

# Predicción de sitios de interacción con MCH en la estructura primaria de los SAGs



Universidad Nacional de General Sarmiento



Omar J. Argañarás<sup>1</sup>  
Marisa M. Fernández<sup>2</sup>  
Mauricio De Marzi<sup>3</sup>

<sup>1</sup> Instituto de Ciencias, Universidad Nacional de General Sarmiento  
Buenos Aires, Argentina  
<sup>2</sup>(IDEHU), FFyB-UBA-CONICET  
Buenos Aires, Argentina  
<sup>3</sup>Laboratorio de Inmunología, Universidad Nacional de Luján  
Buenos Aires, Argentina

mmfernan@ffyb.uba.ar, mdemarzi@unlu.edu.ar,  
omar.arganaras@campus.ungs.edu.ar

## 1. Introducción

Los alineamientos múltiples de secuencias (MSA) nos aportan como mínimo dos líneas de conocimiento, por un lado se encuentran ciertas posiciones de aminoácidos en el alineamiento altamente conservadas y por otro lado existen posiciones vinculadas a la estructura – función de la proteína estudiada [1].

Utilizando la información mutua de la teoría de la información se puede ponderar la relación coevolutiva entre dos posiciones en una familia de proteínas. Con el conocimiento de los aminoácidos que intervienen en la relación MCH-CPA (complejo mayor de histocompatibilidad - célula presentadora del antígeno) de la proteína SEI de la familia Saphylococcus aureus queremos predecir la ubicación en la secuencia de aminoácidos coevolucionados, si se encuentran, el MCH.

Existen aplicaciones disponibles pero nosotros proponemos mejorar el método en sus pasos intermedios.

## 2. El Modelo

Datos de entrada al modelo:

Buscamos secuencias de Superantígenos (SAGs) de la subfamilia Saphylococcus aureus. Y la estructura de SEI (2G9H) obtenida en la base de estructura denominada PDB [2].

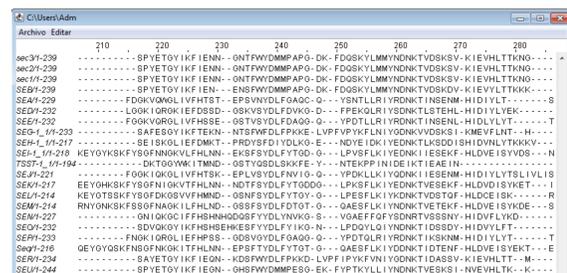


Figura 1: Alineamiento múltiple de secuencia utilizando T-Coffee [3] y visualizado con Jalview [4]

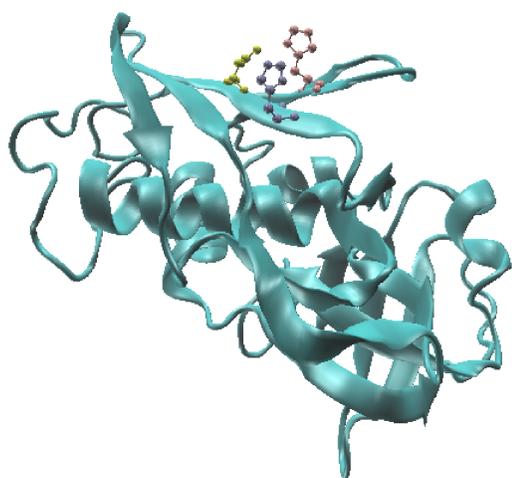


Figura 2: Estructura cristalizada de SEI destacando los 3 aminoácidos que interaccionan formando el CMH (complejo mayor de histocompatibilidad) [5]

Modelo matemático:

Aplicamos el modelo de información mutua [6] entre dos posiciones en un alineamiento múltiple de secuencias de aminoácidos que está descrito por la siguiente ecuación:

$$MI(i, j) = \sum_{i,j} P(a_i, b_j) \cdot \log \left( \frac{P(a_i, b_j)}{P(a_i) \cdot P(b_j)} \right)$$

(1)

donde P(ai, bj) es la frecuencia del aminoácido a en la posición i y el aminoácido b en la posición j en la misma secuencia, P(ai) es la frecuencia del aminoácido a en la posición i y P(bj) es la frecuencia del aminoácido b en la posición j.

