



RIDAA

Repositorio Institucional
Digital de Acceso Abierto de la
Universidad Nacional de Quilmes



Universidad
Nacional
de Quilmes

Vélez Rueda, Ana Julia

Desarrollo de herramientas bioinformáticas para el estudio y predicción de la promiscuidad y multiplicidad de sustrato en enzimas



Esta obra está bajo una Licencia Creative Commons Argentina.

Atribución - No Comercial - Compartir Igual 2.5

<https://creativecommons.org/licenses/by-nc-sa/2.5/ar/>

Documento descargado de RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes de la Universidad Nacional de Quilmes

Cita recomendada:

Vélez Rueda, A. J. (2020). *Desarrollo de herramientas bioinformáticas para el estudio y predicción de la promiscuidad y multiplicidad de sustrato en enzimas. (Tesis de doctorado). Universidad Nacional de Quilmes, Bernal, Argentina. Disponible en RIDAA-UNQ Repositorio Institucional Digital de Acceso Abierto de la Universidad Nacional de Quilmes <http://ridaa.unq.edu.ar/handle/20.500.11807/2772>*

Puede encontrar éste y otros documentos en: <https://ridaa.unq.edu.ar>

Desarrollo de herramientas bioinformáticas para el estudio y predicción de la promiscuidad y multiplicidad de sustrato en enzimas

TESIS DOCTORAL

Ana Julia Vélez Rueda

anavelezrueda@gmail.com

Resumen

A pesar de su robustez a las mutaciones, las enzimas exhiben una notable adaptabilidad evolutiva. Nuevas enzimas han surgido a lo largo de toda la historia del planeta. Un concepto a menudo olvidado de los procesos evolutivos es que ocurren mientras mantienen la aptitud de los organismos al entorno y, en consecuencia, es una suposición razonable que los genes existentes se modificaron para generar nuevas estructuras y funciones proteicas, que pueden coexistir con las primeras. En el contexto de las enzimas, el término promiscuidad se aplica a las proteínas que poseen la capacidad de catalizar reacciones químicamente disímiles y se plantea que la promiscuidad podría funcionar como un reservorio en el metabolismo de los organismos, donde las múltiples capacidades químicas podrían permitir una mayor diversificación del metaboloma. En los últimos años se ha incrementado notablemente el uso de enzimas promiscuas para distintos usos biotecnológicos-industriales. Sin embargo, la predicción de la promiscuidad enzimática y la elección de una enzima para un determinado propósito biotecnológico resultan ser aún los principales obstáculos a vencer para su uso generalizado.

El objetivo general de esta tesis es comprender el origen y las bases estructurales que sustentan el comportamiento promiscuo. Dada la estrecha relación entre la diversidad conformacional proteica y su función, planteamos como hipótesis central de este proyecto que las distintas proteínas que catalizan una determinada reacción comparten estructuras funcionales, como superficies, cavidades, túneles, entornos del sitio activo, como residuos catalíticos y de anclaje del sustrato y de direccionamiento de sustratos y productos, para sustentar una determinada actividad catalítica. Por lo tanto, se propone en el presente trabajo de tesis conocer y determinar aquellas características o determinantes catalíticos que hacen a una proteína promiscua.

Para tal fin hemos desarrollado una base de datos de proteínas que muestran promiscuidad catalítica. Nuestra base de datos ProtMiscuity ofrece una colección de datos derivados experimentalmente, que permiten vincular información sobre las condiciones de reacción que promueven una dada actividad promiscua, y datos estructurales, secuenciales y funcionales de cada proteína promiscua, resultando de gran relevancia para la exploración de las propiedades promiscuas con miras a nuevas aplicaciones biotecnológicas e industriales. Gracias a nuestra recopilación de datos sobre proteínas promiscuas, pudimos además caracterizar la promiscuidad proteica tanto de forma general como particular, en el estudio de las propiedades promiscuas de la familia proteica de las albúminas séricas y de la promiscuidad en medios no acuosos.

Encontramos que la promiscuidad como fenómeno no puede ser descartado a priori en ningún caso y que sólo la indagación estructural y químico-mecanística podrán a futuro predecir dicho comportamiento. Nuestros resultados ofrecen, por lo tanto, nuevas herramientas y perspectivas que esperamos redunden en una mejor comprensión de los mecanismos subyacentes a la función proteica y al fenómeno de la promiscuidad proteica, así como también a la generación de herramientas bioinformáticas más precisas.

Palabras claves: promiscuidad proteica, diversidad conformacional.



TESIS DOCTORAL

“Desarrollo de herramientas bioinformáticas para el estudio y predicción de la promiscuidad y multiplicidad de sustrato en enzimas”

Lic. Ana Julia Velez Rueda

Director: Dr. Parisi, Gustavo

Co- Director: Dr. Iglesias, Luis Emilio

Lugar de trabajo: Grupo de Bioinformática Estructural, UNQ.

Dedicado a Maia, Nora y todas las mujeres del mundo.

Agradecimientos

Cuando el sol se asomó en el horizonte, por fin las ideas tomaban su curso en ese río tormentoso que supo ser su pensamiento. Ni un Dios podría haberlas ordenado con un sentido más lógico que la claridad de quién se sabe libre. Ana levantó la mirada de la hoja y finalmente cerró el cuaderno. No había ya más palabras en su lapicera que quisieran escapar a ese mar de tinta que yacía sobre el escritorio. Pasaron por su mente mil días con sus mil noches. Pasaron la mirada cálida de su madre y los dedos ásperos de Le acariciándole la mejilla, su luz le llenó el alma y la garganta se le hizo un nudo de emoción. Pasaron las sonrisas saltarinas de Mai que tan feliz la hacían y los silencios cómplices de Emi, que tanto se parecían a la hermandad. Paso la mirada amorosa y profunda de Fran, los cantos alegres de José y los experimentos culinarios de Omi. Pasó la infinita distancia que a veces la separaba del Rodri y las pocas palabras que bastaban para traerlo de vuelta. Pasaron los chistes del Juanpi y los mensajes intermitentes de Gra. Pasaron Isabel y Coca, Pilar, Rosario, Belen, Laura y Mariana las mujeres fuertes que siempre admiró. Pasaron esos amigos que nunca fallan y con quienes siempre quería compartir los mates. Una a una desfilaron las personas importantes de su vida ante la mirada perdida en recuerdos. Pasaron las noches en vela, los miedos infundados y los que tenían fundamento. Pasaron los momentos tristes, seguidos por las melodías y aromas de su gente amada. Y cuando por fin había dejado salir todas las lágrimas que reivindicaban las luchas mínimas de cada batalla cotidiana que la habían llevado hasta allí, se decidió a escribir los agradecimientos del trabajo que sólo resumía el esfuerzo de los últimos años:

Agradezco a...

A Gustavo que me abrió las puertas del SBG y me dio la posibilidad de desarrollarme profesionalmente, sin restricciones, ni prejuicios. Un gran maestro, que admiro profesional y humanamente. Me llevo de los años de doctorado bajo tu dirección la mirada crítica y la importancia de la pregunta; la visión curiosa del universo y la admiración hacia los misterios de la Biología, que siento guiarán mi camino profesional.

A mi codirector Luis en quien encontré una gran calidez y empatía, de quien he aprendido la minuciosidad y con quien he re-aprendido la química!

A Nicolás (Nico) que acompañó este proceso, muchas veces angustioso, con la mirada crítica de un colega y la complicidad de un amigo.

A Silvina que siempre sabe poner en acción su ayuda desinteresada y su pensamiento lógico cuando es necesario.

A Dieguito que con la mayor humildad y paciencia del mundo me ayudó a empezar a programar y acompaña hace años y con gran empeño el proyecto de extensión.

A mis amiguescompañecolegue (neologismo implementado por mí para la ocasión) Guille, Pablo, Pato, Mirna, Maia, Ale, Richard que supieron aguantar la parada y compartir de igual modo la ciencia, los mates, las penas y las risas... ¡Gracias!

A mis cumpas y colegas Julia, Ema, Tade, Víctor, Mati, Alex, Mar, Bety, Tincho, Tino, Agus, Elin, Pato, Nahue, Seba, Cris, Carli, Juli.

A todas las persona que creyeron en mí y ayudaron a hacer realidad una de mis más hermosa iniciativa en esta etapa: el proyecto “La Bioinformática va a la escuela”. Gracias por seguirme en esta locura, ser parte y acompañar!

A la UNQ cada una de las personas que hacen de esta universidad un hogar.

A la comunidad de Python Argentina por permitirme cruzar a su vereda y siempre dar una mano cuando hace falta.

Finalmente, pero no menos importante, a mis afectos:

A Le, shekh ma shieraki anni, porque esta tesis es tan mía como tuya, sin tu amor, consejo y apoyo incondicional no estaría hoy aquí. Gracias por acompañarme a soñar!

A Mai que además de ser el motivo para ser mejor cada día, ha acompañado este proceso con sus esperas pacientes y sus besos inflados. A Omi, Fran, Emi, Lili, Edu, Marian, Pupi, Clau, Rochi, Pili, Belu y Juan mi familia a quienes amo y me ayudan cada día a no olvidar lo que es realmente importante. A mis amigos/as del alma Lau, Chiro, Diame, Ale, la Roro, el Ro, el Juanpi, Tai.

Publicaciones

En el marco de la presente tesis doctoral

1. Velez Rueda, A.J., Palopoli, N., Zacarías, M., Sommese, L.M. and Parisi, G. ProtMiscuity: a database of promiscuous proteins. *Database: the Journal of Biological Databases and Curation* 2019.
2. Rueda, A.J.V., Monzon, A.M., Ardanaz, S.M., Iglesias, L.E. and Parisi, G. 2018. Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. *BMC Bioinformatics* 19(1), p. 27.
3. Ardanaz, S.M., Borucki, E.L., Velez Rueda, A.J., Parisi, G., Iribarren, A.M. and Iglesias, L.E. 2019. Bovine serum albumin-catalysed cross aldol condensation: Influence of ketone structure. *Process Biochemistry* 86, pp. 50-57.
4. Ardanaz, S.M., Velez Rueda, A.J., Parisi, G., Iribarren, A.M. and Iglesias, L.E. 2018. A mild procedure for enone preparation catalysed by bovine serum albumin in a green and easily available medium. *Catalysis Letters* 148(6), pp. 1750–1757.

Fuera del marco de la presente tesis doctoral

5. Echeverría, E., Rueda, A.J.V., Cabrera, M., et al. 2019. Identification of inhibitors of the RGS homology domain of GRK2 by docking-based virtual screening. *Life Sciences*, p. 116872.
6. Velez Rueda, A.J., Mistchenko, A.S. and Viegas, M. 2013. Phylogenetic and phylodynamic analyses of human metapneumovirus in Buenos Aires (Argentina) for a three-year period (2009-2011). *Plos One* 8(4), p. e63070.

Artículos de educación y divulgación científica

7. Velez Rueda, A.J. 2019. La bioinformática va a la escuela ...¡y llegó al Liceo! *Hilvanando Experiencias*.
8. Velez Rueda, A.J., Benítez, G.I., Marchetti, J., et al. 2019. Bioinformatics

calls the school: Use of smartphones to introduce Python for bioinformatics in high schools. *PLoS Computational Biology* 15(2), p. e1006473.

Índice

Abreviaturas	12
Introducción	13
1.1 Definición del estado nativo proteico	14
1.2. Diversidad conformacional de las proteínas	18
1.3. Proteínas como catalizadores biológicos	19
1.3.a. Catálisis enzimática	19
1.3.b. Modelos tradicionales de catálisis enzimática	22
1.3.c. Especificidad enzimática	25
1.4. Relación estructura función	27
1.4.a. Hipótesis de relación entre estructura y la capacidad catalítica	27
1.4.b. Relación entre diversidad conformacional y la función proteica	29
1.5. Promiscuidad proteica	31
1.5.a. Noción de promiscuidad	31
1.5.b. Tipos de promiscuidad	33
1.5.c. Evolución de la función proteica promiscua	36
1.5.d. Herramientas bioinformáticas para el estudio de la promiscuidad proteica	38
1.5.e. Aplicaciones biotecnológicas de la promiscuidad proteica	42
Objetivos generales y específicos	44
Objetivos	45
Objetivos específicos	46
Desarrollo de una base de datos para el estudio de la promiscuidad proteica: ProtMiscuity	47
2.1. ProtMiscuity: una base de datos de proteínas promiscuas	48

2.2. Aportes de la base de datos a la caracterización y comprensión de la promiscuidad proteica como fenómeno de interés biológico	49
2.3. Desarrollo de la base de datos ProtMiscuity	51
2.3.a. Reclutamiento bibliográfico por web scraping y minería de textos	51
2.3.b. Curado y extracción de datos de relevancia	52
2.4. Servidor Web ProtMiscuity	53
2.4.a. Descripción general de la base de datos	53
2.4.b. Implementación del Servidor Web ProtMiscuity	54
2.4.c. Vinculación con otras fuentes de información biológica	56
2.4.d. Ejemplo de búsqueda en ProtMiscuity	57
2.4.e. Estadística del contenido de la base de datos	60
2.4.f. Conclusiones destacadas	62
Catálisis y promiscuidad en medios no acuosos: estudio de la diversidad conformacional de proteínas en medio no acuoso	64
3.1. Catálisis y promiscuidad en medios no acuosos	66
3.2. Métodos específicos para el estudio de los cambios conformacionales en medios no acuosos	69
3.2.a. Reclutamiento y filtrado de estructuras cristalográficas para el análisis.	69
3.2.b. Métodos para la estimación de las diferencias estructurales entre conformeros acuosos y no acuosos	71
3.2.c.i Métodos para la estimación de la diversidad conformacional de túneles y cavidades	75
3.2.c.ii Métodos para la estimación de la diversidad conformacional de sitios activos	75
3.3. Análisis de las diferencias estructurales entre conformaciones acuosas y no acuosas	76

3.3.a. Comparación de la diversidad conformacional entre confórmers acuosos y no acuosos	76
3.3.b. Diversidad conformacional en estructuras funcionalmente relacionadas	82
3.3.c. Ejemplos de interés biológico	85
3.4. Conclusiones destacadas	88
Estudio de la actividad promiscua en un sistema específico:	
promiscuidad de albúminas séricas	90
4.1 Caracterización estructural de albúminas séricas	92
4.2. Promiscuidad catalítica en la albúmina sérica bovina	94
4.2.a. Métodos computacionales para el estudio de las bases estructurales del mecanismo de acción la seroalbúmina bovina en la reacción de condensación aldólica cruzada	96
4.2.b. Resultados de la caracterización estructural de la actividad de la seroalbúmina bovina como biocatalizador de la condensación aldólica	98
4.3 Marco teórico para el estudio de las bases evolutivas y estructurales de la reacción promiscua de condensación aldólica	102
4.3.a Métodos para el estudio de selección positiva de albúminas séricas y su relación con los patrones estructurales	104
4.3.b Análisis de las bases evolutivas y estructurales de la actividad promiscua de las albúminas séricas	107
4.4. Conclusiones destacadas	110
Discusión general y perspectivas a futuro	112
6. Apéndice	117
Tablas Suplementarias	117
Figuras Suplementarias	118
7. Bibliografía	123

Abreviaturas

A	Medio acuoso
AA	Medio acuoso - acuoso
AO	Medio acuoso - orgánico
AS	Albúminas séricas
ASA	Superficie de accesibilidad al solventes
ASB	Albúmina sérica bovina
ASC	Albúmina sérica de conejo
ASH	Albúmina sérica humana
CoDNaS	Conformational Diversity of the Native State, Base de datos de Diversidad Conformacional en proteínas
CSA	Catalytic site atlas, Atlas de sitios catalíticos
CSV	Valores separados por comas (comma separated values)
GDP	Guanosina difosfato
GTP	Guanosina trifosfato
IDP	Proteína intrínsecamente desordenada (Intrinsically Disordered Protein)
nr	No redundante
O	Medio orgánico (no acuoso)
OO	Medio orgánico - orgánico
PDB	Protein data bank, Banco de datos de proteínas
RMSD	Desviación media cuadrática de la posición de los carbonos alfa

Introducción

1.1 Definición del estado nativo proteico

Desde los trabajos realizados por Fischer en 1894 (Fischer, 1894) y Kossel en 1898 (Kossel, 1898), y pasando por las primeras reflexiones sobre el problema del plegamiento de proteínas hechos por Wu en 1931 (Wu, 1995), se empieza a intuir una posible relación entre la función proteica, su composición y la disposición espacial de los aminoácidos que la componen. Pauling y Mirsky hacen un avance fundamental en la comprensión de la estructura de las proteínas cuando, en su trabajo publicado en 1936, definen el estado nativo de las proteínas como un plegamiento o conformación característico, que le confiere a las proteínas su función, y cuya pérdida denominaron desnaturalización (Mirsky y Pauling, 1936).

Luego de décadas de experimentos, Anfinsen finalmente confirma que la secuencia aminoacídica contenía la información necesaria para el plegamiento de una proteína en una conformación biológicamente activa (Anfinsen et al., 1961). En dicho estudio, realizado sobre la ribonucleasa y extendido luego a miles de proteínas, Anfinsen y colaboradores postularon, entonces, que a partir de la estructura primaria de la proteína podría ser predicha la conformación o estructura terciaria biológicamente activa. Si bien Karush postula en 1950 la existencia de diferentes conformaciones con energías similares en equilibrio dinámico para explicar el estado nativo de las proteínas, en un trabajo realizado con la seroalbúmina bovina (Karush, 1950), no será hasta una década más tarde que estas ideas cobren relevancia para el entendimiento de la función proteica y la descripción del estado nativo.

La evidencia de flexibilidad funcionalmente relevante, no necesariamente en torno a una estructura única, comienza a construirse a partir del estudio realizado por Koshland (Koshland et al., 1958) sobre la fosfoglucomutasa. Koshland detectó de modo indirecto (por métodos químicos) los cambios conformacionales sufridos por la enzima tras la unión del ligando, introduciendo por primera vez la idea de cambio conformacional inducido. En 1965 Monod hace una descripción del estado nativo a través de un conjunto de conformaciones que constituyen el ensemble del estado nativo, considerando a

la estructura proteica con cierta flexibilidad inherente a su función biológica (Monod et al., 1965). Se establece, entonces, la idea actual del estado nativo proteico como un conjunto de conformaciones en equilibrio dinámico.

La actual descripción del estado nativo proteico involucra consideraciones termodinámicas, explicando la naturaleza y características del paisaje conformacional de una proteína como consecuencia de las barreras energéticas que permiten adoptar una u otra conformación (Bryngelson et al., 1995; Cruzeiro y Degreuve, 2017). El paisaje energético de las proteínas podría ser representado como un embudo de energía (entropía) decreciente, en el que la proteína en cuestión puede adoptar muchas vías de plegamiento posible y aún así alcanzar su estado nativo. El vértice de dicho embudo no es puntual, sino un fondo rugoso con múltiples conformaciones igualmente probables, interconvertibles entre sí con pequeños cambios en la estructura proteica, en un ensamble en equilibrio dinámico (Cieplak y Banavar, 2013) (Fig. 1.1). Existe entonces un gran número de estados y subestados conformacionales a lo largo de la vía de plegamiento de las proteínas, así como fluctuaciones termodinámicas en torno al estado nativo (Frauenfelder et al., 1991; Wei et al., 2016).

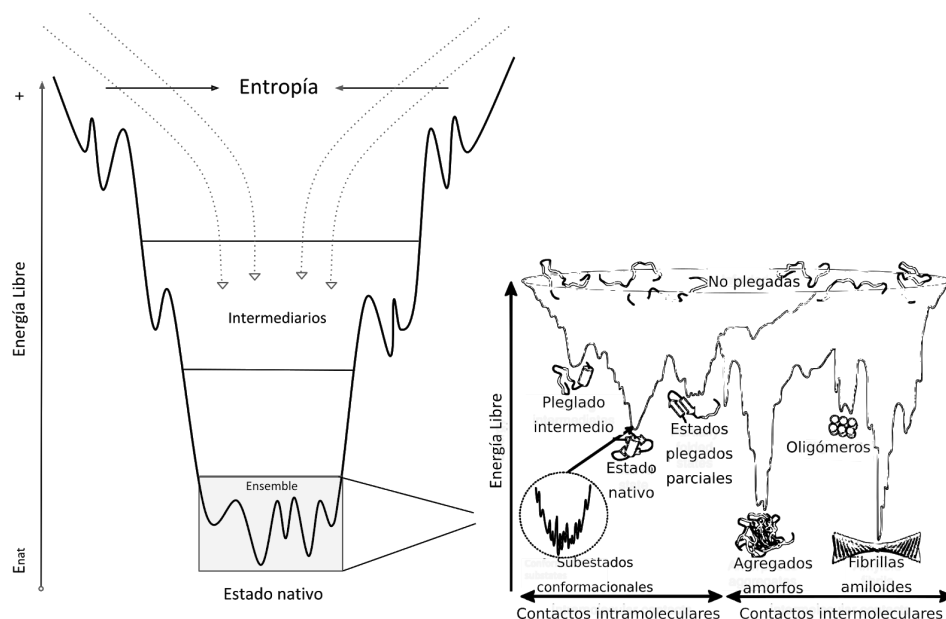


Fig. 1.1. Esquema del paisaje energético en el que se pliega una proteína a menudo se compara con un "embudo". El fondo del embudo rugoso describe el estado nativo de la

proteína formado por un conjunto de conformeros, de energía (entalpía) comparable, en equilibrio dinámico. En una ampliación del fondo del embudo se muestra la multitud de conformaciones "encauzadas" hacia el estado nativo a través de la formación de contactos intramoleculares, o hacia la formación de fibrillas amiloides a través de contactos intermoleculares. Dentro del mínimo del estado nativo, existe una multitud de movimientos posibles ("respiración" proteica). Figura adaptada de *"Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins"* (Raskatov y Teplow, 2017).

La función proteica deriva de su paisaje dinámico de energía. Estructuralmente, cualquier proteína puede describirse como un conjunto modular de conformaciones inducibles semi-estructuradas o completamente desordenadas, que conviven en equilibrio dinámico (Uversky, 2015). Alteraciones transitorias en el entorno proteico se traducen en paisajes energéticos modificados, con diferentes formas y distribuciones de dichas poblaciones de conformeros (Kumar et al., 2000). La función de una proteína y sus propiedades estarán determinadas, entonces, por la distribución de sus subestados conformacionales y las redistribuciones de las poblaciones en diferentes entornos (Zhuravlev et al., 2009).

Max Perutz mediante la determinación de la estructura tridimensional de la hemoglobina, pudo caracterizar experimentalmente dos estados conformacionales de la misma, el estado "oxi" y el "desoxi", siendo esta una de las primeras evidencias experimentales de la diversidad conformacional (Perutz y Mathews 1966). El autor describe en su trabajo que la hemoglobina está formada por cuatro cadenas polipeptídicas cada una de las cuales posee un grupo hemo unido. En su estado desoxi de la hemoglobina posee un mayor número de interacciones débiles y es, por tanto, más rígida. Estas uniones estabilizantes se oponen a la unión del oxígeno. Mientras que en su estado oxi (unido a O_2) las interacciones que estabilizan el tetrámero se reducen, favoreciendo de este modo la unión del oxígeno al grupo hemo (Eaton et al. 1991). Estas diferencias son debidas a que, tras la unión del O_2 al átomo de hierro del grupo hemo, se genera un cambio conformacional que resulta en el desplazamiento y aproximación de un residuo de histidina a la porfirina (Gelin et al. 1983). Este cambio conformacional se traduce en un aumento de la

afinidad de la hemoglobina al O_2 y un desplazamiento del equilibrio entre las conformaciones, que explica las propiedades funcionales de la hemoglobina, como el cooperativismo (Eaton et al. 2007; Eaton et al. 1991) (Ver Fig. 1.2).

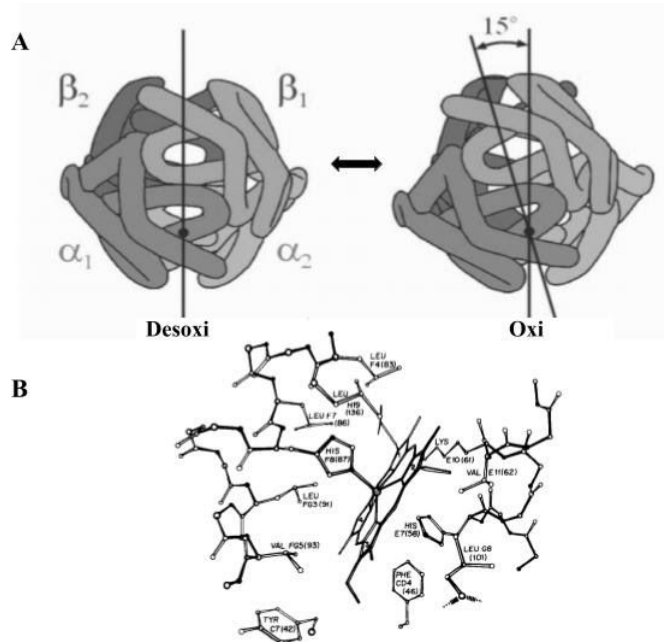


Fig. 1.2. A) Esquema de la transición entre las conformaciones oxi y desoxi de la hemoglobina humana. Modificado de Dickerson y Geis n.d., 1983 . B) Esquema de la relación de los residuos proteicos con el grupo hemo. Esquema adaptado de Gelin et al. 1983.

Los cambios en el entorno proteico pueden, en algunos casos, también llevar a cambios conformacionales que generen a una pérdida de la función proteica. Las proteínas amiloides, por ejemplo, en ciertas condiciones se transforman de sus conformaciones solubles a formas fibrilares agregadas altamente ordenadas, asociadas a múltiples enfermedades (Chiti y Dobson, 2006). Éstas son capaces de generar conjuntos de diversos estados conformacionales y de agregación, incluyendo oligómeros, protofibrillas y fibrillas (Ver Fig.1.1). De esta forma se requiere un equilibrio preciso entre la flexibilidad estructural y la estabilidad proteica para que una proteína sea funcional. Resulta, por tanto, fundamental analizar las bases estructurales para comprender la función biológica, la naturaleza de las interacciones proteína-ligando, la catálisis enzimática y el origen de nuevas actividades.

1.2. Diversidad conformacional de las proteínas

La cristalización es un proceso de purificación que conduce al aislamiento de moléculas conformacional y químicamente homogéneas. Sin embargo, como bien mencionáramos anteriormente, la estructura cristalina de una proteína en solución podría no ser la única conformación. Monod postula por primera vez la idea de la existencia de más de una conformación de la misma proteína en pre-equilibrio en su trabajo realizado en 1965, retomando el trabajo de Perutz (Monod et al., 1965; Perutz y Mathews, 1966). Este equilibrio no es estático, por el contrario y como dijimos anteriormente, la distribución relativa de los distintos conformeros puede variar por cambios en el entorno proteico (presencia de ligandos, fluctuaciones de pH, modificaciones postraduccionales, etc) (Kumar et al., 2000; Rueda et al., 2018). La magnitud de la diversidad conformacional observada en las proteínas puede abarcar desde la fluctuación en las cadenas laterales, hasta reordenamientos globales de la estructura terciaria (Böttcher et al., 2000; Cecchini et al., 2008; Engh and Bossemeyer, 2002; Koshland, 1998; Monzon et al., 2017; Ping et al., 2013; Plapp, 2010). Se han descrito, por ejemplo, casos donde un solo anticuerpo, adoptando cambios radicales en la conformación proteica, reconfigura su sitio de unión a ligandos, modificándose de este modo su especificidad (James et al., 2003). Pero también existen ejemplos en los que la función proteica es mediada por pequeños cambios conformacionales, como el del receptor de aspartato expresado en la superficie celular de las bacterias, que es capaz de proporcionar la señal de quimiotaxis al unirse al ligando, con solo pequeños cambios en su dominio extracelular (Milburn et al., 1991).

Existe una relación estrecha entre la función proteica y su flexibilidad, y un ejemplo que ilustra bien la relación dinámica estructural-función proteica son las proteínas termófilas (Feller, 2010). Los organismos extremófilos se han adaptado para vivir a temperaturas extremas, donde sus proteínas deben mantenerse estables para funcionar. Las proteínas de organismos termófilos son estables a altas temperaturas por un aumento en el número de interacciones débiles, que le permite compensar el aumento de energía térmica

del entorno (Wallon et al. 1997). Cuando estas enzimas son llevadas a temperaturas más bajas, mantienen su estructura tridimensional intacta, pero hay una pérdida de actividad por un exceso de rigidez (Collins et al., 2003); mostrando que, aún cuando la integridad estructural se mantenga, la restricción de la dinámica proteica es capaz de limitar su funcionalidad.

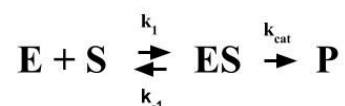
Las enzimas, también, hacen uso de un conjunto de conformaciones para reconocer sus sustratos y estabilizar los estados de transición (Maria-Solano et al., 2018). La estabilización diferencial de distintos conformeros puede producir cambios en el mecanismo de reacción de la enzima o permitir la entrada, tránsito y ubicación de sustratos de distintos tamaños (Kern et al., 2005; Wolf-Watz et al., 2004). Desde el punto de vista mecanístico, la diversidad conformacional provee una explicación para fenómenos relacionados con la diversidad funcional proteica (Monzon et al., 2017; Tsai et al., 1999). Por dicha razón, estudiaremos en el presente trabajo la catálisis y el fenómeno de la promiscuidad proteica contemplando la diversidad conformacional como un aspecto clave para su análisis.

1.3. Proteínas como catalizadores biológicos

1.3.a. Catálisis enzimática

Las enzimas son proteínas capaces de acelerar las velocidades de las reacciones químicas, logrando que éstas se completen en tiempos adecuados para la célula (Radzicka and Wolfenden, 1995). Las reacciones no catalizadas podrían llevar desde días a millones de años, dependiendo de la reacción. Si bien se han presentado distintas hipótesis que intentan racionalizar dicho fenómeno, con variantes que van desde la reducción de la energía de activación del estado de transición, a la desestabilización del intermediario de reacción (Benkovic and Hammes-Schiffer, 2003; Cannon et al., 1996; Menger and Nome, 2019), el mecanismo esencial por el cual las enzimas logran disminuir la velocidad aún continúa en discusión.

Aunque algunos mecanismos intrínsecos de la catálisis enzimática se desconocen, es posible caracterizar cinéticamente las enzimas gracias a los aportes hechos hace cien años por Maud Menten y Leonor Michaelis (Michaelis and Menten, 1913). En su trabajo demostraron que la velocidad de una reacción catalizada por enzimas es proporcional a la concentración inicial del sustrato (Deichmann et al., 2014). A partir de experimentos realizados con la beta-fructofuranosidasa, y retomando las ideas de Brown sobre la formación de un complejo enzima-sustrato para el planteamiento general de una reacción catalizada por una enzima (Brown, 1902), Menten y Michaelis desarrollaron la primera expresión matemática de la velocidad de reacción. Tomando como esquema básico una reacción enzimática planteada en dos etapas, se puede expresar:



donde la enzima (E) cataliza la transformación del sustrato (S) en producto (P); asumiendo el equilibrio de la reacción de formación del complejo enzima-sustrato (ES), con una constante cinética de formación del complejo ES k_1 y de ruptura k_{-1} . Michaelis-Menten observaron que la velocidad de reacción se mantiene constante en condiciones controladas de pH y temperatura. La relación de las constantes cinéticas de formación y ruptura del complejo ES representa la afinidad de la enzima por el sustrato y fue definida como K_s . Asumiendo que el equilibrio de formación del complejo ES se establece de forma rápida, los autores propusieron el siguiente modelo matemático:

$$v = V_{\max} [S] / K_s + [S]$$

el cual expresa la velocidad inicial de una reacción en términos de la concentración de sustrato, la velocidad inicial máxima y la constante de equilibrio K_s . La $V_{\max}$ corresponde a la máxima velocidad inicial alcanzable a concentraciones saturantes de sustrato.

Años más tarde, Briggs y Haldane plantean modificaciones a la ecuación desarrollada por Michaelis y Menten (Briggs and Haldane, 1925). Bajo la teoría del estado estacionario, según la cual la concentración del complejo enzima-sustrato es pequeña y constante a lo largo de la reacción, observan por tanto que la velocidad de formación del complejo ES resultaría igual a la de su disociación. Así, los autores plantean que la velocidad de reacción puede definirse como:

$$v = V_{\max} [S] / K_m [S]$$

definiendo la constante K_m (de Michaelis) como la relación de constantes cinéticas:

$$K_m = (k_{-1} + k_{cat}) / k_1$$

cuyo valor se corresponde con la concentración de S tal que la velocidad v será la mitad de la $V_{\max}$. La constante de catalítica (k_{cat}) o número de recambio representa el número de ciclos catalíticos por unidad de tiempo que la enzima es capaz de producir.

Definidas de este modo, tanto $V_{\max}$ como K_m no dependen de la concentración de S. El valor de K_m , en contraposición a lo encontrado por Michaelis Menten, es una relación compleja de constantes cinéticas que no dan una idea de la afinidad de la enzima por su sustrato. Aún cuando estos parámetros fueron considerados durante mucho tiempo como una medida de la especificidad relativa de una enzimática, es decir de la capacidad de una enzima para diferenciar entre dos sustratos posibles, la especificidad se encuentra mejor representada por el cociente de la constante de recambio (k_{cat}) y el K_m (es decir, especificidad = k_{cat}/K_m) (Brot and Bender, 1969; Eisinger et al., 2007).

Se ha demostrado que una enzima con múltiples conformaciones estables puede comportarse cinéticamente de acuerdo al modelo de Menten-Michaelis (Callender and Dyer, 2015; Min et al., 2006). Las perturbaciones que influyen en los movimientos dinámicos de las proteínas también influyen en las

estructuras intermedias de una manera que implica contribuciones dinámicas al estado de transición (Nashine et al., 2010). Incluso los pequeños movimientos pueden ser importantes para la catálisis (Gutteridge and Thornton, 2004). Desde el punto de vista evolutivo esto podría explicarse a través de pequeños cambios en regiones distantes de una proteína, que a través de cambios incrementales en la actividad y bajo presión evolutiva, pueden favorecer sistemáticamente variantes beneficiosas (selección positiva) y descartar cambios desfavorables (selección negativa) (Koshland, 1998).

Las combinaciones de diferentes residuos forman "unidades catalíticas" que facilitan la catálisis en el sitio activo de la enzima. La gran diversidad de reacciones catalizadas por enzimas resulta de la capacidad de la mayoría de estos grupos para funcionar como catalizadores nucleofílicos, electrófilos o ácidos-base generales (Bartlett et al., 2002). Estos residuos suelen mostrarse conservados y aún mutaciones puntuales podrían resultar perjudiciales para la catálisis (Cao et al., 2019). Se ha observado en los sitios activos el fenómeno de evolución convergente, es decir que la misma solución ha aparecido independientemente en diferentes ramas del árbol de la vida (Gherardini et al., 2007). Esto se entiende en un escenario en el que la evolución favorece la fijación de distintas variantes que aumentan la eficacia de reacciones similares. Sobre la base de lo antes expuesto, resulta fundamental remarcar que el estudio de los procesos catalíticos mediados por enzimas debe contemplar no sólo sus aspectos mecanísticos, sino también explorar la dinámica estructural de las proteínas y las modificaciones fisicoquímicas en el entorno catalítico que estos movimientos producen. Por tanto y como uno de los objetivos del presente trabajo, estudiaremos cómo aspectos dinámicos de las estructuras proteicas explican las nuevas funciones y actividades catalíticas.

1.3.b. Modelos tradicionales de catálisis enzimática

Muchas enzimas exhiben un alto grado de especificidad por las reacciones que catalizan y los sustratos involucrados en estas reacciones. Esta observación inspiró uno de los primeros modelos propuesto para explicar la catálisis

enzimática (Fischer, 1894). En el modelo de “llave-cerradura”, Emil Fischer propone que las enzimas adoptan una conformación estructural única, que le confiere especificidad para la unión a su sustrato. Este modelo hace una analogía entre las enzimas y un candado, cuya llave era su sustrato, mostrando la necesidad de complementariedad de tamaño, forma y la naturaleza química de un sustrato con el sitio activo de una enzima (Fig. 1.3 A). Si bien el modelo llave-cerradura explicaba muchos sistemas enzimáticos, en 1958 David Koshland plantea modificaciones que extendían lo planteado por Fisher, teniendo en cuenta la flexibilidad de las proteínas (Koshland, 1958).

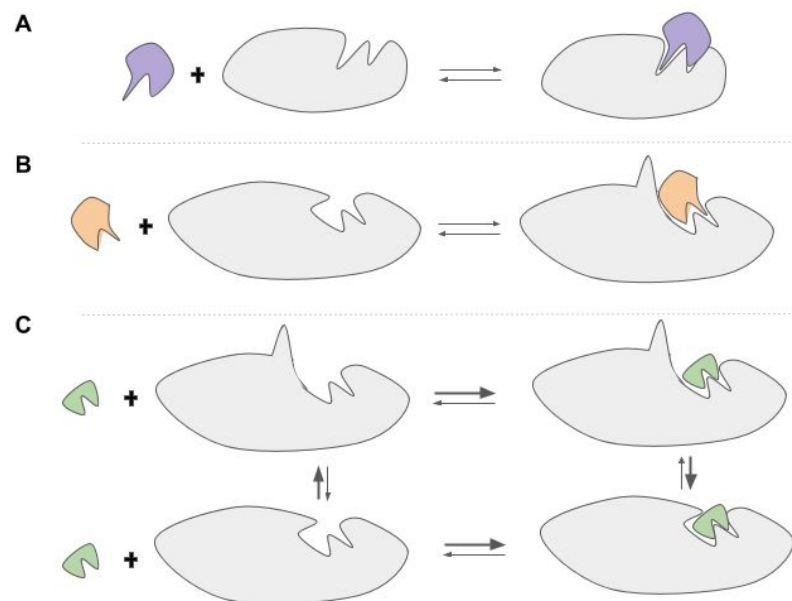


Fig. 1.3. A. Esquema representativo del modelo llave-cerradura, en el que se muestra la enzima (gris) con su sitio activo totalmente complementario al sustrato (violeta). B. Representación esquemática del modelo de ajuste inducido en el que la proteína (gris) sufre un cambio en su forma (estructura) para propiciar la unión del sustrato (naranja). C. Representación de una proteína en sus formas abiertas (arriba) y cerradas (abajo) junto con su ligando (verde). En el modelo de pre-equilibrio todos los conformeros se encuentran presentes en la solución antes de la interacción con el sustrato. El ligando se une al conformero de mayor afinidad, desplazando el equilibrio hacia esta conformación específica.

