



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

**Ingeniería de proteínas para la síntesis de
aminas quirales: optimización de una
 ω -transaminasa de *Capronia semiimmersa***

Ariel Tijman Nuñez

Tesis de Maestría

Presentada como uno de los requisitos para el título de
Magíster en Química

Programa de Posgrado en Química de la Facultad de Química
Universidad de la República

Octubre 2023

**Ingeniería de proteínas para la síntesis de
aminas quirales: optimización de una
 ω -transaminasa de *Capronia semiimmersa***

Luciana Capece

Presidenta del tribunal

Gabriela Irazoqui Duñach

Tribunal

Agustina Vila Grigorio

Tribunal

Paola Panizza Scasso

Directora de tesis

Pablo D. Dans Puiggròs

Director de tesis



*The nitrogen in our DNA, the calcium in our teeth,
the iron in our blood, the carbon in our apple pies
were made in the interiors of collapsing stars.*

We are made of starstuff.

We are a way for the universe to know itself.

Carl Sagan

Agradecimientos

Desde pequeño sentí curiosidad por el mundo que me rodeaba. Los juguetes desarmados que muchas veces no pude volver a ensamblar son mis pobres testigos. Ustedes pueden serlo también: me acabo de acordar que a veces guardaba moscas en el freezer de casa y al sacarlas un rato más tarde veía cómo volvían lentamente a la vida. Así que, como si no tuviese que haber impreso esta tesis hace semanas, me puse a buscar una explicación. Resulta que en algunas especies de moscas, las acuaporinas y las acuagliceroporinas (dos proteínas de membrana) evitan la ruptura celular expulsando agua de la célula e ingresando glicerol, respectivamente, a un ratio mucho mayor que en otras especies.¹ Poder comprender cómo funcionan las cosas no me sería posible si no estuviese parado en hombros de gigantes. Como decía Bernardo de Chartres:

Nos esse quasi nanos, gigantium humeris incidentes, ut possimus plura eis et remotiora videre, non utique proprii visus acumine, aut eminentia corporis, sed quia in altum subvenimur et extollimur magnitudine gigantea.

Somos como enanos a los hombros de gigantes. Podemos ver más, y más lejos que ellos, no por la agudeza de nuestra vista ni por la altura de nuestro cuerpo, sino porque somos levantados por su gran altura.

Yo creo que esta frase no solo aplica al cuerpo de conocimiento, sino también al sistema de soporte que poseemos mientras intentamos mirar más lejos. Deseo aquí, por tanto, agradecer a muchos pequeños gigantes.

Empezando por el soporte más fundamental, quiero agradecer a mis padres. Por criarme, por los valores que decidieron consciente o inconscientemente inculcarme, por siempre buscar brindarme lo que creyeron mejor. Gracias a mamá por introducirme a la pasión de cocinar, mi primera conexión con la química, y enseñarme que “*a través de la cocina se transmiten caricias invisibles*”. Gracias a papá por su perseverancia, por siempre empujar a más, e introducirme a mis primeros biocatalizadores: la levadura y el kéfir.

Gracias a Nadi y Luly por siempre malcriarme un poco y amarme incondicionalmente; también por siempre tener una vela lista para prender ante cada desafío. Gracias, Luly, por traer a Joaco a nuestras vidas. 

Nunca tuve una vida social muy activa, hasta hace un par de años, así que gracias a todos los amigos que supieron esperarme. Tati, Mage, Jessi, Ale, gracias por todos los viajes, recuerdos y paisajes que recorrimos en los últimos tiempos. Por todas las noches de vinito,

charlas y juegos. Tati, gracias por ser la mejor roomie, ¡prometo que termino con esto y nos tomamos un Santa Julia! Niky, kededirte, gracias por siempre levantarme (literal y figurativamente) y ver en mí más de lo que jamás podré ver yo mismo. Cris y Andy, gracias por ser mi hotel favorito, por todas las noches de juegos y traguitos, por saber dar un abrazo y perdonar mis suspiros. Nati M, Nati R, Gabo, Gui y Caro, ¡cuánto partido le sacamos a la AUGM! Gracias por estar presentes y siempre acompañar mi antojo de helado. Gracias Diegui por siempre estar para discutir un experimento o un dilema. Gracias Gabi por tu luz. Gracias Gonza por ser mi fiel compa de aventuras bioinfo. Gracias Bel por ser mi compa de extensión y 3D.

No sé cómo caí en la biocatálisis, pero sé bien por qué me quedé. Gracias, Sonia, la jefa máxima, por ser una excelente mentora y dar el ejemplo tanto en el terreno académico como en el personal. Gracias también por todos esos almuerzos cada vez que hubo algo para festejar o estaba teniendo un dilema existencial. Gracias a mi hermanito mayor, Cesitar. Sos el claro ejemplo de “cuando sea grande quiero ser como vos”. A pesar de tu mesada caótica supiste enseñarme gran parte de lo que sé. Gracias por siempre estar dispuesto a probar cosas nuevas y dar una mano, y también por ser humano, cálido, y tener las palabras justas.

Gracias a los amigos y compas de este trayecto: Vale, Vir, Feli, Mari, Sofi, Agustín, Agus, MIS, a los peones, a las jefas, a los microbios, LBB, Bioquímica y el resto de áreas que anduve pululando a lo largo de los años. Como canta NTVG, “*con hambre no se puede pensar*”, así que gracias también a la gente de la cantina de Medicina por proveernos de altos menús sumamente asequibles siempre con la mejor de las ondas. Mil gracias a las bedelías de grado y posgrado por todos los trámites que aceleraron debido a mi amor por los *deadlines*.

Quiero agradecer especialmente a Paola y Pablo por aceptarme como estudiante y haber depositado su confianza en mí. Gracias también por haberme seguido la corriente con cada cosita que quise probar y apoyarme al 100% en mis ambiciones. Gracias a Max por haberme recibido en Groningen, habiendo aparecido de la nada. Por último, gracias a Agus, Gabi y Lula por haber aceptado ser parte de esta historia. :)

Ingeniería de proteínas para la síntesis de aminas quirales: optimización de una ω -transaminasa de *Capronia semiimmersa*

Ariel Tijman Nuñez, Mag.

Programa de Posgrado en Química de la Facultad de Química
Universidad de la República

2023

DIRECTORES:

Paola Panizza Scasso^a, Pablo D. Dans Puiggròs^b

^a Facultad de Química, Universidad de la República

^b CENUR Litoral Norte, Universidad de la República

Resumen

Las enzimas son catalizadores eficientes y sustentables, capaces de llevar a cabo reacciones tanto quimio- como estereoespecíficas difíciles de lograr mediante métodos químicos tradicionales. La biocatálisis se alinea con los principios de la química verde, lo que redundará en ventajas tanto desde el punto de vista ambiental como económico. Un área en la que

presenta un gran potencial es la síntesis de aminas quirales. Estas se encuentran presentes en un alto porcentaje de fármacos y compuestos bioactivos, por lo que su producción es un objetivo fundamental para la industria. Su síntesis química tradicional en forma enantioméricamente pura es difícil de lograr, requiriendo catalizadores que utilizan metales pesados y por lo tanto generan alto impacto ambiental. Enzimas como imino-reductasas, aminosas reductivas o transaminasas podrían proveer un proceso mejor para este fin, pero se hace necesaria la búsqueda de enzimas de este tipo que sean capaces de cumplir con los exigentes requisitos de los procesos industriales.

En trabajos previos del grupo fue identificada y caracterizada una ω -transaminasa de *Capronia semiimmersa* (TACap). Esta enzima demostró ser capaz de utilizar tanto D-alanina como isopropilamina como donores de amino, siendo esta última más atractiva desde un punto de vista económico. La producción y purificación de la enzima fueron optimizadas y la especificidad de sustrato ensayada. En particular se destaca la capacidad de esta enzima de tomar como sustrato aceptor de amino a la fenilacetona, dando como producto a la L-anfetamina. Este producto es intermediario en la síntesis del [^{11}C]L-deuterodeprenil, un radiofármaco de utilidad en el diagnóstico de enfermedades neurodegenerativas de uso actualmente en el Centro Uruguayo de Imagenología Molecular. Si bien esto resultó promisorio, el grupo se encontró con bajas conversiones debido a una baja estabilidad de la enzima en las condiciones de reacción.

Este trabajo de maestría se planteó aprovechar el avance de los métodos bioinformáticos y de la química computacional para desarrollar, a partir de esta enzima ya caracterizada, variantes que permitiesen cumplir con los requisitos de actividad y estabilidad requeridos para el escalado de la reacción mencionada anteriormente. Partiendo de esta premisa fue recreada una estructura tridimensional de TACap, cuya dinámica molecular fue simulada y analizada para comprender las interacciones moleculares presentes y proponer modificaciones que pudiesen aumentar su estabilidad. Mediante esta estrategia se propuso la eliminación de dos regiones desordenadas. En paralelo fue seguida una estrategia basada en el análisis de la secuencia: fue construida una secuencia consenso utilizando 919 secuencias relacionadas a la TACap. Se ha reportado que los aminoácidos conservados pueden contribuir favorablemente a la estabilidad, por lo que se seleccionó un conjunto de 41 mutaciones puntuales (aminoácido nativo intercambiado por aminoácido consenso) con potencial para estabilizar la proteína.

Las modificaciones propuestas fueron incorporadas sobre el gen codificante de la enzima nativa mediante PCR, y las variantes fueron expresadas para analizar su estabilidad y actividad. La estabilidad fue evaluada mediante la medida de la temperatura de *melting* (desnaturalización) aparente (T_m) de las proteínas, y aquellas con un T_m superior a aquel de la enzima nativa fueron evaluadas en cuanto a su actividad. La prueba de actividad fue llevada a cabo mediante un ensayo espectrofotométrico con un sustrato modelo, y aquellas enzimas con una actividad comparable o superior a la de la enzima nativa fueron consideradas para el siguiente paso. Las tres mejores variantes fueron combinadas para dar lugar a cuatro enzimas que fueron nuevamente evaluadas en cuanto a su estabilidad y actividad. Una de estas variantes, conteniendo las mutaciones L106V y Δ 211-217, presentó un aumento en la T_m de 6°C, junto con una actividad relativa de 213% comparada con la enzima original, lo que la vuelve una opción muy prometedora para su aplicación a la síntesis del radiofármaco de interés.

Protein Engineering for the Synthesis of Chiral Amines: Optimization of an ω -Transaminase from *Capronia semiimmersa*

Ariel Tijman Nuñez, Mag.

Graduate Program in Chemistry, Facultad de Química
Universidad de la República

2023

DIRECTORS:

Paola Panizza Scasso^a, Pablo D. Dans Puiggròs^b

^a Facultad de Química, Universidad de la República

^b CENUR Litoral Norte, Universidad de la República

Abstract

Enzymes are efficient and sustainable catalysts capable of carrying out both chemio- and stereospecific reactions that are difficult to achieve using traditional chemical methods. Biocatalysis aligns with the principles of green chemistry, resulting in environmental and economic advantages. One area where enzymes show great potential is the synthesis of chiral

amines. These are present in a high percentage of pharmaceuticals and bioactive compounds, making their production a fundamental objective for the industry. Traditional chemical synthesis of enantiomerically pure amines is challenging, requiring catalysts that use heavy metals and thus have a high environmental impact. Enzymes such as imine reductases, reductive aminases, or transaminases could provide a better process for this purpose, but the search for such enzymes capable of meeting the demanding requirements of industrial processes is necessary.

In previous work by the group, an ω -transaminase from *Capronia semiimmersa* (TACap) was identified and characterized. This enzyme demonstrated the ability to use both D-alanine and isopropylamine as amino donors, with the latter being more economically attractive. The production and purification of the enzyme were optimized, and substrate specificity was tested. Particularly noteworthy is the enzyme's ability to use phenylacetone as an amino acceptor substrate, resulting in the production of L-amphetamine. This product is an intermediate in the synthesis of [¹¹C]L-deuterodeprenyl, a radiotracer used in the diagnosis of neurodegenerative diseases currently in use at the Uruguayan Center for Molecular Imaging (CUDIM). However, the group encountered low conversions due to the enzyme's low stability under reaction conditions.

This master's thesis aimed to leverage the advancement of bioinformatics and computational chemistry methods to develop variants of the characterized enzyme that would meet the activity and stability requirements necessary for scaling up the aforementioned reaction. Starting from this premise, a three-dimensional structure of TACap was recreated, and its molecular dynamics were simulated and analyzed to understand the present molecular interactions and propose modifications to increase its stability. Using this strategy, two disordered regions were proposed for elimination. Simultaneously, a sequence-based strategy was followed, constructing a consensus sequence based on 919 sequences related to TACap. It is known that conserved amino acids can contribute favorably to stability, so a set of 41 point mutations (exchanging native amino acids for consensus amino acids) with the potential to stabilize the protein was selected.

The proposed modifications were incorporated into the coding sequence of the native enzyme through PCR, and the variants were expressed to analyze their stability and activity. Stability was evaluated by measuring the melting temperature (T_m) of the proteins, and those with a T_m higher than that of the native enzyme were further evaluated for their activity. Activity

testing was performed using a spectrophotometric assay with a model substrate, and those enzymes with comparable or superior activity to the native enzyme were considered for the next step. The three best variants were combined to generate four enzymes, which were then re-evaluated for their stability and activity. One of these variants, containing the mutations L106V and Δ 211-217, showed an increase in T_m of 6°C, along with a relative activity of 213% compared to the wild type enzyme, making it a very promising option for its application in the synthesis of the radiotracer of interest.

Tabla de contenidos

Agradecimientos.....	vii
Resumen	ix
Abstract	xiii
Tabla de contenidos.....	xvii
Lista de tablas.....	xxi
Lista de figuras	xxiii
Lista de abreviaturas.....	xxix
Introducción	1
Biocatálisis	1
Determinación de estructuras tridimensionales de proteínas	8
Estabilidad de enzimas	10
Diseño racional <i>in silico</i>	12
Simulaciones de dinámica molecular	13
Secuencia consenso	16
Cálculo de energía libre de plegamiento tras mutación	17
Aminas quirales y amino transaminasas	17
ω -transaminasa de <i>Capronia semiimmersa</i>	20
Vinculación al CUDIM	21
Objetivos	25
Objetivo general	25
Objetivos específicos.....	25
Capítulo I: Modelado, simulaciones y análisis bioinformático.....	27
1.1. Introducción	29
1.2. Materiales y métodos	29
1.2.1. Modelado por homología.....	29
1.2.2. Modelado con AlphaFold2	30
1.2.3. Parametrización de aldimina interna Lys-PLP (KLP).....	30
1.2.4. Preparación de estructuras para MD	31
1.2.5. Preparación del sistema y corrida de MD de producción	32
1.2.6. Análisis de simulaciones.....	32

1.2.7. Secuencia consenso	33
1.2.8. Cambio de energía libre de plegamiento tras mutación.....	34
1.3. Resultados y discusión	35
1.3.1. Obtención de estructura de TACap.....	35
1.3.2. Generación de librería para la aldimina interna Lys-PLP (KLP)	39
1.3.3. Simulaciones de dinámica molecular	40
1.3.4. Reportes previos	52
1.3.5. Secuencia consenso	52
1.3.6. Cambio de energía libre de plegamiento tras mutación.....	58
1.4. Conclusiones	64
Capítulo II: Clonado, expresión y análisis de variantes enzimáticas	67
2.1. Introducción	69
2.2. Materiales y métodos	69
2.2.1. Cepas bacterianas y plásmidos	69
2.2.2. Medios de cultivo y enzimas	69
2.2.3. Soluciones y materiales	70
2.2.4. Extracción de ADN (miniprep)	71
2.2.5. Electroforesis de ADN en gel de agarosa	71
2.2.6. Purificación de ADN por gel	71
2.2.7. Restriction Free Cloning (RFC).....	72
2.2.8. Concentración de ADN.....	72
2.2.9. Precipitación de ADN con ARNt	73
2.2.10. Golden Gate Cloning	73
2.2.11. Mutagénesis sitio-dirigida	73
2.2.12. Delección de regiones.....	74
2.2.13. Primers utilizados	75
2.2.14. Condiciones de PCR.....	77
2.2.15. Preparación de células electrocompetentes	78
2.2.16. Transformación de células electrocompetentes	79
2.2.17. Preparación y transformación de células quimiocompetentes.....	80
2.2.18. Guardado de cepas en freezer	80

2.2.19. Envío de cepas en papel de filtro	80
2.2.20. Expresión de TACap.....	80
2.2.21. Electroforesis de proteínas PAGE-SDS.....	81
2.2.22. Medida de T_m	81
2.2.23. Medida de actividad enzimática	82
2.2.24. Cinética de inactivación térmica.....	83
2.3. Resultados y discusión	84
2.3.1. Clonado.....	84
2.3.2. Construcción de variantes.....	86
2.3.3. Expresión de proteínas.....	87
2.3.4. Determinación de T_m	89
2.3.5. Análisis de algunas mutaciones estabilizantes	94
2.3.6. Ensayos de actividad y estabilidad	97
2.4. Conclusiones	102
Conclusiones generales y perspectivas	105
Anexos.....	107
Secuencia de ADN de TACap.....	107
Secuencia aminoacídica de TACap.....	107
Modificación de campo de fuerza	108
Librería de parámetros para KLP	109
Códigos empleados	116
Archivos adjuntos.....	123
Bibliografía.....	125

Lista de tablas

Tabla 1 Principios de la química verde y su aplicación en biocatálisis. Adaptado de Sheldon y Woodley. ²⁰	5
Tabla 2 Proteínas homólogas a TACap.....	35
Tabla 3 Valores de coordenadas, tipo de átomo y carga parcial asignada a cada átomo del KLP. Los valores de carga fueron extraídos de Venhorst <i>et al.</i> ⁹⁷	39
Tabla 4 Condiciones de simulación de las distintas enzimas.....	41
Tabla 5 Resumen de valores de RMSD promedio y desviación estándar para las simulaciones corridas.....	42
Tabla 6 Parámetros seleccionados promedio en la ventana de análisis, junto a su desviación estándar. En la última columna se presenta el porcentaje de aumento de SAS respecto a su MD a 298 K.....	47
Tabla 7 Resumen de condiciones empleadas en las simulaciones de MD de mutantes de TACap.....	49
Tabla 8 Porcentaje de aumento del RMSF observado en la MD a 348 K respecto a 298 K, para la enzima nativa y los dos mutantes seleccionados.....	50
Tabla 9 Resumen de los parámetros analizados en las dinámicas de TACap y sus variantes. 50	
Tabla 10 Porcentaje de conservación del aminoácido nativo (col. 3) y el consenso (col. 5), por posición. En la última columna se presenta la diferencia de conservación entre el aminoácido consenso y el nativo.	54
Tabla 11 Resumen de las mutaciones seleccionadas para ensayar experimentalmente, junto con el método y el motivo por el que fueron postuladas.	65
Tabla 12 Resumen de los mutantes puntuales propuestos y obtenidos. En negrita aquellos a partir de reportes previos; en romana las mutaciones consenso; con fondo verde las mutaciones que pudieron ser obtenidas en el laboratorio; con fondo blanco las que no fueron obtenidas; con letra en rojo las que fueron mal diseñadas.	86
Tabla 13 Estrategia seguida para la generación de los mutantes múltiples. Se indica mutante a obtener, molde empleado y primers correspondientes.	87
Tabla 14 Temperatura de melting aparente de todos los mutantes evaluados y ΔT_m respecto a la enzima nativa (wt).	90
Tabla 15 ΔT_m de los mutantes consenso evaluados respecto a la enzima nativa (wt) y porcentaje de conservación del aminoácido consenso en esa posición.	91

Tabla 16 ΔT_m de los mutantes puntuales evaluados respecto a la enzima nativa (wt) y $\Delta\Delta G^{\text{fold}}$ obtenido mediante FoldX (F) y Rosetta-ddG (R) sobre las estructuras apo y holo.	93
Tabla 17 Actividad relativa de las enzimas con un ΔT_m positivo respecto a la enzima nativa.	100
Tabla 18 T_m , ΔT_m y actividad relativa a la enzima nativa de los mutantes combinados.	101
Tabla 19 Lista y descripción de archivos adjuntos.	123

Lista de figuras

Figura 1 Esquema de formación de un enlace peptídico entre dos aminoácidos con cadenas laterales R_1 y R_2 . Extraído de Wikimedia Commons. ⁷	2
Figura 2 Tabla de los 20 aminoácidos canónicos encontrados en la naturaleza. Adaptado de Wikimedia Commons. ⁸	2
Figura 3 Ilustración de los distintos niveles de plegamiento de una proteína. Adaptado de Wikimedia Commons. ⁹	3
Figura 4 a- Obtención enzimática de 6-APA a partir de penicilina. b- Obtención enzimática de antibióticos semisintéticos a partir de 6-APA. Adaptado de Heckmann y Paradisi. ¹¹	4
Figura 5 Principios de la química verde. Extraído de Compound Interest ³⁴	7
Figura 6 Superficie de energía potencial simplificada de una proteína hipotética. Cada punto en el plano definido por los ejes X e Y representa una conformación diferente de la enzima. El valor correspondiente en el eje Z es la energía libre de plegamiento. La conformación catalítica es fácilmente accesible desde el estado nativo. Una transición reversible desde el estado nativo a un estado parcialmente desplegado a través del estado de transición TS_1 se caracteriza por la diferencia de energía libre de plegamiento ΔG_1 y la barrera de energía libre ΔG_1^\ddagger . A partir del estado parcialmente desplegado puede darse una transición irreversible hacia un estado completamente desplegado a través del estado de transición TS_2 . Otra vía irreversible conduce a un estado de agregación. ΔG_1 y ΔG_2 se relacionan con la estabilidad termodinámica, mientras que ΔG_1^\ddagger y ΔG_2^\ddagger se relacionan con la estabilidad cinética. Extraído de Musil <i>et al.</i> ⁴⁰	11
Figura 7 Esquema representativo de los pasos que se dan durante una simulación de dinámica molecular.	14
Figura 8 Reacción esquemática de reacción para una transaminasa, con desplazamiento del equilibrio por evaporación del subproducto.	18
Figura 9 Esquema representativo de la síntesis de sitagliptina mediante procesos de segunda (A) y tercera generación (B). Extraído de Savile <i>et al.</i> ⁸³	19
Figura 10 Mecanismo de reacción de ω -amino transaminasas. Se ejemplifica con isopropilamina como donador de amino y acetofenona como aceptor. Adaptado de Berglund <i>et al.</i> ⁸⁶	20
Figura 11 Estructura del [¹³ C]L-deuterodeprenil.	21

Figura 12 Síntesis tradicional desarrollada para la producción del L-nordeuterodeprenil en el CUDIM, partiendo de L-fenilalanina y con L-anfetamina como intermediario (obtenida en paso v)b)). 6 corresponde al tosilato de propargilo deuterado.	22
Figura 13 Ruta sintética propuesta para el L-nordeuterodeprenil a partir de fenilacetona, utilizando la ω -amino transaminasa de <i>Capronia semiimmersa</i> como biocatalizador.	23
Figura 14 Modelo por homología de TACap. En cyan, extremo C-terminal.....	36
Figura 15 Alineamiento de secuencias molde 4CMD y 4CE5 con TACap. Los puntos representan aminoácidos iguales a TACap y los guiones gaps en el alineamiento.....	37
Figura 16 Predicción de estructura de los 28 aminoácidos terminales de TACap. a- AlphaFold2, b- RoseTTAFold.	37
Figura 17 Predicción de estructura de un monómero de TACap. a- AlphaFold2, b- RoseTTAFold.....	38
Figura 18 Superposición de los modelos alineados de TACap obtenido por homología + <i>ab initio</i> para el extremo C-terminal y el obtenido mediante AlphaFold2. La estructura secundaria se encuentra coloreada según el RMSD entre las estructuras, de menor (azul) a mayor (rojo).	38
Figura 19 Estructura de la aldimina interna Lys-PLP (KLP). Cada átomo figura con su nombre. Extraído de Venhorst <i>et al.</i> ⁹⁷	39
Figura 20 Análisis de MD de 4CE5. RMSD (izq.) y RMSD promedio acumulado (der.)	42
Figura 21 RMSD vs. t para las MD de TACap, alineadas al primer marco a 298 K.	43
Figura 22 RMSF de 4CE5 en las distintas condiciones de MD. Se indica con línea punteada vertical el residuo 130.	44
Figura 23 Estructura de 4CE5. En gris oscuro cadena A, gris claro cadena B. En rojo se señalan aquellas regiones con un RMSF superior a 2 Å en MD sin urea, junto a su nombre. Las regiones pertenecientes a la cadena B se señalan con un asterisco. En verde se señala el loop flexible en MD con urea. Es identificado el loop 3 por su relevancia en TACap.....	44
Figura 24 RMSF de las distintas simulaciones de TACap.....	45
Figura 25 Radio de giro de TACap-HM y TACap-HM- Δ Ctail. La línea anaranjada corresponde al cálculo de ROG sobre la enzima nativa, excluyendo los extremos del análisis (noTail)....	46
Figura 26 SAS vs. t para 4CE5.	47
Figura 27 Estructura de TACap. En gris oscuro cadena A, gris claro cadena B. En rojo se señalan aquellas regiones con un RMSF superior a 2 Å en MD sin urea, junto a su nombre. Las	

regiones pertenecientes a la cadena B se señalan con un asterisco. En verde se señalan las regiones solo flexibles en MD con urea.	48
Figura 28 RMSF de los mutantes seleccionados de TACap, bajo distintas condiciones de simulación. A- N-terminal; B a E- loops 1 a 4; F- C-terminal.	49
Figura 29 Disociación del dímero de una TA de clase I propuesta por Ruggieri <i>et al.</i> ¹²⁴ Adaptado de este reporte.	51
Figura 30 Superposición de la lisina catalítica, luego de la MD, enlazada al PLP (forma holo, carbonos en gris) o libre (forma apo, carbonos en cian).	52
Figura 31 Diagrama del proceso seguido para la obtención del MSA para secuencia consenso.	53
Figura 32 Árbol filogenético de las secuencias seleccionadas para el MSA, empleando la <i>branched-chain amino acid aminotransferase</i> de <i>E. coli</i> (ecBCAT) como outgroup.	54
Figura 33 Secuencia de TACap coloreada de acuerdo con el puntaje de conservación otorgado por ConSurf.	56
Figura 34 Gráfico de barras representando la diferencia de conservación entre el aminoácido consenso y el nativo.	57
Figura 35 Extracto del MSA. La primera secuencia (fondo en amarillo) es TACap y se señala con las flechas el loop 211-217 y el último aminoácido conservado, 332.	58
Figura 36 Superposición de los sitios activos de TACap-AF-holo (carbonos en cian) y 4CE5 (carbonos en gris).	60
Figura 37 A la izquierda se presentan diagramas de Venn con el número de mutaciones predichas como estabilizantes por FoldX (arriba) y Rosetta-ddg (abajo). A la derecha se comparan las energías de cada mutación calculada por FoldX (arriba) y Rosetta-ddg (abajo).	61
Figura 38 Diagrama de Venn presentando el número de mutaciones predichas como estabilizantes de acuerdo con FoldX o Rosetta-ddg, partiendo de la estructura apo (izquierda) u holo (derecha) de TACap.	61
Figura 39 Arriba, mapa de calor de los $\Delta\Delta G^{\text{fold}}$ calculados con FoldX para la forma apo. Se omite el equivalente para la forma holo por ser muy similar. Debajo, mapeo en la estructura 3D de las regiones señaladas con flechas en el mapa de calor.	63
Figura 40 Mapa de calor de los $\Delta\Delta G^{\text{fold}}$ calculados con Rosetta-ddg para la forma apo. Se omite el equivalente para la forma holo por ser muy similar.	63

Figura 41 Mutagénesis sitio-dirigida mediante QuikChange, empleando primers desfasados. Extraído de Fürst <i>et al.</i> ¹⁴¹	74
Figura 42 Distribución de buffers en microplaca de 96 pocillos para ensayo de actividad.	83
Figura 43 Electroforesis en gel de agarosa. a- marcador de PM (2); miniprep del vector pKK223-3 TACap (3). b- amplificación de TACap con primers 2.2.13.1 (1, 2, 3); marcador de PM (5).	85
Figura 44 a- Electroforesis del ADN extraído de algunas de las colonias transformadas con presunto pBAD-TACap; b- PCR sobre dicho ADN con primers linderos a la zona de inserción.	85
Figura 45 Secuencia esperada del vector pBAD-myc-HisA (arriba) y aquella obtenida por secuenciado (abajo).	86
Figura 46 Electroforesis de proteínas en gel de poliacrilamida en condiciones desnaturalizantes. Proteína TACap. Carriles 1-2: fracciones soluble e insoluble luego del lisado por sonicación. Carriles 3-4: fracciones soluble e insoluble luego del lisado por lisozima. Carril 5: marcador de peso molecular.....	88
Figura 47 Electroforesis de proteínas en gel de poliacrilamida en condiciones desnaturalizantes. Carriles 1-2, 3-4 y 5-6: TACap, $\Delta 211-217$ y ΔC -tail, respectivamente, fracciones soluble (s) e insoluble (i). Carril 7: marcador de peso molecular.....	89
Figura 48 Curva de melting para análisis de T_m . Fluorescencia vs. T (izq.) y $-dF/T$ vs. T (der.) TACap nativa en azul, $\Delta 211-217$ en verde y ΔC -tail en rojo.	90
Figura 49 Gráfico de porcentaje de conservación del aminoácido consenso vs. ΔT_m	92
Figura 50 Gráfico de $\Delta\Delta G^{\text{fold}}$ vs. ΔT_m . La línea horizontal cruza en -5 kJ/mol, la cual es la mejora en $\Delta\Delta G^{\text{fold}}$ sugerida por el protocolo FRESCO para seleccionar mutaciones estabilizantes.	94
Figura 51 Comparación entre estructura nativa (izquierda) y mutante P63V (derecha).....	95
Figura 52 Comparación entre estructura nativa (izquierda) y mutante S64H (derecha).....	95
Figura 53 Comparación entre estructura nativa (izquierda) y mutante T101I (derecha).	96
Figura 54 Comparación entre estructura nativa (izquierda) y mutante L106V (derecha).	97
Figura 55 Curva de calibración de acetofenona.	98
Figura 56 Espectro (230-300 nm) de los componentes de reacción en el análisis de actividad.	98

Figura 57 Gráfico de actividad remanente de TACap vs. t al ser incubada a distintas concentraciones de PLP durante 45 min a 40°C.	99
Figura 58 Gráfico de actividad remanente de TACap vs. t al ser incubada a distintas temperaturas.	99
Figura 59 Mutaciones con ΔT_m positivo mapeadas en la estructura de TACap. Vista frontal (izq.) e inferior (der.).....	100

Lista de abreviaturas

6-APA	Ácido 6-aminopenicilánico
ARNt	ARN de transferencia
CASP	Evaluación crítica de predicción de estructuras de proteínas
Cryo-EM	Criomicroscopía electrónica
CUDIM	Centro Uruguayo de Imagenología Molecular
KLP	Aldimina interna Lys-PLP
MAO-B	Monoamino oxidasa B
MD	Dinámica molecular
MSA	Alineamiento múltiple de secuencias
OD ₆₀₀	Densidad óptica medida a $\lambda = 600$ nm
PCR	Reacción en cadena de la polimerasa
PET	Tomografía de emisión de positrones
PLP	Piridoxal-5'-fosfato
PMP	Piridoxamina-5'-fosfato
REMD	Dinámica molecular de intercambio de réplicas
RFC	Clonado sin enzimas de restricción
RMN	Resonancia magnético nuclear
RMSD	Raíz cuadrada del desvío cuadrático medio
RMSF	Raíz cuadrada de la fluctuación cuadrática media
ROG	Radio de giro
TACap	ω -amino transaminasa identificada en <i>Capronia semiimmersa</i>
TACap-AF	Estructura de TACap obtenida mediante AlphaFold2
TACap-HM	Estructura de TACap obtenida mediante modelado por homología
T _m	Temperatura de <i>melting</i>
wt	Enzima nativa
ΔG^{fold}	Energía libre de plegamiento
$\Delta\Delta G^{\text{fold}}$	Cambio de energía libre de plegamiento tras mutación

Introducción

Biocatálisis

Más vieja de lo que parece

Por definición, un biocatalizador es una sustancia biológica que aumenta la velocidad de una reacción química y no es consumida en el proceso. La biocatálisis comprende, por tanto, el estudio y aplicación de estas sustancias. Si consideramos la producción de bebidas fermentadas, pan y queso en la Mesopotamia, el norte de Europa y China durante la Edad Antigua, la biocatálisis es aplicada por los seres humanos desde hace al menos ocho mil años.²⁻⁴ Sin embargo, no es hasta 1833 que los franceses Payen y Persoz⁵ reportan por primera vez la existencia de una sustancia de este tipo: la diástasa, una mezcla de amilasas capaz de degradar azúcares complejos. Con el correr de los años fueron reportadas algunas “sustancias” más (pepsina, invertasa) y se determinó su carácter catalítico, para finalmente en 1877 ser acuñado el término “enzimas”⁶. Vale en este punto aclarar que, si bien las enzimas no son los únicos tipos de biocatalizadores, en este trabajo nos referiremos exclusivamente a ellas o a los sistemas en que se encuentran involucradas.

Las enzimas son proteínas que cumplen la función de catalizar reacciones bioquímicas. Las proteínas constituyen heteropolímeros conformados por unidades moleculares conocidas como aminoácidos, los cuales se ensamblan mediante enlaces amida, específicamente denominados enlaces peptídicos (Figura 1). Estos aminoácidos, de carácter quiral (excepto la glicina), poseen cadenas laterales distintivas y por tanto presentan diversas propiedades fisicoquímicas que confieren a la proteína una extensa diversidad de combinaciones posibles (Figura 2).

La secuencia de aminoácidos, conocida como estructura primaria, en conjunto con su entorno, que comprende factores como el solvente, los iones, otras biomoléculas y modificaciones químicas adicionales, determinan la conformación tridimensional, la dinámica y la función de la proteína. Estos elementos fundamentales convergen en lo que se denomina plegamiento proteico, una organización tridimensional que a su vez se compone de estructuras más pequeñas agrupadas bajo la categoría de estructura secundaria. En esta escala, emergen las hélices alfa y las láminas plegadas beta, aunque también existen regiones menos ordenadas. Cuando estos elementos interactúan y se organizan en una escala más grande se habla de estructura terciaria, nivel en el cual se observan combinaciones recurrentes de elementos de

estructura secundaria, denominadas motivos de plegamiento. Por encima del nivel de organización terciaria existe la posibilidad de formar oligómeros, involucrando dos o más cadenas polipeptídicas iguales o diferentes, dando origen a la estructura cuaternaria. Pueden apreciarse los distintos niveles de plegamiento mencionados en la Figura 3. Los procesos de plegado y desplegado de proteínas resultan de un equilibrio delicado entre entalpía y entropía, lo que las hace particularmente sensibles a variaciones en condiciones como la temperatura, el pH y la fuerza iónica.

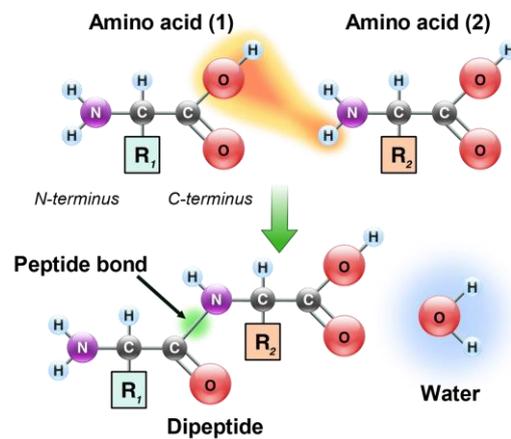


Figura 1 Esquema de formación de un enlace peptídico entre dos aminoácidos con cadenas laterales R₁ y R₂. Extraído de Wikimedia Commons.⁷

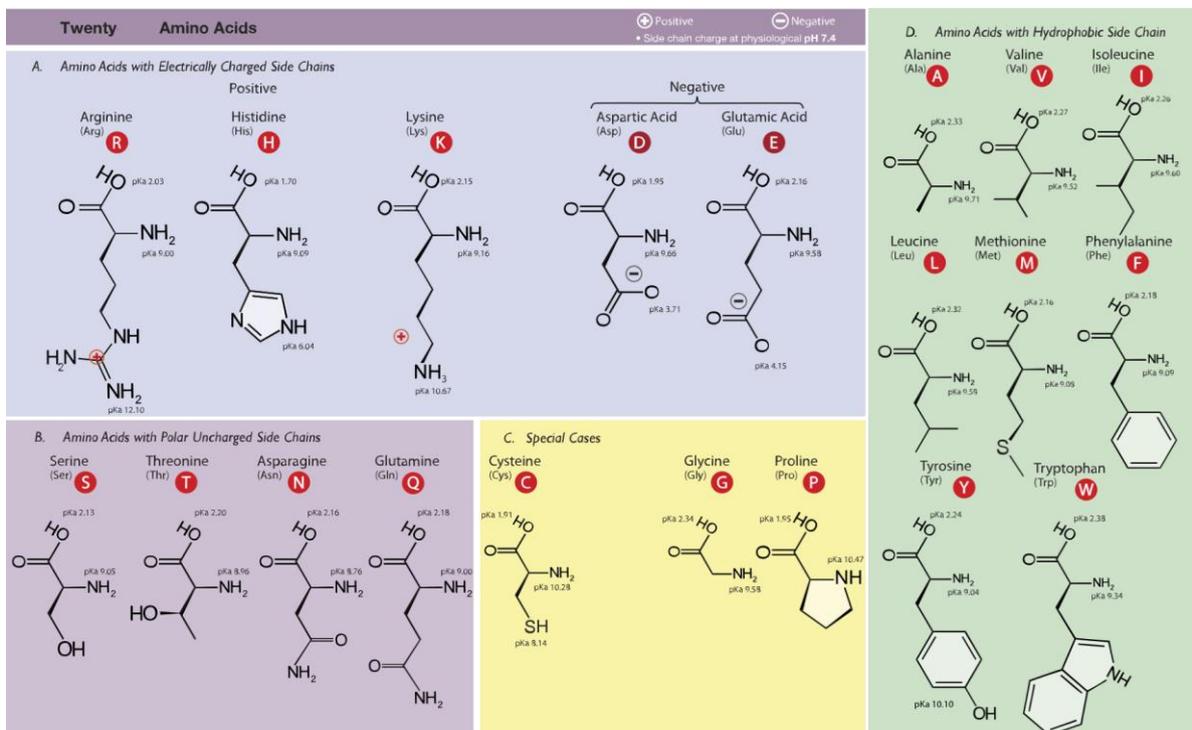


Figura 2 Tabla de los 20 aminoácidos canónicos encontrados en la naturaleza. Adaptado de Wikimedia Commons.⁸

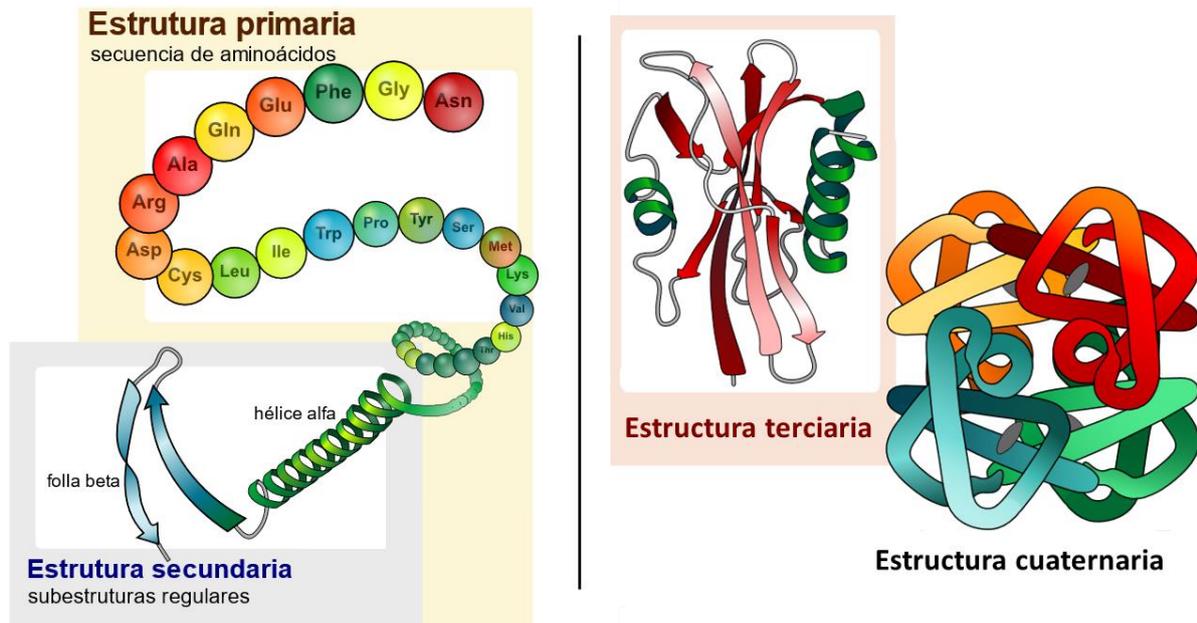


Figura 3 Ilustración de los distintos niveles de plegamiento de una proteína. Adaptado de Wikimedia Commons.⁹

Las proteínas pueden asociarse con otras biomoléculas mediante enlaces covalentes, enlaces de coordinación o interacciones intermoleculares más débiles. De esta manera, algunas proteínas incorporan monosacáridos u oligosacáridos, iones metálicos y grupos prostéticos con una amplia diversidad química. Estas uniones tienen la capacidad de influir en la estructura proteica, la dinámica conformacional, la ubicación celular, el estado oligomérico y, en última instancia, en la función biológica. Las proteínas cumplen una amplia gama de funciones en los organismos vivos, incluyendo la contribución a la organización de las estructuras celulares, el transporte de sustancias hacia y desde las células, así como la mediación de procesos de señalización celular, entre otras tareas fundamentales.

Como ya se mencionó, las enzimas son aquellas proteínas capaces de catalizar reacciones químicas. En el contexto de la síntesis química, la biocatálisis implica el uso de estas enzimas para catalizar reacciones de interés, y presenta algunas ventajas indiscutibles frente a la química “tradicional”, a saber: alta quimio- y regioselectividad, excelente estereoespecificidad y posibilidad de actuar en condiciones de reacción suaves (temperatura cercana a la ambiente, presión atmosférica, solvente acuoso).¹⁰ Como narran Heckmann y Paradisi¹¹, entre fines del siglo XIX y mediados del XX la biocatálisis encontró diversos nichos de aplicación industrial, ya fuese empleando preparados enzimáticos u organismos como bacterias, hongos y levaduras, en lo que constituyó la primera ola de la biocatálisis¹². La comprensión de las cualidades y mecanismos de las enzimas iría aumentando, desde su clasificación primero como proteínas y

luego en clases según su actividad, hasta la determinación de sus parámetros cinéticos. Un paso clave fue la elucidación de la estructura tridimensional de estas moléculas, siendo la de la lisozima la primera en ser reportada por Blake *et al.* en 1962¹³. La observación de los sitios de unión de sustratos e inhibidores, así como de los cambios conformacionales consecuencia de estas interacciones allanó el camino para la comprobación de los mecanismos de acción previamente propuestos, así como la identificación de nuevos.

A pesar de estos descubrimientos su uso industrial se encontraba limitado por la baja cantidad en que las enzimas podían ser producidas, y no sería hasta la revolución del ADN (su secuenciado y las tecnologías para manipularlo) que su uso se volvería más generalizado. El desarrollo de nuevas técnicas de biología molecular, como el clonado en plásmidos y la expresión recombinante de enzimas en *E. coli* u otros hospederos adecuados en la segunda mitad del siglo XX permitiría obtener estos biocatalizadores en cantidades abundantes y en mejor calidad.¹¹ La primera enzima en ser expresada en forma recombinante fue la penicilina amidasa en 1979¹⁴, empleada para la obtención del ácido 6-aminopenicilánico a partir de penicilina, así como para catalizar la reacción inversa en la obtención de antibióticos semisintéticos como la ampicilina y la amoxicilina (Figura 4).

