

Título: Simulación de procesos moleculares de relevancia fisicoquímica y biológica.

Integrantes: Alfonso Hernández, Laura; Barletta, Patricio German; Clavero, Esteban; Cossio; Pérez, Rodrigo; Fornasari, María Silvana; Franklin Mergarejo, Ricardo; Gonzalez, Claudia Roxana; Grosso, Marcos; Hasenauer, Marcia; Marchetti, Julia; Monzon, Alexander; Nardo, Agustina; Oldani, Nicolás; Ondarse Álvarez, Dianelys; Palma, Juliana; Palopoli, Nicolás; Pierdominici Sottile, Gustavo; Rueda, Ana Julia; Saldano, Tadeo

Resumen: Este programa, que ya lleva ocho años de ejecución, articula diferentes proyectos que tienen en común el hecho de utilizar simulaciones, cálculos y algoritmos computacionales para estudiar sistemas de interés fisicoquímico y biológico. Muchos de sus proyectos apuntan a contestar preguntas similares aunque desde perspectivas distintas y complementarias ya sea de la fisicoquímica o la biología. Esta complementariedad queda manifiesta además en la variedad disciplinar de los integrantes del programa, que reúne bioquímicos, químicos, físicos y biotecnólogos. En estos años hemos constituido un grupo de investigación sólido, robustecido en las distintas líneas de investigación, hemos formado recursos humanos y hemos generado numerosas publicaciones (actualmente 10 miembros del programa están realizando sus tesis doctorales mientras que en los últimos 5 años 5 miembros han obtenido su doctorado, además de haber publicado 50 artículos en revistas internacionales con referato).

Mientras las diferentes líneas enfocan sus investigaciones sobre temáticas y sistemas diversos, las herramientas utilizadas para su estudio son similares y/o complementarias. Entre estas herramientas podemos mencionar a los métodos derivados de la dinámica molecular clásica, mixta QM/MM (por Quantum Mechanics/Molecular Mechanics) y semiclásica; los algoritmos de clustering, los modelos de evolución molecular y de estimación filogenética, las bases de datos para estudiar la diversidad conformacional en proteínas (www.codnas.com.ar) y webserver para la evaluación de modelos estructurales de proteínas (<http://www.embnet.qb.fcen.uba.ar/embnet/charge.php?app=BeEP>). Muchas de las herramientas computacionales utilizadas en el programa son códigos estándar, disponibles gratuitamente en el ámbito académico. Pero además, las diferentes líneas utilizan herramientas ad-hoc, desarrolladas por los integrantes del programa para resolver las problemáticas particulares de su línea de investigación. Para el desarrollo de nuestros propios códigos utilizamos los lenguajes FORTRAN, C/C++, R, Python y Perl.

En la actualidad el programa consta de tres grandes líneas de investigación. A saber: relajación y redistribución intramolecular de energía vibracional, dinámica computacional de proteínas y reacciones químicas y métodos bioinformáticos aplicables a la biología estructural de proteínas. A continuación se describen los distintos temas a ser estudiados en cada una de las actuales líneas del programa

Una de las líneas de trabajo concierne a los estudios de fotodinámica de moléculas orgánicas conjugadas. Esta línea involucra los proyectos "Simulación de dinámica fotoinducida en sistemas moleculares conjugados extendidos" y "Simulación de procesos fotofísicos en cicloparafenilenos,

