



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

ANÁLISIS DE LOS CONTACTOS ENTRE AMINOÁCIDOS PARA IDENTIFICAR
MODULADORES DE LA FUNCIÓN FUERA DEL SITIO CATALÍTICO DE
ENZIMAS: EL CASO DE HIDRÓLISIS Y TRANSGLICOSIDACIÓN EN LA
FAMILIA DE LAS GLICÓSIDO HIDROLASAS GH13

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

RODRIGO ALEJANDRO ARREOLA BARROSO

DRA. GLORIA SAAB RINCÓN
[Instituto de Biotecnología, UNAM](#)

DR. LORENZO PATRICK SEGOVIA FORCELLA
[Instituto de Biotecnología, UNAM](#)

DRA. CARMEN NINA PASTOR COLÓN
[CIDC - UAEM](#)

Cuernavaca, Morelos

febrero, 2024



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

AGRADECIMIENTOS

Agradezco a mi tutora, la Dra. Gloria Saab Rincón, por su confianza, apoyo y paciencia durante el desarrollo de este proyecto y la escritura de la tesis, quedo en deuda con usted. Esta aventura no hubiera sido posible sin usted.

Agradezco a la Dra. Nina Pastor Colón y al Dr. Lorenzo Segovia Forcella, miembros de mi comité tutor, por sus valiosas opiniones y guía durante el desarrollo del proyecto.

Agradezco a mi jurado de los exámenes de candidatura y de grado por sus valiosos comentarios que enriquecieron y mejoraron la calidad de esta tesis.

Agradezco a la Ing. Leticia Olvera Rodríguez por su apoyo técnico durante la construcción de los mutantes y la purificación de los mismos.

Agradezco al Dr. Alexey Llopiz Arzuaga por la realización de las simulaciones de dinámica molecular.

Agradezco Dr. Sabino Pacheco Guillen por su asesoría de la técnica de quick change durante la construcción de variantes.

Agradezco al M. en T.I. Juan Manuel Hurtado Ramírez por su apoyo en la instalación de software para la realización de esta tesis

Agradezco a B.A. Dip.Lib. A.L.A. Shirley Elizabeth Ainsworth Gore por su servicio para conseguir de forma rápida y expedita múltiples artículos indispensables para esta investigación.

Agradezco el apoyo los proyectos DGAPA PAPIIT IN-226623 e IN-211020 por el financiamiento para comprar los insumos para este proyecto.

Agradezco a la beca 443241 otorgada por CONACYT, que pagó mi manutención durante el desarrollo del proyecto.

Agradezco al Programa de Apoyo a los Estudios de Posgrado (PAEP) de la UNAM por su financiamiento para asistir a congresos nacionales e internacionales.

Agradezco al Instituto de Biotecnología-UNAM por el acceso a su cluster de computadoras.

DEDICATORIA

A mis padres, Lucio y Lupita, que me han apoyado estando siempre a mi lado aunque estemos lejos.

A mis hermanos, Lucio y Adriana, que han sido mis compañeros desde el inicio.

A Astron Martínez Rosas por acompañarme en este proceso.

Al pueblo de México, cuyos impuestos hicieron posible mi educación superior en la UNAM en la licenciatura y posgrado.

Contenido

RESUMEN	6
1. INTRODUCCIÓN	7
2. HIPÓTESIS	12
3. OBJETIVOS	13
4. MATERIALES Y MÉTODOS	13
4.1. Análisis bioinformático del enriquecimiento diferencial de residuos entre miembros de la familia GH13 con función hidrolítica y transglicosídica	14
4.1.1. Modelo de homología de la enzima <i>TmAmyA</i>	14
4.1.2. Cálculo de factores de enriquecimiento por residuo	14
4.1.3. Clasificación funcional de enzimas	15
4.2. Análisis bioinformático de la conservación de los contactos entre residuos en la familia de enzimas glicósido hidrolasas GH13	16
4.2.1. Determinación de la conservación de los contactos en miembros de la familia de enzimas glicósido hidrolasas GH13	16
4.2.2. Clasificación evolutiva de enzimas mediante la conservación de sus contactos	16
4.3. Investigación del efecto de cambiar residuos en miembros de la familia GH13 con base en los factores de enriquecimiento de los aminoácidos por medio de estudios bioinformáticos y experimentales	17
4.3.1 Análisis bioinformático	17
4.3.2. Estudio experimental para evaluar las mutaciones seleccionadas con los factores de enriquecimiento	18
4.3.3. Simulaciones dinámicas moleculares	21
5. RESULTADOS	22
5.1 Modelo de homología de <i>TmAmyA</i>	22
5.2 Cálculo de factores de enriquecimiento por residuo	24
5.3 Clasificación de proteínas como transglucosidasas o hidrolasas	24
5.4 Determinación de la conservación de los contactos en miembros de la familia de enzimas glicósido hidrolasas GH13	26
5.5 Clasificación evolutiva de enzimas mediante la conservación de sus contactos	27
5.6 Selección de los sitios de mutación.	31

5.7	Caracterización de las variantes de <i>TmAmyA</i>	33
5.8	Caracterización de las variantes de <i>TmGTasa</i>	36
6.	DISCUSIÓN	42
7.	CONCLUSIONES	45
8.	PERSPECTIVAS.....	46
9.	REFERENCIAS BIBLIOGRÁFICAS Y BIBLIOGRAFÍA	47
	Anexo I. Estructuras empleadas en este trabajo	55
	Anexo II Oligonucleótidos empleados en este trabajo.....	69
	Anexo III. Resultados de simulaciones de dinámica molecular.....	70

RESUMEN

En el presente trabajo se desarrolló de una metodología que permitiera encontrar residuos fuera del sitio catalítico y que modificaran la selectividad Transglicosidación/Hidrólisis (T/H) sobre el enlace 1,4- α -glucosídico, aplicándolo en el caso de la familia de enzimas glicósido hidrolasa GH13. Con este último propósito, seleccionamos estructuras del *Protein Data Bank* (PDB), correspondientes a miembros de la familia GH13, con actividad transglicosídica e hidrolítica. Además, usamos el programa CMView, al igual que un programa escrito en R, y comparamos los contactos entre residuos de las hidrolasas contra las transglicosidasas.

Ahora bien, este análisis nos llevó a identificar residuos diferencialmente enriquecidos en los dos grupos, lo que nos permitió proponer sustituciones de residuos en una hidrolasa (la α -amilasa de *Thermotoga marítima* AmyA, *TmAmyA*) y una transferasa (1,4- α -glucanotransferasa de *Thermotoga marítima*, *TmGTasa*). Usando mutagénesis-sitio-dirigida, construimos los vectores que contenían las variantes propuestas y evaluamos la actividad transglicosídica e hidrolítica tanto de las enzimas silvestres como de las variantes sugeridas por nuestra metodología. Con las mutaciones seleccionadas, logramos aumentar la selectividad T/H en la hidrolasa *TmAmyA* y disminuirla en la transferasa *TmGTasa*. En el caso de la variante *TmGTasa* M279N, logramos reducir la actividad transglicosídica y medimos un incremento en la hidrólisis. A pesar de que, en la mayoría de las variantes presentadas, el cambio de selectividad T/H se obtuvo por pérdida de la actividad principal de cada enzima, mientras la actividad secundaria permaneció intacta o se redujo en menor medida, en todos los casos, el cambio de T/H se obtuvo en la dirección deseada.

De esta suerte, concluimos que el análisis de contactos entre residuos basado en mapas de contacto permite identificar sitios blanco de mutagénesis, fuera del sitio catalítico que modulan la relación T/H en miembros de la familia GH13, tanto en transferasas como en hidrolasas. Asimismo, el análisis de los contactos entre residuos permitió clasificar a las enzimas de la familia GH13, de acuerdo con su selectividad como hidrolasas o transferasas, y definir un parámetro que refleja las relaciones evolutivas de la familia GH13, permitiendo, a su vez, identificar eventos de duplicación y transferencia horizontal de genes, siendo estos parámetros más sensibles que el análisis por secuencias, además de poder ser automatizados.

1. INTRODUCCIÓN

Las enzimas son proteínas que aceleran una infinidad de reacciones *in vivo* e *in vitro* [1–4], cada una con muy alta especificidad de reacción y de sustrato. Estas máquinas moleculares tienen un grupo de residuos de aminoácidos dentro de su secuencia, denominados residuos catalíticos, que reaccionan con el sustrato para llevar a cabo la catálisis [5].

Los residuos catalíticos en la función enzimática son importantes: eliminarlos mediante una sustitución por otro aminoácido lleva a la pérdida parcial o completa de la función enzimática, y enzimas homólogas catalizando la misma reacción posicionan y orientan sus residuos de la misma forma a pesar de diferencias en la secuencia [6]. Esta evidencia sugiere que su adición a cualquier otra estructura proteica con el mismo posicionamiento y orientación llevará a la misma función enzimática; sin embargo, parecen no ser el único elemento estructural que permite a las enzimas alcanzar sus altas tasas de conversión, ya que, para producir una enzima que catalizara la eliminación de Kemp (Kemp eliminasa) con actividad similar a las enzimas naturales [7], no bastó insertar los residuos catalíticos en un molde proteico, fue necesario sustituir residuos fuera del sitio catalítico, lo que cambió la posición de los residuos catalíticos y la dinámica de la proteína. Otros experimentos de evolución dirigida han encontrado, en bancos producidos al azar, sustituciones de residuos fuera del sitio catalítico que modifican su función [8,9].

Más evidencia de que la información de la función no está contenida únicamente en el sitio catalítico ha sido generada después de analizar los resultados de secuenciar de forma masiva (deep sequencing) bancos de variantes producidas al azar (cuya actividad puede aumentar hasta 30 veces) y relacionarla con la funcionalidad de cada mutante para generar la estructura terciaria de diversas enzimas al analizar el efecto de todas las mutaciones concurrentes [10,11]. Por otro lado, para la proteína Ras se ha construido un banco de mutantes sencillas sustituyendo, para todas las posiciones, el residuo original por los otros 19 en todas

las posiciones, observándose que hay posiciones fuera del sitio catalítico claves para controlar la velocidad de hidrólisis del GTP [12].

Considerando lo anterior, es necesario desarrollar una metodología que permita explotar el efecto de los residuos fuera del sitio catalítico en el diseño y rediseño de enzimas.

Una estrategia para recuperar la información funcional dispersa en la estructura de las enzimas podría ser el análisis de contactos entre residuos. Una prueba de esto es el algoritmo SCHEMA [13], que aumentó la tasa de éxito para crear quimeras estables y funcionales de enzimas evolutivamente emparentadas al considerar los contactos conservados en la familia de enzimas y realizar los cortes en sitios donde se minimice la ruptura de contactos compartidos por ambas enzimas para formar quimeras en las que los contactos de las enzimas donantes se mantienen.

También el análisis de la covariación de residuos, direct coupling analysis (DCA), en alineamientos múltiples de secuencias (MSA) ha permitido identificar residuos importantes para la función fuera del sitio activo de las enzimas. En estos experimentos, el análisis de múltiples secuencias de proteínas evolutivamente relacionadas identifica posiciones en la secuencia primaria que cambian de forma correlacionada, los cuales están en contacto en la estructura tridimensional; este método considera que cuando un residuo importante para la función muta, los residuos en contacto también cambian para compensar la primera sustitución y así mantener la función [14,15].

La relación entre los residuos que coevolucionan y la especificidad de la proteína se ha evaluado en la familia de las acuaporinas para encontrar los residuos asociados con el transporte de agua o glicerol [16]. Además, se evaluaron residuos cuya coevolución está asociada con la especificidad de la proteína en doce familias de proteínas [17]. Adicionalmente, al sustituir, mediante mutagénesis a saturación, aminoácidos que covarian dentro de una familia de proteínas, se ha observado que dichas sustituciones alteran la función enzimática [18,19]. Sin embargo, en muchos casos, la coevolución de los residuos es difícil de detectar debido a (1) la presencia de cambios en las mutaciones compensatorias de la red más allá de la primera esfera de contactos, (2) la dependencia significativa de las covariaciones con las

distancias evolutivas entre las secuencias usadas, (3) el número de proteínas en un MSA, (idealmente este número debe asegurar que la probabilidad de encontrar un aminoácido en cualquier posición -no conservada- del alineamiento, sea la misma que aquella de todo el universo de proteínas) y (4) la calidad del alineamiento en el entorno de los residuos que coevolucionan [15].

Además de la covariación de residuos, la importancia que tienen los contactos entre residuos para la función, se ve reflejada en su aprovechamiento para predecir la clasificación funcional de enzimas. Esto ha sido posible al comparar los patrones de empaquetamiento característicos de una familia y usarlos para identificar enzimas con la misma función [20,21].

Por otra parte, programas de diseño y predicción de estructura de proteínas se han beneficiado al incluir en su algoritmo a los contactos entre residuos. Por ejemplo, considerar los contactos entre residuos permitió mejorar los resultados del programa de diseño *de novo* de proteínas Rosetta [22,23] y más recientemente al usar inteligencia artificial para analizar las redes de contactos entre residuos [24]. En este mismo sentido, el programa AlphaFold2 superó al resto de programas de predicción de estructuras en el concurso CASP14 al incluir una inteligencia artificial que analiza simultáneamente la información evolutiva y espacial, acoplando los modelos tridimensionales al alineamiento de secuencias mediante un mapa de contactos durante su proceso inductivo-deductivo, permitiendo mejorar su modelado [25,26].

Un mapa de contactos es una representación 2D de la estructura terciaria de una proteína que permite visualizar los ~~Sus~~ contactos entre sus residuos; consisten en una matriz de $N \times N$, donde N corresponde al número de aminoácidos que componen la proteína, que registra, de forma binaria, si existen o no contactos entre ellos. Los mapas de contactos registran la estructura secundaria de las proteínas, hélice α o lámina β , en patrones característicos. La información almacenada en estos mapas, los ha hecho muy importantes para un gran número de metodologías de predicción de estructura de proteínas [27], en las que se da énfasis en la capacidad de producirlos y alinearlos, para obtener los correspondientes a proteínas desconocidas a partir de su estructura primaria [28,29]. El análisis y comparación

de los mapas de contactos también han sido empleados en métodos para la predicción de contactos proteína-proteína [30]. Además, han encontrado aplicación para la clasificación topológica y funcional de proteínas [21,31,32].

La familia de enzimas glicósido hidrolasas GH13 nos pareció de interés para la búsqueda de residuos importantes para la selectividad en la transglicosidación/hidrólisis (T/H) a través del análisis de mapas de contactos. Esta es la familia de enzimas glicósido hidrolasas mejor caracterizada dentro de la base de datos CAZy (base de datos de enzimas activas de carbohidratos) [33,34]; además, a pesar de que sus miembros comparten tres dominios [35] y su maquinaria y mecanismos catalíticos [36], algunos de estos llevan a cabo la reacción de hidrólisis de enlaces α -(1→4)-glucosídicos en polímeros y oligómeros de glucosa, mientras otros catalizan la reacción de transglicosidación en las mismas moléculas. Por estas razones, el análisis comparativo global de homólogos que incluyan ambas funciones podría llevar a la producción de herramientas para la glicosidación de moléculas orgánicas, un área activa de investigación [37–41]. Sin embargo, su análisis usando únicamente sus secuencias se dificulta, ya que, existe un muestreo discontinuo de su espacio de secuencias resultando en miembros de esta familia sin homólogos con identidad de secuencia >40%, a pesar de su similitud estructural con otras proteínas dentro de la familia (lo que dificulta analizar el patrón de coevolución de los residuos de estas proteínas y; además a nivel secuencia solo siete segmentos cortos de aminoácidos se conservan en todas las GH13, en los cuales se ha visto un enriquecimiento diferencial de residuos de acuerdo a su función [42,43]. Además, al igual que para otros grupos de enzimas la evolución dirigida ha revelado sitios fuera del sitio catalítico y de los segmentos de aminoácidos conservados en la familia de enzimas glicósido hidrolasas GH13 que cambian la actividad, por ejemplo: para la isozima 2 de α -amilasa de la cebada se obtuvo una variante 1000 más activa que la silvestre con 5 mutaciones, 2 de las cuales estaban en el dominio de unión a almidón [44]; para la α -amilasa de *Bacillus licheniformis*, una mutación en el dominio A mejoró la actividad de la enzima medida en U/mg un 62 % [45]; la CGTasa de *Bacillus* sp. I-15 logró con 3 mutaciones, una de ellas en el dominio de unión a almidón, reducir la actividad de ciclización 10

veces, mientras aumentó su actividad hidrolítica 15 veces más [46]; por mencionar algunos [47].

Para modificar la proporción de transglicosidación/hidrólisis (T/H) de las enzimas amilolíticas se han utilizado múltiples estrategias [48–51]. Los factores considerados en la selección de aminoácidos a modificar para este propósito incluyen la ubicación de regiones flexibles asociadas con el transporte interno de agua [52,53], cambios en la dinámica ácido-base de residuos [53] y cambios conformacionales de cadena lateral de residuos cerca del centro activo [53,54]. Otras investigaciones han pretendido modificar la especificidad del sitio aceptor hacia diversas moléculas orgánicas [48] y han logrado cambiar la relación T/H; sin embargo, las estrategias exploradas generalmente se han centrado en mutaciones puntuales cerca del sitio activo, identificado por MSA de algunas glucosidasas, presumiblemente con alta especificidad hacia reacciones hidrolíticas o de transferencia [48,49,51,55,56].

Este trabajo analiza los mapas de contacto de miembros de la familia GH13 con herramientas bioinformáticas, comparando enzimas de dos grupos de esta familia con selectividad diferente (hidrolíticas y transglicosídicas), para evaluar si a través del proceso evolutivo se han diferenciado residuos en contacto, enriqueciéndose con aminoácidos de acuerdo a la selectividad de la enzima (hidrolíticas o transglicosídicas). Asimismo, estudiamos si la conservación en sus contactos sigue algún patrón evolutivo. Para valorar si el enriquecimiento de aminoácidos en los contactos de ambos grupos puede ser usado para cambiar la selectividad T/H, con el análisis anterior elegimos mutaciones en dos enzimas de *Thermotoga maritima*, la α -amilasa *TmAmyA* (enzima hidrolítica) y la glucanotransferasa *TmGTasa* (enzima transglicosídica) -que se clonó como parte del trabajo de esta tesis. En el caso de la enzima TmGTasa, para ayudarnos a seleccionar el sitio a mutagenizar, junto con el factor de enriquecimiento, usamos el parámetro de centralidad de intermediación que ha sido reportado como una medida de la importancia de los residuos de aminoácidos en la estructura y función de las proteínas [57]. Medimos el efecto de estas mutaciones en la selectividad T/H después de construir las variantes correspondientes e intentamos racionalizar sus efectos a través de una simulación de dinámica molecular.

Encontramos patrones de conservación y enriquecimiento diferencial de aminoácidos de los contactos entre residuos, los cuales nos permitieron clasificar las enzimas de la familia GH13 de acuerdo a su función y su filogenia, además de facilitarnos la elección de mutaciones que aumentaron la selectividad T/H de la amilasa *TmAmyA* y disminuyeron la de la glucanotransferasa *TmGTasa* a través de mutaciones fuera del sitio catalítico. Aunque estos cambios se deben, principalmente, a la pérdida de actividad hidrolítica en la hidrolasa y de la actividad transglicosídica en la transferasa; de entre las variantes construidas para la *TmGTasa* M279N, observamos la actividad transglucosídica disminuir a un cuarto de la silvestre, y la actividad hidrolítica deseada aumentar en un 25%, resultando en un incremento de cinco veces en la relación H/T con respecto a la enzima silvestre. La comprensión de los determinantes moleculares de la especificidad de las enzimas GH13 contribuirá al desarrollo de enzimas para la glicosíntesis y el diseño de enzimas con mayor capacidad hidrolítica o transglucosídica. De igual manera, el advenimiento de Alpha-Fold2 permitirá el aumento de los modelos estructurales de proteínas para las que solo tenemos la secuencia de aminoácidos como nunca antes, con una confiabilidad similar a la obtenida por la difracción de cristales con rayos Xs, una vez que mejore sus predicciones de las cadenas laterales [58], este hecho nos anima a explorar las estructuras de rayos X actualmente accesibles para comparar e identificar la información que, posteriormente, podremos extraer con Alpha-Fold2.

2. HIPÓTESIS

Al analizar los contactos entre residuos en la familia de enzimas glicósido hidrolasa GH13, contrastando grupos con selectividad por la hidrólisis y la transglucosidación, se podrán identificar contactos cuyos residuos han sido diferencialmente seleccionados por la evolución para realizar su función en cada grupo analizado y podremos usar esta información para modificar su selectividad T/H.

3. OBJETIVOS

- Comparar los contactos entre residuos de enzimas de la familia GH13 con las funciones hidrolítica y transglicosídica, para determinar si contactos equivalentes son ocupados de forma preferencial y diferenciada entre los grupos.
- Analizar los patrones de conservación de contactos entre residuos en los miembros de la familia de las enzimas glicósido hidrolasas GH13.
- Proponer mutaciones encaminadas a cambiar la relación T/H de los miembros de la familia GH13, basadas en el análisis de enriquecimiento de residuos.
- Construir los mutantes propuestos y caracterizarlos.
- Determinar si las mutaciones propuestas tienen efecto sobre la relación de actividades T/H de las enzimas GH13 investigadas.

