

La evolución dirigida y la coevolución tecnologías de frontera para mejorar las funciones en las proteínas

Rodrigo Arreola Barroso¹ , Verónica Quintero-Hernández² , Jesús Muñoz-Rojas³ , América Rivera-Urbalejo⁴ , Víctor Rivelino Juárez-González^{5*} 

¹ Departamento de Ingeniería Celular y Biocatálisis, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos 62210 México.

² CONACYT-Laboratorio de Ecología Molecular Microbiana, Centro de Investigaciones en Ciencias Microbiológicas, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla (BUAP), Puebla, México.

³ Grupo Ecology and Survival of Microorganisms, LEMM, CICM, IC, BUAP, Puebla, México. Edificio 103 J, Ciudad Universitaria, San Manuel, Puebla, México. C. P. 72570.

⁴ Facultad de Estomatología, BUAP, Puebla, México.

⁵ Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos. México. CP: 62210.

*Email autor corresponsal: [*rivelino.juarez@ibt.unam.mx](mailto:rivelino.juarez@ibt.unam.mx)

Recibido: 26 diciembre 2021. **Aceptado:** 11 marzo 2022

RESUMEN

En los últimos 10 años, los científicos han logrado simular la evolución natural en sus laboratorios. Este proceso se denomina evolución dirigida, e inicia con un gen o secuencia diana. A partir de esta secuencia del gen se generan bibliotecas de genes mutantes utilizando mutagénesis aleatoria mediante la técnica de PCR propensa a error: epPCR. Las librerías generadas son sometidas a procesos de selección o tamizaje para identificar las proteínas ganadoras de una función o funciones en particular, con respecto a la proteína de la cual se originaron.

Otra manera de identificar los productos generados en este largo proceso evolutivo es por medio de los análisis de coevolución, los cuales son de mucha ayuda cuando no existen procesos de selección y el método de tamizaje disponible para encontrar mutantes es muy lento, limitando el número de mutantes a muestrear para obtener una variante mejorada. En la coevolución, se analizan los patrones de variación de residuos que forman un contacto entre sí, porque dichas mutaciones cambian el empaquetamiento de la proteína y pueden provocar modificaciones de función, estabilidad, selectividad, incluso cuando están lejos del sitio activo. La presión de selección a la que son sometidas las proteínas hace que dos residuos que se encuentran en contacto varíen de forma correlacionada cuando son importantes para la función.

Por lo anterior, la evolución dirigida aunada a la coevolución son unas de las técnicas más exitosas de la biología molecular y de frontera que nos permiten explorar y mejorar las funciones en una gran cantidad de proteínas de interés médico, biotecnológico, industrial, etc, en un tiempo muy corto.

Palabras clave: evolución dirigida; coevolución; proteínas; despliegue en fagos.

ABSTRACT

In the last 10 years, scientists have been able to simulate natural evolution in laboratories through the process of directed evolution, which begins with a target gene or sequence for which a mutant gene library is generated using random mutagenesis. After producing such libraries through the technique of error-prone PCR (epPCR), they are subjected to selection or screening for the identification of improved proteins towards a particular function or functions when compared to the protein that originated them.

Another methodology to harness the products of the long evolutionary history is the coevolution analysis, which is very helpful when no selection strategy is feasible and the screening method available to sift mutants is very slow, limiting the number of mutants that can be sampled to find an improved variant. In the coevolution analysis, the variation patterns of residues in contact are analyzed, because mutations in these residues change the residue packing in the protein, thus, producing modifications in function, stability, selectivity, even when residue substitution is far from the active site. The selection acting on proteins force two residues in contact to vary in a correlated way when they are important for function.

Therefore, directed evolution coupled with coevolution are some of the most successful techniques in molecular and frontier biology that allow us to explore and improve the functions of a large number of proteins of medical, biotechnological, industrial interest, etc., in a very short time.

Keywords: coevolution; directed evolution; phage display; proteins.

INTRODUCCIÓN

El diseño de proteínas se ha enriquecido al entender los procesos mediante los cuales los seres vivos han logrado que las proteínas trabajen de manera coordinada en un espacio y

tiempo con las demás moléculas (ADN, ARN, lípidos, carbohidratos, agua, etc), que forman parte de la célula en la que se encuentran, cumpliendo funciones que permiten la vida en los organismos (Figura 1; Figura 2).