bases del andamiaje de nanotubos de carbono". La propuesta general considera la simulación computacional de procesos de fotoexcitación y subsecuente relajación electrónica y vibracional en una variedad de moléculas de interés biológico e interés en aplicaciones tecnológicas que van desde antenas recolectoras de luz, generación de imágenes y células fotovoltaicas. En los últimos años hemos desarrollado, en colaboración con los grupos del Prof. Adrian Roitberg (University of Florida) y Sergei Tretiak (Los Alamos National Laboratory) la metodología NA-ESMD (por NonAdiabatic-Excited State Molecular Dynamics) que permite la simulación de dinámica molecular no-adiabática en estados excitados para sistemas moleculares conjugados de cientos de átomos (~300) y varias decenas de estados electrónicos excitados (~40-50) a escalas de tiempos de picosegundos [Tretiak 2014]. Esto nos ha permitido adquirir reconocimiento internacional en el área publicando más de 15 artículos en los últimos 4 años [Tretiak 2011], [Fernandez-Alberti 2012a], [Fernandez-Alberti 2012b], [Fernandez-Alberti 2014a], [Tretiak 2014a], [Tretiak 2014b], [Fernández-Alberti 2014b], [Fernández-Alberti 2014c]. En los próximos años nos proponemos avanzar en el desarrollo, implementación y aplicación de NA-ESMD introduciendo nuevos métodos que nos permitirán ampliar las aplicaciones a una variedad de sistemas moleculares conjugados extendidos con potenciales aplicaciones tecnológicas y que no pueden ser abordados en la etapa actual de desarrollo. Uno de los aspectos importantes a ser tratado en los próximos años concierne a la fusión de nuestro actual código NA-ESMD con subrutinas QM/MM y técnicas de generación de amplios muestreos conformacionales. Estas subrutinas están actualmente disponibles y optimizadas en programas de dinámica molecular de proteínas. En este aspecto metodológico, esta línea de trabajo se fusiona con la línea de dinámica de proteínas y reacciones químicas. Ambas líneas requieren el conocimiento e implementación de técnicas comunes. Por este motivo se hace imprescindible la convivencia de ambas líneas bajo un mismo programa de investigación. Avanzaremos en el tratamiento cuántico de grados de libertad nucleares, reformulación de la propagación del paquete de ondas electrónico con el objeto de incrementar la eficiencia del código, incorporación de solvente implícito y técnicas híbridas QM-MM. Las distintas implementaciones nos permitirán estudiar sistemas moleculares de mayor tamaño, predecir efectos de coherencia en la propagación nuclear, separaciones de carga y estudiar la competencia de procesos intramoleculares e intermoleculares, incluida la relajación inducida por el solvente. De este modo, la propuesta consiste en avanzar en el desarrollo de simulaciones NA-ESMD que nos permitan abordar un número de sistemas moleculares importantes actualmente en el foco de la investigación experimental.

En lo que respecta a los proyectos "Estudio del efecto de mutaciones en la transferencia de energía vibracional de proteínas" y "Redes de residuos dinámicamente importantes para la conservación evolutiva de la diversidad conformacional de proteínas", ambos pertenecen a una línea general de análisis de dinámica molecular de proteínas fundamentada en el estudio del impacto de mutaciones en la diversidad conformacional de una proteína, su unión al ligando, estabilidad y alteración de mecanismos de transferencia intramolecular de energía. La dinámica de

proteínas es abordada utilizando tanto la aproximación de modos normales (AMN) como simulaciones de dinámica molecular (DM). En primer lugar, nos proponemos desarrollar métodos que permitan estudiar cambios en la dinámica esencial de proteínas resultado de mutaciones puntuales, resaltando similitudes y diferencias debido a la unión al ligando, mutaciones, cambios del entorno, etc. En los últimos años hemos desarrollado un método que permite resaltar similitudes y diferencias en patrones de fluctuación debido a la unión al ligando [Fernández-Alberti 2015]. . Nos proponemos aplicar esta metodología y adaptarla para su aplicación a efecto de mutaciones en distintos sistemas utilizando tanto AMN como DM. En segundo lugar, nos proponemos desarrollar métodos computacionales que permitan el estudio de la evolución temporal de la energía vibracional posterior a la introducción de una perturbación inicialmente localizada en una parte específica de la proteína. Los métodos se basarán en el análisis de simulaciones de DM de no-equilibrio. Finalmente, nos proponemos desarrollar métodos que permitan identificar y caracterizar redes de residuos dinámicamente importantes para el mantenimiento de la diversidad conformacional de una proteína. Esperamos poder considerar estos residuos como elementos asociados a la función biológica. Utilizaremos nuestra recientemente desarrollada base de datos de diversidad conformacional CoDNaS [Parisi 2011, Parisi 2013]. Nos proponemos combinar la información obtenida a partir de métodos basados en características estructurales y dinámicas de las proteínas con información relacionada con su conservación evolutiva [Parisi 2012]. Para ello, calcularemos patrones de sustitución conformero-específicos en conjuntos de proteínas con diferentes grados de diversidad conformacional [Parisi 2001]. El proyecto representa un primer paso hacia una meta ulterior de desarrollar métodos de asignación funcional de proteínas que contemplen el requerimiento de la conservación evolutiva de su diversidad conformacional. La propuesta requiere una estrecha colaboración entre las líneas de análisis de dinámica molecular y las líneas de desarrollo de herramientas bioinformáticas del programa. Es necesario el uso de herramientas y bases de datos bioinformáticas para la generación del conjunto de proteínas adecuado para el posterior análisis de sus propiedades dinámicas bajo los distintos enfoques planteados. De este modo, la propuesta permitirá abordar conceptos biológicos generales relacionados con distintos aspectos dinámicos de la diversidad conformacional de proteínas, como ser, unión a ligandos, mutaciones con efectos específicos sobre la actividad biológica, etc. Para ello, el primer paso consiste en la generación de distintos sets de proteínas representativas de la diversidad de estructuras de familias de plegamientos actualmente disponible. La generación, mediante técnicas bioinformáticas, del set de estudio inadecuado, conduciría a resultados generales erróneos. Por este motivo se requiere el trabajo conjunto coordinado y con objetivos comunes de las líneas de análisis de dinámica molecular y las líneas de desarrollo de herramientas bioinformáticas presentadas en este programa de investigación.

