

# Predicción de unión de péptidos de MSP-1 y EBA-140 de *Plasmodium falciparum* al HLA clase II

## Probabilidad, combinatoria y entropía aplicadas a secuencias peptídicas

Javier Rodríguez, Pedro Bernal, Luisa Álvarez, Sandra Pabón, Sandra Ibáñez, Nanyid Chapuel, Héctor Pérez, Alexandra Correa, Luis Carlos Salazar, Raúl Walteros

GRUPO INSIGHT

Universidad Nacional de Colombia, Departamento de Fisiología, Línea de profundización de física y matemáticas aplicadas a la fisiología, Colombia.

*PLASMODIUM FALCIPARUM MSP-1 AND EBA-140 PEPTIDES PREDICTION OF BINDING TO HLA CLASS II PROBABILITY, COMBINATORY AND ENTROPY APPLIED TO PEPTIDE SEQUENCES*

Recibido: 25 mayo 2010

Aceptado: 2 de julio de 2010

### RESUMEN

La proteína de superficie del merozoíto MSP-1 está asociada al fenómeno de invasión de la malaria al glóbulo rojo, mientras que la proteína del antígeno EBA-140 interactúa con las sialo-glicoproteínas de la superficie de los eritrocitos. La identificación de péptidos de estas proteínas que se unan al HLA clase II es de vital importancia para el desarrollo de vacunas.

En nuestro trabajo se aplicó una teoría predictiva de unión al HLA clase II basada en la proporción  $S/k$  de la entropía, para la predicción del fenómeno de unión de péptidos de las proteínas MSP-1 y EBA-140 a la totalidad de secuencias de 20 aminoácidos de dichas moléculas. Se calcularon los valores de probabilidad, combinatoria y entropía de 948 secuencias noámeras sobrelapadas de las proteínas MSP-1 y 732 de EBA-140. Se construyeron computacionalmente tres proteínas teóricas de 500 aminoácidos de longitud cada una, para aplicar la teoría de unión desarrollada en todos los péptidos noámeros sobrelapados de las mismas.

Se predijo que para las dos proteínas del merozoíto estudiadas, 298 secuencias están asociadas al macroestado de unión y 1409 al macroestado de no unión. La predicción teórica desarrollada puede facilitar la selección de péptidos implicados en el desarrollo de vacunas. Para las tres proteínas construidas teóricamente se encontró que las secuencias 111, 84 y 72 son predichas como pertenecientes al macroestado de unión, mientras que las restantes, 381, 408 y 420 se predicen asociadas al macroestado de no unión.

Las predicciones evidencian un orden físico y matemático en la presentación antigénica, útil en el desarrollo de vacunas.

### ABSTRACT

The MSP-1 merozoite surface protein is related to the invasion phenomenon of malaria to red cells, while the EBA-140 antigen protein interacts with the sialoglycoproteins on the surface of erythrocytes. The identification of peptides from both proteins that bind to HLA class II has a particular importance for the development of vaccines.

In the present work, a predictive binding to HLA class II theory, based on  $S/k$  proportion of entropy was applied, for the prediction of proteins MSP-1 and EBA-140 binding phenomenon to the totality of 20 amino acid sequences of both molecules. The probability, combinatorial and entropy values were calculated for 948 and 732 nonamer overlapping sequences of MSP-1 and EBA-140, respectively. Three theoretical proteins of 500 aminoacids of length each were computationally built in order to apply the developed binding theory to all their nonamer overlapping peptides.

It was predicted that for the two studied merozoite proteins, 298 sequences are related to the binding macrostate and 1409 to the not-binding macrostate. The developed theoretical prediction can facilitate the selection of peptides during the process of vaccine development. For the three theoretical proteins built, it was found that 111, 84 and 72 are predicted as included into the binding macrostate, while sequences 381, 408, and 420 are predicted as related to the not-binding macrostate.

The predictions are an evidence of a physical and mathematical order in antigen presentation, which can be useful to the development of vaccines.

KEY WORDS: Binding / Entropy / HLA class II / MSP-1 / EBA-140.

PALABRAS CLAVE: Unión / Entropía / HLA clase II / MSP-1 / EBA-140.

## INTRODUCCIÓN

La malaria constituye un problema de salud pública a nivel mundial, afectando a casi 500 millones de personas al año y generando alrededor de 3.000.000 de muertes solamente en el continente africano<sup>(1,2)</sup>. En este sentido, se han desarrollado gran cantidad de trabajos para conocer el proceso de invasión, que buscan establecer las interacciones ligando-receptor específicas durante todo el proceso, con el fin de desarrollar vacunas<sup>(3)</sup>.

Dentro de los antígenos identificados como potencialmente útiles en el desarrollo de una vacuna contra la malaria, la más estudiada ha sido la proteína Mayor de Superficie del Merozoito MSP-1, sintetizada durante el desarrollo del esquizonte, la cual está presente en la superficie del merozoito como un complejo de fragmentos proteolíticos<sup>(4,5)</sup> al estar directamente involucrada en el proceso de invasión al glóbulo rojo. Investigaciones sugieren que tiene una importante función en las primeras etapas de dicha interacción, siendo por ello un activador de la respuesta inmunitaria humana. Trabajos en este área han demostrado que la activación de la respuesta humoral a diferentes regiones de la proteína disminuye la susceptibilidad a la malaria clínica<sup>(6-9)</sup>. Por otro lado, la EBA-140, una de las proteínas del merozoito de *Plasmodium falciparum* o antígeno de unión eritrocitaria, es también conocida como BAEBL o PfEBP-2; estudios de la localización subcelular sugieren que está localizada en los micronemas del complejo apical, cuyas proteínas interactúan con las sialoglicoproteínas de la superficie de eritrocitos sanguíneos<sup>(10,11)</sup>.