Proteínas en *Escherichia coli*

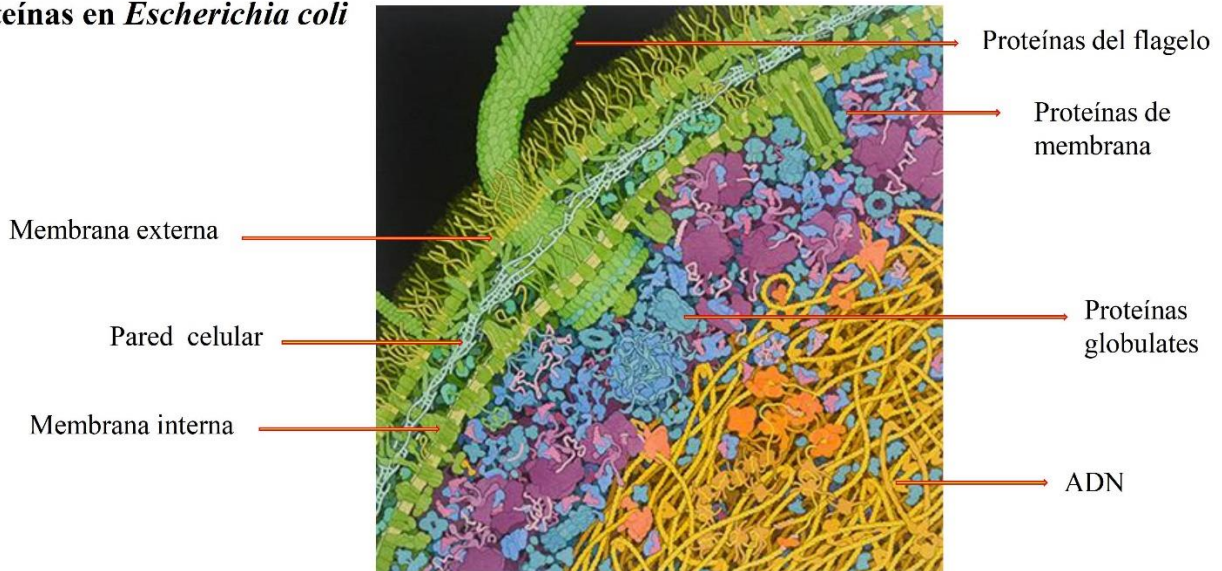


Figura 1. La bacteria *Escherichia coli*. Pintura basada en evidencia experimental sobre el grado de empaquetamiento que tienen las proteínas de la bacteria *E. coli*. Imagen generada por David S. Goodsell. RCSB Protein Data Bank; doi: 10.2210/rcsb_pdb/goodsell-gallery-028.

Proteínas y el ADN en el núcleo de una célula

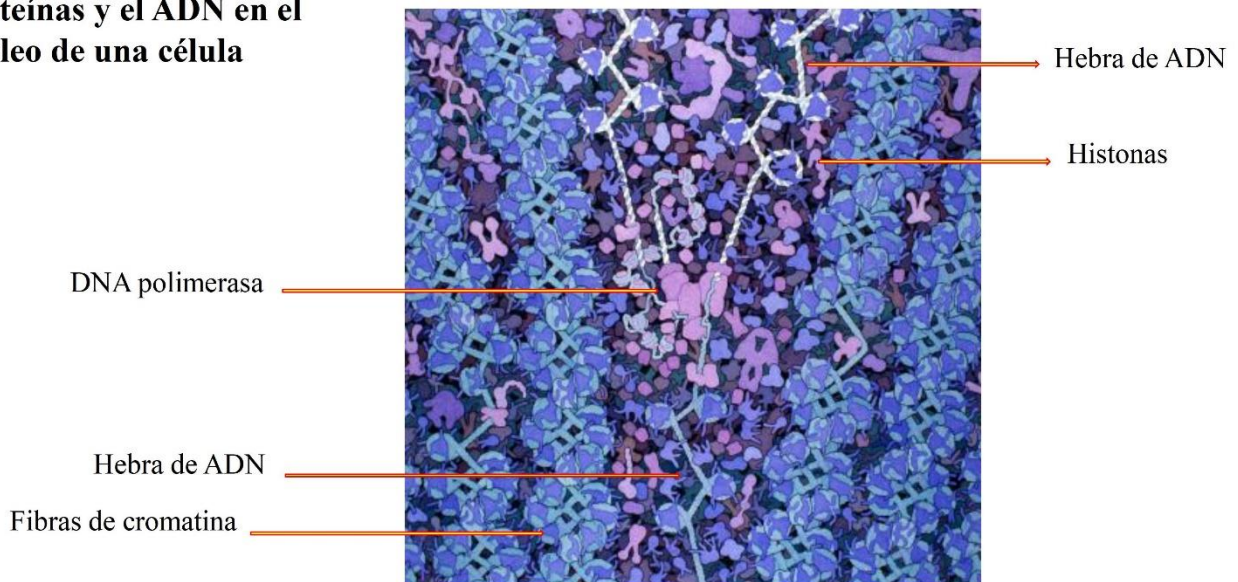


Figura 2. La replicación del ADN en el núcleo de una célula. Pintura basada en evidencia experimental sobre la replicación en una célula eucariote, donde se observa el grado de empaquetamiento de las proteínas y el ADN. Imagen generada por David S. Goodsell. RCSB Protein Data Bank; doi: 10.2210/rcsb_pdb/goodsell-gallery-00.

La estructura primaria o secuencia de aminoácidos de una proteína determina la función cuando la proteína adopta su estructura terciaria o tridimensional. Por ejemplo, en el caso de las enzimas, determinan dos aspectos fundamentales: la posición y orientación correcta en el espacio de los aminoácidos que componen el sitio activo, ambos factores son importantes para poder interactuar con sus sustratos y catalizar la reacción. De igual forma en las interfaces presentes en las proteínas que funcionan formando oligómeros (dímeros, trímeros y tetrameros, etc), son importantes las posiciones y orientaciones de los aminoácidos que constituyen estas interfaces para el adecuado funcionamiento de las proteínas oligoméricas [1–6].

El empaquetamiento de aminoácidos en las proteínas en su interior es fundamental para que cada aminoácido interactúe con los aminoácidos vecinos, tanto al interior de la proteína como al exterior con las moléculas aledañas, para poder cumplir sus funciones celulares (Figura 1; Figura 2). De tal forma, las mutaciones en los aminoácidos empaquetados en las proteínas pueden provocar cambios locales o globales en la estructura de la misma [7].

Los alineamientos múltiples

Una de las formas de visualizar y entender la presencia de mutaciones acumuladas de manera natural en las secuencias de las proteínas, es por medio de los análisis de alineamientos múltiples de secuencias de aminoácidos de proteínas de diferentes especies. Para este fin,

se utilizan diversos servidores para emprender este análisis, como: Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) [8], BLAST (<https://www.ncbi.nlm.nih.gov/home/proteins/>) [9], T-COFFEE (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) [10], entre otros. Con estos análisis se pueden reconocer a las proteínas que llevan a cabo la misma función al comparar porcentajes de identidad entre secuencias de aminoácidos, identificando regiones conservadas en la secuencia de aminoácidos. Estos residuos conservados son especialmente importantes para la función, estabilidad, afinidad, especificidad de las proteínas y las propiedades que podemos medir experimentalmente.