4. MATERIALES Y MÉTODOS

El presente trabajo tiene una etapa bioinformática y otra experimental. Analizamos con herramientas bioinformáticas los mapas de contactos, construidos a partir de modelos estructurales (experimentales y por homología), de enzimas de la familia GH13 para identificar patrones de conservación de los contactos entre residuos, así como preferencias de los aminoácidos para ocupar estos contactos de acuerdo a la función de las enzimas, como resultado de la selección ejercida por la evolución. Usamos esta información para proponer mutaciones que modificaran la selectividad Transglicosidación/Hidrólisis (T/H), aumentándola en la hidrolasa TmAmyA y disminuyéndola en la transferasa TmGTasa, clonada como parte de este trabajo (para la que también se usó el parámetro de centralidad de intermediación para reducir más el número de sitios a evaluar con el factor de enriquecimiento); se construyeron las variantes con las mutantes propuestas para caracterizar

experimentalmente su selectividad Transglicosidación/Hidrólisis (T/H). Finalmente, realizamos una simulación de dinámica molecular buscando racionalizar nuestros resultados.

4.1. Análisis bioinformático del enriquecimiento diferencial de residuos entre miembros de la familia GH13 con función hidrolítica y transglicosídica

4.1.1. Modelo de homología de la enzima *TmAmyA*

El modelo de homología de *TmAmyA* se construyó utilizando la estructura cristalográfica de la amilasa de *Thermotoga petrophila* (PDB ID 5M99, resolución 1,96 Å) como plantilla, con la que guarda 98,4 % de identidad. Para validar este modelo se utilizó la herramienta “Evaluación de estructuras” del servidor Swiss-Model [59,60].

4.1.2. Cálculo de factores de enriquecimiento por residuo

Con el programa CMView [61] creamos el mapa de contactos de 14 miembros de la familia CAZy GH13, 10 hidrolasas y 4 transferasas (Conjunto de datos 1, Tabla S2), cuyas estructuras tomamos del Protein Data Bank (PDB). Con el mismo programa alineamos éstos con el mapa de contactos de la 1,4- α -glucanotransferasa de *Thermotoga maritima* TmGTase (PDB: 1LWJ). Para este análisis, generamos los mapas de contacto únicamente a partir de estructuras unidas al análogo del estado de transición acarbosa, ya que esto asegura una estructura más parecida a la que lleva a cabo la catálisis. Escribimos un programa en R para identificar los contactos entre residuos compartidos por TmGTasa con el resto de las estructuras, registrando los aminoácidos ocupando estas posiciones. Con esta información, nuestro programa obtuvo la frecuencia con la que aparece cada aminoácido en los contactos equivalentes de un grupo de enzimas (f_{aa}^{ij}), ya sea hidrolasa, ($f_{aa}^{ij,H}$) o

transferasa, ($f_{aa}^{ij,T}$). A partir de dichas frecuencias, obtuvimos el factor de enriquecimiento (Δf_{aa}^{ij}) definido como:

$$\Delta f_{aa}^{ij} = f_{aa}^{ij,T} - f_{aa}^{ij,H}$$

4.1.3. Clasificación funcional de enzimas

Usando un programa escrito en el lenguaje de programación R para este fin (disponible en <https://github.com/rhodbacter/ContactMaps-Catalysis/>), clasificamos las 14 estructuras utilizadas en su obtención y un conjunto adicional de 14 enzimas, 12 hidrolasas y 2 transferasas (Conjunto de datos 2, Tabla S2), como transferasas o hidrolasas con estos factores de enriquecimiento, dependiendo del número de contactos hidrolíticos y transglicosídicos en cada enzima. Usamos el factor de enriquecimiento para calificar los residuos en cada contacto de las enzimas bajo análisis como hidrolítico ($\Delta f_{aa}^{ij} < 0$) o transglicosídico ($\Delta f_{aa}^{ij} > 0$), y con estos valores calificamos al contacto considerando la puntuación de ambos residuos en el contacto.

Sin embargo, al aparecer un número mayor de enzimas en las bases de datos, se calcularon los factores de enriquecimiento para todos los contactos en el modelo de homología de la α -amilasa AmyA de *Thermotoga marítima* (TmAmyA). Se obtuvieron los factores de enriquecimiento usando TmAmyA como referencia y 78 estructuras de enzimas GH13, 37 transferasas y 41 hidrolasas (Conjunto de datos 4, Tabla S4), que se clasificaron usando el procedimiento arriba descrito.

4.2. Análisis bioinformático de la conservación de los contactos entre residuos en la familia de enzimas glicósido hidrolasas GH13

4.2.1. Determinación de la conservación de los contactos en miembros de la familia de enzimas glicósido hidrolasas GH13

Tomamos del Protein Data Bank (PDB) 40 enzimas GH13 (con 10-30% de identidad entre sí después de alinear con Clustal Omega), una enzima GH97 (PDB: 2ZQ0) y una enzima GH31 (PDB: 3W37), todas actuando sobre enlaces α -(1,4)-glucosídicos, ya sea como hidrolasa o transferasa (Conjunto de datos 3, Tabla S3), después, con el programa CMView [61], obtuvimos los mapas de contacto de estas estructuras y los alineamos entre sí. Posteriormente, escribimos un programa en R para identificar los contactos entre cada par de estructuras, una vez comparadas todas las estructuras entre sí, comparamos cada enzima con el resto y calculamos, para cada contacto, la fracción de enzimas en las que pudimos identificar el contacto a través de los alineamientos, este cálculo se hizo a través de la siguiente fórmula:

$$\text{fracción de enzimas conservando el contacto } k = \frac{\text{Número de enzimas compartiendo el contacto } k \text{ con enzima } i}{\text{Número total de enzimas}}$$

4.2.2. Clasificación evolutiva de enzimas mediante la conservación de sus contactos

Después de calcular para cada enzima la fracción de enzimas que conservaban los contactos en su estructura, nuestro programa en lenguaje R, uso estos datos para calcular un coeficiente de conservación. Con este fin, en el par de enzimas i y j , el programa identificó los contactos que se encontraban tanto en i como en j ; para esta colección de contactos, C , con contactos individuales c_k , hizo la siguiente operación: gráfico la fracción de enzimas compartiendo c_k con i , contra este mismo valor para

c_k en j . Esto resultó en un gráfico de correlación para los contactos en C , donde un eje correspondía a la fracción de enzimas que compartían con i a c_k , y el otro a la fracción de enzimas que compartían con j a c_k . Al coeficiente de correlación del gráfico lo denominamos coeficiente de conservación porque de ser idénticas las enzimas, los contactos en sus estructuras de i y j también lo serían, y C correspondería a todos los contactos de ambas enzimas en el par y el coeficiente de correlación sería uno, pero al divergir las enzimas en su patrón de contactos, por efecto de la evolución, al tamaño de C disminuye porque es más difícil identificarlos como equivalentes tanto entre i y j , como entre cada enzima del par con las otras en el grupo. Comparamos la sensibilidad del coeficiente de conservación para distinguir la relación filogenética entre enzimas con la alineación de secuencias múltiples con ClustalW [62] y estructuras con DALI [63].

4.3. Investigación del efecto de cambiar residuos en miembros de la familia GH13 con base en los factores de enriquecimiento de los aminoácidos por medio de estudios bioinformáticos y experimentales

4.3.1 Análisis bioinformático

4.3.1.1 Selección de los sitios de mutación

Para evaluar si los factores de enriquecimiento permitirían seleccionar sustituciones de aminoácidos que cambiaran la selectividad en el sentido deseado elegimos dos sistemas de estudio: una hidrolasa (*TmAmyA*) y una transferasa (*TmGTasa*), en los cuales se buscó cambiar su selectividad T/H.

Para la enzima *TmAmyA* encontramos posiciones para sustituir y producir variantes con relación T/H aumentada con base en los resultados de calcular los factores de enriquecimiento por residuo. Los sitios de mutación se seleccionaron en función de su puntuación de conservación ($\geq 10/14$), factor de enriquecimiento (sustitución por un aminoácido con factor de enriquecimiento ≤ 0 por otro con un factor de

enriquecimiento $\geq 0,2$) y su distancia desde el sitio catalítico (se prefirieron los residuos que no estaban en contacto con el análogo del estado de transición).

Para la enzima *TmGTasa* elegimos posiciones para sustituir y producir variantes con relación T/H disminuida con base en los resultados de calcular los factores de enriquecimiento por pares de residuos. Los sitios de mutación se seleccionaron en función de la relación de enzimas que contiene un contacto y las totales, que definió la puntuación de conservación ($\geq 10/14$); el factor de enriquecimiento (contactos ocupados por pares de residuos con factor de enriquecimiento en el 1% más negativo, preferentes en hidrolasas, y el 1% más positivo, preferentes en transferasas) y su puntuación de centralidad por intermediación (20 % superior). Para calcular el valor de centralidad por intermediación (“betweenness centrality”, BC) transformamos el mapa de contactos de *TmGTasa* en una red tridimensional (V) en la que los nodos ocupaban la posición de los carbonos α de los residuos en su estructura y los bordes correspondían a los contactos entre residuos; a partir de esta red calculamos BC para cada nodo, representando un residuo, con la ecuación siguiente:

$$BC_x = \sum_{u,v \in V}^N \frac{\sigma_{u,v}(x)}{\sigma_{u,v}}$$

donde $\sigma_{u,v}$ representa el número de caminos que conectan los nodos u y v , es decir, las colecciones de nodos en contacto que hay que recorrer para ir del nodo u al v ; $\sigma_{u,v}(x)$ es el número de estos caminos que pasan por el nodo x , cuya BC se está calculando [64].

4.3.2. Estudio experimental para evaluar las mutaciones seleccionadas con los factores de enriquecimiento

4.3.2.1. Construcción del vector del gen *TmGTasa* (KEGG ID: Tm0364)

Con la finalidad de modificar una enzima transglicosídica de la familia GH13, construimos un vector de expresión conteniendo el gen Tm0364 (KEGG ID) se amplificó a partir de una biblioteca de ADNc de *T. maritima* mediante una reacción de PCR usando ADN polimerasa pfu, utilizando un par de oligonucleótidos que se unen en los extremos 5' y 3' del gen, y adicionaban los sitios de corte NdeI y Xho I

al gen. El producto de PCR se purificó mediante electroforesis en gel de agarosa y el kit de purificación de PCR de Roche (Roche Diagnostics GmbH, Alemania). El vector pET22a y el producto de PCR fueron digeridos usando las enzimas de restricción NdeI y Xho I. Posteriormente, los fragmentos para la construcción fueron purificados mediante electroforesis en gel de agarosa y el kit de purificación Roche PCR (Roche Diagnostics GmbH, Alemania), y finalmente ligados usando ADN ligasa T4 (ThermoFisher, EE. UU.). El producto de ligación se usó para transformarse en células electro competentes de *Escherichia coli* MC1061. Al insertar el gen Tm0364 en el vector pET22a adicionamos un segmento de DNA que codifica para un conector y seis histidinas en el extremo carboxilo terminal para la posterior purificación de la proteína en el extremo carboxilo terminal de esta.

4.3.2.2. Construcción de variantes de TmAmyA y TmGTase

Las mutaciones se construyeron a través del método Mega Primer (usando la DNA polimerasa PfuS6) [65] o Quick Change [66] (con DpnI y la enzima de alta fidelidad ThermoFisher Scientific phusion) usando oligonucleótidos (Tabla S5) comprados a la Unidad de Síntesis y Secuenciación de ADN del Instituto de Biotecnología, UNAM.

4.3.2.3. Expresión de variantes de TmAmyA y TmGTase

Todos los plásmidos pET22a que contenían los genes se transformaron en células *E. coli* K12 ER2738 competentes en calcio. La inducción se realizó con IPTG 0.5 mM (variantes de TmAmyA) o IPTG 0.1 mM (variantes de TmGTasa) cuando se alcanzó una OD₆₀₀ de 0.4-0.5. Los cultivos se centrifugaron por 10 minutos a 1500 g después de que superaron OD₆₀₀ de 1. Todas las muestras se purificaron en dos pasos después de la lisis celular por sonicación: (1) calentamiento a 70 °C durante 1 h seguido de centrifugación a 1500 g por 20 minutos y filtración a través de una membrana Merck de tamaño de poro de 0.2 µm y (2) cromatografía de afinidad del filtrado con columna de agarosa Ni-NTA (ThermoFisher Scientific, EE. UU.).

4.3.2.4. Caracterización de las variantes *TmAmyA*

Uno de los objetivos para aumentar la actividad transglucosídica en las glucósido hidrolasas es crear biocatalizadores glucosintéticos. Además de su impacto en la síntesis de alquil-glucósidos, la reacción de alcoholisis (es decir, la reacción de un azúcar con un alcohol para producir un alquil-glucósido) puede utilizarse como indicador de la reacción de transferencia a otros azúcares [49,67]. Se utilizó esta aproximación porque los productos de transglicosidación de *TmAmyA* no se acumulan significativamente debido a que la reacción de hidrólisis es la predominante [49,68] y a que la diferenciación de los productos de hidrólisis y transglicosidación requiere el uso de azúcares marcados con el isótopo radioactivo ^{14}C [68].

La reacción de alcoholisis se inició mediante la adición de 20 U de enzima a 1 mL de 100 mg/mL de almidón en butanol al 10 %, Tris 50 mM, NaCl 150 mM, CaCl_2 2 mM, pH 7,0, y se midió después de 12 h de reacción a 85 °C. Los productos de hidrólisis y transglicosidación se midieron como azúcares reductores mediante el reactivo DNS y como la formación de glucósido de butilo mediante HPLC después de la digestión con glucoamilasa (Sigma-Aldrich, EE. UU.), respectivamente. El análisis HPLC se realizó en un sistema HPLC Waters-Millipore 510 equipado con un muestreador automático (modelo 717 Plus, Waters Corp., EE. UU.), un detector de índice de refracción (Waters 410, Waters Corp., EE. UU.) y una columna de HPLC de aminoácidos Hypersil GOLD™ (Thermo Scientific, Reino Unido) que utiliza acetonitrilo:agua (80:20) como fase móvil a un flujo de 1,0 ml/min.

4.3.2.5 Caracterización de las variantes *TmGTase*

La actividad de transglicosidación se midió en las reacciones iniciadas por la adición de 12 µg de enzima a 1 mL de 5 mg/mL de maltoheptosa en Tris 50 mM, NaCl 150 mM, CaCl_2 2 mM pH 7,0 y mantenidas a 70 °C. Finalmente, la cuantificación del complejo de azúcar con reactivo de yodo [69–71] después de 1, 2, 5, 10, 15, 30, 60 y 90 min se midió a 580 nm y se expresó como equivalentes de almidón producidos.

En otro experimento independiente, las reacciones hidrolíticas se iniciaron mediante la adición de 60 µg de enzima a 1 ml de 10 mg/ml de almidón en Tris 50 mM, NaCl 150 mM, CaCl₂ 2 mM, pH 7,0. La hidrólisis a 70 °C, se midió por los azúcares reductores producidos después de 6, 12, 18 y 24 h, usando el reactivo DNS [72].

4.3.2.6 Sobre los reactivos y enzimas usadas en esta tesis

La mayoría de los reactivos se adquirieron de Sigma-Aldrich, con excepciones que se indicaron en el texto. Las enzimas de New England Biolabs, salvo las excepciones indicadas.

4.3.3. Simulaciones dinámicas moleculares

Para explicar los resultados experimentales cuya metodología se describe en secciones subsecuentes, realizamos simulaciones de dinámica molecular de las enzimas silvestres *TmAmyA* y *TmGTasa* y sus variantes.

La simulación y el análisis de la dinámica molecular (MD) se realizaron con GROMACS (GRONingen MACHine for Chemical Simulation) versión 2020.4 con ligeras modificaciones [73] [74]. Las coordenadas de las proteínas se obtuvieron del Protein Data Bank. Todas las modificaciones o visualizaciones de archivos PDB [75] se realizaron con UCSF Chimera [76] o Pymol (Schrödinger, Inc., EE. UU.). Las estructuras incompletas se completaron con Modeller [77]. La hidratación fue con un "Simple Point Charge Extended" (SPCE), en una caja cúbica con un espacio de 1 nm medido desde los bordes de la enzima en las direcciones *x*, *y*, *z*. El campo de fuerza fue OPLS-AA/L (potenciales optimizados para simulaciones de líquidos). El sistema se neutralizó con iones Na⁺ y Cl⁻. Posteriormente, se realizaron pasos de minimización hasta que la energía disminuyó a menos de 1000 kJ/mol/nm. Se realizaron equilibrios a N, V, P y T constantes durante 1000 ps con un paso de 50000. Finalmente, las simulaciones se realizaron para 500 ns. El análisis de los resultados se realizó con herramientas de GROMACS y la visualización con Xmgrace [78]. El pKa de los ácidos aspártico y glutámico se calculó con PROPKA 3.4.0 [79].

5. RESULTADOS

5.1 Modelo de homología de TmAmyA

Decidimos utilizar la α -amilasa AmyA (*TmAmyA*) de *Thermotoga maritima* como modelo de hidrolasa debido a su parecido con la 1,4- α -glucanotransferasa de *T. maritima*, una enzima transglicosídica. Esto hace a esta enzima un buen candidato para transicionar a una enzima más transglicosídica, Además, esta enzima ofrece como ventajas una alta estabilidad térmica, con una temperatura óptima por encima de los 80 °C, y un eficiente patrón de hidrólisis sacarificante del almidón, asociado a una considerable actividad de transglicosidación [49,68], sin embargo, al no existir un modelo estructural basado en datos experimentales, decidimos construir su modelo por homología.

Al construir el modelo por homología de *TmAmyA* utilizamos la estructura cristalográfica de la amilasa de *Thermotoga petrophila* (PDB ID 5M99, resolución 1,96 Å) como plantilla. Se excluyeron treinta residuos adicionales en el extremo N-terminal de *TmAmyA* que no están presentes en la amilasa cristalizada de *Thermotoga petrophila* y que no pertenecen a los dominios centrales de las enzimas GH13. Los 504 residuos restantes tienen una identidad de secuencia del 98,4 % con sólo seis sustituciones. Al validar este modelo (Figura 1) con la herramienta “Evaluación de estructuras” del servidor Swiss-Model [80], el valor QMEAN fue 1,06; se obtuvo un valor similar para la estructura utilizada como plantilla; el MolProbity Score tuvo un valor de 2,96 y 96,63% de los residuos tenían una configuración favorable en el gráfico de Ramachandran (Ramachandran Favored). Adicionalmente, la validación con ProSA-web (Z-Score de -9.95) tuvo un valor similar a otras proteínas de tamaño similar. VERIFY3D sugiere que el modelo tiene una calidad adecuada para el estudio presentado en este trabajo (el 99,21 % de los residuos tenían una puntuación media 3D-1D $\geq 0,2$).

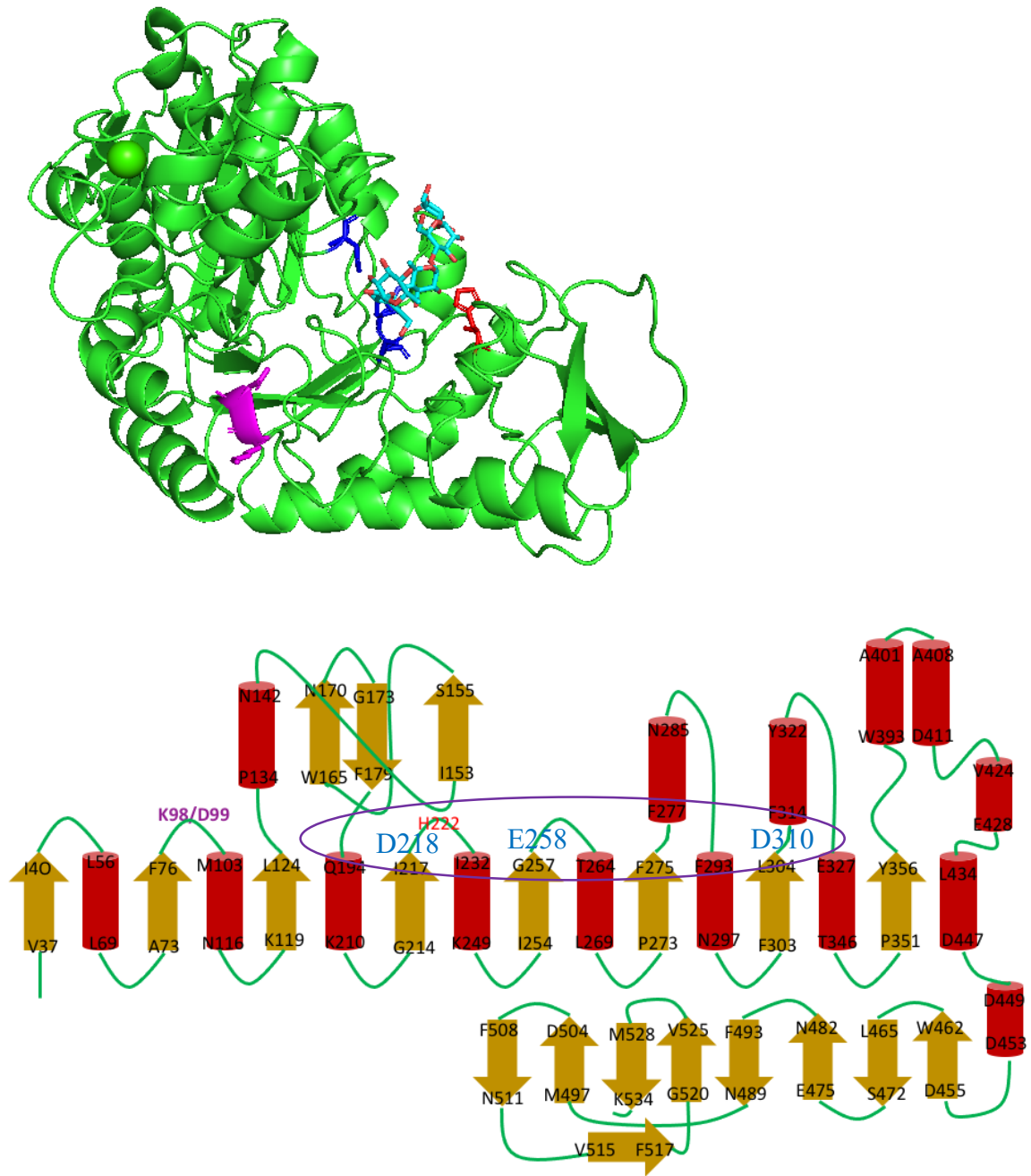


Figura 1. Modelo por homología de TmAmyA. a) estructura tridimensional, b) mapa topológico. Los sitios que se mutaron en el presente trabajo aparecen en violeta (K98/D99) y rojo (H222). Los residuos catalíticos (D218, E258, D310) están marcados en azul. En la estructura, el sitio catalítico está marcado con el análogo del estado de transición acarbosa (en azul cielo), en el diagrama con un óvalo morado.

