UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Instituto de Biotecnología

Estudio del sitio catalítico de una alfa-amilasa por medio de ingeniería de proteínas

Tesis

que para obtener el grado de Doctor en Biotecnología

Presenta:

el M. en C. Gabriel del Río Guerra

Cuernavaca, Morelos Agosto de 1997

RESUMEN

En este trabajo estudiamos la función catalítica, en la alfa-amilasa de Bacillus stearothermophilus (ABST), de un residuo conservado y un dominio variable (dominio B), en la familia de enzimas alfa-amilasas. Para ello generamos un banco de mutantes, para cada región de interés, mediante la técnica de la reacción en cadena de la polimerasa (PCR) y seleccionamos varias mutantes para su ulterior caracterización bioquímica. Para el diseño de la mutagénesis desarrollamos un programa de cómputo. Este nos ayudó a establecer la taza de mutagénesis adecuada para obtener un banco de tamaño mínimo, con una alta proporción de mutantes de nuestro interés. Pudimos observar que la posición Tyr63 participa en la termoestabilidad y en la química de la reacción enzimática de la ABST. Por otra parte, observamos que el dominio B no determina exclusivamente la unión al calcio. Adicionalmente, al comparar las secuencias y estructuras de tridimensionales de varias isoformas de alfa-amilasas y ciclodextrin glicosiltransferasas (CGTasas), observamos que varios residuos de amino ácido en el dominio B participan en las reacciones de transferencia de estas dos enzimas. Al analizar la evolución de la actividad entre estas dos clases de enzimas mediante filogenia molecular, propusimos que el ancestro común a estas dos enzimas debió ser semejante a una alfa-amilasa y que este divergió hacia la actividad de CGTasa mediante la adición de dominios y la alteración del dominio B.

ABSTRACT

In this work we studied the catalytic function, in the alpha-amylase from Bacillus stearothermophilus (ABST), of a conserved amino acid residue and a variable domain (domain B), in the family of enzymes alpha-amylases. For such goal we generated a bank of mutants, for each region of interest, by means of the technique of the polymerase chain reaction (PCR) and we selected several mutants for their biochemical characterization. For mutagenesis design, we developed a computational program. This help us to establish the mutagenesis rate appropriate to obtain a bank of mutants of the minimum size, with a high proportion of mutants of our interest. We could observed that the position Tyr63 participates on the thermostability and on the chemistry of the enzymatic reaction of the ABST. In another hand, we observed that the domain B does not determine exclusively the binding of calcium. Additionally, when comparing the sequences and three dimensional structures of several isoforms of alpha-amylases and cyclodextrin glycosyltransferases (CGTases), we observed the various amino acid residues in domain B participate on the transferring reactions of both enzymes. When analyzing the evolution of the activity between these two classes of enzymes by means of molecular phylogeny, we proposed that the common ancestor to these two enzymes should be more like an alpha-amylase and that it diverged to the CGTase activity by adding domains and altering domain B.

INDICE GENERAL

INDICE DE TABLAS

INDICE DE FIGURAS

I.- ANTECEDENTES

Este proyecto fue concebido a partir de la idea de generar alquil glicósidos, productos de alto potencial comercial (4, 30), utilizando para ello una enzima que pudiera emplear polisacáridos y alcoholes alifáticos como reactantes, ambas sustancias de bajo valor comercial. Un esquema general para una reacción de alcohólisis en carbohidratos se puede plantear como:

$$R1\text{-}O\text{-}R2 + R3\text{-}OH \text{ --------- } R1\text{-}O\text{-}R3 + R2\text{-}OH$$

en donde R1-O-R2 es un éter de carbohidratos, un oligo o polisacárido, y R3-OH es un alcohol primario alifático.

Para que esta reacción se favorezca es necesario utilizar un medio que permita reaccionar a ambas especies, por ejemplo un cosolvente. Adicionalmente, si la reacción se lleva a cabo en presencia de una enzima, esta deberá ser estable a la presencia del cosolvente, y al menos presentar afinidad por uno de los reactantes, es decir, un alcohol alifático y un polisacárido.

Basados en la clasificación de las actividades enzimáticas (oxido-reductasas, transferasas, hidrolasas, liasas, isomerasas y ligasas), consideramos que las enzimas de la clase 2, las transferasas, serían las enzimas adecuadas para producir alquilglicósidos a partir de un polisacárido y un alcohol alifático. Cabe destacar que al igual que las transferasas, las ligasas pudieran servir para este propósito, solo que estas enzimas requieren de cofactores que encarecen su empleo en condiciones industriales. No existe ninguna transferasa o ligasa conocida que reconozca a un alcohol alifático de cadena larga, por lo que las enzimas adecuadas para realizar estudios de alcohólisis de polisacáridos serían las glicosil transferasas (EC 2.4.1).

Sin embargo, estas glicosil transferasas son enzimas que presentaban varias dificultades para su empleo en este proyecto: no se disponía de ningún gene que codificara para una de estas enzimas, estas presentan baja estabilidad, no se conocía la estructura tridimensional ni el mecanismo de catálisis a fondo. Un grupo de hidrolasas que no tienen estos inconvenientes, y que además tienen afinidad por el mismo reactante que las glicosiltransferasas son las α-amilasas (EC 3.2.1.1). De hecho, se ha reportado que la α-amilasa de *Aspergillus oryzae*, taka-

amilasa, puede llevar a cabo reacciones de alcohólisis con alcoholes alifáticos de bajo peso molecular (17). Esta actividad alcoholítica es sin embargo poco eficiente. En este estudio nos enfocamos al estudio del mecanismo de reacción de las α-amilasas con la idea de poder alterar la capacidad alcoholítica de estas enzimas.

Asi pues, el objetivo general de este trabajo es estudiar el mecanismo de catálisis de las α-amilasas, haciendo énfasis en entender de que manera se podría alterar la especificidad, para lograr que la enzima reconociera un alcohol alifático en vez del agua.

Para seleccionar la α-amilasa adecuada para nuestro estudio de alcohólisis consideramos que debía tener las siguientes cualidades:

- Ser de origen bacteriano, para facilitar su manipulación genética en *Escherichia coli*.
- Que fuera termoestable: Esta cualidad en algunos casos correlaciona positivamente con la capacidad de mantener su estructura funcional en solventes no acuosos (1), como por ejemplo altas concentraciones de alcohol alifático o un cosolvente.

- Conocimiento previo del mecanismo de reacción y de la estructura tridimensional de la enzima, para facilitar el estudio molecular de la reacción.

- Disponibilidad del gene.

La mayoría de estas cualidades las reune la α-amilasa de *Bacillus stearothermophilus* ATCC 12980 (ABST), por lo que fue seleccionada para este estudio. Aunque en particular de esta enzima no se conocía su estructura tridimensional, si existía un modelo propuesto de su estructura basado en mutagénesis y modelamiento estructural (11). La taka-amilasa, para la que ya se había reportado su capacidad alcoholítica, no fue seleccionada porque su expresión recombinante en *E. coli* sería un problema, aunque si cumplía con los demás requisitos arriba mencionados.

II.- INTRODUCCION

II.1.- Generalidades Sobre Mecanismos de Reacción.

El estudio cinético y estructural de las enzimas ha permitido determinar que éstas poseen sitios bien definidos para el reconocimiento de su sustrato, y que para transformarlo utilizan grupos químicos diversos, por ejemplo, grupos ácidos y básicos. La mayoría de los conceptos en catálisis están basados en la teoría del estado de transición, que consiste en analizar tres estados químicos del sustrato: su estado original o basal, el estado final o producto de la reacción, y la especie química más inestable durante la reacción, el estado de transición. Al graficar una coordenada de la reacción contra la energía involucrada en la misma, el estado de transición de la reacción se localizaría en el pico más alto en la gráfica, y los intermediarios sobre los valles (23) (véase figura 1). Se acepta que la frecuencia a la cual se descompone un estado de transición es igual a la frecuencia de vibración del enlace que se rompe, la cual a 25° C tiene un valor aproximado de 6 $X10^{12}$ s^{-1}. Es decir, la vida media de un estado de transición es del orden de 10^{-12} s (23).

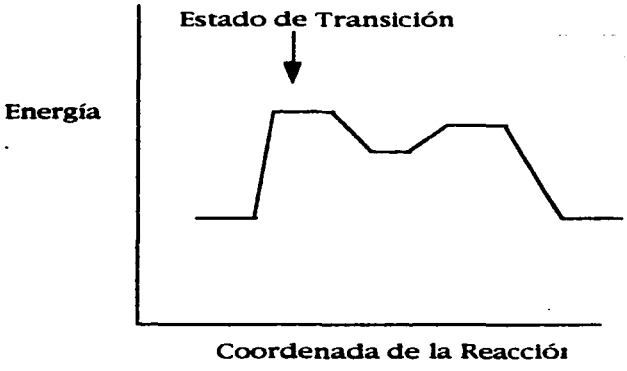


Figura 1. Gráfica de una Reacción Química

Comúnmente los estados iniciales y finales de una reacción enzimática son detectados por técnicas espectrofotométricas, sin embargo, en el caso del estado de transición, los métodos utilizados solamente han logrado intuirlo o sugerirlo, más que detectarlo directamente, debido a su vida media corta. Algunas de estas estrategias son:

- Diseño de inhibidores análogos a un estado de transición supuesto (9)

- Difracción de rayos X tipo Laue (34)

- Cinética de la forma nativa de una enzima y mutantes de la misma, con o sin marcaje radiactivo (9)

- Química Computacional (21)

Gracias a la combinación de estas estrategias es que se ha entendido el papel que juegan los distintos grupos químicos en las enzimas, en la generación y estabilización de un estado de transición particular. De las o-glicosidasas (EC 3.2.1) que se han estudiado a este detalle es la Lisozima (EC 3.2.1.17) (16).

II.2.- Mecanismo de Reacción de o-Glicosidasas

La reacción catalizada por este grupo de enzimas es la hidrólisis de un enlace glicosídico (véase figura 2).

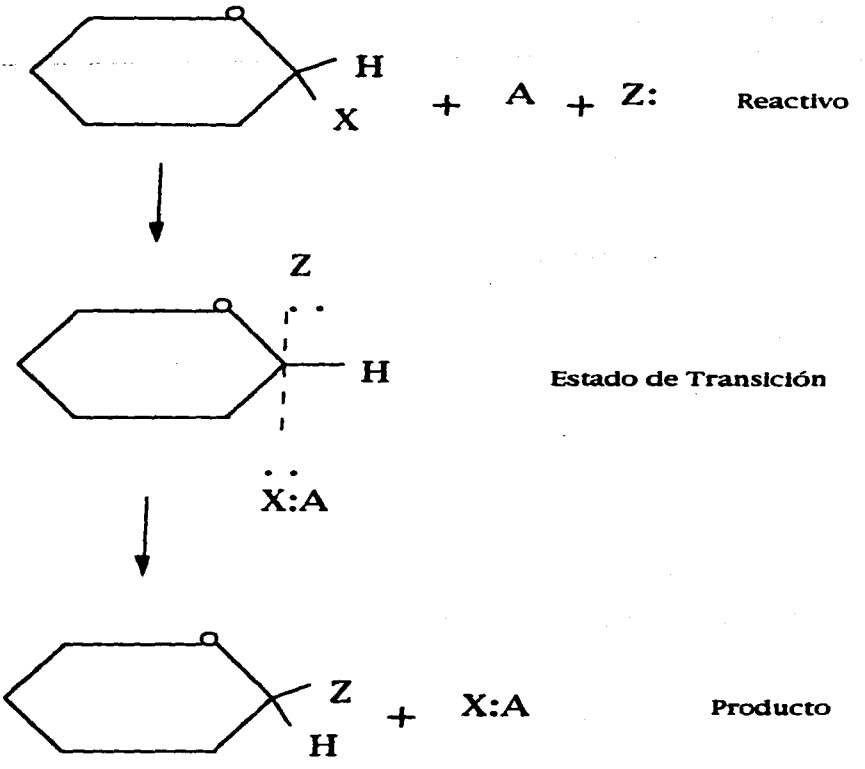


Figura 2. Mecanismo de Reacción para o-Glicosidasas

en donde A es un ácido general o un electrófilo, Z: es una molécula nucleófila, o donadora de electrones (que puede ser agua), y X es cualquier grupo químico eterificado al anillo de piranosa.

Del estudio de la química no enzimática de la hidrólisis del enlace glicosídico, se ha propuesto que durante la reacción se favorece la formación de un catión glicosídico. En solventes menos polares que el agua, este catión es demasiado inestable para existir (15). Estimaciones de la vida media en agua varían de 10^{-10} a 10^{-12} s, lo cual concuerda con el valor estimado a 25° C (véase arriba). Esto sugiere que el sitio activo de enzimas que procesen este enlace deberán poseer una polaridad alta que permita estabilizar a este estado de transición, o al intermediario de la reacción.

En el caso de reacciones enzimáticas de hidrólisis del enlace glicosídico, esta puede llevarse a cabo mediante dos mecanismos (6, 15, 33):

- Con retención de la configuración del carbono anomérico hidrolizado, para lo cual se ha propuesto un doble desplazamiento, que involucra a un grupo nucleófilo en la enzima (véase figura 3). Este mecanismo propuesto de sustitución es similar al denominado de Sustitución

Nucleofílica tipo 1 (SN1), en donde la velocidad de formación del estado de transición solo depende de una de las dos especies en la reacción, en este caso del polímero de glucosa.

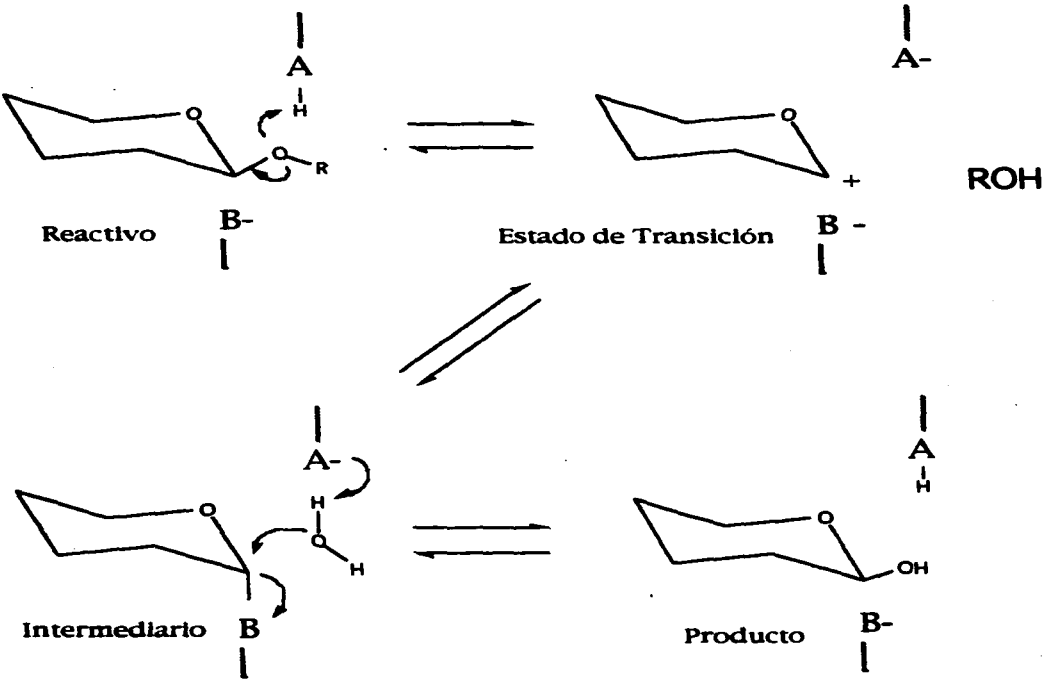


Figura 3. Mecanismo de Reacción para Enzimas Retenedoras

- Con inversión de la configuración del carbono anomérico hidrolizado, que se propone debería proceder mediante un simple paso de sustitución por una molécula nucleófila de agua (véase figura 4). A este tipo de reacciones se les conoce con el nombre de Sustitución Nucleofílica tipo 2 (SN2), porque la velocidad de formación del estado de transición depende de las dos especies químicas reaccionantes, el agua y el polímero de glucosa.

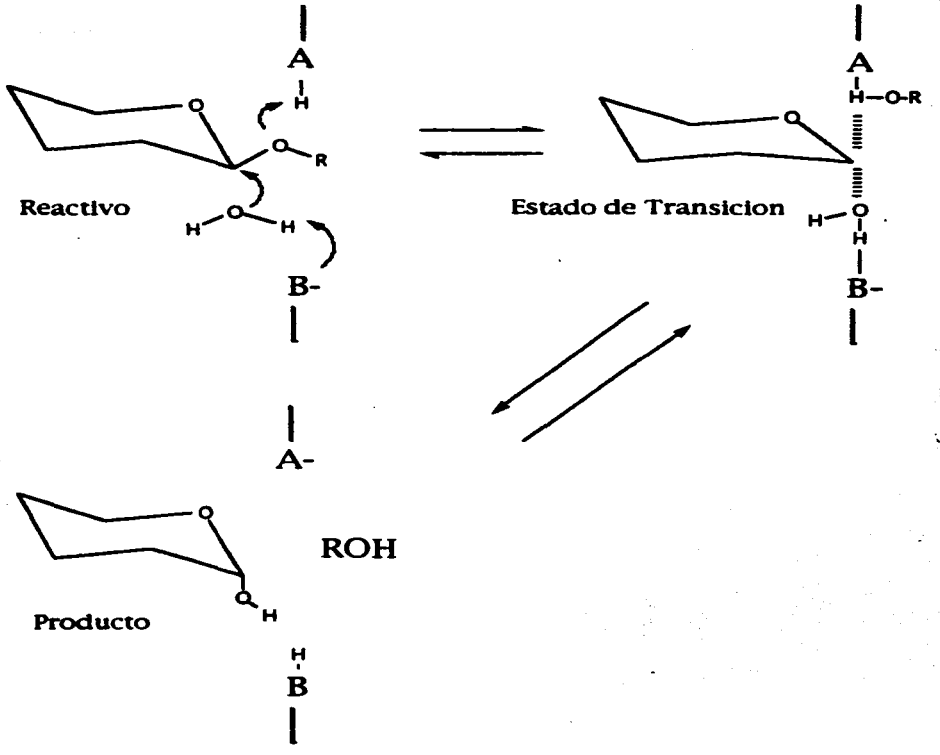


Figura 4. Mecanismo de Reacción para Enzimas Invertidoras

Es importante destacar que las hidrolasas que retienen la configuración del carbono anomérico, podrían transferir residuos glicosídicos a

alcoholes de bajo peso molecular tales como el metanol o al agua, y tal es el caso de la taka-amilasa (17). Esta actividad no podría esperarse en una enzima que invierte la configuración, ya que esto implicaría que estas enzimas pudieran funcionar como transferasas inespecíficas. Supóngase la hidrólisis de un metil β-glicósido por una enzima invertidora, la cual generaría α-glucósido. Si la enzima tuviera actividad de transferasa hacia metanol, podría generar a partir de un α-glucósido y metanol, metil β-glicósido. Por el principio de reversibilidad, tendría que aceptar como substratos tanto a α- y β-glicósi-dos, lo cual implicaría que estas enzimas no discriminan la anomericidad del substrato (33). Visto de otra manera, y de acuerdo a la teoría del estado de transición, podría considerarse que las enzimas que catalizan reacciones del tipo SN2, deberían tener afinidad por los dos reactantes de los que depende la velocidad de formación del estado de transición, mientras que las que catalizan por medio del mecanismo SN1, solo deberían ser afines por la especie química de la que depende la formación del estado de transición. Esto implica que para poder usar una enzima invertidora en nuestro estudio, deberíamos cambiar la afinidad del agua por la de un alcohol alifático, mientras que usando una enzima retenedora tendríamos que generar *de novo* afinidad por un alcohol. Considerando que en el primer caso la nueva afinidad

implicaría hacer huecos en el sitio activo debido a la diferencia de tamaños entre el agua y el alcohol, y esto provocaría una desestabilización de la estructura, decidimos que las α-amilasas, enzimas retenedoras, son adecuadas para nuestro estudio. Para entender el mecanismo molecular de operación de estas enzimas, es útil compararlas con la lisozima, ya que también cataliza la reacción de un enlace glicosídico, y que ha sido más exaustivamente estudiada en ese sentido.

La lisozima cataliza la hidrólisis del componente peptidoglucano de la pared celular de bacterias Gram (+), cortando el enlace entre N-acetil murámico y N-acetilglucosamina. También es capaz de hidrolizar a la quitina, un polímero de residuos N-acetilglucosamina en β(1,4). Se ha observado que para la lisozima de clara de huevo de gallina, el mecanismo de hidrólisis es con retención de la configuración. El estudio detallado del mecanismo de reacción de esta lisozima ha mostrado que su sitio activo posee una polaridad que estabiliza el estado de transición aún en ausencia del nucleófilo en la enzima (29). Esto concuerda con lo esperado por comparación con reacciones hidrolíticas de glicósidos en ausencia de catalizadores enzimáticos (véase arriba).

En lo que respecta a la forma en como se transfiere este catión a un nucleófilo como el agua o algún alcohol, solo se pueden inferir algunas características de este proceso. Por ejemplo, se ha observado que un tipo de o-glicosil transferasas que retiene la configuración, las ciclodextrin glicosil transferasas (CGTasas) (EC 2.4.1.19), transfieren mediante un mecanismo denominado ping pong (22). Es decir, en la transferencia de un glucósido, el catión glucosídico forma un enlace covalente con la enzima. Por el contrario, en la hidrólisis de o-glicósido s no se ha podido detectar la presencia de un intermediario covalente durante la reacción (6, 33). Por otra parte, se ha observado que en presencia de metanol, las CGTasas no generan metil-glicósidos, en cambio las α-amilasas sí (19, 17). Esto implica que las α-amilasas no son específicas por el alcohol a quien transfieren al catión glucosídico, mientras que las CGTasas sí, y que la transferencia implica un intermediario covalentemente unido a la enzima.

Cabe mencionar que algunas α-amilasas pueden transferir un glicósido a otro en vez de al agua, en presencia de altas concentraciones del glicósido. A esta reacción se le conoce como de transglicosidación (33). Para determinar si una enzima transglicosídica es mejor alcoholasa que hidrolasa, realizamos un estudio en colaboración con la QFB. Rosa Isela

Santa María, estudiante de Maestría con el Dr. Agustín López-Munguía en el Instituto de Biotecnología/UNAM. En este estudio comparamos la actividad alcoholítica de 6 α-amilasas con su actividad transglicosídica (trabajo en preparación). Aquí observamos que las α-amilasas capaces de transglicosidar fueron las mejores alcoholasas. Así pues, nuestro estudio se enfocó en el análisis de reacciones de transglicosidación como sistema modelo de las reacciones alcoholíticas, ya que el seguimiento de esta última actividad es más complejo.