La línea de dinámica computacional de proteínas y reacciones químicas se inició hace cuatro años con diversos proyectos dedicados a estudiar reactividad química, tanto en fase gaseosa como en enzimas, utilizando metodologías basadas en el concepto de estado de transición [Palma 2011a, Palma 2011b]. Con los años, la evolución y el crecimiento de estas líneas han dado lugar a varias ramificaciones y extensiones de las temáticas originales. Por un lado, en los estudios de reacciones en fase gaseosa hemos tenido que ampliar el rango de herramientas utilizadas incorporando métodos de scattering reactivo tradicional y métodos para el tratamiento de procesos electrónicamente no-adiabáticos. La necesidad de incorporar estas metodologías proviene de la discrepancia entre nuestras predicciones teóricas y los resultados experimentales correspondientes [Manthe 2014, Manthe 2015]. El plan a desarrollar en este aspecto se encuentra descrito en el proyecto "Influencia de complejos pre-reactivos en la dinámica de reacciones poliatómicas en fase gaseosa". Por otro lado, la experiencia y los conocimientos sobre dinámica molecular de proteínas, adquiridos al estudiar reacciones enzimáticas, nos permitió afrontar el desafío de simular proteínas embebidas en membrana lipídicas, a fin de dilucidar cómo afecta la presencia del sustrato o agonista (según corresponda) los movimientos de estos sistemas. De esta manera iniciamos el estudio de la proteína P2X4, hace aproximadamente un año y medio; y de la proteína LeuT, a fin del año pasado. Ambos trabajos se realizan en el marco del proyecto titulado "Estudios experimentales y teóricos sobre los mecanismos moleculares del funcionamiento de canales y transportadores trans-membrana", que desarrollamos en colaboración con el Dr. Luciano Moffat de la UBA. Además, al revisar la metodología existente para describir cambios conformacionales y dinámicos en proteínas, pudimos deducir las fórmulas que dan fundamento al llamado método de PCA-concatenado [Palma 2014], una técnica que se había usado por muchos años sin un fundamento teórico riguroso. El encontrar dicho fundamento nos permitió plantearnos nuevos usos de esta metodología. La propuesta para el desarrollo de estos nuevos usos está contenida en el proyecto titulado "Algoritmos computacionales para la identificación y caracterización de movimientos funcionales en proteínas". Finalmente, uno de nuestros proyectos anteriores estaba dedicado al estudio de la enzima trans-sialidasa de *Trypanosoma Cruzi* [Roitberg-2014]. El contacto con la bibliografía de las enzimas de *T. Cruzi* nos llevó a tomar conocimiento de la cristalización de la enzima UGM, que es de fundamental importancia para la supervivencia del parásito [Sobrado 2011]. Inmediatamente iniciamos un estudio detallado de la reacción catalizada por UGM [Palma 2014b] y, como una continuación de dicho trabajo, estamos ahora iniciando una línea de investigación tendiente a identificar fármacos que inhiban la acción de esta enzima. Los detalles de ese plan de trabajo se encuentran descritos en el proyecto "Búsqueda y análisis computacional de fármacos destinados a inhibir la función de la enzima UDP-galactopiranososa mutasa de *Trypanosoma cruzi*".

