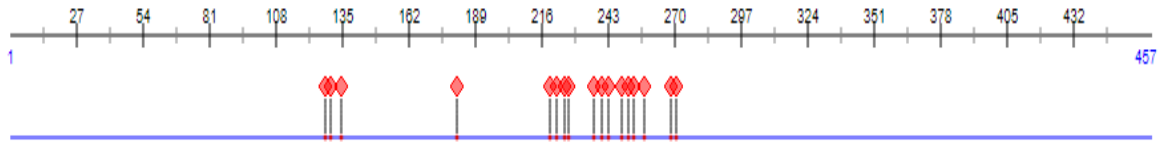


Figura 6. Regiones de unión proteína-proteína. Los rombos rojos indican las regiones de unión proteína-proteína.

El punto isoeléctrico de la proteína PE_PGRS18 es 4.27 y su peso molecular es 42074.79 Da.

8.2 Identificación de péptidos de la proteína PE_PGRS18 que interaccionan con moléculas de HLA clase I y II

Para este análisis se eligieron 19 alelos del HLA clase I (8 HLA-A y 11 HLA-B) y 12 alelos del HLA clase II (HLA-DRB) como representantes de la diversa población humana. En la proteína de la cepa de referencia H37Rv, el análisis del HLA clase II predijo un total de 124 epítopes CD4+ de los cuales 20 son de alta afinidad, mientras que el análisis del HLA clase I predijo un total de 389 epítopes CD8+ los cuales corresponden a 167 epítopes del HLA-A, 67 epítopes de alta afinidad, y 222 a los alelos HLA-B, 48 epítopes de alta afinidad.

El análisis de distribución de los epítopes del HLA clase I reveló que del total de epítopes del HLA-A y HLA-B, 167 pertenecen al dominio PE, 44 de alta afinidad y 105 son de baja afinidad (Fig. 7A). Además, 206 epítopes pertenecen al dominio PGRS, con 64 de alta afinidad y 142 de baja afinidad (Fig. 7B), y 32 epítopes se encuentran en la región no estructural, con 5 epítopes de alta afinidad y 27 de afinidad baja (Fig. 7C).

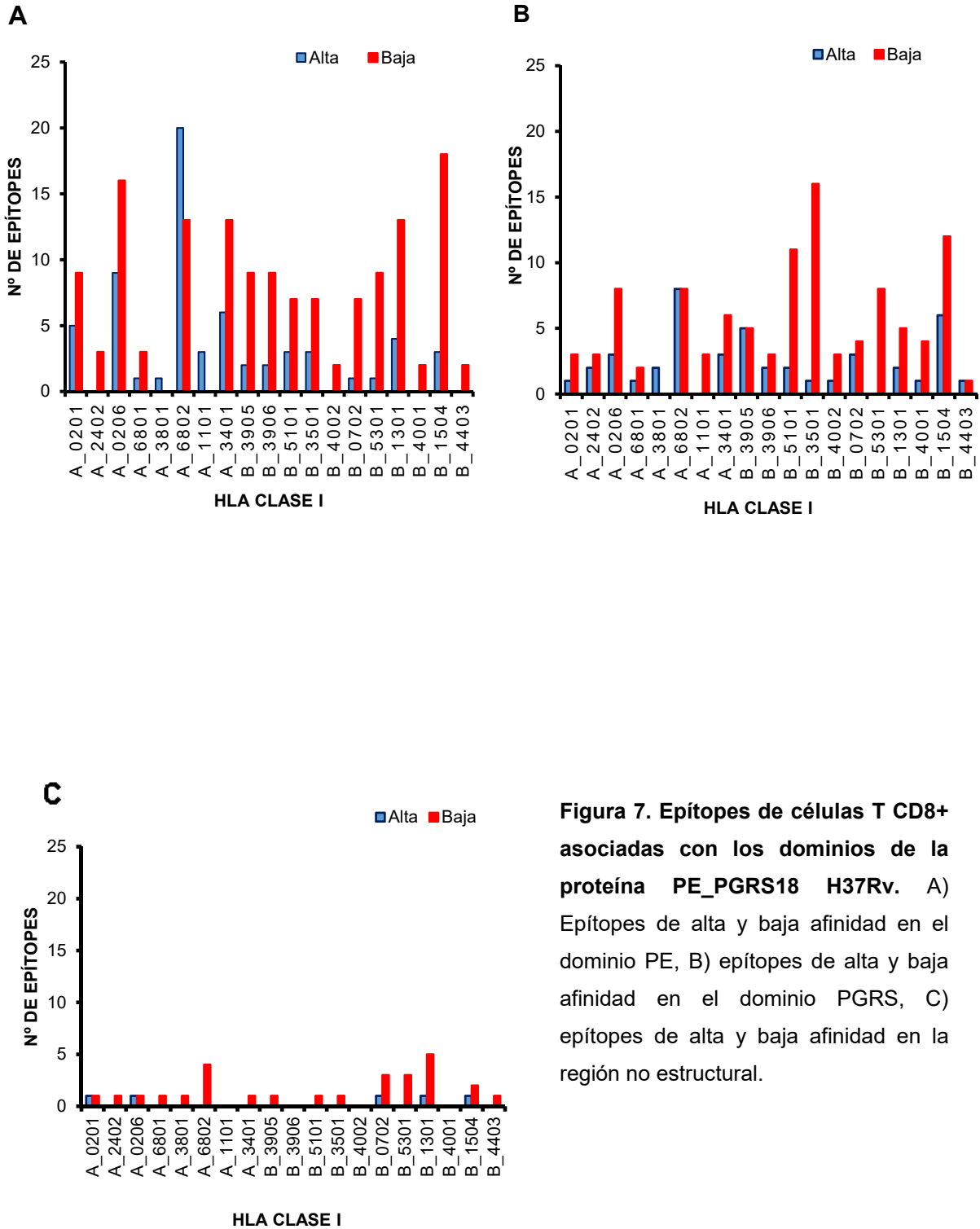


Figura 7. Epítopes de células T CD8+ asociadas con los dominios de la proteína PE_PGRS18 H37Rv. A) Epítopes de alta y baja afinidad en el dominio PE, **B)** epítopes de alta y baja afinidad en el dominio PGRS, **C)** epítopes de alta y baja afinidad en la región no estructural.

El análisis de distribución de los epítopes del HLA clase II reveló que de los 124 epítopes identificados, 62 pertenecen al dominio PE, con 19 epítopes de alta afinidad y 53 de baja afinidad (Fig. 8A). Además, 34 epítopes pertenecen al dominio PGRS de baja afinidad (Fig. 8B) y 16 epítopes se encuentran en la región no estructural, con un epítope de alta afinidad y 15 de afinidad baja (Fig. 8C).

8.2.1 Epítopes de unión a células T CD8+ en las secuencias de estudio

El análisis de péptidos de unión a moléculas de HLA clase I, mostró un total de 16,894 epítopes. El análisis de distribución de los epítopes en los dominios de la proteína de estudio indicó que del total de epítopes encontrados, 6,021 pertenecen al dominio PE con 2,191 epítopes de alta afinidad y 3,850 de baja afinidad (Fig.9A). Además 8,851 epítopes pertenecen al dominio PGRS, de los que 2,654 son de alta afinidad y 6,197 de baja afinidad (Fig.9B); 2,022 epítopes se encuentran en la región no estructural con 269 epítopes de alta afinidad y 1,753 de afinidad baja (Fig.9C).

Se observó que el alelo A*6802 que pertenece solo a la región mexicana reconoce un mayor número de epítopes de alta afinidad, con un total de 889 los cuales se encuentran en el dominio PGRS, seguido del alelo A*0206 el cual reconoce un total de 411 epítopes de alta afinidad los cuales se encuentran en el dominio PGRS. Además, este alelo también pertenece a la región mexicana.

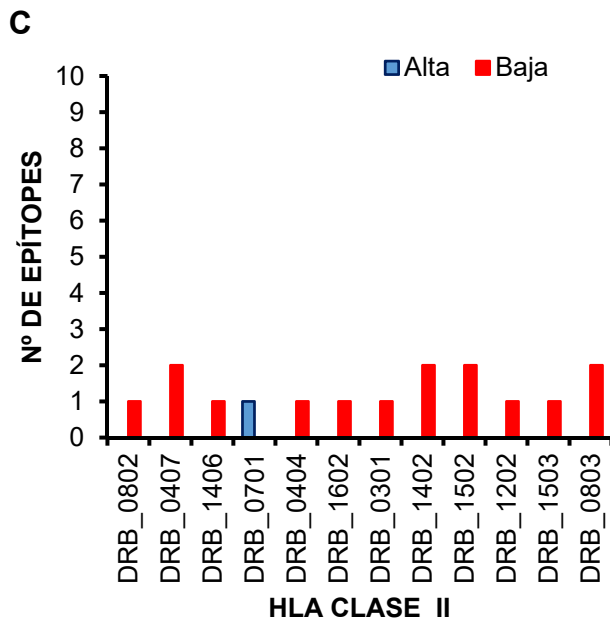
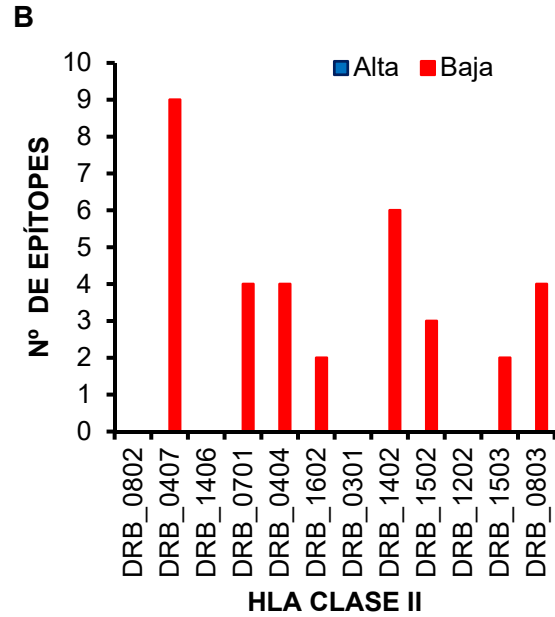
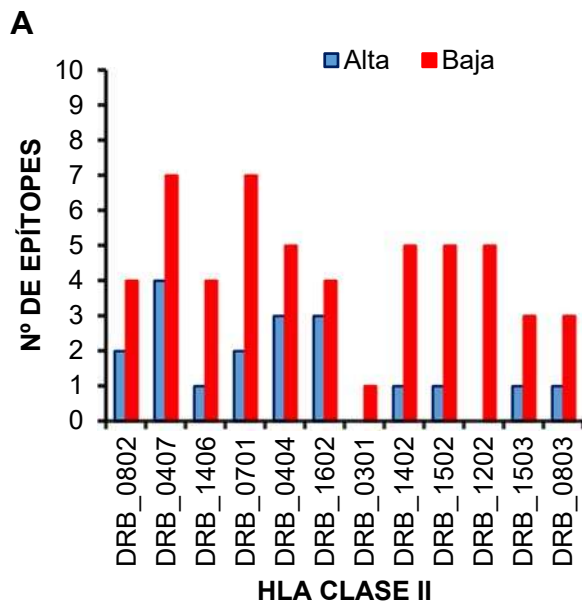


Figura 8. Epítopes de células T CD4+ asociadas con los dominios de la proteína PE_PGRS18 H37Rv. A) Epítopes de alta y baja afinidad en el dominio PE, B) epítopes de alta y baja afinidad en el dominio PGRS, C) epítopes de alta y baja afinidad en la región no estructural.

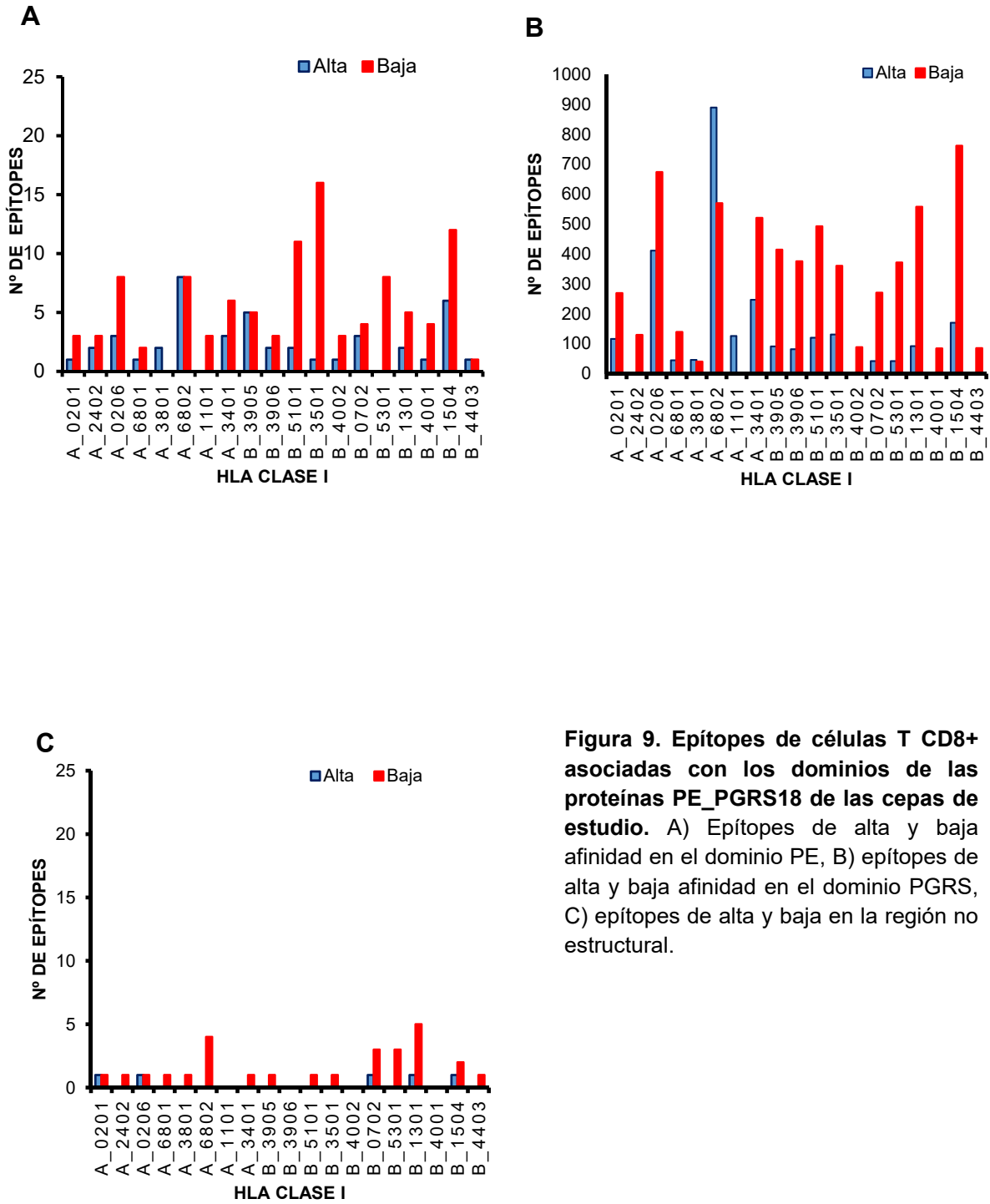


Figura 9. Epítopes de células T CD8+ asociadas con los dominios de las proteínas PE_PGRS18 de las cepas de estudio. A) Epítopes de alta y baja afinidad en el dominio PE, B) epítopes de alta y baja afinidad en el dominio PGRS, C) epítopes de alta y baja en la región no estructural.

8.2.1.a Comparación estadística del número de epítopes en los dominios de la proteína PE_PGRS18 que son reconocidos por alelos de clase I (células T CD8+) de la región mexicana y distintas regiones del mundo.

Por lo observado anteriormente se realizó la relación estadística de los alelos de clase I del HLA que reconocen un mayor número de epítopes comparando los alelos que pertenecen a la región mexicana y a otras regiones, en los distintos dominios de la proteína. Para este análisis se utilizó la prueba t de student para muestras independientes; para ello se utilizó el programa SPSS, el cual nos indica un valor de p menor a 0.05 si el valor es estadísticamente significativo.

En la tabla 3 se muestra los distintos dominios de la proteína, el total de epítopes que se encuentran en cada dominio y el número de epítopes que son reconocidos por alelos de la región mexicana y por distintas regiones.

Tabla 3. Epítopes de alta afinidad reconocidos por alelos de clase I del HLA de la región mexicana y distintas regiones			
Dominio	Total epítopes	México	Distintas regiones
PE	2378	1186	1192
PGRS	2770	1765	1005
No estructural	324	71	253

No se obtuvo un valor de p estadísticamente significativo (Tabla 4), por lo que se acepta la hipótesis nula la cual indica, que los alelos de la región mexicana reconocen el mismo número de epítopes que los alelos de distintas regiones en los distintos dominios de la proteína PE_PGRS18 de las cepas de estudio.

Tabla 4. Probabilidad de que los alelos de clase I de la región mexicana y de distintas regiones, reconozcan el mismo número de epítopes en los dominios de la proteína

Dominios	Valor de <i>p</i>
PE	0.296
PGRS	0.167
No estructural	0.174
Total de epítopes	<u>0.182</u>

8.2.2 Epítopes de unión a células T CD4+ en las secuencias de estudio

Se encontraron un total de 5,395 epítopes de unión a moléculas del HLA clase II, en las 46 secuencias de estudio de la proteína PE_PGRS18. El análisis de distribución de los epítopes en los dominios de la proteína de estudio indicó que del total de epítopes encontrados, 3,560 pertenecen al dominio PE, con 891 epítopes de alta afinidad y 2,669 de baja afinidad (Fig.10A). Además, 1,298 epítopes pertenecen al dominio PGRS, 1 epítope de alta afinidad y 1,297 de baja afinidad (Fig.10B). Se encontraron 537 epítopes en la región no estructural, 103 de los cuales son de alta afinidad y 434 de afinidad baja (Fig.10C).