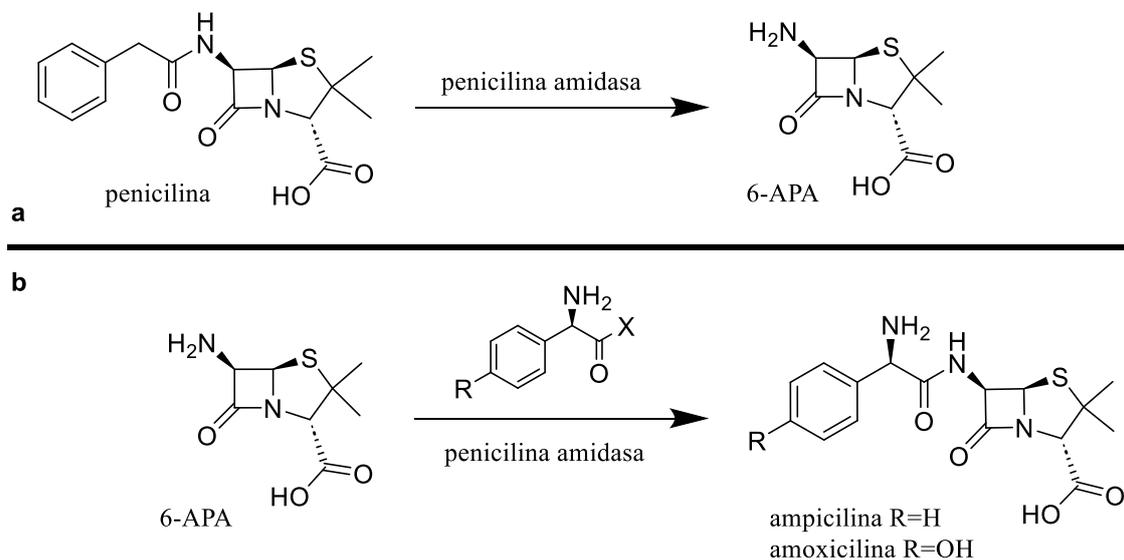


Figura 4 a- Obtención enzimática de 6-APA a partir de penicilina. b- Obtención enzimática de antibióticos semisintéticos a partir de 6-APA. Adaptado de Heckmann y Paradisi.¹¹

Sin embargo, las enzimas nativas por lo general trabajan sobre un limitado espectro de sustratos, tienen un estrecho margen de estabilidad y una baja tolerancia a cosolventes.^{15,16} Por tanto, no resulta inesperado que en las últimas décadas del siglo XX se desarrollasen las técnicas

y tecnologías necesarias para la introducción de mutaciones en el ADN recombinante que se tradujesen en mutaciones de la secuencia aminoacídica de las enzimas, buscando mejorar sus cualidades. Esto también permitió estudiar de forma directa el rol de ciertos residuos en la actividad y estabilidad, dando lugar a una primera etapa en el diseño racional de biocatalizadores y consiguio a la segunda ola de la biocatálisis.¹²

No obstante, los resultados obtenidos con estos diseños eran con frecuencia impredecibles e infructuosos. En consecuencia, el enfoque se tornó hacia aproximaciones que incorporan el azar, primero introduciendo mutaciones aleatorias en la secuencia mediante polimerasas de baja fidelidad, y luego con otras tecnologías como la de barajado de ADN.¹⁷⁻¹⁹ Se generan de esta forma inmensas librerías de mutantes que son depuradas mediante la aplicación de una presión evolutiva. Estos desarrollos, bajo la denominación de evolución dirigida, fueron sumamente exitosos para mejorar estabilidad a temperatura, pH, cosolventes, actividad frente a sustratos no naturales, estereoselectividad, o una combinación de los anteriores, y dieron inicio a lo que sería la tercera ola de la biocatálisis.¹² En paralelo, debido a la creciente preocupación por los impactos negativos de la química tradicional en el ambiente y la salud humana, fue desarrollado el concepto de “química verde”, cuyos principios se describen en la Figura 5. Como señalan Sheldon y Woodley²⁰, la biocatálisis cumple con todos estos principios, a excepción del 4 y el 10 que refieren al producto obtenido, y no al proceso (Tabla 1).

Tabla 1 Principios de la química verde y su aplicación en biocatálisis. Adaptado de Sheldon y Woodley.²⁰

Principio	En biocatálisis...
1 Prevención de residuos	Residuos reducidos significativamente
2 Economía atómica	Mayor economía por átomo y paso de reacción
3 Síntesis menos peligrosas	Generalmente baja toxicidad
4 Diseño de productos más seguros	no aplica
5 Solventes y auxiliares más seguros	Generalmente en agua
6 Eficiencia energética	Condiciones suaves de reacción
7 Uso de materia prima renovable	Las enzimas son renovables
8 Reducción de derivatizaciones	Evita pasos de protección/desprotección
9 Catálisis	Las enzimas son biocatalizadores
10 Diseño de productos degradables	no aplica
11 Análisis en tiempo real para prevenir la contaminación	Aplicable a los procesos biocatalíticos
12 Procesos intrínsecamente más seguros	Condiciones suaves de reacción

La repentina disponibilidad de biocatalizadores con propiedades adecuadas para los procesos industriales, sumado a sus ventajas en el marco de la química verde^{10,21-23}, llevó a un incremento constante en la incorporación de metodologías biocatalíticas en la industria, con cientos de ejemplos al día de hoy^{10,24-26}.

Aun así, el gran cuello de botella de la evolución dirigida es el cribado (*screening* en inglés) que debe hacerse del gran número de variantes generadas. Para enfrentar esta situación, existen dos aproximaciones: desarrollo de métodos de cribado que procesen las variantes en forma más rápida y eficiente; o la generación de librerías reducidas, por ejemplo, a través del estudio de la estructura y la determinación de regiones que al modificarse posean mayor probabilidad de éxito. Si bien la primera aproximación depende en gran parte del tipo de enzima y la propiedad que se desea modificar, la segunda es más plausible de ser aplicada a diversos desarrollos. Bornscheuer, quien sugiriera las distintas “olas” en la historia de la biocatálisis, insinuaba en 2017²⁷ que nos encontrábamos, posiblemente, al borde de la cuarta ola, determinada por:

- la emergencia de nuevas clases de enzimas;
- la catálisis de nuevas reacciones no naturales;
- el aumento exponencial en el número de secuencias de ADN disponibles, especialmente a partir de metagenomas;
- el uso de simulaciones biomoleculares de mecánica cuántica o molecular para el estudio de mecanismos de acción o dinámica de proteínas;
- la posibilidad de predecir la estructura tridimensional de las enzimas;
- la aplicación de la ingeniería metabólica.

Es claro que los avances en el área de la biocatálisis en las últimas décadas han sido gigantescos. Encontrándonos en una época dorada del desarrollo de herramientas de aprendizaje automático, su aplicación al área de la biología computacional ha sido en gran parte exitosa, y es de esperar que surjan más aplicaciones de relevancia para el área.²⁸⁻³³

The 12 Principles of GREEN CHEMISTRY

Green chemistry is an approach to chemistry that aims to maximize efficiency and minimize hazardous effects on human health and the environment. While no reaction can be perfectly 'green', the overall negative impact of chemistry research and the chemical industry can be reduced by implementing the 12 Principles of Green Chemistry wherever possible.

<p>1. WASTE PREVENTION</p>  <p>Prioritize the prevention of waste, rather than cleaning up and treating waste after it has been created. Plan ahead to minimize waste at every step.</p>	<p>7. USE OF RENEWABLE FEEDSTOCKS</p>  <p>Use chemicals which are made from renewable (i.e. plant-based) sources, rather than other, equivalent chemicals originating from petrochemical sources.</p>
<p>2. ATOM ECONOMY</p>  <p>Reduce waste at the molecular level by maximizing the number of atoms from all reagents that are incorporated into the final product. Use atom economy to evaluate reaction efficiency.</p>	<p>8. REDUCE DERIVATIVES</p>  <p>Minimize the use of temporary derivatives such as protecting groups. Avoid derivatives to reduce reaction steps, resources required, and waste created.</p>
<p>3. LESS HAZARDOUS CHEMICAL SYNTHESIS</p>  <p>Design chemical reactions and synthetic routes to be as safe as possible. Consider the hazards of all substances handled during the reaction, including waste.</p>	<p>9. CATALYSIS</p>  <p>Use catalytic instead of stoichiometric reagents in reactions. Choose catalysts to help increase selectivity, minimize waste, and reduce reaction times and energy demands.</p>
<p>4. DESIGNING SAFER CHEMICALS</p>  <p>Minimize toxicity directly by molecular design. Predict and evaluate aspects such as physical properties, toxicity, and environmental fate throughout the design process.</p>	<p>10. DESIGN FOR DEGRADATION</p>  <p>Design chemicals that degrade and can be discarded easily. Ensure that both chemicals and their degradation products are not toxic, bioaccumulative, or environmentally persistent.</p>
<p>5. SAFER SOLVENTS & AUXILIARIES</p>  <p>Choose the safest solvent available for any given step. Minimize the total amount of solvents and auxiliary substances used, as these make up a large percentage of the total waste created.</p>	<p>11. REAL-TIME POLLUTION PREVENTION</p>  <p>Monitor chemical reactions in real-time as they occur to prevent the formation and release of any potentially hazardous and polluting substances.</p>
<p>6. DESIGN FOR ENERGY EFFICIENCY</p>  <p>Choose the least energy-intensive chemical route. Avoid heating and cooling, as well as pressurized and vacuum conditions (i.e. ambient temperature & pressure are optimal).</p>	<p>12. SAFER CHEMISTRY FOR ACCIDENT PREVENTION</p>  <p>Choose and develop chemical procedures that are safer and inherently minimize the risk of accidents. Know the possible risks and assess them beforehand.</p>


 © COMPOUND INTEREST 2015; WWW.COMPOUNDCHEM.COM
 Shared under a CC Attribution-NonCommercial-NoDerivatives licence.



Figura 5 Principios de la química verde. Extraído de Compound Interest³⁴.

Determinación de estructuras tridimensionales de proteínas

¡Modelos, modelos everywhere!

Como se mencionó anteriormente, la función biológica de una proteína depende de su estructura terciaria/cuaternaria, la cual deriva de su secuencia aminoacídica en base a principios fisicoquímicos y el nivel de energía más bajo alcanzable. Es por tanto que el entendimiento del plegamiento de proteínas es uno de los objetivos más importantes en la biología estructural, y también en la biocatálisis.³⁰

Experimentalmente, las estructuras tridimensionales de las proteínas pueden ser determinadas mediante tres métodos: cristalografía de rayos X, criomicroscopía electrónica (cryo-EM) o resonancia magnética nuclear (RMN). La cristalografía de rayos X consiste en hacer pasar rayos X a través de una muestra de proteína cristalizada, de forma que el patrón de difracción obtenido permite determinar la posición de los átomos en el espacio. Este método permite también determinar el movimiento vibracional de cada átomo, parámetro que es llamado factor B. Este parámetro es de especial relevancia en la ingeniería de enzimas para aumentar su estabilidad, ya que se ha demostrado que el aumento de la rigidez en las regiones flexibles conlleva por lo general un aumento de la estabilidad.³⁵ El origen de la técnica de rayos X se remonta a principios del siglo XX, y es la más ampliamente empleada permitiendo obtener una alta resolución. Sin embargo, requiere que la muestra sea cristalizable, lo cual no siempre es posible. En la criomicroscopía electrónica (cryo-EM) la muestra proteica es congelada instantáneamente con nitrógeno líquido, y luego se adquieren varias imágenes mediante un microscopio electrónico; la combinación de estas imágenes permite reconstruir la estructura 3D de la proteína. Este método permite obtener la estructura de proteínas cuya cristalización resulta difícil para una cristalografía de rayos X, y al día de hoy permite resoluciones similares a las de dicha técnica. Algunas de sus desventajas, sin embargo, son los altos niveles de ruido obtenidos en el análisis y la dificultad de obtener buenos modelos para proteínas pequeñas. Las técnicas de rayos X y cryo-EM se usan muchas veces de forma combinada, obteniendo con cryo-EM una estructura de resolución media (7 a 10 Å) de un gran complejo proteico y con rayos X resolviendo con altísima resolución (< 1 Å) cada monómero. Por último, en la resonancia magnética nuclear la muestra proteica (en solución) es sometida a un fuerte campo magnético, y se utilizan ondas de radiofrecuencia para excitar el núcleo de aquellos isótopos que poseen 2 estados de spin, como el ¹H, ¹³C, ¹⁵N o ³¹P. Esto genera una señal que puede ser detectada e interpretada para determinar la localización de estos átomos, cómo se encuentran

enlazados y qué tan rápidamente se mueven unos respecto a los otros, entre otras aplicaciones. Las grandes ventajas de esta técnica son la posibilidad de determinar la estructura de la proteína en solución, en una conformación más cercana a la nativa (aquella de menor energía) que la cristalina, así como de proveer información sobre la dinámica. Algunas de las desventajas son la alta pureza, homogeneidad e integridad de la muestra requeridas, así como la necesidad muchas veces de enriquecer la muestra con los isótopos detectables.

El mayor inconveniente de los métodos experimentales mencionados es que son complejos, económica y temporalmente costosos, y la estructura determinada es frecuentemente distinta a la nativa.³⁰ No sorprende por tanto que, a la fecha, si bien el número de proteínas con secuencia conocida supera los 250 millones (UniProtKB/TrEMBL), el número de estructuras determinadas experimentalmente no llega a los 210 mil (RCSB PDB), dictando una relación aproximada de 1200:1.

En este contexto, se han desarrollado varios métodos para la predicción de estructuras proteicas mediante técnicas computacionales. Estos métodos suelen clasificarse entre aquellos de modelado libre y aquellos basados en un molde, aunque con frecuencia ambas aproximaciones son usadas. De hecho, en la competición CASP (Critical Assessment of protein Structure Prediction) estas son las tres categorías principales. La competición CASP se desarrolla bienalmente, y evalúa los avances alcanzados en la predicción de estructura de proteínas. Para ello, los distintos grupos de investigación participantes deben presentar la predicción lograda por su método propuesto para varias proteínas objetivo cuya estructura cristalográfica no ha sido aún publicada.³⁰ Si bien son varias las herramientas desarrolladas a lo largo de los años, hay dos que empleamos en este trabajo y nos interesa destacar.

La primera herramienta es MODELLER³⁶, cuyo desarrollo comenzó en 1993 y se basa en el modelado por homología; es decir, la secuencia aminoacídica de la proteína de interés es alineada con la estructura conocida de otra proteína homóloga a ésta. Esta técnica, mediante esta herramienta u otras, ha sido la más empleada históricamente y ha permitido lograr resultados notables, siendo aún empleada al día de hoy³⁷.

La segunda herramienta es AlphaFold2³⁸, ganadora de la competición CASP en 2020 gracias a su predicción de estructuras con una exactitud sin precedentes. AlphaFold2 emplea algoritmos de aprendizaje profundo para construir y procesar un múltiple alineamiento de secuencias en base a la secuencia de interés, detectando de esta forma posiciones proclives a mutar y la correlación entre ellas, al mismo tiempo que identifica proteínas homólogas con

estructura reportada. La combinación de esta información, pero especialmente la arquitectura de las redes neuronales utilizadas para procesarla ha permitido un gran avance en la predicción de estructuras de proteínas, lo que tiene importantes implicaciones para la investigación en biología y medicina.

Estabilidad de enzimas

Crónica de una muerte anunciada

Si bien una de las ventajas de la biocatálisis es el poder trabajar en condiciones suaves de reacción, muchas veces los procesos industriales requieren condiciones que son exigentes para los biocatalizadores, como ser temperaturas superiores a la ambiente, uso de cosolventes o pH alejado del neutro.³⁹ En particular, un aumento de temperatura puede inducir la inactivación de la enzima en un proceso de desnaturalización cuya representación más sencilla es un equilibrio entre la forma nativa (N) y la desplegada (*unfolded*, U): $N \rightleftharpoons U$. Como explican Musil *et al.*⁴⁰ y Rahban *et al.*⁴¹, en este proceso se encuentran comprendidas la estabilidad termodinámica y la estabilidad cinética, en tanto la primera refiere a la diferencia de energía libre entre un estado y el otro (ΔG^{unfold}), y la segunda es la barrera energética (o energía de activación) que separa estos estados (ΔG^\ddagger). En cierta forma, la estabilidad termodinámica refiere a la capacidad de la enzima de volver a su estado nativo una vez que la temperatura disminuye, mientras que la estabilidad cinética refiere a la capacidad de la enzima de evitar un desplegamiento irreversible.⁴² Una representación gráfica muy clara de esta situación se presenta en la Figura 6.

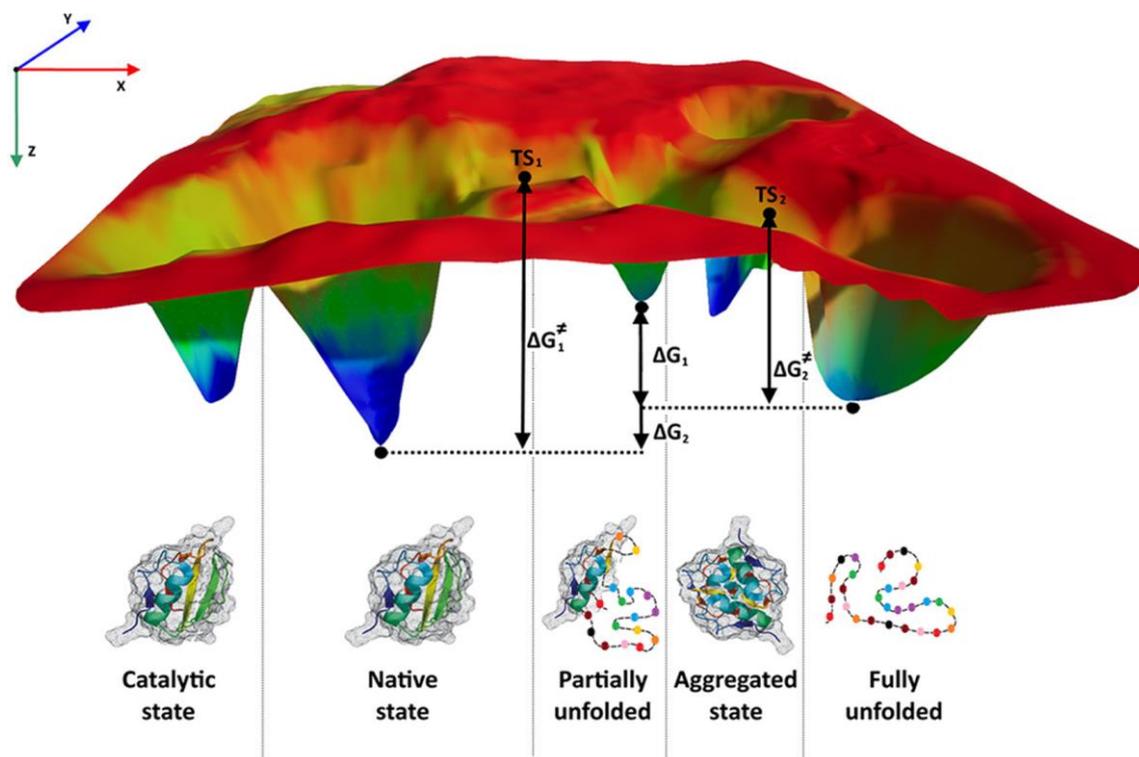


Figura 6 Superficie de energía potencial simplificada de una proteína hipotética. Cada punto en el plano definido por los ejes X e Y representa una conformación diferente de la enzima. El valor correspondiente en el eje Z es la energía libre de plegamiento. La conformación catalítica es fácilmente accesible desde el estado nativo. Una transición reversible desde el estado nativo a un estado parcialmente desplegado a través del estado de transición TS₁ se caracteriza por la diferencia de energía libre de plegamiento ΔG_1 y la barrera de energía libre ΔG_1^\ddagger . A partir del estado parcialmente desplegado puede darse una transición irreversible hacia un estado completamente desplegado a través del estado de transición TS₂. Otra vía irreversible conduce a un estado de agregación. ΔG_1 y ΔG_2 se relacionan con la estabilidad termodinámica, mientras que ΔG_1^\ddagger y ΔG_2^\ddagger se relacionan con la estabilidad cinética. Extraído de Musil *et al.*⁴⁰

La determinación de estas energías es experimentalmente difícil, por lo que generalmente lo que se mide es el cambio en estas energías tras una mutación ($\Delta\Delta G$). Un parámetro relacionado a $\Delta\Delta G^{unfold}$ es ΔT_m , siendo T_m la temperatura a la que la fracción de enzima en estado N y en estado U es igual a 0.5. Este T_m puede ser determinado mediante el ensayo de desplazamiento térmico. En este ensayo un compuesto fluorescente se agrega a la muestra proteica y la misma se somete a una rampa de temperatura; al irse desplegando la proteína, el compuesto fluorescente se va uniendo a algunos aminoácidos del núcleo hidrofóbico, aumentando la fluorescencia emitida. Por su parte, dos parámetros relacionados a $\Delta\Delta G^\ddagger$ son la semivida ($t_{1/2}$) y la temperatura de semi-inactivación (T_{50}), siendo la primera el

tiempo en que la actividad de la enzima decae a la mitad a cierta temperatura, y la segunda la temperatura a la que la actividad enzimática decae a la mitad tras cierto tiempo de incubación.⁴² Si bien la caracterización completa de una enzima se ve favorecida por la determinación de los parámetros de estabilidad tanto termodinámicos como cinéticos, es reconocido que la mejora en unos va de la mano con una mejora en los otros.⁴⁰

Son diversas las estrategias existentes para la mejora de estabilidad de enzimas. Por un lado están aquellas que implementan modificaciones al medio de reacción, como ser la inmovilización sobre soportes, las modificaciones químicas o el agregado de aditivos.⁴² Por otro lado tenemos las estrategias que buscan modificar la secuencia proteica para aumentar la estabilidad intrínseca de la enzima, como ser la evolución dirigida, los diseños racional, semiracional y *de novo*.³⁵ Dentro de los métodos racionales hay múltiples aproximaciones, por ejemplo, el cálculo de $\Delta\Delta G^{\text{fold}}$, la determinación de la secuencia consenso, la reconstrucción de la secuencia ancestral, la inclusión de enlaces disulfuro, la introducción de prolinas o el uso de algoritmos de inteligencia artificial.⁴³

Diseño racional *in silico*

¿Y si mejor probamos a ver qué da?

Los avances en el conocimiento y predicción de la estructura tridimensional de las proteínas han permitido dar grandes pasos en el estudio de la relación entre estructura y actividad. El diseño racional de biocatalizadores es una estrategia que permite la mejora de sus propiedades enzimáticas, tales como una mayor actividad, selectividad y estabilidad, a través del uso de herramientas computacionales para predecir el comportamiento de las enzimas y para guiar el diseño de nuevas variantes con las propiedades deseadas. La clave de esta aproximación es justamente el uso de herramientas computacionales de forma de disminuir los esfuerzos necesarios para el desarrollo y análisis de variantes enzimáticas, aumentando así el índice de éxito.

El diseño racional de enzimas se remonta al comienzo de la química computacional, cuando las primeras técnicas de modelado molecular fueron desarrolladas. En la década de 1960 fueron introducidos los primeros campos de fuerza de mecánica molecular, que permitieron el cálculo de energía de una molécula basado en su geometría. En la década siguiente fueron ejecutadas las primeras simulaciones de dinámica molecular, lo que permitió el estudio del movimiento de las moléculas en el tiempo.⁴⁴ Ya en los 80 fueron desarrollados los primeros

métodos de predicción de estructura de proteínas, que si bien no se comparan a los métodos actuales permitieron en su momento conocer más sobre la estructura de enzimas.⁴⁵ En la década de 1990 surgieron los primeros métodos computacionales para el diseño de proteínas,^{46,47} los cuales no solo fueron mejorados durante la década siguiente, sino que también se hicieron más rápidos y eficientes gracias a los desarrollos tecnológicos en el área computacional. Al día de hoy, el diseño racional de enzimas es un área madura que ha dado diversos frutos en el terreno de los desarrollos biotecnológicos, y cuyas capacidades se encuentran en constante expansión.⁴⁸

A continuación introduciremos tres técnicas que empleamos en este trabajo: simulaciones de dinámica molecular, diseño de secuencia consenso y cálculos de energía libre.

Simulaciones de dinámica molecular

La Matrix posta

La simulación de dinámica molecular (MD) es una metodología en que las leyes del movimiento Newtonianas son extrapoladas a los sistemas moleculares. En esta aproximación se deja de lado la estructura subatómica de la materia, y los parámetros energéticos del sistema pasan a ser función de las coordenadas nucleares. Mediante esta simplificación del sistema aún pueden extraerse datos explicativos y/o predictivos del comportamiento molecular a una fracción del costo computacional de la aproximación cuántica, que sí tiene en consideración a las partículas subatómicas.

El método computacional empleado para estimar las interacciones moleculares se denomina “campo de fuerza”, y comprende al conjunto de tipos de átomos y de ecuaciones y constantes asociadas para el cálculo de energía potencial del sistema (U). La forma más básica y usual de la ecuación de U_{total} está compuesta por cinco términos, tres de los cuales responden a interacciones entre átomos enlazados, y dos entre átomos no enlazados:

$$U_{total} = U_{enlace} + U_{ángulo} + U_{torsión} + U_{VDW} + U_{electrostática}$$

donde los primeros tres términos refieren a las variaciones en la longitud de los enlaces, sus ángulos y torsiones, respectivamente, y los dos últimos a las interacciones de Van der Waals e interacciones electrostáticas. Las expresiones explícitas de cada término, así como los parámetros contenidos en ellos, son lo que definen al campo de fuerza. Es así que podemos encontrar campos de fuerza para proteínas como ff19SB⁴⁹, para ADN como Parmbsc1⁵⁰, para membranas lipídicas como lipid21⁵¹, etc. Calcular todos estos términos para todos los átomos del sistema sería computacionalmente muy costoso e impráctico, por lo que se adoptan dos

estrategias al respecto. Para los términos de enlace (primeros tres) se definen “tipos de átomo” basados en su número atómico y el tipo de enlace en el que se encuentran involucrados, mientras que para los términos de no-enlace (últimos dos) se establece una distancia máxima a partir de la cual no se tendrán en cuenta interacciones.⁵²

La función de energía U es clave en las simulaciones de MD ya que permite el cálculo de las fuerzas sobre cada átomo de acuerdo con $\vec{F}_i = -\frac{\partial U}{\partial \vec{r}_i}$, siendo \vec{F}_i la fuerza sobre el átomo i y \vec{r}_i el vector que representa la ubicación de dicho átomo. Obtenida la fuerza y aplicando la segunda ley de Newton ($\vec{F}_i = m_i \vec{a}_i$) puede hallarse la aceleración del átomo y por tanto su localización en un tiempo $t + \delta t$. Este proceso se ilustra en la Figura 7. Cabe aclarar que no es posible resolver analíticamente las ecuaciones de movimiento para sistemas compuestos por más de dos átomos, por lo que se emplean aproximaciones numéricas obtenidas a partir de algoritmos computacionales.⁵³

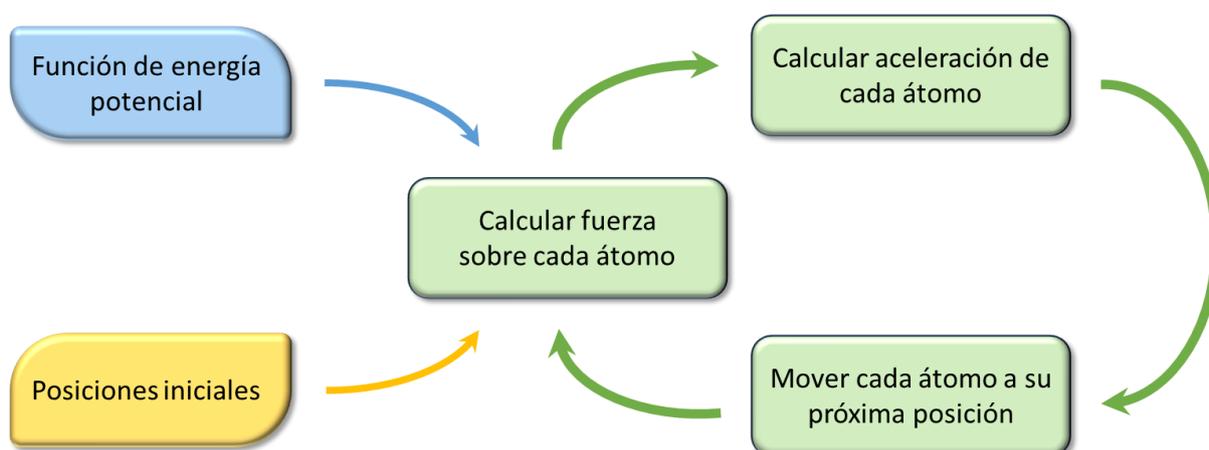


Figura 7 Esquema representativo de los pasos que se dan durante una simulación de dinámica molecular.

El δt mencionado, también llamado “paso de tiempo” (del inglés *time-step*) debe ser lo suficientemente pequeño como para poder describir correctamente el proceso de interés. Dado que en los sistemas biológicos los movimientos más rápidos corresponden a la oscilación de los átomos de hidrógeno en torno a un átomo pesado (como el carbono) y el período de esta oscilación ronda los 10 fs, por defecto se emplea un paso de tiempo de 1 fs,⁵³ aunque algunos métodos permiten aumentar este valor a 2 (usando SHAKE⁵⁴) o incluso 4 fs (usando repartición de masas⁵⁵).

A la hora de analizar las simulaciones de MD son varias las estrategias utilizadas, desde la visualización de la dinámica en forma de película, hasta el cálculo de parámetros estadísticos

y termodinámicos. Los parámetros que analizamos en este trabajo fueron: RMSD, RMSF, radio de giro y superficie accesible al solvente (SAS). El parámetro más conocido es el RMSD, raíz cuadrada del desvío cuadrático medio (del inglés *root of the mean squared deviations*) que mide la similitud global entre dos estructuras, donde una actúa como referencia. La expresión del RMSD viene dada por:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \|x_i - x_i^{ref}\|^2}$$

donde x_i es la posición de la i ésima partícula y x_i^{ref} es la posición de la misma partícula en la estructura de referencia (generalmente la de partida). Es frecuente calcular esta magnitud en proteínas utilizando únicamente las coordenadas de los carbonos alfa, para no tener en cuenta la variabilidad conformacional propia de las cadenas laterales. Usualmente se calcula el RMSD para cada “foto” de una trayectoria y se lo grafica en función del tiempo. Esta medida suele ser utilizada como criterio de convergencia en la dinámica molecular, y se discutirá más sobre esto al final de esta sección.

El RMSF, raíz cuadrada de las fluctuaciones cuadráticas medias (del inglés *root of the mean squared fluctuations*) es de especial interés en el análisis de estabilidad de proteínas, ya que se correlaciona con los valores de factor B mencionados anteriormente.³⁵ Esta medida otorga una noción de la similitud local entre un conjunto de estructuras provenientes de una trayectoria de MD, y se calcula mediante la siguiente ecuación:

$$RMSF_i = \sqrt{\langle r_i - \langle r_i \rangle \rangle^2}$$

donde r_i son las coordenadas del átomo i y $\langle r_i \rangle$ es la posición media de ese átomo a lo largo de la trayectoria. Esta medida informa acerca de la distancia media de cada átomo (o conjunto de átomos) respecto de la estructura promedio. Es frecuente que esta magnitud se calcule solo para los átomos del esqueleto peptídico y se reporte por residuo de forma tal que refleje la flexibilidad de las distintas regiones de la proteína.^{56,57}

El radio de giro es una medida de qué tan compacta es una proteína, y se define como la raíz cuadrada de la distancia media cuadrada de los átomos de la proteína con respecto a su centro de masa: $R_g^2 = \frac{\sum m_i (r_i - R_C)^2}{M}$, donde m es la masa del átomo i , r_i son sus coordenadas, M es la masa de la proteína y R_C son las coordenadas de su centro de masa. Por lo general, un

aumento del radio de giro a lo largo de la simulación implica que la proteína se está haciendo menos compacta, y por tanto es posible que se esté desplegando.^{41,58}

La superficie accesible al solvente (SAS) es, como su nombre indica, la superficie de una biomolécula que se encuentra accesible para un solvente. La SAS se describe generalmente en unidades de ángstroms cuadrados (\AA^2) y se puede calcular mediante métodos que implican el desplazamiento *in silico* de una sonda esférica, que aproxima una molécula de agua, alrededor de la biomolécula (también conocido como superficie de Connolly⁵⁹). Hay varios métodos disponibles para estimar la SAS, como el método LCPO⁶⁰ utilizado en el paquete AMBER.

Ahora bien, una vez corrida una simulación de MD, ¿cómo determinar si el sistema se encuentra equilibrado, y por tanto las propiedades que extraemos del sistema son realistas? En otras palabras, ¿cómo determinar si la simulación ha convergido? Es interesante la lectura de un hilo de White *et al.*⁶¹ en la plataforma X que discute al respecto. Un profesor de la Universidad de Dartmouth opina que “*tu simulación ha convergido cuando todos los enlaces se han roto y los átomos se han dispersado en el cosmos y son consumidos como combustible en el horno de estrellas moribundas*”. Más allá de lo poético, es cierto que es difícil hablar de convergencia cuando al menos en teoría la entropía del sistema siempre se encuentra en aumento. No existe una única respuesta a esta pregunta, y son múltiples los métodos reportados para responderla.^{62–65} Sin embargo, como planteaba el estadístico George E. P. Box, “*todos los modelos son incorrectos, pero algunos son útiles*”. En este contexto, la práctica más usual (no por esto libre de controversia) es medir el RMSD y aguardar por su estabilización. También puede calcularse el RMSD promedio acumulado (es decir, el RMSD promedio calculado para ventanas incrementales de tiempo) y aguardar a que se vuelva asintótico a un valor.

Secuencia consenso

El libro inédito de Mary Shelley

La secuencia consenso es creada a partir de un múltiple alineamiento de secuencias (MSA) homólogas, y se define sencillamente mediante el aminoácido más frecuente en cada posición de dicho MSA. La relativa facilidad con que esta aproximación puede ser implementada ha permitido comprobar, a través del estudio de varias familias de proteínas, que las mutaciones hacia residuos más conservados suelen mejorar la estabilidad de estas.^{66,67} Una posible explicación a esta observación es que a medida que las proteínas evolucionaron a partir de un ancestro común no especializado pero estable, la deriva evolutiva permitió el muestreo

de diferentes mutaciones que mantuviesen una estabilidad adecuada. A través de la evolución de funciones especializadas muchas proteínas ahora existen en un justo equilibrio entre estabilidad y función, y, por esta razón, tienden a conservarse los residuos estabilizantes. El diseño de consenso puede aprovechar, por tanto, millones de años de evolución e identificar residuos conservados a partir de numerosos homólogos de proteínas, reuniendo varias mutaciones estabilizantes que ninguna proteína individual ha necesitado acumular.⁶⁶

Cálculo de energía libre de plegamiento tras mutación

Relax, take it eeeeasy, acá no hay ecuaciones

Esta aproximación se basa en el cálculo de diferencia de energía libre de plegamiento (ΔG^{fold}) entre la estructura predicha de un mutante puntual de una proteína y la estructura nativa de esta, es decir, calcula el $\Delta\Delta G^{fold} = \Delta G_{mut}^{fold} - \Delta G_{wt}^{fold}$. Por tanto, un $\Delta\Delta G^{fold}$ negativo indica que el mutante es termodinámicamente más estable que la proteína nativa y viceversa. De igual forma que en las simulaciones de dinámica molecular, un campo de fuerza y sus algoritmos asociados son los encargados de realizar los cálculos correspondientes a la energía libre de plegamiento de cada estructura. Diversos métodos y algoritmos han sido desarrollados, siendo los más populares FoldX y Rosetta-ddG que cuentan con numerosas aplicaciones exitosas.⁶⁸⁻⁷³

Aminas quirales y amino transaminasas

Imprescindibles para las pastillitas

Las aminas quirales son de gran relevancia para las industrias farmacéutica y agroquímica; más de un 90% de los fármacos de bajo peso molecular más vendidos o recientemente aprobados contienen o se sintetizan a partir de aminas.⁷⁴ Dado que los distintos estereoisómeros pueden tener una diferente actividad farmacológica se ha puesto un gran empeño en el diseño de estrategias de producción de fármacos o precursores enantioméricamente puros, eliminando de esta forma los efectos secundarios asociados a la administración de racematos y el gasto innecesario en la producción del enantiómero que no es de interés.^{75,76} Estos compuestos forman parte de moléculas naturales como aminoácidos, vitaminas y alcaloides, entre otros; la naturaleza posee las herramientas para sintetizarlas, y por lo tanto resulta lógico que la biocatálisis haya surgido como una alternativa exitosa frente a la síntesis química.^{77,78}

Se han aplicado numerosos tipos de enzimas a la síntesis de aminas quirales,⁷⁷ siendo especialmente valiosas aquellas que permiten obtener aminas enriquecidas en un enantiómero a partir de compuestos proquirales. En particular las transaminasas —enzimas PLP-dependientes que catalizan la transferencia de un grupo amino desde un donador de amino a una cetona (Figura 8)— han encontrado amplia aplicación en la síntesis de aminas primarias^{79,80}. Tal es así que hace más de una década se emplean en la producción de ingredientes farmacéuticos activos^{81,82} y es notable ejemplo de ello su aplicación en la producción de sitagliptina, principio activo empleado para el tratamiento de la diabetes tipo II. La transaminasa empleada en dicho proceso es una variante de la transaminasa ATA-117, desarrollada mediante modelado computacional y evolución dirigida. La síntesis resultante (Figura 9) es una producción de tercera generación, luego de dos generaciones empleando catalizadores metálicos convencionales, y representa respecto a estas una mejora del 10% en rendimiento y 53% en productividad (kg/L por día), así como una reducción del 19% de residuos y del costo total de producción.^{83,84}

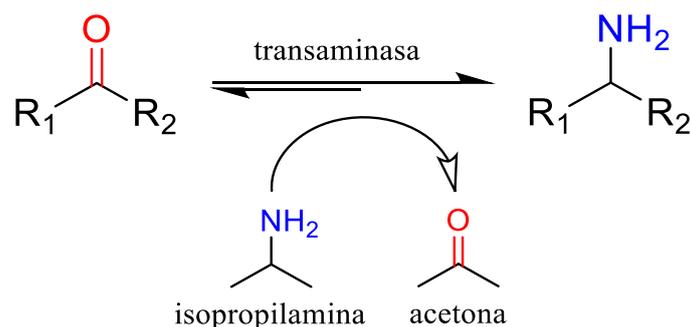


Figura 8 Reacción esquemática de reacción para una transaminasa, con desplazamiento del equilibrio por evaporación del subproducto.

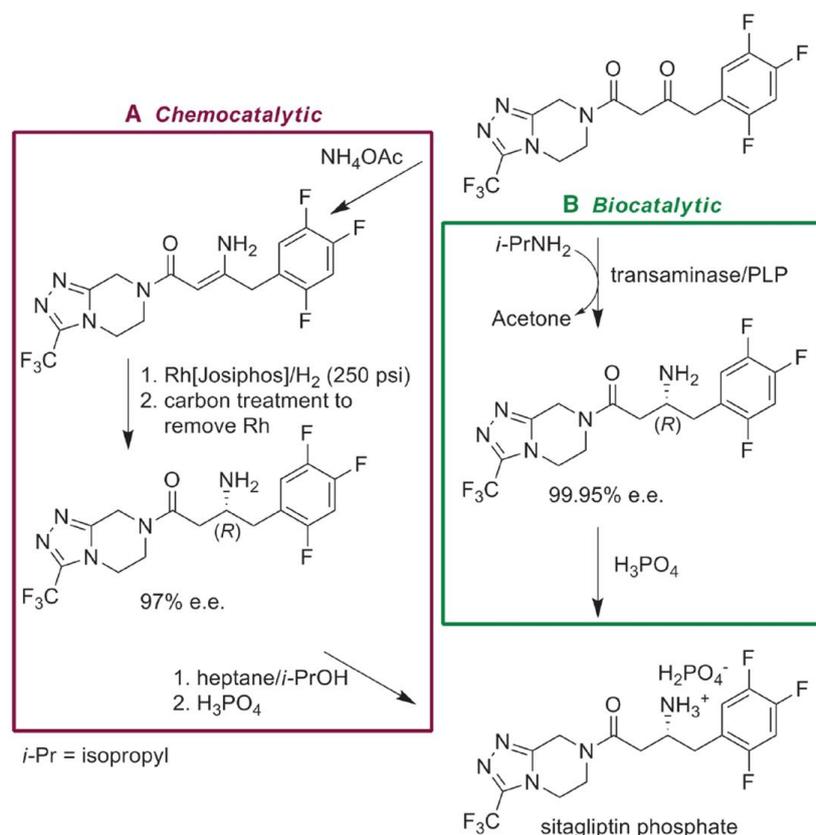


Figura 9 Esquema representativo de la síntesis de sitagliptina mediante procesos de segunda (A) y tercera generación (B). Extraído de Savile *et al.*⁸³

Estructuralmente, las transaminasas forman homodímeros en solución, aunque también algunas han sido reportadas como tetrámeros.⁸⁵ Poseen dos sitios activos por dímero, conformados por residuos de ambas cadenas. Su mecanismo de acción consta de dos partes y se ilustra en la Figura 10. En la primera mitad de la reacción, la lisina conservada del sitio activo, que forma una base de Schiff con el cofactor PLP en la proteína nativa (aldimina interna) es desplazada por el donante de amina. Esto genera una aldimina externa entre el donador y el PLP, y a través de intermediarios subsecuentes quinonoides y cetiminas, el grupo amino del donante se transfiere al PLP, produciendo la piridoxamina 5'-fosfato (PMP). El donante desaminado se libera como una cetona, aldehído o ácido cetónico. Después de esto, en la segunda mitad de la reacción, el aceptor de amina reacciona con el intermedio PMP para formar una cetimina, que posteriormente se descompone a través de un quinonoide y una aldimina externa. Finalmente, la lisina de la enzima reemplaza al aceptor aminado, que se libera, finalizando el ciclo catalítico.⁶⁸

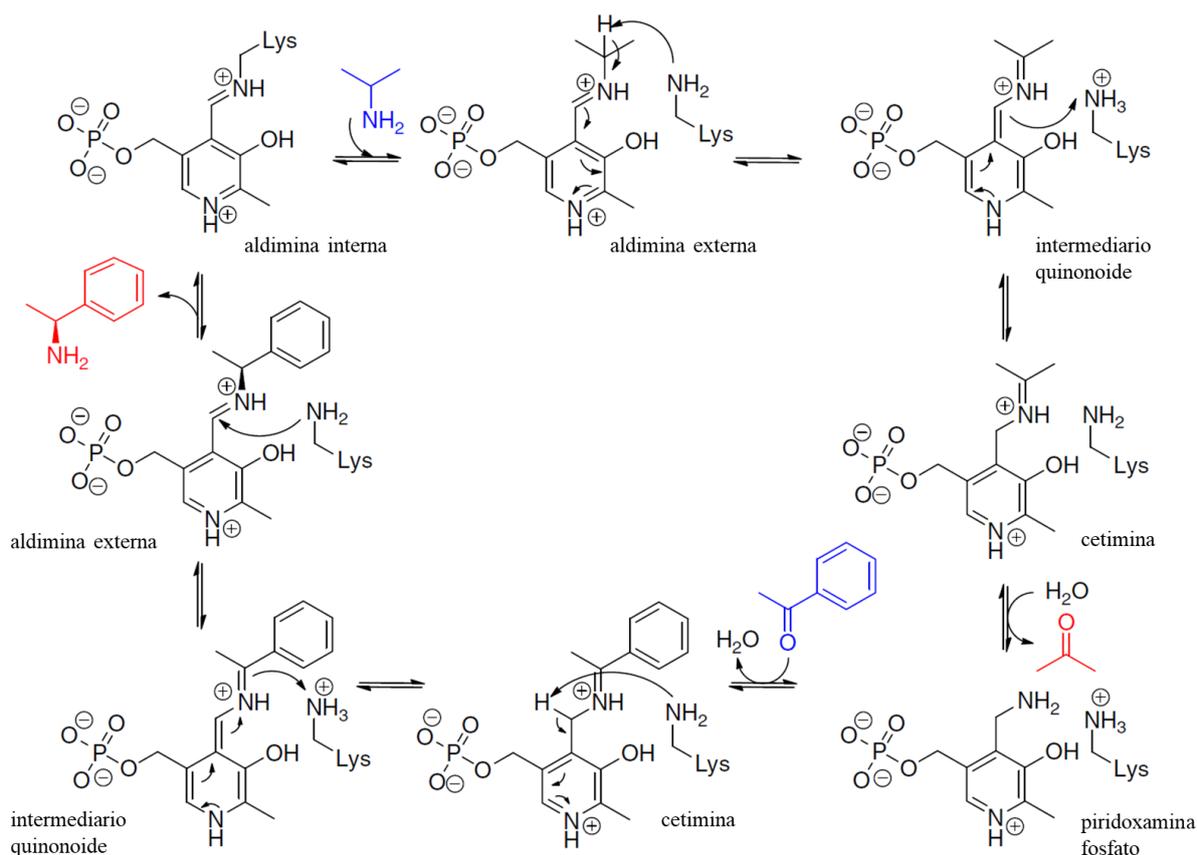


Figura 10 Mecanismo de reacción de ω -amino transaminasas. Se ejemplifica con isopropilamina como donador de amino y acetofenona como aceptor. Adaptado de Berglund *et al.*⁸⁶

ω -transaminasa de *Capronia semiimmersa*

Nuestro caballito de batalla

La *R*- ω -amino transaminasa de *Capronia semiimmersa* es una enzima PLP dependiente de clase IV, que fue identificada por nuestro grupo en trabajos previos.⁸⁷ Buscando identificar nuevas transaminasas con un espectro de sustrato más amplio que lo reportado hasta el momento, se aplicó un algoritmo computacional diseñado por Höhne *et al.*⁸⁸ que considera diversos motivos conservados y aminoácidos clave de transaminasas y enzimas relacionadas para seleccionar o descartar las distintas secuencias. Dentro de las 423 secuencias obtenidas por este método se seleccionó la ω -transaminasa de *Capronia semiimmersa* (TACap) por presentar diferencias frente a otras previamente reportadas.^{87,89} La enzima identificada fue clonada y expresada en pKK223-3, y su especificidad de sustrato y estereoselectividad fueron caracterizadas, obteniendo resultados comparables a aquellos obtenidos con otras transaminasas, con un excelente exceso enantiomérico. Es de destacar que TACap acepta

isopropilamina como dador de amina, una característica deseable para su aplicación industrial por ser este sustrato económico y desplazar el equilibrio de reacción por evaporación del coproducto (acetona).

