



**BENEMÉRITA UNIVERSIDAD AUTÓNOMA DE
PUEBLA**



POSGRADO EN MAESTRÍA EN CIENCIAS MICROBIOLOGÍA

MICROBIOLOGÍA MÉDICA

TESIS PRESENTADA POR:

MARÍA PATRICIA REYES BRAVO

PARA OBTENER EL GRADO DE:

MAESTRA EN CIENCIAS MICROBIOLÓGICAS

**ESTUDIO DE LAS PILINAS MENORES BFPI, BFPJ Y BFPK DEL
PILI FORMADOR DE MECHONES DE *ESCHERICHIA COLI*
ENTEROPATÓGENA**

DIRECTOR DEL PROYECTO:

DC. CLAUDIA FABIOLA MARTÍNEZ DE LA PEÑA

Realizada en las instalaciones del Instituto de Ciencias de la Universidad Autónoma de Puebla. Edificio IC 10. Ciudad Universitaria, Colonia San Manuel, C.P. 72570. Puebla, Puebla México.

PUEBLA, PUEBLA NOVIEMBRE 2017

REVISIÓN

El proyecto de investigación **Estudio de las pilinas menores BfpI, BfpJ y BfpK del pili formador de mechones de *Escherichia coli* enteropatógena** realizado por **María Patricia Reyes Bravo** ha sido revisado y aprobado por el siguiente consejo particular, para obtener el Título de:

MAESTRA EN MICROBIOLOGÍA

Firma

Director del proyecto	Claudia Fabiola Martínez de la Peña	_____
Revisor	Rosa del Carmen Rocha Gracia	_____
Revisor	Miguel Castañeda Lucio	_____
Revisor	Margarita Ma. de la Paz Arenas Hernández	_____

El presente trabajo de investigación fue desarrollado en las instalaciones del **Centro de Investigación en Ciencias Microbiológicas, Instituto de Ciencias BUAP**, en el laboratorio de **Biología Molecular de Enteropatógenos**. Dicho trabajo, fue realizado gracias al apoyo del **Consejo Nacional de Ciencia y Tecnología CONACYT** N° de becario 581485 y al apoyo del **Centro de Investigación en Ciencias Microbiológicas, Instituto de Ciencias BUAP**, bajo la dirección de la **DC. Claudia Fabiola Martínez de la Peña**.

DEDICATORIA

AGRADECIMIENTOS

TABLA DE CONTENIDO

REVISIÓN	i
DEDICATORIA.....	iii
AGRADECIMIENTOS.....	iv
Tabla de contenido.....	v
ÍNDICE DE FIGURAS	vii
ÍNDICE DE TABLAS.....	ix
1 RESUMEN.....	1
2 INTRODUCCIÓN	2
2.1 Pili Tipo IV	2
2.2 Pilina Tipo IV	3
2.3 Pilinas menores.....	9
2.4 <i>E. coli</i> enteropatógena EPEC.....	10
2.5 Operones LEE y BFP.....	12
2.6 Regulación del plásmido EAF	13
2.7 BfpA.....	14
2.8 Cristalización de pilinas.....	15
3 ANTECEDENTES.....	15
4 JUSTIFICACIÓN.....	18
5 OBJETIVOS.....	18
5.1 General.....	18
5.2 Específicos	19
5. HIPÓTESIS.....	19
6 MATERIALES Y MÉTODOS	19
6.1 Material biológico.....	19
6.2 Condiciones de crecimiento.....	21
6.3 Extracción plasmídica por el método Quick prep.....	21
6.4 Extracción plasmídica por Kit Zyppy™ Plasmid Miniprep	22
6.5 Diseño de primers para amplificar los genes <i>bfpI</i> , <i>bfpJ</i> y <i>bfpK</i>	22
6.6 Diseño de primers del vector pET-3a	23

6.7	Amplificación por PCR	23
6.8	Clonación en vector pET-3a	24
6.9	Preparación de células químicamente competentes.....	25
6.10	Transformación por choque térmico.....	26
6.11	Preparación de células electrocompetentes y electroporación.....	26
6.12	Expresión de las proteínas BfpI, BfpJ y BfpK.....	27
6.13	Determinación de solubilidad	28
6.14	Purificación y diálisis de las proteínas BfpI, BfpJ y BfpK.....	29
6.15	Determinación de la concentración de las proteínas.....	30
6.16	Electroforesis en geles de poliacrilamida	30
6.17	Análisis de Western Blot	31
7	RESULTADOS.....	32
7.1	Diseño de primers para amplificar los genes <i>bfpI</i> , <i>bfpJ</i> y <i>bfpK</i>	32
7.2	Amplificación de los genes <i>bfpI</i> , <i>bfpJ</i> y <i>bfpK</i>	37
7.3	Conservación del vector comercial pET-3a en las cepas HB101 y MC4100.....	38
7.4	Doble restricción del plásmido pET-3a	38
7.5	Clonación de BfpI-pET3a y BfpJ-PET3a	39
7.6	Transformación de la cepa HB101 con las construcciones BfpI-pET3a y BfpJ-pET3a40	
7.7	Verificación de las construcciones BfpI-pET3a y BfpJ-pET3a en la cepa HB101 por PCR y doble restricción.....	42
7.8	Secuenciación de las clonas de BfpI-pET3a y BfpJ-pET3a en la cepa HB101.....	49
7.9	Transformación en la cepa BL21 (DE3) con las construcciones BfpI-pET3a y BfpJ-pET3a.....	57
7.10	Transformación en la cepa Rosetta gami B (DE3) con las construcciones BfpI-pET3a y BfpJ-pET3a.....	59
7.11	Expresión de la proteínas BfpI y BfpJ mediante IPTG en las cepas BL21 (DE3) y Rosetta gami B (DE3).....	60
7.12	Determinación de la solubilidad de BfpI y BfpJ.....	62
7.13	Purificación de la proteína BfpI en condiciones nativas.....	62
7.14	Purificación de la proteína BfpI y BfpJ en condiciones desnaturalizantes.....	63
7.15	Western Blot	65

7.16	Renaturalización de proteínas	67
7.17	Análisis bioinformático de las secuencias aminoacídicas de BfpI y BfpJ.....	68
7.18	Análisis informático de probables epitopes	70
8	DISCUSIÓN DE RESULTADOS	74
9	CONCLUSIÓN	82
10	ANEXOS.....	84
10.1	Esquema de trabajo	84
10.2	Epitopes probables	85
11	BIBLIOGRAFÍA.....	107

ÍNDICE DE FIGURAS

Figura 1	Similitudes entre los componentes del Sistema de secreción tipo II, Pili tipo IV y el flagelo de archaeas.....	3
Figura 2	Estructura de la pilina tipo IVa PilE de <i>Neisseria gonorrhoeae</i>	6
Figura 3	Pilinas tipo IVb. <i>S. typhi</i> PilS, <i>V. cholerae</i> TcpA y EPEC BfpA	7
Figura 4	Modelo de la estructura del BFP.....	8
Figura 5	Patogénesis de tEPEC mediante BFP.	12
Figura 6	Estructura tridimensional de la pilina tipo IVb, BfpA de <i>Escherichia coli</i> enteropatógena (EPEC)	14
Figura 7	Mapa del plásmido pET-3a	20
Figura 8	Esquema de doble restricción.....	32
Figura 9	Esquema de expresión y purificación de las proteínas.....	33
Figura 10	Secuencia de BfpI	334
Figura 11	Secuencia de BfpJ.....	35
Figura 12	Secuencia de BfpK.....	¡Error! Marcador no definido.
Figura 13	Secuencias de los primers sintetizados para BfpI, BfpJ y BfpK.....	36
Figura 14	Primers diseñados para el vector pET-3a.....	36
Figura 15	Visualización mediante el programa Snap gene viewer de la posible hibridación de los primers diseñados para el vector pET-3a.....	¡Error! Marcador no definido.
Figura 16	Geles de agarosa al 1%. Productos de amplificado de PCR.	37
Figura 17	Gel de agarosa al 1% transformación de cepa HB101 y MC4100 con pET3a	38
Figura 18	Gel de agarosa al 1%. Doble restricción del vector comercial (VL) y del ADN plasmídico de clonas transformadas con el vector pET-3a.	39
Figura 19	Gel de agarosa al 1%. Perfil de bandeos de clonas transformadas con construcciones BfpI-pET3a y BfpJ-pET3a.....	40

Figura 20 Gel de agarosa al 1%. Extracción plasmídica de las clonas transformadas con la construcción BfpK-pET3a.....	41
Figura 21 Gel de agarosa al 1% corrimiento electroforético con muestras de la reacción de doble restricción y PCR.....	¡Error! Marcador no definido.
Figura 22 Gel de agarosa al 1%. Izquierda extracción plasmídica de las clonas transformadas con la construcción BfpI-pET3a.	44
Figura 23 Doble digestión de la construcción BfpI-pET3a.....	45
Figura 24 Gel de agarosa al 1%. Producto de PCR del ADN plasmídico de las clonas transformadas con la construcción BfpK-pET3a.....	46
Figura 25 Doble restricción de las clonas con las construcciones BfpK-pET3a.....	48
Figura 26 Gel de agarosa al 1%. Amplificado río arriba y río abajo del sitio de integración del inserto de las construcciones BfpI-pET3a y BfpJ-pET3a.....	48
Figura 27 Gel de agarosa al 1%. Amplificado de la región río arriba y río abajo del sitio de integración del inserto de la construcción BfpK-pET3a..	48
Figura 28 Alineamiento entre la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona d (BfpI-pET3a)	52
Figura 29 Alineamiento entre la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona f (BfpI-pET3a)	53
Figura 30 Alineamiento de la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona 10..	54
Figura 31 Alineamiento de la secuencia teórica y la secuenciación con el primer forward del ADN plasmídico de la clona 11	55
Figura 32 Alineamiento de la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona 11 con el primer reverso.	58
Figura 33 Secuencia de la construcción BfpK-pET3a (clona 17).	59
Figura 34 Secuencia de la construcción BfpK-pET3a (clona 18)	57
Figura 35 Gel de agarosa al 1%. Extracción plasmídica de las cepas BL21 (DE3) con la construcción BfpI- pET3a.	58
Figura 36 Gel de agarosa al 1%. Extracción plasmídica de las cepas BL21 (DE3) con la construcción BfpJ- pET3a	59
Figura 37 Gel de agarosa al 1%. Extracción plasmídica de las cepas Rosseta gammi (DE3) transformadas con las construcciones BfpI-pET3a y BfpJ-pET3a.....	¡Error! Marcador no definido.
Figura 38 Corrimiento electroforético de las proteínas de las cepas BI21 y Roseta-gami (DE3).	¡Error! Marcador no definido.
Figura 39 Determinación de la solubilidad de las proteínas BfpI y BfpJ..	¡Error! Marcador no definido.
Figura 40 Corrimiento electroforético de las muestras de purificación de BfpI bajo condiciones nativas.....	¡Error! Marcador no definido.
Figura 41. Corrimiento electroforético de la purificación de BfpI bajo condiciones desnaturalizantes.....	¡Error! Marcador no definido.

Figura 42 Corrimiento electroforético de las muestras de purificación de BfpJ bajo condiciones desnaturalizantes.....**¡Error! Marcador no definido.**
Figura 43 Gel de poliacrilamida utilizado para el Western Blot**¡Error! Marcador no definido.**
Figura 44. Western Blot dirigido hacia las etiquetas de histidina de las proteínas BfpI y BfpJ.....**6¡Error! Marcador no definido.**

ÍNDICE DE TABLAS

Tabla 1 Cepas utilizadas y sus características.	19
Tabla 2 Plásmido utilizado y sus características	20
Tabla 3 Probables epitopes en la secuencia de BfpI, BfpJ y BfpK.	80

1 RESUMEN

Este trabajo tiene por modelo de estudio las pilinas menores del pili formador de mechones de *Escherichia coli* enteropatógena (EPEC) un patógeno diarregénico que afecta a niños menores de seis meses a nivel mundial y principalmente países en vías de desarrollo.

Para conocer la probable función de las pilinas menores BfpI, BfpJ y BfpK es necesario obtener la estructura de las proteínas. Una de las técnicas mediante las que se puede conocer la estructura es mediante cristalografía de rayos X, esta técnica requiere que la proteína de estudio sea soluble. Por tal motivo en este trabajo se diseñaron primers específicos para amplificar cada uno de los genes *bfpI*, *bfpJ* y *bfpK* sin la porción hidrofóbica N-terminal para obtener las proteínas solubles. A continuación el amplificado de cada gen se clonó en el vector pET-3a, se verificó cada construcción (BfpI- pET-3a, BfpJ- pET-3 y BfpK- pET-3a) por PCR, doble restricción y secuenciación. Posteriormente se expresó en la cepa BL21 (DE3) y purificó bajo condiciones desnaturizantes (con el uso de urea) las proteínas BfpI y BfpJ. BfpK a diferencia no se expresó por que las construcciones no se verificaron de manera aprobatoria. Finalmente se dializaron las proteínas BfpI y BfpJ, sin embargo al eliminar la urea las proteínas dializadas se precipitaron, por lo que se propone modificar las condiciones de diálisis usadas en este trabajo adicionando agentes que favorezcan condiciones oxidantes, así como la técnica empleada.

2 INTRODUCCIÓN

2.1 Pili Tipo IV

Los pili tipo IV son apéndices superficiales, formados por subunidades de proteínas denominadas pilinas, son flexibles y son expresados por muchas especies de bacterias Gram negativas, Gram positivas y también por archaeas. Algunas de las bacterias Gram negativas que lo expresan son: *Pseudomonas aeruginosa*, *Vibrio cholerae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, *Salmonella typhi* y *Escherichia coli* enteropatógena (EPEC). Por otra parte algunas de las bacterias Gram positivas que lo expresan son *Clostridium perfringens* y *Clostridium difficile* (Lieberman *et al.*, 2015). Los pili tipo IV son extremadamente delgados con un diámetro aproximado de 60-80 Å y una longitud mayor a 1µm (Craig *et al.*, 2006). Estos pili se distribuyen formando agregados o mechones (Telford *et al.*, 2006). Entre los procesos bacterianos mediados por los pili tipo IV se encuentran adherencia bacteriana a células hospederas, especificidad por células diana, motilidad por twitching, captación de ADN, adsorción de bacteriófagos (McNamara & Donnenberg., 2000), formación de microcolonias, biopelículas, secreción de exoproteínas (Giltner *et al.*, 2012) y transferencia de electrones (Reguera *et al.*, 2005).

Los pili tipo IV poseen similitudes estructurales con el sistema de secreción tipo II y el flagelo de archaeas (Peabody *et al.*, 1993). Estas similitudes son, la subunidad mayor (llamada pilina para los pili tipo IV, pseudopilina para el sistema de secreción tipo II y flagelina para el flagelo de archaeas), una prepilina peptidasa, una ATPasa ensambladora y una proteína de membrana interna llamada secretina (sistema de secreción tipo II y pili tipo

IV). El pili tipo IV además de estas estructuras posee una ATPasa que dirige la despolimerización del pilus (Craig & Li., 2008) (Figura 1).

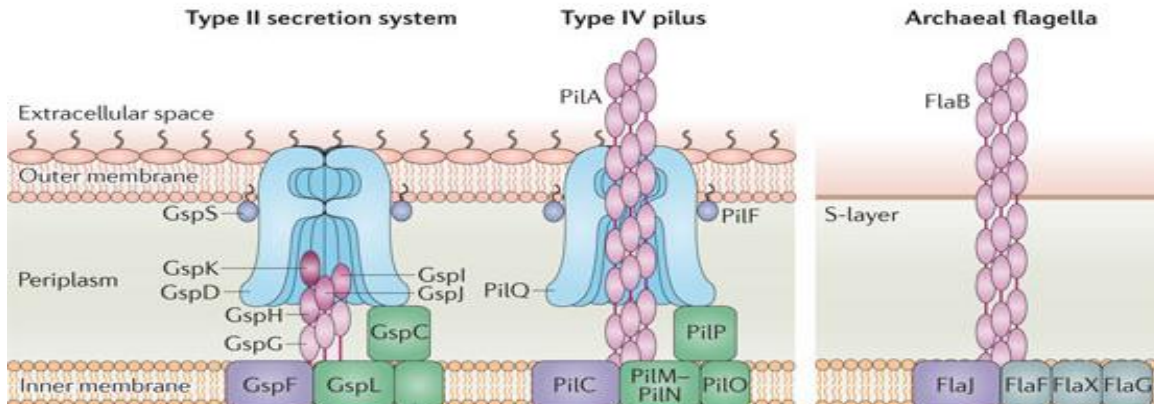


Figura 1. Similitudes entre los componentes del Sistema de secreción tipo II, Pili tipo IV y el flagelo de arqueas. En rosa se observan las subunidades mayores (pseudopilinas, pilinas y flagelina), ATPasas en anaranjado, morado y verde son las proteínas de membrana interna, y proteínas de membrana externa (secretinas) en azul. Tomado de Korotkov *et al.*, 2012.

2.2 Pilina Tipo IV

El pili tipo IV, como anteriormente se mencionó, está formado por subunidades mayores o pilinas y subunidades menores también conocidas como pilinas menores. Las pilinas tipo IV son proteínas de aproximadamente ~15 a 20kDa (Craig & Li., 2008). Tienen la característica de tener un extremo N-terminal altamente hidrofóbico, el cual es un dominio transmembranal y a la vez un dominio de interacción proteína-proteína. Este extremo N-terminal posee una estructura α -helice la cual se divide en dos subdominios, el α 1-N (que abarca desde el aminoácido 1~28) y el α 1-C (desde el 29~52) (Giltner *et al.*, 2012). Otra característica importante de estas proteínas es el poseer una secuencia señal en el extremo N-terminal llamada tipo III la cual es procesada por una peptidasa específica, denominada prepilin-peptidasa (McNamara & Donnenberg., 2000). Las proteínas con este tipo de secuencia señal son llamadas prepilinas y solo son aptas para su uso en la

polimerización del pilus hasta que se remueve esta secuencia señal en la membrana citoplasmática (Strom & Lori., 1991).

Las pilinas tipo IV se dividen en dos tipos: pilina IVa y pilina IVb. Esta clasificación se basa en la longitud del péptido señal, en el aminoácido inicial de la proteína madura y en su organización dentro del genoma. La pilina tipo IVa tiene una secuencia tipo III pequeña de 6 a 7 aminoácidos y una fenilalanina metilada en el extremo N-terminal una vez que el péptido señal es escindido (Giltner *et al.*, 2012). Por otro lado las pilinas tipo IVb tienen una secuencia tipo III larga de 15 a 30 residuos y una vez escindido el péptido señal el primer aminoácido inicial es hidrofóbico (Giltner *et al.*, 2012). Respecto a la organización de los genes implicados en la biosíntesis del pili IVa estos se encuentran distribuidos por todo el genoma; por el contrario los genes implicados en la biosíntesis del pili tipo IVb se encuentran en operones (Lieberman *et al.*, 2015). Las pilinas tipo IVa se han encontrado en patógenos de plantas, animales y de humanos por ejemplo *Pseudomonas*, *Neisseria* y *Dichelobacter*, y en bacterias del ambiente como *Thermus*, *Myxococcus*, *Deinococcus*, *Bdellovibrio* y *Shewanella* (Pelicic, 2012). Por otra parte las pilinas tipo IVb se han identificado comúnmente en bacterias entéricas como *Salmonella entérica serovar Typhi*, (Zhang *et al.*, 2000) *Vibrio cholerae* (Craig *et al.*, 2003) y *Escherichia coli* (enteropatógena) (Donnenberg & Kaper., 1992)

La estructura terciaria de las pilinas tipo IV se asemeja a una paleta o cucharón y se ha estudiado en varios modelos como por ejemplo, Pile de *N. gonorrhoeae* (Figura 2) o en PilA de *P. aeruginosa*, donde la región α 1-N sobresale del dominio globular C-terminal y forma el núcleo central del ensamble del pilus (asemeja al mango de la paleta). Este subdominio es multifuncional debido a que actúa como un segmento transmembranal que retiene las subunidades de pilina ancladas en la membrana citoplasmática antes del

ensamble del pilus, y como dominio de interacción entre subunidad y subunidad en la fibra (Giltner *et al.*, 2012). Este dominio es funcionalmente importante en el ensamble del pili tipo IV ya que mutaciones puntuales en la subunidad mayor pilina PilE de *Neisseria gonorrhoeae*, específicamente en el residuo de glutamina 5 que está altamente conservado en la región α 1-N, demuestran que el ensamble del pili se imposibilita (Aas *et al.*, 2007).

A diferencia del N-terminal, el dominio extremo C-terminal de las pilinas tipo IV conserva un motivo estructural. En las estructuras de las pilinas tipo IVb está formado por láminas- β antiparalelas compuestas de 5 a 7 hebras, como es el caso de *V. cholerae* TcpA (Craig *et al.*, 2003) y de *Salmonella* Typhi PilS, (Xu *et al.*, 2004) así como una mezcla paralela y antiparalela de láminas- β para el caso de BfpA de *Escherichia coli* enteropatógena sin conectividad inmediata (Figura 3) (Ramboarina *et al.*, 2005).

Típicamente las estructuras de las pilinas se obtienen mediante cristalización de rayos X. Para realizar este método se requiere la proteína en forma soluble sin embargo la porción hidrofóbica en el amino terminal dificulta este proceso. A pesar de esto, se han logrado cristalizar algunas estructuras de pilinas cuando se escinde la porción N-terminal hidrofóbica obteniendo así la proteína de forma soluble. Comúnmente se escinden los primeros 28 aminoácidos o incluso más de la secuencia del extremo N-terminal de las pilinas con el fin de eliminar la porción hidrofóbica (PAK, de *P. aeruginosa* trabajo de Helaine *et al.*, 2007. Fim U y Pil E pilinas menores de *P. aeruginosa* trabajo de Nguyen *et al.*, 2015(A). CofB pilina menor de CFA/III ETEC trabajo de Kawanara *et al.*, 2015). Solo existen tres estructuras completas de las pilinas (con el extremo N-terminal sin escindir) estas pertenecen a las tipo IVa: la pilina GC o gonococcal PilE de *N. gonorrhoeae* (Parge *et al.*, 1995), la pilina PAK o PilA de *P. aeruginosa* (Craig *et al.*, 2003) y la pilina DnFimA

de *Dichelobacter nodosus* (Hartung *et al.*, 2011). Estudios realizados por Craig y colaboradores en el 2003 demostraron que el eliminar la región hidrofóbica N-terminal no conlleva cambios en la estructura globular de la pilina PilA (PAK) de *Pseudomonas aeruginosa* puesto que se comparó la estructura completa con la estructura truncada en el extremo N-terminal y estas resultaron ser idénticas (Craig *et al.*, 2003)

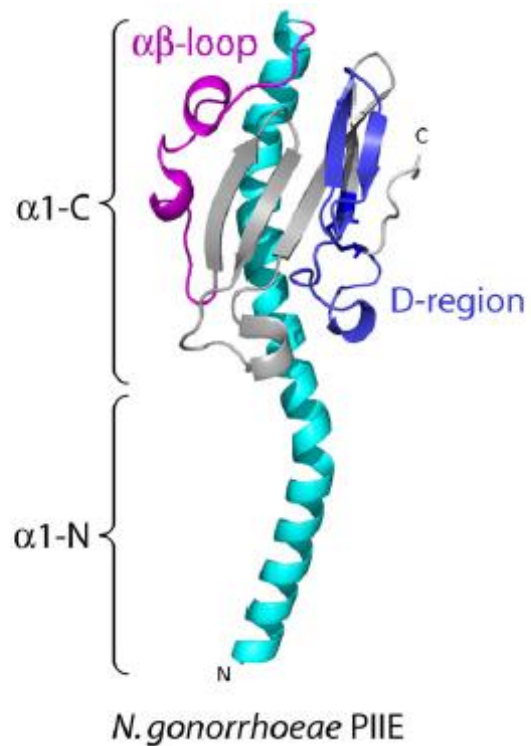


Figura 2. Estructura de la pilina tipo IVa Pile de *Neisseria gonorrhoeae*. Se observan tres regiones conservadas: el extremo N-terminal α -hélice (cian), el extremo C-terminal lámina- β (gris) y el lazo- $\alpha\beta$ (magenta) que conecta el extremo N-terminal y el extremo C-terminal. El extremo N-terminal α -hélice se divide en dos subdominios α 1-N y el α 1-C. La región D (azul) está formada por dos residuos de cisteína que forman un puente disulfuro. Imagen tomada de Giltner CL, 2012

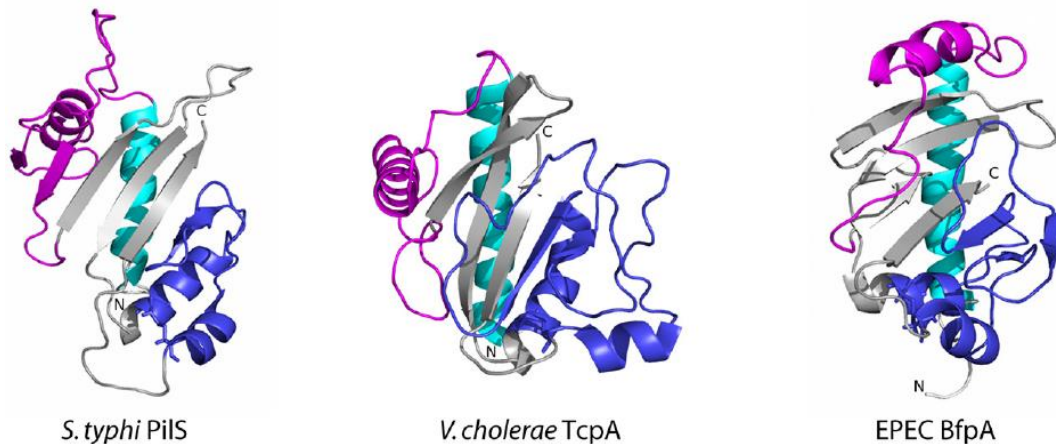


Figura 3. Pilinas tipo IVb. *S. typhi* PilS, *V. cholerae* TcpA y EPEC BfpA. Estructura terciaria de pilinas tipo IVb donde se observa en magenta el lazo- $\alpha\beta$, la región D en azul, en cian el extremo N-terminal y en gris la disposición de la laminas- β . Los tres modelos demuestran diferencias en el plegamiento de las proteínas. Imagen tomada de Giltner CL, 2012

Otro estudio similar, es el realizado por Ramboarina y colaboradores en el 2005 donde eliminaron los primeros 22 aminoácidos de la región hidrofóbica del extremo N-terminal de BfpA, subunidad mayor de BFP de EPEC. Esta región hidrofóbica es importante en el ensamble de BFP y proponen un modelo del ensamble del filamento BFP. Este modelo consiste en un tipo espiral, donde las regiones N-terminal hidrofóbicas se encuentran en la parte interna del filamento y los dominios globulares entre subunidad y subunidad se encuentran más expuestos (Ramboarina *et al.*, 2005) (Figura 4).

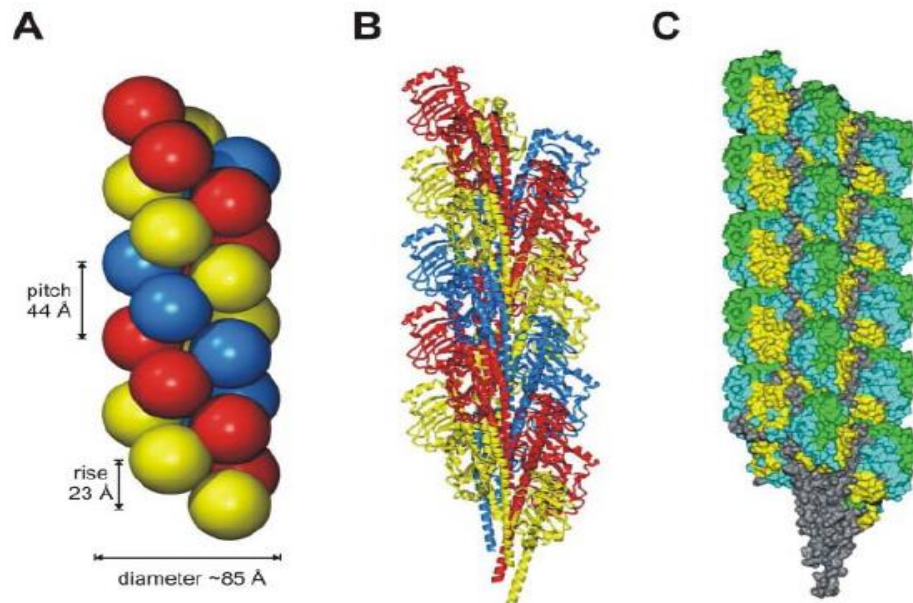


Figura 4. Modelo de la estructura del BFP.

- A) Representación del filamento helicoidal 3 inicios o hebras, en el cual cada subunidad se representa como esfera y las tres hebras en color azul, amarillo y rojo. B) modelo del filamento de BFP en el cual cada subunidad de pilina se representa en cinta. Las hebras helicoidales se muestran en amarillo, azul y rojo. C) Representación del modelo de BFP donde el lazo- $\alpha\beta$, el núcleo y la región D se muestran en verde, azul y amarillo respectivamente.; En gris el N-terminal formando la capa central del filamento. Imagen tomada de Ramboarina *et al.*, 2005)

El dominio globular de las pilinas es el C- terminal. Esta región varía entre las pilinas de diferentes especies, e incluso entre diferentes cepas de la misma especie (Kus *et al.*, 2004). Sin embargo se mantiene una región conservada entre muchas pilinas mayores y pilinas menores. Esta región se conoce como región D la cual está formada por un puente disulfuro formado por un par cisteínas.

Un estudio realizado en *Pseudomonas aeruginosa* sugiere que la región D puede funcionar como el componente de adhesión del pilus (Giltner *et al.*, 2006). Por otro lado, esta misma región es importante en la polimerización del pilus, pues cuando el pilus se somete a un tratamiento con agentes reductores esto provoca la desintegración rápida del pilus (Li *et al.*, 2004).

2.3 Pilinas menores

Así como las pilinas mayores forman parte del pilus tipo IV, las pilinas menores o pilin-like protein también se han encontrado formando parte del filamento. Sin embargo, estas se expresan a niveles bajos comparados con la expresión de la pilinas mayores (Martinez de la Peña *et al.*, 2016). Todos los sistemas pili tipo IV poseen una o más pilinas menores. Por ejemplo, *V. cholerae* TCP y *Escherichia coli* enterotoxigénica (ETEC) CFA/III y el Longus pilus son de la clase tipo IVb, cada uno posee una pilina menor, a diferencia de BFP de EPEC que posee tres pilinas menores codificadas en el extremo 3' del operón *bfp* (Kolappan *et al.*, 2015) y *Clostridium difficile* codifica al menos nueve pilinas putativas menores (Maldarelli *et al.*, 2014). Por otro lado, el sistema de secreción tipo II tiene cuatro pseudopilinas menores (Kolappan *et al.*, 2015).

Las subunidades menores CofB y LngB (del pili CFA/III y el Longus pili IVb respectivamente de ETEC) son muy similares en secuencia y tamaño, y a su vez son similares a las pseudopilinas del sistema de secreción tipo II. La estructura de CofB obtenida mediante cristalización de rayos X, demuestra que el dominio C-terminal se encuentra en la punta del pilus además existe una porción que sobresale de la punta la cual potencialmente se asocia con los canales de secretina para continuar con la polimerización del pilus. CofB representa la primera estructura cristalizada de una pilina menor del pilus

tipo IVb y de cualquier pilina que pueda iniciar el ensamble del pilus (Kolappan *et al.*, 2015).

Las funciones de las subunidades menores son en algunos casos desconocidas; sin embargo, en el caso de *N. gonorrhoeae* y de *P. aeruginosa* se conoce que son necesarias en el ensamble del pili tipo IV porque las mutantes en los genes de las pilinas menores y en la ATPasa de retracción se observa una disminución del ensamblaje (Winther-Larsen *et al.*, 2005) (Giltner *et al.*, 2010). Otras funciones de las pilinas menores se involucran en la autoagregación (Helaine, 2005) y motilidad por swarming (Kuchma *et al.*, 2012). Hasta la fecha las estructuras terciarias de las subunidades menores que se han obtenido han sido sin el extremo N-terminal α -1N y la mayoría de ellas se asemeja a las estructuras de sus subunidades mayores sugiriendo que se incorporan en el pilus. (Kolappan *et al.*, 2015).

La pilina menor PilX de *N. meningitidis* (pilina menor tipo IVa) tiene dos α -hélice en la región D la cual está expuesta en la superficie del pilus (Helaine *et al.*, 2007). Mediante microscopia electrónica de transmisión y marcaje con oro (immunogold) se demuestra el bajo nivel de incorporación de PilX en el pilus tipo IVa de *N. meningitidis*, además PilX participa en la autoagregación mediada por el pilus y en la adhesión a las células huésped, estas funciones requieren de las dos α -hélices (Helaine *et al.*, 2007).

2.4 *E. coli* enteropatógena EPEC

E. coli enteropatógena (EPEC) es uno de los patótipos diarreagénicos más importantes de *Escherichia coli* que infectan a niños pequeños en países en vías de desarrollo (Bieber *et al.*, 1998). EPEC típica a diferencia la EPEC atípica produce un pili tipo IV, denominado pilus formador de mechones por sus siglas en inglés BFP (Bundle-

Forming Pilus) (Girón *et al.*, 1993) el cual es un factor de virulencia importante en la patogénesis de tEPEC.

En la patogénesis de tEPEC se pueden distinguir varias etapas siendo la primera, la adherencia localizada de tEPEC a las células epiteliales del intestino delgado mediada por BFP, seguida por una adherencia íntima que ocasiona la pérdida de microvellosidades (borramiento) y la formación de pedestales ricos en actina situados por debajo de la bacteria adherida y la afectación de las uniones estrechas enterocito-enterocito (Zahavian *et al.*, 2011). Esta patogénesis de tEPEC ocasiona una lesión característica denominada lesión A/E (por sus siglas en inglés Attaching and Effacing). El inicio de dicha lesión comienza cuando los BFP, se unen a receptores de LacNAc (N-acetil-lactosamina) (Humphries *et al.*, 2009) presentes en las células epiteliales. A esta primera adherencia se le denomina adherencia localizada o LA (Localized Adherence) (Figura 5). Este primer contacto ocasiona que BFP se retraiga acercando de esta manera la bacteria a la célula hospedera y, favoreciendo que EPEC inyecte mediante su sistema de secreción tipo III (SST3) las proteínas efectoras que modelaran las estructuras de actina del citoesqueleto celular ocasionando el borramiento de las microvellosidades y la formación del pedestal. (Donnenberg *et al.*, 1997) (Celli & Deng, 2000).

Esta translocación de proteínas efectoras desencadena, además de lo antes mencionado, cascadas de señalización que promueven alteraciones en la secreción de iones Cl^- y HCO_3^- así como deslocalización de las acuaporinas y pérdida de las uniones estrechas entre enterocito-enterocito (Zahavian *et al.*, 2011). Otra proteína traslocada es el receptor Tir, el cual es el receptor de la íntima de EPEC. Tir una vez traslocado es fosforilado en el citoplasma celular y se presenta en la membrana citoplasmática donde se une a la Intimina (adhesina no fimbrial) que está presente en la superficie bacteriana,

generándose la adherencia íntima entre EPEC y el enterocito (Kenny, 1999) (Goosney *et al.*, 2000) Figura 5.

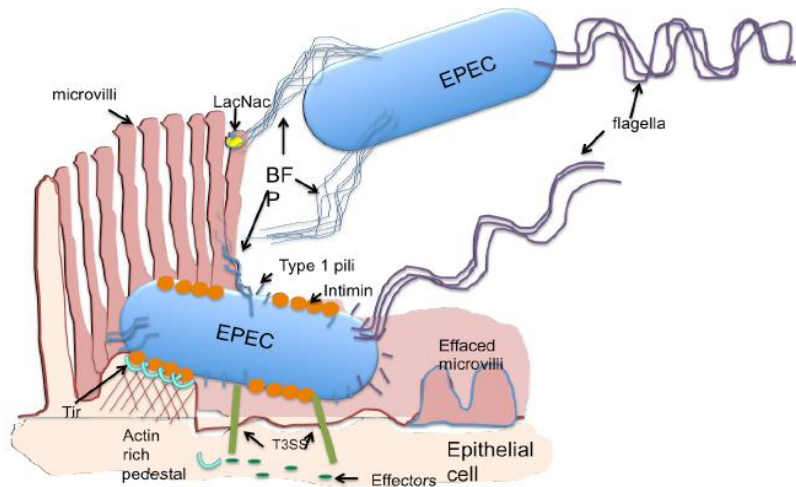


Figura 5. Patogénesis de tEPEC mediante BFP. Por medio de transducción de señales entre el enterocito y la bacteria ocurre una asociación íntima, este proceso es conocido como patogénesis de tEPEC provoca una lesión conocida como ataque y borrado (A/E). Tomado de Martínez de la Peña, 2015 PhD thesis.

2.5 Operones LEE y BFP

Los genes involucrados en el fenotipo de la lesión A/E se encuentran localizados en una isla de patogenicidad genómica de 35Kpb denominado locus de borrado del enterocito o LEE por sus siglas en inglés (locus of enterocyte effacement) (McDaniel *et al.*, 1995). LEE posee 41 marcos abiertos de lectura organizados en cinco operones policistrónicos que van de LEE1-LEE5, los cuales se dividen en tres categorías dependiendo de su función: una codifica los componentes del T3SS, otra para proteínas efectoras y sus chaperonas, una última codifica para la intimina y su receptor Tir (DeVinney *et al.*, 1999) (Elliot *et al.*, 1998)

La adherencia localizada depende de la presencia de un plásmido de 90kb denominado Factor de Adherencia EPEC (EAF), el cual contiene al operón BFP con los 14

genes requeridos para la biosíntesis de BFP (Sohel *et al.*, 1996), y al operón *perABC* con los genes reguladores para la expresión del pilus y de algunos genes de LEE (Puente *et al.*, 1996) (Bustamante *et al.*, 1998). El primer gen del operón *bfp* es *bfpA* que codifica para la pilina mayor Bundlina o BfpA. Tras la escisión de su péptido señal N-terminal (~12 aminoácidos) por la enzima pre-pilin peptidasa BfpP produce una proteína madura de 19 kDa. BfpA posee un motivo transmembranal hidrofóbico de 22 aminoácidos y un dominio globular periplásmico (Jarrell, 2009). En el operón *bfp* además se encuentran el gen *bfpP* que codifican para una prepilin peptidasa, los genes *bfpI*, *bfpJ* y *bfpK* que codifican para las pilinas menores BfpI, BfpJ y BfpK respectivamente, *bfpD* y *bfpF* que codifican para dos ATPasas (necesarias para el ensamble y desensamble del pilus respectivamente) (Jarrell, 2009).

2.6 Regulación del plásmido EAF

La expresión de BFP se regula en respuesta a señales ambientales específicas e involucra procesos transcripcionales (Martinez-Laguna *et al.*, 1999) Muchos patógenos productores del pili tipo IV lo producen en respuesta a crecimiento bajo condiciones específicas de cultivo, temperatura o en respuesta al contacto con las células hospederas. Los genes de *bfp* son expresados por un promotor localizado río arriba de *bfpA* durante la fase exponencial de crecimiento a 37°C, en un medio de cultivo y se regula negativamente por amonio y temperaturas por debajo o arriba de 37°C (Bustamante *et al.*, 1998) (Puente *et al.*, 1996). La activación transcripcional del operón *bfp* requiere el producto del gene *perA*, el primer gen del operón *perABC* que está localizado río abajo de *bfpA* en el plásmido EAF (Tobe *et al.*, 1998) (Puente *et al.*, 1996) (Bustamante *et al.*, 1998). El gen *perA* codifica una

proteína de 274 aminoácidos que pertenece a la familia de reguladores transcripcionales AraC/XylS, se ha demostrado ser un regulador esencial de virulencia para tEPEC en las células hospederas de humanos (Tobe *et al.*, 1998) (Donnenberg *et al.*, 1997)

2.7 BfpA

Un análisis de la variación de las secuencias en *bfpA* revela nueve diferentes alelos que pueden ser divididos en dos clases α y β (Ramboarina *et al.*, 2005). Estas variaciones se concentran en el extremo C-terminal localizadas en la parte interna del pilus, adoptan cuatro α -hélices que rodean un grupo de siete hebras continuas láminas- β dispuestas en el siguiente orden $\beta 2$, $\beta 3$, $\beta 1$, $\beta 7$, $\beta 4$, $\beta 6$ y $\beta 5$ (Ramboarina *et al.*, 2005), Figura 6. Sin embargo se han encontrado características comunes observadas en BFP y otras estructuras de pilinas tipo IV como los dos residuos de cisteína que forman un puente disulfuro en el dominio globular (Jarrell, 2009).

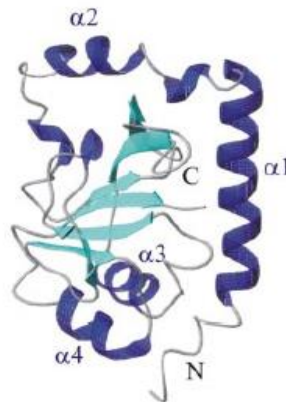


Figura 6. Estructura tridimensional de la pilina tipo IVb, BfpA de *Escherichia coli* enteropatógena (EPEC). Siete láminas- β flanqueadas por cuatro α -hélice. Las α -hélice se representan en color azul, mientras que las láminas- β están representadas en color cian. Imagen tomada de Ramboarina *et al.*, 2005.

2.8 Cristalización de pilinas

La estructura de las subunidades mayores del pili tipo IV han sido obtenidas mediante cristalografía de rayos X. Muchas de las estructuras obtenidas hasta la fecha han sido principalmente subunidades mayores, pero se han obtenido también las estructuras de las pilinas menores de algunos modelos de estudio por ejemplo PilX de *N. meningitidis* (Helaine *et al.*, 2007), CofB de ETEC (Kawahara *et al.*, 2015) y Pile de *P. aeruginosa* (Nguyen *et al.*, 2015(B). La falta del conocimiento de las estructuras de las subunidades menores significa la incomprensión del pili con respecto a su flexibilidad y adaptabilidad en bacterias que lo expresan (Giltner *et al.*, 2012).

Kawahara y colaboradores en el 2015 clonaron, expresaron, purificaron y cristalizaron la proteína CofB de ETEC sin el segmento N-terminal hidrofóbico, este fue truncado en los primeros 28 aminoácidos, para obtener una proteína soluble eficiente para la cristalización (Kawahara *et al* 2015). Sus resultados demostraron que CofB es la pilina tipo IV más grande reportada hasta la fecha (50kDa). A pesar de la búsqueda de homología con otras pilinas usando bases de datos, CofB no mostró homología con otras pilinas, solo el segmento N-terminal hidrofóbico de CofB demostró homología de secuencia con la de otras pilinas tipo IV, incluyendo la glicina de la posición 5 (Kawahara *et al.*, 2015)

3 ANTECEDENTES

Este trabajo tiene por modelo de estudio las pilinas menores BfpI, BfpJ y BfpK de EPEC. Para conocer la probable función de las pilinas menores en la mayoría de los casos se resuelve mediante la obtención de la estructura de las proteínas por cristalografía de

rayos X cuando a las pilinas se les elimina la porción hidrofóbica N-terminal, además de realizar mutaciones en los genes de dichas proteínas.

Los primeros estudios realizados para conocer las funciones de las pilinas menores se conocen en *Neisseria meningitidis* con la pilina PilX. Su estructura terciaria se obtuvo mediante cristalografía de rayos X y para eliminar la porción hidrofóbica truncaron los primeros 28 aminoácidos de la proteína madura (Helaine *et al.*, 2007). Su estructura muestra un plegamiento α/β y se localiza dentro del pili tipo IV, pero se observa en menor abundancia con respecto a la pilina mayor por lo que es una pilina menor. Por otro lado, PilX expone sus motivos estructurales hacia la superficie estabilizando la agregación bacteriana (Helaine *et al.*, 2007).

Por otra parte, *Escherichia coli* enteropatógena típica es otro modelo de estudio interesante debido a que interactúa con las células epiteliales del intestino de manera distinta de lo observado en otras bacterias como *Salmonella* o *Shigella*. tEPEC no es un patógeno invasivo y la interacción inicial mediante la expresión de BFP es el proceso de colonización donde ocurre la adherencia localizada, fenotipo típico de tEPEC (Scaletsky *et al.*, 1984). Se ha demostrado que las proteínas menores de tEPEC, BfpI, BfpJ y BfpK se incorporan dentro del filamento BFP y que tienen relación con la integridad estructural del filamento y función de adherencia en las células hospederas (Martínez de la Peña *et al.*, 2016). Estos resultados se comprobaron mediante la creación de cepas tEPEC mutantes en los genes *bfpI*, *bfpJ* y *bfpK* donde estas no produjeron BFP. Pero al crear dobles mutantes en los genes *bfpI*, *bfpJ*, *bfpK* en un fondo mutante en *bfpF* (este último gen está relacionado con la retracción de BFP), se observaron que estas sí expresan el BFP de diferentes maneras (Martínez de la Peña *et al.*, 2016). Las diferencias en los fenotipos entre el BFP expresado

en las dobles mutantes con respecto a la cepa silvestre muestra que la doble mutante *bfpF/bfpI* expresa un pili más largo y más flexible que la cepa silvestre, la doble mutante *bfpF/bfpJ* expresaba un pili corto fragmentado, solamente detectado por el marcaje con oro usando un anticuerpo anti-BFP y en la doble mutante *bfpF/bfpK*, expresa BFP parecido al expresado en la cepa silvestre (Martínez de la Peña *et al.*, 2016).

El análisis funcional de los pili expresados por las dobles mutantes usando el ensayo de adherencia localizada (eLA) a células HEp-2, revela que la doble mutante en BfpI presenta una adherencia disminuida a pesar de poseer un pili más largo. La mutante *bfpJ/bfpF* también se ve afectada en la adherencia a las células HEp-2 posiblemente por fragilidad observada en la microscopía. La mutante *bfpF/bfpK* se adhirió a las células HEp-2 mejor que *bfpF/bfpI* o *bfpF/bfpJ* pero menos que las mutantes solas en *bfpF*. Por otro lado, la mutante cuádruple en *bfpIJKF* revela producción de BFP y una adherencia disminuida pero al complementarla con la proteína BfpI la adherencia se restaura a niveles parecidos a la adherencia por la cepa silvestre, sugiriendo de esta manera que la proteína BfpI pudiera estar funcionando como adhesina alternativa (Martínez de la Peña *et al.*, 2016).

Estos estudios sugieren que la subunidad BfpI puede tener una función como adhesina accesorio o podría orientar adecuadamente las subunidades BfpA de tal forma que sus dominios de unión al receptor pueden establecer relación con receptores de célula huésped, la función de BfpJ y BfpK no es muy clara pero se propone que quizá BfpJ podría estar involucrado en la flexibilidad o resistencia a la tracción en BFP y a la estabilidad del filamento y BfpK podría tener función reguladora en la biogénesis ya sea a nivel genético y / o en el montaje BFP aunque falta por esclarecer cual pudiera ser el mecanismo (Martínez de la Peña, 2015).

4 JUSTIFICACIÓN

E. coli enteropatógena (EPEC) es un patógeno diarreagénico que afecta a niños menores de seis meses a nivel mundial y principalmente países en vías de desarrollo donde se han observado los casos más graves que se asocian a mortalidad. Por lo tanto, es necesario desarrollar antimicrobianos o vacunas específicas para EPEC que ayuden a disminuir los brotes diarreicos y la tasa de mortalidad asociada a ella. Las pilinas menores, podrían ser la clave para el desarrollo de estas terapias, debido a que estas se encuentran formando el pili tipo IV o BFP y poseen menor variabilidad antigénica en comparación con la pilina mayor. El conocimiento del papel que desempeñan estas subunidades menores en el ensamble y desensamble del pili tipo IV y su función específica, pueden sustentar las bases para el desarrollo una vacuna o un antimicrobiano. En el presente estudio nuestra aportación será el poder purificar en forma soluble a las proteínas menores BfpI, BfpJ y BfpK para su posterior cristalización y análisis de actividad de lectina para ayudar a elucidar su probable función en la patogénesis de EPEC.