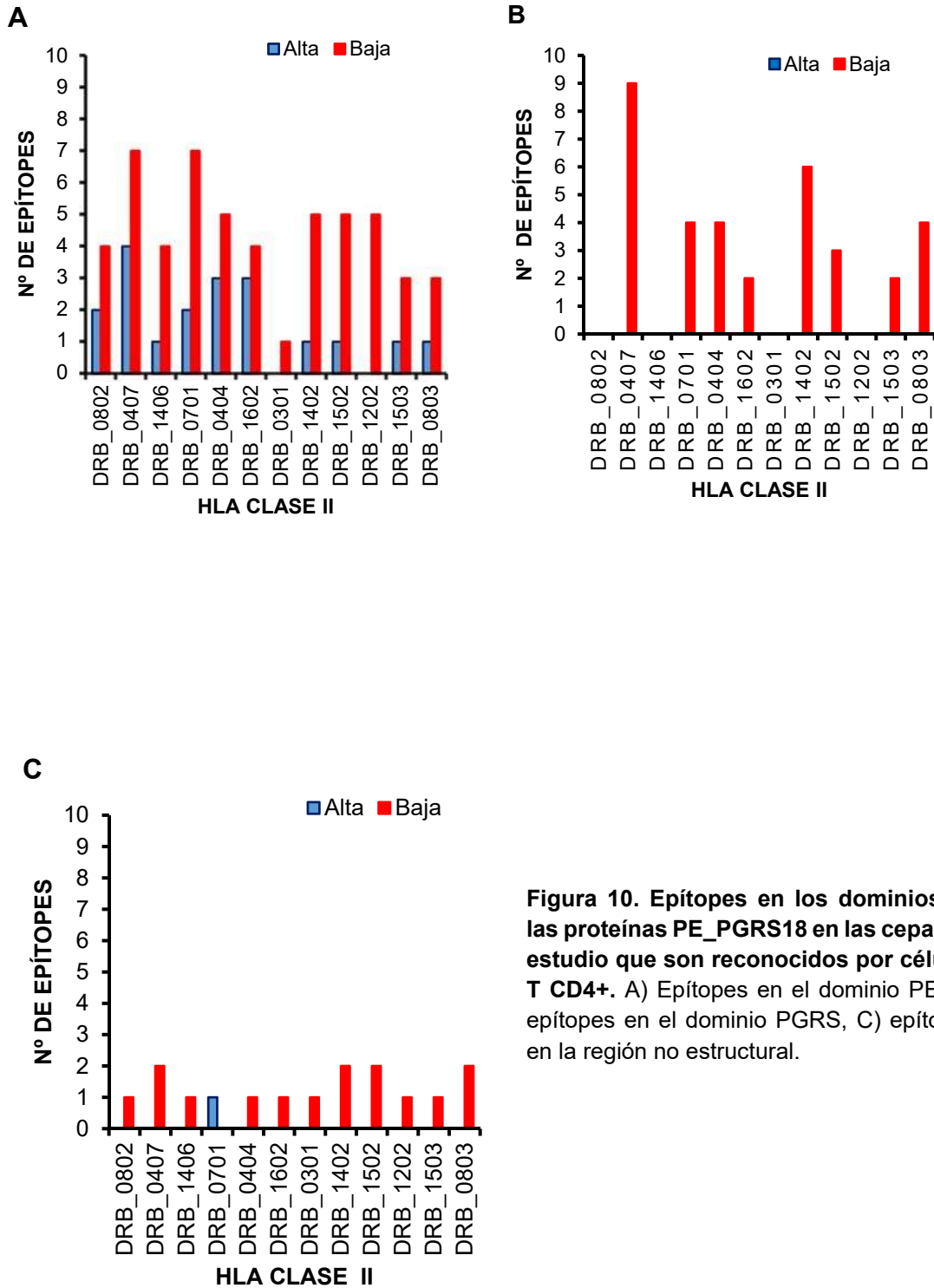


Figura 10. Epítopes en los dominios de las proteínas PE_PGRS18 en las cepas de estudio que son reconocidos por células T CD4+. A) Epítopes en el dominio PE, B) epítopes en el dominio PGRS, C) epítopes en la región no estructural.

En los gráficos se puede observar que el alelo DRB*0407 y el alelo DRB*1602 reconocen un mayor número de epítopes en el dominio PE de la proteína de estudio. Sin embargo, el alelo DRB*0407 pertenece solo a la región mexicana y el alelo DRB*1602 pertenece a la región mexicana y a la región de América del Sur.

a) Epítopes en los dominios de la proteína PE_PGRS18 que son reconocidos por alelos de clase II (células T CD4+) de la región mexicana y distintas regiones del mundo.

Posteriormente se determinó si los alelos de clase II (T CD4+) pertenecientes a la región mexicana reconocen un mayor número de epítopes en los dominios de la proteína PE_PGRS18 comparado con los alelos de distintas regiones del mundo. Para este análisis se utilizó la prueba t de student para muestras independientes, del programa SPSS. Que indica un valor de p menor a 0.05 si el valor es estadísticamente significativo.

En la tabla 5 se puede observar que en el dominio PGRS y la región no estructural se encontraron un mayor número de epítopes que son reconocidos por alelos de la región mexicana.

Tabla 5. Epítopes reconocidos con alta afinidad por alelos de clase II del HLA de diferentes regiones geográficas.

Dominio	Total epítopes	México	Distintas regiones
PE	1029	688	341
PGRS	1	1	0
No estructural	103	102	1

Se obtuvo un valor estadísticamente significativo con una $p = 0.029$ ($p < 0.05$) en la relación estadística de los alelos de México y de distintas regiones del

mundo que reconocen un mayor número de epítopes en el dominio PE. Este resultado nos indica el rechazo de la hipótesis nula y se acepta la hipótesis alternativa la cual indica que los alelos de la región mexicana reconocen un mayor número de epítopes en el dominio PE que los alelos de distintas regiones del mundo.

8.3 Mutaciones que estimulan/evaden la respuesta inmune del hospedero

Se identificaron 332 mutaciones en las 46 secuencias de estudio, de las cuales 244 mutaciones estimulan la respuesta inmune, es decir, provocan que algunos péptidos sean reconocidos por células T CD4+ y CD8+. Así mismo, se identificaron 88 mutaciones que provocan la evasión de la respuesta inmune, es decir, péptidos que no son reconocidos por las células T CD4+ y CD8+. El análisis de distribución reveló que 22 mutaciones pertenecen al dominio PE de las cuales 11 mutaciones (11 SNP's) son capaces de estimular la respuesta inmune mediada por células T CD4+ y CD8+ y 11 mutaciones evaden esta respuesta inmune (11 SNP's) (Fig.11A). Además, 149 mutaciones pertenecen al dominio PGRS de las cuales 144 mutaciones (17 eliminaciones y 127 SNP's) son capaces de estimular la respuesta inmune mediada por células T CD4+ y CD8+ y 5 SNPs evaden esta respuesta inmune (Fig.11B). Conjuntamente, 161 mutaciones se encuentran en la región no estructural de la proteína de las cuales 89 (8 inserciones y 81 SNP's) son capaces de estimular la respuesta inmune mediada por células T CD4+ y CD8+ y 72 mutaciones evaden esta respuesta inmune (8 inserciones, 2 eliminación y 62 SNP's) (Fig.11C).

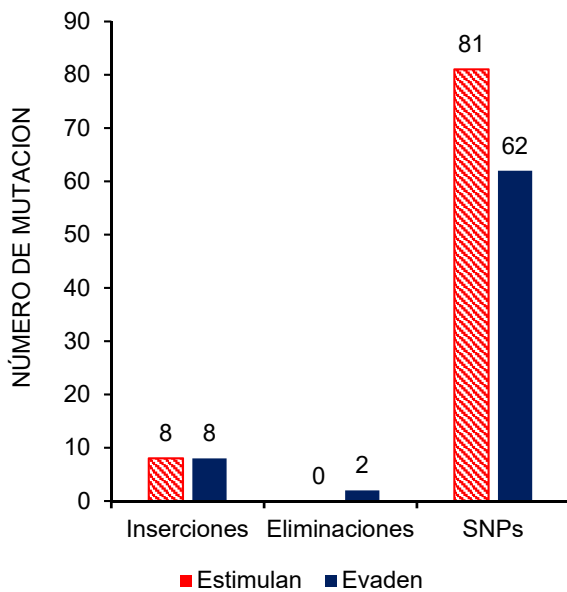
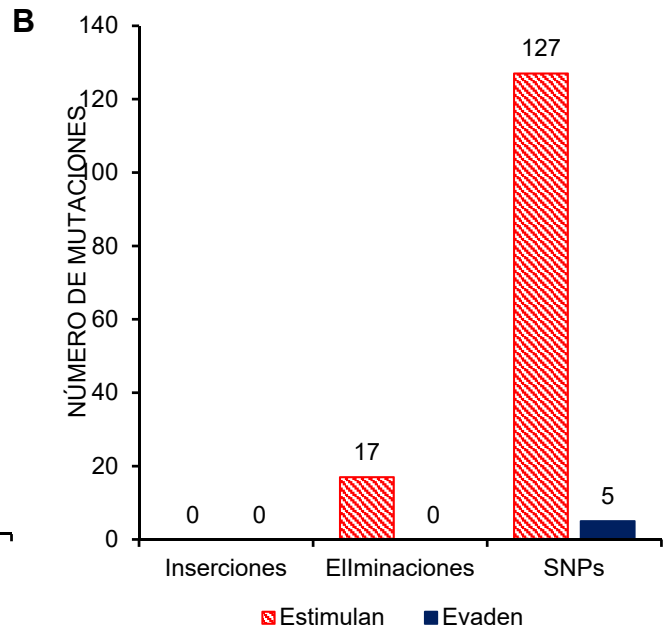
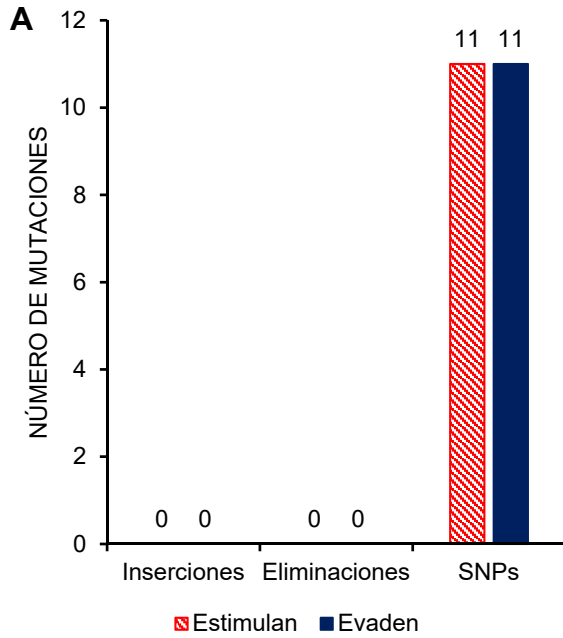


Figura 11. Mutaciones que estimulan o evaden el reconocimiento de péptidos por las células T CD4+ Y CD8+. A) Mutaciones en el dominio PE, B) Mutaciones en el dominio PGRS, C) mutaciones en la región no estructural.

8.3.1 SNPs que estimulan y evaden la respuesta inmune

Los SNPs son el tipo de mutación más abundante que probablemente estimula y evade la respuesta inmune del hospedero en los distintos dominios de la proteína PE_PGRS18 (Fig. 10). En la tabla 6 se muestra el total de SNPs encontrados en cada dominio, los SNPs que estimulan y los que evaden la respuesta inmune celular.

Tabla 6. SNPs que estimulan y evaden la respuesta inmune mediada por células T CD4+ y CD8+			
Dominio	Total SNPs	Estimulan	Evaden
PE	22	11	11
PGRS	132	127	5
No estructural	143	81	62

Se determinó cuáles SNPs en los distintos dominios de la proteína y la región no estructural, tienen mayor probabilidad de estimular o evadir la respuesta inmune del hospedero. Para este análisis se utilizó la prueba de Fisher, que determina si existe relación entre dos variables. Esta prueba indica una relación estadísticamente significativa con un valor de $p < 0.05$ (Hae-Young, 2017).

La prueba indicó que los 127 SNPs en el dominio PGRS, tienen mayor probabilidad de estimular la respuesta inmune de células T comparado con los SNPs del dominio PE y la región no estructural. Mientras tanto, los 11 SNPs en el dominio PE y 62 en la región no estructural tienen mayor probabilidad de evadir la respuesta inmune de células T comparados con los SNPs en el dominio PEGRS.

8.4 Modelado 3D de epítopes de la proteína PE_PGRS18 y modelado 3D de alelos de clase I y II del HLA

8.4.1 Epítopes

Se realizó el modelado de *novo* de diez epítopes de alta afinidad que son reconocidos por un mayor número de alelos CD4+ y CD8+ (anexo 1). Los modelos se seleccionaron de acuerdo a los criterios mencionados en materiales y métodos.

8.4.2 Alelos que reconocen un mayor número de epítopes en la proteína PE_PGRS18

Se realizó la búsqueda de las estructuras cristalizadas en la base de datos PDB de seis de los alelos de estudio, los cuales en el análisis previo indicaron que reconocen un mayor número de epítopes y además pertenecen a la región mexicana. Sin embargo, en PDB no se encuentra la estructura cristalizada de los alelos de estudio, algunas estructuras están disponibles pero unidas a un péptido. Es por ello que se realizó la búsqueda de las secuencias de aminoácidos de cada uno de los alelos de estudio en la base de datos IPD-IMGT/HLA (<https://www.ebi.ac.uk/ipd/imgt/hla/allele.html>), para posteriormente realizar el modelado 3D por homología de cada uno de los alelos mediante el servidor Phyre2. Cada uno de los modelados de los alelos utiliza alrededor de 120 moldes. En la tabla 7 se muestra el nombre de los alelos de clase I y II del HLA, algunos de los moldes que el servidor utilizó para realizar el modelado y los parámetros de robustez de cada molde.

Tabla 7. Información de los moldes representativos utilizados para el modelado 3D por homología de alelos de clase I y II del HLA.

Alelos clase I	Molde	% Confianza	% ID	Cobertura
A*68:020101	c1kj3H	100.0	71	25-302
	c2clvA	100.0	72	25-302
	c2zswE	100.0	71	25-301
	c1kj2I	100.0	71	25-301
	c1i7rA	100.0	96	25-299
	c3tidA	100.0	72	25-300
A*02:060101	c2clvA	100.0	73	25-302
	c1kj3H	100.0	72	25-302
	c1i7rA	100.0	100	25-299
	c2zswE	100.0	73	25-301
	c1kj2I	100.0	73	25-301
	c2zswC	100.0	73	25-301
B*15:040101	c1kj3H	100.0	71	25-302
	c2clvA	100.0	72	25-302
	c3tidA	100.0	72	25-300
	c1kj2I	100.0	71	25-301
	c1kjmA	100.0	75	25-300
	c2zswE	100.0	71	25-301
B*39:050101	c1kj3H	100.0	71	25-302
	c2clvA	100.0	72	25-302
	c1kj2I	100.0	71	25-301
	c2zswE	100.0	71	25-301
	c2zswC	100.0	71	25-301
	c2zswA	100.0	71	25-301
Alelos clase II	Molde	% Confianza	% ID	Cobertura
DRB*04:070101	c1iebD	100.0	75	25-302
	c3lqzB	100.0	66	25-302
	c1iaoB	100.0	62	25-302
	c1sebB	100.0	92	25-301
	c3usaH	100.0	63	25-301
	c4d8pD	100.0	67	25-301
DRB*16:020101	c1iebD	100.0	73	22-217
	c3lqzB	100.0	68	26-220
	c1iaoB	100.0	61	23-219
	c1sebB	100.0	95	31-221
	c3usaH	100.0	63	34-221
	c4d8pD	100.0	66	29-219

% Confianza: probabilidad de 0-100% de que la secuencia de estudio y el molde sean homólogos, %ID: proporción de residuos de la proteína de estudio equivalentes a residuos del molde idénticos en la alineación generada, Cobertura: cobertura de alineación del molde con respecto a la secuencia de estudio.

a) *Validación de los modelos.*

Posteriormente se realizó la validación de cada uno de los modelos de los alelos mediante el servidor Rampage y Prosa web (Tabla 8). Analizando el porcentaje de residuos en las regiones favorables y z-score de cada uno de los modelos de los alelos; se encontró que el porcentaje de residuos en regiones favorable en algunos modelos es menor al 90%.

Tabla 8. Calidad de los modelos 3D de los alelos de clase I y II del HLA.		
Alelos clase I	% Residuos en regiones favorables	z- score
A*68:020101	94.8	- 7.92
A*02:060101	81.3	-7.51
B*15:040101	85.3	-7.85
B*39:050101	85.0	-7.52
Alelos clase II	% Residuos en regiones favorables	z- score
DRB*04:070101	92.4	-4.2
DRB*16:020101	89.8	-5.16

Rampage= % de residuos en regiones favorables, Prosa web= z-score.

Para mejorar la calidad de los modelados 3D de los alelos de la clase I: A*02:060101, B*15:040101 y B*39:050101 y de la clase II: DRB*16:020101 se realizó el refinado de cada uno mediante el servidor Galaxyweb. Posteriormente se verificó la calidad del modelo con el servidor Rampage. En la tabla 9, se muestra el número del modelo elegido en el servidor Galaxyweb, los valores de GDT-HA y la validación de la calidad de los modelos por el servidor Rampage; se puede observar una calidad mayor al 90%.