Vinculación al CUDIM

Ay, celeste, regalame un sol

La instalación en nuestro país del Centro Uruguayo de Imagenología Molecular (CUDIM) ha significado un avance significativo en las posibilidades de diagnóstico y seguimiento de varias enfermedades de carácter oncológico y neurológico. Desde un punto de vista académico, la puesta en marcha del CUDIM representa un desafío y una oportunidad, ya que se deben desarrollar procesos de marcado de radiofármacos que sean eficientes y cumplan con los estándares de calidad requeridos. Asimismo, abre oportunidades al desarrollo de procesos sintéticos propios para las diferentes moléculas utilizadas ya que se debe contar con los precursores fríos adecuados. Un precursor frío es la molécula precursora al compuesto marcado que se administra al paciente. Dado el decaimiento natural en la radiactividad del compuesto marcado, es el precursor frío el que se mantiene en stock, por lo que su disponibilidad y calidad son esenciales.

En particular, el [^{11}C]L-deuterodeprenil (Figura 11) es utilizado en la evaluación *in vivo* de patologías neurológicas que implican una sobreexpresión de la enzima monoaminoxidasa-B (MAO-B) en casos de astrocitosis. Este radiofármaco presenta una estructura basada en el L-deprenil, un inhibidor selectivo e irreversible de la MAO-B. Es un radiofármaco de referencia para la cuantificación de la actividad de la MAO-B mediante tomografía de emisión de positrones (PET).⁹⁰ Este radiofármaco se utiliza actualmente tanto en la clínica del CUDIM, como en la línea de investigación de evaluación *in vivo* de patologías neurológicas que implican astrocitosis, como por ejemplo la Enfermedad de Alzheimer.

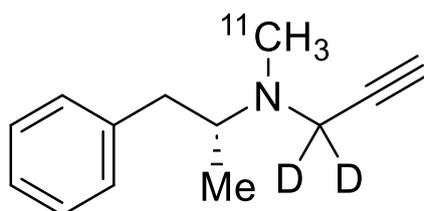


Figura 11 Estructura del [^{11}C]L-deuterodeprenil.

Dado el alto costo en el mercado de los precursores fríos de este tipo de radiofármacos se ha planteado la necesidad de desarrollar metodologías sintéticas eficientes y rápidas para su

producción en laboratorios del CUDIM o en grupos de investigación de la Universidad de la República que trabajen de manera cooperativa en el marco del convenio vigente entre ambas instituciones. En este contexto se ha desarrollado un proceso de 6 pasos para la síntesis del precursor frío L-nordeuterodeprenil partiendo de L-fenilalanina (Figura 12).⁹¹

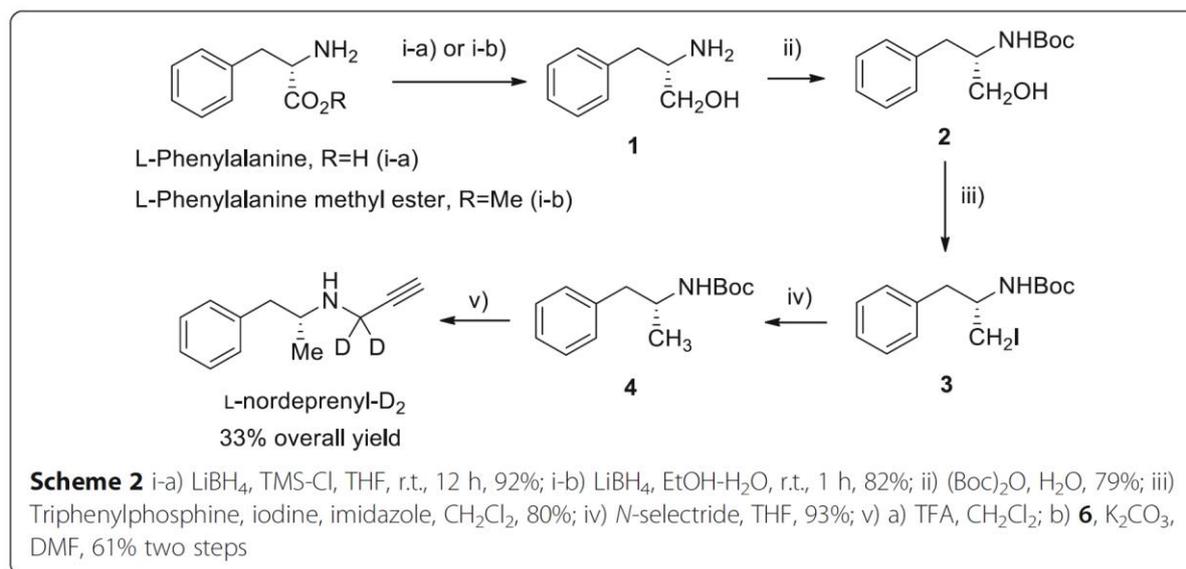


Figura 12 Síntesis tradicional desarrollada para la producción del L-nordeuterodeprenil en el CUDIM, partiendo de L-fenilalanina y con L-anfetamina como intermediario (obtenida en paso v)b)). **6** corresponde al tosilato de propargilo deuterado.

El desarrollo de esta metodología permitió al CUDIM obtener el [¹¹C]L-deuterodeprenil por procesos propios, independizándolo de los altos costos asociados al precursor. No obstante, la ruta desarrollada tiene ciertas desventajas como ser un alto número de pasos de reacción y la obtención del producto final en rendimientos bajos a moderados. Nuestro grupo ha venido trabajando en el desarrollo de una síntesis eficiente de este precursor utilizando metodologías biocatalíticas. Particularmente, se logró aplicar la ω-amino transaminasa de *Capronia semiimmersa*⁸⁷ (TACap) a la obtención de L-anfetamina en un único paso de reacción, obteniendo un rendimiento del 15% con un exceso enantiomérico >99%.. Un paso posterior de propargilación clásico permite obtener el precursor frío de la molécula de interés, de acuerdo con la Figura 13. A pesar del éxito obtenido en los ensayos preliminares, en estos ensayos se observó que la baja estabilidad de la TACap podía ser una de las causas del bajo rendimiento obtenido, que dificulta el escalado de la reacción necesario para su aplicación clínica en el CUDIM. Por este motivo en esta tesis se plantea aportar a la optimización de esta enzima

mediante la aplicación de métodos de diseño racional asistidos por el uso de herramientas computacionales.

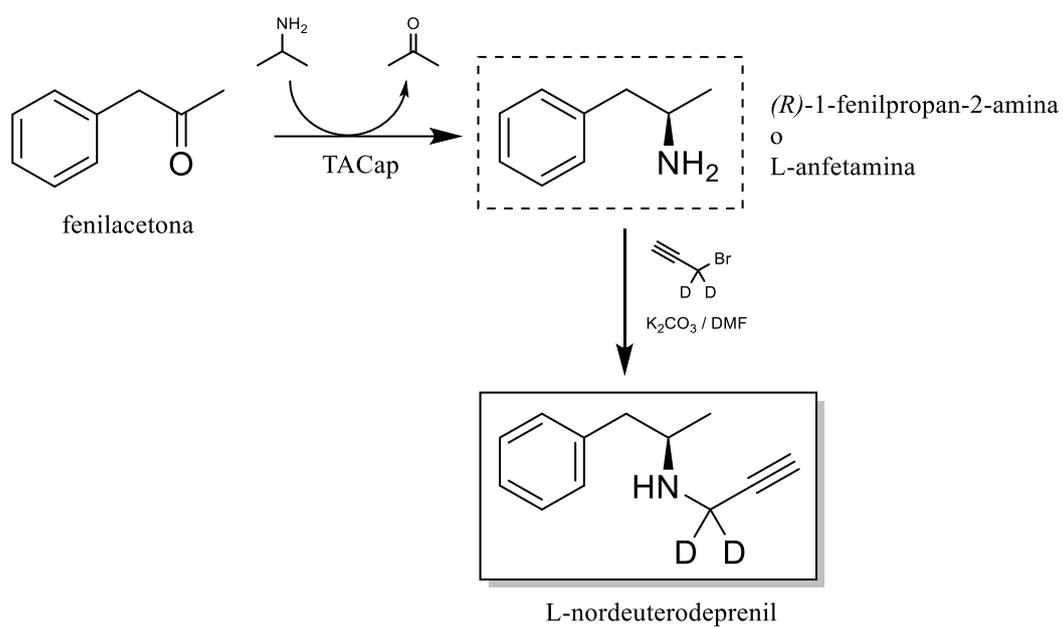


Figura 13 Ruta sintética propuesta para el L-nordeuterodeprenil a partir de fenilacetona, utilizando la ω -amino transaminasa de *Capronia semiimmersa* como biocatalizador.

Objetivos

Objetivo general

Aplicar metodologías bioinformáticas específicas al diseño de variantes más estables de la ω -transaminasa de *Capronia semiimmersa* (TACap) permitiendo su aplicación a la síntesis del radiofármaco [^{11}C]L-deuterodeprenil.

Objetivos específicos

- Determinar las mutaciones candidatas a generar estabilidad en la TACap a partir de distintas aproximaciones bioinformáticas.
- Analizar las interacciones intramoleculares de las mutaciones propuestas y de las simulaciones de dinámica molecular.
- Construir vectores de expresión para las variantes de TACap que incorporen las mutaciones seleccionadas.
- Analizar la estabilidad y actividad de los mutantes generados.
- Analizar los resultados obtenidos con las nuevas enzimas y evaluar la efectividad del proceso de selección *in silico*.

Capítulo I:

Modelado, simulaciones y análisis bioinformático

1.1. Introducción

En este capítulo presentaremos los experimentos computacionales realizados y discutiremos los resultados obtenidos. Más específicamente, hablaremos sobre cómo obtuvimos un modelo tridimensional de nuestra enzima, TACap, y qué desafíos nos encontramos para poder simular su dinámica molecular. Luego, haremos un análisis en profundidad sobre los datos obtenidos en estas simulaciones, y propondremos nuestros primeros mutantes. Adicionalmente, ¿podemos identificar si el mecanismo de desnaturalización de TACap sigue un mecanismo previamente conocido?

En segunda instancia trataremos el diseño de la secuencia consenso de TACap, cómo hicimos la selección de secuencias homólogas, qué información adicional obtuvimos desde el múltiple alineamiento de secuencias, y qué variantes propusimos expresar.

Por último, en el marco del cálculo de energía libre de plegamiento, discutiremos acerca de la importancia de la estructura de entrada sobre la que se procesarán estos cálculos.

1.2. Materiales y métodos

El modelado por homología, la parametrización del KLP, la preparación de las estructuras para MD así como el análisis de los datos resultantes y la generación de la secuencia consenso fueron ejecutados en una computadora local (personal o del grupo de trabajo). El modelado con AlphaFold2 fue ejecutado en los servidores de Google. Las simulaciones de MD y su análisis fueron ejecutados generalmente en un clúster de cómputo de alto rendimiento como ClusterUY⁹² o MareNostrum (Barcelona Supercomputing Center)⁹³.

1.2.1. Modelado por homología

- i. Se realizó una búsqueda de proteínas homólogas a la secuencia de interés con la herramienta BLASTp⁹⁴, seleccionando como base de datos las estructuras cristalinas depositadas en el Protein Data Bank⁹⁵.
- ii. Fueron seleccionadas de 1 a 5 proteínas homólogas con un porcentaje de identidad superior al 55%. Se comprobó que fuesen “buenos modelos” para realizar el modelado por homología: número de cadenas, sustratos o inhibidores presentes, RMSD, etc.
- iii. Seguimos el procedimiento de la herramienta MODELLER³⁶ teniendo en cuenta lo siguiente:

- a. Para convertir las secuencias del formato FASTA al PIR empleamos el script adjunto 'fasta2pir.sh'.
- b. Las secuencias solo pueden contener aminoácidos estándar.
- c. Una vez alineadas las secuencias debe verificarse que el formato del encabezado sea correcto (como se indica en el manual) y debe incorporarse la segunda cadena, de corresponder.
- d. El modelado lo realizamos sin refinamiento ya que esto insume cierto tiempo, y posteriormente largaríamos simulaciones de dinámica.

1.2.2. Modelado con AlphaFold2

Gracias a la implementación de la herramienta AlphaFold2 en Google Colab (cuaderno electrónico que permite correr código utilizando recursos en la nube) es posible obtener predicciones de la estructura de proteínas nativas sin necesidad de instalar programas en el sistema local. Uno de los cuadernos más utilizados y el empleado para este trabajo con los parámetros predeterminados es ColabFold⁹⁶, disponible en el siguiente enlace:

<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/alphaFold2.ipynb>.

1.2.3. Parametrización de aldimina interna Lys-PLP (KLP)

Para la construcción de la librería los datos de carga parcial fueron extraídos de Venhorst *et al.*⁹⁷ A continuación se indican los pasos seguidos para la obtención de los parámetros restantes. Se incluyen en anexos la modificación de campo de fuerza (p. 108) y la librería (p. 109) obtenidas.

- i. Se cargó en VMD un PDB conteniendo la aldimina interna (4CMD).
- ii. Seleccionamos la aldimina y extrajimos sus coordenadas a un nuevo PDB. Alternativamente esto puede hacerse a mano en un editor de texto.
- iii. Editamos las coordenadas de KLP en un editor de texto:
 - a. Eliminamos los hidrógenos presentes.
 - b. Modificamos la primera columna para que indique ATOM para todos los átomos (en lugar de HETATM).
 - c. De la misma forma, la columna de nombre de molécula debe indicar KLP para todos los átomos y el número de residuo debe ser 1 para todos los átomos.
- iv. Abrimos la molécula editada en xleap y la editamos, creando los enlaces entre los átomos.

- v. Agregamos los protones faltantes y verificamos que fuesen correctos de acuerdo con la hibridación de los enlaces.
- vi. Optimizamos la geometría de la molécula y guardamos las coordenadas.
- vii. Mediante el programa antechamber asignamos automáticamente los tipos de átomo de la molécula y los verificamos, obteniendo un archivo de salida mol2.
- viii. Sobre este archivo agregamos manualmente las cargas parciales de cada átomo extraídas de bibliografía, teniendo en cuenta que la carga neta debe ser cero (corregimos con pequeños diferenciales sobre los átomos más electronegativos).
- ix. Mediante el programa parmchk verificamos la disponibilidad de los parámetros necesarios en el campo de fuerza general de Amber (GAFF). Obtuvimos el archivo de modificación del campo de fuerza, FRCMOD.
- x. Generamos la librería mediante el programa tleap, empleando el Código 1 (anexo).
- xi. Modificamos el archivo de librería generado, indicando los átomos que conectan con la cadena peptídica y el tipo de residuo.
- xii. Probamos el sistema con una proteína conocida, por ej. 4cmd, empleando el Código 2 (anexo).
- xiii. Buscamos en los campos de fuerza (e.g. PARM19 y GAFF) los parámetros detectados como faltantes y los añadimos al archivo FRCMOD.

1.2.4. Preparación de estructuras para MD

- i. Depuramos el PDB de forma que solo quede la proteína y el cofactor. Empleamos para ello el Código 3 (anexo).
- ii. Verificamos que no existan residuos repetidos por multiplicidad de posiciones observadas en la cristalografía. En caso de existir, conservamos solo el más abundante.
- iii. Modificamos los residuos de lisina catalítica y PLP para que pasen a llamarse KLP, y de esta forma sean interpretados como un único residuo.
- iv. Cargamos el PDB resultante a la herramienta PDB2PQR⁹⁸ del servidor de Poisson Boltzman que puede encontrarse en <https://server.poissonboltzmann.org/pdb2pqr> para asignar estados de protonación correctos.
- v. Una vez obtenido el PQR (mismo formato que PDB) verificamos si el residuo KLP fue eliminado y en caso de que así fuese volvimos a agregarlo.
- vi. Ejecutamos tleap y procesamos el PQR o PDB obtenido para verificar que no existan errores empleando el Código 4 (anexo).

- vii. Neutralizamos, solvatamos y ajustamos la fuerza iónica del sistema, también en tleap. Se adjunta script empleado para determinar el número de iones a agregar de acuerdo con la concentración deseada, 'addionsrand.py'. Guardamos archivos de topología y coordenadas empleando el Código 5 (anexo). En el caso de simulaciones con urea, empleamos la caja equilibrada de urea 8 M disponible en el paquete Amber.
- viii. Verificamos en VMD que el sistema quedó bien construido.

1.2.5. Preparación del sistema y corrida de MD de producción

- i. Minimizamos el sistema mediante pmemd, en tres secuencias sucesivas, con una restricción sobre la proteína de 25, 15 y 5 kcal/mol respectivamente, empleando 2000 pasos de descenso pronunciado y 500 pasos de gradiente conjugado. Se adjunta script en Bash para este paso, 'minimization.sh'. La forma del script de Amber se presenta en el Código 6 (anexo).
- ii. Calentamos el sistema a la temperatura deseada con una restricción de 5 kcal/mol sobre la proteína, a volumen constante, durante 100 ps. Se adjunta script en Bash para este paso, 'heat.sh'. La forma del script de Amber se presenta en el Código 7 (anexo).
- iii. Equilibramos el sistema a temperatura y presión constante, durante 500 ps. Se adjunta script en Bash para este paso, 'equilibrate.sh'. La forma del script de Amber se presenta en el Código 8 (anexo).
- iv. Largamos la simulación de producción de dinámica molecular. Se adjunta a modo de ejemplo script en Bash para correr en local, 'runMD.sh'. La forma del script de Amber se presenta en el Código 9 (anexo).

1.2.6. Análisis de simulaciones

El análisis de las simulaciones fue realizado mediante los programas VMD (principalmente visualización) y cpptraj (parámetros estadísticos). Las imágenes presentadas fueron generadas con YASARA. Para el gráfico de los datos obtenidos se empleó Jupyter Notebook, utilizando el lenguaje Python y las bibliotecas matplotlib, numpy y pandas. A modo de ejemplo se presentan en anexos los códigos Código 10 a Código 13. Para el caso de TACap el análisis de RMSD fue realizado excluyendo los extremos C-terminal (residuos 330-356), y el análisis de RMSF se hizo sobre una de las cadenas peptídicas, en ambos casos sobre los carbonos alfa.

1.2.7. Secuencia consenso

Fue seguido en general el protocolo reportado por Sternke *et al.*⁹⁹ En forma resumida, involucra los siguientes pasos:

- i. Realizamos una búsqueda de secuencias relacionadas a la proteína de interés. En este caso, dada la abundancia de secuencias reportadas, recurrimos a al servidor BLASTp con los parámetros predeterminados.
- ii. Determinamos la mediana de la muestra de secuencias y eliminamos aquellas que se desvían más de un 30% de esta. En este caso empleamos Geneious¹⁰⁰.
- iii. Clusterizamos las secuencias con un porcentaje de identidad superior al 90% y continuamos el análisis solo con una secuencia representativa de cada clúster. Para este paso empleamos el programa UCLUST¹⁰¹.
- iv. Buscamos en estas secuencias aquellas que contienen los motivos conservados de interés. En este caso nos interesan los motivos reportados por Höhne *et al.*⁸⁸ y Telzerow *et al.*¹⁰² para la identificación de *R*-transaminasas. Empleamos Geneious y la herramienta grep¹⁰³ para la búsqueda y selección de secuencias.
- v. Alineamos las secuencias. En este caso empleamos el algoritmo MUSCLE¹⁰⁴ como se encuentra implementado en el programa MEGA X¹⁰⁵. También construimos un árbol filogenético, para lo que agregamos una enzima distante como *outgroup* y empleamos Geneious con los parámetros predefinidos.
- vi. Depuramos las secuencias por tres criterios más:
 - a. Motivos buscados en (iv) no alineados con los correspondientes en la secuencia de interés.
 - b. Ausencia de la lisina catalítica.
 - c. Presencia de aminoácidos no estándar.
- vii. Cargamos el MSA resultante en la plataforma en línea ConSurf¹⁰⁶ disponible en https://consurf.tau.ac.il/consurf_index.php. Esta plataforma provee el porcentaje de conservación de cada aminoácido para cada posición de la proteína de interés de acuerdo con el MSA cargado, así como el grado de confianza en este valor según el número de secuencias que tengan un aminoácido en esa posición. Con esta información y de acuerdo con el criterio del usuario se construye la secuencia consenso.

1.2.8. Cambio de energía libre de plegamiento tras mutación

Fue seguido el protocolo computacional desarrollado y reportado por Wijma *et al.*¹⁰⁷ para la rápida mejora de la estabilidad de proteínas. Dicho protocolo se adjunta a este trabajo ('FRESCO protocol.pdf'), así como su material suplementario ('FRESCO protocol supp.pdf') y el conjunto de scripts necesarios para su ejecución ('FRESCO.zip'). El protocolo fue seguido tal cual, teniendo en cuenta:

- En una primera etapa se partió de dos estructuras iniciales distintas: por un lado el modelo obtenido mediante AlphaFold2, relajado mediante una breve dinámica; por otro lado este mismo modelo, pero con el PLP incorporado como cosustrato mediante el programa AlphaFill²⁹ y posteriormente relajado. Los resultados obtenidos fueron comparados.
- En una segunda etapa se empleó este protocolo solo para realizar las MD de algunas de las mutaciones puntuales obtenidas experimentalmente, partiendo de la estructura relajada del modelo obtenido por homología, con el cosustrato como aldimina interna.

1.3. Resultados y discusión

1.3.1. Obtención de estructura de TACap

Se hizo una búsqueda de secuencias homólogas a TACap en la base de datos del PDB, seleccionándose aquellas con un porcentaje de identidad superior al 55% y un *query cover* superior al 90%, resultando en siete proteínas homólogas, todas *R*- ω -transaminasas (Tabla 2).

Tabla 2 Proteínas homólogas a TACap.

Estructura	Porcentaje de identidad	Query cover (%)	Resolución (Å)	Organismo
4CMD	65	0.99	1.68	<i>Nectria haematococca</i>
4CE5	64	0.92	1.63	<i>Aspergillus terreus</i>
6XWB	63	0.94	2.20	<i>Thermomyces stellatus</i>
6FTE	63	0.99	1.52	<i>Exophiala Xenobiotica</i>
6SNL	61	0.94	3.13	<i>Exophiala sideris</i>
4CHI	60	0.92	1.27	<i>Aspergillus fumigatus</i>
6XU3	58	0.92	2.10	<i>Shinella sp.</i>

A continuación, teniendo en cuenta los valores indicados en la Tabla 2 decidimos seleccionar 4CMD¹⁰⁸ ('4cmd.pdb') y 4CE5¹⁰⁹ ('4CE5.pdb') para realizar el modelado por homología. Para ello fue necesario preparar los modelos eliminando moléculas orgánicas, buffer e iones, optimizar la protonación de los residuos, y renombrar la lisina catalítica y el PLP para constituirlos en un único residuo "KLP". Luego seguimos las instrucciones de MODELLER (ver sección 1.2.1) para la construcción del modelo por homología de TACap. Como puede observarse en la Figura 14 el extremo C-terminal (aminoácidos 330 a 356) de ambos monómeros no adopta una estructura definida. Esto se debe a que nuestra proteína presenta un extremo más largo que aquella de los moldes empleados y por tanto no hay una estructura sobre la que modelar (ver alineamiento en Figura 15). Si bien otras proteínas (6FTE y 6SNL) tienen un largo parecido al de TACap su estructura cristalográfica no presenta esta región por lo que probablemente nos encontremos ante una sección intrínsecamente desordenada. Las simulaciones posteriores de dinámica molecular parecen confirmar esta hipótesis.

Mediante predicción de estructura intentamos producir una estructura más realista para el extremo C-terminal. Para esto recurrimos a los recientemente presentados (en septiembre 2021) AlphaFold2³⁸ y RoseTTAfold¹¹⁰. En primera instancia modelamos únicamente los

aminoácidos del 330 en adelante (Figura 16). Buscando obtener mejores resultados probamos con el modelado de un monómero entero, seleccionando como mejor modelo el obtenido con RoseTTAFold, por tener una estructura secundaria mejor definida (Figura 17). Para incorporar este extremo C-terminal al modelo por homología obtenido anteriormente sencillamente alineamos las estructuras e intercambiamos las coordenadas del extremo C-terminal, obteniendo el modelo final para simular ('TACap-HM.pdb').

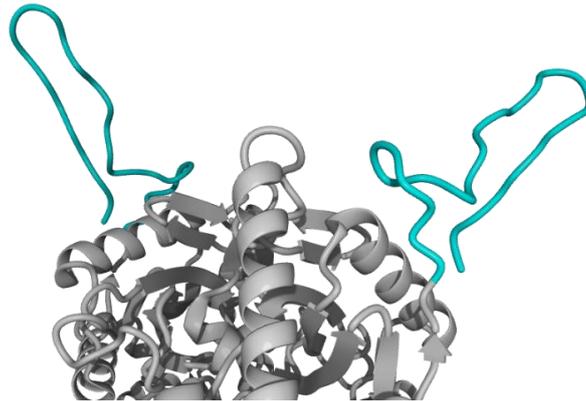


Figura 14 Modelo por homología de TACap. En cyan, extremo C-terminal.

```

          10      20      30      40      50      60
-----|-----|-----|-----|-----|-----|
4ce5  .AS.D.V.AG.A..QAI.ESTETT...K....E.ELV.LA..R..L....M..... 60
4cmd  .A..D.V.AG.AE.QAV.E..K--.L.K.V..I..ELV.L...R..L....M..... 58
TACap MTTMEKIFSAIQARVSTLTASRASNPFDGIAWVQGRVTPIHEAQIPMLDQGFLHSDITY 60

          70      80      90      100     110     120
-----|-----|-----|-----|-----|-----|
4ce5  .....IT...A..T.L...L....D..KQI.VE.V.K..... 120
4cmd  .....E...N...A..K....M....EE.IK..VD.V.K..... 118
TACap DVPSVWDGRFFRLDDHLERLERSCAKMRLRCPLPRAQVRHTLCAMLARSGIRDFAVELIV 120

          130     140     150     160     170     180
-----|-----|-----|-----|-----|-----|
4ce5  ...K...TRPEDI---V.N...FVQ.....DM.RV.-..V....R.V..GAI. 176
4cmd  ...T...AKPE.L--.N.N...F.Q.....D.D..YT.-.R.....R.V..GSI. 175
TACap TRGLRGVRLSAAEVDALPNSLYMWIVPYVWVMEPAVQLAGSGSAIVARTVHRTPPVCM 180

          190     200     210     220     230     240
-----|-----|-----|-----|-----|-----|
4ce5  .....V.....A...T.....-----G..HL.....L...V.Y. 229
4cmd  ..I.....V..L.....T.....-----G...L.....V.LI...V.Y. 228
TACap PTVKNLQWGLTRGMFEANDRGAGYPFLTDRSISEETADANITEGSGFNIVVVKDGTLHT 240

          250     260     270     280     290     300
-----|-----|-----|-----|-----|-----|
4ce5  .D...Q...K..INAAEAF.IEVRVEF...E..YRC.....M. 289
4cmd  .D...Q..I..K..IDAA.SC.YEIRVEH..IEATYQ...L.....D.. 288
TACap PKRGVLEGVTRRESVFECRRRLGVPYALDTPVRLALEADEIFMCTTAGGIMPITTLDGKP 300

          310     320     330     340     350
-----|-----|-----|-----|-----|-----|
4ce5  .NG.QI...KK...G..AM..DAAY..EI..NE-RN..... 327
4cmd  .K..K....KA...R..AM.W..E...KIN.L----- 322
TACap VGDGAVGPITRQIWDVYWRHLHYEDGFSFAVDYEDETGLEGVANGNANGTVSVNGKA 356

```

Figura 15 Alineamiento de secuencias molde 4CMD y 4CE5 con TACap. Los puntos representan aminoácidos iguales a TACap y los guiones gaps en el alineamiento.

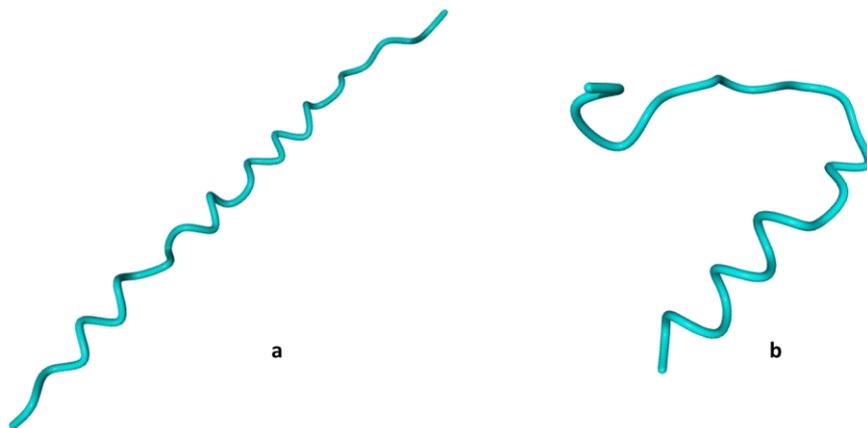


Figura 16 Predicción de estructura de los 28 aminoácidos terminales de TACap. a- AlphaFold2, b- RoseTTAFold.

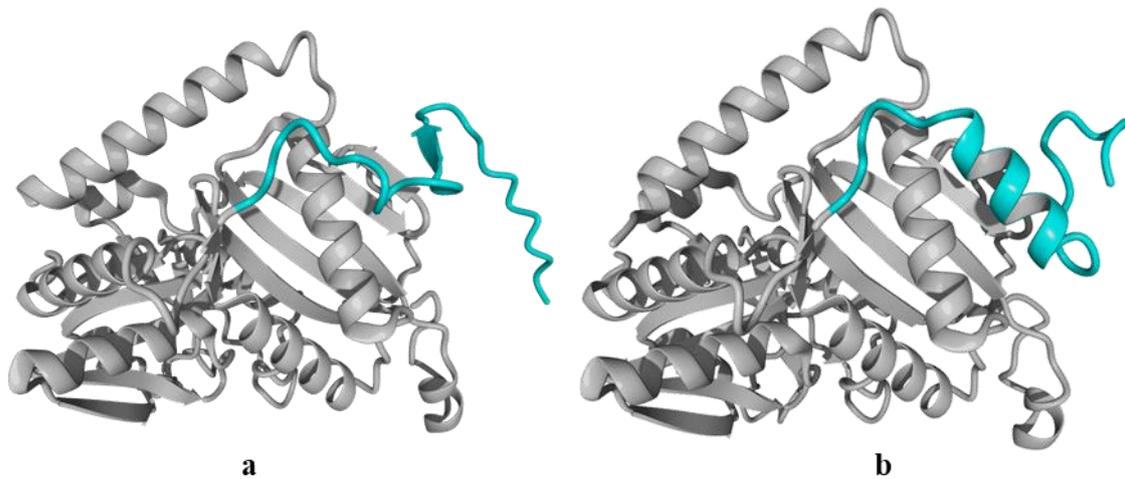


Figura 17 Predicción de estructura de un monómero de TACap. a- AlphaFold2, b- RoseTTAFold.

Durante el desarrollo del trabajo fueron incorporadas nuevas funciones en el ColabFold que implementa AlphaFold2, entre ellas, la posibilidad de predecir la estructura de dímeros. Cuando esto fue posible modelamos la TACap utilizando esta herramienta y el modelo obtenido fue sometido a simulaciones de MD. Una superposición de los modelos obtenidos por homología y por AlphaFold2 se ilustra en la Figura 18. Como puede observarse, las estructuras son muy similares, difiriendo principalmente en los loops y el extremo C-terminal.

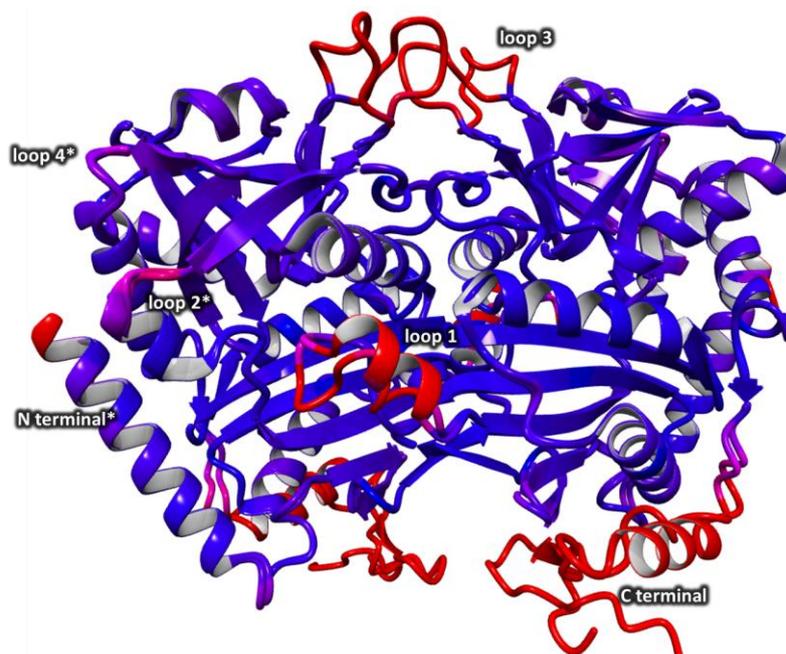


Figura 18 Superposición de los modelos alineados de TACap obtenido por homología + *ab initio* para el extremo C-terminal y el obtenido mediante AlphaFold2. La estructura secundaria se encuentra coloreada según el RMSD entre las estructuras, de menor (azul) a mayor (rojo).

1.3.2. Generación de librería para la aldimina interna Lys-PLP (KLP)

En el ciclo catalítico de las ω -transaminasas en estudio participa el cofactor PLP. Éste se une a la enzima a través de un enlace covalente con una lisina del sitio activo, formando una aldimina interna Lys-PLP, a la cual denominaremos KLP. Para poder realizar simulaciones de una molécula cuyos parámetros no se encuentran en las librerías del campo de fuerza es necesario primero obtener dichos parámetros, *i. e.* coordenadas de átomos, su conectividad e hibridación, tipo de átomos y sus cargas parciales. En general la obtención de las cargas parciales es el paso computacionalmente más costoso ya que involucra cálculos de química cuántica; sin embargo, para esta molécula en particular contamos con los datos de carga proporcionados por Venhorst *et al.*⁹⁷ Los restantes parámetros fueron obtenidos como fuera descrito en la sección 1.2.3. En la Figura 19 puede apreciarse la estructura de la aldimina interna, así como el nombre que recibe cada átomo; las coordenadas, tipo de átomo y carga parcial asignados se indican en la Tabla 3.

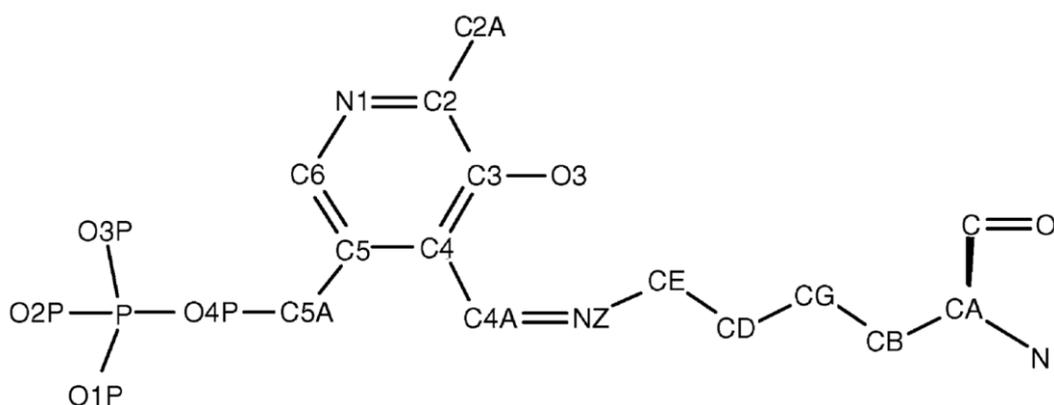


Figura 19 Estructura de la aldimina interna Lys-PLP (KLP). Cada átomo figura con su nombre. Extraído de Venhorst *et al.*⁹⁷

Tabla 3 Valores de coordenadas, tipo de átomo y carga parcial asignada a cada átomo del KLP. Los valores de carga fueron extraídos de Venhorst *et al.*⁹⁷

Nombre de átomo	X	Y	Z	Tipo de átomo	Carga parcial
N	96.438	15.523	44.029	n2	-0.3479
H	96.396	16.470	44.347	hn	0.2747
CA	95.058	15.049	43.680	c3	-0.4419
HA	95.103	14.101	43.365	h1	0.1102
C	94.494	15.910	42.589	c1	0.7341
O	94.445	17.337	43.050	o	-0.5894
CB	94.178	15.130	44.893	c3	0.4995

Nombre de átomo	X	Y	Z	Tipo de átomo	Carga parcial
HB2	94.563	14.559	45.618	hc	-0.1147
HB3	94.125	16.077	45.208	hc	-0.1147
CG	92.805	14.635	44.548	c3	0.0629
HG2	92.406	15.212	43.836	hc	-0.0782
HG3	92.869	13.694	44.217	hc	-0.0782
CD	91.938	14.667	45.772	c3	0.1094
HD2	92.354	14.099	46.482	hc	-0.0058
HD3	91.851	15.606	46.105	hc	-0.0058
CE	90.584	14.119	45.435	c3	0.1156
HE2	90.017	14.096	46.258	h1	-0.0114
HE3	90.147	14.696	44.744	h1	-0.0114
NZ	90.739	12.824	44.928	n3	-0.1399
HZ	91.285	12.677	44.103	hn	0.3819
N1	90.782	8.050	44.076	na	-0.2120
H1	90.938	7.131	43.713	hn	0.3455
C2	91.293	9.124	43.437	c2	0.2058
C2A	92.039	8.954	42.265	c3	-0.2790
H2A1	92.338	9.847	41.928	hc	0.0969
H2A2	91.474	8.508	41.572	hc	0.0969
H2A3	92.839	8.386	42.458	hc	0.0969
C3	91.084	10.362	43.929	c2	0.3557
O3	91.613	11.477	43.267	o	-0.7655
C4	90.364	10.529	45.058	c2	0.1020
C4A	90.171	11.771	45.550	c2	0.1023
H4A	89.636	11.905	46.384	h4	0.2054
C5	89.855	9.456	45.700	c2	-0.1738
C6	90.063	8.216	45.208	c2	-0.0992
H6	89.689	7.420	45.683	h4	0.1919
C5A	89.119	9.630	46.877	c3	0.3499
H5A1	88.823	8.739	47.220	h1	0.0083
H5A2	88.318	10.198	46.689	h1	0.0083
O4P	89.979	10.304	47.906	os	-0.5382
P	89.195	10.491	49.171	p5	1.2439
O1P	90.050	11.182	50.191	o	-0.8970
O2P	87.981	11.326	48.888	o	-0.8970
O3P	88.770	9.153	49.698	o	-0.8970

1.3.3. Simulaciones de dinámica molecular

Los sistemas a simular fueron preparados como fuese indicado en las secciones 1.2.4 y 1.2.5, en las condiciones que se indica en la Tabla 4. Se escogió la enzima 4CE5 como estructura

de referencia dados los numerosos reportes en que se aplicaron distintas aproximaciones para su estabilización.¹¹¹⁻¹¹⁶ Precisamente, la mutación T130M fue una de las primeras en ser reportada como estabilizante para esta enzima; fue hallada a partir del cálculo de $\Delta\Delta G^{\text{fold}}$ sobre las regiones flexibles de la proteína. Por este motivo fue incluida en este estudio para comparar su comportamiento contra la referencia.

Tabla 4 Condiciones de simulación de las distintas enzimas.

Enzima	T (K)	t (ns)	urea	Obs.
4CE5	298	500	X	Enzima empleada como referencia.
	323			
	348			
	298	500	✓	
4CE5-T130M	298	500	X	La estabilidad de este mutante fue reportada como superior a la de la enzima nativa por Huang <i>et al.</i> ¹¹¹
	348			
	298	500	✓	
TACap-HM	298	500	X	TACap obtenida por modelado por homología como fuese indicado en la sección 1.2.1.
	348			
	298	500	✓	

Como fue comentado en la introducción, uno de los métodos para verificar convergencia de una simulación de MD es evaluar el RMSD y/o su promedio acumulado. Como se observa en la Figura 20, para la enzima de referencia 4CE5 estos parámetros parecen haber alcanzado el equilibrio en las simulaciones a 298 y 323 K, no así para 348 K y la simulación con urea. Esto es entendible dado que tanto la alta temperatura como la urea son sumamente desestabilizantes^{117,118}, y es esperable que estas dos simulaciones no converjan hasta que la proteína se haya desplegado. Dadas estas diferencias, decidimos realizar y analizar las siguientes simulaciones solo a dos temperaturas, 298 y 348 K, y en presencia de urea. Además, siendo 4CE5 nuestra referencia y teniendo en cuenta su convergencia, seleccionamos una ventana de análisis para todos los parámetros y todas las simulaciones entre los 300 y 500 ns, descartando así los primeros 300 ns de MD de producción.

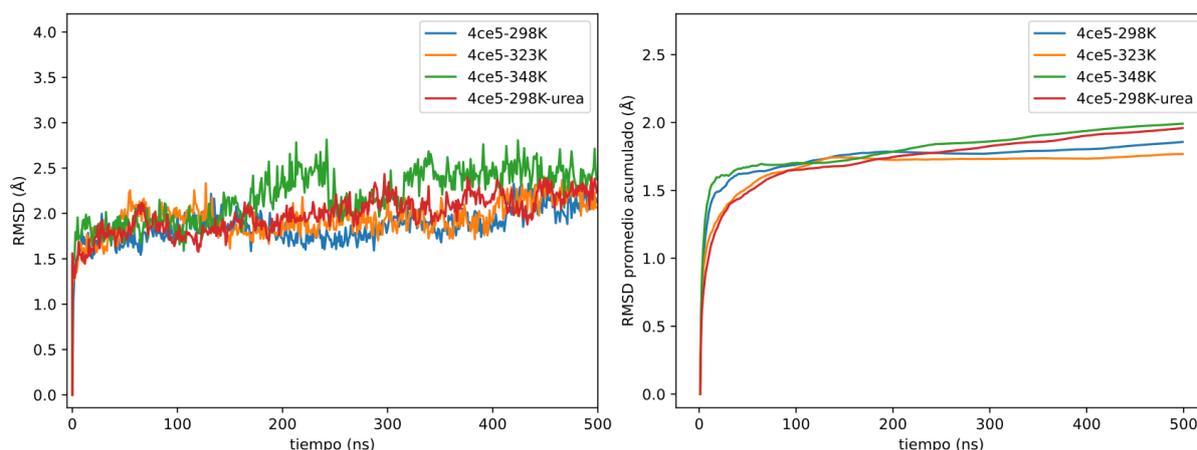


Figura 20 Análisis de MD de 4CE5. RMSD (izq.) y RMSD promedio acumulado (der.)

Los valores promedio de RMSD y su desviación estándar se resumen en la Tabla 5. Como se indica allí y podemos apreciar en la Figura 21, los valores de RMSD de las MD de TACap son más altos que los de 4CE5, lo cual es probable se relacione a la estructura de partida. La estructura de 4CE5 fue determinada experimentalmente, por lo que es probable se encuentre en una conformación cercana a la nativa, mientras que la estructura de TACap fue obtenida por homología y puede tener que recorrer subestados de mayor energía.

Tabla 5 Resumen de valores de RMSD promedio y desviación estándar para las simulaciones corridas.

Enzima	T (K)	Urea	$\overline{\text{RMSD}}$ (Å)	std
4CE5	298	X	1.99	0.14
	348	X	2.42	0.14
	298	✓	2.15	0.13
4CE5-T130M	298	X	2.19	0.12
	348	X	2.33	0.16
	298	✓	2.49	0.20
TACap-HM	298	X	2.89	0.14
	348	X	3.12	0.14
	298	✓	3.12	0.28

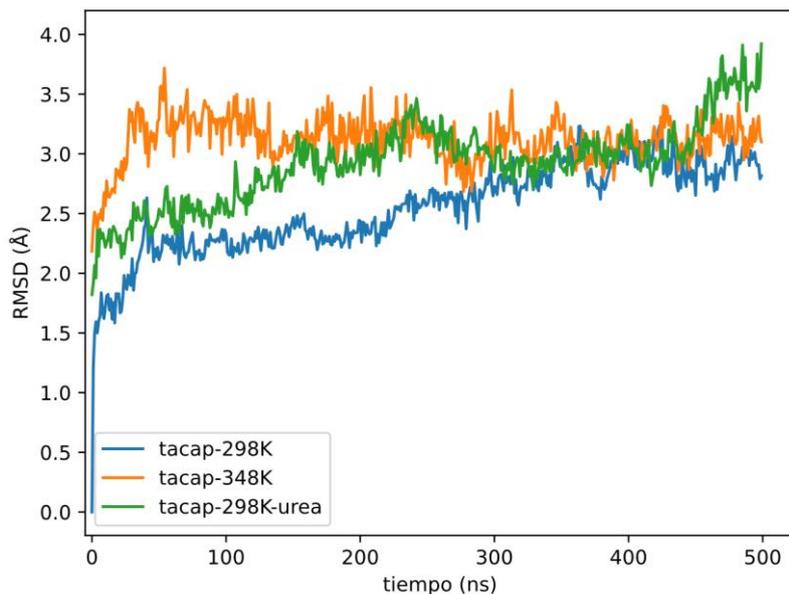


Figura 21 RMSD vs. t para las MD de TACap, alineadas al primer marco a 298 K.