5 OBJETIVOS

5.1 General

Clonación, expresión y purificación de las proteínas BfpI, BfpJ y BfpK, que están presentes en el pili formador de mechones de *Escherichia coli* enteropatógena.

5.2 Específicos

1. Diseñar los cebadores específicos para amplificar los genes de *bfpI*, *bfpJ* y *bfpK* sin la región hidrofóbica N-terminal de la proteína madura.
2. Clonar las proteínas BfpI, BfpJ y BfpK en el vector pET3-a y transformar en la cepa *E. coli* HB101 y BL21.
3. Purificar las proteínas BfpI, BfpJ y BfpK clonadas por medio de una columna de agarosa Ni-NTa en condiciones nativas y desnaturalizantes.
4. Verificación de las proteínas purificadas por Western-Blot usando anticuerpos anti-His.

5. HIPÓTESIS

Al eliminar los primeros 39, 27 y 33 aminoácidos de la región N-terminal de las prepilinas BfpI, BfpJ y BfpK respectivamente se obtienen proteínas solubles para cristalización de rayos X.

6 MATERIALES Y MÉTODOS

6.1 Material biológico

Las cepas y el plásmido que se utilizaron en el estudio se especifican en la **tabla 1** y **2** respectivamente.

Tabla 1 Cepas utilizadas y sus características.

Cepas	Características	Referencias
E2348/69	EPEC cepa silvestre serotipo O127: H6	(Levine <i>et al.</i> , 1978)
HB101	Es un híbrido de la cepa <i>Escherichia coli</i> K12 y la <i>Escherichia coli</i> B, tiene una mutación que minimiza la recombinación.. Esta cepa es sensible a ampicilina,	(Alliance Bio, 2007)

	cloranfenicol, kanamicina y tetraciclina, pero resistente a estreptomicina.	
MC 4100	Es una cepa utilizada como huésped para expresión génica.	(Peters <i>et al.</i> , 2003)
BL21(DE3)	Esta cepa es deficiente en proteasas de membrana externa <i>ompT</i> y proteasa Lon. Tiene un gen que codifica para una RNA polimerasa T7 bajo inducción con IPTG.	(Thermo Fisher Scientific, 2010)
Rosetta-gami B(DE3)	Potencia la expresión de proteínas eucarióticas además de la formación de puentes disulfuro en el citoplasma bacteriano. Esta cepa es compatible con vectores con cassette de resistencia a ampicilina. La cepa tiene resistencia a kanamicina, tetraciclina y cloranfenicol.	(Novagen, 2004)

Tabla 2 Plásmido utilizado y sus características

Plásmido	Características	Referencias
pET-3a	Es un vector de clonación y de expresión de proteínas recombinantes en <i>Escherichia coli</i> . Tiene un promotor T7. La expresión de las proteínas clonadas es inducida con IPTG. Figura 7	(Agilent, 2015)

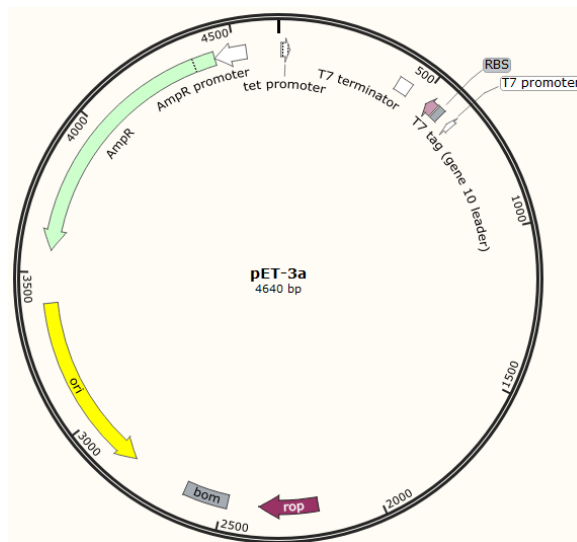


Figura 7. Mapa del plásmido pET-3a. Plásmido con un cassette de resistencia a ampicilina, un origen de replicación y un promotor T7. Imagen obtenida con el programa SnapGene viewer.

6.2 Condiciones de crecimiento

Las cepas *E. coli* E2348/69, HB101 y BL21 se crecieron en caldo y agar Luria-Bertani (LB) medio que contiene peptona de caseína, extracto de levadura y cloruro de sodio y se incubaron a 37°C. Se adicionó ampicilina a una concentración final de 100 µg/ml cuando el ensayo lo requería.

6.3 Extracción plasmídica por el método Quick prep

Se creció la cepa de interés en un tubo con 3 ml de caldo LB a 37°C en agitación a 200 rpm toda la noche. Al día siguiente en un tubo eppendorf se centrifugó 1.5 ml, del caldo crecido, a 12 000 rpm por 2 min, la pastilla obtenida se resuspendió con 150 µL de solución I (glucosa 50 mM, Tris-Cl 25 mM a pH 8 y EDTA 10 mM), 2 µL de RNAsa y 0.0003 g de lisozima y se dejó incubar cinco minutos a temperatura ambiente. Posteriormente se adicionó 150 µL de SDS al 2% y 150 µL de NaOH al 0.4%, se invirtió suavemente y se dejó incubar en hielo cinco minutos. A continuación se agregó 225 µL de acetato de amonio (7.5 M), se invirtió e incubo en hielo por diez minutos; pasando ese tiempo se centrifugó a 12000 rpm por 15 minutos y el sobrenadante se pasó a un tubo eppendorf limpio de 1.5 ml y se le adicionó 600 µL de fenol-cloroformo (en relación 1:1) y se mezcló suavemente por inversión 50 veces, a continuación se centrifugó a 12 000 rpm por 5 minutos, se transfirió la fase superior del centrifugado a un tubo eppendorf con 600 µL de cloroformo y se mezcló suavemente por 100 veces y después se centrifugó a 12 000 rpm por 5 minutos. La fase superior del sobrenadante se transfirió a un tubo limpio eppendorf con 1 ml de etanol absoluto, se llevó a incubar a -70°C (en Revco) por 30 minutos. A continuación se centrifugó a 12 000 rpm por 15 minutos, se desechó el sobrenadante y después se agregó 1 ml de etanol al 70% y se centrifugó a 12 000 rpm por 3

minutos, se desechó el sobrenadante y se invirtió el tubo eppendorf hasta que se evaporó el etanol de las paredes del tubo. Por último se adicionó 50 µL de agua pisa para resuspender el ADN y mantenerse a -20°C.

6.4 Extracción plasmídica por Kit Zyppy™ Plasmid Miniprep

Se realizó conforme al protocolo establecido por el Kit Zyppy™ de Zymo research Plasmid Miniprep.

6.5 Diseño de primers para amplificar los genes *bfpI*, *bfpJ* y *bfpK*

Se obtuvieron las secuencias de los genes *bfpI*, *bfpJ* y *bfpK* del GenBank. A partir de las secuencias se diseñaron los primers, con el propósito de amplificar los genes sin la región hidrofóbica de las proteínas de BfpI, BfpJ y BfpK. Para el caso de BfpI se omitió una región hidrofóbica de 27 aminoácidos de la proteína madura o 39 aminoácidos de la prepilina que son 117pb a partir del ATG. Quedando los primers diseñados a partir del aminoácido 28 de la pilina madura. El gen amplificado esperado fue de 453 pb. Para BfpJ se omitió la región hidrofóbica con una longitud de 20 aminoácidos de la proteína madura o 27 aminoácidos de la prepilina que son 81 pb a partir del ATG. El fragmento esperado del gen amplificado fue de 501 pb. Para BfpK se omitió una región hidrofóbica con una longitud de 23 aminoácidos de la proteína madura o 33 aminoácidos de la prepilina que son 99 pb de la secuencia del gen a partir del ATG. El fragmento amplificado del gen esperado fue de 417 pb. Cada uno de los primers diseñados incluyó sitios de restricción para *NdeI* en el primer delantero y *BamHI* en el reverso (CA↓TATG Y G↓GATCC respectivamente) además de una etiqueta para codificar 6 histidinas en el primer delantero que amplifica la región amino terminal de la secuencia de las proteínas (CACCATCACCATCACCAT). Los

oligos se revisaron mediante el programa SnapGene Viewer de acuerdo a las especificaciones de uso del programa.

6.6 Diseño de primers del vector pET-3a

Se obtuvo la secuencia del vector pET-3a de la página de Addgene The nonprofit plasmid repository (Addgene, 2016). Los primers se diseñaron río arriba y río debajo del sitio de reconocimiento para las enzimas de restricción *NdeI* y *BamHI*, sitio de integración del inserto (*BfpI*, *bfpJ* o *bfpK*) la longitud de los primers es de 23 y 24 bases (forward y reverso respectivamente). Mediante el programa Snap gene viewer se visualizó la posible hibridación de los primers forward y reverso en la secuencia lineal del vector pET-3a. El primer forward se diseñó para alinearse río abajo del sitio de reconocimiento para la enzima *NdeI*, el primer reverso se diseñó para alinearse río arriba del sitio de reconocimiento para la enzima *BamHI*. El amplificado esperado es de 197 pb aproximadamente.

6.7 Amplificación por PCR

Las condiciones de PCR utilizando los primers diseñados fueron las siguientes: una desnaturalización inicial a 94 °C durante 5 minutos, seguida de 30 ciclos iniciando con una desnaturalización a 94 °C durante 40 segundos, alineamiento a 60 °C durante 40 segundos (*bfpI* y *bfpJ* para *bfpK* fue de 70°C), extensión a 72 °C durante 1 minuto y 10 segundos, y un periodo final de alargamiento a 72 °C durante 10 minutos. La mezcla de reacción se preparó usando como templado ADN plasmídico 1 μL , oligo forward 1 μL , oligo reverso 1 μL , dNTPS 1 μL , 5 μL de buffer (MgSO_4), 1 μL de enzima PFU(Thermo Fisher Scientific) y agua PISA 40 μL

6.8 Clonación en vector pET-3a

Por un lado se realizó doble restricción con las enzimas *BamHI* y *NdeI* (Thermo Fisher Scientific FastDigest) de los productos de PCR de *bfpI*, *bfpJ* y *bfpK* y del vector pET3a. Para la restricción del vector se preparó una mezcla de reacción con agua estéril (PISA) Buffer Green FastDigest, 1 μ L de enzima *BamHI* y 1 μ L de enzima *NdeI* y 1 μ L de ADN del vector y se incubó 5 min a 37°C a baño María, según las instrucciones del fabricante (Thermo Fisher Scientific FastDigest). Por otra parte en otra mezcla de reacción con agua estéril (PISA), Buffer Green FastDigest, 1 μ L de enzima *BamHI* y 1 μ L de enzima *NdeI*, se agregó 1 μ L del producto de PCR y se incubó 5min a 37°C a baño María, según las instrucciones del fabricante (Thermo Fisher Scientific FastDigest). Una vez digeridos se midió la concentración del vector y del inserto en el equipo Nano Drop 2000 según las especificaciones del fabricante, para realizar los cálculos para la ligación usando la relación 1:3 (inserto: vector digerido relación molar). $\frac{(ng \text{ de vector } \times \text{ tamaño del inserto en KB})}{\text{tamaño del vector en kb}} \times \frac{3}{1} = ng \text{ del inserto}$, siguiendo la formula, la mezcla de reacción se llevó a cabo con 1 μ L de enzima T4ligasa (ThermoFisher Scientific), 2 μ L buffer 10x, a continuación 14, 13 y 12 μ L de agua estéril (PISA) (para la mezcla de *bfpI*, *bfpJ* y *bfpK* respectivamente), 2 μ L de vector pET-3a y 1, 2 y 3 μ L del inserto (para la mezcla de *bfpI*, *bfpJ* y *bfpK* respectivamente) a un volumen final de 20 μ L cada reacción para BfpI-pET3a, BfpJ-pET3a y BfpK-pET3a a una temperatura de 14 - 25°C toda la noche (Figura 8).

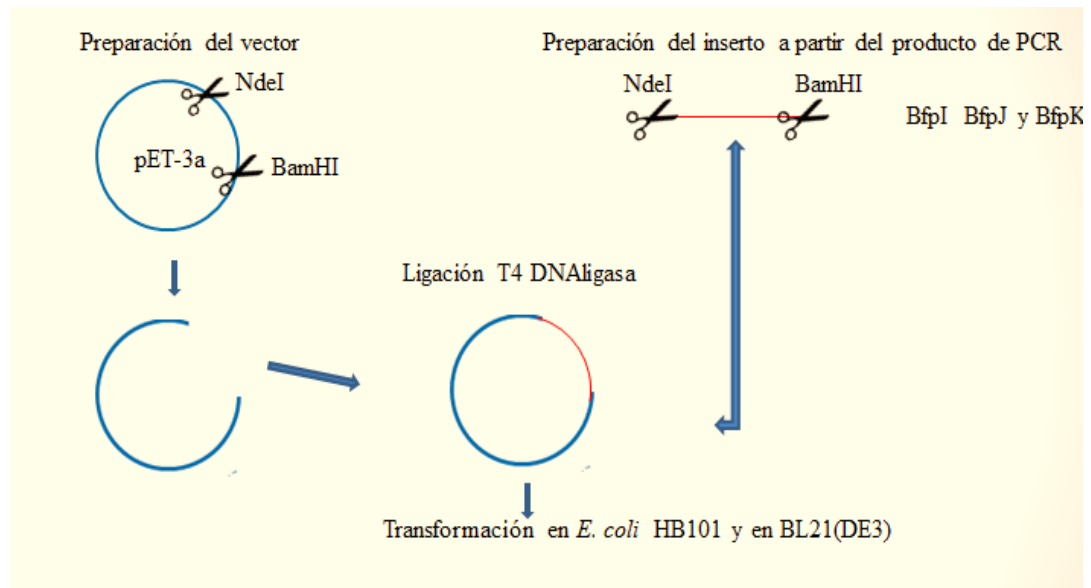


Figura 8. Esquema de doble restricción. Productos de PCR genes *bfpI*, *bfpJ* y *bfpK* así también del vector comercial pET-3a.

6.9 Preparación de células químicamente competentes

Se inoculó un tubo de 5 ml con caldo LB con la cepa HB101 o BL21, según el caso, se dejó crecer toda la noche 37°C con agitación a 200 rpm (preinóculo). Al día siguiente se inoculó un matraz con 100 ml con caldo LB con 1ml del preinóculo y se dejó crecer a 37°C con agitación a 200 rpm hasta alcanzar una D.O.₆₀₀ de 0.4-0.6. Una vez alcanzada la D.O. se recuperaron las células centrifugando el cultivo en tubos para ultracentrífuga a 4 000 rpm por 5 minutos. A continuación se decantó el sobrenadante y la pastilla de células se resuspendió con 30 ml de la solución TFB1 (100 mM RbCl, 50 mM MnCl₂, 30 mM acetato de potasio, 10 mM CaCl₂ y glicerol al 15%, ajustado a pH 5.8, esterilizada por filtración) dejándose incubar en hielo 90 minutos. Pasado ese tiempo se centrifugó a 4000 rpm por 5 minutos, se decantó el sobrenadante y la pastilla de células se resuspendió con 4ml de la solución TFB2 (10 mM MOPS, 10 mM RbCl, 75 mM CaCl₂,

glicerol al 15%, ajustado a pH 6.8 con KOH, la solución se esteriliza por filtración), se hicieron alícuotas de 500 μ L y 500 μ L de glicerol para almacenar a -70°C .

6.10 Transformación por choque térmico

Se descongeló una alícuota de las células químicamente competentes, y en un tubo eppendorf de 1.5 ml se agregaron 100 μ L de células competentes y 5 μ L de mezcla de ligación, esta mezcla se mantuvo en hielo por 20 minutos. A continuación se incubó 20 segundos en baño María a una temperatura de 42°C y se mantuvo inmediatamente en hielo por 2 minutos, pasado ese tiempo se le adicionó 1 ml de medio en caldo LB y se incubó por 90 minutos a 37°C en agitación constante a 200 rpm. A continuación se sembraron con varilla de vidrio en placas de agar LB adicionado con ampicilina a una concentración de 100 $\mu\text{g/ml}$ y se dejaron incubar toda la noche a 37°C .

6.11 Preparación de células electrocompetentes y electroporación

Se inoculó un tubo de 5 ml con caldo LB, con la cepa HB101 o BL21, se dejó crecer toda la noche 37°C con agitación a 200 rpm (preinóculo). Al día siguiente se inoculó un matraz con 100 ml con caldo LB con 1 ml del preinóculo y se dejó crecer a 37°C con agitación a 200 rpm, hasta alcanzar una D.O.₆₀₀ de 0.4 - 0.6, una vez que alcanzó la D.O. se recuperaron las células por centrifugación del cultivo, en tubos para ultracentrífuga, a 4 000 rpm por 5 minutos. A continuación se decantó el sobrenadante y la pastilla de células se resuspendió con 30 ml agua tridestilada estéril, esto se repitió una vez más, posteriormente se centrifugó a 4 000 rpm por 5 minutos, se decantó el sobrenadante y la pastilla de células se resuspendió con 4ml de agua tridestilada estéril, esto se repitió tres veces, finalmente el paquete celular se resuspendió en 200 μ L de agua tridestilada estéril y se transfirió 100 μ L de células directamente a la cubeta de

electroporación estéril junto con 5 μ L de la mezcla de ligación. Esta celda se transfirió al Electroporador, Multiporator®, Eppendorf; donde se electroporó con una tensión de impulsos de 200–2500 V, y una constante de tiempo de 5 ms. A continuación se sembraron con varilla de vidrio en placas de agar LB adicionado con ampicilina a una concentración de 100 μ g/ml y se dejaron incubar toda la noche a 37°C.

6.12 Expresión de las proteínas BfpI, BfpJ y BfpK.

Las clonas con las construcciones pET3a-BfpI, pET-3a-BfpJ se sometieron a extracción de ADN plasmídico, por método Quickprep. La verificación de las construcciones se realizó por PCR, doble restricción con las enzimas de restricción *NdeI* y *BamHI*, y por secuenciación.

Una vez comprobadas las construcciones se transformó en la cepa BL21 (DE3) y se expresaron las proteínas. Se creció la cepa BL21 (DE3) en caldo LB, adicionado con ampicilina a 100 μ g/ml, se dejó crecer de 2 a 3 horas a 37°C con agitación a 200 rpm hasta alcanzar una D.O. de 0.5 - 0.7 a una longitud de onda de 600 nm. Después se indujeron con IPTG a una concentración final de 1 mM y se dejaron crecer cuatro horas más, posteriormente se recolectaron las células por centrifugación a 4 000 rpm por 20 minutos. Las células se lisaron bajo condiciones desnaturalizantes con urea 8 M agregando 1 ml por cada 50 ml de cultivo por 15 minutos seguidamente se sonicaron a una amplitud de 60 volts por 1 minuto a repeticiones de 10 segundos, posteriormente se centrifugó a 10,000 rpm por 25 minutos. El extracto bacteriano se depositó en una columna de resina Ni-NTA, siguiendo las instrucciones del fabricante (The QIAexpressionist™) (Figura 9) y se corroboró la expresión por SDS-PAGE.

6.13 Determinación de solubilidad

Se inoculó un tubo con 5 ml caldo LB conteniendo 100 µg/ml de ampicilina. Se dejó crecer toda la noche a 37°C con agitación constante. Posteriormente se inoculó 50 ml de caldo LB adicionado con ampicilina a una concentración de 100 µg/ml de ampicilina con 2.5 ml del preinóculo que se creció la noche anterior, el cultivo se dejó crecer a 37°C en agitación constante hasta llegar a una D.O.₆₀₀ de 0.5 - 0.7. Inmediatamente se tomó 1 ml de muestra antes de la inducción (el ml se centrifugó a 4 000 rpm por 20 min y el pellet se resuspendió con 50 µL de SDS al 0.1%) que se realizó agregando IPTG a una concentración final de 1 mM. Se dejó crecer 4-5 horas más y se colecta por centrifugación 1 ml de muestra (el ml se centrifugó a 4 000 rpm por 20 min y el pellet se resuspendió con 100 µL de SDS al 0.1%) y el cultivo restante a 4 000 rpm por 20 minutos, el pellet se mantuvo a 4°C.

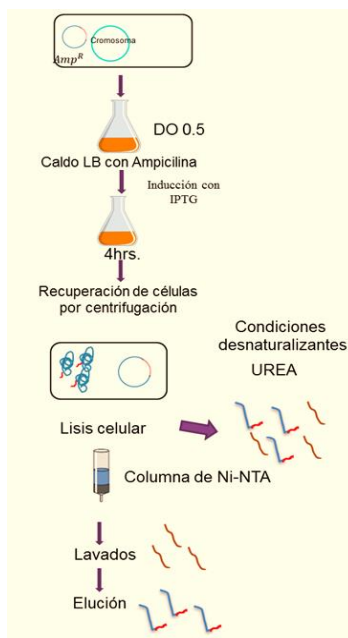


Figura 9. Esquema de expresión y purificación de las proteínas. La inducción se realizó con IPTG y la purificación bajo condiciones desnaturalizantes con urea.

Para la extracción de la proteína, el pellet se resuspendió en 5 ml de buffer de lisis (100 mM NaH_2PO_4 , 10 mM Tris·Cl, 6 M GuHCl, ajustado a pH 10) y a continuación se sonóico 2x 1 minuto con pausas de 10 segundos a 60 volts. El lisado se mantuvo en hielo todo el tiempo. Se centrifugó el lisado a 10 000 rpm por 25 minutos a 4°C y el sobrenadante se decantó (este se guardó a 4°C ya que es la fracción soluble, para correr la muestra en un gel de poliacrilamida). Por otra parte se resuspendió el pellet en 5 ml de buffer de lisis, esta fracción es la materia insoluble o proteína insoluble. Se almacena a 4°C para su corrimiento en gel de poliacrilamida.

6.14 Purificación y diálisis de las proteínas BfpI, BfpJ y BfpK

Las proteínas se purificaron bajo el protocolo de condiciones desnaturizantes esto se realizó de la siguiente manera: en una columna se agregó 2.5 ml de resina Thermo Scientific HisPur Ni-NTA. Una vez que se asentó la resina en la columna se hicieron 5 lavados (cada uno agregando 10 ml) con agua tridestilada. A continuación se hicieron 6 lavados (cada uno agregando 10 ml) con buffer B (100 mM NaH_2PO_4 , 10 mM Tris·Cl, 8 M urea, se llevó a pH final de 8 con NaOH y se aforó a un litro). Después se agregó el lisado (las células que se lisaron con urea 8 M) y se mantuvo 30 minutos en rotación constante para que las proteínas con la etiqueta de histidinas se logaran fijar de mejor manera a la resina. Pasado ese tiempo se recuperó la elución. Después se realizaron 3 lavados con el buffer B (pH 8). En el primer lavado se recuperó la elución de este (1 ml) y se mantuvo a 4°C. Después se hicieron 10 lavados (cada uno agregando 10 ml) con el buffer C (100 mM NaH_2PO_4 , 10 mM Tris·Cl, 8 M urea, se llevó a pH final de 6.3 con NaOH y se aforó a un litro).

En cada lavado se recuperó la elución de este (1ml) y se mantuvo a 4°C. Después se hizo 1 lavado con el buffer E (pH 4.5) (100 mM NaH_2PO_4 , 10 mM Tris-Cl, 8 M urea, se llevó a pH final de 4.5 con HCl y se aforó a un litro). En este lavado se recuperó por completo los 10 ml del lavado (proteína eluída). Después de la purificación en las columnas de resina Ni-NTA, las proteínas se dializaron cuatro veces en un buffer (0.1% Tritón X-100, 0.01 M $MgCl_2$, 0.05 M Tris 1M, 20% glicerol, 0.5 M NaCl) con una concentración inicial de urea de 6 M, 4 M, 1 M y una final 0 M. La membrana se mantuvo por 12 horas en agitación constante con 500 ml del buffer con concentración inicial de urea de 6 M, pasado ese tiempo se cambió el buffer por el de concentración 4M, así sucesivamente hasta eliminar la urea.

6.15 Determinación de la concentración de las proteínas

Se llevó a cabo por la absorbancia de la muestra usando el método de BCA (Ácido Bicinconínico de Thermo Fisher Scientific) (200 μ L de BCA y 4 μ L de reactivo B $CuSO_4$ y 10 μ L de las proteínas, se realiza este paso para cada proteína), tomando como referencia las concentraciones crecientes de albumina sérica bovina (BSA), 0.5 μ g/ml, 20 μ g/ml y 40 μ g/ml a una longitud de onda de 596 nm, el valor obtenido se interpoló en la curva de BSA y se multiplicó por el factor de dilución.

6.16 Electroforesis en geles de poliacrilamida

Los geles de poliacrilamida se prepararon al 12.5% (2.5 ml de Tris pH 8.8 (1.875 M), 4.1 ml de Bisacrilamida (37:1), 3.3 ml de agua tridestilada, 100 μ L de EDTA (0.2 M), 20 μ L de TEMED y 50 μ L de APS al 10%). La electroforesis se llevó a cabo a 120 volts durante 1 hora, en seguida se tiñó con azul de Coomasie al 1%. Para poder observar las bandas de las proteínas, los geles se lavaron con una solución

de desteñido de ácido acético al 10% y metanol al 50% con agitación constante hasta lograr observar el patrón electroforético de las proteínas.

6.17 Análisis de Western Blot

Después de realizar el corrimiento electroforético de las proteínas purificadas en los geles PAGE-SDS al 12.5%, se sumergió en solución de transferencia (Tris-glicina con un 20% de metanol) y la membrana de PVDF se sumergió en metanol durante 10 segundos, las proteínas se transfirieron a membranas empleando el aparato de electrotransferencia (Trans-Blot SD Semi-Dry Transfer-Cell Bio-Rad).

La transferencia se realizó a 15 volts durante 15 minutos, la membrana se bloqueó con leche descremada al 5% en TBST (Tris-buffer salino y tween 20) durante toda la noche. Los anticuerpos anti-His fueron diluidos 1/3000 en TBST y se incubaron junto con la membrana de PVDF por una hora a temperatura ambiente, seguida de 3 lavados con TBST por 10 minutos. El anticuerpo secundario HRP (horseradish peroxidase) conjugado anti-conejo se diluyó 1:80000 en TBST y se incubó 45 minutos a temperatura ambiente. El revelado se efectuó siguiendo las instrucciones del protocolo Amersham ECL prime Western blotting detection reagent.

7 RESULTADOS

7.1 Diseño de primers para amplificar los genes *bfpI*, *bfpJ* y *bfpK*

Con el fin de obtener las proteínas solubles BfpI, BfpJ y BfpK en este trabajo se diseñaron primers específicos para amplificar los genes *bfpI*, *bfpJ* y *bfpK* sin la porción hidrofóbica N-terminal (Materiales y métodos) y posteriormente clonarlos en el vector pET-3a. Mediante el programa SnapGene Viewer se observó la posible hibridación de los primers con la secuencia de los genes a amplificar *bfpI*, *bfpJ* y *bfpK* (Figura 10, 11 y 12 respectivamente).

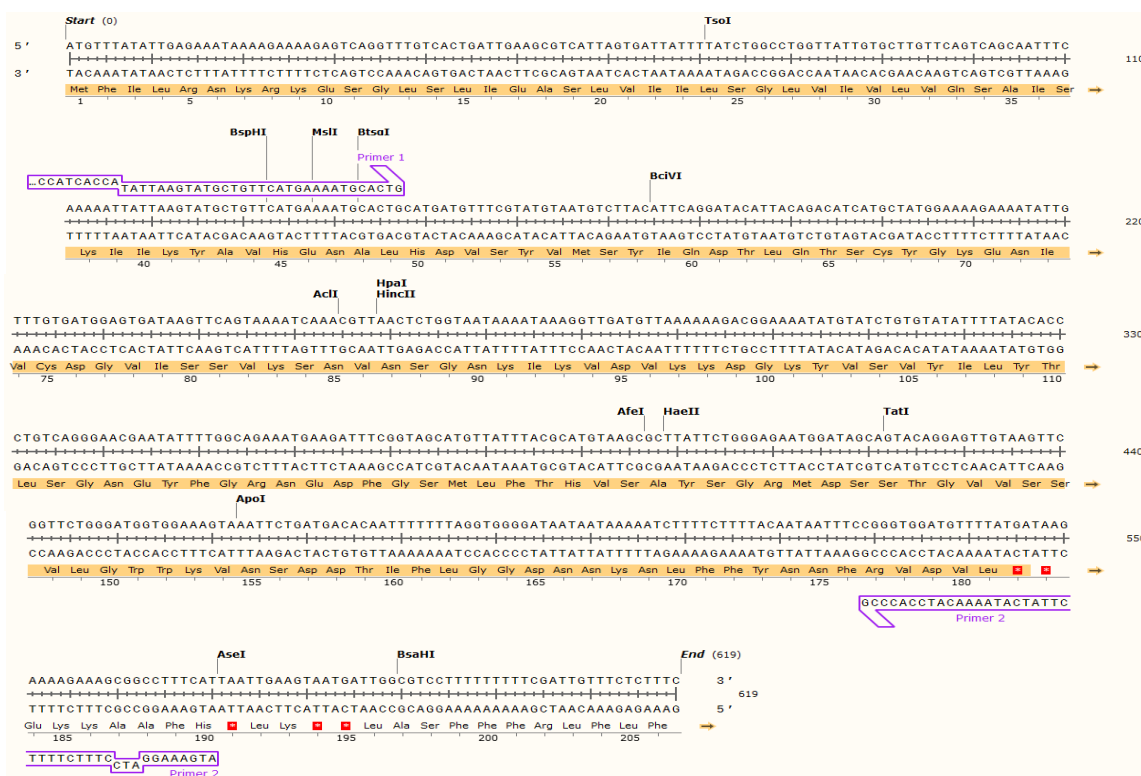


Figura 10. Secuencia de BfpI. Imagen del programa SnapGene Viewer que muestran la secuencia de doble cadena y el marco abierto de lectura de BfpI. Se observa en morado el sitio probable en donde hibridan los primers diseñados para BfpI.

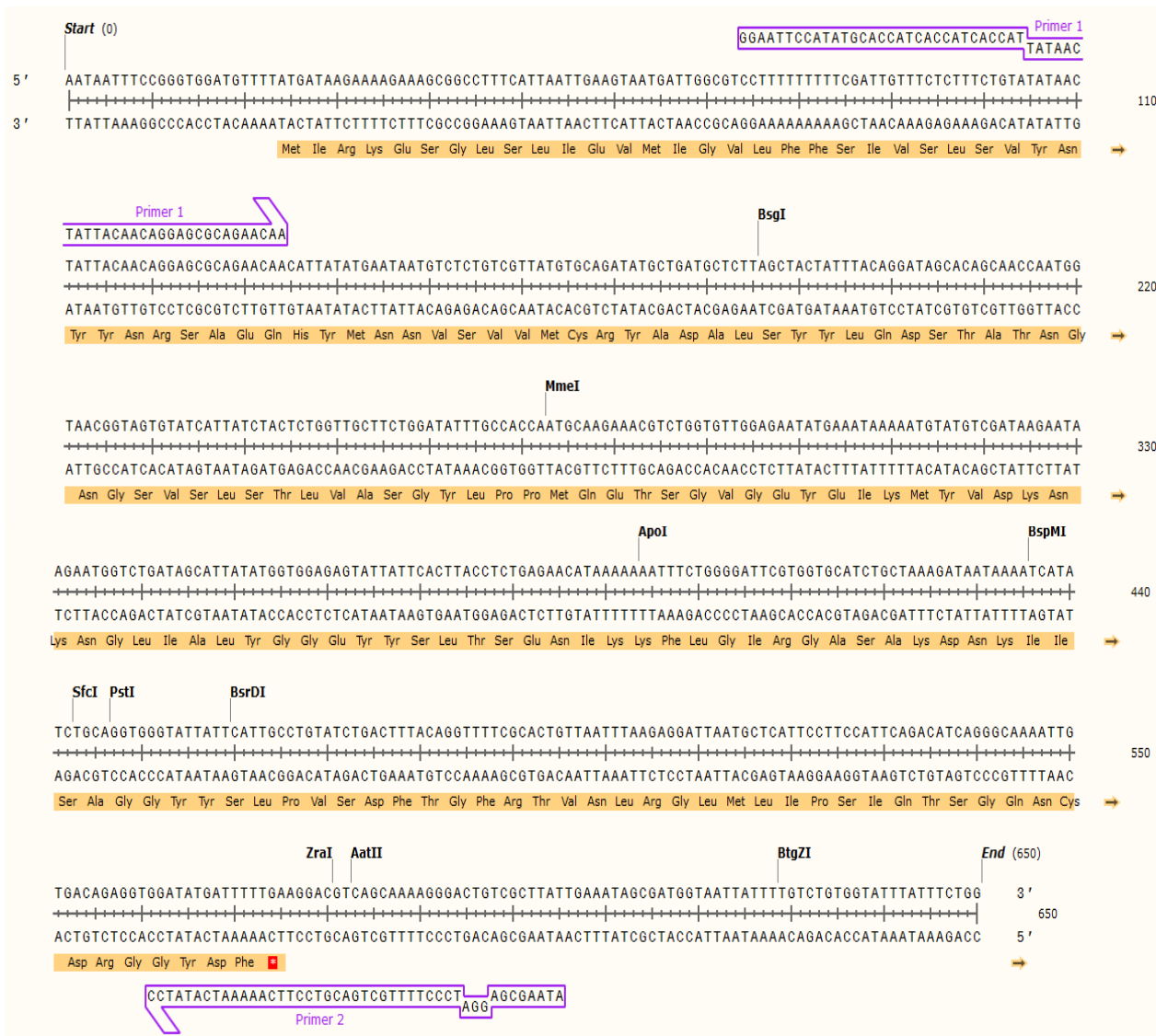


Figura 11. Secuencia de BfpJ. Imagen del programa SnapGene Viewer que muestran la secuencia de doble cadena y el marco abierto de lectura de BfpJ. Se observa en morado el sitio probable en donde hibridan los primers diseñados para BfpJ



Figura 12. Secuencia de BfpK. Imagen del programa SnapGene Viewer que muestran la secuencia de doble cadena y el marco abierto de lectura de BfpK. Se observa en morado el sitio probable en donde hibridan los primers diseñados para BfpK.

Con el objetivo de clonar y purificar las proteínas expresadas los primers se diseñaron con sitios de restricción para las enzimas *NdeI* y *BamHI*, además de una etiqueta de histidinas (Figura 13).

BfpI
FORWARD
5'-GGAATTCCATATGCACCATCACCATCACCATATTAAGTATGCTGTTTCATGAAAATGCACTG-3'
Longitud 61 bases
REVERSE
5'-ATGAAAGGATCCTTTCTTTTCTTATCATAAAACATCCACCCG-3'
Longitud 42 bases

BfpJ
FORWARD
5'-GGAATTCCATATGCACCATCACCATCACCATTATAACTATTACAACAGGAGCGCAGAACAA-3'
Longitud: 61bases
REVERSE 5'-ATAAGCGAGGATCCCTTTTGCTGACGTCCTTCAAAAATCATATCC-3'
Longitud: 45 bases

BfpK
FORWARD
5'-GGAATTCCATATGCACCATCACCATCACCATATATCTCTTTTCAAGGTTTTTGCTCAATCT-3'
Longitud: 61 bases
REVERSE 5'-AAAAACGGATCCAGAAGAATAACATAACTAGTACCGTTTAAACAG-3'
Longitud: 45 bases

Figura 13. Secuencias de los primers sintetizados para BfpI, BfpJ y BfpK. Rojo sitios de restricción para *NdeI*, en azul etiqueta de histidinas y en verde sitio de restricción para la enzima *BamHI*

Por otra parte para verificar las construcciones (BfpI-pET3a, BfpJ-pET3a, y BfpK-pET3a) mediante secuenciación se diseñaron primers para amplificar río arriba y río abajo del sitio de integración del inserto en el vector pET-3a (Materiales y métodos). Los primers diseñados se observan en la Figura 14 y en la Figura 15 se visualiza su posible hibridación con la secuencia del vector mediante el programa SnapGene Viewer.

FORWARD pET3a
 5'-GGGAGACCACAACGGTTTCCCTC-3'
 Long.23 bases
 REVERSE pET3a
 5'-GTCCCGTTTAGAGGCCCAAGGGG-3'
 Long.24 bases

Figura 14. Primers diseñados para el vector pET-3a.

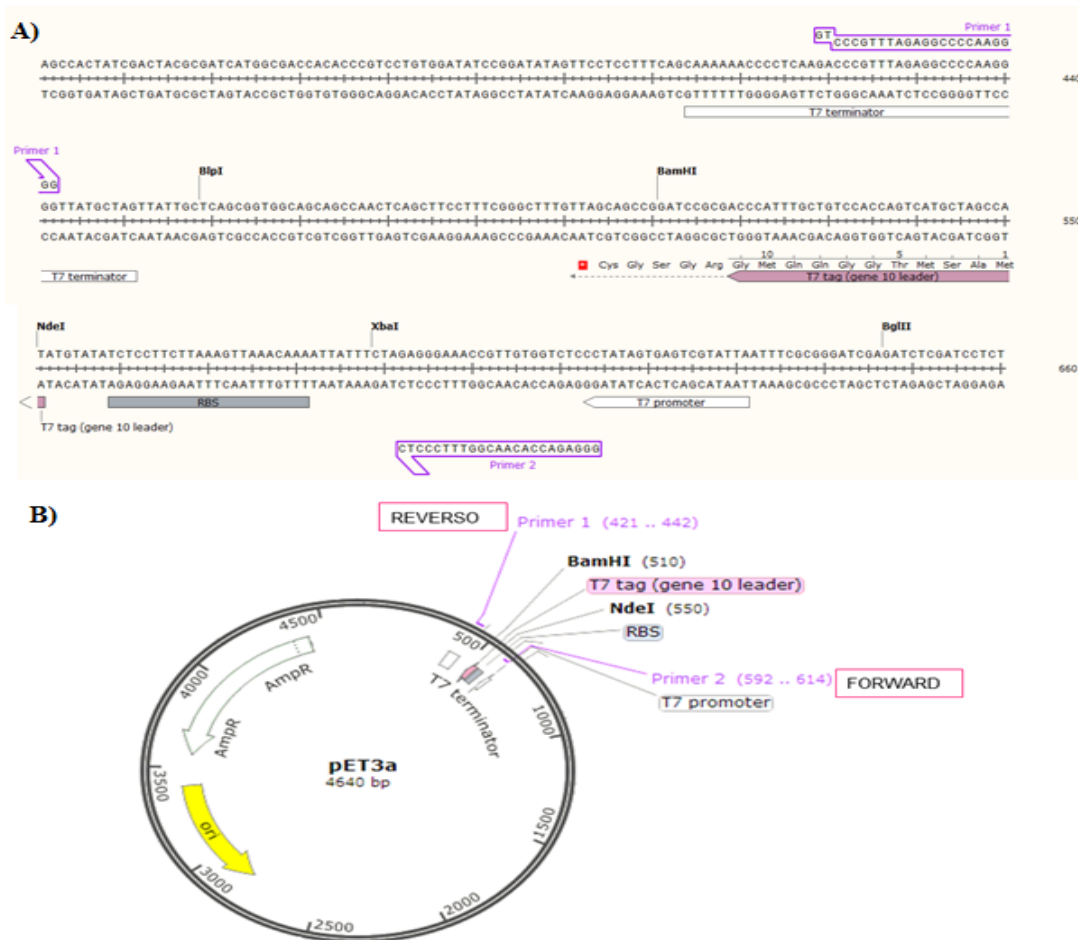


Figura 15. Visualización mediante el programa SnapGene Viewer de la posible hibridación de los primers diseñados para el vector pET-3a

A) Secuencia lineal del vector pET-3a. B) Mapa de la secuencia del vector pET-3a.

7.2 Amplificación de los genes *bfpI*, *bfpJ* y *bfpK*

El principal objetivo de este estudio es obtener las proteínas BfpI, BfpJ y BfpK de manera soluble, para cual se amplificaron los genes de dichas proteínas sin la porción N-terminal hidrofóbica (Materiales y métodos).

Se amplificaron los genes *bfpI*, *bfpJ* y *bfpK* utilizando los oligos, las condiciones y parámetros descritos en Materiales y métodos. Los amplificados se observaron mediante geles de agarosa al 1%. En la Figura 16 (A y B) se observan los amplificados de *bfpI* y *bfpJ* con pesos aproximados de 453 y 501 pb respectivamente. Para *bfpK* las condiciones de amplificación fueron diferentes, se amplificó con una T_m de 70°C, temperatura optima obtenida después de amplificar en gradiente de temperatura desde 58°C a 70°C donde el mejor amplificado fue a la temperatura de 70°C. El producto de PCR se corrió en un gel de agarosa al 1% Figura 16 (C).

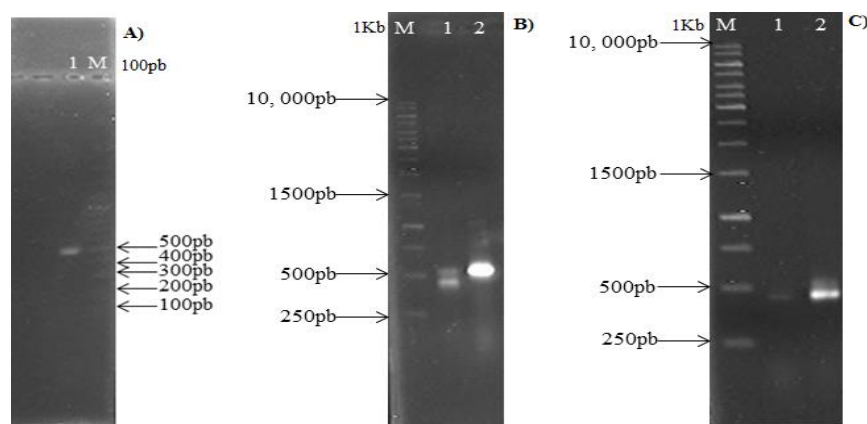


Figura 16. Geles de agarosa al 1%. Productos de amplificado de PCR. A) Amplificado de *bfpI*: se observó una banda del amplificado a una T_m de 60°C, marcador de peso molecular de 100pb. B) Amplificado de *bfpJ*: pozo 1 amplificado inespecífico de *bfpK*, pozo 2 amplificado de *bfpJ* donde se observó una sola banda a una T_m de 60°C. Marcador de peso molecular de 1Kb. C) Amplificado de *bfpK* se observó una sola banda del amplificado a una T_m de 70°C. Marcador de peso de 1Kb para B y C.

7.3 Conservación del vector comercial pET-3a en las cepas HB101 y MC4100

Con el objetivo de mantener en una cepa de *E. coli* el vector pET-3a adquirido comercialmente, se procedió a transformar las cepas HB101 y MC4100 por electroporación. Las clonas que crecieron en agar LB con ampicilina se sometieron a extracción plasmídica por método Quickprep y se escogieron 12 clonas que mostraron un patrón de bandeo igual al del plásmido solo. (Figura 17).

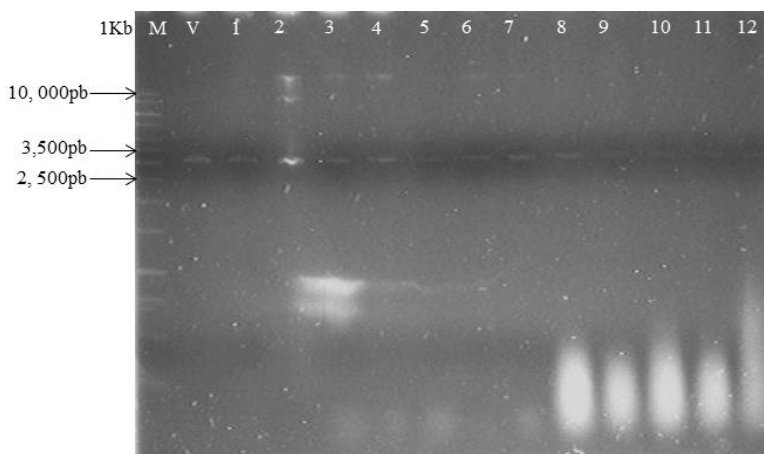


Figura 17. Gel de agarosa al 1% transformación de cepa HB101 y MC4100 con el vector pET-3a. Pozo 1, 5, 6, 7, 10, 11, 12 muestras de clonas transformadas HB101. Pozo 2, 3, 4, 8 y 9 muestras de clonas transformadas MC4100. Se observa un perfil de bandeo similar al del vector solo pozo V, entre 3000 y 3500 pb. Marcador de peso de 1Kb.

7.4 Doble restricción del plásmido pET-3a

Con el objetivo de confirmar que las clonas HB101 y MC4100 se transformaron con el plásmido pET-3a se realizó doble restricción del plásmido solo pET-3a para usarse como control y del ADN plasmídico de las clonas transformadas para comparar con el peso del vector linearizado (control). Las enzimas *NdeI* y *BamHI* se utilizaron conforme a las especificaciones del fabricante (Materiales y métodos). Se hizo extracción plasmídica a dos clonas de la cepa HB101 (Figura 18 pozo 1 y 5) y dos clonas de la cepa MC4100 (Figura 18 pozo 3 y 7) y posteriormente doble restricción para comprobar que los pesos

coincidieran con el vector solo sometido al mismo proceso de restricción (pozo VL) (Figura 18). Se comprobó mediante un gel de agarosa, que las clonas de la cepa HB101 y MC 4100 integraron el vector pET-3a de interés, puesto que los pesos del vector linearizado (control) coinciden con los pesos del ADN plasmídico obtenido de las clonas transformadas. Esto confirmó la obtención del vector pET-3a y se continuó con el proceso de clonación.

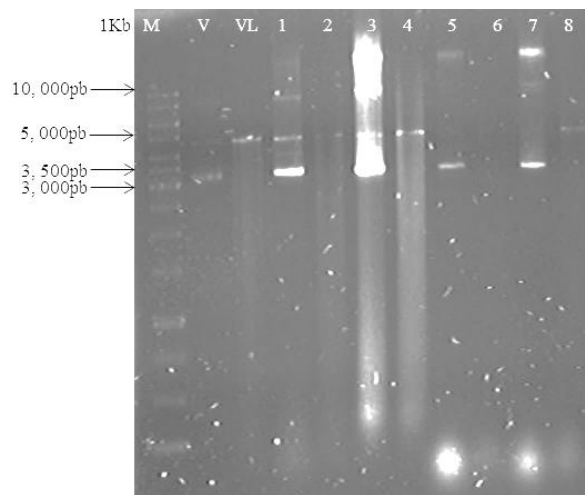


Figura 18. Gel de agarosa al 1%. Doble restricción del vector vacío (VL) y del ADN plasmídico de clonas transformadas con el vector pET-3a. Pozo 2, 4, 6 y 8 similar al del vector vacío linearizado (pozo VL) aproximadamente de 5,000 pb. Pozo V vector solo sin digestión, pozo 1, 3, 5 y 7 ADN plasmídico sin digestión con una peso de 3,500 pb y 3 000 pb, Marcador de peso de 1Kb.

7.5 Clonación de BfpI-pET3a y BfpJ-PET3a

Con el objeto de obtener las proteínas sin el extremo hidrofóbico se inició el proceso de clonación, el cual que consistió por un lado, en someter los productos de PCR de *bfpI* y *bfpJ* y al vector pET-3a a reacciones de doble restricción con las enzimas *BamHI* y *NdeI* (Materiales y métodos). Posteriormente los productos digeridos se sometieron a una reacción de ligación y el producto de ésta se utilizó para transformar en la cepa HB101 por electroporación (Materiales y métodos).