Al empezar nuestro estudio fue reportada una mutación en la α-amilasa de *Saccharomycopsis fibuligera*, que aumentaba la actividad transglicosídica de esta enzima (18), es decir, que generaba más productos de transglicosidación. Con la intención de utilizar esta propiedad en reacciones de alcohólisis, decidimos mutar la posición homóloga en la ABST. Durante el desarrollo de este proyecto se reportó la estructura tridimensional de varias CGTasas (38). Esto nos permitió iniciar un estudio comparativo entre α-amilasas (hidrolasas) y CGTasas (transferasas) (enzimas mecanísticamente similares y estructuralmente homólogas), con la idea de entender como una α-amilasa podría cambiar su preferencia por un nucleófilo distinto del agua. Pensamos que esta comparación sería útil para estos fines, ya que las CGTasas son

capaces de transferir un alcohol (glicósido) en lugar del agua, en solución acuosa.

III.- HIPOTESIS

Las actividades de α-amilasa y ciclodextrin glicosiltransferasas pueden ser interconvertidas con un número limitado de alteraciones. Basados en el estudio del mecanismo de reacción de α-amilasas, y la comparación de éstas con transferasas homólogas, ciclodextrin glicosiltransferasas, se puede diseñar una estrategia de mutagénesis para lograr este objetivo.

IV.- OBJETIVOS

1) Estudiar el mecanismo de reacción de las glicosidasas

2) Generar mutantes en la α-amilasa de *Bacillus stearothermophilus* que permita estudiar la capacidad de transferasa de esta enzima.

3) Comparar a nivel de estructura primaria y terciaria las α-amilasas y las ciclodextrin glicosiltransferasas, para detectar los determinantes estructurales que determinan la capacidad de transferasa.

V.- METODOS

V.1.- Comparación entre α-Amilasas y CGTasas.

Para el presente análisis, se utilizaron las secuencias y estructuras de las siguientes α-amilasas (entre paréntesis se indica el nombre del archivo del Protein Data Bank): de trigo (1amg), de hongo (1amy, 2taa), pancreática de puerco (1ppi), salivar humana (1smd) y bacilar (1bip). Las CGTasas utilizadas en este análisis de secuencia fueron todas bacilares (1cdg, 1ciu, 1cyg).

El análisis de los determinantes estructurales de la posible función de exclusión del agua del sitio activo, se realizó sobre las estructuras de la α-amilasa pancreática de puerco (1ppi) y la CGTasa de *B. circulans* 251 (1cxe), ambas en complejo con acarbosa.

V.2.- Deleción del Dominio B en la α-Amilasa de *B. stearothermophilus.*

V.2.1.- Predicción de Estructura Secundaria.

Para la predicción de la estructura secundaria en la ABST utilizamos los algoritmos desarrollados por Chou-Fasman (25) y Garnier (10). Además, también consideramos predicciones previas hechas para esta enzima mediante otro método denominado Hydrophobic Cluster Analysis (27).

V.2.2.- Diseño de Oligonucleótidos.

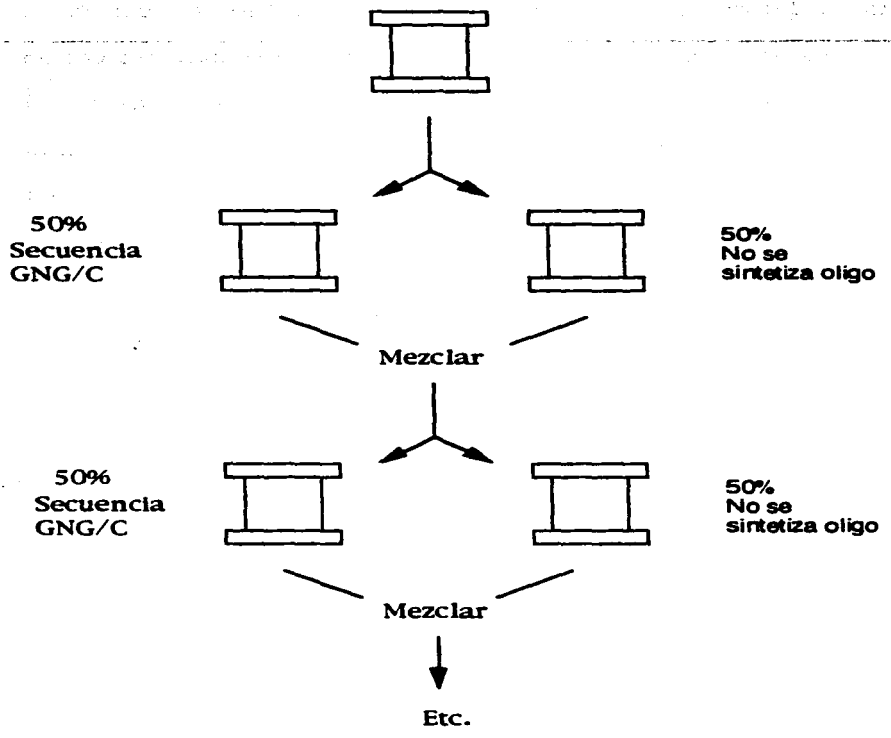
Para determinar el tamaño y la longitud de los conectores que sustituirían al dominio B, se analizaron 10 enzimas con un plegamiento tipo TIM-barrel (véase Resultados, Tabla 1). Dado que el grupo de enzimas analizadas es pequeño, un análisis de composición de amino ácidos no tendría suficiente validez. Así que decidimos utilizar amino ácidos de cadena lateral pequeña (para darle suficiente flexibilidad al conector) y polares (ya que este conector estaría expuesto al solvente). Además, ya que la sustitución iba a ser llevada a cabo a nivel del gene de la ABST, buscamos amino ácidos que compartieran entre si al menos 2 bases en sus codones, con la idea de generar un banco lo mas pequeño posible, a nivel DNA. Así fue como seleccionamos los amino ácidos codificados por los codones GNG/C, en donde G: Guanina, C: Citosina y N: Adenina, Timina, G y C (véase Tabla 2).

Codon	Amino Acido
GAG	Glu
GAC	Asp
GTG	Val
GTC	Val
GGG	Gly
GGC	Gly
GCG	Ala
GCC	Ala

Tabla 2.- Codones usados en la deleción del dominio B de la ABST. Los codones y el amino ácido correspondiente (en código de tres letras) son indicados en la tabla

El diseño de los oligonucleótidos incluyó, además de la variabilidad en secuencia GNG/C, la variabilidad en longitud, de 1 a 8 codones. Para

esto último, durante la síntesis de los oligonucleótidos se sustraía la mitad de la resina de síntesis, cada vez que se sintetizara un codon GNG/C (5, 7) (véase Figura 5). En esta forma buscamos generar un banco de mutantes, a nivel de amino ácidos de 10^5, con una distribución tal que estuviesen altamente representadas las mutaciones múltiples. Las bases para optimizar el diseño de las condiciones de mutagénesis están detalladas en la publicación 1.



El vial en donde está acoplado el oligonucleótido esta representado como un cilindro.

Figura 5. Esquema para la Síntesis de Oligos para la Deleción del Dominio B de la ABST

La secuencia de los oligos empleados para tales fines fue:

oligo 1.- 5´- CAC GAC ATC GGC GTA CAC TTG-3´

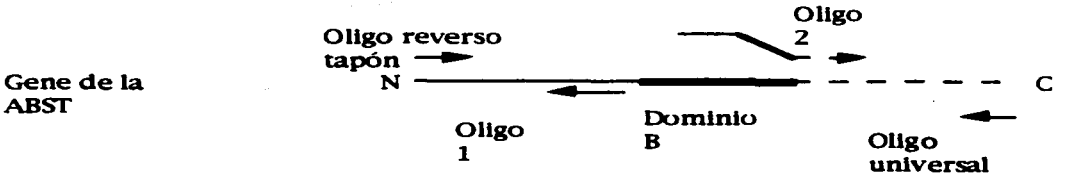
oligo 2.- 5´- GCC GAT GTC GTG GNG/C GNG/C GNG/C GNG/C GNG/C GNG/C GNG/C GNG/C CCC GAA GTC GTG-3´

La región subrayada es complementaria en ambos oligonucleótidos ya que es a partir de esta que se ensamblarían los genes mutantes (véase figura 6).

Para incluir 6 Histidinas en el C-terminal de la enzima mediante la técnica de PCR, utilizamos al primer universal (20) y diseñamos el siguiente oligonucleótido:

oligo 3.- 5´- GGT GGC ATG GGC TCA TCA CCA TCA CCA TCA CTG ATC ACT GCG ATC GCG-3´

Diseño:



Nota: la región con linea gruesa en el Oligo 2 representa la región variable en longitud y secuencia que sustituiría al dominio B, también indicado con linea gruesa.

Protocolo:

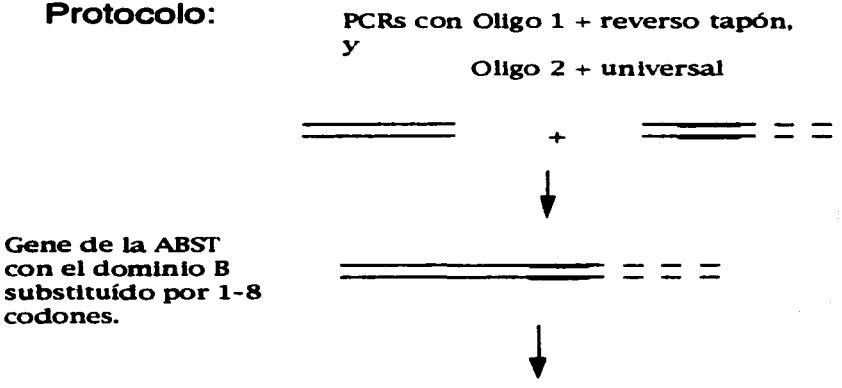


Figura 6. Estrategia para la Deleción del Dominio B de la ABST

V.2.3.- Ensayo de Actividad Amilolítica en Caja.

Las células a ensayar se crecieron en caja petri con medio Luria, agar 1.5% y almidón al 1%, . Después de 12 horas de crecimiento a 37°C se observó que solo las células portadoras del gene nativo de ABST presentaban un halo claro alrededor de ellas, producto de la hidrólisis del almidón. A las 24 Hrs. de crecimiento se presentó un halo en células que portaban solo el plásmido sin el gene de la ABST. De aqui que las células portadoras de un gene variante de ABST con actividad distinguible del ruido de fondo, deberían producir un halo entre las 12 y las 24 horas de crecimiento a 37° C.

V.2.4.- Ensayos de Renaturalización.

Esto se realizó de dos formas distintas y con varias condiciones como se describe a continuación.

a) Renaturalización de las proteínas en la columna.

Una vez adherida las proteínas a la columna de Niquel y lavada de acuerdo al protocolo del fabricante (QIAGEN, USA), se procedió a renaturalizar *in situ* mediante un gradiente, de 0 a 8 M de urea. Esto se intentó con 5, 7 y 10 mL de cada solución en las cámaras, y a una velocidad de caída de la gota a la columna de 1, 1.5 y 2 mL por minuto. Solo una vez se logró recuperar actividad con el protocolo de 7 mL y a 2 mL por minuto. También en ese ensayo la cantidad de enzima que se recuperó de las células fue el doble (60 μg; 20 μg/mL de concentración) respecto a los demás intentos.

b) Renaturalización de la proteínas una vez eluídas de la columna.

Al ser eluídas las proteínas de la columna de Niquel por competencia con 0.2 M de imidazol, de acuerdo a las condiciones descritas por el fabricante (QIAGEN, USA), esta solución se goteó en un vaso de precipitados de 250 mL, en hielo y con agitación, que contenía alguna de las siguientes soluciones: glucosa 200 mM, glicerol 200 mM. En ambos casos las soluciones contenían buffer de acetatos o fosfatos, pH 6.0, a 50 o 100 mM. Solo con la solución con glicerol en buffer de acetatos 100 mM pudimos renaturalizar la enzima para obtener actividad en 5 de 15 intentos. Pudimos observar una correlación entre

la renaturalización exitosa y el extraer de las células una cantidad de proteína mayor de 30 μg (concentración de 10 μg/mL).

Se intentó inducir la expresión del gene de la ABST mediante el promotor lacUV5 en *E. coli* y el promotor de subtilisina en *B. subtilis*, pero ninguno mejoró los rendimientos ya reportados (10 μg/mL de concentración; datos no mostrados). Adicionalmente, la Dra. Gloria Saab Rincón probó la expresión en *E. coli* bajo el promotor T7 con los mismos resultados negativos.

V.2.5.- Medición de Actividad Amilolítica *in vitro*.

La actividad amilolítica fue medida con el reactivo ácido dinitro salicílico o DNS (37). Este compuesto al reaccionar con el extremo reductor de un azúcar, forma un compuesto colorido que se puede cuantificar específicamente a 540 nm de longitud de onda en un espectrofotómetro. Las reacciones fueron llevadas a cabo a 70° C y pH 4.5 (solución buffer de acetatos 50 mM, 1 mM $CaCl_2$). Para quelar el calcio y medir el efecto en la actividad amilolítica se adicionó EDTA a una concentración final de 1 mM. Todos los ensayos de actividad fueron hechos al menos por duplicado.

VI.- RESULTADOS Y DISCUSION

VI.1.- Función de Residuos Conservados en α-Amilasas.

El residuo Tyr63 de la ABST, homólogo al mutado en *S. flivuligera* que aumentaba su actividad transglicosídica, fue mutagenizado a saturación y 4 mutantes fueron seleccionadas y caracterizadas (véase Trabajo sometido a publicación 1). Ninguna de las mutantes presentó una mejora en su actividad transglicosídica. Sin embargo, sí se observaron variaciones en el perfil de pH y k_{cat}, lo que aunado a la conservación de este residuo de amino ácido en α-amilasas y ciclodextrin glicosiltransferasas (CGTasas), y el comportamiento similar observado de otras mutantes en posiciones también conservadas y tridimensionalmente cercanas a la Tyr63 (Arg204 e His296 en la numeración de la taka-amilasa), nos condujeron a proponer que este grupo de residuos participan probablemente en la estabilización de el estado de transición en la reacción catalizada por α-amilasas.

Como no se logró aumentar la actividad transglicosídica en la α-amilasa seleccionada para nuestro estudio, decidimos abordar el problema de la

transferencia de un nucleófilo diferente al agua, a través de la comparación entre una hidrolasa y una transferasa.

VI.2.- Análisis Comparativo del Mecanismo de Acción entre α-Amilasas y CGTasas.

El mecanismo de sustitución del enlace glicosídico puede llevarse a cabo de dos formas, SN1 y SN2. Se destacó el hecho de que las enzimas que usan mecanismos tipo SN2 no serían de utilidad para el proyecto de alcohólisis. Esto nos condujo a analizar un par de grupos de enzimas que usan el mismo mecanismo de catálisis, tipo SN1, una de ellas funciona como hidrolasa (α-amilasas) y la otra como transferasa (CGTasas). Estas últimas pueden prevenir el ataque del agua al catión glicosídico, y en su lugar permiten acceder preferentemente un nucleófilo diferente, en este caso un glucósido. Con base en estas observaciones, buscamos en la comparación estructural de α-amilasas y CGTasas una base conceptual que nos permitiera diseñar una α-amilasa con actividad de transferasa mejorada.

VI.2.1.- Análisis de Secuencia.

Analizando el alineamiento descrito en el trabajo sometido a publicación 1, observamos que los residuos invariantes entre las α-amil asas también lo son entre las CGTasas, lo cual apoya la idea de que ambas enzimas utilizan la misma maquinaria catalítica. Por otra parte, las posiciones Ile52, Val60, Leu194, Gly199, Phe203 y Val293 (numeración referida a la taka-amilasa) que son semiconservadas entre las α-amilasas, aumentan su variación en sequencia cuando se incluyen las CGTasas. Estos amino ácidos son vecinos estructurales de los residuos invariantes Tyr82 (Ile52 y Val60), Arg204 (Leu194, Gly199 y Phe203) e His296 (Val293) (en paréntesis se indican los residuos semiconservados). Además, observamos que algunas posiciones invariantes entre CGTasas, dejan de serlo al incluir a las α-amilasas en el alineamiento. Estas posiciones son Gln13, His108, Asp202, Phe292 y Gly323 (numeración referida a la taka-amilasa). Estas diferencias pueden deberse a que las CGTasas analizadas todas son de origen bacilar, mientras que las α-amilasas alineadas presentan una mayor diversidad filogenética. También se observó que la secuencia Phe-Ala-Pr o en la tercer hebra beta del barril, que había sido propuesta como firma distintiva de las CGTasas (12), no es invariante entre las CGTasas analizadas. Solo esta firma Phe-Ala-Pro y Phe292 (de los residuos

invariantes entre CGTasa) estan cercanos al sitio activo de α-amilasas o CGTasas, pero lo suficientemente lejanos de este como para poder inferir alguna función *a priori*.

En resumen, no encontramos una correlación directa a nivel de secuencia primaria con la capacidad transglicosídica de las CGTasas, que nos permitiera establecer una base para el diseño de mutantes de-α-amilasas con actividad transglicosídica mejorada. Sin embargo, si pudimos observar algunas firmas distintivas entre CGTasas que probablemente esten involucradas en otras funciones especificas para CGTasas. Decidimos entonces analizar la estructura tridimensional de ambas enzimas.

VI.2.2.- Análisis de Estructura.

Las α-amilasas y las CGTasas comparten similitudes estructurales en tres dominios (véase figura 7): Dominio A, cuyo plegamiento es de un barril del tipo $(\beta/\alpha)_8$, también conocido como TIM-barrel porque fue precisamente en la Triosa Fosfato Isomerasa (TIM) donde se observó este plegamiento por primera vez (8). En este dominio se encuentran los residuos invariantes entre ambas enzimas, y a este se le ha asignado

un papel preponderante en la catálisis (38). Dominio B, que se localiza entre la tercera hebra beta y la tercera α-hélice del dominio A (aquí observamos que la longitud de este dominio es mayor en α-amilasas bacilares que en las de origen eucarionte analizadas). Se ha propuesto que la dependencia de calcio para la actividad de α-amilasas reside en la estructuración que el calcio imparte a este dominio (3). Dominio C, cuya topología es de un sandwich de beta plegadas y al cual se le ha asignado el papel de adhesión al almidón (38). Adicionalmente, las CGTasas poseen un par de dominios (comúnmente denominados D y E) que se extienden a partir del dominio C, ambos con una topología de beta plegadas (13).

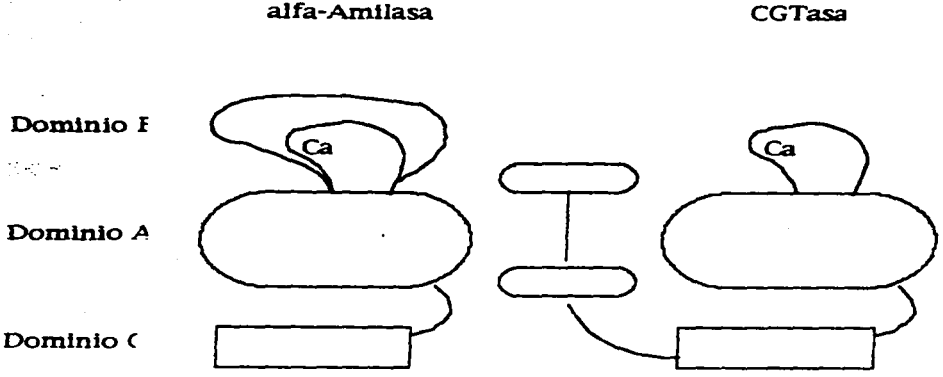


Figura 7. Diagrama de Dominios de alfa-Amilasas y CGTasas

El dominio B se muestra en dos tamaños, de acuerdo a lo observad en nuestro análisis. El dominio B contiene el sitio de unión al calcio (Ca).

Para buscar las diferencias estructurales entre α-amilasas y CGTasas que pudieran ser responsables de la prevención del ataque del agua nucleófilica por parte de las CGTasas, utilizamos las estructuras cristalográficas de estas enzimas en complejo con un inhibidor, acarbosa, que es análogo al estado transición (catión glicosídico) de la reacción catalizada por estas enzimas (véase Métodos). Nuestro criterio de búsqueda fue el siguiente:

- Observar la geometría del enlace glicosídico atacado, es decir, si el enlace está horientado hacia la enzima o bien hacia el lado opuesto (véase figura 8). Es razonable pensar que la dirección en la que se encuentra el enlace a procesar debería ser la misma en la que atacaría el nucleófilo al intermediario de la reacción (véase Figura 3).

- Determinar la posición desde donde el agua nucleofílica accede al catión glicosídico, es decir, desde el seno de la enzima o desde el solvente.

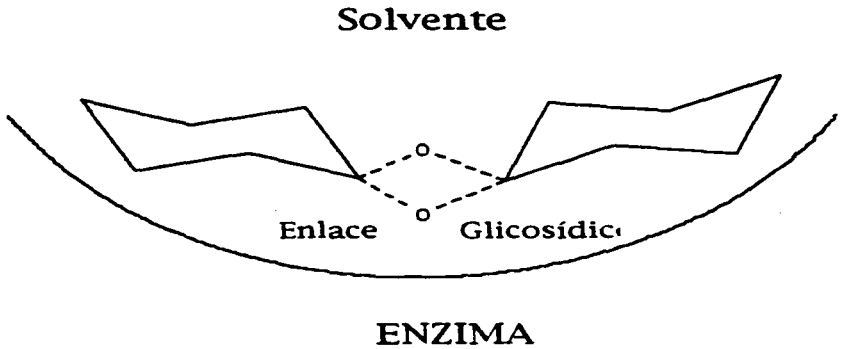


Figura 8. Esquema de la Forma de Unión al Sustrato por alfa-Amilasas

Para ambas enzimas en complejo, se puede observar que el polisacárido sufre una distorsión justo en la zona donde se localizan los residuos esenciales para catálisis (véase figura 9). Esta distorsión posiciona al enlace C-O glicosídico a procesar apuntando ligeramente hacia afuera del seno de la enzima. Esta observación es consistente con la posición del ácido general durante la catálisis en ambas enzimas (Glu230 en taka-amilasa y Glu257 en CGTasas) (26, 36), es decir, el ácido general debe atacar al átomo O glicosídico (véase figura 8), y si el ácido está situado hacia afuera del seno de la enzima, el átomo O glicosídico también deberá estarlo. Si aceptamos que estas enzimas procesan este enlace mediante un mecanismo de doble desplazamiento (véase figura 3), y que la orientación observada en el enlace a procesar será la que adoptará el nucleófilo atacante, la estructura observada sugiere que el agua nucleofílica debe atacar al catión glicosídico desde el solvente, es decir, sin contacto íntimo con la enzima. Esta observación es consistente con lo anteriormente mencionado en la Introducción referente a enzimas que catalizan reacciones mediante mecanismos tipo SN1, es decir, la enzima solo debe tener afinidad por el carbohidrato, de quien depende la velocidad de reacción, y no requiere afinidad por el agua.

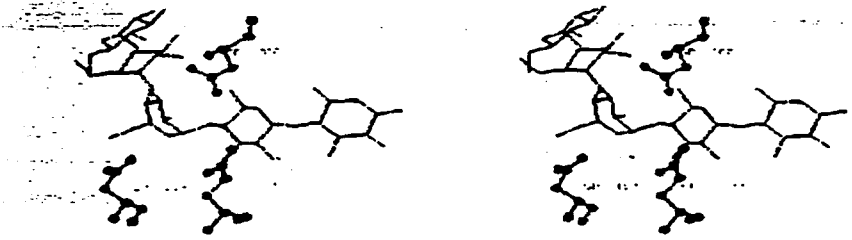


Figura 9. Par Estereoscópico de Acarbosa en el Sitio Activo de la Taka-Amilasa

Código de colores: Carbono, verde; Oxígeno, rojo; Nitrógeno, azı

Al buscar los elementos estructurales que permiten a una CGTasa obstruir la entrada del agua hacia el catión glicosídico, observamos que esto se logra a través de dos formas:

- Un residuo voluminoso (tirosina o fenilalanina) localizado justo encima del enlace glicosídico a procesar

- La forma de unir al sustrato.