Para el análisis de flexibilidad evaluamos el RMSF, clasificando los valores obtenidos como bajos ($<1 \text{ \AA}$), medios (entre 1 y 2 \AA) o altos ($>2 \text{ \AA}$). Analizando primero la referencia, graficamos el RMSF en la Figura 22 y mapeamos las regiones de mayor flexibilidad sobre la estructura en la Figura 23. Podemos observar que en las simulaciones térmicas (sin urea) las regiones más flexibles son dos o tres residuos en los extremos y un loop entre los residuos 125-133 (loop 1), lo cual no resulta extraño dada su naturaleza flexible⁴¹. Esto reproduce los resultados experimentales de cristalografía¹⁰⁹, ya que este loop posee altos valores de factor B. Cabe mencionar que para dicho loop vemos un RMSF superior a 298 K que a 348 K, lo cual no podemos explicar con los datos disponibles. Por otro lado, al observar el RMSF de la simulación con urea el loop 1 es solo moderadamente flexible, y en cambio es el loop 149-157 (loop 2) el de alta flexibilidad. Esto implica que su presencia como agente desnaturante afecta la estructura y estabilidad de la enzima en forma diferente a como lo hace una alta temperatura, indicando que los mecanismos de desnaturación térmica y química siguen caminos diferentes. A pesar de ser este un resultado esperable, ya que la primera debilita todos los tipos de enlace, mientras que la segunda intercambia enlaces de hidrógeno proteína-proteína por proteína-urea, es remarcable su detección *in silico*.

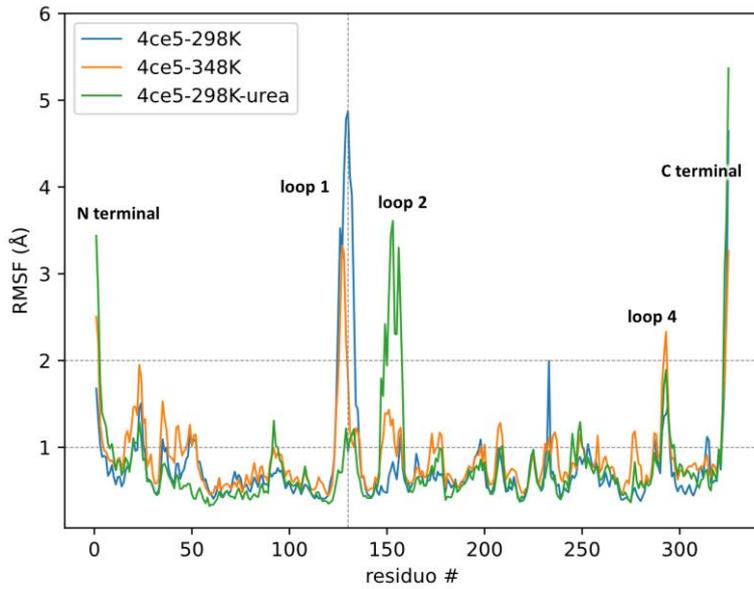


Figura 22 RMSF de 4CE5 en las distintas condiciones de MD. Se indica con línea punteada vertical el residuo 130.

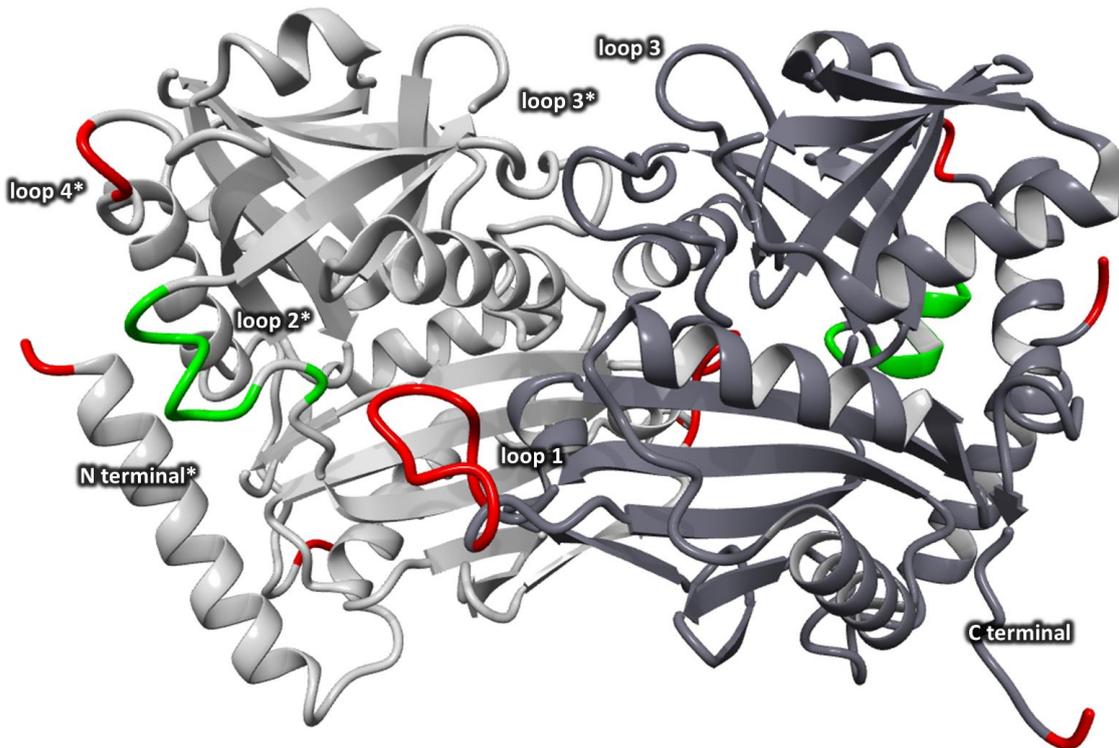


Figura 23 Estructura de 4CE5. En gris oscuro cadena A, gris claro cadena B. En rojo se señalan aquellas regiones con un RMSF superior a 2 Å en MD sin urea, junto a su nombre. Las regiones pertenecientes a la cadena B se señalan con un asterisco. En verde se señala el loop flexible en MD con urea. Es identificado el loop 3 por su relevancia en TACap.

En tanto, al analizar el RMSF de TACap (Figura 24) vemos en las simulaciones térmicas una alta flexibilidad en el loop 1 (residuos 129-134), loop 3 (215-218) y el extremo C-terminal (residuos 336-356). Es de destacar que el loop 3 y el extremo C-terminal son flexibles en TACap, y no así en el modelo de 4CE5, lo cual sea debido seguramente a que en TACap estos segmentos son más largos: loop 3, 7 aminoácidos más; extremo C-terminal 31 aminoácidos más. En la MD con urea vemos acrecentada esta flexibilidad, a no ser en el loop 1 donde es menor; esto mismo ocurriría con 4CE5 y postulamos es debido a una diferencia en el mecanismo de desnaturalización.

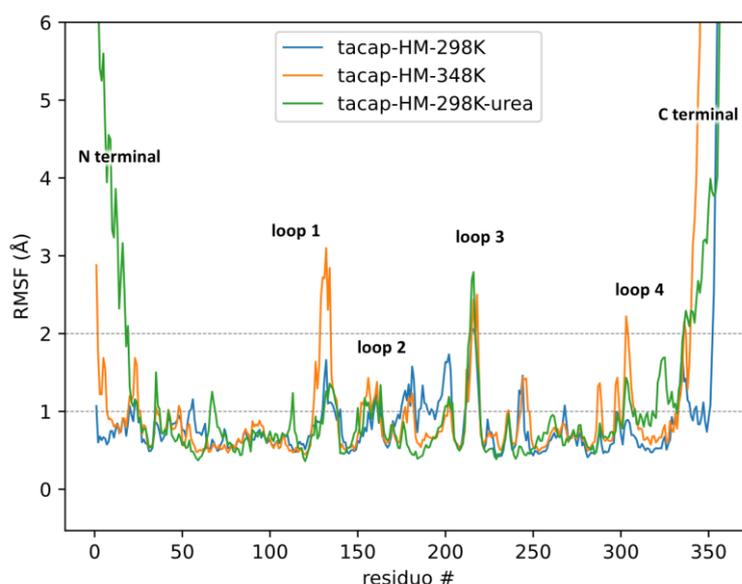


Figura 24 RMSF de las distintas simulaciones de TACap.

El tercer parámetro que analizamos fue el radio de giro, el cual prácticamente no cambia para 4CE5 y su mutante T130M en las distintas condiciones ensayadas (Tabla 6). A temperaturas moderadas (25 a 75°C), un aumento del radio de giro puede ser un indicador de la pérdida de estructura terciaria y secundaria por ruptura de los enlaces de hidrógeno e interacciones de Van der Waals, pero requiere generalmente de simulaciones largas ya que es una propiedad cuyo cambio se empieza a notar cuando el proceso de desnaturalización está en una etapa avanzada. Para el caso de TACap, si bien en la Tabla 6 se observa un pequeño aumento del ROG en la simulación con urea respecto a las otras condiciones, al excluir los extremos C-terminal del análisis esta diferencia desaparece. Esto se ilustra en la Figura 25, e implica que el aumento percibido de ROG es debido exclusivamente al desordenamiento de los extremos C-terminal.

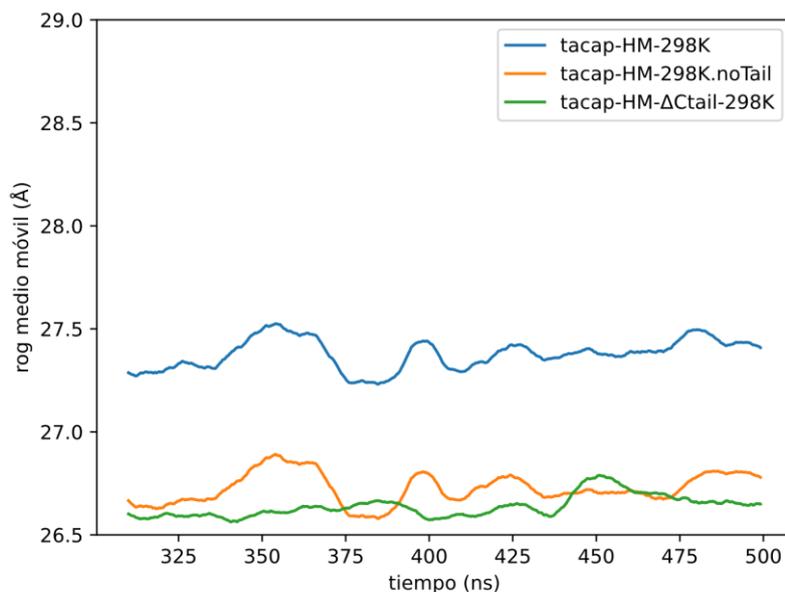


Figura 25 Radio de giro de TACap-HM y TACap-HM- Δ Ctail. La línea anaranjada corresponde al cálculo de ROG sobre la enzima nativa, excluyendo los extremos del análisis (noTail).

A continuación, evaluamos la superficie accesible al solvente. El análisis se hizo sobre los residuos hidrofóbicos, pudiendo de esta forma detectar cambios en la exposición del núcleo de la proteína al entorno, cuyo aumento es un claro signo de desnaturalización. En la Figura 26 graficamos el valor de SAS para las cuatro simulaciones de 4CE5, y podemos verificar dos cosas: existe un aumento del SAS con cada incremento de temperatura, lo cual valida la ventana de análisis empleada; y la urea parece ser un agente mucho más poderoso que la temperatura para inducir el desplegamiento. Si calculamos el porcentaje de aumento de SAS para cada simulación respecto a la media de la simulación a 298 K de la enzima correspondiente (última columna de la Tabla 6) podemos observar que el aumento de temperatura afecta en distinta medida el SAS de 4CE5 (13%) y su mutante T130M (7%), lo cual es coherente con el aumento de estabilidad reportado para este mutante. En el caso de TACap los porcentajes de aumento de SAS son comparables a los de este mutante.

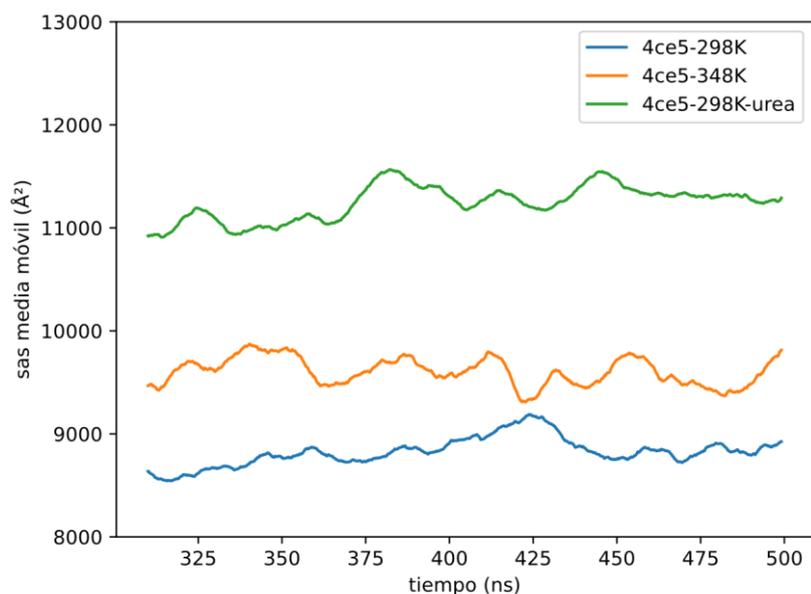


Figura 26 SAS vs. t para 4CE5.

Tabla 6 Parámetros seleccionados promedio en la ventana de análisis, junto a su desviación estándar. En la última columna se presenta el porcentaje de aumento de SAS respecto a su MD a 298 K.

Enzima	T (K)	Urea	ROG (Å)	std	SAS _{apolar} (Å ²)	std	%SAS
4CE5	298	X	26.79	0.07	8736	227	n/a
	348		26.94	0.08	9840	234	13
	298	✓	26.90	0.06	10998	180	26
4CE5-T130M	298	X	26.90	0.08	8808	206	n/a
	348		26.98	0.09	9388	215	7
	298	✓	27.29	0.06	10921	238	24
TACap-HM	298	X	27.44	0.11	9482	196	n/a
	348		27.25	0.08	10021	295	6
	298	✓	28.44	0.06	12366	173	30

n/a – no aplica

Thompson y Eisenberg¹¹⁹ presentaron evidencia, hace más de dos décadas, de la tendencia en proteínas de organismos termófilos a presentar secuencias de loops más cortas que en sus homólogas de organismos mesófilos. Sin embargo, la incorporación de esta observación en una estrategia para la ingeniería de proteínas no parece haber sido extensamente utilizada, o al menos reportada, a no ser por el trabajo de Damjanović *et al.*¹²⁰ concerniente a la delección de un loop de superficie para la mejora de estabilidad de una fosfolipasa.

Teniendo en cuenta esta información, los datos de flexibilidad obtenidos en la dinámica y el alineamiento de secuencias homólogas en el contexto de la determinación de la secuencia consenso (ver sección 1.3.5) propusimos la delección de dos regiones: residuos 211-217 (loop 3) y 332-356 (extremo C-terminal) (Figura 27). Decidimos no modificar los residuos 129-134 (loop 1) ya que esta región ha sido sugerida como muy importante para el reconocimiento de sustrato^{87,121,122}.

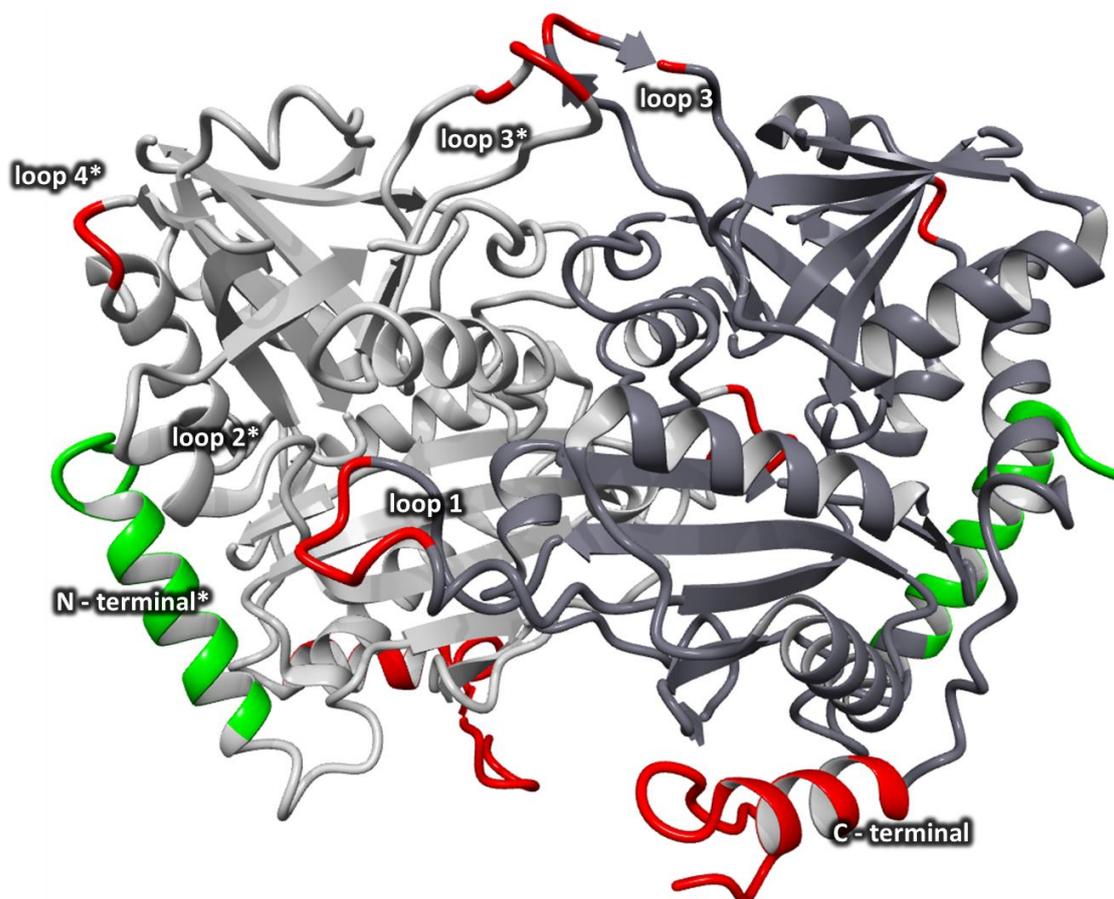


Figura 27 Estructura de TACap. En gris oscuro cadena A, gris claro cadena B. En rojo se señalan aquellas regiones con un RMSF superior a 2 Å en MD sin urea, junto a su nombre. Las regiones pertenecientes a la cadena B se señalan con un asterisco. En verde se señalan las regiones solo flexibles en MD con urea.

Se procedió entonces a la simulación de MD de estas dos variantes de TACap en las condiciones resumidas en la Tabla 7. Por cuestión de tiempo no pudieron ser corridas todas las simulaciones durante el período/en las condiciones deseadas, pero presentamos a continuación los resultados obtenidos para lo ensayado. Para aquellas simulaciones en que no fue posible simular 500 ns se emplearon ventanas más pequeñas, a sabiendas que la simulación puede no haber convergido.

Tabla 7 Resumen de condiciones empleadas en las simulaciones de MD de mutantes de TACap.

Enzima	T (K)	t (ns)	urea	Obs.
TACap - Δ 211-217	298	115	X	TACap-HM con deleción del loop 211-217. Fue seleccionada en base a los resultados de esta sección.
	348	105		
TACap - Δ C-tail	298	500	X	TACap-HM con deleción del extremo C-terminal (residuos 331 en adelante). Fue seleccionada en base a los resultados de esta sección.
	348			
	298	500	✓	

Analizando el RMSF de estos mutantes (Figura 28) y el de la enzima nativa (Figura 24) podemos inferir el efecto de la temperatura sobre las regiones más flexibles. Para esto calculamos el porcentaje de aumento del RMSF a 348 K respecto a 298 K. Como se desprende de la

Tabla 8 ambos mutantes se ven afectados en forma diferente por la temperatura respecto a la enzima nativa. El análisis de los datos de ROG y SAS no arroja diferencias significativas respecto a lo observado para la enzima nativa, por lo que consideramos deseable y necesario proceder a la evaluación *in vitro* de estas variantes.

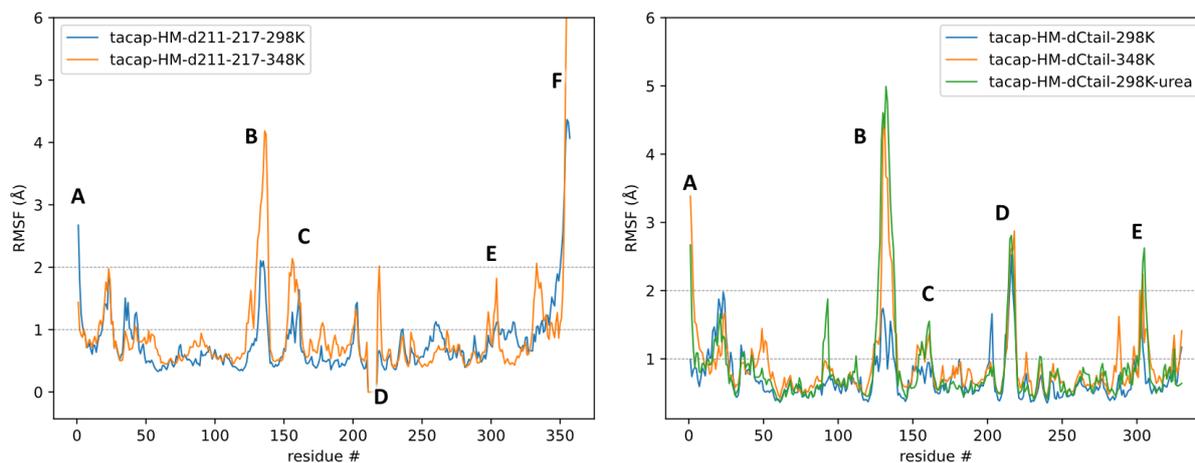


Figura 28 RMSF de los mutantes seleccionados de TACap, bajo distintas condiciones de simulación. A- N-terminal; B a E- loops 1 a 4; F- C-terminal.

Tabla 8 Porcentaje de aumento del RMSF observado en la MD a 348 K respecto a 298 K, para la enzima nativa y los dos mutantes seleccionados.

	loop 1	loop 2	loop 3	loop 4
wt	124	22	17	138
$\Delta 211-217$	80	59	n/a	49
ΔC -tail	156	36	28	20

Tabla 9 Resumen de los parámetros analizados en las dinámicas de TACap y sus variantes.

Enzima	T (K)	Urea	$\overline{\text{RMSD}}$ (Å)	std	rog (Å)	std	SAS _{apolar} (Å ²)	std	%SAS
TACap-HM	298	X	2.89	0.14	27.44	0.11	9482	196	n/a
	348		3.12	0.14	27.25	0.08	10021	295	6
	298	✓	3.12	0.28	28.44	0.06	12366	173	30
TACap-HM $\Delta 211-217^*$	298	X	1.90	0.09	27.00	0.07	9258	223	n/a
	348	X	2.13	0.10	27.07	0.06	9773	232	6
TACap-HM ΔC tail	298	X	2.77	0.10	26.59	0.06	8954	166	n/a
	348		2.92	0.12	26.79	0.09	9301	283	4
	298	✓	3.23	0.12	27.16	0.05	11633	149	30

n/a – no aplica; *tiempos de simulación más cortos

Finalmente, consideramos interesante comentar sobre un aspecto más de índole hipotético-teórico. Como reportaran Börner *et al.*¹²³, la estabilidad del intermediario enzima-PMP es clave en la estabilidad de estas enzimas: su disociación y la pérdida del cofactor, dando lugar a la forma apo, parecen desencadenar los cambios conformacionales que llevarán al desplegamiento de la estructura. Si bien el mecanismo por el que se da este reordenamiento en las transaminasas de clase IV no se encuentra aún dilucidado, Ruggieri *et al.*¹²⁴ sugirieron un posible mecanismo para las transaminasas de clase I: luego de la pérdida del cofactor, la lisina catalítica rota sobre el eje de la estructura primaria debido a la tensión experimentada en la forma holo, provocando el desplazamiento de tres loops y dejando al descubierto hendiduras hidrofílicas, desencadenando de esta forma la disociación del dímero (Figura 29).

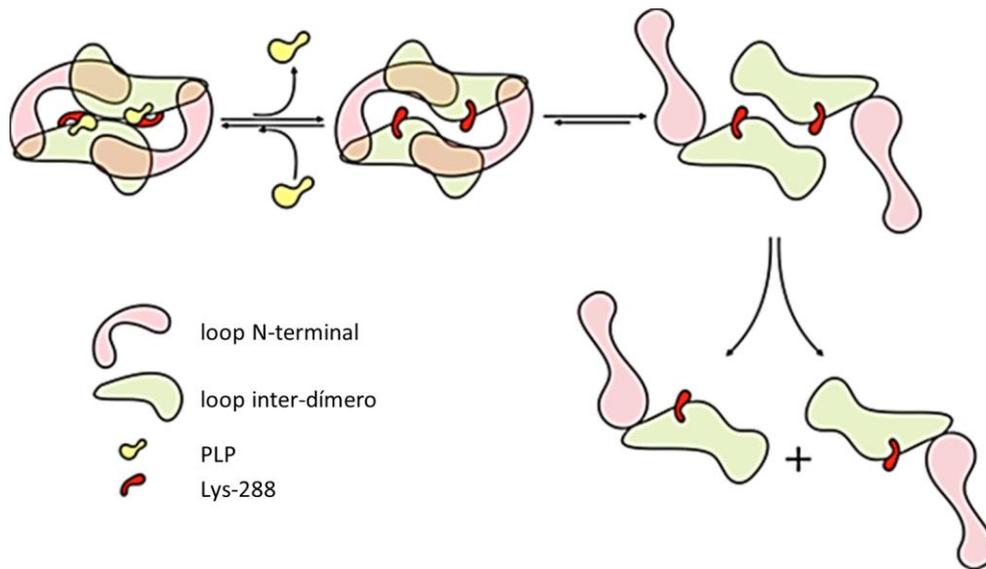


Figura 29 Disociación del dímero de una TA de clase I propuesta por Ruggieri *et al.*¹²⁴ Adaptado de este reporte.

En un estudio de carácter más cualitativo realizamos simulaciones de MD a alta temperatura para verificar la influencia de la presencia de PLP en la estabilidad de TACap. Al observar la película de la dinámica la lisina catalítica libre en la forma apo sufre una reorientación significativa respecto a la situación de la forma holo, como se aprecia en la Figura 30. Si bien haría falta otro tipo de simulaciones y análisis para extraer conclusiones de estos resultados, es esperable que se den rearrreglos por dos motivos: el conjunto de aminoácidos que conforman la cavidad de unión al grupo fosfato del PLP pierden al fosfato como grupo coordinador; y las características electrostáticas del entorno se ven modificadas al pasar de tener al PLP con carga -2 a la lisina con carga +1. Hay por tanto indicios de que el desensamblaje del homodímero podría ocurrir de forma similar que en las transaminasas de clase I.

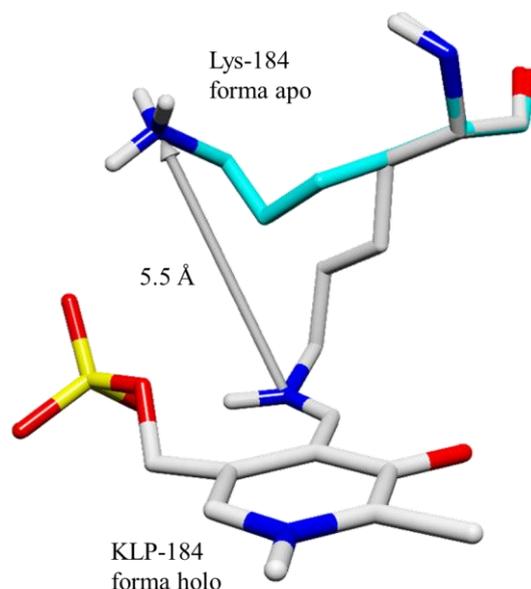


Figura 30 Superposición de la lisina catalítica, luego de la MD, enlazada al PLP (forma holo, carbonos en gris) o libre (forma apo, carbonos en cyan).

1.3.4. Reportes previos

Como fuera mencionado anteriormente existen varios reportes sobre la mejora de estabilidad en la enzima 4CE5 a través de diferentes aproximaciones. Dada la similitud de esta enzima con la nuestra propusimos realizar algunas de las mutaciones que fueron reportadas como exitosas: L118T¹¹⁶, D235K¹¹⁵ y P264A¹¹⁵. La primera mutación reportada, T130M, si bien fue escogida para el análisis bioinformático, no fue propuesta como candidata para TACap ya que en esa posición nuestra enzima tiene una leucina, con características fisicoquímicas similares a la metionina.

1.3.5. Secuencia consenso

Para la obtención de un conjunto de secuencias inicial se hizo una búsqueda en BLASTp con los parámetros predeterminados, solicitando el máximo permitido de 5000 resultados. Estas secuencias fueron depuradas de acuerdo con lo indicado en la sección 1.2.7. La mediana de la longitud de secuencias resultó ser 302 residuos, por lo que aquellas secuencias que diferían en más de un 30% de este valor (<211 o >393 residuos) fueron eliminadas. De esta manera se descartan secuencias truncadas o inusualmente largas (a menudo por largas inserciones internas).

A continuación, se hizo el *clustering* de las secuencias cuya identidad fuese superior al 90% y se conformó un nuevo conjunto de secuencias constituido por una secuencia

representativa de cada *cluster*. Esto permite evitar una sobrerrepresentación de secuencias que tienen un alto porcentaje de identidad entre sí.

Para seleccionar solo aquellas secuencias que sean *R*-amino transaminasas homólogas a TACap realizamos un filtrado⁶⁶ en tres pasos:

- primero se descartaron aquellas secuencias que no contenían los motivos:
 - motivo 1 = [HRK]-x(4)-[YQ]-x-[VTA]-x-[HSAPTG];
 - motivo 2 = [YF]-[VIC]-[EAQSNWMD];
- en segundo lugar, se alinearon las secuencias y se descartaron aquellas que si bien tenían estos motivos, no se alineaban con los correspondientes en TACap;
- en tercer lugar, fueron eliminadas las secuencias que no poseían la lisina catalítica del sitio activo.

Luego de la depuración obtuvimos un alineamiento múltiple (MSA) con 919 secuencias alineadas. Un diagrama del proceso se ilustra en la Figura 31, y el árbol filogenético del MSA se muestra en la Figura 32. Este MSA fue cargado en la plataforma ConSurf para el cálculo del porcentaje de conservación de cada aminoácido en cada posición (Tabla 10, Figura 34). Esta plataforma también provee una escala de conservación del 1 al 9 para cada posición, yendo de más variable a más conservado (Figura 33). A los efectos de este trabajo solo tuvimos en cuenta los resultados con una puntuación del 7 al 9.

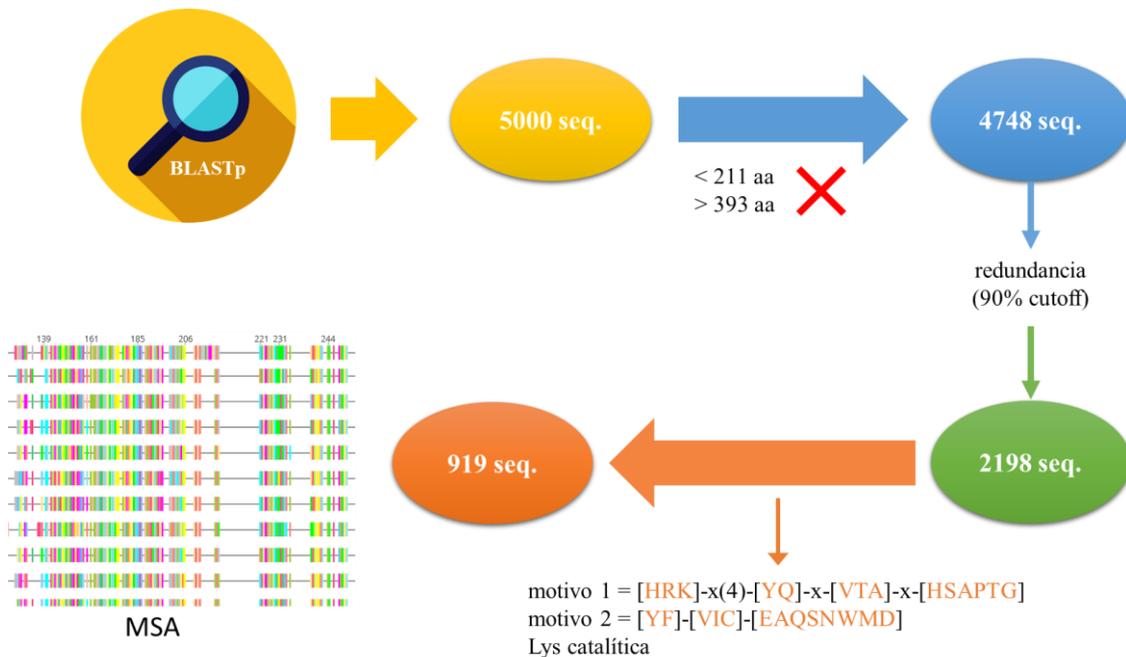


Figura 31 Diagrama del proceso seguido para la obtención del MSA para secuencia consenso.

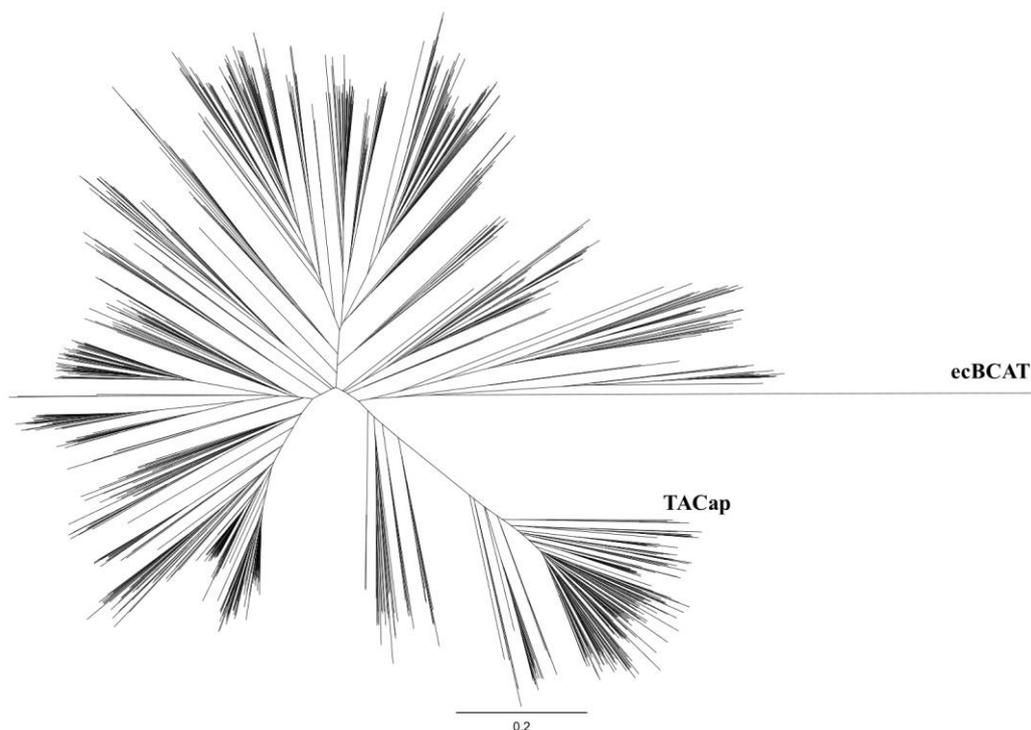


Figura 32 Árbol filogenético de las secuencias seleccionadas para el MSA, empleando la *branched-chain amino acid aminotransferase* de *E. coli* (ecBCAT) como outgroup.

Tabla 10 Porcentaje de conservación del aminoácido nativo (col. 3) y el consenso (col. 5), por posición. En la última columna se presenta la diferencia de conservación entre el aminoácido consenso y el nativo.

Posición	wt	%	consenso	%	diferencia (%)
48	M	3	V	39	36
51	Q	20	W	55	35
58	L	24	A	45	21
63	P	21	V	61	40
64	S	24	H	42	18
80	L	20	F	78	58
87	M	12	L	63	51
101	T	5	I	66	61
105	M	27	C	53	26
106	L	2	V	90	88
111	I	20	L	58	38
115	F	20	Y	80	60
118	L	18	M	62	44
120	V	30	C	46	16
125	R	4	S	31	27

Posición	wt	%	consenso	%	diferencia (%)
130	L	1	S	58	57
131	S	3	R	77	74
132	A	3	D	63	60
133	A	2	P	87	85
135	V	2	R	59	57
142	L	29	F	61	32
144	M	14	A	68	54
145	W	10	F	61	51
146	I	10	A	63	53
153	M	23	A	45	22
163	G	22	L	37	15
165	A	30	L	36	6
168	A	41	S	42	1
174	T	8	I	75	67
178	C	3	S	72	69
179	M	6	V	60	54
186	L	22	Y	56	34
187	Q	27	H	68	41
195	M	16	L	69	53
205	Y	21	T	52	31
206	P	25	A	38	13
207	F	19	V	34	15
220	A	11	G	76	65
226	S	18	P	73	55
231	V	15	F	77	62
249	V	24	I	69	45
253	S	34	T	62	28
257	C	0	L	54	54
258	C	44	A	52	8
281	I	37	V	46	9
284	C	20	T	46	26
285	T	45	S	55	10
293	I	37	V	54	17
313	I	36	L	58	22

Debido a limitaciones prácticas para el análisis experimental seleccionamos como candidatos los 41 mutantes con un mayor porcentaje de diferencia entre la conservación del aminoácido consenso y aquella del nativo.



The conservation scale:



X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figura 33 Secuencia de TACap coloreada de acuerdo con el puntaje de conservación otorgado por ConSurf.

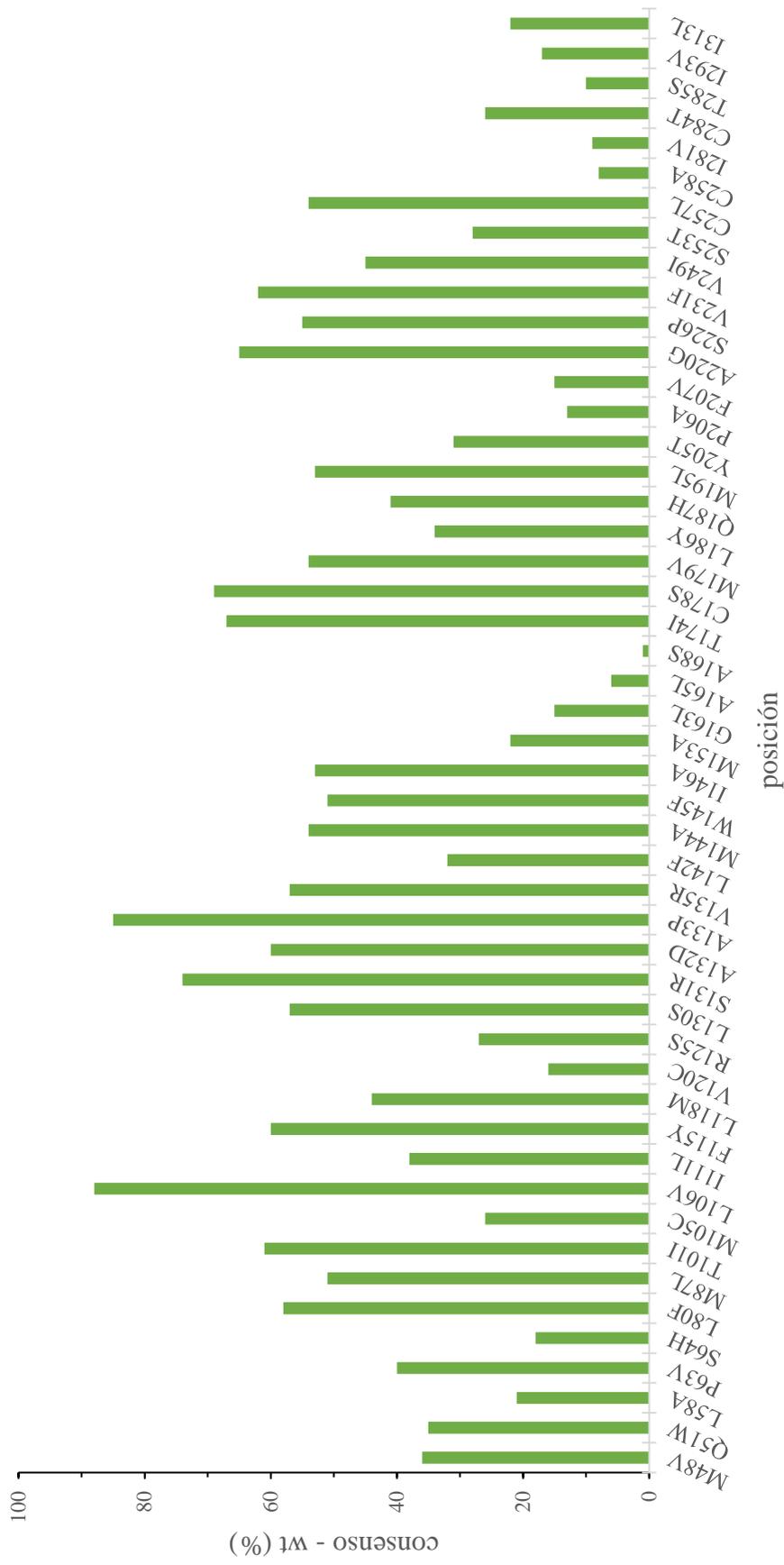


Figura 34 Gráfico de barras representando la diferencia de conservación entre el aminoácido consenso y el nativo.

Adicionalmente, el MSA se empleó para investigar la conservación de aminoácidos en las regiones que decidimos remover en base a las simulaciones de MD. Como puede verse en la Figura 35, el loop que ocupa los residuos 211-217 no se encuentra conservado; de hecho, solo una proteína de las 919 (además de TACap) presenta una inserción en este segmento. En la misma figura se señala también el aminoácido Y332, que resulta ser el último con una alta conservación en el MSA. Por lo anterior, fueron definidos para su remoción los segmentos 211-217 y 333-356.



Figura 35 Extracto del MSA. La primera secuencia (fondo en amarillo) es TACap y se señala con las flechas el loop 211-217 y el último aminoácido conservado, 332.

1.3.6. Cambio de energía libre de plegamiento tras mutación

Esta aproximación fue incorporada en el protocolo FRESCO¹⁰⁷, desarrollado por el grupo del Prof. Fürst donde se realizó una pasantía en el marco de la maestría. El protocolo FRESCO se divide en dos secciones, una primera bioinformática y una segunda experimental. Durante la pasantía se realizó parte de la primera sección, y es sobre lo que se reportará aquí. La sección bioinformática del protocolo FRESCO consiste en:

- i. Preparación de la estructura de partida.
- ii. Selección de residuos a mutar.
- iii. Cálculos de $\Delta\Delta G^{\text{fold}}$.
- iv. Selección de mutaciones a simular por MD.

- v. Simulaciones de MD e inspección de estas.
- vi. Selección de candidatos a expresar experimentalmente.

Si bien esta estrategia ha sido empleada exitosamente para la estabilización de oxidasas⁶⁹⁻⁷², e incluso de una amino transaminasa⁶⁸, todos estos reportes parten de una estructura cristalográfica de la enzima obtenida por medios experimentales y no existen antecedentes publicados en los que se parta de un modelo tridimensional obtenido *in silico*. Por tanto, nuestro interés estuvo focalizado en evaluar esta estrategia partiendo de dos estructuras: TACap-AF-apo y TACap-AF-holo. Decidimos emplear el modelo generado por AlphaFold2 en lugar del de homología ya que, al momento de este análisis, el sistema había sido validado por la comunidad científica¹²⁵. El modelo apo (sin PLP) fue generado por AlphaFold2 y, luego de la optimización de su protonación, fue minimizado mediante YASARA¹²⁶⁻¹³¹. El modelo holo (con PLP) fue obtenido de igual forma, pero antes de ser minimizado fue procesado por AlphaFill, un algoritmo que basado en secuencia y estructura “trasplanta” cofactores y otras pequeñas moléculas desde estructuras determinadas experimentalmente hacia modelos predichos.