Koshland entiende que la descripción hecha por Fisher era insuficiente, en tanto plantea la necesidad de una enzima individual para cada enlace que se forma o se rompe, lo que requeriría una cantidad inconcebiblemente enorme de enzimas para todas las acciones catalíticas de las células. Koshland plantea entonces el modelo de “ajuste inducido”, en el que propone que una enzima es capaz de acomodarse o ajustarse estructuralmente al sustrato, este ajuste permite entonces la orientación precisa de los grupos catalíticos requerida para la acción enzimática. Asimismo, se explica desde este modelo la especificidad catalítica entendiendo que los cambios estructurales causados por la unión del sustrato a la proteína posicionarán los grupos catalíticos en la alineación adecuada solo si este sustrato es el adecuado (Fig. 1.3 B). Para demostrar la importancia de la flexibilidad proteica en el entendimiento de los procesos catalíticos, Koshland estudió la fosfoglucomutasa y detectó que dicha proteína en presencia y ausencia del ligando presentaba una diferencia en el número de tioles libres titulables (Koshland, 1995). Este resultado puso en evidencia ciertas inconsistencias en el modelo de “llave y cerradura” para explicar la relación estructura-función en proteínas. Al modelo de ajuste inducido se plantea uno alternativo con el trabajo de Monod en 1965 (Monod et al., 1965), que explica de forma más acabada la relación entre la estructura y la función proteica. En el modelo de pre-equilibrio de Monod, se asume un equilibrio entre múltiples estados conformacionales, preexistente a la unión del ligando. El ligando es capaz, entonces, de seleccionar y estabilizar una conformación proteica complementaria o específica que favorezca su unión (Fig. 1.3 C).

Estos modelos tradicionales de catálisis enzimática no hacen referencia alguna a la relación entre la estructura y la actividad de la proteína (ver sección 1.4). En general, son modelos del comportamiento cinético de las enzimas michaelianas y no-michaelianas. Con el paso del tiempo y el incremento de estructuras proteicas se hizo evidente que los sitios activos de las proteínas no siempre estaban expuestos al solvente. Estos pueden localizarse en cavidades internas, que se conectan con la superficie de la proteína por medio de túneles por los que deben transitar los sustratos y los productos para llegar y salir del sitio activo (Cojocaru et al., 2007; Coulombe et al., 2001; Lario et al., 2003). En

este proceso de reacción enzimática resulta de vital importancia el dinamismo de la estructura proteica. Fue justamente a partir de los trabajos de Koshland y Monod que se destaca la importancia de la dinámica conformacional para describir los mecanismos subyacentes a la función proteica, aspecto que será la base de nuestro análisis sobre la promiscuidad proteica.

1.3.c. Especificidad enzimática

La especificidad enzimática, a diferencia de la eficiencia catalítica, es un concepto relativo, ya que hace referencia a la capacidad de una enzima de discriminar entre dos o más posibles sustratos, en presencia de estos compuestos (Herschlag, 1988). Sin embargo, la capacidad enzimática de discernir entre sustratos no depende solamente de la afinidad del sustrato por la proteína, si no que suele medirse en términos de la relación entre el K_m y el K_{cat} (ver sección 1.3.a) (Cornish-Bowden and Cárdenas, 2010). Si bien se asume que las primeras enzimas eran generalistas y fueron especializándose con el curso de la evolución (Nam et al., 2012), la mayoría de las enzimas actuales operan lejos de los límites teóricos de especificidad y eficiencia.

En un estudio realizado sobre 2500 proteínas, se observó que la eficiencia catalítica de las enzimas se encuentra en promedio muy por debajo del límite de difusión ($K_{cat} \approx 10 \text{ s}^{-1}$) (Bar-Even et al., 2011). Las enzimas involucradas en el metabolismo central de los organismos parecen ser 30 veces más eficientes que las involucradas en el metabolismo secundario. Asimismo, se encontró una fuerte relación entre las propiedades fisicoquímicas de los sustratos y los parámetros cinéticos, mostrando que la especificidad enzimática se encuentra restringida evolutivamente por las propiedades de los sustratos a los que se encuentra expuesta. Estas no son, sin embargo, las únicas fuerzas que operan en la evolución de la selectividad enzimática.

Existen vastas evidencias que soportan la idea de una fuerte relación entre la función y la estructura proteica (Monzon et al., 2017; Orengo et al., 1999; Schmidt et al., 2003; Tsai et al., 1999a; Wei et al., 2016; Zea et al., 2016). Los movimientos de las proteínas están involucrados en el recambio enzimático y

son limitantes de la velocidad en el proceso catalítico general (Hammes-Schiffer and Benkovic, 2006; Kamerlin and Warshel, 2010). Además, en diferentes escalas temporales, los movimientos son requeridos para unir el sustrato, acompañarlo en su camino al sitio activo y permitir la liberación del producto (Eisenmesser et al., 2002; Nashine et al., 2010). Aún un cambio o mutación puntual puede producir una alteración del proceso catalítico, alterando el equilibrio de conformeros de la proteína y aumentando, por tanto, la barrera energética para la reacción (Thorpe and Brooks, 2003). Resulta lógico entonces pensar que las estructuras proteicas poseen restricciones para los cambios evolutivos en sus secuencias y que éstas se relacionan de algún modo con la especificidad enzimática.

Aunque la divergencia secuencial ha sido acompañada en mayor medida por la divergencia en la función proteica, dentro de muchas superfamilias de proteínas se ha observado que miembros distantes pueden compartir rasgos funcionales (Gerlt and Babbitt, 1998; Todd et al., 2001). Por ejemplo, en el caso de la familia de los citocromos P450, las características funcionales conservadas se deben a un residuo de cisteína fuertemente conservado entre sus miembros. Esta cisteína interactúa con el átomo de hierro coordinado en el grupo hemo a través de un enlace tiolato, conformando así una unidad ferril (Rittle and Green, 2010). La formación de esta unidad está asociada con la estabilización del intermediario de reacción altamente reactivo, que es esencial en el ciclo catalítico P450 para oxidar los enlaces C-H no activados.

Ahora bien, existen también ejemplos en los que reactividades similares se logran simplemente por características estructurales comunes en la región del sitio activo, aún sin compartir un ancestro común (James and Tawfik, 2001). Se han descrito patrones estructurales en las superficies de las cavidades que conforman los sitios activos de las distintas proteínas, que se encuentran relacionados con mecanismos catalíticos en particular (Dundas et al., 2011; Najmanovich et al., 2007; Tseng and Liang, 2007). En este sentido, los cambios conformacionales que impliquen un rearrreglo de estas cavidades podrán generar un cambio en la especificidad enzimática, mediada por las nuevas interacciones químicas accesibles a la enzima en dicha conformación. Un

ejemplo estudiado de este comportamiento es el caso de la transialidasa de *Trypanosoma cruzi*. Colman y Smith, revelaron cómo el sitio catalítico canónico de la enzima sufre un cambio conformacional que crea un bolsillo de unión adyacente para la lactosa, permitiendo de este modo que la actividad promiscua lactonasa pueda ser llevada a cabo (Colman and Smith, 2002). Asimismo, la flexibilidad del sitio es considerada un requisito necesario para la reducción de la barrera de energía libre y la aceleración de la reacción enzimática (Hammes-Schiffer y Benkovic, 2006). Por lo tanto, en los contextos en los que esta flexibilidad de las proteínas se ve restringida, pueden ser favorecidos cambios en la función proteica y la aparición de nuevas actividades. Por ejemplo, en los medios no acuosos, se observa el secuestro de conformaciones específicas con menor desorden y sitios activos más rígidos (Mattos y Ringe, 2001; Rueda et al., 2018). Estos cambios estructurales promueven la reducción de la capacidad catalítica respecto de un medio acuoso y posibilitan la aparición de reacciones promiscuas (Albanese y Gaggero, 2015; Busto et al., 2010; Klibanov, 1989, 2001).

En el presente trabajo se buscará comprender la complejidad de las actividades catalíticas de los distintos sistemas de forma general, profundizando en los complejos mecanismos subyacentes a las actividades promiscuas en particular.

1.4. Relación estructura función

1.4.a. Hipótesis de relación entre estructura y la capacidad catalítica

En las últimas décadas se han identificado un gran número de proteínas con actividades múltiples (Ben-David et al., 2013; Bornscheuer y Kazlauskas, 2004; Miao et al., 2015; Pabis y Kamerlin, 2016). Aun cuando las actividades originales (o canónicas) y promiscuas distan en su química, en la mayoría de los casos ambas actividades involucran una misma configuración del sitio activo. En efecto, es posible encontrar la misma actividad en proteínas

completamente diferentes, que comparten características similares del sitio activo (James y Tawfik, 2001). Estudios sobre sitios activos enzimáticos indican que existe un predominio de un pequeño conjunto de aminoácidos involucrados en la catálisis y que determinan el entorno general del sitio activo (Bartlett et al., 2002). Los residuos catalíticos en general muestran una exposición muy limitada al solvente y su entorno tridimensional local se encuentra más conservado que el resto de la proteína. Existen evidencias de que la evolución opera seleccionando los movimientos dinámicos funcionales o restringiendo los movimientos estructurales que fueran en detrimento de la función proteica (Granata et al., 2017; Marchetti et al., 2019).

Como hemos dicho anteriormente, comprender la relación estructura-función implica un conocimiento integral tanto del mecanismo catalítico como de las características estructurales que hacen posible esa reacción. En un estudio acerca de los dominios PAS ampliamente extendidos en arqueas, bacterias y eucariotas, en que se analiza la evolución funcional de la superfamilia, se observó una conexión directa entre la estructura y la función de dichas proteínas (Dong et al., 2018). En dicho trabajo, las proteínas con la misma función mostraron características secuenciales que permitieron agruparlas, ya que poseían no sólo patrones vibratorios diferenciales y particulares de sus grupos funcionales, sino también residuos conservados (secuencial y estructuralmente) en dichos grupos funcionales cuya fluctuación podía ser correlacionada con la función biológica correspondiente (Fig. 1.4).

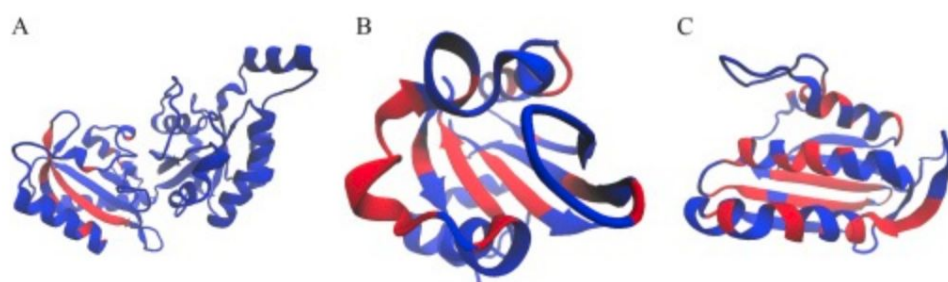


Fig. 1.4. Los residuos conservados en la estructura secundaria de las proteínas mostrados en rojo. A, estructura 2WKP representante del Grupo 1 cuya función es la unión de nucleótidos; B, estructura 1F9I representante del Grupo 2 cuya función se relaciona con la

actividad de fotorreceptores; y C, estructura 3AOT representante del Grupo 3 con actividad transferasa, respectivamente. Figura extraída del trabajo: *Combining protein sequence, structure, and dynamics: A novel approach for functional evolution analysis of PAS domain superfamily*. Dong, Z., Zhou, H., et. al. *Protein Sci.* 27, 421–430. DOI: 10.1002/pro.3329.

En esta tesis tendremos en cuenta la evidencia disponible sobre la relación entre la función y estructura proteica, como base para nuestro estudio sobre proteínas y los determinantes estructurales que subyacen a sus actividades promiscuas.

1.4.b. Relación entre diversidad conformacional y la función proteica

El concepto de diversidad conformacional propone repensar el paradigma genotipo-fenotipo y por tanto fenotipo-función (Nussinov et al., 2019). Un gen codifica no sólo para una, sino para un conjunto de conformaciones proteicas; y una proteína que adopta diferentes conformaciones, en principio, podría presentar diferentes funciones (Ma et al., 2002; Sikosek et al., 2016). Si bien clásicamente se consideraba que existe una relación lineal entre el fenotipo y genotipo, existen evidencias que la diversidad conformacional de proteínas se correlaciona con la tasa evolutiva (Javier Zea et al., 2013) y que las mutaciones alteran las probabilidades relativas de las conformaciones (Jolly et al., 2018). De ahí que la composición genética se exprese a través de un conjunto de conformaciones responsable de la función proteica, dado que una actividad específica es ejecutada por una conformación específica (Christopoulos, 2014), y un cambio en las probabilidades relativas de los distintos conformeros puede favorecer una actividad particular y conducir así a un cambio fenotípico.

Asimismo, se ha demostrado que el equilibrio de conformaciones puede ser desplazado también por la unión a sustratos o ligandos (Kahraman et al., 2007). En contextos celulares, estos hallazgos podrían explicar la multiplicidad de funciones que presentan algunas proteínas y algunas enzimas capaces de catalizar promiscuamente otras reacciones que no son las habituales (James et al., 2003; Kahraman et al., 2007). Un ejemplo estudiado de multi especificidad

mediada por el desplazamiento del equilibrio conformacional se describe en el trabajo de James et al. (James et al., 2003), quienes describen que un solo anticuerpo es capaz de adoptar diferentes conformaciones de su sitio de unión y, por lo tanto, de unir antígenos no relacionados.

Se podría, entonces, describir la estructura de una proteína como la distribución de concavidades o bolsillos de su superficie a lo largo de sus distintas conformaciones. Entre las estructuras posibles que conforman la anatomía de una proteína encontramos túneles o canales, cavidades, bolsillos, poros y bolsillos inaccesibles al solvente (voids); y son estas las que delimitan los ambientes donde pueden ser llevadas a cabo las funciones biológicas (Zhou et al., 1998). Es la conexión entre dichas estructuras la que determina el tránsito diferencial de las distintas moléculas del medio (Kingsley y Lill, 2015; McGovern y Shoichet, 2003), ya que los aminoácidos que las delimitan pueden variar en organización y exposición como consecuencia del propio equilibrio dinámico entre conformaciones de la proteína, e influir por tanto en su funcionalidad (Marques et al., 2018; Monzon et al., 2017; Murga et al., 2019, 2008).

Esta unión diferencial a ligandos y el tránsito de los mismos desde la superficie de la proteína al sitio activo o a los sitios de unión se ha estudiado con gran profundidad en la búsqueda de evidencias estructurales de la relación entre diversidad conformacional y la función especificidad proteica (Biedermannová et al., 2012; Otyepka et al., 2012; Schleinkofer et al., 2005). Un ejemplo bien estudiado de esta relación entre las subestructuras proteicas y la función proteica son las enzimas del citocromo P450 de mamíferos (Yu et al., 2013). La comparación de las estructuras cristalinas de los citocromos de mamíferos en estados cerrados, abiertos y parcialmente cerrados reveló que los cambios conformacionales de las partes flexibles de la proteína generalmente acompañan la entrada y salida de los ligandos. Asimismo, se observó que algunos citocromos adoptan inclusive diferentes conformaciones cuando se unen a diferentes ligandos, lo que sugiere que las rutas de entrada y salida del ligando pueden diferir según las propiedades del ligando.

De los diversos ejemplos antes mencionados se deduce el rol que juega la diversidad conformacional de las proteínas en los mecanismos de diversificación funcional y especificidad proteica. Desde el punto de vista mecanístico, las diferencias conformacionales pueden favorecer la estabilización diferencial de distintos estados de transición en el sitio activo o producir cambios en el mecanismo de reacción de una misma enzima o simplemente permitir la entrada, tránsito y ubicación de sustratos o ligandos de distintos tamaños (Biedermannová et al., 2012; Kingsley y Lill, 2015; Ma et al., 2002; Miller y Raines, 2004; Zhou et al., 1998). Resulta obvio, por tanto, la relevancia de dichas consideraciones en el análisis y estudio de la función proteica y sus posibles aplicaciones. Dichas consideraciones que serán tenidas en cuenta en el presente trabajo para el análisis de los mecanismos evolutivos y estructurales que expliquen la promiscuidad proteica.

1.5. Promiscuidad proteica

1.5.a. Noción de promiscuidad

La forma en que las enzimas han evolucionado a su forma actual en los organismos parece estar vinculada a cómo surgieron las vías metabólicas y se configuraron las redes metabólicas existentes (Carbonell et al., 2011). Se ha mostrado que existe un núcleo de proteínas altamente conservadas involucradas en las vías metabólicas esenciales para la vida de todos los organismos, complementadas con un conjunto de proteínas menos conservadas y asociadas con innovaciones específicas de taxones (Peregrín-Alvarez et al., 2009). Si bien, el término promiscuidad es planteado frecuentemente desde una perspectiva bioquímica como la amplitud de interacciones que puede presentar una proteína, no necesariamente relevantes en sentido fisiológico (Copley, 2003), todos los genomas codifican un gran número de enzimas que no necesariamente participan en rutas metabólicas y que pueden ser clasificadas como promiscuas. Tal es el caso de la seroalbúmina y la hemoglobina humanas, dos proteínas de transporte con

actividades promiscuas demostradas, y cuyas implicancias biológicas aún se desconocen (Jia et al., 2017; Masson et al., 2007). Por tanto y en el contexto de las enzimas, el término promiscuidad se aplica tanto a las proteínas que poseen la capacidad de catalizar reacciones químicamente disímiles, como a aquellas capaces de aceptar como sustratos a una diversidad de compuestos (Hult y Berglund, 2007).

Se plantea que la promiscuidad podría funcionar como un reservorio de múltiples capacidades catalíticas en los organismos (Danchin, 2009), que podrían compensar las deficiencias de las enzimas esenciales o permitir una mayor diversificación del metaboloma (Desai y Miller, 2008; James et al., 2003; Kuznetsova et al., 2006). Si bien, se suele asumir que las actividades promiscuas son menos eficientes en términos cinéticos que las actividades canónicas, las magnitudes de las actividades promiscuas pueden ser muy variables, desde valores apenas detectables, a varios órdenes de magnitud por encima de la actividad canónica (Khersonsky et al., 2006). En este sentido, resulta un ejemplo interesante la extensa familia de las fosfatasas *dehalogenase-like* (HAD) de *E.coli*, en cuyos miembros fueron detectadas actividades secundarias a las tradicionalmente descritas y sin una aparente funcionalidad fisiológica (Kuznetsova et al., 2006). Kuznetsova *et. al.* estudiaron el rol fisiológico y la especificidad enzimática de 23 HADs de *E.Coli*. Los resultados de dicho trabajo indicaron que 20 de las HADs presentaban actividad fosfatasa, con valores fisiológicamente relevantes para una gran variedad de los metabolitos ensayados. Por ejemplo, la proteína YniC mostró actividad 2-deoxiglucosa-6-fosfatasa además de su actividad canónica 1-fosfatasa. Cada uno de los azúcares producto de dicha reacción desempeñan un papel en el metabolismo primario del organismo, sugiriendo que esta enzima podría ser importante en condiciones limitadas de fosfato, al aumentar su disponibilidad para otros procesos. Este trabajo, entre otros, plantea no sólo la posibilidad de funciones canónicas y promiscuas coexistiendo fisiológicamente, sino también el posible papel de algunas enzimas para mejorar la función de la red metabólica general, para lo que pueden requerir una especificidad de sustrato amplia o actividades catalíticas

alternativas (Desai y Miller, 2008; Patrick y Matsumura, 2008; Patrick et al., 2007).

Existen muchos y muy diversos sistemas como el descrito que demuestran la complejidad de la funcionalidad proteica. A lo largo del presente trabajo intentaremos comprender la naturaleza de la promiscuidad proteica, su diversidad y su funcionalidad biológica, para lo que empezaremos por buscar una definición conceptual del fenómeno.

1.5.b. Tipos de promiscuidad

A pesar que la promiscuidad de las proteínas ha sido ampliamente estudiada en las últimas décadas, la definición del término aún está en discusión (Copley, 2015). Se ha utilizado la palabra promiscuidad para describir varios fenómenos desde múltiples perspectivas; y se han propuesto diferentes esquemas de clasificación, que involucran desde la variedad de funciones fisiológicas que puede presentar una misma proteína hasta la capacidad de unir sustratos alternativos (Copley, 2017; Hult y Berglund, 2007; López-Iglesias y Gotor-Fernández, 2015) (Ver Tabla 1.1).

La capacidad de unión a múltiples sustratos en varias proteínas es un rasgo derivado de la presión selectiva durante la evolución para cumplir una tarea biológica determinada (Nath y Atkins, 2008). Existen numerosas enzimas que durante la evolución han ganado una gran plasticidad en términos de los sustratos sobre los que son capaces de catalizar una misma reacción; por ejemplo, el citocromo P450 es capaz de oxidar tanto moléculas endógenas como exógenas, que van desde ácidos grasos de distinta longitud hasta compuestos aromáticos y tioles (Guengerich, 1991; Meunier et al., 2004; Porter y Coon, 1991; Wen y Fulco, 1987). Esta característica suele ser denominada multiplicidad de sustrato y se diferencia de la promiscuidad catalítica, definida por Khersonsky y Tawfik, como la capacidad de una enzima para catalizar una reacción diferente a aquella para la cual la proteína ha evolucionado (Khersonsky y Tawfik, 2010). Si bien esta definición aporta una perspectiva

interesante desde un punto de vista químico y funcional, confiere también una gran complejidad al análisis de casos particulares.

Desde una perspectiva más pragmática, y que enriquece este punto de vista químico-funcional, se puede describir a la promiscuidad catalítica como la capacidad de una enzima para catalizar una reacción secundaria en el mismo sitio activo donde ocurre su actividad primaria. Esta reacción secundaria debe tener un mecanismo químico diferente, generalmente con otro nombre, e involucrando la formación y/o rotura de enlaces distintos (Kazlauskas, 2005). De manera similar, la promiscuidad de sustrato también se ha utilizado para describir la capacidad de la enzima para realizar reacciones químicas comparables utilizando diferentes sustratos (Nath y Atkins, 2008). Bajo esta perspectiva, las actividades promiscuas catalíticas y de sustrato generalmente involucran sustratos y productos que carecen de relevancia fisiológica o bioquímica para los organismos (Copley, 2017; López-Iglesias y Gotor-Fernández, 2015). En consecuencia, debe evitarse el uso de "promiscuidad" para describir proteínas y enzimas con una amplia especificidad para los ligandos biológicamente relevantes.

Las proteínas que cumplen más de una función fisiológica, a menudo reguladora o estructural en lugar de enzimática, y en diferentes momentos o compartimentos celulares, pueden clasificarse de manera más apropiada como proteínas *moonlighting*. Brevemente, este fenómeno se encuentra relacionado estrechamente a la bioquímica de los organismos, y hace referencia a la aquellas proteínas con más de una función metabólica o fisiológica (Jeffery, 2003) (Tabla 1.1). Un ejemplo descrito recientemente es el caso de la proteína Clf1p de *S. cerevisiae*, asociada tradicionalmente al "splicing" o empalme del pre-ARNm, para la que se encontraron evidencias de su rol directo en el inicio de la replicación del ADN (Zhu et al., 2002). Si bien el moonlighting juega un rol importante en la comunicación y cooperación entre las diferentes vías metabólicas dentro de una célula, así como entre diferentes tipos de células dentro de un organismo, las bases evolutivas y estructurales que explican este fenómeno y sus diferencias con la promiscuidad catalítica aún deben ser caracterizadas en profundidad.

Además de sus muchas definiciones y perspectivas, la promiscuidad dista mucho de ser un fenómeno poco común, como se pensaba hasta hace unas décadas. Cada vez está más presente en los protocolos de desarrollo de fármacos o de síntesis de compuestos orgánicos, por citar algunos ejemplos (Bornscheuer y Kazlauskas, 2004; Nobeli et al., 2009). Se han descrito múltiples casos de promiscuidad catalítica, que involucran diferentes mecanismos (Ver Tabla 1.1) (Atkins, 2015; Babbie et al., 2010). Por ejemplo, las metaloenzimas son bien conocidas por ampliar su repertorio catalítico mediante el intercambio de cofactores (Ben-David et al., 2013; Nianios et al., 2015; Zalatan et al., 2008), lo que resulta de gran importancia dada la variabilidad de la concentración iónica tanto en la célula en general como en los distintos compartimentos celulares en particular, capaz de ejercer la presión química necesaria para las distintas funcionalidades bioquímicas (Dupont et al., 2010; Osman et al., 2019; Outten y O'Halloran, 2001). Muchas proteínas no enzimáticas fueron descritas como promiscuas, capaces de catalizar más de una reacción compleja. Tal es el caso de las ya mencionadas albúminas séricas, tanto humana como bovina, que mostraron capacidades catalíticas muy diversas (Ardanaz et al., 2018; di Masi et al., 2011) que van desde reacciones de eliminación de Kemp hasta condensaciones aldólicas cruzadas.

Proteína	Tipo de fenómeno	Actividad nativa	Actividad promiscua	Referencias
Lisozima C	Promiscuidad catalítica	Hidrólisis de enlaces beta (1-> 4) entre ácido N-acetilmurámico y N-acetil-D-glucosamina	Reacción de aza-Diels–Alder	He et al. 2012
D-Alosa kinasa	Promiscuidad de sustrato	Fosforilación de D-alosa	Fosforilación de D-alosa a D-glucosa	Miller y Raines 2004
Glutación transferasas	Moonlighting	Amplia especificidad de sustrato GST de clase A	Acoplamiento de glutación con varios ligandos	Hou et al. 2007
Pepsina	Promiscuidad catalítica	Endopeptidasa	Condensación aldólica	Reetz et al. 2007;

			Reacción de Morita-Baylis-Hillman	He et al. 2016; Xiang et al. 2016
			Adición de Michael	
Glutamato sintasa dependiente de ferredoxina (Fd-GOGAT)	Moonlighting	Biosíntesis de L-glutamato	Biosíntesis de sulfolípidos (dominio de FMN de unión)	Shimojima et al. 2005
Citocromo P450 monooxigenasa	Multiplicidad de sustrato	Actividad monooxigenasa sobre moléculas orgánicas	Actividad monooxigenasa sobre moléculas orgánicas	Porter y Coon 1991

Tabla 1.1. Ejemplos de procesos relacionados con distintas definiciones de promiscuidad, *moonlighting* y multiplicidad de sustrato caracterizados en bibliografía.

Éstos y muchos otros casos interesantes muestran la complejidad de la funcionalidad de la proteína y la necesidad de recopilar información que pueda ayudar a comprender el origen evolutivo y los mecanismos subyacentes a la promiscuidad, así como el desarrollo de nuevas herramientas para la predicción de actividades no canónicas (Copley, 2015; Nath y Atkins, 2008; Newton et al., 2018; Stainbrook et al., 2017).

En este punto es importante aclarar que durante el presente trabajo, nuestra definición operativa de promiscuidad es la recomendada por Kazlauskas, que incorpora una mirada pragmática de clasificación químico-funcional y nos permite, por tanto, una clasificación simple de los casos de estudio (Kazlauskas, 2005).

1.5.c. Evolución de la función proteica promiscua

A pesar de su robustez a las mutaciones, las enzimas exhiben una notable adaptabilidad evolutiva. Un concepto a menudo olvidado de los procesos evolutivos es que ocurren mientras la aptitud del organismo al medio se mantiene. Ohno propuso que las duplicaciones de genes propician la

acumulación de mutaciones requeridas para la aparición de nuevas funciones (Ohno, 1970). Y en este sentido, Hughes propone que la coexistencia de ambas copias de los genes permitiría que cada uno de ellos se especialice en una de las funciones del gen ancestral (Hughes, 1994). A la luz de estas teorías, se plantea que las funciones promiscuas de algunas enzimas podrían proporcionar puntos de partida para nuevas enzimas o especificidades enzimáticas (Bornscheuer y Kazlauskas, 2004; Jensen, 1976). Se propone, entonces, que en la ruta evolutiva más frecuente los miembros más antiguos suelen ser de especificidad amplia o alta promiscuidad y sobre ellos opera la evolución a nivel molecular, retocando o haciendo ajustes. De esta forma, una familia de proteínas se diversifica produciendo las nuevas familias y superfamilias de enzimas, funcionalmente diversas, que vemos hoy (Elias y Tawfik, 2012; van Loo et al., 2010).

Sin embargo, la trayectoria evolutiva de una proteína no siempre es una trayectoria lineal que culmina en la nueva función; ya que los intermedios evolutivos pueden tener nuevas propiedades además de su especificidad relajada, como por ejemplo una estabilidad alterada, y de este modo la trayectoria evolutiva se vería afectada para evitar compensaciones negativas (Dellus-Gur et al., 2013). Es importante recalcar, además, que estos intermediarios promiscuos resultan más fácilmente optimizables que las proteínas específicas y con menos restricciones para el cambio (Nam et al., 2012), lo cual resultaría un rasgo ventajoso para garantizar la adaptabilidad de la célula u organismo manteniendo su viabilidad (Dellus-Gur et al., 2013; Keeling et al., 2008; Wagner, 2011). Además, las nuevas especificidades proteicas deben, de algún modo, proveer de ventajas evolutivas para el organismo que favorezcan su permanencia durante la evolución (Siddiq et al., 2017). En algunos casos la presión de selección es fuerte y las nuevas funciones enzimáticas pueden evolucionar en cuestión de pocas décadas, como ocurriera con las enzimas degradantes de sustancias químicas sintéticas que aparecieran por primera vez en este planeta durante el siglo XX (Janssen et al., 2005; Raushel y Holden, 2000). Tal es el caso de las deshalogenasas bacterianas, que catalizan la degradación de compuestos con enlaces

carbono-halógeno, un paso clave en las vías de mineralización aeróbica de compuestos halogenados que se producen como contaminantes ambientales (Janssen et al., 2005). Se han encontrado en diversos organismos deshalogenasas secuencialmente similares junto con una gran distribución de enzimas deshalogenasas que no parecieran pertenecer a ninguna ruta metabólica conocida. Estos hallazgos podrían mostrar no sólo una respuesta de la naturaleza a la presión ambiental, sino una probable adaptación en curso, con el reclutamiento de secuencias silenciosas en rutas catabólicas funcionales, que contribuyan a la limpieza del ambiente de estos compuestos tóxicos: un ejemplo de cómo operan las fuerzas purificadoras sobre un pool preexistente de variantes génicas.

El patrón de conservación de la función en la evolución parece depender de la superfamilia proteica (Chiang et al., 2008; Giuseppe et al., 2012) ya que algunas parecen haber divergido funcionalmente, mientras otras exhiben menor diversidad funcional. La explicación de dichas diferencias en las trayectorias evolutivas aún se desconoce, en parte debido a la falta de información o sistematización de la información, y en parte debido a la inespecificidad de los métodos de asignación de función (Devos y Valencia, 2000). En este sentido, resulta evidente la necesidad de nuevas herramientas que aporten a la caracterización y sistematización, para lograr una mayor comprensión de la funcionalización de las proteínas y de la evolución de la promiscuidad proteica

1.5.d. Herramientas bioinformáticas para el estudio de la promiscuidad proteica

Como mencionamos en la sección 1.5.b, la promiscuidad catalítica de una proteína o enzima se refiere a la capacidad de un sitio activo para catalizar más de una transformación química (Kazlauskas, 2005). Kazlauskas propone que las transformaciones son diferentes si las actividades canónica (aquella descrita originalmente para la proteína) y promiscua involucran distintos tipos

de enlaces o mecanismos de reacción. Esta perspectiva mecanística podría ser utilizada para el estudio y predicción de actividades promiscuas.

En este sentido, los autores Khersonsky y Tawfik proponen un posible método de clasificación y evaluación del grado de promiscuidad de enzimas, mediante la comparación de los números de la Comisión de Enzimas (EC) para las actividades canónicas y promiscuas (Khersonsky et al., 2006). Este código es un esquema de clasificación de enzimas que utiliza 4 números separados por puntos para codificar el tipo de reacción y las moléculas involucradas en la actividad catalítica de dicha enzima (Bairoch, 2000). El primer número determina la subclase de enzima (liasa, isomerasa, transferasa, etc) o cambio químico global que ésta cataliza; el segundo refiere al tipo de enlace que sobre el que actúa y la transformación que sufre el sustrato; mientras que los últimos dos números identifican de forma precisa los átomos y sustancias sobre las que actúa la enzima.

Por tanto, estos números progresivamente más específicos, deben ser iguales al menos hasta el tercer dígito para enzimas que exhiben la misma capacidad catalítica sobre sustratos distintos (enzimas multi específicas). Sin embargo, en los casos de promiscuidad catalítica, los EC de los diversos sustratos y reacciones catalizadas por la misma enzima diferirán también en el segundo o tercer dígito, o incluso en el primer dígito si la enzima emplea un mecanismo de reacción completamente diferente. Sin embargo, los números EC pueden ser engañosos. Existen casos como el de la anhidrasa carbónica en los que sus reacciones canónica y promiscua (Fernández-Gacio et al., 2006; Gould y Tawfik, 2005), ambas consistentes en el ataque de un ion hidróxido en un carbonilo, distan significativamente en sus ECs (Fig. 1.5). Esto se debe principalmente a que las definiciones de EC se relacionan no solo con la química, sino también con el contexto fisiológico.

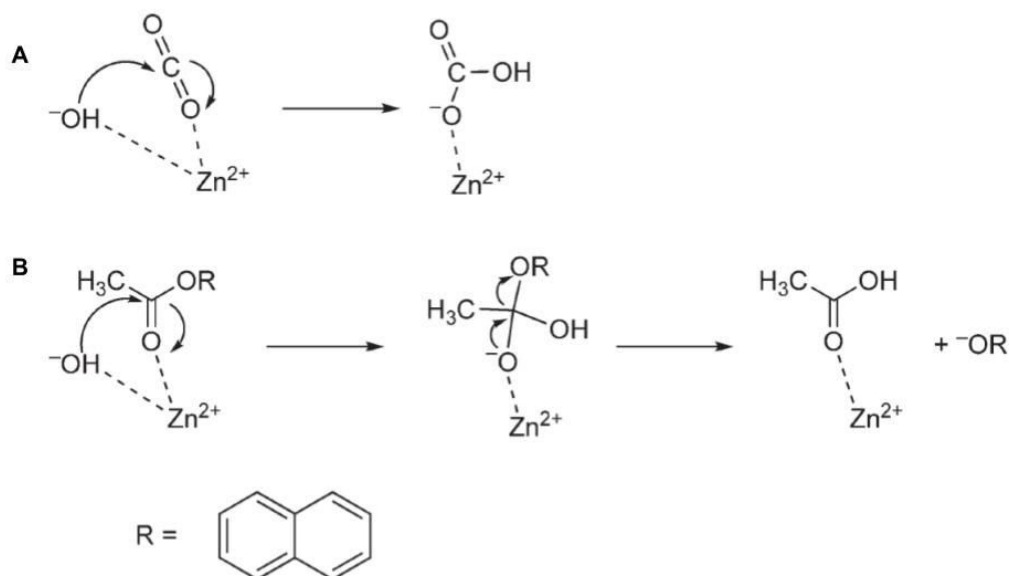


Fig. 1.5. Ejemplo de los posibles inconvenientes en el uso de los números EC como método de clasificación de proteínas promiscuas. A. Reacción nativa de hidratación de la anhidrasa carbónica con CO_2 (EC 4.2.1.1); B. Reacción promiscua arilesterasa de la anhidrasa carbónica, ejemplificada con acetato de naftilo (EC 3.1.1.X). Ambas reacciones implican el ataque de un ión hidróxido sobre un carbonilo, seguido de la estabilización de un oxianión intermedio por el sitio activo Zn^{2+} . A pesar de esta similitud, los números EC de estas reacciones difieren desde el primer dígito. Figura modificada de la referencia “*Enzyme Promiscuity – Evolutionary and Mechanistic Aspects*”. Olga Khersonsky y Dan S. Tawfik.

Es bien sabido que la actividad enzimática se origina a partir de interacciones entre moléculas pequeñas (ligandos) y sus objetivos proteicos, por lo tanto, no resulta extraño pensar en enfoques predictivos que contemplen esta característica. En este sentido, se han desarrollado algunos softwares de predicción de actividad promiscua basados en “firmas” moleculares, es decir descriptores que tienen en cuenta los átomos presentes en una dada molécula y su entorno químico a una distancia fija (Carbonell y Faulon, 2010). El método CLASP (*CataLytic Active Site Prediction*) (Chakraborty, 2012), por ejemplo, define una proteína “incubadora” con estructura tridimensional conocida y sitio activo definido para una dada actividad enzimática, sobre la que se estima un patrón o andamiaje estructural necesario para dicha catálisis y las características químicas de dicho patrón. A través de métodos heurísticos de

superposición y búsqueda de patrones de proteínas problema con proteínas incubadoras, CLASP busca identificar proteínas con posibles sitios promiscuos de igual actividad. En la misma línea y desarrollado por el mismo laboratorio, el método PROMISE amplía el criterio de búsqueda iniciado por CLASP y genera una estimación de un índice de promiscuidad (Chakraborty y Rao, 2012). Éste basa su estimación en las diferencias electroestáticas que genera la presencia de ciertos residuos en el entorno catalítico, asumiendo que la congruencia de la diferencia potencial implica un entorno propicio para la función particular.

Los métodos con basamento en el análisis de similitud estructural son de relativo fácil uso. Sin embargo, como bien mostraran Mu Gao y Jeffrey Skolnick, existe una gran complejidad en este tipo de análisis predictivo que radica en que, si bien existe un conjunto reducido de cavidades en el amplio mundo proteico, un sitio de unión podría unir múltiples ligandos (Gao y Skolnick, 2013). Estos métodos de predicción, por tanto, arrojan falsos positivos y resultan ineficaces para su objetivo.