7.6 Transformación de la cepa HB101 con las construcciones BfpI-pET3a y BfpJ-pET3a

Las clonas que crecieron en agar LB adicionado con ampicilina se utilizaron para extraerles ADN plasmídico por método Quickprep (Materiales y métodos). El ADN plasmídico obtenido se corrió en un gel de agarosa al 1% (Figura 19). Las muestras 4 y 6 presentaron un peso ligeramente mayor al perfil de bandeo del vector pET-3a (Pozo V) por lo que probablemente integraron la construcción obtenida de la ligación BfpI-pET3a.

Para el caso de las clonas que integraron la construcción BfpJ-pET3a, las muestras 10 y 11 se eligieron por tener un peso ligeramente mayor al bandeo del vector pET-3a vacío. Se visualizó el peso de las construcciones ligeramente mayor en comparación con el peso del vector vacío, es probable que se deba a la integración del gen específico en el vector; sin embargo, esto no siempre es así, por lo tanto se confirmó que las construcciones sean correctas mediante doble restricción, PCR y secuenciación.

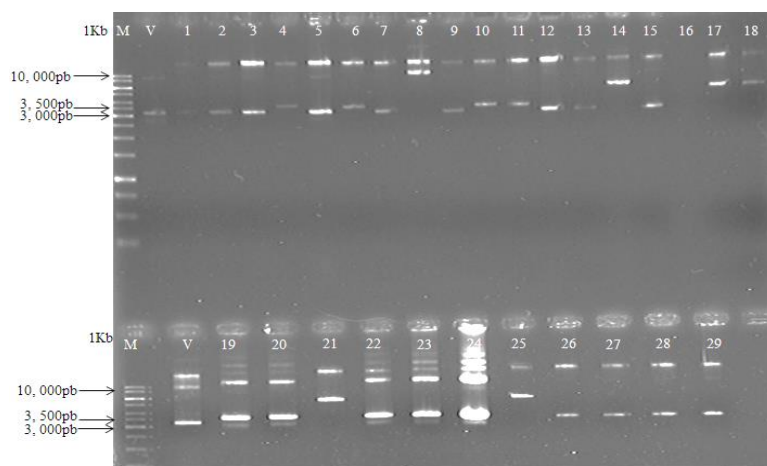


Figura 19. Perfil de bandeo de clonas transformadas con construcciones BfpI-pET3a y BfpJ-pET3a. El perfil de bandeo del ADN de las muestras 1 a la 8 son clonas de BfpI-pET3a, donde las clonas 4 y 6 muestran un perfil de bandeo ligeramente mayor al perfil de bandeo del vector pET-3a solo (pozo V). El perfil de bandeo de las clonas 9 a la 29 son construcciones BfpJ-pET3a, donde las clonas 10, 11, 19, 20, 21, 22, 23 y 24 mostraron un perfil de bandeo ligeramente mayor al del vector pET-3a sin inserto. Marcador de peso de 1Kb

Para obtener la construcción BfpK-pET3a se siguió el mismo procedimiento que se realizó para obtener las construcciones BfpI-pET3a y BfpJ-pET3a arriba descrito (Materiales y métodos). Una vez obtenidas las construcciones, se transformó en la cepa BL21 (DE3). Posteriormente se eligieron las clonas que crecieron en agar LB adicionado con ampicilina para hacerles extracción plasmídica por método Quickprep (Materiales y métodos) (Figura 20).

Las muestras 5, 7, 10, 11, 14, 17 y 18 (Figura 20) presentaron un peso ligeramente mayor al perfil de bandeo del vector pET-3a solo (Pozo V) por lo que era probable que integraran la construcción obtenida de la ligación BfpK-pET3a, por el contrario las clonas restantes tenían un peso similar al del vector pET-3a solo, por lo que se descartaron ya que es probable que no integraran la construcción. Esto fue un primer cribado para elegir las clonas que posiblemente integraron las clonaciones.

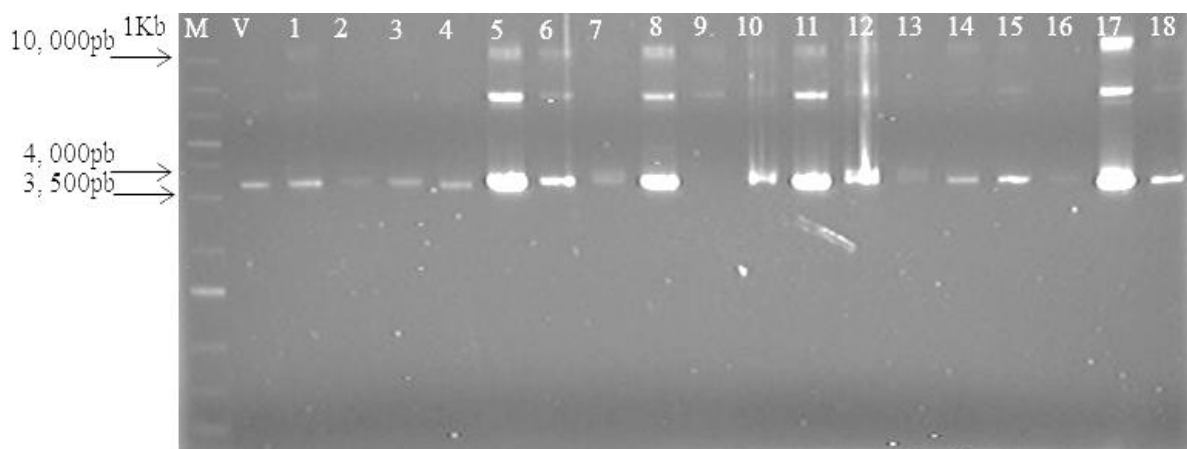


Figura 20. Gel de agarosa al 1% Extracción plasmídica de las clonas transformadas con la construcción BfpK-pET3a. Se eligieron las clonas 5, 7, 10, 11, 14, 17 y 18 que mostraron un perfil de bandeo más pesado con respecto al perfil del vector comercial (pozo V). Marcador de peso de 1Kb.

7.7 Verificación de las construcciones BfpI-pET3a y BfpJ-pET3a en la cepa HB101 por PCR y doble restricción

Una vez que se hizo el primer cribado para elegir las clonas que posiblemente integraron las construcciones BfpI-pET3a y BfpJ-pET3a, se verificó que estas fueran correctas por PCR y doble restricción. El ADN plasmídico de estas clonas se sometió a una reacción de doble digestión con las enzimas *NdeI* y *BamHI* y a continuación PCR. Para esto se crecieron las clonas 4 y 6 (construcción BfpI-pET3a) 10 y 11 (construcción BfpJ-pET3a) en caldo LB adicionado con ampicilina para posteriormente extraer el ADN plasmídico por método Quickprep (Materiales y métodos). A continuación se realizó doble digestión y el producto de esta doble restricción se corrió en un gel de agarosa al 1% (Figura 21) en este se observó la linearización del vector pET-3a y la liberación del inserto (gen *bfpI* de la clona 6, así como el gen *bfpJ* de las clonas 10 y 11). La clona 4 (construcción BfpI-pET3a) no mostró liberación del inserto solo el plásmido linearizado.

Por otra parte, se realizó PCR al ADN plasmídico de las clonas 4 y 6, (BfpI-pET3a) 10 y 11 (BfpJ-pET3a) como otra forma de verificar la integración del gen de interés en el vector. El producto de PCR de las clonas 4 y 6 (BfpI-pET3a) se corrió en un gel de agarosa al 1%, en este se observó amplificado del gen *bfpI* en la muestra de la clona 6 (Figura 21), con un peso aproximado de 500 pb, a diferencia de la clona 4 que no mostró amplificado. Para el caso de la construcción BfpJ-pET3a, las clonas 10 y 11 amplificaron del gen *bfpJ* con un peso aproximado de 500 pb. Por el contrario las clonas 19, 20 y 21 (construcción *bfpJ*-pET3a) que a pesar de mostrar amplificado, su perfil de bandeo en la doble restricción fue muy diferente comparándolo con el vector vacío digerido (Figura 21), por lo tanto se descartaron estas últimas clonas para someter el ADN plasmídico a secuenciación y solo las

clonas 6 (BfpI-pET3a), 10 y 11(BfpJ-pET3a) se mandaron a secuenciar como última etapa para verificar dichas construcciones.

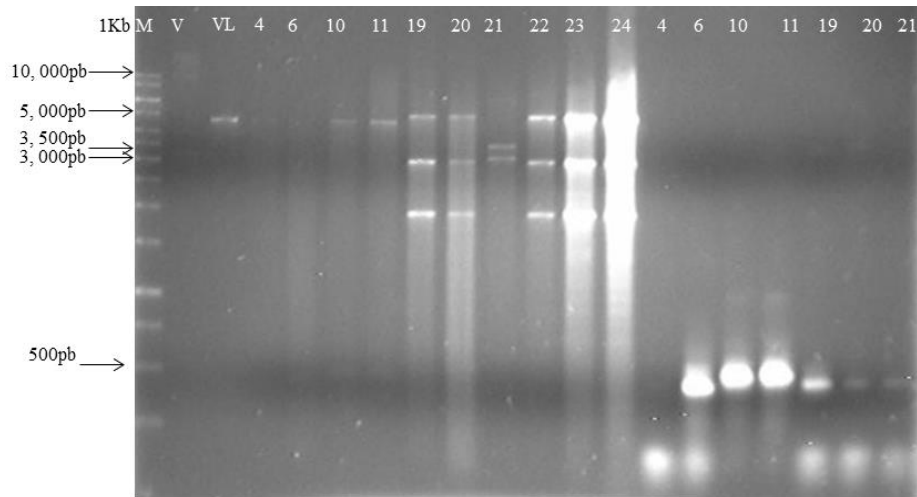


Figura 21. Corrimiento electroforético con muestras de la reacción de doble restricción y PCR. Pozo 6 muestra de ADN digerido de la clona BfpI-pET3a con perfil de bandeo similar al perfil del vector vacío linealizado (VL) y amplificado del producto de PCR (derecha pozo 6), pozo 10 y 11 ADN plasmídico digerido clonas BfpJ-pET3a con perfil de bandeo similar al del vector vacío linealizado (VL) y amplificado del producto de PCR (derecha pozo 10 y 11). Marcador de peso de 1Kb

Por otra parte se transformó nuevamente por electroporación en la cepa HB101, la ligación de la construcción BfpI-pET3a para asegurar más clonas con esta construcción. Las clonas que crecieron en agar LB con ampicilina se sometieron a extracción plasmídica por método Quickprep (Materiales y métodos).

El ADN plasmídico se corrió en un gel de agarosa al 1% y se visualizó en éste el perfil de bandeo de las clonas con los números 4, 5 y 6 como posibles clonas con la construcción BfpI-pET3a (Figura 22 A). El ADN plasmídico de estas clonas se utilizó para la reacción de PCR y se observó un sólo amplificado (clona numerada como 4 y 6 construcción BfpI-pET3a) con el peso aproximado de 500 pb, a diferencia de la clona numerada como 5, que mostró dos amplificados, uno aproximadamente de 500 pb y otro de

mayor tamaño (Figura 22 B), por lo que se descartó la clona 5 (BfpI-pET3a). Para verificar las construcciones por doble restricción sólo se trabajó con las clonas 4 y 6 (BfpI-pET3a), estas se renombraron como d y f respectivamente para no confundir las primeras clonas (mencionadas arriba de la construcción BfpI-pET3a) con estas últimas también de la construcción BfpI-pET3a.

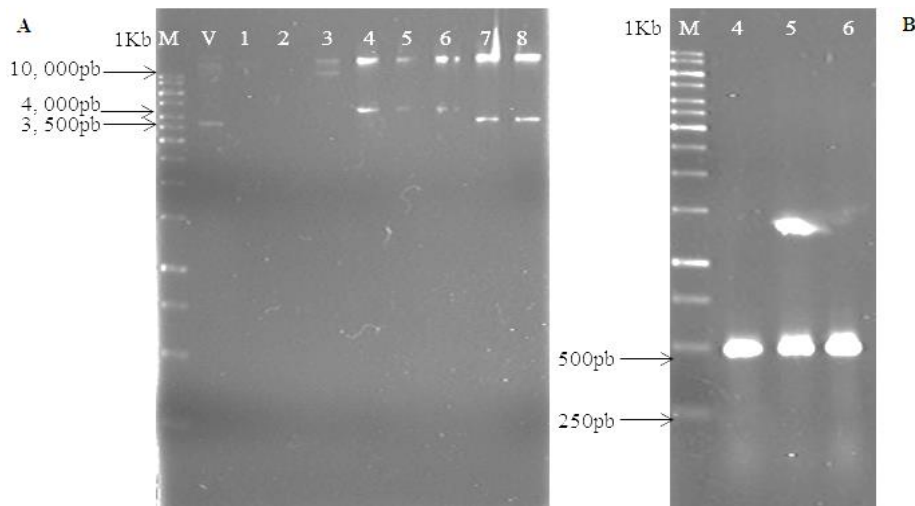


Figura 22. Extracción plasmídica de las clonas transformadas con la construcción BfpI-pET3a. A) pozo 4, 5, 6 perfil de bandeo de las clonas ligeramente mayor al del vector vacío (pozo V). B) gel de agarosa al 1% con productos de PCR donde el pozo 4 y 6 amplificaron *bfpI* de peso aproximado a 500 pb

Una vez que se confirmó por PCR la construcción BfpI-pET3a (clonas d y f) se verificó por doble restricción la construcción BfpI-pET3a (clonas d y f) con las enzimas *NdeI* y *BamHI*. El producto de restricción se corrió en un gel de agarosa al 1% y en este se observó que la clona d fue la única que mostró la liberación del inserto (*bfpI*) y la linearización del vector (Figura 23). La muestra f mostró linearización del vector, pero no se observó liberación del inserto (*bfpI*).

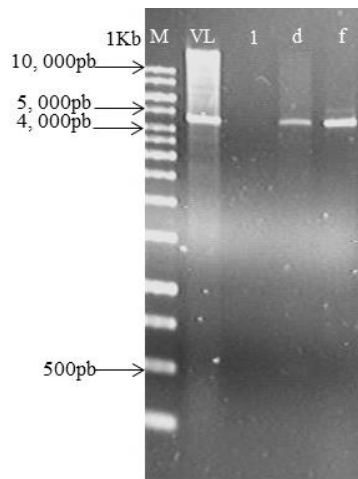


Figura 23. Doble digestión de la construcción BfpI-pET3a clonas d y f. Doble digestión de las muestras d y f (construcción BfpI-pET3a), en la muestra d se observa la linearización del vector con un peso similar al del pozo VL (vector comercial linearizado) y la liberación del inserto *bfpI*. Marcador de peso 1Kb.

Por otro lado la construcción de BfpK-pET3a de las clonas 5, 7, 10, 11, 14, 17 y 18 se verificaron por PCR. Todas las muestras amplificaron un producto de peso aproximado a 500 pb (Figura 24), por lo que la construcción de BfpK-pET3a de las clonas 5, 7, 10, 11, 14, 17 y 18 se verificó por doble digestión (enzimas *NdeI* y *BamHI*). En este ensayo se observó un perfil de bandeo que no corresponde al esperado en ninguna clona (linearización del vector pET-3a y liberación del inserto *bfpK*) (Figura 25).

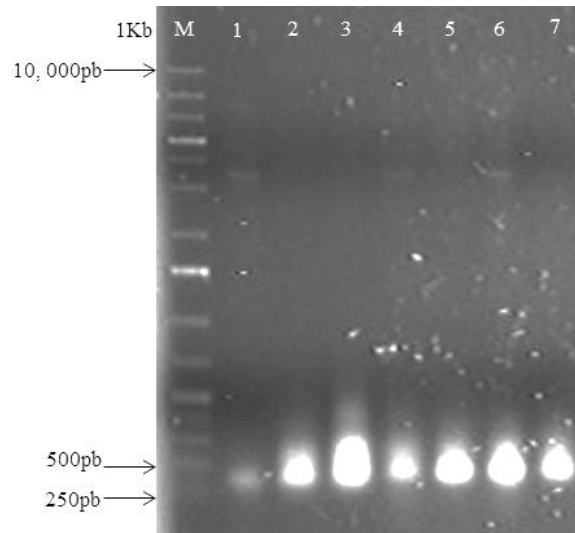


Figura 24. Gel de agarosa al 1%. Producto de PCR del ADN plasmídico de las clonas transformadas con la construcción BfpK-pET3a. Pozo 1, 2, 3, 4, 5, 6 y 7 amplificados con peso aproximado a 500pb. Marcador de peso 1Kb.

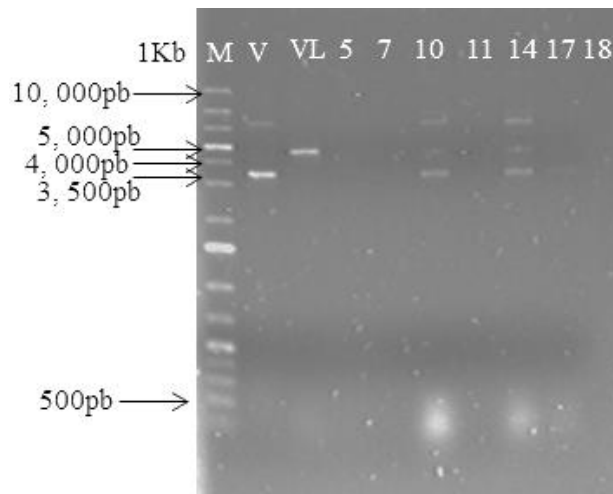


Figura 25. Doble restricción de las clonas con las construcciones BfpK-pET3a. No se visualizó alguna muestra con el perfil de bandeo similar al del vector linealizado (VL), ni la liberación del inserto en el peso aproximado a 500 pb.

Una vez que se verificó cada construcción BfpI-pET3a, BfpJ-pET3a y BfpK-pET3a por PCR y doble restricción se verificaron las construcciones por secuenciación. En este paso se utilizaron los primers del vector pET-3a los cuales amplifican río arriba y río abajo del sitio de integración del inserto (Materiales y métodos) por lo que antes de mandar a secuenciar se usaron en las construcciones para observar el amplificado correspondiente al inserto y a las secuencias río arriba y río abajo del vector pET-3a.

Las construcciones BfpI-pET3a (clona 6, d y f) y BfpJ-pET3a (clona 10 y 11) se sometieron a PCR con los primers diseñados para el vector pET-3a. Se comparó el amplificado del vector vacío (peso del amplificado aproximado de 196pb, Figura 25 A) y el amplificado obtenido de las construcciones BfpI-pET3a (peso aproximado de 649pb) y BfpJ-pET3a (peso aproximado de 697pb Figura 25A).

Como se observó un amplificado inespecífico (Figura 25A) se repitió la PCR con los primers diseñados para el vector pET-3a de las muestras d, f (BfpI-pET3a) y las muestras 10 y 11 (BfpJ-pET3a) en esta repetición desapareció la banda inespecífica (Figura 25B).

Para la construcción BfpK-pET3a las clonas 5, 7, 10, 11, 14, 17 y 18 se sometieron a PCR con los primers diseñados para el vector pET-3a. Se observó mediante corrimiento electroforético el amplificado inespecífico en todas las muestras (el peso esperado del amplificado es de 613 pb aproximadamente, Figura 26).

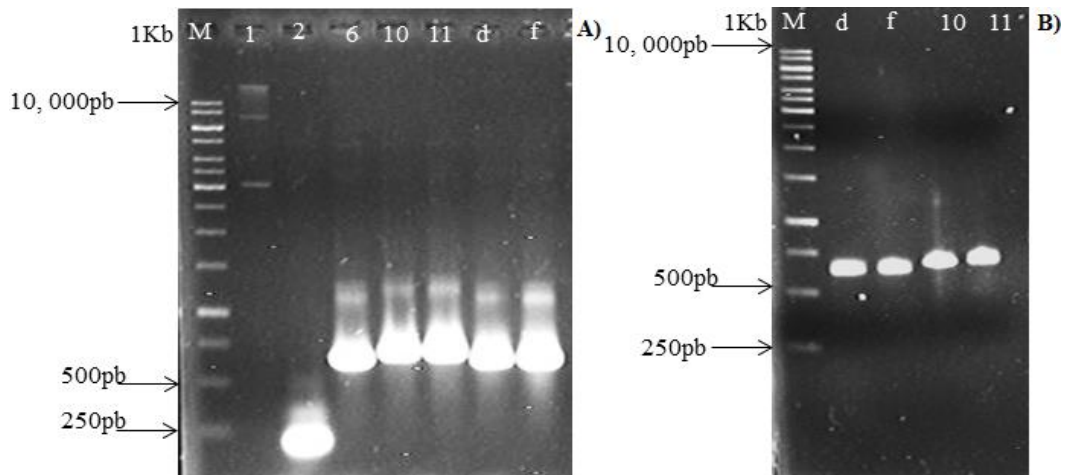


Figura 26. Gel de agarosa al 1%. Amplificado río arriba y río abajo del sitio de integración del inserto de las construcciones BfpI-pET3a y BfpJ-pET3a.

- A) Amplificado de PCR del vector vacío (pozo 2) y amplificado río arriba y río abajo del sitio de integración del inserto de las construcciones BfpI-pET3a (pozo 6, d y f) y BfpJ-pET3a (pozo 10 y 11). B) repetición de la reacción de PCR. Amplificado río arriba y río abajo del sitio de integración del inserto de las construcciones BfpI-pET3a (pozo d y f) y BfpJ-pET3a (pozo 10 y 11).

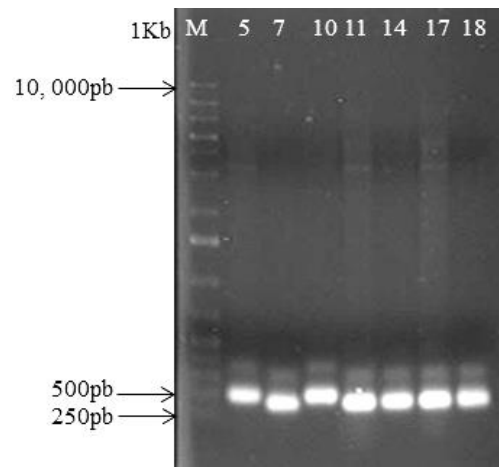


Figura 27. Gel de agarosa al 1%. Amplificado de la región río arriba y río abajo del sitio de integración del inserto de la construcción BfpK-pET3a. Todas las muestras amplificaron el gen *bfpK*.

7.8 Secuenciación de las clonas de BfpI-pET3a y BfpJ-pET3a en la cepa HB101.

Como se mencionó anteriormente, el último proceso por el cual se verificaron las construcciones fue la secuenciación. Una vez que se verifica que las construcciones son correctas (marco de lectura idéntico, integración del inserto en el sentido correcto, etc.) se prosigue a transformar en la cepa BL21 (DE3) para continuar con la expresión de la proteína de interés. Por lo tanto la secuenciación es muy importante para confirmar o descartar las construcciones.

Se mandó a secuenciar el ADN plasmídico BfpI-pET3a (clonas d, f) y BfpJ-pET3a (clonas 10 y 11). Se realizó el alineamiento entre la secuencia teórica del gen *bfpI* y la secuencia proveniente de la secuenciación, con el uso del programa Clustal Omega. El alineamiento de la secuencia de BfpI-pET3a (clona d), fue idéntica en la región del gen, la etiqueta de histidina y las regiones río arriba y río abajo del sitio de integración del inserto en el vector pET-3a se conservaron. (Figura 28). Por otra parte, se alineó la secuencia de BfpI-pET3a (clona f) con la secuencia teórica y se observó la ausencia de 51 pb de la región 510 pb río arriba de la secuencia del vector (Figura 29). Sin embargo la secuencia que codifica para la etiqueta de histidina se mantuvo intacta pero en la secuencia correspondiente a la secuencia del gen *bfpI* se observó un cambio de base de Citosina por Adenina (punto al inicio de la imagen), por tal motivo se descartó el uso de la clona f (BfpI-pET3a).

La secuenciación del ADN plasmídico de BfpJ-pET3a (clona 10) se alineó con la secuencia teórica de BfpJ (Figura 30). Se observa la secuencia que codifica para la etiqueta de histidina, la región río arriba y río abajo del sitio de integración del inserto de BfpJ, así como la parte intermedia que corresponde al gen de *bfpJ*; a pesar de mantenerse

estas regiones, se observan sustituciones o inserciones de bases (que se visualizan como espacios en blanco o guiones intermedios) tanto en las regiones que corresponden al vector pET3a así como en el inserto en la región que corresponde al extremo carboxilo (Figura 30). Por tal motivo se descartó la clona 10 (construcción de BfpJ-pET3a) ya que no es idéntica a la secuencia teórica.

La secuencia de BfpJ-pET3a (clona 11) se alineó con la secuencia teórica del gen *bfpJ* (Figura 31). En este se observaron las secuencias del vector pET-3a y la secuencia que codifica para la etiqueta de histidina. En la parte superior se encuentra la secuencia teórica de *bfpJ*, en la parte inferior se encuentra la secuencia proveniente de la secuenciación del ADN plasmídico de BfpJ-pET3a (clona 11) (Figura 31). Se observaron 19 cambios de bases en la región del gen de *bfpJ* correspondiente al extremo carboxilo de BfpJ, cabe mencionar que este alineamiento se hizo con la secuenciación utilizando el primer forward.

Se repitió la secuenciación pero usando el primer reverso y al realizar el alineamiento con la secuencia teórica de *bfpJ* se visualizó la secuencia que codifica para la etiqueta de histidina (recuadro rojo) y las regiones pertenecientes a las secuencias del vector de clonación pET-3a (recuadro azul) (Figura 32). En la región del gen *bfpJ* se observó un cambio de base de Adenina por Guanina representada con un punto (recuadro morado) (Figura 32). Pero esta sustitución no cambia el aminoácido (glicina (GGU)) o marco de lectura, según el código genético la sustitución por cualquier otra base siempre codifica para glicina (GGU, GGG, GGC y GGA) por lo tanto, el ADN plasmídico de la clona 11 se utilizó como la clona con la construcción correcta para BfpJ- pET3a

CLUSTAL O(1.2.3) multiple sequence alignment

```
SteoricaBfpI   TTTCTGGGCTTTGTTAGCAGCCGGATCC TTTCTTTTCTTATCATAAAACATCCACCCGGAA
SecdI         TTTCTGGGCTTTGTTAGCAGCCGGATCC TTTCTTTTCTTATCATAAAACATCCACCCGGAA
*****

SteoricaBfpI   ATTATTGTA AAAAGAAAAGATTTTTATTATTATCCCCACCTAAAAAAATTTGTGTCATCAGA
SecdI         ATTATTGTA AAAAGAAAAGATTTTTATTATTATCCCCACCTAAAAAAATTTGTGTCATCAGA
*****

SteoricaBfpI   ATTTACTTTCCACCATCCCAGAACCGAACTTACAACCTCTGACTGCTATCCATTCTCCC
SecdI         ATTTACTTTCCACCATCCCAGAACCGAACTTACAACCTCTGACTGCTATCCATTCTCCC
*****

SteoricaBfpI   AGAATAAGCGCTTACATGCGTAAATAACATGCTACCGAAATCTTCATTCTGCCAAAATA
SecdI         AGAATAAGCGCTTACATGCGTAAATAACATGCTACCGAAATCTTCATTCTGCCAAAATA
*****

SteoricaBfpI   TTCGTTCCCTGACAGGGTGTATAAAATATACACAGATACATATTTCCGCTTTTTTAAAC
SecdI         TTCGTTCCCTGACAGGGTGTATAAAATATACACAGATACATATTTCCGCTTTTTTAAAC
*****

SteoricaBfpI   ATCAACCTTTATTTATTACCAGAGTTAACGTTTGATTTACTGAACTTATCACTCCATC
SecdI         ATCAACCTTTATTTATTACCAGAGTTAACGTTTGATTTACTGAACTTATCACTCCATC
*****

SteoricaBfpI   ACAAACAATATTTTCTTTCCATAGCATGATGCTGTAAATGTATCCTGAATGTAAGACAT
SecdI         ACAAACAATATTTTCTTTCCATAGCATGATGCTGTAAATGTATCCTGAATGTAAGACAT
*****

SteoricaBfpI   TACATACGAAACATCATGCAGTGCATTTTCATGAACAGCATACTTAATATGGTCA TGGTG
SecdI         TACATACGAAACATCATGCAGTGCATTTTCATGAACAGCATACTTAATATGGTCA TGGTG
*****

SteoricaBfpI   ATGGTGATATGTATATCTCCTTCTTAAAGTAAACAAAATTATTTCTAGAGGGAAACCG
SecdI         ATGGTGATATGTATATCTCCTTCTTAAAGTAAACAAAATTATTTCTAGAGGGAAACCG
*****

SteoricaBfpI   TTGTGGT-
SecdI         TTGTGGT-
*****
```

Figura 28. Alineamiento entre la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona d (BfpI-pET3a) Se observan en recuadros azules la región perteneciente a la secuencia del vector pET-3a, en rojo se observa la secuencia que codifica la etiqueta de histidina y la región intermedia es la secuencia de BfpI. Los asteriscos de la imagen representan identidad entre los alineamientos, el espacio vacío representa una sustitución o cambio de base, el guion medio representa pérdida de la base. Se observó que la construcción sigue el marco de lectura y tiene las regiones de interés para poder purificar una vez expresada la proteína

CLUSTAL O(1.2.3) multiple sequence alignment

```
SteoricaI      TTTCTGGGCTTTGTTAGCAGCCGGATCCCTTCTTTTTCTTATCATAAAAACATCCACCCGGAA
SfI            -----CNCACCAGGAA
                ****_****

SteoricaI      ATTATTGTAAAAGAAAAGATTTTTATTATTATATCCCCACCTAAAAAAATTGTGTCATCAGA
SfI            ATTATTGTAAAAGAAAAGATTTTTATTATTATATCCCCACCTAAAAAAATTGTGTCATCAGA
                *****

SteoricaI      ATTTACTTTCCACCATCCCAGAACCGAACCTACAACCTCCTGTACTGCTATCCATTCTCCC
SfI            ATTTACTTTCCACCATCCCAGAACCGAACCTACAACCTCCTGTACTGCTATCCATTCTCCC
                *****

SteoricaI      AGAATAAGCGCTTACATGCGTAAATAACATGCTACCGAAATCTTCATTTCTGCCAAAATA
SfI            AGAATAAGCGCTTACATGCGTAAATAACATGCTACCGAAATCTTCATTTCTGCCAAAATA
                *****

SteoricaI      TTCGTTCCCTGACAGGGTGTATAAAATATACACAGATACATATTTTCCGTCTTTTTTAAC
SfI            TTCGTTCCCTGACAGGGTGTATAAAATATACACAGATACATATTTTCCGTCTTTTTTAAC
                *****

SteoricaI      ATCAACCTTTATTTTATTACCAGAGTTAACGTTTGATTTTACTGAACTTATCACTCCATC
SfI            ATCAACCTTTATTTTATTACCAGAGTTAACGTTTGATTTTACTGAACTTATCACTCCATC
                *****

SteoricaI      ACAACAATATTTTCTTTTCCATAGCATGATGTCTGTAATGTATCCTGAATGTAAGACAT
SfI            ACAACAATATTTTCTTTTCCATAGCATGATGTCTGTAATGTATCCTGAATGTAAGACAT
                *****

SteoricaI      TACATACGAAACATCATGCAGTGCATTTTCATGAACAGCATACTTAA ATGGTGATGGTG
SfI            TACATACGAAACATCATGCAGTGCATTTTCATGAACAGCATACTTAA ATGGTGATGGTG
                *****

SteoricaI      ATGGTG CATATGTATATCTCCTTCTTAAAGTTAAACAAAAT TATTTCTAGAGGGGAAACCG
SfI            ATGGTG CATATGTATATCTCCTTCTTAAAGTTAAACAAAAT TATTTCTAGAGGGGAAACCG
                *****

SteoricaI      TTGTGGT-----
SfI            TTGTGGTCCCAAAA
                *****
```

Figura 29. Alineamiento entre la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona f (BfpI-pET3a) Se observan en recuadro rojo la secuencia codificante para la etiqueta de histidina, en recuadro azul las porciones del vector pET-3a y la secuencia intermedia del inserto *bfpI* con un cambio de base representado como punto. En la región del gen *bfpI*, se observa pérdida de bases (guion medio) en el extremo N-terminal. En la región que corresponde al vector se observa pérdida de bases (guion medio). Esta clona se descartó

```

Steorica3      TTTTTTGGGGAGACCACAACGGTTTCCTCTAGAAAATAATTTTGTTTAACTTTAAGAAGG
S103          TTTTTTGGGGACACCAACGGTTTCCTCTAGAAAATAATTTTGTTTAACTTTAAGAAGG
*****

Steorica3      AGATATACATATGCACCATCACCATCACCAATATAACTATTACAACAGGAGCGCAGAACA
S103          AGATATACATATGCACCATCACCATCACCAATATAACTATTACAACAGGAGCGCAGAACA
*****

Steorica3      ACATTATATGAATAATGTCTCTGTCTGTTATGTGCAGATATGCTGATGCTCTTAGCTACTA
S103          ACATTATATGAATAATGTCTCTGTCTGTTATGTGCAGATATGCTGATGCTCTTAGCTACTA
*****

Steorica3      TTTACAGGATAGCACAGCAACCAATGGTAACGGTAGTGTATCATTATCTACTCTGGTTGC
S103          TTTACAGGATAGCACAGCAACCAATGGTAACGGTAGTGTATCATTATCTACTCTGGTTGC
*****

Steorica3      TTCTGGATATTTGCCACCAATGCAAGAAACGTCTGGTGTGGAGAATATGAAATAAAAAAT
S103          TTCTGGATATTTGCCACCAATGCAAGAAACGTCTGGTGTGGAGAATATGAAATAAAAAAT
*****

Steorica3      GTATGTCGATAAAGAATAAGAATGGTCTGATAGCATTATATGGTGGAGAGTATTATTCACT
S103          GTATGTCGATAAAGAATAAGAATGGTCTGATAGCATTATATGGTGGAGAGTATTATTCACT
*****

Steorica3      TACCTCTGAGAACATAAAAAAATTTCTGGGGA-TTCGTGGTGCATCTGCTAAAGATAATA
S103          TACCTCTGAGAACATAAAAAAATTTCTGGGGA-TTCGTGGTGCATCTGCTAAAGATAATA
*****

Steorica3      AAATCATATCTGCA--GGTGGGTATTA--TTCATTGCCTGTATCTGACTTTACAGGTTTT
S103          AAATCATATCTGCA--GGTGGGTATTA--TTCATTGCCTGTATCTGACTTTACAGGTTTT
*****

Steorica3      CGCACTGTTAATTTAAGAGGATTAATGCTCATTCT--CT-TCCATTTCAGACATCAGGGCAA
S103          CGCACTGTTAATTTAAGAGGATTAATGCTCATTCT--CT-TCCATTTCAGACATCAGGGCAA
*****

Steorica3      AATTGTGACAGAGGTGGATATGATTTTTGAAGSACGTGAGCAAAAGGSACAAGACCCTT
S103          AATTGTGACAGAGGTGGATATGATTTTTGAAGSACGTGAGCAAAAGGSACAAGACCCTT
*****

Steorica3      TAGAGGCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCAACTCAGCTT
S103          TAGAGGCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGCAACTCAGCTT
*****

Steorica3      CCTTTCGGGCTTTTGTAGCAGCCG
S103          CCTTTCGGGCTTTTGTAGCAGCCG
*****

```

Figura 30. Alineamiento de la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona 10. Se observa en recuadros azules las secuencias pertenecientes al vector pET-3a, en recuadro rojo la secuencia que codifica para la etiqueta de histidinas. La región intermedia es la secuencia de *bfpJ*. Se observaron sustituciones y deleciones tanto en la secuencia del gen de *bfpJ* como en las regiones del vector pET-3a (espacios en blanco y guion medio). Esta clona se descartó


```

CLUSTAL O(1.2.3) multiple sequence alignment

SteoricaBfpJ S113  ---TCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATC
TGGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAGCAGCCGGATG
*****

SteoricaBfpJ S113  CCCCTTTTGTGACGTCCTTCAAAAATCATATCCACCTCTGTACAAATTTTGCCCTGATG
CCCTTTTGTGACGTCCTTCAAAAATCATATGCCACCTCTGTACAAATTTTGCCCTGATG
***  *

SteoricaBfpJ S113  TCTGAATGGAAGGAATGAGCATTAAATCCTCTTAAATTAACAGTGCGAAAAACCTGTAAAGT
TCTGAATGGAAGGAATGAGCATTAAATCCTCTTAAATTAACAGTGCGAAAAACCTGTAAAGT
*****

SteoricaBfpJ S113  CAGATACAGGCAATGAATAATACCCACCTGCAGATATGATTTTATTATCTTTAGCAGATG
CAGATACAGGCAATGAATAATACCCACCTGCAGATATGATTTTATTATCTTTAGCAGATG
*****

SteoricaBfpJ S113  CACCACGAATCCCAGAAATTTTTTATGTTCTCAGAGGTAAGTGAATAATACTCTCCAC
CACCACGAATCCCAGAAATTTTTTATGTTCTCAGAGGTAAGTGAATAATACTCTCCAC
****

SteoricaBfpJ S113  CATATAATGCTATCAGACCATTCTTATCTTATCGACATACATTTTTATTTCATATTCTC
CATATAATGCTATCAGACCATTCTTATCTTATCGACATACATTTTTATTTCATATTCTC
*****

SteoricaBfpJ S113  CAACACCAGACGTTTCTTGCAATGGTGGCAAATATCCAGAAGCAACCAGAGTAGATAATG
CAACACCAGACGTTTCTTGCAATGGTGGCAAATATCCAGAAGCAACCAGAGTAGATAATG
*****

SteoricaBfpJ S113  ATACACTACC GTTACCATTGGTTGCTGTGCTATCCTGTAAATAGTAGCTAAGAGCATCAG
ATACACTACC GTTACCATTGGTTGCTGTGCTATCCTGTAAATAGTAGCTAAGAGCATCAG
*****

SteoricaBfpJ S113  CATATCTGCACATAACGACAGAGACATTATTCATATAATGTTGTTCTGCGCTCCTGTTGT
CATATCTGCACATAACGACAGAGACATTATTCATATAATGTTGTTCTGCGCTCCTGTTGT
*****

SteoricaBfpJ S113  AATAGTTATATGGTGATGGTGATGGTSCATATGTATATCTCCTTCTTAAAGTTAAACAA
AATAGTTATATGGTGATGGTGATGGTSCATATGTATATCTCCTTCTTAAAGTTAAACAA
*****

SteoricaBfpJ S113  AATTATTTCTAGAGGGAAACCGTTGTGGTCTCCC--
AATTATTTCTAGAGGGAAACCGTTGTGGTCTCCC--
*****

```

Figura 31. Alineamiento de la secuencia teórica y la secuenciación con el primer forward del ADN plasmídico de la clona 11. Los asteriscos de la imagen representan identidad entre los alineamientos, el espacio vacío representa una sustitución o cambio de base, el guion medio representa pérdida de la base. Se observó que la construcción tiene sustituciones de bases en la región N-terminal de BfpJ.


```

CLUSTAL O(1.2.4) multiple sequence alignment

secteoricaBfpJ11      -----TGCAGCGGTGGCAGCAGC
SECMUESTRAFBFPJ11    TTTTAACCGTTTAGAGGCCCAAGGGGTTATGCTAGTTATTGCTCAGCGGTGGCAGCAGC
*****

secteoricaBfpJ11      CAACTCAGCTTCCTTTTCGGGCTTTGTTAGCAGCCGGATCCCTTTTGTGACGTCTTC
SECMUESTRAFBFPJ11    CAACTCAGCTTCCTTTTCGGGCTTTGTTAGCAGCCGGATCCCTTTTGTGACGTCTTC
*****

secteoricaBfpJ11      AAAAAATCATATCCACCTCTGTCACAATTTTGCCCTGATGTCGTAATGGAAGGAATGAGCA
SECMUESTRAFBFPJ11    AAAAAATCATATCCACCTCTGTCACAATTTTGCCCTGATGTCGTAATGGAAGGAATGAGCA
*****

secteoricaBfpJ11      TTAATCCTCTTAAATTAACAGTGCGAAAACCTGTAAAGTCAGATACAGGCAATGAATAAT
SECMUESTRAFBFPJ11    TTAATCCTCTTAAATTAACAGTGCGAAAACCTGTAAAGTCAGATACAGGCAATGAATAAT
*****

secteoricaBfpJ11      ACCCACCTGCAGATATGATTTTATTATCTTTAGCAGATGCCACGATCCCCAGAAATT
SECMUESTRAFBFPJ11    ACCCACCTGCAGATATGATTTTATTATCTTTAGCAGATGCCACGATCCCCAGAAATT
*****

secteoricaBfpJ11      TTTTATGTTCTCAGAGGTAAGTGAATAATACTCTCCACCATATAATGCTATCAGACCAT
SECMUESTRAFBFPJ11    TTTTATGTTCTCAGAGGTAAGTGAATAATACTCTCCACCATATAATGCTATCAGACCAT
*****

secteoricaBfpJ11      TCTTATTCTTATCGACATACATTTTTATTTCATATTCTCCAACACCAGACGTTTCTTGCA
SECMUESTRAFBFPJ11    TCTTATTCTTATCGACATACATTTTTATTTCATATTCTCCAACACCAGACGTTTCTTGCA
*****

secteoricaBfpJ11      TTGGTGCAAATATCCAGAAGCAACCAGAGTAGATAATGATACACTACCGTTACCATTGG
SECMUESTRAFBFPJ11    TTGGTGCAAATATCCAGAAGCAACCAGAGTAGATAATGATACACTACCGTTACCATTGG
*****

secteoricaBfpJ11      TTGCTGTGCTATCCTGTAAGTAGTAGCTAAGAGCATCAGCATATCTGCACATAACGACAG
SECMUESTRAFBFPJ11    TTGCTGTGCTATCCTGTAAGTAGTAGCTAAGAGCATCAGCATATCTGCACATAACGACAG
*****

secteoricaBfpJ11      AGACATTATTCATATAATGTTGTTCTGCGCTCCTGTTGTAATAGTTAATGGTGATGGT
SECMUESTRAFBFPJ11    AGACATTATTCATATAATGTTGTTCTGCGCTCCTGTTGTAATAGTTAATGGTGATGGT
*****

secteoricaBfpJ11      GATGGTGCATATGATATCTCCTTCTTAAAGTTAAACAAAATATTTCTAGAGGGAACC
SECMUESTRAFBFPJ11    GATGGTGCATATGATATCTCCTTCTTAAAGTTAAACAAAATATTTCTAGAGGGAACC
*****

secteoricaBfpJ11      GTTGTGGTCTCCC
SECMUESTRAFBFPJ11    -----

```

Figura 32. Alineamiento de la secuencia teórica y la secuencia proveniente de la secuenciación del ADN plasmídico de la clona 11 con el primer reverso. Se observa en cuadros azules las secuencias pertenecientes al vector pET-3a, en recuadro rojo la secuencia que codifica para la etiqueta de histidinas. La región intermedia es la secuencia de *bfpJ*. Comparando e alineamiento con la secuencia del primer forward de la clona 11, no se observaron deleciones la secuencia del gen de *bfpJ* solo una sustitución (recuadro morado) que no afecta el marco de lectura ni cambia el aminoácido.

Se verificaron por secuenciación las construcciones BfpK-pET3a (clonas 17 y 18) estas secuencias se metieron al programa SnapGene viewer, este programa las identifico como secuencias río abajo del sitio de integración del inserto pertenecientes al vector, en ningún caso se identificó la secuencia del inserto debido a que la secuenciación se realizó río abajo del sitio de integración del inserto (Figura 33 y 34). Por tal motivo no se pueden descartar dichas clonas, sino que se debe repetir la secuenciación para conocer la secuencia del inserto ya que en la verificación por PCR en ambas clonas se obtuvo amplificado de *bfpK*. No se utilizó ninguna construcción para continuar con la sobreexpresión de BfpK sin antes confirmar la secuencia de *bfpK*.

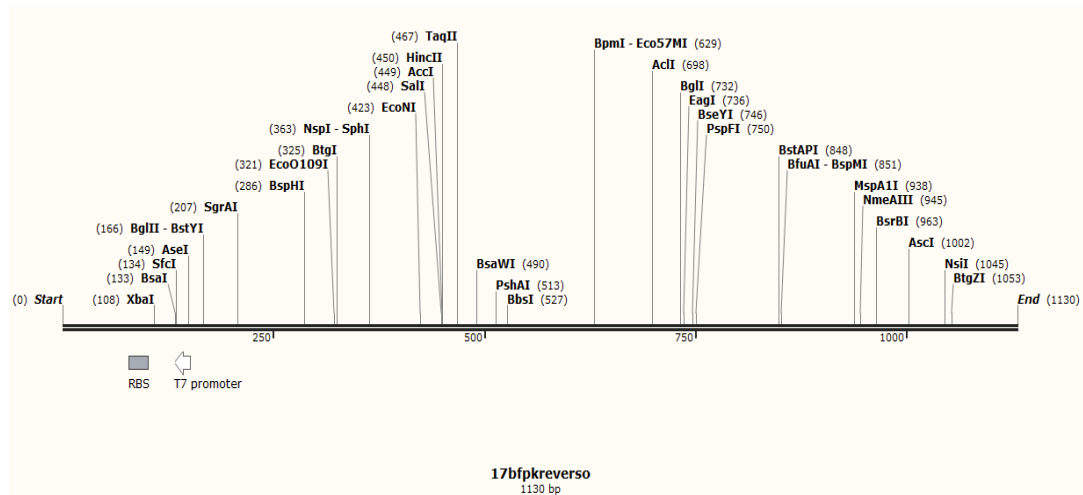


Figura 33. Secuencia de la construcción BfpK-pET3a (clona 17) Se observa la secuencia del vector (región río abajo del sitio de integración del inserto) pero no la secuencia del inserto de *bfpK*.

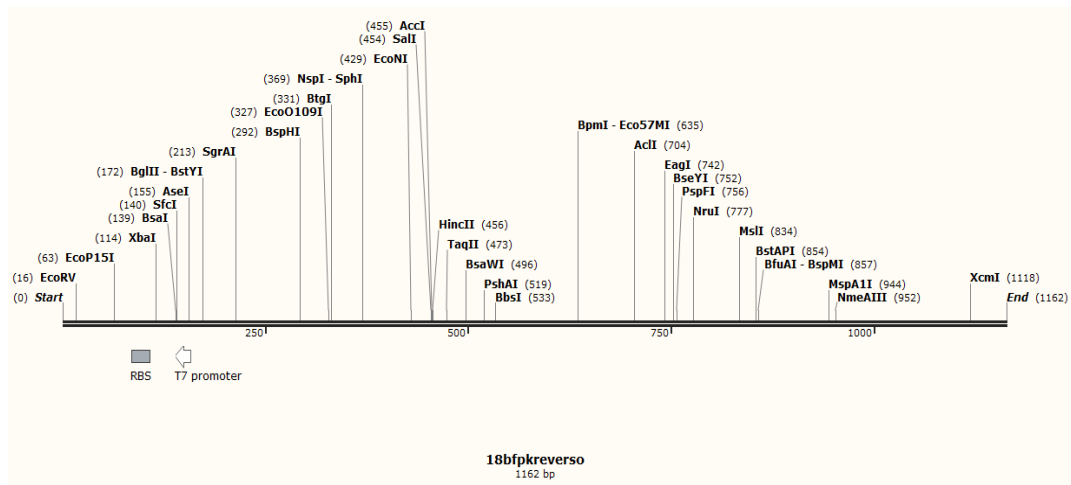


Figura 34. Secuencia de la construcción BfpK-pET3a (clona 18) No se observa la secuencia del inserto *bfpK*, solo la secuencia río abajo del sitio de integración del vector pET-3a.

7.9 Transformación en la cepa BL21 (DE3) con las construcciones BfpI-pET3a y BfpJ-pET3a.

Una vez que se verificaron las construcciones BfpI- pET3a (clona d) y BfpJ- pET3a (clona 11) por doble restricción, PCR y secuenciación se transformaron en la cepa BL21 (DE3) para poder expresar cada proteína de interés induciendo con IPTG. Esta cepa se utilizó pues tiene delectados los genes de las proteasas OmpT y Lon lo cual disminuye la posibilidad de degradación de las proteínas sobreexpresadas (BfpI y BfpJ).