Para verificar un correcto posicionamiento del PLP respecto al sitio activo superpusimos las estructuras de TACap-AF-holo y 4CE5, que fue cristalizado con PLP. Como puede observarse en la Figura 36 los aminoácidos del sitio activo que interactúan con el PLP se encuentran conservados y adoptan la misma conformación en ambas enzimas, por lo que consideramos el modelo obtenido apto para proseguir con los análisis.

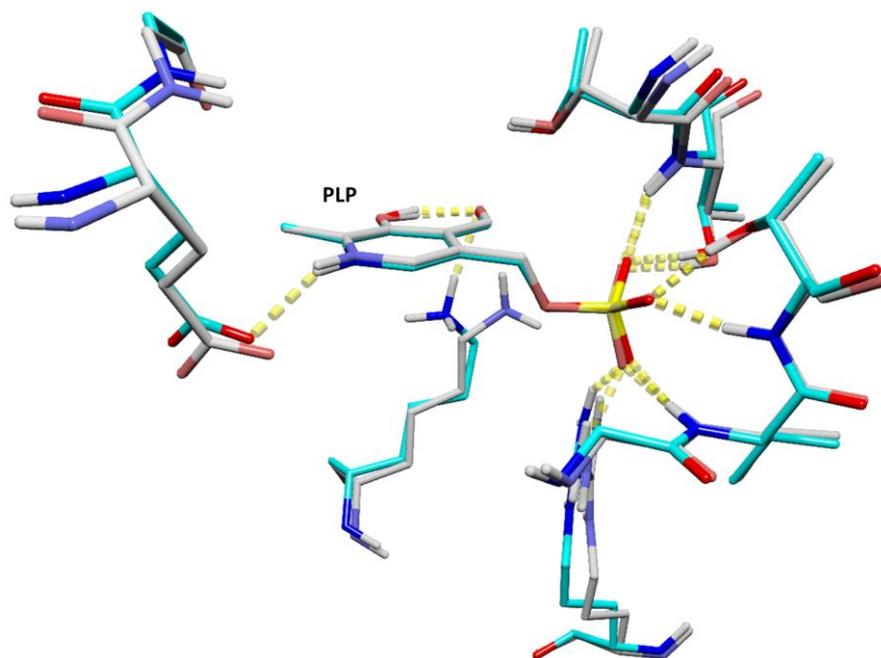


Figura 36 Superposición de los sitios activos de TACap-AF-holo (carbonos en cyan) y 4CE5 (carbonos en gris).

Para el cálculo de $\Delta\Delta G^{\text{fold}}$ permitimos la mutación de todos los residuos, es decir, 356 posiciones multiplicado por 19 aminoácidos distintos al nativo para dar un total de 6764 mutaciones posibles. El ΔG^{fold} de cada variante fue calculado, como indica el protocolo FRESCO, mediante dos algoritmos: FoldX¹³² y Rosetta-ddg¹³³. Obtenidos todos los $\Delta\Delta G^{\text{fold}}$, fueron ordenados por su valor y seleccionadas como favorables todas aquellas mutaciones con una energía menor a -5 kJ/mol. A continuación, se exponen y discuten algunos de los resultados observados.

En primer lugar, se pueden hacer una serie de comparaciones entre las estrategias utilizadas. Rosetta parece proveer valores de $\Delta\Delta G^{\text{fold}}$ mucho más extremos que FoldX, lo que explica el número mucho mayor de mutaciones que predice como estabilizantes (Figura 37). Por otro lado, se puede ver una correlación entre las energías predichas para las estructuras apo y holo por medio de ambos algoritmos, lo cual puede deberse a que no hay cambios sustanciales en la estructura durante la minimización inicial. Si comparamos las mutaciones estabilizantes predichas por uno y otro algoritmo sobre una misma estructura (Figura 38) vemos que coinciden en menos de un 20%, lo cual se encuentra en acuerdo con reportes anteriores^{134,135}.

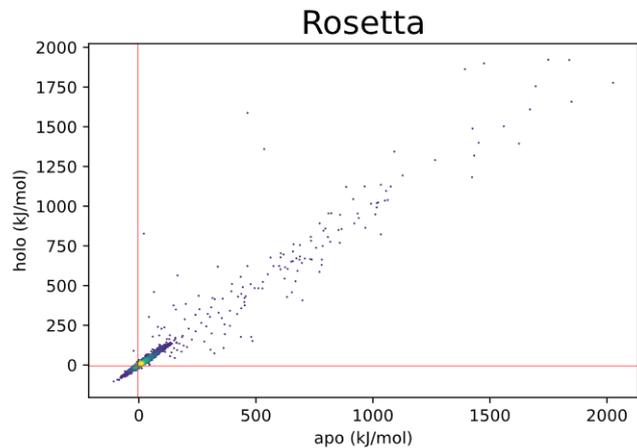
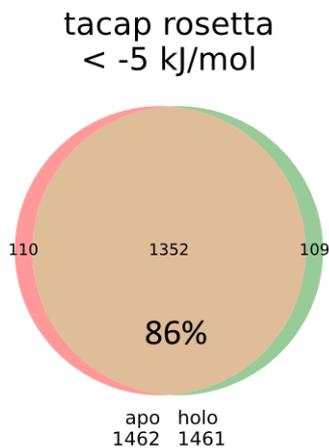
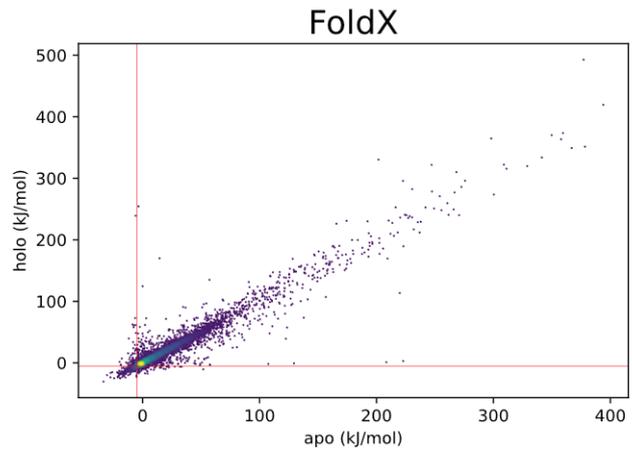
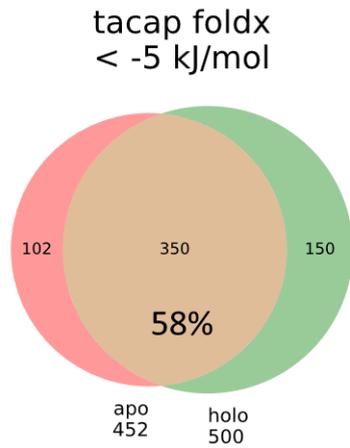


Figura 37 A la izquierda se presentan diagramas de Venn con el número de mutaciones predichas como estabilizantes por FoldX (arriba) y Rosetta-ddg (abajo). A la derecha se comparan las energías de cada mutación calculada por FoldX (arriba) y Rosetta-ddg (abajo).

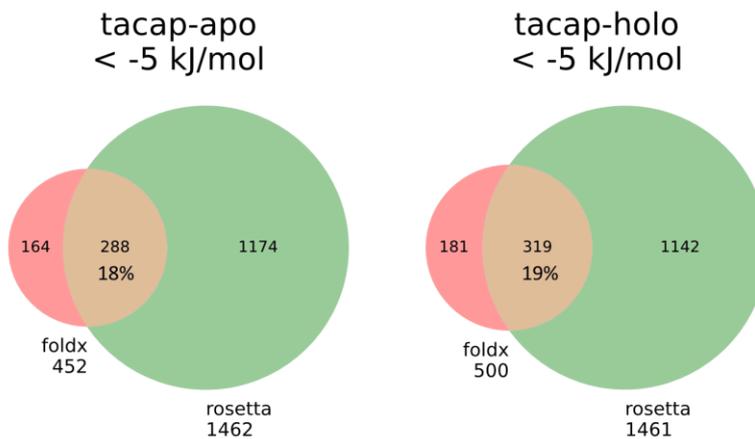


Figura 38 Diagrama de Venn presentando el número de mutaciones predichas como estabilizantes de acuerdo con FoldX o Rosetta-ddg, partiendo de la estructura apo (izquierda) u holo (derecha) de TACap.

En segundo lugar, si representamos todas las energías calculadas por FoldX en un mapa de calor (Figura 39) podemos ver algunas tendencias interesantes:

- Existen al menos tres regiones en las que casi ninguna mutación parece modificar el $\Delta\Delta G^{\text{fold}}$. Estas regiones son muy similares a las que observamos como flexibles en las simulaciones de MD, y puede que esta sea la razón por la cual mutaciones puntuales parecen no tener influencia en la estabilidad.
- Las mutaciones por prolina (P) casi siempre dan un $\Delta\Delta G^{\text{fold}} > 0$ (también en Figura 40). La prolina es un aminoácido particular, ya que es el único cuya cadena lateral se une al nitrógeno de la cadena peptídica formando un anillo pirrolidina y restringiendo así la rotación del enlace $C_{\alpha}-N$, otorgando mayor rigidez al esqueleto. Como explican Jung Choi y Mayo¹³⁶, esto en principio debiera resultar en una mayor estabilidad, sin embargo, no es posible predecir al momento el efecto resultante: una prolina en una hélice alfa o una hoja beta puede romper la estructura secundaria. Por un lado, la prolina carece de un hidrógeno en el nitrógeno del enlace peptídico, perdiendo así la posibilidad de ser donador en un enlace de hidrógeno. Por otro, la restricción estérica generada por el anillo pirrolidina dificulta aún más la formación de estructuras secundarias.
- Las mutaciones por L y M dan en su mayoría $\Delta\Delta G^{\text{fold}} < 0$ (también en Figura 40). Esto puede deberse simplemente a un sesgo del algoritmo por estos residuos. Este tipo de programas pueden tener sesgos inherentes; Usmanova *et al.*¹³⁷ reportaron otro tipo de sesgo en FoldX, Rosetta y otros programas similares.

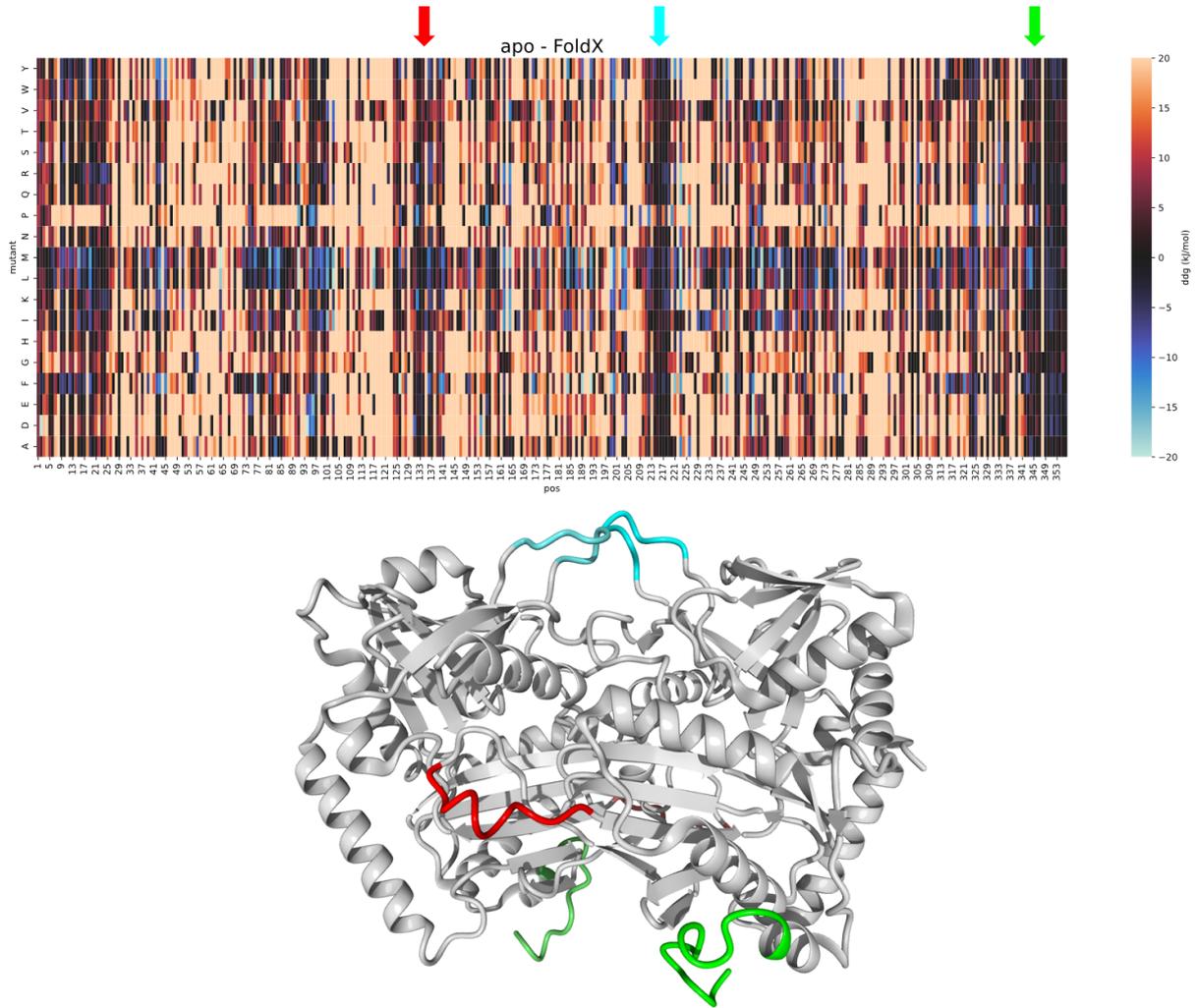


Figura 39 Arriba, mapa de calor de los $\Delta\Delta G^{\text{fold}}$ calculados con FoldX para la forma apo. Se omite el equivalente para la forma holo por ser muy similar. Debajo, mapeo en la estructura 3D de las regiones señaladas con flechas en el mapa de calor.

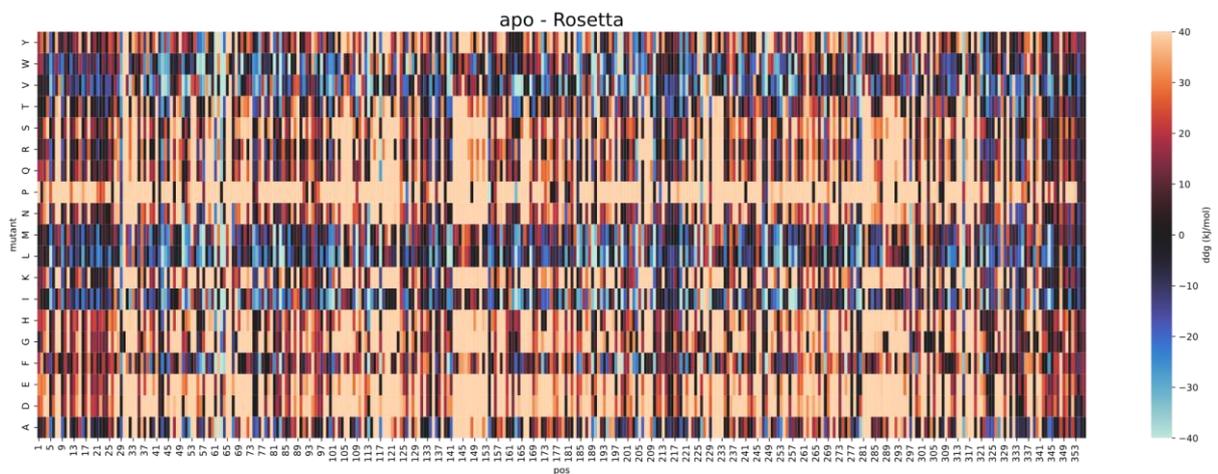


Figura 40 Mapa de calor de los $\Delta\Delta G^{\text{fold}}$ calculados con Rosetta-ddG para la forma apo. Se omite el equivalente para la forma holo por ser muy similar.

1.4. Conclusiones

Se obtuvo un modelo por homología de TACap. La estructura del extremo presentó un desafío ya que no se contaba con estructuras homólogas adecuadas, por lo que finalmente fue modelada *de novo* utilizando RoseTTAFold.

Para realizar ensayos de dinámica molecular fue necesario generar una librería de parámetros para el residuo modificado KLP, correspondiente a la aldimina interna conformada por la unión covalente entre la lisina catalítica y el cofactor PLP, lo que se logró utilizando valores de carga descritos previamente.

Utilizando el modelo de TACap obtenido y la estructura de referencia 4CE5 se corrieron dinámicas de producción a distintas temperaturas. El análisis de estas simulaciones permitió distinguir tres regiones de alta flexibilidad en nuestra enzima: loop 1 (residuos 129-134), loop 3 (215-218) y el extremo C-terminal (residuos 336-356). Se decidió no modificar el loop 1 por encontrarse en la vía de entrada del sustrato, y remover el loop 2 y el extremo C-terminal buscando aumentar la rigidez del sistema. Los puntos de corte fueron determinados por comparación con secuencias similares utilizando un alineamiento múltiple.

Las dinámicas moleculares clásicas permitieron también verificar una diferencia esperable entre los mecanismos de desnaturalización térmica y química.

En forma paralela, se realizó un análisis de secuencia consenso en base a un conjunto de proteínas homólogas a TACap, seleccionadas por criterios de similitud de secuencia y presencia de aminoácidos clave. Se calculó el porcentaje de conservación de cada aminoácido y se comparó con el propio de los aminoácidos presentes en TACap. En base a estos resultados, se propusieron para el análisis experimental 41 mutantes puntuales, los cuales presentan una diferencia de conservación del aminoácido consenso respecto al aminoácido nativo de entre el 15 y el 88%. Adicionalmente, en base a reportes previos de mutaciones estabilizantes sobre la enzima homóloga 4CE5 se seleccionaron tres mutaciones a probar experimentalmente en TACap.

Fue aplicado parte del protocolo computacional FRESCO sobre nuestra proteína. Si bien por motivos de tiempo estos resultados no fueron utilizados para la selección de mutantes, esta información resulta valiosa tanto para analizar otros datos computacionales como para la discusión de los resultados experimentales.

Finalmente, al comparar dinámicas de las formas holo y apo de TACap notamos en la segunda un reordenamiento de la lisina catalítica coincidente con un reporte anterior acerca del

mecanismo de desplegamiento en transaminasas de clase I, y postulamos que en esta enzima puede suceder algo similar.

A modo de resumen, se señalan en la Tabla 11 las modificaciones que fueron seleccionadas para analizar experimentalmente sus efectos sobre la estabilidad de TACap.

Tabla 11 Resumen de las mutaciones seleccionadas para ensayar experimentalmente, junto con el método y el motivo por el que fueron postuladas.

Mutación	Método	Motivo
Δ 211-217 Δ C-tail	MD	Alta flexibilidad. Las posiciones de corte fueron fijadas de acuerdo con MSA.
L118T	reporte	Coevolución de residuos.
D235K P264A		Interacciones carga-carga.
mutaciones consenso	sec. consenso	41 mutaciones.

Capítulo II:

Clonado, expresión y análisis de variantes enzimáticas

2.1. Introducción

Superado el primer capítulo es hora de levantarse y pasar a la mesada. Aquí comentaremos sobre el clonado de TACap en un vector de expresión y cómo fueron obtenidos los mutantes propuestos en el capítulo anterior. Luego, analizaremos los datos de estabilidad termodinámica de estos mutantes y discutiremos sobre la correlación entre los resultados *in silico* e *in vitro* (*spoiler alert*: ni tan buena como se promociona, ni tan fantástica como dice el *reviewer* #2).

Finalmente, conjugamos los resultados de estabilidad y actividad para producir variantes de mejores características que podrán ser empleados para la síntesis del L-nordeuterodeprenil.

2.2. Materiales y métodos

2.2.1. Cepas bacterianas y plásmidos

La cepa utilizada en este trabajo fue *E. coli* TOP10 obtenida de New England Biolabs, cuyo genotipo es *F- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ*. Los vectores utilizados fueron pKK223-3 de GenScript ('pKK223-3.geneious'), pBAD-myc-HisA ('pBAD-myc-HisA.geneious') y pBAD-C-term-His ('pBAD-C-term-His.geneious'), cedidos por el Prof. Fürst (RUG).

Se utilizaron técnicas estándar de microbiología para el manejo de los biocatalizadores¹³⁸. Para su mantenimiento a largo plazo, las cepas fueron conservadas en freezer a -70°C en glicerol 20%, partiéndose para cualquier experimento del primer repique de estos. En el corto plazo fueron mantenidas a 4°C en placas de medio de cultivo sólido (LB-kanamicina), selladas con Parafilm. Los plásmidos construidos son mantenidos en la cepa indicada, pudiendo ser purificados utilizando kits comerciales de *miniprep*, de acuerdo con el protocolo establecido por el proveedor.

2.2.2. Medios de cultivo y enzimas

Triptona, extracto de levadura y agar para la preparación de medios de cultivo fueron adquiridos de Oxoid. La ampicilina, L-arabinosa, sales, glucosa, y otros componentes de los medios de cultivo se adquirieron de Sigma-Aldrich y Droguería Industrial Uruguay. Todas las enzimas fueron obtenidas de Thermo Scientific o New England Biolabs: RNAsa, lisozima,

ADNasa, polimerasas, ligasa, enzimas de restricción, al igual que los marcadores en escalera de ADN (GeneRuler 1 kb DNA ladder SM0311; GeneRuler 1 kb plus DNA ladder SM1331) y proteínas (PageRuler prestained protein ladder 10-180 kDa #26616).

Medio Luria Bertani (LB): triptona 10 g/L, extracto de levadura 5 g/L, NaCl 10 g/L y agua destilada c.s.p. 1 L. Agregar agar (15 g/L) para la preparación de medio sólido. Autoclavar y suplementar con antibiótico al momento de uso de ser necesario.

Medio TB: triptona 12 g/L, extracto de levadura 24 g/L, glicerol 4.8 g/L y agua destilada c.s.p. 900 mL; autoclavar. Por separado preparar fosfato monobásico de potasio (KH_2PO_4) 2.3 g/L, fosfato dibásico de potasio (K_2HPO_4) 12.5 g/L y agua destilada c.s.p. 100 mL; autoclavar. Mezclar antes de usar.

Medio L-broth: triptona 10 g/L, extracto de levadura 5 g/L, NaCl 5 g/L y agua destilada c.s.p. 1 L.

Medio SOB: triptona 20 g/L, extracto de levadura 5 g/L, NaCl 0.5 g/L y agua destilada c.s.p. 1 L.

Medio SOC: suplementar el SOB con MgSO_4 y MgCl_2 ambos en una concentración final de 10 mM.

2.2.3. Soluciones y materiales

Los materiales de trabajo estériles fueron preparados siguiendo técnicas estándar de laboratorio microbiológico (escarbadientes, tips, tubos de microcentrífuga, etc.) Los buffers cuya composición no se indica fueron preparados de acuerdo con lo indicado en el manual Sambrook¹³⁸. Los geles de poliacrilamida para electroforesis de proteínas fueron adquiridos de GenScript (ExpressPlus 4-12%, Bis-Tris M41215), así como el buffer de corrida MOPS.

Buffer HEPES 100 mM: HEPES 26 g/L y agua destilada c.s.p. 1 L. Ajustar pH con HCl a pH 8.

Buffer Tris: Tris base 12 g/L y agua destilada c.s.p. 1 L. Ajustar pH con HCl a pH 8.

Buffer de lisis: lisozima 1 g/L, DNAsa I 5 mg/L, RNAsa A 5 mg/L, SigmaFast (S8830) 10 tabs/L, $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ 10 mM, PLP 0.01 mM, HEPES pH 8 50 mM y agua destilada c.s.p. 1 L.

Buffer isopropilamina 3 M (IPAm): HEPES 26 g y agua destilada 50 mL. Agregar en campana isopropilamina 34 mL. Muy lentamente y con agitación llevar a pH 8-9 con HCl concentrado (aprox. 23 mL).

2.2.4. Extracción de ADN (miniprep)

Se siguieron las instrucciones del proveedor. A modo de ejemplo se adjunta el protocolo empleado para la mayoría de las minipreps de esta tesis del proveedor Qiagen ('Qiagen miniprep.pdf').

2.2.5. Electroforesis de ADN en gel de agarosa

- i. Suspendimos la agarosa en buffer TBE 0.5x para una concentración de 0.8%.
- ii. Calentamos en microondas a máxima potencia a intervalos de 30 segundos, agitando y revisando hasta que la agarosa se haya disuelto.
- iii. Agregamos 2 μL de intercalante cada 100 mL de solución (o lo sugerido por el fabricante).
- iv. Mezclamos y vertimos en placa para gel dentro del molde correspondiente. Incorporamos la peineta y dejamos solidificar durante 20 minutos.
- v. Removimos con cuidado la peineta y transferimos el gel junto con su placa a la cuba para electroforesis. Llenamos la cuba con buffer TBE 0.5x.
- vi. Sobre Parafilm mezclamos 5 μL de cada muestra con 1 μL de loading buffer.
- vii. Sembramos el marcador escalera de ADN (2 μL) y las muestras (6 μL).
- viii. Corrimos la electroforesis a 100 V durante 30 minutos.
- ix. Tomamos foto del gel en cámara UV.

2.2.6. Purificación de ADN por gel

Seguimos en general mismo procedimiento que para la electroforesis normal, pero en lugar de agregar loading buffer, agregamos 1/5 del volumen de muestra a purificar de glicerol 30%. Calculamos el volumen que puede agregarse de muestra por pocillo, y de ser necesario unimos con cinta algunas púas de la peineta para generar pocillos más grandes. Tuvimos la precaución de dejar al menos un pocillo libre entre muestras y entre la muestra y el marcador.

- i. Una vez corrido el gel llevamos a placa UV, tomando las precauciones necesarias para evitar la exposición directa del operador. El gel también debe ser expuesto a la luz UV la menor cantidad de tiempo posible.
- ii. Cortamos con un bisturí limpio la sección de ADN a purificar. Eliminamos de la sección cortada todo el gel posible que no contuviese ADN.

- iii. Transferimos a un tubo de centrifuga y empleamos un kit comercial para eluir el ADN desde la matriz de agarosa. A modo de ejemplo se adjunta el protocolo del kit de Bio-Rad ('Bio-Rad freeze 'n squeeze.pdf').

2.2.7. Restriction Free Cloning (RFC)

- i. Amplificamos la secuencia de ADN de interés por PCR empleando una polimerasa de alta fidelidad.
- ii. Purificamos el producto por gel de agarosa.
- iii. Verificamos la purificación por electroforesis en agarosa. Cuantificamos por gel o nanodrop.
- iv. Amplificamos el vector objetivo utilizando 30 ng de ADN molde y 120 ng del megaprimer o amplicón obtenido en ii. Ajustamos el tiempo de extensión de la PCR de acuerdo con el largo del molde a amplificar. Incluimos un blanco sin primers que será el control negativo hasta la transformación.
- v. Digerimos muestra y blanco con DpnI de acuerdo con las instrucciones del fabricante.
- vi. Precipitamos con ARNt.
- vii. Transformamos células competentes de *E. coli* por electroporación.

2.2.8. Concentración de ADN

- i. Estimamos el volumen de solución a concentrar = V_i .
- ii. Agregamos $V_i \times 0.025$ volúmenes de NaCl 4 M.
- iii. Agregamos $V_i \times 2.5$ volúmenes de EtOH abs. Frío.
- iv. Incubamos a -20°C overnight.
- v. Centrifugamos a 13 000 xg, 20 min, 4°C .
- vi. Transferimos el sobrenadante a otro tubo por vertido por la cara opuesta al pellet y conservamos hasta verificar que se recuperó el ADN.
- vii. Agregamos sobre el pellet un volumen = V_i de EtOH 70% frío.
- viii. Repetimos v-vi.
- ix. Dejamos secar el pellet a temperatura ambiente.
- x. Retomamos en un volumen adecuado de buffer Tris, Tris-EDTA o agua ultrapura.

2.2.9. Precipitación de ADN con ARNt

Cuando se cuenta con muy poca cantidad de ADN (por ej. luego de una ligación o RFC) puede realizarse el procedimiento de concentración de ADN indicado arriba, añadiendo ARNt como co-precipitante. De esta forma suele aumentar significativamente la eficiencia de transformación. Seguimos estos pasos:

- i. Estimamos el volumen de solución a concentrar = V_i .
- ii. Agregamos V_i volúmenes de ARNt de levadura (solución de uso).
- iii. Agregamos $V_i \times 2$ volúmenes de agua ultrapura.
- iv. Agregamos $V_i \times 10$ volúmenes de EtOH abs. Frío.
- v. Incubamos a -20°C 1 h u overnight.
- vi. Centrifugamos a 13 000 xg, 20 min, 4°C .
- vii. Transferimos el sobrenadante a otro tubo por vertido por la cara opuesta al pellet y conservamos hasta verificar que se recuperó el ADN.
- viii. Agregamos sobre el pellet 100 μL de EtOH 70 % frío.
- ix. Repetimos vi-vii.
- x. Dejamos secar el pellet a temperatura ambiente.
- xi. Retomamos en 2 μL de agua ultrapura y utilizamos para transformar.

2.2.10. Golden Gate Cloning

Se adjunta el protocolo empleado ('GoldenGateCloning.pdf').

2.2.11. Mutagénesis sitio-dirigida

Los primers a utilizar deben seguir las recomendaciones habituales para su diseño (longitud, contenido GC, extremos, etc.)¹³⁹ y deben ser complementarios a la región a modificar. Si bien el manual de QuikChange¹⁴⁰ indica que los primers sean completamente complementarios, algunos autores recomiendan cierto grado de desfasaje para lograr un mejor resultado: la región 5' será complementaria y contendrá la mutación a realizar, mientras que la región 3' no será complementaria (primers desfasados, ver Figura 41)^{139,141}.

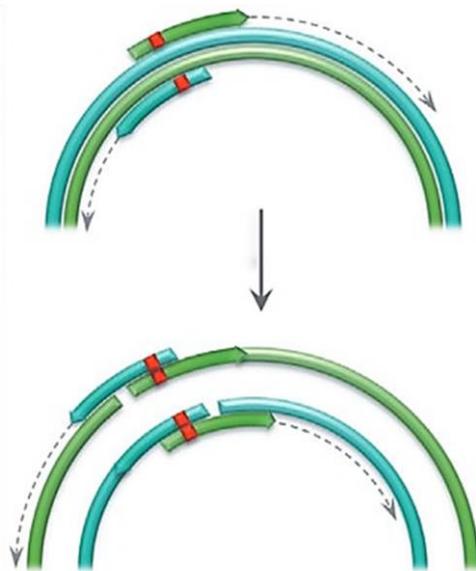


Figura 41 Mutagénesis sitio-dirigida mediante QuikChange, empleando primers desfasados. Extraído de Fürst *et al.*¹⁴¹

Si van a realizarse varias mutaciones puntuales puede utilizarse un programa para el diseño de los primers que tiene en cuenta estos requisitos: Aascan¹⁴². Un manual paso a paso para su uso fue elaborado por Fürst *et al.*¹⁴¹ Este manual fue seguido también para generar los mutantes puntuales en lote, en placas de 96 pocillos para PCR; el procedimiento a continuación es para la realización de mutantes puntuales individuales.

- i. Preparamos la reacción de PCR de forma habitual, siguiendo las instrucciones del proveedor de la polimerasa, pero por duplicado. En una de las reacciones agregamos el primer forward, y en la otra el reverse. Agregamos agua ultrapura para llevar a volumen.
- ii. Corremos un programa adecuado, con 5 ciclos de amplificación.
- iii. Unimos las reacciones forward y reverse, y volvemos a largar el programa esta vez con 25 ciclos de amplificación.
- iv. Verificamos amplificación por electroforesis en agarosa.
- v. Digerimos con DpnI siguiendo las instrucciones del fabricante.
- vi. Purificamos el ADN mediante kit de purificación de PCR.
- vii. Transformamos en células competentes.

2.2.12. Delección de regiones

Los primers deben diseñarse de forma tal de ser complementarios a las secuencias adyacentes a la región a eliminar.¹⁴³ Proceder de igual forma que para la mutagénesis sitio-

dirigida, es decir, comenzando con los primers forward y reverse por separado y luego uniendo las reacciones para la amplificación final. El producto obtenido es luego tratado con NEB HiFi DNA Assembly MasterMix, que contiene una exonucleasa, una polimerasa y una ligasa para el ensamblaje de los fragmentos *in vitro*.

2.2.13. Primers utilizados

2.2.13.1. Subclonado de TACap desde pKK223-3 a pBAD

atnei32-fw	gctaacaggaggaattaacc
atnei32-rev	caaacagccaagctggag

2.2.13.2. Subclonado de TACap desde pKK223-3 a pBAD con delección de extremo C-terminal

atnei33b-fw	gctaacaggaggaattaaccatgaccactatggaaaagatttctctg
atnei33b-rv	tttaaactcaatgatgatgatgatgatgcacggcgaatgaaaaccgt

2.2.13.3. Delección de extremo C-terminal de TACap clonada en pBAD

atnei33a-fw	ttttcattcgccgtgcatcaccatcaccatcac
atnei33a-rv	gtgatggtgatggtgatgcacggcgaatgaaaaac
dCtail-Fw	gtggattatcatcaccatcaccatcac
dCtail-Rv	atggtgatgataatccacggcgaatgaaaaac
dCtail2-fw	accggcctgcatcaccatcaccatcactaacg
dCtail2-rv	atggtgatgcaggccggttcgtcttc

2.2.13.4. Delección de loop 211-217

d211-217-Fw	tgaccgatgcagacgcgaatattactg
d211-217-Rv	cgtctgcatcggtcaggaaggatc

2.2.13.5. Mutaciones puntuales

primer	forward	reverse
M48V	ccctGTGttggaccaaggcttctctg	ggccaaCACaggatctgagcttcatg
Q51W	gttgacTGGggcttctctgcacagc	gccCCAgccaacataggatctgagc
L58A	gcacagcgaatGCGacttacgacgttcc	CGCatcgctgtgcaggaagccttgg

primer	forward	reverse
P63V	cgacgttGTGagtgtgtggatgtag	actCACAacgtcgttaagcaaatcgtgtgc
S64H	gacgttccgCATgtgtggatgtag	cATGcggaacgtcgttaagcaaatcgtgtgc
L80F	ccatctggaacgtTTTgagagaagttgtgc	acgttccagatggatccaggcggaaaaatc
M87L	CTGagactgcgttgcctttgccaag	caacgcagtctCAGctttgcacaacttctc
T101I	cacATCctgtgtgctatgttgccag	cacacagGATgtgacgaacctgtgc
M105C	cctgtgtgtTGcTtgccagatctgg	GCAagcacacagggtgtgacgaacctg
L106V	gGTGgccagatctgtattcgcgac	agatctggcCACcatagcacacagggtg
I111L	gatctggtCTGcgcgacgattgtgg	gcgCAGaccagatctggccaatagc
F115Y	gcgacgcaTATgtggaactgatcgtc	ccacATAtgcgtcgcgaataccagatc
L118M	ggaaATGatcgtcactcgcggcttg	gacgatCATttccacaaatgcgtcgcgaatac
L118T	ggaaACGatcgtcactcgcggcttg	gacgatCGTttccacaaatgcgtcgcgaatac
V120C	gatcTGCactcgcggcttgctg	cgcgagtGCAgatcagttccacaaatgc
R125S	ggcttgAGCggagtcagaggtctg	ctccGCTcaagccgcgagtgacg
L130S	cagaggtAGCtctgctcgcgaggttg	cagaGCTacctctgactccacgcaagc
S131R	ggtctgCGTgctgccgaggtgatg	ggcagcACGcagacctctgactcc
A132D	GATgccgaggttgatgcgctgc	caacctcggcATCagacagacctctgac
A133P	ctgctCCGgaggttgatgcgctgc	acctcCGGagcagacagacctctgac
V135R	gccgagCGTgatgcgctgcctaac	cgcatcACGctcggcagcagacag
L142F	gctgcctaactcTTTtatatgtggattgtg	ggagttaggcagcgcatacaacctc
M144A	aactccttgatGCGtggttgccatacag	Catacaaggagttaggcagcgcatacaacc
W145F	gTTTattgtgccatacgtttgggtatggaac	tggcacaatAAAcatatacaaggagttaggcag
I146A	gtggGCGgtgccatacgtttgggtatgg	tatggcacCGCccacatatacaaggagttagg
M153A	gggttGCGgaaccggctgttcagc	cggttcCGCaacccaacgtatggcac
T174I	ATTccaccggttgatgcatccgac	caaaccggttgAAATctatgactgtacg
C178S	ccggttAGCcatggtaccgacagtg	ccatGCTaaccggtggcgttctatg
M179V	cggtttgcGTGgatccgacagtgaataatc	gatcCACgcaaacgggtggcgttc
L186Y	TATcaatggggcgacttgacgagagg	cgccccattgATAattttcactgtcggatc
Q187H	ctgCATtggggcgacttgacgag	cgccccATGcagattttcactgtgc
M195L	tgacgagaggaCTGtttgaagtaacgatc	AGtctctctgcaagtcgccccattg
Y205T	gccggaACCccttctgaccgatag	aaggGGTtccggcggcggatcgc
A220G	aaccgcagacGGCaatattactgaaggttc	ttGCCgtctcgggttcttcgctaattgatc
D224K	actAAAggttccggcttaacatcgttgtg	ccggaaccTTTtagtaattcgcgtctgc
S226P	ctgaaggtCCGggctttaaactcgttgg	ccCGGaccttcagtaattcgcgtctgc
V231F	cTTTgttggtcaaatggtgacacttcgac	tttgaccacAAAgatgttaaagccggaacc
V249I	cctggagggcATTacaagagaatctgtg	gtAATgcctccaggacaccag
P253A	gagaaGCGgtgttcgagttgccgtag	gaacacCGTtctcttgaacgcctcc
S253T	gagaaACCgtgttcgagttgccgtag	gaacacGGTtctcttgaacgcctcc

primer	forward	reverse
C257L	cgagTTAtgccgtagattgggtgtcc	ctacggcaTAAActcgaacacagattctcttg
C284T	ACCacaacggccggtggtattatgccgatc	cggccgttgTGTcataaaaatctcatcggcttc
I293V	gccgGTGaccactctggatggaaaacc	agtggTCAcCggcataataaccaccggccgttg
V306P	gacggtgccCCGggtccaattactcg	ccCGGggcaccgtctcccacagg
I313L	tcgcaaCTGtgggatgtttattggcgtttg	ccaCAGttggcgagtaattggaccgacg

2.2.13.6. Mutantes múltiples

L106V_fw	ggtggccagatctggtattcgcgac
T101I+L106V_rev	agatctggccaccatagcacacaggatg

2.2.13.7. Secuenciado

pBAD-FP	atgcatagcattttatcc
pBAD-R	atgcatagcattttatcca

2.2.13.8. Verificación de presencia de inserto por PCR

pBAD-Forward	atgcatagcattttatcc
pBAD-Rv	atccgcaaaaacagccaag

2.2.14. Condiciones de PCR

2.2.14.1. Subclonado de pKK223-3 a pBAD

Primera PCR. Inicio: 98°C, 1 min 30 s; **Amplificación x 30:** 98°C, 10 s; 60°C, 30 s, 72°C, 33 s; **Fin:** 72°, 7 min. Phusion DNA polymerase.

Segunda PCR. Inicio: 98°C, 1 min 30 s; **Amplificación x 30:** 98°C, 10 s; 62°C, 30 s, 72°C, 1 min 58 s; **Fin:** 72°, 7 min. Phusion DNA polymerase.

2.2.14.2. Subclonado de TACap desde pKK223-3 a pBAD con delección de extremo C-terminal

Inicio: 98°C, 1 min 30 s; **Amplificación x 30:** 98°C, 10 s; 62°C, 30 s, 72°C, 29 s; **Fin:** 72°, 7 min. Phusion DNA polymerase.

2.2.14.3. Verificación de presencia de inserto

Inicio: 94°C, 5 min; **Amplificación x 25:** 94°C, 30 s; 43°C, 30 s, 68°C, 1 min 14 s; **Fin:** 68°, 5 min. OneTaq Hot Start polymerase.

2.2.14.4. Deleción de extremo C-terminal desde pBAD-TACap

Primera PCR. Inicio: 98°C, 30 s; **Amplificación x 5:** 98°C, 10 s; 61°C, 30 s, 72°C, 2 min 31 s; **Fin:** 72°, 3 min. Q5 Hot Start High Fidelity polymerase.

Segunda PCR. Inicio: 98°C, 30 s; **Amplificación x 30:** 98°C, 10 s; 68°C, 30 s, 72°C, 2 min 31 s; **Fin:** 72°, 3 min. Q5 Hot Start High Fidelity polymerase.

2.2.14.5. Deleción de loop 211-217 de TACap

Primera PCR. Inicio: 98°C, 30 s; **Amplificación x 5:** 98°C, 10 s; 62°C, 30 s, 72°C, 2 min 32 s; **Fin:** 72°, 3 min. Q5 Hot Start High Fidelity polymerase.

Segunda PCR. Inicio: 98°C, 30 s; **Amplificación x 30:** 98°C, 10 s; 71°C, 30 s, 72°C, 2 min 32 s; **Fin:** 72°, 3 min. Q5 Hot Start High Fidelity polymerase.

2.2.14.6. Mutagénesis sitio-dirigida en lote

Inicio: 95°C, 2 min; **Amplificación x 26:** 95°C, 20 s; 55°C, 20 s, 72°C, 2 min 33 s; **Fin:** 72°, 10 min. Pfu Ultra II Hot Start polymerase.

2.2.14.7. Mutagénesis sitio-dirigida individual

Inicio: 98°C, 1 min; **Amplificación x 26:** 98°C, 10 s; 60°C, 10 s, 72°C, 2 min 32 s; **Fin:** 72°, 5 min. Phusion Plus polymerase.

2.2.15. Preparación de células electrocompetentes

Para la producción de células electrocompetentes y su transformación se siguió el protocolo de Bio-Rad ('Bio-Rad electroporation.pdf'), con algunas modificaciones.

- i. Tomamos un inóculo de la cepa correspondiente a partir de su stock conservado a -70°C y realizamos un aislamiento por estrías en placa de LB-agar. Incubar a 37°C *overnight*.

- ii. Inoculamos un tubo con 5 mL de LB con una sola colonia del aislamiento obtenido. Incubamos *overnight* a 37°C con agitación.
- iii. Inoculamos 150 mL de L-broth con 1/100 volúmenes de cultivo fresco crecido de acuerdo con el paso anterior.
- iv. Incubamos a 37°C hasta OD₆₀₀ entre 0.3 y 0.5.
- v. Colocamos el cultivo en hielo durante aproximadamente 20 minutos; de aquí en más mantuvimos las células frías en baño de hielo, al igual que todos los recipientes y soluciones a utilizar. Transferimos el cultivo a un tubo de centrifuga frío y centrifugar a 4000 xg durante 15 minutos a 4°C.
- vi. Descartamos todo el sobrenadante.
- vii. Resuspendimos el *pellet* en 100 mL de glicerol 10%. Centrifugamos a 4000 xg durante 15 minutos a 4°C y descartamos el sobrenadante.
- viii. Resuspendimos el pellet en 50 mL de glicerol 10%. Centrifugamos a 4000 xg durante 15 minutos a 4°C y descartamos el sobrenadante.
- ix. Resuspendimos el pellet en 10 mL de glicerol 10% y pasamos a tubo de centrifuga más pequeño. Centrifugamos a 4000 xg durante 15 minutos a 4°C y descartamos el sobrenadante.
- x. Resuspendimos el pellet en 200 µL de glicerol 10%. Alicuotamos de a 40 µL en tubos de microcentrífuga estériles y usamos en el día o guardamos a -70°C.

2.2.16. Transformación de células electrocompetentes

- i. Agregamos a cada tubo con células competentes hasta 4 µL de solución del ADN que se quiere transformar en la célula. Si las células estaban congeladas primero dejamos descongelar durante algunos minutos en hielo. Incluimos un control positivo con un plásmido de concentración conocida para calcular la eficiencia de transformación.
- ii. Transferimos a una celda de electroporación fría, bajamos las células al fondo asegurándonos que toquen ambas placas metálicas y que no haya presencia de burbujas. Colocamos en el equipo de electroporación.
- iii. Aplicamos pulso de 2.5 V e inmediatamente agregamos 1 mL de SOC.
- iv. Transferimos a tubo de microcentrífuga e incubamos a 37°C durante 1 hora con agitación.
- v. Centrifugamos durante algunos segundos para bajar las células (excepto para el control positivo), descartamos el sobrenadante y resuspendemos en 200 µL de SOC.

- vi. Sembramos en placas de LB-antibiótico 200 μ L de cada muestra y una dilución adecuada del control positivo para obtener placas contables de acuerdo con la concentración de ADN de este. Incubamos 24 horas a 37°C.
- vii. Al día siguiente contamos las colonias obtenidas en el control positivo y calculamos la eficiencia como $\frac{\text{colonias}}{\mu\text{g ADN}}$. Es deseable una eficiencia superior a 1×10^6 .

