

## **PLANTILLA 3**

### **PLAN DE TRABAJO**

#### **TÍTULO:**

*Efecto de mutaciones sobre la conservación de la diversidad conformacional de proteínas*

#### **RESUMEN:**

*El estado nativo de una proteína puede ser representado como una distribución no uniforme de poblaciones de conformeros en equilibrio dinámico. La coexistencia de estos conformeros, con diferentes afinidades por el ligando y/o roles en la función biológica, es la base del modelo de selección conformacional o pre-equilibrio. Las barreras energéticas que separan estos conformeros son comúnmente superadas por las fluctuaciones térmicas. La flexibilidad de la proteína modula la altura de estas barreras y la extensión del conjunto de conformaciones. Bajo este modelo, la diversidad conformacional de proteínas está íntimamente ligada a la función biológica y las diferencias en las propiedades estructurales y dinámicas de los conformeros pueden tener importantes consecuencias funcionales. La dinámica vibracional específica de cada conformero garantiza los cambios conformacionales y, por este motivo, se asocia con determinados rasgos de conservación evolutiva. Es así que las vibraciones asociadas a cambios conformacionales se encuentran conservadas evolutivamente y es posible identificar posiciones claves dentro de la proteína cuyas mutaciones afecten significativamente estas vibraciones. En este marco, el presente proyecto propone analizar el efecto de mutaciones sobre la diversidad conformacional evaluando su impacto sobre las vibraciones involucradas en los cambios conformacionales.*

#### **MARCO TEÓRICO:**

*Los sitios críticos para la función de las proteínas pueden identificarse mediante métodos de alineación estructural y secuencial [Orengo-2007], [Stemberg-2004].*

*Según la teoría neutra de la evolución molecular [Kimura-1974], los residuos más relevantes para la función varían más lentamente que los menos importantes. Sin embargo, estos métodos no proporcionan una información completa sobre la naturaleza de la relación secuencia-estructura-función y se requiere información adicional relacionada con la dinámica de las proteínas [Bahar-2004], [Hammes-Schiffer-2006], [Orengo-2008], [Gerstein-1998], [Uniprot-2010], [Walsh-2010].*

*El estado nativo de una proteína puede ser representado como una distribución no uniforme de poblaciones de conformeros en equilibrio dinámico [Nussinov-2004]. La coexistencia de estos conformeros con diferentes afinidades por el ligando y/o roles en la función biológica es la base del modelo de selección conformacional o pre-equilibrio. Los cambios conformacionales pueden significar una amplia gama de cambios estructurales que van desde distorsiones localizadas espacialmente, por ejemplo, la reorientación de cadenas laterales que actúen como puertas para la entrada de ligandos [Damborsky-2013], a reordenamientos globales colectivos, por ejemplo, cambios en la orientación relativa entre dominios [Gerstein-2004], [Bahar-2005].*

*La existencia de un conjunto de conformeros en equilibrio dinámico valida el modelo de selección conformacional generalizado [Takada-2008], [Wright-2009], [Vakser-2013] originalmente formulado por Monod, Wyman y Changeux para explicar el cooperativismo y el alosterismo [Tawfik-2003], [Changeux-2012]. Según este modelo, la proteína requiere de una multiplicidad conformacional preexistente para que pueda cumplir su función biológica. Actualmente, se han desarrollado diferentes bases de datos de diversidad conformacional en el estado nativo de las proteínas (CoDNaS [Parisi-2016], PDBFlex [Godzik-2016]).*

*Las barreras energéticas que separan estos conformeros son comúnmente superadas por las fluctuaciones térmicas. Por lo tanto, las vibraciones de las proteínas, en condiciones de equilibrio térmico, modulan la diversidad conformacional y la población relativa de los conformeros.*

*Los métodos basados en el análisis de modos normales (NMA) que utilizan modelos de red elástica (ENM) [Orozco-2008] han demostrado ser en gran medida eficientes para explorar la dinámica global (es decir, modos de baja frecuencia) de proteínas comúnmente asociadas a las transiciones*

conformacionales [Chirikjian-2009]. Además, se observa que solo unos pocos modos normales de baja frecuencia son necesarios para lograr una buena descripción de la mayoría de los cambios conformacionales asociados a la función biológica de las proteínas [Sanejouand-2001].