La presentación de antígenos a los linfocitos T es mediada por el HLA: los linfocitos T CD4<sup>+</sup> reconocen los péptidos presentados por el HLA clase II<sup>(12)</sup>. La región central de unión de péptidos del HLA clase II se encuentra abierta en sus extremos y está formada por los dominios  $\alpha 1$  y  $\beta 1$ ; dicha apertura explica la variabilidad en el tamaño de los péptidos presentados<sup>(13-17)</sup>. De manera general se ha encontrado que una región central compuesta por nueve aminoácidos es fundamental para la unión<sup>(18)</sup>.

Dada la importancia del HLA clase II en el proceso de activación antigénica resulta de gran utilidad determinar si existen péptidos de las proteínas asociadas a la invasión del eritrocito por parte del merozoito que puedan ser reconocidos por dicha molécula. Recientemente se desarrolló una teoría para la predicción de péptidos a la región central del HLA clase II<sup>(19)</sup>, con base en las teorías de probabilidad, combinatoria y entropía. Este trabajo partió de una abstracción física teórica basada en la observación de la distribución de aminoácidos en la región central de unión como un fenómeno finito y probabilista donde la distribución de las

probabilidades de aparición de grupos de aminoácidos en las diferentes posiciones permite estudiar matemáticamente las diferencias entre los péptidos de unión y no unión al HLA clase II. Al hacer grupos de aminoácidos, es posible hacer una abstracción de las particularidades de cada péptido para lograr una caracterización más general, y realizar una analogía con los macroestados de la mecánica estadística, en términos del número de posibles formas de organización de moléculas de un gas, que tienen ciertas características comunes por las que son incluidas dentro de un mismo macroestado. Con el fin de establecer el conjunto total de posibilidades del fenómeno, independientemente de la especificidad alélica, en dicho experimento se estudiaron no sólo péptidos sintéticos, naturales y promiscuos, tanto de unión como de no unión, sino lo que ocurriría en el caso de encontrar un péptido teórico con todos los aminoácidos iguales, así como un péptido teórico que presente todos los aminoácidos diferentes<sup>(19)</sup>. De este modo fue posible establecer que los péptidos de unión se caracterizan por ciertas repeticiones y condiciones específicas, al ser estudiada la distribución de grupos de aminoácidos con base en las leyes de probabilidad y combinatoria. Esta caracterización probabilista permitió la evaluación de la autoorganización del sistema con base en la ley de la entropía, estableciendo diferencias matemáticas entre péptidos de unión y no unión con base en la relación S/k. Dicha teoría fue aplicada a 161 péptidos, con una eficacia del 100% en su capacidad de predicción de su estado de unión o no unión a la molécula de HLA II. Posteriormente también fue aplicada en la determinación de péptidos de unión al HLA clase II de las proteínas MSA-2 y AMA-1 de malaria, determinando péptidos que pueden ser útiles en el desarrollo de vacunas<sup>(20)</sup> así como a la proteína API m1 y a dos del HER-2/neu, mejorando medidas de desempeño de los métodos de predicción actuales (Rodríguez J, Bernal P, et al manuscrito en preparación).

Por medio de la probabilidad se puede cuantificar la ocurrencia futura de un evento<sup>(21,22)</sup>, de esta forma la región central fue definida como compuesta por varios eventos probabilísticos con el fin de realizar cuantificaciones que den cuenta de la combinación de los aminoácidos en las secuencias y posteriormente evaluar la probabilidad de la secuencia, respecto a todas las posibles secuencias. La entropía ha recibido varias interpretaciones<sup>(23-26)</sup>, el concepto análogo usado en esta teoría fue el enunciado por Boltzmann, en términos del número de microestados de un sistema equiprobable<sup>(23)</sup>, en donde la entropía de un macroestado es proporcional al número de microestados o configuraciones en las que el macroestado se puede presentar; de esta forma se realiza una analogía entre los péptidos presentados por

la región central y los macroestados y microestados de la mecánica estadística.

En la presente investigación se aplicará la teoría de predicción de unión de péptidos a la región central del HLA clase II<sup>(19)</sup>, a los péptidos noámeros sobrelapados de las 79 secuencias de 20 aminoácidos no sobrelapados que abarcan la proteína MSP-1 y a los péptidos noámeros sobrelapados de las 61 secuencias de 20 aminoácidos no sobrelapados que abarcan la proteína EBA-140.

**Definiciones**

*Macroestado:* conjunto de secuencias que presentan alta unión o no unión al HLA clase II. Por lo tanto existen dos clases de macroestados, asociados a cada una de estas condiciones específicas<sup>(19)</sup>.

*Microestado:* Toda secuencia específica de nueve aminoácidos<sup>(19)</sup>.

*Tipo de secuencia:* Microestados que presenten el mismo valor en su combinatoria y su evaluación con los criterios definidos<sup>(19)</sup>.

*Probabilidad Laplaciana:* La probabilidad de un tipo de secuencia A es definida como la cantidad de microestados N<sub>A</sub> asociados a este tipo de secuencia dividida entre el total de posibles microestados N<sup>(24,19)</sup>:

$$P(A) = \frac{\text{Microestados asociados al tipo de secuencia A}}{\text{Todos los Microestados}} = \frac{N_A}{N}$$

Ecuación 1

*Ley combinatoria para determinar la cantidad de microestados ω:* De acuerdo con la metodología desarrollada por Rodríguez<sup>(19)</sup>, cada posición del péptido tiene asociada un número de posibles aminoácidos. Para calcular el número de posibles microestados asociados a un tipo de secuencia se multiplican los valores asociados a cada posición, que dependen del número de posibles aminoácidos asociados a cada lugar, los cuales a su vez están determinados por el grupo al que pertenezcan y sus repeticiones en el tipo de secuencia. Para las cuantificaciones específicas ver tablas de criterios evaluadores y tablas de grupos de aminoácidos de la teoría realizada<sup>(19)</sup>.

*Entropía:* En un sistema con microestados equiprobables la entropía se define de acuerdo con la ecuación 2:

$$S = KLn(\omega) \tag{Ecuación 2}$$

Donde k es igual a la constante de Boltzmann, 1.38x10<sup>-23</sup> (J/k), ω los posibles microestados y S el valor de la entropía<sup>(19, 23-25)</sup>.