Por otra parte, los alineamientos múltiples de secuencias nos han llevado a inferir que una familia de proteínas comparte un ancestro común, el cual fue divergiendo genéticamente para dar lugar a las proteínas que se encuentran hoy en día en los seres vivos y en virus. Dentro de estos, los genes ortólogos producen proteínas que llevan a cabo la misma función en microorganismos u organismos diferentes y tienen secuencias de aminoácidos similares entre sí [11]. Los alineamientos múltiples también permiten darnos cuenta de los procesos de divergencia de las proteínas. Estos procesos de divergencia inician cuando aparece una nueva especie o después de un proceso de duplicación genética que origina dos copias de un gen para una proteína, una de las cuales adquiere función diferente en el mismo microorganismo u organismo [11].

Los cambios que llevan a la divergencia

ocurren en nuestro material genético. De manera general, sabemos que la secuencia de una proteína está contenida en su ARN (ácido ribonucleico), que a su vez está codificada en un gen de ADN (ácido desoxirribonucleico), el cual es el principal constituyente del material genético de los seres vivos. Esta molécula almacena la secuencia que va a dar origen a cada proteína en fragmentos de la misma molécula llamados genes. Cada vez que se separan las cadenas de ADN para el proceso de replicación (generar copias) se pueden generar mutaciones durante el proceso debido a errores de copiado, los cuales sustituyen unos nucleótidos (formados por una *base* nitrogenada y un azúcar desoxirribosa unida a un grupo fosfato) por otros, dando lugar a la generación de nuevas secuencias de proteínas con cambios de aminoácidos.

Durante la evolución natural las mutaciones se van acumulando por millones de años y restringiendo mediante procesos de selección. A pesar de que las mutaciones ocurren a baja frecuencia y de manera azarosa, pueden mantenerse en los eventos de replicación del ADN de futuras generaciones aquellas que proporcionen una ventaja a la proteína y al organismo sobre los demás organismos. Los procesos de generación de variabilidad y selección de las mutaciones benéficas para una función en particular, es lo que genera la evolución de las especies [12].

La evolución dirigida como fuente de variación en los genes

La evolución dirigida o “directed evolution” en

inglés, está basada en los principios Darwinianos de variación y selección. Esta herramienta inicia con un gen o secuencia diana a partir del cual se generan bibliotecas de genes mutantes utilizando mutagénesis aleatoria por medio de la técnica de PCR propensa a error, epPCR (“error prone PCR” en inglés) (Figura 3). Esto permite generar bibliotecas del gen o genes diana con diferentes tasas de mutación controlando en cada una el grado de variabilidad con lo cual se obtienen bibliotecas de mutagénesis baja, media y alta [13]. Existen otras técnicas para generar variabilidad en las poblaciones de genes tales como: la recombinación de genes mutados por barajeo (“gene shuffling”, en inglés), en donde se crean quimeras con trozos de genes (recombinación) que ya tienen mutaciones que les confieren alguna ventaja; una mezcla que incluya al gen silvestre puede eliminar las mutaciones que no contribuyen al mejoramiento de la función [14–16]. Este proceso no solo se aplica a genes mutantes, también podemos realizar el barajeo de familias de genes (“family shuffling” en inglés), donde, a partir de una mezcla de las secuencias de genes de una familia se crea una biblioteca de genes quiméricos, a partir de los cuales se seleccionan las quimeras deseadas [17]. En todas las técnicas para generar variación, las bibliotecas resultantes son clonadas en vectores de expresión y luego transformadas en bacterias para la expresión de las proteínas mutagenizadas con el fin de identificar mejoras en su función.

