

Predicción computacional cuantitativa de epítomos de células B

Raúl Isea

Fundación Instituto de Estudios Avanzados, Hoyo de la Puerta, Baruta, Venezuela.

email: risea@idea.gob.ve

El presente trabajo desarrolla una metodología computacional para la predicción cuantitativa de epítomos de células B. Para ello, se definió la función $\langle F \rangle$ que reflejará el valor promedio de los epítomos B predichos tras considerar ocho predictores de epítomos B diferentes, así como los factores estructurales y energéticos de la proteína de donde estos se derivan. La metodología propuesta pudiera ser útil para desarrollar nuevas vacunas contra el dengue y el chikungunya.

Palabras clave: epítomo, célula B, dengue, chikungunya.

Introducción

La predicción de epítomos en células B es una de las técnicas computacionales por excelencia que están siendo empleadas para el diseño racional de vacunas, así como en el desarrollo potencial de métodos de diagnóstico, como se ha explicado en esta serie de trabajos científicos (1-3).

Actualmente, existe una gama de programas de computación capaces de realizar la predicción de epítomos de células B, algunos de los cuales se muestran en la Tabla 1. Estos se agrupan según el tipo de predicción: conformacional o discontinua, y lineal o continua.

Tabla 1. Algunos de los programas computacionales empleados para la predicción de epítomos B.

Predictores conformacionales	Predictores lineales
ElliPro	BepiPred
BEpro, antes conocido como PEPITO	BcePred
DiscoTope	ABCPred
SEPPA	BCPreds

Una recopilación más completa están indicadas en (4, 5)

Los programas mencionados permiten la predicción de epítomos B, pero condicionados porque su valor debe superar un valor de corte (del inglés cutoff) que va a variar en cada programa; es decir, y a modo de ejemplo: el programa BepiPred considera que un aminoácido contribuye a un potencial epítomo B si dicho valor es

igual o superior a 0,7, mientras que el criterio de acuerdo al programa DiscoTope es que el resultado sea menor a -7. En el caso de BcePred y ElliPro, rige que deben superar el valor de corte de 1,9 y 0,7, respectivamente.

Conviene recordar que los epítomos B predichos deben estar expuestos al solvente, por ende, debe ser un criterio importante a la hora de seleccionarlos. Entre los programas que se utilizan para determinar el grado de exposición al solvente, cabe citar NetSurfP (6), Vadar (7) y PoPMuSiC (8). Esta idea no es nueva, y se ha empleado como pauta en múltiples trabajos publicados en la literatura científica (9).

Existen además diversas revisiones centradas en realizar comparaciones de resultados entre los distintos predictores de epítomos B; un ejemplo es el estudio publicado por Kringelum (10) donde se analizan los resultados obtenidos con los programas PEPITO (ahora conocido como BEpro), ElliPro, SEPPA, Epitopia, EPCES y EPSVR (10).

Nosotros vamos a definir una función cuantitativa llamada $\langle F \rangle$ que reflejará información de la frecuencia de aparición de los epítomos B predichos por una gama de programas (tanto lineales como conformacionales), y a su vez expresará las contribuciones debido a factores energéticos y estructurales de la proteína de los cuales ellas se derivan. En ese sentido, se determinará la energía libre de Gibbs, que nos indica cuán estable es un aminoácido con respecto a posibles mutaciones en dicha posición. Esta condición no se ha considerado hasta la fecha para cuantificar un epítomo, pero puede ayudar a describir aquellos epítomos B que provienen de secciones de la proteína propensas a cambios por algún

proceso evolutivo donde se suponga que pueden ser los mejores candidatos a la hora de diseñar una vacuna o método de diagnóstico. De modo que valores muy altos muestran la tendencia a que el aminoácido sea propenso a una mutación. Entre los programas que permiten este estudio figuran por ejemplo PoPMuSiC (11), ENCoM (12) y BAPPL (13).

Por otra parte, se determinará el grado de movilidad de cada uno de los aminoácidos dependiendo del análisis de los modos de vibración normal de los átomos que conforman la proteína. De manera que valores muy altos corresponderán a gran movilidad de los átomos que integran esos aminoácidos. Lógicamente, la movilidad de los átomos en el epítipo no se correlacionará con el valor predicho en el epítipo, pero permitirá reconocer regiones lábiles con respecto a las de menor movilidad. Entre los programas empleados para el cálculo de los modos de vibración normal se cuentan elNémo (14) y WEBnm@ (15).