Proporción S/k: Despejando de la ecuación 2 se obtiene<sup>(19)</sup>:

$$S/K = Ln(\omega) \tag{Ecuación 3}$$

**MATERIAL Y MÉTODO**

Inicialmente se determinaron los péptidos noámeros sobrelapados pertenecientes a secuencias no sobrelapadas de 20 aminoácidos, 79 correspondientes a la proteína MSP-1 y 61 a la EBA-140, así como los péptidos noámeros sobrelapados de tres proteínas teóricas compuestas por 500 residuos cada una.

Posteriormente se aplicó la metodología desarrollada por Rodríguez<sup>(19)</sup> para determinar si las secuencias sobrelapadas pertenecen al macroestado de unión o no unión. Para ello inicialmente se define un espacio muestral, donde cada microestado es considerado como un posible evento. Luego se evaluó la probabilidad laplaciana de cada una de las secuencias, aplicando la ley de combinatoria (ver definiciones), para calcular los microestados equivalentes a partir de la cantidad de aminoácidos que tendrían la posibilidad de ocupar cada posición. El número de aminoácidos asociados a cada posición depende del valor del grupo al que pertenezca y sus repeticiones en el tipo de secuencia<sup>(19)</sup>.

Para estos cálculos se realizaron cuatro pasos en donde se agrupan los 20 aminoácidos de forma específica, el procedimiento de cálculo para los pasos uno y tres es asignar a cada posición un número según el aminoácido encontrado en cada lugar dependiendo del conjunto en el que se encuentre; se realiza el mismo procedimiento con cada aminoácido que aparece consecutivamente, considerando además la disminución de sus posibilidades en la medida que van apareciendo aminoácidos de un mismo grupo y las repeticiones de un mismo aminoácido. Posteriormente para calcular la probabilidad laplaciana de un tipo de secuencia (ver definiciones), se multiplican dichos valores hallando el número de microestados asociados al tipo de secuencia<sup>(19)</sup>. Partiendo del número de microestados posibles se realizaron los cálculos de entropía para cada tipo de secuencia aplicando la ley de Boltzmann (Ecuación 2).

La predicción teórica desarrollada se calcula con base en la proporción S/k, ecuación 3, y de los rangos mínimos y máximos asociados a los péptidos de alta unión; de esta forma se encontró previamente que las secuencias noámeras con proporciones S/k entre 8,55 y 12,8 para los primeros pasos y entre 6,98 y 10,5 para los últimos pasos son predichas de unión al HLA clase II de acuerdo con la teoría desarrollada<sup>(19)</sup>.

De este modo, una vez hallado el valor de la entropía de cada secuencia noámera se determina el valor de la proporción S/k y se determina si se encuentra o no dentro de los rangos predichos de unión mencionados previamente (Ecuación 3).

**TABLA I.** Péptidos no superpuestos de 20 aminoácidos de la proteína MSP-1, y secuencias noámeras predichas de unión correspondientes junto con sus valores de la proporción S/k