Desde métodos sencillos que nos permiten diferenciar los tipos de promiscuidad mediante el uso de notación de EC, hasta métodos complejos que combinan la información estructural y fisicoquímica de la arquitectura de los sitios activos, las herramientas disponibles que hemos descrito hasta ahora basan sus predicciones en la existencia de información acumulada de casos de proteínas promiscuas. Sin embargo, esta información no se encuentra disponible de forma sistematizada y en muchos casos resulta incompleta. Los casos de promiscuidad en proteínas no enzimáticas (como las albúminas séricas bovinas) (Drmanovic et al., 1999; Hurst et al., 1999) o sin estructura conocida (Kremer y Li, 2008; Reetz et al., 2007) hacen más complejo aún el desarrollo y la aplicación efectiva de los métodos predictivos, dejando en evidencia la necesidad de bases de datos que nucleen y sistematicen la información experimental hasta ahora disponible.

1.5.e. Aplicaciones biotecnológicas de la promiscuidad proteica

En las últimas décadas se ha incrementado notablemente el uso de enzimas promiscuas para mejorar el rendimiento de determinadas síntesis orgánicas (Hult y Berglund, 2007; Perugini et al., 2004), facilitar la enantioselectividad en reacciones químicas (Magnusson et al., 2005), incrementar el poder catalítico en procesos de biorremediación (Scott et al., 2009) o favorecer la ocurrencia de reacciones no típicas en solventes orgánicos (Busto et al., 2010) no sólo para incrementar el conocimiento académico sobre el funcionamiento de las enzimas, sino también en el plano industrial, farmacéutico y agroquímico (Bornscheuer et al., 2012).

Las lipasas pertenecen al grupo de enzimas que se han estudiado ampliamente y aplicado como catalizadores promiscuos en diversas condiciones y con fines múltiples (Magnusson et al., 2005). Dichas enzimas han mostrado una estabilidad variada hacia la presencia de solventes orgánicos, condiciones extremas de pH y líquidos iónicos (De Diego et al., 2005; Guisan et al., 2001; Lima et al., 2004). Se ha encontrado que los solventes orgánicos influyen en la selectividad de las reacciones catalizadas por lipasas y pueden inducir nuevas actividades (Dwivedee et al., 2018; Hua et al., 2016; Lou et al., 2008). Se ha documentado, por ejemplo, la capacidad catalítica de lipasa B de *Candida antártica* en medio orgánico para la reacción de adición aldólica en medio orgánico (Branneby et al., 2003). Se ha descrito como un conjunto de residuos que podrían facilitar la transferencia de protones y la estabilización de la carga a través de enlaces de hidrógeno durante el proceso catalítico; sin embargo, las bases estructurales que justifican los cambios de funcionalidad en un mismo sitio activo y su relación con los cambios en el entorno proteico aún se desconocen.

Los esfuerzos de las investigaciones actuales en el tema se centran en comprender (i) el mecanismo de la promiscuidad catalítica, es decir cómo las enzimas, que generalmente ejercen gran especificidad, catalizan

promiscuamente otras reacciones; (ii) la evolución de las funciones promiscuas, para establecer qué factores determinan la aparición de estas actividades promiscuas y cómo evitan comprometer gravemente la actividad original. Sin embargo, la predicción de la promiscuidad enzimática y la elección de una enzima para un determinado propósito biotecnológico, resultan ser los principales obstáculos a vencer para su uso generalizado. Con nuestro trabajo buscamos aportar nuevas herramientas metodológicas y conceptuales que permitan extrapolar los conocimientos disponibles hasta el momento sobre el fenómeno de la promiscuidad proteica y sus bases estructurales, de modo de facilitar futuras aplicaciones biotecnológicas.

Objetivos generales y específicos

Objetivos

El objetivo general del presente proyecto es comprender las bases evolutivas y estructurales de la promiscuidad proteica. Creemos que esta comprensión es necesaria para el desarrollo de herramientas bioinformáticas que puedan predecir la ocurrencia de dichas propiedades. Cómo estas propiedades pueden considerarse elementos clave en el desarrollo y selección de biocatalizadores, el mejoramiento de técnicas predictivas podría redundar en la optimización de protocolos para la elección de biocatalizadores y de ahí tener un importante impacto en el plano biotecnológico.

Hemos mencionado que existe una estrecha relación entre la diversidad conformacional de una proteína y su poder catalítico. De esta forma, la hipótesis central de nuestro trabajo es que la diversidad conformacional debería jugar un rol esencial en el origen de la promiscuidad catalítica y/o de sustrato. Si bien los cambios conformacionales son necesarios para acomodar distintos ligandos y/o catalizar distintas reacciones químicas, el otro componente esencial de la promiscuidad consiste en que la enzima disponga de sitios o regiones secuenciales aptas para sustentar un determinado mecanismo catalítico. De esta forma, nuestra hipótesis central es que las distintas enzimas que catalizan una determinada reacción (canónica o promiscua) comparten estructuras “funcionales” como superficies, componentes secuenciales de cavidades y túneles, entornos adecuados del sitio activo, como residuos catalíticos y de anclaje del sustrato y de direccionamiento de sustratos y productos para sustentar una determinada actividad catalítica.

Objetivos específicos

Para estudiar dicha hipótesis, se procederá a:

- a. Generar una base de datos relacional entre diversidad conformacional y promiscuidad.
- b. Estudiar estructural y evolutivamente familias de proteínas que presenten actividades promiscuas descritas en nuestra base de datos, para la caracterización de sistemas promiscuos.
- c. Analizar las bases evolutivas y estructurales de la actividad promiscua de las albúminas séricas de interés biotecnológico.

Desarrollo de una base de datos para el estudio de la promiscuidad proteica: ProtMiscuity

El comportamiento promiscuo en proteínas desafía el concepto central de la relación estructura-función. La promiscuidad descrita como la capacidad de una enzima para catalizar múltiples reacciones químicas en el mismo sitio catalítico, dista mucho de ser un fenómeno poco común como se pensaba anteriormente y sus aplicaciones tecnológicas son múltiples. Si bien se han descrito diversos casos de promiscuidad catalítica, que involucran diferentes mecanismos, estos conocimientos no han sido sistematizados.

Aquí presentamos ProtMiscuity, una base de datos disponible en la web de proteínas que muestran promiscuidad catalítica. ProtMiscuity contiene información sobre actividades canónicas y promiscuas. Las entradas proporcionan información sobre sustratos de reacción, productos y parámetros cinéticos, mapeo de sitios activos en secuencias y estructuras; y el enlace a recursos externos con anotaciones biológicas y funcionales. ProtMiscuity, ofrece una colección de datos derivados experimentalmente, que podría ampliar el conocimiento sobre la función proteica en general y del fenómeno en particular.¹

¹ Este capítulo basa su contenido en la publicación de la autora: "ProtMiscuity: a database of promiscuous proteins". Ana Julia Velez Rueda, Nicolás Palopoli, Matías Zacarías, Leandro Matías Sommesse, Gustavo Parisi. Database. Volume 2019. <https://doi.org/10.1093/database/baz103>.

2.1. ProtMiscuity: una base de datos de proteínas promiscuas

A nivel experimental, los primeros enfoques para la identificación de actividades promiscuas fueron métodos que utilizaban bibliotecas de genes clonados en plásmidos de sobreexpresión. Estos tenían por finalidad identificar actividades promiscuas que pudiesen reemplazar la ausencia de una enzima en particular dentro de una ruta metabólica o en ensayar la resistencia a antibióticos (Colin et al., 2015; Miller y Raines, 2004; Patrick y Matsumura, 2008; Soo et al., 2016). No obstante, desde el punto de vista bioinformático, la predicción e identificación de actividades promiscuas representa un gran desafío aún.

Las perspectivas que se han utilizado para el abordaje de dicha problemática son de lo más variadas y abarcan desde la utilización de métodos de acoplamiento molecular o docking, que enfocan su análisis en la interacción proteína ligando; hasta métodos evolutivos secuenciales, que basan su análisis en la información secuencial de familias proteicas con funciones caracterizadas (Ver sección 1.5.d). A pesar de haberse diseñado un gran número de algoritmos capaces de predecir la promiscuidad y función proteica, existen múltiples trabajos científicos que alertan sobre la baja sensibilidad de dichas inferencias a partir de la información secuencial o estructural (Devos y Valencia, 2000; Gómez et al., 2003; White, 2006). La principal dificultad para el arribo de soluciones más efectivas radica, justamente, en la falta de colecciones de evidencias científica sistematizadas sobre promiscuidad. Aquí presentamos ProtMiscuity, una base de datos en línea que pretende llenar este vacío, proporcionando un conjunto de datos manualmente curados de proteínas promiscuas y la información biológica relacionada a cada una de ellas.

Atento al amplio espectro de significados que se le adjudican al término "promiscuidad" (ver sección 1.5.d), hemos restringido nuestra base de datos a casos de promiscuidad catalítica, descrita como la capacidad de una enzima

para catalizar reacciones secundarias en un sitio activo que se especializa para una reacción primaria diferente (Copley, 2003). Por tanto, ProtMiscuity nuclea información referente a proteínas que sostienen una o más reacciones químicas diferentes de la reacción catalizada canónicamente. Al organizar nuestro conocimiento sobre este tipo específico de promiscuidad, buscamos contribuir a varios proyectos tecnológicos, incluido el diseño de nuevos fármacos dirigidos a sitios activos conocidos para aplicaciones tanto biomédicas como industriales (Hult y Berglund, 2007) y el progreso en la ingeniería de proteínas para modular las funciones catalíticas (Bornscheuer y Kazlauskas, 2004), brindando pautas para la evolución dirigida de estructuras proteicas.

2.2. Aportes de la base de datos a la caracterización y comprensión de la promiscuidad proteica como fenómeno de interés biológico

Cuando hablamos de promiscuidad, abarcar todos los fenómenos que suelen ser englobados bajo esta denominación resulta no sólo imposible, sino además inapropiado, dadas las características tan disímiles que parecen poseer dichos fenómenos (ver sección 1.5.b). En particular, la promiscuidad catalítica es de por sí un fenómeno aún incomprendido tanto desde el punto de vista evolutivo como mecanístico. Se ha intentado caracterizarlo y organizarlo mediante diversos mecanismos más o menos prácticos, que merecen ser revisados si se pretende una sistematización de la información hasta ahora disponible, que permita un análisis más acabado del fenómeno en cuestión.

Aún cuando se han mostrado bases evolutivas para el origen de las actividades promiscuas (Luo et al., 2012; O'Brien y Herschlag, 1999; Zou et al., 2015), se han descrito también actividades promiscuas similares en proteínas no emparentadas cuya relación estructural únicamente se basa en la presencia de un bolsillo apolar en la superficie de dichas proteínas y una lisina reactiva (James y Tawfik, 2001). Tal es el caso de la seroalbúmina humana y el

anticuerpo aldolasa 38C2, que poseen en su superficie un bolsillo cuya base altamente hidrofóbica se encuentra rodeada por varios residuos aromáticos. A pesar de las considerables diferencias estructurales generales, la organización de los aparentes sitios activos de ambas proteínas, en un microambiente heterogéneo de residuos hidrofóbicos junto a residuos polares, parece ser suficiente para inducir efectos específicos que contribuyen a la catálisis (Hotta et al., 2000).

Como mencionamos en la sección 1.5.d, se han propuesto diversas herramientas que predicen la actividad promiscua de proteínas basadas en las características fisicoquímicas del sitio activo y su entorno, las características químicas de los sustratos y productos que intervienen en la reacción, o la comparación de los números de notación de la Comisión de Enzimas (EC) para las actividades canónicas y promiscuas (Carbonell y Faulon, 2010; Chakraborty y Rao, 2012; Khersonsky et al., 2006). Dichos métodos poseen ventajas desde el punto de vista mecanístico en la clasificación y estudio de proteínas promiscuas, ya que proveen una mecánica de comparación rápida, sistematizable y potencialmente escalable. Sin embargo, una de las dificultades que se plantean es la falta de datos que caractericen los sistemas promiscuos, dificultad en parte debida a que la mayoría de las actividades promiscuas no se encuentran anotadas de forma sistemática, o no se han establecido relaciones adecuadas con la información estructural y biológica disponible de cada sistema. Casos como el descrito de dos proteínas, que presentan promiscuidad catalítica con mecanismos similares sin estar emparentadas, nos dan una idea de la necesidad de anotaciones exhaustivas, que permitan el correcto trazado de generalizaciones, que hagan posibles predicciones sobre el comportamiento promiscuo de las proteínas.

Hemos visto que, dada la gran diversidad de casos de proteínas promiscuas, se requieren análisis de gran complejidad para el estudio de la promiscuidad como fenómeno. Éstos análisis deben contemplar información concerniente no sólo a las características químicas de las reacciones catalizadas, sino también información evolutiva y estructural. Tales consideraciones sólo pueden ser realizadas contando con una fuente ordenada de información de los sistemas

promiscuos. La base de datos ProtMiscuity, en tanto una base que sistematiza la información disponible y ofrece una colección única y curada de datos derivados experimentalmente, sienta la base de futuros estudios sobre los mecanismos subyacentes a las reacciones promiscuas.

2.3. Desarrollo de la base de datos ProtMiscuity

2.3.a. Reclutamiento bibliográfico por *web scraping* y minería de textos

Los datos incluidos en ProtMiscuity fueron recopilados a través de la implementación de técnicas de *web scraping* y de minería de texto sobre bibliografía relacionada con la promiscuidad proteica, utilizando bibliotecas estándar del lenguaje de programación Python (Ver Fig. 2.1). La técnica de web scraping permitió realizar búsquedas a gran escala en diferentes bases de datos de publicaciones científicas, como ser PubMed², y sitios web de revistas científicas, simulando la navegación manual en la web y mediante el uso de expresiones regulares para extraer la información contenida en cada sitio. Dichos procedimientos permitieron reclutar más de 500 publicaciones científicas relacionadas con el fenómeno de interés.

Mediante el análisis y *parseo* de dichos textos con herramientas de minería de textos, implementados y adaptados para detectar y extraer terminología relacionada con la promiscuidad catalítica; y con su posterior inspección manual, se identificaron los textos relevantes para la descripción y caracterización de las distintas proteínas promiscuas. De este modo, se seleccionaron 301 publicaciones, basadas en criterios de texto, para la posterior revisión exhaustiva y la extracción de datos de relevancia biológica que ayudaran a caracterizar los sistemas descritos. De esta colección de referencias a probables proteínas promiscuas, fueron descartadas aquellas publicaciones en las cuales la evidencia experimental sobre la promiscuidad no

² PubMed: <https://www.ncbi.nlm.nih.gov/pubmed/>

fuera suficiente; por ejemplo, cuando no se trabajara con la proteína pura, si no se reportan niveles de actividad significativos de la función promiscua, o no fuese posible la asignación confiable de la identidad de la proteína a través de la inferencia de su identificador Uniprot (Pundir et al., 2017; The UniProt Consortium, 2017)). Mediante este análisis obtuvimos un total de 57 proteínas con una o más actividades promiscuas caracterizadas, provenientes de 42 organismos diferentes.

2.3.b. Curado y extracción de datos de relevancia

El proceso de curación manual de la bibliografía derivada del proceso inicial de reclutamiento de casos, incluyó la revisión crítica de dichos artículos y el análisis de los datos experimentales que estos aportan. Para cada proteína fue verificada su secuencia, permitiendo de este modo la asignación de un código Uniprot (Pundir et al., 2017; The UniProt Consortium, 2017), que permitiera la relación posterior con otras bases de datos que aportaran a la caracterización biológica de dicha proteína. Asimismo, se extrajeron, cuando se encontraban disponibles, datos relativos a las reacciones catalizadas por cada proteína, tanto promiscuas como canónicas, como ser: los parámetros cinéticos de dichas reacciones (K_{cat}), los sustratos y productos ensayados para cada reacción y los parámetros relacionados con la afinidad de a los mismos (K_m y K_m/K_{cat}). Además, fueron extraídos los EC asociados a cada actividad enzimática que hubiesen sido anotados bibliográficamente. Por otro lado, se extrajeron, en los casos que hubiesen sido descritos, los residuos pertenecientes al sitio activo de la proteína en cuestión.

Mediante el uso del código Uniprot asignado a cada entrada, mapeamos la información relativa a los sitios activos de cada proteína, anotados por medio de la curación bibliográfica manual, en las estructuras cristalográficas disponibles en la base de datos Protein Data Bank (PDB) (Berman et al., 2000; Touw et al., 2015). Como se detallará más adelante (sección 3.4.c), todos los

datos biológicos, estructurales y químicos de relevancia para caracterizar a cada proteína, fueron complementados desde diversas bases de datos (Fig. 2.1).



Fig. 2.1. Mediante técnicas de web scraping y minería de texto fueron seleccionados los artículos académicos de los casos de interés. Los datos extraídos fueron cruzados con datos provenientes de distintas bases de datos para enriquecer la información de cada entrada de ProtMiscuity.

2.4. Servidor Web ProtMiscuity

2.4.a. Descripción general de la base de datos

En su versión actual, ProtMiscuity cuenta con un total de 57 proteínas documentadas como casos positivos de promiscuidad proteica, provenientes de 42 organismos diferentes, y caracterizadas con una o más actividades promiscuas catalíticas. Esta incluye, además, 2001 diferentes estructuras cristalográficas relacionadas con dichas proteínas. Cada proteína se halla anotada en Protmiscuity por su identificador UniProt (Pundir et al., 2017) y su nombre completo, de modo de que sea posible y rápida su vinculación con otras bases de datos de interés biológico.

Nuestra base de datos cubre un total de 88 reacciones químicas distintas, entre las cuales 54 fueron identificadas como reacciones promiscuas y 38 como canónicas. El 68% de las entradas de la base de datos tiene solo una reacción promiscua descrita, mientras que 20% de las entradas tienen dos y solo un

12% tienen tres o más. Las reacciones, tanto promiscuas como canónicas, fueron caracterizadas con sus correspondientes descriptores cinéticos, con la información obtenida de la literatura. Esta incluye, no sólo los valores de K_m y k_{cat} obtenidos experimentalmente para dichas reacciones, sino también la descripción química de los sustratos y productos, y condiciones de reacción que se utilizaron en dichos ensayos experimentales. Asimismo, los sustratos y productos relacionados con cada reacción descrita fueron vinculados a la información disponible en PDB Ligand Expo (Sitzmann et al., 2012) y PubChem (Kim et al., 2016) para facilitar la identificación de posibles sustratos por similitud química. La información estructural relativa a cada proteína promiscua (dominios estructurales, residuos de sitios activos y datos sobre la diversidad conformacional vinculada a dicha proteína) también se encuentran disponibles para cada entrada de la base de datos.

ProtMiscuity también incluye una sección de tutoriales y de respuestas a las preguntas más frecuentes para facilitar la navegación y el uso por parte de usuarios no experimentados. Todos los datos se pueden descargar como un archivo de valores separados por comas (CSV).

2.4.b. Implementación del Servidor Web ProtMiscuity

En el primer objetivo del proyecto se planteó el desarrollo de una base de datos relacional entre diversidad conformacional y promiscuidad proteica. Para tal fin se construyó una base de datos de proteínas promiscuas: ProtMiscuity³, de acceso libre y gratuito para su uso académico. La base de datos, disponible en la web desde el 2018, suministra al usuario información curada manualmente acerca de proteínas con actividades promiscuas de catálisis, los parámetros cinéticos de dichas reacciones y toda la información biológica y estructural derivada de otras bases de datos.

³ ProtMiscuity: <http://ufq.unq.edu.ar/protmiscuity>

El servidor⁴ fue implementado utilizando Java y SpringBoot⁵. El conjunto de datos curado se almacenó como una base de datos relacional MySQL 5.5.43⁶, compuesta por 8 tablas: 7 tablas principales que almacenan toda la información relativa a cada proteína promiscua (Proteina, Reaccion, Cinetica, Molecula, SitioActivo, Enlace y EstructuraPDB) y una tabla intermedia (MoleculasSimilares) que implementa una relación muchos-a-muchos entre la información cinética de una reacción dada y las moléculas similares asociadas (Fig. 2.2). Se creó una interfaz web responsiva para ProtMiscuity⁷, que brinda soporte para una navegación y visualización más fácil de los contenidos de la base de datos en múltiples dispositivos. Se implementó en HTML, CSS y JavaScript, con Angular4 y NodeJS. ProtMiscuity está alojado en nuestro servidor y se puede acceder libremente en <http://ufq.unq.edu.ar/protmiscuity>.

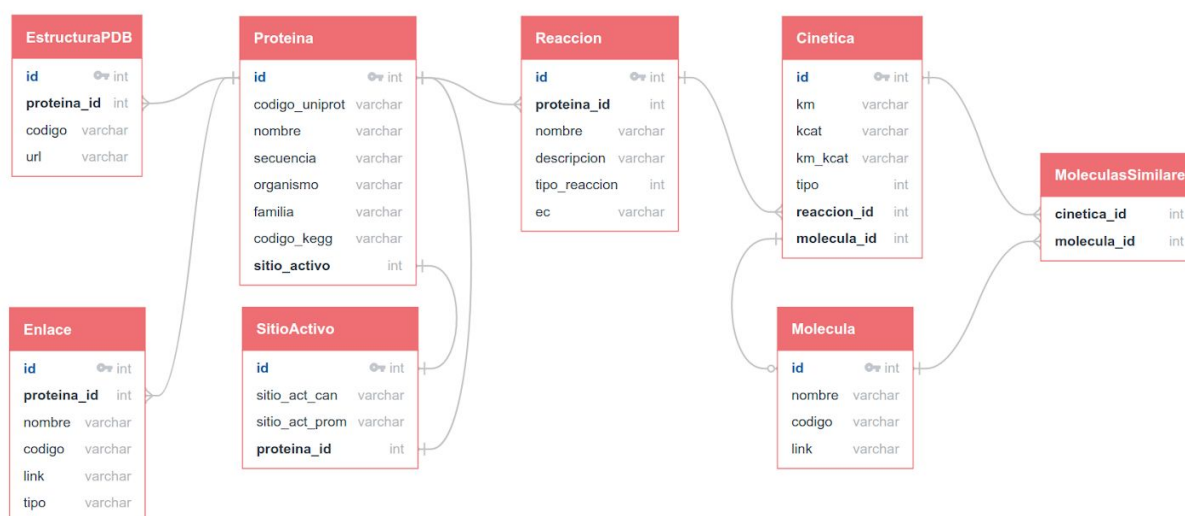


Fig. 2.2. Diagrama entidad-relación (DER) del modelo físico de datos ProtMiscuity, tal como se presenta en la base de datos MySQL. Éste se compone de 8 tablas. En el presente diagrama se observan 7 tablas principales que almacenan toda la información relativa a cada proteína promiscua (Proteina, Reaccion, Cinetica, Molecula, SitioActivo, Enlace y EstructuraPDB) y una tabla intermedia (MoleculasSimilares) que implementa una

⁴ Repositorio GitHub del código fuente del back end ProtMiscuity: <https://github.com/protein-api/protein-manager-api>

⁵ SpringBoot: <https://spring.io/projects/spring-boot>

⁶ MySQL: www.mysql.com

⁷ Repositorio GitHub del código fuente de la aplicación web ProtMiscuity: <https://github.com/protein-api/protmiscuity-front>

relación muchos-a-muchos entre la información cinética de una reacción dada y las moléculas similares asociadas.

Las consultas realizadas por los usuarios se envían a la base de datos por medio del sitio web y se muestran de forma dinámica en el servidor según los criterios definidos por el usuario. Estas consultas son enviadas al motor de base de datos MySQL, para que se ejecute la consulta y devuelvan los datos relativos a las entradas que se vinculen con dichos criterios. El resultado es procesado en el lado del servidor y mostrada en un formato accesible al usuario. Generalmente los datos están organizados en diferentes secciones que permiten ver fácilmente la información de una determinada proteína.

2.4.c. Vinculación con otras fuentes de información biológica

Cada entrada de proteína promiscua asentada en la base de datos ProtMiscuity ha sido asociada con su correspondiente código de UniProt (Pundir et al., 2017), con el fin de derivar toda la información que esta contiene de cada proteína y vincularlos con los datos provenientes de otras bases de datos de uso académico. UniProt, es una base de datos que nuclea las secuencias y la información funcional de proteínas conocidas. Esta identifica a cada producto génico (una determinada proteína de un determinado organismo) asignando un código, por medio del cual puede ser reclutada la información sobre sus actividades promiscuas y canónicas en ProtMiscuity.

Para proporcionar a los usuarios información que permita correlacionar los datos funcionales y cinéticos con las características estructurales y de diversidad conformacional, cada proteína de la base de datos fue vinculada con PDB (Touw et al., 2015), CoDNAS (Monzon et al., 2016) y PFAM (Finn et al., 2016), entre otros. La base de datos PDB es el repositorio de dominio público de estructuras tridimensionales de proteínas y ácidos nucleicos, obtenidas experimentalmente por los métodos de cristalografía de rayos X o resonancia magnética. CoDNAS, por otro lado, es una base de datos de diversidad conformacional, basada en proteínas enteras derivadas de PDB. Esta contiene

la colección redundante de todas las estructuras diferentes (confórmers) e información relativa a la diversidad conformacional correspondientes a cada proteína en la base de datos. Por último, PFAM es una colección de familias de dominios de proteínas, que proporciona al usuario información sobre la funcionalidad de cada proteína.

Asimismo, cada proteína promiscua perteneciente a la base de datos ha sido caracterizada funcionalmente mediante la vinculación con la herramienta QuickGo (Binns et al., 2009). Dicha herramienta proporciona anotaciones funcionales para productos génicos, mediante el uso de un vocabulario controlado que les otorga atributos que describen típicamente tres aspectos: la función molecular, su rol en los procesos biológicos y su localización celular. En este sentido, y para proveer al usuario información que le permita comprender las funciones y utilidades de alto nivel del sistema biológico a partir de información de nivel molecular, se realiza la vinculación de cada entrada de ProtMiscuity con la base de datos KEGG (Kanehisa y Sato, 2020). Toda la información obtenida de la vinculación con otras bases de datos se explicita en las distintas secciones mediante las especificaciones y enlaces correspondientes.

2.4.d. Ejemplo de búsqueda en ProtMiscuity

Se pueden realizar búsquedas en ProtMiscuity tanto por nombre de proteína o Uniprot ID, como por organismo o por nombre de actividad canónica o promiscua. El usuario simplemente debe especificar alguno de estos identificadores y hacer clic en el botón de búsqueda para obtener la información relativa a dicha proteína. Como búsqueda alternativa, el usuario puede buscar todas las proteínas relacionadas con un organismo dado, seleccionando el botón "Organismo" del menú desplegable. Asimismo, es posible buscar todas las proteínas que tengan una dada reacción promiscua o canónica aplicando como motor de búsqueda el nombre de la actividad que sea de interés. También se encuentra disponible un índice de proteínas que

permite navegar por la base de datos, permitiendo acceder a cada entrada de forma automática.

Luego de realizar una búsqueda, ProtMiscuity mostrará en la página los posibles resultados mediante un diseño que representa cada proteína con un tarjeta que resume la información disponible para dicha entrada. Tal información incluye la familia de proteínas a la que pertenece, el organismo de origen, el número de reacciones promiscuas y canónicas en las que está involucrada y el número de estructuras relacionadas con la proteína en cuestión (Fig. 2.3.A). El usuario puede presionar la tarjeta deseada para recuperar en detalle toda la información disponible para la proteína. Una consulta típica que utiliza el nombre de la proteína, recupera información general de la misma agrupada bajo cinco títulos, en forma de solapas desplegadas: Información general de la proteína, Secuencia y sitios activos, Reacciones e información cinética, Visualización de estructuras y Diversidad conformacional, bibliografía y enlaces externos.

La sección de información general incluye: el nombre común de la proteína, el código Uniprot, los identificadores de KEGG, el organismo del que proviene la proteína y el nombre de la familia de la proteína (Fig. 2.3.C). Al hacer clic en la sección de “Secuencia y sitios activos”, se coloca en contexto secuencial el mapeo, no solo de los sitios de reacción promiscuos y canónicos que se describen para dicha proteína, sino también a través de la información proveniente de las principales bases de datos de dominios de proteínas, estructuras secundarias y terciarias, desórdenes intrínsecos y sitios de unión, todos proporcionados por la herramienta de visualización interactiva ProViz (Jehl et al., 2016). ProtMiscuity proporciona, además, una descripción detallada de las reacciones canónicas y/o promiscuas catalizadas por la proteína. Al hacer clic en una tarjeta de reacción, pueden ser visualizadas las listas de sustratos y productos ensayados para dicha reacción, vinculados a PubChem (Kim et al., 2016) y con parámetros cinéticos (K_m , k_{cat}) cuando están disponibles en bibliografía.

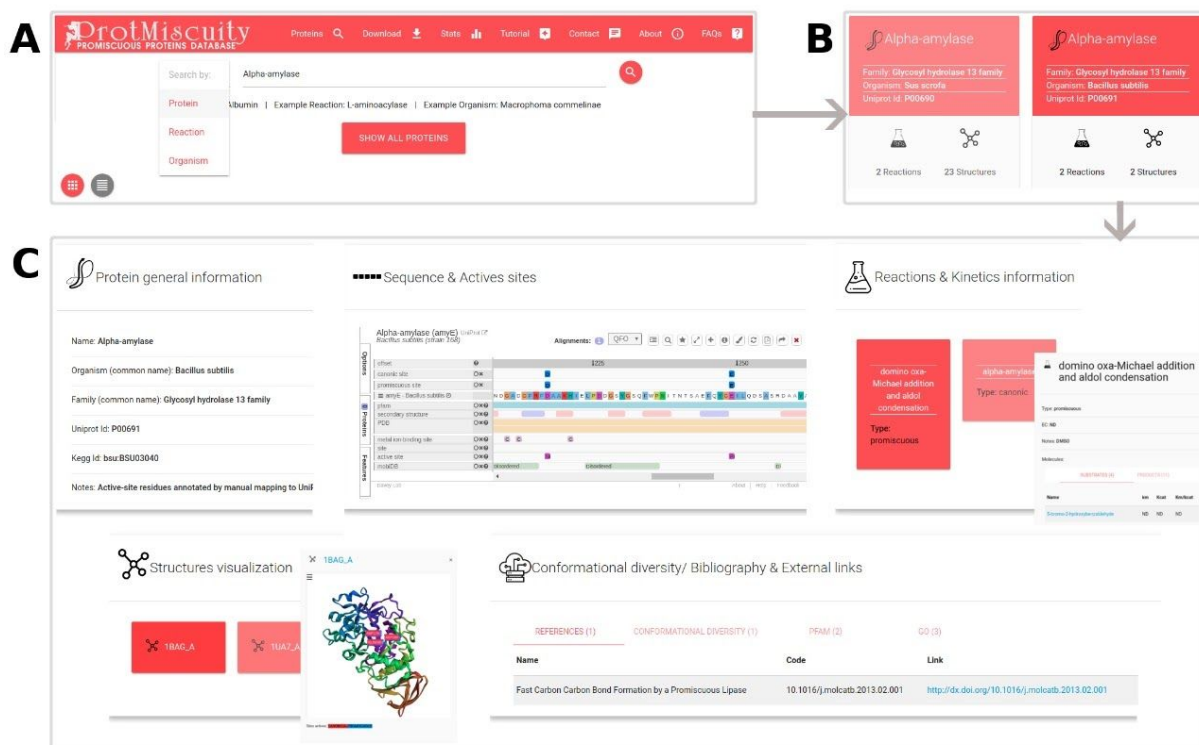


Fig. 2.3. A) Página de inicio de ProtMiscuity. Se puede buscar en la base de datos usando nombres de proteínas, organismos y reacciones objetivo. En este ejemplo, se realiza una búsqueda de la proteína alfa-amilasa. B) Página de resultados. Muestra todas las coincidencias con el término de consulta en forma de tarjetas específicas de proteínas. En este ejemplo, se recuperan las alfa-amilasas de dos organismos distintos. C) Página de información. Al hacer clic en la tarjeta de una proteína se muestra toda la información disponible, organizada en cinco secciones de interés. De arriba a abajo, de izquierda a derecha: una descripción general de la proteína; el mapeo de los sitios activos canónicos y promiscuos, junto con otra fuente de información relevante, sobre la secuencia de la proteína; información sobre actividades canónicas y promiscuas, con sustratos, productos y parámetros cinéticos conocidos (panel superior); una visualización de cada estructura disponible de la proteína, con los sitios catalíticos mapeados en ella; y ejemplos de diversidad conformacional, además de enlaces a bibliografía relevante y otras bases de datos, como pestañas separadas (panel inferior).

Asimismo, es posible visualizar todas las estructuras de PDB relacionadas con la proteína de interés por medio de la herramienta 3Dmol.js (Rego y Koes, 2015) integrada a la base de datos. El usuario puede acceder a todos los PDB disponibles para una dada proteína haciendo clic en las tarjetas, lo que le permitirá la visualización de la estructura de PDB, con marcas en los residuos

correspondientes a los sitios activos canónico y promiscuo, de forma interactiva. ProtMiscuity también proporciona información sobre la diversidad conformacional de las proteínas mediante la referencia cruzada de la proteína buscada con la base de datos CoDNaS. De igual modo, se vincula cada entrada a las publicaciones originales que proporcionaron los datos descritos en la base de datos, a los dominios PFAM anotados para dicha proteína, y a las diversas categorías de Ontología de genes asociadas a la misma.

2.4.e. Estadística del contenido de la base de datos

Comprender el origen y los mecanismos relacionados con la promiscuidad puede ser útil para lograr una interpretación más profunda de la función y evolución de la proteína. La caracterización del comportamiento promiscuo ha ampliado el repertorio químico de las reacciones enzimáticas, descubriendo un gran número de aplicaciones potenciales en biotecnología y áreas relacionadas (Bornscheuer y Kazlauskas, 2004; Hult y Berglund, 2007; Nemoto et al., 2013; Nobeli et al., 2009). En este sentido, ProtMiscuity proporciona un recurso único y útil para explorar nuevas actividades catalíticas y sus mecanismos subyacentes.

Dada la diversidad de organismos y especies representadas en la base de datos, podemos inferir que la promiscuidad catalítica es una propiedad de todos los linajes taxonómicos. En este sentido, existen proteínas pertenecientes a la misma familia proteica que presentan actividad promiscua en miembros pertenecientes a los tres dominios. Por ejemplo, encontramos que la familia de las AB hidrolasa, que tiene la mayoría de los miembros anotados en nuestra base de datos y es la familia más representada en ProtMiscuity, incluye proteínas que se encuentran presentes en varias bacterias y hongos, pero también en el trigo común y el en cerdo. Dentro del total proteínas de la base de datos, el 60.41% de las corresponden a bacterias, el 37.5% a eucariotas y únicamente el 2.08% pertenece a arqueas (Fig. 2.4);

siendo la familia mayoritaria la de las AB hidrolasas, seguida de la familia de las Hidrolasas metalo-dependientes.

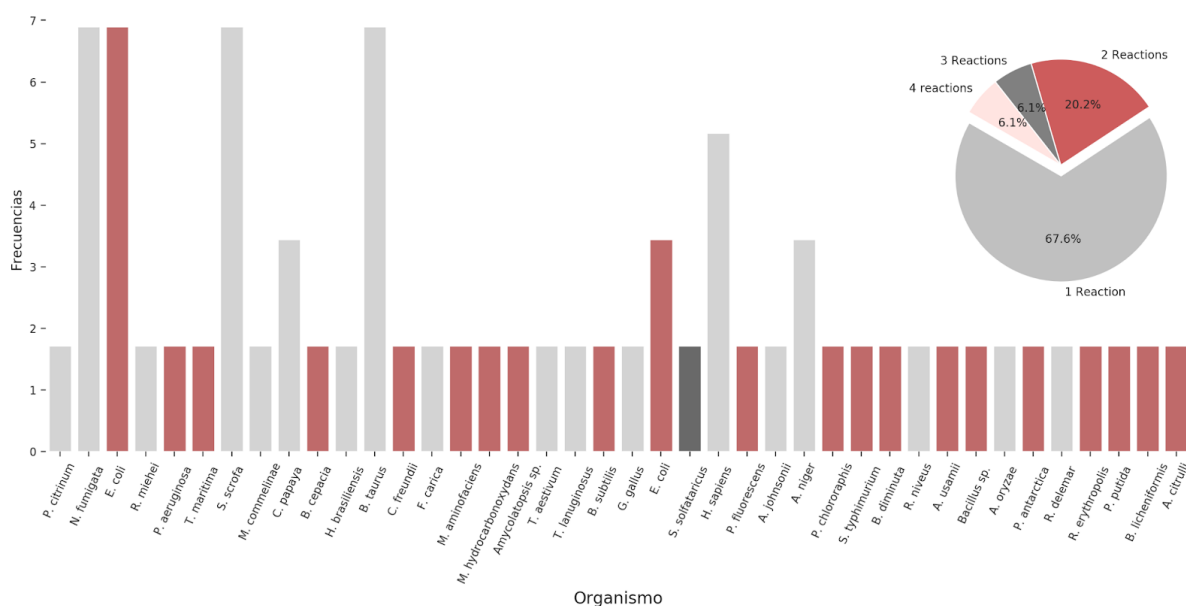


Fig. 2.4. Representación de la distribución taxonómica en ProtMiscuity. En el diagrama de torta se representan las frecuencias de proteínas pertenecientes a los tres dominios Eukarya, Archaea y Bacteria (Color coral, gris claro y gris oscuro respectivamente). Además, en gráfico de barras se muestran las frecuencias de cada organismo en la base de datos cuyos colores representan en la misma gama de colores el dominio al que pertenece cada organismo.