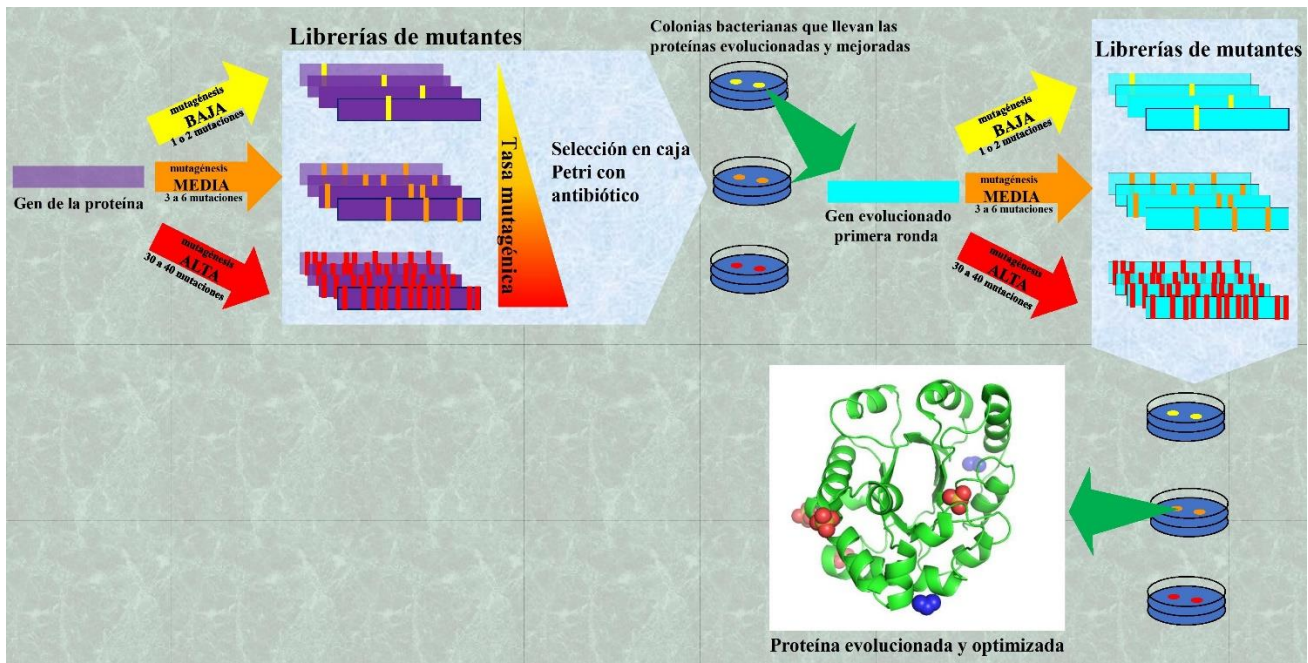


Figura 3. Proceso de evolución dirigida y selección. Se usan librerías con diferentes tasas de mutación para la generación de una proteína mejorada, solo crecen las colonias ganadoras de la función evaluada.

Las proteínas mutantes que han mejorado la función deseada son reconocidas por dos métodos: 1) En el método de selección crecen únicamente las colonias bacterianas que contienen un plásmido con mejoras en el gen deseado; estos genes deben codificar proteínas que desempeñan funciones esenciales en un microorganismo u organismo, o condicionar la función de uno de estos genes; la mayoría de las clonas contenidas en las bibliotecas mutagénicas en este tipo de métodos no son capaces de crecer [16, 17] (Figura 3) y 2) En el método de tamizado o *screening*, la mayoría de las colonias bacterianas que fueron transformadas con plásmidos crecen y los investigadores deben seleccionar las proteínas que han ganado la función que ellos deseen, por ejemplo, por medio del despliegue en fagos o *phage display* [18, 19] (Figura 4). Mediante ambos métodos se han aislado una gran

cantidad de proteínas con funciones mejoradas con respecto a la proteína silvestre o la proteína de la cual partieron. Todo esto ha colocado a la evolución dirigida, en los últimos 10 años, como una de las técnicas más exitosas que se emplean en ciencia de frontera para explorar y mejorar las funciones en una gran cantidad de proteínas de interés médico, biotecnológico, industrial, etc.

Aunado a lo anterior, la evolución dirigida permite acortar los tiempos de la evolución natural para obtener mejoras en la característica deseada. Lo que a la naturaleza le toma millones de años puede realizarse en solo unos cuantos meses de trabajo en el laboratorio, permitiendo a los expertos en esta área generar diversidad en las secuencias nucleotídicas de los genes de las proteínas para incrementar su función.

El premio Nobel de Química por la evolución dirigida y el despliegue en fagos

La primera investigadora en la lista de galardonados fue la Dra. Frances Hamilton Arnold la cual ha desarrollado la evolución dirigida para trabajar durante muchos años el mejoramiento catalítico de las reacciones de enzimas usadas, por ejemplo, para el mejoramiento de biocombustibles. La Dra. Arnold también desarrolló la metodología SCHEMA para generar genes quiméricos con el objetivo de obtener nuevas proteínas [22], e incluso género enzimas capaces de formar enlaces no presentes en las moléculas biológicas como el enlace Si-C [23,24].

Los siguientes galardonados fueron los Dres. George P. Smith [25] y Gregory P. Winter [26], de los cuales Smith es considerado el padre del despliegue en fagos (Phage Display en inglés). Este novedoso y poderoso método de tamizado permitió a los investigadores Smith y Winter evaluar, con una velocidad nunca antes vista, bibliotecas de distintos tipos de moléculas (péptidos, citosinas, anticuerpos, enzimas, proteínas de unión a ADN, etc). Para lograrlo, las diferentes proteínas a probar se fusionaron a las proteínas que conforman la cubierta de los fagos filamentosos tipo M13, lo que permitió identificar a las mutantes más activas frente a un sustrato por competencia y su posterior secuenciación de los fagos ganadores.

Lo descrito en el párrafo anterior, colocó al despliegue en fagos como una de las metodologías más exitosas, económicas, eficientes y fáciles de utilizar para la selección de las proteínas mutantes de interés

biotecnológico, médico, etc [27–30]. Este método de detección puede analizar millones de variantes en pocos días (Figura 4). En el despliegue en fagos, los virus que se utilizan presentan la capacidad de infectar de manera natural a las bacterias como *Escherichia coli* mediante la proteína pIII. Esta proteína interacciona con el pili de *E. coli*, introduciendo su material genético con una alta eficiencia, incluso varios órdenes de magnitud si se compara con la eficiencia de otros métodos tradicionales de transformación como la electroporación, que introduce el ADN a las células bacterianas utilizando celdas especiales y descargas eléctricas, o la transformación química, la cual somete a diferentes choques de calor a las células bacterianas para introducir el ADN de interés.

Nosotros, por evolución dirigida y despliegue en fagos, fuimos capaces de generar varios anticuerpos en los formatos scFv (30 kDa) y Fab (50 kDa), con propiedades neutralizantes de la toxina Cn2 del alacrán *Centruroides noxius*, considerado el alacrán más peligroso de México para los humanos. Estos anticuerpos permitieron neutralizar los efectos de la toxina Cn2 en ratones hembra cepa CD1, de 20 g de peso y utilizando una dosis letal 50 de toxina (LD50) [31,32].

