

Proteína LIC10494 de *Leptospira interrogans* serovar *Copenhageni*: modelo estructural y regiones funcionales asociadas

Leidy Viviana Barreto¹, George Emilio Barreto¹, Ludis Morales¹, Orlando Emilio Acevedo², Janneth González Santos ^{1*}

¹ Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá D.C., Colombia

² Departamento de Física, Facultad de Ciencias, Pontificia Universidad Javeriana, Bogotá D.C., Colombia

*janneth.gonzalez@javeriana.edu.co

Recibido: 30-11-2011; Aceptado: 23-02-2012

Resumen

Objetivo. Predecir computacionalmente la estructura tridimensional de la proteína antigénica LIC10494 e inferir las regiones funcionales asociadas importantes para su antigenicidad e inmunogenicidad. **Materiales y métodos.** Se realizó un análisis computacional de la estructura primaria de LIC10494 a partir de los servidores BLAST, PROTPARAM, PROTSCALE, DAS, SOSUI, TOPPRED, TMAP, TMPRED, SPLIT4, PHDHTM, TMHMM2, HMMTOP2, GLOBPLOT y PROSITE. La estructura secundaria se obtuvo por consenso de los algoritmos SOPM, PREDATOR GOR4, DPM y DSC. La aproximación a la estructura terciaria se obtuvo con el algoritmo MUSTER. La minimización de energía se obtuvo a partir del campo de fuerza AMBER94 de la suite de análisis molecular SCHRODINGER, y la validación tanto estereoquímica como energética del modelo se realizó con el servidor RAMPAGE. El modelo final fue visualizado con el programa PyMol v.0,98. **Resultados.** En el presente estudio se propone un modelo computacional que detalla la estructura tridimensional de la lipoproteína hipotética LIC10494 y está de acuerdo con reportes experimentales previos; el estudio demuestra la existencia de patrones que podrían tener un papel importante en la patogenicidad y la protección de la bacteria frente al sistema inmune del hospedero; la presencia de una región desordenada entre los aminoácidos 80 y 140; y la presencia de un segmento transmembrana entre los aminoácidos 8 y 22. **Conclusión.** La coincidencia entre segmentos estructurales y funcionales sugiere que el modelo puede usarse para predecir ciertos aspectos del comportamiento biológico de la proteína en cuanto a la patogenicidad e inmunogenicidad de la bacteria.

Palabras clave: antígeno, bacteria, leptospirosis, LIC10494, proteína periférica

Abstract

Protein LIC10494 of *Leptospira interrogans* serovar *Copenhageni*: structural model and associated functional regions. Objective. Predict by computational means the 3D structure of the antigenic protein LIC10494 and report associated important functional regions for its pathogenicity and immunogenicity. **Materials and methods.** We performed a computational analysis of the primary structure of LIC10494 using the servers BLAST, PROTPARAM, PROTSCALE, DAS, SOSUI, TOPPRED, TMAP, TMPred, SPLIT4, PHDHTM, TMHMM2, HMMTOP2, GLOBPLOT and PROSITE. The secondary structure was obtained by consensus of the algorithms SOPM, PREDATOR GOR4, DPM and DSC. The approach to the tertiary structure was obtained using the algorithm MUSTER. The energy minimization was done using the AMBER94 force field of the Schrodinger suite of molecular analysis, and the stereochemistry and energy model validation was performed by the RAMPAGE server. The final model was visualized using PyMol V.0,98. **Results.** This study proposes a computational model that describes the 3D structure of the hypothetical lipoprotein LIC10494 and agrees with previous experimental reports; thus, our study demonstrates the existence of patterns that could play an important role in the pathogenicity and protection of the bacteria against the host immune system; the presence of a disorganized region between amino acids 80 and 140, and of a transmembrane segment between amino acids 8 and 22. **Conclusion.** The coincidence between structural and functional segments

suggests that our model can be used to predict certain aspects of the biological behaviour of the protein according to the pathogenic and immunogenic characteristics of the bacteria.

Key words: antigen, bacteria leptospirosis, LIC10494, outer membrane protein.

Resumo

Proteína LIC10494 de *Leptospira interrogans* serovar Copenhagen: modelo estrutural e regiões funcionais associadas. **Objetivo.** Predizer computacionalmente a estrutura tridimensional da proteína antigênica LIC10494 e inferir as regiões funcionais associadas importantes para a sua antigenicidade e imunogenicidade. **Materiais e métodos.** Foi realizada uma análise computacional da estrutura primária da LIC10494 nos servidores, BLAST, PROTPARAM, PROTSCALE, DAS, SOSUI, TOPPRED, TMAP, TMPRED, SPLIT4, PHDHTM, TMHMM2, HMMTOP2, GLOBPLOT e PROSITE. A estrutura secundária foi obtida por consenso dos algoritmos SOPM, PREDATOR GOR4, DPM e DSC. A aproximação para a estrutura terciária foi obtida com o algoritmo MUSTER. A minimização de energia foi obtida a partir do campo de força AMBER94 do conjunto de análise molecular SCHRODINGER, e a validação estereoquímica e energética do modelo foi realizada utilizando o servidor RAMPAGE. O modelo final foi visualizado com o programa PyMol V. 0,98. **Resultados.** Este estudo propõe um modelo computacional que descreve a estrutura tridimensional da lipoproteína hipotética LIC10494 e concorda com anteriores relatórios experimentais; o estudo demonstra a existência de padrões que poderiam desempenhar um papel importante na patogenicidade e na proteção da bactéria ao sistema imune do hospedeiro; a presença de uma região desordenada entre os aminoácidos 80 e 140, e a presença de um segmento transmembrana entre os aminoácidos 8 e 22. **Conclusão.** A coincidência entre os segmentos estruturais e funcionais sugere que o modelo pode ser utilizado para prever determinados aspectos do comportamento biológico da proteína na patogenicidade e imunogenicidade da bactéria.

Palavras-chave: antígeno, bactéria, leptospirose, LIC10494, proteína da membrana externa.

Introducción

La Leptospirosis es una zoonosis infecciosa considerada como una enfermedad reemergente de comportamiento endémico, y de distribución mundial con brotes continuos en varios continentes (1). Datos estadísticos revelaron en el 2009 que la incidencia mundial de leptospirosis es de aproximadamente 1:100.000 habitantes en climas templados, 100:100.000 habitantes en climas húmedos tropicales y más de 100:100.000 habitantes en poblaciones con alto riesgo de exposición (www.who.org). En Latinoamérica, la leptospirosis humana ha alcanzado tasas superiores a 10:100.000 habitantes, y en Colombia se estimó una incidencia de 3:100.000 habitantes en el año 2008 de acuerdo a la notificación de casos realizada al Sistema Nacional de Vigilancia (Sivigila) (www.ins.gov.co). A causa de la gran cantidad de especies de mamíferos reservorios y la presencia de *Leptospira interrogans* en ambientes cálidos y húmedos, la leptospirosis es considerada la zoonosis más importante en el mundo (1).