En su versión actual, ProtMiscuity ofrece 54 ejemplos curados de reacciones químicas promiscuas que involucran ~ 580 productos y sustratos diferentes. Resulta interesante que el 12% de las 58 proteínas anotadas en la base de datos tienen más de dos reacciones promiscuas, aunque todavía no está claro cómo estas reacciones se complementan entre sí (Fig. 2.5). Tal es el caso de la albúmina sérica bovina (ASB) para la que se describen en nuestra base de datos 4 reacciones promiscuas diferentes (reacción de Henry, condensación aldólica, eliminación de Kemp y reacción de Biginelli) asociadas con el mismo sitio activo (Lys 222), ubicado en su cavidad principal (Boucher et al., 2005; Miao et al., 2015; Sharma et al., 2013; Wu et al., 2006). Solo algunas de estas reacciones promiscuas son compartidas por la seroalbúmina humana, de la misma familia proteica, y por otras proteínas no relacionadas como la aldoxima deshidratasa. Estos ejemplos hacen evidente la complejidad subyacente a este

fenómeno y refuerzan, por tanto, la necesidad de ahondar en el estudio de los mecanismos que permiten dichas funcionalidades para comprender mejor la evolución de las proteínas.

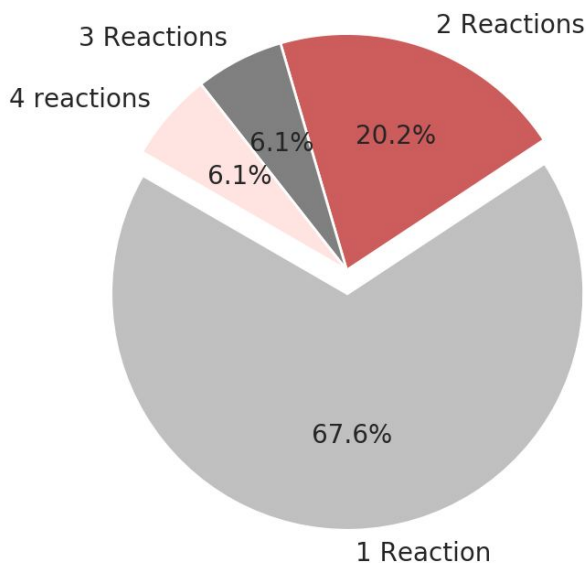


Fig. 2.5. Número de reacciones promiscuas por proteína. Se representan en el diagrama de torta el porcentaje de proteínas promiscuas con 1, 2, 3 y 4 o más reacciones promiscuas descritas, correspondiendo a las porciones gris claro, rojo, gris oscuro y rosa respectivamente.

2.4.f. Conclusiones destacadas

La disponibilidad de ejemplos curados como los ofrecidos por ProtMiscuity podría ser de gran importancia para profundizar en los problemas conceptuales de la promiscuidad proteica, como su origen evolutivo y su posible relación con la dinámica de las proteínas y la versatilidad química y funcional (Zou et al., 2015). Además, los conjuntos de datos curados proveen información sobre las cavidades, superficies y arreglos de aminoácidos relacionados a las distintas actividades (Gutteridge y Thornton, 2005; Holliday et al., 2009), lo que permite a los usuarios recopilar múltiples descriptores de los sitios catalíticos activos, que pueden mejorar el diseño de protocolos de ingeniería de proteínas y el descubrimiento de fármacos in silico. Queda como trabajo a futuro, la incorporación en Protmiscuity de información sobre otras propiedades

estructurales similares, tales como la presencia de túneles y cavidades (Gora et al., 2013) y la descripción de las particularidades fisicoquímicas de los aminoácidos que conforman los sitios activos promiscuos (es decir, cambios de pka) (Gutteridge y Thornton, 2005).

Catálisis y promiscuidad en medios no acuosos: estudio de la diversidad conformacional de proteínas en medio no acuoso

Existen muchos casos descritos de promiscuidad catalítica en solventes orgánicos, que resultan de gran interés en el campo de la industria, ya que amplían las fronteras de la biocatálisis. Sin embargo, poco se han caracterizado los cambios estructurales que adoptan las proteínas en dichos medios no acuosos. Algunos estudios informan que las proteínas co-cristalizadas o embebidas en solventes orgánicos conservan su integridad estructural. Esta observación contrasta con el hecho de que la actividad catalítica de las proteínas en dichos solventes es mucho menor que la observada en agua. Para explicar esta actividad reducida y caracterizar mejor el comportamiento de las proteínas en entornos no acuosos, realizamos un análisis comparativo, a gran escala, de la diversidad conformacional de proteínas cristalizadas en agua y co-cristalizadas o empapadas en medios acuosos.

Sobre la base del estudio de la diversidad conformacional hecho en más de 1000 proteínas distintas, se observó que las proteínas en medios no acuosos muestran una diversidad conformacional mucho menor en comparación a sus conformeros correspondientes obtenidos en agua. Estas características estructurales específicas, con cavidades menos flexibles y menor desorden, muestra la existencia de un repertorio de conformaciones más acotado en el paisaje conformacional posible de la proteína.

Nuestros resultados desafían la idea de “atrapamiento cinético” hasta ahora establecido, y se proporciona un nuevo modo de comprender las funcionalidades diferenciales de las proteínas en los distintos medios⁸.

⁸ Este capítulo basa su contenido en la publicación de la autora: Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. Ana Julia Velez Rueda, Alexander Miguel Monzon, Sebastián M Ardanaz, Luis E Iglesias, Gustavo Parisi. BMC Bioinformatics, 2018. DOI: <https://doi.org/10.1186/s12859-018-2044-2>.

3.1. Catálisis y promiscuidad en medios no acuosos

La biocatálisis en solventes orgánicos es hoy en día una práctica común y con un gran potencial biotecnológico (Arora et al., 2014; Hult y Berglund, 2007). Básicamente, el uso de disolventes orgánicos en la catálisis enzimática puede ofrecer ventajas sobre el uso de un medio acuoso, ya que entre otras cosas: aumenta la solubilidad de muchos sustratos y reactivos orgánicos, disminuye las reacciones secundarias no deseadas frecuentes en agua, y permite también la separación de los biocatalizadores al final de la reacción y una purificación más fácil de la mezcla de reacción, debido a la insolubilidad de la enzima en disolventes orgánicos y puntos de ebullición más bajos de los disolventes orgánicos comunes (Klibanov, 2001). Asimismo, se ha mostrado la inducción de actividades promiscuas en solventes orgánicos, que resultarían imposibles en medios acuosos, y que amplían el repertorio catalítico de las enzimas (Busto et al., 2010; Qian et al., 2007; Xue et al., 2012). En este sentido, múltiples estudios sugieren que el entorno proteico influye en su plegamiento y, por lo tanto, en su actividad biológica. La presencia de ligandos, la concentración de iones, la temperatura, la cantidad de moléculas de agua unidas y la presencia de moléculas orgánicas, como el disolvente, afectan el plegamiento de la proteína y, por tanto, su funcionalidad (Serdakowski y Dordick, 2008).

Al contrario de lo que se puede creer en bioquímica, ya que la mayoría de las enzimas evolucionaron y desempeñan su función en un medio acuoso, varios estudios de investigación han encontrado que las proteínas co-cristalizadas o empapadas (soaked) en solventes orgánicos preservan la integridad del plegamiento proteico (Gao et al., 1999). Se han obtenido varias estructuras de proteínas en diferentes solventes orgánicos, tales como: quimotripsina en hexano (Yennawar et al., 1994), subtilisina en acetonitrilo anhidro (Schmitke et al., 1997), tripsina en ciclohexano (Zhu et al., 1998), lisozima clara de huevo en presencia de alcoholes (Deshpande et al., 2005) y termolisina en isopropanol (English et al., 1999), solo por mencionar algunos ejemplos. Se propone desde la teoría de "atrapamiento cinético" que las proteínas en medios no acuosos

permanecen en su estructura nativa debido a una mayor cantidad de enlaces de hidrógeno entre los átomos de proteínas, lo que resulta en una barrera cinética más alta para los reordenamientos estructurales (Ke y Klibanov, 1998). Este efecto se lo relaciona con la deshidratación y la resuspensión que tienen lugar durante la cristalización (Chang et al., 1988; Ke y Klibanov, 1998; Ru et al., 2000). Se acepta, además, que las proteínas liofilizadas tienen un comportamiento determinado en función del pH de la solución acuosa de la que se liofilizaron, permaneciendo en la misma conformación cuando se transfieren a un entorno no acuoso.

Pero, a pesar de esta "conservación estructural", que se describe en varios artículos, está bien establecido que la actividad catalítica de las proteínas en un medio orgánico es menor que en el agua (Clark, 2004; Klibanov, 1989). Incluso cuando la mayoría de las proteínas co-cristalizadas o empapadas en un medio orgánico pudieran tener casi la misma estructura que la obtenida en agua, únicamente la preservación de la estructura no garantizará la misma actividad de la proteína. Por ejemplo, las enzimas de los organismos termófilos están inactivas a bajas temperaturas debido a la falta de energía térmica suficiente para superar el exceso de rigidez que muestran estas proteínas (Elias et al., 2014). El plegamiento proteico se conserva en su estado "nativo" a bajas temperaturas; sin embargo, la falta de "dinamismo estructural" o cambios conformacionales produce inactivación. Por lo tanto, el término "estado nativo" debe incluir características estructurales y dinámicas de las proteínas. En este sentido, y como mencionamos en la introducción de este trabajo de tesis, está bien establecido que el estado nativo se entiende mejor como un conjunto de múltiples conformeros estructurales que coexisten en equilibrio (Kumar et al., 2000).

Se han caracterizado una amplia gama de diferencias estructurales entre los conformeros en medios acuosos para explicar las funciones de las proteínas, desde los movimientos relativos de dominios (Gerstein et al., 1994), reordenamientos de elementos de la estructura secundaria y terciaria (Gerstein y Krebs, 1998) y los movimientos de bucle (Gu et al., 2015), hasta reestructuraciones de regiones de la proteínas que carecen de una estructura

bien definida (conocidas como regiones intrínsecamente desordenadas (IDP, del inglés)) (van der Lee et al., 2014). Pero, además de estos grandes reordenamientos estructurales, también se observan pequeños movimientos necesarios para la función biológica y que posibilitan la catálisis (Koshland, 1998; Mesecar et al., 1997). En un estudio de diversidad conformacional, realizado en 60 enzimas en sus formas apo y unidas al sustrato en solventes acuosos, Gutteridge y Thornton (Gutteridge y Thornton, 2005) observaron que los movimientos necesarios en las enzimas para unir los sustratos eran muy pequeños, y que las enzimas que requerían movimientos grandes representaban sólo una pequeña proporción. Un 75% de sus datos mostraron una desviación media cuadrática de la posición de los carbonos alfa (RMSD) menor a 1 Å, y el 91% tenía un RMSD menor que 2 Å con un promedio de 0.7 Å. Curiosamente, también observaron que las comparaciones de las estructuras apo para la misma proteína mostraron un RMSD de 0.5 Å, un valor ligeramente por debajo del promedio observado de para las conformaciones de unión al sustrato y apo. Estos hallazgos refuerzan la idea que aún los pequeños cambios entre los conformeros podrían afectar en gran medida el comportamiento de las enzimas (Mesecar et al., 1997). Además, en los últimos años, varios estudios han revelado la importancia de estructuras como los bolsillos, cavidades y túneles, en la función de la proteína (Gora et al., 2013). Brevemente, estas estructuras participan en la canalización de sustratos y otros ligandos (cofactores, productos, etc.) desde la superficie de la proteína a las cavidades internas que probablemente están asociadas con sitios activos o de unión. La apertura y el cierre de estas estructuras a través de movimientos ligeros de muy pocos residuos podrían definir conformaciones activas o inactivas (Chovancova et al., 2012).

Dados todos los cambios estructurales, grandes y pequeños, necesarios para que la catálisis se lleve a cabo, y teniendo en cuenta la característica dinámica del estado nativo proteico, nos proponemos abordar el estudio de las transiciones conformacionales de proteínas desde medios acuosos a medios no acuosos. Para lo cual, tendremos en cuenta no uno, sino todos los aspectos involucrados en la funcionalidad enzimática como posibles causales de la

menor actividad observada en medios orgánicos y de la inducción de nuevas actividades catalíticas.

3.2. Métodos específicos para el estudio de los cambios conformacionales en medios no acuosos

3.2.a. Reclutamiento y filtrado de estructuras cristalográficas para el análisis.

Para examinar los cambios estructurales asociados a la transición desde un medio acuoso a uno no acuoso, se focalizó el análisis en proteínas que tuviesen estructuras cristalinas obtenidas en distintos medios y que permitieran su comparación. Sin embargo, la información acerca de los procedimientos aplicados para la cristalización de las proteínas no siempre se encuentra disponible. Para resolver este problema, creamos una lista consenso de los solventes orgánicos y medios no acuosos comúnmente utilizados en el proceso de cristalización, confeccionada a partir de la exploración de manuales cristalográficos y artículos de investigación relacionados con el tema (Ver Tabla suplementaria 1). Utilizando la lista antes mencionada, se obtuvieron estructuras cristalinas (sin mutaciones y de resolución <4 Å) de la base de datos de Diversidad Conformacional en el estado nativo de las proteínas (CoDNaS) mencionada anteriormente (Monzon et al., 2013).

La presencia de estas moléculas orgánicas en el cristal, indicada en el campo HETATOM de los archivos PDB, se usó para distinguir las estructuras cristalizadas en ambientes acuosos y no acuosos. Se consideraron entonces aquellas proteínas que tuviesen al menos un confórmero en solvente orgánico y uno en agua para construir el conjunto global de datos (dataset grande), que resultó en un conjunto de 1737 proteínas y un total de 3474 estructuras. Asimismo, consideramos otro conjunto de datos resultante del método de web scraping y de la curación manual para la recopilación de estructuras relacionadas con soaking (embebido o empapado) y co-cristalización en

solventes orgánicos, obteniendo un total de 33 proteínas y 2755 estructuras. La recolección de estructuras por el método de web scraping consistió en la exploración de las bases de datos bibliográficas para recopilar artículos de investigación relacionados con los métodos de soaking y co-cristalización en solventes orgánicos y/o medios no acuosos. Luego, utilizando el método de minería de textos se analizaron los artículos y se relacionaron con las estructuras PDB correspondientes. Dichas estructuras fueron vinculadas con CoDNAS, para obtener el total de conformeros para cada proteína. Este conjunto de datos obtenido se consideró como "control" y todas sus tendencias se contrastaron contra el conjunto global de datos.

En ambos conjuntos de datos, se exploraron las conformaciones para detectar la presencia de ligandos unidos, con el fin de obtener pares de conformeros con y sin ligandos, y así evitar sesgos en el análisis de la diversidad conformacional, ya que, como bien se sabe, este es uno de los factores que pueden modificar la diversidad conformacional de las proteínas (Parisi et al., 2015). La presencia de ligandos unidos se evaluó utilizando la base de datos BioLiP (Yang et al., 2013), que proporciona una descripción detallada de las interacciones ligando-proteína asociados a cada estructura cristalográfica. Los datos se presentaron y analizaron teniendo en cuenta tres subgrupos de pares de conformeros:

- a. aquellos en los que ambos conformeros contenían cualquiera de los compuestos orgánicos comúnmente utilizados en las estimaciones de estructuras proteicas relevados en nuestra lista consenso (ver archivo Tabla suplementaria 1) o bien fueron estructuras obtenidas a partir de artículos de investigación relacionados con la co-cristalización y empapado en solventes orgánicos (OO);
- b. aquellos en los que no se encontró ningún solvente orgánico en el cristal (AA);

- c. aquellos en los que solo uno de ellos tenía las moléculas orgánicas en su cristal (AO).

Para cada proteína se consideró en cada grupo sólo el par que maximizara el RMSD entre los conformeros correspondientes (RMSD máximo). Por lo tanto, obtuvimos tres subgrupos de pares máximos para el conjunto grande de datos (AA, AO y OO con 9680, 1737, 2062 pares de conformeros, respectivamente) y tres subgrupos para el conjunto de datos de control (AA, AO y OO con 33, 31, 25 pares de conformeros, respectivamente) (Tabla 3.1), sobre los que se realizaron distintas estimaciones que permitieran contrastar las características estructurales propias de cada subconjunto de datos.

Set de Datos Subgrupo	Pares AA	Pares AO	Pares OO
Grande	9680	1737	2062
Control	33	31	25

Tabla 3.1 Cantidad de pares máximos reclutados por cada conjunto de datos. El conjunto de datos control, curado manualmente utilizando datos bibliográficos, contiene 33, 31, 25 pares de conformeros AA, AO y OO, respectivamente. El conjunto de datos grandes posee un total de 9680, 1737, 2062 pares de conformeros de AA, AO y OO, respectivamente.

3.2.b. Métodos para la estimación de las diferencias estructurales entre conformeros acuosos y no acuosos

Para estimar la disimilitud estructural entre los pares de conformeros clasificados en los subgrupos AA, AO, OO, utilizamos el RMSD de carbonos alfa calculado utilizando MAMMOTH (Ortiz et al., 2002). Este es un método heurístico para encontrar, por superposición rígida, el subconjunto máximo entre dos proteínas con el mismo backbone y conformación tridimensional. El algoritmo funciona descomponiendo la estructura de la proteína en péptidos cortos, de siete residuos, que se comparan con los heptapéptidos de la otra

proteína. Luego, mediante el cálculo del score de similitud entre los heptapéptidos de las dos estructuras, un vector unidad de RMSD, se construye la matriz de similitud con la que se calcula el alineamiento óptimo de cada residuo. Este programa, por tanto, permite obtener el RMSD entre dos estructuras, además de la cantidad de residuos alineados y el alineamiento secuencial a partir del estructural. Aplicando este análisis tanto a nuestros conjuntos de datos control y global se observó la misma tendencia en términos de RMSD del global; y por tanto, los restantes análisis estructurales fueron realizados solo en el conjunto global de datos.

Suponiendo que los confórmeros podían cambiar su superficie al introducirse la proteína en un ambiente apolar, decidimos realizar un análisis que nos permitiese caracterizar dichos cambios entre las conformaciones acuosas y orgánicas. La superficie de accesibilidad al solventes (ASA) es una medida del área de superficie de una biomolécula que es accesible a un solvente. Realizamos, entonces, cálculos de ASA para cada conformación de los tres subgrupos AA, AO y OO, utilizando NACCESS (Hubbard, 1992). Este método, basado en el algoritmo de Lee y Richards (Lee y Richards, 1971), basa su cálculo en el recorrido estimado de una molécula genérica de radio 1.4 Å (definido por el radio de Van der Waals de sus átomos), sobre la superficie de la proteína y cuya trayectoria, teniendo en cuenta el centro de la molécula, determinará la superficie accesible al solvente. El ASA global que devuelve el algoritmo corresponde a la suma de los valores absolutos de ASA de cada residuo y el ASA relativo se calcula para cada aminoácido en la proteína, expresando las diversas superficies accesibles a los residuos sumadas como un porcentaje del observado en un tripéptido ALA-X-ALA. Para obtener una medida de los movimientos de los aminoácidos, hemos calculado la cantidad de aminoácidos enterrados (los ASA inferiores al 25% se consideraron enterrados, y los ASA superiores al 25% se consideraron expuestos). Las comparaciones de ASA se calcularon como diferencias de ASA entre pares de confórmeros, para las tres poblaciones (AA, AO y OO).

Dado que una proteína en solución puede tener múltiples conformaciones, suelen utilizarse métodos probabilísticos para el cálculo de la geometría de la

misma. Si bien una parte de la caracterización de la diversidad de los conformeros se realizó utilizando el RMSD para los carbonos alfa, esta medida sólo es adecuada para estudiar variaciones a nivel de la cadena carbonada de la proteína. Por esta razón decidimos utilizar métodos adicionales que nos permitan profundizar la caracterización de la diversidad conformacional de las proteínas en el pasaje de un medio a otro. Siguiendo con nuestra hipótesis que los conformeros en solventes orgánicos podrían reducir su volumen, calculamos el radio de giro de las proteínas en agua y en medios no acuosos. Este es uno de los descriptores geométricos más utilizados para el cálculo de volumen, se define como la distancia descrita entre el centro de masa de cada residuo de la proteína, al centro de masa de la molécula entera. Para una mejor descripción geométrica de las conformaciones adoptadas por las proteínas en los distintos medios, se estimaron los radios de giro para todas las estructuras de PDB utilizando las herramientas MMTSB⁹. Asimismo se calculó el número de enlaces de hidrógenos de la proteína utilizamos HBPLUS (McDonald y Thornton, 1994), medida que permite tener una idea de la estabilidad de la proteína en un medio dado (Pace et al., 2014).

Como una medida de la restricción estructural de las proteínas en medios no acuosos, y a modo de comprender las diferencias funcionales que pudieran sufrir las proteínas en los distintos medios, decidimos comparar la presencia de regiones desordenadas en los conformeros orgánicos y acuosos. La presencia de regiones intrínsecamente desordenadas (IDPs) se han descrito la relevancia de las transiciones orden-desorden en la modulación de función proteica (Tompa et al., 2005; Zea et al., 2016), así como también se han relacionado con funciones de regulación y señalización celular (Gsponer y Babu, 2009; Wright y Dyson, 2015). El desorden está representado en una estructura cristalográfica por la ausencia de coordenadas de densidad electrónica (Gsponer y Babu, 2009; Tompa et al., 2015). En particular, para definir las regiones intrínsecamente desordenadas en nuestros conjuntos de datos, consideramos aquellos segmentos con cinco o más residuos faltantes consecutivos que no estaban en los extremos amino terminal o carboxilo

⁹ MMTSB: http://blue11.bch.msu.edu/mmtsb/Main_Page

terminal de la secuencia de la proteína. Es decir, que para el análisis se excluyeron los primeros y últimos 20 residuos de la cadena, de modo de evitar el sesgo por errores propios de la determinación cristalográfica (DeForte y Uversky, 2016).

Finalmente, la clase y la superfamilia se estudiaron vinculando las proteínas con la base de datos CATH (Sillitoe et al., 2015), una base de datos que identifica en las estructuras depositadas en PDB dominios estructurales relacionados evolutivamente. Esta vinculación se realizó a través de CoDNaS que posee asignado el/los código/s de la/s superfamilia/s de CATH a la que pertenecen.

Mediante el test de normalidad de D'Agostino y Pearson se confirmó que las distribuciones de los conjuntos de datos eran continuas y no paramétricas. Por este motivo, las comparaciones dentro de los grupos se realizaron mediante el test de Kolmogorov-Smirnov correspondiente. Asimismo, se utilizó ANOVA de una vía para las comparaciones multigrupo. En todos los casos, se tomó un valor de $p < 0,05$ para indicar significación estadística. Todos los datos obtenidos fueron recuperados y procesados utilizando códigos propios en el lenguaje de programación Python.

Un hecho bien establecido es que el tránsito del ligando desde el solvente hasta la cavidad que contiene el sitio activo o de unión posibilita la catálisis enzimática (Kingsley y Lill, 2015). Existen túneles que promueven el tránsito al interior de la proteína y la conexión entre cavidades y también facilitan la salida, como en el caso de los productos de enzimas. Sin embargo, múltiples trabajos científicos han mostrado que la disponibilidad de dichas estructuras es dinámica y puede variar de conformación en conformación (McGovern y Shoichet, 2003; Murga et al., 2008) y podrían, por tanto, dar una idea de los cambios estructurales que sufre una proteína. Nos propusimos, entonces, estudiar la diversidad conformacional de los subgrupos de transición (AA, AO y OO) debido a cambios en estructuras tales como túneles, cavidades y los sitios activos, con el objetivo de analizar los movimientos locales que pudiera sufrir la proteína en la transición de un medio a otro.

3.2.c.i Métodos para la estimación de la diversidad conformacional de túneles y cavidades

El cálculo del número de cavidades y túneles, así como sus propiedades, se estimaron para todos los confórmeros utilizando Fpocket (Le Guilloux et al., 2009) y MOLE (Sehna et al., 2013). Brevemente, tanto el Fpocket como el MOLE basan su algoritmo de cálculo en el concepto de esferas alfa para describir la superficie de una proteína (Edelsbrunner y Mücke, 1992). Esta se describe como una esfera que en su interior no encierra átomo alguno y que se encuentra en contacto con cuatro átomos en su superficie a una distancia r (radio de la esfera) al centro de la esfera alfa. El radio de esta esfera refleja, por lo tanto, la curvatura definida por los cuatro átomos que la rodean, y los algoritmos antes mencionados son capaces de encontrar las alfa esferas del radio que mejor representen las cavidades y túneles de la proteína, utilizando el método de teselación de Delaunay-Voronoi (Kim et al., 2008). Asimismo, permiten conocer el volumen y las propiedades fisicoquímicas de dichas cavidades, teniendo en cuenta la composición de aminoácidos que revisten sus superficies; estas propiedades fueron analizadas comparativamente para los distintos subconjuntos de datos.

3.2.c.ii Métodos para la estimación de la diversidad conformacional de sitios activos

Se estudió la diversidad conformacional de sitios activos en el conjunto de datos grande, teniendo en cuenta los residuos descritos en la base de datos de sitios activos Catalytic site Atlas (Furnham et al., 2014), para los que se calculó el RMSD por posición utilizando ProFit (McLachlan, 1982), un método iterativo rápido para calcular las rotaciones de cuerpo rígido que se necesitan para superponer y comparar dos estructuras de proteínas entre sí.

Para explorar las posibles transiciones de estructuras secundaria que pudieran sufrir la proteínas a nivel local, en los sitios activos, definimos la estructura

secundaria para cada cónfórmero utilizando DSSP (Zhang y Sagui, 2015). Este algoritmo predice estructura secundaria basándose en la identificación de los puentes de hidrógeno entre los grupos carbonilo y amida de la cadena principal. De este modo, el DSSP define la estructura secundaria en 8 estados, que para facilitar el análisis de nuestros datos fueron agrupados en tres, siguiendo la metodología propuesta en otros trabajos científicos (Linding et al., 2003): alfa hélices (H y G), láminas beta (E) y loops (T, S, B, I, ' ').

3.3. Análisis de las diferencias estructurales entre conformaciones acuosas y no acuosas

3.3.a. Comparación de la diversidad conformacional entre cónfórmeros acuosos y no acuosos

Con el fin de estudiar la diversidad conformacional de las proteínas que pasan de entornos no acuosos a acuosos, creamos dos conjuntos de datos de proteínas con diversidad conformacional determinada experimentalmente (ver sección 4.2.a). Los conjuntos de datos resultantes incluyen entradas de CoDNaS que poseen al menos dos estructuras de proteínas en entornos no acuosos y al menos otras dos estructuras obtenidas en medios acuosos, todas ellas para la misma secuencia (100% de identidad de secuencia global). Las diferentes estructuras de la misma proteína fueron tomadas como distintas conformaciones, que en CoDNaS se comparan estructuralmente utilizando RMSD (Monzon et al., 2016). Además, dado que uno de los principales factores que influye en el alcance de la diversidad conformacional es la presencia de ligandos (Zea et al., 2016), ambos pares de cónfórmeros seleccionados correspondían a sus formas no unidas y/o a su forma unida, intentando, de este modo, enfocar nuestro análisis en los cambios estructurales debidos únicamente a las transiciones del medio.

Finalmente, obtuvimos un número total de 1737 proteínas con cónfórmeros en ambos medios (acuoso y no acuoso) para el conjunto grande de datos, y 33

proteínas con confórmeros en ambos medios para el conjunto de datos control. Se contrastaron las tendencias de ambos conjuntos de datos, observando las distribuciones de RMSD para los pares máximos de proteínas cristalizadas en diferentes medios de la Fig. 3.1. Tanto para el control, como para el conjunto grande de datos, se pudo observar que los cambios conformacionales en las transiciones de un medio acuoso a uno acuoso (AA) y de un medio acuoso a uno no acuoso (AO) fueron estadísticamente superiores a los cambios observados para la transición de un medio no acuoso a otro medio no acuoso (OO). Los valores *p* para las comparación de distribuciones entre OO - AO y OO - AA fueron mucho menores a 0.001, mientras que las distribuciones de AO y AA no mostraron diferencias estadísticas significativas.

Las distribuciones de RMSD de los pares máximos mostró el mismo comportamiento en ambos conjuntos de datos: con un RMSD promedio para AA, AO y OO de 0.96, 0.97 y 0.71 Å para el conjunto de datos grande, y de 1.41, 1.02 y 0.50 Å en el conjunto de datos de control (Fig. 3.1 A y B, respectivamente).

Como vemos en la figura Fig. 3.1 C, si tomamos todos los confórmeros por proteína y no sólo los pares máximos, del conjunto de datos control, podemos observar que las proteínas en medios no acuosos no parecen explorar todo el espacio conformacional que exploran en el agua. Las distribuciones de RMSD de los confórmeros no acuosos se observan claramente restringidas a una región de alrededor de 0,5 Å, cercano al error cristalográfico (Burra et al., 2009). Es importante tener en cuenta que estas distribuciones de RMSD son independientes de la superfamilias a las que pertenecen las proteínas; ni el conjunto de datos grande, ni el control muestran un enriquecimiento diferencial en los grupos AA, AO y OO.

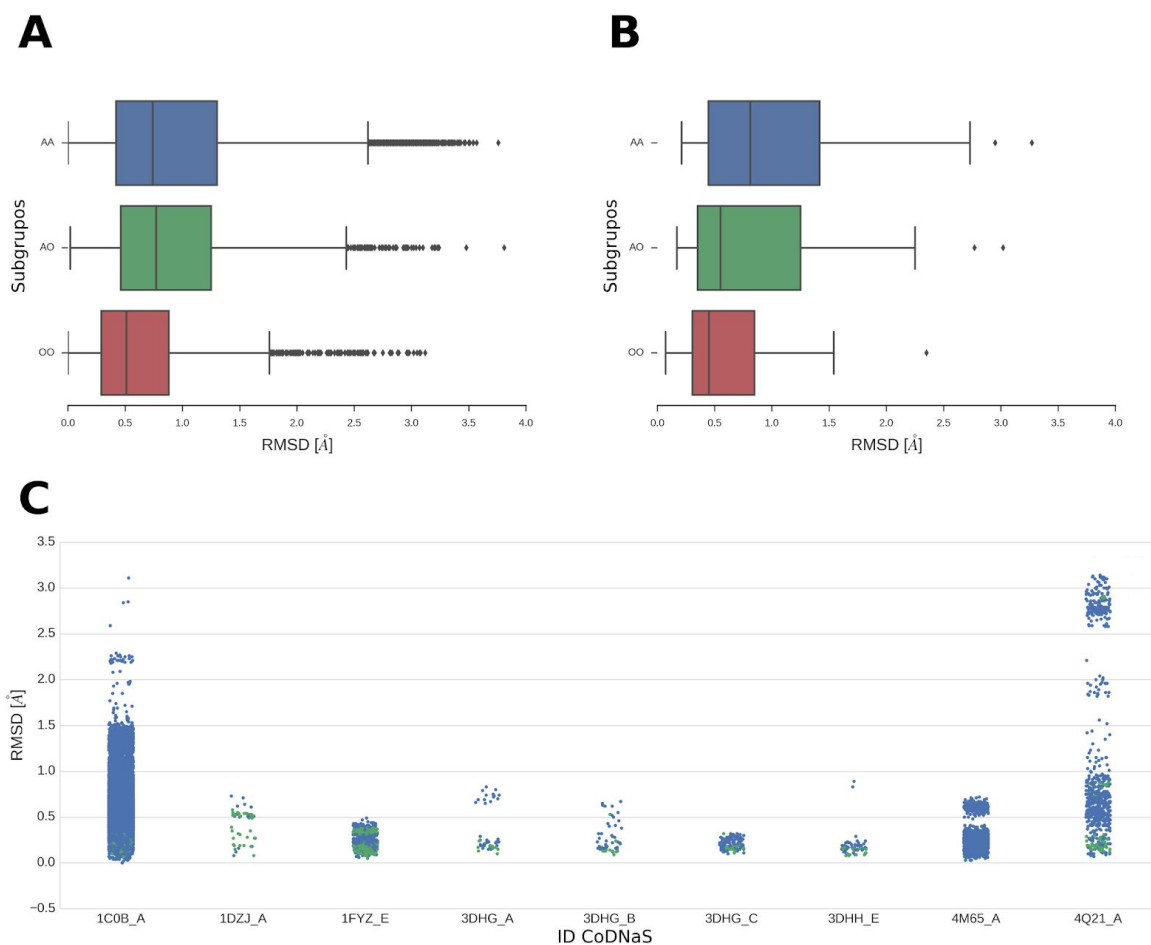


Fig. 3.1 A) Distribuciones de diversidad conformacional para los conjuntos de datos grande. Distribuciones de diversidad conformacional medida en RMSD (Å) para el conjunto grande de datos para los diferentes subgrupos AA (en azul), AO (en verde) y OO (en rojo). Es posible observar que los pares OO tienen una diversidad conformacional mucho menor que los pares de subgrupos AA y AO. Los RMSD promedios fueron de 0.97, 0.94 y 0.68 Å para AA, AO y OO, respectivamente (medianas observadas: 0.74, 0.77 y 0.51 Å, respectivamente). Los valores de p para las comparaciones entre OO - AO y OO - AA fueron <0.001 , mientras que las distribuciones de AO y AA no mostraron diferencias significativas. B) El conjunto de datos control muestra el mismo comportamiento que el conjunto de datos grande (1A). C) Distribuciones de diversidad conformacional para las proteínas más pobladas del conjunto de datos control, teniendo en cuenta todos los pares confórmicos. Distribuciones de diversidad conformacional de un conjunto representativo de proteínas del conjunto de datos de control medido en RMSD (Å), para los subgrupos AA y OO. Las transiciones de ambientes acuosos a acuosos (AA) se muestran en azul y la transición de ambientes no acuosos a no acuosos (OO) se muestran en verde.

Para explorar cómo los distintos solventes orgánicos pueden afectar la diversidad conformacional de las proteínas, en tanto a sus distintas propiedades fisicoquímicas, analizamos las distribuciones OO subdividiendo el conjunto de datos según el medio de cristalización utilizado. Se observó, entonces, que cuando los confórmeros orgánicos difieren en el medio de cristalización utilizado, el RMSD promedio es de 0.82 Å, mientras que cuando difieren en otras condiciones, como por ejemplo la presencia de modificaciones postraduccionales o diferencias en el estado oligomérico, el RMSD es de 0.63 Å. Dichos resultados muestran la gran influencia que puede tener el medio en la estructura proteica.

Para una mayor comprensión de estos cambios estructurales, analizamos en el conjunto grande de datos las diferencias en la estructura secundaria de los confórmeros del par que maximiza la diferencia conformacional. El RMSD promedio por sitio estimado para los bucles (loops), alfa hélices y láminas beta se muestra en la Tabla 3.2. Observamos que para todos los elementos de la estructura secundaria, los pares de confórmeros AO presentan la variación máxima. Como era de esperar, el valor máximo correspondió a la variación en los bucles debido a su flexibilidad intrínseca. Curiosamente, la variación en los pares de AO estuvo por encima inclusive del promedio de la variación estructural en los pares de AA, lo que posiblemente refleja que los confórmeros en ambientes no acuosos muestran algunas características estructurales únicas, distintas a las observadas en agua, lo que magnifica las diferencias cuando se comparan confórmeros provenientes de medios distintos. Analizando las transiciones entre elementos de estructura secundaria, como porcentajes de cambios entre un elemento y otro, no se encontraron cambios significativos en ninguno de los tres subgrupos.

Estructura Secundaria / Subgrupo	AA (Å)	AO (Å)	OO (Å)
Alfa hélice	0.78	0.91	0.62
Lámina beta	0.64	0.75	0.52
Loop	0.93	1.03	0.75

Tabla 3.2. RMSD promedio por estructura secundaria de los distintos subgrupos. Se puede observar, dado los resultados explicitados en la presente tabla, que la variación máxima en la diversidad conformacional correspondió a los bucles (loops), lo que resulta esperable debido a su flexibilidad intrínseca. Los pares de confórmeros AO presentan la variación máxima, para todos los elementos de la estructura secundaria, mostrando la diferencia marcada entre las conformaciones que pueden adoptar las proteínas en medio acuoso, respecto de sus correspondientes en un entorno orgánico.

Los cambios en la superficie accesible al solvente (ASA) entre los pares máximos de los distintos subgrupos siguieron la tendencia general mostrada para RMSD. Observamos que tanto la diferencia en el ASA global como en el ASA relativo son más bajas para los subgrupos OO (en promedio 310.76 Å^2 y 196.33 Å^2 respectivamente) y AA (en promedio 412.78 Å^2 y 261.39 Å^2 respectivamente), que las observadas para el subgrupo AO (en promedio 448.66 Å^2 y 285.51 Å^2) (Fig. 3.2 A y B). Esta misma tendencia de las distribuciones ASA globales y relativas del conjunto grande de datos, fue observada también en el conjunto de datos de control (Figura suplementaria 3S.1 A y B respectivamente). Estas observaciones podrían explicarse por el hecho que los confórmeros de los subgrupos OO y AA se obtuvieron en medios similares, y por tanto las conformaciones presentan una exposición al solvente similar; mientras que en el caso del subgrupo AO estamos observando transiciones de un medio acuoso a uno no acuoso, con sus consiguientes cambios estructurales, que implicarán consecuentemente una mayor diferencia entre confórmeros.

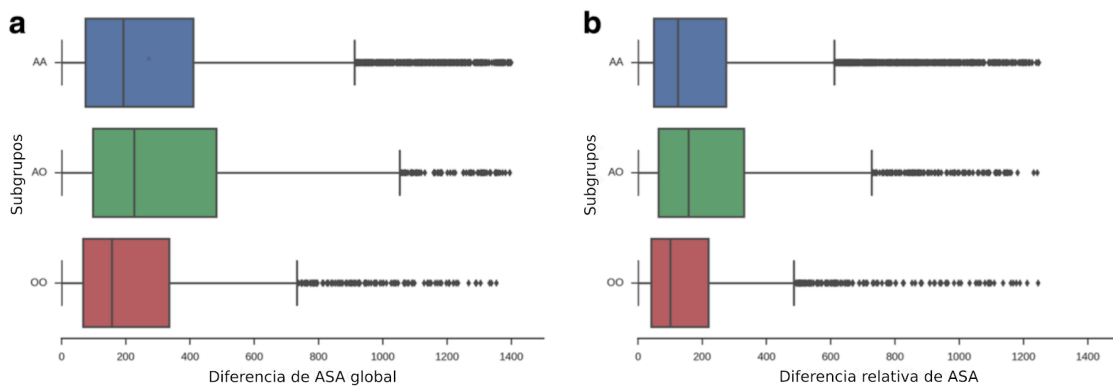


Fig. 3.2. A) Diferencias en el ASA global para los diferentes subgrupos. El subgrupo AO (en verde) muestra las diferencias máximas, evidenciando mayores cambios entre los confórmers obtenidos en diferentes solventes. Los promedios de las diferencias del ASA global fueron 412.78, 448.66 y 310.76 para AA , AO y OO respectivamente (medianas observadas: 237.67, 257.70 y 162.87, respectivamente). Los valores p para las comparaciones fueron en todos los casos <0.001 . B) Diferencias en el ASA relativo para los diferentes subgrupos. El subgrupo AO (en verde) muestra las diferencias máximas, en una tendencia similar a la observada para el ASA global. Los promedios de las diferencias de ASA relativas fueron 261.39, 285.51 y 196.33 para AA, AO y OO, respectivamente (medianas observadas: 154.30, 165.10 y 103.05, respectivamente). Los valores p para las comparaciones fueron en todos los casos <0.001 .