Para el caso del residuo voluminoso (Tyr195), localizado en el dominio B, solo cambia a Fenilalanina entre las CGTasas analizadas, mientras que entre las α-amilasas esta posición es ocupada por amino ácidos más

pequeños como glicinas y serinas. Cuando esta posición fue mutada en la CGTasa de *B. circulans* 251 (24) se observó en las variantes analizadas (Tyr195Phe, Tyr195Trp, Tyr195Leu y Tyr195Gly) que se alteraba la reacción de transferencia. Las CGTasas pueden catalizar cuatro reacciones de transferencia: Formación de ciclodextrinas, transferencia de un extremo reductor a otro no reductor en la misma cadena de un polisacárido, lo cual genera oligosacáridos cíclicos llamados ciclodextrinas; acoplamiento, cuando se combinan una ciclodextrina con un oligosacárido lineal para generar un oligosacárido lineal de mayor longitud; desproporción, es la transferencia de una parte de un oligosacárido lineal a otro lineal; sacarificación, es la transferencia de un nuevo extremo reductor al agua. Todas las mutantes reportadas en esta posición afectaron la capacidad de ciclización y acoplamiento. Solo las mutantes Tyr195Leu y Tyr195Gly mostraron actividad sacarificante ligeramente aumentada. Si esta posición tuviera una función de obstrucción del agua, CGTasas con cadenas laterales pequeñas en esta posición serían mas aptas para sacarificar. Conforme a ello, las mutantes Tyr195Leu y Tyr195Gly mostraron aumento en su capacidad sacarificante y disminución en su capacidad transglicosídica. Sin embargo, se puede concluir que esta posición no determina totalmente la exclusión del agua.

Un factor estructural que consideramos importante para la función de transferencia es la forma de unión del sustrato. ¿Cómo puede la forma de unión al sustrato por CGTasas inducir a la transferencia de un glicósido en lugar del agua? La estructura de una ciclodextrina (véase figura 10) produce una distribución de los hidroxilos de tal forma que el centro de la molécula es hidrofóbico y su exterior hidrofílico, propiedad que ha sido utilizado con diversos fines en la industria y la investigación (39, 31). Así pues, en la zona donde el enlace glicosídico es procesado, la conformación del sustrato favorecería un ambiente hidrofóbico que excluiría al agua.

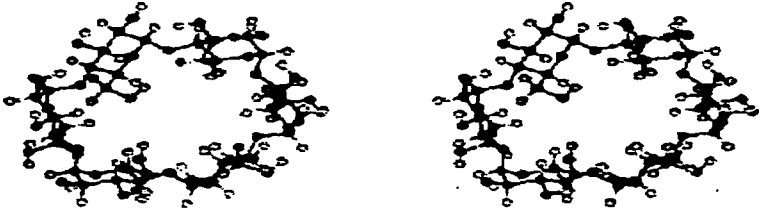


Figura 10. Par Estereoscópico de una alfa-Ciclodextrina

Código de colores: Carbono, verde; Hidrógeno, gris; Oxígeno, rojo.

Ahora bien, ¿Cómo induce la enzima esta conformación en el sustrato? Recientemente se ha resuelto la estructura de la CGTasa de *B. circulans*

251 con un nanomaltosacárido (oligosacárido de 9 glucosas enlazadas en configuración α(1,4)), y se observó que la curvatura en el sustrato es promovida por una región vecina al enlace glicosídico a procesar (35) (véase figura 11). Las regiones estructurales en CGTasas que participan en la generación de tal estructura cíclica en el sustrato, están localizadas una en el dominio A, otra en el dominio B y otra mas en el dominio E (35). Estas observaciones sugieren que las CGTasas son transferasas gracias a la combinación de dos habilidades, la de obstruir el paso del agua al sitio activo (Tyr195 y forma cíclica de unir al substrato), y la de poseer sitios afines al aceptor en una región vecina al sitio de procesamiento del donador. El sitio afín al aceptor en reacciones de ciclización estaría localizado en el dominio B y se ha determinado que es de baja afinidad: 10 veces menor a la afinidad del sitio de unión al donador (2). Alternativamente, el sitio de unión al aceptor en reacciones de desproporción y acoplamiento, estaría situado en el dominio A, justo en el sitio que ocupa el grupo saliente del donador.

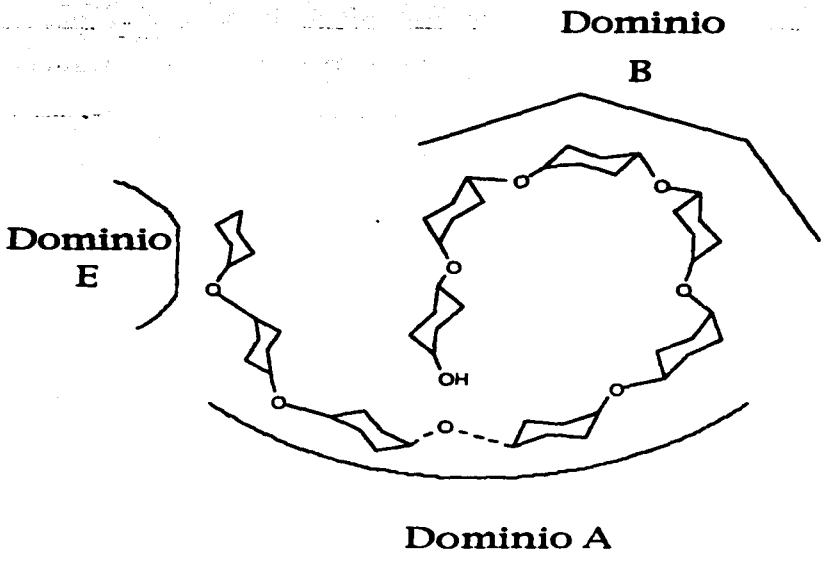


Figura 11. Esquema de la Curvatura Inducida al Sustrato en el Sitio Activo de CGTasas

En linea punteada se indica el enlace glicosídico a procesar

De este análisis concluimos que el dominio B en CGTasas juega un papel importante en la exclusión del agua por estas enzimas, ya sea a través del residuo Tyr195 localizado en este dominio, o a través de la estructuración del substrato, que en ambos casos promueven un

ambiente hidrofóbico en la parte superior del enlace glicosídico a procesar. Cabe señalar que la enzima debe desarrollar una carga complementaria al catión glicosídico, y esta estaría por debajo del enlace a procesar, como lo sugerimos en el trabajo sometido a publicación 1.

Si buscamos producir una α-amilasa que evite la entrada del agua al sitio activo, deberíamos utilizar estas cualidades observadas en las CGTasas para el dominio B, pero sustituyendo la habilidad de reconocer al aceptor desarrollada por las CGTasas, por la habilidad de reconocer a otro aceptor (por ejemplo, un alcohol alifático). Ya que este dominio es particularmente variable en longitud y secuencia mas no en estructura, entre las α-amilasas y CGTasas analizadas, consideramos que mutaciones puntuales en este dominio en α-amilasas, no conducirían eficientemente a transferir la función de exclusión del agua a una α-amilasa. En su lugar proponemos translocar el dominio completo, para posteriormente mutar, si es necesario, el sitio de unión al aceptor en este dominio.

Estos resultados complementan una hipótesis que planteamos sobre el origen de la actividad de CGTasa (véase trabajo sometido a publicación

2). En este estudio, proponemos que las CGTasas divergen de un ancestro con características de α-amilasa, y que es posible que la manera en como divergió la actividad de CGTasa de este ancestro fue mediante la adición de dominios. Nuestras observaciones en la comparación estructural entre CGTasas y α-amilasas indican que, además de la adición de dominios planteada durante el proceso de evolución de la actividad de CGTasa, debe ocurrir una alteración del dominio B.

Adicionalmente, cuando analizamos otros pares de hidrolasas-transfera sas distinto de la α-amilasas-CGTasas, pudimos observar que hay semejanzas en las estrategias usadas para evitar la entrada del agua al sitio activo por las transferasas (trabajo en preparación), en particular, incluir un sitio de afinidad por el aceptor en el sitio activo de las enzimas, que tenga la capacidad de excluir al agua.

VI.3.- Deleción del Dominio B de la α-Amilasa de *B. stearothermophilus*.

Al momento en que llegamos a analizar la relevancia del dominio B en la función de exclusión del agua, no teníamos acceso al gene de ninguna CGTasa, ya que estas enzimas tienen gran valor a nivel industrial (31).

Con la idea de explorar la viabilidad de una translocación en este dominio, iniciamos un proyecto que buscaba sustituir, a nivel genético, este dominio por péptidos conectores de distinto tamaño y longitud (1 a 8 amino ácidos de longitud). La idea era observar si alguna de estas deleciones-sustituciones podía mantener actividad enzimática. Esto implicaría que si una deleción es tolerada, la translocación también podría serlo. Estas construcciones tienen otras ventajas. Por ejemplo, se ha propuesto que la dependencia que las α-amilasas tienen por calcio para su actividad reside en la estructuración que el calcio da al dominio B (3). Si esto fuera correcto, la deleción del dominio eliminaría la necesidad por calcio de las α-amilasas, propiedad deseable en los procesos industriales de producción de jarabes en las α-amilasas (32). Adicionalmente, esta construcción nos permitiría estudiar el papel que este domino tiene en otras funciones de las α-amila-sas, como por ejemplo la cinética de plegamiento.

VI.3.1- Deleción del Dominio B de la ABST por PCR.

Al iniciar este proyecto, solo las estructuras tridimensionales de α-amila sas fungales y animales eran conocidas, las cuales tienen un 25% de identidad en secuencia con la α-amilasa de nuestro estudio. De aquí que

para localizar los límites del dominio B en la ABST, recurrimos a métodos de predicción de estructura y alineamiento de secuencia con las α-amilasas de estructura tridimensional conocida (véase Métodos). De este estudio concluimos que la tercera hoja beta y la tercera alfa-hél ice del TIM-barrel en la ABST están comprendidas entre los amino ácidos Asn97-Val103 y Pro209-Trp221, respectivamente (numeración referida a la ABST).

Por otra parte, ya que la actividad de α-amilasa no es seleccionable, buscamos generar un banco de variantes en la zona del dominio B, que permitiera rastrear todas las variantes dentro de los límites de transformación bacteriana (10^7, véase publicación 1) y dentro de números humanamente ensayables. Para limitar la variabilidad, analizamos la diversidad natural observada en los conectores en el sitio catalítico de 10 enzimas con estructura tipo TIM-barrel (véase Tabla 1). Es importante mencionar que entre todas las enzimas tipo TIM-barrel conocidas (28), el sitio activo es conformado por los conectores en el C-terminal de las hojas beta, como es el caso del dominio B.

Conector

Estructura	1	2	3	4	5	6	7	8	Media
FCB2	15	6	6	9	3	13	5	12	8.6
GO	12	7	5	9	4	13	5	11	8.2
RUBISCO	8	12	3	5	4	15	6	10	7.9
MLE	2	3	6	4	2	4	4	36	7.6
PK	4	6	94	3	6	8	13	6	17.5
ENOL	19	5	2	7	7	6	8	6	7.5
XYI	21	10	18	14	15	10	6	8	12.7
TS	8	10	8	9	5	14	4	2	7.5
TIM	6	4	15	2	3	13	5	8	7
AMYL	24	34	62	6	8	15	13	26	23.5
Media	11.9	9.7	21.9	6.8	5.7	11.1	6.9	12.5	

Tabla 1.- Longitud de conectores catalíticos en TIM-barrels. Las longitudes de los 8 conectores entre C-terminales de hojas beta y N-terminales de α-hélices del dominio de TIM-barrel fueron determinados usando las estructuras de flavocitocromo B2 (FCB2), Glicolato oxidasa (GO), Ribulosa 1,5-bifostafo carboxilasa/ oxigenasa (RUBISCO), Enzima lactonizante del muconato (MLE), Piruvato quinasa (PK), Enolasa (ENOL), Xilosa isomerasa (XYI), subunidad α del complejo Triptofano sintetasa (TS), Triosa fosfato isomerasa (TIM), α-amilasa (AMYL). La longitud promedio de los conectores (Media) es indicada para cada estructura analizada y para cada conector.

De este estudio concluimos que la longitud promedio de un conector es de 8, y que conectores mayores a 8 amino ácidos de longitud eran menos frecuentes que los de menos de 8 amino ácidos de longitud, por lo que nuestro diseño de oligonucleótidos debería incluir estos dos cualidades, es decir, conectores de 1 a 8 amino ácidos de longitud y cuya diversidad sea menor a 10^7 (véase Métodos).

VI.3.2.- Purificación y Renaturalización de Enzimas.

El gene silvestre de la ABST se obtuvo por PCR de genoma con un par de oligonucleótidos que incluían un sitio Shine-Dalgarno consenso para *E. coli* y que cambiaban el GTG inicial en la transcripción por ATG (véase trabajo sometido a publicación 1). Tanto a la enzima nativa como a las mutantes se les añadió en su extremo C-terminal 6 histidinas mediante la técnica de PCR (véase Métodos), las cuales les permitirían adherirse específicamente a la resina de níquel. La adición de estas histidinas se realizó anticipando diferencias en propiedades fisicoquímicas de las mutantes que dificultarían su purificación.

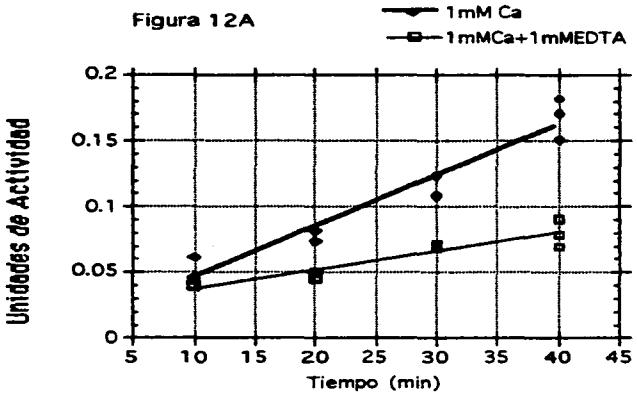
Después de ensayar 30 colonias portadoras de genes mutantes en caja petri para su actividad amilolítica (véase Métodos), seleccionamos 2 de ellas que parecían ser activas. Estas mutantes y la enzima nativa fueron purificadas en forma desnaturalizante, con 8M urea, mediante una columna de afinidad a níquel de acuerdo a las especificaciones del fabricante (QIAGEN, USA). Ninguna de las enzimas mutantes pudo ser obtenida en forma soluble directamente, ya sea mediante afinidad con almidón o bien mediante afinidad con níquel. Cabe señalar que incluso la enzima nativa no pudo purificarse en forma soluble mediante afinidad por la columna de níquel. En los experimentos de replegado en la columna (véase Métodos) observamos que al ir disminuyendo la concentración del agente desnaturalizante (urea) en la columna, la enzima se eluía (datos no mostrados). Este resultado nos indica que tan pronto la enzima se pliega, la cola de Histidinas ya no puede seguir manteniéndola adherida a la columna. Una posible explicación a este fenómeno es que al momento de plegarse la enzima, la cola de histidinas deja de estar expuesta y por tanto ya no puede seguir adherida a la columna. Esto también explicaría el por qué no se purificó la enzima en forma no desnaturalizante mediante este procedimiento.

Después de varios experimentos de replegado (véase Métodos), encontramos una metodología con la que pudimos recuperar actividad tanto en la enzima nativa como en una de las dos mutantes seleccionadas, aunque en ambas enzimas recuperamos poca actividad, 15 y 6 unidades por mL de volumen final de purificación, para enzima nativa y mutante respectivamente (1 unidad de actividad amilolítica la definimos como el número de μg de azúcares reductores producidos cada minuto). En ambos casos, la concentración de proteína fue la misma (10 μg/mL de volumen final de purificación).

VI.3.3.- Efecto del Calcio en la Actividad Amilolítica.

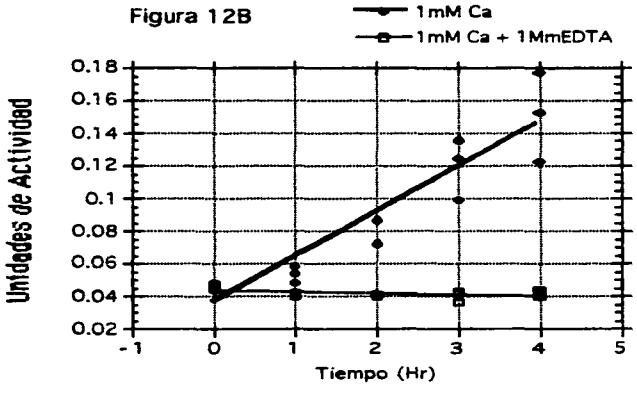
Con estos resultados en la purificación de nuestras enzimas, se realizó una caracterización cinética de la dependencia al calcio para la actividad amilolítica de la enzima nativa y una de las mutantes (deleción 6, véase mas adelante), mediante EDTA como quelante (véase Métodos). Los resultados se muestran en la figura 12.

Figura 12A



Enzima nativa

Figura 12B



Enzima mutante

De estos resultados determinamos que, al secuestrar el calcio, la enzima nativa (figura 12A) retiene un 37% de la actividad, mientras que la mutante (figura 12B) pierde por completo la actividad.

Al analizar la estructura de los sitios de unión a calcio en las α-amilasas de estructura tridimensional conocida, se observa que 2 de los 4 residuos de amino ácido que participan en la unión de un calcio que enlaza al dominio A con el B, están en el dominio A (3). Nuestra observación experimental indica que esos dos residuos son suficientes para mantener un sitio de unión a calcio, aunque de menor afinidad que el original. Consistente con este resultado es un estudio hecho con híbridos en el dominio B de dos isoformas de una α-amilasas de trigo (14). En este trabajo se propuso que no solo el dominio B debería especificar la dependencia al calcio.

De este estudio observamos que nuestra predicción de la localización del dominio fué acertada. La reciente elucidación de la estructura de la α-amilasa de *B. licheniformis*, enzima 80% idéntica a la ABST, nos permitió valorar nuestra predicción: nos equivocamos en la longitud tanto de la tercera hoja beta como la tercera alfa-hélice del TIM-barrel. Sin embargo, los límites C terminal de la tercera hoja beta y el N terminal de la tercera α-hélice los predijimos correctamente, excepto por un aminoácido en el C-terminal de la tercera hoja beta. Esto tuvo como consecuencia que incluimos un amino ácido (de la tercera hoja beta) de

más a cada péptido conector sustituto. Por otra parte, también observamos que el dominio B es dispensable para la actividad amilolítica en la α-amilasa de *B. stearothermophilus* , pero que aún en su ausencia, esta requiere del calcio para su actividad.

VI.3.4.- Secuenciación de las Mutantes.

Al determinar la secuencia nucleotídica de 7 de las mutantes, encontramos que la mayoría poseían un conector sustituto de 6 residuos de longitud, que la variante activa presentó mayor número de amino ácidos polares y que en todas se introdujeron cambios en la secuencia nucleotídica en la tercer alfa-hélice del TIM-barrel, es decir, justo después del dominio B (residuo prolina (P) en letra negrilla):

Secuencia	Actividad	
	en caja	*in vitro*
- Deleción 1: mqvyadvvVIVAGEsevvtelkswg	Si	ND
- Deleción 2: mqvyadvvGGGGGGsevvtelkswg	Si	ND
- Deleción 3: mqvyadvvSCGGG.pevvselkswg	Si	ND
- Deleción 4: mqvyadvvGGGRGGpevvselkswg	Si	ND
- Deleción 5: mqvyadvvGGGSGGsevvtelkswg	Si	ND
- Deleción 6: mqvyadvvRDGGGVsevvtelkswg	Si	Si
- Deleción 7: mqvyadvvIVVAGEsevvtelkswg	Si	Si
- Silvestre: mqvyadvvF... Hpevvtelkaga	Si	Si

Estructura bbbbbbbb hhhhhhhhhhh

La secuencia subrayada corresponde al péptido conector sustituto del domino B. ND: No determinada. La deleción 6 fue la ensayada en presencia de EDTA. b: Tercera beta-plegada en el TIM-barrel, h: Tercera alfa-hélice en el TIM-barrel.

Aún no sabemos si parte de la diferencia observada por la quelación del calcio en la actividad amilolítica, entre la mutante y la enzima nativa, pueda explicarse por los cambios detectados en la secuencia de la tercera α-hélice o por una función *bona fide* de estructuración del Dominio A por el calcio.

VII.- CONCLUSIONES Y PERSPECTIVAS

En este trabajo estudiamos el mecanismo de reacción de α-amilasas, con la idea de proponer una estrategia que permita favorecer las reacciones de transglicosidación o alcohólisis. Esto lo abordamos con tres aproximaciones. La primera fue estudiar la función de un residuo

invariante entre α-amila-sas, el cual se había reportado que mejoraba la actividad transglicosídica en la α-amilasa de *S. flivuligera*. De este estudio concluímos que tanto ese residuo (Tyr82 en la numeración de la taka-amilasa) como los residuos Arg204 y His297 deberían participar en la generación de una carga complementaria al estado de transición (catión glicosídico). Sin embargo, no detectamos ninguna alteración en la actividad transglicosídica en la α-amilasa de nuestro estudio.

La segunda aproximación consistió en comparar a las α-amilasas y las ciclodextrin glicosiltransferasas (CGTasas) a nivel estructura primaria, terciaria y cuaternaria. Esta comparación la realizamos con la finalidad de encontrar los elementos estructurales que pudieran favorecer la exclusión del agua del sitio activo en las CGTasas. De este análisis concluímos que el dominio B en las CGTasas tendría un papel preponderante en la exclusión del agua del sitio activo. Con base en esta observación proponemos que la translocación del dominio B de una CGTasa en una α-amilasa debería generar una enzima con reducida capacidad hidrolítica, pero con actividades transglicosídicas o alcohóliticas mejoradas. Alternativamente se podrían deletar los dominios D y E en una CGTasa (véase Figura 7) y ensayar la actividad alcoholítica de esta construcción.

La tercera estrategia se planeó como una alternativa experimental para valorar la viabilidad de una translocación en el dominio B de la α-amilas a de *B. stearothermophilus* (ABST). Esta estrategia consistió en deletar el dominio B de la ABST. De este estudio concluímos que la deleción del dominio B aunque disminuye la actividad amilolítica de la ABST, aún permite mantener actividad en una de las deleciones analizada. Este resultado nos sugiere que una translocación del domino B de una CGTasa por el de la ABST es viable.