Por evolución dirigida se han explorado muchas funciones en diversas proteínas, como ejemplo se encuentran las enzimas: epóxido hidrolasa, glyfosato, n-acetyltransferasa, xylanasa y fosfotriesterasa, en cuyas variantes se ha observado mejoramiento en la especificidad, las constantes catalíticas, como la constante Km

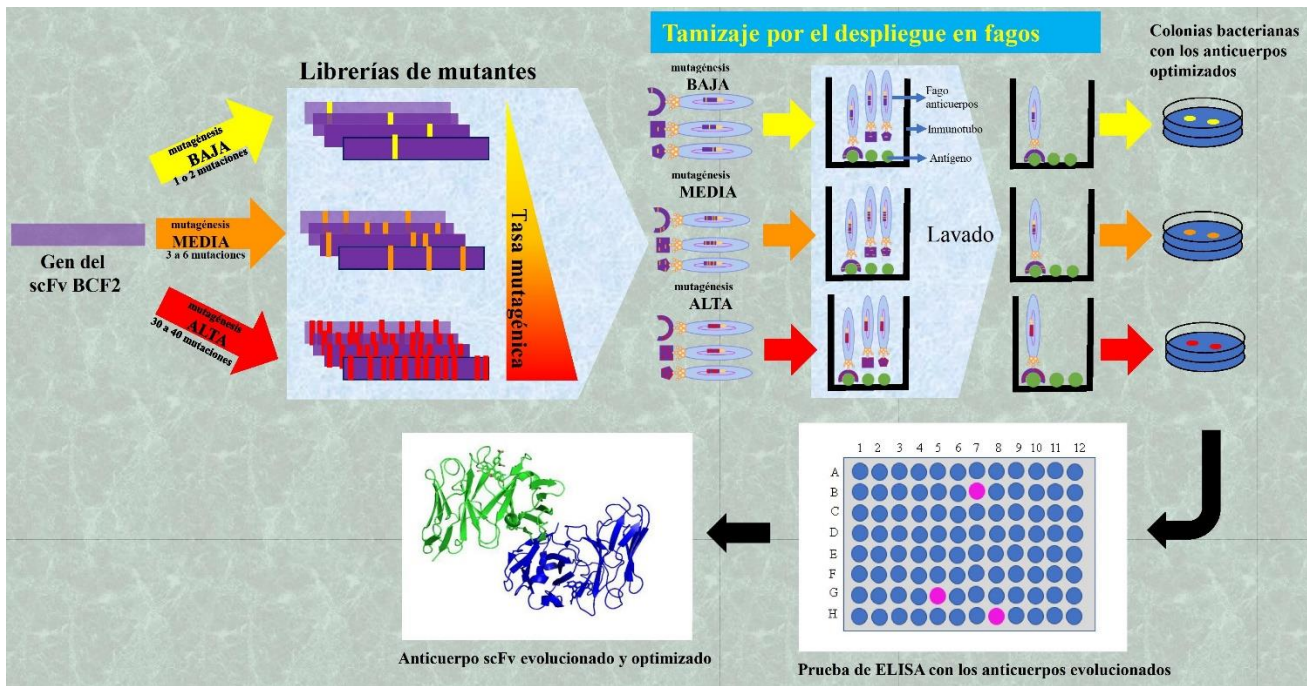


Figura 4. La evolución dirigida con tamizaje por despliegue en fagos. Permite producir anticuerpos neutralizantes de moléculas con interés biotecnológico y médico.

(afinidad de la enzima por su sustrato; a menor valor de K_m mayor es la afinidad de la enzima por su sustrato) y la V_{max} (velocidad máxima de la reacción), la estabilidad y la solubilidad [33]. El mejoramiento de las propiedades de estas enzimas es primordial para su uso en la industria y sus diversas aplicaciones biotecnológicas, médicas, alimenticias, etc [34,35].

Saab-Rincón y coautores (incluyendo al autor correspondiente de esta revisión), por evolución dirigida, se logró obtener una variante que mostró un incremento en la actividad de la enzima monomérica de la triosa fosfato isomerasa de *Trypanosoma brucei*, la cual fue expresada en la bacteria *Escherichia coli*. Para lograr lo anterior, fue necesario interrumpir el gen cromosomal que codificaba

para la triosa fosfato isomerasa de *Escherichia coli*, esto nos permitió seleccionar a las mutantes de la enzima monomérica de *T. brucei*, mejoradas por evolución dirigida, las cuales se identificaron midiendo sus constantes catalíticas [36].

Los aportes generados por la Dra. Frances Hamilton Arnold, el Dr. George P. Smith y el Dr. Gregory P. Winter, revolucionaron los métodos para explotar la variación y para seleccionar a las mutantes de proteínas con funciones mejoradas, las cuales fueron aisladas de bibliotecas integradas por millones de mutantes generadas en tiempos muy cortos, comparados con los millones de años que tardó la evolución natural para formarlas y seleccionarlas.

La coevolución fuente de variación en las proteínas

Otra forma de aprovechar los productos del largo proceso evolutivo son los análisis de coevolución. Dichos análisis son de mucha ayuda cuando no se tienen procesos de selección y el tamizaje de mutantes disponible es lento, pues pueden limitar el número de mutantes a muestrear para obtener una mejorada. Para hacerlo, se analiza la forma en que varían los residuos formando un contacto; las mutaciones en estos residuos cambian el empaquetamiento de la proteína y pueden causar modificaciones de función, estabilidad, selectividad aun cuando estén lejos del sitio activo.

La selección a la que son sometidas las proteínas hace que dos residuos en contacto varíen de forma correlacionada cuando son importantes para la función. Una mutación que produce un efecto desfavorable en una proteína requiere una segunda mutación que compense el efecto de la primera para mantener su estructura y función. El mejor sitio para que ocurra esta mutación compensatoria es en alguno de los residuos en contacto con el primero. Dicho patrón de variación resulta en pares de contacto que covarían. Esta covariación puede detectarse de forma directa, comparando los contactos entre estructuras de ortólogos, o indirecta, analizando los alineamientos múltiples de secuencias de aminoácidos (Figura 5). Rama Ranganathan, en 2018, fue el primero en proponer una función energética para describir esta correlación y predecir residuos interaccionando. Este análisis

ha sido usado para identificar residuos de aminoácidos importantes para la función, redes de interacción alostérica y patrones de contacto característicos de familias de proteínas [37].