5.2 Cálculo de factores de enriquecimiento por residuo

Al analizar los mapas de contactos entre residuos en 14 estructuras en el Conjunto de datos 1 (con 4 transglucosidasas y 10 hidrolasas de la familia GH13) consideramos sólo los monómeros y descartamos las interacciones en la interfase oligomérica porque, de acuerdo con la información disponible, todas las proteínas son monoméricas en solución. Por ejemplo, la *TmGTasa*, cuya estructura cristalina sugiere una proteína dimérica, en solución se encuentra en un equilibrio donde 90 % de la proteína está como monómero y 10 % como oligómero [57]. Además, ya que nuestra metodología identifica contactos entre residuos comunes a todos los miembros de la familia de enzimas GH13, se espera que los efectos de las mutaciones que identificadas, sean también generales y no dependan del estado oligomérico de la enzima.

Los valores del factor de enriquecimiento (Δf_{aa}^{ij}) de los aminoácidos no fueron homogéneos para los residuos de todos los contactos y se alejaron de cero siendo en algunos casos positivos (indicando mayor frecuencia del aminoácido en las enzimas transglicosídicas que en las hidrolíticas) y negativos (indicando que el aminoácido es más común encontrarlo en las enzimas hidrolíticas que en las transglicosídicas), lo que coincide con la noción de que durante la evolución tiende a aumentar la frecuencia de los contactos entre residuos importantes para para mantener la estructura y la función enzimática.

5.3 Clasificación de proteínas como transglucosidasas o hidrolasas

Al calificar los contactos de las enzimas como hidrolíticas (si $\Delta f_{aa}^{ij} < 0$) o transglicosídicas (si $\Delta f_{aa}^{ij} > 0$) con los factores de enriquecimiento obtenidos y graficar la relación de contactos transglicosídicos/hidrolíticos (Figura 2), agrupamos las enzimas de acuerdo a su función, tanto en el Conjunto de datos 1, usado para calcular los valores de Δf_{aa}^{ij} , (Figura 1a) como en el Conjunto de datos 2, que se mantuvieron fuera del cálculo de los valores de Δf_{aa}^{ij} , (Figura 2b), sin embargo,

algunos residuos del Conjunto de datos 2 quedaron sin calificar, señalando un muestreo incompleto del espacio de la secuencia con el Conjunto de datos 1. Lo anterior sugiere que los valores de los factores de enriquecimiento (Δf_{aa}^{ij}) obtenidos reflejan la selección de los residuos para cumplir la función como enzimas hidrolíticas o transglicosídicas.

En la Figura 2b observamos subgrupos de enzimas cuya actividad transglucosídica/hidrolítica sería interesante caracterizar en estudios futuros para ver si la actividad observada el cociente de sus actividades transglicosidación/hidrólisis se corresponde a la clasificación con la relación de contactos T/H encontrados.

También pudimos discriminar la función con los factores de enriquecimiento en el Conjunto de datos 4 (Figura 2c), que contenía un mayor número de enzimas (que se acumularon durante la realización de este estudio). A pesar del éxito para distinguir entre enzimas transglicosídicas e hidrolíticas, notamos que el origen de las enzimas empleadas para obtener el factor de enriquecimiento influye en los valores obtenidos. Por ejemplo, para el Conjunto de datos 4 la hidrolasa con el mayor número de contactos transglicosídicos correspondió a una α -amilasa del grupo Firmicutes (PDB ID:1QHO), mismo grupo al que pertenecen la mayoría de las enzimas transglicosídicas de ese Conjunto de datos (Tabla S4), lo que parece haber generado un sesgo en los aminoácidos considerados transglicosídicos, ya que también tienen un componente Firmicutes. Por tanto, los factores de enriquecimiento nos permiten hacer la clasificación funcional de las enzimas, pero reflejan no solo la especificidad, sino también el origen filogenético de las enzimas empleadas para obtenerlos, sugiriendo que se requiere obtener una muestra que incluya enzimas provenientes de una mayor diversidad de organismos para observar, preferentemente, el sesgo generado en los aminoácidos por la función.

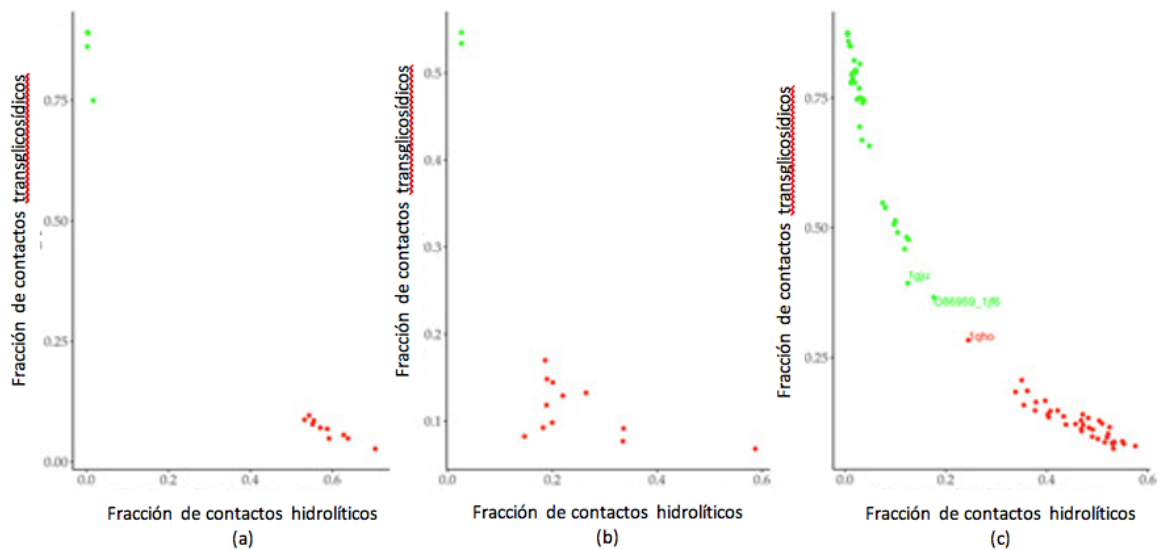


Figura 2 Clasificación de las enzimas de la familia GH13 por funciones de acuerdo a la relación de residuos hidrolíticos y transglucosídicos que contienen. (a) Conjunto de datos 1 (enzimas del grupo interno o de entrenamiento). (b) Conjunto de datos 2 (enzimas del grupo externo o de validación). (c) Conjunto de datos 4 (conjunto de enzimas caracterizadas en el PDB). En los tres casos, las enzimas definidas como transglucosídicas o hidrolíticas se muestran con puntos verdes o rojos, respectivamente.

5.4 Determinación de la conservación de los contactos en miembros de la familia de enzimas glicósido hidrolasas GH13

Cada contacto entre residuos en una enzima puede identificarse como equivalente en otras enzimas y la fracción de enzimas que lo comparten da un valor de conservación para dicho contacto, al graficar estos valores de conservación para un par de enzimas, produjo un gráfico de dispersión donde cada enzima ocupa uno de los ejes (Figura 3). Los datos se comportan como una recta en la que los contactos con valores de conservación similar están más cerca de la diagonal y los residuos más conservados tienen valores de conservación cercanos a uno. Una mayor similitud evolutiva, resulta en un gráfico con coeficiente de correlación más cercano a uno porque en ambas enzimas se podrá identificar los valores de conservación de sus contactos serán más parecidos, debido a que la probabilidad de identificar el contacto como equivalente en otras enzimas es similar en las dos bajo comparación.

5.5 Clasificación evolutiva de enzimas mediante la conservación de sus contactos

Al considerar todas las gráficas producidas, notamos que el coeficiente de correlación del gráfico, que varía entre cero y uno, adquiriría valores más cercanos a uno cuando las dos enzimas comparadas eran más cercanas evolutivamente y más cercano a cero cuando las enzimas eran más distantes, decidimos denominar a este coeficiente como coeficiente de similitud. La Figura 3 muestra la correlación entre todas las proteínas del Conjunto de datos 3 (39 enzimas pertenecientes a la familia GH13, una a la familia GH97 y una perteneciente a la GH31) cuando se le comparó con *TmAmyA*. Las correlaciones más bajas corresponden a las enzimas identificadas con el PDB ID: 2ZQ0 (familia GH97) y 3W37 (familia GH31), resultado esperado ya que estas enzimas no pertenecen a la familia GH13. Por lo tanto, este valor de correlación podría usarse como un coeficiente de similitud de contactos para discriminar entre miembros de diferentes familias de glucósidos hidrolasas. Esto puede adjudicarse a un cambio más lento en los contactos entre residuos que en la naturaleza de los residuos que lo forman.

Correlación de la conservación de contactos entre AmyA y otras enzimas

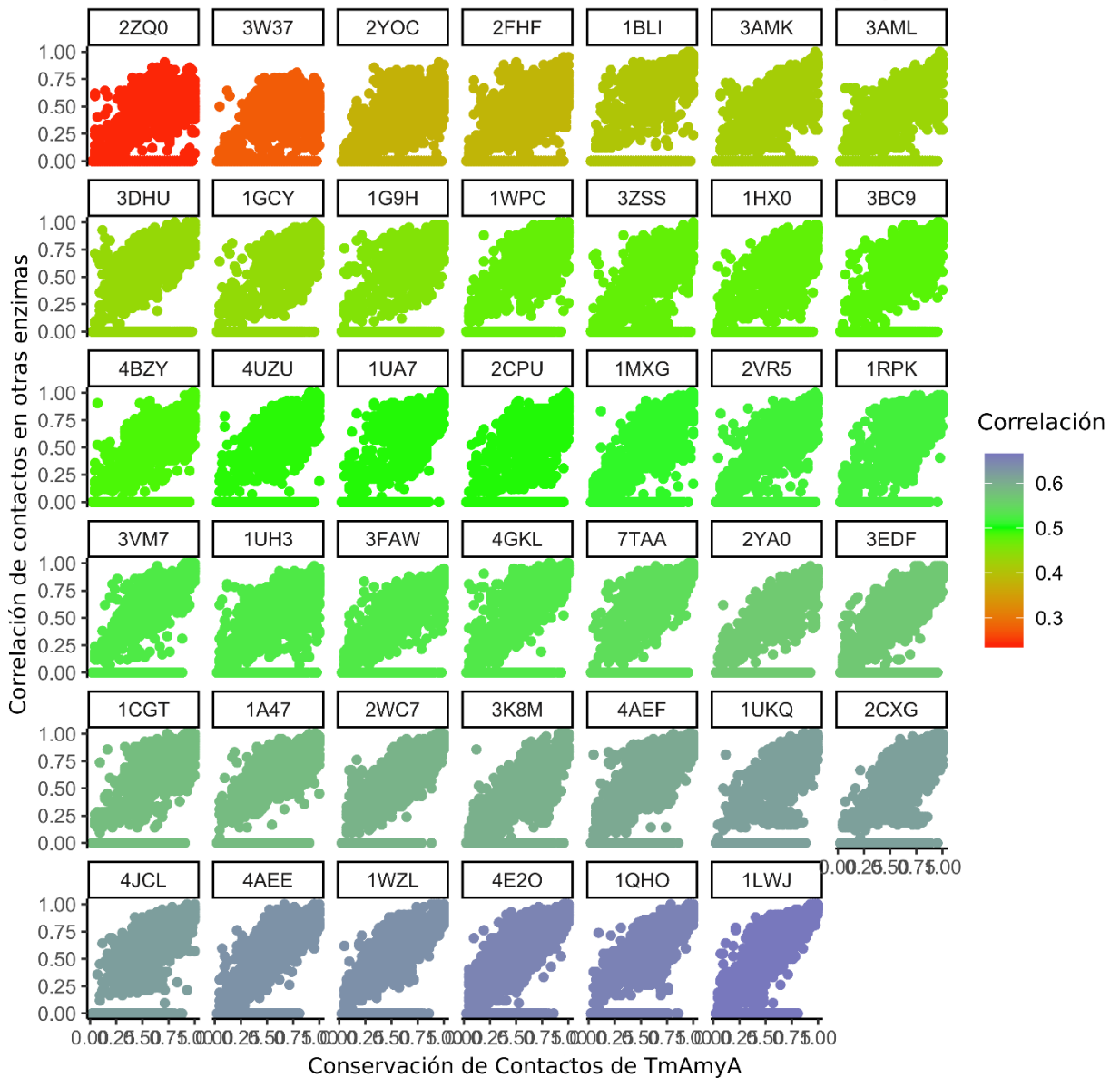


Figura 3 Correlación de la conservación del contacto residuo-residuo entre todas las enzimas y las de TmAmyA, una de las enzimas modificadas en este estudio. La conservación de cada contacto en una proteína se representa frente a su conservación en TmAmyA.

El coeficiente de similitud de contacto se correlacionó con el parámetro estructural DALI Z-value, que corresponde a la suma de las diferencias de distancias entre carbonos alfa identificados como equivalentes entre dos estructuras después de

una optimización conformacional (Figura 4a) y mostró una relación con la identidad de secuencia [81]. Al comparar el coeficiente de similitud con la identidad de secuencias, su resolución para identificar diferencias entre proteínas fue mayor que la de la distancia de secuencia por pares, para la cual los valores son muy similares para las proteínas bajo consideración (Figura 4b), demostrándose que el coeficiente de similitud de contacto mostró más sensibilidad para diferenciar enzimas que los parámetros de similitud de secuencia, para los que, dos enzimas con igual similitud (por debajo del 30%) no pueden diferenciarse sin la presencia de enzimas con similitud intermedia.

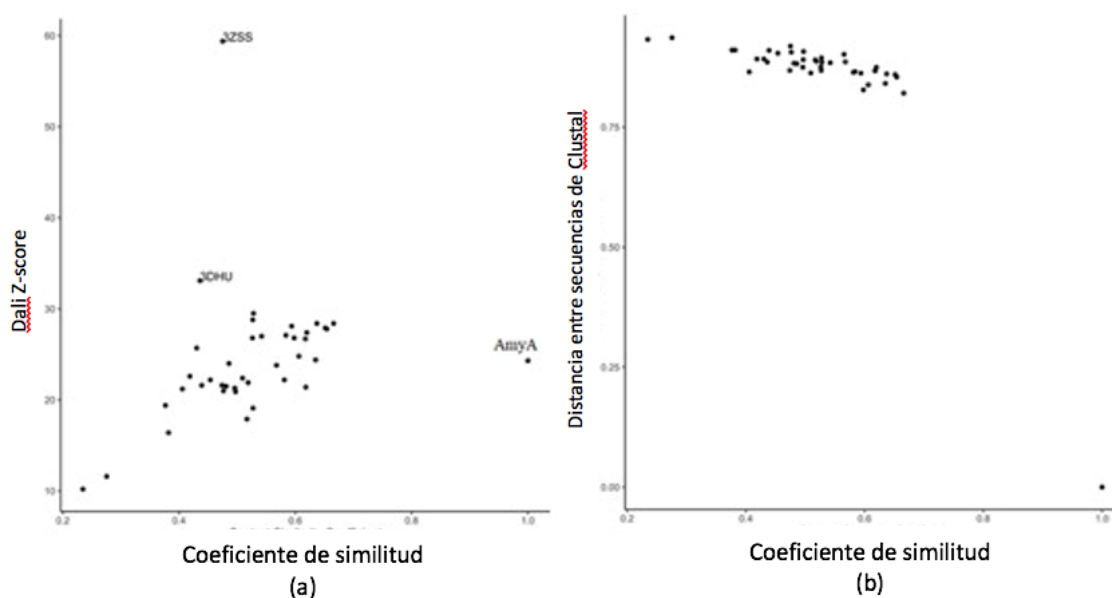


Figura 4 Parámetros de distancia de similitud entre las enzimas evaluadas y *TmAmyA* en función de su correlación entre la conservación de contacto (R -cuadrado) y (a) estructura (Dali) y (b) alineamientos de secuencia (Clustal).

Mientras *TmAmyA* y *TmGTasa* no parecen similares cuando se comparan sus secuencias (26.1% de identidad, 41.7% de similitud), al comparar mapas de contactos encontramos que son similares; esta similitud también se observa en la comparación estructural del modelo cristalográfico 1LWJ (*TmGTasa*) y el modelo por homología de *TmAmyA*. Esta similitud de los contactos y estructuras de ambas

proteínas, pero con una función y secuencia diferenciadas sugiere surgieron de un evento de duplicación. La comparación de contactos y estructuras de las enzimas de la familia GH13 también sugiere un parentesco cercano entre algunas enzimas firmicutes y *TmAmyA* que no revela la comparación de secuencias (identidad <28%). A pesar de la diferencia estructural y de secuencia entre *TmAmyA* y el resto de las enzimas, el coeficiente de similitud de contacto pudo identificar a la GTase de *Thermotoga maritima* (*TmGTase*, PDBID: 1LWJ) como su estructura más cercana. El coeficiente de similitud de contacto también agrupó a algunos miembros de la subfamilia GH13 (subfamilia 2: 4JCL, 2CXG, 1UKQ; subfamilia 13: 2YOC, 2FHF; subfamilia 8: 3AMK, 3AML) e identificó extraños (GH97: 2ZQ0, GH31: 3W37). La semejanza entre las enzimas Firmicutes (de mayor a menor similitud 1QHO, 4E2O, 1WZL, 4JCL, 2CXG, 1UKQ, ...) y las Thermotogales. Esto aboga por una transferencia horizontal entre los dos grupos –la subfamilia 36 (donde se encuentra *AmyA*) y la subfamilia 2 (donde se encuentran las enzimas de origen Firmicutes). Evidencia adicional de esto puede encontrarse al comparar los dominios B de la estructura *TmGTase* con los de otras enzimas del grupo Firmicutes; no obstante que el dominio B varía más entre los tres dominios principales de la familia GH13; los residuos 148–161 en 1LWJ podrían alinearse fácilmente con 180–191 de *Geobacillus thermoleovorans* CCB_US3_UF5 Gtase (PDB ID: 4E2O), rasgo compartido con otras enzimas del grupo. Aunque extravió estructuras, como 3K8M, que deberían haber estado cerca de *TmAmyA*, esto podría deberse a que 3K8M se comparó incluyendo dominios adicionales, que no están presentes en la familia GH13 [82,83]. Como la comparación del conjunto de datos con una sola estructura produjo tales identificaciones, las comparaciones entre el conjunto de datos completo pueden proporcionar una herramienta para realizar la clasificación automatizada de enzimas al realizar el análisis de dominios.

Los resultados anteriores sugieren que el coeficiente de similitud podría usarse como un parámetro para el análisis filogenético.

5.6 Selección de los sitios de mutación.

Para el caso de la enzima hidrolítica *TmAmyA* buscamos introducir mutaciones encaminadas a aumentar su relación Transglicosidación/Hidrólisis (T/H) para hacer esta enzima más parecida a una transglicosidasa.

Para evaluar si los factores de enriquecimiento (Δf_{aa}^{ij}) pueden usarse como una guía para identificar sitios de mutagénesis y sustituciones de aminoácidos que produzcan el cambio de selectividad deseado, elegimos construir la doble mutante K98P/D99A. Dicha mutación fue elegida porque el Δf_{aa}^{ij} de los aminoácidos originales es de cero mientras los aminoácidos que los sustituiría tienen un valor positivo; K98 ($\Delta f_{aa}^{ij} = 0$) por P98 ($\Delta f_{aa}^{ij} = 0.45$) y D99 ($\Delta f_{aa}^{ij} = 0.00$) por A99 ($\Delta f_{aa}^{ij} = 0.4$) (Figura 4a), siendo estos unos de los valores de enriquecimiento encontrado (10 % más alto). Además, las mutaciones seleccionadas están fuera del sitio catalítico, y proponen un cambio drástico en las propiedades fisicoquímicas de las posiciones a sustituir.

Además, los residuos K98/D99 están ubicados en un bucle que une la hebra $\beta 2$ y la hélice-2 en el barril TIM (ver Figura 1), que no forma parte de ninguna de las regiones altamente conservadas de la familia GH13.

Las sustituciones K98P/D99A representan un cambio drástico de las propiedades fisicoquímicas, sin embargo, la prolina elegida para sustituir el residuo K98 favorece el giro observado en el modelo estructural en la posición a mutar, y su sustitución está indicada por un valor positivo de Δf_{aa}^{ij} , lo cual indica la presencia más frecuente de este residuo en las transglicosidasas que en las hidrolasas.