Como extensión del objetivo primordial de este estudio, propusimos una función biológica para las CGTasas y un ancestro común para α-ami lasas y CGTasas (véase trabajo sometido a publicación 2). También observamos que la capacidad de transglicosidar está ligada a la capacidad de alcoholizar, al menos en las 6 α-amilasas estudiadas (trabajo en preparación). Asi mismo, proponemos una estrategia general para diseñar una transferasa a partir de una hidrolasa (trabajo en preparación). Adicionalmente, construimos una mutante de la ABST carente del dominio B cuya actividad es menor a la enzima nativa, pero que puede servir como semilla de futuros experimentos de mutagénesis que busquen eliminar la necesidad de calcio para la estabilidad de la

ABST. Finalmente, desarrollamos un programa de computación útil para diseñar estrategias de mutagénesis óptimas (véase publicación 1) y estos conceptos los empleamos en el diseño de nuestras estrategias experimentales (véase Publicación 1 y Métodos).

De los resultados de este trabajo planteamos las siguientes perspectivas útiles para generar una enzima con capacidades mejoradas para realizar reacciones de transglicosidación o alcohólisis. Primero, para generar la translocación del dominio B de una CGTasa a la ABST, conseguimos que el Dr. Dirk Penninga, de la Universidad de Groningen, Holanda, nos diera el gene de la CGTasa de *B. circulans* 251 y actualmente estamos caracterizando la posible actividad del híbrido. Segundo, del trabajo con el grupo del Dr. Agustín López-Munguía sobre la relación tranglicosidación/ hidrólisis observada en 6 α-amilasas, hemos construído un par de mutantes en el residuo Ala289 en la ABST (Ala289Phe y Ala289Tyr), con lo que buscamos demostrar que la capacidad de transferasa se puede incrementar bloqueando un "canal de aguas" que se dirije al sitio catalítico. Actualmente estamos caracterizando estas dos mutantes.

Bibliografía

1. Arnold, F.H. Engineering Proteins for nonnatural environments. FASEB J. 7:744-749, 1993.

2. Bender, H. Studies of the Mechanism of the Cyclisation Reaction Catalysed by the Wildtype and a Truncated alpha-Cyclodextrin Glycosyltransferase from Klebsiella pneumoniae Strain M 5 a1, and the beta-Cyclodextrin Glycosyltransferase from Bacillus circulans Strain 8. Carbohydr. Res. 206:257-267, 1990.

3. Boel, E., Brady, L., Brzozowski, A.M., Derewenda, Z., Dodson, G.G., Jensen, V.J., Petersen, S.B., Swift, H., Thim, L. and Woldike, H.F. Calcium Binding in alpha-Amylases: An X-ray Diffraction Study at 2.1-Å Resolution of Two Enzymes from Aspergillus. Biochemistry 29:6244-6249, 1990.

4. Chopineau, J., McCafferty, F.D., Therisod, M. and Klibanov, A.M. Production of Biosurfactants from Sugar Alcohols and Vegetable Oils Catalyzed by Lipases in a Nonaqueous Medium. Biotechnol. Bioeng. 3 1:208-214, 1988.

5. Cormack, B.P. and Struhl, K. Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III trascription. Science 262:244, 1993.

6. Davies, G. and Henrissat, B. Structures and mechanisms of glycosyl hydrolases. Structure 3:853-859, 1995.

7. del Río, G., Osuna, J. and Soberón, X. Combinatorial libraries of proteins: analysis on the efficiency of mutagenesis techniques. Biotechniques 17: 1132-1139, 1995.

8. Farber, G.K. and Petsko, G.A. The evolution of alpha/beta barrel enzymes. TIBS 15:228-234, 1990.

9. Fersht, A. Enzyme Structure and Mechanism, USA:W.H. Freeman and Company, 1985. Ed. 2nd pp. 176-245.

10. Garnier, J. and Robson, B. The GOR Method for Predicting Secondary Structures in Proteins. In: Prediction of Protein Structure and the Principles of Protein Conformation, edited by Fasman, G.D. New York, USA: Plenum Press, 1989, p. 417-466.

11. Holm, L., Kolvula, A., Lehtovaara, P., Hemminki, A. and Knowles, J. Random mutagenesis used to probe the structure and function of Bacilus Stearothermophilus alpha-amylase. Protein Eng. 3:181-191, 1990.

12. Janecek, S., MacGregor, E.A. and Svensson, B. Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from alpha-amylases. Biochem. J. 305:685-688, 1995.

13. Jespersen, H.M., MacGregor, E.A., Sierks, M.R. and Svensson, B. Comparison of the domain-level organization of starch hydrolases and related enzymes. Biochem. J. 280:51-55, 1991.

14. Juge, N., Rodenburg, K.W., Guo, X.J., Chaix, J.C. and Svensson, B. Isozyme Hybrids Within the Protruding Third Loop Domain of the Barley alpha-Amylase (beta/alpha)8-Barrel. Implication for BASI Sensitivity and Substrate Affinity. FEBS Lett. 363:299-303, 1995.

15. Koshland, Jr.,D.E. Stereochemistry and the mechanism of enzymatic reactions. In: , edited by Brookhaven National Laboratory, N.Y.,U.S.A. N.Y., U.S.A.: Brookhaven National Laboratory, Upton, 1953, p. 416-436.

16. Kuroki, R., Weaver, L.H. and Matthews, B.W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. Science 262:2030-2033, 1993.

17. Matsubara, S. Studies on Taka-amylase A. VII. Transmaltosidation by Taka-Amylase A. J. Biochem.(Tokyo) 49:226-231, 1961.

18. Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S. and Honda, K. An increase in the transglycosylation activity of Saccharomycopsis alpha-amylase altered by site-directed mutagenesis. Biochim. Biophys. Acta 1077:416-419, 1991.

19. Mattsson, P., Korpela, T., Paavilainen, S. and Mäkelä, M. Enhanced Conversion of Starch to Cyclodextrins in Ethanolic Solutions by Bacillus circulans var alkalophilus Cyclomaltodextrin Glucanotransferase. Appl. Biochem. Biotechnol. 30:17-28, 1991.

20. Merino, E., Osuna, J., Bolívar, F. and Soberón, X. A general, PCR-base d method for single or combinatorial oligonucleotide-directed mutagenesis on pUC-M13 vectors. Biotechniques 12(4):508-510, 1992.

21. Mullholland, A.J., Grant, G,H. and Richards, W.G. Computer Modelling of Enzyme Catalyzed Reaction Mechanisms. Protein Eng. 6:133-147, 1993.

22. Nakamura, A., Haga, K. and Yamane, K. The transglycosylation reaction of cyclodextrin glucanotransferase is operated by a ping-pong mechanism. FEBS Lett. 337:66-70, 1994.

23. Palmer, T. The Chemical Nature of Enzyme Catalysis. In: Understanding Enzymes, edited by Wiseman, A. West Sussex: Ellis Horwood Limited, 1985, p. 208-238.

24. Penninga, D., Strokopytov, B., Rozeboom, H.J., Lawson, C.L., Dijkstra, B.W., Bergsma, J. and Dijkhuizen, L. Site-Directed Mutations in Tyrosine 195 of Cyclodextrin Glycosyltransferase from Bacillus circulans Strain 251 Affect Activity and Product Specificity. Biochemistry 34:3368-3376, 1995.

25. Prevelige, P.,Jr. and Fasman, G.D. Chou-Fasman Prediction of the Secondary Structure of Proteins: The Chou-Fasman-Prevelige Algorithm . In: Prediction of Protein Structure and the Principles of Protein Conformation, edited by Fasman, G.D. New York: Plenum press, 1990, p. 391-416.

26. Qian, M., Haser, R., Buisson, G., Duée, E. and Payan, F. The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-Å resolution. Biochemistry 33:6284-6294, 1994.

27. Raimbaud, E., Buléon, A., Perez, S. and Henrissat, B. Hydrophobic cluster analysis of the primary sequences of alpha-amylases. Int. J. Biol. Macromol. 11:217-225, 1989.

28. Reardon, D. and Farber, G.K. The structure and evolution of alph a/beta barrel proteins. FASEB J. 9:497-503, 1995.

29. Sanz, J.M., García, P. and García, J.L. Role of Asp-9 and Glu-36 in the active site of the pneumococcal CPL1 lysozyme: an evolutionary perspective of lysozyme mechanism. Biochemistry 31:8495-8499, 1992.

30. Sarney, D.B. and Vulfson, E.N. Application of enzymes to the synthesis of surfactants. TIBTECH 13:164-172, 1995.

31. Schmid, G. Cyclodextrin glycosyltranferance production: Yield Enhancement by Overexpression of Cloned Genes. TIBTECH 7:244-248, 1989.

32. Shetty, J.K. and Allen, W.G. TAKA-TERM® II L-170. An Acid Stable, Thermostable alpha-Amylase For Starch Liquefaction. ENZYMES FROM MILES 1:1-31, 1990.

33. Sinnott, M.L. Catalytic mechanisms of enzymic glycosyl transfer. Chem. Rev. 90:1171-1202, 1990.

34. Stoddard, B.L., Dean, A. and Bash, P.A. Combining Laue Diffraction and Molecular Dynamics to Study Enzyme Intermediates. Nature Structural Biology 3:590-595, 1996.

35. Strokopytov, B., Knegtel, R.M.A., Penninga, D., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L. and Dijkstra, B.W. Structure of Cyclodextrin Glycosyltransferase Complexed with a Maltononaose Inhibitor at 2.6 Å

Resolution. Implications for Product Sprecificity. Biochemistry 35:4241-4249, 1996.

36. Strokopytov, B., Penninga, D., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L. and Dijkstra, B.W. X-ray Structure of Cyclodextrin Glycosyltransferase Complexed with Acarbose. Implications for the Catalytic Mechanism of Glycosidases. Biochemistry 34:2234-2240, 1995.

37. Summer, J.B. and Howell, S.F. A method for determination of invertase activity. J. Biol. Chem. 108:51-54, 1935.

38. Svensson, B. Protein engineering in the alpha-amylase family: catalytic mechanism, substrate specificity, and stability. Plant Molecular Biology 25:141-157, 1994.

39. Szejtli, J. Medical Applications of Cyclodextrins. Med. Res. Rev. 1 4:353-386, 1994.

Publicación 1

BioFeature

Combinatorial Libraries of Proteins: Analysis of Efficiency of Mutagenesis Techniques

Gabriel Del Río, Joel Osuna and Xavier Soberón
Instituto de Biotecnología/UNAM, Cuernavaca, Morelos, Mexico

ABSTRACT

A number of considerations are made on the efficiency of mutagenesis techniques used in protein engineering, particularly those that include a random component. The expected outcome of different protocols is analyzed using computer programs. Special emphasis is made on the effect that the degeneracy of the genetic code has on the bias of the representation of amino acid replacements. The consequences of using alternative methods is analyzed in terms of the likelihood of obtaining underrepresented amino acid substitutions in mutant libraries. A consideration is also made of the outcome of combinatorial mutagenesis experiments with regard to the size of the amino acid window and the multiplicity of replacements that could be sampled with different methods. Optimal mutagenesis rates for specific conditions could be derived from the presented data and the computer program made available with this paper.

INTRODUCTION

Since the advent of recombinant DNA technology, the study of protein structure and function has received tremendous impulse. The capacity to implement any desired amino acid replacement on proteins has produced a leap forward in the study of their sequence, structure and function relationship.

Variant proteins have been obtained by geneticists for a long time, with interesting phenotypes being selected out of very large numbers of cells. The lesions, at the DNA level, could then be identified after significant work. Subsequent analysis of the corresponding differences at the polypeptide level provides the bridge to the functional consequences of the mutation. Recombinant DNA permitted the sequence analysis of many mutants obtained in this way, but also took us to a new scenario, in which the production of the mutations is subject to control by the experimenter. The use of site-directed mutagenesis techniques is widespread and has now become routine to approach the study of protein structure and function by protein engineering. Moreover, the power of recombinant DNA-based mutagenesis has driven research to ever more ambitious schemes for the introduction of a great deal of variability, specially since the development of powerful screening systems (15,30,34). Several approaches have been utilized that incorporate such random components. Cassette mutagenesis (29), regional mutagenesis (4,20) "spiked oligonucleotides" (19) and combinatorial mutagenesis (26, 28,29) are terms referring to such approaches. The use of combinatorial libraries in present day biotechnology seems an important way to develop novel molecules (5,10–12).

In cassette mutagenesis, a target region is selected and suitable adjacent restriction sites are found or introduced. A synthetic DNA duplex is made, containing the desired mutation(s) and subsequently cloned, replacing the wild-type sequence. Regional mutagenesis is achieved by several methods. In general, the protein of interest is subjected to any of several treatments (chemical or enzymatic), using conditions that introduce a controlled rate of alterations per nucleotide. The product of such reactions is then cloned. The term "spiked oligonucleotides" has been used to refer to methods employing synthetic oligonucleotides in which some amount of a mixture of different bases is included at predetermined positions, thus "spiking" the wild-type base during the synthesis. The oligonucleotides are then used to prime an extension reaction with its final incorporation into heteroduplex DNA molecules, which are cloned. The goal of combinatorial mutagenesis is to generate mutants with several amino acid replacements per protein. A region or "window" is defined, and the rate of mutagenesis is set as high as necessary for the multiplicity of replacements desired. This has generally been done using cassettes.

Since the usual mutagenesis unit is the nucleotide and the protein-variation unit is the amino acid, the genetic code, which relates the two, poses specific constraints on the process of creating collections of protein variants (33). This facet of mutagenesis is only superficially mentioned in much of the literature, and only a few papers have dealt with it specifically (19,28). These considerations become particularly apparent when one considers the astronomically high numbers of variants that can be easily generated by recombinant DNA methodology. A combinatorial mutagenesis scheme involving only 7 amino acid residues would generate over 10^9 variant proteins under ideal conditions, which is above the practical limits of currently available screening methodologies (3,22,30). Given that the number of residues directly involved in molecular recognition and/or the catalytic function of any given protein is likely to be higher than 7, there is a need for optimizing mutagenesis procedures in order to perform the screening on the best possible collection of variants. Indeed, some recent literature explores various ways to achieve that goal (7,8,15,21,24,31).

In this paper, we analyze the expected outcome of different mutagenesis schemes, as reflected in the frequency, distribution and variety of amino acid replacements. We focus primarily on the effects of various ways of "doping" oligonucleotides to make degenerate mixtures and also on the variables involved in schemes that aim at non-saturating conditions. Such conditions have been utilized and analyzed before from other perspectives (9,17,19) and constitute an attractive way to better sample the sequence space with a collection more likely to contain interesting variants (8,33). The study we present here aims to provide a general reference and some procedures to evaluate and compare mutagenesis regimes in terms of the expected outcome at the level of amino acid replacements, rather than nucleotide changes.

MATERIALS AND METHODS

Computer programs were written in either Turbo PASCAL Version 5.5 (Borland-Osborne, Scott's Valley, CA, USA) or Visual BASIC (Microsoft, Redmond, WA, USA) and run on an IBM® PC-compatible machine or written in ANSI C and run on a DECstation 3100 (Digital Equipment Corporation, Maynard, MA, USA).

The programs considered each of the 61 sense codons that comprise structural genes. The outcome of a mutagenesis experiment was calculated based on the probability of finding any base on each of the three codon positions. Such probability depends on the mutagenesis method employed. The resulting distribution of DNA sequences was then evaluated in terms of the distribution of amino acids for which it coded.

Amino Acid Replacement Rate

The quantification of this parameter was done in terms of the expected frequency of each codon in the mutant populations generated, assuming that incorporation of any base at a given codon position was equally likely and proportional to the intensity of the mutagenic treatment. Therefore, for each base at any codon position:

$$F_b = F_{bw} + ([1-F_w]*0.25) \qquad \text{[Eq. 1]}$$

where F_b = fraction of a specific base at each codon position, F_w = fraction of wild-type sequence (this is the contribution of the starting DNA, in the case of non-oligonucleotide mutagenesis, or the base added from the pure base vial, in the case of spiked oligonucleotides). $1-F_w$ = fraction of mutagenized bases. We designate this as "mutagenesis rate." It corresponds to the targeted base substitution rate from any method. F_{bw} = fraction of the base contributed by the wild-type sequence ($F_{bw} = F_w$ for the base in the wild-type codon, $F_{bw} = 0$ for the other 3 bases).

This formula was used for the so-called NNN mutagenesis scheme at all three positions. For the NNG/C mutagenesis case, the frequencies for the first two bases of a codon were calculated with Equation 1, and a frequency of 0.5 for each of the bases C and G was used at the third position (see Results for considerations on the experimental conditions pertaining to these calculations). A set of calculations were performed for several mutagenesis rates on the two mentioned approaches. For each codon, 64 or 32 frequency values were generated, corresponding to each possible triplet obtainable by the NNN and NNG/C methods, respectively. The values of synonymous triplets were then added together, giving a total of 20 frequencies (A_M), one for each amino acid replacement and one value for the stop codon(s). A frequency (A_W) was also computed for the wild-type amino acid. Fractions $1-A_W$, over all 61 codons, were averaged for each mutagenesis rate, giving a corresponding amino acid replacement rate (A_{RR}).

Window Size

We employed the following equation (6) to relate the library size, N, with the probability, P, of sampling a mutant represented at a fraction S in the library:

$$N = [\ln(1-P)]/[\ln(1-S)] \qquad \text{[Eq. 2]}$$

P was set to 0.90, and values of S could be derived for any value of N. Hence:

$$S = 1 - e^{(\ln(1-P)/N)} \qquad \text{[Eq. 3]}$$

The values for S were then used to calculate a window size, n, using Equation 4:

$$S = A_W^{(n-m)}(1-A_W)^m(A_{MD}/[1-A_W])^m = A_W^{(n-m)}A_{MD}^m \qquad \text{[Eq. 4]}$$

where m = the number of amino acid replacements per window (amino acid replacement multiplicity).

Values for A_{MD}, the frequency of an amino acid replacement requiring two base-pair changes, were calculated as per Equation 1:

$$A_{MD} = [F_{bw} + (1-F_w)*0.25][(1-F_w)*0.25]^2$$

for NNN mutagenesis and

$$A_{MD} = 0.5*[(1-F_w)*0.25]^2$$

for NNG/C mutagenesis. We have analyzed only the case of two base-pair changes in order to directly compare the NNG/C and the NNN approaches in the case of hard-to-get mutations. In the NNG/C approach, the third codon position

is saturated with G/C, and therefore it is meaningless to consider triple base-pair changes.

Corresponding values for $A_W = 1-A_{RR}$ were calculated by interpolation, from a polynomial expression fitted to the data of A_{RR} vs. $1-F_W$ derived previously (see Figure 1). Hence, from Equation 4:

$$n = [\,(\mathrm{Ln}(S) - m^*(\mathrm{Ln}(A_{MD}))\,/(\mathrm{Ln}(A_W))\,] - m \qquad [\mathrm{Eq.}\ 5]$$

These formulas (1–5) were then used to calculate any desirable variable, namely, amino acid replacement multiplicity (m), library size (N), window size (n) or optimal mutagenesis rates.

Optimal Mutagenesis Rate

Using an iterative process and a graphic representation (as implemented in the PC interface to the programs described below), optimal mutagenesis rates were found for specific conditions. A window size, a mutant order and an arbitrary mutagenesis rate are chosen; the program returns the required library size and a graphic representation that indicates the range of mutagenesis rates at which such window and library sizes are attainable. The maximum of the curve constitutes a better mutagenesis rate, from which a smaller library size can be calculated. Repetition of this process rapidly converges to an optimal mutagenesis rate for the specified window size and amino acid replacement multiplicity.

Program for Analyzing Diverse Mutagenesis Conditions

We developed a user interface for the application of the equations described above. The user specifies a set of conditions and the program calculates the unspecified variable or variables. For example, one can ask the following question: "What window size can I sample when looking for triple replacements at 85% probability if I am able to screen 10^5 variants?" Compare the sizes for the NNN or the NNG/C methods and find out the optimal mutagenesis rate for each approach. The program was developed in Visual BASIC (Microsoft) for the PC Windows™ environment, and it is available from the authors.

RESULTS AND DISCUSSION

Computer programs were written for calculating the expected outcome of mutagenesis schemes. In relating our calculations to the actual experimental situations, it should be kept in mind that several assumptions/simplifications were made. For a chemical or enzymatic mutagenesis procedure, a random distribution of base changes was used; that is, replacement of any base by A, G, C or T had equal probability. This similarly applies to a method utilizing oligonucleotides as mutagenic agents. We first considered the average frequency of amino acid replacements expected when each of the 61 amino acid-coding base triplets were mutated, at a given rate, at any of the three codon positions. This can be called NNN mutagenesis, and it is applicable to the idealized forms of methods not employing oligonucleotides as mutagens (chemical or enzymatic, for instance). It would also correspond to a spiked oligonucleotides approach in which the DNA synthesis is performed using mixtures of the wild-type base in one vial and equimolar amounts of the four bases in another vial (19).

As plotted in Figure 1, the frequency of mutant protein vs. mutant DNA rises in a nonlinear fashion and varies depending on the method used. Experiments utilizing all four bases at each of the three codon positions are referred to as NNN; those using only G or C at the third wobble base are called NNG/C. This latter scheme has been utilized (11,29) as a means to reduce the appearance of termination codons (from 3/64 to 1/32) and to make the distribution of some amino acids more even (met and trp from 1/64 to 1/32); the approach requires synthetic DNA. A further variation should be considered here: namely, the experimenter could choose to introduce some proportion of the wild-type base at the third position of the codon or, alternatively, a 100% contamination of the G plus C mixture could be used (both methods would afford a sampling of all amino acid replacements). Only results of the latter approach are given, since results from the former are virtually identical to the NNN alternative. However, note that codon usage would be significantly changed with this approach, as every mutant codon would end with G or C, regardless of the original codon present. The plots show the values as an average of the 61 sense codons grouped according to the amino acid for which they code. Note that NNN will produce an amino acid replacement rate (A_{RR}) greater than NNG/C; thus, for the same A_{RR}, NNG/C will need a greater mutagenesis rate than NNN. In comparing the properties of different methods, we used an approach aimed at simplifying the evaluation of relevant parameters involved in current mutagenesis designs. This approach considered the actual rate

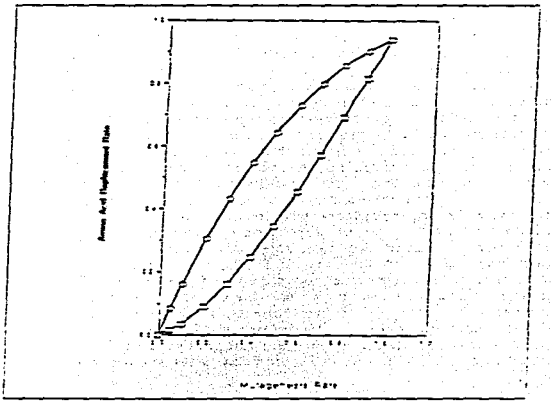


Figure 1. Comparison of the amino acid replacement rates in libraries produced by NNN (circles) and NNG/C (squares) mutagenic approaches.

of amino acid replacement, A_{RR}, rather than the rate of nucleotide replacement. This parameter is therefore used in several of the calculations that follow.

Another important consideration, the bias of the frequency expected for each of the 20 possible amino acids, has been discussed (33). By virtue of the genetic code, amino acids that are translated from 6 or 4 codons turn out more frequently on a mutant collection than those coded for by only one or two codons. This imbalance shows up more strongly at low mutagenesis rates, as illustrated in Figure 2 for the arginine codon AGG. Consequently, conventional low-rate mutagenesis procedures sample sequence space essentially at the more accessible amino acid replacements—those requiring only one base change (19).