Los análisis de coevolución fueron creados con el propósito de identificar residuos que están interaccionando para mantener la estructura o permitir la función. Esto ha permitido que se apliquen en la búsqueda de sitios alostéricos, que afectan el sitio activo por la unión de una molécula en otro sitio. Los sitios alostéricos se unen al sitio activo por medio de residuos cuya interacción y evolución están acopladas.

Dado que al analizar la coevolución de aminoácidos se identifican residuos que están interaccionando, se trató de extender su uso a la detección de contactos entre residuos. Se quería aplicar la detección de interacciones para predecir estructuras tridimensionales. Un problema al que se enfrentó esta aplicación es que esta predicción depende de la calidad del alineamiento: familias de proteínas con alto nivel de variación en los aminoácidos de sus secuencias pueden alinearse con menor nivel de confianza en muchos de sus segmentos de secuencia. Además, los residuos que están coevolucionando pueden encontrarse distantes e interactuar mediante una red de residuos: es necesario diferenciar estos residuos de los que están en contacto directo para predecir los residuos en contacto en la estructura.

La calidad del alineamiento y la detección de residuos en contacto directo han sido problemas que se han atacado con diversos algoritmos. Por ejemplo, el problema de diferenciar residuos en contacto directo de otros acoplamientos tuvo su

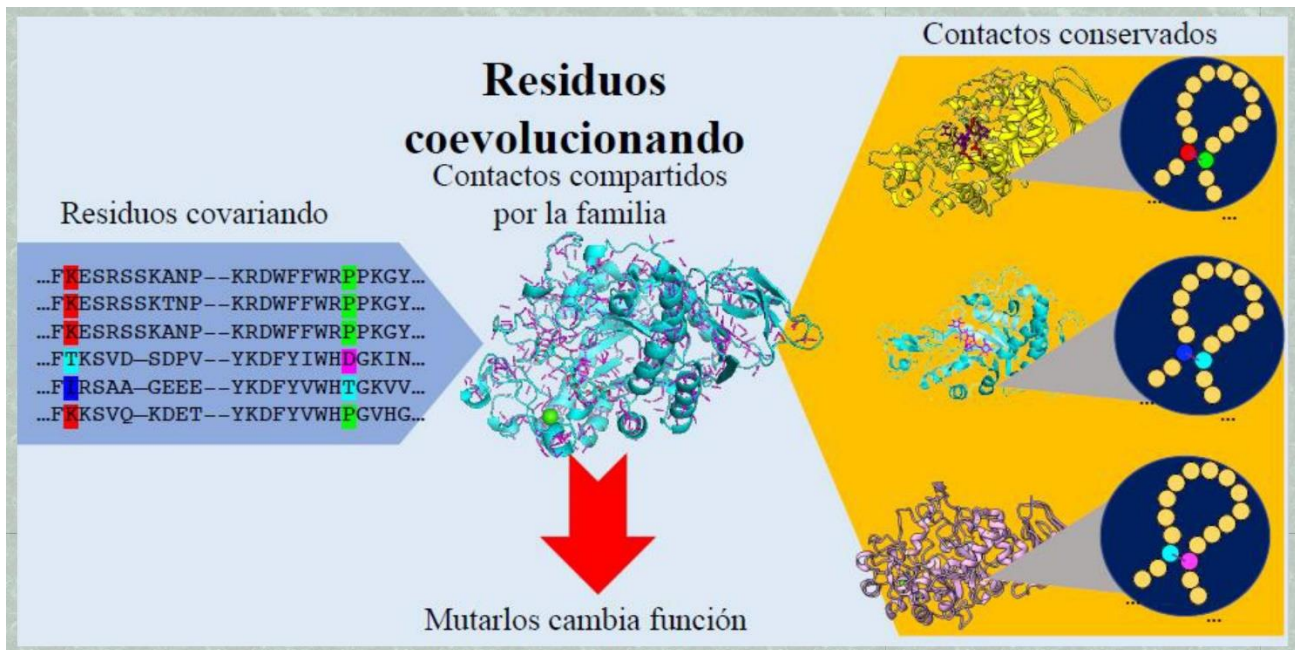


Figura 5. Identificación de residuos coevolucionando. Las parejas de aminoácidos en contacto se conservan cuando son importantes para la función. El cambio de un residuo en el par debe compensarse por la variación del otro y así mantener la función de la proteína.

primer éxito tras el trabajo del grupo liderado por Terence Hwa, Martin Weigt and José N. Onuchic [36, 37]. Este algoritmo calcula el acoplamiento directo aplicando estadística bayesiana (basada en aspectos del método científico para recolectar evidencia consistente o no con una hipótesis dada, a medida que la evidencia se acumula, el grado de creencia en esa hipótesis se va modificando) para crear un modelo estadístico global, cuyos resultados debían ajustarse al análisis de acoplamiento por pares y a propiedades del sistema como la entropía de cada residuo. Recientemente, mediante el uso del programa Alpha Fold, el cual aplica inteligencia artificial para ajustar los resultados del análisis acoplamiento directo a las estructuras tridimensionales (3D) disponibles en el PDB o Protein Data Bank (banco de información estructural de las

macromoléculas a nivel atómico determinadas por la comunidad científica de todo el mundo, las direcciones web de acceso son: www.pdb.org (wwPDB: Worldwide Protein Data Bank y RCSB PDB: Homepage y <https://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/introduction>). Así, Alpha Fold ha sido capaz de calcular estructuras 3D con la misma precisión que la difracción de rayos X, aún para secuencias de proteínas cuya estructura no ha sido determinada y depositada en PDB [40]. Nosotros, por medio de estudios de coevolución, hemos modificado la selectividad de las enzimas glicosido hidrolasas de la familia GH13, también conocida como de las α -amilasas. Logrando aumentar la relación hidrólisis/transglicosidación (ruptura/transferencia de cadenas de azúcares) hasta 5 veces [41]. Además, estos cambios de selectividad se

lograron mutando residuos fuera del sitio catalítico (que es el foco de estudios que no incluyen evolución dirigida). Cambiar la preferencia de una enzima de romper cadenas de azúcares para que pegue azúcares en otras moléculas tiene una gran importancia biotecnológica porque puede permitir la creación de herramientas de glicosíntesis (usadas en la síntesis de fármacos, surfactantes y mejoramiento de probióticos) y ajustar la selectividad para una función específica (generación de fructosa a partir de almidón, tratamiento de textiles, producción de dextrinas).