NMA desacopla los movimientos complejos y las fluctuaciones de las proteínas en una combinación lineal de osciladores armónicos independientes, es decir, los modos normales, cada uno de los cuales implica los movimientos concertados de muchos átomos. De esa manera, se pueden identificar movimientos de dominio a gran escala, involucrados en la conexión de los diferentes estados conformacionales relacionados con la función. El efecto de las mutaciones sobre estos modos colectivos y funcionalmente relevantes ha sido previamente estudiado desde diferentes puntos de vista. Por un lado, se ha informado de la solidez de estos modos a las variaciones de secuencia [Bahar-2000], [Ortiz-2005], [Thirumalai-2006], [Maguid-2005]. Además, se ha demostrado que la conservación en modo normal aumenta linealmente con la colectividad, de modo que los modos más lentos más colectivos son los más conservados [Maguid-2008]. Dado que estos modos son los que más contribuyen a determinar la flexibilidad general de los perfiles del factor B, se puede explicar la conservación observada de la flexibilidad de la cadena principal [Maguid-2006]. Por otro lado, la comprensión molecular de la función biológica requiere la identificación de la red de residuos que participan en la dinámica relacionada con la función, como la unión del sustrato y la liberación de productos, regulaciones alostéricas y plegado. Por ejemplo, se ha demostrado que los residuos que son dinámicamente importantes para la unión del ligando se conservan evolutivamente. Mediante el uso del Método de Perturbación Estructural (SPM) [Thirumalai-2006], [Thirumalai-2005], [Wang-2005], que demuestra la respuesta específica del residuo a las perturbaciones, Zheng et al. fueron capaces de asociar cambios conformacionales de unión a ligando a redes de residuos funcionalmente importantes [Zheng-2009].

En los últimos años, el grupo de investigación de la Universidad de Quilmes, dirigido por el Prof. Sebastián Fernández-Alberti, ha desarrollado un método que permite la identificación de redes de residuos dinámicamente importantes para la conservación evolutiva de la diversidad conformacional de proteínas

*[Fernandez-Alberti-2016]. Brevemente, el método se basa en la previa selección de modos normales involucrados en el cambio conformacional debido a la unión al ligando. El efecto de mutaciones puntuales sobre el subespacio definido por estos modos se analiza introduciendo perturbaciones locales en el modelo de red elástica (ENM) de la proteína. Esto permite identificar posiciones claves cuyas mutaciones perturban significativamente a este subespacio vibracional. Estas posiciones claves representan residuos dinámicamente importantes para el cambio conformacional de unión al ligando. De esta forma, pudimos constatar que los residuos dinámicamente importantes se encuentran conservados evolutivamente. Es así que esta línea de estudio permite identificar críticos para la función de las proteínas y, por lo tanto, explicar su conservación evolutiva a través de su participación en la relación secuencia-estructura-función asociada a la preservación de la diversidad conformacional.*

#### **HIPÓTESIS:**

*La diversidad conformacional de proteínas está íntimamente ligada a la función biológica y las diferencias en las propiedades estructurales y dinámicas de los cónfórmeros pueden tener importantes consecuencias funcionales. La dinámica vibracional específica de cada cónfórmero garantiza los cambios conformacionales y, por este motivo, se asocia con determinados rasgos de conservación evolutiva. Es así que las vibraciones asociadas a cambios conformacionales se encuentran conservadas evolutivamente y es posible identificar posiciones claves dentro de la proteína cuyas mutaciones afecten significativamente estas vibraciones. En este marco, el presente seminario propone analizar el efecto de mutaciones sobre la diversidad conformacional evaluando su impacto sobre las vibraciones involucradas en los cambios conformacionales.*

#### **OBJETIVOS:**

*El objetivo principal de este seminario de investigación es la adaptación y utilización de métodos computacionales que permitan predecir el efecto de mutaciones sobre la dinámica vibracional responsable de la diversidad conformacional de las proteínas. De este modo nos será posible identificar sitios relevantes para preservar la diversidad conformacional de las proteínas, siendo*

*esto un nuevo aporte al momento de entender su grado de conservación evolutiva. La variación de la conservación evolutiva entre los distintos cambios conformacionales posibles permitiría inferir la relevancia biológica de las distintas estructuras proteicas involucradas.*

*Para ello se proponen los siguientes objetivos específicos:*

- a) Definir el subespacio de modos normales que participan de los cambios conformacionales requeridos para garantizar el equilibrio dinámico entre los distintos conformeros que caracterizan al estado nativo de las proteínas.*
- b) Estudiar la respuesta de los subespacios de modos activos ante mutaciones puntuales en los distintos residuos que componen la proteína.*
- c) Explorar las variaciones evolutivas de estos subespacios vibracionales con el objeto de relacionarlos con los cambios conformacionales que existen entre las distintas conformaciones.*