Recordemos que elNémo obtiene los cinco modos normales de baja frecuencia de la proteína para finalmente fijar un coeficiente de correlación entre los factores B observados en la estructura tridimensional de la proteína, y los predichos por esta metodología computacional. El coeficiente debe ser mayor a 0,6 para ser significativo; y es justamente este valor el que determinará la selección de aquellos epítipos que presenten los valores menores.

Este procedimiento se aplicará en dos proteínas independientes entre sí, que han sido empleadas como candidatas a vacuna para posiblemente afrontar la fiebre causada por el virus de chikungunya y la producida por el virus del dengue. Recordemos que el dengue y el chikungunya, transmitidos por la picadura de mosquitos infectados, el *Aedes aegypti* y el *Aedes albopictus*, presentan similar sintomatología. Ambos virus son del tipo arbovirus: el primero pertenece a la familia Flaviviridae, mientras que el del chikungunya pertenece a la familia Togaviridae y se propagan frecuentemente en las regiones tropicales y subtropicales.

En la literatura se consiguen diversos estudios focalizados en la predicción de epítipos en células B en dengue como se muestra en el trabajo de Nevis y colaboradores (16), basados en los resultados de los programas ABCPred y BcePred; así como el recientemente publicado por Isea (17) donde empleó los programas BepiPred, BcePred y BCPreds. Sin embargo, el presente trabajo plantea una nueva función para cuantificar la predicción de los epítipos B según factores estructurales y energéticos, en

vez de indicar únicamente la frecuencia de aparición de esos epítipos mediante una amplia gama de predictores.

Materiales y Métodos

El requisito de la presente metodología para determinar la predicción de epítipos B es conocer tanto su secuencia de aminoácidos (necesario para la predicción de los epítipos lineales), y la información derivada de su conformación tridimensional (requisito para la predicción conformacional), como se describe a continuación:

1.-Se realizan las ocho predicciones de epítipos B empleando para ello todos los programas indicados en la Tabla 1. Es importante reseñar que para obtener los resultados con los programas BCPreds y ABCPred, fue necesario indicar la longitud del epítipo B predicho, el cual se fijó en 10 aminoácidos, con una especificidad del 75% y 0,51, impuestos respectivamente.

2.-A partir de los epítipos predichos, se agrupan los resultados de acuerdo al tipo de predicción, es decir, aquellos que potencialmente pueden presentar una conformación lineal y los otros del tipo conformacional. En cada grupo, se determina la frecuencia de aparición del aminoácido presente en el epítipo, y lo vamos a denotar con las letras WL y WC, que corresponden a los obtenidos por los epítipos lineales (de allí la letra "L") y conformacionales (representado con "C"); con la condición que sí el mismo no es predicho por al menos dos predictores, dicho valor será igual cero (más adelante se visualiza con un ejemplo ilustrativo).

3.-Se determinan computacionalmente los cambios energéticos en cada uno de los aminoácidos que conforman la proteína inicial con la que se desea predecir los epítipos empleando para ello el cálculo de la energía libre de Gibbs. Para ello, se empleó el programa PoPMuSiC. Este criterio no tiene base experimental, es una condición que pretende examinar la posibilidad de considerar la estabilidad del aminoácido predicho en cada epítipo. El programa puede ofrecer un valor igual a cero (0), en cuyo caso, se sustituirá por -0,01, un valor seleccionado arbitrariamente, hasta que se optimice dicho valor a través de ensayos experimentales.

4.-Se predicen los modos normales de vibración de la proteína a partir de un modelo de redes elásticas, que ha sido posible gracias al uso del programa elNémo. Sin embargo, no es una garantía que dicha predicción sea exitosa hasta que se valide con una data experimental, la cual se espera realizar en futuros trabajos.

5.-Se van a definir y determinar los valores de las funciones cuantitativas FL y FC de la siguiente manera:

$$FL = WL \cdot \Delta\Delta G \cdot S1 \cdot S2 / B$$

$$FC = WC \cdot \Delta\Delta G \cdot S1 \cdot S2 / B$$

WL y WC son la frecuencia de aparición de los aminoácidos presentes en dicho epítipo, mientras que $\Delta\Delta G$ es el valor del cambio de energía libre obtenido con PoPMuSiC. Asimismo, debería favorecer a aquellos que están expuestos a la superficie gracias a los resultados logrados mediante el programa PoPMuSiC (representado por S1) y con el programa NetSurfP (S2), e inversamente proporcional a la movilidad resultante de los aminoácidos de la proteína de donde se deriva el epítipo obtenido con ayuda del programa elNémo (señalado con la letra B).