CONCLUSIÓN

Los cambios de función, selectividad, creación de nuevas funciones y mejoramiento de los antivenenos han sido logros de la evolución dirigida y la coevolución. Estos han sido posibles gracias a la inspiración que ha obtenido del estudio del fenómeno de evolución y de sus productos, los ácidos nucleicos (ADN y ARN) y las proteínas. Los investigadores se preguntarán: ¿Qué más podemos aprender del estudio de los procesos evolutivos? solo el tiempo lo dirá, pero hasta ahora la evolución dirigida nos ha permitido generar moléculas en tiempo récord para obtener nuevas funciones y es una poderosa herramienta que nos permite modificar las moléculas creadas por la naturaleza para obtener otras que se ajusten a nuestras necesidades.

Finalmente, hay que resaltar que las novedosas y poderosas técnicas de evolución dirigida y

coevolución nos acercarán más al entendimiento de la evolución y al mejoramiento de las funciones en las proteínas para los fines que la ciencia y la humanidad necesiten.

CONFLICTO DE INTERESES

Los autores declaran la ausencia de conflictos de interés.

AGRADECIMIENTOS

Los autores agradecen al Mtro. Juan Manuel Hurtado Ramírez, M.E.M. David Santiago Castañeda Carreón, Ing. Roberto P. Rodríguez Bahena y L.I. Servando Aguirre Cruz por su asistencia técnica en computación.

REFERENCIAS

- [1]. Brinda K V., Vishveshwara S. Oligomeric protein structure networks: Insights into protein-protein interactions. *BMC Bioinformatics* 2005;6. <https://doi.org/10.1186/1471-2105-6-296>.
- [2]. Mei G, Di Venere A, Rosato N, Finazzi-Agrò A. The importance of being dimeric. *FEBS J* 2005;272:16–27. <https://doi.org/10.1111/j.1432-1033.2004.04407.x>.
- [3]. Mallik S, Kundu S. Topology and Oligomerization of Mono- and Oligomeric Proteins Regulate Their Half-Lives in the Cell. *Structure* 2018;26:869-878.e3. <https://doi.org/10.1016/j.str.2018.04.015>.
- [4]. Kumari N, Yadav S. Modulation of protein

- oligomerization: An overview. *Prog Biophys Mol Biol* 2019;149:99–113. <https://doi.org/10.1016/j.pbiomolbio.2019.03.003>.
- [5]. Roche JV, Törnroth-Horsefield S. Aquaporin protein-protein interactions. *Int J Mol Sci* 2017;18:1–23. <https://doi.org/10.3390/ijms18112255>.
- [6]. Watson A, Phipps MJS, Clark HW, Skylaris CK, Madsen J. Surfactant proteins A and D: Trimerized innate immunity proteins with an affinity for viral fusion proteins. *J Innate Immun* 2018;11:13–28. <https://doi.org/10.1159/000492974>.
- [7]. Brender JR, Zhang Y. Predicting the Effect of Mutations on Protein-Protein Binding Interactions through Structure-Based Interface Profiles. *PLoS Comput Biol* 2015;11:1–25. <https://doi.org/10.1371/journal.pcbi.1004494>.
- [8]. Sievers F, Higgins DG. Clustal Omega. *Curr Protoc Bioinforma* 2014;2014:3.13.1-3.13.16. <https://doi.org/10.1002/0471250953.bi0313s4>.
- [9]. Kent WJ. BLAT —The BLAST -Like Alignment Tool . *Genome Res* 2002;12:656–64. <https://doi.org/10.1101/gr.229202>.
- [10]. Notredame C, Higgins DG, Heringa J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 2000;302:205–17. <https://doi.org/10.1006/jmbi.2000.4042>.
- [11]. Koonin E V. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* 2005;39:309–38. <https://doi.org/10.1146/annurev.genet.39.073003.114725>.
- [12]. Imhof M, Schlötterer C. Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc Natl Acad Sci U S A* 2001;98:1113–7. <https://doi.org/10.1073/pnas.98.3.1113>.
- [13]. Cadwell RC, Joyce GF. Manual Supplement III1 | 11 Mutagenic PCR. *Genome Res* 1994;3–8.
- [14]. Stemmer WP. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc Natl Acad Sci [Internet]*. 1994 Oct 25;91(22):10747–51. Available from: <https://doi.org/10.1073/pnas.91.22.10747>.
- [15]. Mikhailovsky G, Gordon R. Shuffling type of biological evolution based on horizontal gene transfer and the biosphere gene pool hypothesis. *BioSystems* 2020;193–194:104131. <https://doi.org/10.1016/j.biosystems.2020.104131>.
- [16]. Feng H, Wang HY, Zhao HY. Novel random mutagenesis method for directed evolution. *Methods Mol Biol* 2017;1498:483–90. https://doi.org/10.1007/978-1-4939-6472-7_32.
- [17]. Behrendorff JBYH, Johnston WA, Gillam EMJ. Chapter 12 Restriction Enzyme-Mediated DNA Family Shuffling 2014;1179:175–87. <https://doi.org/10.1007/978-1-4939-1053-3>.
- [18]. Yang F, Moss LG, Phillips GN. The Molecular Structure of Green Fluorescent