## **MATERIALES Y MÉTODOS:**

*El método actual [Fernandez-Alberti-2016], desarrollado en el grupo de trabajo, permite predecir sitios dinámicamente relevantes para preservar el cambio conformacional apo-holo. En este proyecto adaptaremos el método para predecir sitios dinámicamente relevantes para conservar el espacio conformacional representado por conjuntos de conformeros existentes obtenidos de la base de datos de diversidad conformacional (CoDNaS) [Parisi-2016]. Básicamente, CoDNaS es una base de datos redundante de distintas estructuras para la misma proteína, curada y ampliamente relacionada con distintas bases de datos. Seleccionaremos sets de proteínas que abarquen distintos grados de extensión de la diversidad conformacional. Exploraremos familias de proteínas homólogas con diversidad conformacional basada experimentalmente. Estas familias se obtendrán de la base de datos CoDNaS utilizando proteínas con un 40% de identidad de secuencia y un 70% de cobertura. Cada proteína en CoDNaS tiene un RMSD máximo asociado derivado de la comparación de todos los conformeros que pertenecen a una proteína dada. Este RMSD máximo se toma como la diversidad conformacional máxima que podría tener la proteína. Las proteínas se clasificarán en homogéneas o heterogéneas de acuerdo con su*

dispersión de RMSD máxima en cada familia ( $\sigma < 2 \text{ \AA}$  para familias homogéneas y  $\sigma > 2 \text{ \AA}$  por las heterogéneas). Si bien CoDNaS tiene ~29000 entradas, y sólo para citar unos pocos ejemplos, estudiaremos familias de proteínas rígidas como las pertenecientes a la familia de celulasas de *C. cellulolyticum* (PDB 1f9d) con un RMSD máximo menor a  $0.5 \text{ \AA}$  o altamente flexibles como la familia de las pirofosfatasas de *B. subtilis* con un RMSD máximo  $\sim 10 \text{ \AA}$ . Entre estas familias existen centenas de proteínas con diversidades conformacionales intermedias.

El hecho de que los modos normales proporcionen una descripción armónica desacoplada de las vibraciones de las proteínas es fundamental para identificar los movimientos vibracionales de equilibrio individuales que participan en los cambios conformacionales. No obstante, la identidad de los modos normales debe rastrearse después de pequeñas perturbaciones y esta no es una tarea sencilla, ya que pueden introducir reordenamientos en su orden de frecuencia. Además, la complejidad de la función energética potencial de una proteína puede hacer que varíen sustancialmente y, eventualmente, que se mezclen fuertemente. Con el fin de minimizar estos efectos, en el presente trabajo no nos ocupamos de los modos normales individuales, sino de los subespacios de modo normal asociados a los cambios conformacionales. Trabajaremos en un procedimiento para definir y comparar subespacios de modo normal asociados a cada cambio conformacional. Nuestra definición de posiciones clave, es decir, aquellas que son dinámicamente importantes para garantizar un conjunto de cambios conformacionales, se basa en el efecto de mutaciones en estos subespacios.

Los resultados serán correlacionados con valores de conservación evolutiva obtenidos a partir de alineamientos múltiples. La correlación entre conservación evolutiva y nuestros parámetros dinámicos, asociados a la conservación de los cambios conformacionales que garantizan la diversidad conformacional, indican la relevancia funcional del espacio conformacional de una proteína durante la evolución y de este modo pueden ser potenciales sitios cuyas mutaciones alteren sustancialmente la función biológica. Los residuos dinámicamente importantes serán cotejados con la base de datos UniProt [UniProt-2015], que proporciona una descripción completa de la información disponible sobre las proteínas, incluida la información relacionada con la función, actividad catalítica y

*mutaciones con efectos informados sobre la afinidad por el sustrato y actividad catalítica.*

### **CRONOGRAMA:**

Actividad	Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
a) Seleccionar el conjunto de proteínas con suficiente diversidad conformacional	x	x	x									
b) Adaptación del método para identificar residuos dinámicamente importantes para preservar la diversidad conformacional				x	x	x						
c) Cálculos sobre el set de proteínas seleccionadas en a)							x	x	x			
d) Análisis de resultados e Informe Final										x	x	x

### **FACTIBILIDAD:**

*En cuanto a infraestructura, el principal requerimiento del proyecto son los recursos computacionales. Actualmente contamos con un equipo Dell PowerEdge R815 2U con procesadores 4x AMD Opteron 6168 (capacidad:48 nodos), un cluster de 120 CPUs con procesadores AMD de 64 bits y un cluster de 6 GPU, el cual tenemos proyectado ampliar próximamente.*