It has been argued (33) that using triplets as mutagenic units would result in significantly better mutant collections. Such an approach would afford an unbiased amino acid composition (by employing, at the DNA synthesis stage, an equimolar mixture of 20 codons, one for each amino acid, in the mutagenic vial), irrespective of the mutagenesis rate. A similar, constantly unbiased mutagenesis could be achieved with NNN or NNG/C approaches, but would use a more elaborate DNA synthesis protocol involving partitioning and remixing of the solid support (7,15,21) (see Conclusion and Recommendations).

Mutagenesis of large DNA segments, such as a regional mutagenesis scheme (4) over the whole gene or a method utilizing long "spiked oligonucleotide" (23), are typically implemented utilizing relatively low mutagenesis rates aimed at obtaining populations of genes with low proportions of multiple amino acid replacements. The question could be formulated in terms of the probability of finding a particular type of mutant, given the fraction it represents in a mixture and the size of the population accessible for screening. With this in mind, we decided to analyze the population sizes required for the sampling of single and triple amino acid replacements, as a function of the length of the target region (window size, see Materials and Methods), with the various mutagenesis methods. We set a cutoff value at 90% probability of finding an underrepresented mutant when looking at various window sizes and either single or triple amino acid replacements; and the library size that would be required was calculated. The results, shown in Figure 3A, demonstrate that there are noticeable differences among the three approaches in all conditions. For larger window sizes and higher multiplicities, the differences become more pronounced. Notice that it is possible to adequately sample single replacements in domain-sized windows with any method by using appropriately low mutagenesis rates. It is clear that mutagenesis rates larger than the optimal would increase the probability of finding multiple replacements, therefore raising the complexity of the library. Furthermore, it could be feasible to sample all possible triple replacements within a ten-residue window by using codon-based mutagenesis and selection (>10^7 clones). In our comparison we utilized the optimal mutagenesis rate for each method and goal, which varies significantly, as shown in Figure 3B.

Our results complement and extend previous studies that consider other aspects of mutagenesis efficiency. Hermes et al. (19) considered the library size and mutagenic conditions for sampling single amino acid replacements in large windows. Our analysis is consistent with their results and also considers other mutagenesis approaches and multiplicity of replacements. In fact, using our programs we established that the optimal mutagenesis rate for sampling single amino acid replacements requiring 2 base changes, within a window of 25 amino acids, is around 4% (the authors implement a 0.67% rate, achieve a 2% rate and consider that, theoretically, the proper rate is between 2% and 4%). Under 2% conditions, a library of 212 676 clones would be required to sample at a probability of 0.9. The authors, who attain 150 000, say that they can sample 60% of the single amino acid replacements. Furthermore, we find that merely by changing to an NNG/C approach and mutating at a 14.6% rate, all amino acid replacements could be sampled with a library of only 14 176 clones. Certainly, the number of mismatches expected for oligonucleotides spiked in this fashion (due to the presence of only G or C at third position) probably requires the use of

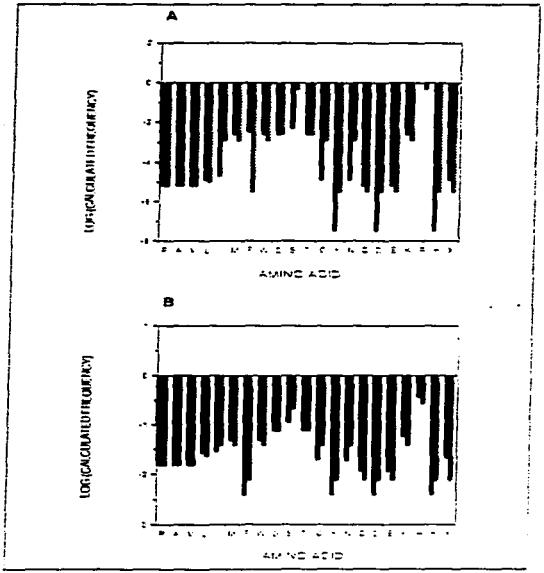


Figure 2. An example of the expected distribution of amino acid replacements at different mutagenesis rates. Values for the AGG arginine codon are shown for both mutagenic approaches, NNN (dark bars) and NNG/C (shadow bars). At A mutagenesis rate of 0.1 and at B mutagenesis rate of 0.4.

protocols involving the polymerase chain reaction (PCR), which might not have been fully developed some years ago. Clackson and Wells (5) analyzed the limits for saturating-condition mutagenesis with the NNG/C approach. We obtain essentially the same results with our programs and extended the analysis to NNN and codon-based mutagenesis and also to non-saturating conditions.

CONCLUSION AND RECOMMENDATIONS

Based on this study, we can draw some conclusions and provide some general recommendations for mutagenesis experiments.

The analysis of the amino acid rather than nucleotide replacement rate, together with useful experimental parameters (window size, mutagenesis rate and library size), suggest ways to improve designs of mutagenesis experiments. For instance, taking into account the differences between the three mutagenesis approaches analyzed here, one could decide the best mutagenesis rate parameter. This analysis could also be helpful to evaluate, with a small accessible library size, the probability of success in looking for desired mutants. These and other experimental strategies could be analyzed by just varying the parameters included in our calculations. The computer program, available from the authors, is intended to support such analysis. In a regional mutagenesis approach, perhaps the main difference between the chemical, enzymatic and the "spiked oligonucleotide" methods is related to the degree of control over the rate of mutagenesis and the degree of true randomization of each base in the mutagenesis window. For both parameters, the results should be better with the latter method. These three methods, as they stand today, suffer from the drawback imposed by the genetic code, which, in practice, limits the representation of amino acid replacements to a subset (e.g., about a third of all possibilities in small libraries aimed at one or a few replacements per gene; Reference 33).

Depending on various considerations, different mutagenesis schemes could be used for different questions and systems. Clearly, in some cases, sufficient knowledge is available to draw meaningful information from site-directed and site saturation mutagenesis experiments (18,25,37,38). In such cases, the methods are well established and do not represent a serious limitation.

On the other hand, our results illustrate that when one is aiming at an efficient sampling of sequence space in large windows and/or searching for multiple amino acid replacements, the choice of methodology does make a significant difference. The comparison in Figure 3 shows that sampling single amino acid replacements requiring two-base changes, at 90% probability, is only possible for small windows or large library sizes with the NNN method. The NNG/C method performs better and, as noted in Figure 1 and Figure 3B, requires higher mutagenesis rates. Codon-level mutagenesis affords a further improvement. Note that for some purposes the optimal mutagenesis rate is quite low, probably not compatible with all DNA synthesis machines (1).

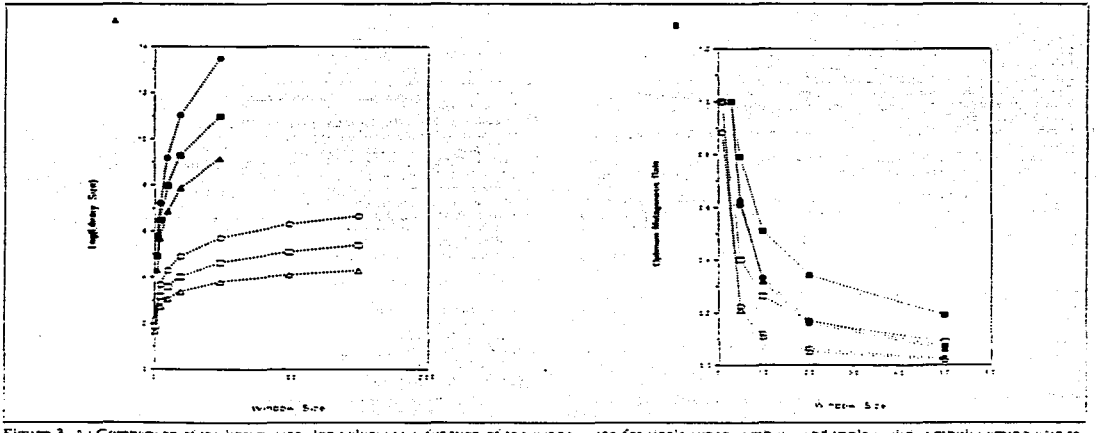


Figure 3. A) Comparison of the library size (log values) as a function of the window size for single (open symbols) and triple (solid symbols) amino acid replacements with diverse mutagenic approaches: NNN (circles), NNG/C (squares) and codon-based mutagenesis (triangles) are shown. B) Optimal mutagenesis rates required for the targeted window sizes and amino acid replacements multiplicity

ACKNOWLEDGMENTS

We are thankful to Sonia Caro and Alma Tremari for secretarial assistance and to the referees and members of our laboratory for helpful discussions and feedback. This work was supported, in part, by grants from PAPIID, DGAPA/UNAM and from PEMEX.

REFERENCES

1. **Arkin A.P. and D.C. Youvan.** 1992. Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis. *Biol/Technology 10*:297-300.
2. **Bowie, J.U., R. Lüthy and D. Eisenberg.** 1992. A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*:164-170.
3. **Bradbury, A., L. Persic, T. Werge and A. Cattaneo.** 1993. Use of living columns to select specific phage antibodies. *Bio/Technology 11*:1565-1569.
4. **Chen, K. and F.H. Arnold.** 1991. Enzyme engineering for nonaqueous solvents: random mutagenesis to enhance activity of subtilisin E in polar organic media. *Bio/Technology 9*:1073-1077.
5. **Clackson, T. and J.A. Wells.** 1994. In vitro selection from protein and peptide libraries. *TIBTEC. 12*:173-184.
6. **Clarke L. and J. Carbon.** 1976. A colony bank containing synthetic ColE1 hybrid plasmids representative of the entire E. coli genome. *Cell 9*:91-99.
7. **Cormack, B.P. and K. Struhl.** 1993. Regional codon randomization: defining a TATA-binding protein surface required for RNA polymerase III transcription. *Science 262*:244.
8. **Delagrave, S. and D.C. Youvan.** 1993. Searching sequence space to engineer proteins: Exponential ensemble mutagenesis. *Bio/Technology 11*:1548.
9. **Dubnick, M., A. Thliveris and D.W. Mount.** 1991. Mixed oligo designer (MOD), a computer program to aid planning of automated, mixed oligodeoxyribonucleotide synthesis for mutagenesis experiments. *Gene 105*:1-7.
10. **Dunn, C.R., H.M. Wilks, D.J. Halsall, T. Atkinson, A.R. Clarke, H. Muirhead and J.J. Holbrook.** 1991. Design and synthesis of new enzymes based on the lactate dehydrogenase framework. *Philos. Trans. R. Soc. Lond. [Biol.] 332*:176-184.
11. **Dunn, I.S., R. Cowan and P.A. Jennings.** 1988. Improved peptide function from random mutagenesis over 'short windows.' *Protein Eng. 2*:253-291.
12. **Edgington, S.M.** 1993. Shape space. Is biopharmaceutical discovery entering a new evolutionary stage? *Bio/Technology 11*:285-289.
13. **Evnin, L.B., J.R. Vásquez and C.S. Craik.** 1990. Substrate specificity of trypsin investigated by using a genetic selection. *Proc. Natl. Acad. Sci. USA 87*:6659-6663.
14. **Fersht, A. and G. Winter.** 1992. Protein engineering. *Trends Biol. Sci. 17*:292-294.
15. **Glaser, S.M., D.E. Yelton and W.D. Huse.** 1992. Antibody engineering by codon-based mutagenesis in a filamentous phage vector system. *J. Immunol. 149*:3903-3913.
16. **Graham, L.D., K.D. Haggett, P.A. Jennings, D.S. Le Brocque and R.G. Whittaker.** 1993. Random mutagenesis of the substrate-binding site of a serine protease can generate enzymes with increased activities and altered primary specificities. *Biochemistry 32*:6250-6258.
17. **Gregoret, L.M. and R.T. Sauer.** 1993. Additivity of mutant effects assessed by binomial mutagenesis. *Proc. Natl. Acad. Sci. USA 90*:4246-4250.
18. **Henderson, G.B., N.J. Murgolo, J. Kuriyan, K. Osapay, D. Kominos, A. Berry, N.S. Scrutton, N.W. Hinchliffe and A. Cerami.** 1991. Engineering the substrate specificity of glutathione reductase toward that of trypanothione reduction. *Proc. Natl. Acad. Sci. USA 88*:8760-8763.
19. **Hermes, J.D., S.M. Parekh, S.C. Blacklow, H. Köster and J.R. Knowles.** 1989. A reliable method for random mutagenesis: the generation of mutant libraries using spiked oligodeoxyribonucleotide primers. *Gene 84*:143-151.
20. **Holm, L., A. Koivula, P. Lehtovaara, A. Hemminki and J. Knowles.** 1990. Random mutagenesis used to probe the structure and function of Bacillus Stearothermophilus alpha-amylase. *Protein Eng. 3*:181-191.
21. **Houft van Huijsduijnen, R.A.M., G. Ayala and J.F. DeLamarter.** 1992. A means to reduce the complexity of oligonucleotides encoding degenerate peptides. *Nucleic Acids Res. 20*:919.
22. **Hoogenboom, H.R. and G. Winter.** 1992. Bypassing immunization: human antibodies from synthetic repertoires of germ line VH-gene segments rearranged in-vitro. *J. Mol. Biol. 227*:381-388.
23. **Hutchison, C.A., III, R. Swanstrom and D.D. Loeb.** 1991. Complete mutagenesis of protein coding domains. p. 356-389. *In* J.J. Langone (Ed.), Methods in Enzymology. Molecular Design and Modeling: Concepts and Applications. Part A: Proteins, Peptides, and Enzymes. Academic Press, San Diego.
24. **LaBean, T.H. and S.A. Kauffman.** 1993. Design of synthetic gene libraries encoding random sequence proteins with desired ensemble characteristics. *Protein Sci. 2*:1249-1254.
25. **Meng, M., C. Lee, M. Bagdasarian and J.G. Zeikus.** 1991. Switching substrate preference of thermophilic xylose isomerase from D-xylose to D-glucose by redesigning the substrate binding pocket. *Proc. Natl. Acad. Sci. USA 88*:4015-4019.
26. **Merino, E., J. Osuna, F. Bolívar and X. Soberón.** 1992. A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC-M13 vectors. *BioTechniques 12*:508-510.
27. **Olesen, K. and M.C. Kielland-Brandt.** 1993. Altering substrate preference of carboxypeptidase Y by a novel strategy of mutagenesis eliminating wild type background. *Protein Eng. 6*:409-415.
28. **Osuna, J., H. Flores and X. Soberón.** 1991. Combinatorial mutagenesis of three major groove-contacting residues of EcoRI: single and double amino acid replacements retaining methyltransferase-sensitive activities. *Gene 106*:7-12.
29. **Reidhaar-Olson, J.F. and R.T. Sauer.** 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science 241*:53-57.
30. **Scott, J.K. and G.P. Smith.** 1990. Searching for peptide ligands with an epitope library. *Science 249*:386-390.
31. **Shortle, D. and J. Sondek.** 1992. A general strategy for random insertion and substitution mutagenesis: Substoichiometric coupling of trinucleotide phosphoroamidites. *Proc. Natl. Acad. Sci. USA 89*:3581-3585.
32. **Sippl, M. and S. Weitckus.** 1992. Detection of native-like models for amino acid sequences of unknown three-dimensional structure in a data base of known protein conformations. *Proteins: Structure, Function, Genet. 13*:255-271.
33. **Sirotkin, K.** 1986. Advantages to mutagenesis techniques generating populations containing the complete spectrum of single codon changes. *J. Theor. Biol. 123*:261-279.
34. **Smith, G.P.** 1991. Surface presentation of protein epitopes using bacteriophage expression systems. *Curr. Opin. Biot. 2*:668-673.
35. **Stemmer, W.P.C.** DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci. USA* (In press)
36. **Weisberg, E.P., R. Giorda and M. Trucco.** 1993. Simultaneous mutagenesis of multiple sites:application of the ligase chain reaction using PCR products instead of oligonucleotides. *BioTechniques 15*:1:172-181.
37. **Wilks, H.M., K.W. Hart, R. Feeney, C.R. Dunn, H. Muirhead, W.N. Chia, D.A. Barstow, T. Atkinson, A.R. Clarke and J.J. Holbrook.** 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. *Science 242*:1541-1544.
38. **Wilks, H.M. and J.J. Holbrook.** 1991. Alteration of enzyme specificity and catalysis by protein engineering. *Curr. Opin. Biot. 2*:561-567.
39. **Zoller, M.J.** 1991. New molecular biology methods for protein engineering. *Curr. Opin. Biot. 2*:526-531.

Received 18 March 1994; accepted 8 September 1994.

Address correspondence to:
Xavier Soberón
Instituto de Biotecnología/UNAM
Apdo. Postal 510-3 Col. Miraval 62071
Cuernavaca, Morelos, Mexico
Internet: soberon@pbr322.ceinpebi.unam.mx

One obvious way to improve mutagenesis would be to develop methods for "spiking" oligonucleotides at the codon level (15,33,39): some steps in this direction have been taken (28,31). We are currently working on synthetic DNA approaches that accommodate this method while maintaining compatibility with existing chemistries and automated synthesis machines. Another way to achieve a similar result is in performing the oligonucleotide synthesis with removal and remixing of the solid support: a fraction of the resin is taken out, reacted for three cycles to achieve a 100% NNN or NNG/C regime, while the remainder is separately reacted with bases comprising the wild-type codon: the two fractions are remixed and the procedure repeated for as many codons as necessary. We, as well as others (7,15,21), have successfully used this approach. Unfortunately, the method is too cumbersome for large mutagenesis windows or low mutagenesis rates.

In addition to the previous improvements, a desirable goal would be the combinatorial mutagenesis of windows of amino acid residues that are close in three-dimensional space, but not contiguous in the sequence (e.g., residues clustered around the active site, a hydrophobic cluster or a protein-protein interface). Recent advances in mutagenesis methods, especially those based on the PCR, have made such a goal feasible (16,26,27,35,36).

As we stand today, it is very hard to formulate precise predictions about the structural and functional consequences of residue substitutions in proteins (14). Rather, it seems that a combination of crude predictions and hindsight could be applied, together with an empirical approach to obtain new or improved functions from existing proteins. This is especially true for proteins whose three-dimensional structure is not known or is only approximately inferred, as it is in the case of models derived from the recently introduced fold-recognition algorithms (2,32). In another exciting development, mutagenesis together with the utilization of the bacteriophage display of the mutant protein libraries (34) allows the screening of vast numbers of protein variants for specific binding attributes. Clearly, the most efficient mutagenesis methods will constitute an important component of the battery of methods required for the artificial evolution of proteins.

PROGRAM DISTRIBUTION

A PC-compatible program, implemented in Visual BASIC, can be obtained by contacting the last author or directly downloaded from BioTechNet®. The program requires Microsoft Windows™ Version 3.1.

Trabajo sometido a publicación 1

"On the role of a conserved Tyrosine in alpha-amylases"
Gabriel del Río, James M. Briggs, Alejandro Garciarrubio, Gloria Saab and Xavier Soberón.

Sometido a la revista Bichemical Journal

Abstract.

A prevalent idea is that the reaction mechanism of alpha-amylases is similar to that of lysozyme, in which two carboxylates are esential to catalyze the hydrolysis of a glycosidic bond. These two carboxylates and several other residues are conserved in alpha-amylases, and the function of many of them has not been precisely determined yet. That is the case of Tyr82, Asp117, Arg204 and His297 (taka-amylase numbering), for which a role in catalysis has been previously proposed (Biochem. Biophys. Res. Com. (1994), 204:297-302). Due to the relevance of these enzymes in basic science and industrial applications, we conducted a study aimed at determining the possible role of one of these residues. We mutated the residue homologous to Tyr82 in the alpha-amylase from *Bacillus stearothermophilus* (Tyr63) and biochemically characterized four mutants. In order to rationalize the effect of these mutations, we carried out a comparative sequence analysis of several alpha-amylases and Cyclodextrin Glycosiltransferases, and performed electrostatic calculations utilizing the available crystallographic structures. The sequence and structure-based analyses show that all of the four residues mentioned above are also conserved in Cyclodextrin Glycosiltransferases. Our results support that Tyr82 participates in the chemical reaction of alpha-amylases even though its function is not essential for catalysis. We propose that the residues Tyr82, Arg204 and His296, are engaged in helping to develop an electrostatic environment complementary to the transition state, and should account for the pH profiles observed in alpha-amylases.

Keywords: alpha-amylases, Polymerase Chain Reaction (PCR), transglycosylation, evolution, pKa, electrostatics.

The alpha-amylases (EC 3.2.4.1) are endodepolymerases which catalyze the cleavage of alpha(1-4) glycosidic bonds. The reaction mechanism of these enzymes has been generally accepted to be similar to that of hen egg white lysozyme (here referred as lysozyme), in which two carboxylate groups perform a general acid-base catalysis on the glycosidic bond. For some lysozymes, as for some alpha-amylases, it has been observed that the reaction products retain the initial anomeric configuration. Hence, an SN1 reaction mechanism is invoked. In the initial step of the reaction, it is presumed that the carboxylate groups generate and stabilize a carbocation species. The reaction proceeds when a hydroxyl nucleophile attacks this reaction intermediate which is located in the glycone binding site. If the hydroxyl group comes from a water molecule, a hydrolysis reaction proceeds, but if it comes from an oligosaccharide, the outcome is transglycosylation [1,2] (see Figure 1A).

It has been proposed by Mazur et al [3] that an alternative reaction mechanism could take place in alpha-amylases (hereafter referred to as a proton channel mechanism). Specifically, they analyzed the structures of the porcine pancreatic and fungal amylases. In this scheme, a proton channel is proposed to be involved in the protonation of the cyclic oxygen atom in the sugar ring (instead of only one carboxylate acting as a general acid). That channel initiates at His296 (throughout this paper all residue numbering is referred to taka-amylase unless specified) and follows to Tyr82, Asp117, Arg204 and Asp206 (see Figure 1B). However, more experimental evidence supports the lysozyme mechanism than the proton channel proposal [4,5].

Additionally, in the lysozyme scheme, the proton channel residues have so far not been implicated in the reaction mechanism. Mutations in the homologous possition of Tyr82 in the alpha-amylase from *Saccharomycopsis fibuligera* [6], have been reported which altered the hydrolysis/transglycosylation ratio of the enzyme. The authors suggested an explanation, based on the lysozyme reaction mechanism, that proposes a geometry perturbation on binding of the oli-

gosaccharide. Such an event should diminish occupancy of a water molecule which is proposed to be the nucleophile. We decided to study the role of this residue (Tyr82) in the context of the alpha-amylase from *Bacillus stearothermophilus* (ABST). Particularly, we wanted to test the effect on the hydrolysis/transglycosylation ratio in a homologous amylase and the electrostatic role of this position (i.e. the proton channel) as well. In order to rationalize the effect of the mutations, we performed electrostatic calculations on porcine pancreatic amylase and taka-amylase. Additionally, aligning 51 alpha-amylases and 26 Cyclodextrin Glycosiltransferases (CGTases), we look for the conservation of these residues in CGTases, that have been proposed to operate with the same cataltytic machinery than alpha-amylases.

METHODS

Sequence and Structure-Based Analysis.