Secuencia	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K
KIIFLCSFLFYIINTQCVT	FLCSFLFYI	8,70	LCSFLFYII	8,70	FLFYIINTQ	9,80	FYIINTQCV	9,98		
HESYQELVKKLEALEDVLT	YQELVKKLE	8,19								
GYSLFQKEKMLVNEGTSFTA	LFQKEKMLV	8,88	VLNEGTSFT	10,09	LNEGTSFTA	9,57				
VTTSTPGSKGSVASGGSGGG	VTTSTPGSK	9,10	GSKGSVASG	8,45						
VASGGSVASGGSVASGGVA										
GGSGNSRRTPNSDSSDSD	GSNSRRTPN	8,45	SRRTNPSDN	8,96	RRTNPSDNS	9,66				
AKSYADLKHRVRNYLLTIKE	LKHRVRNYL	8,41	VRNYLLTIK	10,49						
LKYPQLFDLTHNMLTCDNI	LKYPQLFDL	9,10	LFDLTHNML	9,10	FDLTHNMLT	9,80				
HGFKYLDGYEINEELLYKL	FKYLDGYE	8,88	LIDGYEIN	8,70	IDGYEINE	7,31				
NFYFDLLRAKLNVDVCANDYC	YFDLLRAKL	7,78	FDLLRAKLN	7,50	LLRAKLNVD	7,50	LRAKLNVD	10,49	LNDVCANDY	10,09
QIPYNLKRANELDVLKLV	IRANELDVL	8,19								
FGYRPLDNIKDNVGMEDY	YRKPLDNIK	8,37	LDNIKDNV	9,57						
IKKNKTIENINELIEESK	ENINELIEE	7,58	INELIEESK	8,41						
TIDKNKNATKEEKKLYQA	IDKNKNATK	8,41								
QYDLSIYNKQLEEAHNLSV	LEEAAHNLS	9,80								
LEKRIDLTKKNENIKELYDK	LEKRIDLTK	7,50	IDTLKKNEN	8,41						
INEIKNPPANSNGTNPNTLL	IKNPPANS	7,50								
DKNKKIEEHEKEIEIAKTI	KEIEIEIAK	8,27								
KFNIDSLFTDPLEEYLLRE	FNIDSLFTD	10,09	IDSLFTDPL	10,09	FTDPLEEY	7,31				
KNKNIDISAKVETRESTEPN	VETRESTEP	8,41								
EYPNGVTYPLSYNDINNALN	YPNGVTYPL	10,49	YPLSYNDIN	10,09	LSYNDINNA	10,09	YNDINNALN	8,48		
ELNSFGDLINPFDYTKESK	LNSFGDLIN	10,09	FGDLINPFD	9,57	FDYTKESK	9,80				
NIYTDNERKKFINEIEKIK										
IEKKIESDKKSYEDRSKSL	KKSYEDRSK	7,35								
NDIT KEYEKLLNEIYDSKFN	ITKEYEKLL	8,41	KEYEKLLNE	8,27	YEKLLNEIY	7,50	LLNEIYDSK	9,80		
NNIDLTFEKMGMGRYSYK	FEKMGMGRY	8,88	MMGRYSYK	8,19	MGGRYSYK	9,80				
EKLTHPNTFASYENSKHNLE	YENSKHNLE	9,10								
KLTKALKYMEDYSLRNIIVVE	LTKALKYME	7,90	LKYMEDYSL	7,50	YMEDYSLRN	9,80	YSLRNIIVVE	10,20		
KELKYKNLISKIENIEIETL	KELKYKNL	8,27	LKYKNLISK	8,41	YKKNLISKI	8,41	LISKIENIE	8,41		
VENIKKDEEQLEFKKITKDE										
NKPDEKILEVSDIVKVQVQK	ILEVSDIVK	7,50	LEVSDIVKV	7,50	VSDIVKVQV	7,50				
VLLMNKIDELKKTQLIYKNV	VLLMNKIDE	8,88	LLMNKIDEL	7,50	IDELKKTQL	7,72	LKKTQLIYK	9,10		
ELKHNIHVNPNSYKQENKQEP	LKHNIHVNP	9,57	YKQENKQEP	8,41						
YYLIVLKKKEIDKLVFMPKV	IDKLVFMPK	9,29								
ESLINEEKKNYKTQGSQSDNS										
EPSTEGETGQATTKPGQQA	ITGQATTKP	10,20								
GSALEGDSVQAQAQEQKQAQ	LEGDSVQAQ	9,80								
PPVPPVPEAKAKVPTPPAP	PVPEAKAKV	7,35								
VNNKTENVSKLDYLEKLYQF	VNNKTENV	8,19								
LNTSYICHKYILVSHSTMNE	YICHKYILV	8,59	ICHKYILVS	10,49						
KILQYKITEEESKLSKCD	ILQYKITEK	9,10	EESKLSKCD	8,67						
PLDLLFNIQNNIPVMSMFD	LDLLFNIQN	8,01	LLFNIQNNI	7,31	IQNNIPVMY	10,09	IPVMSMFD	10,49		
SLNNSLSQLFMEIYKEMVC	LNNSLSQLF	8,01	NNSLSQLFM	10,17	FMEIYKEM	7,50	MEIYKEMV	7,50		
NLYLKDNDKIKNLEEAKK	YKLKDNDKI	8,19								
VSTSVKTLSSSSMQPLSLTP	VSTSVKTLS	7,72	VKTLSSSSM	7,50						
QDKPEVSANDDTSHSTNLNN	VSANDDTSH	9,80								
SLKLFENILSLGKNKNIYQE	LKLFENILS	9,10	LFENILSLG	9,39	ILSLGKNKN	8,41	LSLGKNKNI	9,10	LGKNKNIYQ	10,49
FYNESFTNFVSKADDINSL	FYNESFTNF	8,19	YNESFTNFV	9,57	FTNFVKSKA	10,49	FVKSADDI	9,29	VKSADDIN	9,80
NDESKRKKLEEDINKLKKTL										
QLSFDLYNKYKLERLFDK	LSFDLYNKY	9,10	FDLYNKYKL	7,72	KLKLERLFD	7,17				
KKTGVGKYKMQIKLTLLEQ	VGVGKYKMQ	10,49	YKMQIKLTL	10,49	MQIKLTLLE	8,41				
LESKLSLNNPKHVLQYPSV	LNSLNNPKH	8,19	LNNPKHVLQ	9,80						
FYNKKKEAIEAETENTLENT	EAEIAETEN	9,19	AIEAETENT	8,27	IAETENTLE	8,70				
KILLKHYKGLVKYNGESSP	ILLKHYKGL	7,50	LLKHYKGLV	7,50	LKHYKGLVK	8,19	YKGLVKYNG	8,19	VKYNGESS	9,80
LKTLSEESIQTEDNYASLEN	LKTLSEESI	8,41	LSEESIQTE	8,70						
FKVLSKLEGKLDNLEK	FKVLSKLEG	9,57	VLSKLEGK	8,19						
KLSYLSRGLHHLIAELKEVI	KLSYLSRGL	9,66	LSYLSRGLH	8,19	YLSRGLHHL	8,19	LSRGLHHLI	8,19	SRGLHHLIA	10,06
KNKNYTGNSPSVNNTDVNN	VNNTDVNN	7,86	VNNTDVNNA	7,31						
LESYKFLPEGTDVATVVE	LESYKFLP	8,41								
SGSDTYEQSQPKKPASTHVG	YEQSQPKKP	8,41								
AESNTITTSQNVDDEVDDVY	SNTITTSQN	8,27								
IVLIFGESEEDYDDLQVVT	IFGESEEDY	8,70	FGSESEEDY	7,31	YDDLQVVT	10,09				
GEAVTTSVIDNILSKIENEY	VIDNILSKI	9,10	ILSKIENEY	9,80						

**TABLA I.** Continuación

Secuencia	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K
EVLYKPLAGVYRSLKKQLE	VLYLKPLAG	8,59	LYLKPLAGV	8,59	YLKPLAGVY	8,19	LKPLAGVYR	9,98	VYRSLKKQL	9,80	YRSLKKQLE	8,41
NNVMTFNVNVDILNSRYNK	VMTFNVNVK	8,41	FNVNVKDIL	8,59	VNVKDILNS	9,80						
RENFKNVLESDLIPYKDLTS	FKNVLES DL	9,80	VLES DLIPY	9,39	LES DLIPYK	8,41	LIPYKDLTS	9,80				
SNYVVKDPYKFLNKEKRDKF	VVKDPYKFL	8,88	VKDPYKFLN	9,80								
LSSYNYIKDSIDTININFAND	IDTININFAN	8,70										
VLGYKILSEKYKSDLDSIK	VLGYKILS	9,80	LGYYKILSE	9,10	YYKILSEKY	7,50	YKILSEKYK	8,19	YKSDLDSIK	8,19		
KYINDKQGENEKYLPFLNNI												
ETLYKTVNDKIDLFVHLEA	VNDKIDLFV	8,59										
KVLNYTYEKSNEVKIKELN	LNYTYEKS N	9,29	YTYEKSNE	9,57	VEVKIKELN	8,19						
YLKTIQDKLADFKNNNFVG	YLKTIQDKL	9,80	LKTIQDKLA	9,80								
IADLSTDYHNHNNLLTKFLST	LSTDYHNHN	9,29	YHNHNNLLTK	8,41								
GMVYENLLKSVLSNLLDWKL	MVYENLLKS	10,49	VYENLLKSV	9,10	YENLLKSVL	9,10	LLKSVLSNL	7,03	LSNLLDWKL	7,72		
ARYVKHFTTPMRKKTMIQQS	FTTPMRKKT	8,19	MRKKTMIQQ	8,41								
SGCFRHLDEREECKCLLNYK												
QEGSKCVENS NPTCNENNGG	VENS NPTCN	9,80										
CDADAKCTEEDSGSNGKKIT	EEDSGSNGK	8,67										

