



UNIVERSIDAD
DE SANTIAGO
DE CHILE



Gabriel Vallejos-Bacchiere

gvallejos@ug.uchile.cl

Laboratorio de Bioquímica y Biología
Molecular. Departamento de Biología,
Facultad de Ciencias.

Universidad de Chile.

<https://orcid.org/0000-0002-6807-1791>

Artículo recibido: 30 de junio de 2022

Artículo aceptado: 30 de julio de 2022

Artículo publicado: 31 de julio de 2022



[CC BY, Gabriel Vallejos-Bacchiere, 2022]

Artículo de Investigación

<https://doi.org/10.35588/cc.v3i1.5584>

Problemas contemporáneos en la filosofía de la bioquímica: lecciones desde el plegamiento de proteínas y el problema *in-vitro/in-vivo*

Contemporary Problems in the Philosophy of Biochemistry: Lessons From Protein Folding and the In-vitro/in-vivo Problem

Resumen

Si bien en la filosofía de las ciencias ya se han explorado algunos ejemplos provenientes de la bioquímica como casos de estudio, la filosofía de la bioquímica (*i.e.* el estudio filosófico sistemático de los problemas metacientíficos de esta ciencia) es una subdisciplina naciente. En este artículo estudiaremos dos problemas filosóficos de relevancia contemporánea en esta ciencia. Por un lado, examinaremos las bases epistemológicas del problema del plegamiento de proteínas. En particular lo relacionado con la predicción de la estructura tridimensional de las proteínas a partir de su secuencia, asunto que ha dado mucho que hablar debido a los nuevos avances utilizando *deep learning*. Por otro lado, exploraremos el problema *in-vitro/in-vivo* y, más generalmente, el problema de la extrapolación en las ciencias biológicas. Finalmente, a partir de las consecuencias de ambos temas consideraremos algunas reflexiones filosóficas generales acerca del reduccionismo, el pluralismo y el lugar de la bioquímica en las ciencias biológicas.

Palabras clave: Bioquímica, Plegamiento de proteínas, *In-vitro/in-vivo*, Pluralismo, Reduccionismo.

Abstract

Although examples coming from biochemistry have already been explored in the philosophy of sciences as case studies, the philosophy of biochemistry (*i.e.* the systematic philosophical study of the metascientific problems of this science) is a nascent subdiscipline. In this article, two philosophical problems of contemporary relevance within this science will be explored. First, we are going to examine the epistemological bases of the protein folding problem, mostly concerning protein 3D structure prediction from its sequence, which has given much to talk about due to new advances involving *deep learning*. Second, we will explore the *in-vitro/in-vivo* problem and, more generally, the extrapolation problem in biological sciences. Finally, taking into account the consequences of both subjects, we will consider some philosophical reflections about reductionism, pluralism and the place of biochemistry among biological sciences.

Keywords: Biochemistry, Protein folding, *In-vitro/in-vivo*, Pluralism, Reductionism.

1. Introducción¹

La bioquímica es la ciencia que estudia *los fundamentos químicos de la vida*. Sin embargo, y por más simple que parezca, esta caracterización no deja de ser problemática. Su ubicación en el intrincado espacio entre la química y la biología le ha hecho cargar con la ardua responsabilidad de ser la mediadora entre dos ciencias con objetivos, metodologías y lenguajes diferentes. Históricamente, la bioquímica se ha visto en la necesidad de desarrollar diversas metodologías y líneas de investigación que han provocado su ramificación en múltiples subdisciplinas con identidad propia y relativa autonomía. Esta pluralidad trae consigo la dificultad de poder comunicar e integrar los conocimientos provenientes de diversas fuentes. Por otro lado, esto también conlleva la consecuencia de que exista una proliferación de metodologías, cada una con sus propios problemas epistemológicos. Es por esto que el abordaje filosófico de la bioquímica será, a su vez, plural y diverso.

La filosofía de la bioquímica puede considerarse como una subdisciplina emergente. Esto no quiere decir que antes no haya habido reflexión filosófica acerca de la bioquímica y áreas afines. Por ejemplo, Bechtel y Richardson (1993) se basaron, entre otras cosas, en la historia de la bioquímica metabólica para proponer su interpretación filosófica de la investigación científica en términos de la búsqueda y descripción de mecanismos. Esta investigación luego abrió toda una escuela en filosofía de las ciencias que persiste hasta el día de hoy. Así mismo, otros autores se han basado en el desarrollo de la bioquímica y sus conexiones tanto con la biología molecular como con la biología estructural para analizar filosóficamente la experimentación biológica (Rheinberger, 1997) o el uso de modelos en ciencias (de Chadarevian, 2011).

Debido a la pluralidad de la bioquímica, lo que con propiedad podríamos llamar *filosofía de la bioquímica* también ha sido plural y dispersa. Por un lado, dentro de las líneas de estudio que podríamos mencionar se encuentran, por ejemplo, la selección causal en el estudio de redes metabólicas (Ross, 2017); las representaciones y explicaciones en las visiones estática y dinámica del alosterismo (Neal, 2021); las clases naturales y el microestructuralismo en el caso de las macromoléculas (*e.g.* Slater, 2009; Tobin, 2010; Bartol, 2016; Goodwin, 2011; Havstad, 2018; Tahko, 2020); los factores causales que actúan en el plegamiento de las proteínas (Santos et al. 2020); el pluralismo y la integración de conocimiento en los modelos sobre proteínas (Mitchell, 2019); el rol de las metáforas en el nacimiento de la enzimología (Mertens, 2019); la estructura de las teorías cinéticas del alosterismo (Alleva *et al.*, 2017); la proposición de una ontología de procesos para entender las macromoléculas (Guttinger, 2018; Alassia, 2022); el problema *in-vitro/in-vivo* (Strand, 1996, 1999; Jacob, 2002; Ibarra y Mormann, 2006; García, 2015; Esposito y Vallejos, 2020); y algunos aspectos filosóficos de la predicción de estructura de proteínas (Ramsey, 2007; Mitchell y Gronenborn, 2017).

En este artículo desarrollaremos en detalle dos problemas filosóficos propiamente bioquímicos de relevancia contemporánea. Por un lado, analizaremos los principales aspectos epistemológicos del problema del plegamiento de proteínas y los recientes avances en la predicción de sus estructuras

¹ Agradezco los comentarios de las revisoras o revisores que la revista asignó a este artículo y al editor.

tridimensionales basados en la tecnología del *deep learning*. Por otro lado, analizaremos el problema *in-vitro/in-vivo* y el problema general de la extrapolación en la experimentación biológica. En ambos casos intentaremos extraer los principales presupuestos filosóficos presentes en las prácticas de la bioquímica, aquellos con que los científicos se las arreglan para lidiar con estos problemas en el día a día. Finalmente, a partir de estos análisis extraeremos algunas lecciones filosóficas generales que nos permitirán lidiar con problemas filosóficos clásicos que han acompañado históricamente a la bioquímica. Concretamente los problemas del reduccionismo y del pluralismo.

2. El problema del plegamiento de proteínas

Recientemente, en diversos medios de comunicación tales como BBC News y The Guardian, circuló la noticia de que una inteligencia artificial había resuelto uno de los más grandes misterios de la biología: el problema del plegamiento de proteínas (Briggs 2020; Rincon 2021; Sample, 2020). Utilizando *deep learning* (DL), la compañía DeepMind de Google obtuvo un triunfo arrollador en la 13^a (Senior *et al.*, 2020) y 14^a (Jumper *et al.*, 2021) versiones de CASP (Critical Assessment of Techniques for Protein Structure Prediction) con sus softwares AlphaFold (AF) y AlphaFold2 (AF2) respectivamente² (AlQuraishi 2019, 2020; Lupas *et al.*, 2021; Perrakis y Sixma, 2021; Obermayer y Uversky, 2021). Esta es una competencia bianual en donde diversos equipos ponen a prueba metodologías computacionales desarrolladas para predecir la estructura de las proteínas a partir de su secuencia (Kryshtafovych *et al.*, 2019). La magnitud de este gran triunfo dejó atónitos a muchos investigadores del área. Su impacto ha sido tal que incluso se ha llegado a afirmar que en algún futuro los métodos computacionales, principalmente basados en *deep learning*, superarán a los métodos experimentales y los reemplazarán paulatinamente (Lupas *et al.*, 2021; Obermayer y Uversky, 2021).

No obstante, de lo recién expuesto surgen varias preguntas: ¿qué significaría exactamente *resolver* el problema del plegamiento? ¿Qué es lo que AF realmente ha resuelto (si es que ha resuelto algo)? ¿Cómo podemos entender su éxito? ¿Es plausible la afirmación de que los métodos computacionales (como DL) reemplazarán alguna vez a los métodos experimentales? De ser así, ¿cómo lo sabríamos? En esta sección intentaremos hacer un breve recorrido por estas preguntas y sugerir algunos lineamientos para las investigaciones filosóficas que podrían desprenderse de ellas. Para esto, primero haremos un breve recorrido por el problema del plegamiento de proteínas, su importancia científica y sus principales aspectos filosóficos.

2.1. Explicación y predicción en el plegamiento de proteínas

Las proteínas son centrales en el funcionamiento de todos los organismos conocidos. Son las macromoléculas que llevan a cabo funciones tales como la catálisis de reacciones químicas en el metabolismo, la regulación de procesos celulares, el transporte de componentes intra y extracelulares, el movimiento a nivel tanto macro como microscópico de los organismos y un largo etcétera (Voet y Voet, 2010). Sin temor a exagerar demasiado, podría decirse que son las moléculas que hacen posible la vida tal cual la conocemos en la Tierra.

² De aquí en adelante nos referiremos solo a *esté último* por ser la última versión a la fecha y el que ha cosechado el más éxito.

Estas macromoléculas son cadenas lineales de aminoácidos unidos covalentemente a través de un enlace peptídico³. Son sintetizadas en la célula por los ribosomas que catalizan la formación de este enlace entre distintos aminoácidos en un orden de acuerdo con la secuencia nucleotídica de un ARN mensajero (ARNm)⁴. Ambas secuencias se encuentran conectadas a través del *código genético*, que asocia a cada codón de 3 nucleótidos consecutivos algún aminoácido⁵. A su vez, la secuencia este ARNm ha sido generada de acuerdo con la secuencia de un segmento de ADN, vale decir, un gen. Esta descripción corresponde, *grosso modo*, al Dogma Central de la Biología Molecular. Sin embargo, para poder realizar sus funciones biológicas en un medio fisiológico las proteínas deben adquirir una estructura tridimensional determinada: su ‘estructura nativa’^{6,7}. El proceso mediante el cual ésta se alcanza se conoce como ‘plegamiento’. Entender cómo ocurre y cómo se relaciona con la composición de las proteínas ha sido un problema fundamental de la bioquímica desde mucho antes del nacimiento de la biología molecular (Tanford y Reynolds, 2003).

Este problema conlleva una gran importancia no solo para la bioquímica básica, sino que también para aplicaciones tales como la biomedicina, la industria y la farmacia. Por ejemplo, enfermedades neurodegenerativas tales como el mal de Alzheimer y la enfermedad de Creutzfeldt-Jakob son causadas por la formación de agregados proteicos producto del mal plegamiento de ciertas proteínas. En el caso de la industria, entender los factores que determinan la estabilidad de las proteínas sería útil para optimizar procesos industriales en los que se utilizan enzimas en condiciones extremas. Así mismo, la capacidad de predecir la estructura de las proteínas solo considerando la secuencia conlleva la promesa de entender a nivel estructural la acción de cada uno de los genes del organismo. De esta forma, podrían descubrirse blancos terapéuticos para nuevos fármacos y así curar una gran cantidad de enfermedades (Wang *et al.*, 2016).

Dado que las proteínas son extraídas ya plegadas desde las células, en un comienzo se postulaba que éstas debían ser plegadas por algún mecanismo intracelular (Tanford y Reynolds, 2003). Sin

³ Este enlace se forma entre el grupo amino y carboxilo de dos aminoácidos (*c.f.* Voet y Voet, 2010; Fersht, 2017).

⁴ Es usual encontrar en la literatura bioquímica frases como ‘la secuencia del ARN determina la secuencia de la proteína’. Qué es lo que significaría esta determinación es algo difícil de justificar, puesto que no es la secuencia del ARN por sí sola la que causa la secuencia de la proteína, sino una gran cantidad de mecanismos celulares. Este tema ha sido ampliamente discutido en la literatura acerca de la causalidad en sistemas biológicos, donde se han propuesto varias soluciones en términos de generación de diferencias concretas, especificidad causal y otros tratamientos (*c.f.* Waters 2007, 2019; Weber, 2006; Woodward, 2010; Vecchi, 2020) que escapan al objetivo de este artículo.

⁵ La frase ‘código genético’ se deriva de la metáfora de la información y los programas en los sistemas biológicos, ampliamente utilizada en biología molecular. El fenómeno descrito superviene en un mecanismo complejo en el que participan una gran cantidad de enzimas que son capaces de: reconocer cada codón en un ARN de transferencia, unir covalentemente el aminoácido y depositarlo en el ribosoma para que ocurra la formación del enlace peptídico (Krebs *et al.*, 2017).

⁶ En este artículo centraré mi atención en las proteínas globulares y solubles. Además de ellas existen proteínas fibrilares y/o insolubles que juegan otra clase de roles y tanto sus estructuras como métodos de estudio son diferentes.