La Leptospirosis es causada por la bacteria patógena *Leptospira interrogans*, microorganismo espiroquetal con más de 200 variedades o serovares resultantes de la heterogenicidad estructural en el componente carbohidrato del lipopolisacárido de la membrana externa de *Leptospira*. Muchos serovares están adaptados a hospederos mamíferos específicos que albergan la bacteria en sus túbulos renales para después excretarla por la orina. Esta enfermedad es frecuente en sitios urbanos donde

la población puede llegar a ser contagiada por contacto de la piel y/o las mucosas con agua y/o suelos contaminados por la orina de animales infectados, especialmente ratas. La etapa inicial de la enfermedad es caracterizada por síntomas como fiebre, escalofríos, dolores musculares y de cabeza severos. La progresión de la enfermedad genera insuficiencia pulmonar, renal y/o hepática entre el 5 y el 15% de los casos, con tasas de mortalidad del 5 al 40% (1). La principal estrategia de intervención contra la enfermedad consiste en vacunas inactivas que no confieren inmunidad cruzada frente a diferentes serovares, proporcionan una inmunidad de menor intensidad y duración que las vacunas vivas, y necesitan varias dosis de refuerzo para generar protección a largo plazo (2).

La necesidad de crear nuevas vacunas para la prevención de la leptospirosis estimuló el estudio de proteínas antigenicas de membrana que además de estar conservadas a través de varias especies leptospirales y ser aptas para la generación de resistencia a largo plazo, tuviesen funciones involucradas principalmente en la conversión de señales extracelulares en procesos intracelulares importantes para la presentación del antígeno y su interacción con el anticuerpo (3). Varios grupos de investigación encontraron proteínas que pueden estar envueltas en la patogenicidad y la virulencia de leptospira resultando ser las lipoproteínas de membrana las principales involucradas, razón por la cual se han convertido en el objeto de estudio más importante para la creación de vacunas efectivas frente a la enfermedad (4, 5).

La secuenciación del genoma de *Leptospira interrogans* facilitó la identificación de: 15 genes codificantes para probables proteínas de superficie extracelular y 174 lipoproteínas hipotéticas que podrían estar envueltas en la patogénesis de la bacteria siendo por tanto posibles blancos terapéuticos para la protección inmunológica frente a la Leptospirosis. LIC10494 fue identificada como una de éstas lipoproteínas. Según el estudio llevado a cabo por Atzingen y Colaboradores (6) la lipoproteína hipotética LIC10494 logró generar en los animales de experimentación, una respuesta inmune caracterizada por el incremento de anticuerpos, lo cual indujo la protección parcial frente a la enfermedad. Sin embargo, la forma en la cual LIC10494 confiere dicha protección es aún desconocida.

El conocimiento de la estructura tridimensional de las proteínas de membrana, importante para el entendimiento de sus mecanismos de acción, es imposibilitado por dificultades experimentales para la sobreexpresión y cristalización de éstas (7). Los métodos computacionales permiten hacer una aproximación al conocimiento estructural de las proteínas de membrana contrarrestando las dificultades experimentales mencionadas, razón por la cual han tomado importancia para el entendimiento de los mecanismos de acción de las proteínas a partir de su estructura tridimensional (8).

Por lo anterior, el objetivo de este estudio consistió en predecir computacionalmente la estructura tridimensional de la proteína antigénica LIC10494 e inferir regiones funcionales asociadas importantes para su antigenicidad e inmunogenicidad. Los resultados evidencian la primera aproximación existente de la estructura tridimensional de la proteína antigénica LIC10494 y el encuentro de patrones funcionales asociados a ésta que podrían ser importantes para los mecanismos de patogénesis e inmunogenicidad de la bacteria. Tales resultados están de acuerdo con reportes experimentales de Atzingen en el 2009; Atzingen y colaboradores en el 2010; y Barbosa y colaboradores en el 2006 en cuanto a la composición estructural de la proteína (6, 21, 27).

Materiales y métodos

Análisis computacional de la estructura primaria de LIC10494

La secuencia de la proteína hipotética LIC10494 (Acc. No. J455996261 gb |AAS 69115,1|) fue obtenida a través del GenBank en el Centro Nacional de Información para Biotecnología (NCBI) (9). La búsqueda de similitud con otras secuencias reportadas se llevó a cabo con BLAST disponible en NCBI (9).

Las propiedades fisicoquímicas, tales como el peso molecular, el índice de estabilidad, el punto isoeléctrico, el tiempo de vida, el índice alifático y el promedio general de hidropatía, fueron calculadas con PROTPARAM (<http://expasy.org/tools/>). Para el análisis del índice de hidrofobicidad se usó la herramienta PROTSCALE y el algoritmo de Kyte y Doolittle del servidor Expasy proteomic tolos (<http://expasy.org/tools/>). El tamaño de la ventana para los análisis de PROTSCALE fue de 7 aminoácidos por ser el recomendado para encontrar regiones hidrofílicas que están expuestas en la superficie y posiblemente son antigenicas. La predicción de segmentos transmembranales se obtuvo a partir del consenso de 9 programas (<http://expasy.org/tools/>): (DAS, SOSUI y TOPPRED2), los cuales se basan en escalas de hidrofobicidad; TMAP (10), TMPRED (<http://expasy.org/tools/>) y SPLIT4 (10), los cuales se fundamentan en métodos estadísticos de propensión estructural; y PHDHTM (10), TMHMM2 (<http://expasy.org/tools/>) y HMMTOP2 (<http://expasy.org/tools/>), los cuales usan redes neuronales para sus predicciones. La predicción de las regiones globulares y las regiones no estructuradas de la proteína fueron determinadas con GLOBPLOT (11), y la identificación de motivos se realizó con el servidor PROSITE (12).

Predicción de estructuras secundaria y terciaria de LIC10494

Para la predicción de la estructura secundaria se realizó un consenso de cinco algoritmos a partir del servidor NSP@ (<http://www.bioinf.manchester.ac.uk/dbbrowsr/bioactivity/NPS2.html>): Dos métodos fundamentados en el uso de parámetros de probabilidad determinados por las frecuencias relativas de las apariciones de cada aminoácido en cada tipo de estructura secundaria (SOPM y PREDATOR) y tres métodos basados en probabilidad de inferencia bayesiana como son GOR (GOR4, DPM y DSC). La estructura terciaria de la proteína se predijo con el algoritmo MUSTER (13). La minimización de la energía total de interacción entre átomos del modelo estructural previamente obtenido se llevó a cabo con el campo de fuerza AMBER 94 (14) de MAESTRO 8.0 (15). La validación estereoquímica y energética del modelo se realizó con el servidor Rampage (<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>). El modelo final fue visualizado con el programa PyMol v.0,98 (<http://www.pymol.org>).