## 3. Resultados

A partir de un alineamiento múltiple de secuencias de aureus y la estructura (pdb) de SEI obtuve resultados de posibles sitios de CMH en secuencias que no se conocen dichos sitios.

Aureus	207 (MI)	209 (MI)	169	Observaciones
SEC3	9,505427	7,287265		
SEC2	9,511734	7,279851		
SEC1	9,603033	7,289615		
SEB	9,211882			*
SEA	11,114527			*
SED	10,907445			
SEE	10,820602			
SEG	9,883984			
SEH	9,677607	7,145700		
SEI	11,599986	7,632669		
SEJ	10,408566	6,873729		
SEK	8,544084			
SEL	8,517099			
SEM	10,294096	7,387277		
SEN	9,111823			
SEO	9,406819	6,593579		
SEP	10,237132			
SEQ	9,045854			
SER	10,083408	6,541051		
SEU	7,930507			*

Cuadro 1: Valores de información mutua entre sitios de estudio [0-14]. Las secuencias marcadas en el campo Observaciones se podrían incorporar a la predicción con la mejora del algoritmo de alineamiento entre la secuencia y estructura cuaternaria.

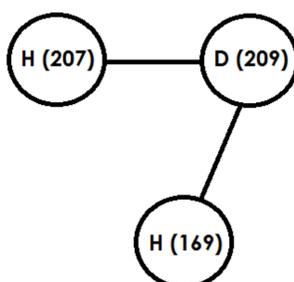


Figura 3: Grafo del sitio MCH con valores significativos.

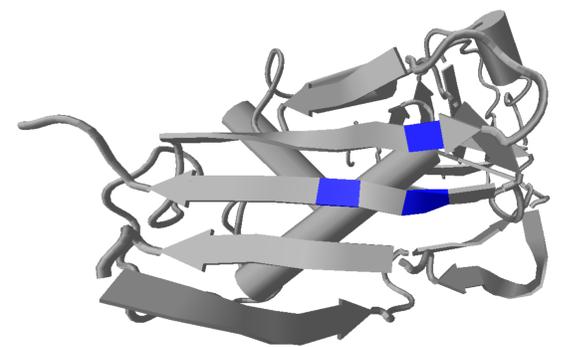


Figura 4: Estructura SEI con las posiciones marcadas en azul.

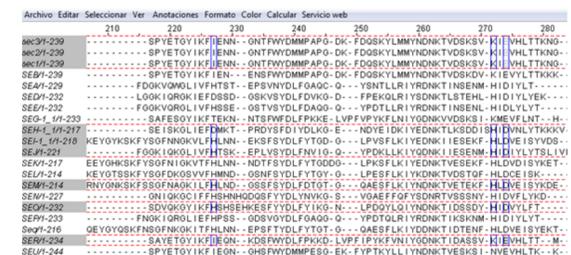


Figura 5: En el MSA las posiciones del MCH de SEI son 227 (H), 272 (H) y 274 (D).

## 4. Conclusiones

Nuestros resultados basados en la teoría de la información estiman el nivel de información mutua para cada par de aminoácidos ubicados en la zona de estudio de la subfamilia de Sags denominada Saphylococcus Aureus. Estas métricas nos ayudan a inferir el sitio del MCH en secuencias no estudiadas (in vitro). Introduciendo una mejora en el proceso de alineamiento de la secuencia contra la estructura se obtienen más posibles MCH.

## Referencias

- [1] Atchley, W.R., Wollenberg, K.R., Fitch, W.M., Terhalle, W. and Dress, A.W. (2000) Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. Mol. Biol. Evol., 17, 164–178.
- [2] Berman, H., Henrick, K., Nakamura, H. (2003). Announcing the worldwide protein data bank. Nature Structural Molecular Biology, 10(12), 980.
- [3] Notredame, C., Higgins, D. G., Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. Journal of molecular biology, 302(1), 205-217.
- [4] Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., Barton, G. J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), 1189-1191.
- [5] Fernández, M. M., Guan, R., Swaminathan, C. P., Malchiodi, E. L., Mariuzza, R. A. (2006). Crystal structure of staphylococcal enterotoxin I (SEI) in complex with a human major histocompatibility complex class II molecule. Journal of Biological Chemistry, 281(35), 25356-25364.
- [6] Thomas M. Cover, Joy A. Thomas (1991) Elements of Information Theory. Print ISBN 0 -471-06259-6 On line 0-4714-20061-1.