Mediante el corrimiento electroforético se observó que la construcción BfpI- pET3a se mantuvo en la clona d1 y d2. Se eligió la clona d2 para expresar la proteína BfpI. Las

clonas d3, d4, d5 y d6 se descartaron ya que no fue visible el plásmido transformado (Figura 35).

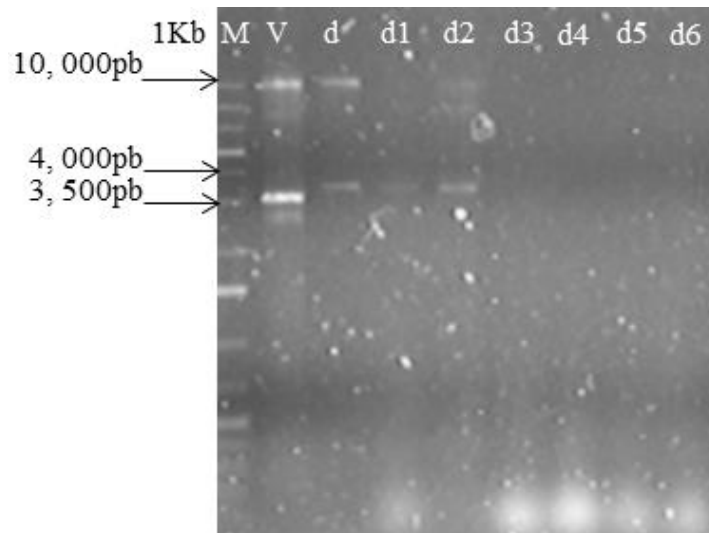


Figura 35. Gel de agarosa al 1%. Extracción plasmídica de las cepas BL21 (DE3) con la construcción BfpI- pET3a. Pozo d1, d2, d3, d4, d5 y d6 cepas BL21 (DE3) transformadas con la construcción BfpI-pET3a. Pozo V es el vector vacío. Pozo d construcción de BfpI-pET3a. Marcador de peso 1Kb.

Por otro lado se hizo la extracción plasmídica de las clonas BL21 (DE3) transformadas con el plásmido BfpJ-pET3a (clona 11). En el corrimiento electroforético se observó que la clona J3 (BfpJ-pET3a) mostró un perfil de bandeo similar al de la construcción BfpJ-pET3a verificada (clona 11, Figura 36), por lo que se utilizó para sobreexpresar la proteína BfpJ. Las muestras J0, J1 y J2 no mostraron un perfil de bandeo, por lo tanto se descartaron.

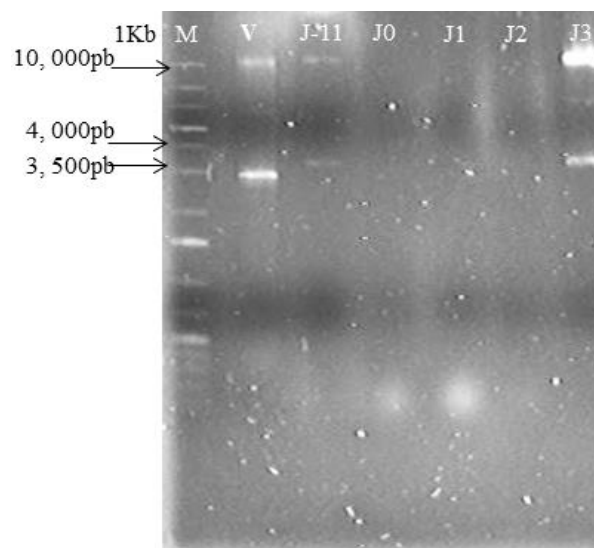


Figura 36. Gel de agarosa al 1%. Extracción plasmídica de las cepas BL21 (DE3) con la construcción BfpJ- pET3a. Pozo J-11(construcción BfpJ-pET3a verificada) J0, J1, J2 y J3 clonas transformadas con la construcción BfpJ-pET3a verificada. Pozo V vector vacío. Marcador de peso 1Kb

7.10 Transformación en la cepa Rosetta gami B (DE3) con las construcciones BfpI-pET3a y BfpJ-pET3a.

Con el fin de comparar que cepa expresa mejor las proteínas BfpI y BfpJ, se transformaron los plásmidos en dos cepas (BL21 (DE3) y Rosetta gami B (DE3)). La cepa Rosetta gami B (DE3) se utilizó pues potencia la formación de puentes disulfuro en el citoplasma bacteriano, favoreciendo de esta manera la formación del puente disulfuro en la región C-terminal de BfpI y de BfpJ.

Las clonas de Rosetta gammi transformadas con la construcción BfpI-pET3a verificada (clonas d1, d3, d4, d5 y d6) mostraron un perfil de bandeo similar al de la construcción BfpI-pET3a verificada (clona d). Por otra parte, las clonas transformadas con la construcción BfpJ-pET3a verificada (clonas J1, J2 y J3) mostraron un perfil de bandeo parecido al de la construcción BfpJ-pET3a verificada (clona J11), Figura 37).

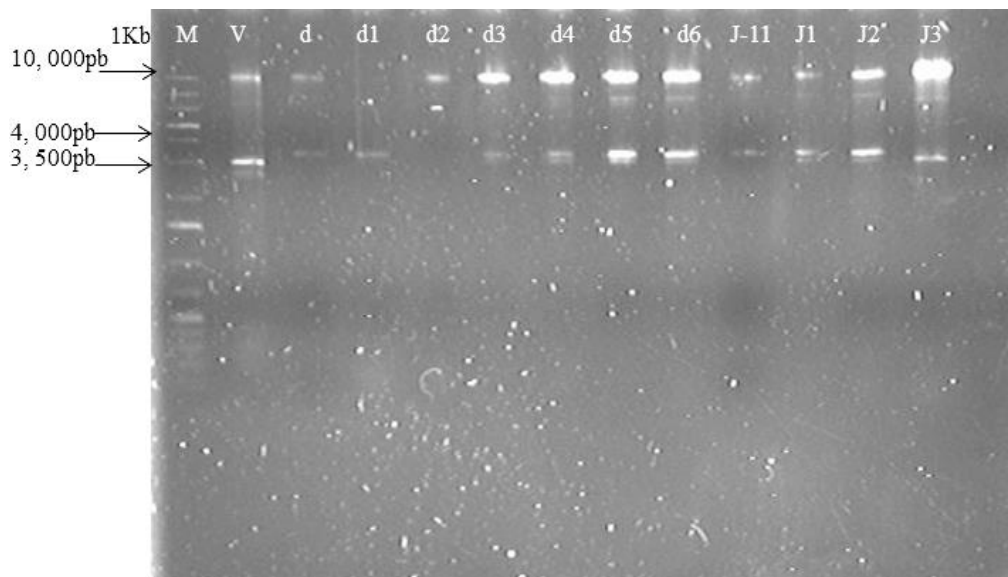


Figura 37. Gel de agarosa al 1%. Extracción plasmídica de las cepas Rosetta gami B (DE3) transformadas con las construcciones BfpI-pET3a y BfpJ-pET3a. Pozos d1, d2, d3, d4, d5 y d6 clonas con la construcción BfpI-pET3a. Pozos J1, J2 y J3 construcciones BfpJ-pET3a. Pozo d construcción verificada BfpI-pET3a. Pozo J11 construcción verificada BfpJ-pET3a.

7.11 Expresión de la proteínas BfpI y BfpJ mediante IPTG en las cepas BL21 (DE3) y Rosetta gami B (DE3)

Con el objetivo de comparar la sobreexpresión de las proteínas y obtener la mayor cantidad posible de proteína se sobreexpresaron las proteínas BfpI y BfpJ en las cepas BL21 (DE3) y Rosetta-gami B (DE3). Esta expresión depende de la inducción mediante la adición de IPTG. Estas cepas tienen integrado en su cromosoma ADN del bacteriófago DE3 el cual codifica para la T7 *RNApolimerasa* que está regulada por el promotor *lacUV5*, la cual se induce mediante IPTG. El vector pET-3a (que se utilizó en las construcciones BfpI-pET3a y BfpJ-pET3a) tiene una región promotora reconocida por la T7 *RNApolimerasa*.

Con el fin de observar la expresión de BfpI y BfpJ de las muestras recolectadas antes y después de la inducción del lisado celular, se calculó mediante el método colorimétrico BCA (Ácido Bicinconínico) la cantidad de cada muestra que se debía de

cargar en un SDS-PAGE para que la concentración final de proteína de cada muestra fuera de 20µg. Posteriormente las muestras se cargaron en un SDS-PAGE y se observó el corrimiento electroforético de las proteínas, de esta manera se visualizó la sobreexpresión de BfpI y de BfpJ (Figura 38).

La sobreexpresión de BfpI y BfpJ en la cepa BL21 fue mayor a las 2 y 4 horas después de la inducción (Figura 38, muestra 4 y 5 (BfpI) respectivamente, 7 y 8 (BfpJ) respectivamente. Se descartó la cepa Rosetta-gami B (DE3) puesto que no se observó la sobreexpresión de las proteínas en ningún caso (2 y 4 horas después de la inducción) (Figura 38).

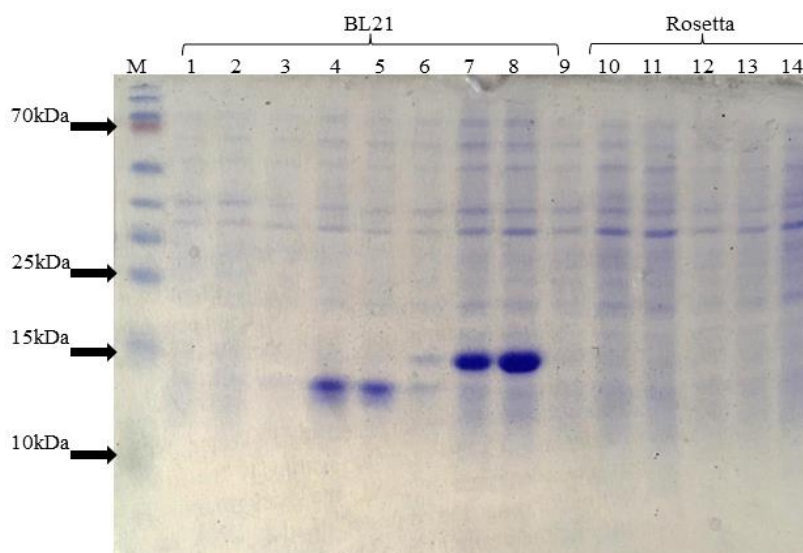


Figura 38. Corrimiento electroforético de las proteínas de las cepas BL21 (DE3) y Rosetta-gami B (DE3). M marcador de peso para proteínas PageRuler Prestained. 1 y 2 cepa BL21 (pET-3a sin inserto) antes y 4 horas después de inducir respectivamente. 3, 4 y 5 cepa BL21 (expresión de BfpI) antes, 2 horas y 4 horas después de inducir respectivamente. 6, 7 y 8 cepa BL21 (expresión de BfpJ) antes, 2 horas y 4 horas después de la inducción respectivamente. Cepa Rosetta-gami (DE3) se numeraron como 9, 10, 11, (expresión de BfpI) antes, 2 horas y 4 horas después de inducir respectivamente. 12, 13 y 14 (expresión de BfpJ) antes, 2 horas y 4 horas después de inducir respectivamente.

7.12 Determinación de la solubilidad de BfpI y BfpJ

Para evaluar si las proteínas de interés podrían estar en cuerpos de inclusión se llevó a cabo el protocolo de la determinación de la solubilidad propuesto por Qiagen (Materiales y métodos). Las muestras se corrieron en un SDS-PAGE donde se observó que en las muestras de BfpI tanto de la fracción soluble como insoluble se recupera la proteína (Figura 39). Por otra parte comparando las muestras 9 y 10 de BfpJ (fracción soluble e insoluble respectivamente) solo la fracción insoluble resulto tener el peso esperado de BfpJ, la parte soluble no coincidía con el peso esperado de BfpJ 17.16 kDa aproximadamente (Figura 39), por lo que se decidió entonces purificar BfpJ en condiciones desnaturalizantes.

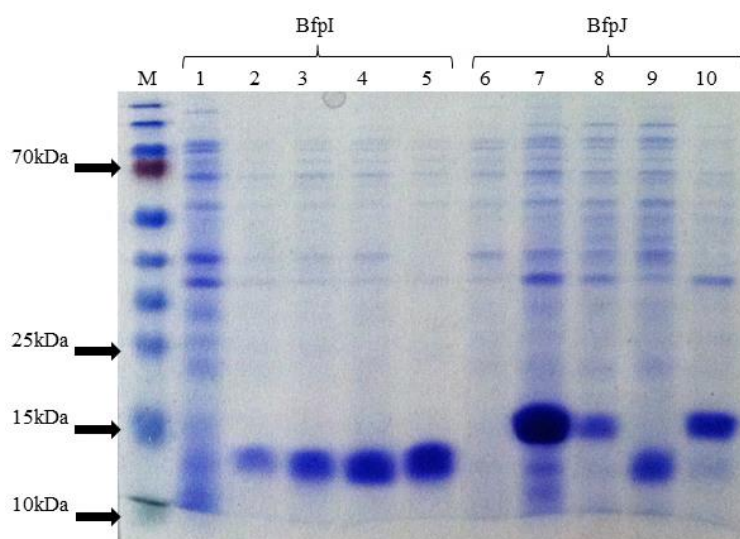


Figura 39. Determinación de la solubilidad de las proteínas BfpI y BfpJ. M marcador de peso para proteínas PageRuler Prestained. 1, 2, 3, 4 y 5 muestras de BfpI, 1 es antes de la inducción; 2, 4 horas después de la inducción del experimento (usado como control); 3 es 4 horas después de la inducción; 4 fracción soluble; 5 fracción insoluble. 6, 7, 8, 9 y 10 (BfpJ), 6 es antes de la inducción, 7 es 4 horas después de la inducción del experimento (usado como control), 8 es 4 horas después de la inducción, 9 fracción soluble, y 10 muestra de fracción insoluble.

7.13 Purificación de la proteína BfpI en condiciones nativas

Como se mencionó anteriormente BfpI se podía recuperar bajo condiciones nativas según el protocolo de determinación de solubilidad. Por lo tanto se siguió el protocolo para

obtener la proteína bajo condiciones nativas (diferentes buffers que contenían imidazol, cuya concentración incrementó durante el proceso de purificación de 10 mM, 20 mM hasta 250 mM), sin embargo al correr las muestras de las fracciones de elución de nuestra proteína BfpI en un gel de poliacrilamida (Figura 40) no se observó la recuperación de las fracciones de elución de la proteína BfpI. Se decidió entonces, realizar la purificación en condiciones desnaturizantes debido probablemente a que la etiqueta de histidinas no está expuesta y afecta la posibilidad de poderla recuperar bajo condiciones nativas.

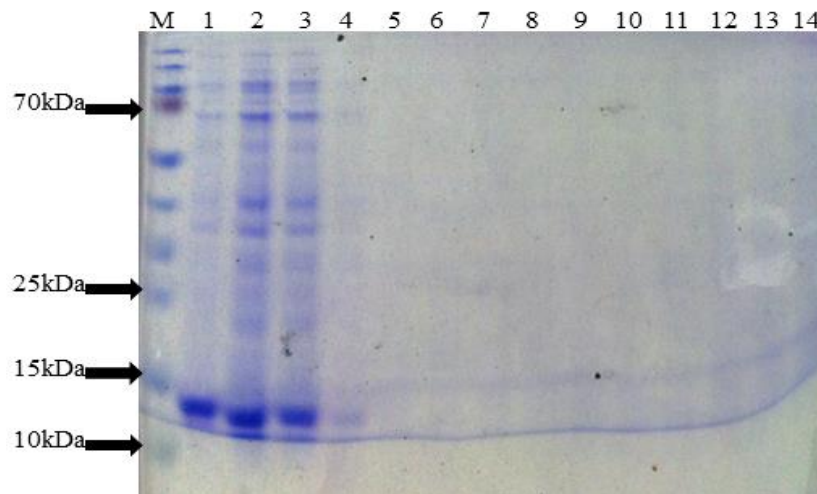


Figura 40. Corrimiento electroforético de las muestras de purificación de BfpI bajo condiciones nativas. M marcador de peso para proteínas PageRuler Prestained. 1 corresponde a la inducción después de 4 horas, la muestra 2 es el lisado bajo condiciones nativas, la muestra 3 es el flujo de las proteínas que no se unieron a la resina, la muestra 4 es el último lavado con el buffer de lavado (20mM imidazol). La muestra 5 hasta la 14 es la elución de la proteína con el buffer de elución (250mM imidazol).

7.14 Purificación de la proteína BfpI y BfpJ en condiciones desnaturizantes

BfpI y BfpJ se purificaron en condiciones desnaturizantes y mediante un SDS-PAGE se observó la proteína en todas las fracciones de elución. La purificación de BfpI se visualizó como bandas que corresponden al peso esperado calculado en 15.10 kDa aproximadamente (Figura 41). A su vez las fracciones de elución de BfpJ se observaron en

SDS-PAGE como bandas correspondientes al peso calculado para BfpJ 17.16kDa aproximadamente (Figura 42).

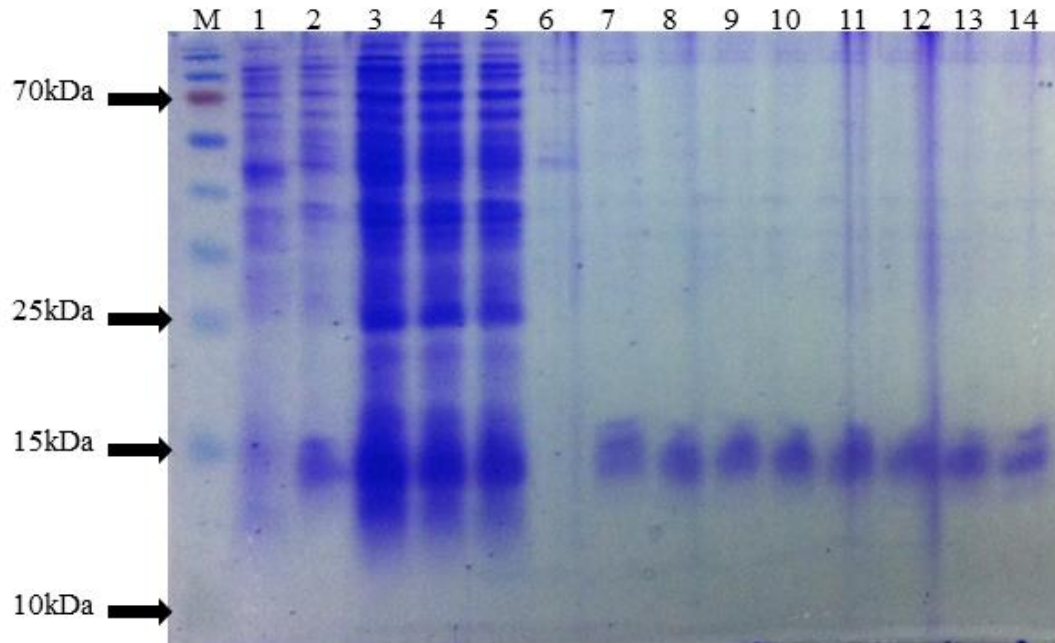


Figura 41. Corrimiento electroforético de las muestras de purificación de BfpI bajo condiciones desnaturizantes. M marcador de peso para proteínas PageRuler Prestained. 1 es la muestra antes de la inducción de la proteína, 2 es la muestra 4 horas después de la inducción, la 3 es el lisado bajo condiciones desnaturizantes, 4 es el flujo de proteínas que no se unieron a la resina de níquel, 5 es el último lavado con urea (pH 8). 6 es el último lavado con urea (pH 6.3). 7, 8, 9, 10, 11, 12, 13 y 14 corresponden a las fracciones de elución de la proteína BfpI.

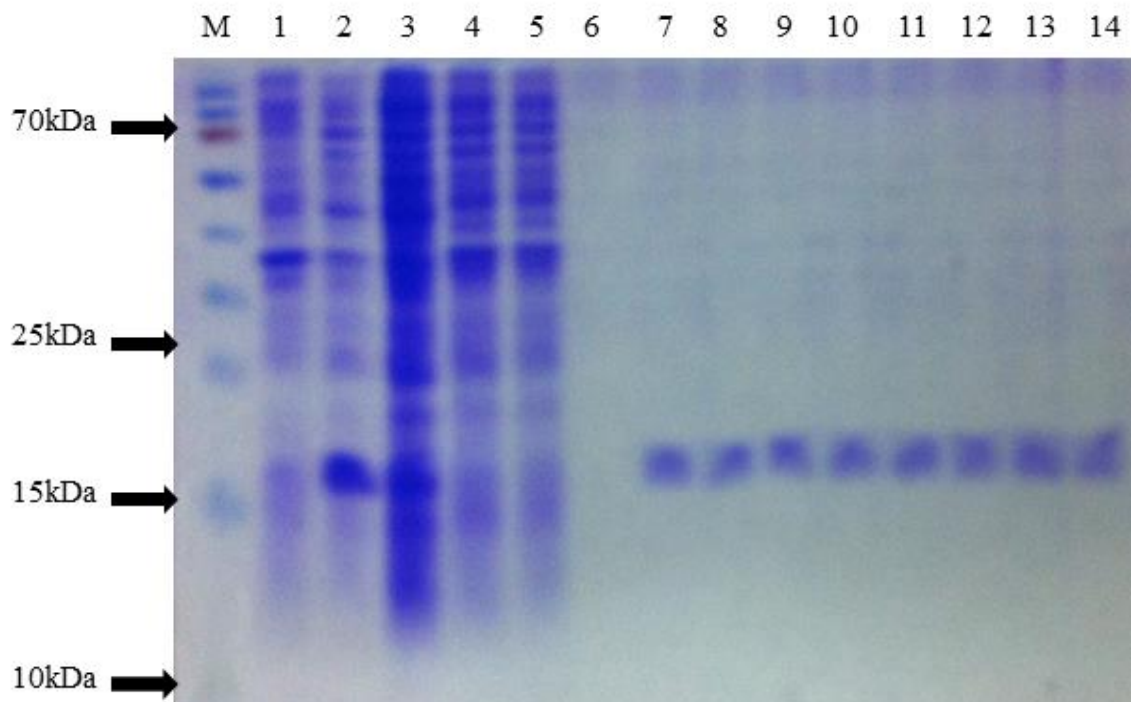


Figura 42. Corrimiento electroforético de las muestras de purificación de BfpJ bajo condiciones desnaturizantes. M marcador de peso para proteínas PageRuler Prestained. 1 es la muestra antes de la inducción de la proteína, 2 es la muestra 4 horas después de la inducción, 3 es el lisado bajo condiciones desnaturizantes, 4 es el flujo de proteínas que no se unieron a la resina de níquel, 5 es el último lavado con urea (pH 8). 6 es el último lavado con urea (pH 6.3). 7, 8, 9, 10, 11, 12, 13 y 14 corresponden a las fracciones de elución de la proteína BfpJ.

7.15 Western Blot

Con el fin de detectar en BfpI y BfpJ que las etiquetas de histidina estuvieran presentes y corresponden a las clonadas se llevó a cabo un ensayo de Western Blot (Materiales y métodos). El Western Blot se realizó con las muestras de purificación en condiciones desnaturizantes de BfpI y BfpJ. Una vez transferidas las proteínas a la membrana, el SDS-PAGE utilizado para la transferencia se tiñó con azul de Coomassie y se observaron unas bandas intensas en los pozos correspondientes a la inducción de BfpI (Pozo 1,2 y 3) y BfpJ (Pozo 5, 6 y 7) (Figura 43).

En el revelado del Western Blot (Figura 44) se observaron unas bandas intensas que corresponden a la detección de las etiquetas de histidina de BfpI y de BfpJ (muestra de la inducción, lisado y en el flujo de proteínas que no se unieron a la resina de níquel).

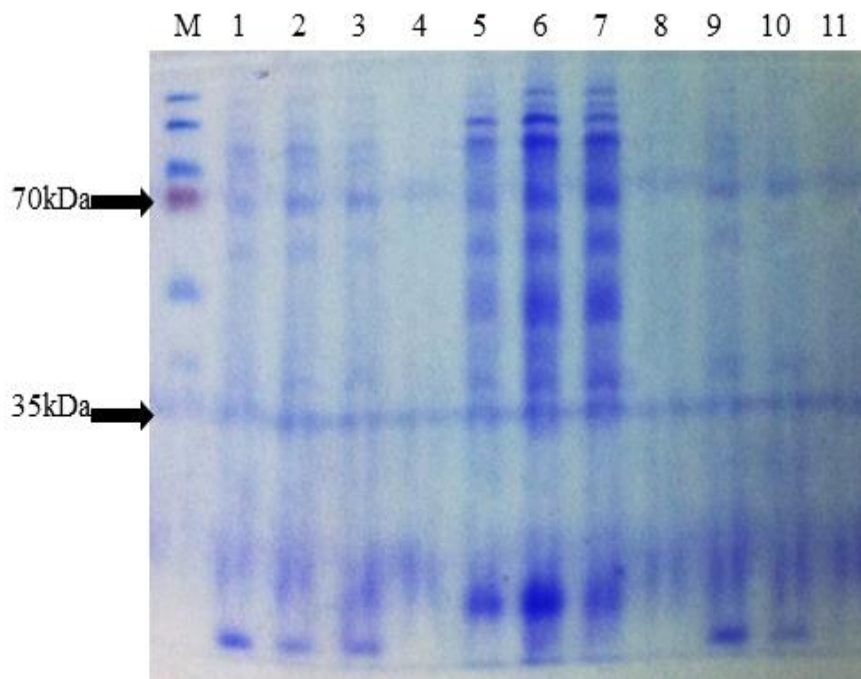


Figura 43. SDS-PAGE utilizado para el Western Blot. M es el marcador de peso para proteínas PageRuler Prestained, 1 es la muestra 4 horas después de la inducción de la expresión (BfpI), 2 lisado bajo condiciones desnaturalizantes, 3 flujo de proteínas que no se unieron a la resina de níquel, 4 elución de la proteína con urea a pH 4.5; 5 es la muestra 4 horas después de la inducción de la expresión (BfpJ), 6 lisado bajo condiciones desnaturalizantes, 7 corresponde al flujo de proteínas que no se unieron a la resina de níquel, 8 elución de la proteína con urea a pH 4.5. La muestra 9 (BfpI purificada por método miniprep) 4 horas después de la inducción, 10 flujo de proteínas que no se unieron a la resina de níquel, 11 elución de BfpI con urea a pH 4.5.

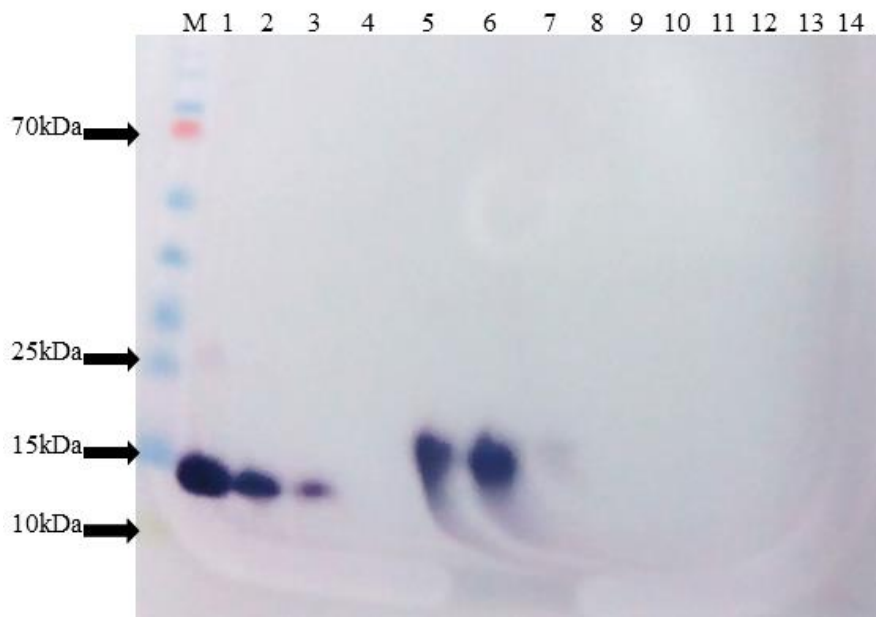


Figura 44. Western Blot dirigido hacia las etiquetas de histidina de las proteínas BfpI y BfpJ. M marcador de peso para proteínas PageRuler Prestained. (BfpI) 1 es la muestra 4 horas después de la inducción de la expresión, 2 es la muestra del lisado bajo condiciones desnaturalizantes, 3 flujo de proteínas que no se unieron a la resina de níquel, 4 elución con urea a pH 4.5. (BfpJ) 5 es la muestra 4 horas después de la inducción de la expresión, 6 lisado bajo condiciones desnaturalizantes, 7 flujo de proteínas que no se unieron a la resina de níquel. 8 elución con urea a pH 4.5. La muestra 9 (BfpI purificada por método miniprep) 4 horas después de la inducción, 10 flujo de proteínas que no se unieron a la resina de níquel, 11 elución de BfpI con urea a pH 4.5.

7.16 Renaturalización de proteínas

Una vez purificadas BfpI y BfpJ continuamos con la diálisis de las proteínas. Esta se realizó con cuatro buffers con una concentración inicial de urea de 6M, 4M, 1M y una final 0M (Materiales y métodos). Con el fin de eliminar la urea y tenerlas de manera soluble, pero una vez finalizado el proceso de diálisis se observó un precipitado blanco en las paredes de la membrana lo que indica que las proteínas que purificamos no son solubles.

7.17 Análisis bioinformático de las secuencias aminoácidas de BfpI y BfpJ.

Se hizo un análisis Bioinformático de las proteínas BfpI y BfpJ mediante el programa Colorseq el cual indica el número de aminoácidos hidrofóbicos, hidrofílicos, aromáticos, con carga positiva o carga negativa presentes en la secuencia de una proteína además de señalar por color y número la posición de estos aminoácidos.

Como se observa en la figura 50 al escindir los 27 aminoácidos de la proteína madura se reduce de 63 a 42 aminoácidos hidrofóbicos (A y B, antes y después de truncar la proteína respectivamente), y también se observan el número de aminoácidos hidrofílicos (C y D, antes y después de truncar la proteína respectivamente) podemos concluir con estos valores que en la secuencia de la proteína truncada 47% de los aminoácidos son hidrofílicos (D) y 30% de los aminoácidos en la proteína truncada son hidrofóbicos (B) por lo tanto en teoría la proteína BfpI truncada de este estudio debería ser soluble, sin embargo no fue así.

Por otra parte los resultados obtenidos mediante el programa Colorseq de la proteína trunca BfpJ (20 aminoácidos extremo N-terminal de la proteína madura) (Figura 51)) arrojaron que en la secuencia completa de aminoácidos (A) 58 son hidrofóbicos y los aminoácidos hidrofóbicos en la proteína truncada (B) son 42. En resumen en la secuencia de aminoácidos de BfpJ truncada un 28% restan de aminoácidos hidrofóbicos y un 44% restan de aminoácidos hidrofílicos. Se debe tener en cuenta que gran parte de estos residuos hidrofóbicos están en la región del dominio globular y se debe analizar la cantidad de aminoácidos hidrofóbicos que se pueden eliminar sin afectar la estructura globular.

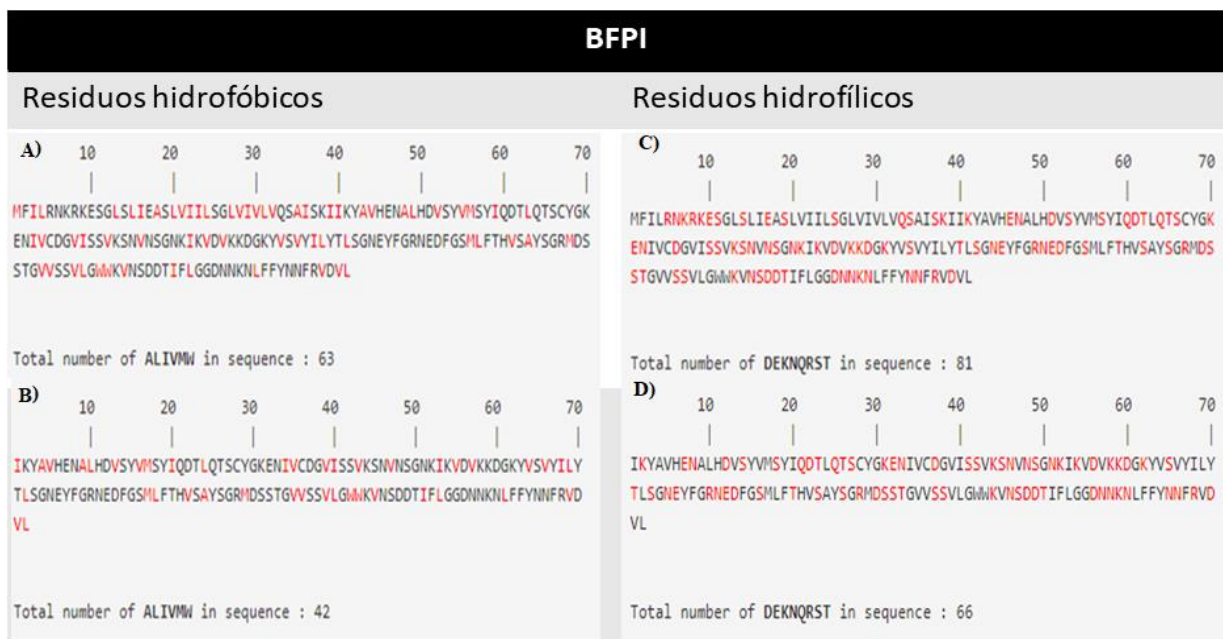


Figura 46. Secuencia de aminoácidos hidrofóbicos e hidrofílicos de BfpI mediante el programa Colorseq. (A) secuencia completa de BfpI, en rojo los aminoácidos hidrofóbicos. (B) secuencia de BfpI truncada en el N-terminal, en rojo los aminoácidos hidrofóbicos. (C) secuencia completa de BfpI, en rojo los aminoácidos hidrofílicos. (D) secuencia de BfpI truncada en el N-terminal, en rojo los aminoácidos hidrofílicos.

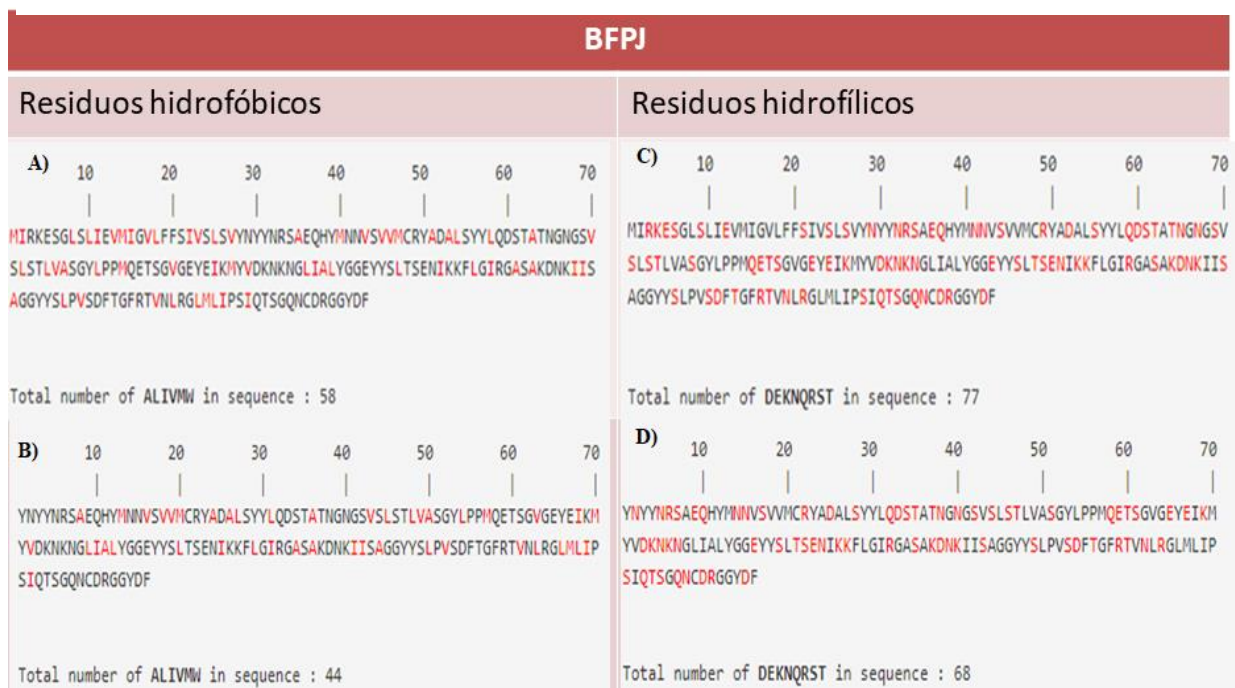


Figura 45. Secuencia de aminoácidos hidrofóbicos e hidrofílicos de BfpJ mediante el programa Colorseq. BfpJ truncada en el N-terminal, en rojo los aminoácidos hidrofóbicos. (A) secuencia completa de BfpJ, en rojo los aminoácidos hidrofóbicos. (B) secuencia de BfpJ truncada en el N-terminal, en rojo los aminoácidos hidrofóbicos. (C) secuencia completa de BfpJ, en rojo los aminoácidos hidrofílicos. (D) secuencia de BfpJ truncada en el N-terminal, en rojo los aminoácidos hidrofílicos.

7.18 Análisis informático de probables epitopes

La importancia de conocer la estructura molecular de las subunidades menores impacta directamente en elucidar la estructura de BFP, su ensamble, desensamble así como comprender como cada subunidad interacciona con las otras. En los modelos de estudio *Streptococcus pyogenes*, *Streptococcus agalactiae* y *Streptococcus pneumoniae*, se han caracterizado al pili como el principal factor de virulencia implicado en la patogénesis de estas bacterias. Los componentes del pilus de *Streptococcus agalactiae* y *Streptococcus pyogenes* (Telford *et al*, 2006) parecen ser buenos candidatos para el desarrollo de vacunas bacterianas o fármacos que tengan la capacidad de inhibir la función del pilus (Craig *et al*, 2003). Por este y otros estudios nos dimos a la tarea de investigar la secuencia de aminoácidos de las subunidades menores de BFP y conocer si servían como probables antígenos y si fuera el caso, cuáles de los epitopes se encuentran en la región globular que es la que estaría expuesta en el filamento. Este análisis complementara en el futuro a los resultados de cristalización de las proteínas BfpI, BfpJ y BfpK para conocer específicamente el lugar de exposición de estos motivos. Esto podría sentar las bases para el desarrollo posterior de alguna vacuna o fármaco dirigido contra BFP.

Mediante el programa NetCTL 1.2 se analizó la secuencia de BfpI, BfpJ y BfpK con la finalidad de identificar la predicción de los péptidos que son reconocidos por las células T, estos péptidos tienen una longitud aproximada de 8 a 11 residuos y también son conocidos como epitopes. Estos epitopes son reconocidos por los linfocitos T citotóxicos mediante el Complejo Mayor de Histocompatibilidad I (por sus siglas en inglés MHC I). El programa arrojó un total de cinco probables epitopes en la secuencia de aminoácidos de BfpI, donde el péptido con un valor más alto en afinidad hacia el MHC I es

MLFTHVSAY (ver anexo) que se sitúa en la región correspondiente al dominio globular como se muestra en la figura 52 en el recuadro rojo, esta figura es la estructura secundaria obtenida mediante el programa Phyre 2, donde se observa con una flecha en rojo los aminoácidos eliminados de la proteína para obtenerla soluble.

Otro programa que se utilizó para conocer los probables epitopes de BfpI, BfpJ y BfpK fue el NetMHCII 2.2. Este programa busca la afinidad de los epitopes y los alelos del Complejo Mayor de Histocompatibilidad II (MHC II): HLA-DR, HLA-DQ, HLA-DP. Para BfpI los resultados que arrojó este programa indicaron un total de 44 péptidos de afinidad alta; sin embargo, la secuencia del péptido con mayor afinidad fue GSMLFTHVSAYSGRM (ver anexo) que se visualiza en la figura 52 en el recuadro morado, este péptido se encuentra en la región globular de BfpI

Por último probamos el programa VaxiJen el cual solo indica la probabilidad de que la secuencia de la proteína sea un antígeno. Para BfpI arrojó un valor de 0.3347 lo que indica este valor es que probablemente no es un buen antígeno(ver anexo).

Los resultados obtenidos con el programa NetCTL 1.2 con BfpJ mostraron 9 ligandos hacia el MHC I (ver anexo), donde el péptido con mayor afinidad fue LIALYGGEY ubicado en el dominio globular de la proteína recuadro rojo de la figura 53, la flecha roja indica los aminoácidos que se eliminaron para obtener la proteína soluble.

En el programa el NetMHCII 2.2. Se observó un total de 76 péptidos de afinidad alta, donde el péptido con mayor afinidad fue ALSYYLQDSTATNGNG (ver anexo) en recuadro morado de la figura 53. El programa VaxiJen para BfpJ arrojó un resultado de 0.5998, lo que indica que es probablemente un buen antígeno (ver anexo).

Para BfpK los resultados obtenidos en el programa NetCTL 1.2 mostraron un total de 10 péptidos de alta afinidad hacia el MHC I (ver anexo) donde el péptido con mayor afinidad fue LLSAYGLNY como se observó en el recuadro rojo de la figura 54. La flecha roja indica los aminoácidos que se eliminaron de BfpK para obtener la proteína soluble.

Con el programa NetMHCII 2.2. Se observó un total de 56 péptidos de afinidad alta para BfpK, donde el péptido con mayor afinidad fue YQVYFGNGYTLISG (ver anexo) en recuadro morado de la figura 54. Finalmente en el programa VaxiJen para BfpK arrojó un resultado de 0.4995, lo que indica que es probablemente un buen antígeno (ver anexo).



Figura 47. Estructura secundaria de BfpI mediante el programa Phyre 2. El recuadro rojo se observa el péptido con mayor afinidad hacia el MHC I. Recuadro morado péptido con mayor afinidad hacia en MHC II

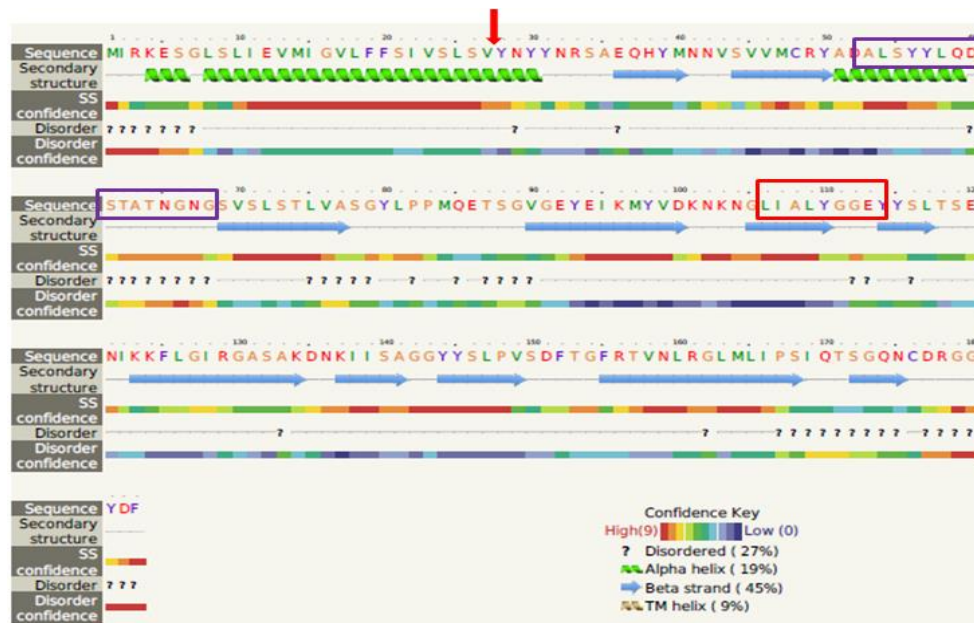


Figura 48. Estructura secundaria de BfpJ mediante el programa de Phyre 2. En el recuadro rojo se observa el péptido con mayor afinidad hacia el MHC I. Recuadro morado péptido con mayor afinidad hacia en MHC II

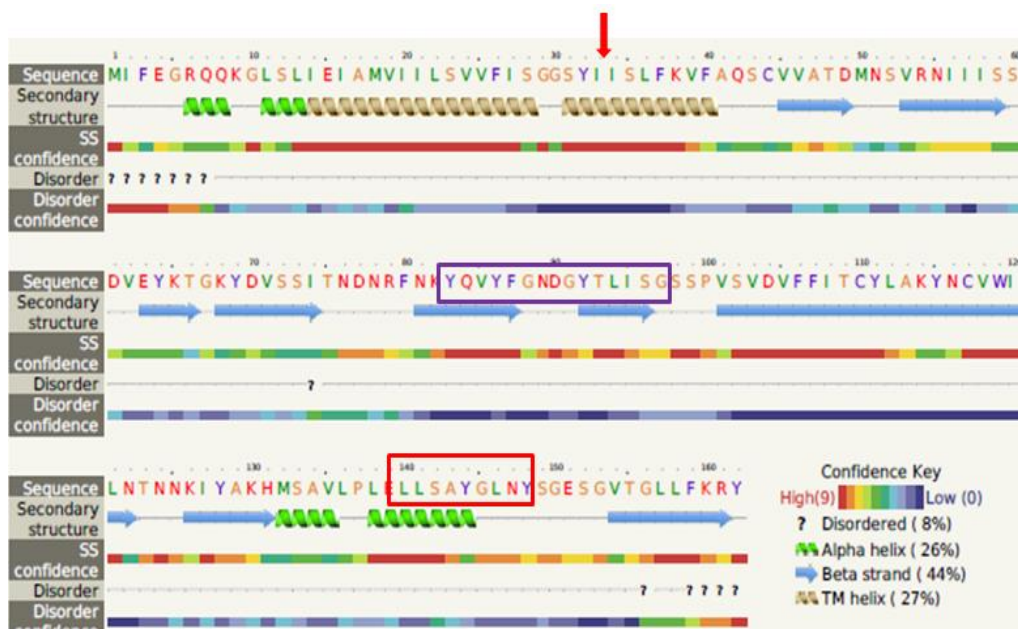


Figura 49. Estructura secundaria de BfpK mediante el programa de Phyre 2. En el recuadro rojo se observa el péptido con mayor afinidad hacia el MHC I. Recuadro morado péptido con mayor afinidad hacia en MHC II.

8 DISCUSIÓN DE RESULTADOS

Las pilinas menores presentes en los pili tipo IV tienen diferentes estructuras que se han resuelto mediante la RMN o cristalografía de rayos X. En el presente trabajo se clonaron, expresaron y purificaron las pilinas menores del BFP de EPEC deletando el extremo N-terminal hidrofóbico, con el fin de obtener de manera soluble estas proteínas para su posterior cristalización.

Muchas de las estructuras de pilinas tipo IV han sido truncadas en los primeros 28 aminoácidos del extremo N-terminal hidrofóbico y posteriormente se han cristalizado. (Craig & Li, 2008). En el año 2003, Craig y colaboradores obtuvieron la estructura de la proteína TcpA (pilina tipo IVb de *V. cholerae*) por cristalización mediante la eliminación de los primeros 28 aminoácidos del extremo N-terminal; a su vez también obtuvieron la estructura de la pilina PAK mediante cristalización truncándola de manera similar a TcpA, lo cual les permitió obtener las proteínas solubles. Por otra parte, los experimentos realizados por Ramboarina y colaboradores en el 2005 permitieron obtener la estructura de BfpA mediante RMN expresando y purificando la proteína BfpA madura, a la cual se le eliminaron los primeros 22 residuos del extremo N-terminal. A pesar de que la cristalografía de rayos X ha sido utilizada como método para resolver la estructura tridimensional de pilinas tipo IV, uno de los problemas que se tiene al resolver la estructura de estas proteínas es que al tener una región hidrofóbica en el N-terminal llega a ser un problema que impide la cristalización (Craig, 2003).