Resultados y discusión

Análisis computacional de la estructura primaria de LIC10494

El alineamiento local de secuencias realizado con BLAST permitió la identificación de una proteína con secuencia de

aminoacidos similar en un 99,9% a la secuencia de LIC10494: La lipoproteína putativa con función desconocida LA3735 (Número de acceso: ref|NP_713915.1) de *Leptospira interrogans serovar Lai*. Teniendo en cuenta que la identidad de cada proteína se basa en el orden preciso de cada uno de sus aminoácidos (16), el porcentaje de identidad o similitud entre las secuencias LIC10494 y LA3735 permite deducir que éstas son probablemente homólogas y casi idénticas en cuanto a la conformación de su estructura tridimensional y por tanto en sus funciones. Lo anterior constituye un buen hallazgo teniendo en cuenta que la proteína putativa LA3735 no tiene información disponible con respecto a su estructura y función.

Las propiedades fisicoquímicas obtenidas mediante PROTPARAM (**Tabla 1**) permitieron establecer el peso molecular, el índice de estabilidad, el punto isoeléctrico, el tiempo de vida, el índice alifático y el promedio general de hidropatía de LIC10494.

Generalmente los compuestos de peso molecular entre 10.000 a 100.000 Da y con índices de estabilidad menores a 40 son inmunógenos potentes (17). La proteína tiene un peso molecular, entre los 10.000 y los 100.000 Da, acorde con la inmunogenicidad descrita por Atzingen (6) y un índice de estabilidad que la clasifica como estable. En consecuencia, los datos obtenidos permiten inferir que LIC10494 es un inmunógeno potente. Varios autores han descrito que el incremento de la estabilidad conformacional de las proteínas antigenicas puede aumentar o disminuir la inmunogenicidad de las mismas (18), razón por la cual es posible que la proteína LIC10494 sea capaz de incrementar su inmunogenicidad en condiciones experimentales.

El punto isoeléctrico, donde las cargas positivas de la proteína se igualan a las negativas anulando la existencia de movimiento en un campo eléctrico representa por tanto el pH al cual LIC10494 podría mostrar un mínimo de solubilidad durante ensayos experimentales facilitando su aislamiento en un campo eléctrico durante los mismos (19).

El tiempo de vida calculado por PROTPARAM hace referencia al tiempo (*in vitro*) que tarda la proteína en desaparecer una vez es sintetizada por la célula. El resultado del tiempo de vida de LIC10494 puede ser útil para la identificación de su residuo N terminal por medio de métodos experimentales (20). El índice alifático, o volumen relativo ocupado por las cadenas laterales alifáticas, permite inferir que la proteína es altamente termoestable obedeciendo a las observaciones de Ikai en las cuales las proteínas de origen mesofílico con alto índice alifático tienden a ser más termoestables (21). Por otra parte, el índice de hidropatía es una medida de la energía [kJ/mol] que se debe usar para

transferir un segmento de secuencia de longitud definida de un medio hidrofóbico a un medio hidrofílico, es así, como los perfiles hidropáticos constituyen una medida de la hidrofobicidad de una secuencia proteica que permite tener indicios sobre la topología, dominios transmembranales y probables sitios antigenicos expuestos en la superficie de la proteína (19). Con el fin de hacer una aproximación a la estructura y topología de LIC10494, PROTPARAM posibilitó el conocimiento de un promedio general de hidropatía (GRAVY) que indica la posibilidad para LIC10494 de establecer interacciones con el agua sugiriendo que puede encontrarse altamente hidratada en medios acuosos, ésto dado que a mas bajo indice general de hidropatia hay mayor posibilidad de establecer dicha interacción (22).

Tabla 1. Resultados de 6 parámetros de evaluación de las propiedades fisicoquímicas de LIC10494 revelados por Protparam.

Propiedades Fisicoquímicas de LIC10494	
Peso molecular (Da)	25811,2
Índice de estabilidad	29,76
Punto isoeléctrico	8,48
Tiempo de vida (Horas)	30
Índice alifático	72,21
Promedio general de hidropatía (GRAVY)	0,483

Para el resultado de ProtScale el tamaño de la ventana usada tuvo un valor de 7, lo cual fue apropiado para encontrar las regiones hidrofilicas expuestas en la superficie y por tanto potencialmente antigenicas (23). El análisis hidrofóbico de ProtScale (**Figura 1**) reveló que LIC10494 es altamente hidrofílica, es decir que gran parte de su estructura esta en contacto con el entorno acuoso, tambien exhibió 4 regiones hidrofóbicas en contacto con la parte interna de la membrana celular (entre los aminoácidos 4-24, 59-70, 182-189 y 219-227).

Tanto los resultados obtenidos a partir de la herramienta computacional Protscale como los referentes al perfil de hidropatía de la proteína se ajustan positivamente a un estudio experimental previo, llevado a cabo con PAGE/SDS, que sugiere la naturaleza altamente soluble de la proteína (24).

El resultado del consenso para la predicción de helices transmembranales coincide con los resultados de Protscale, dado que la region transmembranal predicha en el consenso

