

## Conjuntos conformacionales de Proteínas Intrínsecamente Desordenadas moldean las velocidades de evolución dando origen a patrones conformacionales

Julia Marchetti<sup>1,\*</sup>, Nicolas Palopoli<sup>1,\*</sup>, Alexander Miguel Monzon<sup>2</sup>, Diego Javier Zea<sup>3</sup>, Maria Silvina Fornasari<sup>1</sup>, Silvio C.E. Tosatto<sup>2</sup> and Gustavo Parisi<sup>1</sup>.

<sup>1</sup> Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, CONICET, Bernal, Buenos Aires, Argentina

<sup>2</sup> Department of Biomedical Sciences, University of Padua, Padua, Italy

<sup>3</sup> Sorbonne Université, CNRS, IBPS, Laboratoire de Biologie Computationnelle et Quantitative (LCQB), Paris, France

\* Estos autores contribuyeron de igual manera a la elaboración del trabajo.

### Introducción

Las proteínas intrínsecamente desordenadas (Intrinsically Disordered Proteins, IDPs) son cadenas de aminoácidos que carecen de una estructura estable en condiciones fisiológicas. Esta característica las diferencia de las proteínas más tradicionales y las hace particularmente desafiantes para la biología estructural y la biología evolutiva. En este trabajo se propone usar un enfoque evolutivo para la descripción de la extensa diversidad conformacional presente en cada IDP y comprender cuáles son los aspectos estructurales que impactan de manera directa sobre la velocidad de evolución en cada una de sus posiciones.

Los experimentos de determinación estructural de las proteínas proveen información sobre las coordenadas espaciales de cada uno de los átomos que las conforman. La ventaja de la Resonancia Magnética Nuclear (NMR, por sus siglas en inglés) sobre otras técnicas similares es que permite obtener información estructural de múltiples modelos en simultáneo, que representan en su conjunto el comportamiento de la proteína en solución, cercano a su comportamiento natural en la célula.

### Metodología y herramientas

Para este análisis se recolectó información estructural de *ensembles* o conjuntos de conformaciones alternativas de 311 proteínas intrínsecamente desordenadas, provenientes de experimentos de NMR. Para cada uno de sus modelos estructurales se determinó el carácter de desorden intrínseco posición a posición. También se estimó la tasa sitio-específica de intercambio por otros aminoácidos a lo largo de la evolución, inferida a partir de alineamientos de proteínas vinculadas evolutivamente, con el programa Rate4Site (1). Por lo tanto en este trabajo se contó con información sitio-específica y modelo-específico, tanto estructural como evolutiva, dando lugar a un total de 790,128 datos para analizar.

Utilizamos paquetes estándar de R para llevar a cabo los análisis necesarios para derivar información científica relevante sobre nuestro conjunto de datos. En particular y para cada conformación presente en el conjunto conformacional de una proteína dada, consideramos cada uno de sus aminoácidos por separado y exploramos la relación entre los contactos establecidos con otros aminoácidos de la proteína, su carácter de desorden intrínseco y su velocidad de evolución.

Para el análisis estadístico se utilizaron funciones del paquete tidyverse (2), mientras que la visualización de datos y la generación de gráficos para publicación se realizó con ggplot2 (3) , ggpubr (4), patchwork (5) y ggtext (6). El reporte general del trabajo se realizó con R Markdown. Junto con el repositorio abierto en gitlab (<https://gitlab.com/sbgung/publications/palopoli-marchetti-2020-rates>) esto permitió asegurar la reproducibilidad del análisis, pilar de la ciencia moderna y requisito para su publicación (7)

## Resultados

Encontramos que las IDPs presentan una gran heterogeneidad en su velocidad de evolución sitio-específica, asociadas a diferentes restricciones estructurales impuestas por los contactos establecidos entre residuos. La correlación entre el establecimiento de contactos y las velocidades de evolución mejora cuando se tiene en cuenta información estructural derivada de múltiples cónfórmers, pero no del conjunto redundante de estructuras disponibles. El análisis simultáneo de distintas proteínas permitió identificar perfiles de velocidad característicos y compartidos entre ellas, que se vinculan con la alternancia de regiones ordenadas y desordenadas en las proteínas.

## Bibliografía

1. Pupko T, Bell RE, Mayrose I, Glaser F, Ben-Tal N. Rate4Site: an algorithmic tool for the identification of functional regions in proteins by surface mapping of evolutionary determinants within their homologues. *Bioinformatics*. 2002;18 Suppl 1:S71-7.
2. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *JOSS*. 2019 Nov 21;4(43):1686.
3. H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
4. Alboukadel Kassambara (2020). ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0.<https://CRAN.R-project.org/package=ggpubr>
5. Thomas Lin Pedersen (2020). patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>
6. Claus O. Wilke (2020). ggtext: Improved Text Rendering Support for 'ggplot2'. R package version 0.1.1. <https://CRAN.R-project.org/package=ggtext>
7. Palopoli N, Marchetti J, Monzon AM, Zea DJ, Tosatto SCE, Fornasari MS, et al. Intrinsically disordered protein ensembles shape evolutionary rates revealing conformational patterns. *J Mol Biol*. 2021 Feb 5;433(3):166751.