Para conocer más sobre la relación entre el movimiento de los residuos con estas observaciones, calculamos los porcentajes de aminoácidos enterrados (buried) en los confórmers acuosos y orgánicos de los tres subgrupos. Estos porcentajes muestran, como es de esperar, un mayor número de residuos enterrados en medios no acuosos: un 48.37% para los confórmers en solventes orgánicos, respecto de un 44.63% observado en agua. Es interesante observar que cuando se comparó la superficie accesible al solvente de todos los confórmers para cada proteína, encontramos que los confórmers obtenidos en medios no acuosos muestran valores en torno a la mitad de la distribución de la población acuosa. Este comportamiento (Figura Suplementaria 3S.2) muestra nuevamente la restricción en el paisaje conformacional que sufren las proteínas en ambientes no acuosos, donde las mismas tienden a no explorar conformaciones extremas como sus contrapartes acuosas.

Analizando las diferencias en la cantidad de puentes de hidrógeno entre los pares de todos los subgrupos, encontramos que nuevamente las diferencias principales se observaron en el subgrupo AO, reforzando la idea de una mayor heterogeneidad de las conformaciones acuosas comparadas con la población de orgánicas (Figura suplementaria 3S.3). Observamos en detalle el contenido promedio de puentes de hidrógeno en las conformaciones de ambas condiciones, agrupando todos los confórmeros acuosos (A) y los orgánicos (O) de cada par de los subgrupos. Encontramos que el promedio de puentes de hidrógeno en el subgrupo A es 721.18 mientras que para O es 847.87 (Figura suplementaria 3S.4), asociándose con una menor flexibilidad en los confórmeros orgánicos respecto de los acuosos. Finalmente, para tener una mejor idea del grado de compactación de la proteína en los distintos medios, se analizaron los radios de giro de los pares de confórmeros en todos los subgrupos. Se observó la misma tendencia general, con mayores diferencias en el subgrupo AO (Figura suplementaria 3S.5).

3.3.b. Diversidad conformacional en estructuras funcionalmente relacionadas

Estudiamos la diversidad conformacional debida a la transición de medio, evidenciable en cambios en estructuras funcionales proteicas, como túneles, cavidades y sitios activos. Como se describe en la sección 1.4.b, los túneles y las cavidades son estructuras funcionales que, interconectadas, permiten el tránsito de moléculas de solvente o ligandos desde la superficie de la proteína al sitio activo o sitios de unión. El cambio en dichas estructuras se estudió en el conjunto de proteínas con un sitio activo caracterizado disponibles en nuestro conjunto de datos. También hemos enfocado el análisis en evaluar la presencia de transiciones orden-desorden, debido a su importancia en la actividad biológica y su contribución a la diversidad conformacional (Monzon et al., 2017).

Las diferencias estructurales globales no se correlacionan con el comportamiento de los túneles, ya que no se encontraron diferencias entre los tres subgrupos (AA, AO y OO). Las diferencias en el número y longitud de los túneles en los distintos subgrupos no fueron estadísticamente significativas. Aún cuando la longitud, el número y la distribución de cavidades entre los distintos subconjuntos de datos fueron estadísticamente equivalentes,, sus flexibilidades y volúmenes mostraron diferencias. La flexibilidad media de las mismas (medida como el promedio de los factores B, de todos los átomos de la cavidad) fue menor en el subgrupo OO, en comparación con los subgrupos AA y AO, siendo la flexibilidad media de las cavidades 0.23, 0.25 y 0.30, respectivamente. Encontramos, también, que las cavidades son más grandes en las conformaciones provenientes de medios no acuosos que en sus correspondientes acuosos (volumen medio de cavidades totales 5969.58 y 5762.70 Å³, respectivamente). Esta tendencia fue la misma teniendo en cuenta tanto la cavidad de volumen máximo como el volumen total de todas las cavidades encontradas en un confórmero dado.

Usando los residuos caracterizados como necesarios para mantener la actividad enzimática, extraídos de la base de datos CSA (Furnham et al., 2014), pudimos analizar las diferencias estructurales entre los confórmeros en sus sitios activos. Utilizamos un total de 390 de pares máximos AA, 153 pares AO y 197 pares OO. Encontramos que el RMSD medio de los residuos del sitio activo, así como su ASA media, fueron significativamente más altos para AA y AO que los observados para OO (los valores P para las comparaciones de RMSD entre OO y AO, y entre OO y AA, fueron menores que 0.001; mientras que la comparación de las distribuciones AO y AA no mostró diferencias significativas) (Fig. 3.3. A y B).

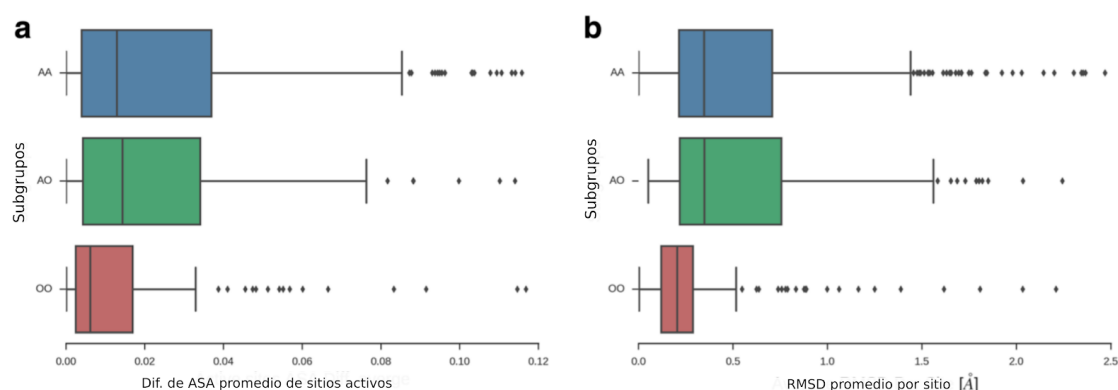


Fig. 3.3 A) Distribución del RMSD promedio para los residuos correspondientes al sitio activo para los diferentes subgrupos: AA (en azul), AO (en verde) y OO (en rojo). Los promedios de RMSD por sitio estimados para los residuos en el sitio activo fueron de 0.68, 0.68 y 0.32 Å para AA, AO y OO respectivamente (medianas observadas: 0.46, 0.41 y 0.20 Å, respectivamente). Los valores p para las comparaciones entre OO y AO y entre OO y AA fueron <0.001 , mientras que las distribuciones de AO y AA no mostraron diferencias significativas. B) Distribución de las diferencias promedio de ASA para los residuos correspondientes al sitio activo, para los diferentes subgrupos AA (en azul), AO (en verde) y OO (en rojo). Las diferencias promedio de ASA en los sitios activos fueron 0.044, 0.043 y 0.019 para AA, AO y OO respectivamente (medianas observadas 0.028, 0.017 y 0.01 respectivamente). Los valores p para comparaciones de diferencias de ASA relativas fueron en todos los casos <0.001 .

Finalmente, después del análisis de la flexibilidad de la proteína (van der Lee et al., 2014), cuantificamos las diferencias en las regiones desordenadas y los residuos faltantes (missing) para los confórmers en cada subgrupo (ver sección 4.2.c). Observamos las mayores diferencias en el subgrupo AO (promedio de 0.67 y 0.03 respectivamente) y las más bajas en el subgrupo OO (promedio de 0.31 y 0.015). Además, los promedios son los más bajos entre los confórmers no acuosos, indicando que las transiciones de orden-desorden están muy afectadas por la presencia de un medio no acuoso.

3.3.c. Ejemplos de interés biológico

Una de las principales conclusiones de nuestro trabajo es que las proteínas en solventes acuosos muestran mayor diversidad conformacional, medida en RMSD, que aquellas en solventes no acuosos. Un ejemplo que muestra este comportamiento está representado por la proteína Ras humana. La proteína Ras pertenece a una gran superfamilia de proteínas conocidas como proteínas G con actividad GTPasa (Bourne et al., 1990). La mayoría de los miembros de esta superfamilia se encuentran unidos a membrana plasmática (Colicelli, 2004). Pueden ser activados por diferentes factores de intercambio de nucleótidos de guanina (GEF), asociados a distintos efectores, y desactivados por la proteína activadora de GTPasas (GAPs) (Cherfils y Zeghouf, 2013). Estas proteínas participan en vías de transducción de señales asociadas a la regulación que controlan diferentes procesos: integridad del citoesqueleto, proliferación, crecimiento celular, diferenciación, adhesión y migración celular y apoptosis; por tanto, su desregulación se encuentra asociada al cáncer (Konstantinopoulos et al., 2007; Oxford, 2003).

Las proteínas Ras funcionan así como interruptores binarios que alternan, durante la transducción de señales, entre los estados de encendido y apagado, según se encuentren unidos a guanosina trifosfato (GTP) o guanosina difosfato (GDP) respectivamente. Estos dos estados poseen distintas características estructurales. Así, el estado nativo de Ras se encuentra descrito por dos formas principales, que se corresponden con las conformaciones inactivas y activas (o 1 y 2, respectivamente) (Shima et al., 2010). Cuando Ras es activada, activa otras proteínas, que en última instancia permiten la expresión de genes implicados en el crecimiento, la diferenciación y la supervivencia celular. La estructura del estado 1 de Ras se distingue del estado 2 por la pérdida de las interacciones de un residuo de treonina (Thr 35) con el fosfato de GTP. Este produce un alejamiento del bucle del interruptor I (residuos 30-40) del nucleótido guanina, lo que induce una conformación inestable y flexible del bucle. Además, un residuo tirosina (Tyr 64) ubicado en la región llamada interruptor II (residuos 60-76), en el estado 1 está demasiado lejos

para ejercer un efecto significativo sobre el gamma-fosfato del GTP a hidrolizar (Fig. 3.4 panel a, ID de PDB 1xd2 y 1ctq).

En CoDNAS, la proteína Ras humana (Uniprot ID P01112) tiene 99 conformeros anotados. El par máximo AA mostró un RMSD de 3.14 Å y el AO 3.01 Å, mientras que el par máximo OO mostró el mínimo RMSD (de 0.82 Å). La misma tendencia se observó para el ASA. Estos resultados reflejan la tendencia ya observada en la Fig. 3.1 para el conjuntos de datos control y global. En la Fig. 3.4 también mostramos en los paneles b, c y d, las representaciones de los pares AA, AO y OO que evidencian las restricciones conformacionales en la diversidad conformacional del par OO. El cambio conformacional acompañado por transiciones de orden-desorden en la proteína Ras también fue descrito por Buhrman et al. (Buhrman et al., 2003). En dicho trabajo estudiaron cómo los solventes orgánicos favorecían las transiciones de segmentos desordenados a ordenados de la proteína Ras, principalmente en la región del interruptor II (Fig. 3.4). Además, coincidentemente con nuestro resultado, los conformeros no acuosos de la proteína Ras presentan menores proporciones en regiones desordenadas. Los solventes hidrófobos podrían favorecer el orden de transiciones de regiones desordenadas cortas en las proteínas, ya que favorecen las interacciones de unión a H entre grupos que son altamente solvatados y móviles en soluciones acuosas. Nuestros resultados muestran que los enlaces de hidrógeno son más numerosos en los conformeros no acuosos, una tendencia que es también observada en los conformeros Ras.

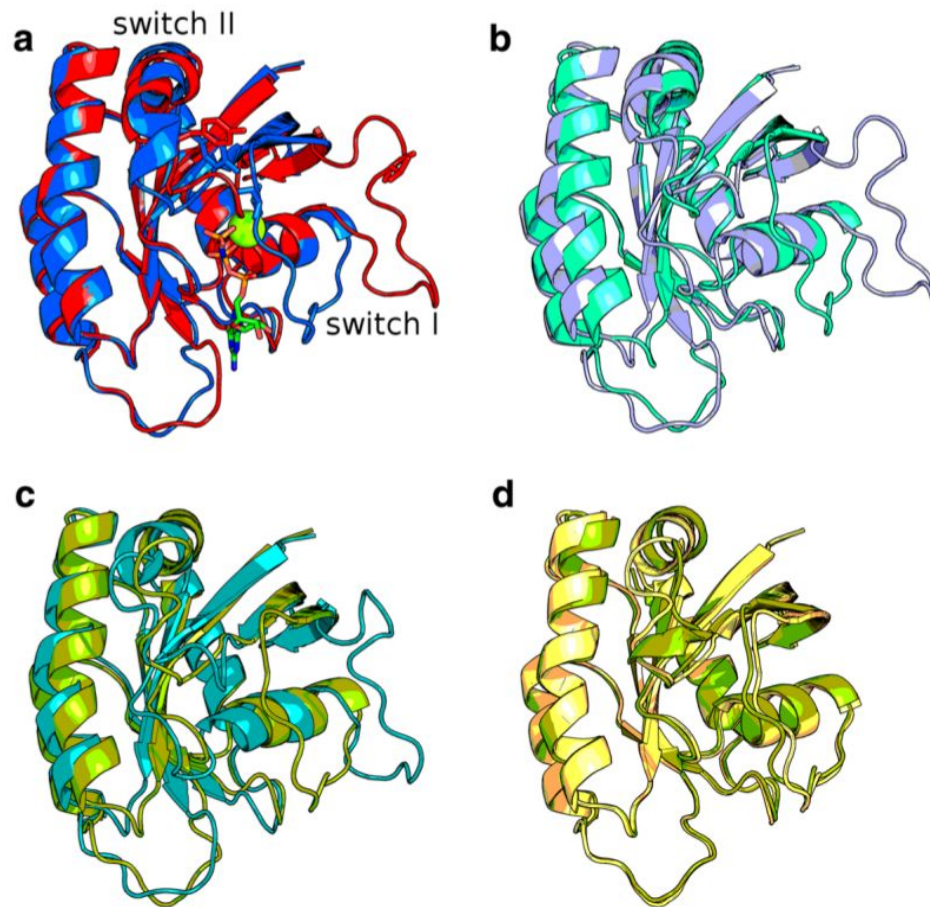


Fig. 3.4. A) Representación estructural de los conformeros de proteínas Ras y todos los componentes esenciales para la actividad de Ras. Se muestra en la figura una representación del estado 1 (rojo, ID de PDB 1x2d_B) y del estado 2 (azul, ID de PDB 1ctq_A) de los conformeros (inactivos y activos respectivamente) de la proteína Ras humana. También, se encuentran representados en los componentes relevantes para la actividad de Ras: el Mg^{++} , los ligandos unidos a GTP, el interruptor I (Thr 35) y el interruptor II (Tyr 64). B) Representación del par máximo de RMSD AA correspondiente a Ras. Se muestran las estructuras cristalinas correspondientes al par máximo, 1xd2_B (púrpura claro) y 4dls_A (cian), que se corresponden al estado 1 y al estado 2 respectivamente. C) Representación de del par máximo de AO de RMSD, donde 1p2s_A (verde claro) y 4nym_R (cian) muestran el estado 1 y el estado 2 respectivamente. D) Representación del par de RMSD máximo OO que muestra 1p2s_A (verde claro) y 3rs5_A (naranja claro) que muestra ambas estructuras en el estado 2. 1p2s_A se resolvió en trifluoroetanol al 50% y 3rs5_A en dimetilformamida al 55%.

3.4. Conclusiones destacadas

Nuestros resultados enfatizan el hecho de que las proteínas en ambientes no acuosos son más rígidas, como lo han demostrado muchos estudios anteriores (Eppler et al., 2006; Klibanov, 2001). Este hallazgo se observa en la distribución OO de RMSD, en comparación con las distribuciones AA y AO, que están ligeramente por encima del rango del error cristalográfico (~ 0.4 Å) (Berman et al., 2000). Aparentemente, nuestros resultados apuntan a que las diferentes estructuras de la misma proteína son casi idénticas en medios no acuosos independientemente de su estado unido o no unido (RMSD OO promedio = 0.68 Å). Sin embargo, bajo la hipótesis del atrapamiento cinético, las proteínas en solventes orgánicos conservarían la misma estructura que tienen en medios acuosos (Klibanov, 2001; Mattos y Ringe, 2001) y, en términos de nuestro conjunto de datos, la distribución de OO debería mostrar casi la misma distribución RMSD que la distribución AA (Fig. 3.1). Se observa el mismo comportamiento para ASA absoluto y relativo (Fig. 3.2 A y B) y cambios en diferentes elementos de la estructura secundaria, donde AO exhibe la mayor variación (Tabla 3.2) en comparación con las distribuciones AA y OO. Aparentemente, los conformeros obtenidos en medios no acuosos cambian a ciertas conformaciones evitando la adopción de conformaciones extremas (apertura/cierre completo), que sí pueden adoptar los conformeros acuosos como se deriva de las distribuciones de ASA (Fig. 3S.2).

Sin embargo, estas diferencias estructurales globales no se correlacionan con el comportamiento de los túneles, donde no se encontraron diferencias entre los tres subgrupos. El número y la longitud de los túneles no muestran diferencias entre los conformeros A y O. Sin embargo, es interesante observar que nuestros resultados muestran que los volúmenes de cavidades son mayores en conformeros O que en conformeros A. Las cavidades que se encuentran normalmente en las proteínas se asocian generalmente con sitios activos de enzimas o sitios de unión en proteínas transportadoras (Liang et al., 1998). Se ha demostrado que a medida que las cavidades no polares se hacen más grandes, se alcanza un punto en que resulta favorable la unión de un

grupo de moléculas de solvente que interactúan entre sí mediante puentes de hidrógeno, y que serían capaces de conferir estabilidad a dicha cavidad (Matthews y Liu, 2009). El aumento del volumen de las cavidades en las proteínas en medios orgánicos se mostró que está relacionada con la entrada de moléculas de solventes orgánicos a las sub estructuras proteicas, sin cambios adicionales en la topología general de la proteína (Stepankova et al., 2013), un hallazgo que va en consonancia con nuestros resultados.

Nuestros hallazgos sugieren algunas discrepancias con las predicciones hechas en base a la hipótesis del atrapamiento cinético (kinetic trapping), ya que encontramos que los confórmeros en medios no acuosos tienen una diversidad conformacional mucho menor que sus contrapartes en medios acuosos. Asimismo, observamos que los confórmeros en medios no acuosos tienen cavidades más grandes, menos superficie expuesta a solventes y menos regiones desordenadas, en contraposición con los confórmeros acuosos. Todas estas parecen ser características propias de un paisaje conformacional restringido respecto del que se observa en agua, que se contrapone a la idea de "conservación estructural" y conservación del estado nativo que propone la teoría del atrapamiento cinético.

Dado que el dinamismo de las proteínas es una característica clave para mantener la función biológica (Callender y Dyer, 2015), así como para garantizar la preservación y el comportamiento dinámico de las cavidades y bolsillos (Desdouits et al., 2015) y las transiciones de orden-desorden (Zea et al., 2016), las características específicas descritas anteriormente para los confórmeros en medios orgánicos podrían contribuir a explicar los cambios en la actividad biológica de las proteínas observados en dichos entornos, como la menor actividad de las reacciones canónicas y/o la adquisición de nuevas funcionalidades.

Estudio de la actividad promiscua en un sistema específico: promiscuidad de albúminas séricas

Un caso interesante de promiscuidad catalítica es el de las albúminas séricas. Varios miembros de esta familia han mostrado diferentes actividades promiscuas, pero desconocemos su importancia en un contexto biológico. En un trabajo en colaboración con el grupo de Biocatálisis de la Universidad Nacional de Quilmes, nos enfocamos en estudiar el comportamiento promiscuo de la seroalbúmina bovina. Describimos las bases estructurales y evolutivas que dan sustento a su actividad catalítica en la reacción de condensación aldólica cruzada. No solo hemos podido identificar la cavidad que contiene el sitio catalítico probable, sino también explicar sobre la base de las características de dicho bolsillo y de la interacción proteína-ligando, las diferencias observadas al utilizar distintos sustratos para su actividad catalítica promiscua.

En este capítulo extendemos nuestro análisis a otros miembros de la familia de la albúmina sérica para estudiar si estos comportamientos promiscuos están implicados en un proceso biológico. Utilizando métodos computacionales de inferencia evolutiva y análisis estructural, analizamos las bases evolutivas del comportamiento promiscuo en la familia. Nuestros resultados ofrecen una explicación mecanicista de dicho comportamiento y muestran que las posiciones responsables de la catálisis evolucionan bajo la presión selectiva típicamente observada durante la divergencia funcional. Según esta visión, las propiedades catalíticas observadas en las albúminas séricas podrían estar involucradas en un proceso biológico desconocido, más que representar un comportamiento promiscuo. Estos resultados abren preguntas

interesantes sobre el origen del comportamiento promiscuo y su evolución¹⁰.

¹⁰ Parte del contenido del presente capítulo se basa en las publicaciones de la autora: "A Mild Procedure for Enone reparation Catalysed by Bovine Serum Albumin in a Green and Easily Available Medium". SM Ardanaz, AJV Rueda, G Parisi, AM Iribarren, LE Iglesias, Catalysis Letters, 2018,148 (6), 1750-1757; y "Bovine serum albumin-catalysed cross aldol condensation: Influence of ketone structure". SM Ardanaz, EL Borucki, AJV Rueda, G Parisi, AM Iribarren, LE Iglesias - Process Biochemistry, 2019. DOI: <https://doi.org/10.1016/j.procbio.2019.08.003>.

4.1 Caracterización estructural de albúminas séricas

Las albúminas séricas (AS) pertenecen a una familia multigénica que incluye la alfa-fetoproteína, alfa-albúmina y la proteína de unión a vitamina D (Peters, 1995). Estas proteínas monoméricas presentan tres dominios alfa-helicoidales homólogos, cada uno de los cuales se divide en subdominios (A y B) (Fig 4.1 A). Las albúminas séricas son estructuralmente conservadas (Peters, 1977; 1995) y de gran tamaño relativo (de alrededor de 67 kDa de peso molecular). Son proteínas con alta solubilidad, lo cual les confiere un rol determinante en la regulación de la presión coloidosmótica de la sangre. También es de gran relevancia su rol en el transporte y unión de diversos ligandos, que incluyen desde ácidos grasos hasta hormonas, minerales y múltiples drogas (Chi y Liu, 2011; Fasano et al., 2005; King y Spencer, 1970; Kosa et al., 1997; Sułkowska, 2002). Esta gran versatilidad de unión de las AS proporciona una amplia gama de posibles aplicaciones farmacéuticas, bioquímicas y biotecnológicas.

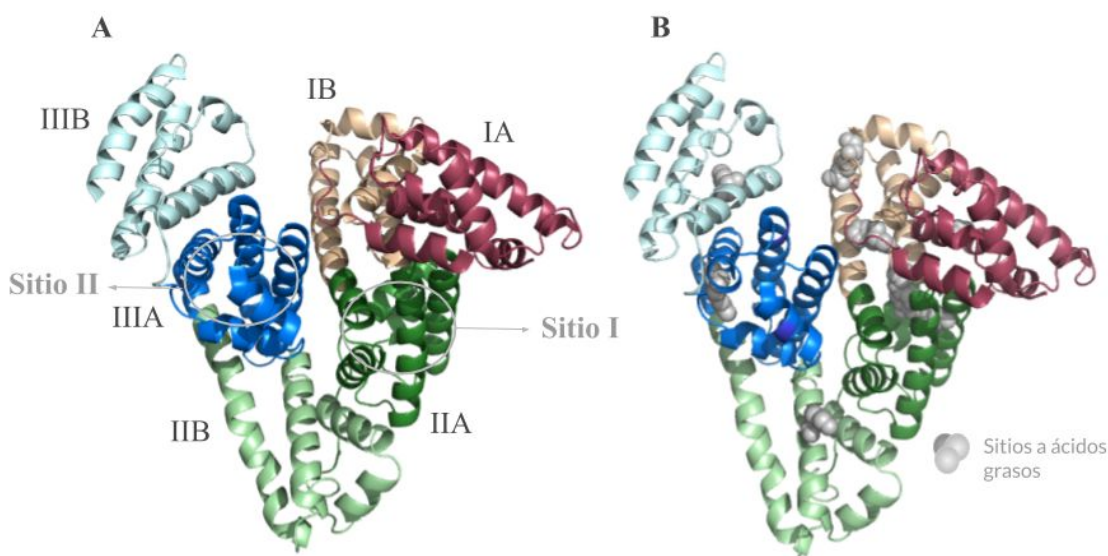


Fig 4.1. A) Subdivisión de dominios de albúminas séricas del I al III y subdominios (A y B), ejemplificado sobre la seroalbúmina humana (ASH). Subdivisión de la albúmina sérica humana en dominios y subdominios, ejemplificados sobre su estructura cristalográfica (PDB ID 1AO6); también se muestran las ubicaciones aproximadas del sitio I y del sitio II de unión a drogas. B) Sitios de unión a ácidos grasos no esterificados y saturados descritos para

seroalbúmina humana (ASH), ejemplificados sobre la estructura cristalográfica de albúmina humana (PDB ID 1AO6).

La seroalbúmina humana (ASH) tiene múltiples sitios de unión bien caracterizados para ácidos grasos no esterificados y para muchos de los ácidos grasos saturados. Mediante estudios cristalográficos se han identificado en la ASH un total de siete sitios de unión a ácidos grasos de cadena media y larga, aunque se han podido identificar también sitios adicionales de unión a ácidos grasos de cadena media en la proteína (Bhattacharya et al., 2000). Estos consisten en un sitio único en el subdominio IB, uno en el subdominio IIIB, un sitio de unión en la interfaz entre los subdominios IA y IIA y múltiples sitios en IIIA (Fig 4.1 B). De este modo, se pudo observar que los principales sitios de unión a drogas, en los subdominios IIA y IIIA, son capaces también de unir ácidos grasos. La cristalización de otras AS de mamíferos permitió identificar sitios de unión a drogas análogos a los observados primeramente en la albúmina humana (Ho et al., 1993; Panjehshahin et al., 1992; Sułkowska, 2002). Y en algunos casos se observó también la unión de ácidos grasos a dichos sitios, con afinidades similares a las observadas en la HSA (Cistola et al., 1987; Spector et al., 1969).

En particular, la albúmina sérica bovina (ASB) posee 8 conformeros en CoDNAS, con un RMSD promedio de 0.76 Å. La predicción de cavidades, hecha con Fpocket (ver descripción del método en sección 4.2.c), arroja un total de 24 cavidades en promedio para dichos conformeros. En contraste con la AS humana, la bovina presenta algunas similitudes en los sitios de unión de los subdominios IA y IIA (Sułkowska, 2002). Estudios de espectros de emisión de fluorescencia en presencia y ausencia de distintos ligandos mostraron que los contactos hidrofóbicos con Leu 203, Phe 211 y Ala 215 y las interacciones electrostáticas con Lys 199 y/o Arg 222 pueden desempeñar un papel clave en la formación de complejos con los mismos. Se ha mostrado el rol de dichos residuos en la catálisis de diversos sustratos y con múltiples mecanismos, tanto para la albúmina humana como para la albúmina bovina (Boucher et al., 2005; Díaz et al., 2001; Fleury et al., 1997).

Asimismo, se han caracterizado inmunológicamente y estructuralmente las AS, mediante estudios de inmunogenicidad, y se encontró un alto grado de reactividad cruzada de los anticuerpos anti-AS (Boutin et al., 1988; Goubran Botros et al., 1996; Hilger et al., 1997; Spitzauer, 1999), inclusive para otras familias proteicas de mamíferos. Dado que se encuentran presentes en los distintos alimentos, y muchas de ellas son también importantes alérgenos, poseen la capacidad de sensibilizar a los individuos predispuestos. Dicha propiedad provee posibles ventajas para su uso clínico en nuevas formas de diagnóstico y tratamiento de alergias a proteínas de mamíferos y otras patologías (Kosa et al., 1997; Varshney et al., 2010).

Dada la conservación estructural y secuencial de las albúminas séricas a lo largo de la familia, las analogías en las cavidades y sitios de unión, resulta esperable, aunque también resulta una incógnita a resolver, la explicación evolutiva de la existencia de dichas similitudes. En el presente capítulo nos proponemos abordar este interrogante desde múltiples perspectivas, tanto experimentales como bioinformáticas, buscando comprender desde las particularidades estructurales y evolutivas del caso de promiscuidad catalítica observado en ASB y extender las observaciones a otros sistemas promiscuos.

4.2. Promiscuidad catalítica en la albúmina sérica bovina

Los procedimientos biocatalíticos se integran actualmente en la química sintética porque proporcionan transformaciones sostenibles y respetuosas con el medio ambiente, permitiendo además ser realizadas en condiciones de reacción moderadas; además, la alta selectividad (regio-, quimio- y estereoselectividad) característica de las enzimas brinda alternativas ventajosas con respecto a numerosas estrategias tradicionales de síntesis orgánica (DeSantis y Davis, 2006; Liese y Filho, 1999). En este sentido, la promiscuidad catalítica ha ampliado las oportunidades sintéticas de los biocatalizadores (Arora et al., 2014). Por ejemplo, varias hidrolasas,

principalmente lipasas y proteasas disponibles comercialmente, han exhibido actividades promiscuas (Busto et al., 2010; López-Iglesias et al., 2012). Sin embargo, además de las enzimas, las proteínas no enzimáticas como la albúmina también pueden promover reacciones químicas.

Cómo ya hemos mencionado anteriormente, las albúminas séricas han sido ensayadas en diferentes tipos de reacciones para las que se encontró que poseían actividad catalítica promiscua (Albanese y Gaggero, 2015; Ardanaz et al., 2018; Reetz et al., 2007). En particular, en la albúmina humana su actividad promiscua también ha sido estudiada (Albanese y Gaggero 2015; Kragh-Hansen 2013). Su actividad tipo esterasa, por ejemplo, ha sido relacionada con el metabolismo de la aspirina, mediante una vía de reacción que involucra la transferencia del grupo acetilo de la misma a Lys 199 de la ASH, para luego ser hidrolizada a ácido salicílico por la ASH. Se estima que este fenómeno podría ser significativo en el síndrome de resistencia a la aspirina caracterizado por asma (Yang et al., 2007), aunque aún no se conocen completamente las consecuencias biológicas de dicho fenómeno. Particularmente la ASB disponible en el mercado se ha definido como un biocatalizador promiscuo, capaz de catalizar la reducción de cetonas, oxidación de sulfuros y reacciones de formación de enlaces carbono-carbono, como la de Morita-Baylis-Hillman (Benedetti et al., 2011; Sharma et al., 2013).

Las características de los medios y sustratos donde las AS son capaces de actuar resultan de lo más variados (Ver Tabla 4.1) y es precisamente dicha versatilidad catalítica la que les confiere su gran potencial como biocatalizadores. En el presente trabajo se intentarán comprender las bases estructurales y evolutivas que explican la acción de la BSA como biocatalizador, mediante el estudio de una reacción en particular catalizada por dicha proteína, y extender las observaciones a toda la familia proteica.

Proteína	Reacciones canónicas	Reacciones promiscuas	Referencias
Albúmina sérica	Ninguna	condensación aldólica	Ardanaz et al., 2019

bovina		eliminación de Kemp	Boucher et al. 2005
		reacción de Biginelli	Sharma et al. 2013
		reacción de Henry	Kühbeck et al. 2013
Albúmina sérica humana	Ninguna	actividad aril-acilamidasa	Masson et al. 2007
		actividad esterasa	Sakurai et al. 2004
		actividad tioesterasa	Kragh-Hansen 2013
		actividad peroxidasa	Hurst et al. 1999
		eliminación de Kemp	Boucher et al. 2005

Tabla 4.1 Resumen de actividades promiscuas de las albúminas séricas humana y bovina, documentadas en la base de datos ProtMiscuity.

4.2.a. Métodos computacionales para el estudio de las bases estructurales del mecanismo de acción la seroalbúmina bovina en la reacción de condensación aldólica cruzada

Como se indicó anteriormente, en una primera aproximación a los casos particulares de promiscuidad se abordó el caso de la ASB. El grupo de Biocatálisis y Biotransformaciones de la Universidad Nacional de Quilmes a cargo del Dr. Luis Emilio Iglesias ha encontrado que la seroalbúmina bovina en medio acuoso o alcohólico es capaz de catalizar la condensación aldólica, una importante reacción de formación de nuevos enlaces carbono-carbono (Ardanaz et al., 2018). En la publicación *“A Mild Procedure for Enone Preparation Catalysed by Bovine Serum Albumin in a Green and Easily Available Medium”* se describe un procedimiento para obtener cetonas α,β -insaturadas (benzalacetonas) a partir de acetona y un conjunto variado de benzaldehídos, en agua o etanol, y catálisis de ASB. Los resultados para cada sustrato ensayado fueron descritos en dicho trabajo y muestran algunas diferencias notorias en el rendimiento de los productos finales resultantes, las cetonas insaturadas. En particular, se observa que los sustratos nitrados

producen una mezcla de cetoles (productos precursores de las cetonas insaturadas) y cetonas insaturadas, en la que los cetoles son mayoritarios, mientras que los sustratos no nitrados sí dan las benzalacetonas correspondientes con muy alta conversión a estos productos.

Para lograr una mejor comprensión de las bases estructurales de dichas reacciones, comenzamos una colaboración con el grupo del Dr. Iglesias. Para tal fin, aplicamos métodos de docking molecular para estimar las energías de unión de los sustratos y productos, y de este modo predecir posibles sitios de interacción con la proteína y las características estructurales de los mismos. El método de docking molecular se llevó a cabo mediante el uso del AutoDock Vina (Trott y Olson, 2010), un programa de acceso libre.

De forma sucinta podemos explicar el método de docking molecular o acoplamiento molecular como la utilización de simulaciones computacionales para predecir la conformación de un complejo receptor-ligando. Dado que el ligando puede adoptar distintas conformaciones dentro de un sitio de unión del receptor (Stockwell y Thornton, 2006), en general las herramientas de docking implementan un proceso en dos etapas: una primera etapa ejecuta un algoritmo de muestreo de conformaciones estructurales posibles para el complejo ligando-receptor, y una segunda etapa ejecuta un algoritmo de puntuación que evalúa energéticamente dichas conformaciones (Cross et al., 2009; Morris y Lim-Wilby, 2008). En particular, el AutoDock Vina es un método que considera al ligando flexible y al receptor rígido, que asume el modelo llave cerradura y que enfoca el análisis en la búsqueda de las conformaciones óptimas del ligando. El método entonces requiere de una estructura cristalográfica, tanto de los ligandos como de los receptores, y una región estimada de unión sobre la cual se realiza el análisis. Este software genera las conformaciones del ligando *in situ*, en las que se adapta de forma continua al sitio de unión de la proteína diana.

Para nuestro sistema, se tomó como referencia una conformación abierta de la ASB (ID: 4Or0), complejada con naproxeno, disponible en la base de datos PDB. Para una mejor comprensión de las particularidades estructurales de la proteína, y en particular de su supuesto sitio activo, se estudiaron además las

cavidades y túneles de la misma y se analizaron sus propiedades fisicoquímicas mediante el uso del software Fpocket (Le Guilloux et al., 2009). Asimismo, se analizó el número de contactos que establece la proteína con los sustratos utilizando el software RING (Piovesan et al., 2016), disponible para su ejecución en el servidor de la Universidad de Padua¹¹. Se estimaron, también, los valores de pKa de los residuos ionizables de la ASB, utilizando el software PROPKA (Olsson et al., 2011).

Todos los datos se manipularon y analizaron mediante la ejecución de scripts propios desarrollados en el lenguaje de programación Python.

4.2.b. Resultados de la caracterización estructural de la actividad de la seroalbúmina bovina como biocatalizador de la condensación aldólica

De la exploración de subestructuras presentes en la ASB, encontramos una cavidad de volumen (de 2849.9 Å³) y polaridad máxima, que coincide aproximadamente con el sitio de unión IIA de la albúmina bovina. Es interesante destacar que en esta cavidad encontramos una lisina anormalmente ácida (Lys 221 según nuestra numeración PDB) de pKa = 7.7, calculado mediante titulación teórica con el PROPKA (recordemos que el pKa de la Lys en agua es aproximadamente 10). La cavidad mayor antes mencionada se fusiona con una más pequeña (de volumen de 1127.1 Å³), para la que observó una alta hidrofobicidad según el Fpocket, y que se encuentra conectada a la superficie de la proteína a través de un gran túnel. Estas características de la proteína nos dan la idea de que posiblemente este túnel permita el tránsito de sustratos y/o productos entre la superficie de la proteína y la mencionada cavidad.

Si bien los resultados del docking no arrojaron diferencias concluyentes en las energías de unión de los distintos sustratos a la cavidad, pudimos obtener la orientación estimada de los ligandos en la cavidad de interés. Los mismos

¹¹ RING: <http://protein.bio.unipd.it/ring/>

mostraron un promedio de 28 contactos interatómicos y de Van der Waals con los residuos de dicha cavidad, siendo los residuos más cercanos la Arg 256, la Arg 217, la Lys 221, el Glu 292 y la Arg 199 (Fig. 4.2). Curiosamente, sólo aquellos sustratos que la ASB no pudo convertir con altos rendimientos a cetonas insaturadas (los nitrobenzaldehídos) mostraron ausencia de contactos con los residuos de las cavidades.