A structure-based alignment was obtained from the FSSP+HSSP database [7] using Protein Database entries 1amg, 1amy, 2aaa, 1smd, 1ppi, 1bpl, 1cdg, 1ciu, and 1cyg. The structural alignment was kindly provided by Dr. Lissa Holm. This was used to align the homologous sequences provided by the HSSP database to all nine structures. This alignment included 51 alpha-amylase sequences, including those with known structures from Pseudomonas (1amg), wheat (1amy), Aspergillus (2aaa), human salivary (1smd), porcine pancreatic (1ppi) and Bacillar (1bpl). It was also included the CGTase sequences with known structures from *B. circulans* (1cdg), *B. thermoanaerobacterium* (1ciu) and *B. stearothermophilus* (1cyg). The alignment was manually edited around the tyrosine of interest (Tyr82 in taka-amylase), in agreement with structural similarities observed in the structural alignment (see Figure 2). Structures were visualized using the InsightII version 2.3 program suite from Biosym Molecular Simulations.

Cloning and Site Directed Mutagenesis of alpha-amylase.

The alpha-amylase gene of Bacillus stearothermophilus ATCC 12980 (ABST) was isolated by PCR amplification from genomic DNA with the following oligonucleotides:

1. 5′-CGG-GCG-AGA-TCT-AGA-AGG-AGT-TAA-ATA-TTA-TGC-TAA-CGT-TTC-ACC-GC-3′

2. 5′-CCC-CGG-AAG-CTT-GGA-TGG-GCG-CCT-TGT-G-3′

The strain was obtained from the American Type Culture Collection (ATCC) directly. The oligonucleotide sequences were based on the upstream (oligonucleotide 1) and downstream (oligonucleotide 2) regions of the ABST sequence previously reported [8]. The PCR product obtained was cloned into the plasmid pGBS18- [9] and sequenced. The amylase gene was expressed into Escherichia coli JM101. A saturation NNS mutagenesis scheme was carried out on the TAT codon, corresponding to the Tyr63 amino acid residue with the following oligonucleotide:

3. 5′-CCT-TTT-TTG-ATT-GAA-CTC-ACC-GAG-GTC-SNN-CAA-GTC-GTA-T-3′

where S= Guanine and Cytosine, N= All four nucleotides. The PCR reaction was conducted as previously described [10].

Protein Purification and Kinetic Characterization.

The wild-type and mutant enzymes were purified following a protocol previously described [11]. All proteins were purified to more than 90 % homogeneity as determined by SDS-PAGE.

Enzyme activities were determined with soluble starch (Sigma Aldrich, USA) according to the 3,5-dinitrosalicylic acid method [12]. Activities were always measured at 0, 3, 6, 9 and 12 min and initial velocities calculated in dupli-

cate. The reactions were conducted at 70°C. For Km calculations, at least 8 substrate concentrations were used (0.5, 1, 3, 5, 7, 10, 15, 20 mg/ml of soluble starch). For the wild-type and mutant enzymes, the initial velocities were determined at pH 6 in a 50 mM Imidazol buffer solution. In every case, 1 mM $CaCl_2$ was added to the reaction solution. The data were fitted to the Michaelis-Menten equation by means of the Kaleidagraph program (Abelbeck Software) in a Macintosh computer. For kcat determination, protein concentration was determined using the Bradford method (BioRad Cia.), and initial velocities were determined at different pH values (see below) at saturating substrate concentration (10 mg/ml).

pH profiles were determined at pH 5.0, 5.4, 5.8, 6.2, 6.6 7.0, 7.4, 7.8 and 8.2, in 50 mM Mes + 50mM Mops buffer. A saturating substrate concentration was also used for this analysis (10 mg/ml). The profiles were adjusted to the kinetic model previously described for porcine pancreatic amylase [13], using the Kaleidagraph program (Abelbeck Software) in a Macintosh computer. Km with this buffer system was determined as mentioned above, at pH values of 5.0, 5.2, 7.2 and 8.2.

Thermal stabilities for the wild-type and the Tyr63Ala and Tyr63Phe mutants were assesed as follows: 0.2 ml of an enzyme solution with 0.04 units of activity (mg of glucose produced per minute per ml), were placed in a 0.5 ml Eppendorff tube and heated in a PCR oven (Perkin Elmer) at 90°C for 5 min.. Aliquots of 0.025 ml were taken every minute and placed on ice for 1 hour. After this incubation time, initial velocities were determined in 50 mM Imidazol buffer, with 1 mM CaCl2, at pH 6 for the wild-type, or pH 7 for the mutants. Substrate concentration was 10 mg/mL.

Maltoheptaose Hydrolysis Products.

The course of hydrolysis of maltoheptaose (Sigma, USA) was analyzed on HPLC in a μBondapack C18 carbohydrate column. The reactions were conducted with 0.04 units of activity at pH 6 and 7. The reactions times analyzed were 0, 1, 3, 5, 10, 15, 60, 180 min.

Electrostatic Calculations.

The electrostatic calculations were conducted on the porcine pancreatic amylase (1ppi PDB entry) and taka-amylase (6taa PDB entry) structures using the UHBD program developed by McCammon and col. [14-16] version 5.1. The calculations were performed on a Power series Silicon Graphics workstation, and were conducted with the following conditions: protein dielectric constant of 2, solvent dielectric constant of 80, ionic strength of 150 mM and ion exclusion radius (Stern layer) of 20 nm. In order to set the protonation state of the Histidines in the molecule, the environment of each nitrogen atom was analyzed: a NE2 atom was considered to have a hydrogen if its nearest surroundings were constituted by deprotonated chemical groups. In the case of His296, the two possible neutral forms (HNE2 or HND1) were used for comparison. The apparent pka values were determined using the algorithm described previously [17] using the output of the UHBD program.

The complex of taka-amylase and maltodecaose was modelled based on the porcine pancreatic amylase complexed with its inhibitor. After the three carboxylic acids essential for catalysis for each amylase were superimposed, the 5-ring molecule of the inhibitor was superimposed to a crystallographic structure determined for maltodecaose (kindly provided by Dr. P.C.E. Moody). This complex is a model of taka-amylase with its natural substrate. The complex was then minimized without water molecules using the CVFF forcefield and the Discover program from Biosym Molecular Symulations.

To estimate the electrostatic interactions between each pair of polar residues in every molecule studied, we used the UHBD program [14-16], with the parameters described above. For instance, taka-amylase has 125 polar residues, and then every polar residue has 124 electrostatic interactions which can be evaluated. The ionic state of a polar residue is a function of these electrostic interactions. The 124 values calculated are then sorted in increasing order, and then the most influencing residues affecting the ionic state of a residue can be determined. The results on the influence on the ionic state of the three essential catalytic residues, in the alpha-amylases studied, are presented in Table 4.

RESULTS.

Sequence and Structure-Based Analysis.

In order to establish the degree of conservation and possible roles of Tyr82 and other residues among the alpha-amylases, we performed multiple alignments and located conserved residues in the taka-amylase structure. The original alignment was kindly provided by Dr. Lissa Holm, and included alpha-amylases and CGTases (see Methods). Sequences for which a structure was availale (including both alpha-amylases and CGTases) were selected and manually edited (see Methods and Figure 2). 11 invariant residues were found in the alignment (hereafter indicated in uppercase), which are included in the four known conserved regions in amylases [18]. Additionally, 12 other positions hold conservative changes (hereafter indicated in lowercase; see Table 1). All of the 11 invariant residues are in the inmediate vecinity of the active site in taka-amylase (see Figure 3), and for most of them a catalytic role has been previously proposed (see Table 1). For instance, H296, D117, R204 and Y82 belong to the proposed proton channel [3]. In short, 5 out of 23 invariant and semi-conserved residues detected in our alignment are in or around the secondary

structures that connect the Tyr82 containing loop, which is the longest loop in the amylases analyzed here.

Site-Directed Mutagenesis of Tyr82.

The homologous position of Tyr82 in the ABST gene was subjected to saturation mutagenesis. Four variants were chosen for analysis and included: a removal of the hydroxyl group (Tyr82Phe), a change in charge (Tyr82Arg), and two small side chain substitutions (Tyr82Ala, Tyr82Ser) (see Table 2) .

Kinetic Characterization.

The Michaelis constants of the wild-type and mutant proteins determined with soluble starch as a substrate and at pH 7 are shown in Table 2. We found that, relative to the wild type enzyme, Km was unaffected in the mutants. However kcat is higher in all four mutant enzymes at pH 7, with smaller differences at pH 6.

To study the effect of pH on the activity, we chose two mutants ,Tyr63Ala and Tyr63Phe, which eliminate the hidroxyl group or the phenol group from the original Tyrosine, respectively (see Figure 4). In the pH range studied none of the enzymes show pH instability, nor the Km values were affected (data not shown, see Methods). Two different pH profiles can be observed, a bell shaped form for the wild type enzyme and the Tyr82Phe mutation, and one broadened form for the Tyr82Ala mutation. Our results fit well with a kinetic model with three ionization states of the enzyme (see figure 4), as has been reported for the porcine pancreatic amylase [13], but not with a two ionization states model for any of the mutants or wild-type enzyme analyzed. Hence, three pKa values were estimated (see Methods) from our data and are presented in Table 3. As can be observed, the Tyr82Ala mutation altered the pKa1 and pKa3 values

relative to the wild type enzyme, while the other mutation did not modify any of the estimated pKa values.

The hydrolysis of maltoheptaose by the wild type and all of the mutant enzymes show that neither transglycosylation products were generated during the processing of the maltoheptaose, nor significant differences were observed in the type of maltooligosaccharides generated during the reaction (data not shown). This last observation is consistent with the unaltered Km value observed for the mutants analyzed.

Tyr82Phe and Tyr82Ala were investigated for thermal stability. They were chosen because we wanted to determine the effect of eliminating the hydroxyl group from the tyrosine residue, and the effect on the reduction of the side chain size. As can be seen in Figure 5, removal of the hydroxyl group (Tyr82Phe mutant) stabilizes the protein slightly, while elimination of the aromatic side chain and the hydroxyl group (Tyr82Ala mutant) does not cause a significant alteration in the half lives of the enzymes at 90°C.

Electrostatic Calculations.

Based in our results, we decided to conduct electrostatic calculations which could help raticnalize the effects observed by the mutations. Additionally, we looked for a possible explanation for the conservation of Tyr82, Asp117, Arg204 and His297. We used the structures of porcine pancreatic and taka amylases for our study. These calculations were conducted with both enzymes studied, and with the complex in one of them (see Methods).

The conservation of Tyr82 suggests that it may have a common function throughout the alpha-amylases and CGTases. According to our results such function should concern the catalytic process (kcat and pH profile alteration), but not binding (Km and hydrolysis products unaffected). Since the experimental data

indicate that the pH profile of one of the mutants is altered, we mainly focused on determining the electrostatic influence exerted by the residues in the proposed proton channel, over those directly involved in catalysis (Asp206, Glu230 and Asp297).

The two possible neutral forms of His296 in taka-amylase were considered for the calculations. Since no significant differences in the pKa values estimated were founded (data not shown), we decided to consider here one neutral form, the one with the ND1 atom protonated.

In order to establish a criterion to help us validate our calculations, we investigated the protonation states of residues E109, K221 and D391 for taka-amylase. The side chains of these residues are completely exposed to the solvent. On the other hand, residues E294 and R300, are ionizable groups facing each other and buried in the interior of the protein. For both groups of residues, the protonation states calculated are in agreement with the values expected: The calculated pKa values for the solvent exposed residues were quite similar to the value determined in solution for these residues. For the buried residues, a pKa shift is observed of the same magnitud as has been previously found for buried residues (data not shown) [19].

From these calculations, shown in Table 4, the strength of the electrostatic interaction was estimated between the three carboxylic acids and residues in the active site crevice. The polar residues which mostly affect the ionization states of the three catalytic carboxylic residues considered are, Arg204 > His296 > Tyr82. For the complex of taka-amylase with its natural substrate, the order is Arg204 > Tyr82 > His296.

DISCUSSION.

Our sequence analysis shows that the residues of the proposed proton chanel are conserved both in alpha-amylases and CGTases, and that they are surrounded by semi-conserved residues, which highligth their relevance. Mutations in the homologous position to Tyr82 in the alpha-amylase from *B. stearothermophilus*, reduce the biological expression of the mutants (data not shown), and alter their half lives at 90°C (see Figure 5). Additionally, the mutants in that position alter kcat and the pH profile (see Table 2 and 3, and Figure 4). A negative influence in stability by residues near the catalytic site, such as Tyr82, has been previously observed for other enzymes [20]. Interestingly, mutant Tyr82Phe showed both higher activity and longer half life at 90°C in ABST, a behaviour predicted for catalytic mutants, but not previously observed [20].

The pH activity shows that three ionizable groups could account for the pH profiles observed. According with the values estimated and previous reports for the porcine pancreatic alpha-amylase [13], two carboxylates and one Histidine could be the amino acids responsible for the pH profiles observed. Asp206 and Glu230 residues should be the carboxylates, and the Histidine is His296, with pKa2=5.1, pKa1=6.6 and pKa3=7.9, respectively, for the wild-type enzyme. According with our calculations (see Table 4), the ionization state of Glu230 and Asp206 is mainly affected by Arg204 and His296, and the ionization state of these last residues is mainly affected by Tyr82 (data not shown). This last residue forms a hydrogen bond with His296, that in Tyr82Ala mutation would be lost, thus increasing the pKa value of His296 (see Table 3). On the other hand, Tyr82 does not form a hydrogen bond with neither Asp206 nor Glu230, so the Tyr82Ala mutation that altered the pKa value of Glu230 might be the result of indirect influence, maybe troughout Arg204 residue, as indicated by our electrostic calculations.

Our finding that changes in Tyr82 in ABST affect neither Km nor hydrolysis/tranglycosylation ratio, is not in agreement with the results observed for

mutations in the homologous position in the alpha-amylase from Saccharomycopsis fibuligera [6]. Then, it seems that Tyr82 is conserved for reasons not directly related to the affinity for the substrate. Instead, this residue and the other ones involved in the proposed proton channel, are invariant and important, but not essential, for the chemical reaction performed by alpha-amylases [3]. So, what is the function of these residues and then, why are they conserved in the alpha-amylases? The simplest explanation would be that due to the ionic nature of these residues, they would have an electrostic role in catalysis, that in the context of the lysozyme mechanism, this would be to assist in developing a complementary charge to the transition state. Consistent with this idea are the effects on kcat and pH profile observed in mutations reported, in the proposed proton channel here, and in previous works [11,18]. Mutations in these positions on CGTases have not been reported, but it would be expected to present a similar behaviour observed for mutations reported for alpha-amylases, based on the sequence conservation of these residues.

In conclusion, we observed that Tyr82 influence the ionization state of Glu230 and His296. We propose that the invariant residues Tyr82, Arg204 and His296, are engaged in helping to develop a complementary charge to the transition state. We also observed that in agreement with a relationship previously observed for catalytic relevant residues [20], mutations in Tyr82 alter the stability of the enzyme (half life at 90°C).

Acknowledgments

We are greatful to Filiberto Sánchez, Fernando González, Paul Gaytán and Eugenio López for technical support, and Alma Tremari for secretarial help. We thank computer facilities provided by the Physics Institute/UNAM. This work was funded in part by PAPIIT project number IN207293 from DGAPA/UNAM.

Bibliography

1. Sinnott, M.L. Catalytic mechanisms of enzymic glycosyl transfer. *Chem.Rev.* 90:1171-1202, 1990.

2. Koshland, Jr.,D.E. Stereochemistry and the mechanism of enzymatic reactions. In: , edited by Brookhaven National Laboratory, N.Y.,U.S.A. N.Y., U.S.A.: Brookhaven National Laboratory, Upton, 1953, p. 416-436.

3. Mazur, A.K., Haser, R. and Payan, F. The catalytic mechanism of alpha-amilases bases upon enzyme crystal structures and model building calculations. *Proc.Natl.Acad.Sci.USA* 204:297-302, 1994.

4. Qian, M., Haser, R., Buisson, G., Duée, E. and Payan, F. The active center of a mammalian alpha-amylase. Structure of the complex of a pancreatic alpha-amylase with a carbohydrate inhibitor refined to 2.2-Å resolution. *Biochemistry* 33:6284-6294, 1994.

5. Machius, M., Vértesy, L., Huber, R. and Wiegand, G. Carbohydrate and protein-based inhibitors of porcine pancreatic alpha-amylase: structure analysis and comparison of their binding characteristics. *J.Mol.Biol.* 260:409-421, 1996.

6. Matsui, I., Yoneda, S., Ishikawa, K., et al. Roles of the aromatic residues conserved in the active center of Saccharomycopsis alpha-amylase for transglycosylation and hydrolysis activity. *Biochemistry* 33:451-458, 1994.

7. Holm, L. and Sander, Ch. The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.* 22:3600-3609, 1994.

8. Suominen, I., Karp, M., Lautamo, J., Knowles, J.K.C. and Mantsaelae, P. Thermostable alpha amylase of Bacillus stearothermophilus: cloning, expression, and secretion by Escherichia coli. In: *Extracellular enzymes of microor-*

ganisms, edited by Chaloupka, J. and Krumphanzl, V. New York: Plenum Press, 1987.

9. del Río, G., Rodríguez, M.A., Munguía, M.A., López-Munguía, A. and Soberón, X. Mutant *E. coli* penicillin acylase with enhanced stability at alkaline pH. *Biotechnol.Bioeng.* 48:141-148, 1995.

10. Merino, E., Osuna, J., Bolívar, F. and Soberón, X. A general, PCR-based method for single or combinatorial oligonucleotide-directed mutagenesis on pUC-M13 vectors. *Biotechniques* 12(4):508-510, 1992.

11. Vihinen, M., Ollikka, P., Niskanen, J., et al. Site-directed mutagenesis of a thermostable alpha-amylase from *Bacillus stearothermophilus*: putative role of three conserved residues. *J.Biochem.* 107:267-272, 1990.

12. Summer, J.B. and Howell, S.F. A method for determination of invertase activity. *J.Biol.Chem.* 108:51-54, 1935.

13. Ishikawa, K., Matsui, I. and Honda, K. Substrate-dependent shift of optimum pH in porcine pancreatic α-amylase-catalyzed reactions. *Biochemistry* 29:7119-7123, 1990.

14. Antosiewicz, J., McCammon, J.A. and Gilson, M.K. Prediction of pH-dependent properties of protein. *J.Mol.Biol.* 238:415-436, 1994.

15. Madura, J.D., Briggs, J.M., Wade, R.C., et al. Electrostatics and diffusion of molecules in solution: simulations with the University of Huston-Brownian Dynamics program. *Comput.Phys.Commun.* 91:57-95, 1995.

16. Davis, M.E., Madura, J.D., Luty, B.A. and McCammon, J.A. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston-Brownian Dynamics program. *Comput.Phys.Commun.* 62:187-197, 1991.

17. Gilson, M.K. Multiple-site titration and molecular modelling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Struct.Fuct.Gen.* 15:266-282, 1993.

18. Svensson, B. Protein engineering in the alpha-amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Molecular Biology* 25:141-157, 1994.

19. Masson, J.M., Mourey, L., Samama, J.P., Swarén, P., Maveyraud, L. and Guillet, V. Electrostatic analysis of TEM1 β-lactamase: effect of substrate binding, steep potential gradients and consequences of site-directed mutations. *Structure* 3:503-613, 1995.

20. Shoichet, B., Baase, W.A., Kuroki, R. and Matthews, B.W. A relationship between protein stability and protein functions. *Proc.Natl.Acad.Sci.USA* 92:452-456, 1995.

Figure Legends

Figure 1.- The reaction mechanism proposed for amylases based on lysozyme (A) and the initial steps in the alternative proton channel mechanism (B). In A, residue numbers correspond to the lysozyme enzyme, while in B to taka-amylase.

Figure 2.- An alignment of representative sequences, showing the conserved Tyrosine. Dots are used to separate each stretch of sequences.

Figure 3.- A ribbon representation of taka-amylase structure showing the 11 invariant (side chains presented) and 16 semi-conserved (identified as white balls) residues.

Figure 4.- The pH profiles for the wild-type (A), Tyr63Phe (B) and Tyr63Ala (C) mutants are presented plotting the activity (mg/mL of glucose by minute) vs. units of pH. Every point in these graphs is the mean of at least three independent experiments. The standard deviation for all the points is less than 5%.

Figure 5.- Effect of preincubation at 90°C of wild-type enzyme (filled squares), Tyr63Ala (empty squares) and Tyr63Phe (empty circles). Every point in this graph is the mean of at least three independent experiments. The standard deviation for all the points is less than 5%.

Residue	Position in structure	Variation	Function
Y82	Connector B2-H2	--	Catalysis*
G56	Connector H1-B2	--	Structural*
P64	B2	--	Structural*
G95	Connector B2-H2	--	Unknown
f13	B1	Ile,Leu	Unknown
w61	B2	Gln,Ile	Unknown
R204	B4	--	Catalysis*
D206	B4	--	Catalysis*
g202	B4	Ala	Unknown
k209	Connector B4-H4	Arg,Thr	Unknown
h210	Connector B4-H4	Gly	Structural*
w242	H5	Tyr	Unknown
H296	Connector B7-H7	--	Catalysis*
D297	Connector B7-H7	--	Catalysis*
f292	B7	Leu	Unknown
D117	B7	--	Unknown
E230	B5	--	Catalysis*
N295	Connector B7-H7	--	Structural*
h122	Connector B3-H3	Ser	Structural*
d175	Connector B3-H3	Asn,Ala	Structural*
g323	Connector H7-B8	Pro	Unknown

Table 1.- The 23 conserved residues (invariant in uppercase and semi-conserved in lowercase) are listed with their corresponding structure positions in the TIM-barrel fold. Those with previous function known [18] are marked with an *. The residues are grouped according with our suggested struttural relationship (see Results and Discussion), including the Y82 group (6 residues), R204-D206 group (6 residues), H296-D297 group (3 residues) and the rest of the residues form the fourth group (6 residues). Hs denote alpha-helices and Bs beta-strands. Catalytic and structural loops are referred to as Connectors. Variations observed in the alignment for the semi-conserved residues are indicated.

	Km (mg/ml)	kcat (1/min) pH 6	kcat (1/min) pH 7
Wild-Type	1.01 ± 0.1	3012 ± 2%	2785 ± 3%
Y63A	1.17 ± 0.2	4624 ± 5%	4874 ± 4%
Y63S	1.50 ± 0.4	4940 ± 7%	4953 ± 6%
Y63F	0.95 ± 0.15	4136 ± 3%	4279 ± 2%
Y63R	1.30 ± 0.2	3910 ± 5%	3989 ± 4%

Table 2.- Kinetic parameters are presented for wild-type and four mutants. kcat values are reported at pH 6 and 7. The standard deviations are shown for Km (in mg/ml) and kcat (in percentage) parameters.