2.2.17. Preparación y transformación de células quimiocompetentes

Ver guía por Fürst *et al.*¹⁴¹

2.2.18. Guardado de cepas en freezer

- i. Inoculamos un tubo de LB con una colonia aislada de la cepa a conservar.
- ii. Incubamos con agitación *overnight* a 37°C.
- iii. Tomamos 1 mL del cultivo y transferimos a un criotubo.
- iv. Agregamos 0.5 mL de una solución estéril de glicerol 60%. Mezclamos bien y llevamos a freezer -70°C.

2.2.19. Envío de cepas en papel de filtro

- i. Se largó un cultivo en tubo de LB-ampicilina de cada cepa a enviar.
- ii. Se prepararon cuadrados de papel de filtro (2-3 mm espesor) de aprox. 5 mm de lado y fueron desinfectados mediante exposición a luz UV.
- iii. Se cargaron 20 μ L del cultivo *overnight* en los papeles de filtro. Pueden dejarse secar en cabina de flujo o guardarse en tubos de microcentrífuga.
- iv. Para la reconstitución de las cepas se cargaron 20 μ L de medio LB estéril en el papel de filtro, y luego de dejar hidratar por unos minutos se estrió en placas de LB-ampicilina.

2.2.20. Expresión de TACap

La producción de la enzima se realizó a partir de la cepa *E. coli* TOP10 pBAD-TACap aplicando un protocolo clásico de crecimiento de la cepa en medio sólido y luego medio líquido. Para los ensayos subsiguientes la biomasa fue lisada y bien el sobrenadante o la enzima purificada fueron utilizados. Como fuera mencionado anteriormente, las técnicas utilizadas, con pequeñas modificaciones, fueron descritas por Sambrook *et al.*¹³⁸

- i. Estriamos una placa de LB-ampicilina con un inóculo de la cepa guardada freezer.
- ii. Incubamos en estufa 37°C *overnight*.

- iii. Inoculamos un tubo de LB-ampicilina con una colonia aislada de la placa anterior.
- iv. Incubamos con agitación 37°C overnight.
- v. Inoculamos un volumen adecuado de medio TB-ampicilina (20 mL para ensayos en pequeña escala) con 1/10 volumen del inóculo anterior.
- vi. Incubamos con agitación 37°C hasta alcanzar un OD₆₀₀ 0.6-0.8 (aprox. 2 h 30 min).
- vii. Inducimos la expresión de la proteína con L-arabinosa a una concentración del 0.02%.
- viii. Incubamos con agitación 24°C, 24 h.
- ix. Medimos el OD y transferimos el cultivo a un tubo de centrifuga. Centrifugamos a 4000 xg, 15 min.
- x. Descartamos el sobrenadante y congelamos el pellet a -70°C, 1 h.
- xi. Descongelamos y resuspendimos el pellet en buffer de lisis (10 mL/g biomasa).
- xii. Incubamos con agitación suave, 30°C, 30 min.
- xiii. Centrifugamos a máxima velocidad, 4°C, 20 min.
- xiv. Filtramos el sobrenadante por un filtro estéril de 0.45 µm hacia un tubo limpio. Este sobrenadante contiene la enzima lista para ensayar como lisado o para purificar.

2.2.21. Electroforesis de proteínas PAGE-SDS

- i. Se preparó una dilución 1/10 del lisado a analizar con agua destilada y buffer de carga en tubos de microcentrífuga o PCR.
- ii. Se desnaturalizaron las proteínas mediante calentamiento a 99°C, 10 min, tomando la precaución de agujerear la tapa de los tubos o agregando un bloque sobre los mismos para evitar que se abra la tapa.
- iii. Fueron cargados 10 µL de esta solución por pocillo del gel de poliacrilamida.
- iv. El gel se corrió a 80 V durante 20 min y a 120 V durante 50 min más. Alternativamente, a 120 V durante 1 h.
- v. Para el revelado se empleó colorante Coomassie blue siguiendo el método estándar.

2.2.22. Medida de T_m

- i. Colocamos en tubos de real-time PCR 20 µL de lisado o enzima purificada.
- ii. Agregamos 5 µL de SYPRO Orange 50x y mezclamos.

- iii. Cargamos los tubos en el equipo de real-time PCR y corrimos una curva de melting desde 25°C hasta 80°C, aumentando 0.5°C cada 30 s, tomando medidas en este mismo intervalo. Las longitudes de excitación y emisión pueden consultarse en la guía de Bio-Rad adjunta ('Bio-Rad fluorophore reference guide.pdf').
- iv. La fluorescencia de cada muestra varía a lo largo del calentamiento, presentando un gran aumento relativo cuando se desnaturaliza la proteína de interés. La derivada primera de esta curva proporciona el punto de pendiente máxima de este proceso, y define la temperatura de melting aparente de la proteína.

2.2.23. Medida de actividad enzimática

El ensayo se basa en la reacción de transaminación entre la metilbencilamina y el piruvato, catalizada por la enzima. La medida de actividad implica la detección espectrofotométrica de uno de los productos de reacción, la acetofenona, a una $\lambda = 245$ nm. Este ensayo se encuentra descrito y reportado por Schätzle *et al.*¹⁴⁴ y fue utilizado en trabajos anteriores para determinar los parámetros cinéticos de TACap^{87,89}. Esta técnica permite medir la actividad en una placa de 96 pocillos de 12 muestras distintas, por triplicado. Definimos la unidad de enzima como la cantidad de enzima que produce 1 mmol de acetofenona por minuto.

Buffer S: *R*- α -metilbencilamina 3 mM; piruvato 6 mM; PLP 0.1 mM; HEPES 50 mM pH 8; agua destilada c.s.p. 8 mL.

Buffers E: Se prepara un buffer por cada muestra a analizar. Lisado o proteína cantidad necesaria (generalmente dilución 1/100), PLP 0.1 mM; HEPES 50 mM pH 8; agua destilada c.s.p. 0.5 mL.

Buffer PLP: PLP 0.1 mM; HEPES 50 mM pH 8; agua destilada c.s.p. 4.8 mL.

- i. Teniendo la precaución de utilizar pipeteo inverso en todos los pasos a continuación para evitar la formación de burbujas, se llenó la microplaca de 96 pocillos como se indica en la Figura 42.
- ii. En el lector de placas fijamos la temperatura en 31°C y termostatizamos la placa durante 3 minutos.
- iii. Transcurrido este tiempo extrajimos la placa y transferimos con pipeta multicanal 100 μ L desde la fila H a las filas B, D y F.
- iv. Se transfirieron 100 μ L desde la fila G a las filas A, C y E.
- v. Inmediatamente se introdujo la placa en el lector y se comenzaron las medidas de

cinética en los pocillos A1:F12.

- vi. Luego de la toma de datos se procesaron los mismos para obtener la pendiente de cada pocillo.

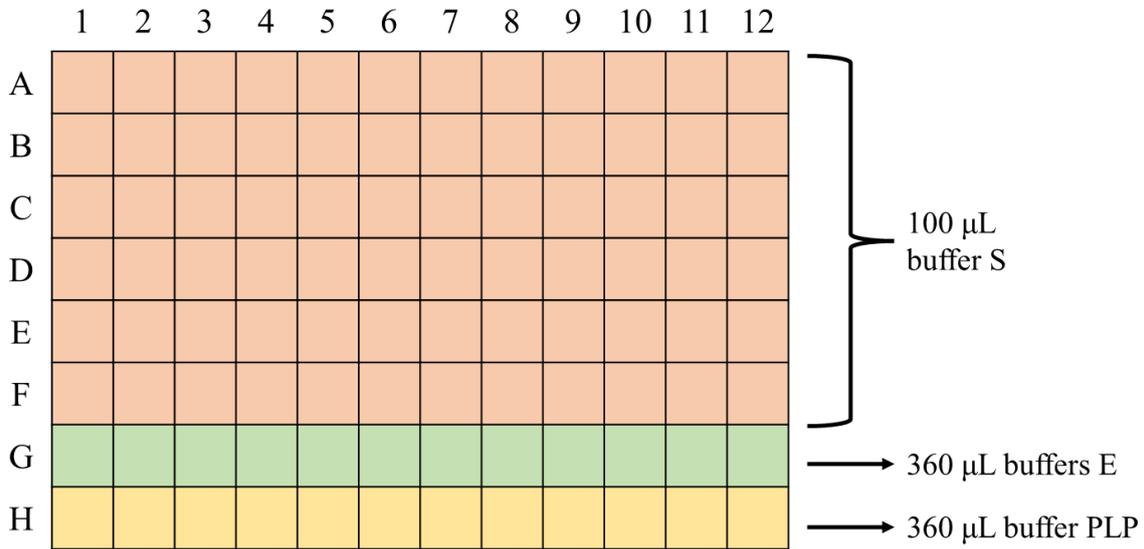


Figura 42 Distribución de buffers en microplaca de 96 pocillos para ensayo de actividad.

2.2.24. Cinética de inactivación térmica

En este ensayo se incubaba la enzima a una temperatura definida y se toman muestras a determinados tiempos. La actividad de estas muestras se mide y, siendo la actividad de la enzima sin incubación el 100%, se grafica el porcentaje de actividad remanente vs. tiempo.

2.3. Resultados y discusión

2.3.1. Clonado

Como fuese indicado en los antecedentes nuestra enzima de interés se encontraba clonada en el vector pKK223-3 ('pKK223-3 TACap.geneious'). Para facilitar la generación, expresión y análisis en lote de mutantes se decidió emplear el sistema pBAD expresado en *E. coli* TOP10. Este sistema permite realizar tanto clonado como expresión en la misma cepa, y la inducción puede ser finamente regulada de acuerdo con la cantidad adicionada de arabinosa (inductor), ya que TOP10 es una cepa incapaz de degradarla (*araABCD*⁻). Por lo anterior se procedió al subclonado de TACap desde pKK223-3 hacia pBAD-myc-HisA. Se aisló el ADN de la cepa original (*E. coli* TOP10 pKK223-3 TACap) mediante *miniprep*, y se realizó la técnica de *Restriction Free Cloning* con *primers* diseñados para el subclonado de la enzima nativa (2.2.13.1), así como para el subclonado y el corte del extremo C-terminal en simultáneo (2.2.14.2). En la primera PCR de estas reacciones (obtención de megaprimer) solo se logró obtener producto con los primeros (Figura 43b). Cumplidos los pasos requeridos para obtener el producto final y transformar células competentes, luego de varios intentos, se logró obtener transformantes, sin que se observara ninguna colonia en el control de digestión con DpnI. Sin embargo, al verificar el plásmido obtenido por PCR (Figura 44) no se detectó ningún inserto. Estas muestras fueron también secuenciadas, verificando así una diferencia de tres pares de bases entre la secuencia esperada del vector pBAD-myc-HisA y aquella secuenciada, en el extremo 3' del primer atnei32-fw, lo que puede haber imposibilitado la segunda PCR del *Restriction Free Cloning* (Figura 45).

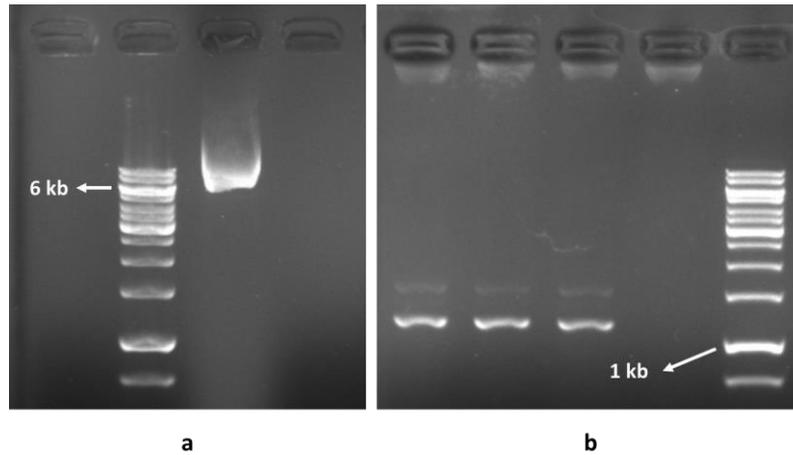


Figura 43 Electroforesis en gel de agarosa. a- marcador de PM (2); miniprep del vector pKK223-3 TACap (3). b- amplificación de TACap con primers 2.2.13.1 (1, 2, 3); marcador de PM (5).

Finalmente, optamos por emplear una estrategia alternativa para el clonado: Golden Gate cloning. El gen de TACap fue encargado a GenScript en dos segmentos, ya que es más económico que encargarlo entero. Se incorporó al gen una mutación silenciosa sustituyendo una G por A en la posición 1005 para eliminar un sitio de restricción para BsaI, con el objetivo de hacer posible el uso de esta enzima en el clonado. El gen comercial fue sometido al protocolo de Golden Gate cloning (2.2.10) junto con el vector pBAD-C-term-His. Con esta estrategia no se presentaron inconvenientes y obtuvimos el plásmido deseado pBAD-TACap ('pBAD-TACap.geneious').

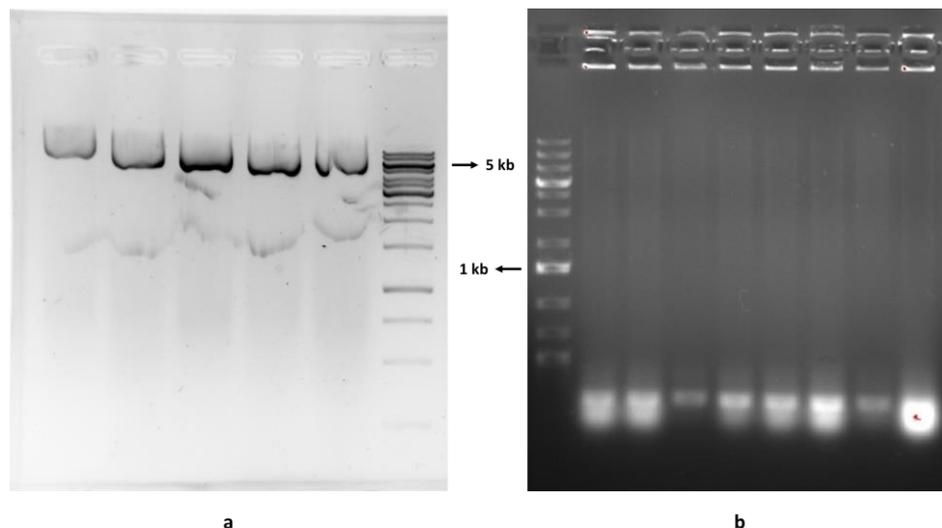


Figura 44 a- Electroforesis del ADN extraído de algunas de las colonias transformadas con presunto pBAD-TACap; b- PCR sobre dicho ADN con primers linderos a la zona de inserción.

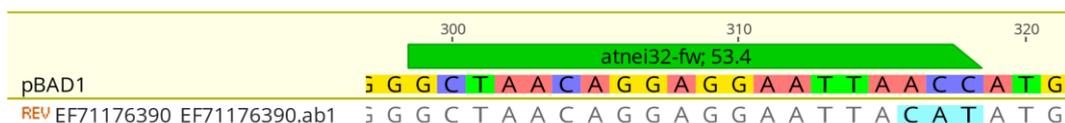


Figura 45 Secuencia esperada del vector pBAD-myc-HisA (arriba) y aquella obtenida por secuenciado (abajo).

2.3.2. Construcción de variantes

2.3.2.1. Mutantes puntuales

Los primers para la mutagénesis sitio-dirigida fueron generados mediante el programa AAScan, como fuese indicado en la sección 2.2.11. Fueron encargados en placa de 96 pocillos y tanto las PCR como su digestión y posterior transformación en *E. coli* fueron realizados en formato placa. En la Tabla 12 se indican las mutaciones propuestas (44) y las que pudieron ser obtenidas en el laboratorio (29). De las mutaciones de reporte, dos fueron diseñadas por error para posiciones/aminoácidos que no correspondían con los planeados. Por otro lado, varias mutaciones propuestas no pudieron ser obtenidas experimentalmente. Esto puede deberse a un pobre diseño de los primers dada la generación automática de los mismos, lo que podría revisarse a futuro para obtener el resto de los mutantes. En esta etapa se decidió continuar con los ensayos experimentales sobre las 29 variantes obtenidas.

Tabla 12 Resumen de los mutantes puntuales propuestos y obtenidos. En **negrita** aquellos a partir de reportes previos; en romana las mutaciones consenso; con fondo verde las mutaciones que pudieron ser obtenidas en el laboratorio; con fondo blanco las que no fueron obtenidas; con letra en **rojo** las que fueron mal diseñadas.

reportes	consenso		obtenido	diseño incorrecto	
M48V	Q51W	L58A	P63V	S64H	L80F
M87L	T101I	M105C	L106V	I111L	F115Y
L118M	L118T	V120C	R125S	L130S	S131R
A132D	A133P	V135R	L142F	M144A	W145F
I146A	M153A	T174I	C178S	M179V	L186Y
Q187H	M195L	Y205T	A220G	D224K	S226P
V231F	V249I	S253A	S253T	C257L	C284T
I293V	I313L				

2.3.2.2. Deleciones

Intentamos emplear los primers atnei33a-fw y rv (2.2.13.3) para la eliminación del extremo C-terminal, sin éxito. Estos primers estaban dirigidos a zonas completamente complementarias de la secuencia de ADN, ya que fueron diseñados inicialmente según las indicaciones del manual QuikChange¹⁴⁵. Dado que esta aproximación no dio resultado, se decidió diseñar primers que fueran parcialmente complementarios en sus extremos 3', de acuerdo con lo recomendado por algunos autores.^{139,141} Finalmente, con los primers parcialmente complementarios dCtail-Fw y Rv (2.2.13.3) logró obtenerse el constructo de interés, lo que fue verificado por secuenciación. Se siguió esta misma estrategia para la deleción del loop 211-217 (primers 2.2.13.4), exitosamente.

2.3.2.3. Mutantes múltiples

De acuerdo a los resultados obtenidos con los mutantes simples (ver 2.3.4) se decidió construir nuevas variantes que combinaran las mejores mutaciones. Dada las posiciones de las mutaciones a combinar (T101I, L106V y Δ 211-217) fue necesario encargar un primer adicional para la generación del mutante T101I+L106V (primer 2.2.13.6). Mediante el protocolo habitual todos los mutantes fueron obtenidos exitosamente. A continuación se indica la estrategia seguida en cada caso.

Tabla 13 Estrategia seguida para la generación de los mutantes múltiples. Se indica mutante a obtener, molde empleado y primers correspondientes.

	Mutante	Molde	primer fw	primer rv
M1	T101I + Δ 211-217	Δ 211-217	T101I_fw	T101I_rv
M2	L106V + Δ 211-217	Δ 211-217	L106V_fw	L106V_rv
M3	T101I + L106V	T101I	L106V_fw	T101I+L106V_rev
M4	T101I + L106V + Δ 211-217	M1	L106V_fw	T101I+L106V_rev

2.3.3. Expresión de proteínas

Para el análisis de la estabilidad y actividad de las variantes enzimáticas de TACap procedimos a expresarlas, para lo cual seguimos el protocolo indicado en la sección 2.2.20. En primera instancia realizamos esto para TACap sin modificar y evaluamos su expresión en estas condiciones, así como la efectividad del método de lisado, utilizando electroforesis de proteínas en condiciones desnaturalizantes (SDS-PAGE). La enzima, con un peso molecular por

monómero de ~ 40 kDa, se expresa correctamente en forma soluble como se observa en la Figura 46. Si bien realizamos un control de lisis mediante sonicado (asumiendo por este método una lisis del 100%), este falló, dando un nivel de lisis inferior a aquel con lisozima. De todas formas, observando solo la expresión de la muestra lisada con lisozima podemos concluir que las condiciones de expresión son satisfactorias para el trabajo a desarrollar. Vale recordar en este momento que los experimentos se llevaron a cabo en placas de 96 pocillos, haciendo inadecuada la lisis por sonicado.

Luego, evaluamos también la expresión de $\Delta 211-217$ y ΔC -tail. Como se aprecia en la Figura 47 poseen una expresión similar a aquella de la enzima nativa. En base a este comportamiento decidimos emplear estas condiciones de expresión y lisado con las subsiguientes variantes a analizar.

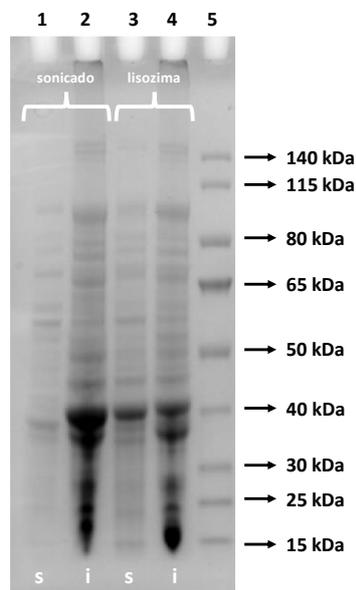


Figura 46 Electroforesis de proteínas en gel de poliacrilamida en condiciones desnaturalizantes. Proteína TACap. Carriles 1-2: fracciones soluble e insoluble luego del lisado por sonicación. Carriles 3-4: fracciones soluble e insoluble luego del lisado por lisozima. Carril 5: marcador de peso molecular.

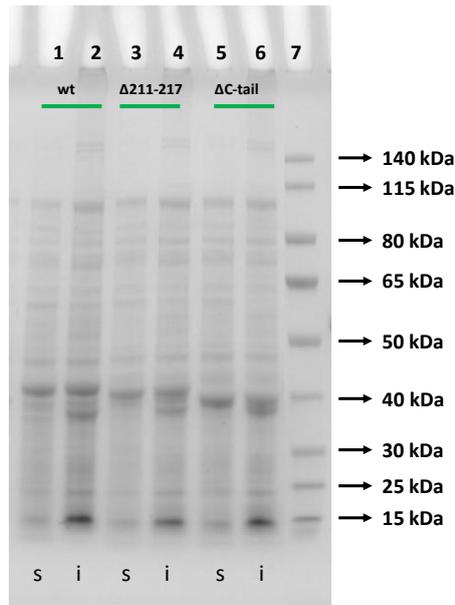


Figura 47 Electroforesis de proteínas en gel de poliacrilamida en condiciones desnaturalizantes. Carriles 1-2, 3-4 y 5-6: TACap, $\Delta 211-217$ y ΔC -tail, respectivamente, fracciones soluble (s) e insoluble (i). Carril 7: marcador de peso molecular.

2.3.4. Determinación de T_m

La determinación de la temperatura de *melting* aparente de los mutantes evaluados fue realizada sobre el extracto libre de células. Si bien pueden existir diferencias considerables entre la T_m determinada sobre este extracto y sobre la enzima purificada, la diferencia de T_m respecto a la enzima nativa es un parámetro que suele conservarse.⁷² A modo de ejemplo, en la Figura 48 se presentan los gráficos de la curva de *melting* a partir de los cuales se calculó el T_m para las distintas enzimas. Para uno de los mutantes consenso, L58A, no pudo obtenerse una curva de *melting*, probablemente debido a una pobre expresión del mutante.

Como podemos ver en la Tabla 14, de los 27 mutantes consenso obtenidos 5 resultaron tener una T_m superior a aquella de la enzima nativa. Esto refleja un índice de éxito de casi 20%, superando el <1% de la evolución dirigida¹³⁵ y siendo comparable al logrado con el protocolo FRESCO⁷¹. Además, concuerda con los hallazgos reportados por Jones *et al.*¹⁴⁶ en cuanto al notable porcentaje de éxito logrado en relación a la eficiencia del diseño y al esfuerzo experimental requerido.

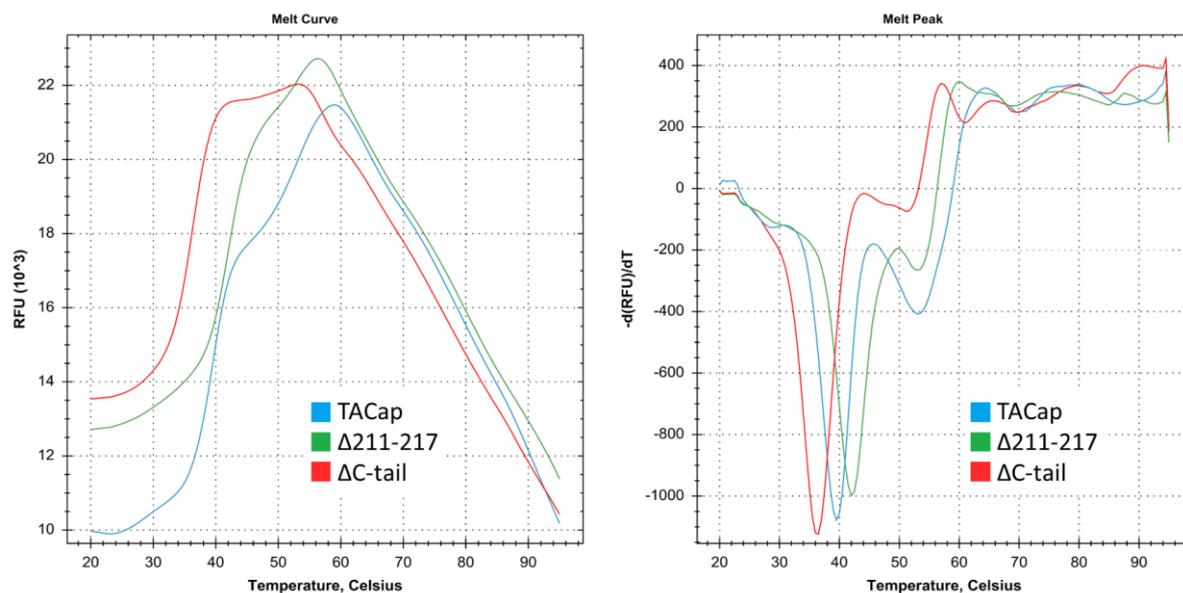


Figura 48 Curva de melting para análisis de T_m . Fluorescencia vs. T (izq.) y $-dF/T$ vs. T (der.) TACap nativa en azul, $\Delta 211-217$ en verde y ΔC -tail en rojo.

Tabla 14 Temperatura de melting aparente de todos los mutantes evaluados y ΔT_m respecto a la enzima nativa (wt).

Mut.	T_m (°C)	ΔT_m (°C)	Mut.	T_m (°C)	ΔT_m (°C)
P63V	44.5	4.5	M179V	38.5	-1.5
S64H	42.5	2.5	I146A	38.5	-1.5
W145F	42.0	2.0	C284T	38.3	-1.8
$\Delta 211-217$	42.0	2.0	S131R	38.0	-2.0
L106V	40.8	0.8	I293V	38.0	-2.0
T101I	40.5	0.5	R125S	37.8	-2.3
wt	40.0	0.0	L80F	37.8	-2.3
F115Y	39.8	-0.3	A220G	37.8	-2.3
M87L	39.5	-0.5	L118M	37.0	-3.0
Q51W	39.0	-1.0	L130S	37.0	-3.0
M48V	39.0	-1.0	V135R	36.5	-3.5
Q187H	39.0	-1.0	ΔC -tail	36.0	-4.0
C178S	39.0	-1.0	S253A	35.5	-4.5
L186Y	38.8	-1.3	C257L	31.0	-9.0
T174I	38.5	-1.5	Y205T	30.0	-10.0
M195L	38.5	-1.5			

n/a- no aplica

Si contemplamos los valores de ΔT_m obtenidos para las mutaciones consenso, los cruzamos con los porcentajes de conservación calculados (Tabla 15) y los graficamos (Figura

49) no podemos observar ningún tipo de correlación entre los mismos. Sullivan *et al.*¹⁴⁷ demostraron que si bien un 50% de las posiciones más conservadas suelen ser estabilizantes, la existencia de interacciones entre la posición mutada y otros aminoácidos de su entorno pueden hacer que se produzca una desestabilización. Sería necesario realizar un estudio de correlación estadística de las posiciones en el MSA de TACap para detectar posibles vinculaciones entre los aminoácidos que permitan explicar este resultado.

Tabla 15 ΔT_m de los mutantes consenso evaluados respecto a la enzima nativa (wt) y porcentaje de conservación del aminoácido consenso en esa posición.

Mut.	ΔT_m (°C)	conservación (%)	Mut.	ΔT_m (°C)	conservación (%)
P63V	4.5	61	M195L	-1.5	69
S64H	2.5	42	M179V	-1.5	60
W145F	2.0	61	I146A	-1.5	63
L106V	0.7	90	C284T	-1.8	46
T101I	0.5	66	S131R	-2.0	77
wt	0.0	n/a	I293V	-2.0	54
F115Y	-0.3	80	R125S	-2.3	31
M87L	-0.5	63	L80F	-2.3	78
Q51W	-1.0	55	A220G	-2.3	76
M48V	-1.0	39	L118M	-3.0	62
Q187H	-1.0	68	L130S	-3.0	58
C178S	-1.0	72	V135R	-3.5	59
L186Y	-1.3	56	C257L	-9.0	54
T174I	-1.5	75	Y205T	-10.0	52

n/a- no aplica

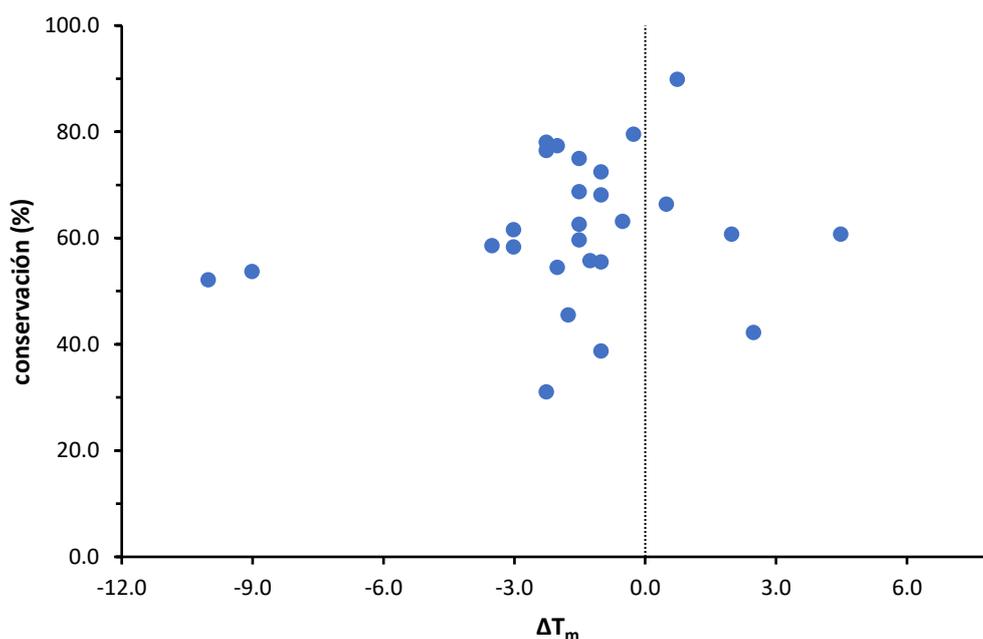


Figura 49 Gráfico de porcentaje de conservación del aminoácido consenso vs. ΔT_m .

Por otra parte, si corroboramos los valores de ΔT_m de los mutantes puntuales evaluados junto a los valores de $\Delta\Delta G^{\text{fold}}$ obtenidos mediante FoldX y Rosetta-ddG sobre las estructuras apo y holo (Tabla 16) y los graficamos (Figura 50), en principio tampoco vemos una correlación clara entre los datos calculados y experimentales. Esto puede atribuirse, por un lado, a que la estructura evaluada no fue determinada experimentalmente; por otro, a las limitaciones inherentes de estos algoritmos, los cuales si bien pueden resultar exitosos cuando se evalúan cientos de mutantes, aún presentan un amplio margen de mejora para ser considerados herramientas efectivas en escalas más reducidas.^{148,149}

A pesar de lo mencionado anteriormente, sí podemos ver cierta correlación entre el ΔT_m y el $\Delta\Delta G^{\text{fold}}$ para las mutaciones desestabilizantes, lo que parece indicar que estos algoritmos presentan un mejor comportamiento al predecir este tipo de mutaciones, lo cual ya ha sido reportado^{73,150}.

La deleción del extremo C-terminal (residuos 333-356), altamente desordenada, al contrario de lo esperado redujo la T_m de la enzima. Nos propusimos otro corte, $\Delta C\text{-tail2}$ (residuos 339-356), pero tampoco tuvimos éxito, con un T_m de 39.0°C (ΔT_m -1.0°C). Si bien estos extremos se observaban móviles en las simulaciones de dinámica molecular, quedan dudas con respecto a la estructura y la función de las mismas, ya que son secuencias poco conservadas. Puede resultar interesante a futuro estudiar más en profundidad estos aspectos.

Tabla 16 ΔT_m de los mutantes puntuales evaluados respecto a la enzima nativa (wt) y $\Delta\Delta G^{\text{fold}}$ obtenido mediante FoldX (F) y Rosetta-ddG (R) sobre las estructuras apo y holo.

Mut.	ΔT_m (°C)	F-apo (kJ/mol)	F-holo (kJ/mol)	R-apo (kJ/mol)	R-holo (kJ/mol)
P63V	4.5	33.4	39.4	-59.6	-66.0
S64H	2.5	-3.8	47.7	-13.8	6.3
W145F	2.0	17.8	14.8	33.8	36.3
L106V	0.7	33.4	36.4	24.5	28.3
T101I	0.5	-16.6	-13.9	-42.7	-44.3
wt	0.0	n/a	n/a	n/a	n/a
F115Y	-0.3	-5.8	-9.2	1.5	-1.0
M87L	-0.5	-1.5	0.1	-12.7	-12.4
Q51W	-1.0	16.9	10.9	0.4	-3.9
M48V	-1.0	20.6	23.3	10.3	15.0
Q187H	-1.0	35.2	35.6	26.8	17.8
C178S	-1.0	-4.0	-2.7	51.5	50.1
L186Y	-1.3	13.6	18.8	40.5	38.3
T174I	-1.5	-19.2	-15.7	-20.9	-25.8
M195L	-1.5	1.0	0.6	-6.2	-7.8
M179V	-1.5	7.6	12.9	6.6	1.5
I146A	-1.5	33.2	33.4	79.8	82.8
C284T	-1.8	10.2	6.2	12.2	6.8
S131R	-2.0	3.3	-0.3	1.1	0.2
I293V	-2.0	7.4	6.6	6.8	5.8
R125S	-2.3	9.8	8.7	17.1	15.4
L80F	-2.3	11.9	17.3	13.4	21.1
A220G	-2.3	10.3	9.6	34.1	30.1
L118M	-3.0	-0.2	0.0	4.7	11.4
L130S	-3.0	26.7	24.1	64.3	67.8
V135R	-3.5	23.4	54.1	59.4	56.9
S253A	-4.5	-1.8	-1.7	-24.8	-25.2
C257L	-9.0	5.7	0.9	-18.8	-17.9
Y205T	-10.0	30.0	32.3	43.0	46.3

n/a- no aplica

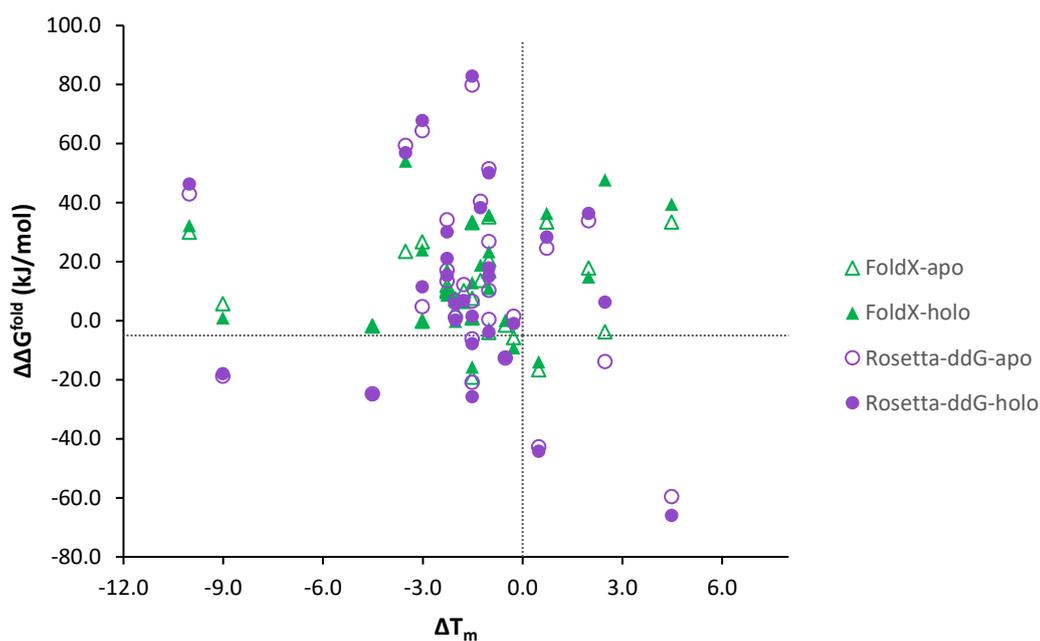


Figura 50 Gráfico de $\Delta\Delta G^{\text{fold}}$ vs. ΔT_m . La línea horizontal cruza en -5 kJ/mol, la cual es la mejora en $\Delta\Delta G^{\text{fold}}$ sugerida por el protocolo FRESCO para seleccionar mutaciones estabilizantes.

2.3.5. Análisis de algunas mutaciones estabilizantes

Mediante inspección visual intentamos dar una explicación racional a algunos de los resultados obtenidos con las mutaciones puntuales. Para la mutación P63V (Figura 51) la sustitución de prolina por valina elimina la tensión resultante del anillo pirrolidina y permite al nitrógeno de la cadena peptídica establecer un enlace de hidrógeno con Val-116 en la hoja β vecina. Este es un tipo de sustitución que ya habíamos mencionado como potencialmente estabilizante en la sección 1.3.6.

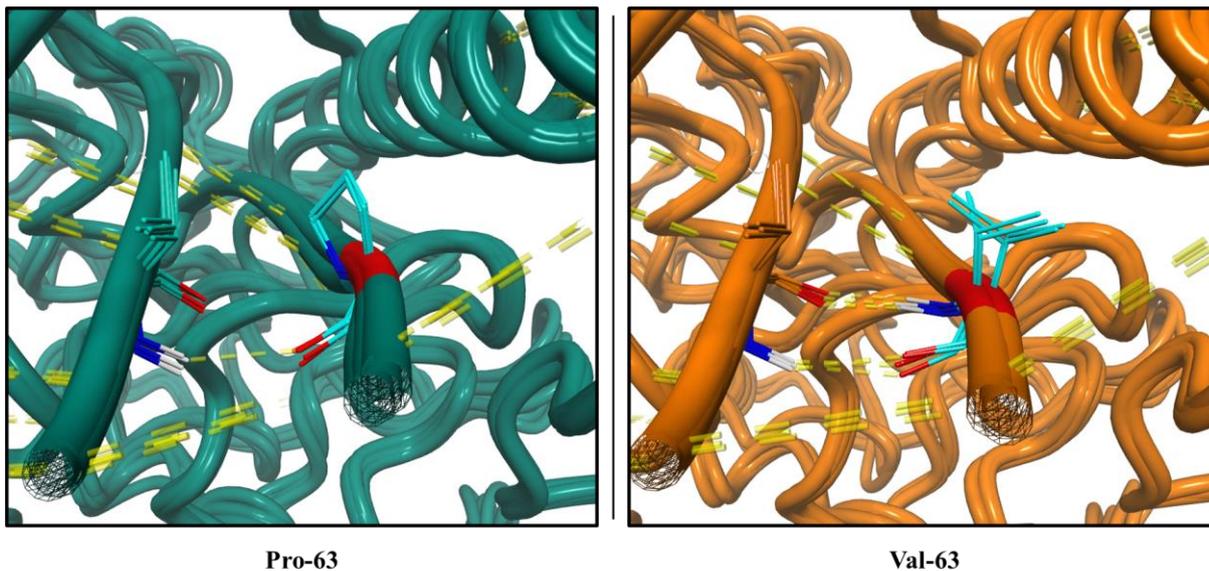


Figura 51 Comparación entre estructura nativa (izquierda) y mutante P63V (derecha).

En el caso de la mutación S64H (Figura 52) la serina nativa establece enlaces de hidrógeno con Tyr-149 e His-321, mientras que la histidina resultante pierde estas interacciones. Dado que no parecen haber otros cambios, en principio no es posible definir el motivo por el cual el mutante es más estable que la enzima nativa.

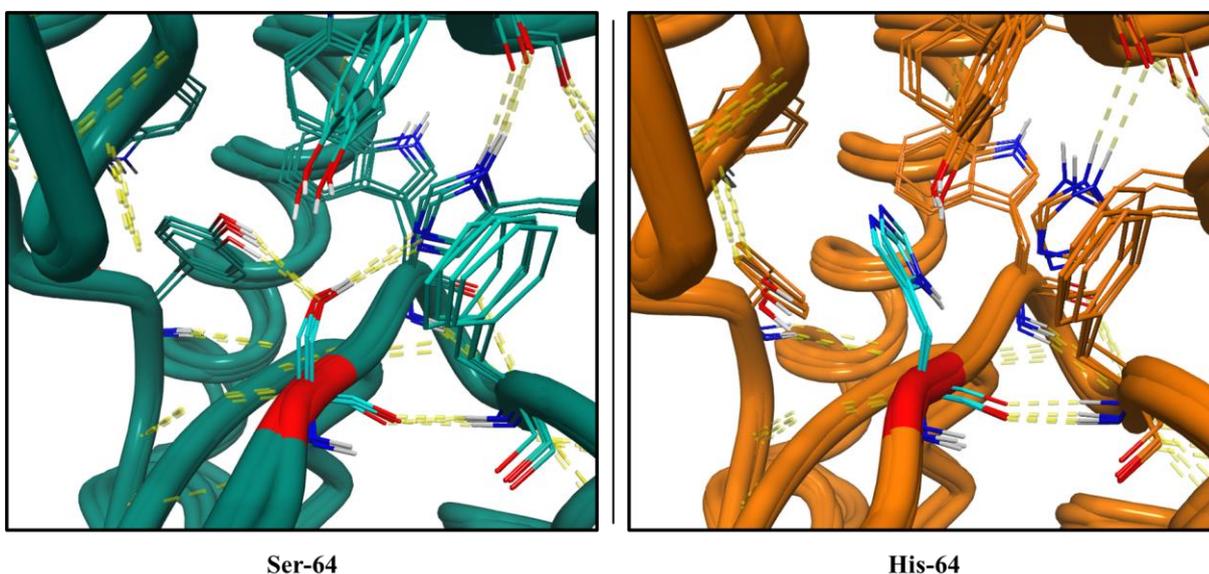


Figura 52 Comparación entre estructura nativa (izquierda) y mutante S64H (derecha).

En la enzima nativa la Thr-101 es capaz de formar un enlace de hidrógeno con el anterior residuo de la hélice alfa (Q97) tanto a través del oxígeno lateral como del nitrógeno de la cadena

peptídica, lo cual puede alterar la conformación de la hélice¹⁵¹ (Figura 53). Cuando la treonina es sustituida por isoleucina, la interacción por enlace de hidrógeno entre el residuo 101 y el anterior en la hélice (Q97) pasa a ser exclusivamente a través del nitrógeno de la cadena peptídica, recuperando la hélice su conformación normal. Vale aclarar que la conformación alterada de la hélice no implica per se una desestabilización de la proteína, y es más una observación de las diferencias entre la estructura nativa y la de T101I.

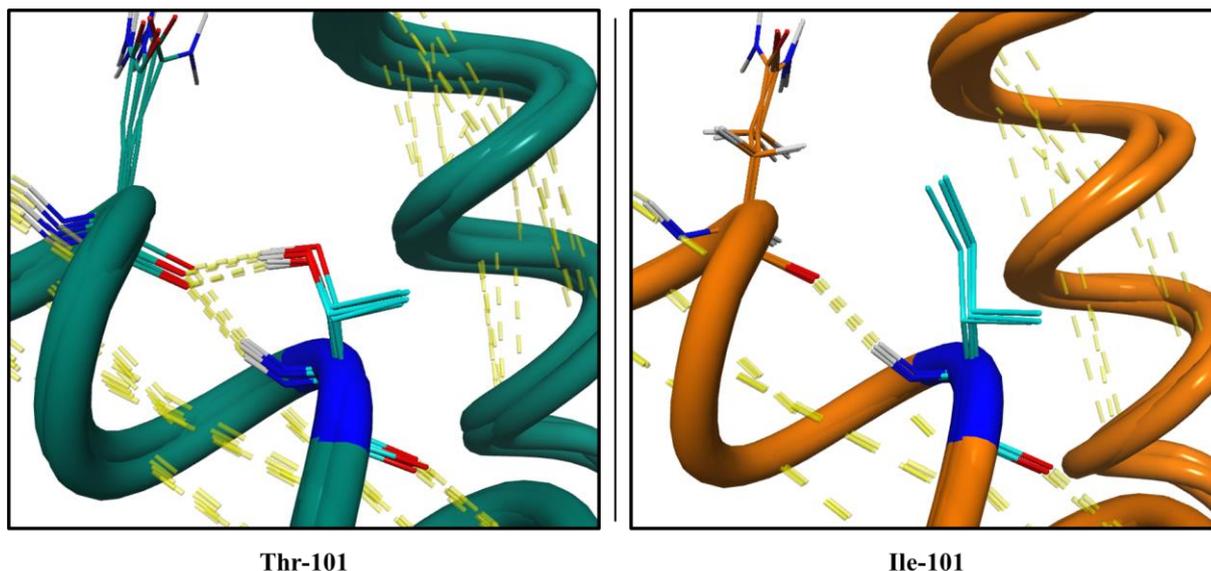


Figura 53 Comparación entre estructura nativa (izquierda) y mutante T101I (derecha).