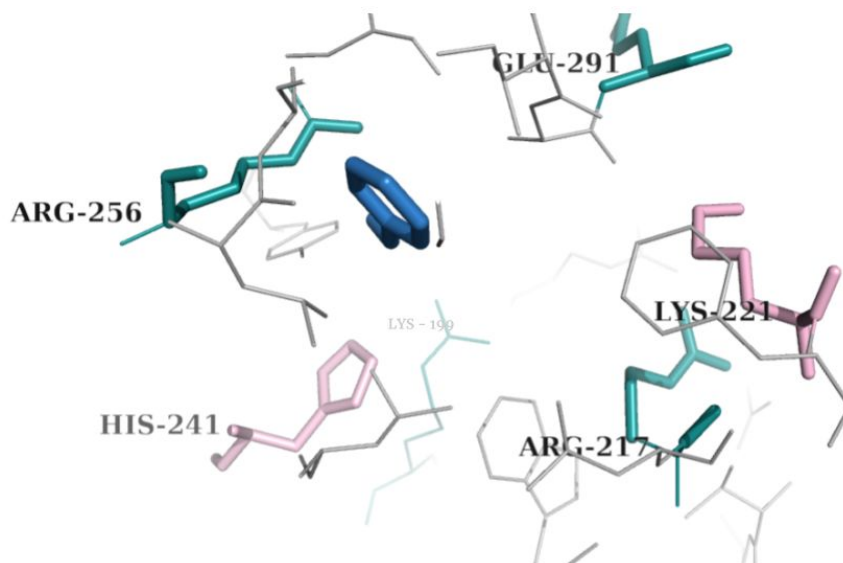


Fig. 4.2. Ejemplo de un sustrato de la reacción de condensación aldólica en la cavidad correspondiente al sitio de unión a drogas IIA de la ASB. En azul se muestra al benzaldehído unido a la cavidad principal del sitio de unión IIA de ASB (PDB ID 4Or0). Los residuos de la cavidad en contacto directo con los sustratos, que incluyen Arg 256, Arg 217, Glu 292 y Arg 199, se muestran resaltados. La Lys 222 (Lys 221 según la numeración de nuestro PDB) anormalmente ácida ($pK_a = 7.7$) y la His 242 también ácida se muestran coloreadas en rosa, mientras que los residuos básicos se muestran coloreados en verde.

Por otra parte, en el laboratorio de Biocatálisis se extendió el alcance estructural de la condensación aldólica catalizada por ASB, ensayándose diferentes cetonas y obteniéndose las correspondientes cetonas α,β -insaturadas, con rendimientos altos. Para tener un enfoque mecanístico de estas biotransformaciones, se estudió la cinética de la reacción entre ciclohexanona y *p*-nitrobenzaldehído en etanol catalizada por ASB y se propuso un mecanismo de reacción para dicha reacción, en colaboración con el grupo del Dr. Luis Emilio Iglesias. En base a las medidas de V_{max} para la

reacción de condensación aldólica estudiada y teniendo en cuenta que un trabajo anterior (Benedetti et al., 2011) postula que la Lys 222 participa en la formación de una enamina intermediaria, proponemos un mecanismo bi-bi secuencial aleatorio o secuencial ordenado (Fig. 4.3), siguiendo el siguiente esquema:

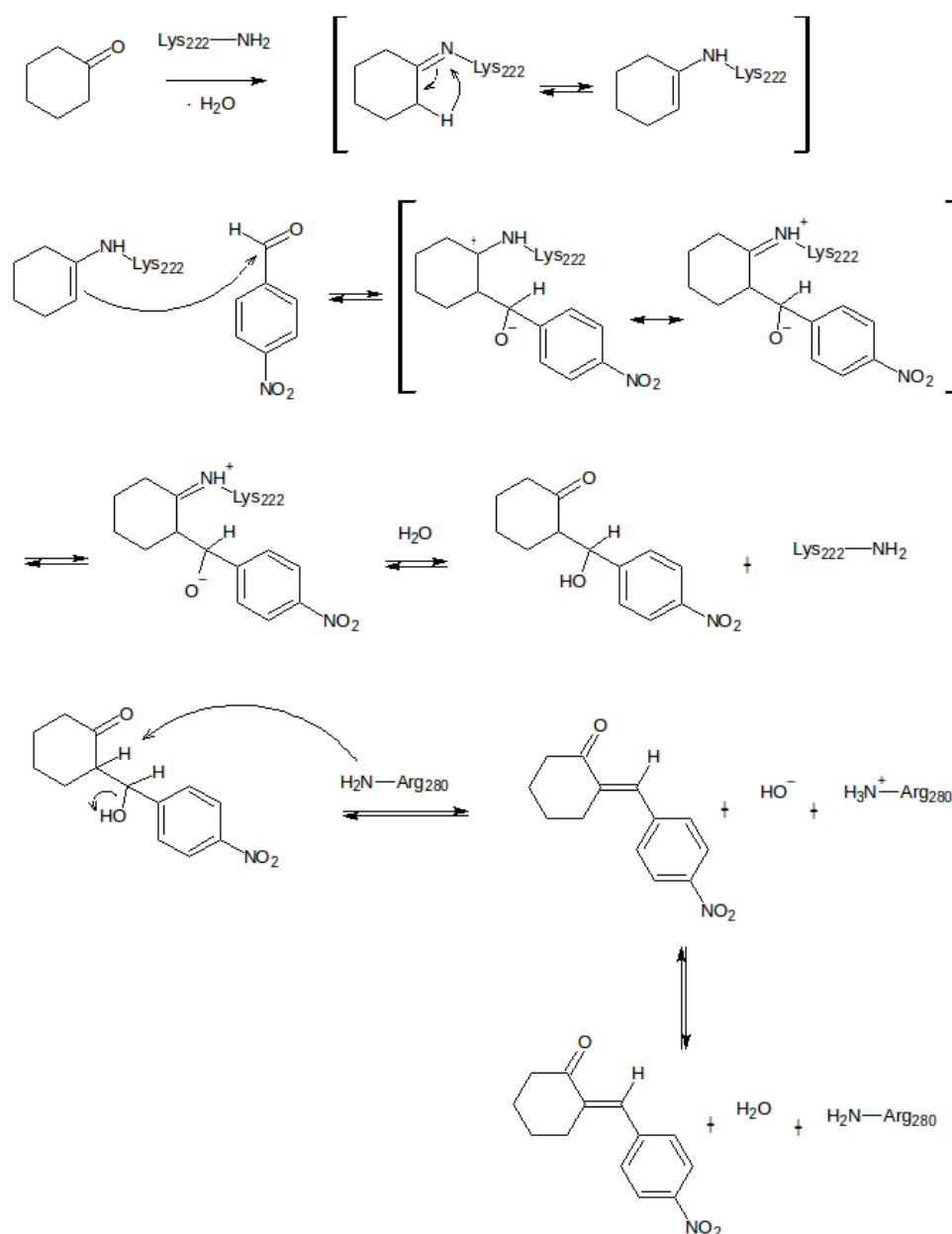


Fig. 4.3 Esquema propuesto para la condensación aldólica catalizada por albúmina entre ciclohexanona y p-nitrobenzaldehído. Figura extraída de “*Bovine serum albumin-catalysed cross aldol condensation: Influence of ketone structure*”. SM Ardanaz, EL Borucki, AJV

Como el mecanismo de reacción probable postula la existencia de un ácido y una base anómalamente fuertes para que esta pueda desarrollarse, exploramos en la cavidad sujeta al docking molecular, la existencia de residuos con un pK anómalo. Para este objetivo, se realizó una titulación teórica de la ASB utilizando el software Propka, con todos los conformeros disponibles de la proteína en la base de datos PDB. Los pKas encontrados para Lys 222 y Arg 280 fueron 7.78 ± 0.83 y 14.07 ± 0.19 respectivamente. Aunque se encontraron otros residuos con corrimientos de pKa, los indicados mostraron los corrimientos de pKa más extremos. La proximidad y la colocación en la misma cavidad de una arginina inusualmente básica y una lisina inusualmente ácida podrían explicar el mecanismo propuesto. En nuestro mecanismo propuesto, la Lys 222 anormalmente ácida podría actuar como nucleófilo en la formación del cetol, a través de una enamina intermediaria. La deshidratación subsiguiente del cetol, para obtener la cetona insaturada, podría ser favorecida por un residuo básico, en este caso ejemplificado en el mecanismo con la arginina inusualmente básica 280 (Fig. 4.3).

4.3 Marco teórico para el estudio de las bases evolutivas y estructurales de la reacción promiscua de condensación aldólica

La primera hipótesis del origen de las actividades promiscuas las propone como punto de partida para nuevas actividades proteicas (Copley et al., 2004; Jensen, 1976). Muchas de las mutaciones acumuladas por las proteínas son neutrales, es decir que no alteran significativamente la capacidad de la misma para realizar su función biológica primaria (Lipman y Wilbur, 1991). Si las mutaciones que son neutrales con respecto a la función biológica primaria de una proteína causan cambios sustanciales en las funciones promiscuas, estas mutaciones podrían permitir una futura evolución funcional (Amitai et al., 2007).

Estos “cambios latentes”, como llaman Amitai et. al., pueden facilitar la adaptación en circunstancias cambiantes, al expandir el rango de actividad de las enzimas existentes y, por lo tanto, proporcionar una ventaja inmediata y reduciendo número de mutaciones requeridas para la divergencia de nuevas funciones.

Muchos estudios de evolución dirigida mostraron que la plasticidad de las proteínas es un factor importante para que mutaciones aleatorias bajo selección perduren favoreciendo cambios funcionales (Aharoni et al., 2005). Incluso los cambios en el esqueleto de la proteína podrían ser responsables de los cambios en su funcionalidad (Spiller et al., 1999). Estos resultados pueden no ser sorprendentes a la luz de los conceptos de diversidad conformacional y su relación con la función proteica, que entiende el estado nativo de una proteína como un conjunto de conformaciones (Tsai et al., 1999b), en el que cualquier cambio en la estructura (incluso pequeño) afecta su funcionalidad (Koshland, 1998). Múltiples trabajos científicos han descrito cómo la diversidad conformacional podría modular el patrón de sustitución de la secuencia proteica (Echave et al., 2016; Marsh y Teichmann, 2014; Tokuriki y Tawfik, 2009) y cómo la extensión del paisaje conformacional está relacionada con la divergencia de la secuencia de proteínas (Aloy et al., 2003; Javier Zea et al., 2013; Juritz et al., 2013; Paoli, 2001; Saldaño et al., 2016). La relación entre estructura-función y la nueva funcionalización debe pensarse teniendo en cuenta estos conceptos (Tokuriki y Tawfik, 2009).

Nuestra base de datos ProtMiscuity nuclea todos los casos de promiscuidad catalítica documentados hasta el momento. Como se mencionara anteriormente, aun no estando clasificadas como enzimas, las albúminas séricas humana y bovina han sido ensayadas en múltiples reacciones (Albanese y Gaggero, 2015; Ardanaz et al., 2018, 2019; Boucher et al., 2005; Sharma et al., 2013; Wu et al., 2006). Tanto la ASB como la ASH poseen más de tres tipos de actividades catalíticas descritas y comparten algunas de estas vías de reacción (Boucher et al., 2005). Los tipos de sustratos y los medios donde estas se llevan a cabo resultan de lo más variados. Existen algunas características estructurales que determinan que una cavidad se constituya

como sitio activo, como la presencia de aminoácidos activados (pKas anómalos), en entornos generalmente hidrofóbicos, que favorecen el ataque de los sustratos y permiten crear el microambiente propicio para la catálisis (Bartlett et al., 2002; Gutteridge y Thornton, 2005a). En el caso de las albúminas séricas humana y bovina, se ha descrito un mecanismo de acción que involucra la intervención de los residuos Lys 199 en ASH y Lys 222 en ASB como nucleófilos para las distintas vías de reacción (Albanese y Gaggero, 2015; Ardanaz et al., 2019; Boucher et al., 2005; Sharma et al., 2013; Wu et al., 2006). Estos residuos se encuentran ubicados en una cavidad hidrofóbica del sitio IIA que, como caracterizamos en la sección 4.2.b, posee las características de un sitio activo, en tanto a su disposición y conexión con la superficie de la proteína. Sin embargo, aún se desconocen los determinantes estructurales que permiten que tal diversidad de reacciones pueda ser llevada a cabo, aparentemente, en un mismo sitio activo que presenta mínimas variaciones entre las especies. Tampoco son claras las implicancias biológicas de dicha actividad, en el caso de que las hubiere, o el origen evolutivo de dichas actividades. ¿Se trata de mecanismos de reacción de relevancia fisiológica? ¿Es, por tanto, la albúmina una enzima de actividad aún no caracterizada? En el presente trabajo nos proponemos analizar los patrones estructurales y evolutivos que permitan responder a dichos interrogantes.

4.3.a Métodos para el estudio de selección positiva de albúminas séricas y su relación con los patrones estructurales

Basamos el presente estudio en la idea de promiscuidad proteica propuesta por Tawfik (Khersonsky et al., 2006), para quien una actividad promiscua es aquella para la cual la proteína no ha evolucionado. Por tanto, si las actividades promiscuas surgen espontáneamente y/o por similitud estructural o mecanística, no debería existir impronta evolutiva que demuestre una adaptación funcional. Dado esto y debido a las numerosas reacciones

promiscuas descritas para la albúmina, nos preguntamos si dichas actividades son meramente azarosas o si en realidad poseen bases fisiológicas/funcionales y/o adaptativas que permanecen aún sin caracterizar. Con esta hipótesis decidimos estudiar las bases evolutivas de la reacción aldólica, ya que para esta transformación el Laboratorio de Biocatálisis ensayó la catálisis de ASB (Ardanaz et al. 2018, 2019), tal como se presentó en páginas anteriores. Para tal fin, se construyó un alineamiento secuencial múltiple de albúminas séricas con el algoritmo de alineamiento secuencial Clustal (Thompson et al., 1994). Las secuencias de genes codificantes y de proteínas homólogas utilizadas para tal fin fueron obtenidas por medio del algoritmo de búsqueda de similitud secuencial BLAST (Altschul et al., 1990), utilizando como secuencia de referencia la albúmina bovina (Uniprot ID: P02769). Dicha búsqueda fue realizada con los parámetros de corrida default del servidor de NCBI¹², contra la base de datos no redundante de secuencias proteicas (la base de datos no redundante (nr), que reúne principalmente secuencias génicas codificantes de GenBank¹³, las secuencias de estructuras depositadas en la PDB¹⁴ y las secuencias proteicas anotadas de SWISS-PROT¹⁵). De los resultados obtenidos del BLAST se buscó incluir secuencias representativas de todos los organismos, que tuvieran estructuras asociadas según las anotaciones de GenBank. Se obtuvieron de este modo 45 secuencias de albúminas séricas provenientes de diversos organismos.

Como bien sabemos, la teoría darwiniana de la descendencia con modificación a partir de ancestros comunes ha sido comprobada gracias a los avances de la biología molecular (Iwabe et al., 1989; Miyata et al., 1991). La combinación entre mutaciones al azar y selección natural explican la adaptación de las especies, propuesta por Darwin (Oldroyd 1986). Aquellas mutaciones que no afectan el fenotipo evolucionan al azar (por deriva genética), las perjudiciales son purgadas por la selección natural negativa y las beneficiosas prosperan por selección positiva. La selección positiva puede, a su vez, subdividirse en una

¹² NCBI: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

¹³ GenBank: <https://www.ncbi.nlm.nih.gov/genbank/>

¹⁴ PDB: <https://www.rcsb.org/>

¹⁵ SWISS-PROT: <http://www.ebi.ac.uk/swissprot/>

selección direccional, que tiende a la fijación de un alelo ventajoso, y una selección que mantiene un polimorfismo. La teoría neutral de la evolución molecular (Kimura, 1991) predice que la selección purificadora es omnipresente, es decir que la mayoría de los cambios evolutivos a nivel molecular se deben a la fijación casual de mutaciones selectivamente neutrales. Análisis genómicos recientes han demostrado que la selección positiva está más extendida a nivel molecular de lo esperado bajo una interpretación restrictiva de la teoría neutral (Eyre-Walker, 2006; Studer et al., 2008). Se espera que ésta produzca cambios en la función y una mejor adaptación del organismo a su entorno y, por lo tanto, identificar sitios de selección positiva es un paso importante no solo en estudios de biología evolutiva, sino también en estudios funcionales. Para la estimación de los sitios sujetos a selección positiva, en nuestro alineamiento se utilizó el algoritmo MEME (Murrell et al., 2012), disponible para su ejecución en el servicio Datamonkey¹⁶ (Weaver et al., 2018). El software MEME es capaz de detectar la probabilidad de que existan instancias de selección positiva a nivel de un sitio individual, utilizando modelos filogenéticos para describir la evolución de los codones a lo largo de una rama en una filogenia dada.

Utilizando los valores default que ofrece el programa, se detectaron 38 posiciones sujetas a selección positiva. Estos fueron mapeados a sus correspondientes secuencias codificantes en las proteínas del alineamiento analizado. Con la información de los residuos seleccionados, tratamos de estimar relaciones estructurales y funcionales. De esta forma, se mapearon los residuos seleccionados evolutivamente en las cavidades predichas con Fpocket (Le Guilloux et al., 2009), para todos los conformeros de las AS humana, de vaca y conejo (ASC) disponibles en PDB. El criterio de selección para dichas especies proteicas dentro de la familia fue la disponibilidad comercial, ya que era nuestra intención contrastar las predicciones bioinformáticas con observaciones experimentales en nuestra colaboración con el Dr. Iglesias. Asimismo, se estimaron los valores de pKa de los residuos ionizables de dichas proteínas, en todos sus conformeros, utilizando el software

¹⁶ Datamonkey: <http://www.datamonkey.org/>

PROPKA (Olsson et al., 2011), con el fin de identificar a los aminoácidos relevantes para la función biológica o actividad promiscua en dichas albúminas. El análisis y manipulación de los datos fue hecho con scripts propios, desarrollados con el lenguaje de programación Python. La visualización de las estructuras se realizó con el software libre Pymol¹⁷.

4.3.b Análisis de las bases evolutivas y estructurales de la actividad promiscua de las albúminas séricas

Nos planteamos como primer paso del análisis, realizar una estimación de las posiciones de la proteína que se encuentran sujetas a fuerzas de selección purificadoras. Para tal fin construimos un alineamiento secuencial múltiple de 44 secuencias de albúminas provenientes de 38 especies distintas, con estructuras cristalográficas disponibles, reclutadas por el algoritmo de búsqueda de similitud secuencial BLAST (ver sección 4.3.a). Sobre dicho alineamiento se realizó una estimación de sitios sujetos a selección positiva utilizando el software MEME del paquete HyPhy (Pond et al., 2005), disponible para su ejecución en el servidor Datamonkey. Se detectaron 38 posiciones sujetas a selección positiva. Para un análisis más profundo de las implicancias de estos resultados se mapearon dichos residuos en las cavidades predichas por el Fpocket para los confórmeros de las distintas proteínas con estructura conocida. Se relacionaron estos datos con la información acerca de los residuos y cavidades funcionalmente importantes, relevados desde bibliografía (Curry et al., 1999; Ha y Bhagavan, 2013; Li et al., 2014; Yang et al., 2014). A partir de esta contrastación de información, se encontró que el 90% de las cavidades de unión a ligandos naturales se encuentran seleccionadas, entre ellas el sitio de unión a drogas IIA, descrito como sitio activo promiscuo de las albúminas humana y bovina (Fig 4.4). Incluso, muchos de los sitios de unión a ligandos se encuentran evolutivamente restringidos, mostrando que las AS "evolucionaron" tanto para la unión diversos ligandos, como para la catálisis.

¹⁷ Pymol: <https://pymol.org/2/>

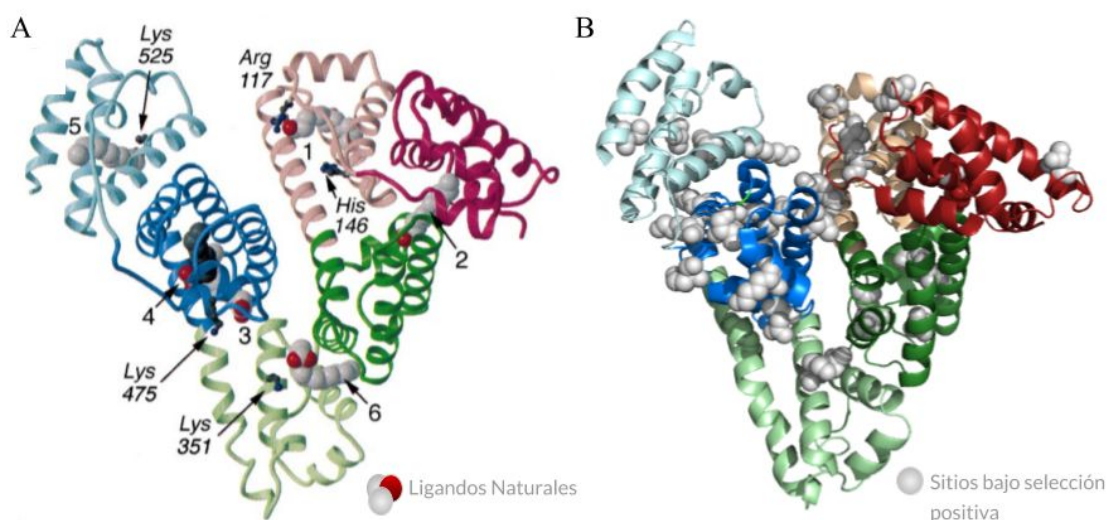


Fig. 4.4. A) Sitios de unión a ligandos naturales ejemplificados sobre la seroalbúmina humana (HSA). Los dominios de albúminas séricas del I al III y sus subdominios (A y B), se muestran coloreados diferencialmente. Adicionalmente se destacan las ubicaciones aproximadas de los sitios de unión a ligandos naturales de la albúmina humana. Figura tomada del trabajo “*Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin*” (Bhattacharya et al., 2000). B) Sitios sujetos a selección positiva predichos con el software MEME, ejemplificados sobre la estructura cristalográfica de albúmina humana (PDB ID 1A06).

Del análisis de los valores de pKa de los residuos ionizables en todos los confórmeros de las albúminas con estructura conocida, predichos con PROPKA, pudimos detectar la presencia de aminoácidos activados en la cavidad que se describe como sitio activo de las AS (correspondiente al sitio de unión IIA). En particular, los aminoácidos equivalentes 199 y 222, descritos como esenciales en las actividades catalíticas promiscuas en humano y vaca respectivamente, mostraron un pKa anormalmente ácido. Esta activación pareciera relacionarse con la presencia de residuos anormalmente básicos que se encuentran en contacto con dichos residuos (ver Fig. 4.5). Vemos, por ejemplo, que mientras la posición 199 de la Lys de la ASH es ácida (pKa promedio de los confórmeros es 7.51), tanto en la ASB como en la ASC esa posición correspondiente a una arginina es alcalina (pKa promedio de los confórmeros 9.23 y 16.52 para vaca y conejo respectivamente). Vemos asimismo que esta situación se invierte en la posición 257 y 222, donde tanto

ASC como ASB presentan un corrimiento de pKa hacia valores más ácidos y la albúmina humana muestra un corrimiento de esas posiciones hacia pKas más alcalinos (Fig. 4.5). Aun cuando la posición de los residuos ácidos se encuentran intercambiada en ASH respecto de otras AS, el entorno básico (principalmente conformado por argininas extremadamente básicas) parece desplazarse y conservarse igualmente.

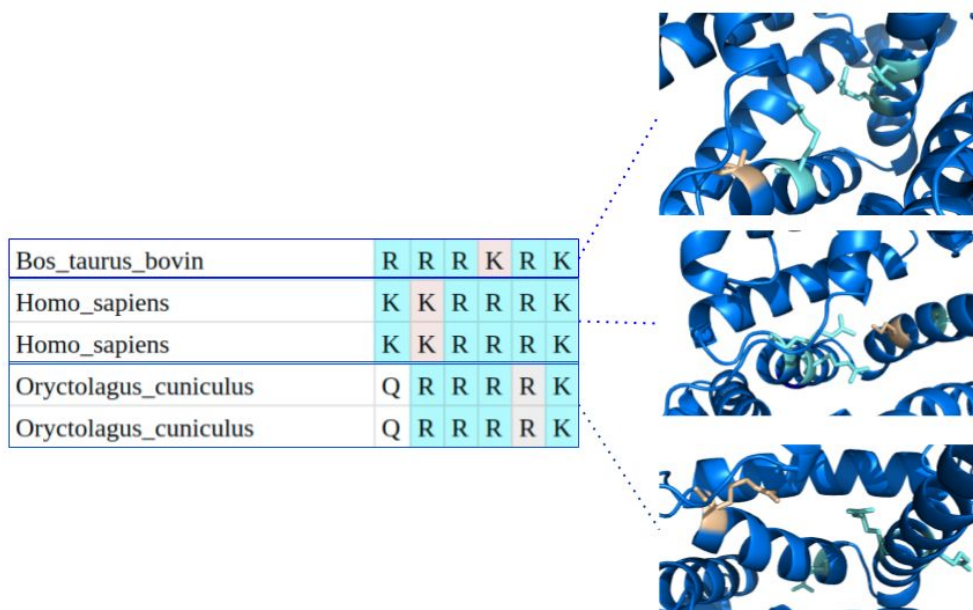


Fig. 4.5. Sitio de unión a drogas All a lo largo del árbol filogenético de las AS. Los residuos activados se muestran en las estructuras coloreados en rosa y celeste, para los residuos anormalmente ácidos y alcalinos, respectivamente. La posición de los residuos ácidos se encuentra intercambiada en ASH respecto de las otras AS y el entorno básico (principalmente conformado por argininas extremadamente básicas) parece desplazarse y conservarse igualmente.

Según nuestros resultados, las albúminas séricas mostraron patrones estructurales similares en las cavidades correspondientes al sitio de unión a drogas IIA, lo que permitiría suponer mecanismos de acción similares para dichas proteínas en las vías de reacción ya descritas para las mismas. Bajo este supuesto, se testearon las albúminas humana y de conejo para la reacción aldólica cruzada, utilizando las condiciones experimentales puestas a punto para la ASB en el trabajo “A Mild Procedure for Enone Preparation Catalysed by Bovine Serum Albumin in a Green and Easily Available Medium” (Ardanaz et

al., 2018), ya mencionado en secciones anteriores; específicamente, se ensayó la reacción entre acetona y *p*-cianobenzaldehído. Ambas proteínas mostraron actividad catalítica para la reacción testeada (ver tabla 4.2), confirmando las predicciones hechas bajo el supuesto estructural.

	ASB	ASH	ASC
% de conversión a cetona insaturada	92.7 ± 9.9	69.5 ± 3.7	63.3 ± 5.3

Tabla 4.2 Condensación aldólica de acetona con *p*-cianobenzaldehído catalizada por albúmina de distintas fuentes

4.4. Conclusiones destacadas

En resumen, de nuestros resultados se desprende que la ASB puede catalizar la reacción de obtención de cetonas aromáticas α,β -insaturadas, a través de una reacción aldólica cruzada en forma promiscua. Este caso particular de promiscuidad catalítica encuentra sus bases mecanísticas en las características estructurales de las albúminas séricas, que según lo observado aportan un entorno fisicoquímico propicio para que la reacción se lleve a cabo. En este sentido, la Lys 222 en la ASB, correspondiente al sitio de unión IIA se encuentra activada, mostrando un corrimiento ácido de su pKa, respecto del esperado para ese aminoácido. Esta observación indicaría un papel esencial en la actividad catalítica "promiscua" de la condensación aldólica cruzada según el mecanismo de reacción propuesto. Dicho aminoácido, se ubica en el interior de una cavidad (sitio de unión AII), que se conecta con la superficie mediante un túnel por la que podrían ingresar los sustratos, típicamente observado en los sitios activos de las enzimas (Pravda et al., 2014).

Del análisis hecho sobre otras proteínas de la familia de las AS, se pudo observar los residuos del sitio IIA que conservaban el patrón de corrimiento de pKa observado en la ASB, donde un aminoácido anormalmente ácido se

encuentra circundado por residuos marcadamente alcalinos. Estos residuos importantes se intercambian de posición en la ASH con referencia a la ASB y ASC. Sin embargo, el entorno básico (principalmente residuos positivos como Arg) en contacto con el residuo anormalmente ácido se desplaza de su posición relativa pero se conserva. El análisis evolutivo respalda la idea de que dicha cavidad podría ser un sitio activo, que evoluciona bajo presión selectiva para mantener la actividad catalítica y la unión.

Nuestros resultados ofrecen una explicación mecanicista para una actividad catalítica observable en las albúminas séricas. Además, se ofrece una nueva perspectiva en la clasificación de las albúminas como posibles enzimas, a la luz de la observación de la presión selectiva en las posiciones clave para la catálisis, típicamente observada durante la divergencia funcional (Drummond et al., 2005). Bajo este punto de vista, las propiedades catalíticas de las albúminas séricas documentadas en el presente trabajo de tesis podrían estar involucradas en un proceso biológico desconocido más que representar un comportamiento promiscuo. Estos resultados no solo nos permiten caracterizar mejor a la familia de las AS, sino que también abren preguntas interesantes sobre el origen del comportamiento promiscuo y sobre la evolución de la función proteica.

Discusión general y perspectivas a futuro

Cuando ya poco faltaba para terminar, Balaám hizo una pausa. Soltó el cincel, miró a su alrededor y contempló el trabajo que le había tomado tanto tiempo y esfuerzo. Los eventos se apelmazaban, no había lugar libre de fechas. Manos y manos de registros, tallados con sus propias patas y garras, cubrían toda la superficie. Uxe-Uxenó, con ojos ajados, casi ciegos, sobre los que descansaban una vida, la observaba en silencio. Y entonces ella, que no veía en Balaám a una pantera sino a una hija, con voz débil, habló.

Felicitaciones Balaám, es la cronología más completa que nuestro pueblo ha jamás conocido, dijo Uxe-Uxenó. Todas las noches siento que está incompleta, que faltan detalles. Miro hacia adelante y pienso en que antes del inicio hubo otro inicio, y antes de este, otros más, respondió Balaám. Y tras una breve pausa, continuó: Y como bien me enseñaste, en cuanto miremos hacia atrás veremos un presente repleto de nuevos hechos, hechos que no habremos incluido. ¿Realmente ésta será mi obra, la obra que me defina?

Para nosotras, las cronistas, nuestras crónicas jamás serán nuestra obra. Deberemos buscarla en otros lugares. O aceptar que es incompleta. O aceptar que tal obra, no existe, dijo Uxe-Uxenó.

*leyenda tradicional Camboré
F.L. Bulgarelli*

Este trabajo presenta tres ejes principales: el desarrollo de una base de datos de proteínas promiscuas, la caracterización de las particularidades dinámicas que sustentan la actividad catalítica en distintos medios y el estudio de la promiscuidad proteica en sistemas particulares. Todos ellos vinculados con el estudio de la función proteica en general, y la actividad catalítica en particular, desde la óptica del análisis estructural.

En primer lugar, desarrollamos la primera base de datos de proteínas promiscuas, incorporando una mirada pragmática de clasificación químico-funcional (Kazlauskas, 2005). Hemos presentado ProtMiscuity como una herramienta fundamental para el análisis de la promiscuidad proteica, tanto de casos particulares como de generalidades. ProtMiscuity ofrece, no sólo datos curados relativos a cada reacción caracterizada y a los residuos proteicos descritos como relevantes para la actividad promiscua, sino que también vincula dichos datos con los provenientes de otras bases de datos de interés biológico (ver sección 2.4.c). Esto permite a los usuarios recopilar múltiples descriptores de los sitios activos catalíticos, como propiedades fisicoquímicas de los residuos involucrados y sus afinidades a los receptores, que podrían mejorar el diseño de protocolos de ingeniería de proteínas y el descubrimiento de fármacos *in silico*, por ejemplo.

A partir del análisis de nuestra base de datos, y según lo descrito en el capítulo 2, podemos observar que existe una amplia distribución y variedad de reacciones promiscuas en los distintos reinos. También, vemos que existen actividades promiscuas descritas en distintas familias proteicas de características estructurales diversas (Orengo et al., 1993), e incluso hemos comprobado que proteínas que no se describen como enzimas, pueden presentar actividades enzimáticas promiscuas (Albanese y Gaggero, 2015). Resulta evidente, a partir de dichas observaciones, que es imposible circunscribir el fenómeno de la promiscuidad proteica a un solo grupo filogenético o un tipo de proteína. En cambio, la promiscuidad es una característica que no puede ser descartada *a priori* en ningún caso y su identificación deberá centrarse, entonces, en la dinámica que mostrara la proteína en un dado contexto.

En ProtMiscuity cada proteína se encuentra incorporada a la base de datos a través de su código Uniprot y se la vincula con las estructuras cristalográficas existentes para dicha proteína. Esto permite la relación directa entre la funcionalidad proteica y las características estructurales de la misma. Además, la base de datos recopila información sobre las condiciones de reacción que promueven una determinada actividad promiscua, datos que resultan relevantes a la hora de diseñar nuevos experimentos y en la exploración de nuevas propiedades promiscuas en proteínas de interés biotecnológico e industrial. Este aporte de ProtMiscuity resulta de particular importancia, ya que como vimos en el presente trabajo, los diversos entornos a los que se encuentran sometidas las proteínas, como por ejemplo solventes no acuosos, pueden generar cambios conformacionales específicos en las proteínas que alteran su funcionalidad.

Lamentablemente, la escasa información en muchos casos y la accesibilidad restringida a los trabajos científicos (Albert, 2006; Björk, 2012; Tennant et al., 2016) hacen que la recuperación de datos relativos a la promiscuidad proteica, de forma sistemática y a gran escala, se convierta en un desafío de gran complejidad. Asimismo, y por lo antes expuesto, vemos que la identificación de controles negativos de promiscuidad resulta, si no imposible, muy difícil de concretar. Si bien no se realizó en el marco de la presente tesis un análisis de la robustez de los distintos métodos predictivos de promiscuidad proteica, ProtMiscuity ofrece la posibilidad a futuro de contrastar el comportamiento de dichos métodos con información curada, haciendo un aporte importante en el campo de predicción computacional. Será necesaria la implementación de algoritmos de búsqueda más sofisticados, que permitan acumular un número mayor de casos. En este sentido, creemos que el conjunto de artículos cuidadosamente seleccionados para la construcción de la base de datos ProtMiscuity nos resultará de gran utilidad para la aplicación de métodos computacionales de aprendizaje automático, que permitan extender y enriquecer los datos recolectados e incrementar así nuestro conocimiento de la promiscuidad proteica.

Nuestros resultados en el estudio de las características estructurales de la proteínas en medios no acuosos, desafían la denominada teoría de atrapamiento cinético, generalmente aceptada, que postula la conservación de la estructura tal y como se observa en agua. Muy por el contrario a lo que se propone en dicha teoría, pudimos observar que los confórmers orgánicos poseen cavidades más voluminosas y una restricción en su movilidad, evidenciada con la reducción de la diversidad conformacional y del desorden. Estos resultados se sustentan en la distribuciones de RMSD de los confórmers orgánicos, que son levemente superiores al error cristalográfico y muy por debajo de las distribuciones observadas para los confórmers en agua (como se muestra en la sección 3.3.a). Dichos cambios podrían explicar no sólo la reducción de actividad de las reacciones canónicas de las proteínas en medios orgánicos, sino que podrían justificar los comportamientos promiscuos en dichos contextos y los cambios en la especificidad proteica observada habitualmente en medios no acuosos.

Finalmente, en este trabajo hemos abordado también el estudio de la promiscuidad proteica desde sistemas particulares, como son las albúminas séricas. Como se explicitara en el capítulo 4, las albúminas séricas son una familia proteica para la que se encuentran descritas un gran número de reacciones promiscuas. En nuestra base de datos poseen anotadas, tanto la albúmina bovina como la humana, más de tres reacciones promiscuas distintas. En resumen, nuestro trabajo brinda una racionalización de la actividad promiscua de la ASB en la reacción aldólica cruzada, que permite obtener cetonas aromáticas α,β -insaturadas. Sobre esta reacción hemos aportado mediante el uso de métodos computacionales una explicación del mecanismo probable de acción enzimática, contemplando las características estructurales de dicha proteína. Pudimos observar la activación de los residuos comúnmente propuestos como catalíticamente relevantes para la actividad promiscua de dichas proteínas (Boucher et al., 2005; Goncharov et al., 2017; Reetz et al., 2007), evidenciada a través del corrimiento de los pKas de dichos residuos y un entorno particularmente alcalino que pareciera provocar dicha activación (sección 4.3.b).

Asimismo, a partir de la descripción de las subestructuras proteicas y las restricciones evolutivas a lo largo de toda la familia, pudimos observar que este entorno propicio para la catálisis se conserva en otras proteínas de la familia. La cavidad correspondiente al sitio de unión IIA se encuentra conectado con la superficie proteica a través de un túnel hidrofóbico que permitiría el ingreso de sustratos al sitio activo, en una configuración observada típicamente en las enzimas. Además, dicha cavidad, junto con casi la totalidad de cavidades descritas como importantes en la unión de ligandos naturales, se encuentra sujeta a presión de selección evolutiva. Este análisis respalda la idea de que dicha cavidad podría ser un sitio activo, que evoluciona bajo presión selectiva para mantener tanto la actividad catalítica como la unión. Estos hallazgos refuerzan nuestra idea de que la promiscuidad como fenómeno no puede ser descartado a priori en ningún caso. En el caso particular de las albúminas séricas vemos, a través de las evidencias evolutivas, que existe una neo-funcionalización de las AS que les confiere su capacidad catalítica. Aun cuando no nos es posible distinguir cuál de las reacciones catalizadas por las mismas corresponde a la vía que sería llevada a cabo in vivo, o cuál es su sustrato natural, nuestros resultados dan idea de la existencia de una actividad enzimática de relevancia fisiológica, que aún no ha sido caracterizada pero que persiste en los miembros de dicha familia.