⁷ Las proteínas son realmente entidades altamente dinámicas y, más que de estructuras fijas, deberíamos hablar de ensamblajes conformacionales (Neal, 2021). Además, existen proteínas ‘intrínsecamente desordenadas’ que pueden ejercer su función sin adquirir una estructura tridimensional determinada. En este artículo no entraré en detalles sobre estos temas; y cuando hable de “estructura nativa” nos estaremos refiriendo simplemente al ensamblaje estructural que permite a una proteína realizar su función biológica.

embargo, esta idea contaba con bastantes problemas. Primero que todo, debía existir un molde que condujera dicho proceso. Dada su diversidad y ubicuidad, las mismas proteínas eran buenos candidatos para ocupar este lugar. No obstante, y si éstas son plegadas por otras proteínas, ¿qué pliega a estas últimas? Por otro lado, cada proteína tiene una secuencia distinta, por lo que en principio debería haber tantos moldes como proteínas a plegar o éstos ser lo suficientemente complejos como para abarcar grandes familias de secuencias. Posteriormente, en el contexto de las investigaciones acerca de la síntesis de proteínas, se planteó que ese molde podría ser provisto por el ARN ribosomal. No obstante, todos los ribosomas de una célula contienen la misma secuencia de ARN, por lo que nuevamente surgía el problema de cómo podrían conducir el plegamiento de proteínas tan variadas.

El problema dio un giro radical a partir de la segunda mitad del siglo XX gracias a las investigaciones de Christian Anfinsen y sus colaboradores, que lo hicieron ganador del premio Nobel de Química en 1972 (Anfinsen *et al.*, 1961; Anfinsen, 1973). En una serie de experimentos mostraron que una proteína purificada (la ribonucleasa A), luego de ser denaturada con urea y agentes reductores, podía replegarse reversiblemente al extraer del medio los denaturantes, recuperando así su actividad biológica en ausencia de otros componentes celulares. Vale decir, la proteína solo se bastaría a sí misma para poder alcanzar su estructura en un medio adecuado. A partir de estos resultados, Anfinsen propuso la hipótesis de que es la secuencia aminoacídica lo que determina la estructura nativa de una proteína, que correspondería a la conformación de mínima energía de la cadena polipeptídica en un medio determinado^{8,9}. Entonces, para predecir la estructura de una proteína solo bastaría con encontrar este mínimo energético conformacional. Sin embargo, las cosas resultaron ser más complejas

Si consideramos una proteína de 100 aminoácidos donde cada enlace peptídico solo pudiese explorar dos conformaciones, ésta tendría que explorar un espacio conformacional de 10^{30} posibilidades para llegar a su estado nativo. Si cada cambio conformacional ocurriese en solo un picosegundo, el proceso demoraría más que la edad del universo¹⁰ si es que el proceso ocurriese en forma aleatoria. Sin embargo, las proteínas suelen plegarse en ordenes de magnitud temporales de entre milisegundos y segundos, siendo raros los casos de ordenes de magnitud mayores (Gomes y Faísca, 2019). De este experimento mental, conocido como la ‘Paradoja de Levinthal’ (Levinthal, 1968, 1969), se desprende una conclusión obvia: el plegamiento no ocurre mediante una búsqueda al azar de conformaciones, sino que debe existir alguna clase de ruta —en términos de una cadena

⁸ Esta hipótesis tiene dos posibles interpretaciones, una puramente termodinámica respecto a los estados conformacionales de la cadena y otra de naturaleza causal respecto a que la secuencia determina la estructura nativa. Para una discusión filosófica ver (Santos *et al.*, 2020).

⁹ A esta hipótesis se le podría llamar también como ‘Dogma Anfinsen-Crick’ (Morange 2006). Previo a los experimentos de Anfinsen ya se había especulado acerca de la posibilidad de que sea la secuencia la que determine la estructura y la función de una proteína de acuerdo con un ‘código del plegamiento’ que supervendría en las propiedades físicas y químicas de la cadena aminoacídica.

¹⁰ Esto se obtiene considerando una proteína ultra simplificada como la del ejemplo. En el caso de una proteína natural, en donde son posibles muchas más que dos conformaciones por aminoácido, la cifra sería aún más grande. A esto habría que sumar las posibles conformaciones de las cadenas laterales de cada aminoácido.

causal de interacciones organizadas espacial y temporalmente— que lo conduzca. Esto supone una complicación adicional: el estado plegado no tiene por qué ser el mínimo conformacional global, sino que simplemente podría ser el estado al que la proteína pueda llegar más rápido (un mínimo local) que sea lo suficientemente estable como para poder realizar su función biológica. Por lo tanto, para predecir la estructura de una proteína no bastaría encontrar el mínimo energético conformacional, también se hace necesario conocer la ruta mediante la cual una proteína se pliega.

Considerando lo dicho hasta ahora, podemos dividir al problema del plegamiento en dos aspectos: uno explicativo y otro predictivo. El primero corresponde a la descripción del mecanismo mediante el cual una proteína alcanza su estado nativo. Por ejemplo, en términos de los factores que determinan su estabilidad o en términos de la cadena causal que lleva del estado desplegado al estado nativo. El segundo corresponde a la predicción de la estructura tridimensional del estado nativo de una proteína.

En las líneas de investigación derivadas de Anfinsen y Levinthal la predicción aparece como un subproducto de la explicación. Disponiendo de una explicación de los principios que guían el plegamiento sería posible determinar la evolución temporal de una cadena proteica en un medio determinado y conocer su estado final considerando solo su secuencia, las propiedades de sus partes y las leyes o principios físicos y químicos que rigen al sistema¹¹. Estas líneas de investigación han logrado grandes éxitos explicativos a la hora de describir en detalle los mecanismos de plegamiento para varias proteínas. Sin embargo, este gran éxito explicativo no ha conllevado un éxito predictivo. Es más, se ha logrado a costo del sacrificio de la predicción. Para poder interpretar los resultados experimentales y generar tanto modelos como representaciones que permitan explicaciones con gran nivel de detalle es necesario recurrir a información estructural previamente disponible para la proteína bajo estudio. Mientras más detalles se deseen explicar, más dependencia habrá de dicha información estructural. Por lo tanto, el éxito explicativo está supeditado a casos en donde el *explanandum* (i.e. el estado nativo de una proteína) es previamente conocido¹².

Con el advenimiento de la química computacional, la ‘caja de herramientas’ de la bioquímica se vio considerablemente aumentada. Además de proveer de importantes complementos a las líneas de

¹¹ En la literatura científica es frecuente encontrar alusiones a términos como ‘leyes’, ‘principios’ u otros en el contexto de explicaciones acerca del comportamiento de algún sistema, sobre todo en contextos donde se presupone cierto determinismo en los procesos que subyacen al fenómeno a ser explicado. Filosóficamente, estas leyes y principios podrían ser descritos de diversas formas, tales como la manifestación de disposiciones causales, la existencia de una relación estable e invariante o alguna otra forma de dependencia contrafáctica. En este trabajo no pretendo entrar a discutir cual es la mejor caracterización metafísica de estas ‘leyes’ o ‘principios’ aludidos por los científicos. Sin embargo, tiendo a creer que esto variará dependiendo del contexto explicativo que se considere, habiendo muchos casos en los que existirá una subdeterminación en cuanto a cuál será la descripción más adecuada. En nuestro caso de estudio nos basta con especificar que las ‘leyes’ o ‘principios’ aludidos suelen extraerse de la química y la física, tales como la ley de Coulomb, las fuerzas de dispersión de London, la segunda ley de la termodinámica, etc. Usualmente aspectos propios de la proteína, tales como sus conformaciones, aspectos topológicos u otros son relegados a condiciones iniciales y de borde, formando parte de la descripción del sistema en la que se instancian los ‘principios’ o ‘leyes’ físicos y químicos que hacen el trabajo explicativo. Vale decir, estaríamos frente a lo que Weber (2005) llama ‘heteronomía explicativa’. Sin embargo, la discusión de este punto excede el alcance de este artículo.

¹² Esta información estructural es provista por la evidencia obtenida mediante técnicas independientes, como la cristalografía de rayos X. Este punto será importante más adelante.

investigación puramente explicativas, también propició el nacimiento de nuevas líneas de investigación en las que la predicción pasaba nuevamente al primer lugar en conjunto con la explicación en base a las propiedades físicas de la secuencia. Sin embargo, debido entre otras cosas al alto costo computacional, el éxito predictivo se vio limitado solo a modelos altamente idealizados y simplificados de proteínas pequeñas (Tanford y Reynolds, 2003; Fasman, 1989; Wooley y Ye, 2007; Lee *et al.*, 2017). Estas dificultades condujeron a que, en orden a aumentar el éxito predictivo, paulatinamente se comenzara a sumar información tanto proveniente de resultados experimentales como también de estructuras ya conocidas de otras proteínas. Esto llevo a que los principios físicos pasaran de jugar un rol explicativo a uno constructivo en el contexto de la elaboración, calibración y refinamiento de modelos. De esta forma, poco a poco comenzó a sacrificarse la explicación con el objetivo de lograr alcanzar el éxito predictivo.

Vemos entonces que en el problema del plegamiento se produjo una disociación entre las metas predictivas y explicativas. Por un lado, cuando la meta es explicativa se sacrifica la predicción para aumentar el detalle y el éxito de las explicaciones. Por otro lado, cuando la meta es predictiva las dificultades prácticas generan que, para aumentar el éxito predictivo, deba relegarse a los aspectos teóricos y mecánicos a un rol netamente constructivo.

Esta disociación se vio consolidada luego del nacimiento de la bioinformática y las bases de datos de secuencias y estructuras de proteínas (Berman, 2008; Strasser, 2010). La gran explosión en el campo de la determinación experimental de estructuras macromoleculares provocó un incremento exponencial en la cantidad públicamente disponible de estructuras tridimensionales de proteínas, las que comenzaron a ser depositadas en bases de datos especializadas (Xu y Xu, 2004; Laskowski, 2011) como el Protein Data Bank (PDB) (Berman HM y Gierasch, 2021). Con el aumento de esta información y el desarrollo de diversas herramientas estadísticas e informáticas comenzó a describirse la frecuencia de patrones estructurales, tales como tendencias a formar estructuras secundarias, ángulos entre residuos, ordenamientos de cadenas laterales, entre otras. Con esto comenzaron a surgir nuevas metodologías computacionales en las que las metas explicativas eran ya totalmente descartadas en favor de las predictivas. Ahora la meta ya no era inferir la estructura solo considerando las propiedades físicas de la cadena polipeptídica en un ambiente determinado, sino que hacerlo con base en los patrones y regularidades encontrados en las bases de datos estructurales (Lee *et al.*, 2017; Abeln *et al.*, 2019; Pearce y Zhang, 2021).

2.2. Predicción sin explicación

En 1994 se realiza la primera versión de CAPS (Moult *et al.*, 1995), un concurso bianual en donde diversos grupos de investigación ponen a prueba sus metodologías para predecir estructuras tridimensionales de proteínas. En este concurso a cada participante se le otorga un conjunto de secuencias de proteínas cuya estructura ya se ha determinado, pero que aún no se ha hecho pública. De esta forma, las predicciones realizadas por cada equipo son evaluadas de acuerdo con su porcentaje de similitud respecto a ellas

Estas metodologías se clasifican en tres grupos: modelamiento por homología, *threading* y modelamiento *ab initio*. La primera utiliza como molde una proteína de estructura ya conocida que presente homología de secuencia con la proteína objetivo (Abeln *et al.*, 2019), lo que se basa en la idea de que la similitud entre secuencias implica similitud entre estructuras. Por otro lado, uno de

los grandes descubrimientos de la biología estructural es que existe una cantidad finita y extremadamente pequeña de clases de estructuras frente a la gigantesca diversidad de secuencias (Kaczanowski y Zielenkiewicz, 2010). Este es el fundamento del método *threading* (Kelley, 2009), que se basa en probar distintas clases estructurales hasta encontrar la que mejor podría servir de molde para enhebrar la secuencia aminoacídica objetivo. El modelamiento *ab initio*, en cambio, prescinde del uso de moldes. Las metodologías clásicas de este grupo se basan en la búsqueda de los patrones estructurales más frecuentes para diversos segmentos con secuencias determinadas, tales como tendencias a formar cierta estructura secundaria (*e.g.* hélices, hebras β), ángulos entre enlaces, giros u otros (Rohl *et al.*, 2004; Xu y Zhang, 2012; Lee *et al.*, 2017; Abeln *et al.*, 2019; Pearce y Zhang, 2021). A esto se suman restricciones tales como que los modelos sean físicamente posibles preocupándose de que, por ejemplo, no existan átomos ocupando el mismo lugar, ángulos imposibles entre enlaces, etc.

En este contexto, el objetivo epistémico inmediato es la construcción de un modelo estructural que prediga la estructura nativa de una proteína. Éste no debe ser visto como una hipótesis que busque corroborar alguna teoría. Si bien en el proceso participan varias teorías, ya sean físicas, químicas, bioquímicas o evolutivas, éstas no juegan un rol explicativo, sino meramente constructivo (Ramsey, 2007; Mitchell y Gronenborn, 2017). Por ejemplo, en la restricción de posibles estructuras, el descarte de estructuras imposibles, el refinamiento de modelos, la minimización de energía conformacional, etc.