6.-Finalmente, las funciones <FL> y <FC> corresponde al valor promedio de cada epítipo B de acuerdo a cada uno de los valores de FL y FC descrito en los puntos anteriores, con la condición adicional que la extensión del mismo debe ser igual o superior a cinco aminoácidos. Dicho rango se podrá ajustar tras ensayos estadísticos que se realizará a futuro.

Resultados y discusión

En la Tabla 2 se muestra una región ilustrativa con los resultados obtenidos tras evaluar la glicoproteína del virus del chikungunya (cuyo identificador en la base de datos PDB es 2XFB) desde el aminoácido Cisteína (posición 330) hasta la Valina (347), donde se indican los valores del factor de temperatura conseguido con elNémo (abreviado como B), y el valor del cálculo de energía libre ($\Delta\Delta G$) y la accesibilidad del solvente (S1) logrados con el programa PoPMuSiC y NetSurfP (S2). El resultado de la predicción de los epítopos B fue posible con los programas BcePred (abreviado como Bce), ABCPred (ABC), BCPreds (BCP), BepiPred (Bepi), DiscoTope (Dis), ElliPro (Elli) y SEPPA (SEP). Se resaltaron en negritas aquellos aminoácidos que de acuerdo a los programas de predicción son candidatos a formar parte de un epítipo B. Posteriormente, se observan los valores de FL y FC que reflejan el valor de la función correspondiente a una predicción lineal y conformacional, respectivamente. Consideremos por ejemplo la Alanina en la posición 335. Este aminoácido está presente en la predicción de los epítopos B del tipo lineal (BCPreds y BepiPred), y conformacional (DiscoTope), de allí que los valores de WL y WC sean

Tabla 2. Región ilustrativa con los resultados obtenidos tras evaluar la glicoproteína del virus del chikungunya desde el aminoácido Cisteína (posición 330) hasta la Valina (347).

AA	ID	B	$\Delta\Delta G$	S1	S2	Bce	ABC	BCP	Bepi	Dis	Elli	SEP	FL	FC
C	330	0,30	-0,01	16,90	0,04	-0,32	0	0	0,07	-12,91	0,2	0,06	0	0
R	331	0,31	-0,13	32,50	0,265	-0,97	0	0	0,16	-7,72	0,51	0,07	0	0
V	332	0,30	-0,01	10,50	0,067	-1,18	0	1	0,28	-7,19	0,33	0,08	0	0
P	333	0,28	-15,62	22,00	0,245	-1,47	0	1	0,70	-5,56	0,38	0,08	-595,0	-595,0
K	334	0,24	-1,96	57,00	0,185	-1,32	0	1	0,81	-4,07	0,45	0,09	-174,3	-174,3
A	335	0,19	-0,27	35,80	0,101	-0,86	0	1	1,15	-4,20	0,51	0,06	-10,1	0
R	336	0,24	-0,01	42,20	0,444	0,05	0	1	1,08	-3,97	0,66	0,11	-1,5	-2,3
N	337	0,25	-0,73	45,10	0,404	0,66	0	1	1,29	-5,37	0,66	0,09	-108,1	-108,1
P	338	0,29	-3,80	20,40	0,268	0,93	0	1	1,14	-6,87	0,70	0,09	-145,2	0
T	339	0,49	-0,08	60,30	0,442	0,42	0	1	1,12	-5,53	0,82	0,08	-8,7	-13,1
V	340	0,49	-0,01	25,60	0,206	-0,15	0	1	1,14	-6,12	0,79	0,1	-0,2	-0,3
T	341	0,71	-0,01	47,40	0,427	-0,86	0	1	1,12	-6,07	0,88	0,13	-0,6	-1,2
Y	342	0,82	-0,01	17,40	0,267	-0,75	0	0	0,91	-7,26	0,87	0,12	0	-0,2
G	343	1,04	-3,38	18,70	0,437	-0,49	0	0	0,60	-7,41	0,94	0,13	0	-80,0
K	344	1,28	-0,72	81,50	0,433	0,10	0	0	0,03	-5,69	0,99	0,17	0	-79,4
N	345	1,04	-0,06	40,90	0,369	0,23	0	0	-0,33	-8,84	0,97	0,17	0	-1,8
Q	346	0,78	-0,55	38,10	0,368	-0,23	0	0	-0,89	-8,95	0,93	0,15	0	-19,8
V	347	0,54	-0,01	0,00	0,045	-1,32	0	0	-1,14	-9,55	0,82	0	0	0