Estas mutaciones han sido evaluadas en el fondo silvestre y en combinación con la mutación H222Q, la cual está localizada en el sitio catalítico y aumentaron la relación transglicosidación/hidrólisis [48,49]. Esta mutación se eligió para comparar el efecto de una mutación en el sitio activo con las de aquellas fuera del sitio activo, además de ver si existen interacciones entre las mutaciones en el sitio catalítico y las de la periferia.

Para complementar la evaluación de nuestro algoritmo investigamos si el factor de enriquecimiento podría usarse para cambiar la selectividad en la otra dirección:

convertir una enzima transferasa en una hidrolasa. Para ello, seleccionamos la *GTasa* de *T. maritima* (*TmGTasa*).

En este caso, analizamos pares de residuos en lugar de residuos individuales dentro de un par de contactos, ya que, la relación de contactos transglicosídicos/hidrolíticos permitió clasificar las enzimas GH13 según su función (Figura 2), adicionalmente, buscamos aumentar la posibilidad de seleccionar mutaciones que disminuyeran la selectividad T/H considerando otro parámetro, la centralidad de intermediación, que ha sido reportada como una medida de la importancia de los residuos de aminoácidos en la estructura y función de las proteínas [57] y como medida del papel de un nodo en la transferencia de información dentro de una red [64]. Este parámetro de centralidad se calcula como la suma de la fracción de los caminos entre todos los pares de nodos i y j (definidos los caminos como los nodos contiguos que debe recorrerse para ir del nodo i al j) que contienen el nodo v , distinto de i y j [84].

Para los residuos con el valor de centralidad de intermediación en el 10 % más alto, identificamos contactos agrupados en clusters ocupados por residuos con factores de enriquecimiento positivo. Nos enfocamos en los clusters de residuos alrededor de los residuos F72 y F273 para los que nuestro algoritmo proponía sustituciones con factores de enriquecimiento negativos, y por tanto enriquecidos en las hidrolasas. Del cluster alrededor del residuo F72 tomamos la pareja de residuos F72-V86 ((Figura 5b), y la pareja T274-M279 del cluster alrededor del residuo F273 como blancos de mutagénesis.

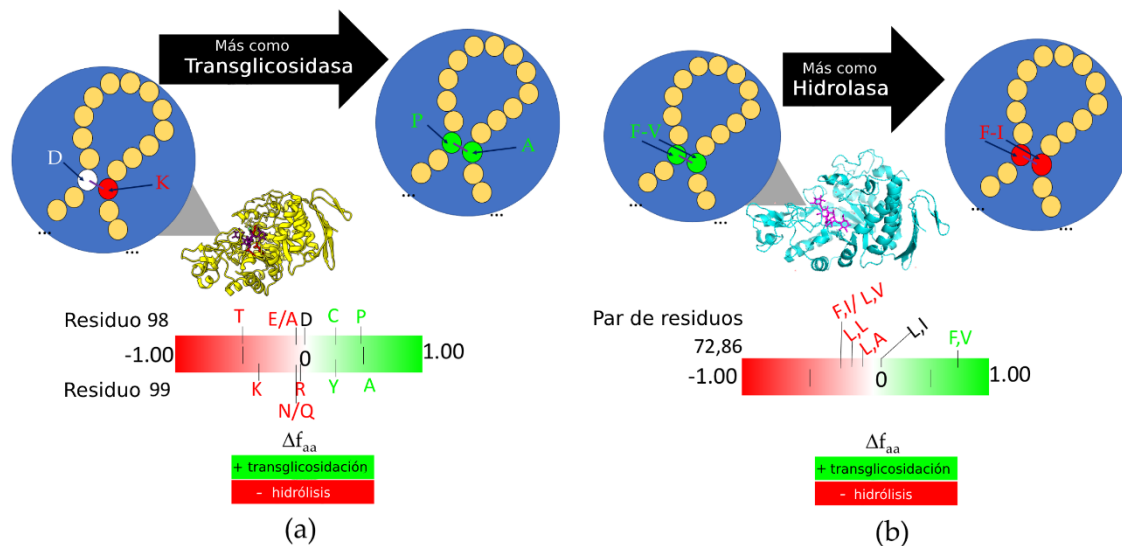


Figura 5 Representación esquemática de sustituciones en glicosidasas basadas en factores de enriquecimiento. (a) K98P/D99A en TmAmyA (b) F72L/V86I para TmGTasa. Los residuos encontrados con mayor frecuencia en hidrolasas y transglucosidasas se muestran en la parte inferior de la figura para cada par en rojo y verde, respectivamente.

En tanto el cluster de residuos alrededor de F72 se encuentra alejado del sitio activo la posición 279 del cluster alrededor del residuo F273 forma parte de las regiones altamente conservada en la secuencia de las proteínas en la familia GH13.

5.7 Caracterización de las variantes de TmAmyA

La caracterización de las variantes de TmAmyA se muestran en la Tabla 1. Para una mejor visualización del efecto de las mutaciones en el contexto de nuestro método, normalizamos los valores de la Tabla 1, usando como referencia a la enzima silvestre, y graficamos los valores de transglucosidación contra los de hidrólisis (Figura 6).

Tabla 1 Producción de variantes de TmAmyA después de 12 h de reacción

Variante	Hidrólisis (mEq Dextrosa/ μ g Proteína $\times 10^{-2}$)	Transglucosidación (mEq Butil Glucósido/ μ g Proteína $\times 10^{-4}$)	Transglucosidación/Hidrólisis (T/H) Relación $\times 10^{-2}$
Silvestre	2.8 \pm 0.2	7 \pm 1	2.5 \pm 0.6
K98A/D99P	1.93 \pm 0.09	5.4 \pm 0.5	2.8 \pm 0.4
H222Q	2.2 \pm 0.3	7.9 \pm 0.8	3.6 \pm 0.4
H222Q/K98A/D99P	1.8 \pm 0.1	10 \pm 2	5.6 \pm 0.4

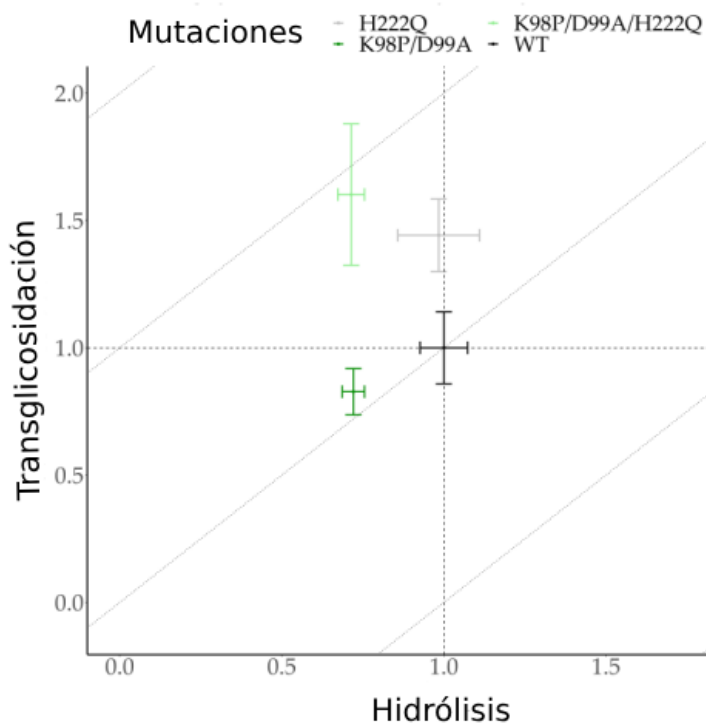


Figura 6 Relación transglucosilación/hidrólisis (T/H) de la α -amilasa TmAmyA y sus variantes. Los valores de transglucosidación e hidrólisis se normalizaron para cada enzima silvestre, que tiene las coordenadas (1,1). Las líneas discontinuas indican relaciones T/H equivalentes. La diagonal central corresponde a la enzima silvestre. Las enzimas con una mejor relación T/H están sobre la línea diagonal central y las enzimas con una mejor relación H/T están debajo de la línea diagonal central.

En el gráfico de la Figura 6, el cociente correspondiente a la selectividad Transglucosidación/Hidrólisis (T/H) se representa como una diagonal, donde todos los puntos en la diagonal tienen el mismo valor T/H y todos los puntos arriba de una diagonal tienen un mayor valor de transglucosidación y por tanto mayor T/H, mientras los puntos por debajo de esta corresponderán a valores de hidrólisis mayores y por

tanto, menor T/H. Nuestro objetivo para la α -amilasa *TmAmyA* con las sustituciones de aminoácidos para esta enzima hidrolítica era hacerla más parecida a las enzimas transglucosídicas y estas variantes aparecerán por arriba de la diagonal que pasa por la enzima silvestre.

La mutante doble K98P/D99A logró aumentar la relación T/H en un factor de 1,17 (Tabla 1) al decrementar la hidrólisis en mayor proporción a la transglucosidación. Aunque todas las mutaciones redujeron la producción hidrolítica con respecto a la enzima silvestre, esta reducción fue mayor en las variantes que contenían la mutación propuesta por los factores de enriquecimiento K98P/D99A, con una reducción de aproximadamente un 25% (Tabla 1). Por otro lado, la mutación redujo el rendimiento de alcoholólisis un 17 % cuando fue introducida a la enzima silvestre, mientras que la aumentó cerca del 10 % en la variante H222Q.

Al comparar los resultados de las mutaciones K98P/D99A con los de la mutación H222Q, vemos que son semejantes, a pesar de que H222Q está en el sitio activo mientras que K98P/D99A están a 15 Å de este sitio (Figura 7).

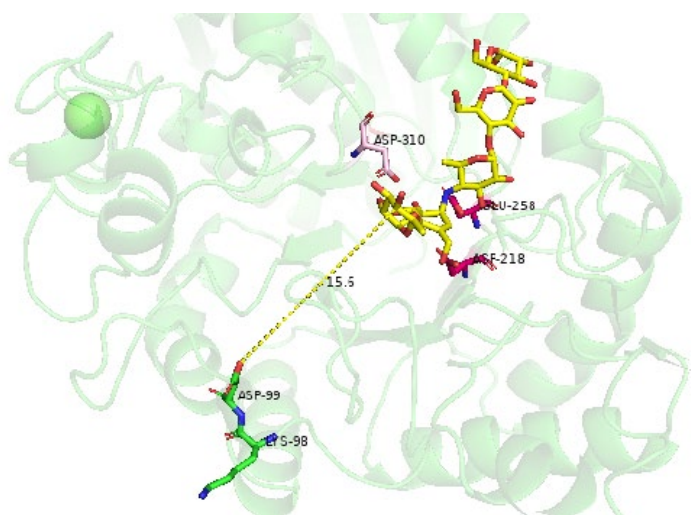


Figura 7 Distancias en Ångstroms entre los sitios de mutación (verde y cian) y los residuos catalíticos (rojo y naranja) en las variantes de *TmAmyA*. El centro activo está delimitado por el inhibidor competitivo acarbosa (amarillo). En rosa se muestra el residuo D310 que actúa como estabilizador del estado de transición.

5.8 Caracterización de las variantes de *TmGTasa*

Investigamos el papel de las sustituciones individuales y combinadas F72L, V86I, T274V, M279N en las reacciones de transglicosidación e hidrólisis (Figura 8). Aunque observamos una disminución en la actividad hidrolítica para muchos mutantes probados en *TmGTasa*, las mutaciones produjeron una mayor reducción de la actividad transglucosídica en todas las variantes, eliminando casi por completo su actividad en algunas de ellas (Tabla 2). Como resultado, se logró el objetivo de aumentar la relación H/T en una enzima transglicosídica, esto es, todas las enzimas con las mutaciones propuestas por nuestro método se encuentran por debajo de la diagonal que pasa por el punto correspondiente a la enzima silvestre (1,1). De particular interés es la variante M279N, que, además de disminuir la actividad transglucosídica a un cuarto de la original, aumentó la actividad hidrolítica deseada en un 25%, lo que produjo un incremento de cinco veces en la relación H/T. Vale la pena mencionar que, durante la caracterización de las variantes, detectamos una variante con una actividad general más alta que, además de las mutaciones diseñadas F72L/T274V, contenía dos modificaciones no deseadas: E77G/E226K. Esta variante no cambia la relación H/T porque ambas reacciones se vieron favorecidas por las sustituciones adicionales, sin embargo, muestra un aumento de alrededor del 40 % de la actividad general en relación con la proteína silvestre (Figura 8).

Es importante resaltar que el aumento de selectividad H/T en las variantes de *TmGTasa* es producto de mutaciones fuera del sitio catalítico, siendo M279N la sustitución más cercana, donde M279N no está en contacto directo con el sustrato (Figura 8b); en tanto el resto de las mutaciones están en el extremo más alejado del sitio catalítico de una de las hebras β del barril $(\beta\alpha)_8$ (Figura 9).

Tabla 2 Actividad específica de las variantes *TmGTasa*

Variante	Actividad hidrolítica ($\times 10^{-5}$ mg Almidón/ μ g Proteína/min)	Actividad transglucosídica ($\times 10^{-3}$ mg almidón/ μ g Proteína/min)	Hidrólisis/Transglucosidación (H/T) Ratio ($\times 10^{-2}$)
Silvestre	5.1 \pm 0.3	4.0 \pm 0.8	1.3 \pm 0.3
V86I	2.9 \pm 0.3	1.3 \pm 0.2	2.2 \pm 0.6
M279N	6.3 \pm 0.8	1.01 \pm 0.03	6 \pm 1
V86I/M279N	4.2 \pm 0.2	ND	NA
T274V/M279N	4.2 \pm 0.3	ND	NA
F72L/V86I/T274V	4.4 \pm 0.2	0.51 \pm 0.07	9 \pm 2
F72L/V86I/T274V/M297N	4.1 \pm 0.5	0.65 \pm 0.03	6 \pm 1
F72L/E77G/E226K/T274V	7.0 \pm 0.9	5 \pm 1	1.4 \pm 0.5
F72L/T274V	4.9 \pm 0.5	1.7 \pm 0.2	2.9 \pm 0.6

ND: no detectada; NA: no disponible.

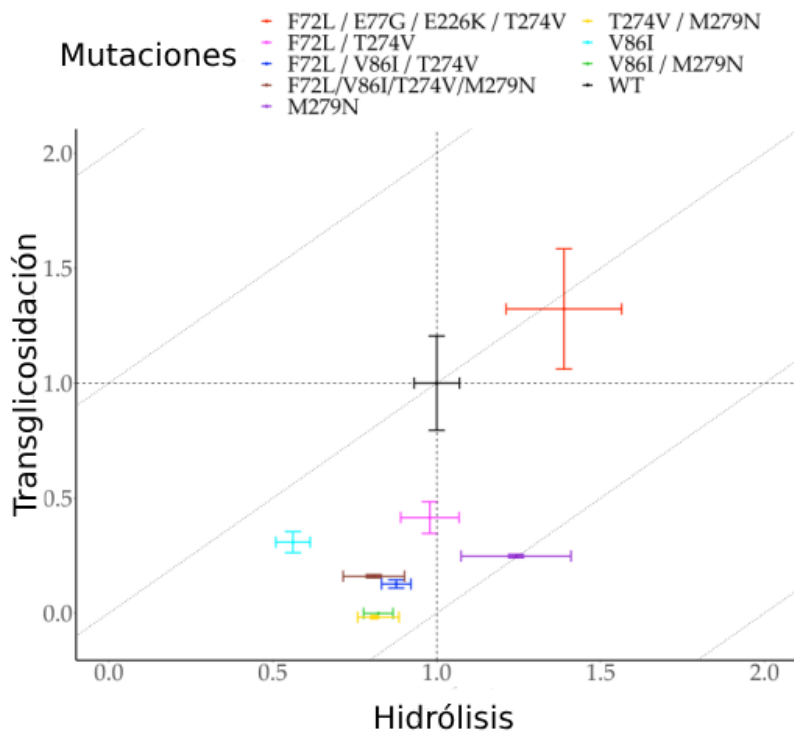


Figura 8 Relación transglucosidación/transglucosidación/hidrólisis (T/H) de la glucatransferasa *TmGTasa* y sus variantes. Los valores de transglucosidación e hidrólisis se normalizaron con la enzima silvestre, que tiene las coordenadas (1,1). Las líneas discontinuas indican relaciones T/H equivalentes. La diagonal central corresponde a la enzima silvestre. Las enzimas con una mejor relación T/H están sobre la línea diagonal central y las enzimas con una mejor relación H/T están debajo de la línea

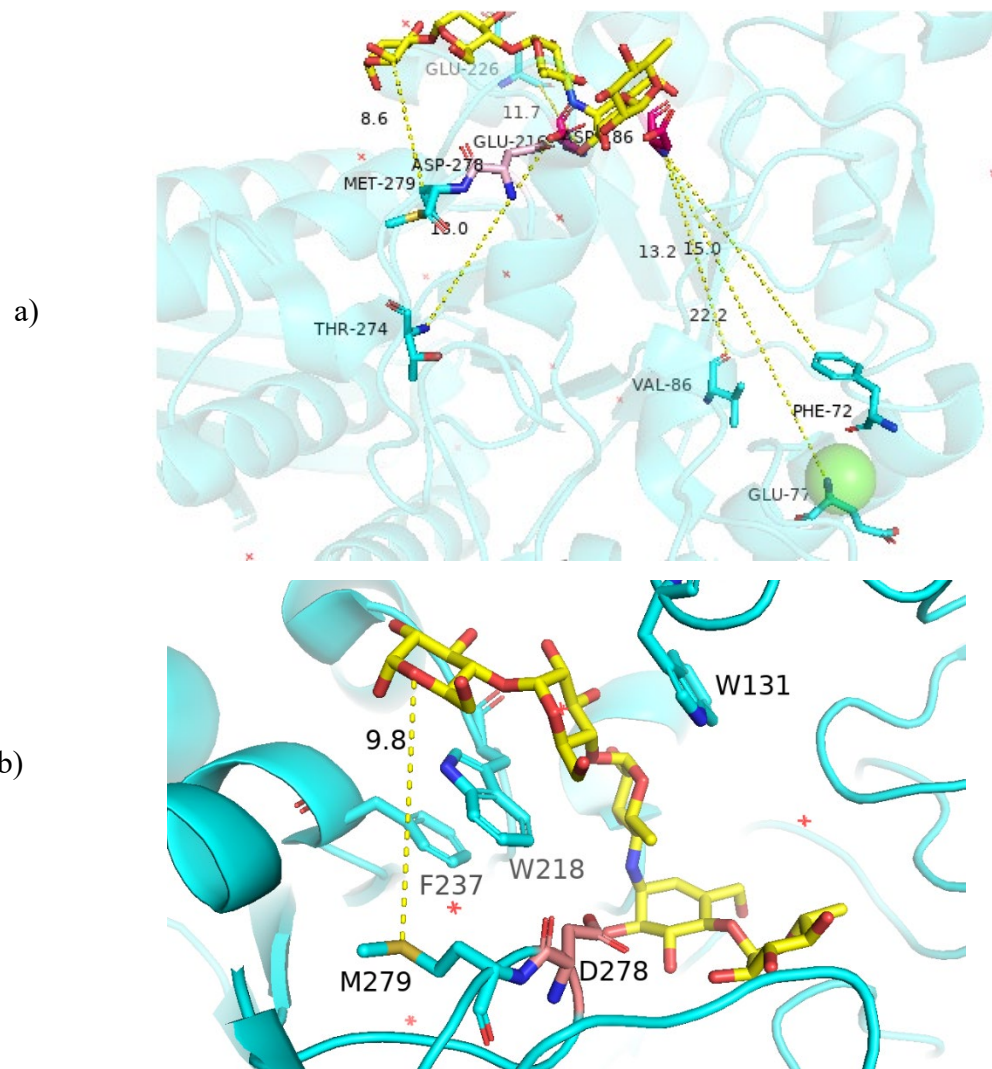


Figura 9 a) Distancias en Ångstroms entre los sitios de mutación (verde y cian) y los residuos catalíticos (rosa mexicano y rosa) de la TmGTasa. El centro activo está delimitado por el inhibidor competitivo acarbosa (amarillo). En rosa se muestra el residuo D-278 que actúa como estabilizador del estado de transición. b) Vecindad del residuo M279 en el sitio catalítico.

El papel de la dinámica de proteínas en la catálisis enzimática es bien reconocido [64,65]. Por lo tanto, llevamos a cabo simulaciones de dinámica molecular para algunas variantes de proteínas para explicar en términos estructurales, el efecto de las mutaciones sobre las actividades de las enzimas en términos estructurales. La simulación de *TmAmyA* muestra que la triple sustitución K98P/D99A/H222Q aumentó la flexibilidad de la enzima exclusivamente en la hélice $\alpha 4$ (dominio catalítico), a la que pertenece el residuo 247, señalado en la Figura 9a, marcado por

una diferencia de fluctuación cuadrática media (RMSF) positiva entre la triple mutante y la enzima silvestre (Figura 10a). Esta hélice está cerca de los bucles que contienen el aspartato catalítico (nucleófilo) y el glutamato (ácido-base). Por el contrario, en el mutante K98P/D99A/H222Q se rigidizaron los bucles y hélices que contienen los residuos 325, 387 y 415. El aumento de movilidad de los residuos catalíticos podría ser la causa de la pérdida de la actividad en las variantes de *TmAmyA*, teniendo un mayor impacto en la función hidrolítica que en la transglicosídica, función para la cual la hélice $\alpha 4$ podría rigidizarse un poco por las interacciones con el aceptor.