Los estudios antes mencionados junto con el análisis de las estructuras secundarias de BfpI, BfpJ y BfpK (mediante el programa Phyre 2 Figura 47, 48 y 49) nos orientaron a eliminar parte de las pilinas menores (extremo N-terminal) para la obtención de estas

proteínas de forma soluble. En la proteína BfpI se deletaron 27 aminoácidos de la región hidrofóbica de la proteína madura o 39 aminoácidos de la secuencia de la prepilina (117 pb a partir del ATG). Para BfpJ se eliminaron los primeros 20 aminoácidos de la proteína madura o 27 aminoácidos de la prepilina (81 pb a partir del ATG) y para BfpK se eliminaron los primeros 23 aminoácidos de la proteína madura o 33 aminoácidos de la prepilina (99 pb de la secuencia del gen a partir del ATG).

En el presente estudio se diseñaron primers específicos para amplificar la región de cada gen (*bfpI*, *bfpJ* y *bfpK*) sin la porción N-terminal hidrofóbica, se logró amplificar la región del gen correspondiente a la porción globular de las proteínas. Para amplificar el gen *bfpK* fue necesario probar en repetidas ocasiones gradientes de temperatura para obtener el mejor amplificado. Posteriormente se clonó cada gen en el vector pET-3a y se verificó cada construcción por doble restricción (*NdeI* y *BamHI*), PCR y secuenciación, así se obtuvo una clona (d2 y 11 BfpI y BfpJ respectivamente) con cada construcción en las cepas de *E. coli* HB101 y BL21 (DE3).

Se expresaron las proteínas BfpI y BfpJ en la cepa BL21 (DE3). BfpK no se expresó porque para obtener el amplificado del gen los primers diseñados se sintetizaron a una concentración muy baja (2415.6 ng/μL y 44.3ng/μL) fue necesario sintetizarlos de nuevo para continuar con los gradientes de temperatura y obtener el mejor amplificado. (7504.8 ng/μL y 5461.2 ng/μL). Esto retrasó los posteriores ensayos de clonación, además de que la verificación por secuenciación (construcción BfpK-pET3a) no se confirmó porque en los resultados de secuenciación, la secuencia no pertenece al inserto sino a la región río abajo del vector pET-3a, por lo que no se analizó el marco de lectura del inserto.

La secuenciación es muy importante para confirmar la fidelidad de la proteína que se pretende expresar, por tal motivo no se realizó la expresión de BfpK.

Las proteínas BfpI y BfpJ se lograron purificar solo en condiciones desnaturizantes a pesar de intentar purificarlas en condiciones nativas. Esto pudo deberse a que en condiciones nativas no se retuvo la proteína en la columna de níquel debido a que la etiqueta de histidinas se encontraba en una parte inaccesible para interactuar con la columna de níquel.

En condiciones desnaturizantes BfpI y BfpJ podrían tener expuesta la etiqueta de histidinas y al interactuar esta con la resina de níquel las proteínas se retienen en la matriz, por consiguiente las proteínas que no tienen la etiqueta se eluyen. Por lo que la purificación de BfpI y BfpJ bajo condiciones desnaturizantes fue eficiente.

No obstante, el inconveniente de purificar las proteínas bajo condiciones desnaturizantes es que estas pierden su conformación tridimensional y su plegamiento característico por lo que es necesario renaturalizarlas para que adquieran su conformación nativa. El proceso de renaturalización consiste en eliminar paulatinamente el agente desnaturizante bajo condiciones que minimicen la agregación y permitan la formación de puentes disulfuro (Wingfield 2016).

Las proteínas BfpI y BfpJ se renaturalizaron mediante diálisis, sin embargo estas se precipitaron a pesar de haber eliminado la región hidrofóbica N-terminal. La secuencia de aminoácidos de BfpI y BfpJ sin la porción de aminoácidos hidrofóbicos obtenida mediante el programa Colorseq nos demuestra que la cantidad de aminoácidos hidrofílicos presentes en cada proteína es mayor en comparación con los hidrofóbicos (47% hidrofílicos y 30% hidrofóbicos para BfpI y 44% hidrofílicos y 28% hidrofóbicos para BfpJ) por lo que el

precipitado de las proteínas no significa que el problema sea la porción de aminoácidos hidrofóbicos deletados, sino más que podría ser el método de purificación y/o diálisis.

El resultado de la precipitación de BfpI y BfpJ posiblemente se deba a algún factor no tomado en cuenta durante el proceso de purificación y renaturalización, debido a que se determinan de manera empírica. Para la purificación elegimos la cromatografía de afinidad (Ni-NTA) dirigida hacia etiquetas de histidinas por ser uno de los métodos más utilizados para la purificación de proteínas recombinantes por la facilidad de purificación. Sin embargo la desventaja que tiene es que se requiere más experiencia y tiempo para desarrollar protocolos de purificación en proteínas donde no se ha estandarizado el protocolo de purificación. Por lo que para la proteína de interés se deben de conocer de manera detallada sus características químicas y físicas importantes para mantenerlas en cada paso de la purificación hasta su renaturalización con el fin de mantener la estabilidad de la proteína, por ejemplo el pH es un factor importante que afecta de manera irreversible su plegamiento, solubilidad y función (Ritchie, 2012).

Durante el proceso de purificación en condiciones desnaturizantes (en la columna de Ni-NTA) se manejan buffers de diferentes pH. En la elución de la proteína de interés el cambio de pH puede afectar la estabilidad de la proteína, por lo que se sugiere que la proteína eluida se neutralice inmediatamente o se diluya con el fin de minimizar el daño (Ritchie 2012). Durante esta etapa se recuperó a BfpI y BfpJ y se mantuvo en el buffer de elución a pH 4.5, por lo que tal vez de debió mantener una vez eluidas en un buffer de pH neutro.

Por otro lado la renaturalización de proteínas de un solo dominio y pequeñas (10 a 20 kDa) generalmente no es problemático (Wingfield, 2016) por lo que BfpI de un peso

aproximado de 15.10kDa y BfpJ de un peso de 17.16 kDa son proteínas pequeñas de un solo dominio por lo cual no habría inconveniente en la renaturalización a menos que el problema sea relacionado a las condiciones que se utilizaron en la renaturalización donde estas no fueran las más apropiadas durante el proceso de diálisis.

La renaturalización para cada proteína debe determinarse empíricamente por lo que con BfpI y BfpJ se siguieron recomendaciones para determinar las mejores condiciones para renaturalizarlas, por ejemplo eliminar los desnaturalizantes por diálisis y mantener la solubilidad de las proteínas durante el plegado mediante el uso de codisolventes tales como concentraciones no desnaturalizantes de urea, el uso de glicina que tiene propiedades solubilizantes, el uso de glicerol y etilenglicol así como de glucosa que estabilizan la estructura de las proteínas, el mantenimiento del pH neutro (pH 7.5 se manejó en este experimento) e incluir sales como KCL, NaCl o MgCl₂(The QIAexpressionist™)

Además de lo anterior también se recomienda manejar durante la renaturalización una concentración baja de la proteína (10-50µg/ml) y agregar un par redox de glutatión reducido y glutatión oxidado (para crear condiciones oxidantes para favorecer enlaces disulfuro) así como llevar la diálisis en frío (The QIAexpressionist™). Estas recomendaciones no las seguimos por lo que es un factor importante para que BfpI y BfpJ no se hayan plegado de la forma adecuada, por lo que es importante repetir la diálisis apegándose a estas recomendaciones. Sin embargo realizar las recomendaciones antes mencionadas no asegura recuperar la forma nativa y soluble de las proteínas, sino que todo depende de la naturaleza de las proteínas, pues se ha reportado que algunas proteínas, una vez plegadas, pueden tener bajas solubilidades y ser especialmente susceptibles a la agregación (Wingfield, 2016).

Un posible método recomendable para purificar BfpI y BfpJ es seguir el de Holzinger y colaboradores (1996) donde se menciona que las proteínas con etiquetas de histidinas purificadas bajo condiciones desnaturalizantes pueden usarse para renaturalizarse directamente en la columna de Ni-NTA antes de la elución usando un gradiente de urea inverso (el gradiente se puede llevar acabo en el equipo de cromatografía líquida rápida en proteínas (FPLC). Muchas proteínas que son insolubles cuando se repliegan en solución pueden ser replegadas exitosamente mientras se mantienen inmovilizadas a la matriz de Ni-NTA porque al inmovilizarse un extremo de la proteína durante la renaturalización parece evitar las interacciones intermoleculares que provocan la formación de agregados (Holzinger *et al.*, 1996) y una vez que se renaturaliza la proteína se puede eluir adicionando imidazol.

Por otro lado otro método para purificar BfpI y BfpJ y posiblemente BfpK consiste en usar el amplificado de cada gen sin porción la hidrofóbica (inserto que se amplificó con primers específicos) para clonarlo en un vector de secreción donde la señal de secreción en el N-terminal dirige a la proteína al espacio periplasmico (donde se acelera el plegamiento y la formación de enlaces disulfuro) y la translocación a través de la membrana plasmática resultaría en la escisión de la secuencia líder de secreción (Beacham, 1979). De esta forma se evita la formación de cuerpos de inclusión y la purificación también simplifica porque el contenido de proteína en el espacio periplasmico es del 4% del total de proteínas de *E. coli* (Beacham, 1979). Por lo tanto se simplifica la purificación de BfpI, BfpJ y BfpK en su estructura nativa, posteriormente se puede usar la cromatografía de exclusión molecular para la etapa final de purificación es un método más veloz y confiable donde las moléculas eluyen por orden decreciente de tamaño molecular.

Por otra parte los resultados obtenidos del análisis informático de posibles epitopes en BfpI, BfpJ y BfpK sirven de manera presuntiva para seleccionarlas como candidatas para el desarrollo de vacunas. Existen vacunas anti-pilus diseñadas para diversas especies como *Neisseria gonorrhoeae*, *Escherichia coli* enterotoxigenica, *Haemophilus influenzae*, *Vibrio cholerae* y *Moraxella bovis* (Cehovin *et al.*, 2011).

Los resultados obtenidos mediante el análisis informático demostraron presuntivamente que BfpJ y BfpK son buenos antígenos por tener un número mayor de posibles epitopes según los programas NetCTL 1.2, NetMHCII 2.2, excepto BfpI que no tiene la misma cantidad de epitopes y el programa Vaxijen no lo reporta como posible antígeno ya que tiene un valor de probabilidad muy bajo para considerarse así. Como se puede ver en la tabla 3 la comparación entre los resultados de los probables epitopes de las tres proteínas observamos que los mejores candidatos son BfpJ y BfpK.

Tabla 3 Probables epitopes en la secuencia de BfpI, BfpJ y BfpK.

Programa/ Proteína	NetCTL 1.2	NetMHCII	Vaxijen
BFPI	5 péptidos de alta afinidad	44 péptidos de alta afinidad	0.3347(No antígeno)
BFPJ	9 péptidos de alta afinidad	76 péptidos de alta afinidad	0.5998(probable antígeno)
BFPK	10 péptidos de alta afinidad	56 péptidos de alta afinidad	0.4995(probable antígeno)

La cristalización de rayos X de estas proteínas esclarecería si estos probables epitopes se encuentran expuestos y por lo tanto funcionar como antígenos, por el contrario si se encuentran los epitopes en la parte no expuesta de la proteína aunque los programas los identifiquen como péptidos con alta afinidad hacia el MHC I o el MHC II estos no pueden ser reconocidos por no estar expuestos. Como ejemplo los estudios realizados por Hartung y colaboradores donde hacen mención de cuál es la importancia de conocer la relación entre las pilinas y la estructura del pilus pues esta radica en entender su función en la patogénesis y su potencial como vacuna. Sus estudios revelan la estructura de la pilina DnFimA de *D. nodosus* mediante cristalización de rayos X. La DnFimA tiene un puente disulfuro altamente conservado que conecta el lazo- $\alpha\beta$ y las hojas $\beta 1$, $\beta 2$ lo que provoca una flexibilidad conformacional. Esta flexibilidad es importante para el diseño de vacunas porque esta región flexible y expuesta es fácil de reconocer por anticuerpos anti-peptidos.

Los resultados obtenidos de los programas informáticos NetCTL 1.2, NetMHCII y Vaxijen arrojan una aproximación para usar a BfpJ y BfpK como antígenos, sin embargo no podemos afirmar que lo sean del todo debido a que se necesita realizar otros estudios complementarios que respalden estos resultados.

9 CONCLUSIÓN

Se logró clonar, expresar y purificar las proteínas BfpI y BfpJ bajo condiciones desnaturizantes sin embargo se formaron precipitados después de dializar, por lo que no se pudo continuar con la cristalización.

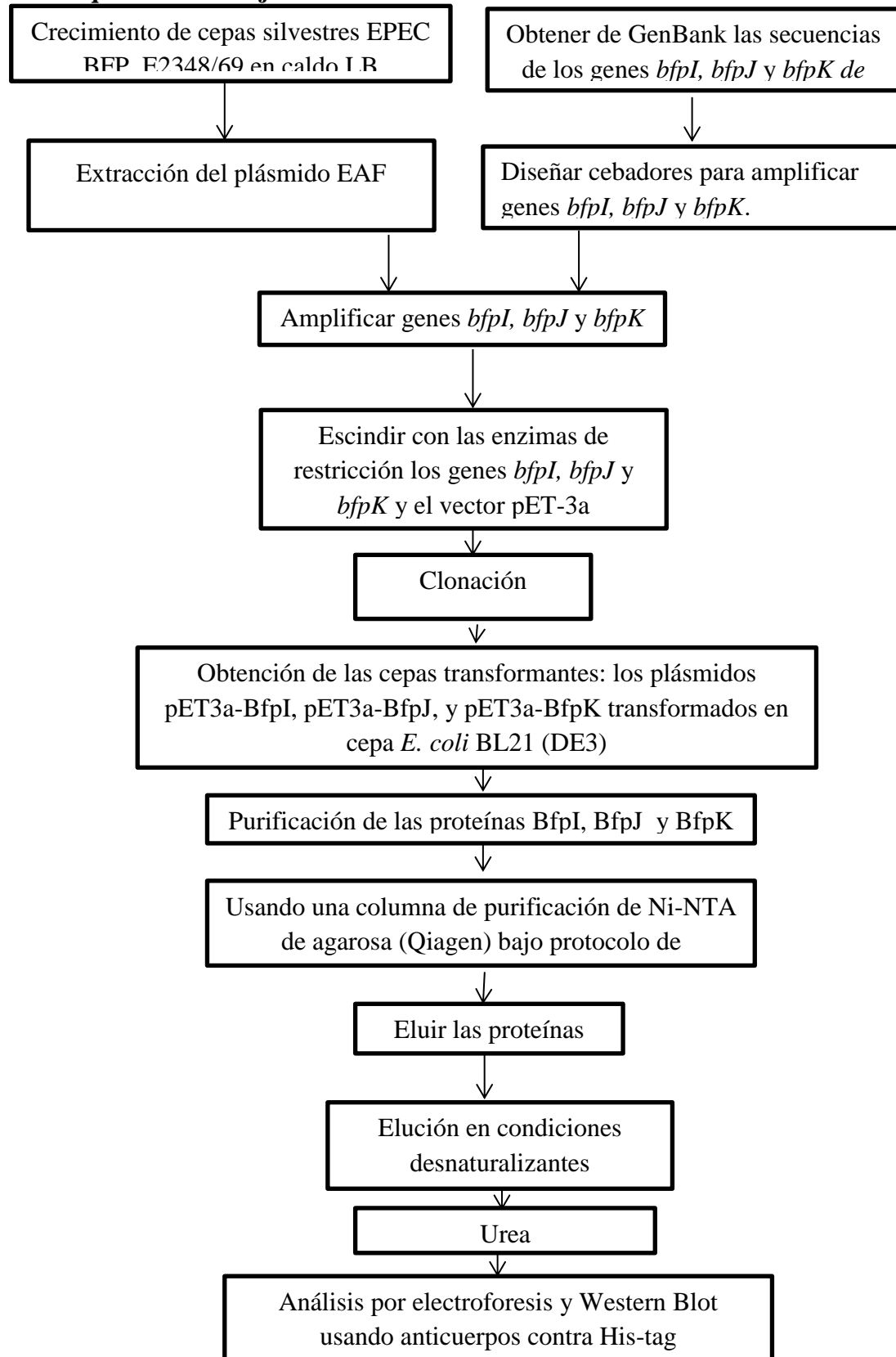
La renaturalización para cada proteína debe determinarse empíricamente por lo que con BfpI y BfpJ es importante repetir la diálisis probando otras condiciones para optimizar el proceso ya que es probable que se pasara por alto algún factor que influyó en la formación de precipitados.

En caso de no obtener la proteína de forma soluble al dializar, se optaría por usar algunas modificaciones al proceso de purificación de BfpI y BfpJ en el que las proteínas purificadas bajo condiciones desnaturizantes pueden usarse para renaturalizarse directamente en la columna de Ni-NTA antes de la elución usando un gradiente de urea inverso. Se espera que este método sea afectivo para purificar y renaturalizar las proteínas, sin embargo otra estrategia que podría funcionar para expresar y purificar las proteínas es el de clonar el gen de interés en un vector de secreción lo que simplifica la purificación de las proteínas en condiciones nativas y posteriormente usar algún método cromatográfico (cromatografía de afinidad, de intercambio iónico, o de exclusión molecular) para purificar las proteínas. Por ultimo falta secuenciar las construcciones BfpK-pET3a y analizar el marco de lectura del inserto. Una vez verificada la construcción continuar con la expresión y purificación de la pilina BfpK, si la purificación se realiza bajo condiciones nativas directamente se usaría la cromatografía de intercambio iónico para eliminar impurezas y continuar con la cromatografía de exclusión molecular. Por otra parte si la purificación se realiza bajo condiciones desnaturizantes las recomendaciones antes mencionadas serian

de gran utilidad para optimizar el proceso renaturalización y purificación bajo condiciones desnaturalizantes. La purificación continuaría con el uso de cromatografía de intercambio iónico y la cromatografía de exclusión molecular.

10 ANEXOS

10.1 Esquema de trabajo

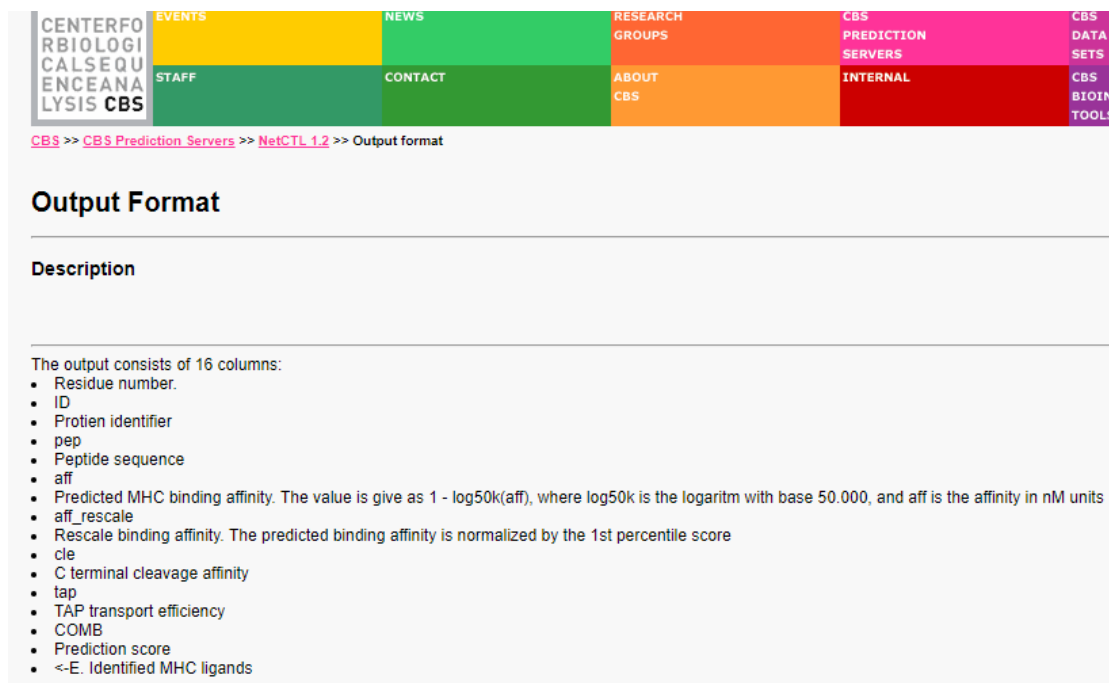


10.2 Epitopes probables

BFPI SECUENCIA COMPLETA

M F I L R N K R K E S
G L S L I E A S L V I I L S G L V I V L
V Q S A I S K I I K Y A V H E N A L H D
V S Y V M S Y I Q D T L Q T S C Y G K E
N I V C D G V I S S V K S N V N S G N K
I K V D V K K D G K Y V S V Y I L Y T L
S G N E Y F G R N E D F G S M L F T H V
S A Y S G R M D S S T G V V S S V L G W
W K V N S D D T I F L G G D N N K N L F
F Y N N F R V D V L

NetCTL 1.2 server predicts CTL epitopes in protein sequences.



The screenshot shows the CBS website navigation menu with the following items: EVENTS, NEWS, RESEARCH GROUPS, CBS PREDICTION SERVERS, CBS DATA SETS, STAFF, CONTACT, ABOUT CBS, INTERNAL, and CBS BIOIN TOOLS. Below the menu is a breadcrumb trail: CBS >> CBS Prediction Servers >> NetCTL 1.2 >> Output format. The main content area is titled "Output Format" and contains a "Description" section. The description states: "The output consists of 16 columns:" followed by a list of 16 items: Residue number, ID, Protein identifier, pep, Peptide sequence, aff, Predicted MHC binding affinity. The value is given as $1 - \log_{50k}(\text{aff})$, where \log_{50k} is the logarithm with base 50,000, and aff is the affinity in nM units, aff_rescale, Rescale binding affinity. The predicted binding affinity is normalized by the 1st percentile score, cle, C terminal cleavage affinity, tap, TAP transport efficiency, COMB, Prediction score, and <-E. Identified MHC ligands.

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

1	ID	Sequence	pep	MFILRNKRK	aff	0.0572	aff_rescale	0.2430	c1e	0.6426	tap	0.7460	COMB	0.3767
2	ID	Sequence	pep	FILRNKRKE	aff	0.0526	aff_rescale	0.2233	c1e	0.0268	tap	-1.4780	COMB	0.1535
3	ID	Sequence	pep	ILRNKRKES	aff	0.0520	aff_rescale	0.2206	c1e	0.1486	tap	-2.2590	COMB	0.1299
4	ID	Sequence	pep	LRNKRKESG	aff	0.0521	aff_rescale	0.2214	c1e	0.0380	tap	-1.2700	COMB	0.1636
5	ID	Sequence	pep	RNKRKESGL	aff	0.0474	aff_rescale	0.2013	c1e	0.3880	tap	1.0920	COMB	0.3141
6	ID	Sequence	pep	NKRKESGLS	aff	0.0540	aff_rescale	0.2293	c1e	0.0242	tap	-1.9780	COMB	0.1340
7	ID	Sequence	pep	KRKESGLSL	aff	0.0534	aff_rescale	0.2267	c1e	0.9447	tap	1.4070	COMB	0.4388
8	ID	Sequence	pep	RKESGLSLI	aff	0.0659	aff_rescale	0.2799	c1e	0.5921	tap	0.7850	COMB	0.4080
9	ID	Sequence	pep	KESGLSLIE	aff	0.0528	aff_rescale	0.2242	c1e	0.0297	tap	-1.4660	COMB	0.1554
10	ID	Sequence	pep	ESGLSLIEA	aff	0.0674	aff_rescale	0.2861	c1e	0.1335	tap	-0.7610	COMB	0.2680
11	ID	Sequence	pep	SGLSLIEAS	aff	0.0572	aff_rescale	0.2429	c1e	0.0249	tap	-2.6080	COMB	0.1162
12	ID	Sequence	pep	GLSLIEASL	aff	0.0577	aff_rescale	0.2451	c1e	0.9077	tap	0.8450	COMB	0.4235
13	ID	Sequence	pep	LSLIEASLV	aff	0.1050	aff_rescale	0.4459	c1e	0.1309	tap	0.3470	COMB	0.4829
14	ID	Sequence	pep	SLIEASLVI	aff	0.0749	aff_rescale	0.3181	c1e	0.8602	tap	0.7240	COMB	0.4833
15	ID	Sequence	pep	LIEASLVII	aff	0.0717	aff_rescale	0.3043	c1e	0.0978	tap	0.5770	COMB	0.3478
16	ID	Sequence	pep	IEASLVIIIL	aff	0.0614	aff_rescale	0.2607	c1e	0.3434	tap	0.9070	COMB	0.3576
17	ID	Sequence	pep	EASLVIIILS	aff	0.0751	aff_rescale	0.3191	c1e	0.0311	tap	-2.3920	COMB	0.2041
18	ID	Sequence	pep	ASLVIIISLG	aff	0.0629	aff_rescale	0.2671	c1e	0.0282	tap	-1.3600	COMB	0.2034
19	ID	Sequence	pep	SLVIIISGL	aff	0.0616	aff_rescale	0.2614	c1e	0.8922	tap	1.2590	COMB	0.4582
20	ID	Sequence	pep	LVIIISGLV	aff	0.0765	aff_rescale	0.3247	c1e	0.1860	tap	0.5800	COMB	0.3816
21	ID	Sequence	pep	VIIISGLVI	aff	0.0628	aff_rescale	0.2667	c1e	0.2556	tap	0.8420	COMB	0.3471
22	ID	Sequence	pep	ILSGLVIV	aff	0.0624	aff_rescale	0.2648	c1e	0.3752	tap	0.5310	COMB	0.3476
23	ID	Sequence	pep	ILSGLVIVL	aff	0.0673	aff_rescale	0.2857	c1e	0.9711	tap	1.0020	COMB	0.4815
24	ID	Sequence	pep	LSGLVIVLV	aff	0.1046	aff_rescale	0.4441	c1e	0.7980	tap	0.2050	COMB	0.5741
25	ID	Sequence	pep	SGLVIVLVQ	aff	0.0548	aff_rescale	0.2326	c1e	0.0358	tap	-0.2530	COMB	0.2254
26	ID	Sequence	pep	GLVIVLVQS	aff	0.0530	aff_rescale	0.2249	c1e	0.2062	tap	-2.2730	COMB	0.1422
27	ID	Sequence	pep	LVIVLVQSA	aff	0.0598	aff_rescale	0.2538	c1e	0.0472	tap	-0.3540	COMB	0.2432
28	ID	Sequence	pep	VIVLVQSAI	aff	0.0593	aff_rescale	0.2519	c1e	0.1226	tap	0.8510	COMB	0.3129
29	ID	Sequence	pep	IVLVQSAIS	aff	0.0563	aff_rescale	0.2390	c1e	0.0272	tap	-2.0190	COMB	0.1421
30	ID	Sequence	pep	VLVQSAISK	aff	0.0651	aff_rescale	0.2763	c1e	0.2672	tap	0.6330	COMB	0.3480
31	ID	Sequence	pep	LVQSAISKI	aff	0.0719	aff_rescale	0.3051	c1e	0.2576	tap	0.6920	COMB	0.3783
32	ID	Sequence	pep	VQSAISKII	aff	0.0560	aff_rescale	0.2376	c1e	0.1876	tap	0.7160	COMB	0.3016
33	ID	Sequence	pep	QSAISKIIK	aff	0.1088	aff_rescale	0.4619	c1e	0.2263	tap	0.5820	COMB	0.5240
34	ID	Sequence	pep	SAISKIIKY	aff	0.1300	aff_rescale	0.5519	c1e	0.9704	tap	3.0500	COMB	0.8499 <-E
35	ID	Sequence	pep	AISKIIKYA	aff	0.0606	aff_rescale	0.2575	c1e	0.1337	tap	-0.2130	COMB	0.2669
36	ID	Sequence	pep	ISKIIKYAV	aff	0.0649	aff_rescale	0.2754	c1e	0.8706	tap	0.5300	COMB	0.4325
37	ID	Sequence	pep	SKIIKYAVH	aff	0.0511	aff_rescale	0.2170	c1e	0.1935	tap	-0.3550	COMB	0.2283
38	ID	Sequence	pep	KIIKYAVHE	aff	0.0602	aff_rescale	0.2557	c1e	0.0450	tap	-1.2070	COMB	0.2021
39	ID	Sequence	pep	IIKYAVHEN	aff	0.0511	aff_rescale	0.2170	c1e	0.0718	tap	-1.0650	COMB	0.1746
40	ID	Sequence	pep	IKYAVHENA	aff	0.0535	aff_rescale	0.2271	c1e	0.0593	tap	-0.2670	COMB	0.2227
41	ID	Sequence	pep	KYAVHENAL	aff	0.0492	aff_rescale	0.2088	c1e	0.5340	tap	1.3520	COMB	0.3565
42	ID	Sequence	pep	YAVHENALH	aff	0.0937	aff_rescale	0.3977	c1e	0.0355	tap	-0.3240	COMB	0.3868
43	ID	Sequence	pep	AVHENALHD	aff	0.0619	aff_rescale	0.2630	c1e	0.0357	tap	-1.5550	COMB	0.1906
44	ID	Sequence	pep	VHENALHDV	aff	0.0616	aff_rescale	0.2615	c1e	0.2794	tap	0.3940	COMB	0.3231
45	ID	Sequence	pep	HENALHDVS	aff	0.0571	aff_rescale	0.2425	c1e	0.0235	tap	-2.5190	COMB	0.1201
46	ID	Sequence	pep	ENALHDVSY	aff	0.0626	aff_rescale	0.2656	c1e	0.6361	tap	2.5620	COMB	0.4892
47	ID	Sequence	pep	LALHDVSYV	aff	0.0581	aff_rescale	0.2467	c1e	0.3690	tap	0.2550	COMB	0.3148
48	ID	Sequence	pep	ALHDVSYVM	aff	0.0649	aff_rescale	0.2757	c1e	0.9752	tap	0.6390	COMB	0.4539
49	ID	Sequence	pep	LHDVSYVMS	aff	0.0577	aff_rescale	0.2451	c1e	0.0325	tap	-2.2560	COMB	0.1372
50	ID	Sequence	pep	HDVSYVMSY	aff	0.0772	aff_rescale	0.3279	c1e	0.9549	tap	2.6520	COMB	0.6038
51	ID	Sequence	pep	DVSYVMSYI	aff	0.0744	aff_rescale	0.3161	c1e	0.6827	tap	0.3300	COMB	0.4350
52	ID	Sequence	pep	VSYVMSYIQ	aff	0.0683	aff_rescale	0.2898	c1e	0.0812	tap	0.0450	COMB	0.3043
53	ID	Sequence	pep	YSYVMSYIQD	aff	0.0469	aff_rescale	0.1990	c1e	0.0305	tap	-1.3730	COMB	0.1350
54	ID	Sequence	pep	YMSYIQDTL	aff	0.0856	aff_rescale	0.3635	c1e	0.0714	tap	-0.4680	COMB	0.3509
55	ID	Sequence	pep	VMSYIQDTL	aff	0.0819	aff_rescale	0.3477	c1e	0.9259	tap	1.0820	COMB	0.5407
56	ID	Sequence	pep	MSYIQDTLQ	aff	0.0905	aff_rescale	0.3844	c1e	0.1154	tap	0.2160	COMB	0.4125
57	ID	Sequence	pep	SYIQDTLQT	aff	0.0513	aff_rescale	0.2180	c1e	0.4442	tap	-0.2830	COMB	0.2705
58	ID	Sequence	pep	YIQDTLQTS	aff	0.0745	aff_rescale	0.3163	c1e	0.1925	tap	-2.1150	COMB	0.2394
59	ID	Sequence	pep	IQDTLQTS	aff	0.0736	aff_rescale	0.3127	c1e	0.0355	tap	-0.1150	COMB	0.3123
60	ID	Sequence	pep	QDTLQTS	aff	0.0584	aff_rescale	0.2477	c1e	0.7712	tap	2.6030	COMB	0.4936
61	ID	Sequence	pep	DTLQTS	aff	0.0575	aff_rescale	0.2439	c1e	0.0486	tap	-1.7640	COMB	0.1630
62	ID	Sequence	pep	TLQTS	aff	0.0689	aff_rescale	0.2927	c1e	0.6128	tap	0.2920	COMB	0.3992
63	ID	Sequence	pep	TLQTS	aff	0.0557	aff_rescale	0.2366	c1e	0.0280	tap	-1.5120	COMB	0.1652
64	ID	Sequence	pep	QTS	aff	0.0882	aff_rescale	0.3744	c1e	0.0292	tap	-1.2800	COMB	0.3148
65	ID	Sequence	pep	TSCY	aff	0.0685	aff_rescale	0.2910	c1e	0.0468	tap	0.5140	COMB	0.3238
66	ID	Sequence	pep	SCY	aff	0.0562	aff_rescale	0.2386	c1e	0.4509	tap	0.5850	COMB	0.3355
67	ID	Sequence	pep	CY	aff	0.0528	aff_rescale	0.2242	c1e	0.0253	tap	0.1290	COMB	0.2345
68	ID	Sequence	pep	Y	aff	0.0535	aff_rescale	0.2271	c1e	0.0264	tap	-2.1110	COMB	0.1255
69	ID	Sequence	pep	KENI	aff	0.0537	aff_rescale	0.2282	c1e	0.0418	tap	-1.7210	COMB	0.1484
70	ID	Sequence	pep	KENI	aff	0.0536	aff_rescale	0.2275	c1e	0.2589	tap	0.1620	COMB	0.2744
71	ID	Sequence	pep	ENI	aff	0.0540	aff_rescale	0.2293	c1e	0.0599	tap	0.3480	COMB	0.2557
72	ID	Sequence	pep	NI	aff	0.0526	aff_rescale	0.2234	c1e	0.0261	tap	-2.1690	COMB	0.1189
73	ID	Sequence	pep	IV	aff	0.0605	aff_rescale	0.2569	c1e	0.0430	tap	-1.9340	COMB	0.1667
74	ID	Sequence	pep	CD	aff	0.0762	aff_rescale	0.3235	c1e	0.9137	tap	0.2300	COMB	0.4720
75	ID	Sequence	pep	DG	aff	0.0542	aff_rescale	0.2302	c1e	0.3103	tap	-0.0000	COMB	0.2767
76	ID	Sequence	pep	DG	aff	0.0486	aff_rescale	0.2063	c1e	0.0486	tap	-2.7570	COMB	0.0758
77	ID	Sequence	pep	VI	aff	0.0525	aff_rescale	0.2227	c1e	0.0266	tap	-1.4260	COMB	0.1554
78	ID	Sequence	pep	VI	aff	0.0737	aff_rescale	0.3130	c1e	0.4237	tap	0.4710	COMB	0.4001
79	ID	Sequence	pep	ISS	aff	0.0666	aff_rescale	0.2827	c1e	0.0510	tap	-1.2090	COMB	0.2299

80	ID	Sequence	pep	SSVKSNNVNS	aff	0.0692	aff_rescale	0.2939	cle	0.0253	tap	-2.0650	COMB	0.1944
81	ID	Sequence	pep	SVKSNVNSG	aff	0.0571	aff_rescale	0.2423	cle	0.0609	tap	-1.1520	COMB	0.1939
82	ID	Sequence	pep	VKSNNVNSGN	aff	0.0577	aff_rescale	0.2451	cle	0.0283	tap	-1.1540	COMB	0.1917
83	ID	Sequence	pep	KSNNVNSGNK	aff	0.0753	aff_rescale	0.3197	cle	0.1287	tap	0.5650	COMB	0.3672
84	ID	Sequence	pep	SNVNSGNKI	aff	0.0587	aff_rescale	0.2491	cle	0.5998	tap	0.6140	COMB	0.3698
85	ID	Sequence	pep	MVNSGNKIK	aff	0.0580	aff_rescale	0.2462	cle	0.2677	tap	0.6000	COMB	0.3163
86	ID	Sequence	pep	VNSGNKIKV	aff	0.0599	aff_rescale	0.2545	cle	0.8332	tap	0.3610	COMB	0.3975
87	ID	Sequence	pep	NSGNKIKVD	aff	0.0516	aff_rescale	0.2189	cle	0.0285	tap	-1.8540	COMB	0.1305
88	ID	Sequence	pep	SGNKIKVDV	aff	0.0618	aff_rescale	0.2625	cle	0.7120	tap	0.1480	COMB	0.3767
89	ID	Sequence	pep	GNKIKVDVK	aff	0.0430	aff_rescale	0.1827	cle	0.1235	tap	0.1850	COMB	0.2105
90	ID	Sequence	pep	NKIKVDVKK	aff	0.0613	aff_rescale	0.2601	cle	0.6697	tap	0.5900	COMB	0.3900
91	ID	Sequence	pep	KIKVDVKKD	aff	0.0463	aff_rescale	0.1965	cle	0.0240	tap	-1.4410	COMB	0.1281
92	ID	Sequence	pep	IKVDVKKDG	aff	0.0473	aff_rescale	0.2008	cle	0.0283	tap	-1.1440	COMB	0.1478
93	ID	Sequence	pep	KVDVKKDKG	aff	0.0664	aff_rescale	0.2817	cle	0.0648	tap	0.5430	COMB	0.3186
94	ID	Sequence	pep	VDVKKDKGY	aff	0.0577	aff_rescale	0.2452	cle	0.4863	tap	2.8580	COMB	0.4610
95	ID	Sequence	pep	DVKKDKGYV	aff	0.0499	aff_rescale	0.2118	cle	0.1470	tap	0.1820	COMB	0.2429
96	ID	Sequence	pep	VKKDKGYVS	aff	0.0436	aff_rescale	0.1852	cle	0.0258	tap	-2.2750	COMB	0.0753
97	ID	Sequence	pep	KDKGYVSV	aff	0.0651	aff_rescale	0.2762	cle	0.7260	tap	0.3420	COMB	0.4022
98	ID	Sequence	pep	KDKGYVSVY	aff	0.0599	aff_rescale	0.2544	cle	0.9335	tap	2.6620	COMB	0.5275
99	ID	Sequence	pep	DGKYVSVYI	aff	0.0466	aff_rescale	0.1979	cle	0.3356	tap	0.0510	COMB	0.2508
100	ID	Sequence	pep	GKYVSVYIL	aff	0.0508	aff_rescale	0.2155	cle	0.9241	tap	0.8730	COMB	0.3978
101	ID	Sequence	pep	KYVSVYILY	aff	0.0884	aff_rescale	0.3755	cle	0.9768	tap	3.3440	COMB	0.6892
102	ID	Sequence	pep	YVSVYILYT	aff	0.1498	aff_rescale	0.6360	cle	0.0923	tap	-0.4900	COMB	0.6253
103	ID	Sequence	pep	VSVYILYTL	aff	0.0635	aff_rescale	0.2696	cle	0.9414	tap	1.0830	COMB	0.4650
104	ID	Sequence	pep	SVYILYTLS	aff	0.0670	aff_rescale	0.2846	cle	0.0474	tap	-1.8200	COMB	0.2007
105	ID	Sequence	pep	VYLYLTLSG	aff	0.0535	aff_rescale	0.2271	cle	0.0872	tap	-0.9460	COMB	0.1929
106	ID	Sequence	pep	YLYLTLSGN	aff	0.0622	aff_rescale	0.2641	cle	0.0254	tap	-1.1420	COMB	0.2108
107	ID	Sequence	pep	ILYTLSGNE	aff	0.0587	aff_rescale	0.2491	cle	0.0466	tap	-1.3880	COMB	0.1867
108	ID	Sequence	pep	LYTLSGNEY	aff	0.1044	aff_rescale	0.4431	cle	0.0732	tap	3.1660	COMB	0.7474
109	ID	Sequence	pep	YTLSGNEYF	aff	0.1718	aff_rescale	0.7294	cle	0.9451	tap	2.5590	COMB	0.9991 <-E
110	ID	Sequence	pep	TLSGNEYFG	aff	0.0570	aff_rescale	0.2420	cle	0.0569	tap	-1.4530	COMB	0.1779
111	ID	Sequence	pep	LSGNEYFGR	aff	0.0656	aff_rescale	0.2786	cle	0.3454	tap	1.3760	COMB	0.3992
112	ID	Sequence	pep	SGNEYFGRN	aff	0.0616	aff_rescale	0.2614	cle	0.0335	tap	-1.5750	COMB	0.1877
113	ID	Sequence	pep	GNEYFGRNE	aff	0.0427	aff_rescale	0.1812	cle	0.0225	tap	-1.9870	COMB	0.0852
114	ID	Sequence	pep	NEYFGRNED	aff	0.0515	aff_rescale	0.2188	cle	0.0879	tap	-1.7570	COMB	0.1441
115	ID	Sequence	pep	EYFGRNEDF	aff	0.0448	aff_rescale	0.1902	cle	0.7809	tap	2.8480	COMB	0.4498
116	ID	Sequence	pep	YFGRNEDFG	aff	0.0578	aff_rescale	0.2452	cle	0.0221	tap	-1.5780	COMB	0.1696
117	ID	Sequence	pep	FGRNEDFGS	aff	0.0540	aff_rescale	0.2294	cle	0.0233	tap	-2.6100	COMB	0.1024
118	ID	Sequence	pep	GRNEDFGSM	aff	0.0492	aff_rescale	0.2089	cle	0.9633	tap	0.1330	COMB	0.3600

119	ID	Sequence	pep	RNEDFGSML	aff	0.0647	aff_rescale	0.2749	cle	0.7222	tap	0.8720	COMB	0.4268
120	ID	Sequence	pep	NEDFGSMFL	aff	0.0771	aff_rescale	0.3273	cle	0.8754	tap	2.3920	COMB	0.5782
121	ID	Sequence	pep	EDFGSMLFT	aff	0.0535	aff_rescale	0.2273	cle	0.0395	tap	-0.9640	COMB	0.1850
122	ID	Sequence	pep	DFGSMLFTH	aff	0.0560	aff_rescale	0.2376	cle	0.0327	tap	-1.0140	COMB	0.1918
123	ID	Sequence	pep	FGSMLFTHV	aff	0.0670	aff_rescale	0.2845	cle	0.6543	tap	-0.1900	COMB	0.3732
124	ID	Sequence	pep	GSMFLFTHV	aff	0.0704	aff_rescale	0.2988	cle	0.2756	tap	-2.4240	COMB	0.2189
125	ID	Sequence	pep	SMLFTHVSA	aff	0.0604	aff_rescale	0.2563	cle	0.9170	tap	-0.3730	COMB	0.3752
126	ID	Sequence	pep	MLFTHVSAY	aff	0.2568	aff_rescale	1.0904	cle	0.9778	tap	3.1570	COMB	1.3949 <-E
127	ID	Sequence	pep	LFTHVSAYS	aff	0.0482	aff_rescale	0.2047	cle	0.0876	tap	-2.0740	COMB	0.1142
128	ID	Sequence	pep	FTHVSAYS	aff	0.0711	aff_rescale	0.3021	cle	0.1343	tap	-1.4340	COMB	0.2505
129	ID	Sequence	pep	THVSAYS	aff	0.0511	aff_rescale	0.2169	cle	0.2088	tap	1.4440	COMB	0.3205
130	ID	Sequence	pep	HVSAYS	aff	0.1193	aff_rescale	0.5066	cle	0.9372	tap	0.3580	COMB	0.6651
131	ID	Sequence	pep	VSAYS	aff	0.0606	aff_rescale	0.2573	cle	0.0228	tap	-1.7190	COMB	0.1747
132	ID	Sequence	pep	SAYS	aff	0.0635	aff_rescale	0.2695	cle	0.0546	tap	-1.9060	COMB	0.1824
133	ID	Sequence	pep	AYS	aff	0.0489	aff_rescale	0.2077	cle	0.0263	tap	-1.7870	COMB	0.1223
134	ID	Sequence	pep	YS	aff	0.0831	aff_rescale	0.3528	cle	0.0259	tap	-0.6940	COMB	0.3220
135	ID	Sequence	pep	SGR	aff	0.0496	aff_rescale	0.2105	cle	0.0306	tap	-1.6120	COMB	0.1345
136	ID	Sequence	pep	GR	aff	0.0584	aff_rescale	0.2479	cle	0.8429	tap	0.4610	COMB	0.3973
137	ID	Sequence	pep	R	aff	0.1653	aff_rescale	0.7020	cle	0.8360	tap	0.4740	COMB	0.8511 <-E
138	ID	Sequence	pep	DS	aff	0.0559	aff_rescale	0.2375	cle	0.0306	tap	-2.3780	COMB	0.1232
139	ID	Sequence	pep	DS	aff	0.0710	aff_rescale	0.3013	cle	0.0476	tap	-2.4890	COMB	0.1840
140	ID	Sequence	pep	S	aff	0.0853	aff_rescale	0.3621	cle	0.8912	tap	0.2330	COMB	0.5074
141	ID	Sequence	pep	ST	aff	0.0933	aff_rescale	0.3961	cle	0.8563	tap	0.8770	COMB	0.5684
142	ID	Sequence	pep	T	aff	0.0532	aff_rescale	0.2260	cle	0.0293	tap	-1.5510	COMB	0.1528
143	ID	Sequence	pep	GV	aff	0.0566	aff_rescale	0.2404	cle	0.6386	tap	0.9320	COMB	0.3828
144	ID	Sequence	pep	V	aff	0.0678	aff_rescale	0.2880	cle	0.8624	tap	1.0510	COMB	0.4680
145	ID	Sequence	pep	V	aff	0.1709	aff_rescale	0.7256	cle	0.1452	tap	0.6100	COMB	0.7779 <-E
146	ID	Sequence	pep	SV	aff	0.1282	aff_rescale	0.5444	cle	0.9280	tap	0.5330	COMB	0.7102
147	ID	Sequence	pep	SV	aff	0.0549	aff_rescale	0.2333	cle	0.0527	tap	-1.0520	COMB	0.1886
148	ID	Sequence	pep	VL	aff	0.0601	aff_rescale	0.2551	cle	0.0303	tap	-2.2730	COMB	0.1460
149	ID	Sequence	pep	LG	aff	0.0528	aff_rescale	0.2243	cle	0.1152	tap	-1.8590	COMB	0.1487
150	ID	Sequence	pep	G	aff	0.0519	aff_rescale	0.2205	cle	0.0967	tap	-1.5280	COMB	0.1586
151	ID	Sequence	pep	W	aff	0.0470	aff_rescale	0.1997	cle	0.0237	tap	-0.5180	COMB	0.1774
152	ID	Sequence	pep	W	aff	0.0602	aff_rescale	0.2556	cle	0.2310	tap	0.5790	COMB	0.3192
153	ID	Sequence	pep	K	aff	0.0660	aff_rescale	0.2803	cle	0.7103	tap	2.6300	COMB	0.5184

```

154 ID Sequence pep VNSDDTIFL aff 0.0580 aff_rescale 0.2463 cle 0.8351 tap 0.9610 COMB 0.4196
155 ID Sequence pep NSDDTIFLG aff 0.1405 aff_rescale 0.5966 cle 0.0465 tap -1.4470 COMB 0.5312
156 ID Sequence pep SDDTIFLGG aff 0.0717 aff_rescale 0.3043 cle 0.0513 tap -1.6850 COMB 0.2277
157 ID Sequence pep DDTIFLGGD aff 0.0477 aff_rescale 0.2023 cle 0.0253 tap -2.4210 COMB 0.0851
158 ID Sequence pep DTIFLGGDN aff 0.0619 aff_rescale 0.2629 cle 0.0234 tap -1.7230 COMB 0.1803
159 ID Sequence pep TIFLGGDNN aff 0.0544 aff_rescale 0.2309 cle 0.0234 tap -1.2350 COMB 0.1726
160 ID Sequence pep IFLGGDNNK aff 0.0511 aff_rescale 0.2169 cle 0.3137 tap 0.6830 COMB 0.2981
161 ID Sequence pep FLGGDNNKN aff 0.0637 aff_rescale 0.2705 cle 0.0400 tap -1.5450 COMB 0.1993
162 ID Sequence pep LGGDNNKNL aff 0.0485 aff_rescale 0.2057 cle 0.3199 tap 0.4240 COMB 0.2749
163 ID Sequence pep GGDNNKNLF aff 0.0620 aff_rescale 0.2633 cle 0.2735 tap 1.9610 COMB 0.4024
164 ID Sequence pep GDNNKNLFF aff 0.0569 aff_rescale 0.2417 cle 0.2551 tap 1.9240 COMB 0.3761
165 ID Sequence pep DNNKNLFFY aff 0.0761 aff_rescale 0.3230 cle 0.3612 tap 2.3320 COMB 0.4938
166 ID Sequence pep NKNLFFYN aff 0.0501 aff_rescale 0.2126 cle 0.0286 tap -1.5720 COMB 0.1383
167 ID Sequence pep NKNLFFYNN aff 0.0516 aff_rescale 0.2192 cle 0.0471 tap -1.1820 COMB 0.1671
168 ID Sequence pep KNLFFYNNF aff 0.0501 aff_rescale 0.2129 cle 0.5540 tap 2.6480 COMB 0.4284
169 ID Sequence pep NLFYNNFR aff 0.0629 aff_rescale 0.2671 cle 0.3922 tap 1.7540 COMB 0.4136
170 ID Sequence pep LFFYNNFRV aff 0.0561 aff_rescale 0.2382 cle 0.7306 tap 0.7390 COMB 0.3847
171 ID Sequence pep FFYNNFRVD aff 0.0524 aff_rescale 0.2224 cle 0.1478 tap -1.5080 COMB 0.1691
172 ID Sequence pep FYNNFRVDV aff 0.0664 aff_rescale 0.2819 cle 0.8776 tap 0.2850 COMB 0.4277
173 ID Sequence pep YNNFRVDVL aff 0.0591 aff_rescale 0.2511 cle 0.7030 tap 0.5830 COMB 0.3857

```

Number of MHC ligands 5 identified. Number of peptides 173. Protein name Sequence

NUMERO DE RESIDUO	SECUENCIA
34	S A I S K I I K Y
109	Y T L S G N E Y F
126	M L F T H V S A Y
137	R M D S S T G V V
145	V S S V L G W W K

Secuencia de BfpI con los epitopes según el programa NetCTL 1.2 en recuadros de colores

M F I L R N K R K E S
G L S L I E A S L V I I L S G L V I V L
V Q S A I S K I I K Y A V H E N A L H D
V S Y V M S Y I Q D T L Q T S C Y G K E
N I V C D G V I S S V K S N V N S G N K
I K V D V K K D G K Y V S V Y I L Y T L
S G N E Y F G R N E D F G S M L F T H V
S A Y S G R M D S S T G V V S S V L G W
W K V N S D D T I F L G G D N N K N L F
F Y N N F R V D V L

NetMHCII 2.2 server predicts binding of peptides to HLA-DR, HLA-DQ, HLA-DP and mouse MHC class II alleles using artificial neuron networks

DESCRIPTION

In the header of the out is indicated which prediction method and HLA allele were selected, as well as the corresponding two threshold values defining high binding peptides, and weak binding peptides. High binding peptides have an IC50 value below 50 nM, and weak binding peptides an IC50 values below 500 nM. For Artificial Neural network (ANN) prediction the output consists of 6 columns. For weight matrix predictions the output consists of 5 columns:

- Residue number
- Peptide sequence (9mer)
- Prediction score (called $-\log_{50k}(\text{aff})$ for ANN predictions)
- Affinity as IC50 value in nM (only for ANN predictions)
- Bind Level (WB for weak binder, SB for strong binder)
- Sequence name

The predictions for each protein are summarized with a line stating the number of high and weak binding peptides identified.