En lo que respecta a la línea de bioinformática estructural de proteínas, la propuesta de investigación está formada por seis proyectos de investigación. El tema central en la mayoría de

estos proyectos es la diversidad conformacional de las proteínas. Con este término nos referimos a las diferencias estructurales que definen los distintos conformeros que representan el estado nativo de la proteína. Este ha sido el tema central de esta línea en los últimos años. Para su estudio hemos desarrollado dos bases de datos de proteínas con distintas extensiones de diversidad conformacional PCDB [Parisi 2012a], una base de datos de dominios proteicos y CoDNaS, una base de datos de proteínas completas [Parisi 2013a]. La base de datos CoDNaS (actualmente con más de 17000 proteínas y más de 190000 conformeros) nutre a casi todas las líneas de investigación de esta línea de investigación con ejemplos y sets de estudios. En particular, usando la base de datos PCDB hemos demostrado que la diversidad conformacional influencia el patrón de sustitución [Parisi 2012b] y la velocidad de evolución de las proteínas [Parisi 2013b]. Recientemente publicamos un review donde exponemos nuestros principales resultados sobre cómo la diversidad conformacional condiciona la divergencia secuencial durante la evolución [Parisi 2015]. En este punto el presente programa continúa estos resultados en el proyecto “Estudio sobre la extensión, distribución y rol de la diversidad conformacional en proteínas” donde proponemos extender nuestro conocimiento sobre el efecto de la diversidad conformacional en cuanto a la divergencia secuencial, su relación con la función biológica y realizar un estudio sobre la extensión y diversidad de movimientos que impliquen la diversidad conformacional en el espectro de proteínas conocido en la base de datos CoDNaS. En este proyecto incluimos el estudio de regiones desordenadas y de regiones que sufren transiciones orden/desorden. Esta línea muestra una fuerte colaboración con el Dr. Silvio Tosatto de la Universidad de Pádova y recientemente hemos enviado un trabajo para su publicación [Parisi 2015]. Esta línea de investigación posee también una fuerte interacción con la línea de análisis de dinámica de proteínas. El tema de la diversidad conformacional es también central en el proyecto “Desarrollo de herramientas bioinformáticas para el estudio y predicción de la promiscuidad y multiplicidad de sustrato en enzimas”, donde proponemos desarrollar herramientas para la predicción y estudio de promiscuidad y multiplicidad de sustrato en enzimas. Este proyecto implica el desarrollo de una base de datos de proteínas con diversidad conformacional y promiscuidad/multiplicidad de sustrato que aportará sets de estudio para desarrollar conceptos generales en el tema y que nutrirán a las herramientas que proponemos desarrollar. Es importante destacar que en este proyecto colaboramos con el Dr. Luis Iglesias de UNQ para la identificación y caracterización de proteínas promiscuas que tengan importancia en la síntesis orgánica y que sirvan como potenciales biocatalizadores. En el proyecto “Desarrollo de métodos bioinformáticos basados en métodos evolutivos y estructurales para el estudio de proteínas” continuamos con nuestros estudios sobre los condicionamientos estructurales a la divergencia secuencial durante la evolución de proteínas. Mayormente proponemos actualizar el modelo de evolución molecular SCPE desarrollado anteriormente [Parisi 2001, Fornasari 2002, Parisi 2004, Parisi 2005, Fornasari 2007], para considerar la diversidad conformacional, nuevos modelos mutacionales y potenciales de energía para hacer al modelo más realista. En el proyecto “Incorporación de información evolutiva,