Entre los años 2002 y 2012 se produjo un estancamiento en los puntajes obtenidos por los participantes de CASP (AlQuraishi, 2019), principalmente porque había una ‘saturación’ de metodologías y los avances se concentraban principalmente tanto en su optimización como refinamiento, además de la inclusión de nueva data estructural. Esto cambió en el año 2014 con la introducción de nuevas metodologías y supuestos teóricos. Por un lado, se implementó el *deep learning* (Schaarschmidt *et al.*, 2018), lo que aumentó considerablemente la eficiencia en la extracción de patrones estructurales y de secuencia. Por otro lado, también se introdujo el análisis de correlación evolutiva entre residuos. Esta última se basa en la inferencia de contactos debido a que se espera que cuando hay una interacción, si durante la evolución un residuo es sustituido por otro con una propiedad distinta, debiese existir una mutación en el otro residuo a modo de compensación. Durante el CASP14, desarrollado en 2020, el software Alphadold2, que involucraba ambas metodologías, derrotó con creces a todos los otros participantes obteniendo récords nunca vistos en la predicción de estructuras (AlQuraishi 2020; Jumper *et al.*, 2021).

2.2.1. AlphaFold2

El funcionamiento de AF2 es complejo y contiene muchos sub-procesos (Jumper *et al.*, 2021), pero básicamente se trata de una red neuronal con varios módulos entrenada con información estructural proveniente del PDB. Por otro lado, a partir de información de miles de secuencias provenientes de bases de datos metagenómicos se extraen patrones de mutaciones correlacionadas entre proteínas homologas cercanas y lejanas a partir de los cuales se infieren posibles contactos entre residuos y sus distancias, en lo que se conoce como un distograma. Con base en esto se genera un modelo tridimensional que pasa por sucesivos procesos de refinamiento, de descarte de modelos físicamente imposibles y varias iteraciones.

El éxito de AF2 ha traído grandes expectativas y promesas (Lupas *et al.*, 2021; Perrakis y Sixma, 2021; Obermayer y Uversky, 2021), sobre todo en lo relacionado con la biomedicina (Rincon, 2021). Utilizando sus predicciones se ha generado la base de datos *AlphaFold Protein Structure Database*, cuyo fin es disponer de las estructuras inferidas para cada una de las proteínas del humano y otros organismos modelo (Tunyasuvunakool *et al.*, 2021). Uno de los campos donde estos avances han generado mayor expectativa es en el diseño de fármacos (Mullard, 2021), ya que con esta información sería posible encontrar grandes cantidades de posibles blancos terapéuticos tales como sitios específicos de receptores, enzimas, etc. y diseñar nuevas moléculas que puedan intervenir en procesos fisiológicos y/o patológicos uniéndose a éstos.

La expectativa ha llegado al punto de que se ha llegado a afirmar que en un futuro metodologías como AF2 reemplazarían a las metodologías experimentales en la determinación de estructuras (Obermayer y Uversky, 2021). Se destacan ventajas tales como el poco tiempo que tomaría obtener una estructura con AF2 en comparación con otras técnicas como la cristalografía, que suelen demorar semanas o incluso años para lograr obtener una sola estructura. Sin embargo, ¿qué significaría esto? ¿Es realmente posible hablar en dichos términos?

Primero que todo, y como es de esperar, AF2 no está exento de limitaciones científicas. Su rango de acción está limitado a modelos monoméricos y no considera grupos prostéticos ni ligandos (AlQuraishi, 2020). Por otro lado, no es que AF2 haya tenido un éxito completo, hubo blancos en los que su desempeño fue bajo (AlQuraishi, 2020). AF2 extrae patrones de coevolución de residuos desde una gigantesca cantidad de secuencias. Sin embargo, el éxito de esta estrategia depende de la existencia de una cantidad suficiente de secuencias conocidas de proteínas homólogas, lo que representa una dificultad en caso de que la proteína objetivo no presente homología suficientemente cercana con las presentes en las bases de datos¹³. Por otro lado, si bien la coevolución de residuos es la principal fortaleza de AF2, puede ser también su mayor debilidad. Mutantes en los residuos que presenten correlación evolutiva podrían hacer fallar la predicción de contactos y producir modelos con estructuras diferentes, como de hecho ha ocurrido (Lowe, 2022; Jha, 2021). Sin embargo, los problemas más importantes de AF2 —y metodologías relacionadas— son principalmente epistemológicos.

Un problema central es la existencia de sesgos en la información considerada para extraer patrones estructurales. Las estructuras que figuran en las bases de datos como el PDB han sido determinadas principalmente mediante cristalografía de rayos X¹⁴. Para determinar una estructura mediante esta técnica las proteínas deben encontrarse en estado cristalino y condiciones artificiales que permitan realizar la difracción de los rayos X, pero para esto primero deben ser proteínas susceptibles de ser cristalizadas. A todo esto hay que sumar que la gran mayoría de las proteínas que figuran en el PDB son producidas recombinantemente en organismos ingenierizados como bacterias o levaduras, de lo que se desprende que además deben ser susceptibles de ser expresadas y

¹³ De hecho, se ha acuñado el concepto de ‘proteoma oscuro’ para clasificar a estas proteínas que no poseen suficientes homólogos y que su estructura aún no se ha determinado (Perdigão y Agostinho, 2019)

¹⁴ Aunque también mediante NMR y criomicroscopía electrónica, pero en mucho menor proporción, siendo la cristalografía la más representativa con aproximadamente un 80%. Por esta razón centraré mi análisis en los sesgos provenientes de ella, aunque el mismo análisis podría aplicarse a los sesgos que pudiesen provenir de las otras metodologías.

purificadas. Esta forma de producción podría generar condiciones de plegamiento diferentes a las de su medio nativo, en el que las proteínas generan interacciones con toda una multitud de moléculas y donde usualmente sufren modificaciones postraduccionales. A todo esto hay que sumar que en las bases de datos existe una sobrerrepresentación de proteínas que pertenecen a ‘organismos modelo’ de uso común en laboratorios y, obviamente, de humanos. En general, las proteínas que no se pueden expresar recombinantemente, purificar, cristalizar, difractar o analizar son descartadas, por lo que finalmente, en las bases de datos solo tenemos una muestra para las cuales todo este proceso fue exitoso.

Por otro lado, la obtención de estructuras a partir de cristalografía es un proceso complejo, largo y tedioso en el que se utiliza mucha información proveniente de otros medios para interpretar los datos, refinar y validar los modelos estructurales producidos. Esto incluye también patrones de estructuras previamente determinadas (Wlodawer *et al.*, 2013; Shi, 2014). Esto implica que la determinación experimental de estructuras también conlleva sesgos, los que son absorbidos por AF2 y las técnicas computacionales en general.

Considerando todos estos sesgos, cabe la posibilidad de que en las bases de datos exista una sobrerrepresentación artificial de algunas clases de estructuras o de ciertos patrones estructurales. De hecho, existen casos de predicciones en que, si bien varios softwares predijeron la misma estructura tridimensional para una secuencia, finalmente el resultado obtenido mediante cristalografía fue totalmente distinto (Figueroa *et al.*, 2016).

Por lo tanto, siendo precisos, lo que nos otorga AF2 es la predicción de cómo sería la estructura de una proteína si fuese expresada recombinantemente, purificada, cristalizada y difractada exitosamente y su estructura determinada mediante ese método con todos los sesgos que eso implica (AlQuraishi, 2020).

2.2.2. Predicción computacional vs determinación experimental

Reconsideremos ahora la pregunta acerca de si las metodologías computacionales para predecir estructuras podrán alguna vez superar y reemplazar a las metodologías experimentales. ¿Qué significaría esto? Consideremos la posibilidad de un futuro donde efectivamente metodologías computacionales como AF2 hubiesen reemplazado a las técnicas experimentales. ¿Cómo podríamos justificar que estas metodologías están funcionando adecuadamente? Pues evaluando si los resultados que nos proporcionan son los correctos. Ahora, ¿cómo podemos justificar que estos resultados son los correctos? Pues argumentando que se obtuvieron usando una metodología que funciona adecuadamente. Este razonamiento circular, imposible de romper si consideramos una metodología en forma aislada, es lo que se ha llamado el argumento del ‘regreso del experimentador’ (Collins 1987; Zuppone 2010) (pues sus primeros análisis fueron en el caso de las ciencias experimentales, aunque también se ha analizado en el caso de las simulaciones computacionales [Gelfert, 2011]). Para romperlo es necesario obtener evidencia externa que nos permita validar una metodología obteniendo un mismo resultado en forma independiente (Culp, 1995). ¿Cómo podríamos hacer esto en un mundo donde solo hubiese metodologías computacionales? La respuesta obvia es que utilizando métodos computacionales independientes. Sin embargo, en un caso como el nuestro esto no sería posible. Todos los métodos para determinar estructura de proteínas se generan considerando el mismo cuerpo de conocimientos de base, consistente principalmente en patrones

encontrados en bases de datos estructurales y de secuencia. Esto implica que las metodologías predictivas de estructuras de proteínas no son independientes al punto que puedan justificarse entre ellas y romper de esta forma el círculo vicioso de la justificación. Para poder validarlas habrá que recurrir entonces a métodos experimentales o, dicho de otro modo, sus resultados deben ser corroborados experimentalmente para su validación. De lo contrario, no podremos saber si estamos frente a una predicción novedosa o ante un artefacto producido por los sesgos propios de la metodología¹⁵.

Finalmente, si las metodologías computacionales desplazaran a las experimentales, entonces el conocimiento de base del que se nutren dejaría de crecer. En este escenario, por más que se perfeccionen los algoritmos, siempre estará el fantasma de las posibilidades que no hayan sido concebidas dentro del conocimiento inicial, y la única forma de luchar contra él es confrontando la realidad material mediante la experimentación. Para aumentar la capacidad predictiva y el ámbito de aplicación de metodologías computacionales, tales como el *Deep learning*, se requiere producir cada vez más datos lo más diversos posibles, por lo que quienes trabajan en cristalografía y otras metodologías experimentales para determinar estructuras de proteínas pueden estar seguros de que tendrán ‘trabajo para rato’¹⁶.

2.3. Promesas y riesgos

Como se dijo más arriba, el diseño de fármacos es uno de los campos donde las promesas de AF2 han causado más expectación. Sin embargo, dichas promesas deben ser asumidas considerando el contexto en donde AF2 realizaría realmente sus aportes. El diseño de fármacos posee muchas más etapas que el diseño de moléculas en base a posibles blancos terapéuticos a nivel estructural de proteínas. En este proceso existen múltiples cuellos de botella, tales como la utilización de modelos animales adecuados y estudios clínicos controlados (Dahlin *et al.*, 2015). En ese sentido, el contar con las estructuras nativas de las proteínas, incluso si fuesen perfectamente determinadas, no supondría un cambio tan significativo en cuanto a la efectividad y rapidez del proceso. Además, en la generación de nuevos fármacos el estándar evidencial no se encuentra a nivel molecular. El objetivo final es que el fármaco produzca el efecto deseado y no presente toxicidad. De hecho, podría ser que el mecanismo subyacente que se consideró para diseñarlo o para explicar su acción esté equivocado y esta información se corrija inocuamente luego de que el medicamento ya está en el mercado, como efectivamente ya ha ocurrido. Un ejemplo interesante es el caso de la pregabalina, que fue diseñada como un análogo del neurotransmisor GABA con el fin de que interactuase con su receptor. El fármaco fue lanzado exitosamente al mercado y su efecto era consistente con el

¹⁵ Podría argumentarse que si utilizamos metodologías computacionales predictivas que acudan a principios físicos este círculo sí podría romperse sin utilizar experimentación. Sin embargo, estas metodologías no son posibles en la práctica, y especular respecto a futuros donde éstas sí fuesen exitosas carecería de sentido en el contexto de este artículo. Igualmente, de ser este el caso, también existirían problemas relacionados con los sesgos del conocimiento inicial, lo que se traduciría en que sin la interacción con el mundo material de los fenómenos que se desean estudiar no habría cómo sacar de la regresividad a programas de investigación solo basados en computadores. De la misma forma nos enfrentaríamos a la incertidumbre, a la posibilidad de sesgos y a las alternativas no concebidas.

¹⁶ Tomo esta idea de la divulgadora científica Sabine Hossenfelder (2021)

mecanismo de acción que se suponía que tenía. Sin embargo, luego se descubrió que su efecto no tenía que ver con este mecanismo, sino que con la inhibición de un canal de calcio (Sills, 2006).

Considerando todos estos problemas científicos y epistemológicos de técnicas como AF2 y similares, podemos concluir que estamos frente a una situación de alto riesgo epistémico en que cabe la posibilidad errores considerables que deberán ser adecuadamente evaluados en cada situación. Esto es especialmente apremiante si se consideran ejemplos como los mencionados en el campo de la salud de las personas, puesto que podría conllevar importantes problemas éticos, políticos y económicos. Sin embargo, esta es la misma situación que ha existido siempre. AF2 no suma más riesgos de los que ya había y ha habido siempre en la bioquímica, la biología estructural y la biomedicina. De hecho, el mismo tipo de críticas científicas y epistemológicas que se han hecho aquí a AF2 y métodos similares en cuanto a sus sesgos y limitaciones se podría hacer también a cualquier otra metodología bioquímica considerada en forma aislada, lo que incluye la determinación de estructuras mediante metodologías experimentales. No existe tal cosa como una metodología perfecta, sin límites y libre de sesgos. Cada una tiene sus propios problemas a ser considerados y susceptibles de ser estudiados.