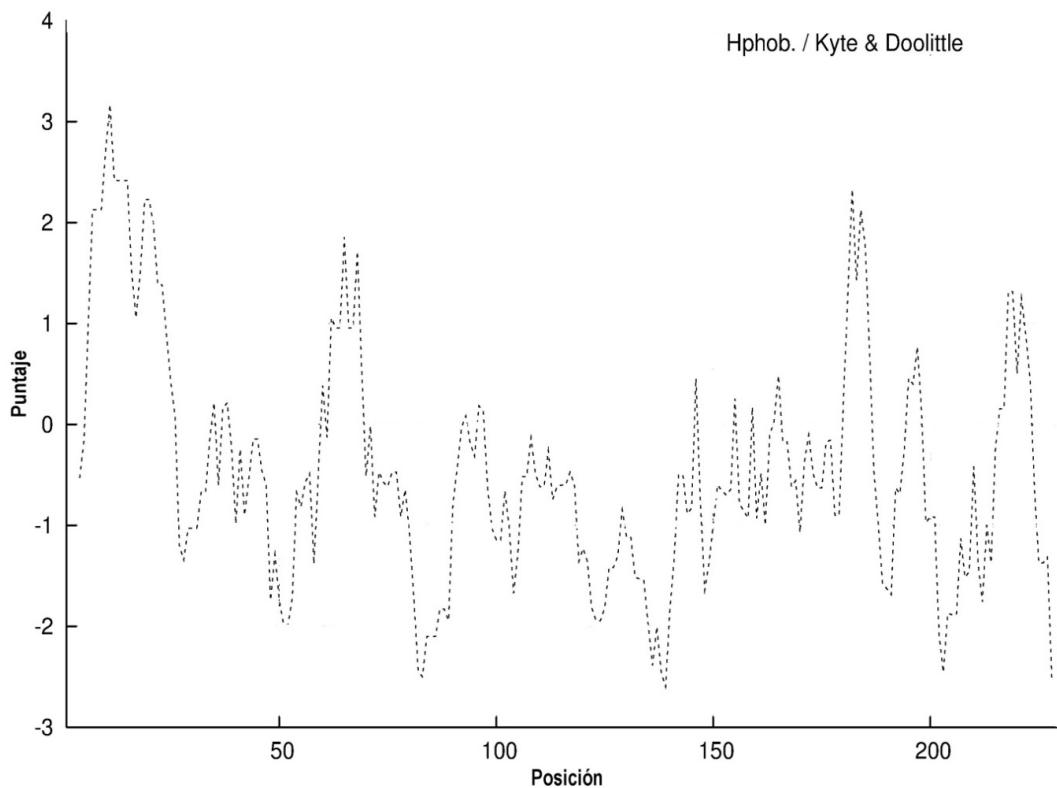


Figura 1. Análisis hidropático de LIC10494 con base en los valores de Kyte y Doolittle y usando una ventana de 7 residuos.

se encuentra entre los aminoacidos 8 y 22 (**Figura 2**), segmento de aminoacidos que corresponde con la region altamente hidrofóbica revelada por ProtScale.

Se ha descrito que las regiones donde hay bajos niveles de aminoácidos hidrofóbicos junto con largos tramos de residuos hidrofílicos pero con cargas netas altas y a pH neutro, están teóricamente asociadas con la falta de compactación de las proteínas bajo condiciones fisiológicas normales resultando en estructuras nativas estructuralmente desordenadas (25). Se ha descubierto que los segmentos desordenados de las proteínas contienen importantes sitios funcionales predichos como motivos lineales (25). Las regiones desordenadas de LIC10494 fueron analizadas para predecir su comportamiento en solución e identificar regiones que pueden ser funcionalmente importantes para su inmunogenicidad. Teniendo en cuenta la ausencia de proteinas homólogas a LIC10494, la predicción de dominios globulares y regiones desordenadas de la proteína se realizo con GLOBPLOT, un servicio web que permite predecir dominios estructurales por medio de herramientas basadas

en la escala de propension de los aminoacidos y, por lo tanto, no esta limitado por la búsqueda de secuencias homólogas estructuralmente conocidas (26). La predicción de regiones desordenadas reveladas por GLOBPLOT (**Figura 3**) sugieren que LIC10494 tiene 2 regiones flexibles y desordenadas (residuos entre las posiciones 80-140 y 163-172) que podrían estar involucradas en funciones importantes para *Leptospira interrogans*. Los resultados combinados de GLOBPLOT y PROSCALE coinciden en cuanto a la predicción de una region desordenada entre los aminoacidos 80 y 140.

En cuanto a los motivos se identificó una región rica en treonina (PS50325), 2 regiones de N-miristilación, 6 sitios de N-glicosilación, 8 sitios de C fosforilación y 6 sitios de fosforilación para la casein-quinasa II (**Tabla 2**). Como se sugirió en un estudio realizado con el hongo *Candida albicans*, la existencia de regiones ricas en treonina para proteinas como la Als5p, una adhesina extracelular de la cual depende el estado patogénico del hongo, puede tener un papel importante en la adhesión a glicoproteínas de membrana celular y proteinas de la matriz extracelular de éste (27);

	10	20	30	40	50-231
	MNQLFYK III	L FFTS II LNG	CV I FGNKTAR	AAEKAATVKA	NYQIGKSENK
DAS	***	***** ***	* *		
PHDHTM	***	***** ***	* ****		
TMAP	** ***	***** ***	* ***** ***	*	
TMHMM2	*** ***	***** ***	* ***** *		
TMPRED	* ***	***** ***	* ****		
TOPPRED2	** ***	***** ***	* ***** *		
SPLIT4	* ***** ***	***** ***	* *****		
HMMTOP2	***	***** ***	* *		
SOSUI	** ***	***** ***	* ***** **		
	- - - - h h HHH HHHHH HHH		H H h hh - - - -		

Figura 2. Resultado del consenso de 9 programas que predicen segmentos transmembranales. Los asteriscos indican los residuos que hacen parte de la helice transmembrana. La predicción consenso se da en la parte inferior: H. predicción consenso de los 9 programas (residuo 8 al 22), h. predicción consenso de 6 programas (residuo 6 al 25).