Tabla 9. Términos del refinado de algunos alelos de clase I y II del HLA

Alelos clase I y II	Galaxyweb		Rampage
	# Modelo	GDT-HA	% residuos en regiones favorables
A*02:060101	3	0.8836	87.6
B*15:040101	5	0.9599	95.0
B*39:050101	3	0.9544	93.9
DRB*16:020101	3	0.9398	93.6

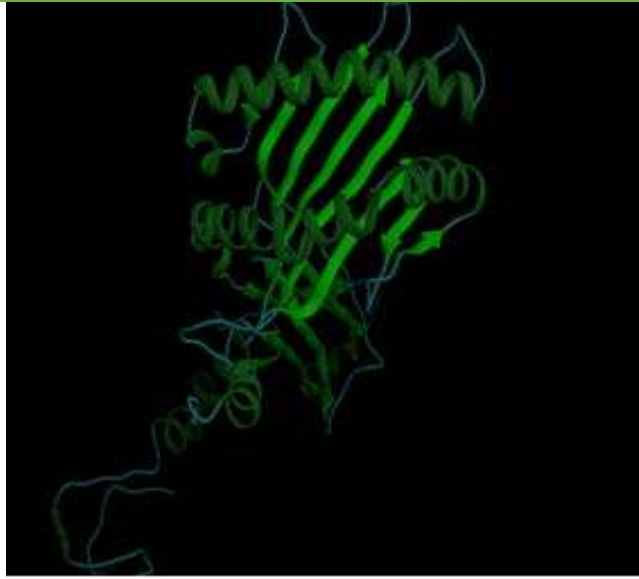
GDT-HA= Prueba de distancia global de modelos basados en plantados de alta precisión

Como se puede observar en el porcentaje de residuos en regiones favorables hay una mejoría de la calidad de los modelados 3D con porcentajes mayores al 90% de después del refinado (Tabla 9). En la tabla 10 se muestran los modelos 3D de los alelos de clase I y II del HLA. En cada uno de los modelos 3D se observan sus dominios y el surco de unión al péptido.

Tabla 10. Modelados 3D de los alelos de clase I y II del HLA

Alelos de clase I del HLA	Modelo 3D
A*68:020101	
A*02:060101	

B*15:040101



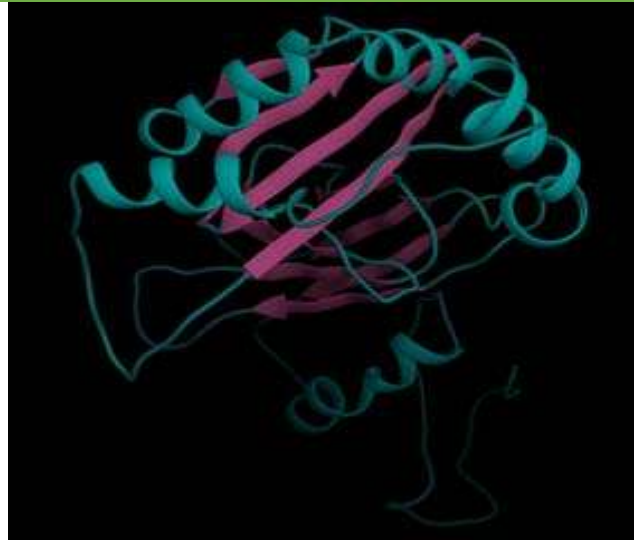
B*39:050101



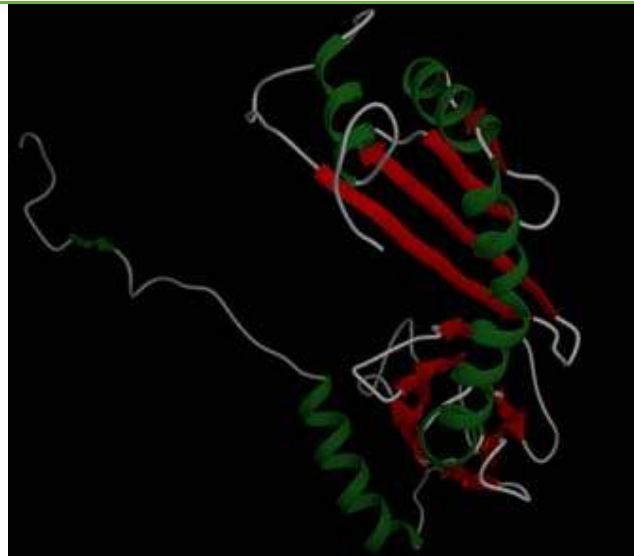
**Alelos clase II
del HLA**

Modelado 3D

DRB*04:070101



DRB*16:020101



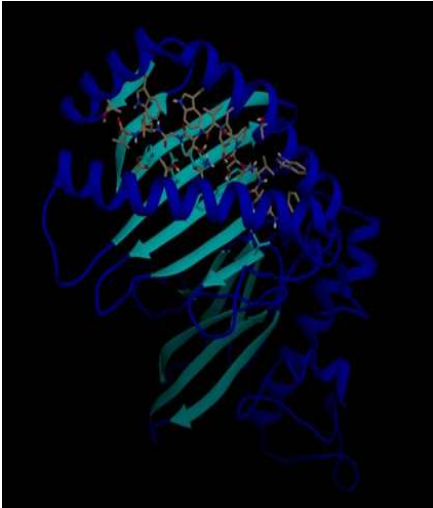
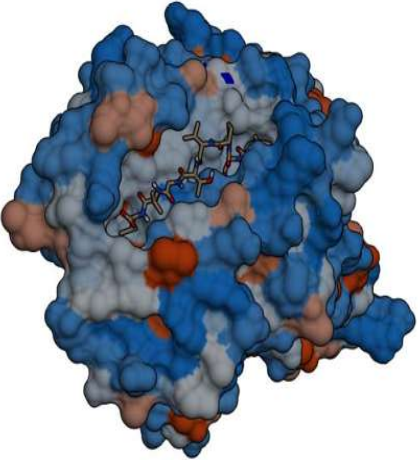
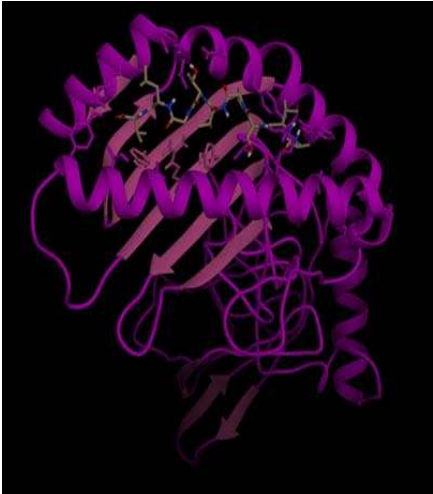
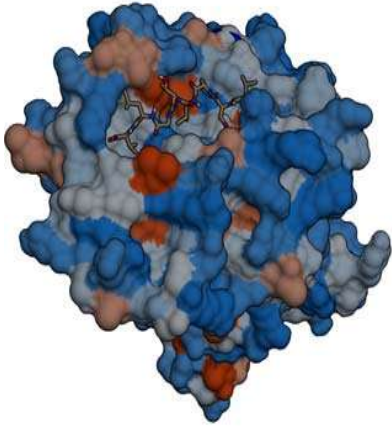
8.5 Docking molecular entre los epítopes y los alelos de estudio

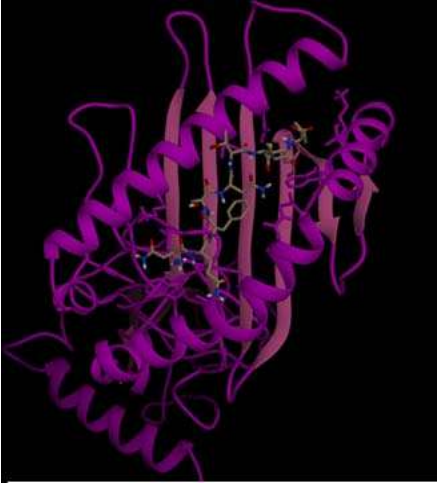
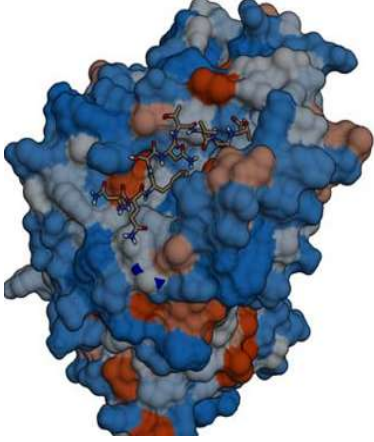
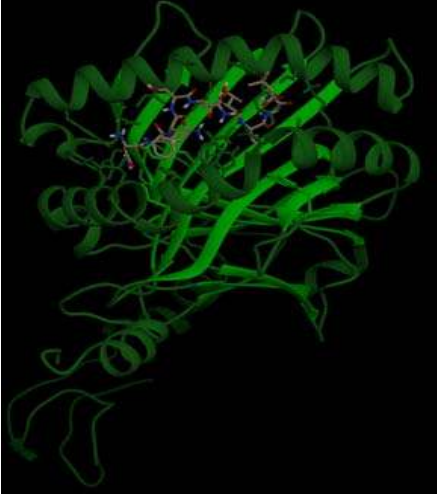
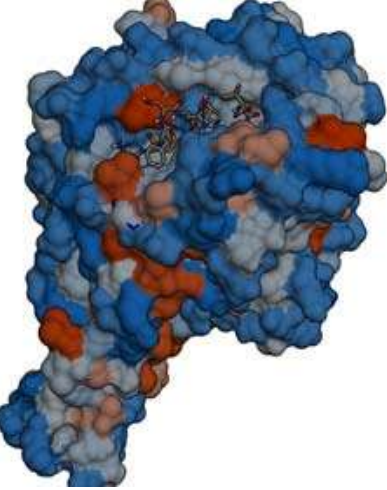
Se realizó el docking molecular, utilizando los epítopes y los alelos de clase I y II modelados previamente, para verificar si los epítopes se acoplaban de manera adecuada a las regiones de unión del péptido de las cadena alfa y beta del HLA I y II. En la tabla 11 se muestran los epítopes y los alelos específicos que los reconocen, según los servidores NetMHCpan y NetMHCIIpan. Posteriormente se verificó la interacción de los epítopes con sus respectivos alelos (tabla 12) mediante el servidor Galaxyepdock.

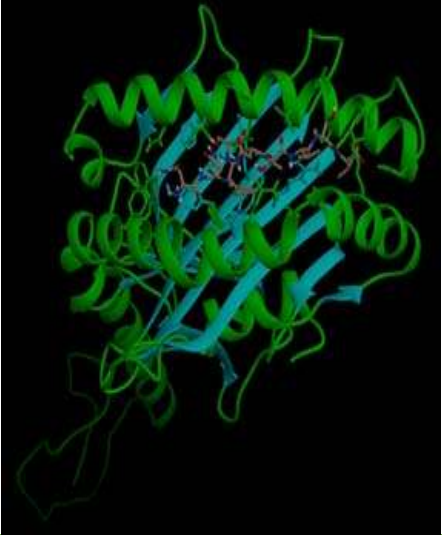
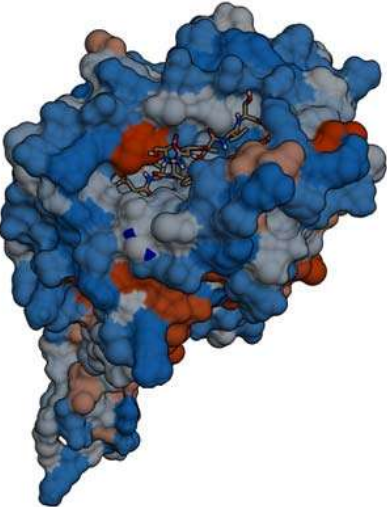
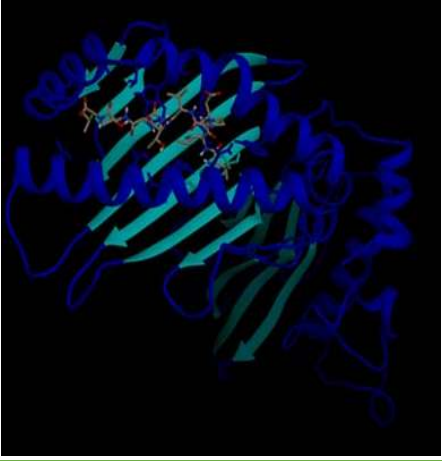
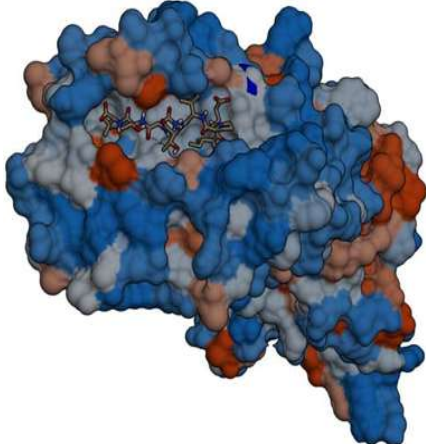
Tabla 11. Epítopes reconocidos por los alelos de clase I y II del HLA							
Nº	Epítope	Alelos clase I				Alelos clase II	
		A*68:02	A*02:06	B*15:04	B*39:05	DRB*04:07	DRB*16:02
1	AVSPVTGLV	X	x				
2	NQFSNTVV		x	x	x		
3	VIDPVTNTV	X	x				
4	SQASSTYAV		x	x			
5	VIDPNTNTV	X	x				
6	VIDPATNTV	X	x				
7	FVLALSQAS					x	x
8	RFVLALSQA					x	x
9	FVNVAPQLV	X	x			x	x
10	YQAISAQVA			x		x	

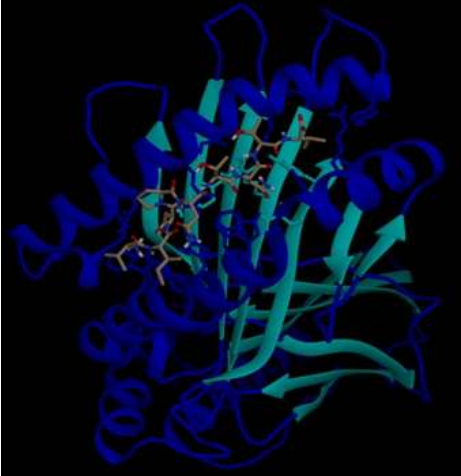
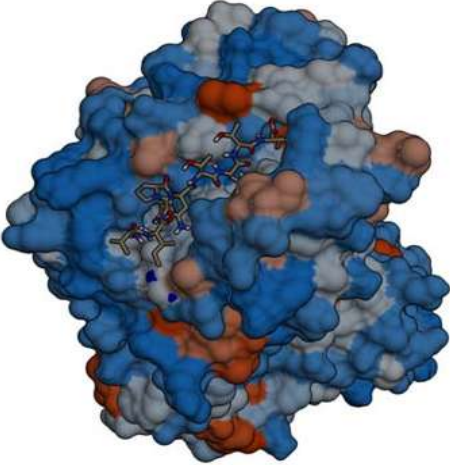
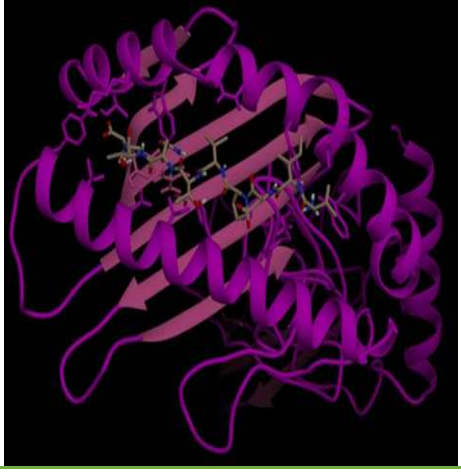
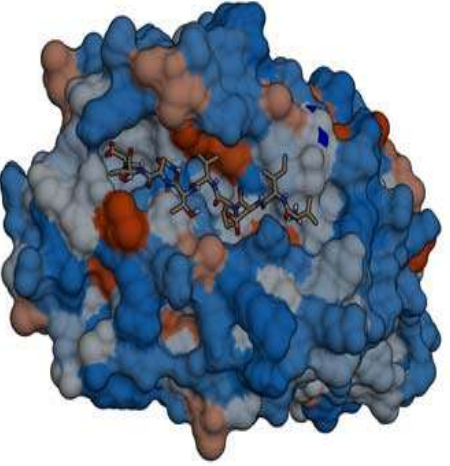
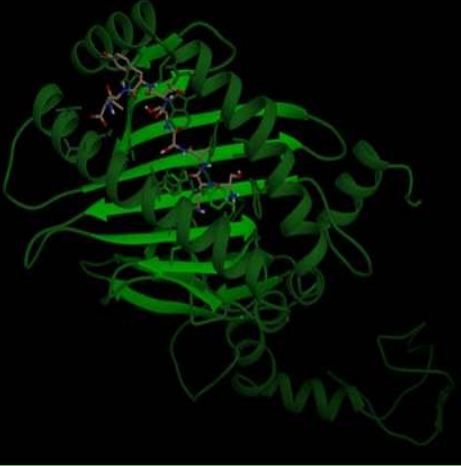
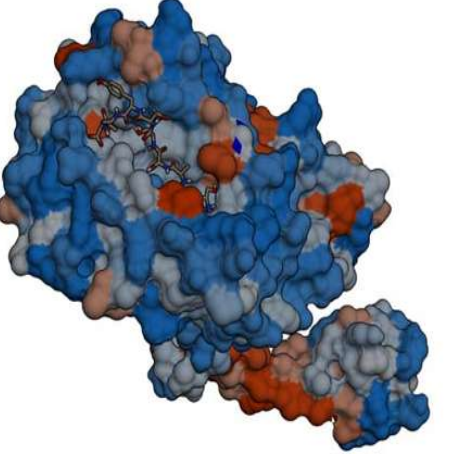
A*68:020101, A*02:060101, B*15:040101, B*39:050101, DRB*04:070101, DRB*16:020101