**RESULTADOS**

De los 79 péptidos de 20 aminoácidos analizados de MSP-1, se encontraron 179 secuencias no ámeras que corresponden al macroestado de unión, mientras que 769 corresponden al macroestado de no unión (Tabla I). De los 61 péptidos de 20 aminoácidos analizados para EBA-140, se encontraron 119 secuencias no ámeras que corresponden al macroestado de unión, mientras que 613 corresponden al macroestado de no unión (Tabla II). Las proporciones S/k de los péptidos asociados al macroestado de unión oscilaron entre 7,03 y 10,49 para MSP-1 y entre 6,98 y 10,49 para EBA-140 (Tablas I y II).

Para los 492 péptidos analizados en cada una de las proteínas teóricas, se encontró que 111, 84 y 72 corresponden respectivamente al macroestado de unión de cada una, mientras que las restantes 381, 408 y 420 respectivamente corresponden al macroestado de no unión (Tabla III). El porcentaje de unión respecto a la totalidad de péptidos corresponde al 22,2% para la secuencia teórica 1, al 16,8% para la 2 y al 14,4% para la 3.

**DISCUSIÓN**

Este es el primer trabajo donde se realiza una predicción de la totalidad de péptidos no ámeros de las proteínas MSP1 y EBA 140 de *Plasmodium falciparum* que pueden unirse a la región central de unión al HLA clase II con base en una teoría basada en las leyes de probabilidad, combinatoria y entropía. Se determinó que existen 179 péptidos incluidos dentro del macroestado de unión y 769 que no se unen para la proteína MSP-1, mientras que se hallaron 119 péptidos asociados al microestado de unión y 613 al de no unión para el

caso del antígeno EBA-140, péptidos que deberán ser sometidos a validación experimental. Además se predijeron péptidos de unión en tres proteínas teóricas, 111, 84 y 72 respectivamente, con base en la metodología implementada; con esto se corrobora una alta especificidad característica del fenómeno de unión, útil en la predicción de péptidos al HLA clase II, a partir de leyes físicas y matemáticas.

Existen cuantificaciones de la unión de péptidos al HLA en trabajos anteriores en predicción de unión de péptidos a partir de ANN (Artificial Neural Networks), programación lineal, o vectores de máquina<sup>(27-29)</sup>. Tales métodos buscan patrones de unión caracterizando las posiciones y sus respectivos motivos; su generalización es limitada pues depende de la muestra usada. En contraposición, la teoría aquí utilizada no depende de la población de estudio al estar basada en una abstracción físico teórica que permite cuantificar desde teorías físicas y matemáticas la totalidad de posibilidades existentes de péptidos. Su aplicación permitirá una disminución significativa de los costos necesarios para la determinación de la capacidad de unión de péptidos con el método que Berzofsky denominó “de fuerza bruta”<sup>(30)</sup>, el cual se basa en la evaluación in vitro de la capacidad de estimulación de respuestas en linfocitos T por parte de péptidos sobrelapados que son sintetizados en el laboratorio. La teoría planteada en este trabajo economiza tiempo y simplifica el procedimiento hasta ahora empleado, ya que permite evaluar todas las secuencias no ámeras sobrelapadas para cualquier secuencia proteica, sustituyendo la utilización de los procesos experimentales de ensayo y error por predicciones numéricas rigurosas, evitando la realización de un análisis experimental de la totalidad de las secuencias no ámeras sobrelapadas del MSP-1 y de EBA-

**TABLA II.** Péptidos no superpuestos de 20 aminoácidos de la proteína EBA-140, y secuencias noámeras predichas de unión correspondientes junto con sus valores de la proporción S/k.

