



Avance Doctorado en Ciencias Aplicadas y de la Ingeniería

- Nombre del becario: María Laura del Valle Marcos.
- Nombre del director: Julián Echave.

Introducción y estado de avance de la investigación

Las proteínas divergen durante la evolución biológica. A nivel de secuencia, diferentes sitios evolucionan a diferentes velocidades, lo que pudimos explicar usando un modelo que desarrollamos recientemente, el “Stress Model” (Huang TT, Marcos ML, Hwang JK, Echave J., 2014). Según este modelo, la velocidad de evolución de cada sitio está determinada por el empaquetamiento del sitio ya que, mutar un sitio empaquetado estresa y desestabiliza la conformación activa de la proteína. Más aún, encontramos que el empaquetamiento de las cadenas laterales de los aminoácidos es el mejor predictor de la variación secuencial (Marcos ML, Echave J., 2015). En contraste con la divergencia de secuencias proteicas, a nivel de estructura y dinámica no parece haber rastros de selección natural, ya que aplicando el modelo mutacional “Linearly Forced Elastic Network Model” (LFENM) se pudieron reproducir los patrones de divergencia estructural y dinámica observados en las proteínas. De todas formas, es de esperar que la selección restrinja en alguna medida la divergencia estructural y dinámica. Por ejemplo, si la estructura y movimientos del sitio activo de una enzima son importantes para su actividad, se esperaría que las mismas se conserven más que lo esperado por un modelo puramente mutacional que no considera la selección natural. Es por esto que en esta etapa de la investigación buscamos evidencia de selección natural modelando mutantes de múltiples sitios con el modelo LF - ENM y seleccionando las mismas con el modelo “Stress Model”.

Objetivos de esta etapa y principales resultados

Luego de culminar con el análisis de la evolución de las secuencias proteicas, analizamos en profundidad la divergencia estructural. Para esto, seleccionamos 8 familias de proteínas representativas de las principales clases “Structural Classification Of Proteins” (SCOP). Luego, para cada familia seleccionamos a una proteína de referencia y representamos su estructura con el modelo “Anisotropic Network Model” (ANM), el cual considera a las proteínas como una red de sitios conectados por resortes. Posteriormente, simulamos mutantes múltiples de la proteína de referencia utilizando el LF - ENM como modelo mutacional. Este modelo simula mutaciones puntuales aplicando fuerzas a lo largo de los contactos del sitio a mutar. Por último, utilizamos el ya mencionado “Stress Model” como modelo de selección. Generamos 4 sets de mutantes teóricas con distintos regímenes de selección: sin selección, con selección baja, media o alta. Para esto, le dimos diferentes valores al parámetro β del “Stress Model” con el fin de aceptar cada mutación con probabilidades de 1, 0.8, 0.5 y 0.2 respectivamente. Luego analizamos la divergencia estructural en coordenadas cartesianas tanto de los conjuntos teóricos como del conjunto experimental mediante el cálculo de variación cuadrática de la estructura de equilibrio de cada mutante / proteína experimental con respecto a la proteína de referencia. Por otro lado, para analizar la divergencia estructural a lo largo de los modos normales de la proteína de referencia obtuvimos sus autovectores \mathbf{q}_n , que representan los movimientos independientes de la proteína, y sus correspondientes energías de deformación λ_n . Posteriormente, calculamos una medida de la contribución de los modos normales de la proteína de referencia a la variación estructural proteica entre la misma y cada mutante, ya sea teórico o experimental:

$$P_n \equiv \frac{(q_n^T \Delta \bar{r})^2}{\sum_n (q_n^T \Delta \bar{r})^2}$$

siendo $\Delta \bar{r}$ la variación en la estructura de equilibrio de las proteínas comparadas.

Luego, para cada familia promediamos los valores obtenidos para cada medida tanto para los conjuntos de mutantes teóricas como para el conjunto de proteínas experimental, obteniendo perfiles de variación estructural en coordenadas cartesianas, "Mean Square Deviation of site i " (MSD_i) y perfiles de variación estructural proyectada sobre los modos normales, $\langle P_n \rangle$. Por último comparamos perfiles teóricos y experimentales mediante el cálculo del coeficiente de correlación de Pearson. Los resultados obtenidos se muestran en la figura 1.