- Protein. *Nat Biotechnol* 1996;14:1246–51. <https://doi.org/10.1038/nbt1096-1246>.
- [19]. Kuchner O, Arnold FH. Directed evolution of enzyme catalysts. *Trends Biotechnol* 1997;15:523–30. [https://doi.org/10.1016/S0167-7799\(97\)01138-4](https://doi.org/10.1016/S0167-7799(97)01138-4).
- [20]. Lemire S, Yehl KM, Lu TK. Phage-Based Applications in Synthetic Biology. *Annu Rev Virol* 2018;5:453–76. <https://doi.org/10.1146/annurev-virology-092917-043544>.
- [21]. Løset GÅ, Sandlie I. Next generation phage display by use of pVII and pIX as display scaffolds. *Methods* 2012;58:40–6. <https://doi.org/10.1016/j.ymeth.2012.07.005>.
- [22]. Meyer MM, Silberg JJ, Voigt CA, Endelman JB, Mayo SL, Wang Z-G, *et al.* Library analysis of SCHEMA-guided protein recombination. *Protein Sci* 2003;12:1686–93. <https://doi.org/10.1110/ps.0306603>.
- [23]. Romero PA, Arnold FH. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 2009;10:866–76. <https://doi.org/10.1038/nrm2805>.
- [24]. Smith MA, Arnold FH. Noncontiguous SCHEMA protein recombination. *Methods Mol Biol* 2014;1179:345–52. https://doi.org/10.1007/978-1-4939-1053-3_23.
- [25]. Smith GP. Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture). *Angew Chemie - Int Ed* 2019;58:14428–37. <https://doi.org/10.1002/anie.201908308>.
- [26]. Winter G. Harnessing Evolution to Make Medicines (Nobel Lecture). *Angew Chemie - Int Ed* 2019;58:14438–45. <https://doi.org/10.1002/anie.201909343>.
- [27]. Petrenko VA. Landscape phage: Evolution from phage display to nanobiotechnology. *Viruses* 2018;10:1–17. <https://doi.org/10.3390/v10060311>.
- [28]. Ledsgaard L, Kilstrup M, Karatt-Vellatt A, McCafferty J, Laustsen AH. Basics of antibody phage display technology. *Toxins (Basel)* 2018;10. <https://doi.org/10.3390/toxins10060236>.
- [29]. Wahlfors J, Loimas S, Pasanen T, Hakkarainen T. Green fluorescent protein (GFP) fusion constructs in gene therapy research. *Histochem Cell Biol* 2001;115:59–65. <https://doi.org/10.1007/s004180000219>.
- [30]. Tan Y, Tian T, Liu W, Zhu Z, J. Yang C. Advance in phage display technology for bioanalysis. *Biotechnol J* 2016;11:732–45. <https://doi.org/10.1002/biot.201500458>.
- [31]. Quintero-Hernández V, Juárez-González VR, Ortiz-León M, Sánchez R, Possani LD, Becerril B. The change of the scFv into the Fab format improves the stability and *in vivo* toxin neutralization capacity of recombinant antibodies. *Mol Immunol* 2007;44:1307–15. <https://doi.org/10.1016/j.molimm.2006.05.009>.
- [32]. Juárez-González VR, Riaño-Umbarila L, Quintero-Hernández V, Olamendi-Portugal T, Ortiz-León M, Ortiz E, *et al.* Directed evolution, phage display and combination of evolved mutants: A strategy to recover the neutralization properties of the scFv version of

BCF2 a neutralizing monoclonal antibody specific to scorpion toxin Cn2. *J Mol Biol* 2005;346:1287–97.

<https://doi.org/10.1016/j.jmb.2004.12.060>.

[33]. Johannes TW, Zhao H. Directed evolution of enzymes and biosynthetic pathways. *Curr Opin Microbiol* 2006;9:261–7. <https://doi.org/10.1016/j.mib.2006.03.003>.

[34]. Kumar A, Singh S. Directed evolution: Tailoring biocatalysts for industrial applications. *Crit Rev Biotechnol* 2013;33:365–78. <https://doi.org/10.3109/07388551.2012.716810>.

[35]. Porter JL, Rusli RA, Ollis DL. Directed Evolution of Enzymes for Industrial Biocatalysis. *ChemBioChem* 2016;17:197–203. <https://doi.org/10.1002/cbic.201500280>.

[36]. Saab-Rincón G, Juárez VR, Osuna J, Sánchez F, Soberón X. Different strategies to recover the activity of monomeric triosephosphate isomerase by directed evolution. *Protein Eng* 2001;14:149–55. <https://doi.org/10.1093/protein/14.3.149>.

[37]. Salinas VH, Ranganathan R. Coevolution-based inference of amino acid interactions

underlying protein function. *Elife* 2018;7. <https://doi.org/10.7554/eLife.34300>.

[38]. Weigt M, White RA, Szurmant H, Hoch JA, Hwa T. Identification of direct residue contacts in protein-protein interaction by message passing. *Proc Natl Acad Sci* 2009;106:67–72. <https://doi.org/10.1073/pnas.0805923106>.

[39]. Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, et al. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci* 2011;108:E1293–301. <https://doi.org/10.1073/pnas.1111471108>.

[40]. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>.

[41]. Arreola-Barroso RA, Llopiz A, Olvera L, Saab-Rincón G. Modulating Glycoside Hydrolase Activity between Hydrolysis and Transfer Reactions Using an Evolutionary Approach. *Molecules* 2021;26:6586. <https://doi.org/10.3390/molecules26216586>.