Secuencia	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K	Péptidos de unión	S/K
MKGYFNIYFLIPLIFLYNVI	MKGYFNIYF	9,80	YFNIYFLIP	8,19	FNIYFLIPL	8,70	LIPLIFLYN	8,70	IPLIFLYNV	9,80		
RINESIIGRTLYNRQDESSD	INESIIGRT	9,80	IIGRTLYNR	10,49	YNRQDESSD	8,19						
ISRVNSPELNNNHKTNIYDS	ISRVNSPEL	10,49	VNSPELNNN	7,09								
DYEDVNNKLINSFVENKSVK	YEDVNNKLI	9,29	VNNKLINSF	10,20	INSFVENKS	9,29						
KKRSLSFINNKTYSYDIIPP	FINNKTYSY	9,98	INNKTYSYD	9,29								
YSYRNDKFNLSSENEEDNSG	YSYRNDKFN	8,88	YRNDKFNLS	8,59	FNSLSENEE	7,31	LSENEEDNSG	7,09				
NTNSNFNANTSEISIGKDNK	NSNNFANTS	8,78										
QYTFIQKRTHLFACGIKRKS	FIQKRTHLF	9,80										
IKWICRENSEKITVCPDRK	ICRENSEKI	8,88										
IQLCIANFLNSRLETMEKFK	IQLCIANFL	10,09	LCIANFLNS	9,39	IANFLNSRL	10,49	LNSRLETME	9,57				
EIFLSVNTAKLLYNKNEG	IFLSVNTAK	10,49										
KDPSIFCNELRNSFSDFRNS	FCNELRNSF	9,80	LRNSFSDFR	8,41								
FIGDDMDFGGNTDRVKGYN												
KKFSDYYKEKNVEKLNNIKK	FSDYYKEKN	8,19	YKEKNVEKL	8,19	KEKNVEKLN	8,27						
EWWEKKNKANLWNHMIVNHKG	EWWEKKNKAN			8,67	WWEKKNKANL	8,19	WEKKNKANLW	8,19	LWNHMIVNH	9,29		
NISKECAIIPAEQPQLNLI	ISKECAIIP	9,80	IPAEQPQLN	8,70								
KEWVENFLMEKKRFLFLNIKD	KKRFLFLNIK	9,66										
KCVENKKYEAACFGGCRPCS	VENKKYEAC	9,80	YEAACFGGCR	9,57								
SYTSFMKSKTQMEVLTNLY	FMKSKTQME	9,10	MKSKTQME	8,19								
KKKNSGVDKNNFLNDFLFFKN	VDKNNFLND	8,19	DKNNFLNDF	8,67								
NKNDLDDFFKNEKEYDDLCD	DDFFKNEKE	7,06										
CRYTATIISFLNGPAKNDV	YTATIISFL	10,20										
DIASQINVNDRGFCGNYKS												
NNEKSWNCTGTFITNKPFGTC	WNCTGTFITN	9,39	FTNKPFGTC	10,49								
EPPRRQTLCLGRTYLLHRGH	RRQTLCLGR	10,35	LCLGRTYLL	8,19	LGRTYLLHR	8,19	GRTYLLHRG	9,14				
EEDYKEHLLGASIEAQLLK	EEDYKEHLL	7,76	YKEHLLGAS	9,98	LLGASIEA	9,80						
YKYKEKDENALCSIIQNSYA	LCSIIQNSYA	9,80										
DLADIIKGSIIKDYGGKKM	LADIIKGS	9,29	IKDYGGKKM	8,19								
EENLNKVNKDKKRNEESLKI												
FREKWWDENKENVWKMVASV	FREKWWDEN			7,27								
LKNKETCKDYDKFQKIPQFL	YDKFQKIPQ	9,29										
RWFKEWGDDFCEKRKEKIYS	FKEWGDDFC	8,19										
FESFKVECKKDCDENTCKN	FESFKVECK	8,19	KKDCDENTC	8,05								
KCSEYKWKIDLKKSEYEKQV	YKWKIDLKK	7,50	WIDLKKSEY	9,80	LKKSEYEKQ	8,19						
DKYTKDKNKKMYDNIIDEVKN	YTKDKNKKM	8,41	DKNKKMYDN	8,67	MYDNIIDEVK	9,80	YDNIIDEVKN	7,68				
KEANVYLKEKSECKDVFND	VYLKEKSEK	8,41	YLKEKSECK	8,41								
DKIFNESPNEYEDMCKKCDE	IFNESPNEY	10,09	FNESPNEYE	8,70								
IKYLNEIKYPKTKHDIYDID	IKYLNEIKY	7,72	YLNEIKYPK	8,19	YPKTKHDIY	8,88						
TFSDTFGDGTPIISINANINE	FSDTFGDGT	7,31	FGDGTPISI	10,09	ISINANINE	7,31						
QQSGKDTNNTGNSSETSDSPV												
SHEPESDAAINVEKLSGDES	VEKLSGDES	8,19										
SSETRGILDINDPSTNNVN	ILDINDPSV	10,09	INDPSTNN	10,09								
EVHDASNTQGSVNTSDITN	VSNTSDITN	8,48										
GHSESSLNRTTNAQDIKIGR												
SGNEQSDNQENSSHSDNSG												
SLTIGQVPSDNTQNTYDSQ												
NPHRTPNALASLPSDDKIN	ALASLPSDD	7,53	LASLPSDDK	7,68								
EIEGFDSSRDSENGRGDTTS	SSRDSENGR	7,58	SRDSENGRG	7,58								
NTHDVRRTNIVSERRVNSHD	VRRTNIVSE	9,57	IVSERRVNS	8,19	VSERRVNSH	7,50						
FIRNGMANNNAHHQYITQIE	FIRNGMANN	9,98	IRNGMANNN	6,98								
NNGIIRGQEEASAGNSVNYKD	IIRGQEEAS	9,80	IRGQEEASG	9,80								
NPKRSNFSSENDHKKNIQYEQ	KRSNFSSEN	8,96										
NSRDTKRVREEIHLKSKQNK	RVREEIHLK	8,56	VREEIHLK	8,88	IHLKSKQNK	8,41						
CNNEYSMEYCTYSDERNSSP	YSMEYCTYS	8,70	YCTYSDERN	9,29	YSDERNSSP	8,41						
GPCSREERKLCQISDYCL	ERKLCQISDY	8,58	LCCQISDYC	9,80								
KYFNFYSIEYINCISEKSEIS	YFNFYSIEY	7,50	FNFYSIEYY	7,78	FYSIEYYNC	8,88	YSIEYYNCI	7,78	IEYYNCISK	9,29	YYNCISEK	9,57
PEYKCFKSEGSIPYFAAG	YKCFKSEGS	9,80										
GILVVVILLSSASRMGKSN	LLSSASRMG	7,68	LSSASRMGK	8,37								
EYDIGESNIEATFEENNYL	YDIGESNIE	8,70	IGESNIEAT	10,09	IEATFEENN	8,70						
NKLSRIFNQEVQETNISDY	FNQEVQETN	8,48										
VQETNISDYSEYNYNEKNMY	ISDYSEYNY	7,09										



