Implementación de una herramienta bioinformática para el diseño de secuencias linker

# 1.- Introducción.

En términos generales, podemos definir la ingeniería de proteínas como la modificación estructural realizada mediante la manipulación del gen codificante, por la cual se consigue un cambio funcional determinado. La máxima expresión de la ingeniería de proteínas es el diseño de proteínas, que consiste en la definición y creación de novo de una secuencia proteica cuya estructura la haga apta para desempeñar determinada función. En el otro extremo se encuentran las modificaciones puntuales de la secuencia de una proteína conocida, con el objetivo de modificar sus propiedades.

Un caso particular de diseño muy utilizado es la generación de proteínas recombinantes de fusión, o quiméricas. Es una técnica ampliamente utilizada hoy en día, que implica la combinación de dominios o motivos estructurales provenientes de distintas proteínas mediante la ligación de las secuencias a través de una secuencia linker. La correcta selección de esta secuencia linker no es un paso trivial y es de gran importancia para la construcción de proteínas con las características deseadas, ya que afectan al plegamiento, la dinámica y la función de la estructura resultante)[1,2,3].

Actualmente, el diseño de los linkers se realiza de forma manual e individualmente para cada fusión que se desea realizar[4]. En este contexto, la optimización de los linkers[5] provista por el diseño racional puede contribuir a mejorar la eficiencia de expresión, el correcto plegamiento y actividad de los dominios de la proteína creada. Una solución automatizada permitiría, además, realizar este proceso de manera escalable para una gran cantidad de combinaciones de proteínas/dominios

A pesar de los beneficios de contar con una herramienta de este tipo, no se han realizado hasta el momento desarrollos que provean esta funcionalidad. El presente trabajo se enfoca, entonces, en obtener una herramienta computacional que permita obtener de manera automática la secuencia linker a partir de un conjunto de propiedades que el usuario ha seleccionado.

El presente trabajo se enmarca dentro de un proyecto de investigación asociado al subsidio PIP 2013-2015 GI, titulado “Desarrollo de sensores fluorescentes para la medición del hacinamiento molecular en células”. En este proyecto, a cargo del Dr. Julio Javier Caramelo y con participación del director de esta propuesta, se intenta desarrollar una sonda que permita medir en forma directa los niveles de hacinamiento molecular tanto in vitro como in vivo.

Como parte de este proyecto se planea diseñar, en base a los estudios teóricos, secuencias conectoras que cumplan con las propiedades estructurales óptimas para lograr la función de los sensores. Es en este punto donde se centra el plan de trabajo aquí propuesto. La herramienta a implementar permitirá diseñar las secuencias conectoras automáticamente, permitiendo expandir el conjunto de conectores disponibles para utilizar en las sondas.

Si bien la herramienta que se desarrollará puede ser aplicada a un conjunto más general de problemas, realizarla en un contexto como el que provee el proyecto recién mencionado permite tener un marco directo de aplicación para poder probarla y obtener una realimentación acerca de su eficacia. Por otro lado, la interacción con otros investigadores y la participación de un proyecto de mayor envergadura aporta en gran parte a la formación profesional de quien realizará este trabajo.

# 2- Objetivos.

El objetivo principal del trabajo es desarrollar una herramienta capaz de generar una secuencia con las propiedades necesarias para cumplir la función de linker en el diseño de una nueva proteína.

Como entrada, el algoritmo recibe las características de una secuencia inicial: longitud, secuencias flanqueantes, etc. En caso de no ingresar una secuencia completa, se generará una secuencia random de la longitud indicada. A partir de esta secuencia inicial, se itera analizando características deseables para la función de linker (propiedades estructurales, fisicoquímicas, etc). Para análisis de las secuencias se utilizarán desarrollos propios y herramientas disponibles a través de la web. Los resultados obtenidos en cada propiedad analizarán las pautas para realizar luego mutaciones puntuales sobre la secuencia en la que se esta trabajando. La implementación o no de estas mutaciones se decide en base a un algoritmo de montecarlo, buscando llevar a la secuencia hacia un óptimo global determinado por el conjunto de propiedades deseables.

El resultado será, entonces, una secuencia que cumpla con las condiciones iniciales planteadas por el usuario y que tenga un valor óptimo de la función objetivo, la cual representa las propiedades deseables.

Dentro del objetivo principal de desarrollar la herramienta, se encuentra un segundo objetivo que consiste en poner a disposición del público el recurso obtenido. Para esto, se implementará una interfaz que permita utilizar la herramienta remotamente una vez alojada en un servidor web.

La herramienta desarrollada formará parte de EMBnet, una red de nodos distribuidos que colaboran para proveer recursos bioinformáticos a la comunidad científica. Se podrá acceder a nuestra herramienta desde cualquier sitio con acceso a la web, a través de BiFe(Bioinformática Federal), el nodo argentino de EMBnet alojado en los servidores del departamento de Química Biológica - FCEN - UBA (<http://www.embnet.qb.fcen.uba.ar/embnet/index.php>).