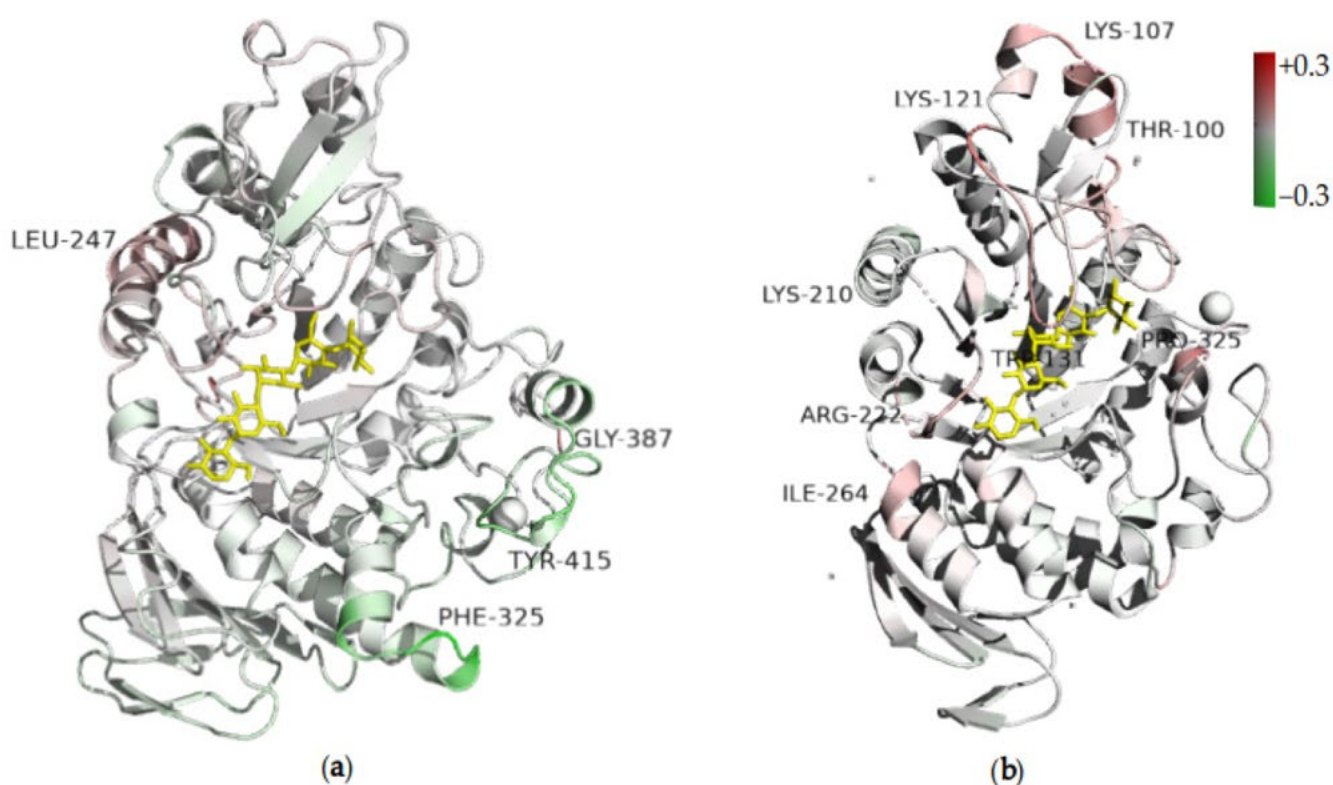


Figura 10 Diferencia en flexibilidad entre las enzimas silvestres *TmAmyA* y *TmGTasa* y sus variantes. Diferenciada RMSF (en nanómetros) a lo largo de la estructura de (a) la hidrolasa *TmAmyA* y su variante la mutante triple K98P/D99A/H222Q (menos hidrolítica que la enzima silvestre). (b) de la transglicosidasa *TmGTasa* silvestre frente a T274V/M279N (menos transglicosídica). Las regiones donde el RMSF no se modificaron en las variantes aparecen en blanco, mientras aquellas con incrementos o decrementos de RMSF están coloreadas en rojo y verde, respectivamente. El centro activo está delimitado por la acarbosa (en amarillo).

Los cambios en la flexibilidad también se estudiaron para las variantes de la enzima *TmGTasa*, para la cual comparamos el efecto de la sustitución simple M279N y

doble T274V/M279N con la *TmGTasa* silvestre. Fue evidente un aumento en los valores de RMSF del doble mutante T274V/M279N en comparación con los de la *TmGTasa* silvestre, alrededor de los residuos 100, 107, 121 (dominio B), 131 (subsitio +2, dominio B), 210 (bucle $\alpha 4\beta 5$), 222 (hélice $\alpha 5$), 264 (hélice $\alpha 6$) y 325 (bucle $\alpha 4\beta 5$, cerca del subsitio -3; Figura 10b). Si bien las fluctuaciones de RMS para el mutante sencillo M279N fueron similares al mutante doble alrededor de los residuos 100, 107, 121 y 131, aumentaron en relación con las otras proteínas cerca del residuo 222. Por lo tanto, el dominio B, que es importante para la especificidad del sustrato en otras enzimas GH13, es más móvil en ambos mutantes [57], esto podría dificultar la interacción con la molécula aceptora necesaria para la reacción de transglicosidación.

Se analizaron los ángulos diedros (χ_{1-3}) del residuo ácido-base catalítico para evaluar su movilidad tanto en la enzima silvestre como en los mutantes de *TmAmyA* y *TmGTase* (datos no mostrados), como en otros reportes. Observamos una distribución más difusa en las variantes más hidrolíticas de ambas enzimas (en *TmAmyA* silvestre vs. K98P/D99A/H222Q, y en M279N vs. *TmGTase* silvestre). Lo anterior, se contrapone con la movilidad mostrada para el residuo ácido-base catalítico mostrado por David et al., que también empleó el ángulo χ_3 como indicador de actividad transglicosídica [85] y lo descrito por Kong et al., quien definió un nuevo ángulo para medir esta movilidad [86], en cuyos trabajos atribuyen una mayor movilidad del residuo ácido base catalítico a una mayor actividad transglicosídica. Estas diferencias podrían justificarse considerando que a pesar de la menor movilidad local del residuo ácido-base catalítico (medido con χ_{1-3}), la estructura secundaria que lo contiene tiene mayor movilidad en las mutantes aquí reportadas. Adicionalmente, debe considerarse que los resultados en este aspecto no son definitivos, al no evaluar el intermediario covalente.

Las mutaciones introducidas produjeron cambios en la dinámica estructural que modifican el pKa de E216 en *TmGTasa*. Los cálculos de PROPKA [79] de este residuo se realizaron en las estructuras correspondientes a tres momentos diferentes en la simulación: 200, 300 y 400 ns, cuando los valores de RMSD ya se habían estabilizado. El pKa del residuo ácido-base catalítico fue mayor para ambas

variantes que para la *TmGTasa* silvestre. Para este parámetro, el promedio fue de $3,0 \pm 0,97$ para la enzima silvestre y $6,1 \pm 0,54$ y $4,8 \pm 0,43$ para T274V/M279N y M279N, respectivamente. Estos resultados concuerdan con la noción de que la hidrólisis requiere un residuo más básico que la transglicosidación [87]. Aunque los valores de pKa obtenidos no reflejan directamente el pKa del E216 durante la reacción de hidrólisis o transglicosidación (la enzima aún no ha formado el intermediario azúcar-enzima unido covalentemente y el sitio catalítico está vacío), estos resultados indican un desplazamiento del pKa de este residuo catalítico por las mutaciones en la dirección que favorece la catálisis de la reacción deseada. Estos cambios de pKa reflejan el cambio en la distancia promedio entre D278 y E216, que fue 1 Å más cercana en ambos mutantes que en la enzima *TmGTasa* silvestre, aumentando el pKa de uno de los aminoácidos participantes, para evitar la repulsión electrostática. En el caso del residuo equivalente en *TmAmyA*, E258, el triple mutante K98P/D99A/H222Q tiene un valor de pKa similar al de la enzima silvestre para su residuo ácido-base catalítico (alrededor de 4,8 para ambos). Este resultado sugiere un mecanismo diferente para el cambio en la especificidad de la reacción, que afecta principalmente a la reacción de hidrólisis.

Otro aspecto a resaltar es el enlace de hidrógeno formado entre el residuo D278, que funciona como estabilizador del estado de transición, y el residuo K324, que interacciona con el glucósido en el subsitio catalítico +2 (donde se une el aceptor). Esta interacción solo se observa en la variante *TmGTase* T274V/M279N y podría generar diferencias de actividad de la enzima en la reacción de hidrólisis y la transglicosidación porque debe romperse al unirse un azúcar en el subsitio +2. En *TmAmyA*, el residuo equivalente a K324 es una glicina, un residuo menos apto para formar un enlace de hidrógeno con el aspartato catalítico correspondiente.

6. DISCUSIÓN

Implementamos una metodología para identificar sitios blanco de mutagénesis para modular la relación transglicosidación/hidrólisis (T/H) en dos miembros de la familia GH13. Esta metodología seleccionó residuos a mutagenizar lejos del sitio activo (entre 11.1 y 22.2 Å de distancia) (Figura 7 y Figura 9), que modificaron la especificidad de la reacción. Fue interesante notar que también podría seleccionar residuos cercanos al sitio catalítico, como el residuo 279 de *TmCGTase*, que está al lado del residuo catalítico D278 (2,7 Å; Figura 9). Esta modulación de la especificidad de la reacción parecía funcionar en ambas direcciones; sin embargo, la contribución más significativa fue la reducción de la reacción no deseada y, en menor medida, el aumento de la reacción deseada, en la mayoría de los casos.

Al implementar nuestra metodología, generamos dos parámetros, el factor de enriquecimiento y el coeficiente de similitud (de los contactos entre enzimas); el primero nos permitió separar a las enzimas GH13 según su especificidad, después de calificar con este los contactos como hidrolíticos o transglicosídicos, (Figura 2) y el segundo, agruparlas en subfamilias (Figura 3).

Nuestro coeficiente de similitud muestra resultados consistentes con trabajos previos en los que se han utilizado patrones de contactos para distinguir familias de proteínas [26] y se suman a su uso en la clasificación funcional y evolutiva de subfamilias, detectando eventos de duplicación de genes (*TmGTase* y *TmAmyA*) y de transferencia horizontal entre los dos grupos –la subfamilia 36 (donde se encuentra *AmyA*) y la subfamilia 2 (donde se encuentran las enzimas de origen Firmicutes). Estos resultados recomiendan un uso más extendido de contactos de residuos en la clasificación de enzimas, complementarios al análisis de secuencias y estructuras, ya que su análisis es más sensible que el análisis de secuencias y puede automatizarse.

A pesar de que nuestro trabajo no logró con todas las mutaciones incrementar la actividad transglicosídica en *TmAmyA* y la actividad hidrolítica en *TmGTase*, con todas las mutaciones el cambio de la selectividad T/H deseado a través de

decrementos de la función hidrolítica en la amilasa *TmAmyA* y de la transglicosídica en la glucanotransferasa *TmGTasa*. Además, nuestra metodología propone explorar los residuos fuera del sitio catalítico para modificar su especificidad o función a través del análisis de contactos, pudiendo ser estos una guía en la selección de blancos de mutagénesis que extienda los residuos estudiados hasta ahora.

Nuestra metodología encuentra paralelo en algunas metodologías de análisis de coevolución que han sido reportadas [88,89]. Aún así, requirió menos datos para identificar sitios para modificar la función, ya que, los contactos pueden ser observados directamente en la estructura mientras el análisis de secuencias solo permite inferirlos después de analizar los patrones de coevolución de distintas posiciones en un alineamiento múltiple de secuencias (que requieren de correcciones para evitar la predicción de contactos inexistentes), además, ya que la estructura se conserva más que la secuencia, es posible considerar enzimas cuya similitud de secuencia es baja por lo que se puede lograr una mejor exploración del espacio de secuencias con un menor número de enzimas muestreadas. Ambas perspectivas destacan la importancia de analizar los contactos de los residuos cuando se estudia una función, incluso lejos del sitio catalítico. Trabajos anteriores destinados a clasificar enzimas de función desconocida y predecir y verificar modelos 3D también han sugerido la importancia de los contactos de residuos [21], donde la predicción del mapa de contactos ha jugado un papel importante.

El método presentado en este trabajo está limitado por la disponibilidad de estructuras 3D, que suelen ser menos abundantes que las secuencias empleadas en estudios coevolutivos. Esta limitación podría superarse con el uso de inteligencia artificial, como Alpha-Fold2, en la predicción de estructuras, sin embargo, este aumento de estructuras debe ir acompañada de un incremento de las enzimas caracterizadas experimentalmente, ya que, contar con enzimas caracterizadas y con estructura conocida es un requisito indispensable para poder determinar los factores de enriquecimiento y usarla como herramienta para clasificación funcional. Por ejemplo, para la familia GH13 se contaba solo con 31 enzimas transglicosídicas caracterizadas según lo informado en la base de datos CAZy.

Otra limitación de nuestra metodología para el caso de la familia GH13 es que a pesar que sus miembros realizan hidrólisis y transglucosidación en diferentes grados, su caracterización está generalmente centrada en la reacción que realiza cada enzima en mayor proporción. Una caracterización de ambas reacciones, facilitaría el análisis de la información que proporcionan las caracterizaciones experimentales con vistas a hacer diseño de enzimas. Esta caracterización también determinaría si existe un significado funcional para la agrupación de enzimas GH13 en función de sus contactos hidrolíticos o transglucosídicos (Figura 2).

Nuestra metodología tuvo como objetivo crear un parámetro para guiar la elección de mutantes que permitieran la modulación de la especificidad transglucosidación/hidrólisis. Como señalaron Hleap y Blouin [90] en su estudio sobre la estabilidad de la evolución de la familia GH13, una predicción cuantitativa de la especificidad requiere información catalítica con un sólo sustrato para todos los miembros de la familia bajo análisis. Por el contrario, la familia GH13 se ha caracterizado por una gran variedad de sustratos tanto para la actividad hidrolítica como para la transglucosídica y, en muchos casos, solo por su actividad más destacada. Incluso cuando se ha utilizado el sustrato más común, el almidón, las variaciones en la ramificación 1,6 podrían impedir esta caracterización. Superamos este obstáculo contrastando enzimas con una especificidad distinta, creando así un parámetro (el factor de enriquecimiento) que ofrece una guía cualitativa para transformar la especificidad de la reacción, que tal vez podría requerir más datos experimentales y el uso de inteligencia artificial para volverse cuantitativa. Sin embargo, esta guía podría reducir el espacio de secuencia que se necesita explorar para lograr un cambio de especificidad, tanto en estudios de diseño racional como de evolución dirigida.

Un resultado valioso de este enfoque fue la identificación de dianas para la mutagénesis más allá del sitio catalítico, cuya relevancia en términos de su especificidad y actividad ha sido demostrada por la evolución dirigida ; éstas corresponden a residuos formando contactos altamente conservados en la familia de proteínas, que son ocupados preferencialmente por un aminoácido, de acuerdo a la función de la enzima[91]. La exploración de dichos sitios está limitada por la

disponibilidad de métodos de selección o de detección de alto rendimiento; por lo tanto, restringir el espacio de secuencia que se va a evaluar se vuelve primordial. En este trabajo identificamos la presencia de clusters conformados por residuos en contacto cuyos aminoácidos fueron enriquecidos tanto en el grupo de enzimas hidrolíticas como en las transglicosídicas de la familia GH13, además logramos clasificar funcionalmente a las enzimas al calificar todos sus contactos como hidrolíticos y transglicosídicos. Ambos resultados sugieren que nuestro método podría tener un mayor éxito con la propuesta de un mayor número de mutaciones simultáneas. La evaluación para su selección se beneficiará con el uso de inteligencia artificial, la cual ya ha sido empleada para proponer enzimas *de novo* con actividad similar a las naturales, al aplicar estrategias empleadas en el análisis de lenguaje y considerar las secuencias de residuos de las enzimas como un enunciado que sigue reglas gramaticales y de semántica [92].

7. CONCLUSIONES

Propusimos y validamos un método basado en mapas de contacto para identificar sitios blanco de mutagénesis que modulen la relación transglicosidación/hidrólisis (T/H) en miembros de la familia GH13, tanto en transferasas como hidrolasas. Con este método logramos modificar la selectividad T/H de dos enzimas, la hidrolasa *TmAmyA* y la transferasa *TmGTasa*, para aumentarla en la primera enzima y disminuirla en la segunda. Para la variante *TmGTasa* M279N, redujimos la actividad transglicosídica, que se deseaba eliminar, y medimos un incremento de la hidrólisis, que se buscaba aumentar. Aunque en la mayoría de las variantes presentadas el cambio de selectividad T/H se obtuvo con la pérdida de la actividad principal de cada enzima, mientras la actividad secundaria permaneció intacta o se redujo en menor medida.

El algoritmo propuesto en este trabajo permitió el cambio de selectividad a través de sustituciones de residuos fuera del sitio catalítico que no se eligieron al azar. Sin

embargo, está limitado por la cantidad de estructuras disponibles de un grupo de enzimas.

Igualmente, al analizar los contactos entre residuos generamos parámetros no solo para elegir sitios de mutagénesis que cambien la selectividad de una enzima, también definimos parámetros para agrupar las estructuras de las enzimas de la familia GH13 según su especificidad y subfamilias, identificando con ello eventos de duplicación y transferencia horizontal de genes; estos parámetros son más sensibles que el análisis por secuencias y son susceptibles de automatización.

8. PERSPECTIVAS

Para hacer más eficiente el método se requieren más estructuras, lo cual puede subsanarse a través del uso de inteligencia artificial, la cual puede ser usada para predecir estructuras tridimensionales de secuencias, expandiendo las proteínas a las que se puede aplicar nuestra metodología. Asimismo, el uso de esta herramienta también podría incrementar el poder predictivo, permitiendo el desarrollo de un método que haga predicciones cuantitativas en el efecto que las mutaciones introducidas tendrán en la función.

Asimismo, nuestros resultados sugieren que podría lograrse un cambio mayor en la selectividad T/H al considerar un mayor número de mutaciones. Entre estos resultados podemos mencionar la presencia de clusters con residuos con factores de enriquecimiento altos y nuestra capacidad de clasificar proteínas al calificar todos los residuos como hidrolíticos o transglicosídicos

La comparación de contactos de residuos entre grupos podría encontrar aplicaciones fuera de la función, como sugiere un trabajo en el que la estabilidad de una enzima GH13 se modificó mediante mutagénesis en puntos críticos identificados por análisis de acoplamiento directo [93], creando y eliminando enlaces de hidrógeno entre residuos. También podría identificar sitios donde se pueden insertar sondas fluorescentes mientras se minimiza el deterioro funcional.

La presencia de clusters con factores de enriquecimiento alto, así como la capacidad de clasificar funcionalmente a las enzimas de la familia GH13 a través del análisis conjunto de varios residuos de forma simultánea y nuestros resultados experimentales con una cantidad limitada de mutaciones, indican que una mayor cantidad de residuos debe ser incluidos como blanco de mutagénesis para lograr el cambio de función. La evaluación para su selección se beneficiará con el uso de inteligencia artificial, como lo ha hecho en la predicción de estructuras.

El análisis de contactos usando la relación Función1/Función2 y el coeficiente de similitud encontrados en este trabajo pueden automatizarse para analizar bancos de proteínas y clasificarlas de acuerdo a su función y relaciones evolutivas.

9. REFERENCIAS BIBLIOGRÁFICAS Y BIBLIOGRAFÍA

- 1 Bell, E.L., Finnigan, W., France, S.P., Green, A.P., Hayes, M.A., Hepworth, L.J., Lovelock, S.L., Niikura, H., Osuna, S., Romero, E., Ryan, K.S., Turner, N.J. and Flitsch, S.L. (2021) Biocatalysis. *Nature Reviews Methods Primers*, **1**, 46. <https://doi.org/10.1038/s43586-021-00044-z>.
- 2 Chapman, J., Ismail, A. and Dinu, C. (2018) Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks. *Catalysts*, **8**, 238. <https://doi.org/10.3390/catal8060238>.
- 3 Farooq, M.A., Ali, S., Hassan, A., Tahir, H.M., Mumtaz, S. and Mumtaz, S. (2021) Biosynthesis and Industrial Applications of α -Amylase: A Review. *Archives of Microbiology*, **203**, 1281–1292. <https://doi.org/10.1007/s00203-020-02128-y>.
- 4 Sheldon, R.A. and Brady, D. (2021) Streamlining Design, Engineering, and Applications of Enzymes for Sustainable Biocatalysis. *ACS Sustainable Chemistry & Engineering*, **9**, 8032–8052. <https://doi.org/10.1021/acssuschemeng.1c01742>.
- 5 Silverman, R. (2002) Organic Chemistry of Enzyme-Catalyzed Reactions. 2nd Editio., Elsevier, Austin. <https://doi.org/10.1016/C2009-0-22837-6>.
- 6 Ribeiro, A.J.M., Holliday, G.L., Furnham, N., Tyzack, J.D., Ferris, K. and Thornton, J.M. (2018) Mechanism and Catalytic Site Atlas (M-CSA): A Database of Enzyme Reaction Mechanisms and Active Sites. *Nucleic Acids Research*, **46**, D618–D623. <https://doi.org/10.1093/nar/gkx1012>.
- 7 Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S. and Baker, D. (2008) Kemp Elimination Catalysts by Computational Enzyme Design. *Nature*, **453**, 190–195. <https://doi.org/10.1038/nature06879>.
- 8 Mendonça, L.M.F. and Marana, S.R. (2011) Single Mutations Outside the Active Site Affect the Substrate Specificity in a β -Glycosidase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1814**, 1616–1623. <https://doi.org/10.1016/j.bbapap.2011.08.012>.