Input is in FSA format

Peptide length 15

NetMHCII version 2.2.

Strong binder threshold 50.00. Weak binder threshold 500.00.

Allele	pos	peptide	core	1-log50k(aff)	affinity(nM)	%Rank	Identity	Bind Level
HLA-DRB10101	1	MFILRNKRKESGLSL	MFILRNKRK	0.4874	256.2	50.00	Sequence	WB
HLA-DRB10101	2	FILRNKRKESGLSLI	KRKESGLSL	0.5532	125.7	50.00	Sequence	WB
HLA-DRB10101	3	ILRNKRKESGLSLIE	KRKESGLSL	0.5866	87.6	32.00	Sequence	WB
HLA-DRB10101	4	LRNKRKESGLSLIEA	KRKESGLSL	0.6589	40.1	32.00	Sequence	SB
HLA-DRB10101	5	RNKRKESGLSLIEAS	KRKESGLSL	0.5541	124.5	50.00	Sequence	WB
HLA-DRB10101	6	NKRKESGLSLIEASL	KRKESGLSL	0.4701	309.0	50.00	Sequence	WB
HLA-DRB10101	7	KRKESGLSLIEASLV	LSLIEASLV	0.7668	12.5	8.00	Sequence	SB
HLA-DRB10101	8	RKESGLSLIEASLVI	LSLIEASLV	0.7919	9.5	8.00	Sequence	SB
HLA-DRB10101	9	KESGLSLIEASLVII	LSLIEASLV	0.8090	7.9	8.00	Sequence	SB
HLA-DRB10101	10	ESGLSLIEASLVIII	LSLIEASLV	0.8244	6.7	4.00	Sequence	SB
HLA-DRB10101	11	SGLSLIEASLVIIIS	LSLIEASLV	0.7997	8.7	8.00	Sequence	SB
HLA-DRB10101	12	GLSLIEASLVIIISG	LSLIEASLV	0.7633	13.0	8.00	Sequence	SB
HLA-DRB10101	13	LSLIEASLVIIISGL	LSLIEASLV	0.7230	20.0	16.00	Sequence	SB
HLA-DRB10101	14	SLIEASLVIIISGLV	LVIIISGLV	0.7430	16.1	16.00	Sequence	SB
HLA-DRB10101	15	LIEASLVIIISGLVI	VIIISGLVI	0.8002	8.7	8.00	Sequence	SB
HLA-DRB10101	16	IEASLVIIISGLVIV	VIIISGLVI	0.8132	7.5	4.00	Sequence	SB
HLA-DRB10101	17	EASLVIIISGLVIVL	VIIISGLVI	0.8262	6.6	4.00	Sequence	SB
HLA-DRB10101	18	ASLVIIISGLVIVLV	VIIISGLVI	0.8190	7.1	4.00	Sequence	SB
HLA-DRB10101	19	SLVIIISGLVIVLVQ	VIIISGLVI	0.7888	9.8	8.00	Sequence	SB
HLA-DRB10101	20	LVIIISGLVIVLVQS	VIIISGLVI	0.7528	14.5	16.00	Sequence	SB
HLA-DRB10101	21	VIIISGLVIVLVQSA	VIIISGLVI	0.6968	26.6	16.00	Sequence	SB
HLA-DRB10101	22	IISGLVIVLVQSAI	VIVLVQSAI	0.6700	35.5	32.00	Sequence	SB
HLA-DRB10101	23	ILSGLVIVLVQSAIS	IVLVQSAIS	0.7627	13.0	16.00	Sequence	SB
HLA-DRB10101	24	LSGLVIVLVQSAISK	IVLVQSAIS	0.7941	9.3	8.00	Sequence	SB
HLA-DRB10101	25	SGLVIVLVQSAISKI	IVLVQSAIS	0.8131	7.6	4.00	Sequence	SB
HLA-DRB10101	26	GLVIVLVQSAISKII	IVLVQSAIS	0.8126	7.6	4.00	Sequence	SB
HLA-DRB10101	27	LVIIVLVQSAISKIIK	IVLVQSAIS	0.8000	8.7	8.00	Sequence	SB

HLA-DRB10101	28	VIVLVQSAISKIIKY	IVLVQSAIS	0.7799	10.8	8.00	Sequence	SB
HLA-DRB10101	29	IVLVQSAISKIIKYA	VQSAISKII	0.7679	12.3	8.00	Sequence	SB
HLA-DRB10101	30	VLVQSAISKIIKYAV	VQSAISKII	0.6948	27.2	16.00	Sequence	SB
HLA-DRB10101	31	LVQSAISKIIKYAVH	VQSAISKII	0.6178	62.5	32.00	Sequence	WB
HLA-DRB10101	32	VQSAISKIIKYAVHE	VQSAISKII	0.5061	209.2	50.00	Sequence	WB
HLA-DRB10101	33	QSAISKIIKYAVHEN	ISKIIKYAV	0.4587	349.5	50.00	Sequence	WB
HLA-DRB10101	34	SAISKIIKYAVHENA	ISKIIKYAV	0.4554	362.4	50.00	Sequence	WB
HLA-DRB10101	35	AISKIIKYAVHENAL	IIKYAVHEN	0.4668	320.4	50.00	Sequence	WB
HLA-DRB10101	36	ISKIIKYAVHENALH	IIKYAVHEN	0.5034	215.5	50.00	Sequence	WB
HLA-DRB10101	37	SKIIKYAVHENALHD	IIKYAVHEN	0.4913	245.8	50.00	Sequence	WB
HLA-DRB10101	38	KIIKYAVHENALHDV	YAVHENALH	0.5315	159.0	50.00	Sequence	WB
HLA-DRB10101	39	IIKYAVHENALHDVSV	YAVHENALH	0.5685	106.5	50.00	Sequence	WB
HLA-DRB10101	40	IKYAVHENALHDVSY	YAVHENALH	0.4803	276.6	50.00	Sequence	WB
HLA-DRB10101	41	KYAVHENALHDVSYV	VHENALHDV	0.4522	375.2	50.00	Sequence	WB
HLA-DRB10101	42	YAVHENALHDVSYVM	VHENALHDV	0.3770	846.2	50.00	Sequence	WB
HLA-DRB10101	43	AVHENALHDVSYVMS	LHDVSYVMS	0.3856	770.7	50.00	Sequence	WB
HLA-DRB10101	44	VHENALHDVSYVMSY	LHDVSYVMS	0.4376	439.3	50.00	Sequence	WB
HLA-DRB10101	45	HENALHDVSYVMSYI	LHDVSYVMS	0.4656	324.3	50.00	Sequence	WB
HLA-DRB10101	46	ENALHDVSYVMSYIQ	LHDVSYVMS	0.4702	308.7	50.00	Sequence	WB
HLA-DRB10101	47	NALHDVSYVMSYIQD	LHDVSYVMS	0.4619	337.6	50.00	Sequence	WB
HLA-DRB10101	48	ALHDVSYVMSYIQDT	VSYVMSYIQ	0.4804	276.3	50.00	Sequence	WB
HLA-DRB10101	49	LHDVSYVMSYIQDTL	YVMSYIQDT	0.5518	127.7	50.00	Sequence	WB
HLA-DRB10101	50	HDSVSYVMSYIQDTLO	YVMSYIQDT	0.6032	73.2	32.00	Sequence	WB
HLA-DRB10101	51	DVSYVMSYIQDTLQT	YVMSYIQDT	0.6610	39.2	32.00	Sequence	SB
HLA-DRB10101	52	VSYVMSYIQDTLQTS	YVMSYIQDT	0.6501	44.0	32.00	Sequence	SB
HLA-DRB10101	53	SYVMSYIQDTLQTSY	YVMSYIQDT	0.5780	96.2	32.00	Sequence	WB
HLA-DRB10101	54	YVMSYIQDTLQTSYCY	YIQDTLQTS	0.5677	107.5	50.00	Sequence	WB
HLA-DRB10101	55	VMSYIQDTLQTSYCYG	YIQDTLQTS	0.5615	115.0	50.00	Sequence	WB
HLA-DRB10101	56	MSYIQDTLQTSYCYGK	YIQDTLQTS	0.4719	303.0	50.00	Sequence	WB
HLA-DRB10101	57	SYIQDTLQTSYCYGKE	YIQDTLQTS	0.3857	770.3	50.00	Sequence	WB
HLA-DRB10101	58	YIQDTLQTSYCYGKEN	YIQDTLQTS	0.3610	1005.7	50.00	Sequence	WB
HLA-DRB10101	59	IQDTLQTSYCYGKENI	DTLQTSYCYG	0.2932	2094.5	50.00	Sequence	WB
HLA-DRB10101	60	QDTLQTSYCYGKENIV	DTLQTSYCYG	0.2771	2492.8	50.00	Sequence	WB
HLA-DRB10101	61	DTLQTSYCYGKENIVC	SCYGKENIV	0.2552	3160.7	50.00	Sequence	WB
HLA-DRB10101	62	TLQTSYCYGKENIVCD	SCYGKENIV	0.2385	3784.9	50.00	Sequence	WB
HLA-DRB10101	63	LQTSYCYGKENIVCDG	SCYGKENIV	0.2294	4176.7	50.00	Sequence	WB
HLA-DRB10101	64	QTSYCYGKENIVCDGV	CYGKENIVC	0.2211	4569.7	50.00	Sequence	WB
HLA-DRB10101	65	TSCYGKENIVCDGVI	YGKENIVCD	0.2190	4675.6	50.00	Sequence	WB
HLA-DRB10101	66	SCYGKENIVCDGVIS	NIVCDGVIS	0.2399	3728.2	50.00	Sequence	WB
HLA-DRB10101	67	CVGKENIVCDGVISS	NIVCDGVIS	0.2608	2974.1	50.00	Sequence	WB
HLA-DRB10101	68	YGKENIVCDGVISSV	NIVCDGVIS	0.3224	1527.8	50.00	Sequence	WB
HLA-DRB10101	69	GKENIVCDGVISSVK	NIVCDGVIS	0.4018	647.2	50.00	Sequence	WB
HLA-DRB10101	70	KENIVCDGVISSVKS	CDGVISSVK	0.3916	722.9	50.00	Sequence	WB
HLA-DRB10101	71	ENIVCDGVISSVKS	CDGVISSVK	0.3929	712.8	50.00	Sequence	WB
HLA-DRB10101	72	NIVCDGVISSVKS	CDGVISSVK	0.4669	320.1	50.00	Sequence	WB
HLA-DRB10101	73	IVCDGVISSVKS	DGVISSVK	0.4906	247.6	50.00	Sequence	WB
HLA-DRB10101	74	VCDGVISSVKS	CDGVISSVK	0.4096	594.7	50.00	Sequence	WB
HLA-DRB10101	75	CDGVISSVKS	ISSVKS	0.3857	770.1	50.00	Sequence	WB
HLA-DRB10101	76	DGVISSVKS	ISSVKS	0.4146	563.6	50.00	Sequence	WB
HLA-DRB10101	77	GVISSVKS	ISSVKS	0.4020	645.7	50.00	Sequence	WB
HLA-DRB10101	78	VISSVKS	ISSVKS	0.3357	1322.3	50.00	Sequence	WB
HLA-DRB10101	79	ISSVKS	SNVNSG	0.3588	1030.9	50.00	Sequence	WB
HLA-DRB10101	80	SSVKS	SNVNSG	0.4123	577.7	50.00	Sequence	WB
HLA-DRB10101	81	SVKS	SNVNSG	0.4140	567.0	50.00	Sequence	WB
HLA-DRB10101	82	VKS	SNVNSG	0.3614	1002.0	50.00	Sequence	WB
HLA-DRB10101	83	KS	SNVNSG	0.3028	1889.2	50.00	Sequence	WB
HLA-DRB10101	84	SNVNSG	SNVNSG	0.2482	3410.8	50.00	Sequence	WB
HLA-DRB10101	85	MVNSG	VNSG	0.1642	8464.8	50.00	Sequence	WB
HLA-DRB10101	86	VNSG	VNSG	0.1629	8583.2	50.00	Sequence	WB
HLA-DRB10101	87	NSG	IKVDVKKDG	0.1494	9932.9	50.00	Sequence	WB
HLA-DRB10101	88	SG	IKVDVKKDGK	0.1849	6764.0	50.00	Sequence	WB
HLA-DRB10101	89	GK	IKVDVKKDGK	0.2105	5125.0	50.00	Sequence	WB
HLA-DRB10101	90	NK	IKVDVKKDGK	0.2230	4478.0	50.00	Sequence	WB
HLA-DRB10101	91	K	IKVDVKKDGK	0.2616	2947.9	50.00	Sequence	WB
HLA-DRB10101	92	IK	VDVKKDGK	0.3367	1308.7	50.00	Sequence	WB
HLA-DRB10101	93	KV	VDVKKDGK	0.2780	2472.2	50.00	Sequence	WB
HLA-DRB10101	94	VD	VKKDGK	0.4507	381.3	50.00	Sequence	WB
HLA-DRB10101	95	DV	VKKDGK	0.4766	288.1	50.00	Sequence	WB
HLA-DRB10101	96	VK	DGK	0.5259	169.0	50.00	Sequence	WB
HLA-DRB10101	97	KD	GK	0.5553	122.9	50.00	Sequence	WB
HLA-DRB10101	98	KD	GK	0.5662	109.3	50.00	Sequence	WB
HLA-DRB10101	99	DG	KYVSYVYLYT	0.6928	27.8	16.00	Sequence	SB
HLA-DRB10101	100	GK	YVSYVYLYT	0.7163	21.5	16.00	Sequence	SB
HLA-DRB10101	101	KY	VSYVYLYT	0.7302	18.5	16.00	Sequence	SB
HLA-DRB10101	102	YV	SYVYLYT	0.7721	11.8	8.00	Sequence	SB
HLA-DRB10101	103	VSY	VYLYT	0.7657	12.6	8.00	Sequence	SB
HLA-DRB10101	104	SV	YLYT	0.7588	13.6	16.00	Sequence	SB
HLA-DRB10101	105	VY	LYT	0.7859	10.1	8.00	Sequence	SB
HLA-DRB10101	106	YI	LYT	0.7473	15.4	16.00	Sequence	SB
HLA-DRB10101	107	IL	YLYT	0.6856	30.0	32.00	Sequence	SB

HLA-DRB10101	108	LYTLSGNEYFGRNED	LYTLSGNEY	0.6090	68.7	32.00	Sequence	WB
HLA-DRB10101	109	YTLSGNEYFGRNEDF	YTLSGNEYF	0.3172	1615.4	50.00	Sequence	
HLA-DRB10101	110	TLSGNEYFGRNEDFG	EYFGRNEDF	0.1742	7590.2	50.00	Sequence	
HLA-DRB10101	111	LSGNEYFGRNEDFGS	FGRNEDFGS	0.2273	4274.1	50.00	Sequence	
HLA-DRB10101	112	SGNEYFGRNEDFGSM	FGRNEDFGS	0.2454	3514.0	50.00	Sequence	
HLA-DRB10101	113	GNEYFGRNEDFGSML	FGRNEDFGS	0.2706	2675.2	50.00	Sequence	
HLA-DRB10101	114	NEYFGRNEDFGSMLF	FGRNEDFGS	0.3088	1770.1	50.00	Sequence	
HLA-DRB10101	115	EYFGRNEDFGSMLFT	FGRNEDFGS	0.2993	1962.1	50.00	Sequence	
HLA-DRB10101	116	YFGRNEDFGSMLFTH	NEDFGSMLF	0.3287	1426.6	50.00	Sequence	
HLA-DRB10101	117	FGRNEDFGSMLFTHV	FGSMLFTHV	0.4849	263.3	50.00	Sequence	WB
HLA-DRB10101	118	GRNEDFGSMLFTHVS	FGSMLFTHV	0.5074	206.3	50.00	Sequence	WB
HLA-DRB10101	119	RNEDFGSMLFTHVSA	FGSMLFTHV	0.5295	162.5	50.00	Sequence	WB
HLA-DRB10101	120	NEDFGSMLFTHVSAY	FGSMLFTHV	0.5528	126.3	50.00	Sequence	WB
HLA-DRB10101	121	EDFGSMLFTHVSAYS	FGSMLFTHV	0.5472	134.2	50.00	Sequence	WB
HLA-DRB10101	122	DFGSMLFTHVSAYS	FTHVSAYS	0.8060	8.2	8.00	Sequence	SB
HLA-DRB10101	123	FGSMLFTHVSAYS	FTHVSAYS	0.8338	6.0	4.00	Sequence	SB
HLA-DRB10101	124	GSMLFTHVSAYS	FTHVSAYS	0.8505	5.0	1.25	Sequence	SB
HLA-DRB10101	125	SMLFTHVSAYS	FTHVSAYS	0.8487	5.1	1.50	Sequence	SB
HLA-DRB10101	126	MLFTHVSAYS	FTHVSAYS	0.8268	6.5	4.00	Sequence	SB
HLA-DRB10101	127	LTHVSAYS	FTHVSAYS	0.7898	9.7	8.00	Sequence	SB
HLA-DRB10101	128	FTHVSAYS	FTHVSAYS	0.7304	18.5	16.00	Sequence	SB
HLA-DRB10101	129	THVSAYS	HVSAYS	0.3141	1670.6	50.00	Sequence	
HLA-DRB10101	130	HVSAYS	SGRMDSS	0.3319	1377.9	50.00	Sequence	
HLA-DRB10101	131	VSAYS	SGRMDSS	0.3974	678.2	50.00	Sequence	
HLA-DRB10101	132	SAYS	SGRMDSS	0.4621	336.9	50.00	Sequence	WB
HLA-DRB10101	133	AYS	GRMDSS	0.4357	448.3	50.00	Sequence	WB
HLA-DRB10101	134	YS	GRMDSS	0.3987	669.1	50.00	Sequence	
HLA-DRB10101	135	SGRMDSS	MDSSTG	0.4800	277.6	50.00	Sequence	WB
HLA-DRB10101	136	GRMDSS	MDSSTG	0.4327	463.1	50.00	Sequence	WB
HLA-DRB10101	137	RMDSS	STGVSS	0.4557	361.2	50.00	Sequence	WB
HLA-DRB10101	138	MDSSTG	STGVSS	0.4965	232.3	50.00	Sequence	WB
HLA-DRB10101	139	DSSTG	TVSS	0.5055	210.6	50.00	Sequence	WB
HLA-DRB10101	140	SSTG	TVSS	0.4974	230.0	50.00	Sequence	WB
HLA-DRB10101	141	STGVSS	TVSS	0.4676	317.4	50.00	Sequence	WB
HLA-DRB10101	142	TGVSS	TVSS	0.4402	426.9	50.00	Sequence	WB
HLA-DRB10101	143	GVSS	TVSS	0.3434	1217.0	50.00	Sequence	
HLA-DRB10101	144	VSS	GMKVS	0.2572	3094.7	50.00	Sequence	
HLA-DRB10101	145	VSS	GMKVS	0.3042	1860.1	50.00	Sequence	
HLA-DRB10101	146	SS	GMKVS	0.3251	1484.0	50.00	Sequence	
HLA-DRB10101	147	S	GMKVS	0.3646	968.2	50.00	Sequence	
HLA-DRB10101	148	VL	GMKVS	0.4356	448.9	50.00	Sequence	WB
HLA-DRB10101	149	L	GMKVS	0.4006	655.8	50.00	Sequence	
HLA-DRB10101	150	GMKVS	GMKVS	0.3413	1245.6	50.00	Sequence	
HLA-DRB10101	151	W	KNVSD	0.3187	1590.9	50.00	Sequence	

HLA-DRB10101	152	W	KNVSD	0.2925	2110.6	50.00	Sequence	
HLA-DRB10101	153	KNVSD	SD	0.2353	3919.8	50.00	Sequence	
HLA-DRB10101	154	V	NSD	0.2367	3860.0	50.00	Sequence	
HLA-DRB10101	155	NSD	IFL	0.3355	1325.5	50.00	Sequence	
HLA-DRB10101	156	S	DD	0.4085	602.0	50.00	Sequence	
HLA-DRB10101	157	DD	IFL	0.4964	232.6	50.00	Sequence	WB
HLA-DRB10101	158	DT	IFL	0.5954	79.7	32.00	Sequence	WB
HLA-DRB10101	159	T	IFL	0.6088	68.9	32.00	Sequence	WB
HLA-DRB10101	160	I	FL	0.5527	126.5	50.00	Sequence	WB
HLA-DRB10101	161	FL	GG	0.4646	328.1	50.00	Sequence	WB
HLA-DRB10101	162	L	GG	0.2678	2758.1	50.00	Sequence	
HLA-DRB10101	163	GG	NN	0.3047	1850.1	50.00	Sequence	
HLA-DRB10101	164	G	NN	0.4659	323.4	50.00	Sequence	WB
HLA-DRB10101	165	D	NN	0.4652	325.8	50.00	Sequence	WB
HLA-DRB10101	166	N	NN	0.5505	117.4	50.00	Sequence	WB
HLA-DRB10101	167	N	NN	0.6552	41.7	32.00	Sequence	SB

Allele: HLA-DRB10101. Number of high binders 44. Number of weak binders 59. Number of peptides 167

VaxiJen RESULTS

Model selected: bacteria

Threshold for this model: 0.4

Your Sequence:

```
MFILRNKRKESGLSLIEASLVIIISGLVIVL  
VQSAISKI IKYAVHENALHDVSYVMSYIQDT  
LQTSCYGKENIVCDGVISSVKSNNVNSGKIK  
VDVKKDGKYVSVYIILYTLGNEYFGRNEDFG  
SMLFTHVSAYSGRMDSSTGVVSSVLGWWKVN  
SDDTIFLGGDNNKNLFFYNNFRVDVL
```

Overall Prediction for the Antigen = **0.3347** (Probable **NON-ANTIGEN**).

BFPJ SECUENCIA COMPLETA

M I R K E S G L S L

I E V M I G V L F F S I V S L S V Y N Y

Y N R S A E Q H Y M N N V S V V M C R Y

A D A L S Y Y L Q D S T A T N G N G S V

S L S T L V A S G Y L P P M Q E T S G V

G E Y E I K M Y V D K N K N G L I A L Y

G G E Y Y S L T S E N I K K F L G I R G

A S A K D N K I I S A G G Y Y S L P V S

D F T G F R T V N L R G L M L I P S I Q

T S G Q N C D R G G Y D F

En gris el péptido señal de la proteína

NetCTL 1.2 server predicts CTL epitopes in protein sequences.

The screenshot shows the CBS website navigation menu with the following items: EVENTS, NEWS, RESEARCH GROUPS, CBS PREDICTION SERVERS, CBS DATA SETS, STAFF, CONTACT, ABOUT CBS, and INTERNAL. Below the menu is a breadcrumb trail: CBS >> CBS Prediction Servers >> NetCTL 1.2 >> Output format. The main content area is titled "Output Format" and contains a "Description" section. The description states: "The output consists of 16 columns:" followed by a list of 16 items: Residue number, ID, Protein identifier, pep, Peptide sequence, aff, Predicted MHC binding affinity (value given as 1 - log50k(aff), where log50k is the logarithm with base 50,000, and aff is the affinity in nM units), aff_rescale, Rescale binding affinity (predicted binding affinity normalized by the 1st percentile score), cle, C terminal cleavage affinity, tap, TAP transport efficiency, COMB, Prediction score, and <-E. Identified MHC ligands.

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

1	ID	Sequence	pep	MIRKESGLS	aff	0.0659	aff_rescale	0.2797	cle	0.0244	tap	-2.0230	COMB	0.1822
2	ID	Sequence	pep	IRKESGLSL	aff	0.0531	aff_rescale	0.2255	cle	0.9105	tap	1.2180	COMB	0.4230
3	ID	Sequence	pep	RKESGLSLI	aff	0.0659	aff_rescale	0.2799	cle	0.0749	tap	0.6930	COMB	0.3258
4	ID	Sequence	pep	KESGLSLIE	aff	0.0528	aff_rescale	0.2242	cle	0.0236	tap	-1.4660	COMB	0.1545
5	ID	Sequence	pep	ESGLSLIEV	aff	0.0818	aff_rescale	0.3471	cle	0.2704	tap	0.0890	COMB	0.3921
6	ID	Sequence	pep	SGLSLIEVM	aff	0.0571	aff_rescale	0.2422	cle	0.2691	tap	-0.0580	COMB	0.2797
7	ID	Sequence	pep	GLSLIEVMI	aff	0.0620	aff_rescale	0.2632	cle	0.5375	tap	0.4250	COMB	0.3650
8	ID	Sequence	pep	LSLIEVMIG	aff	0.0586	aff_rescale	0.2490	cle	0.0244	tap	-1.3630	COMB	0.1845
9	ID	Sequence	pep	SLIEVMIGV	aff	0.0787	aff_rescale	0.3340	cle	0.6352	tap	0.5040	COMB	0.4544
10	ID	Sequence	pep	LIEVMIGVL	aff	0.0813	aff_rescale	0.3453	cle	0.6801	tap	0.9970	COMB	0.4971
11	ID	Sequence	pep	IEVMIGVLF	aff	0.0549	aff_rescale	0.2329	cle	0.1387	tap	2.5480	COMB	0.3811
12	ID	Sequence	pep	EVMIGVLFV	aff	0.0838	aff_rescale	0.3558	cle	0.7102	tap	2.5430	COMB	0.5895
13	ID	Sequence	pep	VMIGVLFVFS	aff	0.0674	aff_rescale	0.2864	cle	0.0261	tap	-2.1600	COMB	0.1823
14	ID	Sequence	pep	MIGVLFVSI	aff	0.0692	aff_rescale	0.2940	cle	0.0989	tap	0.5940	COMB	0.3385
15	ID	Sequence	pep	IGVLFVFSIV	aff	0.0567	aff_rescale	0.2409	cle	0.5170	tap	0.1850	COMB	0.3277
16	ID	Sequence	pep	GVLFVFSIVS	aff	0.0557	aff_rescale	0.2365	cle	0.0716	tap	-2.2590	COMB	0.1343
17	ID	Sequence	pep	VLFVFSIVSL	aff	0.0594	aff_rescale	0.2521	cle	0.9597	tap	1.1180	COMB	0.4519
18	ID	Sequence	pep	LFFSIVSLVS	aff	0.0573	aff_rescale	0.2432	cle	0.0275	tap	-1.9150	COMB	0.1516
19	ID	Sequence	pep	FFSIVSLSV	aff	0.0720	aff_rescale	0.3058	cle	0.8259	tap	0.4310	COMB	0.4512
20	ID	Sequence	pep	FSIVSLSVY	aff	0.3087	aff_rescale	1.3108	cle	0.7162	tap	2.8340	COMB	1.5599 <-E
21	ID	Sequence	pep	SIVSLSVYN	aff	0.0612	aff_rescale	0.2597	cle	0.0311	tap	-1.1610	COMB	0.2063
22	ID	Sequence	pep	IVSLSVYNY	aff	0.2197	aff_rescale	0.9328	cle	0.9625	tap	3.1360	COMB	1.2340 <-E
23	ID	Sequence	pep	VSLSVYNYN	aff	0.4098	aff_rescale	1.7399	cle	0.9741	tap	3.0540	COMB	2.0387 <-E
24	ID	Sequence	pep	SLSVYNYYN	aff	0.0840	aff_rescale	0.3568	cle	0.2451	tap	-1.2000	COMB	0.3336
25	ID	Sequence	pep	LSVYNYYNR	aff	0.0584	aff_rescale	0.2481	cle	0.1645	tap	1.6610	COMB	0.3558
26	ID	Sequence	pep	SVYNYYNRS	aff	0.0729	aff_rescale	0.3095	cle	0.3308	tap	-1.8190	COMB	0.2682
27	ID	Sequence	pep	VYNYYNRSA	aff	0.0457	aff_rescale	0.1942	cle	0.0975	tap	-0.2490	COMB	0.1964
28	ID	Sequence	pep	YNYYNRSAE	aff	0.0513	aff_rescale	0.2177	cle	0.0230	tap	-1.5040	COMB	0.1459
29	ID	Sequence	pep	NYYNRSAEQ	aff	0.0504	aff_rescale	0.2142	cle	0.1472	tap	0.3310	COMB	0.2528
30	ID	Sequence	pep	YYNRSAEQH	aff	0.0584	aff_rescale	0.2481	cle	0.0542	tap	-0.2440	COMB	0.2440
31	ID	Sequence	pep	YNRSAEQHY	aff	0.0650	aff_rescale	0.2759	cle	0.2838	tap	2.7490	COMB	0.4559
32	ID	Sequence	pep	NRSAEQHYM	aff	0.0563	aff_rescale	0.2389	cle	0.4754	tap	0.4920	COMB	0.3348
33	ID	Sequence	pep	RSAEQHYMN	aff	0.0785	aff_rescale	0.3334	cle	0.0315	tap	-0.8730	COMB	0.2945
34	ID	Sequence	pep	SAEQHYMNN	aff	0.0710	aff_rescale	0.3016	cle	0.0229	tap	-1.2220	COMB	0.2439
35	ID	Sequence	pep	AEQHYMNNV	aff	0.0620	aff_rescale	0.2634	cle	0.1883	tap	0.3450	COMB	0.3089
36	ID	Sequence	pep	EQHYMNNVS	aff	0.0513	aff_rescale	0.2178	cle	0.0229	tap	-2.1900	COMB	0.1118
37	ID	Sequence	pep	QHYMNNVSV	aff	0.0545	aff_rescale	0.2316	cle	0.5907	tap	0.3550	COMB	0.3380
38	ID	Sequence	pep	HYMNNVSVV	aff	0.0617	aff_rescale	0.2619	cle	0.9043	tap	0.5040	COMB	0.4227
39	ID	Sequence	pep	YMNNVSVVM	aff	0.0986	aff_rescale	0.4187	cle	0.9739	tap	0.2890	COMB	0.5793
40	ID	Sequence	pep	MNNVSVVMC	aff	0.0589	aff_rescale	0.2503	cle	0.0257	tap	-0.2370	COMB	0.2423
41	ID	Sequence	pep	MNNVSVVMCR	aff	0.0515	aff_rescale	0.2188	cle	0.2814	tap	1.4830	COMB	0.3352
42	ID	Sequence	pep	NVSVVMCRY	aff	0.3960	aff_rescale	1.6812	cle	0.8933	tap	3.2370	COMB	1.9770 <-E
43	ID	Sequence	pep	VSVVMCRYA	aff	0.0562	aff_rescale	0.2384	cle	0.0322	tap	-0.2140	COMB	0.2325
44	ID	Sequence	pep	SVMCRYAD	aff	0.0540	aff_rescale	0.2294	cle	0.0236	tap	-1.4990	COMB	0.1580

45	ID	Sequence	pep	VVMCRYADA	aff	0.0621	aff_rescale	0.2635	cle	0.0600	tap	-0.2160	COMB	0.2617
46	ID	Sequence	pep	VMCRYADAL	aff	0.0785	aff_rescale	0.3332	cle	0.3792	tap	1.1940	COMB	0.4498
47	ID	Sequence	pep	MCRYADALS	aff	0.0563	aff_rescale	0.2391	cle	0.0248	tap	-2.0820	COMB	0.1387
48	ID	Sequence	pep	CRYADALS	aff	0.0753	aff_rescale	0.3198	cle	0.9720	tap	3.3130	COMB	0.6313
49	ID	Sequence	pep	RYADALSYL	aff	0.0809	aff_rescale	0.3102	cle	0.9720	tap	3.4130	COMB	0.7303
50	ID	Sequence	pep	YADALSYLL	aff	0.2104	aff_rescale	0.8932	cle	0.9553	tap	0.9430	COMB	1.0837 <-E
51	ID	Sequence	pep	ADALSYLLQ	aff	0.0535	aff_rescale	0.2272	cle	0.0456	tap	-0.3040	COMB	0.2188
52	ID	Sequence	pep	DALSYLLQD	aff	0.0607	aff_rescale	0.2577	cle	0.1197	tap	-1.9070	COMB	0.1803
53	ID	Sequence	pep	ALSYLLQDS	aff	0.0657	aff_rescale	0.2790	cle	0.0858	tap	-2.2250	COMB	0.1807
54	ID	Sequence	pep	LSYLLQDST	aff	0.0604	aff_rescale	0.2565	cle	0.0459	tap	-0.3200	COMB	0.2474
55	ID	Sequence	pep	SYLLQDSTA	aff	0.0526	aff_rescale	0.2232	cle	0.8789	tap	0.0170	COMB	0.3559
56	ID	Sequence	pep	YLQDSTAT	aff	0.0533	aff_rescale	0.2262	cle	0.1081	tap	-0.3890	COMB	0.2230
57	ID	Sequence	pep	YLQDSTATN	aff	0.0825	aff_rescale	0.3503	cle	0.3833	tap	-1.4000	COMB	0.3378
58	ID	Sequence	pep	LQDSTATNG	aff	0.0711	aff_rescale	0.3020	cle	0.0453	tap	-1.5440	COMB	0.2316
59	ID	Sequence	pep	QDSTATNGN	aff	0.0561	aff_rescale	0.2380	cle	0.0418	tap	-1.5660	COMB	0.1660
60	ID	Sequence	pep	DSTATNGNS	aff	0.0604	aff_rescale	0.2566	cle	0.0335	tap	-1.7560	COMB	0.1738
61	ID	Sequence	pep	STATNGNGS	aff	0.1164	aff_rescale	0.4943	cle	0.0239	tap	-2.3360	COMB	0.3811
62	ID	Sequence	pep	TATNGNGSV	aff	0.0779	aff_rescale	0.3309	cle	0.9590	tap	0.3640	COMB	0.4930
63	ID	Sequence	pep	ATNGNGSVS	aff	0.0689	aff_rescale	0.2926	cle	0.0431	tap	-2.2070	COMB	0.1887
64	ID	Sequence	pep	TNGNGSVSL	aff	0.0578	aff_rescale	0.2455	cle	0.9459	tap	0.7720	COMB	0.4260
65	ID	Sequence	pep	NGNGSVSLS	aff	0.0537	aff_rescale	0.2280	cle	0.0274	tap	-2.5590	COMB	0.1042
66	ID	Sequence	pep	GNGSVLSST	aff	0.0568	aff_rescale	0.2412	cle	0.0768	tap	-1.0590	COMB	0.1997
67	ID	Sequence	pep	NGSVLSSTL	aff	0.0552	aff_rescale	0.2344	cle	0.9398	tap	0.6910	COMB	0.4099
68	ID	Sequence	pep	GSVLSSTLV	aff	0.0868	aff_rescale	0.3685	cle	0.9212	tap	0.3950	COMB	0.5265
69	ID	Sequence	pep	VSVLSSTLV	aff	0.0838	aff_rescale	0.3556	cle	0.8843	tap	-0.3780	COMB	0.4693
70	ID	Sequence	pep	VSVLSSTLVA	aff	0.0623	aff_rescale	0.2643	cle	0.0355	tap	-2.0710	COMB	0.1661
71	ID	Sequence	pep	SVLSSTLVAS	aff	0.0690	aff_rescale	0.2991	cle	0.9303	tap	1.8080	COMB	0.2677
72	ID	Sequence	pep	LSTLVASGY	aff	0.4251	aff_rescale	1.8049	cle	0.9432	tap	2.9870	COMB	2.0957 <-E
73	ID	Sequence	pep	STLVASGYL	aff	0.0897	aff_rescale	0.3809	cle	0.7894	tap	1.0520	COMB	0.5519
74	ID	Sequence	pep	TLVASGYLP	aff	0.0635	aff_rescale	0.2695	cle	0.0399	tap	0.2350	COMB	0.2872
75	ID	Sequence	pep	LVASGYLPP	aff	0.0707	aff_rescale	0.3004	cle	0.0337	tap	0.2740	COMB	0.3191
76	ID	Sequence	pep	VASGYLPPM	aff	0.0633	aff_rescale	0.2688	cle	0.9552	tap	0.4630	COMB	0.4352
77	ID	Sequence	pep	ASGYLPPMQ	aff	0.0543	aff_rescale	0.2305	cle	0.0384	tap	-0.0370	COMB	0.2344
78	ID	Sequence	pep	SGYLPPMQE	aff	0.0468	aff_rescale	0.1987	cle	0.0434	tap	-1.4380	COMB	0.1334
79	ID	Sequence	pep	GYLPPMQET	aff	0.0452	aff_rescale	0.1918	cle	0.4871	tap	-0.5480	COMB	0.2375
80	ID	Sequence	pep	YLPPMQETS	aff	0.0773	aff_rescale	0.3283	cle	0.0770	tap	-2.2910	COMB	0.2253
81	ID	Sequence	pep	LPPMQETSG	aff	0.0496	aff_rescale	0.2105	cle	0.0715	tap	-1.7630	COMB	0.1331
82	ID	Sequence	pep	PPMQETSGV	aff	0.0511	aff_rescale	0.2170	cle	0.7951	tap	-0.3570	COMB	0.3184
83	ID	Sequence	pep	PMQETSGVG	aff	0.0529	aff_rescale	0.2246	cle	0.0485	tap	-1.9310	COMB	0.1353
84	ID	Sequence	pep	MQETSGVGE	aff	0.0635	aff_rescale	0.2696	cle	0.0227	tap	-1.7320	COMB	0.1864
85	ID	Sequence	pep	QETSGVGEY	aff	0.0962	aff_rescale	0.4082	cle	0.8919	tap	2.8020	COMB	0.6821
86	ID	Sequence	pep	ETSGVGEYI	aff	0.0690	aff_rescale	0.2928	cle	0.0231	tap	-1.8440	COMB	0.2041
87	ID	Sequence	pep	TSGVGEYIE	aff	0.0720	aff_rescale	0.3059	cle	0.3424	tap	0.2240	COMB	0.3685
88	ID	Sequence	pep	SGVGEYIEIK	aff	0.0546	aff_rescale	0.2316	cle	0.1143	tap	0.3390	COMB	0.2657

89	ID	Sequence	pep	GVGEYEIKM	aff	0.0618	aff_rescale	0.2623	cle	0.9186	tap	0.1550	COMB	0.4079
90	ID	Sequence	pep	VGEYEIKMY	aff	0.0695	aff_rescale	0.2952	cle	0.2548	tap	2.5460	COMB	0.4667
91	ID	Sequence	pep	GEYEIKMYV	aff	0.0535	aff_rescale	0.2270	cle	0.9595	tap	0.2340	COMB	0.3826
92	ID	Sequence	pep	EYEIKMYVD	aff	0.0431	aff_rescale	0.1828	cle	0.0348	tap	-1.9290	COMB	0.0916
93	ID	Sequence	pep	YEIKMYVDK	aff	0.0725	aff_rescale	0.3078	cle	0.3140	tap	0.2110	COMB	0.3655
94	ID	Sequence	pep	EIKMYVDKN	aff	0.0511	aff_rescale	0.2171	cle	0.0256	tap	-1.4450	COMB	0.1487
95	ID	Sequence	pep	IKMYVDKNK	aff	0.0509	aff_rescale	0.2160	cle	0.0707	tap	0.4350	COMB	0.2484
96	ID	Sequence	pep	KMYVDKNKN	aff	0.0542	aff_rescale	0.2303	cle	0.0308	tap	-0.8910	COMB	0.1904
97	ID	Sequence	pep	MYVDKNKNG	aff	0.0442	aff_rescale	0.1879	cle	0.0765	tap	-0.9050	COMB	0.1541
98	ID	Sequence	pep	YVDKNKNGL	aff	0.1273	aff_rescale	0.5406	cle	0.6813	tap	0.9330	COMB	0.6895
99	ID	Sequence	pep	VDKNKNGLI	aff	0.0545	aff_rescale	0.2315	cle	0.4834	tap	0.2350	COMB	0.3157
100	ID	Sequence	pep	DKNKNGLIA	aff	0.0561	aff_rescale	0.2384	cle	0.0699	tap	-0.8490	COMB	0.2064
101	ID	Sequence	pep	KNKNGLIAL	aff	0.0525	aff_rescale	0.2227	cle	0.8102	tap	0.7680	COMB	0.3827
102	ID	Sequence	pep	NKNGLIALY	aff	0.0915	aff_rescale	0.3886	cle	0.8743	tap	3.0180	COMB	0.6706
103	ID	Sequence	pep	KNGLIALYG	aff	0.0569	aff_rescale	0.2415	cle	0.0260	tap	-1.4630	COMB	0.1722
104	ID	Sequence	pep	NGLIALYGG	aff	0.0497	aff_rescale	0.2108	cle	0.0262	tap	-1.5060	COMB	0.1395
105	ID	Sequence	pep	GLTALYGGF	aff	0.0549	aff_rescale	0.2330	cle	0.0261	tap	-1.5150	COMB	0.1611
106	ID	Sequence	pep	LIALYGGGEY	aff	0.4454	aff_rescale	1.8912	cle	0.8511	tap	3.0440	COMB	2.1711 <-E
107	ID	Sequence	pep	IALLYGGEY	aff	0.1101	aff_rescale	0.4673	cle	0.9358	tap	3.0660	COMB	0.7610 <-E
108	ID	Sequence	pep	ALYGGEYYS	aff	0.0565	aff_rescale	0.2398	cle	0.2789	tap	-1.8860	COMB	0.1873
109	ID	Sequence	pep	LYGGEYYSL	aff	0.0492	aff_rescale	0.2089	cle	0.9142	tap	1.2140	COMB	0.4068
110	ID	Sequence	pep	YGGEYYSLT	aff	0.0722	aff_rescale	0.3065	cle	0.0340	tap	-1.1070	COMB	0.2563
111	ID	Sequence	pep	GGEYYSLTS	aff	0.0577	aff_rescale	0.2449	cle	0.0307	tap	-2.8870	COMB	0.1052
112	ID	Sequence	pep	GEYYSLTSE	aff	0.0442	aff_rescale	0.1878	cle	0.0431	tap	-1.7230	COMB	0.1081
113	ID	Sequence	pep	EYYSLTSEN	aff	0.0506	aff_rescale	0.2149	cle	0.0508	tap	-1.0980	COMB	0.1676
114	ID	Sequence	pep	YYSLTSENI	aff	0.0596	aff_rescale	0.2531	cle	0.5586	tap	0.6080	COMB	0.3673
115	ID	Sequence	pep	YSLTSENIK	aff	0.0911	aff_rescale	0.3866	cle	0.6237	tap	0.4510	COMB	0.5027
116	ID	Sequence	pep	SLTSENIKK	aff	0.0801	aff_rescale	0.3400	cle	0.9542	tap	0.4490	COMB	0.5055
117	ID	Sequence	pep	TSENIKKKF	aff	0.1053	aff_rescale	0.4471	cle	0.7929	tap	2.5670	COMB	0.6944
118	ID	Sequence	pep	TSENIKKFL	aff	0.0890	aff_rescale	0.3781	cle	0.3854	tap	0.8780	COMB	0.4798
119	ID	Sequence	pep	SENIKKFLG	aff	0.0522	aff_rescale	0.2218	cle	0.0300	tap	-1.5620	COMB	0.1482
120	ID	Sequence	pep	ENIKKFLGI	aff	0.0568	aff_rescale	0.2413	cle	0.5458	tap	0.3010	COMB	0.3383
121	ID	Sequence	pep	NIKKFLGIR	aff	0.0584	aff_rescale	0.2478	cle	0.2683	tap	1.4810	COMB	0.3621
122	ID	Sequence	pep	IKKFLGIRG	aff	0.0488	aff_rescale	0.2072	cle	0.0694	tap	-1.1840	COMB	0.1584
123	ID	Sequence	pep	KKFLGIRGA	aff	0.0503	aff_rescale	0.2138	cle	0.5557	tap	-0.2580	COMB	0.2842
124	ID	Sequence	pep	KFLGIRGAS	aff	0.0520	aff_rescale	0.2207	cle	0.0304	tap	-1.8520	COMB	0.1326
125	ID	Sequence	pep	FLGIRGASA	aff	0.0631	aff_rescale	0.2678	cle	0.2309	tap	-0.6730	COMB	0.2688
126	ID	Sequence	pep	LGRGASAKA	aff	0.0583	aff_rescale	0.2474	cle	0.3080	tap	0.1660	COMB	0.3019
127	ID	Sequence	pep	GIRGASAKD	aff	0.0496	aff_rescale	0.2108	cle	0.0407	tap	-1.7690	COMB	0.1284
128	ID	Sequence	pep	IRGASAKDN	aff	0.0514	aff_rescale	0.2181	cle	0.0254	tap	-1.2790	COMB	0.1579
129	ID	Sequence	pep	RGASAKDNK	aff	0.0570	aff_rescale	0.2422	cle	0.0513	tap	0.4680	COMB	0.2733
130	ID	Sequence	pep	GASAKDNKI	aff	0.0629	aff_rescale	0.2672	cle	0.4257	tap	0.5360	COMB	0.3579