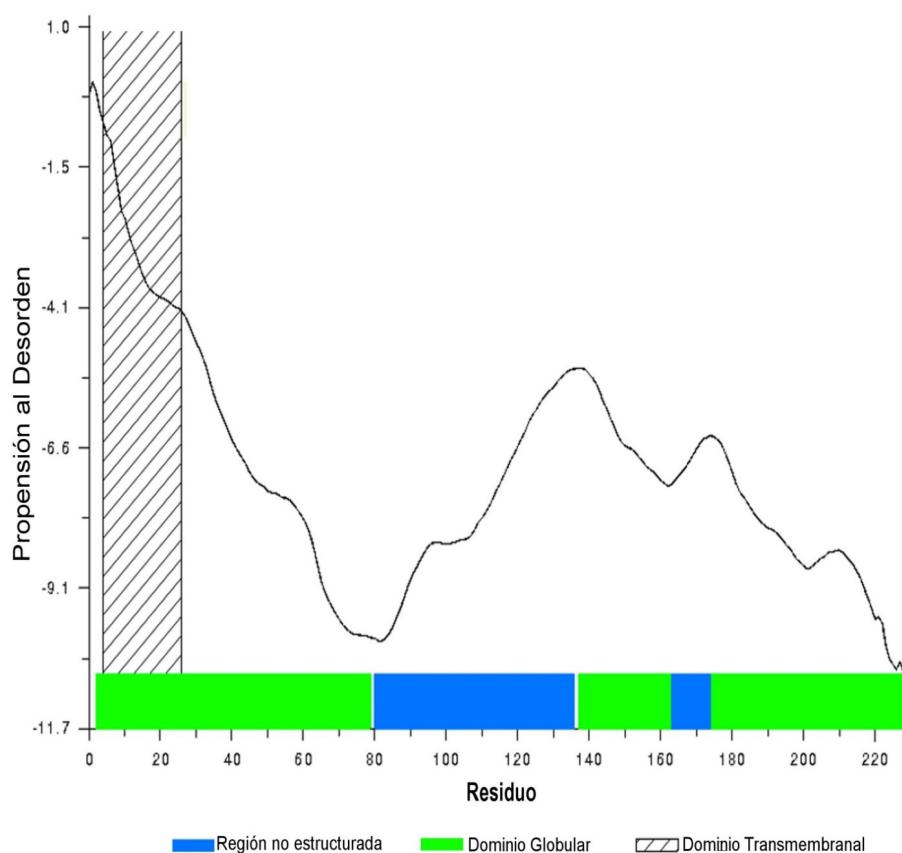


Figura 3. Análisis de GLOBPLOT para LIC10494. Las regiones con desórdenes intrínsecos corresponden a las regiones de color azul (aminoácidos 80-140 y 163-172).

Tabla 2. Patrones funcionales encontrados mediante PROSITE para la proteína hipotética LIC10494.

Patrón	ID	Descripción	Número de aminoácidos	Aminoácidos involucrados
N-{P}-[ST]-{P}	ASN-GLYCOSILATION	Sitio de N-glicosilación	26-29	NKTA
			135-138	NSSK
			151-154	NLSS
			171-174	NGSI
			178-181	NFTT
			206-209	NFSD
[ST]-X-[RK]	PKC_PHOSPHO_SITE	Sitio de Fosforilación de proteína quinaza C	28-30	TAR
			37-39	TVK
			84-86	TSK
			123-125	TPK
			136-138	SSK
			140-142	SKR
[ST]-X(2)-[DE]	CK2_PHOSPHO_SITE	Sitio de fosforilación de casein quinasa II	153-155	SSK
			191-193	TGK
			119-122	TTPE
			130-133	TSTE
			136-139	SSKD
			173-176	SILD
G-{EDRKHPFYW}-X(2)-[STAGCN]-{P}	MYRISTYL	Sitio de N-miristilación	208-211	SDAD
			219-222	SLVD
			25-30	GNKTAR
			167-172	GMSDNG
			84-134	TSKTNPPPTPVATSNPMLQNPEATTAA PSTPTAPTTTPETPKPTTTSTET

El ID corresponde al nombre del sitio funcional encontrado. Se especifica las regiones entre los aminoácidos en la cual está presente el patrón funcional y los aminoácidos involucrados en dicha región.

lo anterior permitió inferir que dicha región en la proteína LIC10494 podría ser de alta importancia, sin embargo, es conveniente estudiar a fondo esta probabilidad. Los 2 sitios potenciales de N-miristilación pueden tener un rol importante en la lipidación de la proteína, la cual es necesaria para el transporte de ésta por el plasma sanguíneo. Por otra parte, los sitios de glicosilación en proteínas de superficie de bacterias gram negativas son importantes para la patogénesis de la bacteria por su papel en la adhesión, en la protección contra la degradación proteolítica, en el incremento de la solubilidad y en la variación antigenica. Teniendo en cuenta que LIC10494 es una proteína de superficie y *Leptospira interrogans* serovar *Copenhageni* es una bacteria Gram-negativa, es posible inferir la importancia de los 6 sitios potenciales de N-Glicosilación para generar protección de la bacteria frente al sistema inmune del hospedero. Considerando dicha importancia en cuanto a la patogénesis y las interacciones con el sistema inmunológico, los sitios de

glicosilación de LIC10494 podrían representar en el futuro nuevas dianas terapéuticas para el desarrollo de antibióticos por alteración de los sitios de glicosilación mediante la inhibición de las interacciones carbohidrato-receptor. Sin embargo las interacciones carbohidrato-receptor que se puedan dar en los sitios de glicosilación de LIC10494 aún se deben estudiar. El encuentro de sitios de fosforilación para LIC10494 puede ser importante para la regulación de la transcripción y la replicación de la bacteria llevada a cabo por medio de Quinazas Celulares (28).

Estructura secundaria

Debido a que la calidad de predicción de la estructura secundaria puede ser mejorada por clasificaciones basadas en consensos, donde los diferentes métodos pueden ser combinados para obtener una predicción más confiable

que a partir de uno de los métodos (29), la predicción de la estructura secundaria se realizó con el servidor NPSA (<http://www.bioinf.manchester.ac.uk/dbbrowser/bioactivity/NPS2.html>) , un servidor que arroja los resultados promedio de la predicción realizada con los algoritmos SOPM; PREDATOR; GORIV; DSC Y DPM; SOPM y PREDATOR fundamentados en la probabilidad condicional de que determinado aminoácido haga parte de una estructura secundaria considerando que los aminoácidos adyacentes adquieran la misma estructura, mientras que GORIV, DSC y DPM utilizan conjuntos de entrenamiento cuyos elementos son estructuras resueltas para identificar secuencias motivo comunes asociadas a disposiciones particulares de estructuras secundarias.

El resultado del consenso de NPSA para la estructura secundaria de LIC10494 (**Figura 4**) reveló la existencia de 5 hélices α entre los aminoácidos: 28-41, 60-69, 141-150 y 210-227; y 4 láminas α entre los aminoácidos: 7-17,

158-162, 181-186 y 194-200. Los aminoácidos restantes fueron regiones no estructuradas. Éstos resultados coinciden con el reporte experimental basado en el método de CD espectroscopía que reveló la predominancia de hélices α y láminas α en la estructura secundaria de LIC10494 (30).

Estructura terciaria

La predicción de la estructura terciaria de la proteína se realizó con MUSTER, un programa de modelamiento por threading disponible en el meta-servidor LOMETS, que difiere del modelamiento por homología en que no se cuenta con estructuras similares reportadas, por lo tanto el modelo se deriva del alineamiento de fragmentos de la secuencia sobre diferentes plantillas estructurales con pliegues reconocidos y reportados en bases de datos biológicas. La minimización de energía, obtenida a partir del software Macromodel V9.7 disponible en la suite de análisis molecular de Schrodinger, fue de primer orden, usando el campo de fuerzas AMBER94

Figura 4. Consenso de la estructura secundaria de LIC10494 generado por NPSA con base en 5 programas de predicción. La secuencia consenso se puede ver en la parte inferior (Sec. Cons) indicando las regiones con hélices α (h), láminas β (e) y regiones no estructuradas (c).

con 50 iteraciones, donde la energía mínima obtenida para el modelo fue de -6300,2251 kJ/mol.