estructural y dinámica para el estudio de los cambios observados en proteínas en el proceso evolutivo y en estados patológicos” proponemos estudiar los determinantes estructurales de mutaciones asociadas a enfermedades, desde una perspectiva evolutiva, estructural y energética. En particular uno de los sistemas que estudiaremos es el EGFR, proteína asociada a cáncer de pulmón. Es importante mencionar que en este proyecto colaboramos con el Dr. Bramuglia, UBA, que nos ha aportado más de 2500 mutantes provenientes de pacientes argentinos, para caracterizar y comprender el efecto de esas mutaciones en dichos pacientes [Fornasari 2015]. En estos proyectos colaboramos con el Dr. Emidio Capriotti, Universidad de Alabama y la Dra. Rita Casadio, Universidad de Bologna. También contamos con un proyecto para el desarrollo de una plataforma bioinformática para la caracterización e identificación de péptidos bioactivos en estrecha colaboración con el grupo de la Dra. Cristina Añón de CIDCA (“Desarrollo de una plataforma bioinformática para la predicción de péptidos bioactivos en proteínas de importancia alimenticia”). Otro de nuestros proyectos implica el uso de técnicas de caracterización evolutiva, secuencial y estructural para comprender la fisiología de las interfaces proteína-proteína mediadas por motivos lineales cortos o SLiMs (“Análisis estructural de interacciones entre proteínas mediadas por motivos lineales cortos”). Con la integración de los intereses y experiencia del grupo de trabajo y en colaboración con el Dr. Richard J Edwards de la Universidad de New South Wales (Australia), se plantea realizar un análisis completo de SLiMs involucrados en interacciones entre proteínas que involucren las tendencias de conservación de residuos en la interfaz según sus condicionamientos estructurales. Como parte de la línea de desarrollo de métodos bioinformáticos en biología estructural de proteínas implementaremos HMM (Hidden Markov Models) evolutivos construidos a partir de un motivo lineal determinado y representativos del mismo, capaces de identificar nuevas instancias del mismo SLiM en bases de datos secuenciales. De esta manera aumentaremos el pool de SLiMs que pueden ser estudiados en profundidad. El trabajo colaborativo entre proyectos nos permitirá estudiar también, los aspectos de transición orden/desorden en la formación de complejos y el estudio general de sus características dinámicas. La flexibilidad proteica tiene un rol decisivo en la evolución de su secuencia y estructura [Marsh 2014]. La consideración de la heterogeneidad estructural (representada en particular por segmentos ordenados y otros de variable desorden, intrínseco o transitorio, codificado diferencialmente en su secuencia [Uversky 2013]) ha permitido asociar distintos niveles de estabilidad conformacional a un amplio espectro de funciones biológicas [Uversky 2015]. Numerosos ejemplos describen los cambios conformacionales asociados con el establecimiento de interacciones proteína-proteína, como la descripción de la tendencia a encontrar residuos rígidos en regiones centrales de la interfaz, con otros residuos periféricos más flexibles [Lin 2015]. Un reciente review de Van Roey et al. [Van Roey 2014] expone aspectos conformacionales y dinámicos esenciales para diversos roles funcionales de los motivos lineales cortos. La cooperatividad de numerosos motivos en la formación de complejos meta-estables, usualmente con un aumento de afinidad más que aditivo respecto de las interacciones relativamente débiles de los SLiMs individuales, facilita la regulación

de señales celulares de acuerdo al contexto. Esta multiplicidad puede incluir varias copias del mismo o de diversos motivos, presentes en uno o más de los monómeros que interactúan, lo que otorga gran variabilidad conformacional para el establecimiento de complejos. Estudios como los mencionados sugieren que las propiedades dinámicas de la interfaz podría ser un aspecto esencial en la formación de complejos proteicos mediados por SLiMs. No obstante, la información acerca de la dinámica de la interfaz proteína-proteína, en particular la flexibilidad de los residuos que la integran, su grado de ordenamiento y la diversidad conformacional vinculada a su función, aún no ha sido evaluada sistemáticamente. Como un primer acercamiento, los datos de diversidad conformacional generados por el proyecto “Estudio sobre la extensión, distribución y rol de la diversidad conformacional en proteínas” serán inspeccionados para extraer información sobre la dinámica estructural de proteínas portadoras de SLiMs. El conjunto de sub-estructuras de SLiMs constituido a partir de CoDNaS se estudiará comparativamente en sus diversas variables estructurales. Es de esperar que emerjan tendencias particulares capaces de determinar la adopción de diversas conformaciones cuya proporción estará definida por los condicionamientos particulares de cada motivo. Exploraremos esta posibilidad y de hallar evidencia experimental complementaria, su dinámica nativa se ajustaría a la noción, ya sugerida, de un estado libre compatible con la teoría de pre-equilibrio en el que coexisten poblaciones de múltiples conformeros nativos en proporciones variables [Dogan 2013]. La técnica de PCA-concatenado vinculada al proyecto “Algoritmos computacionales para la identificación y caracterización de movimientos funcionales en proteínas”, representa una herramienta útil para definir la funcionalidad de los residuos que integran SLiMs particulares en el contexto de su flexibilidad nativa.