- 9 Guo, H.H., Choe, J. and Loeb, L.A. (2004) Protein Tolerance to Random Amino Acid Change. *Proceedings of the National Academy of Sciences*, **101**, 9205–9210. <https://doi.org/10.1073/pnas.0403255101>.
- 10 Schmiedel, J.M. and Lehner, B. (2019) Determining Protein Structures Using Deep Mutagenesis. *Nature Genetics*, **51**, 1177–1186. <https://doi.org/10.1038/s41588-019-0431-x>.
- 11 Rollins, N.J., Brock, K.P., Poelwijk, F.J., Stiffler, M.A., Gauthier, N.P., Sander, C. and Marks, D.S. (2019) Inferring Protein 3D Structure from Deep Mutation Scans. *Nature Genetics*, **51**, 1170–1176. <https://doi.org/10.1038/s41588-019-0432-9>.
- 12 Bandaru, P., Shah, N.H., Bhattacharyya, M., Barton, J.P., Kondo, Y., Cofsky, J.C., Gee, C.L., Chakraborty, A.K., Kortemme, T., Ranganathan, R. and Kuriyan, J. (2017) Deconstruction of the Ras Switching Cycle through Saturation Mutagenesis. *eLife*, **6**. <https://doi.org/10.7554/eLife.27810>.
- 13 Bedbrook, C.N., Rice, A.J., Yang, K.K., Ding, X., Chen, S., LeProust, E.M., Gradinaru, V. and Arnold, F.H. (2017) Structure-Guided SCHEMA Recombination Generates Diverse Chimeric Channelrhodopsins. *Proceedings of the National Academy of Sciences*, **114**. <https://doi.org/10.1073/pnas.1700269114>.
- 14 Salinas, V.H. and Ranganathan, R. (2018) Coevolution-Based Inference of Amino Acid Interactions Underlying Protein Function. *eLife*, **7**. <https://doi.org/10.7554/eLife.34300>.
- 15 Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D.S., Sander, C., Zecchina, R., Onuchic, J.N., Hwa, T. and Weigt, M. (2011) Direct-Coupling Analysis of Residue Coevolution Captures Native Contacts across Many Protein Families. *Proceedings of the National Academy of Sciences*, **108**, E1293–E1301. <https://doi.org/10.1073/pnas.1111471108>.
- 16 Lin, X., Hong, T., Mu, Y. and Torres, J. (2012) Identification of Residues Involved in Water versus Glycerol Selectivity in Aquaporins by Differential Residue Pair Co-Evolution. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, **1818**, 907–914. <https://doi.org/10.1016/j.bbamem.2011.12.017>.
- 17 Chakrabarti, S. and Panchenko, A.R. (2009) Coevolution in Defining the Functional Specificity. *Proteins: Structure, Function, and Bioinformatics*, **75**, 231–240. <https://doi.org/10.1002/prot.22239>.
- 18 Wang, X., Jing, X., Deng, Y., Nie, Y., Xu, F., Xu, Y., Zhao, Y., Hunt, J.F., Montelione, G.T. and Szyperski, T. (2020) Evolutionary Coupling Saturation Mutagenesis: Coevolution-guided Identification of Distant Sites Influencing *Bacillus Naganoensis* Pullulanase Activity. *FEBS Letters*, **594**, 799–812. <https://doi.org/10.1002/1873-3468.13652>.
- 19 Tondnevis, F., Dudenhausen, E.E., Miller, A.M., McKenna, R., Altschul, S.F., Bloom, L.B. and Neuwald, A.F. (2020) Deep Analysis of Residue Constraints (DARC): Identifying Determinants of Protein Functional Specificity. *Scientific Reports*, **10**, 1691. <https://doi.org/10.1038/s41598-019-55118-6>.
- 20 Lee, E.-S.A., Fung, S., Sze-To, H.-Y. and Wong, A.K.C. (2014) Discovering Co-Occurring Patterns and Their Biological Significance in Protein Families. *BMC Bioinformatics*, **15**, S2. <https://doi.org/10.1186/1471-2105-15-S12-S2>.
- 21 Bandyopadhyay, D., Huan, J., Prins, J., Snoeyink, J., Wang, W. and Tropsha, A. (2009) Identification of Family-Specific Residue Packing Motifs and Their Use for Structure-Based Protein Function Prediction: II. Case Studies and Applications. *Journal of*

- Computer-Aided Molecular Design*, **23**, 785–797. <https://doi.org/10.1007/s10822-009-9277-0>.
- 22 O’Meara, M.J., Leaver-Fay, A., Tyka, M.D., Stein, A., Houlihan, K., DiMaio, F., Bradley, P., Kortemme, T., Baker, D., Snoeyink, J. and Kuhlman, B. (2015) Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *Journal of Chemical Theory and Computation*, **11**, 609–622. <https://doi.org/10.1021/ct500864r>.
- 23 Schmitz, S., Ertelt, M., Merkl, R. and Meiler, J. (2021) Rosetta Design with Co-Evolutionary Information Retains Protein Function. Wallqvist, A., Ed., *PLOS Computational Biology*, **17**, e1008568. <https://doi.org/10.1371/journal.pcbi.1008568>.
- 24 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., van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N. V., Adams, P.D., Read, R.J. and Baker, D. (2021) Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science*, **373**, 871–876. <https://doi.org/10.1126/science.abj8754>.
- 25 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Židek, 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., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P. and Hassabis, D. (2021) Highly Accurate Protein Structure Prediction with AlphaFold. *Nature*, **596**, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- 26 Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Židek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D. and Velankar, S. (2022) AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Research*, **50**, D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- 27 Vassura, M., Di Lena, P., Margara, L., Mirto, M., Aloisio, G., Fariselli, P. and Casadio, R. (2011) Blurring Contact Maps of Thousands of Proteins: What We Can Learn by Reconstructing 3D Structure. *BioData Mining*, **4**, 1. <https://doi.org/10.1186/1756-0381-4-1>.
- 28 Hu, Jingjing, M.J. (2002) Mining Protein Contact Maps. In: Zaki, Mohammed J., Wang, Jason T. L., Toivonen, H.T.T., Ed., Proceedings of the 2nd ACM SIGKDD Workshop on Data Mining in Bioinformatics, Edmonton, 3–10.
- 29 Gupta, N., Mangal, N. and Biswas, S. (2005) Evolution and Similarity Evaluation of Protein Structures in Contact Map Space. *Proteins: Structure, Function, and Bioinformatics*, **59**, 196–204. <https://doi.org/10.1002/prot.20415>.
- 30 Pulim, V., Berger, B. and Bienkowska, J. (2008) Optimal Contact Map Alignment of Protein–Protein Interfaces. *Bioinformatics*, **24**, 2324–2328. <https://doi.org/10.1093/bioinformatics/btn432>.

- 31 Heal, J.W., Bartlett, G.J., Wood, C.W., Thomson, A.R. and Woolfson, D.N. (2018) Applying Graph Theory to Protein Structures: An Atlas of Coiled Coils. Valencia, A., Ed., *Bioinformatics*, **34**, 3316–3323. <https://doi.org/10.1093/bioinformatics/bty347>.
- 32 Chou, K.-C. (1999) A Key Driving Force in Determination of Protein Structural Classes. *Biochemical and Biophysical Research Communications*, **264**, 216–224. <https://doi.org/10.1006/bbrc.1999.1325>.
- 33 Glycogenomics Group at AFMB Carbohydrate-Active enZymes Database. <http://www.cazy.org/>.
- 34 Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. (2014) The Carbohydrate-Active Enzymes Database (CAZy) in 2013. *Nucleic Acids Research*, **42**, D490–D495. <https://doi.org/10.1093/nar/gkt1178>.
- 35 Møller, M.S. and Svensson, B. (2016) Structural Biology of Starch-Degrading Enzymes and Their Regulation. *Current Opinion in Structural Biology*, **40**, 33–42. <https://doi.org/10.1016/j.sbi.2016.07.006>.
- 36 Uitdehaag, J.C.M., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G. and Dijkstra, B.W. (1999) X-Ray Structures along the Reaction Pathway of Cyclodextrin Glycosyltransferase Elucidate Catalysis in the α -Amylase Family. *Nature Structural Biology*, **6**, 432–436. <https://doi.org/10.1038/8235>.
- 37 Danby, P.M. and Withers, S.G. (2016) Advances in Enzymatic Glycoside Synthesis. *ACS Chemical Biology*, **11**, 1784–1794. <https://doi.org/10.1021/acscchembio.6b00340>.
- 38 Smith, P.J., Ortiz-Soto, M.E., Roth, C., Barnes, W.J., Seibel, J., Urbanowicz, B.R. and Pfrenge, F. (2020) Enzymatic Synthesis of Artificial Polysaccharides. *ACS Sustainable Chemistry & Engineering*, **8**, 11853–11871. <https://doi.org/10.1021/acssuschemeng.0c03622>.
- 39 Ouadhi, S., López, D.M.V., Mohideen, F.I. and Kwan, D.H. (2023) Engineering the Enzyme Toolbox to Tailor Glycosylation in Small Molecule Natural Products and Protein Biologics. *Protein Engineering, Design and Selection*, **36**. <https://doi.org/10.1093/protein/gzac010>.
- 40 Mészáros, Z., Nekvasilová, P., Bojarová, P., Křen, V. and Slámová, K. (2021) Reprint of: Advanced Glycosidases as Ingenious Biosynthetic Instruments. *Biotechnology Advances*, **51**, 107820. <https://doi.org/10.1016/j.biotechadv.2021.107820>.
- 41 Saab-Rincon, G., Llopiz, A. and Arreola-Barroso, R. (2023) The Use of Biocatalysis for Biosurfactant Production. *Biosurfactants*, Elsevier, 265–301. <https://doi.org/10.1016/B978-0-323-91697-4.00012-0>.
- 42 Janecek, S. (1992) New Conserved Amino Acid Region of α -Amylases in the Third Loop of Their (β/α)₈-Barrel Domains. *Biochemical Journal*, **288**, 1069–1070. <https://doi.org/10.1042/bj2881069>.
- 43 Janeček, Š. (2002) How Many Conserved Sequence Regions Are There in the α -Amylase Family? *Biologia*, **57**, 29–41.
- 44 Wong, D.W.S., Batt, S.B., Lee, C.C. and Robertson, G.H. (2004) High-Activity Barley α -Amylase by Directed Evolution. *The Protein Journal*, **23**, 453–460. <https://doi.org/10.1007/s10930-004-5221-2>.
- 45 Liu, Y., Fan, S., Liu, X., Zhang, Z., Wang, J., Wang, Z. and Lu, F. (2014) A Highly Active Alpha Amylase from *Bacillus Licheniformis*: Directed Evolution, Enzyme Characterization and Structural Analysis. *Journal of Microbiology and Biotechnology*, **24**, 898–904. <https://doi.org/10.4014/jmb.1402.02004>.

- 46 Shim, J.-H., Kim, Y.-W., Kim, T.-J., Chae, H.-Y., Park, J.-H., Cha, H., Kim, J.-W., Kim, Y.-R., Schaefer, T., Spender, T., Moon, T.-W. and Park, K.-H. (2004) Improvement of Cyclodextrin Glucanotransferase as an Antistaling Enzyme by Error-Prone PCR. *Protein Engineering Design and Selection*, **17**, 205–211. <https://doi.org/10.1093/protein/gzh035>.
- 47 Kelly, R.M., Dijkhuizen, L. and Leemhuis, H. (2009) Starch and α -Glucan Acting Enzymes, Modulating Their Properties by Directed Evolution. *Journal of Biotechnology*, **140**, 184–193. <https://doi.org/10.1016/j.jbiotec.2009.01.020>.
- 48 Tran, P.L., Cha, H.-J., Lee, J.-S., Park, S.-H., Woo, E.-J. and Park, K.-H. (2014) Introducing Transglycosylation Activity in *Bacillus Licheniformis* α -Amylase by Replacement of His235 with Glu. *Biochemical and Biophysical Research Communications*, **451**, 541–547. <https://doi.org/10.1016/j.bbrc.2014.08.019>.
- 49 Damián-Almazo, J.Y., Moreno, A., López-Munguía, A., Soberón, X., González-Muñoz, F. and Saab-Rincón, G. (2008) Enhancement of the Alcoholic Activity of α -Amylase AmyA from *Thermotoga Maritima* MSB8 (DSM 3109) by Site-Directed Mutagenesis. *Applied and Environmental Microbiology*, **74**, 5168–5177. <https://doi.org/10.1128/AEM.00121-08>.
- 50 Rivera, M.H., Lopez-Munguia, A., Soberon, X. and Saab-Rincon, G. (2003) Alpha-Amylase from *Bacillus Licheniformis* Mutants near to the Catalytic Site: Effects on Hydrolytic and Transglycosylation Activity. *Protein Engineering Design and Selection*, **16**, 505–514. <https://doi.org/10.1093/protein/gzg060>.
- 51 Saab-Rincón, G., Del-Río, G., Santamaría, R.I., López-Munguía, A. and Soberón, X. (1999) Introducing Transglycosylation Activity in a Liquefying α -Amylase. *FEBS Letters*, **453**, 100–106. [https://doi.org/10.1016/S0014-5793\(99\)00671-7](https://doi.org/10.1016/S0014-5793(99)00671-7).
- 52 Teze, D., Hendrickx, J., Dion, M., Tellier, C., Woods, V.L., Tran, V. and Sanejouand, Y.-H. (2013) Conserved Water Molecules in Family 1 Glycosidases: A DXMS and Molecular Dynamics Study. *Biochemistry*, **52**, 5900–5910. <https://doi.org/10.1021/bi400260b>.
- 53 David, B., Arnaud, P., Tellier, C. and Sanejouand, Y.-H. (2019) Toward the Design of Efficient Transglycosidases: The Case of the GH1 of *Thermus Thermophilus*. Daggett, V., Ed., *Protein Engineering, Design and Selection*. <https://doi.org/10.1093/protein/gzz032>.
- 54 Lundemo, P., Karlsson, E.N. and Adlercreutz, P. (2017) Eliminating Hydrolytic Activity without Affecting the Transglycosylation of a GH1 β -Glucosidase. *Applied Microbiology and Biotechnology*, **101**, 1121–1131. <https://doi.org/10.1007/s00253-016-7833-9>.
- 55 Kelly, R.M., Leemhuis, H. and Dijkhuizen, L. (2007) Conversion of a Cyclodextrin Glucanotransferase into an α -Amylase: Assessment of Directed Evolution Strategies. *Biochemistry*, **46**, 11216–11222. <https://doi.org/10.1021/bi701160h>.
- 56 Leemhuis, H., Rozeboom, H.J., Wilbrink, M., Euverink, G.-J.W., Dijkstra, B.W. and Dijkhuizen, L. (2003) Conversion of Cyclodextrin Glycosyltransferase into a Starch Hydrolase by Directed Evolution: The Role of Alanine 230 in Acceptor Subsite +1. *Biochemistry*, **42**, 7518–7526. <https://doi.org/10.1021/bi034439q>.
- 57 Koutrouli, M., Karatzas, E., Paez-Espino, D. and Pavlopoulos, G.A. (2020) A Guide to Conquer the Biological Network Era Using Graph Theory. *Frontiers in Bioengineering and Biotechnology*, **8**. <https://doi.org/10.3389/fbioe.2020.00034>.

- 58 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., Frost, A., Basquin, J., Lindorff-Larsen, K., Bateman, A., Kajava, A. V., Valencia, A., Ovchinnikov, S., Durairaj, J., Ascher, D.B., Thornton, J.M., Davey, N.E., Stein, A., Elofsson, A., Croll, T.I. and Beltrao, P. (2022) A Structural Biology Community Assessment of AlphaFold2 Applications. *Nature Structural & Molecular Biology*, **29**, 1056–1067. <https://doi.org/10.1038/s41594-022-00849-w>.
- 59 Swiss-Model Modelling. <https://swissmodel.expasy.org/interactive>.
- 60 Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L. and Schwede, T. (2017) The SWISS-MODEL Repository—New Features and Functionality. *Nucleic Acids Research*, **45**, D313–D319. <https://doi.org/10.1093/nar/gkw1132>.
- 61 Vehlou, C., Stehr, H., Winkelmann, M., Duarte, J.M., Petzold, L., Dinse, J. and Lappe, M. (2011) CMView: Interactive Contact Map Visualization and Analysis. *Bioinformatics*, **27**, 1573–1574. <https://doi.org/10.1093/bioinformatics/btr163>.
- 62 Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Research*, **22**, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
- 63 Holm, L. and Rosenström, P. (2010) Dali Server: Conservation Mapping in 3D. *Nucleic Acids Research*, **38**, W545–W549. <https://doi.org/10.1093/nar/gkq366>.
- 64 Freeman, L.C. (1978) Centrality in Social Networks Conceptual Clarification. *Social Networks*, **1**, 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7).
- 65 Sarkar, G. and Sommer, S.S. (1990) The “Megaprimer” Method of Site-Directed Mutagenesis. *BioTechniques*, England, **8**, 404–407.
- 66 Liu, H. and Naismith, J.H. (2008) An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnology*, **8**, 91. <https://doi.org/10.1186/1472-6750-8-91>.
- 67 Saab-Rincón, G., Del-Río, G., Santamaría, R.I., López-Munguía, A. and Soberón, X. (1999) Introducing Transglycosylation Activity in a Liquefying α -Amylase. *FEBS Letters*, **453**, 100–106. [https://doi.org/10.1016/S0014-5793\(99\)00671-7](https://doi.org/10.1016/S0014-5793(99)00671-7).
- 68 Liebl, W., Stemplinger, I. and Ruile, P. (1997) Properties and Gene Structure of the Thermotoga Maritima Alpha-Amylase AmyA, a Putative Lipoprotein of a Hyperthermophilic Bacterium. *Journal of Bacteriology*, **179**, 941–948. <https://doi.org/10.1128/jb.179.3.941-948.1997>.
- 69 Xiao, Z., Storms, R. and Tsang, A. (2006) A Quantitative Starch–Iodine Method for Measuring Alpha-Amylase and Glucoamylase Activities. *Analytical Biochemistry*, **351**, 146–148. <https://doi.org/10.1016/j.ab.2006.01.036>.
- 70 Mould, D.L. (1954) Potentiometric and Spectrophotometric Studies of Complexes of Hydrolysis Products of Amylose with Iodine and Potassium Iodide. *Biochemical Journal*, **58**, 593–600. <https://doi.org/10.1042/bj0580593>.
- 71 Xiao, Z., Storms, R. and Tsang, A. (2006) A Quantitative Starch–Iodine Method for Measuring Alpha-Amylase and Glucoamylase Activities. *Analytical Biochemistry*, **351**, 146–148. <https://doi.org/10.1016/j.ab.2006.01.036>.