*Los recursos financieros requeridos para la realización del plan propuesto provienen del subsidio UNQ: “Simulación de procesos moleculares de relevancia fisicoquímica y biológica” dirigido por la Dra. M. S. Fornasari y co-dirigido por el Prof. S. Fernández-Alberti (\$631.031,94 durante el período 2019- 2023).*

*El grupo también cuenta con financiamiento de dos proyectos provenientes de la Comunidad Económica Europea. Horizon 2020: MSCA-RISE-2017 (Call: H2020-MSCA-RISE-2017 (Marie Skłodowska-Curie Research and Innovation Staff Exchange ) Proposal number: 778247, Proposal acronym: IDPfun (durante el periodo 2018-2021). De 1.233.000,00 €, la parte asignada a la UNQ es de*

499.000,00 €. Y Horizon 2020: MSCA-RISE-2018 (Call: H2020-MSCA-RISE-2018 (Marie Skłodowska-Curie Research and Innovation Staff Exchange) Proposal number: 823886, Proposal acronym: REFRACT (durante el periodo 2019 -2022). De 2.226.400,00 €, la parte asignada a la UNQ es de 772.800,00 €.

#### **PERTINENCIA DISCIPLINAR:**

*La carrera Licenciatura en Biotecnología de la Universidad Nacional de Quilmes (UNQ) tiene como objetivo la formación de profesionales con habilidades en la generación de conocimientos biológicos básicos y aplicados, con una fuerte formación en Biología Molecular.*

*La diversidad conformacional de proteínas está íntimamente ligada a la función biológica, y las diferencias en las propiedades estructurales y dinámicas de los conformeros pueden tener importantes consecuencias funcionales. Esta línea de estudio permite identificar residuos críticos para la función de las proteínas y, por lo tanto, explicar su conservación evolutiva a través de su participación en la relación secuencia-estructura-función asociada a la preservación de la diversidad conformacional.*

*Poder identificar y determinar los efectos dinámicos de estos residuos es un paso clave en la generación de nuevas metodologías de mutagénesis dirigida que contemplen información dinámica y vibracional de la proteína. Ya que el método podría proporcionar un conjunto de residuos o posiciones a mutar con el fin de estabilizar o afectar la estructura y funcionalidad de una proteína.*

#### **BIBLIOGRAFÍA:**

*[Bahar-2000] Keskin O, Jernigan RL, Bahar I. Proteins with Similar Architecture Exhibit Similar Large-Scale Dynamic Behavior. 2000; 78:2093–106*

*[Bahar-2004] Chen S-C, Bahar I. Mining frequent patterns in protein structures: a study of protease families. Bioinformatics. 2004; 20 Suppl 1:i77–85.*

*[Bahar-2005] Tobi D. and Bahar I. Structural Changes Involved in Protein Binding Correlate with Intrinsic Motions of Proteins in the Unbound State. Proc. Natl. Acad. Sci. 2005, 102 (52), 18908-18913.*



[Changeux-2012] Changeux J.-P. *Allostery and the Monod-Wyman-Changeux Model After 50 Years*. *Annu. Rev. Biophys.* 2012, 41 (1), 103-133.

[Chirikjian-2009] Schuyler A. D., Jernigan R. L., Qasba P. K., Ramakrishnan B. and Chirikjian G. S. *Iterative Cluster-NMA: A Tool for Generating Conformational Transitions in Proteins*. *Proteins* 2009, 74 (3), 760-776.

[Damborsky-2013] Gora A., Brezovsky J. and Damborsky J. *Gates of Enzymes*. *Chem. Rev.* 2013, 113 (8), 5871.

[Gerstein-1998] Gerstein M, Krebs W. *A database of macromolecular motions*. *Nucleic Acids Res.* 1998; 26:4280–90.

[Gerstein-2004] Goh C.-S., Milburn D. and Gerstein M. *Conformational Changes Associated with Protein–Protein Interactions*. *Curr. Opin. Struct. Biol.* 2004, 14 (1), 104–109.

[Fernandez-Alberti-2016] Saldaña T. E., Monzon A. M., Parisi G. and Fernandez-Alberti S. *Evolutionary Conserved Positions Define Protein Conformational Diversity*. *PLOS Comput. Biol.* 2016, 12(3): e1004775.

[Godzik-2016] Hrabe T., Li Z., Sedova M., Rotkiewicz P., Jaroszewski L. and Godzik A. *PDBFlex: Exploring Flexibility in Protein Structures*. *Nucleic Acids Res.* 2016, 44 (D1), D423-D428.