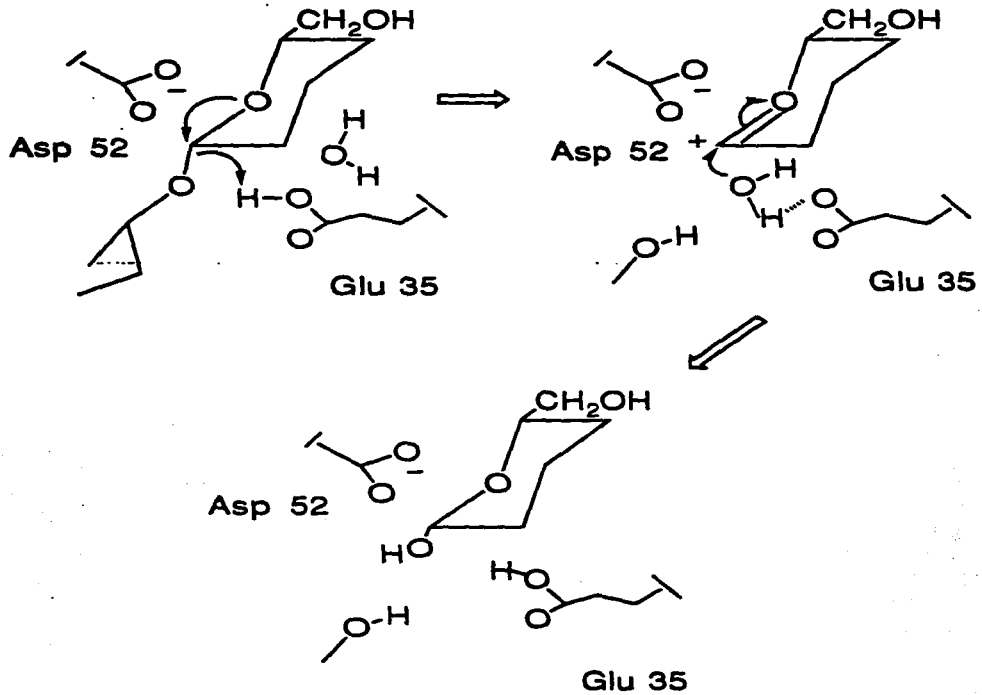
	Wild-Type	Tyr63Phe	Tyr63Ala
pKa1	6.6 ± 0.4	6.8 ± 0.2	5.7 ± 0.04
pKa2	5.1 ± 0.1	5.1 ± 0.07	5.1 ± 0.1
pKa3	7.9 ± 0.7	8.0 ± 0.4	9.0 ± 0.1

Table 3.- pKa value estimation. The three pKa values estimated (see Methods) for the wild-type, Tyr63Phe and Tyr63Ala muatations are presented.

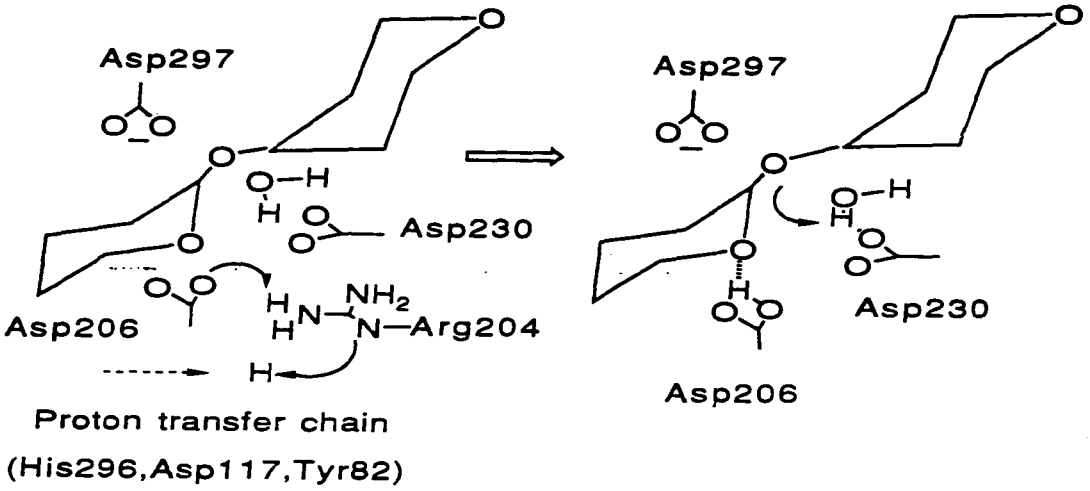
	pKa_app	pKa model
Glu109	4	4.40
Lys221	10	10.40
Asp391	4	4.00
Glu294	0.4	4.40
Arg300	14	12

Table 4.- The pKa (relative acidity) calculated are shown for residues selected to test the calculations. pKa_app is the pH value at which the residue is half-ionized and pKa model refer to the expected pKa value of an amino acid residue in solution.

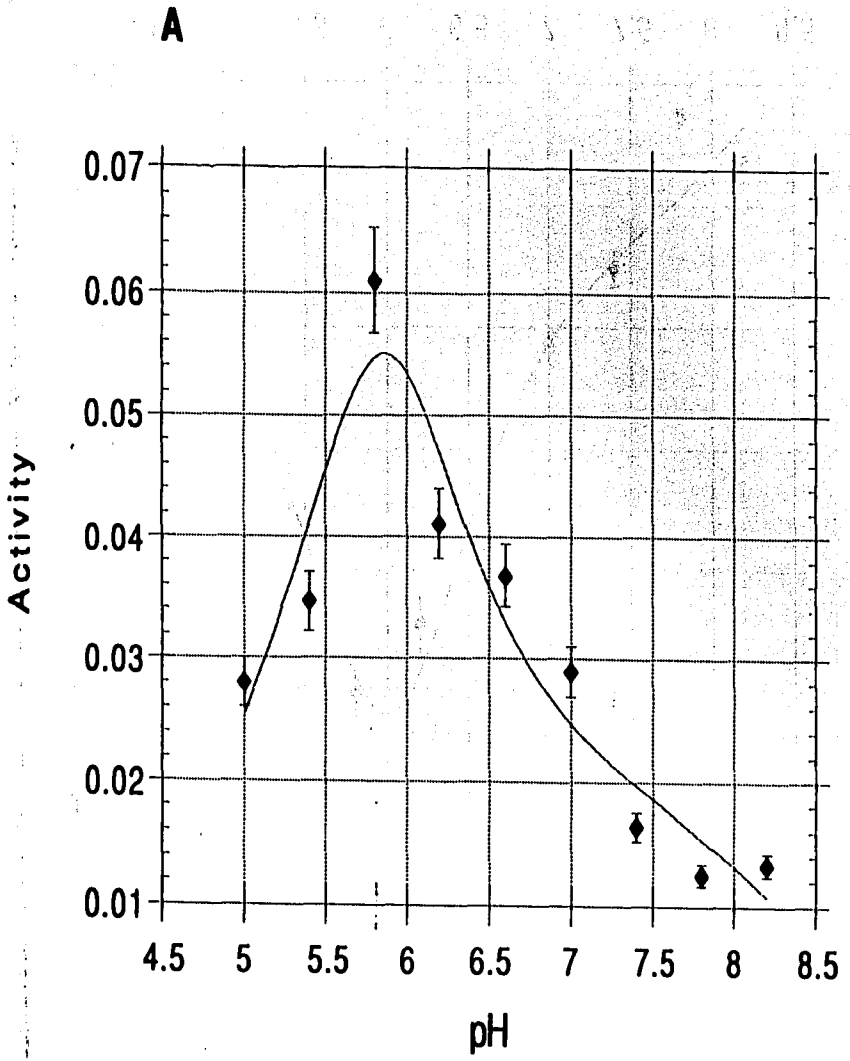
A

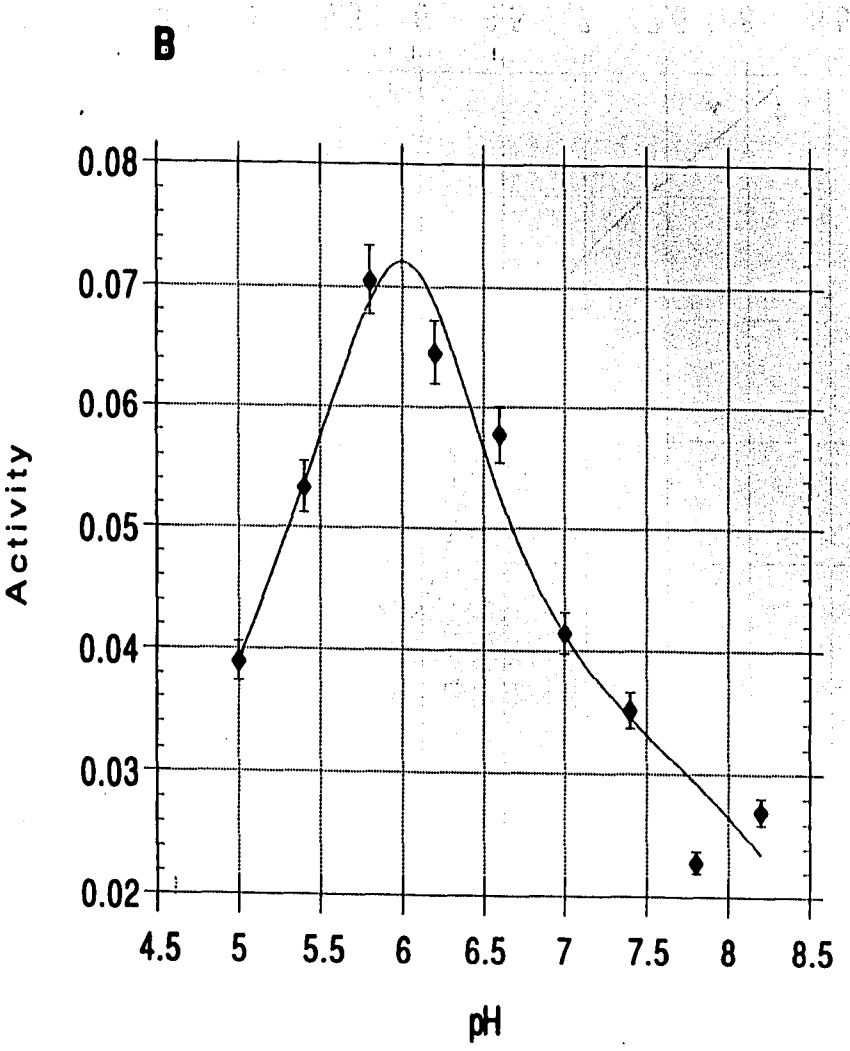


B

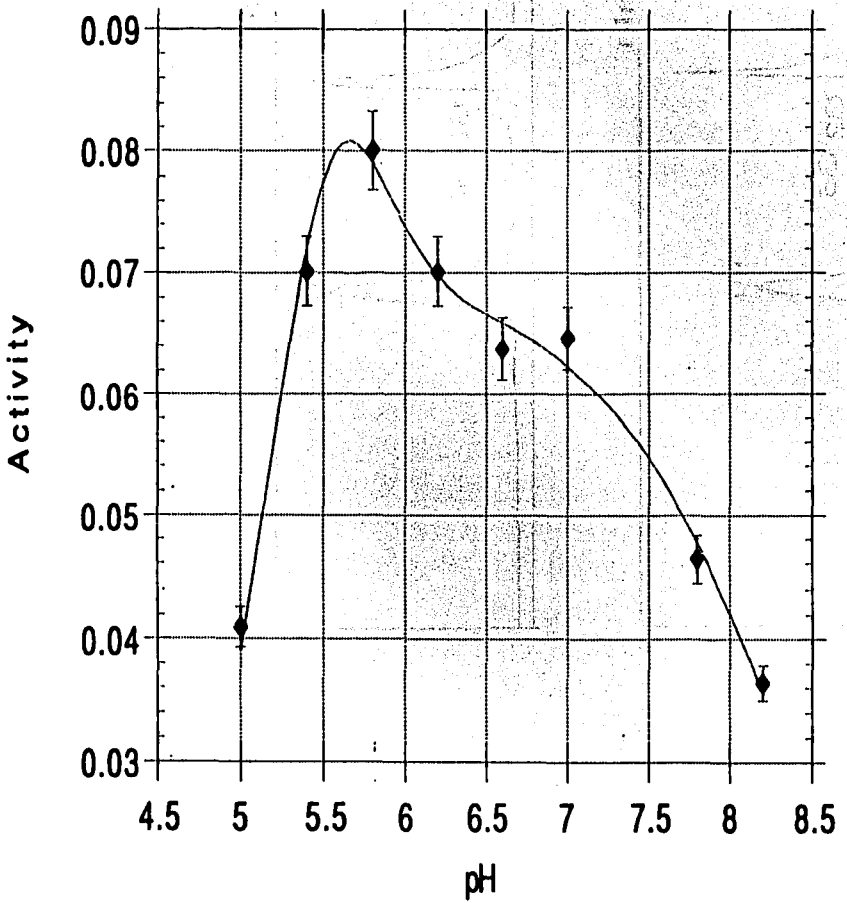

Asp297
Asp230
Asp206
Arg204
Proton transfer chain
(His296,Asp117,Tyr82)
Asp297
Asp230
Asp206

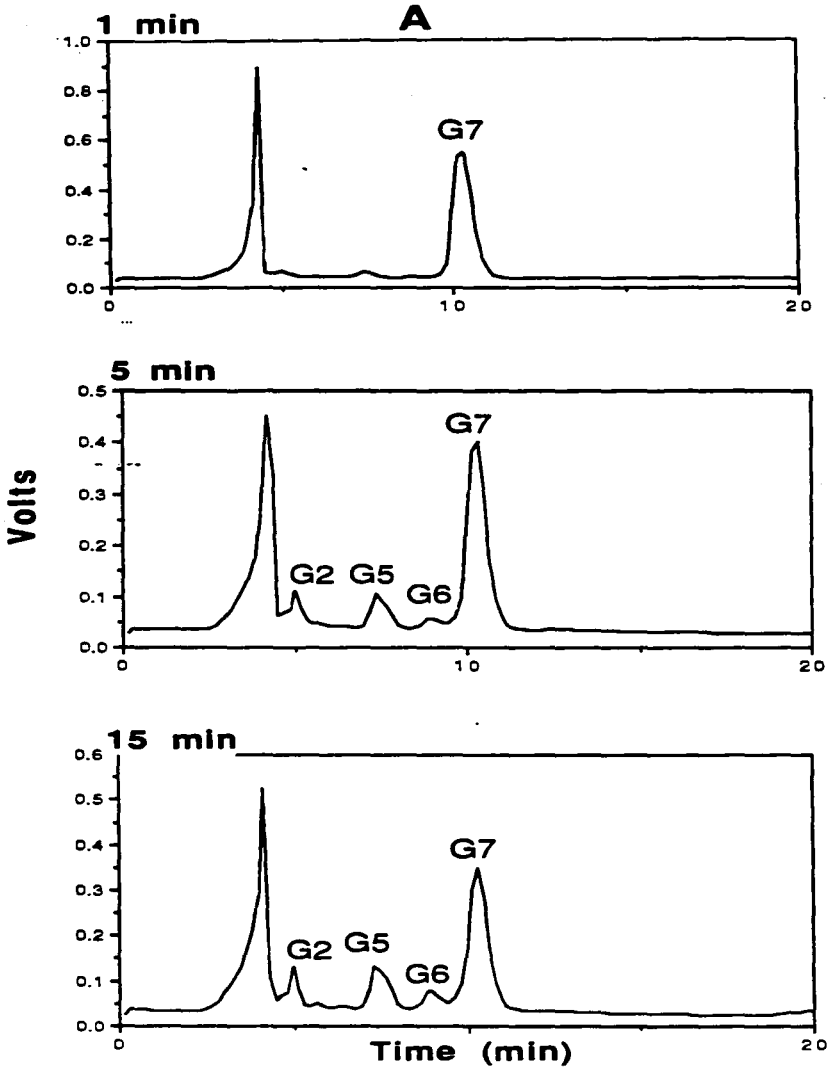
```
Swiss Prot PDB

amys_human 1smd   lapkGfggvqvspPnenvpww.rY  riyvDavinH  gvagfRiDaskh  kpfiyqE  alvfvdNHD

amyp_pig   1ppi   lgpkGfggvqvspPnenvpww.rY  riyvDavinH  gvagfRlDaskh  rpfifqE  alvfvdNHD

amy2_horvu 1amy   iaaaGithvwlp.Pasqsvaeqg Y  kaiaDivinH  gfdgwRfDfakg  psfavaE  avtfvdNHD

amt4_psest 1amg   iaadGfsaiwmpvPwrdfgge.gY  kvlyDvvpnH  gaggfRfDfvrg  nsfcvgE  avtfvdNHD

amy_bacli  1bpl   laehGitavwip.Paykgtadvg Y  nvygDvvinH  .LdgfRlDavkh  emftvaE  avtfvdNHD

amya_aspor 2taa   iqgmGftaiwit.Pvtaqlayhg Y  ylmvDvvanH  sidglRiDtvkh  gvycigE  lgtfveNHD

cdgu_bacci 1cdg   ltgmGvtaiwisqPveni.ayhg Y  kviiDfapnH  gidgiRmDavkh  pvfttfgE  qvtfidNHD

cdgt_bacci 1ciu   fsdlGvtalwisqPveni.ayhg Y  kiviDfapnH  gvdgiRvDavkh  pvfttfgE  qvtfidNHD

cdgt_bacst 1cyg   ltdmGvtaiwisqPvenvs.yhg Y  kviiDfapnH  gidgiRmDavkh  pvfttfgE  qvtfidNHD

                    G    w   P          Y       D   nH  g    R D  kh        E     f  NHD
                   56   61  64         82     117 121 202  204  209       230   292
                                              122        206 210                 295
                                                                                 296
                                                                                 297
```

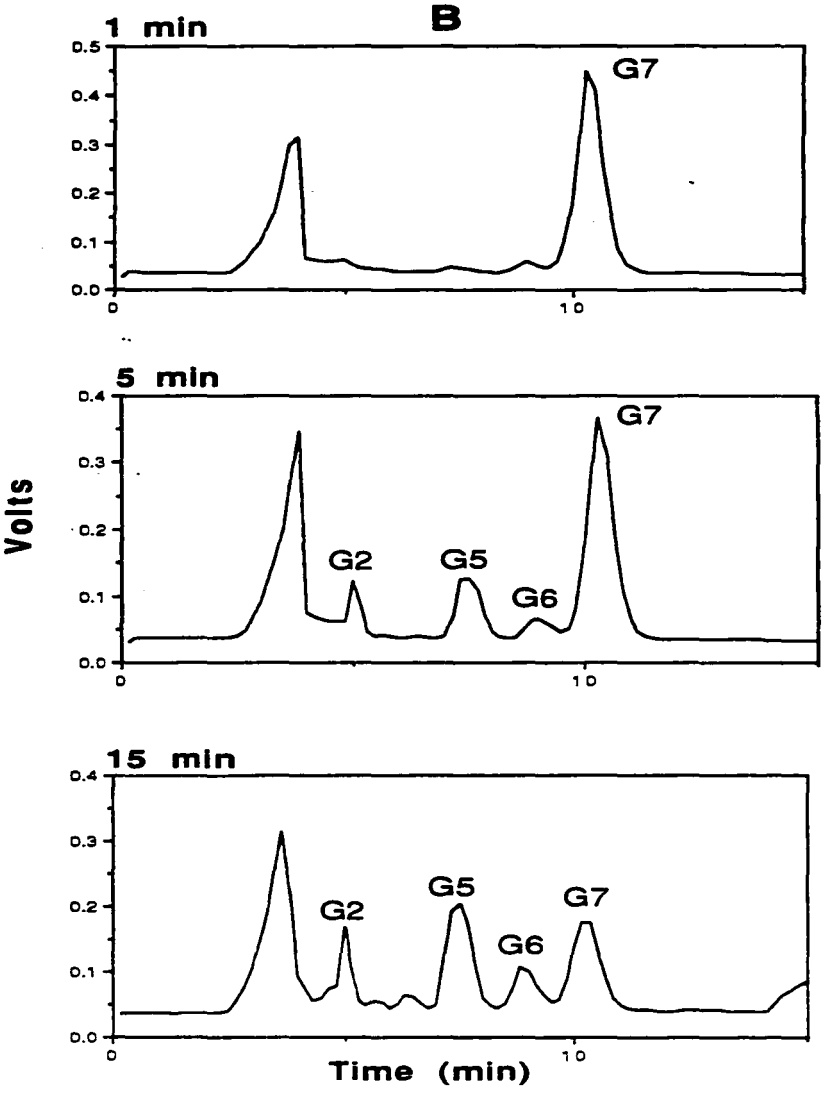

A
0.07
0.06
0.05
0.04
0.03
0.02
0.01
Activity
4.5
5
5.5
6
6.5
7
7.5
8
8.5
pH


B
0.08
0.07
0.06
0.05
0.04
0.03
0.02
Activity
4.5
5
5.5
6
6.5
7
7.5
8
8.5
pH

C




A
1 min
5 min
15 min
G7
G2
G5
G6
Volts
Time (min)


B
1 min
G7
5 min
G7
G2
G5
G6
15 min
G2
G5
G6
G7
Volts
Time (min)

Trabajo sometido a publicación 2

"Did cyclodextrin glycosyltransferases evolved from alpha-amylases?"
Gabriel del Río, Enrique Morett and Xavier Soberón.

Sometido a la revista FEBS Letters

Did Cyclodextrin Glycosyltransferases Evolved from alpha-Amylases?

Gabriel del-Rio, Enrique Morett and Xavier Soberon*
Instituto de Biotecnologia/UNAM
Apdo. Postal 510-3, Cuernavaca, Morelos 62271, Mexico.

Abstract

The hydrolytic enzymes, alpha-amylases, and the cyclodextrin glycosyltransferases (CGTases) are key enzymes in the depolymerization of starch. These two groups of enzymes are evolutionary related. We propose that the transferase activity is likely to have evolved from an ancestral hydrolase. Sequence analysis provided some support to this hypothesis. Consequently, we conducted an experimental study to test the possible adaptive value for evolving a CGTase. We found that when an alpha-amylase and a CGTase are combined more glucose is generated from starch than would be expected from the independent action of any of these enzymes. Thus, we propose that the biological role of CGTases is to work in concert with alpha-amylases for the efficient saccharification of starch. This observation can be useful in industrial processes aimed at producing syrups with high content of glucose or maltose.

Keywords: alpha-amylase, cyclodextrin glycosyltransferase, evolution, divergence.

---

* To whom correspondence should be addressed:
Dr. Xavier Soberon
Departamento de Reconocimiento Molecular y Bioestructura, Instituto de Biotecnologia/UNAM. Apdo. Postal 510-3, Cuernavaca, Morelos 62271, Mexico. Telephone (52)(73)114900. FAX (52)(73)172388.
e-amil: soberon@ibt.unam.mx

1. Introduction

Starch is one of the most abundant biopolymers on earth and it is the main source of energy for a wide variety of living organisms [1]. This molecule is built of units of glucose, linked by either α(1,4) or α(1,6) glycosidic bonds. To obtain glucose from starch different enzymes are used, most of which are grouped in the alpha-amylase family [2] (see Table 1). This group includes enzymes that display specificity for either of the two types of bonds, have either retaining or inverting mechanisms (depending on whether they retain or invert the anomeric structure of the substrate [3]), and whose activities classify them as either hydrolases (EC 3.2.1) or transferases (EC 2.4.1) [4]. By combining these properties, there would be 8 theoretically types of hydrolases and 4 of transferases. However, these numbers must be reduced considering the nature of the reactions involved [3], so only 3 types of hydrolyses and 2 types of transferases would be expected to exist in nature (not including dextranases). All these five activities have been found (see Table 1). Thus, the alpha-amylase family is well suited for studying the way different enzyme activities give rise to each other, particularly the retaining enzymes (alpha-amylases and CGTases), since they are biochemically and structurally well characterized (see below).