Se ha establecido que una estructura estable depende de parámetros estereoquímicos correspondientes a los ángulos de torsión que determinan su plegamiento proteico. Dado que el rango permitido de ángulos dentro de una estructura es muy restrictivo, las variaciones en el ángulo de torsión son muy pocas y han hecho posible la identificación de ángulos posibles que permiten una aproximación a la probabilidad de

que la conformación estructural de la proteína sea correcta (31). La validación del modelo de LIC10494 evaluado con RAMPAGE cumplió esencialmente los parámetros estereoquímicos de una estructura estable con 197 (86%) residuos en regiones favorecidas, 28 (12,2%) residuos en regiones permitidas y 4 (1,7%) residuos fuera de regiones favorecidas (**Figura 5**). De acuerdo a la probabilidad de que los pares de ángulos sean posibles dentro de la estructura de la proteína, Las regiones favorecidas corresponden a los sitios donde hay mayor probabilidad de que los ángulos que

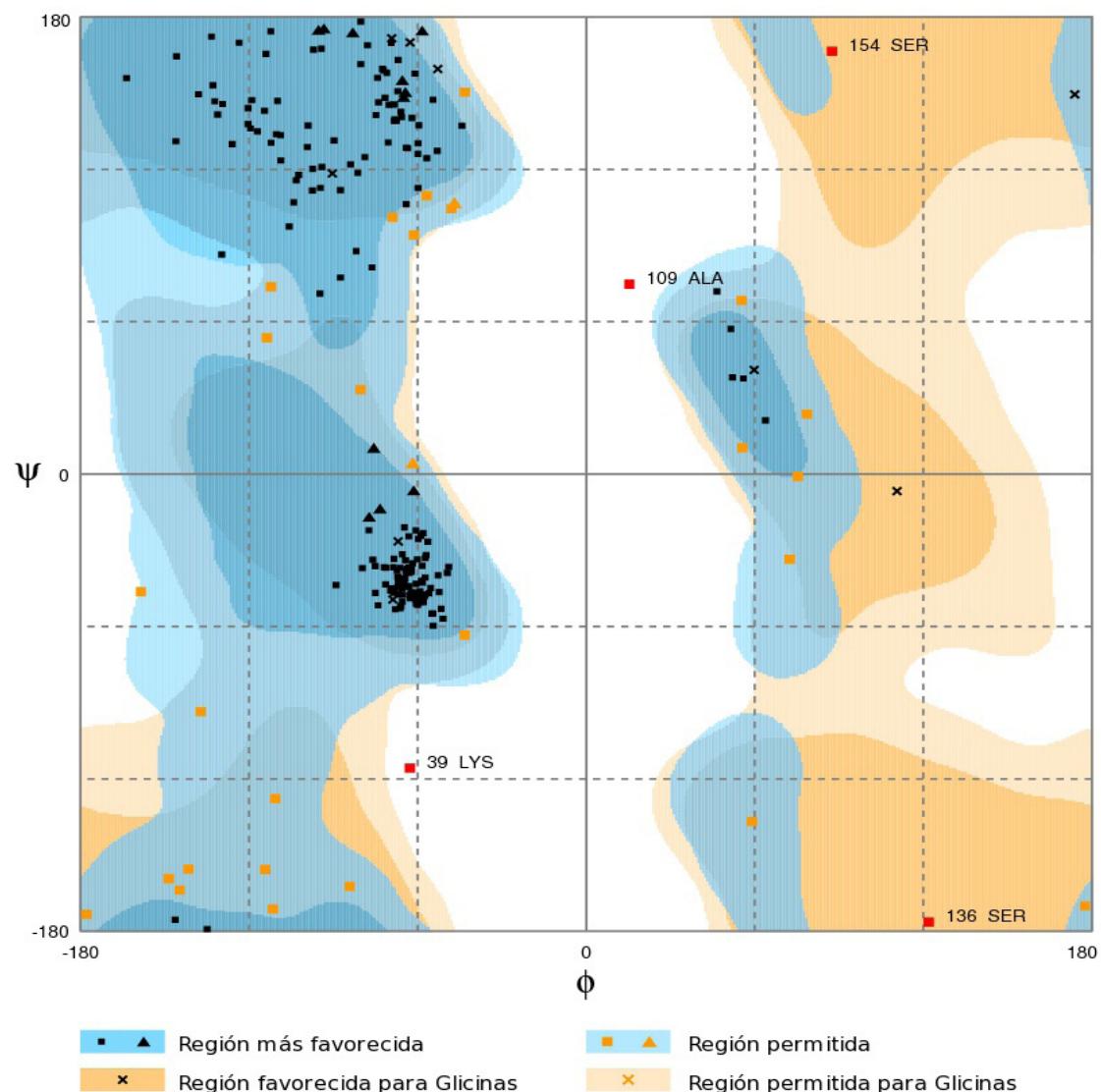


Figura 5. Plot de Ramachandran que muestra la distribución de las combinaciones de los ángulos para la estructura de LIC10494 con alta resolución. El contorno azul oscuro comprende las regiones más favorecidas, el contorno azul claro muestra las regiones permitidas. El contorno naranja y lila muestran las regiones favorecidas y las regiones permitidas para el aminoácido de la Glicina, el cual generalmente se encuentra altamente conservado en las secuencias proteicas y permite la combinación de ángulos que no son posibles para otros aminoácidos.

conforman la estructura proteica sean posibles, mientras que las regiones permitidas y las no permitidas corresponden a sitios con menor probabilidad de acercamiento al ángulo correcto. Por lo anterior y tomando como base los resultados de la evaluación, es posible deducir que el modelo constituye un buen acercamiento a la estructura real de la proteína.

El modelo tridimensional predicho coincide en gran parte de su estructura con la estructura secundaria predicha en este estudio, así mismo coincide con los estudios de hidrofobicidad y desordenes intrínsecos asociados a la secuencia de la proteína (**Figura 6**).