- 72 Miller, G.L. (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*, **31**, 426–428. <https://doi.org/10.1021/ac60147a030>.
- 73 Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B. and Lindahl, E. (2015) GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX*, **1–2**, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- 74 Sefidbakht, Y., Ranaei Siadat, O. and Taheri, F. (2017) Homology Modeling and Molecular Dynamics Study on Schwanniomyces Occidentalis Alpha-Amylase. *Journal of Biomolecular Structure and Dynamics*, **35**, 574–584. <https://doi.org/10.1080/07391102.2016.1154892>.
- 75 Protein Data Bank. <https://www.rcsb.org/>.
- 76 Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera-A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry*, **25**, 1605–1612. <https://doi.org/10.1002/jcc.20084>.
- 77 Fiser, A., Do, R.K.G. and Šali, A. (2000) Modeling of Loops in Protein Structures. *Protein Science*, **9**, 1753–1773. <https://doi.org/10.1110/ps.9.9.1753>.
- 78 Grace. <http://plasma-gate.weizmann.ac.il/Grace/>.
- 79 Li, H., Robertson, A.D. and Jensen, J.H. (2005) Very Fast Empirical Prediction and Rationalization of Protein pKa Values. *Proteins: Structure, Function, and Bioinformatics*, **61**, 704–721. <https://doi.org/10.1002/prot.20660>.
- 80 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R. and Schwede, T. (2018) SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research*, **46**, W296–W303. <https://doi.org/10.1093/nar/gky427>.
- 81 Holmes, J.B. and Tsai, J. (2005) Characterizing Conserved Structural Contacts by Pair-Wise Relative Contacts and Relative Packing Groups. *Journal of Molecular Biology*, **354**, 706–721. <https://doi.org/10.1016/j.jmb.2005.09.081>.
- 82 Koropatkin, N.M. and Smith, T.J. (2010) SusG: A Unique Cell-Membrane-Associated α -Amylase from a Prominent Human Gut Symbiont Targets Complex Starch Molecules. *Structure*, **18**, 200–215. <https://doi.org/10.1016/j.str.2009.12.010>.
- 83 Shipman, J.A., Cho, K.H., Siegel, H.A. and Salyers, A.A. (1999) Physiological Characterization of SusG, an Outer Membrane Protein Essential for Starch Utilization by Bacteroides Thetaiotaomicron. *Journal of Bacteriology*, **181**, 7206–7211. <https://doi.org/10.1128/JB.181.23.7206-7211.1999>.
- 84 Doshi, U., Holliday, M.J., Eisenmesser, E.Z. and Hamelberg, D. (2016) Dynamical Network of Residue–Residue Contacts Reveals Coupled Allosteric Effects in Recognition, Catalysis, and Mutation. *Proceedings of the National Academy of Sciences*, **113**, 4735–4740. <https://doi.org/10.1073/pnas.1523573113>.
- 85 David, B., Arnaud, P., Tellier, C. and Sanejouand, Y.-H. (2019) Toward the Design of Efficient Transglycosidases: The Case of the GH1 of Thermus Thermophilus. Daggett, V., Ed., *Protein Engineering, Design and Selection*. <https://doi.org/10.1093/protein/gzz032>.
- 86 Kong, D., Wang, L., Su, L. and Wu, J. (2021) Effect of Leu 277 on Disproportionation and Hydrolysis Activity in Bacillus Stearothermophilus NO2 Cyclodextrin

- Glucosyltransferase. Kelly, R.M., Ed., *Applied and Environmental Microbiology*, **87**. <https://doi.org/10.1128/AEM.03151-20>.
- 87 Geronimo, I., Payne, C.M. and Sandgren, M. (2018) The Role of Catalytic Residue p K a on the Hydrolysis/Transglycosylation Partition in Family 3 β -Glucosidases. *Organic & Biomolecular Chemistry*, **16**, 316–324. <https://doi.org/10.1039/C7OB02558K>.
- 88 Wang, X., Jing, X., Deng, Y., Nie, Y., Xu, F., Xu, Y., Zhao, Y., Hunt, J.F., Montelione, G.T. and Szyperski, T. (2020) Evolutionary Coupling Saturation Mutagenesis: Coevolution-guided Identification of Distant Sites Influencing *Bacillus Naganoensis* Pullulanase Activity. *FEBS Letters*, **594**, 799–812. <https://doi.org/10.1002/1873-3468.13652>.
- 89 Tondnevis, F., Dudenhausen, E.E., Miller, A.M., McKenna, R., Altschul, S.F., Bloom, L.B. and Neuwald, A.F. (2020) Deep Analysis of Residue Constraints (DARC): Identifying Determinants of Protein Functional Specificity. *Scientific Reports*, **10**, 1691. <https://doi.org/10.1038/s41598-019-55118-6>.
- 90 Hleap, J.S. and Blouin, C. (2018) The Response to Selection in Glycoside Hydrolase Family 13 Structures: A Comparative Quantitative Genetics Approach. Silman, I., Ed., *PLOS ONE*, **13**, e0196135. <https://doi.org/10.1371/journal.pone.0196135>.
- 91 Geronimo, I., Payne, C.M. and Sandgren, M. (2018) The Role of Catalytic Residue p K a on the Hydrolysis/Transglycosylation Partition in Family 3 β -Glucosidases. *Organic & Biomolecular Chemistry*, **16**, 316–324. <https://doi.org/10.1039/C7OB02558K>.
- 92 Madani, A., Krause, B., Greene, E.R., Subramanian, S., Mohr, B.P., Holton, J.M., Olmos, J.L., Xiong, C., Sun, Z.Z., Socher, R., Fraser, J.S. and Naik, N. (2023) Large Language Models Generate Functional Protein Sequences across Diverse Families. *Nature Biotechnology*. <https://doi.org/10.1038/s41587-022-01618-2>.
- 93 Wang, C., Huang, R., He, B. and Du, Q. (2012) Improving the Thermostability of Alpha-Amylase by Combinatorial Coevolving-Site Saturation Mutagenesis. *BMC Bioinformatics*, **13**, 263. <https://doi.org/10.1186/1471-2105-13-263>.

Anexo I. Estructuras empleadas en este trabajo

Tabla S1. Conjunto de datos 1. Grupo Interno: Grupo de estructuras del PDB empleadas para calcular el factor de enriquecimiento

PDB ID	Resolución (Å)	Actividad reportada	Actividad sobre enlaces α -(1-4)-glucosídicos	Sub-familia CAZy	Organismo
1A47	2.56	CGTasa	Transferasa	GH13_2	<i>Thermoanaerobacterium thermosulfurigenes</i> (<i>Clostridium thermosulfurogenes</i>)
1UKQ	2.00	CGTasa	Transferasa	GH13_2	<i>Bacillus sp.</i> 1011
2CXG	2.50	CGTasa	Transferasa	GH13_2	<i>Niallia circulans</i> 251
2VR5	2.80	Enzima desramificante	Transferasa	GH13_11	<i>Saccharolobus solfataricus</i> P2
1MXG	1.60	α -Amilasa	Hidrolasa	GH13_7	<i>Pyrococcus woesei</i>
1HX0	1.38	α -Amilasa	Hidrolasa	GH13_24	<i>Sus scrofa</i>
1RPK	2.00	α -Amilasa	Hidrolasa	GH13_6	<i>Hordeum vulgare</i>
1UH3	2.60	α -Amilasa	Hidrolasa	GH13_21	<i>Thermoactinomyces vulgaris</i> R-47
1WPC	1.90	α -Amilasa	Hidrolasa	GH13_5	<i>Bacillus sp.</i> 707
2CPU	2.00	α -Amilasa	Hidrolasa	GH13_24	<i>Homo sapiens</i>
2WC7	2.37	Enzima desramificante	Hidrolasa	GH13_20	<i>Nostoc punctiforme</i> PCC 73102
3BC9	1.35	α -Amilasa	Hidrolasa	GH13_5	<i>Halothermothrix orenii</i> H 168
3K8M	2.50	α -Amilasa	Hidrolasa	GH13_36	<i>Bacteroides thetaiotaomicron</i> VPI-5482
7TAA	1.98	α -Amilasa	Hidrolasa	GH13_1	<i>Aspergillus oryzae</i> DSM63303

Table S2. Conjunto de datos 2. Grupo externo: conjunto de estructuras 3D usadas para evaluar la capacidad de los factores de enriquecimiento para clasificar enzimas de la familia GH13 por función.

PDB ID	Resolución (Å)	Actividad reportada	Actividad sobre enlaces α -(1-4)-glucosídicos	Sub-familia CAZy	Organismo
1CGT	2.00	Glucano-Transferasa	Transferasa	GH13_2	<i>Niallia circulans</i> 8
4JCL	1.70	Glucano-Transferasa	Transferasa	GH13_2	<i>Paenibacillus macerans</i> IAM1243/ IB7 / IFO 3490 (NRRL B-388) / JFB05-01 (CCTCC M203062)
1BLI	1.90	α -Amilasa	Hidrolasa	GH13_5	<i>Bacillus licheniformis</i> 584 / ATCC 27811
1GCY	1.60	α -Amilasa productora de maltotetra-sacáridos	Hidrolasa	NA	<i>Pseudomonas stutzeri</i> MO-19
1WZL	2.00	Neo-pulunasa Ciclomalto dextrinasa	Hidrolasa	GH13_20	<i>Thermoactinomyces vulgaris</i> R-47
AmyA	Homology model	α -Amilasa	Hidrolasa	GH13_36	<i>Thermotoga maritima</i> MSB8
3BC9	1.35	α -Amilasa	Hidrolasa	GH13_5	<i>Halothermothrix orenii</i> H 168
3DHU	2.00	α -Amilasa	Hidrolasa	NA	<i>Lacticaseibacillus plantarum</i> WCFS1
3EDF	1.65	Ciclomalto dextrinasa	Hidrolasa	NA	<i>Flavobacterium</i> sp. 92

Tabla S2. Conjunto de datos 2. Grupo externo (Continuación)

PDB ID	Resolución (Å)	Actividad reportada	Actividad sobre enlaces α- (1-4)- gluco- sídicos	Sub- familia CAZy	Organismo
3VM7	2.25	α -Amilasa	Hidrolasa	GH13_ 1	<i>Malbranchea cinnamomea</i>
4AEE	2.28	α -Amilasa malto- génica Ciclomalto dextrinasa	Hidrolasa	GH13_ 20	<i>Staphylothermus marinus F1</i>
4AEF	2.34	α -Amilasa Ciclomalto dextrinasa	Hidrolasa	GH13_ 20	<i>Pyrococcus furius DSM 3638</i>
4GKL	2.40	α -Amilasa malto- génica	Hidrolasa	NA	<i>Thermotoga neapolitana DSM 4359</i>
4UZU	1.90	α -Amilasa	Hidrolasa	GH13_ 5	<i>Geobacillus stearothermophilus DY5 / PHI300 / NZ-3</i>

Tabla S3. Conjunto de datos 3. Enzimas empleadas para analizar la conservación de contactos y su relación con la historia evolutiva

PDB ID	Resolución (Å)	Familia CAZy	Subfamilia CAZy	Organismo
3VM7	2.25	GH13	1	<i>Malbranchea cinnamomea</i>
7TAA	1.98	GH13	1	<i>Aspergillus oryzae</i> DSM63303
1A47	256	GH13	2	<i>Thermoanaerobacterium thermosulfurigenes</i> EM1
1CGT	2.00	GH13	2	<i>Niallia circulans</i> 8
1QHO	1.70	GH13	2	<i>Geobacillus stearothermophilus</i> C599
2CXG	2.50	GH13	2	<i>Niallia circulans</i> 251
1UKQ	2.00	GH13	2	<i>Bacillus</i> sp. 1011
4JCL	1.70	GH13	2	<i>Paenibacillus macerans</i> IAM1243/ IB7 / IFO 3490 (NRRL B-388) / JFB05-01 (CCTCC M203062)
3ZSS	1.80	GH13	3	<i>Streptomyces coelicolor</i> A3(2)
1WPC	1.90	GH13	5	<i>Bacillus</i> sp. 707
1BLI	1.90	GH13	5	<i>Bacillus licheniformis</i> 584 / ATCC 27811
3BC9	1.35	GH13	5	<i>Halothermothrix orenii</i> H 168
4UZU	1.90	GH13	5	<i>Geobacillus stearothermophilus</i> DY5 / PHI300 / NZ-3
1RPK	2.00	GH13	6	<i>Hordeum vulgare</i>
1MXG	1.60	GH13	7	<i>Pyrococcus woesei</i>
3AMK	1.90	GH13	8	<i>Oryza sativa</i> Japonica Group
3AML	1.70	GH13	8	<i>Oryza sativa</i> Japonica Group
4BZY	2.75	GH13	8	<i>Homo sapiens</i>

Tabla S3. Conjunto de datos 3 (Continuación)

PDB ID	Resolución (Å)	Familia CAZy	Subfamilia CAZy	Organismo
2VR5	2.80	GH13	11	<i>Saccharolobus solfataricus</i> P2
1WZL	2.00	GH13	20	<i>Thermoactinomyces vulgaris</i> R-47
2YA0	1.85	GH13	12	<i>Streptococcus pneumoniae</i> TIGR4
3FAW	2.10	GH13	12	<i>Streptococcus agalactiae</i> COH1
2YOC	2.88	GH13	13	<i>Raoultella ornithinolytica</i> 10-5246
2FHF	1.65	GH13	13	<i>Klebsiella pneumoniae</i> UNF5023
2WC7	2.37	GH13	20	<i>Nostoc punctiforme</i> PCC 73102
4AEE	2.28	GH13	20	<i>Staphylothermus marinus</i> F1
4AEF	2.34	GH13	20	<i>Pyrococcus furiosus</i> DSM 3638
1UH3	2.60	GH13	21	<i>Thermoactinomyces vulgaris</i> R-47
1HX0	1.38	GH13	24	<i>Sus scrofa</i>
2CPU	2.00	GH13	24	<i>Homo sapiens</i>
1UA7	2.21	GH13	28	<i>Bacillus subtilis</i> 2633
1G9H	1.80	GH13	32	<i>Pseudoalteromonas haloplanktis</i> TAB23 / A23
3K8M	2.00	GH13	36	<i>Bacteroides thetaiotaomicron</i> VPI-5482
AmyA		GH13	36	<i>Thermotoga maritima</i> MSB8
1GCY	1.60	GH13	NA	<i>Pseudomonas stutzeri</i> MO-19

Tabla S3. Conjunto de datos 3 (Continuación)

PDB ID	Resolución (Å)	Familia CAZy	Subfamilia CAZy	Organismo
1LWJ	2.50	GH13	NA	<i>Thermotoga maritima</i> MSB8
3DHU	2.00	GH13	NA	<i>Lactocaseibacillus</i> <i>plantarum</i> WCFS1
3EDF	1.65	GH13	NA	<i>Flavobacterium</i> sp. 92
4E2O	2.10	GH13	NA	<i>Geobacillus</i> <i>thermoleovorans</i> CCB_US3_UF5
4GKL	2.40	GH13	NA	<i>Thermotoga neapolitana</i> DSM 4359
2ZQ0	1.60	GH97	NA	<i>Bacteroides</i> <i>thetaiotaomicron</i> VPI- 5482
3W37	1.70	GH31	NA	<i>Beta vulgaris</i>

Tabla S4. Conjunto de datos 4. Enzimas GH13 caracterizadas en CAZy: Enzimas que cuya función (hidrolítica o transglicosídica) se ha reportado como caracterizada en la base de datos CAZy (2019). Para las transferasas todas las secuencias fueron consideradas y modeladas usando Swiss model cuando fue necesario. Para las hidrolasas solo se tomaron enzimas con estructura tridimensional conocida. Para las estructuras tomadas del PDB, reportamos su PDB ID como identificador, mientras para las secuencias modeladas con Swiss model reportamos la clave de Uniprot o GenBank ID, seguida de la PDB ID de la estructura empleada como molde En la columna Resolución registramos el valor de Q-mean para las estructuras producidas por modelado por homología, las cuales usamos después de minimizar su energía con Rosetta.

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α -(1-4)-glucosídicos
1CDG	2.00	β -Ciclodextrin glucanotransferasa (<i>Niallia circulans</i> 251)	Transferasa
1CGT	2.00	β -Ciclodextrin glucanotransferasa (<i>Niallia circulans</i> 8)	Transferasa
1CYG	2.50	α/β -Ciclodextrin glucanotransferasa (<i>Geobacillus stearothermophilus</i>)	Transferasa
1GJU	2.40	4- α -Glucanotransferasa / maltosiltransferasa (<i>Thermotoga maritima</i>)	Transferasa
1LWJ	2.50	4- α -Glucanotransferasa (<i>Thermotoga maritima</i>)	Transferasa
1V3M	2.00	Ciclodextrin glicósido transferasa (<i>Bacillus sp.</i> 1011)	Transferasa
2VR5	2.80	Isoamilase / 4- α -glucanotransferasa (<i>Saccharolobus solfataricus</i> P2)	Transferasa

Tabla S4. Conjunto de datos 4. Enzimas GH13 caracterizadas en CAZy
(Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α -(1-4)-gluco- sídicos
3BMV	1.60	α / β -Ciclodextrin glicósido transferasa (<i>Thermoanaerobacterium thermosulfurigenes</i> EM1)	Transferasa
4JCL	1.70	α -Ciclodextrin glucanotransferasa (<i>Paenibacillus macerans</i> IAM1243)	Transferasa
4JCM	1.65	γ -Ciclodextrin glucanotransferasa (<i>Evansella clarkii</i>)	Transferasa
A2QTS4_1cgt	QMEAN -2.64	4- α -glucanotransferasa de pared celular (<i>Aspergillus niger</i> CBS 513.88)	Transferasa
A2QYT9_2aaa	QMEAN -2.48	4- α -Glucanotransferasa (<i>Aspergillus niger</i> CBS 513.88)	Transferasa
P26827_1uks (1a47)	QMEAN -1.85	Ciclodextrin glucanotransferasa (<i>Thermoanaerobacterium thermosulfurigenes</i>)	Transferasa
A0A077D499_3bmv	QMEAN -0.30	Ciclodextrin glicósido- transferasa (uncultured <i>Carboxydocella</i>)	Transferasa
A0A0C4WMI 5_1ciu	QMEAN 0.72	Ciclodextrin glicósido transferasa (<i>Thermoanaerobacter</i> sp. P4)	Transferasa
B2D1U4_1ukt	QMEAN 0.33	Ciclodextrin glucanotransferasa (<i>Paenibacillus</i> sp. JB-13)	Transferasa

Tabla S4. Conjunto de datos 4 Enzimas GH13 caracterizadas en CAZy
(Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α-(1→4) gluco-sídicos
B2XY82_4jcl	QMEAN -0.05	Ciclodextrin glicósido transferasa (<i>Paenibacillus graminis</i> MC22.13)	Transferasa
C9WB02_1cgt	QMEAN -0.01	β -Ciclodextrin glicósidotransferasa (<i>Paenibacillus illinoisensis</i> ZY-08 / ZY-8)	Transferasa
B1VC16_6cgt	QMEAN 0.07	β -Ciclodextrin glicósidotransferasa (<i>Paenibacillus pabuli</i> US132)	Transferasa
O82984_1ukq	QMEAN -0.77	β -Ciclodextrin glicósidotransferasa (<i>Bacillus sp.</i> A2-5A)	Transferasa
O86956_1lwh	QMEAN -0.70	4- α -glucanotransferasa (<i>Thermotoga neapolitana</i>)	Transferasa
O86959_1jfg	QMEAN -2.52	Ciclomaltodextrinasa glucanotransferasa (<i>Thermotoga neapolitana</i>)	Transferasa
P08704_4jcm	QMEAN -3.00	α -Ciclodextrin glucanotransferasa (<i>Klebsiella pneumoniae</i> M5a1)	Transferasa
P17692_1cdg	QMEAN 0.19	β -cyclomaltodextrin glucanotransferasa (<i>Bacillus sp.</i> B1018)	Transferasa
P31746_1pj9	QMEAN -0.59	β -Ciclodextrin glucanotransferasa (<i>Bacillus sp.</i> 1-1)	Transferasa
P31747_6cgt	QMEAN 0.04	Ciclodextrin glucanotransferasa (<i>Bacillus sp.</i> 6.3.3)	Transferasa

Tabla S4. Conjunto de datos Enzimas GH13 caracterizadas en CAZy (Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α -(1→4) gluco-sídicos
P31835_6aij	QMEAN -0.30	α -Ciclodextrin glucanotransferasa (<i>Paenibacillus macerans</i>)	Transferasa
Q3HUR2_6aij	QMEAN -3.55	β -Ciclodextrin glucanotransferasa (<i>Pyrococcus furiosus</i> DSM 3638)	Transferasa
Q8X268_1uks	QMEAN -1.82	β -Ciclodextrin glucanotransferasa (<i>Thermococcus kodakarensis</i> KOD1)	Transferasa
Q9UWN2_1qho	QMEAN -1.51	α -Ciclodextrin glucanotransferasa (<i>Thermococcus sp.</i> B1001)	Transferasa
Q9ZAQ0_3bmv	QMEAN -0.63	β -Ciclodextrin glucanotransferasa (<i>Geobacillus stearothermophilus</i> ET1)	Transferasa
Q25CB6_4jcm	QMEAN -0.45	γ -ciclomaltodextrin glucanotransferasa (<i>Bacillus sp.</i> G-825-6)	Transferasa
Q53I75_1ukq	QMEAN -0.07	α -ciclomaltodextrin glucanotransferasa (<i>Haloferax mediterranei</i>)	Transferasa
1AVA	1.9	α -Amilasa (<i>Hordeum vulgare</i>)	Hidrolasa
1EA9	3.20	Ciclomaltodextrinasa (<i>Bacillus sp.</i>)	Hidrolasa
1G94	1.74	α -Amilasa (<i>Pseudoalteromonas haloplanktis</i>)	Hidrolasa

Tabla S4. Conjunto de datos Enzimas GH13 caracterizadas en CAZy (Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α -(1→4) gluco-sídicos
1GCY	1.60	Amilasa formadora de maltotetraosa (<i>Pseudomonas stutzeri</i>)	Hidrolasa
1HT6	1.50	α -Amilasa (<i>Hordeum vulgare</i>)	Hidrolasa
1HX0	1.38	α -Amilasa (<i>Sus scrofa</i>)	Hidrolasa
1J0H	1.90	Ciclomaltodextrinasa / neopululanasa (<i>Geobacillus stearothermophilus</i> TRS40)	Hidrolasa
1JAE	1.65	α -Amilasa (<i>Tenebrio molitor</i>)	Hidrolasa
1JI1	1.60	α -Amilasa (<i>Thermoactinomyces vulgaris</i> R-47)	Hidrolasa
1MWO	2.20	α -Amilasa (<i>Pyrococcus woesei</i>)	Hidrolasa
1QHO	1.70	α -amilase maltogénica (<i>Geobacillus stearothermophilus</i> C599)	Hidrolasa
1SMA	2.80	α -amilase maltogénica (<i>Thermus sp.</i> IM6501)	Hidrolasa
1SMD	1.60	α -Amilasa (<i>Homo sapiens</i> , salivary)	Hidrolasa
1UA7	2.21	α -Amilasa (<i>Bacillus subtilis</i> 2633)	Hidrolasa
1UD2	2.13	α -Amilasa (<i>Bacillus sp.</i> KSM-K38)	Hidrolasa
1VJS	1.70	α -Amilasa (<i>Bacillus licheniformis</i> 584)	Hidrolasa

Tabla S4. Conjunto de datos 4 Enzimas GH13 caracterizadas en CAZy
(Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α-(1→4) gluco-sídicos
1WPC	1.90	Maltohexaose-forming Amilasa (<i>Bacillus sp.</i> 707)	Hidrolasa
1WZA	1.60	α -Amilasa (<i>Halothermothrix orenii</i> H 168)	Hidrolasa
2D2O	2.10	Ciclomaltodextrinasa / neopululanasa (<i>Thermoactinomyces vulgaris</i> R-47)	Hidrolasa
2DIE	2.10	α -Amilasa (<i>Bacillus sp.</i> KSM-1378)	Hidrolasa
2GJP	1.90	α -Amilasa (<i>Sutcliffiella halmapala</i>)	Hidrolasa
2YA0	1.85	Enzima degradadora de glucógeno (<i>Streptococcus pneumoniae</i> TIGR4)	Hidrolasa
2ZE0	2.00	α -Glucosidasa (<i>Geobacillus sp.</i> HTA- 462)	Hidrolasa
3BC9	1.35	α -Amilasa (<i>Halothermothrix orenii</i> H 168)	Hidrolasa
3BH4	1.40	α -Amilasa (<i>Bacillus amyloliquefaciens</i>)	Hidrolasa
3DC0	2.78	α -Amilasa (<i>Bacillus sp.</i> KR8104)	Hidrolasa
3EDJ	1.69	Ciclomaltodextrinasa (<i>Flavobacterium sp.</i> 92)	Hidrolasa