[Hammes-Schiffer-2006] Hammes-Schiffer S, Benkovic SJ. *Relating protein motion to catalysis*. *Annu Rev Biochem.* 2006. 75:519–41.

[Kimura-1974] Kimura M. *On Some Principles Governing Molecular Evolution*. 1974; 71(7):2848–52.

[Maguid-2005] Maguid S, Fernandez-Alberti S, Ferrelli L, Echave J. *Exploring the common dynamics of homologous proteins. Application to the globin family*. *Biophys J.* 2005; 89(1):3–13.

[Maguid-2006] Maguid S, Fernández-Alberti S, Parisi G, Echave J. *Evolutionary conservation of protein backbone flexibility*. *J Mol Evol.* 2006; 63(4):448–57.

[Maguid-2008] Maguid S, Fernandez-Alberti S, Echave J. Evolutionary conservation of protein vibrational dynamics. *Gene*. 2008; 422(1–2):7–13.

[Nussinov-2004] Gunasekaran K. and Ma B. Nussinov, R. Is Allostery an Intrinsic Property of All Dynamic Proteins? *Proteins Struct. Funct. Bioinforma*. 2004, 57 (3), 433–443

[Ortiz-2005] Leo-Macias A, Lopez-Romero P, Lupyan D, Zerbino D, Ortiz AR. An analysis of core deformations in protein superfamilies. *Biophys J*. 2005; 88(2):1291–9.

[Orengo-2007] Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol*. 2007; 8(12):995–1005. PMID: 18037900

[Orengo-2008] Redfern OC, Dessailly B, Orengo C, Exploring the structure and function paradigm. *Curr Opin Struct Biol*. 2008; 18(3):394–402.

[Orozco-2008] Emperador A., Carrillo O., Rueda M. and Orozco M. Exploring the Suitability of Coarse-Grained Techniques for the Representation of Protein Dynamics. *Biophys. J*. 2008, 95 (5), 2127–2138.

[Parisi-2016] Monzon A. M., Rohr C. O., Fornasari M. S. and Parisi G. CoDNaS 2.0: A Comprehensive Database of Protein Conformational Diversity in the Native State. *Database* 2016, 2016, 1–8.

[Sanejouand-2001] Tama F. and Sanejouand Y.-H. Conformational Change of Proteins Arising from Normal Mode Calculations. *Protein Eng. Des. Sel*. 2001, 14 (1), 1–6.

[Sternberg-2004]. Pazos F, Sternberg MJE. Automated prediction of protein function and detection of functional sites from structure. *Proc Natl Acad Sci U S A*. 2004; 101(41):14754–9. PMID: 15456910

[Takada-2008] Okazaki K. and Takada S. Dynamic Energy Landscape View of Coupled Binding and Protein Conformational Change: Induced-Fit versus Population-Shift Mechanisms. *Proc. Natl. Acad. Sci*. 2008, 105 (32), 11182–11187.

[Tawfik-2003] James L. C. and Tawfik D. S. Conformational Diversity and Protein Evolution – a 60-Year-Old Hypothesis Revisited. *Trends Biochem. Sci.* 2003, 28 (7), 361-368.

[Thirumalai-2005] Zheng W, Brooks BR, Doniach S, Thirumalai D. Network of dynamically important residues in the open/closed transition in polymerases is strongly conserved. *Structure*. 2005; 13(4):565–77.

[Thirumalai-2006] Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc Natl Acad Sci U S A*. 2006; 103 (20):7664–9.

[Uniprot-2010] Hinz U, The UniProt Consortium. From protein sequences to 3D-structures and beyond: the example of the UniProt Knowledgebase. *Cell Mol LIFE Sci*. 2010; 67:1049–64.

[Vakser-2013] Ruvinsky A. M., Kirys T., Tuzikov A. V. and Vakser I. A. Ensemble-Based Characterization of Unbound and Bound States on Protein Energy Landscape. *Protein Sci*. 2013, 22 (6), 734-744.

[Wang-2005] Su JG, Xu XJ, Li CH, Chen WZ, Wang CX. Identification of key residues for protein conformational transition using elastic network model. *J Chem Phys*. 2011; 135(17):174101.

[Walsh-2010] Sleator RD, Walsh P. An overview of in silico protein function prediction. *Arch Microbiol*. 2010; 192 (3):151–5.

[Wright-2009] Boehr D. D., Nussinov R. and Wright P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol*. 2009, 5 (11), 789-796.

[Zheng-2009] Zheng W, Tekpinar M. Large-scale evaluation of dynamically important residues in proteins predicted by the perturbation analysis of a coarse-grained elastic model. *BMC Struct Biol*. 2009; 9:45.