Estos resultados están respaldados por su concordancia con los reportes experimentales de estudios anteriores que se aproximaron a la estructura de la proteína pero no revelaron los residuos específicos involucrados en la conformación de las hélices α , las láminas β y las regiones no estructuradas (23, 6, 30). Dentro de las regiones que coinciden con la estructura secundaria y con los demás análisis del presente estudio, están involucrados varios de los patrones funcionales encontrados durante este estudio, tales patrones involucrados en la estructura final son los relacionados con los sitios de N-Glicosilación (entre los aminoácidos 28-39, 178-180 y 206-209) y fosforilación por enzimas quinazas entre los

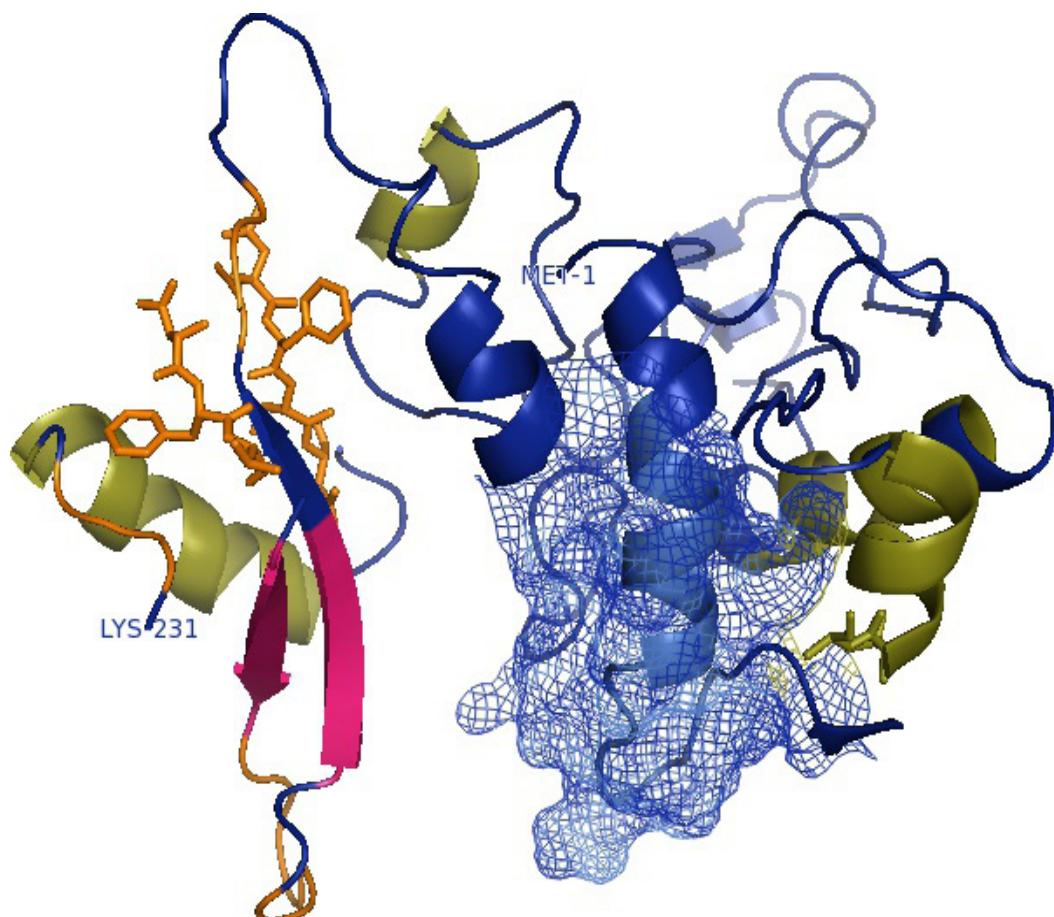


Figura 6. Modelo tridimensional generado por MUSTER y visualizado con PyMol v. 0,98. Las hélices color verde (residuos 28-36; 60-69; 146-150; 214-226) y las láminas color magenta (182-186; 196-200) corresponden a las regiones donde hay coincidencia con la aproximación a la estructura secundaria de LIC10494. Las regiones no estructuradas color naranja (187-193; 228-230) corresponden a los segmentos que coinciden con la aproximación a la estructura secundaria y a las regiones desordenadas predichas por GLOBPLOT. Las regiones mostradas como “sticks” corresponden a los sitios de N-Glicosilación que coinciden tanto con la estructura secundaria como con la estructura terciaria y con los análisis de GLOBPLOT reportados en este estudio (175-180; 204-209). La región color azul claro rodeada por una malla corresponde a la región transmembrana reportada en este estudio por consenso de 9 programas (8-22).

aminoácidos 28-30, 84-86, 191-193, 208,209, 219-222). La predicción de dichos residuos, generada por este estudio, es importante para el conocimiento de las regiones potencialmente importantes para la patogenicidad e inmunogenicidad de la bacteria y por tanto para el futuro desarrollo de estudios enfocados hacia la creación de vacunas a largo plazo efectivas contra varios serovares y varias especies leptospirales, pues la conservación de la proteína LIC10494 ha sido reportada experimentalmente para varios serovares (20, 6), además el encuentro de homología de LIC10494 con la lipoproteína putativa LA3735 de *Leptospira interrogans* serovar *Lai* permite generar una aproximación importante a la estructura de ésta última. Aunque se reportan varios patrones funcionales importantes para la antigenicidad e inmunogenicidad de la bacteria y estos se encuentran claramente asociados con la estructura tridimensional predicha en el presente estudio, aún faltan estudios que investiguen a fondo la importancia de estos patrones y los posibles mecanismos de inhibición de los sitios de N-Glicosilación para desarrollo de antibióticos y/o vacunas efectivas contra la enfermedad. También merece especial atención llevar a cabo estudios que revelen el mecanismo exacto por el cual cada uno de los patrones, reportados en este estudio, intervienen en la funcionalidad de LIC10494 como antígeno para entender el mecanismo de interacción antígeno-anticuerpo generado por LIC10494 en la respuesta inmune de ratones infectados con *Leptospira interrogans* serovar *Copenhageni*.

Conclusiones

La asociación de la estructura, establecida en el presente estudio, con las regiones funcionales permite proponer un modelo tridimensional, para la proteína LIC10494, que coincide con estudios experimentales previos y puede usarse para inferir varios aspectos del comportamiento de la misma en cuanto a la patogenicidad e inmunogenicidad de *Leptospira interrogans* serovar *Copenhageni*.

Financiación

Agradecemos al Departamento de Nutrición y Bioquímica de la Pontificia Universidad Javeriana por el soporte académico e infraestructura en la realización de este trabajo.

Conflictos de Intereses

No existen conflictos de interés en la realización de este trabajo.