# 3- Plan de Actividades Experimentales.

* En una primera etapa se analizarán las herramientas bioinformáticas disponibles que serán utilizadas para el análisis de la secuencia en cada iteración. Estas herramientas proveen información sobre propiedades de interés en la secuencia que se esta generando. Es necesario conocer el análisis que realizan, para luego poder reunir la información de todas las herramientas en una única función global que informe sobre el estado de la secuencia en cada etapa de la iteración.
* Luego de esta etapa se realizará un diseño abstracto de la implementación a realizar y de las pruebas y controles necesarios para evaluar el funcionamiento de la herramienta generada.
* Una vez realizado el análisis de las herramientas que se incorporaran para obtener información de las propiedades, se implementará el algoritmo central que conduce a la generacion de la secuencia objetivo.
* La última etapa en el desarrollo corresponde a las pruebas para medir el correcto funcionamiento y realizar las modificaciones necesarias. Para realizar las pruebas se utilizarán conjuntos de secuencias de entrada que contengan propiedades no deseadas y que el algoritmo debería eliminar. Se evaluarán también casos en los que la entrada cumple con las condiciones de salida necesarias. Asimismo se intentará a evaluar formalmente la capacidad global de la herramienta para lograr obtener una secuencia con las caracteristicas del objetivo.
* Una vez finalizado y probado el desarrollo se diseñará e implementará la interfaz gráfica, junto con la documentación necesaria para poner a disposición la herramienta. Dentro de esta documentación se encuentra el manual de usuario con toda la información necesaria para poder utilizar el recurso, junto con casos de prueba y resultados esperados.
* El último paso, una vez finalizado el desarrollo, consiste en plantear dentro de un informe el trabajo realizado junto con la fundamentación teórica de las decisiones tomadas en cada paso, incluyendo una conclusión acerca del alcance, limitaciones y la forma en que deben interpretarse los resultados de esta herramienta.

# 4 - Cronograma.

Análisis de las herramientas disponibles………………..45 horas

Diseño de la solución……………………………………..50 horas

Implementación del algoritmo……………………………30 horas

Pruebas de la funcionalidad……………………………...34 horas

Implementación de la interfaz, documentación pertinente y alojamiento en el servidor…25 horas

Redacción del informe…………………………………….40 horas

# 5 - Referencias.

**[1]** [**Fusion protein linkers: Property, design and functionality**](http://www.sciencedirect.com/science/article/pii/S0169409X12003006?np=y) **Chen X, Zaro J, Shen W-C. Advanced drug delivery reviews 2012**

**[2]** [**Fusion Protein Linkers: Effects on Production, Bioactivity, and Pharmacokinetics Chen, Xiaoying, Jennica Zaro, and Wei‐Chiang Shen. *Fusion Protein Technologies for Biopharmaceuticals: Applications and Challenges* (2013): 57-73.**](http://onlinelibrary.wiley.com/doi/10.1002/9781118354599.ch4/summary)

**[3]** [**An analysis of protein domain linkers: their classification and role in protein folding**](http://peds.oxfordjournals.org/content/15/11/871.abstract)**. George, Richard A., and Jaap Heringa. *Protein Engineering* 15.11 (2002): 871-879.**

# **[4]** [**Design of the linkers which effectively separate domains of a bifunctional fusion protein**](http://peds.oxfordjournals.org/content/14/8/529.long)**. Arai, Ryoichi, et al. *Protein engineering* 14.8 (2001): 529-532.**

**[5]** [**Design and optimization of a linker for fusion protein construction**](http://www.sciencedirect.com/science/article/pii/S100200710900210X)**. Zhang, Jianhua, et al. *Progress in Natural Science* 19.10 (2009): 1197-1200.**

La Plata, 12 de Septiembre de 2014

Sr. Decano de la Facultad de Ciencias Exactas

de la Universidad Nacional de La Plata (UNLP).

Prof. Dr. Carlos Naón

S/D

Me dirijo a Ud. a efectos de presentarle el plan de trabajo para alcanzar el grado de Licenciado en Biotecnología y Biología Molecular que corresponde a la asignatura Laboratorio de Procesos Biotecnológicos (trabajo final) perteneciente al último año de la carrera. El mismo se realizará bajo la dirección de Ignacio E. Sánchez como director en el Laboratorio de Fisiología de Proteínas, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Sin más, saludo a Ud. muy cordialmente.

Datos del alumno:

Nombre y Apellido: Ignacio Eguinoa

Nº alumno: 56112/7

DNI: 32393079

e-mail: ignacio.eguinoa@gmail.com

Firma Alumno.

Nombre y Apellido del Director: Ignacio E. Sánchez Firma:

Ciudad Autónoma de Buenos Aires, 12 de Septiembre de 2014

Sr. Decano de la Facultad de Ciencias Exactas

de la Universidad Nacional de La Plata (UNLP).

Prof. Dr. Carlos Naón

S/D

Me dirijo a Ud. a efectos de presentarle conformidad para la realización del trabajo final correspondiente a la carrera de Lic. en Biotecnologia y Biologia Molecular por parte del alumno Ignacio Eguinoa, a desarrollarse en el laboratorio de Fisiología de Proteínas, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

Sin más, saludo a Ud. muy cordialmente.