El cambio de Leu-106 por valina (Figura 54) disminuye el largo de la cadena lateral hidrofóbica en un carbono, lo cual no debiera tener mayores consecuencias en el entorno relativamente hidrofóbico en que se encuentra este residuo. Es cierto también, sin embargo, que este residuo no se encuentra demasiado lejano a la superficie, por lo que quizás una cadena lateral apolar más pequeña resulta beneficiosa para la estabilidad.

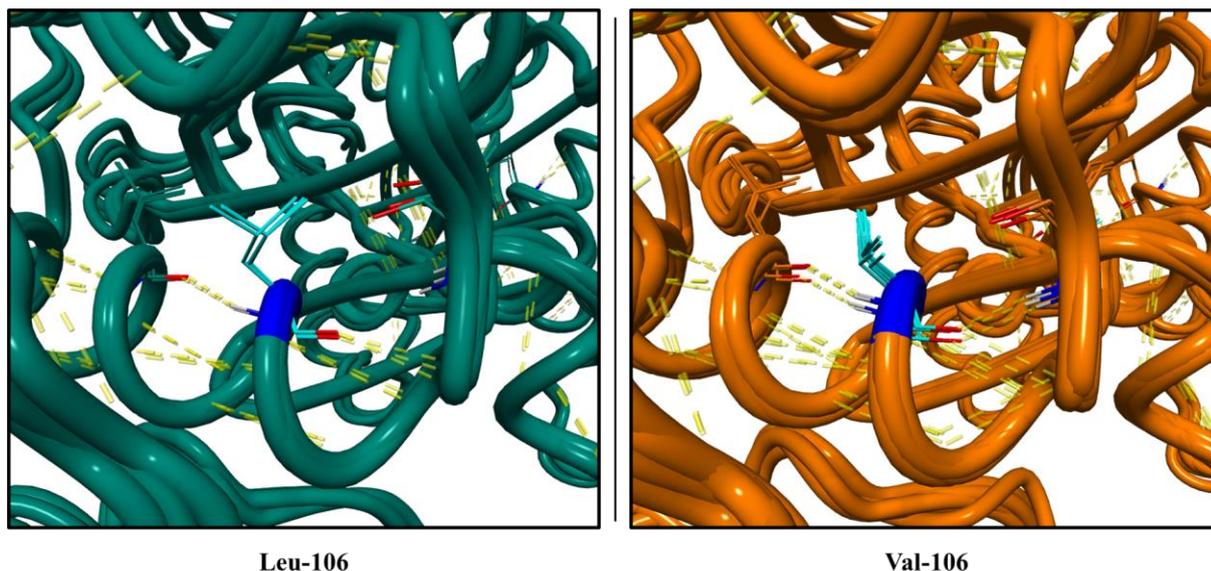


Figura 54 Comparación entre estructura nativa (izquierda) y mutante L106V (derecha).

2.3.6. Ensayos de actividad y estabilidad

Una vez determinados los ΔT_m de las variantes obtenidas procedimos al análisis de la actividad de aquellos que demostraron una mejora de estabilidad respecto a la enzima nativa. Para la medida de la actividad enzimática empleamos el ensayo desarrollado por Schätzle *et al.*¹⁴⁴ Previo a esto fue necesario realizar una curva de calibración para la acetofenona simulando las condiciones de reacción, es decir, variando la concentración de acetofenona desde 0 mM hasta 0.5 mM, y las correspondientes de piruvato, D-alanina y (*R*)-metilbencilamina (Figura 55). También hicimos un escaneo del espectro de los componentes de la reacción para verificar posibles interferencias a la longitud de onda de medida, las cuales no fueron detectadas (Figura 56).

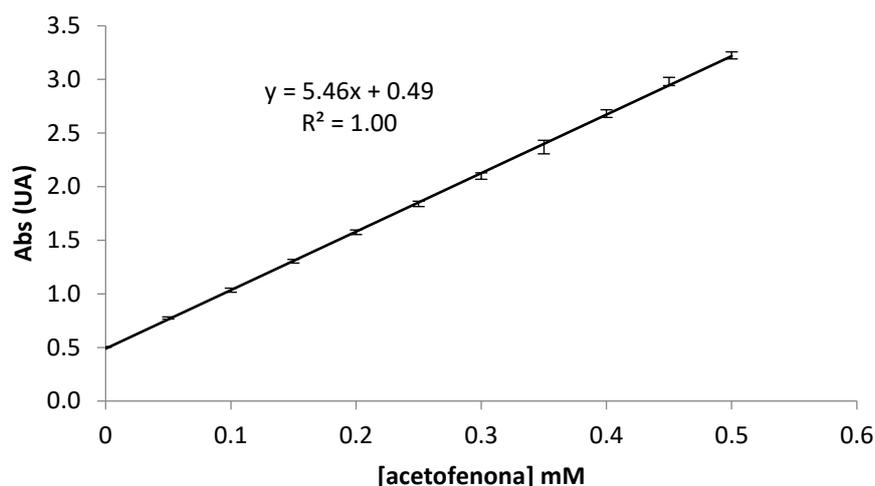


Figura 55 Curva de calibración de acetofenona.

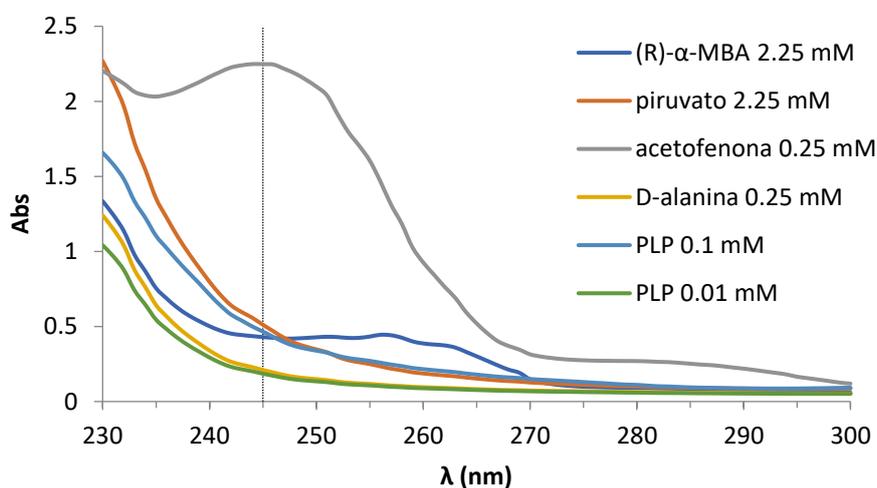


Figura 56 Espectro (230-300 nm) de los componentes de reacción en el análisis de actividad.

Previo al análisis de los mutantes evaluamos algunas características de la enzima nativa. Dado un reporte reciente sobre la necesidad de tener PLP en el medio para mantener la estabilidad de transaminasas¹⁵², quisimos evaluar el comportamiento de nuestra enzima. Por tanto, la incubamos a 40°C durante 45 min en presencia de diferentes concentraciones de PLP y luego evaluamos la actividad remanente respecto a la enzima sin incubación. Los resultados se observan en la Figura 57 y son sumamente claros en cuanto reflejan una notable influencia de la concentración de PLP en la estabilidad de la enzima.

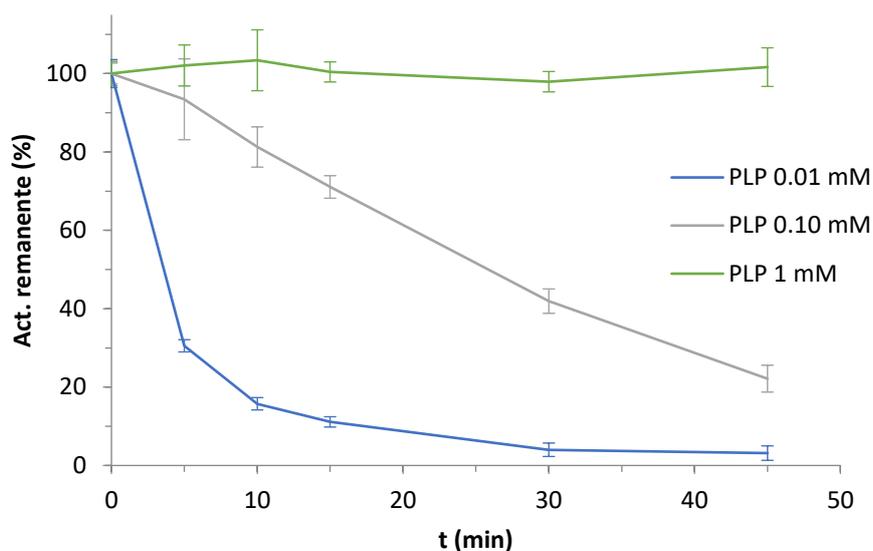


Figura 57 Gráfico de actividad remanente de TACap vs. t al ser incubada a distintas concentraciones de PLP durante 45 min a 40°C.

A continuación, evaluamos en las mejores condiciones del ensayo anterior ($[PLP] = 1$ mM) la estabilidad de TACap frente a la incubación a distintas temperaturas. Como podemos corroborar en la Figura 58, midiendo hasta 45 min solo a 40°C mantiene un 100% de actividad, mientras que a 50 y 60°C la actividad se pierde casi o totalmente apenas transcurridos 5 minutos. Esto es concordante con la T_m obtenida para la enzima nativa.

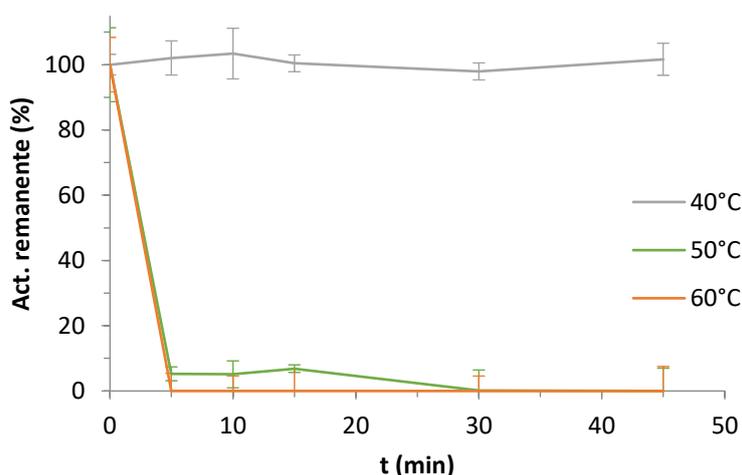


Figura 58 Gráfico de actividad remanente de TACap vs. t al ser incubada a distintas temperaturas.

Pasando al análisis de mutantes, realizamos la determinación de actividad sobre el lisado enzimático, asumiendo un nivel de expresión similar para todas las enzimas. La actividad de la

enzima nativa fue considerada como un 100%, mientras que para los mutantes se calculó la actividad relativa a esta. En la Tabla 17 podemos ver los resultados de las enzimas ensayadas, donde se consideró una actividad relativa superior al 50% como aceptable para pasar a la etapa de combinación de mutantes.

Tabla 17 Actividad relativa de las enzimas con un ΔT_m positivo respecto a la enzima nativa.

Mutante	ΔT_m (°C)	act. relativa (%)
P63V	4.5	35
S64H	2.5	13
T101I	0.5	73
L106V	0.7	91
W145F	2.0	59
$\Delta 211-217$	2.0	139

Si mapeamos las mutaciones con un ΔT_m positivo en la estructura tridimensional de TACap (Figura 59) podemos observar que P63 y S64 se encuentran en la entrada al sitio activo. Esto, sumado a su baja actividad relativa nos llevó a descartar las mutaciones en estos residuos. En la misma situación se encuentra W145, que, si bien tuvo una actividad relativa superior al 50%, es inferior al de los restantes mutantes. Para la combinación de mutaciones se decidió tomar hasta tres candidatos con el fin de mantener un número bajo de variantes para su screening, por lo que estos resultaron ser: T101I, L106V y $\Delta 211-217$.

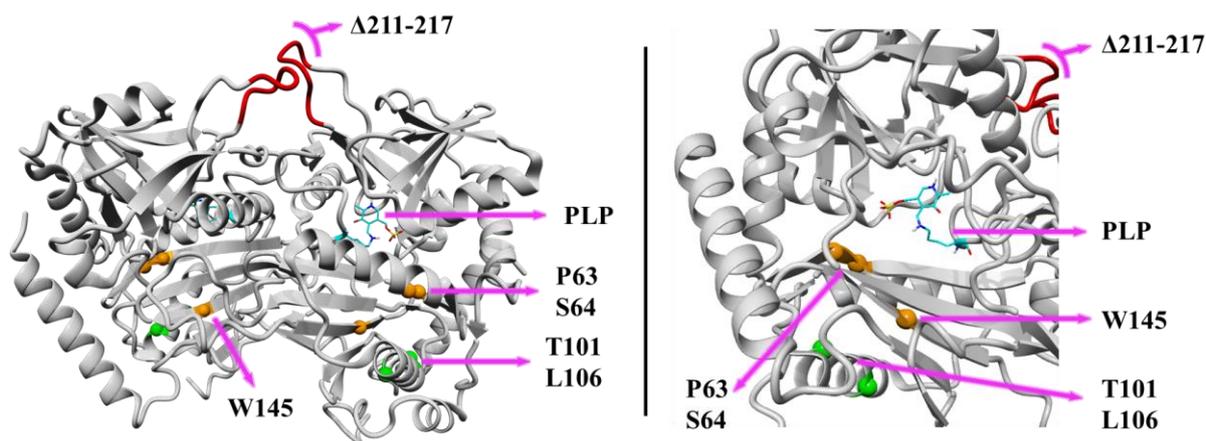


Figura 59 Mutaciones con ΔT_m positivo mapeadas en la estructura de TACap. Vista frontal (izq.) e inferior (der.)

Los cuatro mutantes resultantes fueron: M1 (T101I + $\Delta 211-217$) M2 (L106V + $\Delta 211-217$), M3 (T101I + L106V) y M4 (T101I + L106V + $\Delta 211-217$). Estas variantes fueron

construidas exitosamente y expresadas para evaluar su estabilidad y actividad, resultando los ΔT_m y actividades relativas que se observan en la Tabla 18. Los valores obtenidos parecen indicar que la combinación de un mutante puntual con la delección tiene un efecto sinérgico, mejorando tanto estabilidad como actividad respecto a la enzima nativa, mientras que la combinación de los mutantes puntuales tiene un efecto perjudicial. Se observa que los mutantes M1 y M2 presentan un aumento en su T_m de 4.7 y 6.0 °C respectivamente, mostrando ambos un aumento en la actividad relativa. En el caso de M2, la actividad aumenta a más del doble de lo observado para la WT. Este resultado resulta muy prometedor para su futura aplicación a la síntesis de radiofármacos.

Tabla 18 T_m , ΔT_m y actividad relativa a la enzima nativa de los mutantes combinados.

Mut.	T_m (°C)	ΔT_m	act. relativa (%)
M2	46.1	6.0	213
M1	44.8	4.7	123
M3	43.4	3.4	39
M4	43.4	3.4	88

M1: T101I + Δ 211-217; M2: L106V + Δ 211-217; M3: T101I + L106V; M4: T101I + L106V + Δ 211-217

2.4. Conclusiones

Se intentó subclonar la enzima TACap desde el vector pKK22-3 hacia el vector pBAD mediante *Restriction free cloning*, sin embargo, probablemente debido a diferencias entre la secuencia esperada del vector de destino y la secuencia real, determinada por secuenciado más tarde, esto no fue posible. Alternativamente, a partir del gen sintético de TACap aplicamos la técnica *Golden gate cloning*, con lo cual logramos el clonado exitosamente.

Fueron desarrollados los mutantes puntuales propuestos en el Capítulo I mediante mutagénesis QuikChange, incorporando las recomendaciones de algunos autores respecto al desfase de primers. Se lograron obtener 29 de los 44 mutantes seleccionados, lo cual marca un índice de éxito del 66%. Los mutantes que involucraban una delección fueron obtenidos sin inconvenientes empleando una PCR en dos etapas, la primera de ellas con los primers forward y reverse por separado, y la segunda con ambos. Lamentablemente, por errores en el diseño de primers o por no haber logrado amplificación, no fue posible obtener los mutantes seleccionados en base a reportes previos. Dada la similitud de nuestra enzima con 4CE5 sería conveniente en el futuro analizar la influencia de estas mutaciones en la estabilidad.

Fue evaluada la expresión de la enzima nativa y dos mutantes en condiciones habituales de trabajo, obteniendo niveles de expresión y solubilidad adecuados, por lo que se decidió mantener estas condiciones para todas las variantes.

La temperatura de *melting* de todas las variantes desarrolladas fue determinada mediante el ensayo de desplazamiento térmico, un ensayo de alta eficiencia que facilita enormemente la selección de mutaciones estabilizantes. La generación de esta capacidad en nuestro grupo es un saldo positivo de esta tesis, ya que podrá aplicarse a futuro para el mejoramiento de otras enzimas de interés.

De las 30 variantes ensayadas seis resultaron tener una T_m superior a la de la enzima nativa, lo cual representa un índice de éxito del 20%. Si bien no pudo establecerse una correlación entre el porcentaje de conservación de los aminoácidos consenso y el T_m obtenido, ni una correlación general entre esta y el $\Delta\Delta G^{\text{fold}}$ calculado, se corroboró que los cálculos de $\Delta\Delta G^{\text{fold}}$ suelen ser mejores predictores de mutaciones desestabilizantes que de aquellas estabilizantes.

En base a las simulaciones de MD corridas en el marco del protocolo FRESCO postulamos algunos motivos por los que las mutaciones estabilizantes resultaron serlo, lo que se convierte en un insumo para trabajos futuros.

Logramos obtener tres mutantes con una estabilidad y actividad superiores a las de la enzima nativa: $\Delta 211-217$ (ΔT_m 2.0°C, act. rel. 139%), M1 (ΔT_m 4.7°C, act. rel. 123%) y M2 (ΔT_m 6.0°C, act. rel. 213%).

En total, entre mutantes puntuales y múltiples obtuvimos diez variantes con una estabilidad superior a la de la enzima nativa, la mayoría de ellas presentando una actividad comparable o incluso mejor, lo cual confirma el éxito de las estrategias de diseño racional en el desarrollo de biocatalizadores. Las herramientas adquiridas en este trabajo, tanto intelectuales como técnicas, podrán ser aplicadas al diseño y aplicación de nuevas enzimas de relevancia en el marco de nuestro país.

Conclusiones generales y perspectivas

Una de las virtudes de la escritura de una tesis de maestría es que lo obliga a uno a revisar todo el trabajo realizado durante un lapso de, siendo conservativo, dos años. Con el diario del lunes es sencillo afirmar que si tuviésemos que hacerlo todo de vuelta, no nos llevaría más de seis meses.¹⁵³ Más allá de lo ilusorio de esta afirmación, no podemos dejar pasar que si llegamos a pensarlo es porque adquirimos un vagón de conocimiento, en base a errores y aciertos (especialmente los primeros), que nos permite haber llegado a este punto.

Hemos comprobado a través de este trabajo que el empleo de herramientas bioinformáticas para el diseño racional de mutantes estabilizados de TACap, así como la capacidad de análisis eficaz de estos mutantes en el laboratorio húmedo permiten, en conjunto, el desarrollo de variantes enzimáticas mejoradas a una fracción del costo y esfuerzo requerido en otras aproximaciones. En este caso, reportamos la obtención de al menos tres variantes enzimáticas de TACap con una estabilidad y actividad superiores a las de la enzima nativa. Teniendo esto en cuenta, hay algunos *hotspots* que consideramos de relevancia para la continuación de este trabajo.

Un estudio más completo a nivel molecular del sistema TACap podría lograrse mediante un método de *Replica Exchange Molecular Dynamics* (REMD). Esta metodología simula varias copias de la proteína a distintas temperaturas en paralelo, y estas copias se intercambian periódicamente, permitiendo así una mayor exploración de la superficie de energía potencial. Esto podría darnos pistas, por ejemplo, sobre los resultados obtenidos con Δ C-tail o el camino de desnaturalización de la enzima en su forma apo. Debe tenerse en cuenta, sin embargo, que lamentablemente en Uruguay no contamos con los recursos de cómputo suficientes para correr este tipo de simulaciones. Alternativamente, al menos para comprender la mecánica de disociación del dímero podrían emplearse simulaciones de grano grueso. Este método representa la estructura proteica a un nivel más abstracto, de menor detalle, en el que las partículas involucradas pasan a ser residuos, en lugar de átomos. De esta forma el costo computacional es muy inferior al de las dinámicas clásicas, y permite simular escalas de tiempo bastante más grandes.¹⁵⁴

Sería deseable verificar la expresión, mediante SDS-PAGE, de aquellos mutantes con un ΔT_m positivo para los que se halló una baja actividad. Es posible que estos mutantes tengan una expresión diferencial respecto a las variantes cuya expresión evaluamos, y este sea el motivo de la menor actividad medida. También sería ideal purificar algunas de las enzimas y,

por un lado, verificar los valores de T_m , y por el otro determinar la actividad específica y continuar con su caracterización.

Debe ensayarse la actividad de las mejores variantes en la síntesis de L-anfetamina, precursor del marcador frío L-nordeuterodeprenil. Para esto será primero necesario poner a punto la analítica de la reacción. Aquella/s variantes con actividad frente al sustrato de interés deberán ser evaluadas en cuanto a su estabilidad cinética, para finalmente optimizar las condiciones de reacción. Cumplidos estos pasos debe ser prioritaria la evaluación de su estabilidad en el guardado (congelado, liofilizado). En última instancia, desarrollado el proceso de síntesis de L-anfetamina utilizando una variante estabilizada de TACap, este método estará listo para su implementación a nivel de producción en el CUDIM, constituyendo una síntesis alineada con los principios de la química verde y con un costo inferior para el país.

Anexos

Secuencia de ADN de TACap

La siguiente secuencia corresponde a la enzima nativa, pero los codones fueron optimizados para *E. coli* por la empresa GenScript. Adicionalmente fueron incorporadas dos mutaciones silenciosas para permitir el clonado por Golden Gate (eliminación de sitios para BsaI) y una cola de histidinas C-terminal.

>TACap_wt_optimized_*E. coli*

```
atgaccactatggaaaagattttctctgcttatcaggccccgctgtcaacactgacggcaagtcgtgcgtctaaccattcgcatggta
tcgctgggtccaaggcagagttacaccgattcatgaagctcagatccctatgttgaccaaggcttctgcacagcgattgacttacg
acgttccgagtggtgggatggtagattttccgctggatgaccatctggaacgtttggagagaagttgtgcaaagatgagactgcgtt
gccctttccaagagcacaggttcgtcacaccctgtgtgctatgttgccagatctggtattcgcgacgcatttgggaactgatcgtcac
tcgcggttcgctggagtcagaggtctgtctgctgccgaggttgatgcgctgcctaactcctgtatatgtggattgtccatacgtttggg
ttatggaaccggctgttcagctggccggtatctggtccgcaatcgttgcgcgtacagtcatagaacgccaccggttgcattggatccg
acagtgaaaaatctgcaatggggcgacttgacgagaggaatgttgaagctaacgatcgcggcgccggatattccttctgaccgata
gatcaattagcgaagaaccgcagacgcgaatattactgaaggttccggcttaacatcgttgggtcaaagatggcaccttgacactc
caaagcgtggtgtcctggagggcggttacaagagaatctgtgttcgagtggtgccgtagattgggtgtcccttacgcactggacacggtc
ccagttcgctggctttggaagccgatgagattttatgtgtacaacggccgggtggtattatgccgatcaccactctggatggaaaacctg
tgggagacgggtccgctcggccaattactcgccaaatctgggatgtttattggcggttgattacgaagacggttttcattcggcgtggat
tatgaagacgaaaccggcctggagggagttgcaaacggaaatgcgaacggfactgtgagcgtcaatggtaaagcacatccatcac
catcactaa
```

Secuencia aminoacídica de TACap

>TACap_wt_HisCterm

```
MTTMEKIFSAYQARVSTLTASRASNPFDGIAWVQGRVTPIHEAQIPMLDQGFLHSD
LTYDVPSVWDGRFFRLDDHLERLERSCAKMRLRCPLPRAQVRHTLCAMLARSGIRD
AFVELIVTRGLRGVRLSAAEVDALPNSLYMWIVPYVWVMEPAVQLAGSGSAIVAR
TVHRTPPVCMDPTVKNLQWGDLTRGMFEANDRGAGYPFLTDRSISEETADANITEGS
GFNIVVVKDGLHTPKRGVLEGVTTRESVFECRRRLGVPYALDTPVRLALEADEIFM
CTTAGGIMPITLDGKPVGDGAVGPITRQIWDVYWRLHYEDGFSFAVDYEDETGLEG
VANGNANGTVSVNGKAHHHHHHH
```

Modificación de campo de fuerza

FRCMOD for internal aldimine Lys-PLP
 MASS

BOND

c3-ne	334.70	1.456		same as c3-nc, penalty score= 10.0
hn-ne	385.60	1.023		same as hn-n2, penalty score= 11.1
hn-nc	385.60	1.023		same as hn-n2, penalty score= 8.2
cc-o	622.90	1.225		same as c2-o, penalty score= 4.9
C-n	490.0	1.335		same as ff19SB C-N
C-N	490.0	1.335		same as ff19SB C-N

ANGLE

c3-c3-ne	66.700	108.800		same as c3-c3-n2, penalty score= 2.8
c3-ne-hn	46.109	115.350		Calculated using c3-n2-hn, penalty score= 30.5
c2-ne-c3	66.811	114.440		Calculated using c3-n2-c2, penalty score= 30.5
h1-c3-ne	50.100	108.570		same as h1-c3-nd, penalty score= 2.1
cc-c2-ne	71.200	126.010		same as c2-c2-n2, penalty score= 5.8
c2-ne-hn	50.918	119.090		Calculated using hn-n2-c2, penalty score= 30.5
cc-nc-hn	50.904	116.100		Calculated using ca-n2-hn, penalty score= 22.6
cc-cc-o	17.800	136.060		same as cd-cc-o, penalty score= 1.1
O-C-n	80.0	122.90		same as ff19SB N-C-O
C-n-c3	50.0	121.90		same as ff19SB C-N-XC
C-n-hn	50.0	120.00		same as ff19SB C-N-H
XC-C-n	70.0	116.60		same as ff19SB XC-C-N
O-C-N	80.0	122.90		same as ff19SB N-C-O
C-N-H	50.0	120.00		same as ff19SB C-N-H
C-N-XC	50.0	121.90		same as ff19SB C-N-XC
C3-C-N	70.0	116.60		same as ff19SB XC-C-N

DIHE

ne-c2-cc-cc	4	26.600	180.000	2.000	same as X-c2-ce-X, penalty score=136.0
h4-c2-cc-cc	4	26.600	180.000	2.000	same as X-c2-ce-X, penalty score=136.0
O-C-n-c3	4	10.00	180.0	2.	same as ff19SB X-C-N-X
O-C-N-XC	4	10.00	180.0	2.	same as ff19SB X-C-N-X

```

O -C -n -hn 1 2.50 180.0 -2. same as ff19SB H -N -C -O
O -C -n -hn 1 2.00 0.0 1. same as ff19SB H -N -C -O
O -C -N -H 1 2.50 180.0 -2. same as ff19SB H -N -C -O
O -C -N -H 1 2.00 0.0 1. same as ff19SB H -N -C -O
XC-C -n -c3 4 10.00 180.0 2. same as ff19SB X -C -N -X
XC-C -n -hn 4 10.00 180.0 2. same as ff19SB X -C -N -X
c3-c -N -H 4 10.00 180.0 2. same as ff19SB X -C -N -X
c3-c -N -XC 4 10.00 180.0 2. same as ff19SB X -C -N -X

```

IMPROPER

```

c3-cc-cc-nc 1.1 180.0 2.0 Using the default value
cc-cc-cc-o 1.1 180.0 2.0 Using the default value
c2-cc-cc-cc 1.1 180.0 2.0 Using the default value
cc-h4-c2-ne 1.1 180.0 2.0 Same as X -X -ca-ha, penalty score= 75.4 (use general term))
c3-cc-cc-cc 1.1 180.0 2.0 Using the default value
cc-h4-cc-nc 1.1 180.0 2.0 Same as X -X -ca-ha, penalty score= 67.2 (use general term))

```

NONBON

Librería de parámetros para KLP

```

!!index array str
"KLP"
!entry.KLP.unit.atoms table str name str type int
typex int resx int flags int seq int elimnt dbl chg
"N" "n" 0 1 131072 1 7 -0.347900
"CA" "c3" 0 1 131072 2 6 -0.441900
"C" "c" 0 1 131072 3 6 0.734100
"O" "o" 0 1 131072 4 8 -0.589400
"CB" "c3" 0 1 131072 5 6 0.499500
"CG" "c3" 0 1 131072 6 6 0.062900
"CD" "c3" 0 1 131072 7 6 0.109400
"CE" "c3" 0 1 131072 8 6 0.115600
"NZ" "ne" 0 1 131072 9 7 -0.139900
"H" "hn" 0 1 131072 10 1 0.274700
"HA" "h1" 0 1 131072 11 1 0.110200
"HB2" "hc" 0 1 131072 12 1 -0.114700
"HB3" "hc" 0 1 131072 13 1 -0.114700
"HG2" "hc" 0 1 131072 14 1 -0.078200

```

```

HG3 "hc" 0 1 131072 15 1 -0.078200
HD2 "hc" 0 1 131072 16 1 -0.005800
HD3 "hc" 0 1 131072 17 1 -0.005800
HE2 "h1" 0 1 131072 18 1 -0.011400
HE3 "h1" 0 1 131072 19 1 -0.011400
HZ "hn" 0 1 131072 20 1 0.381900
N1 "nc" 0 1 131072 21 7 -0.212000
C2 "cc" 0 1 131072 22 6 0.205800
C2A "c3" 0 1 131072 23 6 -0.279000
C3 "cc" 0 1 131072 24 6 0.355700
O3 "o" 0 1 131072 25 8 -0.765500
C4 "cc" 0 1 131072 26 6 0.102000
C4A "c2" 0 1 131072 27 6 0.102300
C5 "cc" 0 1 131072 28 6 -0.173800
C6 "cc" 0 1 131072 29 6 -0.099200
C5A "c3" 0 1 131072 30 6 0.349900
O4P "os" 0 1 131072 31 8 -0.538200
P "p5" 0 1 131072 32 15 1.243900
O1P "o" 0 1 131072 33 8 -0.897000
O2P "o" 0 1 131072 34 8 -0.897000
O3P "o" 0 1 131072 35 8 -0.897000
H1 "hn" 0 1 131072 36 1 0.345500
H2A1 "hc" 0 1 131072 37 1 0.096900
H2A2 "hc" 0 1 131072 38 1 0.096900

```

```

H2A3 "hc" 0 1 131072 39 1 0.096900
H4A "h4" 0 1 131072 40 1 0.205400
H5A1 "h1" 0 1 131072 41 1 0.008300
H5A2 "h1" 0 1 131072 42 1 0.008300
H6 "h4" 0 1 131072 43 1 0.191900
lentry.KLP.unit.atomsperinfo table str pname str
ptype int ptypex int pelmnt dbl pchg
"N" "n" 0 -1 0.0
"CA" "c3" 0 -1 0.0
"C" "c" 0 -1 0.0
"O" "o" 0 -1 0.0
"CB" "c3" 0 -1 0.0
"CG" "c3" 0 -1 0.0
"CD" "c3" 0 -1 0.0
"CE" "c3" 0 -1 0.0
"NZ" "ne" 0 -1 0.0
"H" "hn" 0 -1 0.0
"HA" "h1" 0 -1 0.0
"HB2" "hc" 0 -1 0.0
"HB3" "hc" 0 -1 0.0
"HG2" "hc" 0 -1 0.0
"HG3" "hc" 0 -1 0.0
"HD2" "hc" 0 -1 0.0
"HD3" "hc" 0 -1 0.0

```

```

"HE2" "h1" 0 -1 0.0
"HE3" "h1" 0 -1 0.0
"HZ" "hn" 0 -1 0.0
"N1" "nc" 0 -1 0.0
"C2" "cc" 0 -1 0.0
"C2A" "c3" 0 -1 0.0
"C3" "cc" 0 -1 0.0
"O3" "o" 0 -1 0.0
"C4" "cc" 0 -1 0.0
"C4A" "c2" 0 -1 0.0
"C5" "cc" 0 -1 0.0
"C6" "cc" 0 -1 0.0
"C5A" "c3" 0 -1 0.0
"O4P" "os" 0 -1 0.0
"P" "p5" 0 -1 0.0
"O1P" "o" 0 -1 0.0
"O2P" "o" 0 -1 0.0
"O3P" "o" 0 -1 0.0
"H1" "hn" 0 -1 0.0
"H2A1" "hc" 0 -1 0.0
"H2A2" "hc" 0 -1 0.0
"H2A3" "hc" 0 -1 0.0
"H4A" "h4" 0 -1 0.0
"H5A1" "h1" 0 -1 0.0

"H5A2" "h1" 0 -1 0.0
"H6" "h4" 0 -1 0.0
!entry.KLP.unit.boundingBox array dbl
-1.000000
0.0
0.0
0.0
0.0
!entry.KLP.unit.childsequence single int
2
!entry.KLP.unit.connect array int
1
3
!entry.KLP.unit.connectivity table int atom1x int
atom2x int flags
1 2 1
1 10 1
2 3 1
2 5 1
2 11 1
3 4 1
5 6 1
5 12 1
5 13 1

```

6 7 1
6 14 1
6 15 1
7 8 1
7 16 1
7 17 1
8 9 1
8 18 1
8 19 1
9 20 1
9 27 1
21 22 1
21 29 1
21 36 1
22 23 1
22 24 1
23 37 1
23 38 1
23 39 1
24 25 1
24 26 1
26 27 1
26 28 1
27 40 1

28 29 1
28 30 1
29 43 1
30 31 1
30 41 1
30 42 1
31 32 1
32 33 1
32 34 1
32 35 1

lentry.KLP.unit.hierarchy table str abovetype int
abovex str belowtype int belowx
"u" 0 "R" 1
"R" 1 "A" 1
"R" 1 "A" 2
"R" 1 "A" 3
"R" 1 "A" 4
"R" 1 "A" 5
"R" 1 "A" 6
"R" 1 "A" 7
"R" 1 "A" 8
"R" 1 "A" 9
"R" 1 "A" 10
"R" 1 "A" 11

```

"R" 1 "A" 12
"R" 1 "A" 13
"R" 1 "A" 14
"R" 1 "A" 15
"R" 1 "A" 16
"R" 1 "A" 17
"R" 1 "A" 18
"R" 1 "A" 19
"R" 1 "A" 20
"R" 1 "A" 21
"R" 1 "A" 22
"R" 1 "A" 23
"R" 1 "A" 24
"R" 1 "A" 25
"R" 1 "A" 26
"R" 1 "A" 27
"R" 1 "A" 28
"R" 1 "A" 29
"R" 1 "A" 30
"R" 1 "A" 31
"R" 1 "A" 32
"R" 1 "A" 33
"R" 1 "A" 34
"R" 1 "A" 35

"R" 1 "A" 36
"R" 1 "A" 37
"R" 1 "A" 38
"R" 1 "A" 39
"R" 1 "A" 40
"R" 1 "A" 41
"R" 1 "A" 42
"R" 1 "A" 43

lentry.KLP.unit.name single str
"KLP"

lentry.KLP.unit.positions table dbl x dbl y dbl z
96.440000 15.537000 44.028000
95.053000 15.086000 43.675000
94.524000 15.930000 42.691000
94.386000 17.251000 42.937000
94.184000 15.132000 44.897000
92.808000 14.645000 44.548000
91.943000 14.665000 45.773000
90.588000 14.118000 45.434000
90.743000 12.823000 44.927000
96.409000 16.471000 44.382000
95.087000 14.149000 43.327000
94.574000 14.542000 45.603000
94.133000 16.071000 45.238000

```

92.410000 15.232000 43.844000
 92.869000 13.708000 44.204000
 92.360000 14.092000 46.479000
 91.856000 15.602000 46.112000
 90.021000 14.094000 46.257000
 90.153000 14.696000 44.744000
 91.289000 12.675000 44.102000
 90.779000 8.047000 44.077000
 91.292000 9.120000 43.438000
 92.038000 8.949000 42.266000
 91.085000 10.360000 43.929000
 91.618000 11.472000 43.268000
 90.364000 10.528000 45.058000
 90.173000 11.770000 45.549000
 89.854000 9.456000 45.700000
 90.060000 8.215000 45.208000
 89.117000 9.631000 46.877000
 89.976000 10.302000 47.906000
 89.192000 10.490000 49.171000
 90.049000 11.176000 50.193000
 87.981000 11.329000 48.889000
 88.763000 9.152000 49.697000
 90.933000 7.128000 43.714000
 92.338000 9.841000 41.929000

91.472000 8.504000 41.572000
 92.837000 8.380000 42.460000
 89.638000 11.906000 46.382000
 88.818000 8.740000 47.219000
 88.317000 10.200000 46.688000
 89.683000 7.420000 45.684000
 lentry.KLP.unit.residueconnect table int c1x int c2x
 int c3x int c4x int c5x int c6x
 1 3 0 0 0
 lentry.KLP.unit.residues table str name int seq int
 childseq int startatomx str restype int imagingx
 "KLP" 1 44 1 "p" 0
 lentry.KLP.unit.residuesPdbSequenceNumber array int
 0
 lentry.KLP.unit.solventcap array dbl
 -1.000000
 0.0
 0.0
 0.0
 0.0
 0.0
 lentry.KLP.unit.velocities table dbl x dbl y dbl z
 0.0 0.0 0.0
 0.0 0.0 0.0
 0.0 0.0 0.0

Códigos empleados



```
1 source leaprc.protein.ff19SB
2 source leaprc.gaff
3 loadamberparams 4_klp.frcmod
4 KLP = loadmol2 3_klp_ch.mol2
5 check KLP
6 saveoff KLP 5_klp.lib
7 quit
```

Código 1 Generación de la librería con tleap.



```
1 source leaprc.protein.ff19SB
2 source leaprc.gaff
3 loadoff 5_klp_mod.lib
4 loadamberparams 4_klp.frcmod
5 4cmd = loadpdb 4cmd.pqr
6 check 4cmd
7 savepdb 4cmd 4cmd.pdb
8 saveamberparm 4cmd 4cmd_dry.prmtop 4cmd_dry.inpcrd
9 quit
```

Código 2 Detección de los parámetros faltantes en el campo de fuerza con tleap.



```
1 mv 4cmd.pdb 4cmd_ori.pdb
2 grep -E "^ATOM|^HETATM.*PLP|^TER|^END" 4cmd_ori.pdb | sed
  's/HETATM/ATOM /g' > 4cmd.pdb
```

Código 3 Curado de PDB en lenguaje Bash para conservar las entradas ATOM, las HETATM que contienen al PLP y los finales de cadena.



```
1 source leaprc.protein.ff19SB
2 source leaprc.gaff
3 loadoff klp.lib
4 loadamberparams klp.frcmod
5 4cmd = loadpdb 4cmd.pqr
6 check 4cmd
7 savepdb 4cmd 4cmd_leap.pdb
8 quit
```

Código 4 Verificación de errores en el PDB mediante tleap.



```
1 source leaprc.protein.ff19SB
2 source leaprc.gaff
3 source leaprc.water.tip3p
4 loadoff klp.lib
5 loadamberparams klp.frcmod
6 4cmd = loadpdb 4cmd_leap.pdb
7 addions 4cmd Na+ 0
8 solvateoct 4cmd TIP3PBOX 10
9 addionsrand 4cmd Na+ 54 Cl- 54
10 savepdb 4cmd s4cmd.pdb
11 saveamberparm 4ce5 s4ce5.prmtop s4ce5.inpcrd
12 quit
```

Código 5 Generación de archivos de topología y coordenadas del sistema mediante tleap.



```
1 Minimizing the system with 25 kcal/mol restraints on PROTEIN,  
2 2000 steps of steepest descent and 500 of conjugated gradient.  
3 &cntrl  
4   imin=1,ntmin=1,ncyc=2000,maxcyc=2500,  
5   ntr=1,restraintmask=':1-{$res}',restraint_wt=25.0,  
6   cut=9.0,  
7   nsnb=10,  
8   ntp=50,  
9 /
```

Código 6 Minimización del sistema con pmemd.



```
1 Heating the system with 5 kcal/mol restraints on PROTEIN,  
2 V = const., 100 ps  
3 &cntrl  
4   t=0.0,dt=0.002,nstlim=50000,  
5   ntr=1,restraintmask=':1-{$res}',restraint_wt=5.0,  
6   tempi=100.0,ntt=3,ig=-1,gamma_ln=5.0,  
7   ntc=2,ntf=2,  
8   nscm=5000,  
9   cut=9.0,  
10  ntp=500,ntwr=0,ntwx=500,  
11  nmropt=1,  
12 /  
13 &wt  
14   type='TEMP0',  
15   istep1=0,istep2=5000,  
16   value1=100.0,value2={$temp},  
17 /  
18 &wt  
19   type='TEMP0',  
20   istep1=5001,istep2=50000,  
21   value1={$temp},value2={$temp},  
22 /  
23 &wt type='END',  
24 /
```

Código 7 Calentamiento del sistema con pmemd.



```
1 Equilibrating the system during 500 ps at constant T,  
2 P = 1 bar and coupling = 0.2 ps.  
3 &cntrl  
4   ntx=5,irest=1,  
5   t=0.0,dt=0.002,nstlim=250000,  
6   temp0=${temp},ntt=3,ig=-1,gamma_ln=5.0,  
7   pres0=1.0,ntp=1,barostat=1,taup=0.2,  
8   ntc=2,ntf=2,  
9   nscm=5000,  
10  cut=9.0,  
11  ntp=500,ntwr=500,ntwx=500,  
12 /
```

Código 8 Equilibración del sistema con pmemd.



```
1 Molecular dynamics production of 50 ns  
2 &cntrl  
3   ntx=5,irest=1,  
4   t=0.0,dt=0.002,nstlim=25000000,  
5   temp0=${temp}.0,ntt=3,ig=-1,gamma_ln=5.0,  
6   pres0=1.0,ntp=1,barostat=1,taup=5.0,  
7   ntc=2,ntf=2,  
8   iwrap=1,  
9   nscm=5000,  
10  cut=9.0,  
11  ntp=500,ntwr=500000,ntwx=5000,  
12 /
```

Código 9 Corrida de la MD del sistema con pmemd.



```
1 #####
2 prmtop= './4ce5/4ce5.prmtop'
3 traj= './4ce5/4ce5-298K-4000ns.nc'
4 ns='last'
5 res='1-650'
6 name="rmsd-`basename $traj .nc`"
7 #####
8
9 cat << EOF > $name.in
10 parm $prmtop
11 trajin $traj 1 $ns
12 autoimage :$res
13 rmsd :$res@CA \
14     first \
15     out $name.dat
16 run
17 EOF
18
19 cpptraj -i $name.in
```

Código 10 Cálculo del RMSD de la simulación en base al movimiento de los carbonos alfa con cpptraj.



```
1 #####
2 prmtop= './4ce5/4ce5'
3 traj= './4ce5/4ce5-298K-4000ns.nc'
4 from=51
5 to=100
6 res='1-325'
7 name="rmsf-`basename $traj .nc`"
8 #####
9
10 cat << EOF > $name.in
11 parm $prmtop.prmtop
12 trajin $traj $from $to
13 autoimage
14 rmsd :$res@CA first
```



```
15 average :$res@CA \  
16     crdset AVG  
17 run  
18 rmsd :$res@CA \  
19     ref AVG  
20 rmsf :$res@CA \  
21     byres \  
22     out $name.dat  
23 run  
24 EOF  
25  
26 cpptraj -i $name.in
```

Código 11 Cálculo del RMSF por residuo en una ventana dada de la simulación.



```
1 from matplotlib import pyplot as plt  
2 import pandas as pd  
3  
4 # Title  
5 title='ca.rmsd-tacap-apo'  
6  
7 # Set sources  
8 files = {'tacap-HM-holo-400K': './tacap/rmsd/rmsd-tacap-HM-  
400K.ref.400.dat',  
9         'tacap-HM-apo-400K': './tacap/rmsd/rmsd-tacap-HM-apo-  
400K.ref.holo-400.dat'}  
10  
11 # Plot data  
12 for label, file in files.items():  
13     df = pd.read_csv(file, delim_whitespace=True)  
14     start = 0  
15     step = 1  
16     stop = len(df)  
17     averages = []  
18     for i in range(start, stop, step):  
19         averages.append(df.iloc[:,1].iloc[:i].mean())  
20     plt.plot(range(start, stop, step), averages, label=label)
```



```
21 # Add legend
22 plt.legend()
23
24 # Set axes label
25 plt.xlabel("tiempo (ns)")
26 plt.ylabel("RMSD promedio acumulado (Å)")
27
28 plt.ylim(top=None, bottom=None)
29 plt.xlim(left=None, right=None)
30
31 # Save fig
32 plt.savefig(title+'.svg')
```

Código 12 Gráfico del RMSD promedio acumulado.



```
1 from matplotlib import pyplot as plt
2 import pandas as pd
3
4 # Title
5 title='rmsd-4ce5.ref.298'
6
7 # Set sources
8 files = {'4ce5-298K': './4ce5/rmsd/rmsd-4ce5-298.ref.298.dat',
9          '4ce5-323K': './4ce5/rmsd/rmsd-4ce5-323.ref.298.dat',
10         '4ce5-348K': './4ce5/rmsd/rmsd-4ce5-348.ref.298.dat',
11         '4ce5-298K-urea': './4ce5/rmsd/rmsd-4ce5-298-
12         urea.ref.298.dat'}
13 # Plot data
14 for label, file in files.items():
15     df = pd.read_csv(file, header=0, delim_whitespace=True)
16     plt.plot(df.iloc[:,0], df.iloc[:,1], label=label)
17     print(df.iloc[301:500,1].mean(), df.iloc[301:500,1].std())
18
19 # Add legend
20 plt.legend()
```



```
21 # Set axes label
22 plt.xlabel("tiempo (ns)")
23 plt.ylabel("RMSD (Å)")
24
25 # Set Y axis limits
26 plt.ylim(top=4.2, bottom=None)
27 #plt.xlim([-5,500])
28
29 # Save fig
30 plt.savefig(title+'.svg')
```

Código 13 Gráfico de datos de RMSD.