Otras investigaciones desarrolladas desde una perspectiva física y matemática acausal han producido resultados efectivos en áreas como la biología molecular, la morfología celular y la morfología y fisiología cardíaca. En el caso de la inmunología se han desarrollado caracterizaciones del fenómeno de unión de la proteína MSP-1 al receptor de glóbulo rojo, realizadas con la teoría de conjuntos<sup>(36)</sup>, y posteriormente con las teorías de probabilidad y entropía<sup>(37)</sup>, que permiten una diferenciación matemática objetiva de péptidos de unión y no unión. Del mismo modo se realizó una caracterización del fenómeno de unión de péptidos nonaméricos al HLA clase II con la teoría de conjuntos<sup>(38)</sup>. En esta línea de trabajo, también se desarrolló previamente una caracterización de péptidos de alta unión de la proteína EBA-140 (39), en el que se logró diferenciar péptidos de alta unión y de no unión con 100% de acierto para los casos estudiados; este resultado evidencia la existencia de un orden subyacente en el fenómeno de unión, que se puede caracterizar de una manera objetiva y reproducible a partir de leyes físicas y matemáticas.

Es así como la presente investigación hace parte de una serie de trabajos fundamentados en una teoría desarrollada<sup>(19)</sup>, con la cual se simplifica el fenómeno de unión, en donde a partir de la representación numérica de tipos de secuencias y obviando la especificidad alélica, se hace posible predecir los péptidos de unión, viabilizando el desarrollo de vacunas en ausencia de ensayos de prueba y error experimentales. El descubrimiento de órdenes físicos y matemáticos acausales desde las actuales teorías fundamentales de la física<sup>(40)</sup>, evita la necesidad del uso de grandes cantidades de datos experimentales.

Esta concepción acausal ha permitido obtener resultados en otros campos de la medicina, tales como la dinámica cardíaca, donde se realizó una nueva metodología de evaluación del Holter<sup>(41)</sup>. Recientemente, Rodríguez<sup>(42)</sup> desarrolló un método diagnóstico de aplicación clínica de dicho examen con base en la determinación de órdenes entrópicos proporcionales entre regiones geométricas definidas dentro del atractor. Del mismo modo, el establecimiento de órdenes acausales permitió el desarrollo de diferentes métodos predictivos de la dinámica anual en Colombia de enfermedades como malaria, estudiada mediante el análisis probabilista de rangos de infectados<sup>(43)</sup>, y dengue, mediante una analogía de la dinámica con una caminata al azar probabilista<sup>(44)</sup>; en el primero se desarrolló una predicción con un 100% de acierto y en el segundo se realizó una predicción, que al ser comparada con los casos registrados por el INS se encontró como resultado un valor que corresponde al 90,4 % respecto al valor predicho. La aplicación de la ley de entropía al análisis de la dinámica de malaria en todos

los municipios de Colombia permitió realizar predicciones semanales, estableciendo matemáticamente la aparición de brotes<sup>(45)</sup>.

## AGRADECIMIENTOS

Dedicamos este trabajo a nuestros hijos y al ahijado del primer autor, Joshua.

A los doctores Luis Eduardo Cruz y Miguel Martínez del departamento de fisiología y a la facultad de medicina de la Universidad Nacional por apoyar nuestras investigaciones.

## CORRESPONDENCIA:

Javier Rodríguez Velásquez.

Departamento de Fisiología, Universidad Nacional de Colombia Ctra 79 B # 51-16 Sur Interior 5 Apto 102. Bogotá, Cundinamarca-Colombia.

Teléfono: (57-1) 4527541. E-mail: grupoinight2025@yahoo.es

## BIBLIOGRAFÍA

1. WHO. Data World malaria report. Hallado en: <http://www.who.int/malaria/wmr2008/malaria2008.pdf>. Acceso el 10 de Octubre de 2009.
2. Instituto Nacional de Salud. Asi Vamos en Salud. Hallado en: <http://www.asivamosensalud.org>. Acceso el 9 de Octubre de 2009.
3. Cowman A, Crabb B. Invasion of Red Blood Cells by Malaria Parasites. *Cell* 2006;124:755-766.
4. Blackman MJ, Heidrich HG, Donachie S, McBride JS, Holder AA. A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. *J Exp Med* 1990;17:379-382.
5. Holder AA, Freeman RR. Byosynthesis and processing of a Plasmodium falciparum scizont antigen recognized by immune serum and a monoclonal antibody. *J Exp Med* 1982;156:1528-1538.
6. Woehlbier U, Epp C, Kauth CW, Lutz R, Long CA, Coulibaly B, et al. Analysis of Antibodies Directed against Merozoite Surface Protein 1 of the Human Malaria Parasite Plasmodium. *Infect Immun* Feb 2006;1313-1322.
7. Cavanagh DR, Dodoo D, Hviid L, Kurtzhals JAL, Theander TG, Akanmori BD, et al. Antibodies to the N-terminal block 2 of Plasmodium falciparum merozoite surface protein 1 are associated with protection against clinical malaria. *Infect. Immun* 2004;72: 6492-6502.
8. Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihamma N, et al. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. *Nat Med* 2000;6:689-692.
9. Perraut R, Marrama L, Diouf B, Sokhna C, Tall A, Nabeth P, et al. Antibodies to the conserved C-terminal domain of the Plasmodium falciparum merozoite surface protein 1 and to the merozoite extract and their relationship with in vitro inhibitory anti-