131	ID	Sequence	pep	ASAKDNKII	aff	0.0690	aff_rescale	0.2930	cle	0.0975	tap	0.7060	COMB	0.3429
132	ID	Sequence	pep	SAKDNKIIS	aff	0.0533	aff_rescale	0.2265	cle	0.0641	tap	-1.9880	COMB	0.1367
133	ID	Sequence	pep	AKDNKIISA	aff	0.0646	aff_rescale	0.2741	cle	0.5642	tap	-0.4640	COMB	0.3356
134	ID	Sequence	pep	KDNKIISAG	aff	0.0549	aff_rescale	0.2331	cle	0.0427	tap	-1.5360	COMB	0.1627
135	ID	Sequence	pep	DNKIISAGG	aff	0.0450	aff_rescale	0.1911	cle	0.0405	tap	-1.7650	COMB	0.1090
136	ID	Sequence	pep	KNKIISAGG	aff	0.0607	aff_rescale	0.2831	cle	0.4739	tap	2.3410	COMB	0.5022
137	ID	Sequence	pep	KIISAGGYY	aff	0.1158	aff_rescale	0.4919	cle	0.8549	tap	3.3700	COMB	0.7886 <-E
138	ID	Sequence	pep	IISAGGYY	aff	0.0622	aff_rescale	0.2842	cle	0.0460	tap	-1.9970	COMB	0.1713
139	ID	Sequence	pep	ISAGGYYSL	aff	0.0712	aff_rescale	0.3025	cle	0.9662	tap	1.0440	COMB	0.4996
140	ID	Sequence	pep	SAGGYYSLP	aff	0.0733	aff_rescale	0.3114	cle	0.0394	tap	0.0560	COMB	0.3201
141	ID	Sequence	pep	AGGYYSLPV	aff	0.0625	aff_rescale	0.2655	cle	0.1554	tap	0.0830	COMB	0.2929
142	ID	Sequence	pep	GGYYSLPVS	aff	0.0455	aff_rescale	0.1931	cle	0.0319	tap	-2.3500	COMB	0.0803
143	ID	Sequence	pep	GYSLPVSDF	aff	0.0462	aff_rescale	0.1963	cle	0.0726	tap	-1.5730	COMB	0.1286
144	ID	Sequence	pep	YSLPVSDF	aff	0.0643	aff_rescale	0.2732	cle	0.7502	tap	2.7450	COMB	0.5229
145	ID	Sequence	pep	YSLPVSDFG	aff	0.0874	aff_rescale	0.3712	cle	0.0290	tap	-0.7190	COMB	0.3396
146	ID	Sequence	pep	SLPVSDFTG	aff	0.0638	aff_rescale	0.2709	cle	0.1093	tap	-1.3900	COMB	0.2178
147	ID	Sequence	pep	LPVSDFTFG	aff	0.0482	aff_rescale	0.2045	cle	0.9568	tap	2.3660	COMB	0.4663
148	ID	Sequence	pep	PVSDFTFGR	aff	0.0632	aff_rescale	0.2683	cle	0.3994	tap	1.1870	COMB	0.3875
149	ID	Sequence	pep	VSDFTGFR	aff	0.1576	aff_rescale	0.6692	cle	0.0333	tap	-0.9240	COMB	0.6280
150	ID	Sequence	pep	SDFTFGRV	aff	0.0576	aff_rescale	0.2446	cle	0.9519	tap	0.2190	COMB	0.3984
151	ID	Sequence	pep	DFTGFRVNL	aff	0.0478	aff_rescale	0.2028	cle	0.0995	tap	-1.4460	COMB	0.1454
152	ID	Sequence	pep	FTGFRVNL	aff	0.0778	aff_rescale	0.3302	cle	0.8934	tap	0.5360	COMB	0.4910
153	ID	Sequence	pep	TGFRVNLRL	aff	0.0565	aff_rescale	0.2400	cle	0.4284	tap	1.2040	COMB	0.3645
154	ID	Sequence	pep	GFRVNLRLG	aff	0.0555	aff_rescale	0.2357	cle	0.0344	tap	-1.3480	COMB	0.1734
155	ID	Sequence	pep	RVNLRLRGL	aff	0.0522	aff_rescale	0.2215	cle	0.5455	tap	1.0000	COMB	0.3533
156	ID	Sequence	pep	FTVNLRLGL	aff	0.0821	aff_rescale	0.3485	cle	0.4609	tap	0.5000	COMB	0.4427
157	ID	Sequence	pep	VNLRLRGLML	aff	0.0857	aff_rescale	0.3640	cle	0.6604	tap	1.1100	COMB	0.5186
158	ID	Sequence	pep	VNLRLRGLMLI	aff	0.0582	aff_rescale	0.2473	cle	0.2845	tap	0.4720	COMB	0.3136
159	ID	Sequence	pep	NLRGLMLIP	aff	0.0609	aff_rescale	0.2588	cle	0.0398	tap	0.2570	COMB	0.2776
160	ID	Sequence	pep	LRGLMLIPS	aff	0.0561	aff_rescale	0.2381	cle	0.0289	tap	-2.0570	COMB	0.1396
161	ID	Sequence	pep	RGLMLIPSI	aff	0.0531	aff_rescale	0.2254	cle	0.6042	tap	0.5780	COMB	0.3449
162	ID	Sequence	pep	GLMLIPSIQ	aff	0.0551	aff_rescale	0.2338	cle	0.1735	tap	-0.0680	COMB	0.2565
163	ID	Sequence	pep	LMLIPSIQT	aff	0.0577	aff_rescale	0.2452	cle	0.3010	tap	-0.5920	COMB	0.2607
164	ID	Sequence	pep	LIPSIQTS	aff	0.0689	aff_rescale	0.2924	cle	0.4898	tap	-2.0620	COMB	0.2627
165	ID	Sequence	pep	LIPSIQTSG	aff	0.0599	aff_rescale	0.2542	cle	0.7378	tap	-1.2420	COMB	0.3027
166	ID	Sequence	pep	IPSIQTSGQ	aff	0.0446	aff_rescale	0.1892	cle	0.7453	tap	-0.3430	COMB	0.2839
167	ID	Sequence	pep	PSIQTSGQN	aff	0.0546	aff_rescale	0.2318	cle	0.0343	tap	-1.6770	COMB	0.1531
168	ID	Sequence	pep	SIQTSGQNC	aff	0.0639	aff_rescale	0.2712	cle	0.0279	tap	0.0190	COMB	0.2763
169	ID	Sequence	pep	IQTSGQNC	aff	0.0490	aff_rescale	0.2082	cle	0.0770	tap	-1.6990	COMB	0.1348
170	ID	Sequence	pep	QTSGQNC	aff	0.0847	aff_rescale	0.3597	cle	0.3070	tap	1.4810	COMB	0.4798
171	ID	Sequence	pep	TSGQNC	aff	0.0601	aff_rescale	0.2551	cle	0.0416	tap	-1.5880	COMB	0.1819
172	ID	Sequence	pep	SGQNC	aff	0.0533	aff_rescale	0.2264	cle	0.0237	tap	-1.6310	COMB	0.1484
173	ID	Sequence	pep	GQNC	aff	0.0700	aff_rescale	0.2971	cle	0.7942	tap	2.7810	COMB	0.5553

174 ID Sequence pep QNCDRGGYD aff 0.0558 aff_rescale 0.2367 cle 0.0359 tap -1.9760 COMB 0.1433
 175 ID Sequence pep NCDRGGYDF aff 0.0724 aff_rescale 0.3075 cle 0.6381 tap 2.3670 COMB 0.5216

 Number of MHC ligands 9 identified. Number of peptides 175. Protein name Sequence

NUMERO DE RESIDUO	SECUENCIA
20	F S I V S L S V Y
22	I V S L S V Y N Y
23	V S L S V Y N Y Y
42	N V S V V M C R Y
50	Y A D A L S Y Y L
72	L S T L V A S G Y
106	L I A L Y G G E Y
107	I A L Y G G E Y Y
137	K I I S A G G Y Y

Secuencia de BfpJ con los epitopes según el programa NetCTL 1.2 en recuadros de colores

M I R K E S G L S L
I E V M I G V L F F S I V S L S V Y N Y
Y N R S A E Q H Y M N N V S V V M C R Y
A D A L S Y Y L Q D S T A T N G N G S V
S L S T L V A S G Y L P P M Q E T S G V
G E Y E I K M Y V D K N K N G L I A L Y
G G E Y Y S L T S E N I K K F L G I R G
A S A K D N K I I S A G G Y Y S L P V S
D F T G F R T V N L R G L M L I P S I Q
T S G Q N C D R G G Y D F

En gris el péptido señal de la proteína

NetMHCII 2.2 server predicts binding of peptides to HLA-DR, HLA-DQ, HLA-DP and mouse MHC class II alleles using artificial neuron networks

DESCRIPTION

In the header of the out is indicated which prediction method and HLA allele were selected, as well as the corresponding two threshold values defining high binding peptides, and weak binding peptides. High binding peptides have an IC50 value below 50 nM, and weak binding peptides an IC50 values below 500 nM. For Artificial Neural network (ANN) prediction the output consists of 6 columns. For weight matrix predictions the output consists of 5 columns.

- Residue number.
- Peptide sequence (9mer)
- Prediction score (called 1-log50k(aff) for ANN predictions)
- Affinity as IC50 value in nM (only for ANN predictions)
- Bind Level (WB for weak binder, SB for strong binder)
- Sequence name

The predictions for each protein are summarized with a line stating the number of high and weak binding peptides identified.

it is in FSA format

side length 15

NetMHCII version 2.2.

; binder threshold 50.00. Weak binder threshold 500.00.

Allele	pos	peptide	core	1-log50k(aff)	affinity(nM)	%Rank	Identity	Bind Level
HLA-DRB10101	1	MIRKESGLSLIEVMI	IRKESGLSL	0.7438	16.0	16.00	Sequence	SB
HLA-DRB10101	2	IRKESGLSLIEVMIG	IRKESGLSL	0.6768	33.0	32.00	Sequence	SB
HLA-DRB10101	3	RKESGLSLIEVMIGV	GLSLIEVMI	0.5007	221.9	50.00	Sequence	WB
HLA-DRB10101	4	KESGLSLIEVMIGVL	GLSLIEVMI	0.5188	182.4	50.00	Sequence	WB
HLA-DRB10101	5	ESGLSLIEVMIGVLF	IEVMIGVLF	0.6173	62.9	32.00	Sequence	WB
HLA-DRB10101	6	ESGLSLIEVMIGVLF	IEVMIGVLF	0.6362	51.9	32.00	Sequence	WB
HLA-DRB10101	7	GLSLIEVMIGVLFSS	IEVMIGVLF	0.6527	42.9	32.00	Sequence	SB
HLA-DRB10101	8	LSLIEVMIGVLFSS	IEVMIGVLF	0.6667	36.8	32.00	Sequence	SB
HLA-DRB10101	9	SLIEVMIGVLFSSIV	IEVMIGVLF	0.6165	63.4	32.00	Sequence	WB
HLA-DRB10101	10	LIEVMIGVLFSSIVS	IEVMIGVLF	0.5778	96.3	32.00	Sequence	WB
HLA-DRB10101	11	IEVMIGVLFSSIVSL	IEVMIGVLF	0.5493	131.1	50.00	Sequence	WB
HLA-DRB10101	12	EVIGVLFSSIVSLS	IGVLFSSIV	0.5327	157.0	50.00	Sequence	WB
HLA-DRB10101	13	VMIGVLFSSIVSLSV	FFSIVSLSV	0.8080	8.0	8.00	Sequence	SB
HLA-DRB10101	14	MIGVLFSSIVSLSVY	FFSIVSLSV	0.8256	6.6	4.00	Sequence	SB
HLA-DRB10101	15	IGVLFSSIVSLSVYV	FFSIVSLSV	0.8338	6.0	4.00	Sequence	SB
HLA-DRB10101	16	GVLFSSIVSLSVYNY	FFSIVSLSV	0.8400	5.2	1.50	Sequence	SB
HLA-DRB10101	17	VLFSSIVSLSVYNY	FFSIVSLSV	0.8386	5.7	2.00	Sequence	SB
HLA-DRB10101	18	LFFSIVSLSVYNYN	FFSIVSLSV	0.8180	7.2	4.00	Sequence	SB
HLA-DRB10101	19	FFSIVSLSVYNYNR	IVSLSVYNY	0.7946	9.2	8.00	Sequence	SB
HLA-DRB10101	20	FSIVSLSVYNYNRS	IVSLSVYNY	0.7073	23.7	16.00	Sequence	SB
HLA-DRB10101	21	SIVSLSVYNYNRS	IVSLSVYNY	0.6260	57.2	32.00	Sequence	WB
HLA-DRB10101	22	IVSLSVYNYNRS	IVSLSVYNY	0.5733	101.2	32.00	Sequence	WB
HLA-DRB10101	23	VLSVYNYNRS	SLSVYNYN	0.4568	356.8	50.00	Sequence	WB
HLA-DRB10101	24	SLSVYNYNRS	VYNYNRS	0.4949	236.3	50.00	Sequence	WB
HLA-DRB10101	25	LSVYNYNRS	VYNYNRS	0.4420	418.9	50.00	Sequence	WB
HLA-DRB10101	26	SVYNYNRS	NYNRS	0.5076	205.9	50.00	Sequence	WB
HLA-DRB10101	27	VYNYNRS	YNYNRS	0.4966	232.0	50.00	Sequence	WB
HLA-DRB10101	28	YNYNRS	NYNRS	0.4367	443.4	50.00	Sequence	WB
HLA-DRB10101	29	NYNRS	NYNRS	0.3995	663.5	50.00	Sequence	WB
HLA-DRB10101	30	YNYNRS	NYNRS	0.3528	1099.2	50.00	Sequence	WB
HLA-DRB10101	31	NYNRS	NYNRS	0.3696	916.5	50.00	Sequence	WB
HLA-DRB10101	32	NRS	QHYMNVSV	0.4233	512.6	50.00	Sequence	WB
HLA-DRB10101	33	RS	QHYMNVSV	0.6389	54.2	32.00	Sequence	WB
HLA-DRB10101	34	SAEQHYMNVSVVM	YMNVSVM	0.6561	41.3	32.00	Sequence	SB

HLA-DRB10101	35	AEQHYMNNVSVVMCR	YMNNSVVM	0.6990	26.0	16.00	Sequence	SB
HLA-DRB10101	36	EQHYMNNVSVVMCRY	YMNNSVVM	0.7103	23.0	16.00	Sequence	SB
HLA-DRB10101	37	QHYMNNVSVVMCRYA	YMNNSVVM	0.6720	34.8	32.00	Sequence	SB
HLA-DRB10101	38	HVMNNVSVVMCRYAD	YMNNSVVM	0.6121	66.5	32.00	Sequence	WB
HLA-DRB10101	39	YMNNSVVMCRYADA	YMNNSVVM	0.5701	104.7	50.00	Sequence	WB
HLA-DRB10101	40	MNNVSVVMCRYADAL	VVMCRYADA	0.4691	312.4	50.00	Sequence	WB
HLA-DRB10101	41	NNVSVVMCRYADALS	VVMCRYADA	0.5004	222.7	50.00	Sequence	WB
HLA-DRB10101	42	NVSVVMCRYADALS	VVMCRYADA	0.6241	58.4	32.00	Sequence	WB
HLA-DRB10101	43	VSVVMCRYADALSYY	CRYADALS	0.6947	27.2	16.00	Sequence	SB
HLA-DRB10101	44	SVMCRYADALSYYL	CRYADALS	0.7758	11.3	8.00	Sequence	SB
HLA-DRB10101	45	VVMCRYADALSYYLQ	CRYADALS	0.8038	8.4	8.00	Sequence	SB
HLA-DRB10101	46	VVMCRYADALSYYLQD	CRYADALS	0.7643	12.8	8.00	Sequence	SB
HLA-DRB10101	47	MCRYADALSYYLQDST	CRYADALS	0.7244	19.7	16.00	Sequence	SB
HLA-DRB10101	48	CRYADALSYYLQDSTA	YADALSYYL	0.6460	46.1	32.00	Sequence	SB
HLA-DRB10101	49	RYADALSYYLQDSTA	YADALSYYL	0.5754	98.9	32.00	Sequence	WB
HLA-DRB10101	50	YADALSYYLQDSTAT	YYLQDSTAT	0.7950	9.2	8.00	Sequence	SB
HLA-DRB10101	51	ADALSYYLQDSTATN	YYLQDSTAT	0.8207	7.0	4.00	Sequence	SB
HLA-DRB10101	52	DALSYYLQDSTATNG	YYLQDSTAT	0.8378	5.8	4.00	Sequence	SB
HLA-DRB10101	53	ALSYYLQDSTATNGN	YYLQDSTAT	0.8526	4.9	1.25	Sequence	SB
HLA-DRB10101	54	LSYYLQDSTATNGNG	YYLQDSTAT	0.8355	5.9	4.00	Sequence	SB
HLA-DRB10101	55	SYLQDSTATNGNGS	YYLQDSTAT	0.7907	9.6	8.00	Sequence	SB
HLA-DRB10101	56	YYLQDSTATNGNGSV	YYLQDSTAT	0.7546	14.2	16.00	Sequence	SB
HLA-DRB10101	57	YLQDSTATNGNGSVS	YLQDSTATN	0.3378	1292.5	50.00	Sequence	WB
HLA-DRB10101	58	LQDSTATNGNGSVSL	STATNGNGS	0.3076	1792.8	50.00	Sequence	WB
HLA-DRB10101	59	QDSTATNGNGSVSLS	STATNGNGS	0.3053	1839.1	50.00	Sequence	WB
HLA-DRB10101	60	DSTATNGNGSVLSL	TNGNGSVSL	0.3352	1329.5	50.00	Sequence	WB
HLA-DRB10101	61	STATNGNGSVLSLTL	TNGNGSVSL	0.4614	339.3	50.00	Sequence	WB
HLA-DRB10101	62	TATNGNGSVLSLTLV	TNGNGSVSL	0.3919	720.5	50.00	Sequence	WB
HLA-DRB10101	63	ATNGNGSVLSLTLVA	SVLSLTLVA	0.5481	132.9	50.00	Sequence	WB
HLA-DRB10101	64	TNGNGSVLSLTLVAS	SVLSLTLVA	0.5829	91.2	32.00	Sequence	WB
HLA-DRB10101	65	NGNGSVLSLTLVASS	SVLSLTLVA	0.6062	70.9	32.00	Sequence	WB
HLA-DRB10101	66	GNGSVLSLTLVASSY	SVLSLTLVA	0.6544	42.0	32.00	Sequence	SB
HLA-DRB10101	67	NGSVLSLTLVASSYL	LSTLVASSY	0.7007	25.5	16.00	Sequence	SB
HLA-DRB10101	68	GSVLSLTLVASSYLP	LSTLVASSY	0.6797	32.0	32.00	Sequence	SB
HLA-DRB10101	69	SVLSLTLVASSYLP	LSTLVASSY	0.6574	40.7	32.00	Sequence	SB
HLA-DRB10101	70	VLSLTLVASSYLP	LSTLVASSY	0.6026	73.7	32.00	Sequence	WB
HLA-DRB10101	71	SLSLTLVASSYLP	TLVASSYLP	0.5758	98.5	32.00	Sequence	WB
HLA-DRB10101	72	LSTLVASSYLP	TLVASSYLP	0.5317	158.7	50.00	Sequence	WB
HLA-DRB10101	73	STLVASSYLP	TLVASSYLP	0.5125	195.3	50.00	Sequence	WB
HLA-DRB10101	74	TLVASSYLP	TLVASSYLP	0.4670	319.4	50.00	Sequence	WB
HLA-DRB10101	75	LVASSYLP	SGYLP	0.4462	400.0	50.00	Sequence	WB
HLA-DRB10101	76	VASSYLP	ASGYLP	0.3877	753.4	50.00	Sequence	WB
HLA-DRB10101	77	ASSYLP	ASGYLP	0.3668	944.5	50.00	Sequence	WB

HLA-DRB10101	78	SGYLP	YLP	0.2960	2033.1	50.00	Sequence	WB
HLA-DRB10101	79	GYLP	PMQETS	0.2568	3105.7	50.00	Sequence	WB
HLA-DRB10101	80	YLP	PMQETSGVGE	0.2290	4195.3	50.00	Sequence	WB
HLA-DRB10101	81	LPPMQETSGVGEY	PMQETSGV	0.1776	7315.7	50.00	Sequence	WB
HLA-DRB10101	82	PPMQETSGVGEY	PMQETSGV	0.1411	10859.6	50.00	Sequence	WB
HLA-DRB10101	83	PMQETSGVGEY	TSGVGEY	0.1675	8163.1	50.00	Sequence	WB
HLA-DRB10101	84	MQETSGVGEY	TSGVGEY	0.2327	4033.3	50.00	Sequence	WB
HLA-DRB10101	85	QETSGVGEY	GVGEY	0.2445	3548.5	50.00	Sequence	WB
HLA-DRB10101	86	ETSGVGEY	EY	0.2658	2817.3	50.00	Sequence	WB
HLA-DRB10101	87	TSGVGEY	Y	0.3508	1123.6	50.00	Sequence	WB
HLA-DRB10101	88	SGVGEY	Y	0.3760	855.0	50.00	Sequence	WB
HLA-DRB10101	89	GVGEY	Y	0.4074	608.8	50.00	Sequence	WB
HLA-DRB10101	90	VGEY	Y	0.4166	551.3	50.00	Sequence	WB
HLA-DRB10101	91	GEY	Y	0.3937	706.5	50.00	Sequence	WB
HLA-DRB10101	92	EY	Y	0.3788	830.2	50.00	Sequence	WB
HLA-DRB10101	93	Y	Y	0.4060	618.2	50.00	Sequence	WB
HLA-DRB10101	94	EIKMYVDK	VDK	0.4781	283.4	50.00	Sequence	WB
HLA-DRB10101	95	IKMYVDK	VDK	0.6140	65.1	32.00	Sequence	WB
HLA-DRB10101	96	KMYVDK	VDK	0.7076	23.7	16.00	Sequence	WB
HLA-DRB10101	97	MYVDK	VDK	0.6184	62.1	32.00	Sequence	WB
HLA-DRB10101	98	YVDK	VDK	0.5239	172.6	50.00	Sequence	WB
HLA-DRB10101	99	VDK	Y	0.4864	259.1	50.00	Sequence	WB
HLA-DRB10101	100	DKN	Y	0.5541	124.6	50.00	Sequence	WB
HLA-DRB10101	101	KN	Y	0.5986	77.0	32.00	Sequence	WB
HLA-DRB10101	102	N	Y	0.6174	62.8	32.00	Sequence	WB
HLA-DRB10101	103	KN	Y	0.7134	22.2	16.00	Sequence	WB
HLA-DRB10101	104	N	Y	0.6515	43.4	32.00	Sequence	WB
HLA-DRB10101	105	GL	Y	0.6186	62.0	32.00	Sequence	WB
HLA-DRB10101	106	L	Y	0.5648	110.9	50.00	Sequence	WB
HLA-DRB10101	107	I	Y	0.4828	269.4	50.00	Sequence	WB
HLA-DRB10101	108	ALY	Y	0.7191	20.9	16.00	Sequence	WB
HLA-DRB10101	109	LY	Y	0.7774	11.1	8.00	Sequence	WB
HLA-DRB10101	110	Y	Y	0.8192	7.1	4.00	Sequence	WB
HLA-DRB10101	111	G	Y	0.8546	4.8	1.10	Sequence	WB
HLA-DRB10101	112	E	Y	0.8448	5.4	2.00	Sequence	WB
HLA-DRB10101	113	E	Y	0.8312	6.2	4.00	Sequence	WB
HLA-DRB10101	114	Y	Y	0.8188	7.1	4.00	Sequence	WB
HLA-DRB10101	115	Y	Y	0.3400	1263.4	50.00	Sequence	WB
HLA-DRB10101	116	SL	Y	0.4767	287.7	50.00	Sequence	WB
HLA-DRB10101	117	LT	Y	0.5393	146.1	50.00	Sequence	WB
HLA-DRB10101	118	T	Y	0.5827	91.3	32.00	Sequence	WB
HLA-DRB10101	119	SEN	Y	0.8665	4.2	0.50	Sequence	WB
HLA-DRB10101	120	ENT	Y	0.8765	3.8	0.20	Sequence	WB

HLA-DRB10101	121	NIKKFLGIRGASAKD	FLGIRGASA	0.8773	3.8	0.20	Sequence	SB
HLA-DRB10101	122	IKKFLGIRGASAKDN	FLGIRGASA	0.8817	3.6	0.10	Sequence	SB
HLA-DRB10101	123	KKFLGIRGASAKDNK	FLGIRGASA	0.8751	3.9	0.30	Sequence	SB
HLA-DRB10101	124	KFLGIRGASAKDNKI	FLGIRGASA	0.8657	4.3	0.60	Sequence	SB
HLA-DRB10101	125	FLGIRGASAKDNKIT	FLGIRGASA	0.8502	5.1	1.25	Sequence	SB
HLA-DRB10101	126	LGIRGASAKDNKIIS	IRGASAKDN	0.6177	62.6	32.00	Sequence	WB
HLA-DRB10101	127	GIRGASAKDNKIISA	IRGASAKDN	0.4972	230.4	50.00	Sequence	WB
HLA-DRB10101	128	IRGASAKDNKIISAG	IRGASAKDN	0.4298	478.2	50.00	Sequence	WB
HLA-DRB10101	129	RGASAKDNKIISAGG	GASAKDNKI	0.2301	4145.1	50.00	Sequence	WB
HLA-DRB10101	130	GASAKDNKIISAGGY	NKIISAGGY	0.3262	1466.3	50.00	Sequence	WB
HLA-DRB10101	131	ASAKDNKIISAGGY	NKIISAGGY	0.4095	595.0	50.00	Sequence	WB
HLA-DRB10101	132	ASAKDNKIISAGGYG	NKIISAGGY	0.5450	100.0	50.00	Sequence	WB
HLA-DRB10101	133	AKDNKIISAGGYYSL	IISAGGYYS	0.6386	49.9	32.00	Sequence	SB
HLA-DRB10101	134	KDNKIISAGGYYSLP	IISAGGYYS	0.6012	74.8	32.00	Sequence	SB
HLA-DRB10101	135	DNKIISAGGYYSLPV	IISAGGYYS	0.6813	31.5	32.00	Sequence	SB
HLA-DRB10101	136	NKIISAGGYYSLPVS	IISAGGYYS	0.6685	36.1	32.00	Sequence	SB
HLA-DRB10101	137	IISAGGYYSLPVSD	IISAGGYYS	0.6106	67.6	32.00	Sequence	SB
HLA-DRB10101	138	IISAGGYYSLPVSDF	YISLPVSDF	0.7629	13.0	16.00	Sequence	SB
HLA-DRB10101	139	ISAGGYYSLPVSDFT	YISLPVSDF	0.7995	8.8	8.00	Sequence	SB
HLA-DRB10101	140	SAGGYYSLPVSDFTG	YISLPVSDF	0.8176	7.2	4.00	Sequence	SB
HLA-DRB10101	141	AGGYYSLPVSDFTGF	YISLPVSDF	0.8417	5.5	2.00	Sequence	SB
HLA-DRB10101	142	GGYYSLPVSDFTGFR	YISLPVSDF	0.8219	6.9	4.00	Sequence	SB
HLA-DRB10101	143	GYYSLPVSDFTGFRT	YISLPVSDF	0.7956	9.1	8.00	Sequence	SB
HLA-DRB10101	144	YYSLPVSDFTGFRTV	YISLPVSDF	0.7591	13.6	16.00	Sequence	SB
HLA-DRB10101	145	YSLPVSDFTGFRTVN	VSDFTGFRT	0.3840	784.7	50.00	Sequence	WB
HLA-DRB10101	146	SLPVSDFTGFRTVNL	FTGFRTVNL	0.5618	114.6	50.00	Sequence	WB
HLA-DRB10101	147	LPVSDFTGFRTVNLR	FTGFRTVNL	0.6417	48.2	32.00	Sequence	SB
HLA-DRB10101	148	PVSDFTGFRTVNLRL	FTGFRTVNL	0.6768	33.0	32.00	Sequence	SB
HLA-DRB10101	149	VSDFTGFRTVNLRLG	FTGFRTVNL	0.7681	12.3	8.00	Sequence	SB
HLA-DRB10101	150	SDFTFRTVNLRLGLM	FRTVNLRLG	0.7786	11.0	8.00	Sequence	SB
HLA-DRB10101	151	DFTGFRTVNLRLGLM	FRTVNLRLG	0.8139	7.5	4.00	Sequence	SB
HLA-DRB10101	152	FTGFRTVNLRLGLML	FRTVNLRLG	0.8313	6.2	4.00	Sequence	SB
HLA-DRB10101	153	TGFRTVNLRLGLMLIP	FRTVNLRLG	0.8029	8.4	8.00	Sequence	SB
HLA-DRB10101	154	GFRTVNLRLGLMLIPS	FRTVNLRLG	0.7884	9.9	8.00	Sequence	SB
HLA-DRB10101	155	FRTVNLRLGLMLIPSI	FRTVNLRLG	0.7550	14.2	16.00	Sequence	SB
HLA-DRB10101	156	RTVNLRLGLMLIPSIQ	LRGLMLIPS	0.6738	34.1	32.00	Sequence	SB
HLA-DRB10101	157	TVNLRLGLMLIPSIQT	LMLIPSIQT	0.7454	15.7	16.00	Sequence	SB
HLA-DRB10101	158	VNLRLGLMLIPSIQTS	LMLIPSIQT	0.7539	14.3	16.00	Sequence	SB
HLA-DRB10101	159	NLRLGLMLIPSIQTS	LMLIPSIQT	0.7667	12.5	8.00	Sequence	SB
HLA-DRB10101	160	LRGLMLIPSIQTS	LMLIPSIQT	0.7792	10.9	8.00	Sequence	SB
HLA-DRB10101	161	RGLMLIPSIQTS	LMLIPSIQT	0.7360	17.4	16.00	Sequence	SB
HLA-DRB10101	162	GLMLIPSIQTS	LMLIPSIQT	0.6699	35.6	32.00	Sequence	SB
HLA-DRB10101	163	LMLIPSIQTS	LMLIPSIQT	0.6173	62.8	32.00	Sequence	WB
HLA-DRB10101	164	MLIPSIQTS	LMLIPSIQT	0.3135	1681.4	50.00	Sequence	WB

HLA-DRB10101	165	LIPSIQTS	SIQTS	0.2279	4247.1	50.00	Sequence	WB
HLA-DRB10101	166	IPSIQTS	SIQTS	0.1926	6220.1	50.00	Sequence	WB
HLA-DRB10101	167	PSIQTS	SIQTS	0.1713	7834.4	50.00	Sequence	WB
HLA-DRB10101	168	SIQTS	SIQTS	0.1790	7210.5	50.00	Sequence	WB
HLA-DRB10101	169	IQTSG	QNC	0.1472	10165.3	50.00	Sequence	WB

Allele: HLA-DRB10101. Number of high binders 76. Number of weak binders 55. Number of peptides 169

VaxiJen RESULTS

Model selected: **bacteria**

Threshold for this model: **0.4**

Your Sequence:

MIRKESGLSLIEVMIGVLFVSIVSLSVYNY
 NRSAEQHYMNNVSVVMCRYADALSYYLQDST
 ATNGNGSVSLSLTVASGYLPPMQETSGVGEY
 EIKMYVDKNKNGLIALYGGGEYSLTSENIAK
 FLGIRGASAKDNKIISAGGYYSLPVSDFTGF
 RTVNLRLGLMLIPSIQTSQNCDRGGYDF

Overall Prediction for the Antigen = **0.5998** (Probable **ANTIGEN**).

BFPK SECUENCIA COMPLETA

M I F E G R Q Q K G

L S L I E I A M V I I L S V V F I S G G

S Y I I S L F K V F A Q S C V V A T D M

N S V R N I I I S S D V E Y K T G K Y D

V S S I T N D N R F N K Y Q V Y F G N D

G Y T L I S G S S P V S V D V F F I T C

Y L A K Y N C V W I L N T N N K I Y A K

H M S A V L P L E L L S A Y G L N Y S G

E S G V T G L L F K R Y

En gris el péptido señal de la proteína

NetCTL 1.2 server predicts CTL epitopes in protein sequences.

The screenshot shows the NetCTL 1.2 server interface. At the top, there is a navigation bar with several colored buttons: EVENTS (yellow), NEWS (green), RESEARCH GROUPS (orange), CBS PREDICTION SERVERS (pink), and CBS DATA SETS (purple). Below the navigation bar, there are more buttons: STAFF (green), CONTACT (green), ABOUT CBS (orange), INTERNAL (red), and CBS BIOIN TOOLS (purple). The main content area is titled 'Output Format' and contains a 'Description' section. The description states: 'The output consists of 16 columns:' followed by a list of 16 items: Residue number, ID, Protein identifier, pep, Peptide sequence, aff, Predicted MHC binding affinity (with a detailed explanation of the formula), aff_rescale, Rescale binding affinity (with a detailed explanation of the formula), cle, C terminal cleavage affinity, tap, TAP transport efficiency, COMB, Prediction score, and <-E. Identified MHC ligands.

NetCTL-1.2 predictions using MHC supertype A1. Threshold 0.750000

1	ID	Sequence	pep	MIFEGRQQK	aff	0.0610	aff_rescale	0.2591	c1e	0.9541	tap	0.7500	COMB	0.4398
2	ID	Sequence	pep	IFEGRQQKG	aff	0.0492	aff_rescale	0.2090	c1e	0.0337	tap	-1.3130	COMB	0.1484
3	ID	Sequence	pep	FEGRQQKGL	aff	0.0547	aff_rescale	0.2323	c1e	0.4059	tap	0.5700	COMB	0.3217
4	ID	Sequence	pep	EGRQQKGLS	aff	0.0480	aff_rescale	0.2038	c1e	0.0230	tap	-2.7720	COMB	0.0686
5	ID	Sequence	pep	GRQKGLSL	aff	0.0542	aff_rescale	0.2303	c1e	0.4158	tap	0.7830	COMB	0.3318
6	ID	Sequence	pep	RQKGLSLI	aff	0.0663	aff_rescale	0.2814	c1e	0.3156	tap	0.7920	COMB	0.3683
7	ID	Sequence	pep	QKGLSLIE	aff	0.0516	aff_rescale	0.2189	c1e	0.0228	tap	-1.3470	COMB	0.1550
8	ID	Sequence	pep	QKGLSLIEI	aff	0.0634	aff_rescale	0.2692	c1e	0.1275	tap	0.3850	COMB	0.3076
9	ID	Sequence	pep	KGLSLIEIA	aff	0.0553	aff_rescale	0.2347	c1e	0.0359	tap	-0.7260	COMB	0.2038
10	ID	Sequence	pep	GLSLIEIAM	aff	0.0615	aff_rescale	0.2612	c1e	0.2102	tap	0.2420	COMB	0.3049
11	ID	Sequence	pep	LSLIEIAMV	aff	0.0798	aff_rescale	0.3388	c1e	0.1701	tap	0.3470	COMB	0.3817
12	ID	Sequence	pep	SLIEIAMVI	aff	0.0681	aff_rescale	0.2892	c1e	0.6767	tap	0.7240	COMB	0.4269
13	ID	Sequence	pep	LIEIAMVII	aff	0.0832	aff_rescale	0.3532	c1e	0.0540	tap	0.5770	COMB	0.3901
14	ID	Sequence	pep	IEIAMVIIL	aff	0.0591	aff_rescale	0.2508	c1e	0.5221	tap	0.9590	COMB	0.3771
15	ID	Sequence	pep	EIAMVIILS	aff	0.0731	aff_rescale	0.3105	c1e	0.0254	tap	-2.2950	COMB	0.1996
16	ID	Sequence	pep	IAMVIILSV	aff	0.0744	aff_rescale	0.3158	c1e	0.7801	tap	0.2880	COMB	0.4472
17	ID	Sequence	pep	AMVIILSVV	aff	0.0661	aff_rescale	0.2808	c1e	0.9617	tap	0.6670	COMB	0.4584
18	ID	Sequence	pep	MVIILSVVF	aff	0.0671	aff_rescale	0.2850	c1e	0.5511	tap	2.9430	COMB	0.5148
19	ID	Sequence	pep	VIILSVVFI	aff	0.0610	aff_rescale	0.2588	c1e	0.1932	tap	0.8300	COMB	0.3293
20	ID	Sequence	pep	IILSVVVIS	aff	0.0621	aff_rescale	0.2638	c1e	0.0395	tap	-2.0290	COMB	0.1683
21	ID	Sequence	pep	ILSVVVISG	aff	0.0646	aff_rescale	0.2741	c1e	0.1375	tap	-1.3480	COMB	0.2273
22	ID	Sequence	pep	LSVVVISGG	aff	0.0650	aff_rescale	0.2759	c1e	0.0281	tap	-1.2640	COMB	0.2169
23	ID	Sequence	pep	SVVVISGGG	aff	0.0600	aff_rescale	0.2540	c1e	0.0280	tap	-1.0280	COMB	0.1630
24	ID	Sequence	pep	VVVISGGSY	aff	0.1322	aff_rescale	0.5615	c1e	0.8088	tap	3.2910	COMB	0.8474 <-E
25	ID	Sequence	pep	VFISGGSYI	aff	0.0551	aff_rescale	0.2338	c1e	0.0521	tap	0.8800	COMB	0.2857
26	ID	Sequence	pep	FISGGSYII	aff	0.0753	aff_rescale	0.3196	c1e	0.8013	tap	0.6110	COMB	0.4703
27	ID	Sequence	pep	ISGGSYIIS	aff	0.0658	aff_rescale	0.2793	c1e	0.0261	tap	-2.4500	COMB	0.1607
28	ID	Sequence	pep	SGGSYIISL	aff	0.0660	aff_rescale	0.2803	c1e	0.7257	tap	0.5870	COMB	0.4185
29	ID	Sequence	pep	GGSYIISLF	aff	0.0560	aff_rescale	0.2379	c1e	0.6468	tap	2.1390	COMB	0.4418
30	ID	Sequence	pep	GSYIISLFL	aff	0.0842	aff_rescale	0.3576	c1e	0.7552	tap	0.4830	COMB	0.4950
31	ID	Sequence	pep	SYIISLFLK	aff	0.0559	aff_rescale	0.2373	c1e	0.8864	tap	0.6710	COMB	0.4038
32	ID	Sequence	pep	YIISLFLKVF	aff	0.0652	aff_rescale	0.2768	c1e	0.4175	tap	2.7660	COMB	0.4777
33	ID	Sequence	pep	IISLFLKVFA	aff	0.0665	aff_rescale	0.2824	c1e	0.1684	tap	-0.4400	COMB	0.2857
34	ID	Sequence	pep	ISLFLKVFQA	aff	0.0570	aff_rescale	0.2421	c1e	0.0480	tap	-0.0150	COMB	0.2485
35	ID	Sequence	pep	SLFKVFAQS	aff	0.0655	aff_rescale	0.2782	c1e	0.5416	tap	-2.0520	COMB	0.2568
36	ID	Sequence	pep	LFKVFAQSC	aff	0.0499	aff_rescale	0.2119	c1e	0.0369	tap	0.2380	COMB	0.2293
37	ID	Sequence	pep	FKVFAQSCV	aff	0.0548	aff_rescale	0.2328	c1e	0.2415	tap	0.3950	COMB	0.2887
38	ID	Sequence	pep	KVFAQSCVV	aff	0.0626	aff_rescale	0.2658	c1e	0.9467	tap	0.5670	COMB	0.4361
39	ID	Sequence	pep	VFAQSCVVA	aff	0.0516	aff_rescale	0.2190	c1e	0.0272	tap	-0.1880	COMB	0.2137
40	ID	Sequence	pep	FAQSCVVAT	aff	0.0705	aff_rescale	0.2994	c1e	0.0264	tap	-0.7330	COMB	0.2667
41	ID	Sequence	pep	AQSCVVATM	aff	0.0594	aff_rescale	0.2522	c1e	0.0611	tap	-1.7050	COMB	0.1761
42	ID	Sequence	pep	QSCVVATDM	aff	0.0956	aff_rescale	0.4059	c1e	0.3288	tap	0.5210	COMB	0.4812
43	ID	Sequence	pep	SCVVATDMN	aff	0.0558	aff_rescale	0.2371	c1e	0.0237	tap	-1.2130	COMB	0.1800
44	ID	Sequence	pep	CVVATDMNS	aff	0.0609	aff_rescale	0.2585	c1e	0.0233	tap	-2.0160	COMB	0.1612
45	ID	Sequence	pep	VVATDMNSV	aff	0.0876	aff_rescale	0.3721	c1e	0.1799	tap	0.5250	COMB	0.4254
46	ID	Sequence	pep	VATDMNSVR	aff	0.0595	aff_rescale	0.2524	c1e	0.1984	tap	1.6100	COMB	0.3627
47	ID	Sequence	pep	ATDMNSVRN	aff	0.1362	aff_rescale	0.5782	c1e	0.0249	tap	-1.3220	COMB	0.5158
48	ID	Sequence	pep	TDMNSVRNI	aff	0.0525	aff_rescale	0.2231	c1e	0.0903	tap	0.4250	COMB	0.2579
49	ID	Sequence	pep	DMNSVRNII	aff	0.0685	aff_rescale	0.2909	c1e	0.3780	tap	0.2560	COMB	0.3604
50	ID	Sequence	pep	MNSVRNIII	aff	0.0650	aff_rescale	0.2761	c1e	0.2158	tap	0.2860	COMB	0.3228
51	ID	Sequence	pep	NSVRNIIIS	aff	0.0703	aff_rescale	0.2987	c1e	0.0271	tap	-2.0540	COMB	0.2000
52	ID	Sequence	pep	SVRNIIISS	aff	0.0625	aff_rescale	0.2653	c1e	0.2079	tap	-1.8640	COMB	0.2033
53	ID	Sequence	pep	VRNIIISSD	aff	0.0512	aff_rescale	0.2172	c1e	0.0800	tap	-1.5870	COMB	0.1499
54	ID	Sequence	pep	RNIIISSDV	aff	0.0578	aff_rescale	0.2454	c1e	0.5482	tap	0.5220	COMB	0.3537
55	ID	Sequence	pep	NIISSDVE	aff	0.0515	aff_rescale	0.2187	c1e	0.0244	tap	-1.3080	COMB	0.1634
56	ID	Sequence	pep	IIISSDVEY	aff	0.1067	aff_rescale	0.4531	c1e	0.9599	tap	3.2860	COMB	0.7614 <-E
57	ID	Sequence	pep	IISSDVEYK	aff	0.0621	aff_rescale	0.2639	c1e	0.5209	tap	0.6210	COMB	0.3730
58	ID	Sequence	pep	ISSDVEYKT	aff	0.0674	aff_rescale	0.2861	c1e	0.2274	tap	-0.6370	COMB	0.2884
59	ID	Sequence	pep	SSDVEYKTG	aff	0.0823	aff_rescale	0.3493	c1e	0.0415	tap	-1.4690	COMB	0.2820
60	ID	Sequence	pep	SDVEYKTGK	aff	0.0559	aff_rescale	0.2357	c1e	0.1720	tap	0.3120	COMB	0.3237
61	ID	Sequence	pep	DVEYKTGKY	aff	0.3358	aff_rescale	1.4257	c1e	0.9594	tap	2.6570	COMB	1.7024 <-E
62	ID	Sequence	pep	VEYKTGKYD	aff	0.0505	aff_rescale	0.2145	c1e	0.0281	tap	-1.8410	COMB	0.1267
63	ID	Sequence	pep	EYKTGKYDV	aff	0.0475	aff_rescale	0.2015	c1e	0.8118	tap	0.4200	COMB	0.3442
64	ID	Sequence	pep	YKTGKYDVS	aff	0.0579	aff_rescale	0.2459	c1e	0.0268	tap	-2.4190	COMB	0.1289
65	ID	Sequence	pep	KTGKYDVSS	aff	0.0698	aff_rescale	0.2965	c1e	0.0344	tap	-2.3750	COMB	0.1829
66	ID	Sequence	pep	TGKYDVSSI	aff	0.0444	aff_rescale	0.1884	c1e	0.1385	tap	0.3460	COMB	0.2264
67	ID	Sequence	pep	GKYDVSSIT	aff	0.0510	aff_rescale	0.2164	c1e	0.0993	tap	-0.6380	COMB	0.1994
68	ID	Sequence	pep	KYDVSSITN	aff	0.0571	aff_rescale	0.2425	c1e	0.0940	tap	-1.1500	COMB	0.1991
69	ID	Sequence	pep	YDVSSITND	aff	0.0583	aff_rescale	0.2477	c1e	0.0332	tap	-1.9640	COMB	0.1545
70	ID	Sequence	pep	VSSITNDNR	aff	0.0654	aff_rescale	0.2776	c1e	0.0430	tap	-1.5490	COMB	0.2066
71	ID	Sequence	pep	VSSITNDRF	aff	0.0753	aff_rescale	0.3196	c1e	0.0346	tap	1.4440	COMB	0.3970
72	ID	Sequence	pep	SSITNDRNF	aff	0.1085	aff_rescale	0.4605	c1e	0.3655	tap	2.7440	COMB	0.6525
73	ID	Sequence	pep	SITNDRNRF	aff	0.0552	aff_rescale	0.2342	c1e	0.0233	tap	-1.1160	COMB	0.1819
74	ID	Sequence	pep	ITNDRNRFN	aff	0.0760	aff_rescale	0.3267	c1e	0.2222	tap	0.4260	COMB	0.3813
75	ID	Sequence	pep	TNDRNRFNK	aff	0.3010	aff_rescale	1.2781	c1e	0.9612	tap	2.5840	COMB	1.5514 <-E
76	ID	Sequence	pep	NDNRNRFNK	aff	0.0480	aff_rescale	0.1951	c1e	0.0308	tap	-0.4600	COMB	0.1788
77	ID	Sequence	pep	DNRFNKYQV	aff	0.0447	aff_rescale	0.1897	c1e	0.6770	tap	0.0360	COMB	0.2931
78	ID	Sequence	pep	NRFNKYQVY	aff	0.0559	aff_rescale	0.2374	c1e	0.9728	tap	3.1410	COMB	0.5404
79	ID	Sequence	pep	RFNKYQVYF	aff	0.0569	aff_rescale	0.2418	c1e	0.8020	tap	2.9930	COMB	0.5117
80	ID	Sequence	pep	FNKYQVYFG	aff	0.0478	aff_rescale	0.2030	c1e	0.0236	tap	-1.5130	COMB	0.1309
81	ID	Sequence	pep	NKYQVYFGN	aff	0.0526	aff_rescale	0.2234	c1e	0.0268	tap	-1.1470	COMB	0.1701
82	ID	Sequence	pep	KYQVYFGND	aff	0.0491	aff_rescale	0.2085	c1e	0.1061	tap	-1.3350	COMB	0.1577
83	ID	Sequence	pep	YQVYFGNDG	aff	0.0645	aff_rescale	0.2739	c1e	0.0274	tap	-1.1610	COMB	0.2200
84	ID	Sequence	pep	QVYFGNDGY	aff	0.1001	aff_rescale	0.4248	c1e	0.9634	tap	3.2110	COMB	0.7299
85	ID	Sequence	pep	VYFGNDGYT	aff	0.0481	aff_rescale	0.2044	c1e	0.3404	tap	-0.3160	COMB	0.2397
86	ID	Sequence	pep	YFGNDGYTL	aff	0.0652	aff_rescale	0.2768	c1e	0.9356	tap	0.9860	COMB	0.4664
87	ID	Sequence	pep	FGNDGYTLI	aff	0.0666	aff_rescale	0.2828	c1e	0.9235	tap	0.0270	COMB	0.4227
88	ID	Sequence	pep	GNDGYTLIS	aff	0.0702	aff_rescale	0.2980	c1e	0.1769	tap	-2.8580	COMB	0.1816