En las ciencias no existe tal cosa como metodologías aisladas, sino grandes redes interconectadas de metodologías interdependientes en las que todas contribuyen de alguna forma al éxito de algún objetivo epistémico determinado. Como dijimos, así como las metodologías como AF2 y similares dependen en última instancia del conocimiento experimental, las metodologías experimentales para determinación de estructuras, a su vez, también dependen en parte de conocimiento experimental previo. Por otro lado, éstas últimas también se han visto beneficiadas gracias a la aparición de herramientas como AF2. Por ejemplo, en gran parte de los casos el punto de partida para la determinación de una estructura mediante cristalografía es la técnica del reemplazo molecular (Evans y McCoy, 2008). Ésta consiste en asumir una estructura aproximada para la proteína objetivo, la que luego se modifica y refina hasta que el modelo se ajusta a la densidad electrónica obtenida mediante la difracción cristalográfica. Para este proceso usualmente se utilizan estructuras de proteínas homólogas, pero gracias a AF2 se ha conseguido obtener mejores puntos de partida, incluso cuando no existen estructuras de homólogos para la proteína (Masrati *et al.*, 2021; McCoy *et al.*, 2022). Por lo tanto, AF2 puede ser visto como una herramienta entre tantas en la biología estructural, que permite sintetizar y gestionar la información estructural ya disponible, la que puede ser luego utilizada para guiar la determinación de estructuras a nivel experimental y otros procesos. Vale decir, actuando como un proveedor de inputs epistémicos para una gran multitud de prácticas.

Si bien queda claro que no tiene sentido afirmar AF2 y otras metodologías similares vayan a reemplazar a la experimentación, hay quienes dicen que igualmente el surgimiento de estos poderosos aparatos predictivos producirá una revolución en parte importante de la bioquímica y la biología estructural, desplazando a la determinación experimental de estructuras en varios escenarios. Esta sí es una afirmación plausible y en muchos casos es cierto. La pregunta entonces es ¿En qué circunstancias se considerará como suficiente una predicción realizada por AF2? Si el contexto es la biología estructural pura y el objetivo epistémico es la determinación de la estructura de una proteína, entonces claramente AF2 no será suficiente y el estándar evidencial corresponderá a técnicas experimentales para realizar dicha determinación. Sin embargo, si el objetivo es la

caracterización de aspectos fisicoquímicos no directamente relacionados con la estructura de una proteína, como la caracterización cinética de una enzima, entonces probablemente no será necesario corroborar la estructura predicha mediante cristalografía o alguna metodología afín si se utiliza una predicción de AF2 como input, pero tal vez sí será necesario corroborar algunos de sus aspectos a través de otros experimentos tales como modificación química, mutaciones puntuales, espectroscopia, etc. Lo mismo podría decirse de casos como el estudio de las redes de transducción de señales, donde la estructura de una proteína en particular juega un rol relativamente periférico a la hora de explicar un fenómeno a nivel celular.

De esta forma, AF2 debe ser vista como una nueva metodología que pasa formar parte de las nutridas y complejas redes metodológicas que conforman la bioquímica y áreas afines. AF2 cumple la función de un poderoso generador de *inputs* epistémicos que permiten catalizar el surgimiento de nuevas hipótesis e investigaciones. Los sesgos que se han mencionado para AF2 no son nuevos ni propios de esta metodología y han estado allí desde que existen las áreas de investigación que hemos utilizado como ejemplo. Por otro lado, debemos dejar de lado la visión de que existe tal cosa como los fines últimos de la investigación científica. El producto epistémico de toda metodología puede ser visto también como una herramienta a ser utilizada como *input* para otras metodologías y así sucesivamente.

3. El problema *in-vitro/in-vivo* y la extrapolación

Suele decirse que la bioquímica es la ciencia que para estudiar un sistema biológico primero lo destruye. Tomada literalmente, esta frase es una descripción bastante acertada para muchas prácticas dentro de esta disciplina. Gran parte de los objetos que se investigan en los laboratorios de bioquímica no corresponden a organismos o sistemas biológicos completos, sino que a partes de éstos en contextos artificiales; por ejemplo, proteínas purificadas, extractos celulares, cultivos celulares, órganos aislados, etc. Justificar cómo podemos obtener conocimiento acerca de la naturaleza de los sistemas biológicos de esta forma es lo que se conoce como el problema *in-vitro/in-vivo* (Nagel, 1961; Strand, 1996, 1999; Jacob, 2002). Este es uno de los problemas epistemológicos más importantes de la bioquímica.

Este problema es una instancia de un problema mucho mayor, a saber, el problema de la extrapolación, que consiste básicamente en justificar cómo podemos obtener conocimiento acerca de una cosa estudiando otra diferente. En el caso de la bioquímica y la biología experimental consistiría en cómo podemos justificar que estamos conociendo la ‘naturaleza biológica’ si lo que realmente estudiamos son sistemas artificiales en un laboratorio.

En los laboratorios no solo se estudian partes purificadas de organismos, también encontramos entidades biológicas y químicas especialmente producidas, seleccionadas y estandarizadas para poder ser utilizadas adecuadamente en un laboratorio como, por ejemplo, enzimas, plásmidos, drogas y también líneas celulares comerciales o cepas de bacterias, levaduras, plantas, mamíferos, insectos u otros; todos mantenidos en condiciones altamente artificiales, como regímenes estandarizados de nutrientes, temperatura, luz, esterilidad, etc. Más que representantes ideales de sus especies o de sus familias biológicas, corresponden a líneas genéticas estables, estandarizadas y seleccionadas de acuerdo con su utilidad para la producción de cierto tipo de resultados. Esto se

complica aún más si se considera que en muchos casos estos organismos no son utilizados para conocer aspectos de sus respectivas especies, sino que de otras, como ocurre en gran parte de los laboratorios de biomedicina donde se suelen utilizarse ratas como modelo de estudio para conocer aspectos de la salud humana (en este punto podríamos preguntarnos insidiosamente por qué no simplemente decimos que son laboratorios de medicina veterinaria).

En esta sección se hará un breve recorrido por los principales aspectos de este problema, por cuáles son las estrategias que han sido utilizadas para lidiar con él en la práctica científica y cómo podemos analizarlo filosóficamente.

3.1. Sistemas experimentales

Si el objetivo final de la biología experimental realmente es conocer los sistemas biológicos tal cual existen en la naturaleza, entonces es correcto afirmar que nunca tenemos real acceso a lo que pretendemos estudiar. Por otro lado, efectivamente los sistemas que se estudian en los laboratorios son entidades artificiales construidas con el fin de ser operadas para producir cierto tipo de resultados. Estas afirmaciones tan taxativas podrían parecer propias de un escepticismo respecto a las ciencias biológicas. Sin embargo, no cabe duda de que la biología experimental, y la bioquímica en particular, son ciencias exitosas. No solo nos brindan explicaciones acerca de los sistemas biológicos, sino que además nos han permitido realizar predicciones novedosas, intervenir en diversos fenómenos, proveer herramientas a una gran cantidad de otras disciplinas, generar importantes aplicaciones tecnológicas, curar enfermedades, etc. ¿Cómo podemos explicar este éxito?

Para entender esto primero debemos considerar lo que ocurre en los laboratorios y el conocimiento que allí se produce. A lo que tenemos acceso en estos contextos es a sistemas experimentales (SEs). Éstos corresponden a la unidad mínima de investigación en biología experimental (Rheinberger, 1997; Weber, 2005). Son sistemas de manipulación consistentes en un conjunto de artefactos, materiales químicos y/o biológicos, instrumentos de medición, etc. ensamblados en cierta configuración con el fin de producir un fenómeno determinado y controlarlo a voluntad. Los SEs son sometidos a un riguroso proceso de estandarización, lo que permite la obtención de resultados confiables y replicables (Esposito y Vallejos, 2020). Vale decir, si se ensambla siguiendo correctamente ciertas instrucciones, un usuario con las habilidades necesarias debería ser capaz de producir el mismo fenómeno independiente del lugar y tiempo donde se ejecute. Podemos ver entonces que en los laboratorios se genera conocimiento práctico altamente sólido acerca de la confección, estandarización y operación de sistemas experimentales. Sin embargo, los objetivos de quienes operan un SE van más allá de lo meramente práctico. Un SE se utiliza para generar conocimiento acerca de ciertas entidades o procesos.

Un sistema experimental de interés biológico contiene, valga la redundancia, algún tipo de entidades biológicas de las que se desea obtener conocimiento, las que llamaremos ‘objetos epistémicos’ (OEs) (Rheinberger 1997)¹⁷. Éstas se encuentran en un medio artificial que les otorgue estabilidad y las condiciones susceptibles para ser manipuladas. Por ejemplo, una proteína *in vitro* se encuentra en una concentración tal que no precipite ni se disocie (en caso de ser un oligómero) y

¹⁷ Esta es una traducción de lo que Rheinberger llama ‘epistemic things’ (Rheinberger, 1997]

en una solución con pH estable, sales, agentes reductores y otros componentes para mantenerla soluble y estable. Una rata de laboratorio es criada aislada de posibles infecciones y perturbaciones ambientales, además de ser sometida a un régimen de vida (alimentación, luz, temperatura, etc.) controlado y ordenado.

Un concepto que surge inmediatamente en el caso de un SE *in-vitro* es el de ‘similitud con las condiciones fisiológicas’. Asumiendo que dichas condiciones se conocen, ¿qué significaría que un sistema tenga condiciones similares a estas? Podríamos fijar la temperatura, el pH, osmolaridad, etc. de forma que sean similares a dichas condiciones, pero estaría lejos de ser suficiente, puesto que en las condiciones fisiológicas el OE se encuentra envuelto en una multitud de conexiones causales con otras moléculas y en un ambiente que ni siquiera se asemeja a una solución diluida como suelen ser los medios experimentales (Ross, 2016). Podría sugerirse entonces que se agreguen componentes adicionales, como otras proteínas, azúcares, solutos pequeños, etc. de forma que nos acerquemos cada vez más a las condiciones fisiológicas. Pero esto tendría dos consecuencias: la primera es que a medida que hacemos esto cada vez habrá más factores causales en juego, lo que ocasionaría una pérdida de control de SE. La segunda consecuencia es que, pese a todo, seguiríamos estando frente a un sistema *in-vitro*, solo que más complejo.

El control que se ejerce sobre un fenómeno en un SE se realiza mediante su manipulación a través de un mecanismo causal subyacente (Baetu, 2019) en el que participan varias entidades, entre las cuales se encuentra el OE. Éstas han sido colocadas en este contexto deliberadamente con el fin de que, a través de factores causales específicos asociados a sus propiedades, interaccionen con el OE, logrando así manipularlo para que manifieste una influencia causal sobre algún medio de detección (ya sea en forma directa o mediante la interacción con otros componentes intermediarios). Gracias a este mecanismo, al manipular un SE podemos aislar factores causales¹⁸ de nuestro OE, logrando de esta forma caracterizar sus propiedades.

En un sistema natural existe una gigantesca cantidad de factores causales en acción, algunos de los cuales podrían incluso ser totalmente desconocidos. Sin embargo, en la investigación científica solo se seleccionan algunos de ellos, relegando otros a las condiciones de trasfondo a través de su control o aleatorización. Por ejemplo, un experimento en el que se busca caracterizar alguna propiedad de una proteína es realizado manteniendo fija la temperatura, el pH, la fuerza iónica, procurando condiciones de solubilidad y estabilidad, etc. Por otro lado, en todo experimento se realizan controles para descartar la influencia causal de diversas variables. Además, siempre es necesario replicar el experimento para descartar artefactos, posibles errores y/o la acción de algún factor causal no considerado (o no conocido). En este sentido, un SE puede ser visto como un artefacto diseñado específicamente para el aislamiento de cierta clase de factores causales

¹⁸ El término ‘factores causales’ lo utilizaremos sin entrar en detalles metafísicos acerca de la causalidad y las propiedades. En general, con él nos referiremos a disposiciones causales provistas por ciertas propiedades de una entidad, aunque en la práctica bioquímica también suele indicar procesos tales como catálisis e interacciones, lo que usualmente es identificado como ‘actividad’ en la filosofía. Metafísicamente hablando, a través de los factores causales de las entidades que pusimos deliberadamente en el sistema experimental buscamos la manifestación de ciertas disposiciones causales de la entidad a ser estudiada, de forma que luego podamos caracterizar la influencia causal que ésta ejerza sobre ciertos componente del sistema, como otras moléculas o un detector.

específicos, para lo que se utilizan otros factores causales previamente aislados de entidades cuyas propiedades son ya conocidas y caracterizadas de tal forma que pueden ser utilizadas para producir y controlar diversos tipos de fenómenos.

Sin embargo, sigue siendo el caso que los SE son artificiales. Si bien se está obteniendo exitosamente conocimiento acerca de factores causales sobre diversas entidades, cabe preguntarse si éstos no son más que artificios producidos deliberadamente en este contexto artificial. Vale decir, en este punto aún no estamos justificados en decir que estos factores causales se manifiestan en un medio natural donde hay una cantidad gigantesca de otros factores en acción. Sin embargo, la única forma que tenemos de generar algún conocimiento confiable, estable y replicable sobre las entidades biológicas es mediante este proceso de aislamiento y control (Esposito y Vallejos, 2020). De lo contrario, nos veríamos enfrentados a una situación donde difícilmente podríamos identificar cuáles son las causas de los fenómenos observados ni cuales son los aspectos que a éste subyacen puesto que para esto siempre necesitamos intervenir sobre el sistema. Nos encontramos entonces frente un conflicto epistemológico: solo podemos conocer la naturaleza a partir de la artificialidad. A continuación, revisaremos cómo podemos abordar este problema de forma de encontrar alguna solución.