Tabla S4. Conjunto de datos 4 Enzimas GH13 caracterizadas en CAZy
(Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α -(1→4) gluco- sídicos
3VGF	2.30	α -Amilasa / Maltooligosiltrehalose (<i>Saccharolobus solfataricus</i> KM1)	Hidrolasa
3WN6	2.16	α -Amilasa (<i>Oryza sativa</i> Japonica Group)	Hidrolasa
3WY2	1.47	α -Glucosidasa (<i>Halomonas</i> sp. H11)	Hidrolasa
4AEE	2.28	α -Amilasa maltogénica / Ciclomaltodextrinasa (<i>Staphylothermus marinus</i> F1)	Hidrolasa
4AEF	2.34	α -Amilasa maltogénica/ Ciclomaltodextrinasa (<i>Pyrococcus furiosus</i> DSM 3638)	Hidrolasa
4E2O	2.10	α -Amilasa (<i>Geobacillus</i> <i>thermoleovorans</i> CCB_US3_UF5)	Hidrolasa
4GKL	2.40	α -Amilasa maltogénica (<i>Thermotoga neapolitana</i> DSM 4359)	Hidrolasa
4UZU	1.90	α -Amilasa (<i>Geobacillus</i> <i>stearothermophilus</i> DY5)	Hidrolasa
5A2B	1.85	α -Amilasa (<i>Anoxybacillus</i> sp. SK3-4)	Hidrolasa
5H06	1.95	α -Amilasa (bacteria no cultivable)	Hidrolasa

Tabla S4. Conjunto de datos 4 Enzimas GH13 caracterizadas en CAZy
(Continuación)

Identificador	Resolución	Actividad reportada (especie)	Actividad sobre enlaces α-(1-4)- gluco-sídicos
6A0J	1.60	Ciclo maltosil- maltose Hidrolasa (<i>Arthrobacter globiformis</i> M6)	Hidrolasa
6BS6	2.17	α -Amilasa / neopululanasa (<i>Bacteroides thetaiotaomicron</i> VPI- 5482)	Hidrolasa
7TAA	1.98	α -Amilasa (<i>Aspergillus oryzae</i> DSM63303)	Hidrolasa

Anexo II Oligonucleótidos empleados en este trabajo

Tabla S5. Cebadores empleados para introducir las mutaciones en este estudio

Enzima mutada	Mutación	Secuencia (5'-3')
AmyA	K98P/D99A	ACAGACTACTACAACGTCGAGC CGGCGTACGGCACCATGGAAGATCTC
glucanotransferasa	F72L	CGATCATCTCTTTCAACTCTCTC TCACTACC
glucanotransferasa	V86I	GGGAAGGTCAAGAACGATTTTT ATTCCGC
glucanotransferasa	T274V	CCCGGTGAATTTTGTTTCGAA TCACG
glucanotransferasa	M279N	CGAATCACGACAACCTCGAGGC TTGCAAGC

Anexo III. Resultados de simulaciones de dinámica molecular

Sección 1. TmAmyA y sus variantes

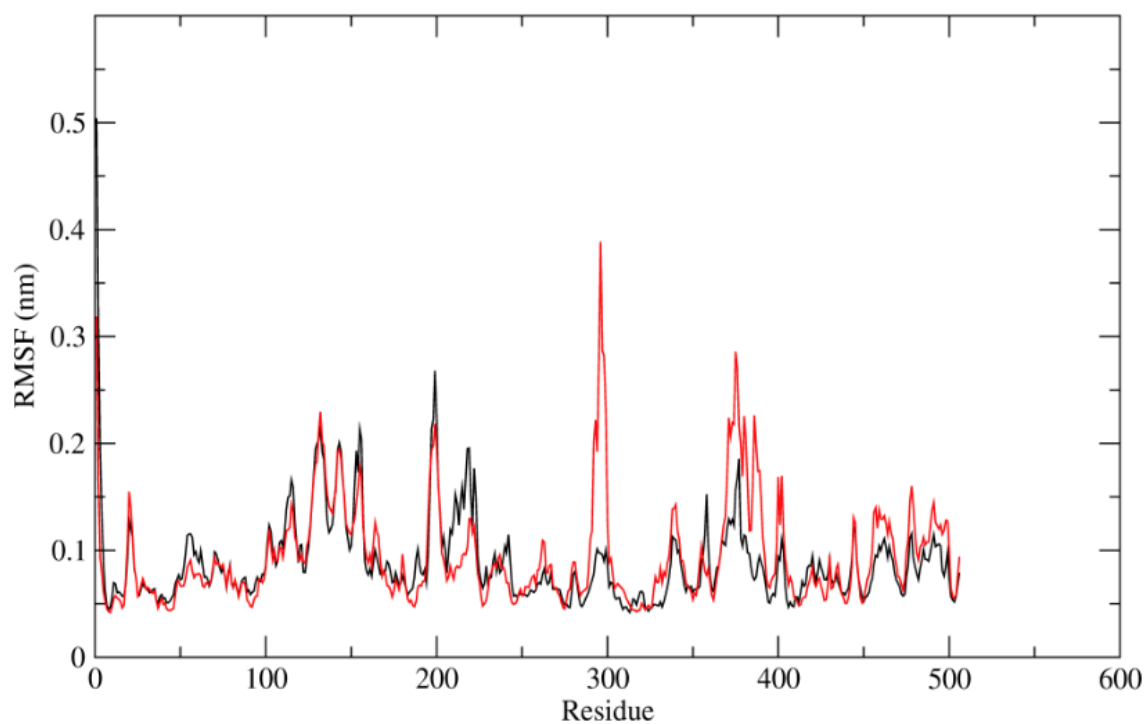
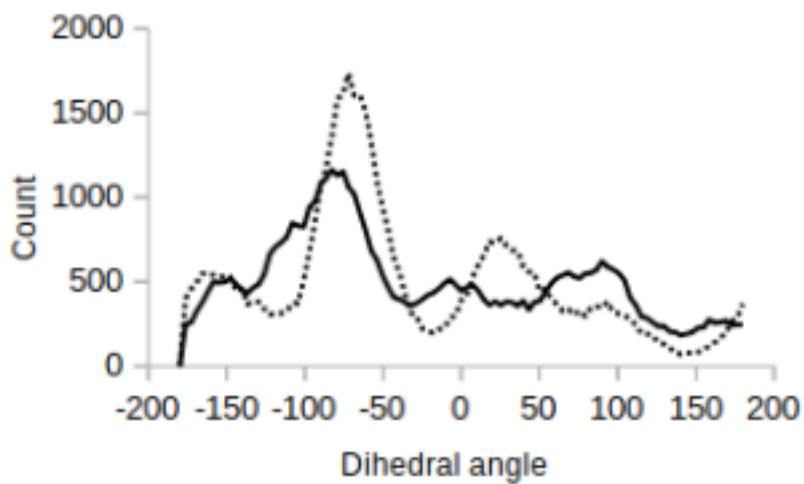
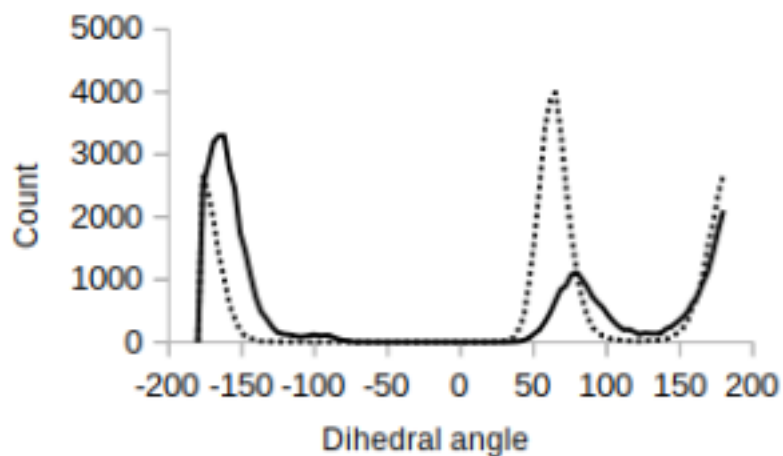


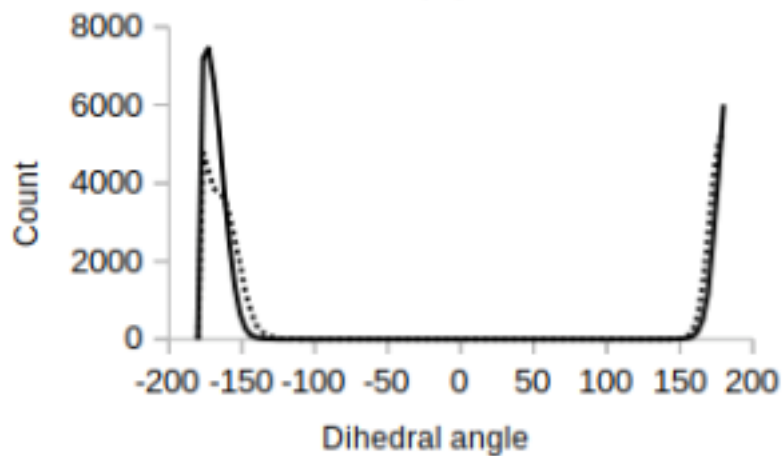
Figura S 1. RMSF para una simulación de Dinámica Molecular (MD) de 500 ns para TmAmyA silvestre (línea roja) y la variante K98P/D99A/H222Q (línea negra). La numeración de los residuos está desfasada -29 con respecto a la de Liebl et al. [36].



(a)



(b)



(c)

Figura S 2 Análisis conformacional de los ángulos diedro χ del residuo ácido base (Glu258) de TmAmyA para la enzima silvestre (línea continua) y la variante K98P/D99A/H222Q (línea punteadas). (a) ángulos diedraless χ_3 (b) ángulos diedrales χ_2 (c) ángulos diedrales χ_1 .

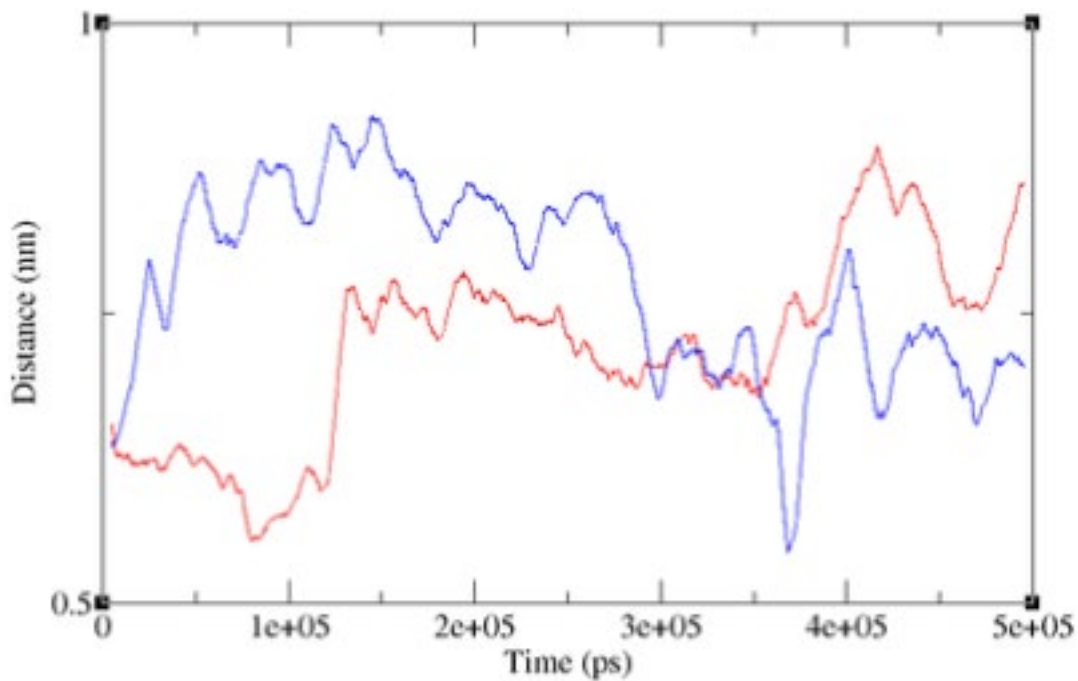


Figura S 3 Cambio en la distancia D310 -E258 de TmAmyA durante la simulación MD para la enzima silvestre (línea azul) y la variante D98P/D99A/H222Q (línea roja).

Sección 2. TmGTasa y

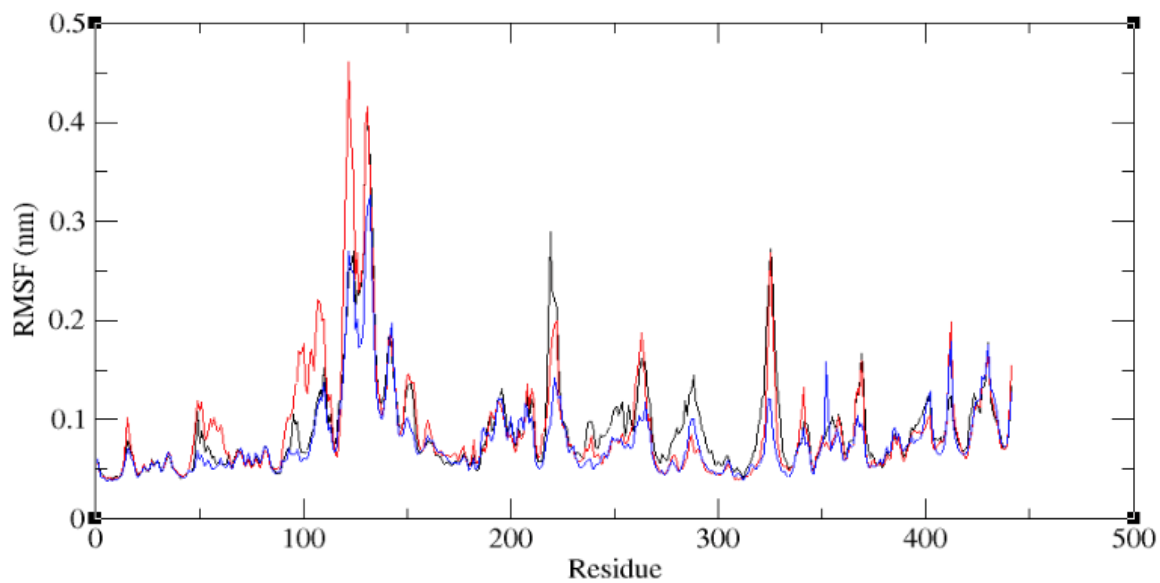


Figura S 4 RMSF para una simulación de Dinámica Molecular (MD) de 500 ns para TmGTasa silvestre (línea roja), la variante M279N (línea negra) y la variante T274V/M279N (línea azul).

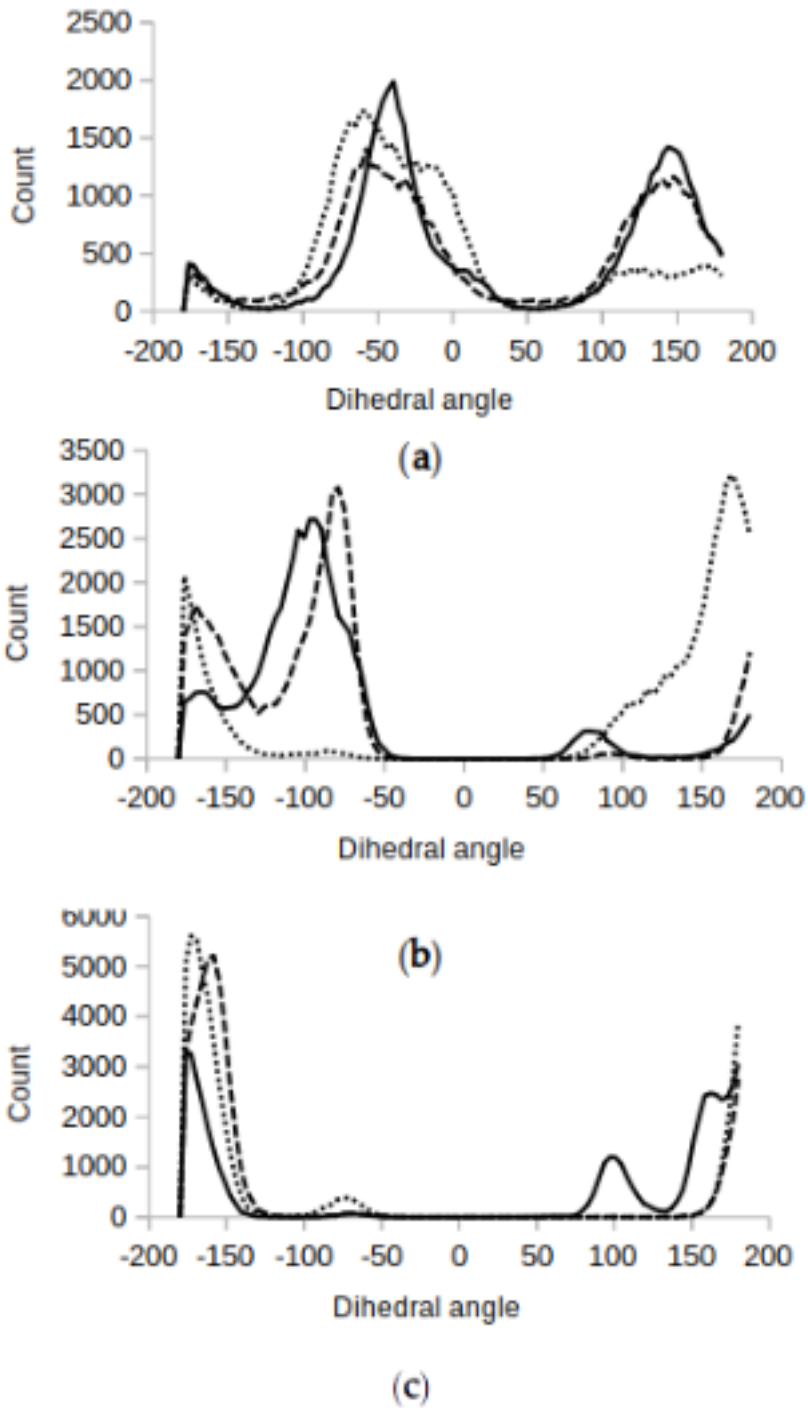


Figura S 5 Análisis conformacional de los ángulos diedro χ del residuo ácido base (Glu216) de TmGTase para la enzima silvestre (línea continua), la variante M279N (línea discontinua), y la variante T274V/M279N (línea punteada). (a) ángulos diedraless χ_3 (b) ángulos diedrales χ_2 (c) ángulos diedrales χ_1

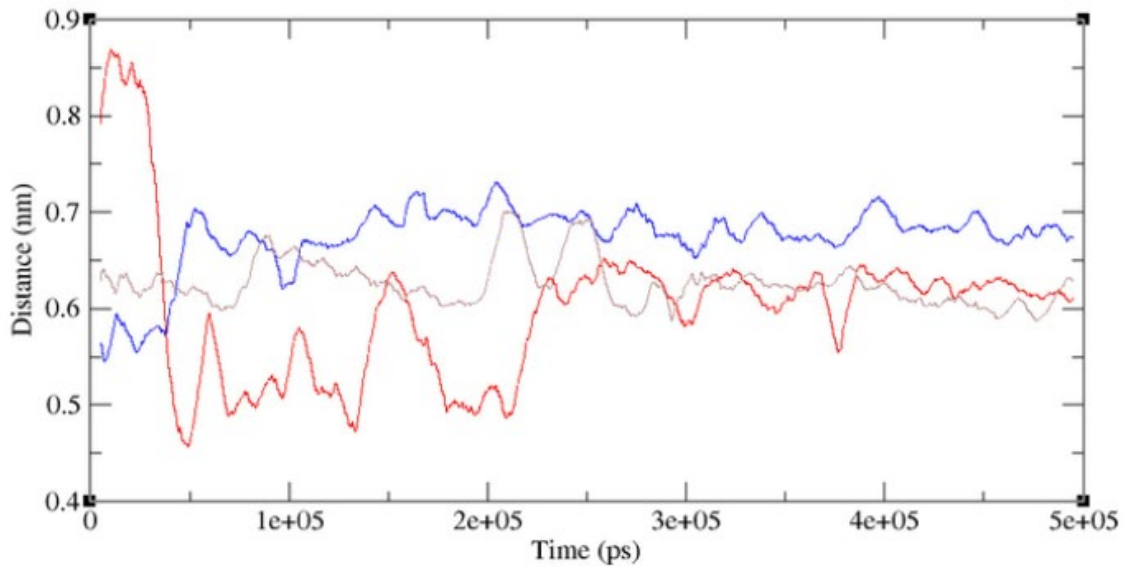


Figura S 6 Cambio en la distancia D278-E216 de TmGTasa durante la simulación MD para la enzima silvestre (línea azul) y la variante M279N (línea roja) y la variante T274V/M279N (línea gris)