89	ID	Sequence	pep	NDGYTLISG	aff	0.0502	aff_rescale	0.2133	c1e	0.0279	tap	-1.8590	COMB	0.1245
90	ID	Sequence	pep	DGYTLISGS	aff	0.0549	aff_rescale	0.2329	c1e	0.0288	tap	-2.5000	COMB	0.1123
91	ID	Sequence	pep	GYTLISGSS	aff	0.0461	aff_rescale	0.1957	c1e	0.0467	tap	-2.3350	COMB	0.0860
92	ID	Sequence	pep	YTLISGSSP	aff	0.0775	aff_rescale	0.3292	c1e	0.0260	tap	0.0510	COMB	0.3357
93	ID	Sequence	pep	TLISGSSPV	aff	0.0717	aff_rescale	0.3045	c1e	0.8109	tap	0.3300	COMB	0.4427
94	ID	Sequence	pep	LISGSSPV	aff	0.0564	aff_rescale	0.2393	c1e	0.0257	tap	-2.1070	COMB	0.1378
95	ID	Sequence	pep	ISGSSPVSV	aff	0.0663	aff_rescale	0.2815	c1e	0.2996	tap	0.2630	COMB	0.3396
96	ID	Sequence	pep	SGSSPVSD	aff	0.0518	aff_rescale	0.2198	c1e	0.0285	tap	-2.0240	COMB	0.1229
97	ID	Sequence	pep	GSSPVSDV	aff	0.0868	aff_rescale	0.3683	c1e	0.8394	tap	0.2260	COMB	0.5055
98	ID	Sequence	pep	SSPVSDV	aff	0.0951	aff_rescale	0.4036	c1e	0.2323	tap	2.5830	COMB	0.5676
99	ID	Sequence	pep	SPVSDV	aff	0.0572	aff_rescale	0.2428	c1e	0.8316	tap	2.4150	COMB	0.4883
100	ID	Sequence	pep	PVSDV	aff	0.0647	aff_rescale	0.2748	c1e	0.1697	tap	0.2430	COMB	0.3125
101	ID	Sequence	pep	VSDV	aff	0.0671	aff_rescale	0.2849	c1e	0.0581	tap	-0.6700	COMB	0.2601
102	ID	Sequence	pep	SDV	aff	0.1208	aff_rescale	0.5130	c1e	0.1460	tap	0.0770	COMB	0.5387
103	ID	Sequence	pep	VDV	aff	0.0561	aff_rescale	0.2381	c1e	0.8585	tap	2.8110	COMB	0.5074
104	ID	Sequence	pep	DV	aff	0.0558	aff_rescale	0.2368	c1e	0.9491	tap	0.9360	COMB	0.4259
105	ID	Sequence	pep	VFFITCYLA	aff	0.0493	aff_rescale	0.2094	c1e	0.1963	tap	-0.3330	COMB	0.2222
106	ID	Sequence	pep	FFITCYLA	aff	0.0659	aff_rescale	0.2707	c1e	0.1220	tap	0.5630	COMB	0.3308
107	ID	Sequence	pep	FITCYLAKY	aff	0.2315	aff_rescale	0.9829	c1e	0.9421	tap	2.8170	COMB	1.2651 <-E
108	ID	Sequence	pep	ITCYLAKYN	aff	0.0600	aff_rescale	0.2548	c1e	0.0223	tap	-1.4100	COMB	0.1877
109	ID	Sequence	pep	TCYLAKYNC	aff	0.0524	aff_rescale	0.2225	c1e	0.0231	tap	0.2100	COMB	0.2365
110	ID	Sequence	pep	CYLAKYNCV	aff	0.0502	aff_rescale	0.2133	c1e	0.1315	tap	0.5660	COMB	0.2614
111	ID	Sequence	pep	YLAKYNCVW	aff	0.1069	aff_rescale	0.4538	c1e	0.7910	tap	0.8940	COMB	0.6171
112	ID	Sequence	pep	LAKYNCVWI	aff	0.0515	aff_rescale	0.2186	c1e	0.0589	tap	0.5370	COMB	0.2543
113	ID	Sequence	pep	AKYNCVWIL	aff	0.0640	aff_rescale	0.2719	c1e	0.6215	tap	1.3830	COMB	0.4343
114	ID	Sequence	pep	KYNCVWILN	aff	0.0544	aff_rescale	0.2308	c1e	0.0262	tap	-0.8840	COMB	0.1905
115	ID	Sequence	pep	YNCVWILNT	aff	0.0715	aff_rescale	0.3037	c1e	0.0256	tap	-0.7680	COMB	0.2692
116	ID	Sequence	pep	NCVWILNTN	aff	0.0554	aff_rescale	0.2354	c1e	0.0252	tap	-1.2320	COMB	0.1776
117	ID	Sequence	pep	CVWILNTNN	aff	0.0575	aff_rescale	0.2439	c1e	0.0431	tap	-0.9050	COMB	0.2051
118	ID	Sequence	pep	VWILNTNNK	aff	0.0534	aff_rescale	0.2266	c1e	0.2398	tap	0.8350	COMB	0.3043
119	ID	Sequence	pep	WILNTNNKI	aff	0.0698	aff_rescale	0.2965	c1e	0.8529	tap	0.7010	COMB	0.4595
120	ID	Sequence	pep	ILNTNNKIY	aff	0.1895	aff_rescale	0.8046	c1e	0.8491	tap	2.8170	COMB	1.0729 <-E
121	ID	Sequence	pep	LNNTNNKIY	aff	0.0323	aff_rescale	0.1223	c1e	0.1207	tap	0.7110	COMB	0.2893
122	ID	Sequence	pep	NTNNKIYAK	aff	0.0950	aff_rescale	0.4032	c1e	0.8782	tap	0.4340	COMB	0.5566
123	ID	Sequence	pep	TNNKIYAKH	aff	0.0548	aff_rescale	0.2325	c1e	0.0761	tap	-0.6620	COMB	0.2108
124	ID	Sequence	pep	NNKIYAKHM	aff	0.0450	aff_rescale	0.1912	c1e	0.6466	tap	0.1970	COMB	0.2980
125	ID	Sequence	pep	NKIYAKHMS	aff	0.0497	aff_rescale	0.2111	c1e	0.0252	tap	-1.9490	COMB	0.1175
126	ID	Sequence	pep	KIYAKHMSA	aff	0.0583	aff_rescale	0.2474	c1e	0.8430	tap	0.0280	COMB	0.3752
127	ID	Sequence	pep	IYAKHMSAV	aff	0.0608	aff_rescale	0.2584	c1e	0.2172	tap	0.7200	COMB	0.3269
128	ID	Sequence	pep	YAKHMSAVL	aff	0.0649	aff_rescale	0.2757	c1e	0.8642	tap	0.9940	COMB	0.4551
129	ID	Sequence	pep	AKHMSAVLP	aff	0.0558	aff_rescale	0.2370	c1e	0.0246	tap	0.2760	COMB	0.2545
130	ID	Sequence	pep	KHMSAVLPL	aff	0.0606	aff_rescale	0.2573	c1e	0.9282	tap	1.2880	COMB	0.4609
131	ID	Sequence	pep	HMSAVLPLE	aff	0.0360	aff_rescale	0.1330	c1e	0.0306	tap	1.4380	COMB	0.3556
132	ID	Sequence	pep	MSAVLPLEL	aff	0.2044	aff_rescale	0.8679	c1e	0.7611	tap	1.0360	COMB	1.0338 <-E
133	ID	Sequence	pep	SAVLPLELL	aff	0.0623	aff_rescale	0.2645	c1e	0.8990	tap	1.1740	COMB	0.4581
134	ID	Sequence	pep	AVLPLELLS	aff	0.0643	aff_rescale	0.2730	c1e	0.0383	tap	-1.8910	COMB	0.1842
135	ID	Sequence	pep	VLPLELLSA	aff	0.0633	aff_rescale	0.2688	c1e	0.5443	tap	-0.3150	COMB	0.3347
136	ID	Sequence	pep	LPLELLSAY	aff	0.0569	aff_rescale	0.2415	c1e	0.9712	tap	2.6890	COMB	0.5217
137	ID	Sequence	pep	PLELLSAYG	aff	0.0545	aff_rescale	0.2315	c1e	0.0665	tap	-1.9520	COMB	0.1438
138	ID	Sequence	pep	LELLSAYGL	aff	0.0536	aff_rescale	0.2274	c1e	0.9698	tap	0.6960	COMB	0.4077
139	ID	Sequence	pep	ELLSAYGLN	aff	0.0659	aff_rescale	0.2797	c1e	0.0404	tap	-1.4340	COMB	0.2140
140	ID	Sequence	pep	LLSAYGLNY	aff	0.5758	aff_rescale	2.4447	c1e	0.9613	tap	2.7770	COMB	2.7277 <-E
141	ID	Sequence	pep	LSAYGLNYS	aff	0.0800	aff_rescale	0.3396	c1e	0.0425	tap	-2.1360	COMB	0.2392
142	ID	Sequence	pep	SAYGLNYSG	aff	0.0601	aff_rescale	0.2553	c1e	0.2607	tap	-1.0550	COMB	0.2417
143	ID	Sequence	pep	AYGLNYSGE	aff	0.0455	aff_rescale	0.1932	c1e	0.0249	tap	-1.2660	COMB	0.1336
144	ID	Sequence	pep	YGLNYSGES	aff	0.0759	aff_rescale	0.3224	c1e	0.0680	tap	-2.3600	COMB	0.2146
145	ID	Sequence	pep	GLNYSGESG	aff	0.0512	aff_rescale	0.2172	c1e	0.0260	tap	-1.7010	COMB	0.1360
146	ID	Sequence	pep	LNYSGESGV	aff	0.0542	aff_rescale	0.2301	c1e	0.1496	tap	0.3230	COMB	0.2687
147	ID	Sequence	pep	NYSGESGVT	aff	0.0469	aff_rescale	0.1993	c1e	0.0661	tap	-0.3600	COMB	0.1912
148	ID	Sequence	pep	YSGESGVTG	aff	0.0717	aff_rescale	0.3045	c1e	0.0508	tap	-1.3970	COMB	0.2423
149	ID	Sequence	pep	SGESGVTGL	aff	0.0608	aff_rescale	0.2582	c1e	0.8259	tap	0.5450	COMB	0.4093
150	ID	Sequence	pep	GESGVTGLL	aff	0.0621	aff_rescale	0.2637	c1e	0.9674	tap	0.6900	COMB	0.4434
151	ID	Sequence	pep	ESGVTGLLF	aff	0.1510	aff_rescale	0.6409	c1e	0.6329	tap	2.1780	COMB	0.8448 <-E
152	ID	Sequence	pep	SGVTGLLFF	aff	0.0638	aff_rescale	0.2710	c1e	0.6587	tap	0.1620	COMB	0.3779
153	ID	Sequence	pep	GVTGLLFFK	aff	0.0547	aff_rescale	0.2320	c1e	0.8774	tap	1.4620	COMB	0.4367
154	ID	Sequence	pep	VTGLLFFKRY	aff	0.2313	aff_rescale	0.9822	c1e	0.9356	tap	2.7720	COMB	1.2611 <-E

Number of MHC ligands 10 identified. Number of peptides 154. Protein name Sequence

NUMERO DE RESIDUO	SECUENCIA
24	V V F I S G G S Y
56	I I I S S D V E Y
61	D V E Y K T G K Y
75	T N D N R F N K Y
107	F I T C Y L A K Y
120	I L N T N N K I Y
132	M S A V L P L E L
140	L L S A Y G L N Y
151	E S G V T G L L I
154	V T G L L F K R Y

Secuencia de BfpK con los epitopes según el programa NetCTL 1.2 en recuadros de colores.

M I F E G R Q Q K G
 L S L I E I A M V I I L S V V F I S G G
S Y I I S L F K V F A Q S C V V A T D M
 N S V R N I I I S S D V E Y K T G K Y D
 V S S I T N D N R F N K Y Q V Y F G N D
 G Y T L I S G S S P V S V D V F F I T C
Y L A K Y N C V W I L N T N N K I Y A K
 H M S A V L P L E L L L S A Y G L N Y S G
E S G V T G L L F K R Y

En gris el péptido señal de la proteína

NetMHCII 2.2 server predicts binding of peptides to HLA-DR, HLA-DQ, HLA-DP and mouse MHC class II alleles using artificial neuron networks

DESCRIPTION

In the header of the out is indicated which prediction method and HLA allele were selected, as well as the corresponding two threshold values defining high binding peptides, and weak binding peptides. High binding peptides have an IC50 value below 50 nM, and weak binding peptides an IC50 values below 500 nM. For Artificial Neural network (ANN) prediction the output consists of 6 columns. For weight matrix predictions the output consists of 5 columns:

- Residue number.
- Peptide sequence (9mer)
- Prediction score (called 1-log50K(aff) for ANN predictions)
- Affinity as IC50 value in nM (only for ANN predictions)
- Bind Level (WB for weak binder, SB for strong binder)
- Sequence name

The predictions for each protein are summarized with a line stating the number of high and weak binding peptides identified.

is in FSA format

de length 15
NetMHCII version 2.2.

binder threshold 50.00. Weak binder threshold 500.00.

Allele	pos	peptide	core 1-log50k(aff)	affinity(nM)	%Rank	Identity	Bind Level	
HLA-DRB10101	1	MIFEGROOKGLSLIE	GROOKGLSL	0.6378	50.3	32.00	Sequence	WB
HLA-DRB10101	2	IFEGROQKGLSLIEI	GRQOKGLSL	0.7145	21.9	16.00	Sequence	SB
HLA-DRB10101	3	FEGROOKGLSLIEIA	GROOKGLSL	0.6432	47.5	32.00	Sequence	SB
HLA-DRB10101	4	EGRQOKGLSLIEIAM	GRQOKGLSL	0.5368	150.1	50.00	Sequence	WB
HLA-DRB10101	5	GRQOKGLSLIEIAMV	LSLIEIAMV	0.5638	112.1	50.00	Sequence	WB
HLA-DRB10101	6	RQOKGLSLIEIAMVI	LSLIEIAMV	0.5854	88.7	32.00	Sequence	WB
HLA-DRB10101	7	QOKGLSLIEIAMVIT	LSLIEIAMV	0.6127	66.0	32.00	Sequence	WB
HLA-DRB10101	8	QKGLSLIEIAMVIL	LSLIEIAMV	0.6407	48.8	32.00	Sequence	SB
HLA-DRB10101	9	KGLSLIEIAMVILS	LSLIEIAMV	0.6080	69.5	32.00	Sequence	WB
HLA-DRB10101	10	GLSLIEIAMVILSV	LSLIEIAMV	0.5572	120.4	50.00	Sequence	WB
HLA-DRB10101	11	LSLIEIAMVILSVV	IAMVILSV	0.5392	146.3	50.00	Sequence	WB
HLA-DRB10101	12	SLIEIAMVILSVVF	IAMVILSV	0.5394	145.9	50.00	Sequence	WB
HLA-DRB10101	13	LIEIAMVILSVVFI	VILSVVFI	0.5704	104.4	50.00	Sequence	WB
HLA-DRB10101	14	IEIAMVILSVVVIS	VILSVVFI	0.5709	103.9	50.00	Sequence	WB
HLA-DRB10101	15	EIAMVILSVVFISG	VILSVVFI	0.5795	94.6	32.00	Sequence	WB
HLA-DRB10101	16	IAMVILSVVFISGG	VILSVVFI	0.5804	93.7	32.00	Sequence	WB
HLA-DRB10101	17	AMVILSVVFISGGS	VILSVVFI	0.5503	129.8	50.00	Sequence	WB
HLA-DRB10101	18	MVILSVVFISGGSY	VVFISGGSY	0.5825	91.6	32.00	Sequence	WB
HLA-DRB10101	19	VILSVVFISGGSYI	VVFISGGSY	0.6616	38.9	32.00	Sequence	SB
HLA-DRB10101	20	IILSVVFISGGSYII	FISGGSYII	0.7851	10.2	8.00	Sequence	SB
HLA-DRB10101	21	ILSVVFISGGSYIIS	FISGGSYII	0.8132	7.5	4.00	Sequence	SB
HLA-DRB10101	22	LSVVFISGGSYIISL	FISGGSYII	0.8235	6.7	4.00	Sequence	SB
HLA-DRB10101	23	SWVFISGGSYIISLF	FISGGSYII	0.8200	7.0	4.00	Sequence	SB
HLA-DRB10101	24	VVFISGGSYIISLKF	FISGGSYII	0.7859	10.1	8.00	Sequence	SB
HLA-DRB10101	25	VFISGGSYIISLKFV	FISGGSYII	0.7463	15.6	16.00	Sequence	SB
HLA-DRB10101	26	FISGGSYIISLKFVF	FISGGSYII	0.7017	25.2	16.00	Sequence	SB
HLA-DRB10101	27	ISGGSYIISLKFVFA	SYIISLKFV	0.5320	158.1	50.00	Sequence	WB
HLA-DRB10101	28	SGGSYIISLKFVFAQ	SYIISLKFV	0.5512	128.5	50.00	Sequence	WB
HLA-DRB10101	29	GGSYIISLKFVFAQS	IISLKFVFA	0.5577	119.8	50.00	Sequence	WB
HLA-DRB10101	30	GSYIISLKFVFAQSC	IISLKFVFA	0.5672	108.1	50.00	Sequence	WB
HLA-DRB10101	31	SYIISLKFVFAQSCV	LFKVFQAQSC	0.6213	60.2	32.00	Sequence	WB
HLA-DRB10101	32	YIISLKFVFAQSCVV	FKVFAQSCV	0.6789	32.3	32.00	Sequence	SB
HLA-DRB10101	33	IISLKFVFAQSCVVA	FKVFAQSCV	0.7341	17.8	16.00	Sequence	SB
HLA-DRB10101	34	ISLKFVFAQSCVVAT	FKVFAQSCV	0.7672	12.4	8.00	Sequence	SB
HLA-DRB10101	35	SIFKFVFAQSCVATD	FKVFAQSCV	0.7374	17.1	16.00	Sequence	SB

HLA-DRB10101	36	LFKVFQAQSCVVATDMN	FAQSCVVAT	0.7241	19.8	16.00	Sequence	SB
HLA-DRB10101	37	FKVFAQSCVVATDMN	FAQSCVVAT	0.7170	21.4	16.00	Sequence	SB
HLA-DRB10101	38	KVFQAQSCVVATDMNS	FAQSCVVAT	0.8191	61.6	32.00	Sequence	WB
HLA-DRB10101	39	VFAQSCVVATDMNSV	FAQSCVVAT	0.5334	155.8	50.00	Sequence	WB
HLA-DRB10101	40	FAQSCVVATDMNSVR	FAQSCVVAT	0.4710	306.0	50.00	Sequence	WB
HLA-DRB10101	41	AQSCVVATDMNSVRN	VVATDMNSV	0.4279	487.6	50.00	Sequence	WB
HLA-DRB10101	42	QSCVVATDMNSVRNI	VVATDMNSV	0.5124	195.5	50.00	Sequence	WB
HLA-DRB10101	43	SCVVATDMNSVRNII	VVATDMNSV	0.4336	458.7	50.00	Sequence	WB
HLA-DRB10101	44	CVVATDMNSVRNIII	MNSVRNIII	0.4588	349.4	50.00	Sequence	WB
HLA-DRB10101	45	VVATDMNSVRNIIIS	TMNSVRNI	0.5043	213.6	50.00	Sequence	WB
HLA-DRB10101	46	VATDMNSVRNIIISS	MNSVRNIII	0.5186	182.8	50.00	Sequence	WB
HLA-DRB10101	47	ATDMNSVRNIIISSD	MNSVRNIII	0.5151	189.9	50.00	Sequence	WB
HLA-DRB10101	48	TMNSVRNIIISSDV	MNSVRNIII	0.4596	346.1	50.00	Sequence	WB
HLA-DRB10101	49	DMNSVRNIIISSDVE	MNSVRNIII	0.4129	573.7	50.00	Sequence	WB
HLA-DRB10101	50	MNSVRNIIISSDVEY	MNSVRNIII	0.4311	471.2	50.00	Sequence	WB
HLA-DRB10101	51	NSVRNIIISSDVEYK	RNIIISSDV	0.4392	431.9	50.00	Sequence	WB
HLA-DRB10101	52	SVRNIIISSDVEYKT	RNIIISSDV	0.3994	663.8	50.00	Sequence	WB
HLA-DRB10101	53	VRNIIISSDVEYKTK	RNIIISSDV	0.3511	1119.6	50.00	Sequence	WB
HLA-DRB10101	54	RNIIISSDVEYKTKG	NIIISSDVE	0.2906	2153.9	50.00	Sequence	WB
HLA-DRB10101	55	NIIISSDVEYKTKGY	NIIISSDVE	0.2232	4468.9	50.00	Sequence	WB
HLA-DRB10101	56	IISSDVEYKTKGYD	IISDVEYK	0.1819	6984.1	50.00	Sequence	WB
HLA-DRB10101	57	IISDVEYKTKGYDV	IISDVEYK	0.1443	10497.3	50.00	Sequence	WB
HLA-DRB10101	58	ISSDVEYKTKGYDVS	YKTGKYDVS	0.1504	9825.4	50.00	Sequence	WB
HLA-DRB10101	59	SSDVEYKTKGYDVSS	YKTGKYDVS	0.1676	8158.2	50.00	Sequence	WB
HLA-DRB10101	60	SDVEYKTKGYDVSSI	YKTGKYDVS	0.2248	4390.2	50.00	Sequence	WB
HLA-DRB10101	61	DVEYKTKGYDVSSIT	YKTGKYDVS	0.2868	2246.7	50.00	Sequence	WB
HLA-DRB10101	62	VEYKTKGYDVSSITN	YKTGKYDVS	0.2913	2138.7	50.00	Sequence	WB
HLA-DRB10101	63	EYKTKGYDVSSITND	YKTGKYDVS	0.2389	3769.7	50.00	Sequence	WB
HLA-DRB10101	64	YKTKGYDVSSITNDN	YKTGKYDVS	0.2378	3816.7	50.00	Sequence	WB
HLA-DRB10101	65	KTKGYDVSSITNDNR	KYDVSSITN	0.2667	2791.7	50.00	Sequence	WB
HLA-DRB10101	66	TGKYDVSSITNDNRF	YDVSSITND	0.3092	1761.3	50.00	Sequence	WB
HLA-DRB10101	67	GKYDVSSITNDNRFN	YDVSSITND	0.2621	2934.6	50.00	Sequence	WB
HLA-DRB10101	68	KYDVSSITNDNRFNK	YDVSSITND	0.2348	3942.7	50.00	Sequence	WB
HLA-DRB10101	69	YDVSSITNDNRFNKY	SSITNDNRF	0.2166	4800.7	50.00	Sequence	WB
HLA-DRB10101	70	DVSSITNDNRFNKYQ	SSITNDNRF	0.1841	6820.1	50.00	Sequence	WB
HLA-DRB10101	71	VSSITNDNRFNKYQV	SSITNDNRF	0.1816	7011.0	50.00	Sequence	WB
HLA-DRB10101	72	SSITNDNRFNKYQVY	SSITNDNRF	0.1777	7314.0	50.00	Sequence	WB
HLA-DRB10101	73	SITNDNRFNKYQVYF	NRFNKYQVY	0.1899	6406.9	50.00	Sequence	WB
HLA-DRB10101	74	ITNDNRFNKYQVYFG	FNKYQVYFG	0.3931	711.2	50.00	Sequence	WB
HLA-DRB10101	75	TNDNRFNKYQVYFGN	FNKYQVYFG	0.4169	549.6	50.00	Sequence	WB
HLA-DRB10101	76	NDNRFNKYQVYFGND	FNKYQVYFG	0.4172	547.5	50.00	Sequence	WB
HLA-DRB10101	77	DNRFNKYQVYFGNDG	FNKYQVYFG	0.4390	432.4	50.00	Sequence	WB
HLA-DRB10101	78	NRFNKYQVYFGNDGY	FNKYQVYFG	0.4608	341.7	50.00	Sequence	WB

HLA-DRB10101	78	NRFNKYQVYFGNDGY	FNKYQVYFG	0.4608	341.7	50.00	Sequence	WB
HLA-DRB10101	79	RNFNKYQVYFGNDGYT	YQVYFGNDG	0.4989	223.8	50.00	Sequence	WB
HLA-DRB10101	80	FNKYQVYFGNDGYTL	YFGNDGYTL	0.7190	20.9	16.00	Sequence	SB
HLA-DRB10101	81	NKYQVYFGNDGYTLI	YFGNDGYTL	0.7769	11.2	8.00	Sequence	SB
HLA-DRB10101	82	KYQVYFGNDGYTLIS	YFGNDGYTL	0.8126	7.6	4.00	Sequence	SB
HLA-DRB10101	83	YQVYFGNDGYTLISG	YFGNDGYTL	0.8308	6.2	4.00	Sequence	SB
HLA-DRB10101	84	QVYFGNDGYTLISGS	YFGNDGYTL	0.7974	9.0	8.00	Sequence	SB
HLA-DRB10101	85	VYFGNDGYTLISGSS	YFGNDGYTL	0.7592	13.5	16.00	Sequence	SB
HLA-DRB10101	86	YFGNDGYTLISGSSP	YFGNDGYTL	0.7035	24.7	16.00	Sequence	SB
HLA-DRB10101	87	FGNDGYTLISGSSPV	YTLISGSSP	0.6527	42.9	32.00	Sequence	SB
HLA-DRB10101	88	GNDGYTLISGSSPVS	YTLISGSSP	0.7164	21.5	16.00	Sequence	SB
HLA-DRB10101	89	NDGYTLISGSSPVS	YTLISGSSP	0.7846	10.3	8.00	Sequence	SB
HLA-DRB10101	90	DGYTLISGSSPVSVD	YTLISGSSP	0.7645	12.8	8.00	Sequence	SB
HLA-DRB10101	91	GYTLISGSSPVSVDV	LISGSSPVS	0.7843	10.3	8.00	Sequence	SB
HLA-DRB10101	92	YTLISGSSPVSVDVF	LISGSSPVS	0.7599	13.4	16.00	Sequence	SB
HLA-DRB10101	93	TLISGSSPVSVDVFF	LISGSSPVS	0.6626	38.5	32.00	Sequence	SB
HLA-DRB10101	94	LISGSSPVSVDVFFI	ISGSSPVS	0.5810	93.1	32.00	Sequence	WB
HLA-DRB10101	95	ISGSSPVSVDVFFIT	PVSVDVFFI	0.4301	476.2	50.00	Sequence	WB
HLA-DRB10101	96	SGSSPVSVDVFFITC	PVSVDVFFI	0.4220	520.0	50.00	Sequence	WB
HLA-DRB10101	97	GSSPVSVDVFFITCY	PVSVDVFFI	0.4545	365.7	50.00	Sequence	WB
HLA-DRB10101	98	SSPVSVDVFFITCYL	PVSVDVFFI	0.4582	351.7	50.00	Sequence	WB
HLA-DRB10101	99	SPVSVDVFFITCYLA	VFFITCYLA	0.4735	297.8	50.00	Sequence	WB
HLA-DRB10101	100	PVSVDVFFITCYLAK	VFFITCYLA	0.4830	268.7	50.00	Sequence	WB
HLA-DRB10101	101	VSDVFFITCYLAKY	DVFFITCYL	0.4962	232.9	50.00	Sequence	WB
HLA-DRB10101	102	SDVFFITCYLAKYNC	VFFITCYLA	0.4991	225.7	50.00	Sequence	WB
HLA-DRB10101	103	VDVFFITCYLAKYN	VFFITCYLA	0.4979	228.8	50.00	Sequence	WB
HLA-DRB10101	104	DVFFITCYLAKYNCV	FITCYLAKY	0.5655	110.1	50.00	Sequence	WB
HLA-DRB10101	105	VFFITCYLAKYNCVW	CYLAKYNCV	0.5900	84.5	32.00	Sequence	WB
HLA-DRB10101	106	FFITCYLAKYNCVWI	CYLAKYNCV	0.5861	88.0	32.00	Sequence	WB
HLA-DRB10101	107	FITCYLAKYNCVWIL	CYLAKYNCV	0.6123	66.3	32.00	Sequence	WB
HLA-DRB10101	108	ITCYLAKYNCVWILN	CYLAKYNCV	0.5712	103.5	50.00	Sequence	WB
HLA-DRB10101	109	TCYLAKYNCVWILNT	CYLAKYNCV	0.5372	149.6	50.00	Sequence	WB
HLA-DRB10101	110	CYLAKYNCVWILNTN	CYLAKYNCV	0.5259	169.0	50.00	Sequence	WB
HLA-DRB10101	111	YLAKYNCVWILNTNN	YNCVWILNT	0.5186	182.8	50.00	Sequence	WB
HLA-DRB10101	112	LAKYNCVWILNTNNK	YNCVWILNT	0.5567	121.1	50.00	Sequence	WB
HLA-DRB10101	113	AKYNCVWILNTNNK	WILNTNNKI	0.5729	101.6	32.00	Sequence	WB
HLA-DRB10101	114	KYNCVWILNTNNKIY	WILNTNNKI	0.6729	34.4	32.00	Sequence	SB
HLA-DRB10101	115	YNCVWILNTNNKIYA	WILNTNNKI	0.7368	17.2	16.00	Sequence	SB
HLA-DRB10101	116	NCVWILNTNNKIYAK	WILNTNNKI	0.7902	9.7	8.00	Sequence	SB
HLA-DRB10101	117	CVWILNTNNKIYAKH	WILNTNNKI	0.7751	11.4	8.00	Sequence	SB
HLA-DRB10101	118	VWILNTNNKIYAKH	WILNTNNKI	0.7491	15.1	16.00	Sequence	SB
HLA-DRB10101	119	WILNTNNKIYAKHMS	WILNTNNKI	0.6885	29.1	32.00	Sequence	SB
HLA-DRB10101	120	ILNTNNKIYAKHMSA	NNKIYAKHM	0.5885	85.8	32.00	Sequence	WB
HLA-DRB10101	121	LNTNNKIYAKHMSAV	NNKIYAKHM	0.6692	35.8	32.00	Sequence	SB
HLA-DRB10101	122	NTNNKIYAKHMSAVL	YAKHMSAVL	0.6499	44.2	32.00	Sequence	SB

HLA-DRB10101	123	TNKKIYAKHMSAVLP	YAKHMSAVL	0.6439	47.1	32.00	Sequence	SB
HLA-DRB10101	124	NKKIYAKHMSAVLPL	YAKHMSAVL	0.6907	28.4	16.00	Sequence	SB
HLA-DRB10101	125	NKIYAKHMSAVLPLE	YAKHMSAVL	0.6870	29.6	32.00	Sequence	SB
HLA-DRB10101	126	KIYAKHMSAVLPLEL	YAKHMSAVL	0.6955	27.0	16.00	Sequence	SB
HLA-DRB10101	127	IYAKHMSAVLPLELL	KHMSAVLPL	0.6882	29.2	32.00	Sequence	SB
HLA-DRB10101	128	YAKHMSAVLPLELLS	KHMSAVLPL	0.6274	56.3	32.00	Sequence	WB
HLA-DRB10101	129	AKHMSAVLPLELLSA	VLPLELLSA	0.6042	72.4	32.00	Sequence	WB
HLA-DRB10101	130	KHMSAVLPLELLSAY	VLPLELLSA	0.6363	51.2	32.00	Sequence	WB
HLA-DRB10101	131	HMSAVLPLELLSAYG	VLPLELLSA	0.6504	43.9	32.00	Sequence	SB
HLA-DRB10101	132	MSAVLPLELLSAYGL	LELLSAYGL	0.7753	11.4	8.00	Sequence	SB
HLA-DRB10101	133	SAVLPLELLSAYGLN	LELLSAYGL	0.7701	12.0	8.00	Sequence	SB
HLA-DRB10101	134	AVLPLELLSAYGLNY	LELLSAYGL	0.7829	10.5	8.00	Sequence	SB
HLA-DRB10101	135	VLPLELLSAYGLNYS	LELLSAYGL	0.7923	9.5	8.00	Sequence	SB
HLA-DRB10101	136	LPLELLSAYGLNYSG	LELLSAYGL	0.7635	12.9	8.00	Sequence	SB
HLA-DRB10101	137	PLELLSAYGLNYSGE	LELLSAYGL	0.7277	19.0	16.00	Sequence	SB
HLA-DRB10101	138	LELLSAYGLNYSGES	LELLSAYGL	0.6962	26.8	16.00	Sequence	SB
HLA-DRB10101	139	ELLSAYGLNYSGESG	LLSAYGLNY	0.4434	412.6	50.00	Sequence	WB
HLA-DRB10101	140	LLSAYGLNYSGESGV	YGLNYSGES	0.4095	595.6	50.00	Sequence	WB
HLA-DRB10101	141	LSAYGLNYSGESGVT	YGLNYSGES	0.4298	477.8	50.00	Sequence	WB
HLA-DRB10101	142	SAYGLNYSGESGVTG	YSGESGVTG	0.4709	306.3	50.00	Sequence	WB
HLA-DRB10101	143	AYGLNYSGESGVTGL	YSGESGVTG	0.5940	80.9	32.00	Sequence	WB
HLA-DRB10101	144	YGLNYSGESGVTGLL	YSGESGVTG	0.7087	23.4	16.00	Sequence	SB
HLA-DRB10101	145	GLNYSGESGVTGLLF	YSGESGVTG	0.7778	11.1	8.00	Sequence	SB
HLA-DRB10101	146	LNYSGESGVTGLLFFK	YSGESGVTG	0.7409	16.5	16.00	Sequence	SB
HLA-DRB10101	147	NYSGESGVTGLLFFKR	YSGESGVTG	0.6529	42.7	32.00	Sequence	SB
HLA-DRB10101	148	YSGESGVTGLLFFKRY	YSGESGVTG	0.6018	74.3	32.00	Sequence	WB

Allele: HLA-DRB10101. Number of high binders 56. Number of weak binders 64. Number of peptides 148

VaxiJen RESULTS

Model selected: **bacteria**

Threshold for this model: **0.4**

Your Sequence:

MIFEGRQQKGLSLIEIAMV IILSVVFISGGS
 YIISLFKVFAQSCV VATDMNSVRNII ISSDV
 EYKTGKYDVSS ITNDRFNKYQVYFGNDGYT
 LISGSSPVSVDVFFITCYLAKYNCVWILNTN
 NKIYAKHMSAVLPLELLSAYGLNYSGESGVT
 GLLFKRY

Overall Prediction for the Antigen = **0.4995** (Probable **ANTIGEN**).

11 BIBLIOGRAFÍA

Lieberman JA., Petro CD., Thomas S., Yang A., & Donnenberg MS. (2015). Type IV Pilus Secretins Have Extracellular C Termini. *mBio*, 6, 1-13.

Craig L., Volkmann N., Arvai AS., Pique ME., Yeager M., Egelman EH., & Tainer JA. (2006). Type IV Pilus Structure by Cryo-Electron Microscopy and Crystallography: Implications for Pilus Assembly and Functions; *Molecular Cell*, 23, 651–662.

Telford JL., Barocchi MA., Margarit I., Rappuoli R., & Grandi G. (2006). Pili in Gram-positive pathogens. *Nature reviews*, 4, 509-519.

McNamara BP. & Donnenberg MS. (2000). Evidence for specificity in type 4 pilus biogenesis by enteropathogenic *Escherichia coli*. *Microbiology*, 146, 719–729.

Giltner CL., Nguyen Y., & Burrows LL. (2012). Type IV Pilin Proteins: Versatile Molecular Modules. *Microbiology and Molecular Biology Reviews*, 76, 740–772.

Peabody CR., Chung YJ., Yen MR., Vidal D., Pugsley AP. & Saier MH. (2003). Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology*, 149, 3051–3072.

Craig L., & Li J. (2008). Type IV pili: paradoxes in form and function, Elsevier; *Current Opinion in Structural Biology* 2008, 18:267–277.

Korotkov KV., Sandkvist M. & Hol WG. (2012). The T2SS, the T4PS, archaeal flagella and the transformation system of Gram-positive bacteria. *Nature*, 10, 336-351.

Strom MS & Lori S. (1991). Amino acid substitutions in pilin of *Pseudomonas aeruginosa* effect on leader peptide cleavage, amino-terminal methylation, and pilus assembly. *The Journal of Biological Chemistry*, 266, 1656- 1664.

Pelicic V. (2008). Type IV pili: e pluribus unum?. *Molecular microbiology*, 68, 827- 837.

Zhang XL., Tsui IS., Yip CM. Fung AW. Wong DK. Dai X., Yang Y., Hackett J & Morris C. (2000). *Salmonella enterica* serovar Typhi uses type IVb Pili to enter human intestinal epithelial cells. *Infection and Immunity*, 68, 3067–3073.

Craig L., Taylor RK., Pique ME., Adai BD., Arvai AS., Singh M., Lloyd SJ., Shin DS., Getzoff ED., Yeager M., Forest KT., & Tainer JA. (2003). Type IV Pilin Structure and Assembly: X-Ray and EM Analyses of *Vibrio cholerae* Toxin-Coregulated Pilus and *Pseudomonas aeruginosa* PAK Pilin. *Molecular cell*, 11, 1139-1150.

Donnenberg MS. & Kaper JB. (1992). Enteropathogenic *Escherichia coli* Infection and Immunity, 60, 3953-3961.

Aas FE., Winther HC., Wolfgang M., Frye S., Lovold C., Roos N., Putten JP. & Koomey M. (2007). Substitutions in the N-terminal alpha helical spine of *Neisseria gonorrhoeae* pilin affect Type IV pilus assembly, dynamics and associated functions. Molecular Microbiology, 63, 69-85.

Xu XF., Tan YW., Lam L., Hackett J., Zhang M. & Mok YK. (2004). NMR structure of a type IVb pilin from *Salmonella typhi* and its assembly into pilus. The Journal of Biological Chemistry, 279, 31599–31605.

Ramboarina S., Fernandes PJ., Daniell S., Islam S., Simpson P., Frankel G. Booy F., Donnenberg MS. & Matthews S. (2005). Structure of the Bundle-forming Pilus from Enteropathogenic *Escherichia coli*. 2005. The Journal of Biological Chemistry. The American Society for Biochemistry and Molecular Biology, 280, 40252–40260.

Hélaine S., Dyer DH., Nassif X., Pelicic V., & Forest KT. (2007). 3D structure/function analysis of PilX reveals how minor pilins can modulate the virulence properties of type IV pili. PNAS, 104, 15888–15893.

Nguyen Y., Sugiman-Marangos S., Harvey H., Bell S., Charlton CL., Junop MS., & Burrows LL. (A). (2015). *Pseudomonas aeruginosa* minor pilins prime type IVa pilus assembly and promote surface display of the PilY1 adhesin; The Journal of Biological Chemistry, 290, 601-611.

Kawahara K., Oki H., Shunsuke F., Takahiro M., Yuji K., Daisuke M., Tooru T., Takeshi H., Tetsuya L., Shota N., & Tadayasu O. (2015). Cloning, expression, purification, crystallization and X-ray crystallographic analysis of CofB, the minor pilin subunit of CFA/III from human enterotoxigenic *Escherichia coli*. Crossmark, 71, 663–667.

Kus JV., Tullis E., Cvitkovitch DG. & Burrows LL. (2004). Significant differences in type IV pilin allele distribution among *Pseudomonas aeruginosa* isolates from cystic fibrosis (CF) versus non-CF patients. Microbiology, 150, 1315–1326.

Giltner CL., Schaik EJ., Audette GF., Kao D. Hodges RS., Hassett DJ. & Irvin RT. (2006). The *Pseudomonas aeruginosa* type IV pilin receptor binding domain functions as an adhesin for both biotic and abiotic surfaces. Molecular Microbiology, 59, 1083–1096.

Li J., Egelman EH. & Craig L. (2004). Structure of the *Vibrio cholerae* type IVb pilus and stability comparison with the *Neisseria gonorrhoeae* type IVa pilus.

Martínez de la Peña CF., De Masi L., Nisa S., Mulvey G., Tong J., Donnenberg MS., & Armstrong GD. (2016). BfpI, BfpJ, and BfpK Minor Pilins are Important

for the function and biogenesis of Bundle Forming Pili Expressed by Enteropathogenic *Escherichia coli*, Journal of bacteriology, 198, 846-856.

Kolappan S., Ng D., Yang G., Harn T., & Craig L. (2015). Crystal Structure of the Minor Pilin CofB, the Initiator of CFA/III Pilus Assembly in Enterotoxigenic *Escherichia coli*. JBC, 290, 25805-25818.

Maldarelli GA., De Masi L., Rosenvinge EC., Carter M. & Donnenberg MS. (2014). Identification, Immunogenicity and Crossreactivity of Type IV Pilin and Pilin-like Proteins from *Clostridium difficile*. NIH-PA, 71, 302-314.

Winther HC., Wolfgang M., Dunham S., Putten JP., Dorward D., Lovold C., Aas FE. & Koomey M. (2005). A conserved set of pilin-like molecules controls type IV pilus dynamics and organelle-associated functions in *Neisseria gonorrhoeae*. Molecular Microbiology, 56, 903–917.

Giltner CL. Lori M. & Burrows L. (2010). *Pseudomonas aeruginosa* Minor Pilins Are Incorporated into Type IV Pili. Elsevier, 398, 444-461.

H elaine S., Carbonnelle E., Prouvensier L., Beretti JL., Nassif X. & Pelicic V. (2005) PilX, a pilus-associated protein essential for bacterial aggregation, is a key to pilus facilitated attachment of *Neisseria meningitides* to human cells. Molecular Microbiology, 55, 65–77.

Kuchma SL., Griffin, EF. & O'Toole GA. (2012). Minor Pilins of the Type IV Pilus System Participate in the Negative Regulation of Swarming Motility. Journal of Bacteriology, 194, 5388–5403.

Bieber D., Ramer SW., Wu CY., Murray WJ., Tobe T., Fernandez R., & Schoolnik GK. (1998). Type IV Pili, Transient Bacterial Aggregates, and Virulence of Enteropathogenic. Sciences, 280, 2114-2118.

Gir on JA., Yue Ho AS. & Schoolnik GK. (1993). Characterization of Fimbriae Produced by Enteropathogenic *Escherichia coli*. Journal OF Bacteriology, 175, 7391-7403.

Zahavi EE., Lieberman JA., Donnenberg MS., Nitzan M, Baruch K, Rosenshine I, Turner JR., Melamed-Book N., Feinstein N, Zlotkin-Rivkin E, Aroeti B. (2011). Bundle-forming pilus retraction enhances enteropathogenic *Escherichia coli* infectivity. Molecular Biology of the Cell, 22, 36-47.

Humphries RM., Griener TP., Vogt SL., Mulvey GL., Donnenberg MS., Kitov PI., Surette M., & Armstrong GD. (2010). *N*-acetyllactosamine-induced retraction of bundle-forming pili regulates virulence-associated gene expression in enteropathogenic *Escherichia coli* Mol Microbiol, 76, 1111–1126.

Jarrell. KF. (2009). Type IV Pilus structure. En Pili and Flagella(1-18). Canadá: CaisterAcademi Press

Bustamante VH., Calva E. & Puente JL. (1998). Analysis of *cis*-Acting Elements Required for *bfpA* Expression in Enteropathogenic *Escherichia coli*. Journal of Bacteriology, 180, 3013–3016.

Nguyen Y., Harvey H., Sugiman-Marangos S., Bell SD., Buensuceso RN., Junop MS., & Burrows LL. (B). (2015). Structural and Functional Studies of the *Pseudomonas aeruginosa* Minor Pilin, PilE. The Journal of Biological Chemistry, 290, 26856–26865.

Schreiber W., Stone KD., Strong MA., DeTolla LJ., Hoppert M. & Donnenberg MS. (2002). BfpU, a soluble protein essential for type IV pilus biogenesis in enteropathogenic *Escherichia coli*. Microbiology, 148, 2507–2518.

Martínez de la Peña CF. (2015). Role of Type IV Pili in the Pathogenesis of Enterovirulent Bacteria, Calgary, Alberta.

Levine MM., Bergquist EJ., Nalin DR., Waterman DH., Hornick RB., Young CR., Sotman S. (1978). *Escherichia coli* strains that cause diarrhoea but do not produce heat-labile or heat-stable enterotoxins and are non-invasive. Lancet, 1, 1119–1122.

Wingfield PT. (2016). Overview of the Purification of Recombinant Proteins. HHS Public Access. 1-50.

Hartung S., Arvai AS., Wood T., Kolappan S., Shin DS., Craig L. & Tainer JA. (2011). Ultrahigh Resolution and Full-length Pilin Structures with Insights for Filament Assembly, Pathogenic Functions, and Vaccine Potential. The journal of biological chemistry, 286, 44254–44265.

Celli J., Deng W. & Finlay BB. (2000). Enteropathogenic *Escherichia coli* (EPEC) attachment to epithelial cells: exploiting the host cell cytoskeleton from the outside. Cell Microbiol, 2, 1-9.

Donnenberg MS., Kaper JB., Finlay BB. (1997). Interactions between enteropathogenic *Escherichia coli* and host epithelial cells. Trends Microbiol, 5, 109-114.

Kenny B. (1999). Phosphorylation of tyrosine 474 of the enteropathogenic *Escherichia coli* (EPEC) Tir receptor molecule is essential for actin nucleating activity and is preceded by additional host modifications. Mol Microbiol, 31, 1229–1241.

Goosney DL., DeVinney R., Pfuetzner RA., Frey EA, Strynadka NC., Finlay BB.(2000). Enteropathogenic *E. coli* translocated intimin receptor, Tir, interacts directly with alpha-actinin. Current biology , 10, 735-738

Elliott SJ., Wainwright LA., McDaniel TK., Jarvis KG., Deng YK., Lai LC., McNamara BP., Donnenberg MS. & Kaper JB. (1998). The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol*, 28, 1-4.

Puente JL., Bieber D., Ramer SW., Murray W. & Schoolnik GK. (1996). The bundle-forming pili of enteropathogenic *Escherichia coli* : transcriptional regulation by environmental signals. *Mol Microbiol*, 20, 87-100.

Bustamante VH., Calva E. & Puente JL. (1998) Analysis of the promoter and cis-acting elements involved in the BfpT-dependent expression of bfpA in enteropathogenic *Escherichia coli*. *J Bacteriol*, 180, 3013- 3016.

McDaniel TK., Jarvis KG., Donnenberg MS. & Kaper JB. (1995). A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl. Acad. Sci. USA*, 92, 1664–1668.

Rapid Protocol for Preparation of Electrocompetent *Escherichia coli* and *Vibrio cholerae*. 06-06-2017, de Journal of visualized Experiments Sitio web: <https://www.jove.com/pdf/50684/jove-protocol-50684-rapid-protocol-for-preparation-electrocompetent-escherichia-coli?&language=Spanish>

Addgene The nonprofit plasmid repository. (2016). Analyze sequence:pET3a. 06-06-2017, de Addgene Sitio web: https://www.addgene.org/browse/sequence_vdb/2637/

Novagen. (2004). User protocol TB009 Competent cells. 05 Junio 2017, de Novagen Sitio web: <https://www.researchgate.net/file.PostFileLoader.html?id=584f083f93553b9ac410e521&assetKey=AS%3A438569241190400%401481574463351>

Alliance Bio. (2007). HB101 competent cells. Alliance Bio, 1-3.

Thermo Fisher Scientific. (2010). Manual One shot BL21 (DE3). 05-02-17, de Thermo Fisher Scientific Sitio web: <https://www.thermofisher.com/order/catalog/product/C600003>

Agilent. (2015). pET System Vector and Hosts Instruction manual. 28-01-16, de Agilent Sitio web: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.612.5161&rep=rep1&type=pdf>.

The QIAexpressionist™ handbook for high-level expression and purification of 6xHis-tagged proteins **March 2001 Qiagen 5 edición**

Ritchie C.(2012). Purificación de proteínas. 22-08-2016, de Labome Sitio web: <http://www.labome.es/method/Protein-Purification.html>