3.2. Extrapolación

Cuando el objetivo epistémico es obtener conocimiento acerca de un sistema objetivo, como por ejemplo un sistema biológico ‘tal cual se da en la naturaleza’, decimos que los SE utilizados para esto son modelos subrogados de dicho sistema (Bolker, 2009; Baetu, 2016). Vale decir, el SE se utiliza como un sustituto en el que es posible realizar manipulaciones y obtener conocimiento acerca de factores causales específicos. Esto a diferencia de lo que ocurriría simplemente observando el sistema ‘tal cual existe en la naturaleza’, puesto que en ese caso sería imposible obtener conocimiento sin intervención. El desafío entonces es justificar la extrapolación entre un modelo subrogado y un sistema objetivo (Baetu, 2016). Dependiendo del tipo de subrogación, la extrapolación tendrá características epistemológicas diferentes a ser exploradas para justificarla. Usualmente, en la literatura el problema de la extrapolación ha sido tratado en forma general sin grandes distinciones entre distintos tipos de subrogación (Steel, 2008; Baetu, 2016). La atención ha sido puesta principalmente en el uso de organismos modelo y en la posibilidad de extrapolar aspectos relacionados con la salud humana a partir de éstos (LaFollette y Shanks 1993, 1995; Culp, 1997; Douglas, 2000; Ankeny, 2001; Weber, 2005; Steel, 2008; Bolker, 2009; Ankeny y Leonelli, 2011, 2021; Baetu, 2016; Tee, 2019). Con algunas excepciones encontramos también tratamientos filosóficos de extrapolaciones no directamente relacionadas con organismos modelo, como el problema *in-vitro/in-vivo* cuyos aspectos epistémicos son diferentes (Nagel, 1961; Strand, 1996, 1999; Jacob, 2002; Douglas, 2000).

En general podemos distinguir dos tipos de extrapolación:

- **Extrapolación composicional:** el Modelo subrogado consiste en partes aisladas del sistema objetivo (*e.g.* el problema *in-vitro/in-vivo*).
- **Extrapolación Comparativa:** el modelo subrogado no es una parte ni tiene una relación material con el sistema objetivo.

3.2.1. Extrapolación Composicional

La meta de este tipo de extrapolación es la generación de algún modelo acerca del funcionamiento de un sistema objetivo a partir de la integración de los diversos factores causales de sus partes. Para esto es necesario justificar que dichos factores causales son efectivamente relevantes o se manifiestan realmente en el sistema objetivo y permiten explicar algún fenómeno de interés. El primer tratamiento filosófico sistemático de este problema lo podemos encontrar en la obra clásica de Ernst Nagel (1961) *La estructura de la ciencia* de 1961. Sin embargo, fue mayormente descuidado hasta que fue brevemente retomado en los noventa por bioquímicos preocupados por el problema *in-vitro/in-vivo* (Strand 1996, 1999; Jacob, 2002), quienes sostuvieron posiciones más bien escépticas al respecto.

Obviamente la distinción entre extrapolación composicional y comparativa no es absoluta y siempre cabrá la posibilidad de casos límite, como podría ser, por ejemplo, el uso de líneas celulares humanas como modelo subrogado de aspectos fisiológicos humanos. Estas células, si bien son humanas, no son extraídas de un humano cada vez que se utilizan en un SE. Se trata de células comerciales almacenadas y reproducidas en el laboratorio que corresponden a un linaje proveniente de una célula que fue extraída alguna vez desde un humano. Lo mismo podría decirse de proteínas producidas recombinantemente y no directamente aisladas de nuestro sistema objetivo. En todos esos casos el modelo subrogado no posee continuidad material con el sistema objetivo, sino que corresponde a un sistema que ha sido estandarizado para facilitar su utilización en el contexto experimental. Lo importante es que, independiente de la continuidad material, estemos justificados para decir que esta entidad material es lo suficientemente similar a una parte del sistema objetivo (a una célula, una proteína, etc.), de forma que es posible un razonamiento composicional en el que ciertas propiedades de los modelos subrogados puedan ser integradas en un modelo acerca de mecanismos del sistema objetivo y su funcionamiento. El problema es entonces justificar que una propiedad presente en un SE será relevante en este mecanismo del sistema objetivo.

Una forma de justificar la relevancia de una propiedad es acceder a ella mediante el uso de diversos medios independientes. Cuando varios SEs poseen factores causales comunes, vale decir, cuando estamos frente a una propiedad robusta, entonces tenemos buenas razones para confiar en su existencia (Wimsatt, 1981; Culp, 1995; Nederbragt, 2003; Soler *et al.*, 2012; Eronen 2015). Por otro lado, cuando un factor causal ha sido aislado mediante múltiples métodos y está bien caracterizado es posible usarlo para construir nuevos sistemas experimentales y generar nuevos fenómenos. Un ejemplo de esto es la carga eléctrica del ADN, que la utilizamos para separar fragmentos de distintos tamaños en geles de agarosa, fijar un plásmido en una superficie y detectarlo mediante microscopía de fuerza atómica, insertar un plásmido dentro de una bacteria y generar resistencia a antibióticos, etc. Gracias a esta robustez es que podemos justificar, por ejemplo, que se está trabajando con la misma propiedad de un mismo objeto en más de un SE (Burian, 1995).

Interesantemente, el recurrir a entidades purificadas en un contexto *in-vitro*, además de presentar problemas epistemológicos a ser justificados, también nos brinda poderosas herramientas para resolverlos. La mayoría de los tratamientos filosóficos sobre el tema parten de la entidad ya purificada (Strand 1996, 1999; Jacob, 2002; García, 2015; Ibarra y Mormann, 2006) pasando por

alto el arduo proceso de purificación y preparación que fue necesario para poder confeccionar el SE *in-vitro*. Un proceso de purificación consiste en una secuencia de SEs interconectados materialmente en que los productos de uno son utilizados como componentes del siguiente. En todas estas etapas se deben aislar factores causales específicos para manipular la entidad a ser purificada y así cambiarla sucesivamente de ambiente, desde su medio biológico original hasta un medio *in-vitro* final. Si tomamos como ejemplo la purificación de una proteína, se comienza con la lisis celular de un cultivo bacteriano para producir un extracto crudo, el que luego es sometido a diversos SEs tales como cromatografías, precipitación selectiva, diálisis, concentración, etc. hasta obtener una muestra final de proteína pura en un medio que propicia su estabilidad. En cada etapa de la purificación se utilizan propiedades tales como su tamaño, forma, carga eléctrica, solubilidad, estabilidad térmica, capacidad de interacción con otras moléculas, actividad enzimática, etc. Todas estas propiedades garantizan el control causal de cada SE que forma parte del proceso de purificación.

Un proceso de purificación es un tipo de experimentación preparativa (Weber, 2005; Waters, 2008), que consiste en la clase de procedimientos que se llevan a cabo en un laboratorio para producir materiales para construir otros sistemas experimentales. Esto incluye también el crecimiento de células y organismos, la selección de cepas, el almacenamiento de muestras, etc. En todos estos procesos se utilizan una gran cantidad de factores causales que no solo son robustos, sino que, además, para garantizar el éxito experimental, deben ser más confiables que los que luego se desean estudiar a través de un SE. De esta forma, cada entidad que conforma un SE es producto de un proceso de experimentación preparativa (ya sea realizado en el mismo laboratorio o en alguna industria). Esto tiene como consecuencia que los SE están conectados a través de vastas redes materiales, en donde un tipo de material puede ser utilizado para construir muchos SEs y, así mismo, cada SE es la conjunción de una gran cantidad de elementos provenientes de otros SEs. De esta forma, dudar de la existencia y de la eficiencia causal de las propiedades que aparecen en múltiples contextos experimentales dentro de esta red material conllevaría el costo de dudar de una gran cantidad de estrategias exitosas de manipulación y control que éstas posibilitan (Esposito y Vallejos, 2020).

A la luz de lo expuesto, podemos considerar que, para que sea confiable, una extrapolación composicional debe estar basada en conocimiento robusto acerca de propiedades de entidades. De esta forma, éstas pueden ser situadas en un contexto teórico más amplio y/o dentro de modelos que pretenden describir mecanismos de sistemas biológicos (Nederbragt, 2003). Sin embargo, esto aún debe ser puesto a prueba. La forma de hacerlo, nuevamente, es mediante el control de fenómenos a través de sistemas experimentales, pero esta vez utilizando estas propiedades robustas y los modelos mecanísticos donde ellas han sido situadas para manipular sistemas más complejos, tales como células y organismos. Vale decir, utilizar la célula y/o el organismo modelo como tubo de ensayo para poner a prueba una hipótesis proveniente de una extrapolación composicional.

A medida que vamos ‘subiendo de nivel de organización’¹⁹, van apareciendo factores causales sobre los cuales no tenemos un control total, por lo que las estrategias para controlar el fenómeno

¹⁹ En este artículo no entraré en detalles respecto al concepto de ‘niveles de organización’. En este caso solo bastaría con considerar jerarquías composicionales que tengan alguna relevancia explicativa para un tipo de fenómenos.

serán diferentes. Por otro lado, también habrá un aumento en la dependencia respecto a teorías biológicas en los mecanismo subyacente del SE (Baetu, 2019). Sin embargo, como en todo SE, lo que estamos haciendo es generar y controlar un fenómeno utilizando para esto factores causales de los que sí poseemos amplio conocimiento, para así aislar algún factor causal de la entidad interés en un nuevo contexto. De esta forma, el control exitoso de un sistema experimental 'de alto nivel de organización' utilizando factores causales descritos usando sistemas experimentales 'de bajo nivel' nos da buenas razones para afirmar que este factor causal también será relevante en los sistemas biológicos 'fuera del laboratorio'. Por ejemplo, luego de caracterizar las propiedades fisicoquímicas y estructurales de una enzima, podemos generar una mutación genética que disminuya su capacidad catalítica y de esta forma controlar un SE consistente en un cultivo bacteriano produciendo que las bacterias no puedan crecer en cierto tipo de medios.

Con la estrategia antes expuesta los científicos logran establecer la relevancia de varias propiedades de diversas entidades en variados contextos, observando la manifestación de factores causales asociados a éstas frente a distintos tipos de manipulación. Luego, considerando varias entidades y diversos factores causales asociados a ellas es posible extrapolar composicionalmente, integrándolas en contextos teóricos más amplios y generando modelos acerca de mecanismos del sistema biológico objetivo. Finalmente, éstos pueden ser evaluados considerando SEs donde se utiliza el organismo (o una célula, tejido, etc.) como tubo de ensayo para analizar la manifestación de dichos factores causales. Si bien no existe tal cosa como una extrapolación definitiva y no hay alguna consideración metafísica o epistemológica que nos lo garantice, vemos que sí es posible generar conocimiento altamente confiable en términos de que dudar de éste conllevaría grandes costos epistemológicamente hablando.

3.2.2. Extrapolación comparativa

El uso de organismos modelos es una práctica ampliamente difundida en la biología, sobre todo en biomedicina. El estudio de los procesos fisiológicos, y más delante de los procesos bioquímicos, trajo consigo el costo de sacrificar la diversidad que había caracterizado a la biología hasta el siglo XIX en cuanto a sus objetos de estudio. De la pluralidad de ejemplos se pasó al estudio a fondo de los mecanismos de unos pocos organismos a modo de ejemplares susceptibles de ser utilizados en el laboratorio. Es así como disciplinas completas han nacido gracias al estudio de unos pocos organismos modelo, como la genética moderna gracias a la mosca *Drosophila* (Kohler, 1994), gran parte de la fisiología gracias a las ratas y ratones, la biología del desarrollo contemporánea gracias *C. elegans* y el pez cebra, la neurociencia gracias a los calamares, la microbiología gracias a *E. coli*, la biología vegetal gracias a *Arabidopsis thaliana* y así un largo etcétera.

La elección de un organismo modelo para el estudio de cierto tipo de mecanismos es un aspecto tremendamente importante, puesto que, si el objetivo es generalizar el conocimiento para toda una clase de sistemas biológicos, entonces se esperaría que éste tuviese características que lo hagan generalizable. Sin embargo, los criterios que se utilizan para elegir organismos modelo son principalmente prácticos y tienen que ver con la idoneidad para su uso en el contexto experimental (Clarke y Fujimura, 1992; Burian, 1993; Bolker, 1995). Esto se traduce no solo en la idoneidad de dicho organismo para poder estudiar cierta clase de mecanismos específicos, sino que también en tiempos adecuados de desarrollo, rapidez de reproducción, facilidad de generar líneas genéticas

puras, estabilidad en las condiciones del laboratorio o incluso en características tales como facilidad de ser analizado con microscopios. Esto claramente tiene como consecuencia la posibilidad de sesgos en las teorías que surjan a la hora de generalizar el conocimiento así obtenido. Cuando un organismo modelo está siendo usado como modelo subrogado respecto de otro sistema, como por ejemplo el ser humano, entonces surgen problemas epistemológicos adicionales. Extrapolar entre dos organismos de distintas especies no es una tarea fácil ni directa, por lo que debe ser adecuadamente justificada.

Una posibilidad es el uso de criterios filogenéticos. Básicamente, si observamos que un mismo mecanismo es encontrado en organismos bastante distantes filogenéticamente, esto es evidencia de que debe tratarse de un rasgo conservado en varias ramas del árbol de la vida, concretamente en las que se agrupan los organismos investigados y las ramas intermedias entre éstos (Weber, 2005). El ejemplo por excelencia es el código genético. Así mismo, también habrá rasgos mecánicos propios de clados más limitados, por ejemplo, solo aplicables a ciertos peces, a mamíferos o a primates. Este razonamiento se basa en criterios de parsimonia que son propios de la filogenética y la evolución.