El trabajo realizado para esta tesis ha sido una experiencia enriquecedora, que nos ha permitido comprender mejor la función proteica y su relación con la promiscuidad. Creemos necesario resignificar las definiciones que enmarcan este fenómeno a la luz de la evolución y del análisis estructural, para avanzar así el conocimiento científico en este área. Hemos podido también vislumbrar la complejidad del estudio de este fenómeno, con alcance irrestricto en el universo proteico y presencia transversal a organismos de complejidades y entornos tan disímiles como pueda imaginarse. Nos queda un largo camino por recorrer hasta comprender la promiscuidad proteica en su totalidad y reconocerla como piedra fundamental de los complejos mecanismos moleculares que gobiernan la historia evolutiva de la vida en nuestro planeta.

6. Apéndice

Tablas Suplementarias

Tabla Suplementaria 1

Código PDB	Solvente / Molécula Orgánica
09V	ciclopropilmetanol
3GR	gliceraldehído
4PB	butanamida
8CL	clorobenceno
ACN	acetona
ANF	antrona
ANL	anilina
ARF	formamida
BML	4-bromofenol
BU1	1,4-butanodiol
BXO	butiraldoxima
BZF	benzofurano
CCN	acetonitrilo
CXL	ciclohexanol
CYH	ciclohexanona
D12	dodecano
DIO	1,4-dioxano
DMF	N,N-dimetilformamida
DMS	dimetilsulfóxido
EOH	etanol
ETF	trifluoroetanol
ETI	yoduro de etilo
GOL	glicerol
HEX	hexano
HEZ	1,6-hexanodiol
HP6	heptano
HY1	fenilacetaldehído
IPA	isopropanol
IPH	fenol
MBN	tolueno

MBR	tribromometano
MCH	cloroformo
MEE	metanotiol
MOH	metanol
NBU	butano
NBZ	nitrobenceno
NEH	etilamina
OCT	octano
PCI	pentaclorofenol
PHZ	fenilhidrazina
POL	propanol
PTD	pentanodial
PYJ	etilbenceno
RSF	R,S,R-bisfuranol
TBU	terbutanol
TMO	trimetilamina
TOU	tiourea
UNA	undecanal

Figuras Suplementarias

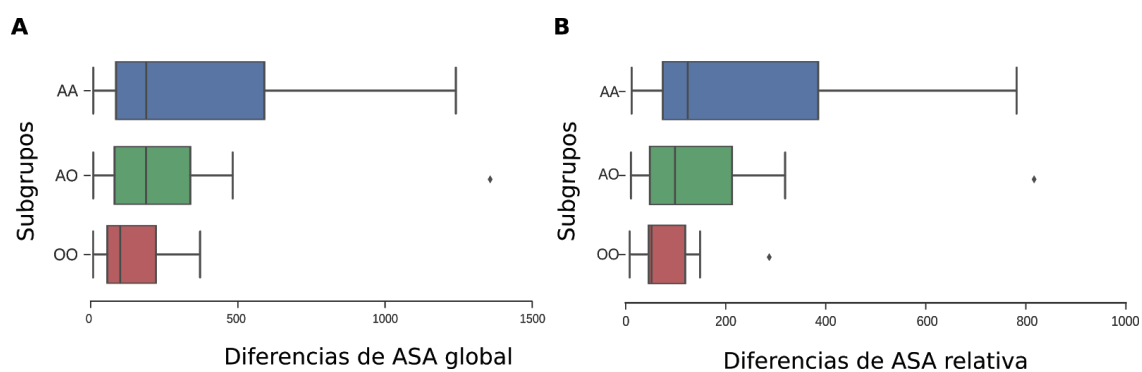


Fig. 3S.1. A) Diferencias en el ASA global para los diferentes subgrupos AA (azul), AO (verde) y OO (rojo) para el conjunto de datos de control. El subgrupo AO muestra las diferencias máximas, evidenciando cambios más grandes entre los confórmeros obtenidos en diferentes medios. Los promedios de diferencias globales de ASA fueron 424.11, 328.12 y 140.62 para AA, AO y OO respectivamente (medianas observadas: 196.33, 183.05 y 94.81 respectivamente). Valores $p < 0.05$. B). Las diferencias relativas de ASA para el conjunto de datos de control muestran el mismo comportamiento que las diferencias globales de ASA.

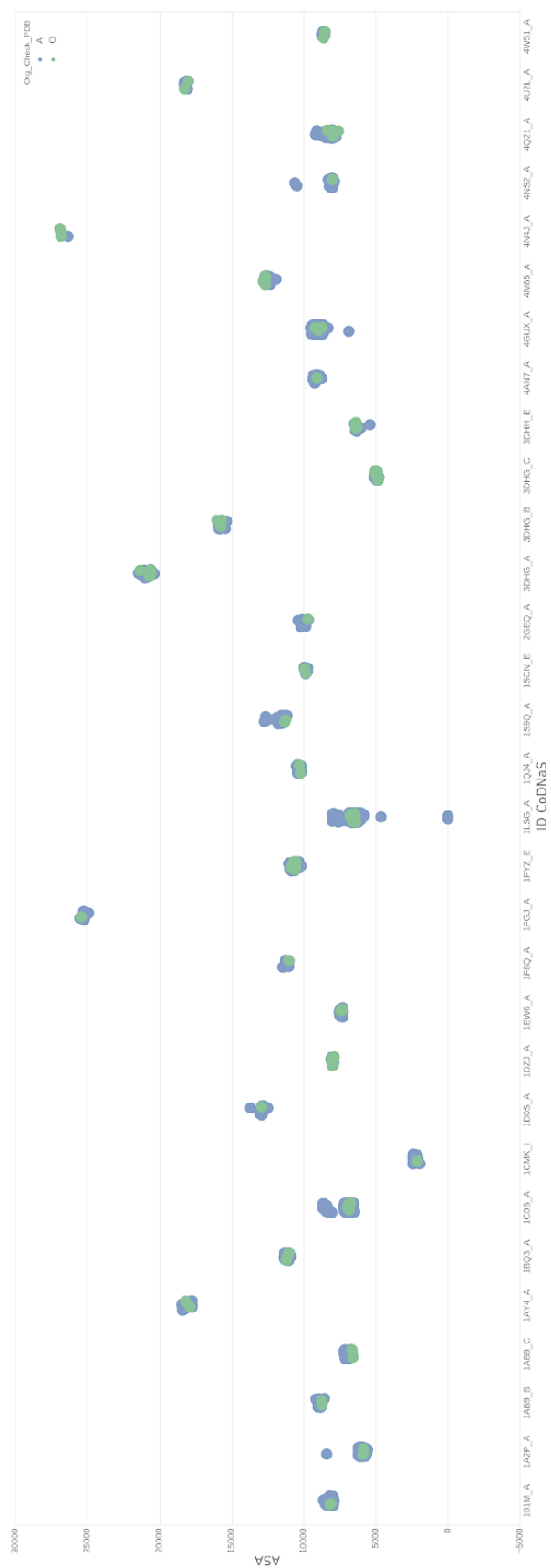


Fig. 3S.2. Distribuciones globales de ASA teniendo en cuenta todos los pares de confórmers. Distribuciones del área de superficie accesible al solvente de un conjunto representativo de proteínas del

conjunto de datos de control medido en ASA ($\text{\AA}^3$) para los confórmers A y O. El ASA para los confórmers acuosos (A) se muestra en azul y el para los confórmers no acuosos (O) se muestra en verde.

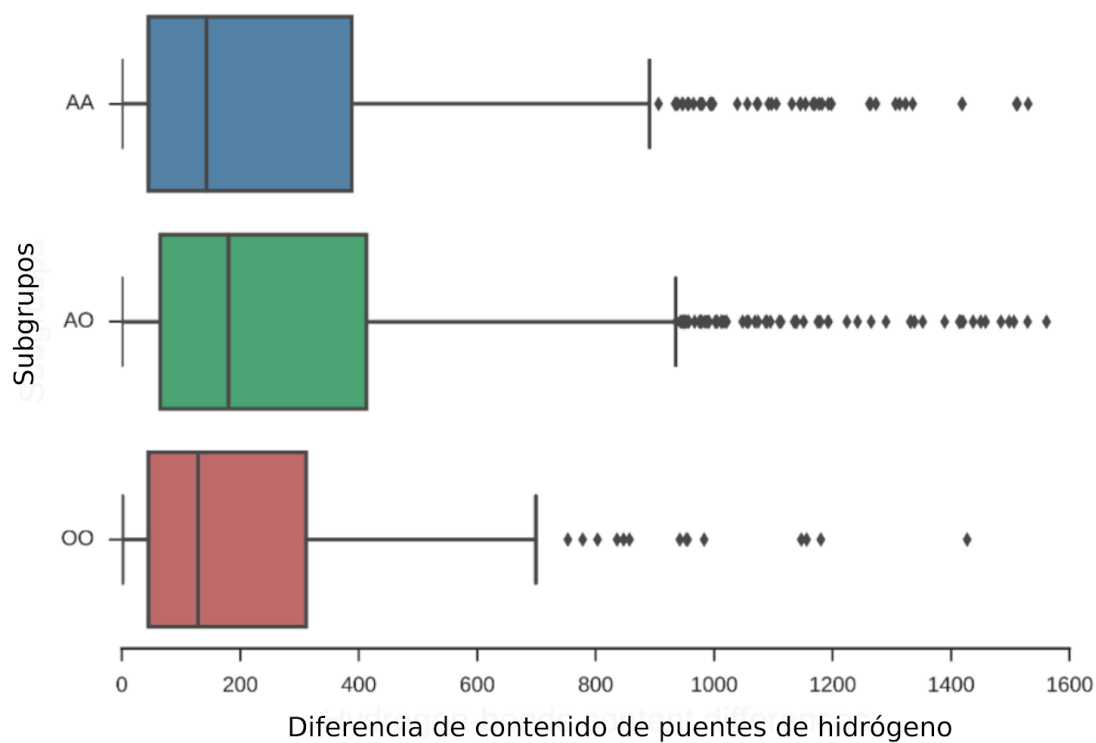


Fig. 3S.3. Las distribuciones de la diferencia en los puentes de hidrógeno para cada subgrupo del conjunto grande de datos. El promedio de diferencias para AA (que se muestra en azul) fue de 315.87 (mediana de 143), para el subgrupo AO (que se muestra en verde) el valor medio fue de 312.71 (mediana de 180), y para la OO (mostrada en rojo) el valor medio fue 246.57 (mediana 129). Los valores P de las comparaciones entre los subgrupos fueron < 0.01 al comparar AA vs AO y OO vs AO. No se encontraron diferencias estadísticas entre las distribuciones AA y OO.

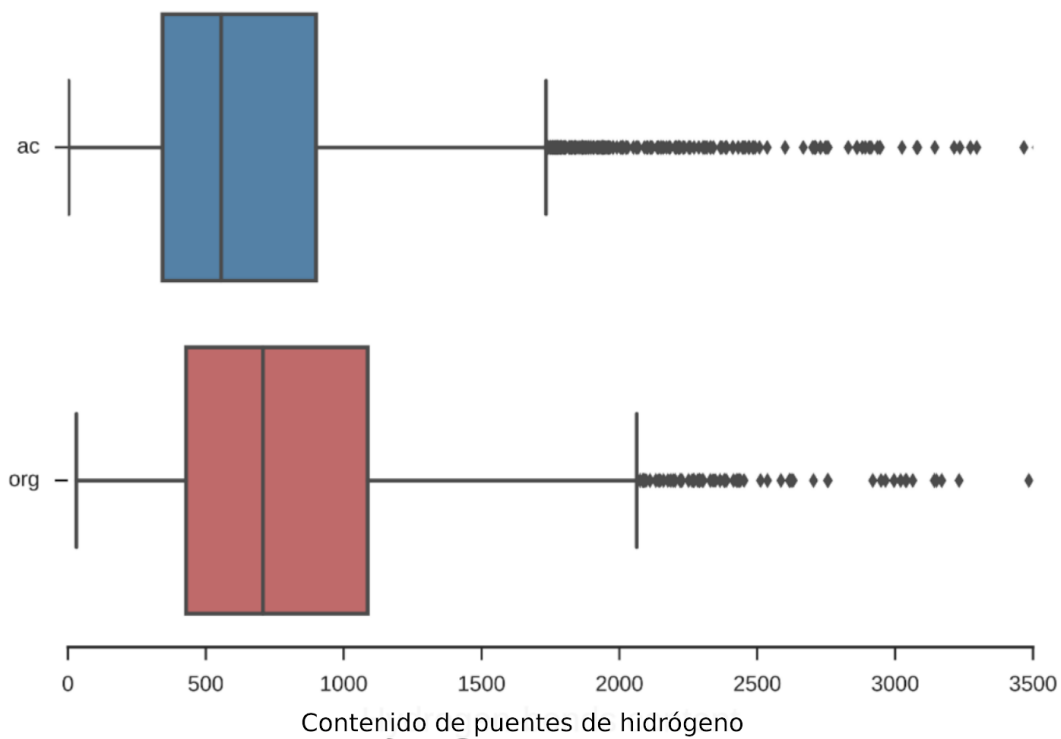


Fig. 3S.4. Distribuciones de enlaces de hidrógeno para los conformeros acuosos y no acuosos en el conjunto grande de datos. Para los conformeros acuosos (mostrados en azul) el promedio de enlaces de hidrógeno es 721.18, mientras que para los conformeros no acuosos es 847.87.

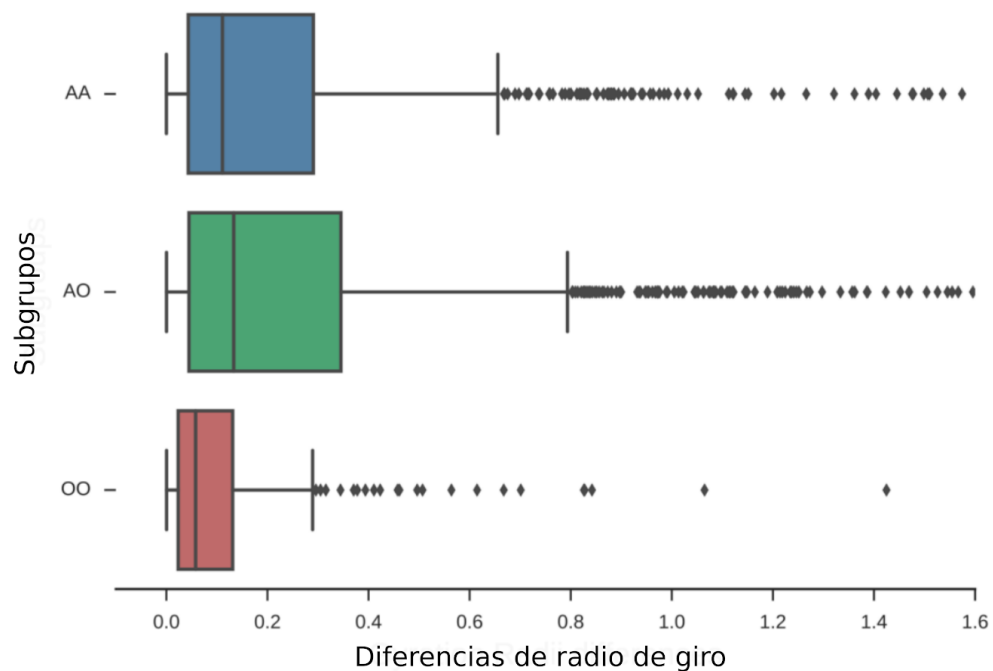


Fig. 3S.5. Distribuciones de las diferencias de giro para los pares máximos de cada subgrupo del conjunto de datos grande. La diferencia media de los radios de giro fue de 0.275 en el subgrupo AA (que se

muestra en azul), 0.392 para el subgrupo AO (que se muestra en verde) y 0.146 para el subgrupo OO (que se muestra en rojo). Valor $P \ll 0.01$ para todas las comparaciones.

7. Bibliografía

- Aharoni, A., Gaidukov, L., Khersonsky, O., McQ Gould, S., Roodveldt, C., and Tawfik, D.S. (2005). The “evolvability” of promiscuous protein functions. *Nat. Genet.* 37, 73–76.
- Albanese, D.C.M., and Gaggero, N. (2015). Albumin as a promiscuous biocatalyst in organic synthesis. *RSC Adv.* 5, 10588–10598.
- Albert, K.M. (2006). Open access: implications for scholarly publishing and medical libraries. *J. Med. Libr. Assoc.* 94, 253–262.
- Aloy, P., Ceulemans, H., Stark, A., and Russell, R.B. (2003). The relationship between sequence and interaction divergence in proteins. *J. Mol. Biol.* 332, 989–998.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410.
- Amitai, G., Gupta, R.D., and Tawfik, D.S. (2007). Latent evolutionary potentials under the neutral mutational drift of an enzyme. *HFSP J.* 1, 67–78.
- Anfinsen, C.B., Haber, E., Sela, M., and White, F.H. (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA* 47, 1309–1314.
- Ardanaz, S.M., Velez Rueda, A.J., Parisi, G., Iribarren, A.M., and Iglesias, L.E. (2018). A mild procedure for enone preparation catalysed by bovine serum albumin in a green and easily available medium. *Catal. Lett* 148, 1750–1757.
- Ardanaz, S.M., Borucki, E.L., Velez Rueda, A.J., Parisi, G., Iribarren, A.M., and Iglesias, L.E. (2019). Bovine serum albumin-catalysed cross aldol condensation: Influence of ketone structure. *Process Biochemistry*.
- Arora, B., Mukherjee, J., and Gupta, M.N. (2014). Enzyme promiscuity: using the dark side of enzyme specificity in white biotechnology. *Sustain. Chem. Process.* 2, 25.
- Atkins, W.M. (2015). Biological messiness vs. biological genius: Mechanistic

aspects and roles of protein promiscuity. *J. Steroid Biochem. Mol. Biol.* **151**, 3–11.

Babtie, A., Tokuriki, N., and Hollfelder, F. (2010). What makes an enzyme promiscuous? *Curr. Opin. Chem. Biol.* **14**, 200–207.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Res.* **28**, 304–305.

Baldwin, A.J., and Kay, L.E. (2009). NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* **5**, 808–814.

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D.S., and Milo, R. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410.

Bartlett, G.J., Porter, C.T., Borkakoti, N., and Thornton, J.M. (2002). Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.* **324**, 105–121.

Ben-David, M., Wieczorek, G., Elias, M., Silman, I., Sussman, J.L., and Tawfik, D.S. (2013). Catalytic metal ion rearrangements underline promiscuity and evolvability of a metalloenzyme. *J. Mol. Biol.* **425**, 1028–1038.

Benedetti, F., Berti, F., and Bidoggia, S. (2011). Aldolase activity of serum albumins. *Org. Biomol. Chem.* **9**, 4417–4420.

Benkovic, S.J., and Hammes-Schiffer, S. (2003). A perspective on enzyme catalysis. *Science* **301**, 1196–1202.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. (2000). The protein data bank. *Nucleic Acids Res.* **28**, 235–242.

Bhattacharya, A.A., Grüne, T., and Curry, S. (2000). Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *J. Mol. Biol.* **303**, 721–732.

Biedermannová, L., Prokop, Z., Gora, A., Chovancová, E., Kovács, M., Damborsky, J., and Wade, R.C. (2012). A single mutation in a tunnel to the

active site changes the mechanism and kinetics of product release in haloalkane dehalogenase LinB. *J. Biol. Chem.* 287, 29062–29074.

Binns, D., Dimmer, E., Huntley, R., Barrell, D., O'Donovan, C., and Apweiler, R. (2009). QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* 25, 3045–3046.

Björk, B.-C. (2012). The hybrid model for open access publication of scholarly articles: A failed experiment? *J. Am. Soc. Inf. Sci.* 63, 1496–1504.

Bornscheuer, U.T., and Kazlauskas, R.J. (2004). Catalytic promiscuity in biocatalysis: using old enzymes to form new bonds and follow new pathways. *Angew Chem Int Ed Engl* 43, 6032–6040.

Bornscheuer, U.T., Huisman, G.W., Kazlauskas, R.J., Lutz, S., Moore, J.C., and Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature* 485, 185–194.

Böttcher, B., Bertsche, I., Reuter, R., and Gräber, P. (2000). Direct visualisation of conformational changes in EF(0)F(1) by electron microscopy. *J. Mol. Biol.* 296, 449–457.

Boucher, G., Robin, S., Fargeas, V., Dintinger, T., Mathé-Allainmat, M., Lebreton, J., and Tellier, C. (2005). Serum albumin-catalyzed trigger system by using a tandem kemp elimination/beta-elimination reaction. *ChemBiochem* 6, 807–810.

Bourne, H.R., Sanders, D.A., and McCormick, F. (1990). The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* 348, 125–132.

Boutin, Y., Hébert, H., Vrancken, E.R., and Mourad, W. (1988). Allergenicity and cross-reactivity of cat and dog allergenic extracts. *Clin. Allergy* 18, 287–293.

Branneby, C., Carlqvist, P., Magnusson, A., Hult, K., Brinck, T., and Berglund, P. (2003). Carbon-carbon bonds by hydrolytic enzymes. *J. Am. Chem. Soc.* 125, 874–875.

Briggs, G.E., and Haldane, J.B. (1925). A note on the kinetics of enzyme action.

Biochem. J. *19*, 338–339.

Brot, F.E., and Bender, M.L. (1969). Use of the specificity constant of .alpha.-chymotrypsin. J. Am. Chem. Soc. *91*, 7187–7191.

Brown, A.J. (1902). XXXVI.—Enzyme action. J. Chem. Soc., Trans. *81*, 373–388.

Bryngelson, J.D., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. Proteins *21*, 167–195.

Buhrman, G., de Serrano, V., and Mattos, C. (2003). Organic solvents order the dynamic switch II in Ras crystals. Structure *11*, 747–751.

Burra, P.V., Zhang, Y., Godzik, A., and Stec, B. (2009). Global distribution of conformational states derived from redundant models in the PDB points to non-uniqueness of the protein structure. Proc Natl Acad Sci USA *106*, 10505–10510.

Busto, E., Gotor-Fernández, V., and Gotor, V. (2010). Hydrolases: catalytically promiscuous enzymes for non-conventional reactions in organic synthesis. Chem. Soc. Rev. *39*, 4504–4523.

Callender, R., and Dyer, R.B. (2015). The dynamical nature of enzymatic catalysis. Acc. Chem. Res. *48*, 407–413.

Cannon, W.R., Singleton, S.F., and Benkovic, S.J. (1996). A perspective on biological catalysis. Nat. Struct. Biol. *3*, 821–833.

Cao, L., Luo, Q., Liu, W., Lin, Y., Liu, X., Cao, Y., Zhang, W., Wu, Y., Yang, J., Yao, T., et al. (2019). Identification of single-atom active sites in carbon-based cobalt catalysts during electrocatalytic hydrogen evolution. Nat. Catal. *2*, 134–141.

Carbonell, P., and Faulon, J.-L. (2010). Molecular signatures-based prediction of enzyme promiscuity. Bioinformatics *26*, 2012–2019.

Carbonell, P., Lecointre, G., and Faulon, J.-L. (2011). Origins of specificity and promiscuity in metabolic networks. J. Biol. Chem. *286*, 43994–44004.

Cecchini, M., Houdusse, A., and Karplus, M. (2008). Allosteric communication in myosin V: from small conformational changes to large directed movements. *PLoS Comput. Biol.* 4, e1000129.

Chakraborty, S. (2012). An automated flow for directed evolution based on detection of promiscuous scaffolds using spatial and electrostatic properties of catalytic residues. *PLoS ONE* 7, e40408.

Chakraborty, S., and Rao, B.J. (2012). A measure of the promiscuity of proteins and characteristics of residues in the vicinity of the catalytic site that regulate promiscuity. *PLoS ONE* 7, e32011.

Chang, B.S., Park, K.H., and Lund, D.B. (1988). Thermal inactivation kinetics of horseradish peroxidase. *J. Food Sci.* 53, 920–923.

Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol. Rev.* 93, 269–309.

Chiang, R.A., Sali, A., and Babbitt, P.C. (2008). Evolutionarily conserved substrate substructures for automated annotation of enzyme superfamilies. *PLoS Comput. Biol.* 4, e1000142.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* 75, 333–366.

Chi, Z., and Liu, R. (2011). Phenotypic characterization of the binding of tetracycline to human serum albumin. *Biomacromolecules* 12, 203–209.

Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B., Gora, A., Sustr, V., Klvana, M., Medek, P., et al. (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* 8, e1002708.

Christopoulos, A. (2014). Advances in G protein-coupled receptor allostery: from function to structure. *Mol. Pharmacol.* 86, 463–478.

Cieplak, M., and Banavar, J.R. (2013). Energy landscape and dynamics of proteins: an exact analysis of a simplified lattice model. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 88, 040702.

- Cistola, D.P., Small, D.M., and Hamilton, J.A. (1987). Carbon 13 NMR studies of saturated fatty acids bound to bovine serum albumin. I. The filling of individual fatty acid binding sites. *J. Biol. Chem.* 262, 10971–10979.
- Clark, D.S. (2004). Characteristics of nearly dry enzymes in organic solvents: implications for biocatalysis in the absence of water. *Philos. Trans. R. Soc. Lond. B. Biol. Sci* 359, 1299–1307; discussion 1307, 1323.
- Cojocaru, V., Winn, P.J., and Wade, R.C. (2007). The ins and outs of cytochrome P450s. *Biochim. Biophys. Acta* 1770, 390–401.
- Colicelli, J. (2004). Human RAS superfamily proteins and related GTPases. *Sci. STKE* 2004, RE13.
- Colin, P.-Y., Kintsjes, B., Gielen, F., Miton, C.M., Fischer, G., Mohamed, M.F., Hyvönen, M., Morgavi, D.P., Janssen, D.B., and Hollfelder, F. (2015). Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* 6, 10008.
- Collins, T., Meuwis, M.-A., Gerday, C., and Feller, G. (2003). Activity, stability and flexibility in glycosidases adapted to extreme thermal environments. *J. Mol. Biol.* 328, 419–428.
- Colman, P.M., and Smith, B.J. (2002). The trypanosomal trans-sialidase: two catalytic functions associated with one catalytic site. *Structure* 10, 1466–1468.
- Copley, S.D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* 7, 265–272.
- Copley, S.D. (2015). An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci.* 40, 72–78.
- Copley, S.D. (2017). Shining a light on enzyme promiscuity. *Curr. Opin. Struct. Biol.* 47, 167–175.
- Copley, S.D., Novak, W.R.P., and Babbitt, P.C. (2004). Divergence of function in the thioredoxin fold suprafamily: evidence for evolution of peroxiredoxins from a thioredoxin-like ancestor. *Biochemistry* 43, 13981–13995.
- Cornish-Bowden, A., and Cárdenas, M.L. (2010). Specificity of

non-Michaelis-Menten enzymes: necessary information for analyzing metabolic pathways. *J. Phys. Chem. B* **114**, 16209–16213.

Coulombe, R., Yue, K.Q., Ghisla, S., and Vrielink, A. (2001). Oxygen access to the active site of cholesterol oxidase through a narrow channel is gated by an Arg-Glu pair. *J. Biol. Chem.* **276**, 30435–30441.

Cross, J.B., Thompson, D.C., Rai, B.K., Baber, J.C., Fan, K.Y., Hu, Y., and Humblet, C. (2009). Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **49**, 1455–1474.

Cruzeiro, L., and Degrève, L. (2017). Exploring the Levinthal limit in protein folding. *J. Biol. Phys.* **43**, 15–30.

Curry, S., Brick, P., and Franks, N.P. (1999). Fatty acid binding to human serum albumin: new insights from crystallographic studies. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1441**, 131–140.

Danchin, A. (2009). Myopic selection of novel information drives evolution. *Curr. Opin. Biotechnol.* **20**, 504–508.

DeForte, S., and Uversky, V.N. (2016). Resolving the ambiguity: Making sense of intrinsic disorder when PDB structures disagree. *Protein Sci.* **25**, 676–688.

Deichmann, U., Schuster, S., Mazat, J.-P., and Cornish-Bowden, A. (2014). Commemorating the 1913 Michaelis-Menten paper Die Kinetik der Invertinwirkung: three perspectives. *FEBS J.* **281**, 435–463.

Dellus-Gur, E., Toth-Petroczy, A., Elias, M., and Tawfik, D.S. (2013). What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. *J. Mol. Biol.* **425**, 2609–2621.

Desai, K.K., and Miller, B.G. (2008). A metabolic bypass of the triosephosphate isomerase reaction. *Biochemistry* **47**, 7983–7985.

DeSantis, G., and Davis, B.G. (2006). The expanding roles of biocatalysis and biotransformation. *Curr. Opin. Chem. Biol.* **10**, 139–140.

Desdouits, N., Nilges, M., and Blondel, A. (2015). Principal Component Analysis reveals correlation of cavities evolution and functional motions in proteins. *J.*

Mol. Graph. Model. 55, 13–24.

Deshpande, A., Nimsadkar, S., and Mande, S.C. (2005). Effect of alcohols on protein hydration: crystallographic analysis of hen egg-white lysozyme in the presence of alcohols. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1005–1008.

Devos, D., and Valencia, A. (2000). Practical limits of function prediction. *Proteins* 41, 98–107.

De Diego, T., Lozano, P., Gmouh, S., Vaultier, M., and Iborra, J.L. (2005). Understanding structure-stability relationships of *Candida antarctica* lipase B in ionic liquids. *Biomacromolecules* 6, 1457–1464.

Díaz, N., Suárez, D., Sordo, T.L., and Merz, K.M. (2001). Molecular dynamics study of the IIA binding site in human serum albumin: influence of the protonation state of Lys195 and Lys199. *J. Med. Chem.* 44, 250–260.

Dong, Z., Zhou, H., and Tao, P. (2018). Combining protein sequence, structure, and dynamics: A novel approach for functional evolution analysis of PAS domain superfamily. *Protein Sci.* 27, 421–430.

Drmanovic, Z., Voyatzi, S., Kouretas, D., Sahpazidou, D., Papageorgiou, A., and Antonoglou, O. (1999). Albumin possesses intrinsic enolase activity towards dihydrotestosterone which can differentiate benign from malignant breast tumors. *Anticancer Res.* 19, 4113–4124.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., and Arnold, F.H. (2005). Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA* 102, 14338–14343.

Dundas, J., Adamian, L., and Liang, J. (2011). Structural signatures of enzyme binding pockets from order-independent surface alignment: a study of metalloendopeptidase and NAD binding proteins. *J. Mol. Biol.* 406, 713–729.

Dupont, C.L., Butcher, A., Valas, R.E., Bourne, P.E., and Caetano-Anollés, G. (2010). History of biological metal utilization inferred through phylogenomic analysis of protein structures. *Proc Natl Acad Sci USA* 107, 10567–10572.

Dwivedee, B.P., Soni, S., Sharma, M., Bhaumik, J., Laha, J.K., and Banerjee,

U.C. (2018). Promiscuity of Lipase-Catalyzed Reactions for Organic Synthesis: A Recent Update. *ChemistrySelect* 3, 2441–2466.

Echave, J., Spielman, S.J., and Wilke, C.O. (2016). Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* 17, 109–121.

Edelsbrunner, H., and Mücke, E.P. (1992). Three-dimensional alpha shapes. In *Proceedings of the 1992 Workshop on Volume Visualization - VVS '92*, (New York, New York, USA: ACM Press), pp. 75–82.

Eisenmesser, E.Z., Bosco, D.A., Akke, M., and Kern, D. (2002). Enzyme dynamics during catalysis. *Science* 295, 1520–1523.

Eisenthal, R., Danson, M.J., and Hough, D.W. (2007). Catalytic efficiency and k_{cat}/K_M : a useful comparator? *Trends Biotechnol.* 25, 247–249.

Elias, M., and Tawfik, D.S. (2012). Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J. Biol. Chem.* 287, 11–20.

Elias, M., Wieczorek, G., Rosenne, S., and Tawfik, D.S. (2014). The universality of enzymatic rate-temperature dependency. *Trends Biochem. Sci.* 39, 1–7.

Engh, R.A., and Bossemeyer, D. (2002). Structural aspects of protein kinase control-role of conformational flexibility. *Pharmacol. Ther.* 93, 99–111.

English, A.C., Done, S.H., Caves, L.S., Groom, C.R., and Hubbard, R.E. (1999). Locating interaction sites on proteins: the crystal structure of thermolysin soaked in 2% to 100% isopropanol. *Proteins* 37, 628–640.

Eppler, R.K., Komor, R.S., Huynh, J., Dordick, J.S., Reimer, J.A., and Clark, D.S. (2006). Water dynamics and salt-activation of enzymes in organic media: mechanistic implications revealed by NMR spectroscopy. *Proc Natl Acad Sci USA* 103, 5706–5710.

Eyre-Walker, A. (2006). The genomic rate of adaptive evolution. *Trends Ecol Evol (Amst)* 21, 569–575.

Fasano, M., Curry, S., Terreno, E., Galliano, M., Fanali, G., Narciso, P., Notari, S., and Ascenzi, P. (2005). The extraordinary ligand binding properties of

human serum albumin. *IUBMB Life* 57, 787–796.

Feller, G. (2010). Protein stability and enzyme activity at extreme biological temperatures. *J. Phys. Condens. Matter* 22, 323101.

Fernández-Gacio, A., Codina, A., Fastrez, J., Riant, O., and Soumillion, P. (2006). Transforming carbonic anhydrase into epoxide synthase by metal exchange. *Chembiochem* 7, 1013–1016.

Finn, R.D., Coghill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., et al. (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279-85.

Fischer, E. (1894). Einfluss der Configuration auf die Wirkung der Enzyme. *Ber. Dtsch. Chem. Ges.* 27, 2985–2993.

Fleury, F., Ianoul, A., Berjot, M., Feofanov, A., Alix, A.J., and Nabiev, I. (1997). Camptothecin-binding site in human serum albumin and protein transformations induced by drug binding. *FEBS Lett.* 411, 215–220.

Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* 254, 1598–1603.

Furnham, N., Holliday, G.L., de Beer, T.A.P., Jacobsen, J.O.B., Pearson, W.R., and Thornton, J.M. (2014). The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes. *Nucleic Acids Res.* 42, D485-9.

Gao, M., and Skolnick, J. (2013). A comprehensive survey of small-molecule binding pockets in proteins. *PLoS Comput. Biol.* 9, e1003302.

Gao, X.G., Maldonado, E., Pérez-Montfort, R., Garza-Ramos, G., de Gómez-Puyou, M.T., Gómez-Puyou, A., and Rodríguez-Romero, A. (1999). Crystal structure of triosephosphate isomerase from *Trypanosoma cruzi* in hexane. *Proc Natl Acad Sci USA* 96, 10062–10067.

Gerlt, J.A., and Babbitt, P.C. (1998). Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* 2, 607–612.

- Gerstein, M., and Krebs, W. (1998). A database of macromolecular motions. *Nucleic Acids Res.* 26, 4280–4290.
- Gerstein, M., Lesk, A.M., and Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry* 33, 6739–6749.
- Gherardini, P.F., Wass, M.N., Helmer-Citterich, M., and Sternberg, M.J.E. (2007). Convergent evolution of enzyme active sites is not a rare phenomenon. *J. Mol. Biol.* 372, 817–845.
- Giuseppe, M., Luigia, M., Elena, P., Yan, F., and Luigi, M. (2012). Enzyme promiscuity in the hormone-sensitive lipase family of proteins. *Protein Pept. Lett.* 19, 144–154.
- Gómez, A., Domedel, N., Cedano, J., Piñol, J., and Querol, E. (2003). Do current sequence analysis algorithms disclose multifunctional (moonlighting) proteins? *Bioinformatics* 19, 895–896.
- Goncharov, N.V., Belinskaia, D.A., Shmurak, V.I., Terpilowski, M.A., Jenkins, R.O., and Avdonin, P.V. (2017). Serum Albumin Binding and Esterase Activity: Mechanistic Interactions with Organophosphates. *Molecules* 22.
- Gora, A., Brezovsky, J., and Damborsky, J. (2013). Gates of enzymes. *Chem. Rev.* 113, 5871–5923.
- Goubran Botros, H., Gregoire, C., Rabillon, J., David, B., and Dandeu, J.P. (1996). Cross-antigenicity of horse serum albumin with dog and cat albumins: study of three short peptides with significant inhibitory activity towards specific human IgE and IgG antibodies. *Immunology* 88, 340–347.
- Gould, S.M., and Tawfik, D.S. (2005). Directed evolution of the promiscuous esterase activity of carbonic anhydrase II. *Biochemistry* 44, 5444–5452.
- Granata, D., Ponzoni, L., Micheletti, C., and Carnevale, V. (2017). Patterns of coevolving amino acids unveil structural and dynamical domains. *Proc Natl Acad Sci USA* 114, E10612–E10621.
- Gsponer, J., and Babu, M.M. (2009). The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* 99, 94–103.

Guengerich, F.P. (1991). Reactions and significance of cytochrome P-450 enzymes. *J. Biol. Chem.* **266**, 10019–10022.

Guisan, J.M., Sabuquillo, P., Fernandez-Lafuente, R., Fernandez-Lorente, G., Mateo, C., Halling, P.J., Kennedy, D., Miyata, E., and Re, D. (2001). Preparation of new lipases derivatives with high activity–stability in anhydrous media: adsorption on hydrophobic supports plus hydrophilization with polyethylenimine. *Journal of Molecular Catalysis B: Enzymatic* **11**, 817–824.

Gutteridge, A., and Thornton, J.M. (2005a). Understanding nature's catalytic toolkit. *Trends Biochem. Sci.* **30**, 622–629.

Gutteridge, A., and Thornton, J. (2004). Conformational change in substrate binding, catalysis and product release: an open and shut case? *FEBS Lett.* **567**, 67–73.

Gutteridge, A., and Thornton, J. (2005b). Conformational changes observed in enzyme crystal structures upon substrate binding. *J. Mol. Biol.* **346**, 21–28.

Gu, Y., Li, D.-W., and Brüschweiler, R. (2015). Decoding the mobility and time scales of protein loops. *J. Chem. Theory Comput.* **11**, 1308–1314.

Hammes-Schiffer, S., and Benkovic, S.J. (2006). Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75**, 519–541.