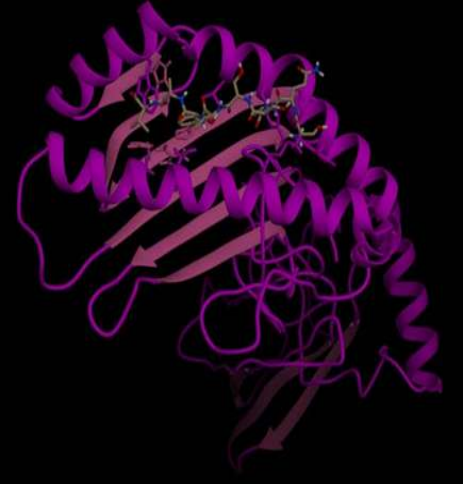
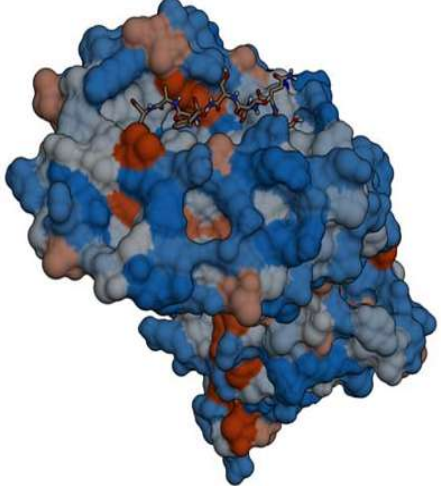
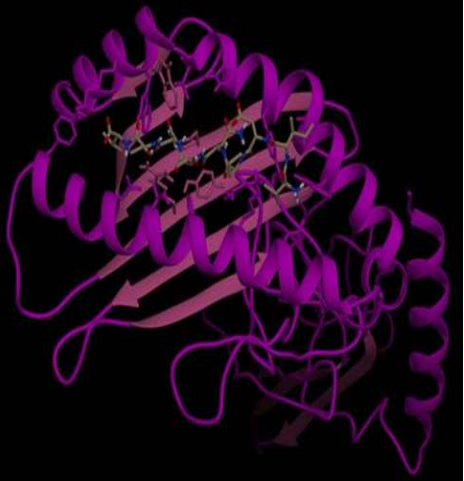
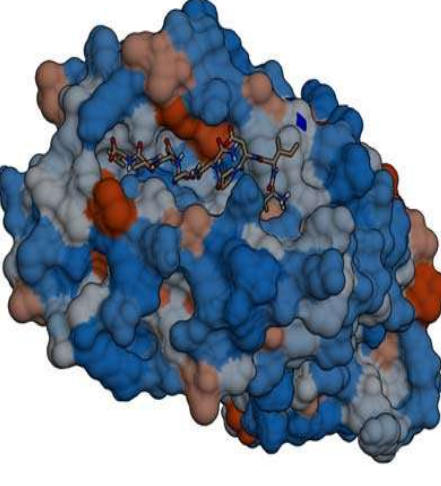
Tabla 12. Docking molecular entre los epítopes y alelos de clase I y II del HLA


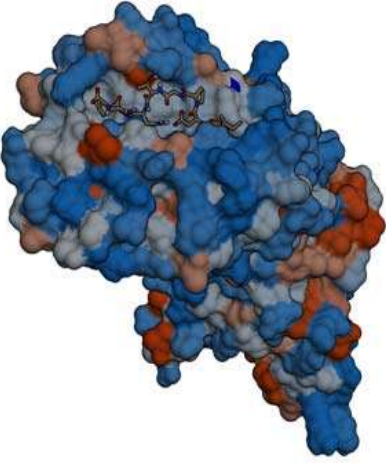

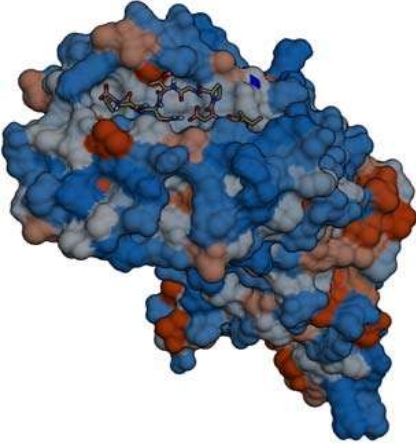
Modelado 3D en cintas	Modelado 3D en superficie hidrofóbica	Precisión estimada
<p data-bbox="235 338 574 373">A*68:020101-epítotope 1</p> 		1.0
<p data-bbox="235 890 574 926">A*02:060101-epítotope 1</p> 		1.0


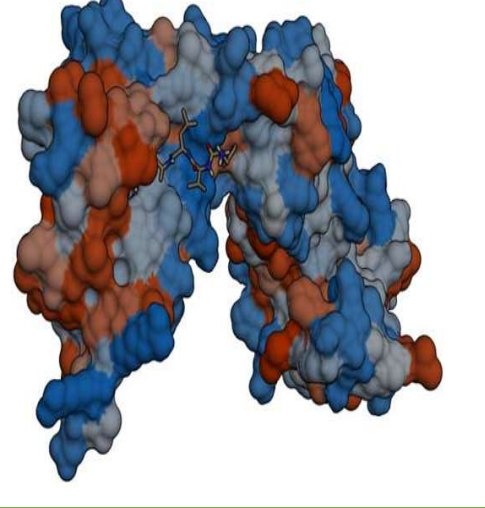
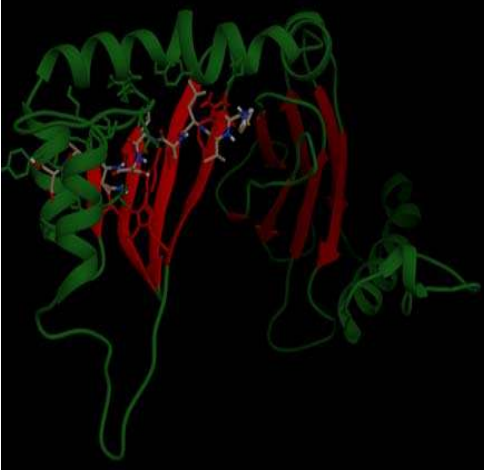
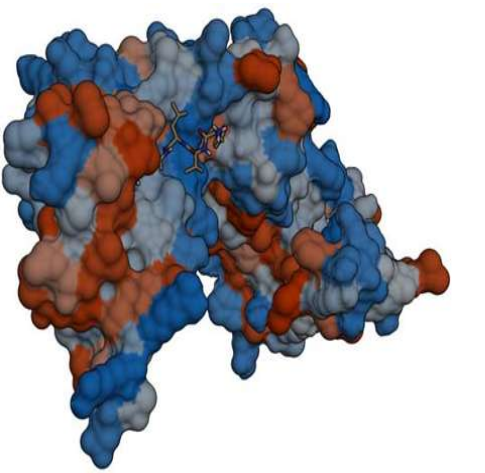
<p>A*02:060101-epítotope 2</p> 		<p>1.0</p>
<p>B*15:040101-epítotope 2</p> 		<p>0.923</p>

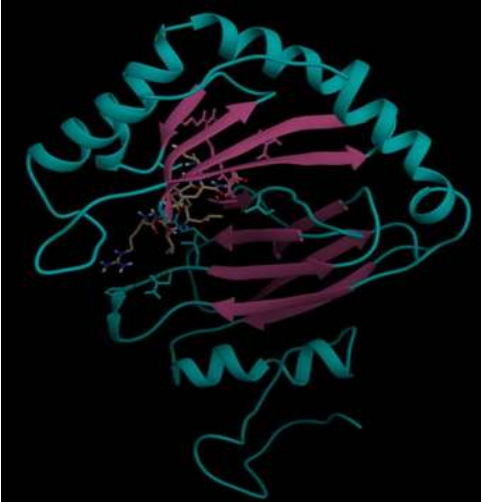
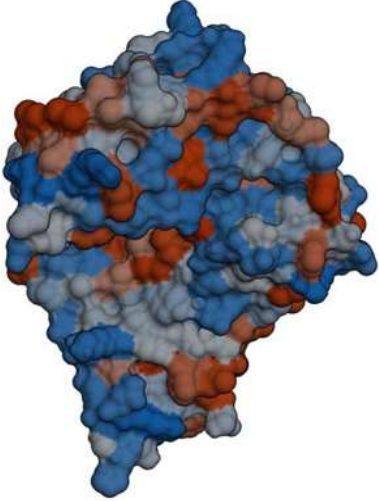

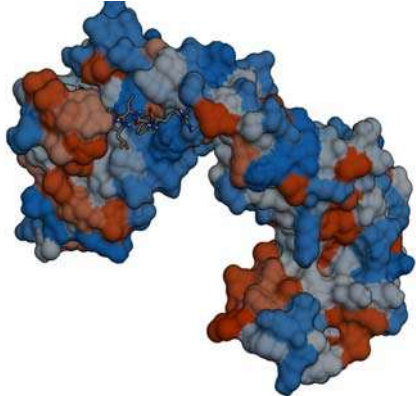
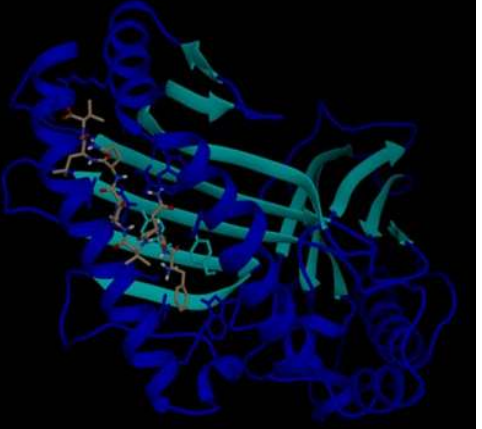
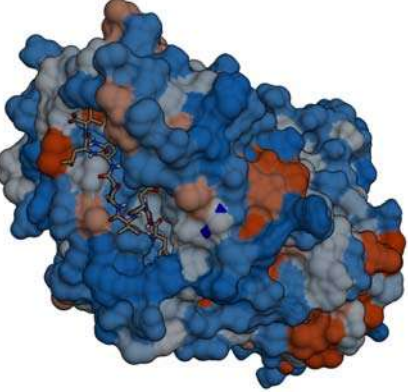
<p>B*39:050101-epítpe 2</p>  A ribbon diagram of the B*39:050101-epítpe 2 protein structure. The protein is shown in shades of green and cyan, with a complex, multi-domain structure. The background is black.	 A surface representation of the B*39:050101-epítpe 2 protein structure. The surface is colored in shades of blue, orange, and white, highlighting different regions. The protein is shown in a perspective view.	<p>0.938</p>
<p>A*68:020101-epítpe 3</p>  A ribbon diagram of the A*68:020101-epítpe 3 protein structure. The protein is shown in shades of blue and cyan, with a complex, multi-domain structure. The background is black.	 A surface representation of the A*68:020101-epítpe 3 protein structure. The surface is colored in shades of blue, orange, and white, highlighting different regions. The protein is shown in a perspective view.	<p>1.0</p>


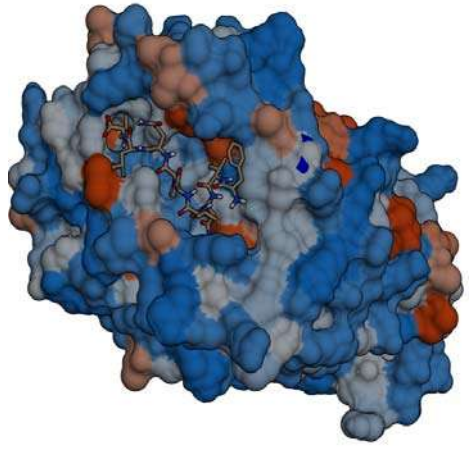

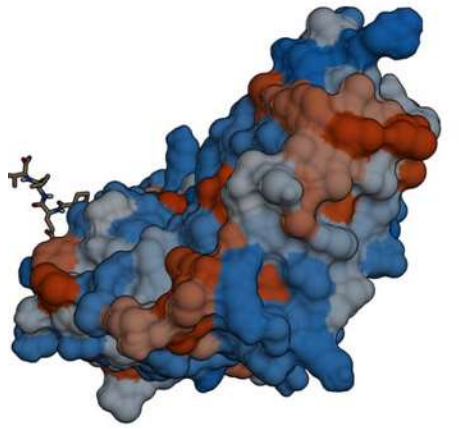
<p>A*68:020101-epítoto 5</p> 		<p>1.0</p>
<p>A*02:060101-epítoto 3</p> 		<p>1.0</p>
<p>B*15:040101-epítoto 4</p> 		<p>1.0</p>

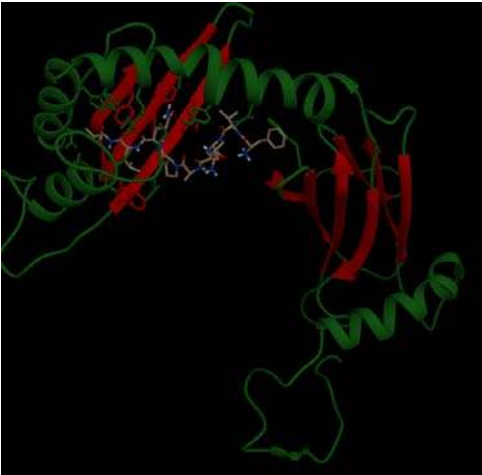
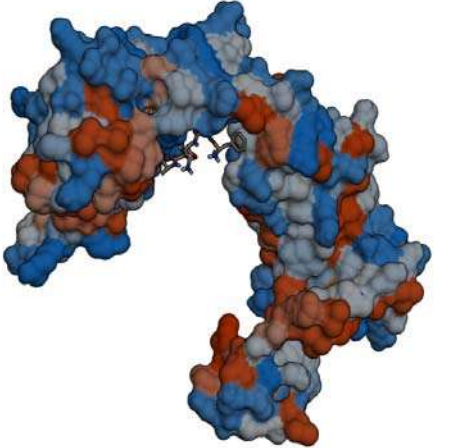

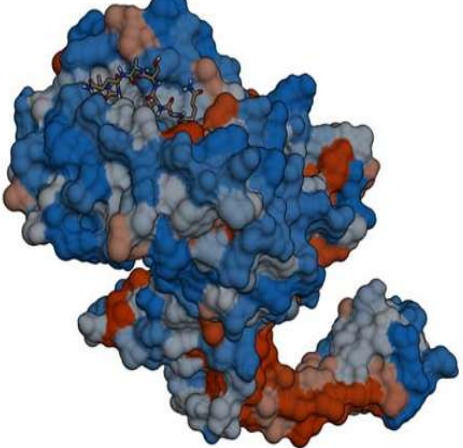

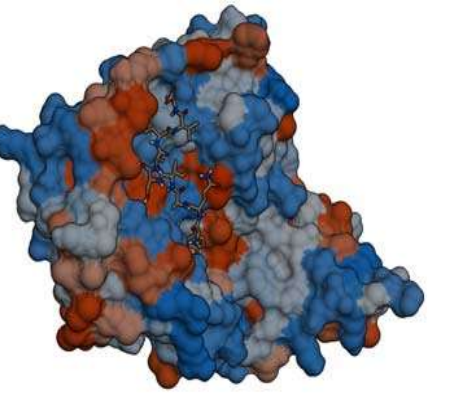
<p>A*02:060101-epítpe 4</p> 		<p>1.0</p>
<p>A*02:060101-epítpe 5</p> 		<p>1.0</p>

<p>A*68:020101-epítpe 6</p> 		<p>1.0</p>
<p>A*02:060101-epítpe 6</p> 		<p>1.0</p>

<p>DRB1*04:070101-epítpe 7</p>  A ribbon diagram of the DRB1*04:070101-epítpe 7 protein structure. The protein is shown in cyan and pink ribbons against a black background. The structure is a complex, multi-domain protein with several alpha-helices and beta-sheets.	 A surface representation of the DRB1*04:070101-epítpe 7 protein structure. The surface is colored in blue, orange, and white, showing the protein's interaction with an epitope (stick model) in the center.	<p>0.493</p>
<p>DRB1*16:020101-epítpe 7</p>  A ribbon diagram of the DRB1*16:020101-epítpe 7 protein structure. The protein is shown in green and red ribbons against a black background. The structure is a complex, multi-domain protein with several alpha-helices and beta-sheets.	 A surface representation of the DRB1*16:020101-epítpe 7 protein structure. The surface is colored in blue, orange, and white, showing the protein's interaction with an epitope (stick model) in the center.	<p>0.774</p>

<p>DRB1*04:070101-epítpe 8</p>  <p>A ribbon diagram of the DRB1*04:070101-epítpe 8 protein structure. The protein is shown in cyan and magenta ribbons against a black background. A small molecular model of the epitope is visible in the center of the protein structure.</p>	 <p>A surface representation of the DRB1*04:070101-epítpe 8 protein structure. The surface is colored in shades of blue, orange, and white, representing different chemical environments. A small molecular model of the epitope is shown bound to the protein surface.</p>	<p>0.503</p>
<p>DRB1*16:020101-epítpe 8</p>  <p>A ribbon diagram of the DRB1*16:020101-epítpe 8 protein structure. The protein is shown in green and red ribbons against a black background. A small molecular model of the epitope is visible in the center of the protein structure.</p>	 <p>A surface representation of the DRB1*16:020101-epítpe 8 protein structure. The surface is colored in shades of blue, orange, and white. A small molecular model of the epitope is shown bound to the protein surface.</p>	<p>0.809</p>
<p>A*68:020101-epítpe 9</p>  <p>A ribbon diagram of the A*68:020101-epítpe 9 protein structure. The protein is shown in blue and cyan ribbons against a black background. A small molecular model of the epitope is visible in the center of the protein structure.</p>	 <p>A surface representation of the A*68:020101-epítpe 9 protein structure. The surface is colored in shades of blue, orange, and white. A small molecular model of the epitope is shown bound to the protein surface.</p>	<p>1.0</p>

<p>A*02:060101-epítpe 9</p> 		<p>1.0</p>
<p>DRB1*04:070101-epítpe 9</p> 		<p>0.506</p>

<p>DRB1*16:020101-epítoto 9</p> 		<p>0.830</p>
<p>B*15:040101-epítoto 10</p> 		<p>0.972</p>
<p>DRB1*04:070101-epítoto 10</p> 		<p>0.502</p>

Probabilidad estimada=1.00 indica interacción del epítoto con el alelo correspondiente del HLA.