Esta estrategia no está exenta de dificultades. Primero que todo, usualmente en un proceso de extrapolación comparativa son dos organismos los que se están comparando. Pero para justificar esta extrapolación en términos filogenéticos, deberán ser muchos los organismos que se comparen. Mientras más se consideren, más robustas serán las inferencias. El resultado será una triangulación cuyo éxito justificativo dependerá de que la información filogenética disponible sea suficiente. Sin embargo, esto implicaría que se tengan que analizar la presencia y conservación del mecanismo de interés para muchas especies dentro de un clado (o dentro de múltiples clados), lo que se vuelve extremadamente difícil. Además del número de casos a registrar, también se hará necesario contar con evidencia suficiente para caracterizar dicho mecanismo en cada caso. Paradójicamente, el uso de organismos modelo como subrogados para generar extrapolaciones ha surgido precisamente para evitar esta ardua tarea. Si bien los criterios filogenéticos sí aportan a la hora de justificar algunas extrapolaciones, sobre todo acerca de mecanismos tan conservados como el código genético o algunos procesos de desarrollo, usualmente serán otros los criterios relevantes a tener en consideración.

Otra estrategia que se ha discutido ampliamente en la literatura es el criterio de la similitud de mecanismos (LaFollette y Shanks, 1993, 1995; Ankeny, 2001; Steel, 2008, Baetu; 2016). Básicamente, si queremos extrapolar entre dos organismos debe existir evidencia previa de que existe una similitud suficiente entre ambos respecto a los mecanismos relevantes que realizan el fenómeno que se desea extrapolar. Sin embargo, esta estrategia también conlleva problemas. Primero que todo, si ya conocemos suficiente acerca del mecanismo de interés en el sistema objetivo, ¿por qué entonces querríamos extrapolar desde otro organismo? (Steel, 2008). Por otro lado, tener suficiente conocimiento sobre el mecanismo de interés en el sistema objetivo implica que debió existir un proceso previo para su obtención. Dado que es necesario acudir a modelos subrogados (si no fuera el caso, entonces no tendríamos por qué preocuparnos de la extrapolación y no estaríamos dando esta discusión), entonces este conocimiento disponible debe provenir, a su vez, de extrapolaciones previas, las que también deberán estar justificadas. De esta forma caemos en un

círculo vicioso del cual se vuelve difícil salir (Steel, 2008; Baetu, 2016). Para romper este círculo se requeriría de evidencia independiente que nos permita justificar las extrapolaciones desde algún punto.

Dado que lo que se está comparando son mecanismos y que estos están compuestos tanto de entidades como procesos que realizan un fenómeno de interés, una estrategia que podría utilizarse para justificar la extrapolación comparativa es acudir a la extrapolación composicional. Vale decir, una estrategia rigurosa y sistemática para extrapolar entre organismos de distintas especies es atender a la presencia de partes equivalentes que tienen propiedades compartidas y reconstruir los mecanismos de ambos organismos mediante extrapolación composicional, así como se completa un rompecabezas. Esto, además, permitiría también inferir la presencia de desanalogías causales entre el modelo subrogado y el sistema objetivo al encontrar propiedades que no son comunes y que podrían afectar la realización del fenómeno de interés. Vale decir, utilizando la analogía del organismo como ‘tubo de ensayo’, lo que estamos haciendo es aislar factores causales asociados a entidades del organismo modelo en un contexto fisiológico, de forma que con este conocimiento se pueda reconstruir el mecanismo de interés en el sistema objetivo basándose en la evidencia acerca de la presencia de partes con determinadas propiedades. De esta forma, gracias a la distinción que hemos hecho entre los dos tipos de extrapolación, es posible concebir una herramienta para romper el círculo de la extrapolación. Sin embargo, aún nos falta un paso que dar, puesto que para que esta justificación sea exitosa debemos tener un amplio conocimiento acerca de la presencia de partes con ciertas propiedades tanto del organismo modelo como del sistema objetivo, lo que implica que deberíamos, dentro de lo posible, caracterizar todas estas partes para ambos casos, lo que debería extenderse para todo organismo modelo que se esté utilizando como subrogado para algún sistema biológico.

Para resolver este problema debemos regresar a la inferencia filogenética, pero esta vez ya no a nivel de organismos, sino que a nivel (macro) molecular. Hoy en día existe una amplia gama de metodologías que permiten comparar secuencias de genes de proteínas y así establecer filogenias. Esto implica que, una vez caracterizada una proteína a nivel estructural y fisicoquímico, es posible obtener información acerca de residuos relevantes para diversas funciones asociadas a ésta. Por ejemplo, podemos establecer cuáles son los motivos de secuencia relevantes para la unión de un sustrato en una enzima. Con esta información es posible rastrear la presencia de un rasgo molecular analizando la conservación de residuos aminoacídicos en posiciones específicas de la secuencia. Incluso es posible inferir si un rasgo, por ejemplo, la capacidad de catalizar una reacción, corresponde a una novedad evolutiva o si es ancestral y se ha perdido en algún linaje. De esta forma es posible inferir las propiedades de las macromoléculas presentes en distintos organismos, pudiendo así brindar una poderosa herramienta para la justificación de una extrapolación composicional en un organismo del cual aún no hemos sido capaces de estudiar sus partes en forma aislada. Y, más importante aún, es posible inferir así sus similitudes y posibles desanalogías.

Si bien nunca obtendremos una extrapolación completa y no existen principios epistemológicos y/o metafísicos que nos permitan justificarla definitivamente, dependiendo del contexto epistémico que se considere se podrá analizar si las justificaciones son o no suficientes. Vale decir, estamos en la misma situación que expusimos en el caso de AF2, dependerá del contexto investigativo y de

consideraciones pragmáticas hasta dónde deseamos justificar. Por ejemplo, si el contexto es la generación de políticas públicas y el objetivo es decidir si se lanza o no al mercado un producto que podría ser cancerígeno, entonces claramente la justificación requerirá una gran cantidad de evidencia proveniente de muchas metodologías distintas y de diversas extrapolaciones de todo tipo (Douglas, 2000; Douglas, 2009). Por otro lado, si el contexto tiene que ver con la ‘ciencia básica’, tal vez las extrapolaciones podrán quedar como hipótesis y caer en un ciclo iterativo de pruebas y errores donde no existe más apremio que la generación de conocimiento. Por último, si el ejemplo es una aplicación industrial, tal vez la justificación de una extrapolación será lo que menos importe mientras se logre optimizar y maximizar la producción en cierto bioproceso²⁰. En este caso, la opacidad respecto a los mecanismos subyacentes y la justificación de la extrapolación podrá relegarse a un segundo plano mientras las cosas funcionen como se espera. Así nos encontramos también frente a un caso donde lo importante es una red interconectada de experimentación que permita triangular mecanismos a través de distintos tipos de extrapolación.

4. Lecciones filosóficas: Reduccionismo y pluralismo

Hemos revisado dos casos de estudio de problemas filosóficos que surgen en el seno de la investigación científica en bioquímica. Debido a la naturaleza plural y dispersa de esta disciplina, es natural que su tratamiento filosófico también resulte plural y disperso, debiendo enfocarse en temas específicos que surgen en contextos determinados de investigación científica. Sin embargo, existen problemas filosóficos que históricamente han acompañado a la bioquímica desde su mismo nacimiento. Uno de ellos es el problema del reduccionismo.

Una de las principales acusaciones que se realizan a la bioquímica es el ser una ciencia reduccionista, en el sentido que pretende imponer una visión química y molecular a todas las ciencias de la vida. Muchas veces el término ‘reduccionista’ es utilizado como una especie de arma arrojadiza que busca anular o disminuir el peso de cierto tipo de razonamientos o investigaciones. Sin embargo, también existen quienes han utilizado el término como sinónimo de una virtud en el sentido de que no estaría más que describiendo una de las grandes ventajas de las ciencias químico-biológicas: el brindarnos explicaciones moleculares de diversos fenómenos además de abrir constantemente nuevas posibilidades para el surgimiento nuevas líneas de investigación en una gran variedad de campos (Ureta, 2003). En esta sección mostraremos que existe un sentido en el que cierto tipo de reduccionismo es inevitable, pero que, sin embargo, es perfectamente compatible con el pluralismo de las ciencias biológicas.

Por varias décadas, en la filosofía de la biología existió un consenso anti reduccionista (Brigandt y Love, 2017) que, o bien sostenía la imposibilidad de la reducción o, más débilmente, que ésta no es la mejor opción para la biología debido a sus limitaciones. De esto se desprende que las metodologías reduccionistas deberían paulatinamente reemplazarse para así obtener mejor conocimiento biológico. Mientras más complejo e integrado el sistema, más inadecuadas son las explicaciones reductivas (Kaiser, 2011). Pero ¿qué significaría esto? ¿Qué quiere decir que la

²⁰ Lo mismo ocurriría en el caso del diseño de fármacos, donde, como vimos, lo importante es que una droga funcione adecuadamente.

bioquímica sea reduccionista? Primero es necesario analizar qué entendemos en este caso por ‘reduccionismo’.

No cabe duda de que ‘reduccionismo’ es un concepto con gran ambigüedad. A los bioquímicos se nos suele hacer la acusación de querer ‘reducir toda la biología a la química y la física’. Este enunciado puede interpretarse al menos de tres formas distintas (Suárez y Martínez, 1998; Brigandt y Love, 2017). Primero, podría interpretarse ontológicamente como que todas las entidades y procesos biológicos son de naturaleza química y física. Esto está lejos de ser problemático y, más que eso, es algo aceptado en toda la biología, de lo contrario se caería en vitalismo y/o pensamiento mágico. Otra cosa sería aseverar que podemos conocer las entidades y procesos biológicos a partir del estudio su naturaleza física y química. Esta segunda interpretación sería epistemológica y, a la luz de lo expuesto en las secciones anteriores, podemos ver que podría ser problemática puesto que el conocer todos los aspectos químicos y físicos de los constituyentes de alguna entidad no garantiza que podamos conocerla en cuanto a sus propiedades sistemáticas como un ‘todo’. Para esto siempre sería necesario considerar las condiciones de borde del sistema, lo que va más allá de la naturaleza física de sus partes. Sin embargo, en todo sistema debemos considerar dichas condiciones, por lo que la posibilidad queda abierta. Finalmente, aseverar la interpretación anterior no implica afirmar que la única (o mejor) forma de conocer las entidades y procesos biológicos sea estudiando su naturaleza física y química. Esta ya sería una afirmación normativa altamente problemática y debatible de la que se entendería que surjan detractores. ¿Es esta última afirmación representativa de la bioquímica? Veremos que esta afirmación, que se suele atribuir a la bioquímica, en realidad carece de sentido si consideramos seriamente la práctica científica. Lo que realmente existe son explicaciones reductivas que tendrán distinta naturaleza si consideramos cada contexto donde éstas surgen (Sarkar, 1992).

Por ‘explicación reductiva’ podemos entender varias cosas (Hüttemann y Love, 2011). Podrían estar hablando de explicaciones:

- a) Intrinsicalistas: que están centradas en aspectos internos del sistema investigado.
- b) Fundamentalistas: que están centradas en niveles de organización más fundamentales que el fenómeno a explicar.
- c) Monistas: que dan primacía a teorías (o causas) por sobre otras para explicar un conjunto amplio de fenómenos.
- d) Aislacionistas: que estudian partes aisladas del sistema investigado.

También podrían darse explicaciones que tengan más de una de estas características. Sin embargo, existe un tipo de práctica reductiva independiente de las acá mencionadas que es ubicua en toda la biología experimental. Hablamos de la selección de ciertos factores causales para la explicación de algún fenómeno, relegando los otros al trasfondo (Gannett 1999; Waters 2007). En otras palabras, el aislamiento de factores causales que hemos estado discutiendo.

Como vimos al comienzo, aislar factores causales es una condición ineludible si se desea obtener conocimiento estable, replicable y robusto²¹. La elección de qué factores seleccionar como relevantes y cuales relegar al trasfondo dependerá de alguna perspectiva particular que dictamina los objetivos epistémicos de la investigación y que nos otorga un marco de partición que identifica qué consideraremos como las partes relevantes de un sistema para proceder a su decomposición (Winther, 2011; Kaiser, 2018), ya sea teórica o materialmente. Por ejemplo, un mismo fenómeno celular puede decomponerse considerando las redes de transducción de señales que lo regulan, las redes metabólicas, los genes involucrados, los aspectos físicos que realizan lo realizan, aspectos fisicoquímicos de los distintos ambientes celulares, etc. (Wimsatt, 1974) Sin embargo, es frecuente que, dentro de un mismo programa de investigación, o incluso dentro de un mismo laboratorio, convivan múltiples perspectivas y que muchas veces sea difícil distinguir entre ellas. Por lo tanto, si bien se trata de una práctica reduccionista metodológica ineludible para hacer posible el conocimiento, se trata de un *reduccionismo local* que solo surge en contextos específicos.