Archivos adjuntos

Tabla 19 Lista y descripción de archivos adjuntos.

klp.frmod	Archivo de modificación del campo de fuerza para el procesamiento de KLP.
klp.lib	Librería con parámetros del residuo KLP.
4ce5.pdb	Estructura procesada de 4CMD.
4cmd.pdb	Estructura procesada de 4CE5.
addionsrand.py	Script para calcular el número de iones a agregar a un sistema de simulación dados el número de moléculas de agua en el sistema y la molaridad deseada.
Bio-Rad electroporation.pdf	Guía de Bio-Rad para la preparación de células electrocompetentes, así como para su transformación.
Bio-Rad fluorophore reference guide.pdf	Guía de fluoróforos disponibles para el marcado de proteínas.
Bio-Rad freeze 'n squeeze.pdf	Kit para la purificación de ADN desde gel de agarosa.
equilibrate.sh	Script para la equilibración de un sistema antes de la MD de producción.
fasta2pir.sh	Script para la conversión de secuencia del formato FASTA al PIR.
FRESCO protocol supp.pdf	Guía suplementaria al protocolo FRESCO.

FRESCO protocol.pdf	Protocolo computacional para la mejora de estabilidad de proteínas mediante cálculos de energía libre.
FRESCO.zip	Conjunto de scripts para la aplicación del protocolo computacional FRESCO.
GoldenGateCloning.pdf	Protocolo para el clonado por Golden Gate.
heat.sh	Script para el calentamiento del sistema luego de la minimización.
minimization.sh	Script para la minimización del sistema luego de su creación.
pBAD-C-term-His.geneious	Vector pBAD empleado para el Golden Gate cloning.
pBAD-myc-HisA.geneious	Vector pBAD donado por el Prof. Fürst (RUG) para el subclonado por <i>Restriction Free Cloning</i> .
pBAD-TACap.geneious	Vector pBAD conteniendo el inserto TACap de interés.
pKK223-3 TACap.geneious	Vector pKK223-3 conteniendo TACap, obtenido en un trabajo anterior.
pKK223-3.geneious	Vector pKK223-3.
Qiagen miniprep.pdf	Protocolo del kit de purificación de ADN a partir de <i>E. coli</i> de Qiagen.
runMD.sh	Script para correr una MD de un sistema ya equilibrado.
TACap-HM.pdb	Archivo PDB de la proteína TACap obtenido por modelado por homología.

Bibliografía

1. Philip, B. N., Kiss, A. J., & Lee, R. E., Jr. (2011). The protective role of aquaporins in the freeze-tolerant insect *Eurosta solidaginis*: Functional characterization and tissue abundance of EsAQP1. *Journal of Experimental Biology*, *214*(5), 848-857.
<https://doi.org/10.1242/jeb.051276>
2. Arranz-Otaegui, A., Gonzalez Carretero, L., Ramsey, M. N., Fuller, D. Q., & Richter, T. (2018). Archaeobotanical evidence reveals the origins of bread 14,400 years ago in northeastern Jordan. *Proceedings of the National Academy of Sciences*, *115*(31), 7925-7930. <https://doi.org/10.1073/pnas.1801071115>
3. McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nuñez, A., Butrym, E. D., Richards, M. P., Wang, C., Cheng, G., Zhao, Z., & Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(51), 17593-17598.
<https://doi.org/10.1073/pnas.0407921102>
4. Salque, M., Bogucki, P. I., Pyzel, J., Sobkowiak-Tabaka, I., Grygiel, R., Szmyt, M., & Evershed, R. P. (2013). Earliest evidence for cheese making in the sixth millennium BC in northern Europe. *Nature*, *493*(7433), Article 7433. <https://doi.org/10.1038/nature11698>
5. Payen, A., & Persoz, J.-F. (1833). Memoire sur la Diastase, les principaux Produits de ses Réactions, et leurs applications aux arts industriels. *Annales de chimie et de physique*, *53*, 73-92.
6. Kühne, W. (1877). Ueber das Verhalten verschiedener organisirter und sog ungeformter Fermente. *Verhandlungen des Naturhistorisch-medizinischen Vereins zu Heidelberg*, 190-193.
7. Mrabet, Y. (2007). *Peptide bond formation*.
<https://commons.wikimedia.org/wiki/File:Peptidformationball.svg>
8. Hubert, B. (2017). *Amino acids*. https://commons.wikimedia.org/wiki/File:Amino_Acids-wide.svg
9. Ferig, M. (2013). *Protein structures*.
https://commons.wikimedia.org/wiki/File:Main_protein_structure_levels_gl.svg

10. Patel, R. N. (2018). Biocatalysis for synthesis of pharmaceuticals. *Bioorganic Med. Chem.*, 26, 1252-1274. <https://doi.org/10.1016/j.bmc.2017.05.023>
11. Heckmann, C. M., & Paradisi, F. (2020). Looking Back: A Short History of the Discovery of Enzymes and How They Became Powerful Chemical Tools. *ChemCatChem*, 12(24), 6082-6102. <https://doi.org/10.1002/cctc.202001107>
12. Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C., & Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*, 485(7397), Article 7397. <https://doi.org/10.1038/nature11117>
13. Blake, C. C. F., Fenn, R. H., North, A. C. T., Phillips, D. C., & Poljak, R. J. (1962). Structure Of Lysozyme: A Fourier Map of the Electron Density at 6 Å Resolution obtained by X-ray Diffraction. *Nature*, 196(4860), Article 4860. <https://doi.org/10.1038/1961173a0>
14. Mayer, H., Collins, J., & Wagner, F. (1980). Cloning of the Penicillin G-Acylase Gene of Escherichia Coli ATCC 11105 on Multicopy Plasmids. En H. H. Weetall & G. P. Royer (Eds.), *Enzyme Engineering: Volume 5* (pp. 61-69). Springer US. https://doi.org/10.1007/978-1-4684-3749-2_7
15. Woodley, J. M. (2019). Accelerating the implementation of biocatalysis in industry. *Appl. Microbiol. Biotechnol.*, 103, 4733-4739. <https://doi.org/10.1007/s00253-019-09796-x>
16. Wenda, S., Illner, S., Mell, A., & Kragl, U. (2011). Industrial biotechnology—The future of green chemistry? *Green Chemistry*, 13, 3007. <https://doi.org/10.1039/c1gc15579b>
17. Chen, K., & Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: Sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proceedings of the National Academy of Sciences*, 90(12), 5618-5622. <https://doi.org/10.1073/pnas.90.12.5618>
18. Arnold, F. H., & Moore, J. C. (1997). Optimizing industrial enzymes by directed evolution. *Advances in Biochemical Engineering/Biotechnology*, 58, 1-14. Scopus. <https://doi.org/10.1007/bfb0103300>
19. Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by DNA shuffling. *Nature*, 370(6488), 389-391. Scopus. <https://doi.org/10.1038/370389a0>

20. Sheldon, R. A., & Woodley, J. M. (2018). Role of Biocatalysis in Sustainable Chemistry. *Chemical Reviews*, *118*(2), 801-838. <https://doi.org/10.1021/acs.chemrev.7b00203>
21. Sheldon, R. A., & Brady, D. (2019). Broadening the scope of biocatalysis in sustainable organic synthesis. *ChemSusChem*, *12*, 2859-2881. <https://doi.org/10.1002/cssc.201900351>
22. Sun, H., Zhang, H., Ang, E. L., & Zhao, H. (2018). Biocatalysis for the synthesis of pharmaceuticals and pharmaceutical intermediates. *Bioorganic Med. Chem*, *26*, 1275-1284. <https://doi.org/10.1016/j.bmc.2017.06.043>
23. Anastas, P. T., & Warner, J. C. (1998). *Green chemistry: Theory and practice*. Oxford University Press.
24. Woodley, J. M. (2020). New frontiers in biocatalysis for sustainable synthesis. *Curr. Opin. Green Sustain. Chem*, *21*, 22-26. <https://doi.org/10.1016/j.cogsc.2019.08.006>
25. Sheldon, R. A., Brady, D., & Bode, M. L. (2020). The Hitchhiker's guide to biocatalysis: Recent advances in the use of enzymes in organic synthesis. *Chemical Science*, *11*(10), 2587-2605. <https://doi.org/10.1039/C9SC05746C>
26. Zawodny, W., & Montgomery, S. L. (2022). Evolving New Chemistry: Biocatalysis for the Synthesis of Amine-Containing Pharmaceuticals. *Catalysts*, *12*(6), Article 6. <https://doi.org/10.3390/catal12060595>
27. Bornscheuer, U. T. (2017). The fourth wave of biocatalysis is approaching. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2110), 20170063. <https://doi.org/10.1098/rsta.2017.0063>
28. Kunka, A., Marques, S. M., Havlasek, M., Vasina, M., Velatova, N., Cengelova, L., Kovar, D., Damborsky, J., Marek, M., Bednar, D., & Prokop, Z. (2023). Advancing Enzyme's Stability and Catalytic Efficiency through Synergy of Force-Field Calculations, Evolutionary Analysis, and Machine Learning. *ACS Catalysis*, 12506-12518. <https://doi.org/10.1021/acscatal.3c02575>
29. Hekkelman, M. L., de Vries, I., Joosten, R. P., & Perrakis, A. (2023). AlphaFill: Enriching AlphaFold models with ligands and cofactors. *Nature Methods*, *20*(2), Article 2. <https://doi.org/10.1038/s41592-022-01685-y>

30. Bertoline, L. M. F., Lima, A. N., Krieger, J. E., & Teixeira, S. K. (2023). Before and after AlphaFold2: An overview of protein structure prediction. *Frontiers in Bioinformatics*, 3. <https://doi.org/10.3389/fbinf.2023.1120370>
31. Lu, H., Diaz, D. J., Czarnecki, N. J., Zhu, C., Kim, W., Shroff, R., Acosta, D. J., Alexander, B. R., Cole, H. O., Zhang, Y., Lynd, N. A., Ellington, A. D., & Alper, H. S. (2022). Machine learning-aided engineering of hydrolases for PET depolymerization. *Nature*, 604(7907), Article 7907. <https://doi.org/10.1038/s41586-022-04599-z>
32. Patsch, D., & Buller, R. (2023). Improving Enzyme Fitness with Machine Learning. *CHIMIA*, 77(3), Article 3. <https://doi.org/10.2533/chimia.2023.116>
33. Mazurenko, S., Prokop, Z., & Damborsky, J. (2020). Machine Learning in Enzyme Engineering. *ACS Catalysis*, 10(2), 1210-1223. <https://doi.org/10.1021/acscatal.9b04321>
34. The Twelve Principles of Green Chemistry: What it is, & Why it Matters. (2015, septiembre 24). *Compound Interest*. <https://www.compoundchem.com/2015/09/24/green-chemistry/>
35. Sun, Z., Liu, Q., Qu, G., Feng, Y., & Reetz, M. T. (2019). Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chemical Reviews*, 119(3), 1626-1665. <https://doi.org/10.1021/acs.chemrev.8b00290>
36. Webb, B., & Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, 54(1), 5.6.1-5.6.37. <https://doi.org/10.1002/cpbi.3>
37. Yu, Y., Rué Casamajo, A., Finnigan, W., Schnepel, C., Barker, R., Morrill, C., Heath, R. S., De Maria, L., Turner, N. J., & Scrutton, N. S. (2023). Structure-Based Design of Small Imine Reductase Panels for Target Substrates. *ACS Catalysis*, 12310-12321. <https://doi.org/10.1021/acscatal.3c02278>
38. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 1-11. <https://doi.org/10.1038/s41586-021-03819-2>

39. Liu, Q., Xun, G., & Feng, Y. (2018). The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnology Advances*, *October*.
<https://doi.org/10.1016/j.biotechadv.2018.10.011>
40. Musil, M., Konegger, H., Hon, J., Bednar, D., & Damborsky, J. (2019). Computational Design of Stable and Soluble Biocatalysts. *ACS Catalysis*, *9*(2), 1033-1054.
<https://doi.org/10.1021/acscatal.8b03613>
41. Rahban, M., Zolghadri, S., Salehi, N., Ahmad, F., Haertlé, T., Rezaei-Ghaleh, N., Sawyer, L., & Saboury, A. A. (2022). Thermal stability enhancement: Fundamental concepts of protein engineering strategies to manipulate the flexible structure. *International Journal of Biological Macromolecules*, *214*, 642-654. <https://doi.org/10.1016/j.ijbiomac.2022.06.154>
42. Polizzi, K. M., Bommarius, A. S., Broering, J. M., & Chaparro-Riggers, J. F. (2007). Stability of biocatalysts. *Current Opinion in Chemical Biology*, *11*(2), 220-225.
<https://doi.org/10.1016/j.cbpa.2007.01.685>
43. Pongsupasa, V., Anuwat, P., Maenpuen, S., & Wongnate, T. (2022). Rational-Design Engineering to Improve Enzyme Thermostability. En F. Magnani, C. Marabelli, & F. Paradisi (Eds.), *Enzyme Engineering: Methods and Protocols* (pp. 159-178). Springer US.
https://doi.org/10.1007/978-1-0716-1826-4_9
44. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, *99*(6), 1129-1143. <https://doi.org/10.1016/j.neuron.2018.08.011>
45. Pirovano, W., & Heringa, J. (2010). Protein Secondary Structure Prediction. En O. Carugo & F. Eisenhaber (Eds.), *Data Mining Techniques for the Life Sciences* (pp. 327-348). Humana Press. https://doi.org/10.1007/978-1-60327-241-4_19
46. Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., Shapovalov, M. V., Renfrew, P. D., Mulligan, V. K., Kappel, K., Labonte, J. W., Pacella, M. S., Bonneau, R., Bradley, P., Dunbrack, R. L. Jr., Das, R., Baker, D., Kuhlman, B., Kortemme, T., & Gray, J. J. (2017). The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation*, *13*(6), 3031-3048. <https://doi.org/10.1021/acs.jctc.7b00125>

47. Voigt, C. A., Martinez, C., Wang, Z.-G., Mayo, S. L., & Arnold, F. H. (2002). Protein building blocks preserved by recombination. *Nature Structural Biology*, 9(7), Article 7. <https://doi.org/10.1038/nsb805>
48. Schlick, T., & Portillo-Ledesma, S. (2021). Biomolecular modeling thrives in the age of technology. *Nature Computational Science*, 1(5), Article 5. <https://doi.org/10.1038/s43588-021-00060-9>
49. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguetto, L., Huang, H., Miguez, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., & Simmerling, C. (2020). ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation*, 16(1), 528-552. <https://doi.org/10.1021/acs.jctc.9b00591>
50. Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpí, J. L., González, C., Vendruscolo, M., Laughton, C. A., Harris, S. A., Case, D. A., & Orozco, M. (2016). Parmbsc1: A refined force field for DNA simulations. *Nature Methods*, 13(1), Article 1. <https://doi.org/10.1038/nmeth.3658>
51. Dickson, C. J., Walker, R. C., & Gould, I. R. (2022). Lipid21: Complex Lipid Membrane Simulations with AMBER. *Journal of Chemical Theory and Computation*, 18(3), 1726-1736. <https://doi.org/10.1021/acs.jctc.1c01217>
52. *Molecular Mechanics (MM)*. (2022, julio 23). CompChems. <https://www.compchems.com/molecular-mechanics-mm/>
53. *Molecular Dynamics: Equations of motion*. (2022, julio 27). CompChems. <https://www.compchems.com/molecular-dynamics-equations-of-motion/>
54. Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *Journal of Computational Physics*, 23(3), 327-341. [https://doi.org/10.1016/0021-9991\(77\)90098-5](https://doi.org/10.1016/0021-9991(77)90098-5)
55. Hopkins, C. W., Le Grand, S., Walker, R. C., & Roitberg, A. E. (2015). Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation*, 11(4), 1864-1874. <https://doi.org/10.1021/ct5010406>

56. Bringas, M. (2020). *Simulación multiescala de procesos de larga escala temporal y espacial en proteínas*. Universidad de Buenos Aires.
57. *How to compute the RMSF using GROMACS*. (2022, octubre 8). CompChems. <https://www.compchems.com/how-to-compute-the-rmsf-using-gromacs/>
58. Ivankov, D. N., Bogatyreva, N. S., Lobanov, M. Y., & Galzitskaya, O. V. (2009). Coupling between Properties of the Protein Shape and the Rate of Protein Folding. *PLOS ONE*, 4(8), e6476. <https://doi.org/10.1371/journal.pone.0006476>
59. Connolly, M. L. (1983). Solvent-Accessible Surfaces of Proteins and Nucleic Acids. *Science*, 221(4612), 709-713. <https://doi.org/10.1126/science.6879170>
60. Weiser, J., Shenkin, P. S., & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *Journal of Computational Chemistry*, 20(2), 217-230. [https://doi.org/10.1002/\(SICI\)1096-987X\(19990130\)20:2<217::AID-JCC4>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1096-987X(19990130)20:2<217::AID-JCC4>3.0.CO;2-A)
61. White, E. (2022, julio 15). *How do you know when your MD simulation has converged?* #AcademicTwitter [Tweet]. Twitter. <https://twitter.com/elisew00/status/1548045696552513539>
62. Sawle, L., & Ghosh, K. (2016). Convergence of Molecular Dynamics Simulation of Protein Native States: Feasibility vs Self-Consistency Dilemma. *Journal of Chemical Theory and Computation*, 12(2), 861-869. <https://doi.org/10.1021/acs.jctc.5b00999>
63. Grossfield, A., Feller, S. E., & Pitman, M. C. (2007). Convergence of molecular dynamics simulations of membrane proteins. *Proteins: Structure, Function, and Bioinformatics*, 67(1), 31-40. <https://doi.org/10.1002/prot.21308>
64. Chodera, J. D. (2016). A Simple Method for Automated Equilibration Detection in Molecular Simulations. *Journal of Chemical Theory and Computation*, 12(4), 1799-1805. <https://doi.org/10.1021/acs.jctc.5b00784>
65. Knapp, B., Frantal, S., Cibena, M., Schreiner, W., & Bauer, P. (2011). Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *Journal of Computational Biology*, 18(8), 997-1005. <https://doi.org/10.1089/cmb.2010.0237>

66. Porebski, B. T., & Buckle, A. M. (2016). Consensus protein design. *Protein Engineering, Design and Selection*, 29(7), 245-251. <https://doi.org/10.1093/protein/gzw015>
67. Nakano, S., Motoyama, T., Miyashita, Y., Ishizuka, Y., Matsuo, N., Tokiwa, H., Shinoda, S., Asano, Y., & Ito, S. (2018). Benchmark Analysis of Native and Artificial NAD⁺-Dependent Enzymes Generated by a Sequence-Based Design Method with or without Phylogenetic Data. *Biochemistry*, 57(26), 3722-3732. <https://doi.org/10.1021/acs.biochem.8b00339>
68. Meng, Q., Capra, N., Palacio, C. M., Lanfranchi, E., Otzen, M., van Schie, L. Z., Rozeboom, H. J., Thunnissen, A.-M. W. H., Wijma, H. J., & Janssen, D. B. (2020). Robust ω -Transaminases by Computational Stabilization of the Subunit Interface. *ACS Catalysis*, 10(5), 2915-2928. <https://doi.org/10.1021/acscatal.9b05223>
69. Aalbers, F. S., Fürst, M. J., Rovida, S., Trajkovic, M., Gómez Castellanos, J. R., Bartsch, S., Vogel, A., Mattevi, A., & Fraaije, M. W. (2020). Approaching boiling point stability of an alcohol dehydrogenase through computationally-guided enzyme engineering. *eLife*, 9, e54639. <https://doi.org/10.7554/eLife.54639>
70. Lončar, N., van Beek, H. L., & Fraaije, M. W. (2019). Structure-Based Redesign of a Self-Sufficient Flavin-Containing Monooxygenase towards Indigo Production. *International Journal of Molecular Sciences*, 20(24), Article 24. <https://doi.org/10.3390/ijms20246148>
71. Fürst, M. J. L. J., Boonstra, M., Bandstra, S., & Fraaije, M. W. (2019). Stabilization of cyclohexanone monooxygenase by computational and experimental library design. *Biotechnology and Bioengineering*, 116(9), 2167-2177. <https://doi.org/10.1002/bit.27022>
72. Martin, C., Ovalle Maqueo, A., Wijma, H. J., & Fraaije, M. W. (2018). Creating a more robust 5-hydroxymethylfurfural oxidase by combining computational predictions with a novel effective library design. *Biotechnology for Biofuels*, 11(1), 56. <https://doi.org/10.1186/s13068-018-1051-x>
73. Buß, O., Rudat, J., & Ochsenreither, K. (2018). FoldX as Protein Engineering Tool: Better Than Random Based Approaches? *Computational and Structural Biotechnology Journal*, 16, 25-33. <https://doi.org/10.1016/j.csbj.2018.01.002>

74. Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., & Bornscheuer, U. T. (2021). Biocatalysis: Enzymatic Synthesis for Industrial Applications. *Angewandte Chemie International Edition*, 60(1), 88-119. <https://doi.org/10.1002/anie.202006648>
75. Brooks, W. H., Guida, W. C., & Daniel, K. G. (2011). The Significance of Chirality in Drug Design and Development. *Current Topics in Medicinal Chemistry*, 11(7), 760-770. <https://doi.org/10.2174/156802611795165098>
76. Ceramella, J., Iacopetta, D., Franchini, A., De Luca, M., Saturnino, C., Andreu, I., Sinicropi, M. S., & Catalano, A. (2022). A Look at the Importance of Chirality in Drug Activity: Some Significant Examples. *Applied Sciences (Switzerland)*, 12(21). Scopus. <https://doi.org/10.3390/app122110909>
77. Grogan, G. (2018). Synthesis of chiral amines using redox biocatalysis. *Curr. Opin. Chem. Biol.*, 43, 15-22. <https://doi.org/10.1016/j.cbpa.2017.09.008>
78. Patil, M. D., Grogan, G., Bommarius, A., & Yun, H. (2018). Oxidoreductase-catalyzed synthesis of chiral amines. *ACS Catal.*, 8, 10985-11015. <https://doi.org/10.1021/acscatal.8b02924>
79. Gomm, A., & O'Reilly, E. (2018). Transaminases for chiral amine synthesis. *Curr. Opin. Chem. Biol.*, 43, 106-112. <https://doi.org/10.1016/j.cbpa.2017.12.007>
80. Kelly, S. A., Mix, S., Moody, T. S., & Gilmore, B. F. (2020). Transaminases for industrial biocatalysis: Novel enzyme discovery. *Applied Microbiology and Biotechnology*, 104(11), 4781-4794. <https://doi.org/10.1007/s00253-020-10585-0>
81. Kelly, S. A., Pohle, S., Wharry, S., Mix, S., Allen, C. C. R., Moody, T. S., & Gilmore, B. F. (2018). Application of ω -Transaminases in the Pharmaceutical Industry. *Chemical Reviews*, 118(1), 349-367. <https://doi.org/10.1021/acs.chemrev.7b00437>
82. Fuchs, M., Farnberger, J. E., & Kroutil, W. (2015). The Industrial Age of Biocatalytic Transamination. *European Journal of Organic Chemistry*. <https://doi.org/10.1002/ejoc.201500852>
83. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W., & Hughes, G. J. (2010). Biocatalytic Asymmetric Synthesis of Chiral Amines from Ketones Applied to

- Sitagliptin Manufacture. *Science*, 329(5989), 305-309.
<https://doi.org/10.1126/science.1188934>
84. Desai, A. A. (2011). Sitagliptin Manufacture: A Compelling Tale of Green Chemistry, Process Intensification, and Industrial Asymmetric Catalysis. *Angewandte Chemie International Edition*, 50(9), 1974-1976. <https://doi.org/10.1002/anie.201007051>
85. Heckmann, C. M., Gourlay, L. J., Dominguez, B., & Paradisi, F. (2020). An (R)-Selective Transaminase From *Thermomyces stellatus*: Stabilizing the Tetrameric Form. *Frontiers in Bioengineering and Biotechnology*, 8. <https://doi.org/10.3389/fbioe.2020.00707>
86. Berglund, P., Humble, M. S., & Branneby, C. (2012). 7.18 C–X Bond Formation: Transaminases as Chiral Catalysts: Mechanism, Engineering, and Applications. En E. M. Carreira & H. Yamamoto (Eds.), *Comprehensive Chirality* (pp. 390-401). Elsevier. <https://doi.org/10.1016/B978-0-08-095167-6.00723-0>
87. Iglesias, C., Panizza, P., & Rodriguez Giordano, S. (2017). Identification, expression and characterization of an R - ω—Transaminase from *Capronia semiimmersa*. *Applied Microbiology and Biotechnology*, 101(14), 5677-5687. <https://doi.org/10.1007/s00253-017-8309-2>
88. Höhne, M., Schätzle, S., Jochens, H., Robins, K., & Bornscheuer, U. T. (2010). Rational assignment of key motifs for function guides in silico enzyme identification. *Nature Chemical Biology*, 6(11), 807-813. <https://doi.org/10.1038/nchembio.447>
89. Iglesias Santos, C. A. (2018). *Nuevas herramientas biocatalíticas para la síntesis asimétrica: Iminoreductasas, monoaminoxidasas y transaminasas* [Universidad de la República]. <https://www.colibri.udelar.edu.uy/jspui/handle/20.500.12008/32205>
90. Santillo, A. F., Gambini, J. P., Lannfelt, L., Långström, B., Ulla-Marja, L., Kilander, L., & Engler, H. (2011). In vivo imaging of astrocytosis in Alzheimer's disease: An 11C-L-deuteriodiprenyl and PIB PET study. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(12), 2202-2208. <https://doi.org/10.1007/s00259-011-1895-9>
91. Zirbesegger, K., Buccino, P., Kreimerman, I., Engler, H., Porcal, W., & Savio, E. (2017). An efficient preparation of labelling precursor of [11C]L-diprenyl-D2 and automated radiosynthesis. *EJNMMI Radiopharmacy and Chemistry*. <https://doi.org/10.1186/s41181-017-0029-5>

92. Nesmachnow, S., & Iturriaga, S. (2019). Cluster-UY: Collaborative Scientific High Performance Computing in Uruguay. En M. Torres & J. Klapp (Eds.), *Supercomputing* (pp. 188-202). Springer International Publishing. https://doi.org/10.1007/978-3-030-38043-4_16
93. *Barcelona Supercomputing Center*. (s. f.). BSC-CNS. Recuperado 15 de septiembre de 2023, de <https://www.bsc.es/marenostrum/marenostrum/technical-information>
94. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/s0022-2836\(05\)80360-2](https://doi.org/10.1016/s0022-2836(05)80360-2)
95. wwPDB consortium. (2019). Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Research*, 47(D1), D520-D528. <https://doi.org/10.1093/nar/gky949>
96. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), Article 6. <https://doi.org/10.1038/s41592-022-01488-1>
97. Venhorst, J., ter Laak, A. M., Meijer, M., van de Wetering, I., Commandeur, J. N. M., Rooseboom, M., & Vermeulen, N. P. E. (2003). Modeling and molecular dynamics of glutamine transaminase K/cysteine conjugate β -lyase. *Journal of Molecular Graphics and Modelling*, 22(1), 55-70. [https://doi.org/10.1016/s1093-3263\(03\)00136-0](https://doi.org/10.1016/s1093-3263(03)00136-0)
98. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., ... Baker, N. A. (2018). Improvements to the APBS biomolecular solvation software suite. *Protein Science*, 27(1), 112-128. <https://doi.org/10.1002/pro.3280>
99. Sternke, M., Tripp, K. W., & Barrick, D. (2020). The use of consensus sequence information to engineer stability and activity in proteins. En *Methods in Enzymology* (Vol. 643, pp. 149-179). Elsevier. <https://doi.org/10.1016/bs.mie.2020.06.001>
100. *Geneious* (2023.1). (s. f.). [Software]. Biomatters. <https://www.geneious.com>
101. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460-2461. <https://doi.org/10.1093/bioinformatics/btq461>

102. Telzerow, A., Paris, J., Håkansson, M., González-Sabín, J., Ríos-Lombardía, N., Gröger, H., Morís, F., Schürmann, M., Schwab, H., & Steiner, K. (2021). Expanding the Toolbox of R-Selective Amine Transaminases by Identification and Characterization of New Members. *ChemBioChem*, 22(7), 1232-1242. <https://doi.org/10.1002/cbic.202000692>
103. *Grep*. (s. f.). [Software]. Free Software Foundation.
<https://www.gnu.org/software/grep/manual/grep.html>
104. Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792-1797.
<https://doi.org/10.1093/nar/gkh340>
105. Kumar, S., Stecher, G., Li, M., Niyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6), 1547-1549. <https://doi.org/10.1093/molbev/msy096>
106. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Research*, 44(W1), W344-W350.
<https://doi.org/10.1093/nar/gkw408>
107. Wijma, H. J., Fürst, M. J. L. J., & Janssen, D. B. (2018). A Computational Library Design Protocol for Rapid Improvement of Protein Stability: FRESCO. En U. T. Bornscheuer & M. Höhne (Eds.), *Protein Engineering: Methods and Protocols* (pp. 69-85). Springer. https://doi.org/10.1007/978-1-4939-7366-8_5
108. Sayer, C., Martinez-Torres, R. J., Richter, N., Isupov, M. N., Hailes, H. C., Littlechild, J. A., & Ward, J. M. (2014). The substrate specificity, enantioselectivity and structure of the (R)-selective amine: Pyruvate transaminase from *Nectria haematococca*. *The FEBS Journal*, 281(9), 2240-2253. <https://doi.org/10.1111/febs.12778>
109. Łyskowski, A., Gruber, C., Steinkellner, G., Schürmann, M., Schwab, H., Gruber, K., & Steiner, K. (2014). Crystal Structure of an (R)-Selective ω -Transaminase from *Aspergillus terreus*. *PLOS ONE*, 9(1), e87350. <https://doi.org/10.1371/journal.pone.0087350>
110. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G. R., Wang, J., Cong, Q., Kinch, L. N., Schaeffer, R. D., Millán, C., Park, H., Adams, C., Glassman, C. R., DeGiovanni, A., Pereira, J. H., Rodrigues, A. V., Dijk, A. A. van, Ebrecht, A. C., ...

- Baker, D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*. <https://doi.org/10.1126/science.abj8754>
111. Huang, J., Xie, D.-F., & Feng, Y. (2017). Engineering thermostable (R)-selective amine transaminase from *Aspergillus terreus* through in silico design employing B-factor and folding free energy calculations. *Biochemical and Biophysical Research Communications*, 483(1), 397-402. <https://doi.org/10.1016/j.bbrc.2016.12.131>
112. Xie, D.-F., Fang, H., Mei, J.-Q., Gong, J.-Y., Wang, H.-P., Shen, X.-Y., Huang, J., & Mei, L.-H. (2018). Improving thermostability of (R)-selective amine transaminase from *Aspergillus terreus* through introduction of disulfide bonds. *Biotechnology and Applied Biochemistry*, 65(2), 255-262. <https://doi.org/10.1002/bab.1572>
113. Xie, D.-F., Yang, J.-X., Lv, C.-J., Mei, J.-Q., Wang, H.-P., Hu, S., Zhao, W.-R., Cao, J.-R., Tu, J.-L., Huang, J., & Mei, L.-H. (2019). Construction of stabilized (R)-selective amine transaminase from *Aspergillus terreus* by consensus mutagenesis. *Journal of Biotechnology*, 293, 8-16. <https://doi.org/10.1016/j.jbiotec.2019.01.007>
114. Zhu, W.-L., Hu, S., Lv, C.-J., Zhao, W.-R., Wang, H.-P., Mei, J.-Q., Mei, L.-H., & Huang, J. (2019). A Single Mutation Increases the Thermostability and Activity of *Aspergillus terreus* Amine Transaminase. *Molecules*, 24(7), Article 7. <https://doi.org/10.3390/molecules24071194>
115. Cao, J.-R., Fan, F.-F., Lv, C.-J., Wang, H.-P., Li, Y., Hu, S., Zhao, W.-R., Chen, H.-B., Huang, J., & Mei, L.-H. (2021). Improving the Thermostability and Activity of Transaminase From *Aspergillus terreus* by Charge-Charge Interaction. *Frontiers in Chemistry*, 9, 224. <https://doi.org/10.3389/fchem.2021.664156>
116. Liu, C.-Y., Cecylia Severin, L., Lyu, C.-J., Zhu, W.-L., Wang, H.-P., Jiang, C.-J., Mei, L.-H., Liu, H.-G., & Huang, J. (2021). Improving thermostability of (R)-selective amine transaminase from *Aspergillus terreus* by evolutionary coupling saturation mutagenesis. *Biochemical Engineering Journal*, 167, 107926. <https://doi.org/10.1016/j.bej.2021.107926>
117. Nnyigide, O. S., Lee, S.-G., & Hyun, K. (2018). Exploring the differences and similarities between urea and thermally driven denaturation of bovine serum albumin: Intermolecular forces and solvation preferences. *Journal of Molecular Modeling*, 24(3), 75. <https://doi.org/10.1007/s00894-018-3622-y>

118. Zhang, D., & Lazim, R. (2017). Application of conventional molecular dynamics simulation in evaluating the stability of apomyoglobin in urea solution. *Scientific Reports*, 7(February), 1-12. <https://doi.org/10.1038/srep44651>
119. Thompson, M. J., & Eisenberg, D. (1999). Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability¹ Edited by I. B. Honig. *Journal of Molecular Biology*, 290(2), 595-604. <https://doi.org/10.1006/jmbi.1999.2889>
120. Damnjanović, J., Nakano, H., & Iwasaki, Y. (2014). Deletion of a dynamic surface loop improves stability and changes kinetic behavior of phosphatidylinositol-synthesizing *Streptomyces* phospholipase D. *Biotechnology and Bioengineering*, 111(4), 674-682. <https://doi.org/10.1002/bit.25149>
121. Guan, L.-J., Ohtsuka, J., Okai, M., Miyakawa, T., Mase, T., Zhi, Y., Hou, F., Ito, N., Iwasaki, A., Yasohara, Y., & Tanokura, M. (2015). A new target region for changing the substrate specificity of amine transaminases. *Scientific Reports*, 5(1), 10753. <https://doi.org/10.1038/srep10753>
122. Bezsudnova, E. Yu., Popov, V. O., & Boyko, K. M. (2020). Structural insight into the substrate specificity of PLP fold type IV transaminases. *Applied Microbiology and Biotechnology*, 104(6), 2343-2357. <https://doi.org/10.1007/s00253-020-10369-6>
123. Börner, T., Rämisch, S., Reddem, E. R., Bartsch, S., Vogel, A., Thunnissen, A.-M. W. H., Adlercreutz, P., & Grey, C. (2017). Explaining Operational Instability of Amine Transaminases: Substrate-Induced Inactivation Mechanism and Influence of Quaternary Structure on Enzyme–Cofactor Intermediate Stability. *ACS Catalysis*, 7(2), 1259-1269. <https://doi.org/10.1021/acscatal.6b02100>
124. Ruggieri, F., Campillo-Brocal, J. C., Chen, S., Humble, M. S., Walse, B., Logan, D. T., & Berglund, P. (2019). Insight into the dimer dissociation process of the *Chromobacterium violaceum* (S)-selective amine transaminase. *Scientific Reports*, 9(1), Article 1. <https://doi.org/10.1038/s41598-019-53177-3>
125. Akdel, M., Pires, D. E. V., Pardo, E. P., Jänes, J., Zalevsky, A. O., Mészáros, B., Bryant, P., Good, L. L., Laskowski, R. A., Pozzati, G., Shenoy, A., Zhu, W., Kundrotas, P., Serra, V. R., Rodrigues, C. H. M., Dunham, A. S., Burke, D., Borkakoti, N., Velankar, S., ... Beltrao, P. (2022). A structural biology community assessment of AlphaFold2

- applications. *Nature Structural & Molecular Biology*, 29(11), Article 11.
<https://doi.org/10.1038/s41594-022-00849-w>
126. Ozvoldik, K., Stockner, T., Rammner, B., & Krieger, E. (2021). Assembly of Biomolecular Gigastructures and Visualization with the Vulkan Graphics API. *Journal of Chemical Information and Modeling*, 61(10), 5293-5303.
<https://doi.org/10.1021/acs.jcim.1c00743>
127. Krieger, E., & Vriend, G. (2015). New ways to boost molecular dynamics simulations. *Journal of Computational Chemistry*, 36(13), 996-1007. <https://doi.org/10.1002/jcc.23899>
128. Krieger, E., Dunbrack, R. L., Hooft, R. W. W., & Krieger, B. (2012). Assignment of Protonation States in Proteins and Ligands: Combining pKa Prediction with Hydrogen Bonding Network Optimization. En R. Baron (Ed.), *Computational Drug Discovery and Design* (pp. 405-421). Springer. https://doi.org/10.1007/978-1-61779-465-0_25
129. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9), 1157-1174. <https://doi.org/10.1002/jcc.20035>
130. Jakalian, A., Jack, D. B., & Bayly, C. I. (2002). Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16), 1623-1641. <https://doi.org/10.1002/jcc.10128>
131. Konagurthu, A. S., Whisstock, J. C., Stuckey, P. J., & Lesk, A. M. (2006). MUSTANG: A multiple structural alignment algorithm. *Proteins: Structure, Function, and Bioinformatics*, 64(3), 559-574. <https://doi.org/10.1002/prot.20921>
132. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., & Serrano, L. (2005). The FoldX web server: An online force field. *Nucleic Acids Research*, 33(Web Server), W382-W388. <https://doi.org/10.1093/nar/gki387>
133. Kellogg, E. H., Leaver-Fay, A., & Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*, 79(3), 830-838.
<https://doi.org/10.1002/prot.22921>

134. Wijma, H. J., Floor, R. J., Jekel, P. A., Baker, D., Marrink, S. J., & Janssen, D. B. (2014). Computationally designed libraries for rapid enzyme stabilization. *Protein Engineering, Design and Selection*, 27(2), 49-58. <https://doi.org/10.1093/protein/gzt061>
135. Bednar, D., Beerens, K., Sebestova, E., Bendl, J., Khare, S., Chaloupkova, R., Prokop, Z., Brezovsky, J., Baker, D., & Damborsky, J. (2015). FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLOS Computational Biology*, 11(11), e1004556. <https://doi.org/10.1371/journal.pcbi.1004556>
136. Choi, E. J., & Mayo, S. L. (2006). Generation and analysis of proline mutants in protein G. *Protein Engineering, Design and Selection*, 19(6), 285-289. <https://doi.org/10.1093/protein/gzl007>
137. Usmanova, D. R., Bogatyreva, N. S., Ariño Bernad, J., Eremina, A. A., Gorshkova, A. A., Kanevskiy, G. M., Lonishin, L. R., Meister, A. V., Yakupova, A. G., Kondrashov, F. A., & Ivankov, D. N. (2018). Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation. *Bioinformatics*, 34(21), 3653-3658. <https://doi.org/10.1093/bioinformatics/bty340>
138. Sambrook, J., & Russell, D. W. (2001). *Molecular cloning: A laboratory manual* (3rd ed). Cold Spring Harbor Laboratory Press.
139. Zheng, L., Baumann, U., & Reymond, J.-L. (2004). An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 32(14), e115. <https://doi.org/10.1093/nar/gnh110>
140. *QuikChange / Site directed mutagenesis kit*. (s. f.). Recuperado 23 de septiembre de 2023, de <https://www.agilent.com/en/product/mutagenesis-cloning/mutagenesis-kits/site-directed-mutagenesis-kits/quikchange-233116#literature>
141. Fürst, M. J. L. J., Martin, C., Lončar, N., & Fraaije, M. W. (2018). Chapter Seven— Experimental Protocols for Generating Focused Mutant Libraries and Screening for Thermostable Proteins. En N. Scrutton (Ed.), *Methods in Enzymology* (Vol. 608, pp. 151-187). Academic Press. <https://doi.org/10.1016/bs.mie.2018.04.007>
142. Sun, D., Ostermaier, M. K., Heydenreich, F. M., Mayer, D., Jaussi, R., Standfuss, J., & Veprintsev, D. B. (2013). AAscan, PCRdesign and MutantChecker: A Suite of Programs

- for Primer Design and Sequence Analysis for High-Throughput Scanning Mutagenesis. *PLOS ONE*, 8(10), e78878. <https://doi.org/10.1371/journal.pone.0078878>
143. Hansson, M. D., Rzeznicza, K., Rosenbäck, M., Hansson, M., & Sirijovski, N. (2008). PCR-mediated deletion of plasmid DNA. *Analytical Biochemistry*, 375(2), 373-375. <https://doi.org/10.1016/j.ab.2007.12.005>
144. Schätzle, S., Höhne, M., Redestad, E., Robins, K., & Bornscheuer, U. T. (2009). Rapid and Sensitive Kinetic Assay for Characterization of ω -Transaminases. *Analytical Chemistry*, 81(19), 8244-8248. <https://doi.org/10.1021/ac901640q>
145. Agilent Technologies. (2023). *QuikChange Site-Directed Mutagenesis Kit Instruction Manual*.
146. Jones, B. J., Lim, H. Y., Huang, J., & Kazlauskas, R. J. (2017). Comparison of Five Protein Engineering Strategies for Stabilizing an α/β -Hydrolase. *Biochemistry*, 56(50), 6521-6532. <https://doi.org/10.1021/acs.biochem.7b00571>
147. Sullivan, B. J., Nguyen, T., Durani, V., Mathur, D., Rojas, S., Thomas, M., Syu, T., & Magliery, T. J. (2012). Stabilizing Proteins from Sequence Statistics: The Interplay of Conservation and Correlation in Triosephosphate Isomerase Stability. *Journal of Molecular Biology*, 420(4), 384-399. <https://doi.org/10.1016/j.jmb.2012.04.025>
148. Carlin, D. A., Hapig-Ward, S., Chan, B. W., Damrau, N., Riley, M., Caster, R. W., Bethards, B., & Siegel, J. B. (2017). Thermal stability and kinetic constants for 129 variants of a family 1 glycoside hydrolase reveal that enzyme activity and stability can be separately designed. *PLOS ONE*, 12(5), e0176255. <https://doi.org/10.1371/journal.pone.0176255>
149. Huang, P., Chu, S. K. S., Frizzo, H. N., Connolly, M. P., Caster, R. W., & Siegel, J. B. (2020). Evaluating Protein Engineering Thermostability Prediction Tools Using an Independently Generated Dataset. *ACS Omega*, 5(12), 6487-6493. <https://doi.org/10.1021/acsomega.9b04105>
150. Patsch, D., Eichenberger, M., Voss, M., Bornscheuer, U. T., & Buller, R. M. (2023). LibGENiE – A bioinformatic pipeline for the design of information-enriched enzyme libraries. *Computational and Structural Biotechnology Journal*, 21, 4488-4496. <https://doi.org/10.1016/j.csbj.2023.09.013>

151. Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. J., & Pardo, L. (2000). Serine and Threonine Residues Bend α -Helices in the $\chi_1 = g^-$ Conformation. *Biophysical Journal*, 79(5), 2754-2760. [https://doi.org/10.1016/S0006-3495\(00\)76514-3](https://doi.org/10.1016/S0006-3495(00)76514-3)
152. Merz, L. M., van Langen, L. M., & Berglund, P. (2022). The Role of Buffer, Pyridoxal 5'-Phosphate and Light on the Stability of the *Silicibacter Pomeroyi* Transaminase. *ChemCatChem*. <https://doi.org/10.1002/cctc.202201174>
153. Stella, G. (2023). Sharing Failure as a Graduate Student. *ChemBioChem*, n/a(n/a), e202300104. <https://doi.org/10.1002/cbic.202300104>
154. Voth, G. A. (Ed.). (2009). *Coarse-graining of condensed phase and biomolecular systems*. Taylor & Francis.