De acuerdo a la probabilidad estimada que indicó el servidor, se observó que la mayor parte de la interacción entre epítopes y alelos predichos previamente es buena, con una probabilidad estimada entre 0.700 a 1.0. Sin embargo, las interacciones DRB1*04:070101-epítope 7, DRB1*04:070101-epítope 8, DRB1*04:070101-epítope 9 y DRB1*04:070101-epítope 10 muestran una probabilidad estimada < 0.503. Como se observa en la tabla 12, los epítopes 7, 8, 9 y 10 no se acoplan de manera adecuada al sitio de unión del alelo DRB1*04:070101, por lo tanto, esta interacción predicha queda descartada para futuros experimentos.

8.6 Modelado 3D por homología de la proteína PE_PGRS18.

La secuencia de aminoácidos de la proteína PE_PGRS18-H37Rv, obtenida previamente, se utilizó para realizar el modelado 3D de dicha proteína mediante el servidor Phyre2 en modo intensivo. El servidor utilizó 120 moldes para el modelado de la proteína de estudio. En la tabla 13 se muestran algunos moldes que cubren la mayor parte de la secuencia, utilizados para la construcción del modelado 3D de la proteína, el % de confianza, el % de identidad (% ID) y la cobertura.

Tabla 13. Características de los moldes representativos utilizados para el modelado 3D por homología de la proteína PE_PGRS18-H37Rv.			
Molde	% Confianza	% ID	Cobertura
c5xfsA	100.0	43	7-83
d2g38a1	100.0	28	8-83
c2g38A	100.0	28	8-83
c1gniE	99.9	19	281-452
c6gc1A	99.9	26	280-457
c1fwbB	99.9	18	279-472
d1gnia2	99.9	15	279-448
c3sbrF	99.9	15	279-448
c5i5iA	99.9	20	286-447
c3hrpA	99.9	18	284-447

% Confianza: probabilidad de 0-100% de que la secuencia de estudio y el molde sean homólogos, %ID: proporción de residuos de la proteína de estudio equivalentes a residuos del molde idénticos en la alineación generada, Cobertura: cobertura de alineación del molde con respecto a la secuencia de estudio.

Posteriormente se verificó la calidad de la proteína mediante el servidor Rampage el cual indica que el 71.4% de los residuos se encuentran en regiones

favorables según los criterios de Ramachandra. Como la calidad de la proteína es muy baja, se realizó un refinamiento mediante el servidor Galaxyrefine, se eligió el modelo número cuatro con un GDT-HA de 0.9365; posteriormente se verificó la calidad del modelado el cual indicó un 87.7 % de residuos en regiones favorables. La calidad del modelo mejoró en un 16.1% (Fig. 12).

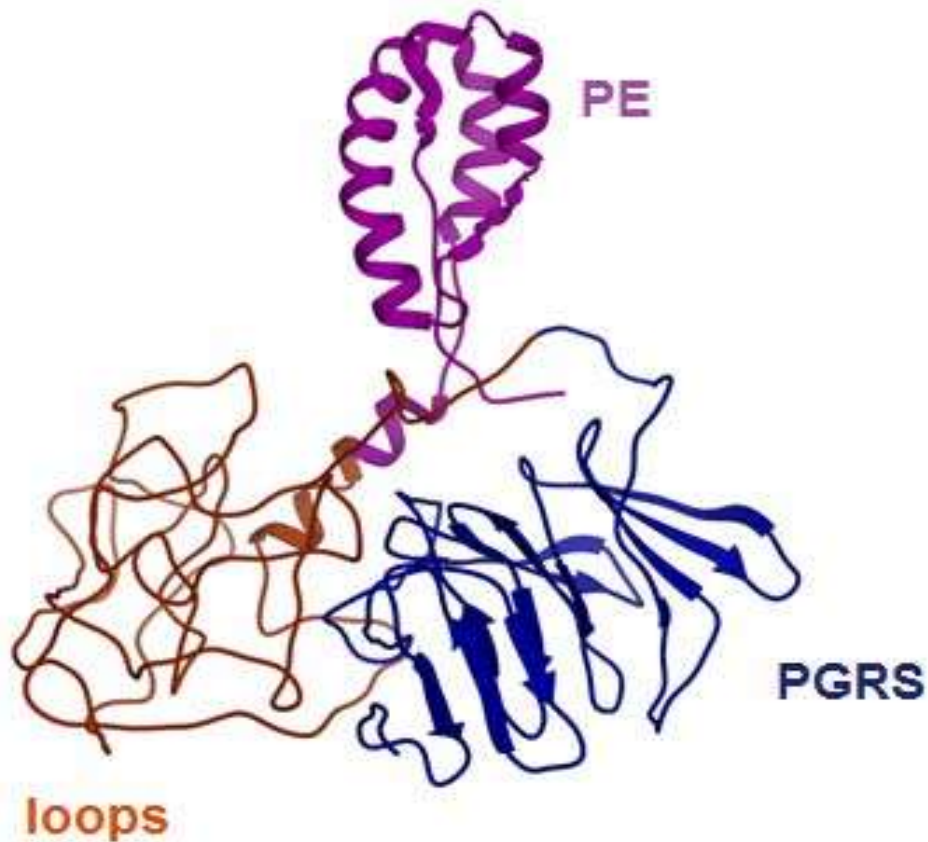


Figura 12. Modelado 3D de la proteína PE_PGRS18 H37Rv. Lila= dominio PE, azul= dominio PGRS y naranja= loops o regiones desordenadas.

En seguida se realizó el modelado 3D por homología de la proteína PE_PGRS18 de seis de las cepas de estudio las cuales comparten mutaciones con distintas cepas (Tabla 14). Para ello se utilizó el servidor Robetta ya que permite realizar el modelado 3D por homología de cada cepa utilizando como molde el modelado de la proteína PE_PGRS18 H37Rv

previamente obtenido. Posteriormente se verificó la calidad de los modelos (tabla 14) en la que se observan porcentajes de residuos en regiones favorables mayores al 90%.

Tabla 14. Cepas modeladas 3D con mutaciones frecuentes en distintas cepas y criterios de validación de la calidad de cada modelo.		
Cepas (MYC) modeladas 3D	Cepas que presentan las mismas mutaciones	% Residuos en regiones favorables
047	048, 064, 068, 069, 076, 087, 089, 090, 091, 093, 097, 108, 109, 112, 116, 100, 101, 103, 106, 107, 113, 114, 115, 117, 120	92.2
051	052, 053, 058, 061, 062, 070, 096	92.1
055	078	92.5
065	80, 86	92.5
071	79	92.0
098		90.6

8.6.1 Comparaciones globales de la estructura tridimensional

Una vez obtenidos los modelos se realizó un análisis comparativo global y local entre las estructuras obtenidas de cada una de las cepas y la estructura de referencia PE_PGRS18 H37Rv, con el objetivo de comparar si los cambios presentes en cada una de las secuencias de las cepas provocan un cambio en la estructura de la proteína, utilizando como control el modelo de 3D de la proteína de referencia PE_PGRS18 H37Rv. En la figura 13 se observa la superposición de las estructuras de las cepas con respecto a la de referencia.

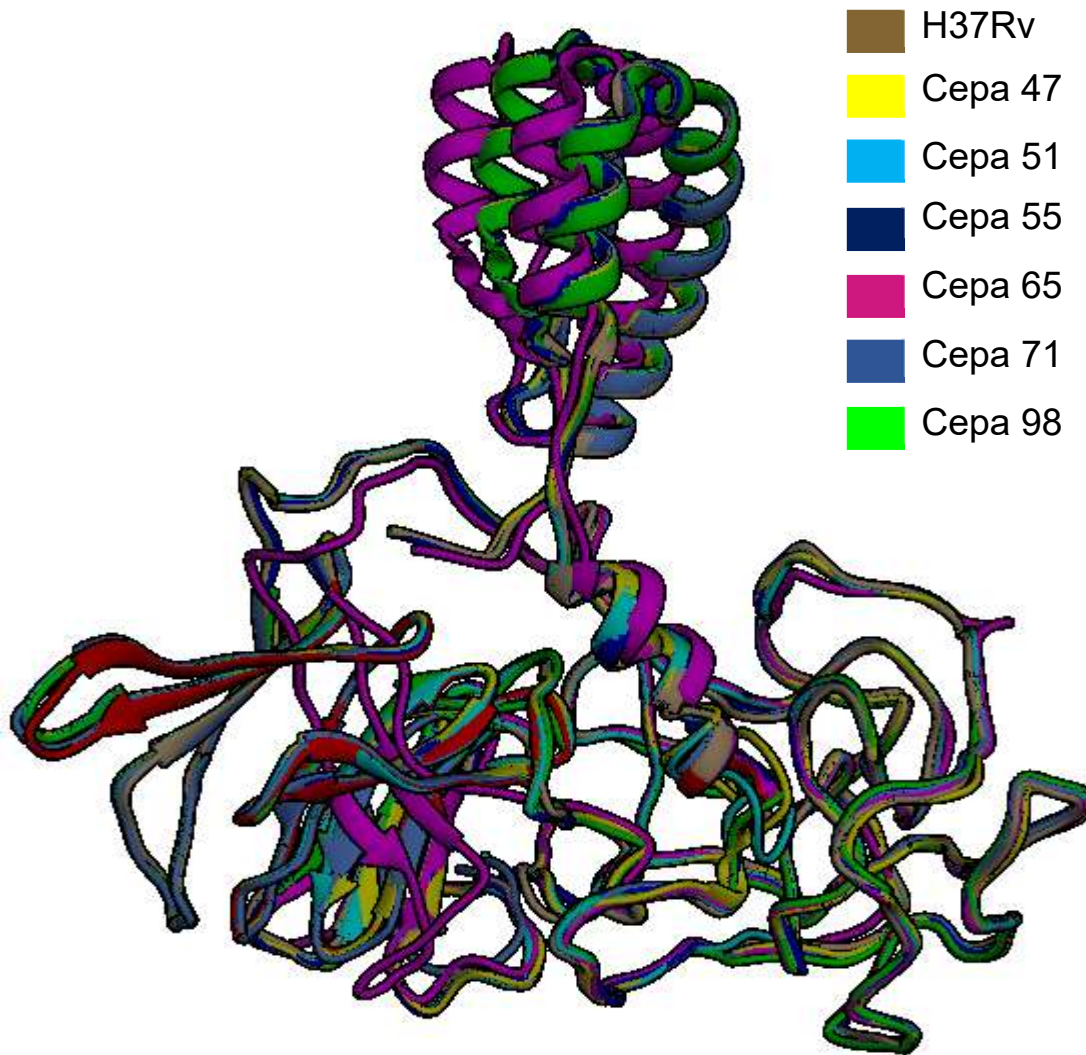


Figura 13. Superposición de las estructuras de las cepas modeladas y la estructura de referencia PE_PGRS18 H37Rv. Los distintos colores mostrados en la superposición representan la estructura específica de cada una de las cepas mostradas en el panel del lado izquierdo.

Se observa que las cepas MYC 047, 051, 055, 071 y 098 presentan una mayor similitud a la estructura de referencia con valores de RMSD similares (cerca de cero). Sin embargo, la cepa MYC 065 presenta cambios en su estructura comparada con la estructura de referencia, con valores de RMSD (alejados de cero), MYC 065= 0.819.

8.7 Docking molecular entre la proteína PE_PGRS18 y los receptores TLRs 2 y 4

Una vez obtenido el modelo de la cepa de referencia y de las cepas de estudio, además de haber validado la calidad de cada modelo se realizó el docking molecular con el modelo de la estructura de referencia PE_PGRS18 H37Rv y los TLRs 2 y 4, mediante el servidor Prism2.0. Para ello se realizó la búsqueda de las estructuras cristalizadas de los TLRs 2 y 4 en la base de datos PDB. Sin embargo, las estructuras estaban unidas a un ligando y el modelado estaba incompleto, por ello se decidió realizar el modelado 3D por homología de cada receptor mediante el servidor Phyre2. Posteriormente se realizó el refinado de cada modelo y se verificó la calidad mediante los servidores mencionados en materiales y métodos (Tabla 15).

Tabla 15. Criterios de refinamiento y validación de la calidad de los TLRs 2 y 4			
TLRS	Galaxyrefine		Rampage
	# Modelo	GDT-HA	% Residuos en regiones favorables
2	2	0.9477	92.7
4	4	0.9511	91.6

GDT-HA= Prueba de distancia global de modelos basados en templados de alta precisión

Los modelos 3D de los TLRs 2 y 4 refinados y validados se observan en la figura 14. No se observó interacción entre la proteína de estudio y los TLRs 2 y 4.

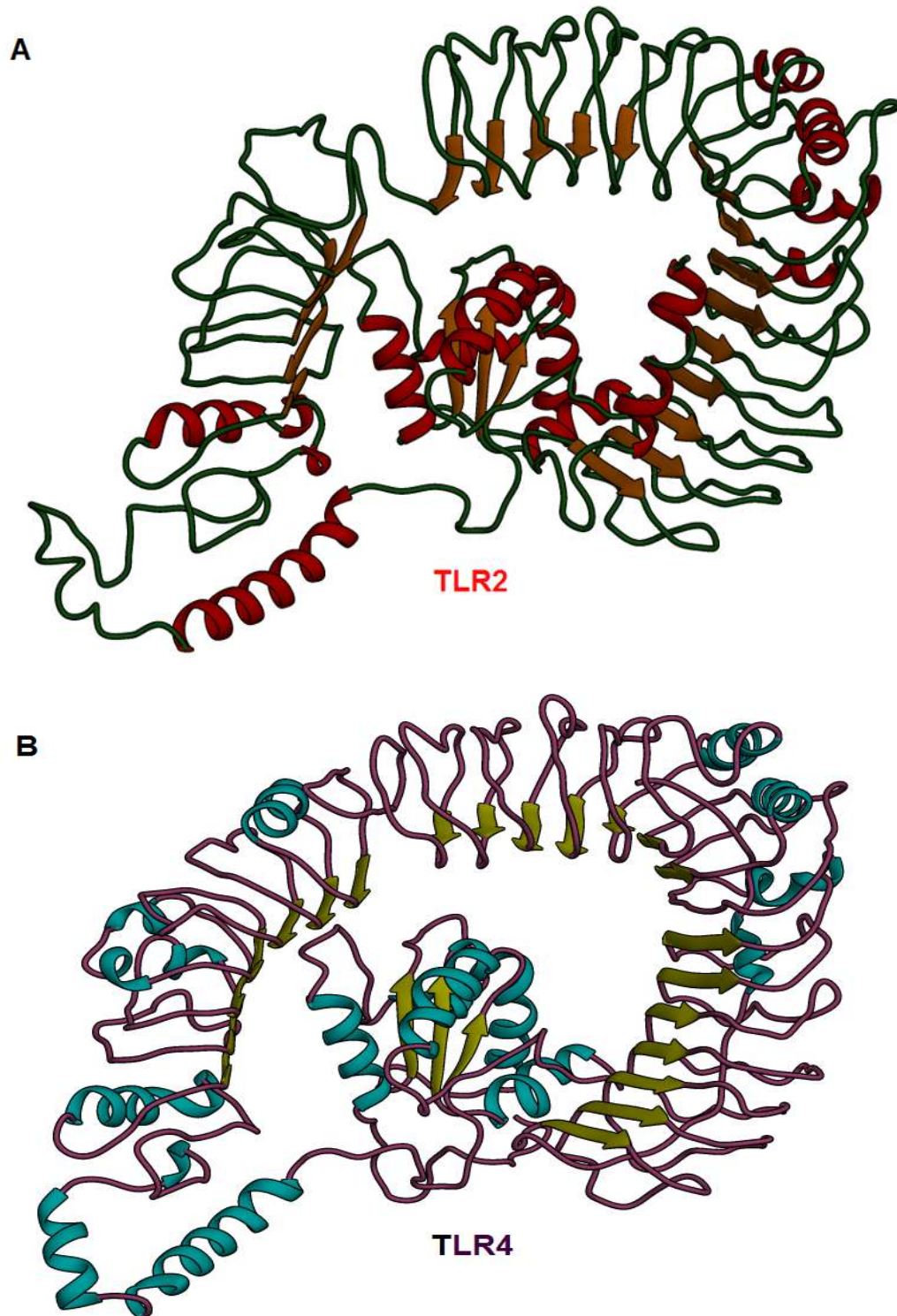


Figura 14. Modelados 3D de los receptores TLR2 y TLR4. A) receptor TLR2 las estructuras de color rojo= hélices α , naranja= laminas β y verde= loops. B) receptor TLR4 las estructuras de color azul= hélices α , amarillo= laminas β y rosa= loops.

8.8 Predicción de la interacción de la proteína PE_PGRS18 H37Rv

Ya que no se encontró interacción de la proteína de referencia con los TLRs 2 y 4, se predijo la interacción de la proteína PE_PGRS18 H37Rv con otras proteínas mediante el servidor String (<http://string-db.org>), con la finalidad de identificar si la proteína PE_PGRS18 es capaz de interactuar con otra proteína para ejercer alguna función.