Obviamente, la elección de una clase de perspectivas conlleva la posibilidad de sesgos a la hora de entender los sistemas biológicos. Es más, podría conllevar que desde un cierto programa de investigación se pretenda imponer dichas perspectivas a otros programas diferentes, o incluso a toda la biología, y que en parte lo logre si es que adquiere suficiente poder y aceptación dentro de la comunidad científica, desacreditando investigaciones que asuman perspectivas distintas. En un caso como este estaríamos frente a un *reduccionismo global*. Ejemplos como el determinismo genético cabrían dentro de esta categoría. Sin embargo, es raro que esto ocurra. Los científicos suelen estar conscientes acerca de las limitaciones de las perspectivas que asumen, así como de los riesgos epistémicos que implicaría el imponerlas a toda la biología. Sin embargo, esto no evita que se produzcan malentendidos, como que se tilde de ‘reduccionistas genéticos’ a quienes estudian factores causales genéticos por el hecho de tener que aislar causas genéticas en sus investigaciones, o que se trate peyorativamente de ‘reduccionistas’ a quienes practican la bioquímica. Estas acusaciones suelen hacerse a investigaciones que asuman perspectivas intrinsicalistas, fundamentalistas y/o aislacionistas. Sin embargo, bajo nuestra definición, una perspectiva que no posea estas características también podría caer en un reduccionismo global, como por ejemplo si se tratase de imponer alguna teoría sobre ‘la vida’ a toda la biología, o si se eligiera considerar como relevantes solo los factores externos en desmedro de los internos de un sistema en fenómenos tanto biológicos como sociales. Usualmente las discusiones que surgen dentro de la biología no buscan establecer cuál es la perspectiva ‘verdadera’, sino cual es la relevancia de algún tipo de causas o la frecuencia con la que éstas son las más relevantes para un conjunto de fenómenos (debates de relevancia significativa, como los llama Beatty [1997]). Así podemos mencionar las contraposiciones entre genes vs ambiente en el fenotipo, células vs estructuras de tejidos en el desarrollo, factores innatos vs factores ambientales y sociales en el comportamiento humano y un gran etcétera.

De esta forma, los sesgos que emergen de las perspectivas consideradas en forma individual y los intentos de reduccionismos globales que podrían aparecer en ciertas circunstancias se

²¹ Esta estrategia no solo se restringe a la práctica experimental, sino que también ocurre en el contexto de teorías y modelos matemáticos que pretendan describir o representen algún sistema, ya sea en términos mecanísticos y/o dinámicos.

contrapesan y diluyen gracias al pluralismo inherente de la biología, en la que conviven diariamente una gran cantidad de perspectivas en lo que Sandra Mitchell ha caracterizado como un pluralismo integrativo (Mitchell, 2003), en el que información aparentemente contradictoria proveniente de múltiples perspectivas es integrada (con más o menos éxito) utilizando diversas estrategias. Así, pese a la existencia de un reduccionismo metodológico inevitable y necesario a nivel local, en la biología encontramos un anti-reduccionismo a nivel global, consistente con la naturaleza del conocimiento biológico como un mosaico de conocimientos en continuo cambio y actualización (Baetu, 2016) provenientes de una gigantesca multitud de diversas perspectivas y metodologías. Como vimos, esta pluralidad de metodologías es lo que surge naturalmente en el caso de la biología estructural, donde metodologías predictivas computacionales se mezclan e interaccionan con metodologías experimentales para la determinación de estructuras macromoleculares, además de proveer de insumos (de mayor o menor utilidad dependiendo de cada caso) a diversas áreas de investigación. Lo mismo ocurre en la extrapolación desde el conocimiento obtenido a partir de una gigantesca variedad de sistemas experimentales a ‘la naturaleza biológica’. Si bien nunca habrá una extrapolación completamente justificada, siempre podemos aspirar a aumentar nuestra confianza en las extrapolaciones mediante la proliferación de nuevas metodologías y sistemas experimentales que nos permitan acceder al mismo fenómeno desde distintas perspectivas, o que nos permitan descartar extrapolaciones incorrectas y detectar artefactos.

La producción de conocimiento biológico depende de una intrincada y sólida red metodológica donde conviven múltiples perspectivas para estudiar el mismo conjunto de sistemas. Si bien cada una de ellas debe inevitablemente proceder de manera reduccionista al aislar factores causales específicos, no debemos ver este reduccionismo como un ‘problema a ser superado’, sino que debemos aceptarlo como una característica intrínseca de las ciencias. Una característica a la que no tenemos por qué temer una vez que hayamos entendido sus distintas manifestaciones y analizado cómo es que los sesgos que inevitablemente conlleva tienden a anularse gracias al pluralismo intrínseco de la biología en general y de la bioquímica en particular.

Referencias

- Abeln, S., Feenstra, K. A. y Heringa, J. (2019). “Protein Three-Dimensional Structure Prediction”. *Encyclopedia of Bioinformatics and Computational Biology*, 2: 497-511. <https://doi.org/h6mf>
- Alassia, F. (2022). “¿Es posible una ontología procesual de las entidades bioquímicas? Consideraciones a partir del caso de los receptores celulares y la señalización celular”. *Estudios De Filosofía*, (65): 153–175. <https://doi.org/h6mg>
- Alleva, K., Díez, J. y Federico, L. (2017). “Models, theory structure and mechanisms in biochemistry: The case of allosterism”. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 63: 1-14. <https://doi.org/h6mh>
- AlQuraishi, M. (2019). “AlphaFold at CASP13”. *Bioinformatics*, 35(22): 4862–4865, <https://doi.org/gh6r9r>

- AlQuraishi, M. (2020). *AlphaFold2 at CASP14: "It feels like one's child has left home"*. <https://moalquraishi.wordpress.com/2020/12/08/alphafold2-casp14-it-feels-like-ones-child-has-left-home/>
- Anfinsen C. B., Haber, E., Sela, M. y White F. H. Jr. (1961). "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain". *PNAS*, 47(9): 1309–1314. <https://doi.org/crb44z>
- Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains". *Science*, 181(4096): 223–230. <https://doi.org/fhq24>
- Ankeny, R. (2001). "Model Organisms as Models: Understanding the 'Lingua Franca' of the Human Genome Project". *Philosophy of Science*, 68(S3): S251-S261. <https://doi.org/csq853>
- Ankeny, R. y Leonelli, S. (2021). *Model Organisms*. Cambridge University Press.
- Ankeny, R., y Leonelli, S. (2011). "What's so special about model organisms?". *Studies in History and Philosophy of Science Part A*, 42: 313–323. <https://doi.org/c2z88z>
- Baetu, T. (2019). *Mechanisms in molecular biology*. Cambridge University Press.
- Baetu, T. 2016. "The 'Big Picture': The Problem of Extrapolation in Basic Research". *British Journal for the Philosophy of Science*, 67(4): 941-964. <https://doi.org/gdxrgk>
- Bartol, J. (2016). "Biochemical Kinds". *The British Journal for the Philosophy of Science*, 67(2): 531–551. <https://doi.org/f3r4v2>
- Beatty, J. (1997) "Why do biologists argue like they do?". *Philosophy of Science* 64 (S4): s432-s443. <https://doi.org/cdr4>
- Bechtel, W. y Richardson, R. C. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge: MIT Press.
- Berman H. M. (2008). "The Protein Data Bank: a historical perspective". *Acta Crystallogr A.*, 64(Pt 1): 88-95. <https://doi.org/chm79x>
- Berman H. M. y Gierasch L. M. (2021). "How the Protein Data Bank changed biology: An introduction to the JBC Reviews thematic series, part 1". *Journal of Biological Chemistry*. 296: 100608. <https://doi.org/h6mk>
- Bolker, J. (1995). "Model systems in developmental biology". *BioEssays*, 17: 451–455. <https://doi.org/dtgp7f>
- Bolker, J. (2009). "Exemplary and surrogate models: Two modes of representation in biology". *Perspectives in Biology and Medicine*, 52; 485–499. <https://doi.org/gmv82t>
- Brigandt, I. y Love, A. (2017). "Reductionism in Biology". En E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/sum2022/entries/reduction-biology/>

- Briggs, H. (2020). One of biology's biggest mysteries 'largely solved' by AI. *BBC News*. <https://www.bbc.com/news/science-environment-55133972>
- Burian, R. (1993). “How the Choice of Experimental Organism Matters: Epistemological Reflections on an Aspect of Biological Practice”. *Journal of the History of Biology*, 26: 351–367. <https://doi.org/dccgkp>
- Burian, R. (1995). “Comments on Rheinberger”. En G. Wolters, J. G. Lennox y P. McLaughlin (eds.). *Concepts, Theories, and Rationality in the Biological Sciences*. Pittsburgh: University of Pittsburgh Press.
- Clarke, A. E. y Fujimura J. H. (1992). *The Right Tools for the Job. At Work in Twentieth-century Life Sciences*. Princeton: Princeton University Press.
- Collins, H. M. (1987). *Changing order: Replication and induction in scientific practice*. Chicago: Chicago University Press.
- Culp, S. (1995). “Objectivity in Experimental Inquiry: Breaking Data-Technique Circles”, *Philosophy of Science*, 62: 430–450. <https://doi.org/cnmfg4>
- Culp, S. (1997). “Establishing genotype/phenotype relationships: Gene targeting as an experimental approach”. *Philosophy of Science*, 64(4): S268–S278. <https://doi.org/fmsecjj>
- Dahlin, J., Inglese, J. y Walters, M. (2015) “Mitigating risk in academic preclinical drug discovery”. *Nature Reviews Drug Discovery*, 14: 279–294. <https://doi.org/10.1038/nrd4578>
- de Chadarevian, S. (2011). *Designs for Life: Molecular Biology after World War II*. Cambridge University Press.
- Douglas, H. (2000). “Inductive risk and values in science”. *Philosophy of Science*, 67: 559–579. <https://doi.org/djjsrh>
- Douglas, H. (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Eronen, M. (2015). “Robustness and reality”. *Synthese*. 192: 3961–3977. <https://doi.org/gns9h2>
- Esposito, M. y Vallejos, G. (2020). “Performative Epistemology and the Philosophy of Experimental Biology: A Synoptic Overview”. En Baravalle, L. y Zaterka, L. (eds.). *Life and Evolution. Latin American Essays on the History and Philosophy of Biology*. Springer.
- Evans P, McCoy A. (2008). “An introduction to molecular replacement”. *Acta Crystallographica Section D*, 64(Pt 1): 1-10. <https://doi.org/br4pq6>
- Fasman, G. D. (1989). “The Development of the Prediction of Protein Structure”. En G.D Fasman (ed.). *Prediction of Protein Structure and the Principles of Protein Conformation*. Boston: Springer. <https://doi.org/dwgxck>
- Fersht, A. (2017). *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. Singapore: World Scientific. <https://doi.org/h6nf>

- Figuerola, M., Sleutel, M., Vandevenne, M., Parvizi, G., Attout, S., Jacquin, O., Vandenameele, J., Fischer, A. W., Damblon, C., Goormaghtigh, E., Valerio-Lepiniec, M., Urvoas, A., Durand, D., Pardon, E., Steyaert, J., Minard, P., Maes, D., Meiler, J., Matagne, A., Martial, J. A. y Van de Weerd, C. (2016). "The unexpected structure of the designed protein Octarellin V.1 forms a challenge for protein structure prediction tools". *Journal of Structural Biology*, 195(1): 19-30. <https://doi.org/f8rjfk>
- Gannett, L. (1999). "What's in a cause? The pragmatic dimensions of genetic explanations". *Biol. Philos.*, 14(3): 349–373. <https://doi.org/ccj9cc>
- García P. (2015). "Computer simulations and experiments: *in vivo*–*in vitro* conditions in biochemistry". *Foundations of Chemistry*, 17(1): 49-65. <https://doi.org/gns9kd>
- Gelfert, A. (2011). "Scientific models, simulation, and the experimenter's regress". En P. Humphreys y C. Imbert (eds.). *Models, Simulations, and Representations*. Routledge.
- Gomes, C. y Faísca, P. (2019). *Protein Folding: An Introduction*. Springer.
- Goodwin, W. (2011). "Structure, function, and protein taxonomy". *Biology & Philosophy*, 26: 533–545. <https://doi.org/cft53x>
- Guttinger, S. (2018). "A Process Ontology for Macromolecular Biology". En D. J. Nicholson y J. Dupré (eds.). *Everything Flows: Towards a Processual Philosophy of Biology*. Oxford: Oxford University Press.
- Havstad, J. C. (2018). "Messy chemical kinds". *The British Journal for the Philosophy of Science*, 69(3): 719–743. <https://doi.org/gd7gk6>
- Hossenfelder, S. (2021). *Has Protein Folding Been Solved?* [Archivo de video]. Youtube. <https://youtu.be/yhJWAdZl-Ck>
- Hüttemann, A., y Love, A. C. (2011). "Aspects of reductive explanation in biological science: intrinsicity, fundamentality, and temporality". *British Journal for the Philosophy of Science*, 62(3): 519–549. <https://doi.org/bqt6s8>
- Ibarra, A. y Mormann, T. (2006). "Scientific Theories as Intervening Representations" *Theoria*, 21(55): 21-38.
- Jacob, C. (2002). "Philosophy and biochemistry: Research at the interface between chemistry and biology". *Foundations of Chemistry*, 4(2): 97-125. <https://doi.org/dw74mq>
- Jha, S. K., Ramanathan, A., Ewetz, R., Velasquez, A., Jha, A. (2021). "Protein Folding Neural Networks Are Not Robust". arXiv. <https://doi.org/h6nh>
- Jumper, J., Evans, R., Pritzel, A. et al. (2021). "Highly accurate protein structure prediction with AlphaFold". *Nature*, 596: 583–589. <https://doi.org/gk7nfp>