Ha, C.-E., and Bhagavan, N.V. (2013). Novel insights into the pleiotropic effects of human serum albumin in health and disease. *Biochim. Biophys. Acta* **1830**, 5486–5493.

Herschlag, D. (1988). The role of induced fit and conformational changes of enzymes in specificity and catalysis. *Bioorg. Chem.* **16**, 62–96.

Hilger, C., Kohnen, M., Grigioni, F., Lehnert, C., and Hentges, F. (1997). Allergic cross-reactions between cat and pig serum albumin. Study at the protein and DNA levels. *Allergy* **52**, 179–187.

Holliday, G.L., Mitchell, J.B.O., and Thornton, J.M. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.* **390**, 560–577.

Hotta, K., Lange, H., Tantillo, D.J., Houk, K.N., Hilvert, D., and Wilson, I.A. (2000). Catalysis of decarboxylation by a preorganized heterogeneous microenvironment: crystal structures of abzyme 21D8. *J. Mol. Biol.* 302, 1213–1225.

Ho, J.X., Holowachuk, E.W., Norton, E.J., Twigg, P.D., and Carter, D.C. (1993). X-ray and primary structure of horse serum albumin (*Equus caballus*) at 0.27-nm resolution. *Eur. J. Biochem.* 215, 205–212.

Hua, X., Xing, Y., and Zhang, X. (2016). Enhanced Promiscuity of Lipase-Inorganic Nanocrystal Composites in the Epoxidation of Fatty Acids in Organic Media. *ACS Appl. Mater. Interfaces* 8, 16257–16261.

Hubbard (1992). www.bioinf.manchester.ac.uk/naccess/nac_intro.html.

Hughes, A.L. (1994). The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256, 119–124.

Hult, K., and Berglund, P. (2007). Enzyme promiscuity: mechanism and applications. *Trends Biotechnol.* 25, 231–238.

Hurst, R., Bao, Y., Ridley, S., and Williamson, G. (1999). Phospholipid hydroperoxide cysteine peroxidase activity of human serum albumin. *Biochem. J.* 338 (Pt 3), 723–728.

Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* 86, 9355–9359.

James, L.C., and Tawfik, D.S. (2001). Catalytic and binding poly-reactivities shared by two unrelated proteins: The potential role of promiscuity in enzyme evolution. *Protein Sci.* 10, 2600–2607.

James, L.C., Roversi, P., and Tawfik, D.S. (2003). Antibody multispecificity mediated by conformational diversity. *Science* 299, 1362–1367.

Janssen, D.B., Dinkla, I.J.T., Poelarends, G.J., and Terpstra, P. (2005). Bacterial degradation of xenobiotic compounds: evolution and distribution of

novel enzyme activities. *Environ. Microbiol.* 7, 1868–1882.

Javier Zea, D., Miguel Monzon, A., Fornasari, M.S., Marino-Buslje, C., and Parisi, G. (2013). Protein conformational diversity correlates with evolutionary rate. *Mol. Biol. Evol.* 30, 1500–1503.

Jeffery, C.J. (2003). Moonlighting proteins: old proteins learning new tricks. *Trends Genet.* 19, 415–417.

Jehl, P., Manguy, J., Shields, D.C., Higgins, D.G., and Davey, N.E. (2016). ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Res.* 44, W11-5.

Jensen, R.A. (1976). Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30, 409–425.

Jia, H.-Y., Zong, M.-H., Yu, H.-L., and Li, N. (2017). Dehydrogenase-Catalyzed Oxidation of Furanics: Exploitation of Hemoglobin Catalytic Promiscuity. *ChemSusChem* 10, 3524–3528.

Jolly, M.K., Kulkarni, P., Weninger, K., Orban, J., and Levine, H. (2018). Phenotypic Plasticity, Bet-Hedging, and Androgen Independence in Prostate Cancer: Role of Non-Genetic Heterogeneity. *Front. Oncol.* 8, 50.

Juritz, E., Palopoli, N., Fornasari, M.S., Fernandez-Alberti, S., and Parisi, G. (2013). Protein conformational diversity modulates sequence divergence. *Mol. Biol. Evol.* 30, 79–87.

Kahraman, A., Morris, R.J., Laskowski, R.A., and Thornton, J.M. (2007). Shape variation in protein binding pockets and their ligands. *J. Mol. Biol.* 368, 283–301.

Kamerlin, S.C.L., and Warshel, A. (2010). At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins* 78, 1339–1375.

Kanehisa, M., and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28–35.

Karush, F. (1950). Heterogeneity of the Binding Sites of Bovine Serum Albumin. *J. Am. Chem. Soc.* 72, 2705–2713.

- Kazlauskas, R.J. (2005). Enhancing catalytic promiscuity for biocatalysis. *Curr. Opin. Chem. Biol.* 9, 195–201.
- Keeling, C.I., Weisshaar, S., Lin, R.P.C., and Bohlmann, J. (2008). Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proc Natl Acad Sci USA* 105, 1085–1090.
- Kern, D., Eisenmesser, E.Z., and Wolf-Watz, M. (2005). Enzyme dynamics during catalysis measured by NMR spectroscopy. *Meth. Enzymol.* 394, 507–524.
- Ke, T., and Klibanov, A.M. (1998). On enzymatic activity in organic solvents as a function of enzyme history. *Biotechnol. Bioeng.* 57, 746–750.
- Khersonsky, O., and Tawfik, D.S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79, 471–505.
- Khersonsky, O., Roodveldt, C., and Tawfik, D.S. (2006). Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol.* 10, 498–508.
- Kimura, M. (1991). The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet* 66, 367–386.
- Kim, D., Cho, C.-H., Cho, Y., Ryu, J., Bhak, J., and Kim, D.-S. (2008). Pocket extraction on proteins via the Voronoi diagram of spheres. *J. Mol. Graph. Model.* 26, 1104–1112.
- Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., et al. (2016). PubChem Substance and Compound databases. *Nucleic Acids Res.* 44, D1202-13.
- Kingsley, L.J., and Lill, M.A. (2015). Substrate tunnels in enzymes: structure-function relationships and computational methodology. *Proteins* 83, 599–611.
- King, T.P., and Spencer, M. (1970). Structural studies and organic ligand-binding properties of bovine plasma albumin. *J. Biol. Chem.* 245, 6134–6148.
- Klibanov, A.M. (1989). Enzymatic catalysis in anhydrous organic solvents.

Trends Biochem. Sci. 14, 141–144.

Klibanov, A.M. (2001). Improving enzymes by using them in organic solvents. *Nature* 409, 241–246.

Konstantinopoulos, P.A., Karamouzis, M.V., and Papavassiliou, A.G. (2007). Post-translational modifications and regulation of the RAS superfamily of GTPases as anticancer targets. *Nat. Rev. Drug Discov.* 6, 541–555.

Kosa, T., Maruyama, T., and Otagiri, M. (1997). Species differences of serum albumins: I. Drug binding sites. *Pharm. Res.* 14, 1607–1612.

Koshland, D.E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc Natl Acad Sci USA* 44, 98–104.

Koshland, D.E. (1995). The key–lock theory and the induced fit theory. *Angew. Chem. Int. Ed. Engl.* 33, 2375–2378.

Koshland, D.E. (1998). Conformational changes: how small is big enough? *Nat. Med.* 4, 1112–1114.

Koshland, D.E., Ray, W.J., and Erwin, M.J. (1958). Protein structure and enzyme action. *Fed. Proc.* 17, 1145–1150.

Kossel, A. (1898). Ueber die Constitution der einfachsten Eiweissstoffe. *Hoppe-Seyler's Zeitschrift Für Physiologische Chemie* 25, 165–189.

Kremer, A., and Li, S.-M. (2008). Tryptophan aminopeptidase activity of several indole prenyltransferases from *Aspergillus fumigatus*. *Chem. Biol.* 15, 729–738.

Kumar, S., Ma, B., Tsai, C.J., Sinha, N., and Nussinov, R. (2000). Folding and binding cascades: dynamic landscapes and population shifts. *Protein Sci.* 9, 10–19.

Kuznetsova, E., Proudfoot, M., Gonzalez, C.F., Brown, G., Omelchenko, M.V., Borozan, I., Carmel, L., Wolf, Y.I., Mori, H., Savchenko, A.V., et al. (2006). Genome-wide analysis of substrate specificities of the *Escherichia coli* haloacid dehalogenase-like phosphatase family. *J. Biol. Chem.* 281, 36149–36161.

Lario, P.I., Sampson, N., and Vrielink, A. (2003). Sub-atomic resolution crystal structure of cholesterol oxidase: what atomic resolution crystallography reveals

about enzyme mechanism and the role of the FAD cofactor in redox activity. *J. Mol. Biol.* **326**, 1635–1650.

Lee, B., and Richards, F.M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.

van der Lee, R., Buljan, M., Lang, B., Weatheritt, R.J., Daughdrill, G.W., Dunker, A.K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D.T., et al. (2014). Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114**, 6589–6631.

Le Guilloux, V., Schmidtke, P., and Tuffery, P. (2009). Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168.

Liang, J., Edelsbrunner, H., and Woodward, C. (1998). Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* **7**, 1884–1897.

Liese, A., and Filho, M.V. (1999). Production of fine chemicals using biocatalysis. *Curr. Opin. Biotechnol.* **10**, 595–603.

Lima, V.M.G., Krieger, N., Mitchell, D.A., and Fontana, J.D. (2004). Activity and stability of a crude lipase from *Penicillium aurantiogriseum* in aqueous media and organic solvents. *Biochem. Eng. J.* **18**, 65–71.

Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., and Russell, R.B. (2003). Protein disorder prediction: implications for structural proteomics. *Structure* **11**, 1453–1459.

Lipman, D.J., and Wilbur, W.J. (1991). Modelling neutral and selective evolution of protein folding. *Proc. Biol. Sci.* **245**, 7–11.

Li, M., McAuley, E., Zhang, Y., Kong, L., Yang, F., Zhou, Z., Wu, X., and Liang, H. (2014). Comparison of binding characterization of two antiviral drugs to human serum albumin. *Chem. Biol. Drug Des.* **83**, 576–582.

van Loo, B., Jonas, S., Babbie, A.C., Benjdia, A., Berteau, O., Hyvönen, M., and Hollfelder, F. (2010). An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily. *Proc Natl Acad Sci USA* **107**, 2740–2745.

López-Iglesias, M., and Gotor-Fernández, V. (2015). Recent Advances in Biocatalytic Promiscuity: Hydrolase-Catalyzed Reactions for Nonconventional Transformations. *Chem. Rec.* **15**, 743–759.

López-Iglesias, M., Busto, E., Gotor, V., and Gotor-Fernández, V. (2012). Stereoselective synthesis of 2,3-disubstituted indoline diastereoisomers by chemoenzymatic processes. *J. Org. Chem.* **77**, 8049–8055.

Lou, F.-W., Liu, B.-K., Wu, Q., Lv, D.-S., and Lin, X.-F. (2008). *Candida antarctica* Lipase B (CAL-B)-Catalyzed Carbon-Sulfur Bond Addition and Controllable Selectivity in Organic Media. *Adv. Synth. Catal.* **350**, 1959–1962.

Luo, J., van Loo, B., and Kamerlin, S.C.L. (2012). Catalytic promiscuity in *Pseudomonas aeruginosa* arylsulfatase as an example of chemistry-driven protein evolution. *FEBS Lett.* **586**, 1622–1630.

Magnusson, A.O., Takwa, M., Hamberg, A., and Hult, K. (2005). An S-selective lipase was created by rational redesign and the enantioselectivity increased with temperature. *Angew Chem Int Ed Engl* **44**, 4582–4585.

Marchetti, J., Monzon, A.M., Tosatto, S.C.E., Parisi, G., and Fornasari, M.S. (2019). Ensembles from Ordered and Disordered Proteins Reveal Similar Structural Constraints during Evolution. *J. Mol. Biol.* **431**, 1298–1307.

Maria-Solano, M.A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J., and Osuna, S. (2018). Role of conformational dynamics in the evolution of novel enzyme function. *Chem Commun (Camb)* **54**, 6622–6634.

Marques, S.M., Bednar, D., and Damborsky, J. (2018). Computational Study of Protein-Ligand Unbinding for Enzyme Engineering. *Front. Chem.* **6**, 650.

Marsh, J.A., and Teichmann, S.A. (2014). Parallel dynamics and evolution: Protein conformational fluctuations and assembly reflect evolutionary changes in sequence and structure. *Bioessays* **36**, 209–218.

di Masi, A., Gullotta, F., Bolli, A., Fanali, G., Fasano, M., and Ascenzi, P. (2011). Ibuprofen binding to secondary sites allosterically modulates the spectroscopic and catalytic properties of human serum heme-albumin. *FEBS J.*

278, 654–662.

Masson, P., Froment, M.-T., Darvesh, S., Schopfer, L.M., and Lockridge, O. (2007). Aryl acylamidase activity of human serum albumin with *p*-nitrotrifluoroacetanilide as the substrate. *J. Enzyme Inhib. Med. Chem.* **22**, 463–469.

Matthews, B.W., and Liu, L. (2009). A review about nothing: are apolar cavities in proteins really empty? *Protein Sci.* **18**, 494–502.

Mattos, C., and Ringe, D. (2001). Proteins in organic solvents. *Curr. Opin. Struct. Biol.* **11**, 761–764.

Ma, B., Shatsky, M., Wolfson, H.J., and Nussinov, R. (2002). Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. *Protein Sci.* **11**, 184–197.

McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777–793.

McGovern, S.L., and Shoichet, B.K. (2003). Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **46**, 2895–2907.

McLachlan, A.D. (1982). Rapid comparison of protein structures. *Acta Cryst. A* **38**, 871–873.

Menger, F.M., and Nome, F. (2019). Interaction vs preorganization in enzyme catalysis. A dispute that calls for resolution. *ACS Chem. Biol.* **14**, 1386–1392.

Mesecar, A.D., Stoddard, B.L., and Koshland, D.E. (1997). Orbital steering in the catalytic power of enzymes: small structural changes with large catalytic consequences. *Science* **277**, 202–206.

Meunier, B., de Visser, S.P., and Shaik, S. (2004). Mechanism of oxidation reactions catalyzed by cytochrome p450 enzymes. *Chem. Rev.* **104**, 3947–3980.

Miao, Y., Rahimi, M., Geertsema, E.M., and Poelarends, G.J. (2015). Recent developments in enzyme promiscuity for carbon-carbon bond-forming reactions.

Curr. Opin. Chem. Biol. 25, 115–123.

Michaelis, L., and Menten, M.M.L. (2013). The kinetics of invertin action. 1913. FEBS Lett. 587, 2712–2720.

Milburn, M.V., Privé, G.G., Milligan, D.L., Scott, W.G., Yeh, J., Jancarik, J., Koshland, D.E., and Kim, S.H. (1991). Three-dimensional structures of the ligand-binding domain of the bacterial aspartate receptor with and without a ligand. Science 254, 1342–1347.

Miller, B.G., and Raines, R.T. (2004). Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. Biochemistry 43, 6387–6392.

Min, W., Gopich, I.V., English, B.P., Kou, S.C., Xie, X.S., and Szabo, A. (2006). When does the michaelis–menten equation hold for fluctuating enzymes? J. Phys. Chem. B 110, 20093–20097.

Mirsky, A.E., and Pauling, L. (1936). On the structure of native, denatured, and coagulated proteins. Proc Natl Acad Sci USA 22, 439–447.

Miyata, T., Iwabe, N., Kuma, K., Kawanishi, Y., Hasegawa, M., Kishino, H., Mukohata, Y., Ihara, K., and Osawa, S. (1991). Evolution of archaebacteria: phylogenetic relationships among archaebacteria, eubacteria, and eukaryotes. In Evolution of Life, S. Osawa, and T. Honjo, eds. (Tokyo: Springer Japan), pp. 337–351.

Monod, J., Wyman, J., and Changeux, J.P. (1965). On the nature of allosteric transitions: a plausible model. J. Mol. Biol. 12, 88–118.

Monzon, A.M., Juritz, E., Fornasari, M.S., and Parisi, G. (2013). CoDNaS: a database of conformational diversity in the native state of proteins. Bioinformatics 29, 2512–2514.

Monzon, A.M., Rohr, C.O., Fornasari, M.S., and Parisi, G. (2016). CoDNaS 2.0: a comprehensive database of protein conformational diversity in the native state. Database (Oxford) 2016.

Monzon, A.M., Zea, D.J., Fornasari, M.S., Saldaño, T.E., Fernandez-Alberti, S.,

- Tosatto, S.C.E., and Parisi, G. (2017). Conformational diversity analysis reveals three functional mechanisms in proteins. *PLoS Comput. Biol.* 13, e1005398.
- Morris, G.M., and Lim-Wilby, M. (2008). Molecular docking. *Methods Mol. Biol.* 443, 365–382.
- Murga, C., Arcones, A.C., Cruces-Sande, M., Briones, A.M., Salaices, M., and Mayor, F. (2019). G Protein-Coupled Receptor Kinase 2 (GRK2) as a Potential Therapeutic Target in Cardiovascular and Metabolic Diseases. *Front. Pharmacol.* 10, 112.
- Murga, L.F., Ondrechen, M.J., and Ringe, D. (2008). Prediction of interaction sites from apo 3D structures when the holo conformation is different. *Proteins* 72, 980–992.
- Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8, e1002764.
- Najmanovich, R.J., Allali-Hassani, A., Morris, R.J., Dombrovsky, L., Pan, P.W., Vedadi, M., Plotnikov, A.N., Edwards, A., Arrowsmith, C., and Thornton, J.M. (2007). Analysis of binding site similarity, small-molecule similarity and experimental binding profiles in the human cytosolic sulfotransferase family. *Bioinformatics* 23, e104-9.
- Nam, H., Lewis, N.E., Lerman, J.A., Lee, D.-H., Chang, R.L., Kim, D., and Palsson, B.O. (2012). Network context and selection in the evolution to enzyme specificity. *Science* 337, 1101–1104.
- Nashine, V.C., Hammes-Schiffer, S., and Benkovic, S.J. (2010). Coupled motions in enzyme catalysis. *Curr. Opin. Chem. Biol.* 14, 644–651.
- Nath, A., and Atkins, W.M. (2008). A quantitative index of substrate promiscuity. *Biochemistry* 47, 157–166.
- Nemoto, W., Saito, A., and Oikawa, H. (2013). Recent advances in functional region prediction by using structural and evolutionary information - Remaining problems and future extensions. *Comput. Struct. Biotechnol. J.* 8, e201308007.

- Newton, M.S., Arcus, V.L., Gerth, M.L., and Patrick, W.M. (2018). Enzyme evolution: innovation is easy, optimization is complicated. *Curr. Opin. Struct. Biol.* **48**, 110–116.
- Nianios, D., Thierbach, S., Steimer, L., Lulchev, P., Klostermeier, D., and Fetzner, S. (2015). Nickel quercetinase, a “promiscuous” metalloenzyme: metal incorporation and metal ligand substitution studies. *BMC Biochem.* **16**, 10.
- Nobeli, I., Favia, A.D., and Thornton, J.M. (2009). Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **27**, 157–167.
- Nussinov, R., Tsai, C.-J., and Jang, H. (2019). Protein ensembles link genotype to phenotype. *PLoS Comput. Biol.* **15**, e1006648.
- Ohno, S. (1970). *Evolution by Gene Duplication* (Berlin, Heidelberg: Springer Berlin Heidelberg).
- Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M., and Jensen, J.H. (2011). PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK Predictions. *J. Chem. Theory Comput.* **7**, 525–537.
- Orengo, C.A., Flores, T.P., Taylor, W.R., and Thornton, J.M. (1993). Identification and classification of protein fold families. *Protein Eng. Des. Sel.* **6**, 485–500.
- Orengo, C.A., Todd, A.E., and Thornton, J.M. (1999). From protein structure to function. *Curr. Opin. Struct. Biol.* **9**, 374–382.
- Ortiz, A.R., Strauss, C.E.M., and Olmea, O. (2002). MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. *Protein Sci.* **11**, 2606–2621.
- Osman, D., Martini, M.A., Foster, A.W., Chen, J., Scott, A.J.P., Morton, R.J., Steed, J.W., Lurie-Luke, E., Huggins, T.G., Lawrence, A.D., et al. (2019). Bacterial sensors define intracellular free energies for correct enzyme metalation. *Nat. Chem. Biol.* **15**, 241–249.
- Otyepka, M., Berka, K., and Anzenbacher, P. (2012). Is there a relationship between the substrate preferences and structural flexibility of cytochromes

P450? *Curr. Drug Metab.* **13**, 130–142.

Outten, C.E., and O'Halloran, T.V. (2001). Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science* **292**, 2488–2492.

Oxford, G. (2003). Ras superfamily monomeric G proteins in carcinoma cell motility. *Cancer Lett.* **189**, 117–128.

O'Brien, P.J., and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105.

Pabis, A., and Kamerlin, S.C.L. (2016). Promiscuity and electrostatic flexibility in the alkaline phosphatase superfamily. *Curr. Opin. Struct. Biol.* **37**, 14–21.

Pace, C.N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S.R., Schell, D., Thurlkill, R.L., Imura, S., Scholtz, J.M., Gajiwala, K., et al. (2014). Contribution of hydrogen bonds to protein stability. *Protein Sci.* **23**, 652–661.

Panjehshahin, M.R., Yates, M.S., and Bowmer, C.J. (1992). A comparison of drug binding sites on mammalian albumins. *Biochem. Pharmacol.* **44**, 873–879.

Paoli, M. (2001). Protein folds propelled by diversity. *Prog. Biophys. Mol. Biol.* **76**, 103–130.

Parisi, G., Zea, D.J., Monzon, A.M., and Marino-Buslje, C. (2015). Conformational diversity and the emergence of sequence signatures during evolution. *Curr. Opin. Struct. Biol.* **32**, 58–65.

Patrick, W.M., and Matsumura, I. (2008). A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* **377**, 323–336.

Patrick, W.M., Quandt, E.M., Swartzlander, D.B., and Matsumura, I. (2007). Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* **24**, 2716–2722.

Peregrín-Alvarez, J.M., Sanford, C., and Parkinson, J. (2009). The conservation and evolutionary modularity of metabolism. *Genome Biol.* **10**, R63.

- Perugino, G., Trincone, A., Rossi, M., and Moracci, M. (2004). Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol.* 22, 31–37.
- Perutz, M.F., and Mathews, F.S. (1966). An x-ray study of azide methaemoglobin. *J. Mol. Biol.* 21, 199–202.
- Peters, T. (1977). Serum albumin: recent progress in the understanding of its structure and biosynthesis. *Clin. Chem.* 23, 5–12.
- Peters, T. (1995). Genetics. In *All About Albumin*, (Elsevier), pp. 133–187.
- Ping, J., Hao, P., Li, Y.-X., and Wang, J.-F. (2013). Molecular dynamics studies on the conformational transitions of adenylate kinase: a computational evidence for the conformational selection mechanism. *Biomed Res. Int.* 2013, 628536.
- Piovesan, D., Minervini, G., and Tosatto, S.C.E. (2016). The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 44, W367-74.
- Plapp, B.V. (2010). Conformational changes and catalysis by alcohol dehydrogenase. *Arch. Biochem. Biophys.* 493, 3–12.
- Pond, S.L.K., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Porter, T.D., and Coon, M.J. (1991). Cytochrome P-450. Multiplicity of isoforms, substrates, and catalytic and regulatory mechanisms. *J. Biol. Chem.* 266, 13469–13472.
- Porter, C.T., Bartlett, G.J., and Thornton, J.M. (2004). The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.* 32, D129-33.
- Pravda, L., Berka, K., Svobodová Vařeková, R., Sehnal, D., Banáš, P., Laskowski, R.A., Koča, J., and Otyepka, M. (2014). Anatomy of enzyme channels. *BMC Bioinformatics* 15, 379.
- Pubmeddev Structural basis of thermostability in hyperthermophilic proteins, or “there’s more than one way to skin a cat”.
- Pundir, S., Martin, M.J., and O’Donovan, C. (2017). UniProt Protein

Knowledgebase. *Methods Mol. Biol.* **1558**, 41–55.

Qian, C., Xu, J.-M., Wu, Q., Lv, D.-S., and Lin, X.-F. (2007). Promiscuous acylase-catalyzed aza-Michael additions of aromatic N-heterocycles in organic solvent. *Tetrahedron Lett.* **48**, 6100–6104.

Radzicka, A., and Wolfenden, R. (1995). A proficient enzyme. *Science* **267**, 90–93.

Raskatov, J.A., and Teplow, D.B. (2017). Using chirality to probe the conformational dynamics and assembly of intrinsically disordered amyloid proteins. *Sci. Rep.* **7**, 12433.

Raushel, F.M., and Holden, H.M. (2000). Phosphotriesterase: an enzyme in search of its natural substrate. *Adv. Enzymol. Relat. Areas Mol. Biol.* **74**, 51–93.

Reetz, M.T., Mondière, R., and Carballeira, J.D. (2007). Enzyme promiscuity: first protein-catalyzed Morita–Baylis–Hillman reaction. *Tetrahedron Lett.* **48**, 1679–1681.

Rego, N., and Koes, D. (2015). 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324.

Rittle, J., and Green, M.T. (2010). Cytochrome P450 compound I: capture, characterization, and C-H bond activation kinetics. *Science* **330**, 933–937.

Rueda, A.J.V., Monzon, A.M., Ardanaz, S.M., Iglesias, L.E., and Parisi, G. (2018). Large scale analysis of protein conformational transitions from aqueous to non-aqueous media. *BMC Bioinformatics* **19**, 27.

Ru, M.T., Hirokane, S.Y., Lo, A.S., Dordick, J.S., Reimer, J.A., and Clark, D.S. (2000). On the Salt-Induced Activation of Lyophilized Enzymes in Organic Solvents: Effect of Salt Kosmotropicity on Enzyme Activity. *J. Am. Chem. Soc.* **122**, 1565–1571.

Saldaño, T.E., Monzon, A.M., Parisi, G., and Fernandez-Alberti, S. (2016). Evolutionary conserved positions define protein conformational diversity. *PLoS Comput. Biol.* **12**, e1004775.

Schleinkofer, K., Sudarko, Winn, P.J., Lüdemann, S.K., and Wade, R.C. (2005).

Do mammalian cytochrome P450s show multiple ligand access pathways and ligand channelling? *EMBO Rep.* 6, 584–589.

Schmidt, D.M.Z., Mundorff, E.C., Dojka, M., Bermudez, E., Ness, J.E., Govindarajan, S., Babbitt, P.C., Minshull, J., and Gerlt, J.A. (2003). Evolutionary potential of (beta/alpha)₈-barrels: functional promiscuity produced by single substitutions in the enolase superfamily. *Biochemistry* 42, 8387–8393.

Schmitke, J.L., Stern, L.J., and Klibanov, A.M. (1997). The crystal structure of subtilisin Carlsberg in anhydrous dioxane and its comparison with those in water and acetonitrile. *Proc Natl Acad Sci USA* 94, 4250–4255.

Scott, C., Jackson, C.J., Coppin, C.W., Mourant, R.G., Hilton, M.E., Sutherland, T.D., Russell, R.J., and Oakeshott, J.G. (2009). Catalytic improvement and evolution of atrazine chlorohydrolase. *Appl. Environ. Microbiol.* 75, 2184–2191.

Sehna, D., Svobodová Vařeková, R., Berka, K., Pravda, L., Navrátilová, V., Banáš, P., Ionescu, C.-M., Otyepka, M., and Koča, J. (2013). MOLE 2.0: advanced approach for analysis of biomacromolecular channels. *J. Cheminform.* 5, 39.

Serdakowski, A.L., and Dordick, J.S. (2008). Enzyme activation for organic solvents made easy. *Trends Biotechnol.* 26, 48–54.

Sharma, U.K., Sharma, N., Kumar, R., and Sinha, A.K. (2013). Biocatalysts for multicomponent Biginelli reaction: bovine serum albumin triggered waste-free synthesis of 3,4-dihydropyrimidin-2-(1H)-ones. *Amino Acids* 44, 1031–1037.

Shima, F., Ijiri, Y., Muraoka, S., Liao, J., Ye, M., Araki, M., Matsumoto, K., Yamamoto, N., Sugimoto, T., Yoshikawa, Y., et al. (2010). Structural basis for conformational dynamics of GTP-bound Ras protein. *J. Biol. Chem.* 285, 22696–22705.

Siddiq, M.A., Hochberg, G.K., and Thornton, J.W. (2017). Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* 47, 113–122.

Sikosek, T., Krobath, H., and Chan, H.S. (2016). Theoretical Insights into the Biophysics of Protein Bi-stability and Evolutionary Switches. *PLoS Comput.*

Biol. 12, e1004960.

Sillitoe, I., Dawson, N., Thornton, J., and Orengo, C. (2015). The history of the CATH structural classification of protein domains. *Biochimie* 119, 209–217.

Sitzmann, M., Weidlich, I.E., Filippov, I.V., Liao, C., Peach, M.L., Ihlenfeldt, W.-D., Karki, R.G., Borodina, Y.V., Cachau, R.E., and Nicklaus, M.C. (2012). PDB ligand conformational energies calculated quantum-mechanically. *J. Chem. Inf. Model.* 52, 739–756.

Soo, V.W.C., Yosaatmadja, Y., Squire, C.J., and Patrick, W.M. (2016). Mechanistic and Evolutionary Insights from the Reciprocal Promiscuity of Two Pyridoxal Phosphate-dependent Enzymes. *J. Biol. Chem.* 291, 19873–19887.

Spector, A.A., John, K., and Fletcher, J.E. (1969). Binding of long-chain fatty acids to bovine serum albumin. *J. Lipid Res.* 10, 56–67.

Spiller, B., Gershenson, A., Arnold, F.H., and Stevens, R.C. (1999). A structural view of evolutionary divergence. *Proc Natl Acad Sci USA* 96, 12305–12310.

Spitzauer, S. (1999). Allergy to mammalian proteins: at the borderline between foreign and self? *Int. Arch. Allergy Immunol.* 120, 259–269.

Stainbrook, S.C., Yu, J.S., Reddick, M.P., Bagheri, N., and Tyo, K.E.J. (2017). Modulating and evaluating receptor promiscuity through directed evolution and modeling. *Protein Eng. Des. Sel.* 30, 455–465.

Stepankova, V., Khabiri, M., Brezovsky, J., Pavelka, A., Sykora, J., Amaro, M., Minofar, B., Prokop, Z., Hof, M., Ettrich, R., et al. (2013). Expansion of access tunnels and active-site cavities influence activity of haloalkane dehalogenases in organic cosolvents. *Chembiochem* 14, 890–897.

Stockwell, G.R., and Thornton, J.M. (2006). Conformational diversity of ligands bound to proteins. *J. Mol. Biol.* 356, 928–944.

Studer, R.A., Penel, S., Duret, L., and Robinson-Rechavi, M. (2008). Pervasive positive selection on duplicated and non duplicated vertebrate protein coding genes. *Genome Res.* 18, 1393–1402.

Sułkowska, A. (2002). Interaction of drugs with bovine and human serum

albumin. *J. Mol. Struct.* **614**, 227–232.

Tennant, J.P., Waldner, F., Jacques, D.C., Masuzzo, P., Collister, L.B., and Hartgerink, C.H.J. (2016). The academic, economic and societal impacts of Open Access: an evidence-based review. [version 3; peer review: 4 approved, 1 approved with reservations]. *F1000Res.* **5**, 632.

The UniProt Consortium (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **45**, D158–D169.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

Thorpe, I.F., and Brooks, C.L. (2003). Barriers to Hydride Transfer in Wild Type and Mutant Dihydrofolate Reductase from *E. coli*. *J. Phys. Chem. B* **107**, 14042–14051.

Todd, A.E., Orengo, C.A., and Thornton, J.M. (2001). Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **307**, 1113–1143.

Tokuriki, N., and Tawfik, D.S. (2009). Protein dynamism and evolvability. *Science* **324**, 203–207.

Tompa, P., Szász, C., and Buday, L. (2005). Structural disorder throws new light on moonlighting. *Trends Biochem. Sci.* **30**, 484–489.

Tompa, P., Schad, E., Tantos, A., and Kalmar, L. (2015). Intrinsically disordered proteins: emerging interaction specialists. *Curr. Opin. Struct. Biol.* **35**, 49–59.

Touw, W.G., Baakman, C., Black, J., te Beek, T.A.H., Krieger, E., Joosten, R.P., and Vriend, G. (2015). A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* **43**, D364–8.

Trott, O., and Olson, A.J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461.

- Tsai, C.J., Kumar, S., Ma, B., and Nussinov, R. (1999a). Folding funnels, binding funnels, and protein function. *Protein Sci.* 8, 1181–1190.
- Tsai, C.J., Ma, B., and Nussinov, R. (1999b). Folding and binding cascades: shifts in energy landscapes. *Proc Natl Acad Sci USA* 96, 9970–9972.
- Tseng, Y.Y., and Liang, J. (2007). Predicting enzyme functional surfaces and locating key residues automatically from structures. *Ann. Biomed. Eng.* 35, 1037–1042.
- Uversky, V.N. (2015). Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* 282, 1182–1189.
- Varshney, A., Sen, P., Ahmad, E., Rehan, M., Subbarao, N., and Khan, R.H. (2010). Ligand binding strategies of human serum albumin: how can the cargo be utilized? *Chirality* 22, 77–87.
- Wagner, A. (2011). The molecular origins of evolutionary innovations. *Trends Genet.* 27, 397–410.
- Wallon, G., Kryger, G., Lovett, S.T., Oshima, T., Ringe, D., and Petsko, G.A. (1997). Crystal structures of *Escherichia coli* and *Salmonella typhimurium* 3-isopropylmalate dehydrogenase and comparison with their thermophilic counterpart from *Thermus thermophilus*. *J. Mol. Biol.* 266, 1016–1031.
- Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., and Kosakovsky Pond, S.L. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777.
- Wei, G., Xi, W., Nussinov, R., and Ma, B. (2016). Protein ensembles: how does nature harness thermodynamic fluctuations for life? the diverse functional roles of conformational ensembles in the cell. *Chem. Rev.* 116, 6516–6551.
- Wen, L.P., and Fulco, A.J. (1987). Cloning of the gene encoding a catalytically self-sufficient cytochrome P-450 fatty acid monooxygenase induced by barbiturates in *Bacillus megaterium* and its functional expression and regulation in heterologous (*Escherichia coli*) and homologous (*Bacillus megaterium*) hosts.

J. Biol. Chem. 262, 6676–6682.

White, R.H. (2006). The difficult road from sequence to function. J. Bacteriol. 188, 3431–3432.

Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E.Z., and Kern, D. (2004). Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair. Nat. Struct. Mol. Biol. 11, 945–949.

Wright, P.E., and Dyson, H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18–29.

Wu, H. (1995). Studies on denaturation of proteins XIII. A theory of denaturation†. In Advances in Protein Chemistry Volume 46, (Elsevier), pp. 6–26.

Wu, W.-B., Xu, J.-M., Wu, Q., Lv, D.-S., and Lin, X.-F. (2006). Promiscuous Acylases-Catalyzed Markovnikov Addition of N-Heterocycles to Vinyl Esters in Organic Media. Adv. Synth. Catal. 348, 487–492.

Xue, Y., Li, L.-P., He, Y.-H., and Guan, Z. (2012). Protease-catalysed direct asymmetric Mannich reaction in organic solvent. Sci. Rep. 2, 761.

Yang, F., Bian, C., Zhu, L., Zhao, G., Huang, Z., and Huang, M. (2007). Effect of human serum albumin on drug metabolism: structural evidence of esterase activity of human serum albumin. J. Struct. Biol. 157, 348–355.

Yang, F., Zhang, Y., and Liang, H. (2014). Interactive association of drugs binding to human serum albumin. Int. J. Mol. Sci. 15, 3580–3595.

Yang, J., Roy, A., and Zhang, Y. (2013). BioLiP: a semi-manually curated database for biologically relevant ligand-protein interactions. Nucleic Acids Res. 41, D1096-103.

Yennawar, N.H., Yennawar, H.P., and Farber, G.K. (1994). X-ray crystal structure of gamma-chymotrypsin in hexane. Biochemistry 33, 7326–7336.

Yu, X., Cojocaru, V., and Wade, R.C. (2013). Conformational diversity and ligand tunnels of mammalian cytochrome P450s. Biotechnol. Appl. Biochem. 60, 134–145.

- Zalatan, J.G., Fenn, T.D., and Herschlag, D. (2008). Comparative enzymology in the alkaline phosphatase superfamily to determine the catalytic role of an active-site metal ion. *J. Mol. Biol.* **384**, 1174–1189.
- Zea, D.J., Monzon, A.M., Gonzalez, C., Fornasari, M.S., Tosatto, S.C.E., and Parisi, G. (2016). Disorder transitions and conformational diversity cooperatively modulate biological function in proteins. *Protein Sci.* **25**, 1138–1146.
- Zhang, Y., and Sagui, C. (2015). Secondary structure assignment for conformationally irregular peptides: comparison between DSSP, STRIDE and KAKSI. *J. Mol. Graph. Model.* **55**, 72–84.
- Zhou, H.X., Wlodek, S.T., and McCammon, J.A. (1998). Conformation gating as a mechanism for enzyme specificity. *Proc Natl Acad Sci USA* **95**, 9280–9283.
- Zhuravlev, P.I., Materese, C.K., and Papoian, G.A. (2009). Deconstructing the native state: energy landscapes, function, and dynamics of globular proteins. *J. Phys. Chem. B* **113**, 8800–8812.
- Zhu, G., Huang, Q., Wang, Z., Qian, M., Jia, Y., and Tang, Y. (1998). X-ray studies on two forms of bovine beta-trypsin crystals in neat cyclohexane. *Biochim. Biophys. Acta* **1429**, 142–150.
- Zhu, W., Rainville, I.R., Ding, M., Bolus, M., Heintz, N.H., and Pederson, D.S. (2002). Evidence that the pre-mRNA splicing factor Clf1p plays a role in DNA replication in *Saccharomyces cerevisiae*. *Genetics* **160**, 1319–1333.
- Zou, T., Risso, V.A., Gavira, J.A., Sanchez-Ruiz, J.M., and Ozkan, S.B. (2015). Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol. Biol. Evol.* **32**, 132–143.
- (1995). *All About Albumin* (Elsevier).