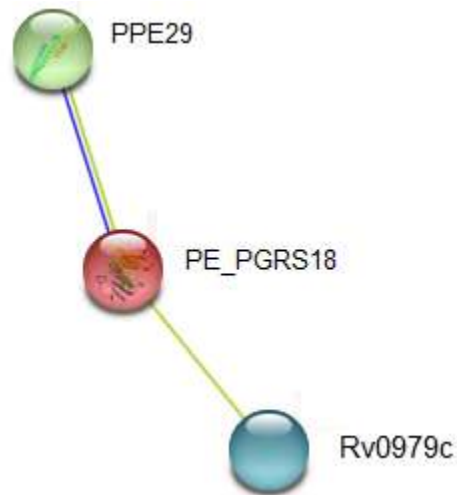


Figura 15. Interacción de la proteína PE_PGRS18

La búsqueda mediante un mapeo de interacción de dominios de las proteínas que interactúan con la proteína objetivo de *M. tuberculosis* (Rashid *et al.*, 2010), indicó que la proteína PE_PGRS18 interactúa con las proteínas PPE29 (Rv1801) y PE (Rv0979c) (Fig.15). Posteriormente se buscó la función específica de estas proteínas mediante el servidor Mycobrowser (<https://mycobrowser.epfl.ch/>). Sin embargo, ambas proteínas presentan una función desconocida.

9. DISCUSIÓN

La tuberculosis es una enfermedad infecciosa, prevenible y curable, que cada año causa la muerte de aproximadamente 2 millones de personas en todo el mundo, por lo que ha sido nombrado el patógeno humano más exitoso (OMS, 2018). La circulación de cepas de *Mycobacterium tuberculosis* multidrogorresistentes y extremadamente drogorresistentes, además de la epidemia de VIH-SIDA, disminuyen en gran medida la capacidad de control y la efectividad del tratamiento de la enfermedad, haciendo necesaria la búsqueda de nuevas técnicas de diagnóstico, la creación de fármacos exitosos y un mejor entendimiento del estilo de vida intracelular de este patógeno.

Por la evidencia *in vitro* que apoya la asociación de la familia de proteínas PE/PPE y la subfamilia PE_PGRS con la micomembrana o la interacción hospedero-patógeno, se decidió estudiar mediante herramientas bioinformáticas a la proteína PE_PGRS18 cuyo gen resultó ser el más polimórfico en cepas de Michoacán (Frutos-Hernández, 2015). La propuesta es que algunos de los polimorfismos son capaces de permitir o interrumpir el reconocimiento de los receptores TLR2 y TLR4 con relación a la respuesta inmune innata, y que además son capaces de estimular o evadir la respuesta inmune adaptativa mediada por los linfocitos T CD4+ y CD8+. Para ello, se determinó inicialmente la localización y posible función de la proteína PE_PGRS18 de la cepa de referencia H37Rv de *M. tuberculosis*.

La subfamilia de proteínas PE_PGRS tienen una longitud que va de 300 a 600 aa, como las proteínas PE_PGRS33, PE_PGRS29, PE_PGRS35 y PGRS17 que tienen una longitud de 498 aa, 570 aa, 331 aa y 370 respectivamente, incluyendo ahora la PE_PGRS18 que tiene una longitud de 457 aa.

La proteína PE_PGRS18 tiene una posible estructura secundaria constituida por hélices (hélice α , pi- y 310) en un 16.85 %, láminas β en un 40.04% y bucles (*loops*) en un 43.5%. Estos resultados proponen que la mayor parte de la estructura de la proteína PE_PGRS18 está constituida de *loops* y láminas β . Dentro de los *loops* se encontraron posibles regiones flexibles y desordenadas. Lo anterior coincide con lo reportado para las proteínas de la subfamilia PE_PGRS, las cuales

tienen un mayor número de regiones desordenadas comparadas con otras familias de proteínas patógenas de *M. tuberculosis* (Ahmad *et al.*, 2019).

Las regiones flexibles y desordenadas pueden adoptar diferentes conformaciones para interactuar con distintos ligandos y otras biomoléculas, conservando su especificidad (Papaleo *et al.*, 2016). Lo anterior es interesante porque dichas regiones se encuentran distribuidas en los dominios de la proteína PE_PGRS18 y la región no estructural. Nuestros resultados sugieren que la proteína PE_PGRS18 posiblemente ejerce una variedad de funciones al interactuar con algún ligando.

Ahmand y colaboradores (2019) proponen que es probable que *M. tuberculosis* utilice la dinámica estructural orden-desorden como estrategia para establecer una respuesta pro-patógena y subvertir la defensa del hospedero para una infección productiva. Esta ganancia funcional también sirve como un medio para compensar la pérdida de contenido genómico debido a la evolución reductiva.

La proteína PE_PGRS18 carece de puentes disulfuro, sugiriendo que tiene una estructura terciaria inestable ya que los enlaces disulfuro desempeñan un papel importante en la estabilización del plegamiento de la estructura terciaria. La predicción de puentes disulfuro es útil para el estudio de las propiedades estructurales y funcionales de proteínas específicas (Ceroni *et al.*, 2006). Los resultados de accesibilidad al solvente y la predicción de la localización y función de la proteína PE_PGRS18 indicaron que se trata de una proteína integral de membrana con una región expuesta en la superficie celular y que posiblemente sea una proteína de unión a proteínas o a complejos proteicos. Delugu y colaboradores (2004) reportaron que la proteína PE_PGRS33 codificada por RV1818c utiliza el sistema de secreción *esx 5* y se encuentra expuesta en la superficie celular, tiene regiones hidrofóbicas que influyen en la estructura de las células bacterianas. Por otro lado, Sakamuri y colaboradores (2019), tras un análisis *in silico* e *in vivo* de la proteína PE_PGRS51, reportaron que dicha proteína también se localiza en la superficie celular. Estos reportes concuerdan con los resultados obtenidos en

nuestro estudio, aunque la función específica de la proteína PE_PGRS18 sigue siendo desconocida.

También se encontró un sitio de unión a un péptido señal de 19 aa localizada en la posición 38-56. Daleke y colaboradores (2011) reportan que los dominios conservados PE y PPE dirigen la lipasa (lypY) de las micobacterias patógenas *M. tuberculosis* y *M. marium*, a la superficie celular a través de la vía ESX 5. Además, indican que cuando las proteínas lypY llegan a la superficie celular, pierden el dominio PE y PPE, por lo que proponen que en dichos dominios se encuentra una secuencia señal que permite el transporte de estas proteínas. Sakamuri y colaboradores (2019), tras realizar un análisis bioinformático de la proteína PE_PGRS51, reportaron la presencia de un péptido señal en el extremo N-terminal. Con el resultado obtenido en este trabajo se corrobora la presencia de una posible secuencia señal en el dominio PE que probablemente dirige a la proteína PE_PGRS18 mediante el sistema de secreción ESX 5 a la superficie celular. Sería conveniente corroborar *in vitro*, en análisis posteriores, la presencia del péptido señal para conocer su función.

Resulta interesante que las 17 regiones identificadas de unión a proteína, se encuentren en la parte hidrofóbica de la proteína PE_PGRS18. Ates y colaboradores (2018) mostraron que mutaciones en el gen *ppe38* bloquean completamente la secreción por el sistema de secreción tipo V (ESX-5) de las proteínas PE_PGRS y PPE-MPTR, provocando un incremento de la virulencia de las cepas de *M. tuberculosis*. Ellos proponen que es posible que las proteínas PPE38 tengan la función de chaperonas secretadas, quizás ayudando en la solubilidad de los dominios hidrofóbicos PGRS y MPTR.

Nuestros resultados sugieren que probablemente la proteína PPE38, alguna otra proteína o un complejo proteico, se une a la proteína PE_PGRS18 en una de estas regiones de unión, quizá ayudando a la solubilidad del dominio PGRS (Ates *et al.*, 2018), contribuyendo al reconocimiento por las células T y TLRs.

Con respecto al análisis de epítopes en la proteína PE_PGRS18 que son reconocidos por alelos de clase I y II del HLA que representan la diversidad de la

población mundial, se identificaron 2,191 epítopes de alta afinidad en el dominio PE, 2, 654 en el dominio PGRS y 269 en la región no estructural. Siendo el alelo A_6802 de HLA perteneciente a la región mexicana el que reconoce un mayor número de epítopes. Con respecto al análisis de epítopes de clase II del HLA, se encontraron 891 epítopes de alta afinidad en el dominio PE, 1 en el PGRS y 103 en la región no estructural. Siendo el alelo DRB_0407 del HLA el cual pertenece a la región mexicana el que reconoce un mayor número de epítopes.

Ya que no hay reportes en la literatura que indiquen específicamente la capacidad de la proteína PE_PGRS18 para estimular la respuesta inmune adaptativa mediada por los linfocitos T CD4+ y CD8+, en este trabajo se buscaron epítopes de la proteína PE_PGRS18 de las 46 cepas de estudio reconocidos por alelos de clase HLA clase I y II. Los resultados aquí encontrados contradicen lo reportado por Copin y colaboradores (2014), quienes reportan que en el dominio PE altamente conservado hay un número mayor de epítopes para células T CD4+ y CD8+, y que la variación en el dominio PGRS es independiente del reconocimiento de células T humanas, por lo que la variación genética en el dominio PGRS está relacionado con la evasión de la respuesta inmune del hospedero.

En el presente análisis, la variación genética en el dominio PGRS de la proteína PE_PGRS18 estimula principalmente la respuesta inmune mediada por los linfocitos T CD8+ y el dominio PE la mediada por los linfocitos T CD4+ principalmente. La diferencia entre nuestros resultados y los obtenidos por Copin y colaboradores (2014), puede explicarse por el origen de las secuencias analizadas. La proteína PE_PGRS18 tiene regiones únicas que dependen de la región geográfica a la que pertenece cada cepa. Las 64 proteínas estudiadas por Copin y colaboradores provienen de cepas de referencia elegidas de una colección mundial del CMTB y de pacientes nepalenses, mientras que las proteínas PE_PGRS18 estudiadas en el presente trabajo pertenecen a cepas mexicanas de los estados de Michoacán y Querétaro.

Por otra parte, Delugu y Brennan (2001) reportan, tras la inmunización con ratones, que el dominio PE de la proteína PE_PGRS33 induce una buena respuesta

inmune celular y el dominio PGRS es responsable de inducir una buena respuesta inmune humoral. Mientras tanto Sakamuri y colaboradores (2019) observaron la presencia de anticuerpos anti-UR contra epítopes en el dominio PE de la proteína PE_PGRS51 en las distintas etapas de la tuberculosis.

De acuerdo al presente análisis, el dominio PE es capaz de estimular una buena respuesta inmune celular mediada principalmente por los linfocitos T CD4+ y posiblemente una respuesta inmune humoral; el dominio PGRS es capaz de estimular una buena respuesta inmune celular mediada principalmente por los linfocitos T CD8+. Lo anterior indica que la proteína PE_PGRS18 tras estudios posteriores puede formar parte de una buena vacuna contra la tuberculosis ya que es capaz de estimular la respuesta inmune celular mediada por los linfocitos T CD4+ como lo hace la vacuna BCG (Bacilo de Calmette y Guérin). Además, estimula la respuesta inmune mediada por los linfocitos T CD8+ que la BCG no estimula y posiblemente la respuesta inmune humoral.

La comparación estadística del número de epítopes en los dominios de la proteína PE_PGRS18 que son reconocidos por alelos de clase I (células T CD8)+ de la región mexicana y distintas regiones del mundo, indicó valores de $p > 0.05$, es decir no se encontró una relación estadísticamente significativa. Sin embargo, la comparación estadística del número de epítopes en los dominios de la proteína PE_PGRS18 que son reconocidos por alelos de clase II (células T CD4+) de la región mexicana y distintas regiones del mundo se obtuvo una $p < 0.05$ en los epítopes del dominio PE. Esto indica que los alelos de clase I pertenecientes a la región mexicana y a distintas regiones del mundo reconocen el mismo número de epítopes de las proteínas PE_PGRS18 de las cepas de estudio.

Estos datos son interesantes ya que distintos alelos del mundo son capaces de reconocer epítopes de clase I de las proteínas PE_PGRS18 de las cepas de Michoacán y Querétaro. Tras análisis posteriores se podrían verificar estos resultados aumentando el número de cepas de Michoacán, Querétaro y otros estados de México en general e incluso el número de alelos frecuentes en distintas regiones del mundo, para verificar los datos obtenidos en este estudio tanto *in silico*

como *in vitro*. Corroborado lo anterior, podría proponerse a la proteína PE_PGRS18 como un buen candidato vacunal.

Los resultados obtenidos en este trabajo indican que los alelos de clase II de la región mexicana reconocen un mayor número de epítopes, exclusivamente los que se encuentran en el dominio PE de la proteína PE_PGRS18 de las cepas de estudio comparado con los alelos de clase II pertenecientes a distintas regiones del mundo. Esto es interesante porque el dominio PE es altamente conservado y nuestras cepas pertenecen a los estados de Michoacán y Querétaro y los epítopes que se encuentran en este dominio son reconocidos por los alelos exclusivamente mexicanos. No hay reportes en la literatura sobre epítopes de proteínas de *M. tuberculosis* o sobre epítopes encontrados en la proteína PE_PGRS18 pertenecientes a cepas de pacientes de Michoacán y Querétaro.

Lo anterior nos permite proponer que los epítopes encontrados en el dominio PE de la proteína de estudio, podrían estimular principalmente la respuesta inmune mediada por los linfocitos T CD4+ reconocidos por alelos de la región mexicana, y que el dominio PGRS es capaz de estimular la respuesta inmune mediada por las células T CD8+ y T CD4+ reconocidos por alelos de México y distintas regiones del mundo.

El docking molecular realizado con los epitopes y alelos seleccionados corroboró la eficiencia del servidor NetMHCpan y NetMHCpan II ya que se observó interacción de los epítopes en la región de unión de los distintos alelos. Sin embargo, los epítopes que indicó el servidor que interaccionan con el alelo DRB*0407 mostraron una precisión estimada cercana a cero y la interacción no se observa en el sitio de unión, por lo que en análisis posteriores esta interacción queda descartada.

Por otro lado, se encontraron 332 mutaciones que estimulan la respuesta inmune mediada por los linfocitos T CD4+ y CD8+. Así mismo se identificaron 88 mutaciones que provocan la evasión de dicha respuesta. El análisis de distribución reveló 11 mutaciones en el dominio PE capaces de estimular la respuesta inmune celular y 11 mutaciones que evaden esta respuesta. Además en el dominio PGRS

se encontraron 144 mutaciones capaces de estimular la respuesta inmune celular y 5 que la evaden. En la región no estructural se identificaron 80 mutaciones capaces de estimular la respuesta inmune celular y 72 mutaciones que la evaden. El 89% de estas mutaciones son SNPs.

Se encontró una relación estadísticamente significativa ($p < 0.05$) entre los dominios PE – PGRS y los dominios PGRS- región no estructural. Estos resultados indican que los 11 SNPs en el dominio PE son capaces de evadir la respuesta inmune mediada por las células T en comparación con los SNPs encontrados en el dominio PGRS. Sin embargo, los 127 SNPs encontrados en el dominio PGRS son capaces de estimular la respuesta inmune de células T en comparación con los SNPs encontrados en el dominio PE.

Los 127 SNPs encontrados en el dominio PGRS son capaces de estimular la respuesta inmune de células T en comparación con los SNPs encontrados en la región no estructural. Mientras tanto los 62 SNPs pertenecientes a la región no estructural son capaces de evadir la respuesta inmune mediada por células T en comparación con el dominio PGRS.

Esto corrobora que el dominio PE es altamente conservado y el dominio PGRS polimórfico (Fishbein *et al.*, 2015). Con lo anterior proponemos que los SNPs encontrados en el dominio PE y en la región no estructural evitan el reconocimiento por las células T CD4+ y CD8+; mientras tanto, los SNPs encontrados en el dominio PGRS estimulan el reconocimiento por las células T CD4+ y CD8+. Los SNPs que permiten la evasión de la respuesta inmune, mediada por células T en el dominio PE y la región no estructural, pueden ser un mecanismo que la bacteria está utilizando para evitar su reconocimiento por el sistema inmune, aumentar su sobrevivencia en el hospedero y su infectividad.

El hecho de que la proteína PE_PGRS18 presente un gran número de mutaciones y que algunas de ellas sean reconocidas por el HLA, podría indicar que *M. tuberculosis* desarrolló este mecanismo para sobrevivir en el hospedero. Por una

parte, la bacteria ha tenido que evolucionar para poder resistir al sistema inmune del huésped, pero manteniendo su capacidad de transmitirse en la población y no extinguirse. Los cambios evolutivos en el patógeno para aumentar su infectividad han sido contrarrestados por cambios evolutivos en la respuesta inmune del huésped para incrementar la resistencia a la infección (co-evolución o evolución conjunta) (Brites and Gagneux, 2015).

Ates y colaboradores (2018) sugieren un papel importante para las proteínas PE_PGRS y PPE_MPTR en reducir la virulencia de *M. tuberculosis* y de este modo permitir la supervivencia prolongada del huésped infectado. Además, Sakamuri y colaboradores (2019) observaron la presencia de distintas proteínas PE_PGRS en distintas etapas de la infección por tuberculosis.

Los resultados obtenidos en este trabajo y lo reportado en estudios previos sugieren que la proteína PE_PGRS18 podría disminuir la infección de *M. tuberculosis* estimulando la respuesta inmune celular mediada por células T y posiblemente la respuesta inmune humoral. Sin embargo, se requiere de estudios posteriores para confirmar estos resultados.

Con respecto a la respuesta inmune innata, se obtuvo el modelado 3D de la proteína PE_PGRS18 de la cepa H37Rv y de seis de las cepas de estudio ya que estas proteínas comparten mutaciones en distintas cepas. La superposición reveló una modificación de la estructura de la proteína PE_PGRS18 en la cepa 65, correspondiente a una eliminación de 78 aa.