dos y cero, respectivamente (no se muestran en la Tabla 2). WC es cero porque solo es predicho por un método conformacional (ver sección anterior para detalles). FL será resultado de multiplicar los siguiente cuatro términos, WL (2), el valor del cambio de energía libre de Gibbs (-0,27), y los valores de exposición del solvente 35,80 y 0,101 (correspondientes a S1 y S2, respectivamente), dividido por el factor de temperatura en dicha posición (0,19). Finalmente, el valor de FL será -10,1, mientras que el valor de FC para dicho aminoácido es cero al ser WC igual a cero (véase la posición 335 en la Tabla 2).

Solo resta indicar como se determinó <FL> que refleja el valor promedio de la predicción de epítomos B sopesado con contribuciones estructurales y energéticas del epítomo predicho per se. Del ejemplo anterior se desprende que el epítomo está entre la posición 333 hasta 341, cuyo valor medio de los valores FL en dicho intervalo (es decir, -595,0, -174,3, -10,1, -1,5, -108,1, -145,2, -8,7, -0,2 y -0,6) es igual a -115,96. De manera que según la metodología que proponemos en este trabajo, se predice el epítomo B lineal — PKARNPTVT — presenta un valor <FL> de -115,96 tras considerar propiedades estructurales y fisicoquímicas.

Tabla 3(a). Epítomos B lineales predichos en la glicoproteína del virus de chikungunya.

Posición	Secuencia	<FL>
121-128	KTDDSHDW	-110,87
136-143	NHMPADAE	-645,13
151-160	TSAPCTITGT	-27,81
182-188	SRKISHS	-703,56
190-201	THPFHHDPPVIG	-54,32
208-217	RPQHKGELPC	-199,54
227-248	TTEEIEVHMPDTPDRTLMSQQ	-153,91
266-276	NCGGSNEGLTT	-110,01
333-341	PKARNPTVT	-115,96
393-402	TWGNNEPYKY	-34,06

En la Tabla 3, se muestra la lista completa de los epítomos B obtenidos tras analizar la glicoproteína del virus del chikungunya cadena E con predictores del tipo lineal (3a) y conformacional (3b), y en la Tabla 4, los derivados de la glicoproteína del virus del dengue posiblemente presente con una conformación lineal (4a) y conformacional (4b).

Conclusiones

El presente trabajo plantea una nueva metodología computacional para determinar epítomos lineales y

Tabla 3(b). Epítomos B conformacionales predichos en la glicoproteína del virus de chikungunya.

Posición	Secuencia	<FC>
123-127	DDSHD	-80,09
202-213	REKFHSRPQHGK	188,40
246-251	SQQSGN	-99,69
280-286	VINNCKV	-64,23
307-318	PRNAELGDRKGGK	-170,46
333-346	PKARNPTVTYGKNQ	-76,80
360-374	SYRNMGEPPNYQEEW	-71,72
376-388	MHKKEVVLTPTE	-142,69
393-405	TWGNNEPYKYWPQ	-47,60

Tabla 4(a). Epítomos B lineales predichos en la glicoproteína del dengue.

Posición	Epítomo	<FL>
34-41	MAKNKPTL	-203,77
45-51	LQKTEAT	-132,88
67-90	NITDSRCPTQGEAILPEEQDQNY	-270,68
98-107	DRGWGNGCGL	-13,21
149-158	HQVGNETQGV	-211,81
180-189	GLECSPRTGL	-522,89
218-232	WTSGATTKTPTWNRK	-510,59
241-246	AHAKKQ	-82,97
309-314	EVSETQ	-581,52
324-329	YKGEDA	-550,36
336-345	STEDGQGKAH	-383,11
358-365	KKEEPVNI	-233,60
367-374	AEPFPGES	-765,72

Tabla 4(b). Epítomos B conformacionales predichos en la glicoproteína del dengue.

Posición	Epítomo	<FC>
96-110	YVDRGWGNGCGLFGK	-15,66
241-245	AHAKK	-30,06
338-346	EDGQGKAHN	-658,28
370-374	PFGES	-97,69
388-392	NWYRK	-682,25

conformacionales tras considerar los resultados de ocho predictores de epítomos de células B teniendo presente factores energéticos y estructurales de la proteína que los deriva. De hecho, la variable WL y WC expresan la frecuencia de aparición de los epítomos B según sea el método predicho, pero no contemplan otros factores que permitan dar valor agregado a los epítomos B predichos.