- Kaczanowski, S., Zielenkiewicz, P. (2010). "Why similar protein sequences encode similar three-dimensional structures?". *Theoretical Chemistry Accounts*, 125: 643–650. <https://doi.org/brn227>
- Kaiser, M. I. (2011). "The limits of reductionism in the life sciences". *History and Philosophy of the Life Sciences*, 33: 453–476.
- Kaiser, M. I. (2018). "Individuating Part-whole Relations in the Biological World". En O. Bueno, R. L. Chen y M. B. Fagan (eds.). *Individuation Across Experimental and Theoretical Sciences*. Oxford University Press.
- Kelley, L. A. (2017). "Fold Recognition". En D. J. Rigden (ed.) *From Protein Structure to Function with Bioinformatics*. Dordrecht: Springer. <https://doi.org/h6nj>
- Kohler, R. (1994). *Lords of the Fly: Drosophila genetics and the experimental life*. Chicago: Chicago University Press.
- Krebs, J., Goldstein, E., Kilpatrick, S. (2017). *Lewin's GENES XII*. Jones & Bartlett Learning
- Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., Moulton, J. (2019). "Critical assessment of methods of protein structure prediction (CASP)—Round XIII". *Proteins*, 87(12): 1011-1020. <https://doi.org/ggjr6z>
- LaFollette, H., y Shanks, N. (1993). "Animal models in biomedical research: Some epistemological worries". *Public Affairs Quarterly*, 7: 113–130.
- LaFollette, H., y Shanks, N. (1995). "Two models of models in biomedical research". *The Philosophical Quarterly*, 45: 141–160.
- Laskowski, R. A. (2011). "Protein Structure Databases". *Molecular Biotechnology*, 48: 183–198. <https://doi.org/fds976>
- Lee, J., Freddolino, P. L., Zhang, Y. (2017). "Ab Initio Protein Structure Prediction". En D, J. Rigden (ed.). *From Protein Structure to Function with Bioinformatics*. Dordrecht: Springer. <https://doi.org/h6nm>
- Levinthal, C. (1968). "Are there pathways for protein folding?" *J. Chim. Phys.*, 65: 44–45. <https://doi.org/gfzq6b>
- Levinthal, C. (1969). "How to fold graciously". En J. T. P Debrunner, E. Munck (eds.). *Mossbauer spectroscopy in biological systems: proceedings of a meeting held at Allerton House*. Monticello: University of Illinois Press.
- Lowe, D. (2022). "Fooling the Protein Folding Software". En *Science (Commentary)*. <https://www.science.org/content/blog-post/fooling-protein-folding-software>
- Lupas, A., Pereira, J., Alva, V., Merino, F., Coles, M. y Hartmann, M. D. (2021). "The breakthrough in protein structure prediction". *Biochemical Journal*, 478(10): 1885–1890. <https://doi.org/gkqqv5>

- Masrati, G., Landau, M., Ben-Tal, N., Lupas, A., Kosloff, M. y Kosinski, J. (2021). “Integrative Structural Biology in the Era of Accurate Structure Prediction”. *Journal of Molecular Biology*, 433(20): 167127. <https://doi.org/gk8xcc>
- McCoy A. J., Sammito M. D., Read R. J. (2022). “Implications of AlphaFold2 for crystallographic phasing by molecular replacement”. *Acta Crystallographica Section D*, 78(Pt.1): 1-13. <https://doi.org/h6np>
- Mertens, R. (2019). *The Construction of Analogy-Based Research Programs: The Lock-and-Key Analogy in 20th Century Biochemistry*. Transcript publishing.
- Mitchell, S. (2003). *Biological Complexity and Integrative Pluralism*. Cambridge: Cambridge University Press.
- Mitchell, S. D., Gronenborn, A. G. (2017). “After fifty years, why are protein X-ray crystallographers still in business?”. *The British Journal for the Philosophy of Science*, 68(3): 703–723. <https://doi.org/h6nq>
- Mithcell, S. (2019) “Perspectives, Representation, and Integration”. En M. Massimi y MCoy, C. *Understanding Perspectivism: Scientific Challenges and Methodological Prospects*. New York: Routledge.
- Morange, M. (2006). “The protein side of the central dogma: permanence and change”. *History and philosophy of the life sciences*, 28(4): 513-24.
- Moult, J., Pedersen J. T., Judson, R. y Fidelis, K. (1995). “A large-scale experiment to assess protein structure prediction methods”. *Proteins*, 23(3): ii-v. <https://doi.org/ddzct4>
- Mullard, A. (2021). “What does AlphaFold mean for drug discovery?”. *Nature reviews drug discovery*, 20(10): 725-727. <https://doi.org/gnn5nw>
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. Brace and World: Harcourt.
- Neal, J. (2021). *Protein Structure, Dynamics, and Function: A Philosophical Account of Representation and Explanation in Structural Biology*. Doctoral dissertation Dietrich School of Arts and Sciences.
- Nederbragt, H. (2003). “Strategies to improve the reliability of a theory: The experiment of bacterial invasion into cultured epithelial cells”. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 34: 593–614. <https://doi.org/dgcqnh>
- Obermayer, A. y Uversky V. (2021). “Solving Protein Structure with AI: Viva AlphaFold and Co.!” *Current Protein y Peptide Science 2021*, 22(12): 826 – 826. <https://doi.org/h6nt>
- Pearce, R. y Zhang, Y. (2021). “Toward the solution of the protein structure prediction problem”. *Journal of Biological Chemistry*, 297(1): 100870. <https://doi.org/h6nv>

- Perdigão, N. y Agostinho, R. (2019). "Dark Proteome Database: Studies on Dark Proteins". *High-Throughput*, 8(2): 8. <https://doi.org/h6nw>
- Perrakis, A. y Sixma, T. K. (2021). "AI revolutions in biology: The joys and perils of AlphaFold". *EMBO Reports*, 22: e54046. <https://doi.org/h6nx>
- Ramsey, J. L. (2007). "Calibrating and constructing models of protein folding". *Synthese*, 155: 307–320. <https://doi.org/dtv3rw>
- Rheinberger, H. J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford University Press.
- Rincon, P. (2021). AI breakthrough could spark medical revolution. *BBC News*. <https://www.bbc.com/news/science-environment-57929095>
- Rohl, C. A., Strauss, C. E, Misura, K. M. y Baker, D. (2004). "Protein structure prediction using Rosetta". *Methods in Enzymology*, 383: 66-93. <https://doi.org/btmt9z>
- Ross, J. L. (2016). "The Dark Matter of Biology". *Biophysical perspective*, 111(5): 909-916. <https://doi.org/f83v9x>
- Ross, L. N. (2017). "Causal selection and the pathway concept". *PhilSci Archive*. <http://philsci-archive.pitt.edu/14361/>
- Sample, I. (2020). DeepMind AI cracks 50-year-old problem of protein folding. *The Guardian*. <https://www.theguardian.com/technology/2020/nov/30/deepmind-ai-cracks-50-year-old-problem-of-biology-research>
- Santos, G., Vallejos, G. y Vecchi, D. (2020). "A relational-constructionist account of protein macrostructure and function". *Foundations of Chemistry*, 22: 363–382. <https://doi.org/h6n2>
- Sarkar, S. (1992). "Models of reduction and categories of reductionism". *Synthese*, 91: 167–194. <https://doi.org/crbspz>
- Schaarschmidt, J., Monastyrskyy, B., Kryshchuk, A. y Bonvin, A. M. J. J. (2018). "Assessment of contact predictions in CASP12: Co-evolution and deep learning coming of age". *Proteins*, 86(S1): 51-66. <https://doi.org/gc5db4>
- Senior, A.W., Evans, R., Jumper, J. *et al.* (2020) "Improved protein structure prediction using potentials from deep learning". *Nature*, 577: 706–710. <https://doi.org/ggjfcc>
- Shi, Y. (2014). "A glimpse of structural biology through X-ray crystallography". *Cell*, 159(5): 995-1014. <https://doi.org/gj9vnk>
- Sills, G. J. (2006). "The mechanisms of action of gabapentin and pregabalin". *Current Opinion in Pharmacology*, 6(1): 108-13. <https://doi.org/dw88vh>
- Slater, M. (2009). "Macromolecular pluralism". *Philosophy of Science*, 76(5): 851–863. <https://doi.org/gcpcgj>

- Soler, L., Trizio, E., Nickles, T. y Wimsatt, W. (2012). *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Springer
- Steel, D. (2008). *Across the boundaries. Extrapolation in biology and social science*. Oxford: Oxford University Press.
- Strand, R. (1999). "Towards a useful philosophy of biochemistry: Sketches and examples". *Foundations of Chemistry*, 1(3): 269-292.
- Strand, R., Fjelland, R. y Flatmark, T. (1996). "In vivo interpretation of in vitro effect studies with a detailed analysis of the method of in vitro transcription in isolated cell nuclei". *Acta Biotheoretica*, 44: 1–21. <https://doi.org/ctx2cn>
- Strasser, B.J. (2010) "Collecting, Comparing, and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequence and Structure, 1954–1965". *Journal of the History of Biology*, 43: 623–660. <https://doi.org/bq9n9t>
- Suárez, E. y Martínez S. (1998). "El problema del reduccionismo en biología: tendencias y debates actuales". En S. Martínez y A. Barahona. *Historia y explicación en Biología*. Fondo de Cultura Económica.
- Tahko, T. E. (2020). "Where Do You Get Your Protein? Or: Biochemical Realization". *The British Journal for the Philosophy of Science*, 71 (3): 799-825. <https://doi.org/ghqh22>
- Tanford, C. y Reynolds, J. A. (2003). *Nature's Robots: A History of Proteins*. New York; Oxford: Oxford University Press.
- Tee, S-H. (2019). "Model Organisms as Simulators: The Context of Cross-Species Research and Emergence". *Axiomathes*, 29(4): 363-382. <https://doi.org/gns9nw>
- Tobin, E. (2010). "Microstructuralism and macromolecules: the case of moonlighting proteins". *Foundations of Chemistry*, 12: 41–54. <https://doi.org/cspgp7>
- Tunyasuvunakool, K., Adler, J., Wu, Z. et al. (2021). "Highly accurate protein structure prediction for the human proteome". *Nature*, 596: 590–596. <https://doi.org/gk9kp7>
- Ureta, T. (2003). *En el filo de la navaja de Occam: reflexiones reduccionistas sobre algunos problemas del ser humano*. Editorial Universitaria.
- Vecchi, D. (2020). "DNA is not an ontologically distinctive developmental cause". *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 81: 101245. <https://doi.org/h6n3>
- Voet, D. y Voet, J. (2010). *Biochemistry (4th Edition)*. Wiley
- Wang, J., Luttrell IV, J., Zhang, N., Khan, S., Shi, N., Wang, M., Kang, J-Q., Wang, Z. y Xu, D. (2016). "Exploring Human Diseases and Biological Mechanisms by Protein Structure Prediction and Modeling". En: Shen, B., Tang, H., Jiang, X. (eds). *Advances in Experimental Medicine and Biology* (vol. 939). Singapore: Springer. <https://doi.org/h6n4>

- Waters, C. K. (2007). "Causes that Make a Difference". *The Journal of Philosophy*, 104(11): 551-579. <https://doi.org/f3f94g>
- Waters, C. K. (2008). "How Practical Know-How Contextualizes Theoretical Knowledge: Exporting Causal Knowledge from Laboratory to Nature". *Philosophy of Science*, 75(5): 707-719. <https://doi.org/bpkqs4>
- Waters, C. K. (2019). "An epistemology of scientific practices". *Philos. Sci.* 86(4): 585–611. <https://doi.org/gmjm8s>
- Weber, M. (2005). *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.
- Weber, M. (2006). "The Central Dogma as a Thesis of Causal Specificity". *History and Philosophy of the Life Sciences*, 28(4): 595-610.
- Wimsatt, W. (1981). "Robustness. Reliability and Overdetermination", En M. Brewer y B. Collins. (eds.), *Scientific Inquiry and the Social Sciences* (p. 124-163). San Francisco: Jossey-Bass.
- Wimsatt, W. C. (1974). "Complexity and organization". En K.F. Schaffner y R. S. Cohen (eds.). *Proceedings of the 1972 meeting of the Philosophy of Science Association* (p. 67–86.). Dordrecht: D. Reidel.
- Winther, R.G. (2011), "Part-whole science". *Synthese*, 178: 397–427. <https://doi.org/dj83ks>
- Wlodawer, A., Minor, W., Dauter, Z. y Jaskolski, M. (2013). "Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination". *The FEBS Journal*, 280(22): 5705-36. <https://doi.org/f5hr29>
- Woodward, J. (2010). "Causation in biology: stability, specificity, and the choice of levels of explanation". *Biology & Philosophy*, 25: 287–318. <https://doi.org/b2wbmr>
- Wooley, J. C. y Ye, Y. (2007). "A Historical Perspective and Overview of Protein Structure Prediction". En Xu, Y., Xu, D. y Liang, J. (eds.), *Computational Methods for Protein Structure Prediction and Modeling. BIOLOGICAL AND MEDICAL PHYSICS BIOMEDICAL ENGINEERING*. New York: Springer. <https://doi.org/fgmwhw>
- Xu, D. y Xu, Y. (2004). "Protein databases on the internet". *Current Protocols in Molecular Biology*. 68: 19.4.1-19.4.15. <https://doi.org/ffj9h9>
- Xu, D. y Zhang, Y. (2012) "Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field". *Proteins*, 80: 1715-1735. <https://doi.org/h6n7>
- Zuppone, R. (2010). "El argumento del regreso del experimentador y la replicación de experimentos". *Scientiae Studia*, 8(2): 243-271. <https://doi.org/fdkg2w>