- bodies and protection against clinical malaria in a Senegalese village. *J Infect Dis* 2005;191:264-271.
10. Adams JH, Blair PL, Kaneko O, Peterson DS. An expanding EBL family of Plasmodium falciparum. *Trends Parasitol* 2001;7:297-299.
  11. Gilberger TW, Thompson JK, Triglia T, Good RT, Duraisingh MT, Cowman AF. A novel erythrocyte binding antigen-175 paralogue from Plasmodium falciparum defines a new trypsin-resistant receptor on human erythrocytes. *J Biol Chem* 2003;278:14480-14486.
  12. Wang JH, Reinherz E. Structural basis of T-cell recognition of peptides bound to MHC molecules. *Mol Immunol* 2001;38:1039-1049.
  13. Abbas A, Lichtman A, Pober J. *Inmunología celular y molecular 4ª edición*. McGraw Hill 2002.
  14. Madden DR. The three-dimensional structure of peptide-MHC complexes. *Annu. Rev. Immunol* 1995;13:587-622.
  15. Fremont DH, Hendrickson WA, Marrack P, Kappler J. Structures of an MHC class II molecule with covalently bound single peptides. *Science* 1996;272:1001-1004.
  16. Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC. 3-Dimensional structure of the human class-II histocompatibility antigen HLADR1. *Nature* 1993;364:33-39.
  17. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* 1994;368:215-221.
  18. Rammensee HG, Friede T, Stevanovic S. MHC ligands and peptide motifs: first listing. *Immunogenetics* 1995;41:178-228.
  19. Rodríguez J. Teoría de unión al HLA clase II teorías de Probabilidad Combinatoria y Entropía aplicadas a secuencias peptídicas. *Inmunología* 2008;27(4):151-166.
  20. Rodríguez J, Bernal P, Correa C, Prieto S, Benítez L, Vitery S, et al. Predicción de unión de péptidos de MSA-2 y AMA-1 de Plasmodium Falciparum al HLA clase II. *Inmunología* 2009;28(3):115-124.
  21. Laplace Pierre. *Ensayo filosófico sobre las probabilidades*. Barcelona: Altaya 1995.
  22. Feynman RP, Leighton RB, Sands M. Probabilidad. En: Feynman RP, Leighton RB, Sands M. *Física*. Vol. 1. Wilmington: Addison-Wesley Iberoamericana SA; 1964;6-1:6-16.
  23. Feynman RP, Leighton RB, Sands M. Leyes de la Termodinámica. En: Feynman RP, Leighton RB, Sands M. *Física*. Vol. 1. Wilmington: Addison-Wesley Iberoamericana SA 1964;44-1:44-19.
  24. Matvéev A. *Física molecular*. Moscú: MIR 1987.
  25. Tolman R. *Principles of statistical mechanics*. New York: Dover Publications 1979.
  26. Shannon Claude. *The mathematical theory of communication*. Chicago: University of Illinois Press 1980.
  27. Nielsen M, Lundegaard C, Worning P, Sylvester S, Lamberth K, Buus S, et al. Improved prediction of MHC class I and II epitopes using a novel Gibbs sampling approach. *Bioinformatics* 2004;20:1388-1397.
  28. Southwood S, Sidney J, Kondo A, del Guercio MF, Appella E, Hoffman S, et al. Several Common HLA-DR Types Share Largely Overlapping Peptide Binding Repertoires. *J Immunol* 1998;160:3363-3373.
  29. Brusica V, Rudy G, Honeyman M, Hammer J, Harrison L. Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. *Bioinformatics* 1998;14:121-130.
  30. Meister GE, Roberts CGP, Berzofsky JA, De Groot AS. Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from Mycobacterium tuberculosis and HIV protein sequences. *Vaccine* 1995;13:581-591.
  31. Hunt DF, Michel H, Dickinson TA, Shabanowitz J, Cox AL, Sakauchi K, Appella E. Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. *Science* 1992;256:1817-1820.
  32. Chiciz RM, Urban RG, Gorga JC, Vignali DAA, Lane WS, Strominger JL. Specificity and promiscuity among naturally processed peptides bound to HLA-DR alleles. *J Exp Med* 1993;178:27-47.
  33. Rudensky AY, Preston-Hurlburt P, Hong S-C, Barlow A, Janeway CA. Sequence analysis of peptides bound to MHC class II molecules. *Nature* 1991;353:622-627.
  34. Chiciz RM, Urban RG, Lane WS, Gorga JC, Stern LJ, Vignali DAA, Strominger JL. Predominant naturally processed peptides bound to HLA DR1 are derived from MHC-related molecules and are heterogeneous in size. *Nature* 1992;358:764-768.
  35. Sinigaglia F, Hammer J. Motifs and Supermotifs for MHC Class II Binding Peptides. *J. Exp Med* 1995;181:449-451.
  36. Rodríguez J. Diferenciación matemática de péptidos de alta unión de MSP-1 mediante la aplicación de la teoría de conjuntos. *Inmunología* 2008;27(2):63-68.
  37. Rodríguez J. Caracterización física y matemática de péptidos de alta unión de MSP-1 mediante la aplicación de la teoría de la probabilidad y la entropía. *Archivos de Alergia e Inmunología Clínica* 2008;39(2):74-82.
  38. Rodríguez J. Teoría de conjuntos aplicada a la caracterización matemática de unión de péptidos al HLA clase II. *Rev Cienc Salud* 2008;6(1):9-15.
  39. Rodríguez J, Correa C, Prieto S, Puerta G, Vitery S, Bernal P, et al. Aplicación de la probabilidad y la entropía a la proteína EBA-140. Caracterización matemática de péptidos de alta unión. *Inmunología* 2009;28(2):65-73.
  40. Fernandez Rañada. *Orden y Caos*. Introducción, Barcelona: Prensa científica 1990.
  41. Rodríguez J, Prieto S, Avilán N, Correa C, Bernal P, Ortiz L, Ayala J. Nueva metodología física y matemática de evaluación del holter. *Rev Colomb Cardiol* 2008;15:50-54.
  42. Rodríguez J. Entropía proporcional de los sistemas dinámicos cardiacos. Predicciones físicas y matemáticas de la dinámica cardiaca de aplicación clínica. *Rev Col Cardiol* 2010;17(3):115-129.
  43. Rodríguez J. Dinámica probabilista temporal de la epidemia de malaria en Colombia. *Revista Med* 2009;17(2):214-22.
  44. Rodríguez J, Correa C. Predicción Temporal de la Epidemia de Dengue en Colombia: Dinámica probabilista de la epidemia. *Rev salud pública* 2009;11(3):443-453.
  45. Rodríguez J. Método para la predicción de la dinámica temporal de la malaria en los municipios de Colombia. *Rev Panam Salud Pública* 2010;27(3):211-218.