Por ello, se emplearon factores energéticos así como el grado de exposición del solvente para complementar y posiblemente dar un factor de peso para la selección del epítipo predicho de acuerdo a la movilidad y estimaciones energéticas de la proteína problema que produce los epítopos que serán escogidos. Sin embargo, es necesario realizar estudios adicionales para poder comparar entre sí los epítopos predichos antes de poder seleccionar y diseñar, por ejemplo, candidatos vacunales contra enfermedades virales como el dengue y el chikungunya.

Agradecimientos

Queremos expresar nuestro reconocimiento a los árbitros del trabajo por sus sugerencias. Este trabajo está dedicado a la memoria del Dr. Rafael Horacio Borges García.

Referencias

1. Yang X, Yu X. An introduction to epitope prediction methods and software. *Rev Med Virol.* 2009;19(2):77-96.
2. Isea, R. Designing a peptide-dendrimer for use as a synthetic vaccine against *Plasmodium falciparum*. *Am J Bioinform Comput Biol.* 2013;1:1-8.
3. Isea, R. Vacunología inversa aplicada en malaria: del genoma a los antígenos. En: De la Iglesia D, Aguiló J, Freire A, López V, Pazos A, editores. *Tecnologías NBIC en salud: El papel protagonista de la Nanociencia. Aplicación de especial interés al cáncer colorrectal.* España: CYTED; 2012: p. 34-40.
4. EL-Manzalawy Y, Honavar V. Recent advances in B-cell epitope prediction methods. *Immunome Res.* 2010;6(Suppl 2):S2. Disponible en: doi: 10.1186/1745-7580-6-S2-S2
5. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X et al. *Bioinformatics Resources and Tools for Conformational B-Cell Epitope Prediction.* *Comput Math Methods Med.* 2013; 943636. Disponible en: doi: 10.1155/2013/943636
6. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Structural Biology* 2009;9(51):1-10.
7. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD et al. VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res.* 2003;31(13):3316-9.
8. Kwasigroch JM, Gilis D, Dehouck Y, Rooman M. PoPMuSiC, rationally designing point mutations in protein structures. *Bioinformatics* 2002;18(12):1701-2.
9. Isea R. Predicción de epítopos consenso de células B lineales en *Plasmodium falciparum* 3D7. *VacchiMonitor* 2013; 22(1):43-6.
10. Kringelum JV, Lundegaard C, Lund O, Nielsen M. Reliable B cell epitope predictions: impacts of method development and improved benchmarking. *PLoSComput Biol.* 2012;8(12):e1002829. Disponible en: doi: 10.1371/journal.pcbi.1002829
11. Gilis D, Rooman M. PoPMuSiC, an algorithm for predicting protein mutant stability changes. Application to prion proteins. *Protein Eng.* 2000;13(12):849-56.
12. Frappier V, Najmanovich RJ. A Coarse-Grained Elastic Network Atom Contact Model and Its Use in the Simulation of Protein Dynamics and the Prediction of the Effect of Mutations. *PLoSComput Biol.* 2014;10(4):e1003569. Disponible en: doi: 10.1371/journal.pcbi.1003569
13. Jain T, Jayaram B. An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. *FEBS Lett.* 2005;579:6659-66.
14. Suhre K, Sanejouand YH. ElNémo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* 2004;32(suppl 2) 610-4: disponible en: doi: 10.1093/nar/gkh368
15. Hollup SM, Salensminde G, Reuter N. WEBnm@: a web application for normal mode analyses of proteins. *BMC Bioinformatics* 2005;6(52):1-8.
16. Nevis A, Reyes F, Calero R, Camacho F, Acosta A. Predicción de epítopos T y B de la proteína NS4b del virus dengue tipo 3. *VacchiMonitor* 2013;22(3):14-21.
17. Isea R. Mapeo computacional de epítopos de células B presentes en el virus del dengue. *INHRR* 2013;44(1):25-9.

In silico quantitative prediction of B-cell epitope

Abstract

This paper shows a computational approach for quantitative prediction of B cell epitopes. The function <F> was defined, which reflects the average value of B epitopes, according to eight predictors of different B epitopes, as well as structural and energetic considerations of the origin protein. The proposed methodology could be useful to develop both dengue and chikungunya vaccines.

Key words: epitope, B cell, dengue, chikungunya.