Referencias

1. Adler B, De la Peña A. Leptospira and leptospirosis. *Veterinary Microbiology* 2010; **140** (3-4): 287-296.
2. Thongboonkerd V. Proteomics in leptospirosis research: towards molecular diagnostics and vaccine development. *Expert Review of Molecular Diagnostics* 2008; **8** (1): 53-61.
3. Nam HJ, Jeon J, Kim S. Bioinformatic approaches for the structure and function of membrane proteins. *BMB Reports* 2009; **42** (11): 697-704.
4. Chang YF, Chen CS, Palaniappan R, He H, McDonough SP, Barr SC, Yan W, Faisal SM, Pan MJ. Immunogenicity of the recombinant leptospiral putative outer membrane proteins as vaccine candidates. *Vaccine* 2007; **25** (48): 8190-8197.
5. Faisal SM, Yan W, Chen CS, Palaniappan R, McDonough SP, Chang YF. Evaluation of protective immunity of *Leptospira* immunoglobulin like protein A (LigA) DNA vaccine against challenge in hamsters. *Vaccine* 2008; **26** (2): 277-287.
6. Atzingen MV, Goncales AP, de Moraes ZM, Araújo ER, De Brito T, Vasconcellos SA, Nascimento A. Characterization of leptospiral proteins that afford partial protection in hamsters against lethal challenge with *Leptospira interrogans*. *Journal of Medical Microbiology* 2010; **59** (9): 1005-1015.
7. Cherezov V, Abola E, Stevens RC. Recent progress in the structure determination of GPCRs, a membrane protein family with high potential as pharmaceutical targets. *Methods in Molecular Biology* 2010; **654**, 141-168.
8. Xu Y, Xu D, Liang D. Computational Methods for Protein Structure Prediction and Modeling. *Springer*. Springer. USA. 2007, 335 p.
9. Pruitt K, Tatusova T, Maglott D. NCBI Reference Sequence (RefSeq): a curated non redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research* 2007; **35**, D61-65.
10. Alterovitz G, Benson R, Ramoni M. Automation In Proteomics and Genomics. an Engineering Case-Based Approach. John Wiley & Sons. New Delhi., India. 2009, 340 p.
11. Imer O, Cavas L. The Bioinformatics Tools for the Estimation of Disordered. *IEEE* 2009; 1-4.
12. Sigrist C, Cerutti L, Castro E, Langendijk GP, Bulliard V, Bairoch A, Hulo N. PROSITE, a protein domain database for functional characterization and annotation. *Nucleic Acids Research* 2010; **38** (1): D161-D166.

13. Wu S, Zhang Y. MUSTER: Improving protein sequence profile-profile alignments by using multiple resources of structure information. *Proteins* 2008; **72**, 547-556.
14. Scott M, Ritterson R, Dill, KA. A Test on Peptide Stability of AMBER Force Fields with Implicit Solvation. *The Journal of Physical Chemistry* 2008; **112** (22): 6878-6886.
15. Schrödinger, LLC. Maestro, version 8.0, Schrödinger, LLC: New York, NY
16. Claverie J, Notredame, C. Bioinformatics for Dummies. Segunda Edición. Wiley Publishing, Inc. Indianapolis., Indiana. 2007, 457 p.
17. Khan FH. The Elements of Immunology. Pearson Education. Nueva Delhi. 2009, 473 p.
18. Arancibia S, Del Campo M, Nova E, Salazar F, Becker M. Enhanced structural stability of Concholepas hemocyanin increases its immunogenicity and maintains its non-specific immunostimulatory effects. *European Journal of Immunology* 2012; **42** (3): 688-699.
19. Werner MS. Biología: Fundamentos para Medicina y Ciencias de la Vida. Reverté. Barcelona, España. 2007, 661p.
20. Jinling C, Dandan Z, Pei S, Wei S, Gengfu X, Yinong D. Bioinformatics analysis on ORF1 protein of Torque teno virus (SANBAN isolate). *Asian Pacific Journal of Tropical Medicine* 2011; **4** (11): 850-856.
21. Hooda V. Physicochemical, Functional and structural characterization of wheat germin using in silico methods. *Current Research Journal of Biological Sciences* 2011; **3** (1): 35-41.
22. Lieberman M, Marks AD, Smith CM. Marks' basic medical biochemistry: a clinical approach. Tercera edición. Wolters Kluwer Health. China. 2009, 997 p.
23. Saha, B. Antigens. eLS. John Wiley & Sons Ltd. Ganeshkhind Pune, India. 2010, 8 p.
24. Atzingen MV. Expressao, purificacao e caracterizacao de proteínas de superficie de Leptospira interrogans. **Tesis Doctoral**. Instituto de Ciencias Biomedicas. Universidad de Sao Paulo, Sao Paulo. 2009. 166p.
25. Uversky VN, Dunker AK. Understanding protein non-folding. *Biochimica et Biophysica Acta* 2010; **1804** (6): 1231-1264.
26. Uversky VN, Radivojac P, Lakoucheva LM, Obradovic Z, Dunker AK. Prediction of intrinsic disorder and its use in functional proteomics. *Methods in Molecular Biology* 2007; **408** (1): 69-92.
27. Hoyer LL, Green CB, Oh SH, Zhao X. Discovering the secrets of the *Candida albicans* agglutinin-like sequence (ALS) gene family- a sticky pursuit. *TOC* 2008; **46** (1): 1-15.
28. Li T, Li F, Zhang X. Prediction of kinase-specific phosphorylation sites with sequence features by log-odds ratio approach. *Bioinformatics* 2008; **70** (2): 404-414.
29. Pollastri G, Martin AJM, Mooney C, Vullo A. Accurate prediction of protein secondary structure and solvent accessibility by consensus combiners of sequence and structure information. *BMC Bioinformatics* 2007; **8**, 201.
30. Barbosa AS, Abreu AE, Neves FO, Atzingen MV, Watanabe MM, Vieira ML, Morais, ZM, Vasconcellos SA, Nascimento A. A newly identified leptospiral adhesin mediates attachment to laminin. *Infection and Immunity* 2006; 6356-6364.
31. Tosatto S, Battistutta R. TAP score: torsion angle propensity normalization applied to local protein structure evaluation. *BMC Bioinformatics* 2007; **8**, 155

Direcciones de Internet

- DeLano WL. The PyMOL molecular graphics system. <http://www.pymol.org>. Consultado el 18 de Julio de 2011.
- De Bakker P, Lovell S. RAMPAGE. <http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>. Consultado el 23 de Julio de 2011.
- Sapay N, Guermeur Y, Deleage G. Pole Bio-Informatique Lyonnais. Network protein Sequence @nalysis. Consensus Secondary Structure Prediction. <http://www.bioinf.manchester.ac.uk/dbrowser/bioactivity/NPS2.html>. Consultado el 14 de Mayo de 2011.
- Swiss Institute of Bioinformatics. <http://expasy.org/tools/>. Consultado el 13 de Febrero de 2011.
- WHO. International Leptospirosis Society. www.who.org. Consultado el 20 de Marzo de 2011
- Instituto Nacional de Salud. Sistema Nacional de Vigilancia. www.ins.gov.co. Consultado el 20 de Marzo de 2011.