El docking molecular indicó que no hay una interacción directa entre la proteína PE_PGRS18 H37Rv y los receptores TLR2 y TLR4. Este análisis también se realizó con la proteína PE_PGRS18 de la cepa 65 para corroborar si los cambios en su estructura podrían modificar la interacción con los TLR2 y TLR4. Sin embargo, no se observó ninguna interacción. Esto puede ser porque de acuerdo a nuestros resultados, probablemente la proteína PE_PGRS18 requiera unirse a una proteína o a un complejo proteico para ser reconocido por los TLRs. A este respecto, Ates y colaboradores (2018) reportaron que las proteínas PE_PGRS dependen de las proteínas PPE38 para ser secretadas por el ESX V. Otra de nuestras hipótesis es

que la proteína PE_PGRS18 no es reconocida por la respuesta inmune innata ya que se ha reportado la expresión de la subfamilia de proteína PE_PGRS cuando la Tb se encuentra activa (Kruh *et al.*, 2010).

La predicción de la interacción de la proteína PE_PGRS18 con otras proteínas indicó que probablemente interacciona con las proteínas PPE29 (Rv1801) y PE (Rv0979c) para ejercer alguna función aún desconocida.

De acuerdo a nuestros resultados y lo reportado por Ates y colaboradores (2018), proponemos que probablemente la proteína PE_PGRS18 interacciona con otra proteína, quizá de esta misma familia de proteínas para poder ser reconocida por estos receptores; o bien, la proteína PE_PGRS18 no es reconocida por la respuesta inmune innata ya que, tras los análisis mencionados anteriormente, sólo se ha visto la expresión de esta subfamilia de proteínas cuando la tuberculosis se encuentra activa (Kruh *et al.*, 2010).

10. RESUMEN DE RESULTADOS

1. En relación a la respuesta inmune adaptativa se encontraron epítopes que son reconocidos por alelos de clase I y II del HLA representativos de distintas regiones del mundo en la proteína PE_PGRS18 de las distintas cepas de *M. tuberculosis* incluidas en este estudio.
2. esta proteína se puede probar en estudios futuros en formar parte de una vacuna contra *M. tuberculosis*.
3. Los polimorfismos encontrados en las secuencias de las proteínas PE_PGRS18 pueden alterar el reconocimiento de un epítope es decir, que este sea reconocido por los distintos alelos (estimule la respuesta inmune) o que no sea reconocido (evada la respuesta inmune).
4. Con respecto a la respuesta inmune innata no se observó interacción de la proteína PE_PGRS18 con los receptores TLR2 y TLR4.

11. CONCLUSIÓN

Se acepta parcialmente la hipótesis planteada en este trabajo.

12. BIBLIOGRAFÍA

- Aidar, MA (2013). Guías de diagnóstico, tratamiento y prevención de la tuberculosis. *Hospital muñiz - instituto vaccarezza*, 1- 42.
- Abbas, L (2014). *Inmunología Básica Funciones y Transtornos del sistema inmunitario*. Barcelona, España: Editorial: Elsevier.
- Abou-Zeid C, Garbe T, Lathigra R, Wiker HG, Harboe M, Rook GA, Young DB. (1991). Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infection and Immunity*. 59(8): 2712-2718.
- Adams HP, Koziol JA (1995). Predicción de unión a moléculas MHC clase I. *J Immunol Methods*. 185: 181-90.
- Ahmad J, Khubaid M, Sheikh JA, Panca R, Kumar S, Srinivasan A, Babu MM, Hasnain SE, Ehtesham N. (2019). Disorder-to-order transition in PE-PPE proteins of *Mycobacterium tuberculosis* augments the pro-pathogen immune response. *Federation of European Biochemical Societies (FEBS) Open Bio*, 1-28. doi: 10.1002/2211-5463.feb412749
- Ates LS, Dippenaar A, Ummels R, Piersma SR, Van der Woude AD, Van der Kuij K, Le Chevalier F, Espinosa DM, Barrios J, Castillo BM, Guapillo C, Jiménez CR, Pain A, Houben E, Warren R, Brosch R, Hernández R and Bitter W (2018). Mutations in ppe38 block PE_PGRS secretion and increase virulence of *Mycobacterium tuberculosis*. *Nature microbiology*. 3: 181-188.
- Bansal, ES (2010). PE_PGRS antigens of *Mycobacterium tuberculosis* induce maturation and activation of human dendritic cells. *The Journal of Immunology*, 3495–3504.
- Barathy, Deivanayaga V, Kaza S (2013). “Crystal structure of a putative aspartic proteinase domain of the *Mycobacterium tuberculosis* cell surface antigen PE_PGRS16.” *Federation of European Biochemical Societies (FEBS) Open Bio*. 3: 256–262.
- Basu S, Pathak SK, Banerjee A, Pathak S, Bhattacharyya A, Yang Z, Talarico S, Kundu M, Basu J (2007). “Execution of macrophage apoptosis by PE_PGRS33 of *Mycobacterium tuberculosis* is mediated by Toll-like receptor 2-dependent release of tumor necrosis factor- α .” *Journal of Biological Chemistry*. 282(2): 1039–1050.
- Biologuías, (2019). Universida Autonoma de Barcelona: <https://www.biologuias.com/base-molecular/estructura-primaria-y-secundaria>
- Bottai D, Di LM, Majlessi L, Frigui W, Simeone R, Sayes F, Bitter W, Brennan MJ,

- Leclerc C, Batoni G, Campa M, Brosch R, Esin S (2012). "Disruption of the ESX-5 system of *Mycobacterium tuberculosis* causes loss of PPE protein secretion, reduction of cell wall integrity and strong attenuation." *Molecular Microbiology*. 83(6): 1195–1209.
- Bouhout T, Serji B, Usman E, El amri B, Bouhout I, Soudi M, El harroudi M (2017). Tuberculose mammaire: à propos d'un cas. *Pan African Medical Journal*. 8688: 1-5.
- Borrero R, Álvarez N, Reyes F, Sarmiento ME, Acosta A (2011). "Mycobacterium Tuberculosis: Factores de Virulencia ." *VacciMonitor* . 20(1): 34–38.
- Brennan, GD (2001). Comparative Immune Response to PE and PE_PGRS Antigens. *American Society for Microbiology*, 5606–5611.
- Bretl DJ, Demetriadou C, Zahrt TC (2011). Adaptation to environmental stimuli within the host: twocomponent signal transduction systems of *Mycobacterium tuberculosis*. *Microbiology and Molecular Biology Rev* .75: 566–582.
- Brites D, Gagneux S (2015). Co .evolution of Mycobacterium tuberculosis and Homo sapiens. *Inmunological Rev*. 264(1): 6-24.
- Bustos L, Moreno R, Duque N (2011). Modelo de una bodega de datos para el soporte a la investigación bioinformática. *Scientia et Technica*. 3:145-152.
- Calamita H, Ko C, Tyagi S, Yoshimatsu T, Morrison NE, Bishai WR (2005). The *Mycobacterium tuberculosis* SigD sigma factor controls the expression of ribosome-associated gene products in stationary phase and is required for full virulence. *Cellular Microbiology*.7: 233–244.
- Cascioferro A, Delogu G, Colone M, Sali M, Stringaro A, Arancia G, Fadda G, Palù G, Manganelli R (2007). "PE is a functional domain responsible for protein translocation and localization on mycobacterial cell wall." *Molecular Microbiology*.66(6): 1536–47.
- CENAPRECE. 1990-2016. "Casos nuevos de tumerculosis otras formas". Estados Unidos Mexicanos.
- Ceroni A., Passerini A., Vullo A., Frasconi P. *Nucleic Acids Res* (2006). DISULFIND: un servidor de predicción de conectividad de cisteína y estado de enlace disulfuro. 34: 177-181.
- Chaitra MG, Hariharaputran S, Chandra NR, Shaila MS, Nayak R (2005). Defining putative T cell epitopes from PE and PPE families of. *Vaccine*, 1265–1272.
- Chatrath, Shweta, Vineet Kumar Gupta, Aparna Dixit, Lalit C. Garg (2011). "The Rv1651c-Encoded PE-PGRS30 protein expressed in *Mycobacterium*

- smegmatis* exhibits polar localization and modulates its growth profile.” *Federation of European Microbiological Societies (FEMS) Microbiology Letters*. 322(2): 194–99.
- Chatrath S, Gupta VK, Dixit A, Garg LC (2014). “The PGRS domain is responsible for translocation of PE-PGRS30 to cell poles while the PE and the C-terminal domains localize it to the cell wall.” *Federation of European Biochemical Societies (FEBS) Letters*. 588(6): 990–94.
- Cole S, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon S, Eiglmeier K, Barry C, Tekaia F, Bashman D, Brown D, Chillingworth T, Connor R, Davies R, Devilin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jabs K, Krogh A, McLean J, Moule S, Murphy L, Oliver K, Osborne J, Quail M, Rajandres M (1998). Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 393: 537-544.
- Cuevas CL, Covarrubias AA (2011). Las proteínas desordenadas y su función: una nueva forma de ver la estructura de las proteínas y la respuesta de las plantas al estrés. *Revista Especializada en Ciencias Químico-Biológicas*. 14(2):97-105,
- Daleke MH, Cascioferro A, de Punder K, Ummels R, Abdallah AM, van der Wel N, Peters PJ, Luirink J, Manganelli R, Bitter W (2011). “Conserved Pro-Glu (PE) and Pro-Pro-Glu (PPE) protein domains target LipY lipases of pathogenic mycobacteria to the cell surface via the ESX-5 pathway.” *Journal of Biological Chemistry*. 286(21): 19024–34.
- Donà, V (2013). “The PPE domain of PPE17 is responsible for its surface localization and can be used to express heterologous proteins on the mycobacterial surface.” *Public Library of Science (PLoS) ONE*. 8(3): 1–8.
- Ekiert DC, Cox JS (2014). “Structure of a PE-PPE-EspG complex from *Mycobacterium tuberculosis* reveals molecular specificity of ESX protein secretion.” *Proceeding of the National Academy Sciences of the United States of America*. 111(41): 14758–63.
- Ferreira RJ, Pogo GA (2008). Vacunas derivadas del análisis de los genomas: vacunología inversa. *Interciencia*. 33(5): 353-358.
- Fishbein S, van Wyk N, Warren RM, Sampson SL (2015). “Phylogeny to function: PE/PPE protein evolution and impact on *Mycobacterium tuberculosis* pathogenicity.” *Molecular Microbiology*. 96(5): 901–16.
- Fontan PA, Voskuil MI, Gomez M, Tan D, Pardini M, Manganelli R (2009). The *Mycobacterium tuberculosis* sigma factor sigmaB is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for in vivo growth. *Journal of Bacteriology*. 191: 5628–5633.
- Forrellad MA, Klepp LI, Gioffré A, Sabio y Garcia J, Morbidoni HR, Santangelo M,

- Cataldi AA, Bigi F (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*. 4(1): 3-66
- Frutos- Hernandez, E (2015). Polimorfismos de los genes *pe* y *ppe* presentes en cepas de *Mycobacterium tuberculosis* provenientes de Michoacán. Universidad Michoacana de San Nicolás de Hidalgo .
- Gama, ET (2004). Complejo mayor de histocompatibilidad y desarrollo de vacunas. *NOVA*. 2(002): 59-70.
- Gehring A, Dobon K, Belisle J, Harding C, Boom H (2004). *Mycobacterium tuberculosis* LprG (Rv1411c): a novel TLR-2 ligand that inhibits human macrophage class II MHC antigen processing. *The journal of immunology*. 173(4):2660-2668.
- Gonzales-Mañas, JM (13 de junio 2019). Departamento de Bioquímica y Biología de País Vasco: <http://www.ehu.es/biomoleculas/proteinas/prot42b.htm>.
- Grover S, Sharma T, Singh Y, Kohli S, Aditi M, Semmler T, Wieler L, Tedin K, Ehtesham N, Hasnain S (2018). The PGRS Domain of *Mycobacterium tuberculosis* PE_PGRS protein Rv0297 is involved in endoplasmic reticulum stress mediated apoptosis through toll-like receptor 4. *American Society for Microbiology*. 9(3): 1-18.
- Guillen Nepita, AL (2014) . Epidemiología de la tuberculosis humana y tipificación genética molecular del agente causal en el estado de Michoacán. Universidad Michoacana de San Nicolás de Hidalgo.
- Heo L, Park H, Seok C (2013). Galaxyrefine: protein structure refinement drive by side-chain repacking. *Nucleic acids research*.41:w384-w388.
- Janeway CA Jr, Golstein P (2001). Lymphocyte activation and effector functions. *Current opinion immunol*. 5(3): 313-23
- Jimenez-Ruiz, EN (2017). Polimorfismos en el gen *pe_pgrs18* en cepas de *Mycobacterium tuberculosis* aisladas en Michoacán y Querétaro y su asociación con el genotipo y datos clínicos. Universidad Michoacana de San Nicolás de Hidalgo.
- Jurtz V , Paul S, Andreatta M , Marcatili P , Peters B, Nielsen M (2017). NetMHCpan 4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *The Journal of Immunology*, 1-21.
- Karp, G (2014). Biología celular y molecular. *McGRAW-HILL Interamericana*, 30-74
- Karboul A, Gey van Pittiu NC, Namouchi A (2006). "Insights into the evolutionary

- history of tubercle bacilli as disclosed by genetic rearrangements within a PE_PGRS duplicated gene pair." *BMC evolutionary biology*, 6: 107.
- Karosiene E, RM (2013). NetMHCIIpan-3.0, un método de predicción MHC clase II común y pan-específico que incluye los tres isotipos MHC clase II humanos, HLA-DR, HLA-DP y HLA-DQ. *Inmunogenética*, 711-24.
- Kaufmann S, Ladel C (1994). Role of T Cell Subsets in Immunity against Intracellular Bacteria: Experimental Infections of Knock-Out Mice with *Listeria monocytogenes* and *Mycobacterium bovis* BCG. *Immunobiology*.191(4): 509-519.
- Keane J, Gershon S, Pharm D, Wise R, Levens E, Kasznica J, Schwieterman W, Siegel S, Braun M (2001). Tuberculosis associated with infliximab, a tumor necrosis factor A –neutralizing agent. *The New England Journal of Medicine*. 545 (15): 1098-5.
- Kim, HY (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry and Endodontics*. 42(2):152-155.
- Kopp J, Bordoli L, Battey J, Kiefer F, Schwed T (2007). Assessment of CASP7 predictions for template based modeling targets. *Proteins: structure, function and bioinformatics*. 69 (8): 38-56.
- Korotkova N, Freire D, Phan TH, Ummels R, Creekmore CC, Evans TJ, Wilmanns M, Bitter W, Parret AH, Houben EN, Korotkov KV (2014). "Structure of the *Mycobacterium tuberculosis* type VII secretion system shaperone EspG5 in complex with PE25-PPE41 dimer." *Molecular Microbiology*.94(2): 367–82.
- Kruh NA, Troudt O, Izzo A, Prenni J, Dobos K (2010). "Portrait of a pathogen: the *Mycobacterium tuberculosis* proteome in vivo." *Public Library of Science (PLoS) ONE*. 5(11). <https://doi.org/10.1371/journal.pone.0013938>
- Lantomasi R, Sali M, Cascioferro A, Palucci I, Zumbo A, Soldini S, Rocca S, Greco E, Maulucci G, De Spirito M, Fraziano M, Fadda G, Manganelli R, Delogu G (2012). "PE_PGRS30 is required for the full virulence of *Mycobacterium tuberculosis*." *Cellular Microbiology*. 14(3): 356–67.
- Lee H, Heo L, Lee M, Seok C (2015). GalaxyPepDock: a protein-peptide docking tool based on interaction similarity. *Nucleic Acids Research*.43(1): w431-w435.
- Lee JH, Karakousis PC, Bishai WR (2008). Roles of SigB and SigF in the *Mycobacterium tuberculosis* sigma factor network. *Journal of Bacteriology*. 190: 699–707.
- Lochta C, Hougardyb J, Rouaneta C, Place S, Mascart F (2006). Heparin-binding hemagglutinin, from an extrapulmonary dissemination factor to a powerful

- diagnostic and protective antigen against tuberculosis. *Tuberculosis*. 86(4): 303–309.
- Lovell S, Davis I, Arendall W, Bakker P, Word J, Prisant M, Richardson J, Richardson D (2003). Structure validation by C α geometry: Φ , Ψ and C β deviation. *Proteins: structure, function, and genetics*. 50(1): 437-450.
- Målen H, Berven FS, Fladmark KE, Wiker HG (2007). Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. *Proteomics*.7:1702–1718.
- Manzalawy Y, Honavar V (2010). Recent advances in B-cell epitope prediction methods. *Immunome Research*, 1-6.
- Martínez R, Artiles S, Carpio S, Vidal P, Álvarez S (2015). Tuberculosis pulmonar y cerebral en un paciente inmunocompetente. *Patología Respiratoria*. 18(1): 35-37.
- Matsuyama W, Hashiguchi T, Matsumuro K, Iwami F, Hirotsu Y, Kawabata M, Arimura K, Osame M (2000). Increased serum level of vascular endothelial growth factor in pulmonary tuberculosis. *American Journal of Respiratory and Critical Care Medicine*. 162 (3): 1120-1122.
- Maulén, NP (2011). Virulence factors of *Mycobacterium tuberculosis*. *Revista médica de Chile*. 139(12): 1605–1610.
- McEvoy CR, Cloete R, Müller B, Schürch AC, van Helden PD, Gagneux S, Warren RM, Gey van Pittius NC (2012). Comparative analysis of *Mycobacterium tuberculosis* *pe* and *ppe* genes reveals high sequence variation and an apparent absence of selective constraints. *Public Library of Science (PLoS) ONE*. 7(4). doi: [10.1371/journal.pone.0030593](https://doi.org/10.1371/journal.pone.0030593)
- Mehrotra, J (2001). Regulation of virulence genes in *Mycobacterium tuberculosis*. *International Journal of Medical Microbiology*. 291(2): 171-182.
- Meng E, Pettersen E, Couch G, Huang C, Ferrin T (2006). Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC bioinformatics*. 7: 1-10.
- Meszaros B, Toth J, Vertessy BG, Dosztanyi Z, Simon I (2011). Proteins with complex architecture as potential targets for drug design: a case study of *Mycobacterium tuberculosis*. *Public Library of Science (PLoS) Computational Biology*. 7, e1002118.