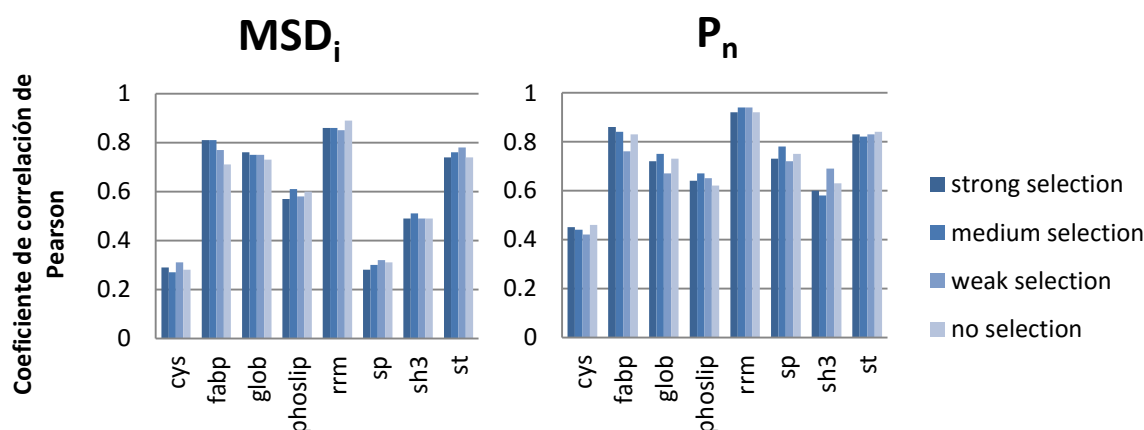


Figura 1: Coeficiente de correlación de Pearson entre los perfiles P_n promedio y MSD_i teóricos y experimentales de cada familia de proteínas. *Cys*: Cysteins, *fabp*: Fatty acid binding proteins, *glob*: globins, *phoslip*: phospholipase A2, *rrm*: RNA recognition motif, *sp*: serin proteases, *sh3*: src homology 3 domains, *st*: snake toxins.

En la figura 1 se observa que la correlación entre medias experimentales y teóricas es alta, considerando o no a la selección natural. Estos resultados contradicen la hipótesis de trabajo de que la selección natural debería afectar a la divergencia estructural, sugiriendo que, al menos para este caso, los patrones generales de divergencia estructural observados experimentalmente pueden explicarse con un modelo de mutaciones no seleccionadas.

Próximos objetivos

Nuestro próximo paso es focalizarnos en el análisis de divergencia estructural de sitios con importancia funcional. Luego, si bien ya hemos comenzado a analizar la divergencia estructural proteica utilizando modelos basados en 2 nodos por sitio; el carbono alfa de los aminoácidos (el nodo utilizado en este reporte) y el centro de masa de las cadenas laterales de los aminoácidos, profundizaremos y culminaremos este análisis. Por último, estudiaremos medidas de divergencia dinámica proteica usando los mismos modelos de red elástica, de modelado de mutaciones y de selección de las mismas que implementamos y utilizamos hasta el momento.

Presentaciones a congresos

- Congreso Argentino de Físicoquímica y Química Inorgánica de la Asociación Argentina de Investigación Físicoquímica (AAIFQ) de 2015.



Materias cursadas

- Fisicoquímica de biomoléculas en ECYT de UNSAM.
- Curso transporte óptimo y análisis de datos FCEyN - UBA.

Publicaciones

- Huang TT, Marcos ML, Hwang JK, Echave J. (2014) *A mechanistic stress model of protein evolution accounts for site-specific evolutionary rates and their relationship with packing density and flexibility*. BMC Evol Biol. doi: 10.1186/1471-2148-14-78
- Marcos ML, Echave J.(2015) *Too packed to change: side-chain packing and site-specific substitution rates in protein evolution*. PeerJ 3e911