- Mortier MC, Jongert E, Mettens P, Ruelle JL (2015). Sequence conservation analysis and in silico human leukocyte antigen-peptide binding. *BMC Immunology*. 16(63). doi: [10.1186/s12865-015-0119-7](https://doi.org/10.1186/s12865-015-0119-7).
- Mukhopadhyay S, Balaji KN (2011). The PE and PPE proteins of *Mycobacterium tuberculosis*. *Tuberculosis*. 91(5): 441-7.
- Nair S, Ramaswamy PA, Ghosh S, Joshi DC, Pathak N, Siddiqui I, Sharma P, Hasnain SE, Mande SC, Mukhopadhyay S (2009). “The PPE18 of *Mycobacterium tuberculosis* interacts with TLR2 and activates IL-10 induction in macrophage.” *The Journal of Immunology*. 183(10): 6269–81.
- O’Garra A, Redford P, Finlay W, Bloom C, Wilkinson R, Berry M (2013). The immune response in tuberculosis. *Annual Review of Immunology*. 31: 475–527.
- Olof E, Brunak S, Heijne G, Nielsen H (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*. 2(4): 953-971.
- Papaleo E, Saladino G, Lambrughl M, Lindorff K, Gervasion FL, Nussinov Ruth (2016). The role of protein loops and linkers in conformational dynamics and allostery. *American Chemical Society*, 6391-6423.
- Parker, KC (1994). Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Journal of Immunology*, 152:163.
- Pérez-Reyes L (2014). Estudio de los determinantes genéticos asociados a la resistencia a antibióticos en cepas de *M. tuberculosis* aisladas en Michoacán. Universidad Michoacana de San Nicolás de Hidalgo.
- Pizza M, Scarlato V, Massignani V, Giuliani M, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B, Ieotti CL, Luzzi E, Manetti R, Marchetti E, Mora M, Nuti S, Ratti G, Santini L, Savino S, Scarselli M, Storni E, Zuo P, Broeker M, Hundt E, Knapp B, Blair E, Mason T, Tettelin H, Hood DW, Jeffries AC, Saunders NJ, Granoff DM, Venter JC, Moxon ER, Grandi G, Rappuoli R (2000). Identification of Vaccine Candidates Against Serogroup B Meningococcus by Whole-Genome Sequencing. *Science*. 286: 1816-1820
- Polena, H (2016). *Mycobacterium tuberculosis* exploits the formation of new blood vessels for its dissemination. *Scientific Reports*. 6-11.
- Raman S, Hazra R, Dascher CC, Husson RN (2004). Transcription regulation by the *Mycobacterium tuberculosis* alternative sigma factor SigD and its role in virulence. *Journal of Bacteriology*. 186: 6605–6616.

- Ramírez NA, Cocotle BE, Méndez A, Arenas J (2005). *Mycobacterium tuberculosis*: su pared celular y la utilidad diagnóstica de las proteínas 16 y 38 kDa. *Médica de la Universidad Veracruzana*. 2(002): 39–43.
- Rashid M, Ramasamy S, Raghava G (2010). A simple approach for predicting protein-protein interactions. *Current protein and peptide science*.11(7):589-600.
- Russell DG, Cardona MJ (2009). Foamy macrophages and the progression of the human tuberculous granuloma. *Nature Immunology*. 10: 943-948.
- Sachdeva P, Misra R, Tyagi AK, Singh Y (2010). The sigma factors of *Mycobacterium tuberculosis*: regulation of the regulators. *Federation of European Biochemical Societies (FEBS) Journal*. 277: 605–626.
- Sakamuri RM, Singh KK, Ryndak MB, Laal S (2019). Evolution of antibodies to epitopes of PE_PGRS51 in the spectrum of active pulmonary tuberculosis. *Oxford University Press for the Infectious Diseases Society*, 1-25.
- Sampson, SL (2011). Mycobacterial PE/PPE proteins at the host-pathogen interface. *Journal of Immunology Research*, 11.
- Sánchez D, Rojas M, Hernández I, Radzioch D, García LF, Barrera LF (2010). Role of TLR2- and TLR4-mediated signaling in *Mycobacterium tuberculosis*-induced macrophage death. *Cellular Immunology*, 128–136.
- Sayes F, Sun L, Di Luca M, Simeone R, Degaiffier N, Fiette L, Esin S, Brosch R, Bottai D, Leclerc C, Majlessi L (2012). Strong immunogenicity and cross-reactivity of *Mycobacterium tuberculosis* ESX-5 type VII secretion: encoded PE-PPE proteins predicts vaccine potential. *Cell host and microbe*. 11(4): 352-363.
- Seiscento M, Vargas F S, Acencio M, Teixeira LR, Capelozzi VL, Sales RK, Antonangelo L (2010). Pleural fluid cytokines correlate with tissue inflammatory expression in tuberculosis. *The International Journal of Tuberculosis and Lung Disease protective*. 14(9):1153–1158.
- Singh KK, Zhang X, Patibandla AS, Chien P Jr, Laal S (2001). Antigens of *Mycobacterium tuberculosis* expressed during preclinical. *Infection and immunity*. 69(6): 4185–4191.
- Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 308–311.
- Sonam G, Tarina S, Yadvir S, Sakshi K, Manjunath P, Aditi S, Torsten S, Lothar H, Karsten T, Nasreen Z, Seyed E (2018). The PGRS domain of *Mycobacterium*

tuberculosis PE_PGRS protein Rv0297 is involved in endoplasmic reticulum stress-mediated apoptosis through Toll-Like receptor 4. *mBio*. 9(3): 1017-1018. doi:10.1128/mBio.01017-18.

Strong M, Sawaya MR, Wang S, Phillips M, Cascio D, Eisenberg D (2006). "Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*." *Proceedings of the National Academy of Sciences of the United States of America*. 103(21): 8060–65.

Sturniolo T, Bono E, Ding J, Raddrizzani L, Tuereci O, Sahin U, Braxenthaler M, Gallazzi F, Protti MP, Sinigaglia F, Hammer J (1999). Generation of tissue-specific and promiscuous HLA ligand database using DNA microarrays and virtual HLA class II matrices. *Nature Biotechnology*, 555-561.

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta J, Simonovic M, Roth A, Santos A, Tsafou K, Kuhn M, Bork P, Jensen L, Mering C (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 447–452.

Taraco AG, Arce Y (2007). Tuberculosis: mecanismos de defensa, inmunopatogenesis y biomarcadores. *Revista de salud pública y nutrición*, 1-7.

Thi EP, Lambertz U, Reiner NE (2012). Sleeping with the Enemy: how intracellular pathogens cope with a macrophage lifestyle. *Public Library of Science (PLoS) Pathogens*. 8(3): 1–4. doi: <https://doi.org/10.1371/journal.ppat.1002551>.

Tiwari BM, Kannan N, Vemu L, Raghunand T (2012). The *Mycobacterium tuberculosis* PE proteins Rv0285 and Rv1386 modulate innate immunity and mediate bacillary survival in macrophages. *Public Library of Science (PLoS) ONE*. 7(12): 1–13. doi: <https://doi.org/10.1371/journal.pone.0051686>.

Tsolaki AG, Hirsh AE, DeRiemer K, Enciso JA, Wong MZ, Hannan M, Goguet de la Salmoniere YO, Aman K, Kato-Maeda M, Small PM (2004). Functional and evolutionary genomics of *Mycobacterium tuberculosis*: insights from genomic deletions in 100 strains. *Proceedings of the National Academy of Sciences of the United States of America*. 101(14): 4865–70.

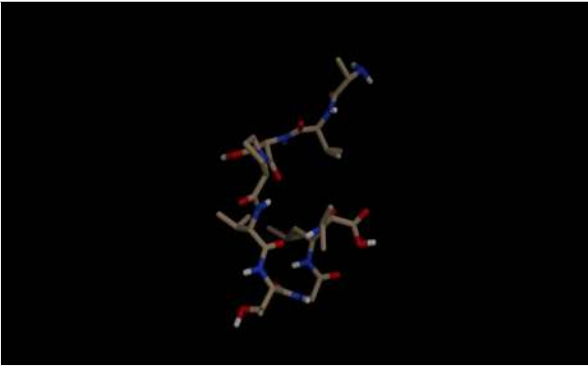
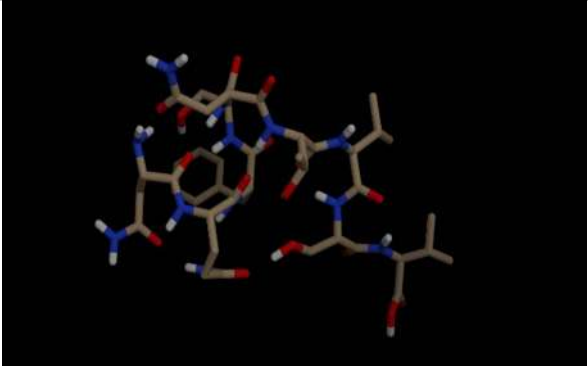
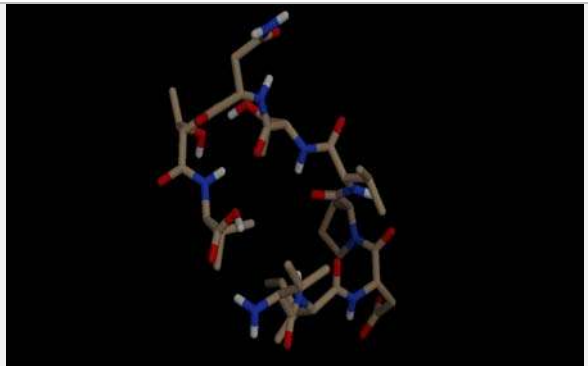
Vázquez-Heredia, MR (2014). Tipificación molecular de micobacterias obtenidas de pacientes con tuberculosis en el periodo 2011-2014. Universidad Michoacana de San Nicolás de Hidalgo.

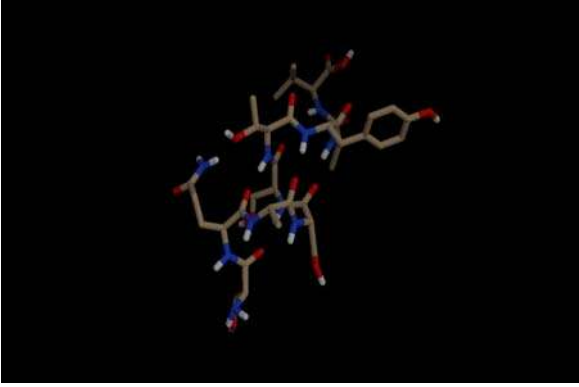
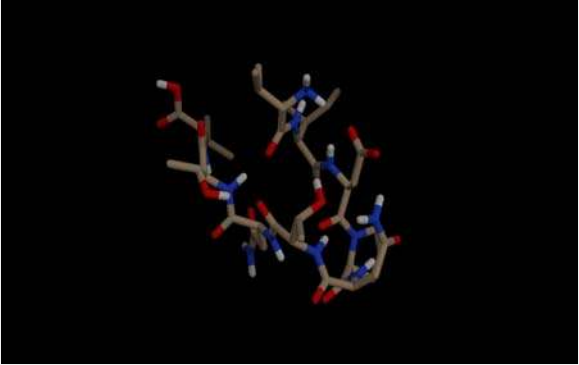
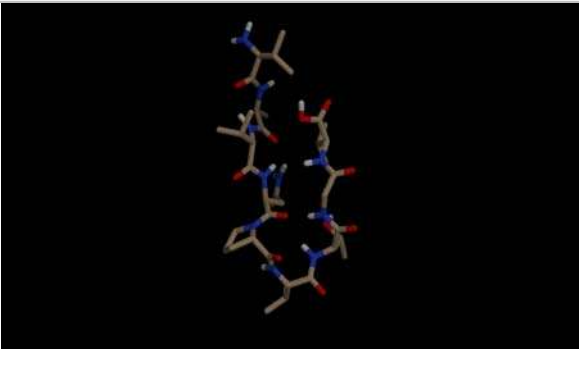
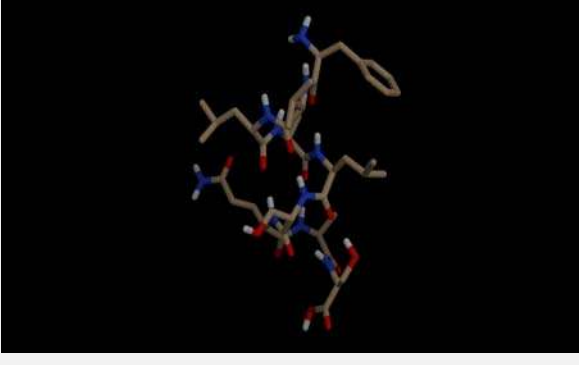
Vielma, JR (2010). Moléculas del complejo mayor de histocompatibilidad. *Revista de la facultad de Medicina UNAM*. 52(2) :1-47.

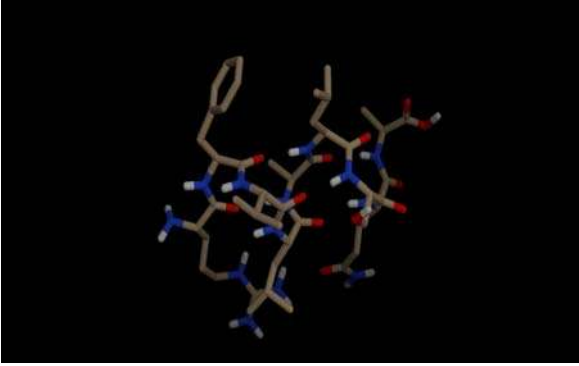
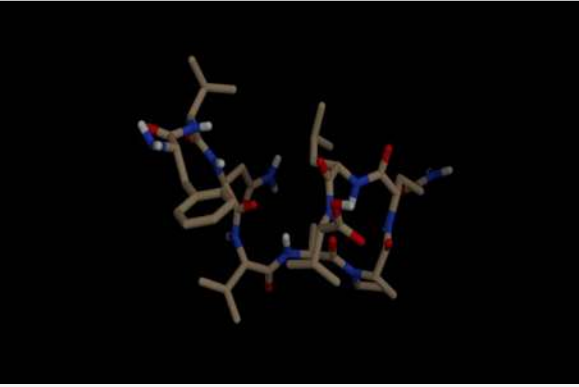
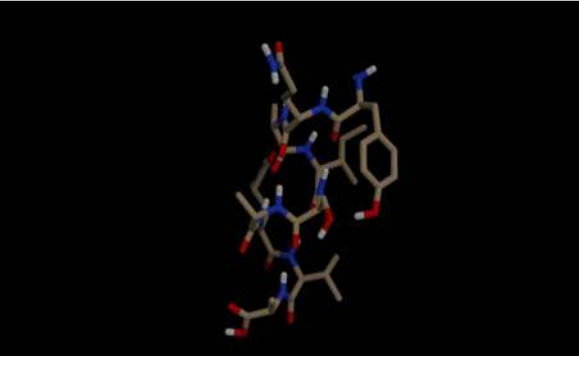
- Volkman H, Beery C, Chang D (2004). Tuberculous granuloma formation is enhanced by a *Mycobacterium* virulence determinant. *Public Library of Science (Plos) Biology*. 2(11). doi: [10.1371/journal.pbio.0020367](https://doi.org/10.1371/journal.pbio.0020367).
- Voskuil MI, Schnappinger D, Rutherford R, Liu Y, Schoolnik, GK (2004). Regulation of the *Mycobacterium tuberculosis* PE/PPE genes. *Tuberculosis*. 84: 256–262.
- Walters SB, Dubnau E, Kolesnikova I, Laval F, Daffe M, Smith I (2006). The *Mycobacterium tuberculosis* PhoPR two-component system regulates genes essential for virulence and complex lipid biosynthesis. *Molecular Microbiology*. 60: 312–330.
- WHO. 2015. "Global tuberculosis report. 2015". *Pan American Health Organization*.
- WHO. 2017. "Global tuberculosis report. 2017." *Pan American Health Organization*.
- WHO. 2018. "Global tuberculosis report. 2017." *Pan American Health Organization*.
- Wiederstein M, Sippl M (2007). Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. 35: w407-w410.
- Williams EP, Lee JH, Bishai WR, Colantuoni C, Karakousis PC (2007). *Mycobacterium tuberculosis* SigF regulates genes encoding cell wall-associated proteins and directly regulates the transcriptional regulatory gene *phoY1*. *Journal Bacteriology*. 189: 4234–4242.
- Trujillo-Alvarez, Y (2017). El complejo mayor de histocompatibilidad. Organización genética. *Panorama*. 13(1): 53-57.
- Zumbo A, Palucci I, Cascioferro A, Sali M, Ventura M, D'Alfonso P, Iantomasi R, Di Sante G, Ria F, Sanguinetti M, Fadda G, Manganelli R, Delogu G (2013). Functional dissection of protein domains involved in the immunomodulatory properties of PE_PGRS33 of *Mycobacterium tuberculosis*. *Pathogens and Disease*. 69(3): 232–39.

ANEXOS

ANEXO 1. Modelado 3D de los epítopes reconocidos por el mayor número de alelos y los parámetros de robustez del servidor

Nº	Epítope	Modelo 3D	sOPEP	Tm
1	AVSPVTGLV		-6.94	0.232
2	NQFSNTVSV		-5.60	0.450
3	VIDPVTNTV		-4.81	0.321

4	SQASSTYAV		-5.37	0.391
5	VIDPNTNTV		-5.29	0.353
6	VIDPATNTV		-3.90	0.335
7	FVLALSQAS		-12.46	0.367

8	RFVLALSQA		-12.39	0.486
9	FVNVAPQLV		-7.96	0.485
10	YQAISAQVA		-8.38	0.576