Through the analysis of the primary structure of the retaining enzymes, it has been found that although there is limited conservation on their overall primary structure, both hydrolases and transferases share sequence motifs which contain the catalytic residues. Thus, it has been assumed that all of them use the same catalytic machinery [4]. Additionally, the three dimensional structure of 5 alpha-amylases and 3 CGTases (see Methods) has already been determined (see legend in Table 1), showing that they share a $(\beta/\alpha)_8$-barrel domain [5]. Such evidence led different authors to suggest that these enzymes probably diverged from a common ancestor [2,5,6]. Taking advantage that this is one of the few examples of two different,

well characterized, enzyme activities related by the same catalytic machinery and fold [7,8], it is possible to carry out a detailed molecular phylogenetic analysis that can shed some light on the evolution of their activities.

Many successful examples of *in vitro* alteration of enzyme specificity have recently been reported [9-11], but few of a change to a new activity [12,13]. This may reflect the scarcity of our knowledge about this last subject and that is more difficult to evolve a new enzyme activity than a new specificity [11]. In this study we propose an evolutionary model to account for the divergence of two enzyme activities in the alpha-amylase family: hydrolase and transferase. We also propose a possible biological role for CGTases.

2. Methods

Sequence Analysis

A multiple sequence alignment including 49 sequences of alpha-amylases and 22 of CGTases was obtained from the FSSP+HSSP databases (http://www.embl-heidelberg.de/dali/dali.html). This alignment was kindly provided by Dr. Lissa Holm and it is based on 6 structures of alpha-amylases (from fungi (2aaa), bacteria (1amg and 1bpl), mammals (1smd and 1ppi) and plants (1amy)) and 3 of CGTases (from bacillus (1cdg, 1ciu, 1cyg)). Known sequence homologous to these structures (but having less than 25% identity) were included. The SwissProt entry of alpha-amylase sequences employed are: amy2_horvu, amy3_horvu, amy4_horvu, amy5_horvu, amy6_horvu, amy1_horvu (*Hordeum vulgare*); amy1_orysa, am2a_orysa, amc2_orysa, am3d_orysa, am3b_orysa, am3c_orysa, am3e_orysa, am3a_orysa, amc1_orysa (*Oryza sativa*); amy3_wheat (*Triticum aestivum*); amya_vigmu (*Vigna mungo*); amy2_ecoli (*Escherichia coli*); amt4_psest (*Pseudomonas*

stutzeri); amt4_psesa (*Pseudomonas saccharophila*); amyp_pig (*Sus scrofa*); amyc_human, amyp_human, amys_human (*Homo sapiens*); amyp_mouse, amys_mouse (*Mus musculus*); amyp_rat (*Rattus norvegicus*); amya_drome (*Drosophila melanogaster*); amy_trica (*Tribolium castaneum*); amy_altha (*Alteromonas haloplanktis*); amya_aerhy (*Aeromonas hydrophila*); amy_thecu (*Thermonospora curvata*); amy_strtl, amy_strvl (*Streptomyces thermoviolaceus*); amy_strgr (*Streptomyces griceus*); amy_strlm (*Streptomyces limosus*); amy_strhy (*Streptomyces hygroscopicus*); amy_strli (*Streptomyces lividans*); amya_aspor (*Aspergillus oryzae*); amy_aspsh (*Aspergillus shirousami*); amy1_schoc (*Schwanniomyces occidentalis*); amy1_sacfi (*Saccharomycopsis fibuligera*); amy1_schpo (*Schizosaccharomyces pombe*); amyb_bacpo (*Bacillus polymixa*); amy_bacli (*Bacillus licheniformis*); amy_bacam (*Bacillus amyloliquefaciens*); amt6_bacs7 (*Bacillus sp.* strain 707); amy_bacst (*Bacillus stearothermophilus*); amy2_salty (*Salmonella typhimurium*). For CGTases: cdgu_bacci,amy_bacci, cdgt_bacci (*Bacillus circulans*); amyr_bacs8 (*Bacillus sp.* strain B1018); cdgt_bacsp (*Bacillus sp.* strain 17-1); cdgt_bacs0 (*Bacillus sp.* strain 1011); cdgt_bacs3 (*Bacillus sp.* strain 38-2); cdgt_bacli (*Bacillus licheniformis*); cdgt_bacss (*Bacillus sp.* strain 6.6.3); amy_thetu (*Thermoanaerobacter thermosulfurogenes*); cdg2_bacma, cdg1_bacma (*Bacillus macerans*); cdgt_bacst, amym_bacst (*Bacillus stearothermophilus*); cdgt_bacoh (*Bacillus ohbensis*); cdgt_bacs2 (*Bacillus sp.* strain 1-1); cdgt_klepn (*Klebsiella pneumoniae*).

The alignment was manually edited in regions where ambiguity was observed. In order to include the neopcllulanase sequence from *Bacillus stearothermophilus*, we used the alignments provided by two different fold recognition algorithms, TOPITS (http://www.embl-heidelberg.de/predictprotein) and UCLA-DOE (http://www.mbi.ucla.edu/people/frsvr/frsvr.html). This neopcllulanase

was chosen because it presents both specificities and activities found in the alpha-amylase family [14].

Phylogenetic Analysis

Phylogenies were reconstructed using distance-based and parsimony methods. The phylogenetic inference programs employed are part of the Phylip package, version 3.5, for UNIX (Felsentein J. 1993. Department of Genetics, University of Washington, Seattle. Distributed by the author). We used the programs PROTDIST with the PAM250 matrix, and FITCH to produce a tree of genetic distance. Global rearrangements and 10 jumbles were used to minimize the effect of sequence order. To generate the most parsimonious tree we used the program PROTPARS. When several equally parsimonious trees were found, a consensus tree was proposed using the program CONSENSE. Bootstrapping analysis was performed using the SEQBOOT program, generating 1000 permutation resamplings of our sequence alignment. Global rearrangements and 10 jumbles were also performed for each of the 1000 bootstrapped data sets. The above mentioned programs were run in a Silicon Graphics workstation. For phylogenetic tree representations, we used the program TreeView version 1.4 [15] in a Macintosh computer.

Enzyme Activity

The depolymerization of starch was followed by either determination of reducing sugars using the reagent 3,5-Dinitrosalicylic acid [16], or by thin layer chromatography on silica gel aminated plaques, using the solvent system water/ethanol/butanol (2:5:3 by vol.). The Glucose concentration was determined with a kit obtained from Boehringer Mannheim (USA), which includes two enzymes, Hexokinase and Glucose 6-Phosphate dehydrogenase.

Alpha-amylase from *Bacillus licheniformis* was obtained from BIOTECSA S.A. C.V. (Mexico), and the CGTase from *Bacillus macerans* was obtained from AMANO International Enzyme Co., Inc. (USA). Soluble starch used in this study was from Sigma (USA). Protein was determined by the Bradford method (Bio-Rad laboratories, USA). The reactions of starch depolymerization with alpha-amylase and/or CGTase were conducted at pH 6.0 (acetate buffer, 30 mM, $CaCl_2$ 3.0 mM), and at 60°C. Such conditions were optimal for both enzymes. These reactions were carried out for 5 hours, and samples were taken every hour. Since the alpha-amylase and CGTase activities are not easily comparable, we used different enzyme mixtures for our study: alpha-amylase alone, 142 µg/mL; CGTase alone, 70 µg/mL; alpha-amylase(1):CGTase(1) mix, 142 µg/mL:70 µg/mL; alpha-amylase(1):CGTase(2) mix, 142 µg/mL:140 µg/mL; and alpha-amylase(2):CGTase(1) mix, 142 µg/mL:35 µg/mL. The final reaction volume was 10ml, and contained 1.2 mg/mL of soluble starch.

3. Results

Phylogeny

The distance tree generated with all the alpha-amylases and CGTases sequences resulted in two well defined groups comprising each type of enzyme (data not shown), except for three putative alpha-amylases which were included with CGTases: amyr_bacs8, amy_thetu and amy_bacci. Two of them (amyr_bacs8 and amy_thetu) might actually be CGTases since they harbor a signature of CGTases [17], and have sequence lengths similar to CGTases. On the other hand, amy_bacci (528 aa) has the sequence length of alpha-amylases (around 500 aa). This amylase aligns only with the three domains shared by alpha-amylases and CGTases (this includes the $(\beta/\alpha)_8$-barrel domain, B domain, and the C-terminal domain in alpha-amy-

lases [18]). It has been shown, that the two additional domains found in CGTases determine their ability to make cyclodextrins [19,20]. Accordingly, the amylase from *Bacillus circulans* (amy_bacci) does not form cyclodextrins. Hence, we believe that the sequence signature criterium might not be sufficient to distinguish between alpha-amylases from CGTases.

To perform a more exhaustive phylogenetic analysis, we selected as representative members those for which structures were available (see Methods). Based on a published thermodynamic study of enzyme activity evolution [21], which considers the independent evolution of Km and kcat, and applying it to the case of alpha-amylases, we concluded that the ancestor of this group must have been an enzyme with low specificity and low activity. These are characteristics present in neopollulanases. These enzymes, based on other arguments, had already been proposed as the ancestor for all of the activities in the alpha-amylase family [14]. As shown in Figure 1A and 1B, the topologies of phylogenetic trees obtained through genetic distance and parsimony analyses were similar, and with both methods alpha-amylases grouped apart from CGTases. The neopollulanase from *B. stearothermophilus* (neop_bacst) is grouped with alpha-amylases, from which it appears to be a relatively recent derivation. Thus, instead of neopollulanases being the ancestor of both alpha-amylases and CGTases as previously proposed, it seems that neop_bacst comes from an alpha-amylase. On the other hand, our phylogenetic analysis suggest that alpha-amylase group is more diverse, and thus, older than the CGTase group. This is consistent with there being alpha-amylases in a much broader range of taxonomic groups. This notion would imply that CGTases, too, evolved from alpha-amylases.

Additionally, we carried out a similar phylogenetic analyses using an alignment previously published for alpha-amylases and CGTases [17], that

proposed a couple of amylases as *intermediates* between these two groups. According to the fold recognition algorithms used (see Methods), both *intermediate* enzymes present at their N-terminus the three domains shared by alpha-amylases and CGTases (data not shown), but posses sequence lengths similar to or larger than CGTases (anc_dth, 685 aa; anc_bfi, 976 aa). As shown in Figure 1C, one of the *intermediate* amylases (anc_dth, from *Dictyoglomus thermophilum*) is indeed placed apart from alpha-amylases and CGTases, but the other one (anc_bfi, from *Butyrivibrio fibrisolvens*) has close relatives within the alpha-amylases group. This indicates that anc_bfi and anc_dth originated independently. Since the most likely ancestor of anc_bfi was an alpha-amylase, by analogy we believe that the ancestor of anc_dth was also an alpha-amylase-like enzyme. This is consistent with the proposal that alpha-amylases gave rise to CGTase by the acquisition of domains, through an intermediate that had similarity to anc_dth.

In *Escherichia coli* [22], *Hemophilus influenza* [23] and *Bacillus subtilis* [24] the operons for the synthesis of starch always include two depolymerases: glycogen phosphorylase and alpha-amylase. At least in *E. coli*, the starch and glycolytic operons are genetically controlled by the same regulator, csrA [25]. This indicates that alpha-amylases evolved as part of the intermediate metabolism. The fact that alpha-amylases are present in more diverse organisms than CGTases, that their sequences also show larger diversity, and that alpha-amylases belong to the intermediate metabolism, all point out to alpha-amylases having appeared before CGTases.

Assuming a scenario where CGTases appeared in an organism already possessing an alpha-amylase, what could be the adaptive value for their emergence? We figured out that the combined action of alpha-amylases and CGTases may produce more glucose in less time than the simple action of

any of these enzymes. This assumption is based on two observations. First, alpha-amylases, when depolymerizing oligosaccharides containing less than the number of glucoses they can bind at their active sites [26], tend to generate products whose further hydrolysis is inefficient. Second, CGTases produce mainly uniform size cyclodextrins [27]). Consequently, alpha-amylase would efficiently linearize and depolymerize the cyclodextrins, to glucose or maltose and a complementary small oligosaccharides (i.e. a 6-glucoses cyclodextrin would be split to glucose and a maltopentaose, or to maltose and maltotetraose). In order to test this model, we performed the depolymerization of starch employing mixtures of an alpha-amylase with a CGTase (see Methods).

Depolymerization of Starch

The mono and oligosaccharides produced by the action of alpha-amylase, CGTase and different combinations of them are presented in Figure 2. As expected, the action of alpha-amylase on starch produced more small oligosaccharides than CGTase, while this last one produced mainly cyclodextrins. As predicted, the combined action of these enzymes always produce more glucose and small oligosaccharides (maltose, see Figure 2) than the single action of any of them. Additionally, all of the enzymatic mixtures assayed, produced more glucose (see Table 2) and reducing sugars (data not shown) than the action of either alpha-amylase or CGTase alone. Furthermore, the glucose (and reducing sugars, data not shown) produced by any of the enzymatic mixtures was more than the sum of the glucose produced by alpha-amylase and CGTase by themselves (see Table 2). This implies that these enzymes cooperate in the depolymerization of starch, as we predicted. In this regard, it is interesting to note that all known CGTases come from organisms possessing alpha-amylase, which suggest that a CGTase activity by itself might not be of sufficient adaptive value.

4. Conclusions

The aim of this study was to test the hypothesis that CGTase activity have evolved from an ancestral hydrolase (alpha-amylase-like). We found some evidence in our molecular phylogenetic analysis. To test the possible selective advantage of evolving a CGTase activity from an alpha-amylase (to accelerate the saccharification of starch), we measured the production of glucose from starch by enzymatic mixtures of alpha-amylase and CGTase. We found that the combined action of these enzyme activities produce glucose from starch faster and more efficiently. Besides supporting our evolutionary hypothesis, this result explains the biological role of CGTases, and could be useful in biotechnological processes aimed at producing glucose or maltose from starch [28].

Acknowledgments

We want to specially thank Agustin Lopez-Munguia for fruitful discussions, and Alejandro Garciarrubio for critical reading. We also thank Fernando Gonzalez for technical support and Alma Tremari for secretarial assistance. This work was funded in part by PAPIIT project number IN207293 from DGAPA/UNAM.

References

1. Janecek, S. and Saláz, S. Evolution of parallel β/α-barrel enzyme family lightened by structural data on starch-processing enzymes. *J.Protein Chem.* 12:509-514, 1993.

2. Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R. and Svensson, B. Starch-and glycogen-debranching and branching enzymes: prediction of structural features of catalytic (β/alpha)8-Barrel Domain and evolutionary relationship to other amylolytic enzymes. *J.Protein Chem.* 12:791-805, 1993.

3. Sinnott, M.L. Catalytic mechanisms of enzymic glycosyl transfer. *Chem.Rev.* 90:1171-1202, 1990.

4. Svensson, B. Protein engineering in the alpha-amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Molecular Biology* 25:141-157, 1994.

5. Henrissat, B. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem.J.* 280:309-316, 1991.

6. Holm, L. and Sander, Ch. The FSSP database of structurally aligned protein fold families. *Nucleic Acids Res.* 22:3600-3609, 1994.

7. Knox, J.R., Moews, P.C. and Frere, J.M. Molecular evolution of bacterial β-lactam resistance. *Chem.& Biol.* 3:937-947, 1995.

8. Schenkman, S., Eichinger, D., Pereira, M.A.E. and Nussenzweig, V. Structural and functional properties of Trypanosoma trans-sialidase. *Annu.Rev.Microbiol.* 48:499-523, 1994.

9. Stemmer, W.P.C. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. *Proc.Natl.Acad.Sci.USA* 91:10747-10751, 1994.

10. Kauffman, S.A. Applied molecular evolution. *J.Theor.Biol.* 157:1-7, 1992.

11. Shao, Z. and Arnold, F.H. Engineering new functions and altering existing functions. *Current Opinion and Chemical Engineering* 210:513-518, 1997.

12. Kuroki, R., Weaver, L.H. and Matthews, B.W. Structure-based design of a lysozyme with altered catalytic activity. *Nature Structural Biology* 2:1007-1011, 1995.

13. Margulis, L. Lynn Margulis: science's unruly earth mother. *Science* 252:378-381, 1991.

14. Takata, H., Kuriki, T., Okada, S., et al. Action of neopullulanase. *J.Biol.Chem.* 267:18447-18452, 1992.

15. Roderic, D.M. Tree view: an application to display phylogenetic trees on personal computers. *Computer Applications in Biosciences* 12:357-358, 1996.

16. Summer, J.B. and Howell, S.F. A method for determination of invertase activity. *J.Biol.Chem.* 108:51-54, 1935.

17. Janecek, S., MacGregor, E.A. and Svensson, B. Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from α-amylases. *Biochem.J.* 305:685-688, 1995.

18. Jespersen, H.M., MacGregor, E.A., Sierks, M.R. and Svensson, B. Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem.J.* 280:51-55, 1991.

19. Kaneko, T., Song, K.B., Hamamoto, T., Kudo, T. and Horikoshi, K. Construction of a chimeric series of Bacillus cyclomaltodextrin glucanotransferases and analysis of the thermal stabilities and pH optima of the enzymes. *J.Gen.Microbiol.* 135:3447-3457, 1989.

20. Fujiwara, S., Kakihara, H., Woo, K.B., et al. Cyclization characteristics of cyclodextrin glucanotransferase area conferred by the NH_2-terminal region of the enzyme. *Appl.Environm.Microbiol.* 58:4016-4025, 1992.

21. Demetrius, L. Evolutionary dynamics of enzymes. *Protein Eng.* 8:791-800, 1995.

22. Romeo, T., Kumar, A. and Preiss, J. Analysis of the Escherichia coli glycogen gene cluster suggests that catabolic enzymes are encoded among the biosynthetic genes. *Gene* 70:363-376, 1988.

23. Tatusov, R., Mushegian, A.R., Bork, P., et al. Metabolism and evolution of Haemophilus influenzae deduced from a whole-genome comparison with Escherichia coli. *Curr.Biol.* 6:279-291, 1996.

24. Kiel, J.A.K.W., Boels, J.M., Beldman, G. and Venema, G. Glycogen in Bacillus subtilis: molecular characterization of an operon encoding enzymes involved in glycogen biosynthesis and degradation. *Mol.Microbiol.* 11:203-218, 1994.

25. Sabnis, N.A., Yang, H. and Romeo, T. Pleiotropic regulation of central carbohydrate metabolism in Escherichia coli via the Gene csrA. *J.Biol.Chem.* 270:29096-29104, 1995.

26. Allen, J.D. and Thoma, J.A. Multimolecular substrate reactions catalyzed by carbohydrases. Aspergillus oryzae α-amylase degradation of maltooligosaccharides. *Biochemistry* 1997.

27. Nakamura, A., Haga, K. and Yamane, K. The transglycosylation reaction of cyclodextrin glucanotransferase is operated by a ping-pong mechanism. *FEBS Lett.* 337:66-70, 1994.

28. Godfrey, T. and Reichelt, J. The application of enzymes in industry. In: *Industrial Enzymology*, edited by The Nature Press, 1983, p. 375-390.

29. Reardon, D. and Farber, G.K. The structure and evolution of alpha/β barrel proteins. *FASEB J.* 9:497-503, 1995.

Class of Enzyme Act.	Specificity	Mechanism involved in amylases found in Nature (&)
Hydrolases	α(1,4)	Inverting (beta-amylase) Retaining (alpha-amylase)
Hydrolases	α(1,6)	Retaining (oligo α(1,6)-glycosidase)
Transferases	α(1,4)	Retaining (CGTase)
Transferases	α(1,6)	Retaining

Table 1. Starch processing enzymes. & The name of a representative enzyme with a known structure [29] is given in parenthesis. The type of mechanism, inverting or retaining, is also indicated.

Enzymatic Mixture	Glucose (mg/mL) at 0 Hrs	at 5 Hrs	Glucose produced (mg/mL)
Alpha-amylase(1)	1.74	3.19	1.45
CGTase(1)	0.016	0.076	0.050
Alpha-amylase(1):CGTase(1)	1.49	5.71	4.22
Alpha-amylase(1):CGTase(2)	1.37	6.69	5.32
Alpha-amylase(2):CGTase(1)	1.46	5.09	3.63

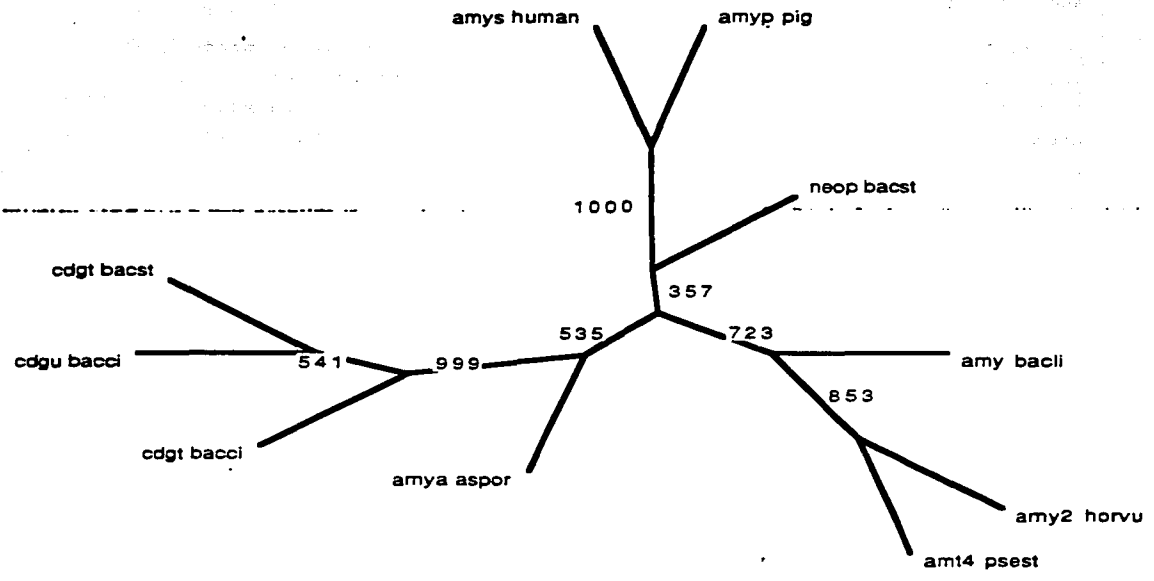
Table 2. Glucose produced from starch by different depolymerization mixtures. Alpha-amylase from *B. licheniformis*, and CGTase: Cyclodextrin Glycosyltransferase from *B. macerans*. The concentration of each enzyme is given in parenthesis as a multiple of standarized concentrations: 142 µg/mL alpha-amylase and 70 µg/mL CGTase.

Figure Legends.

Figure 1.- Phylogenetic threes for alpha-amylase and CGTases. A) genetic distance three, B) parsimony three. The acronyms presented in these trees are explain in Methods. C) genetic distance tree that includes alpha-amylases (amy_*) and CGTases (cgt_*) previously aligned by Svensson and col. [17] (see text).

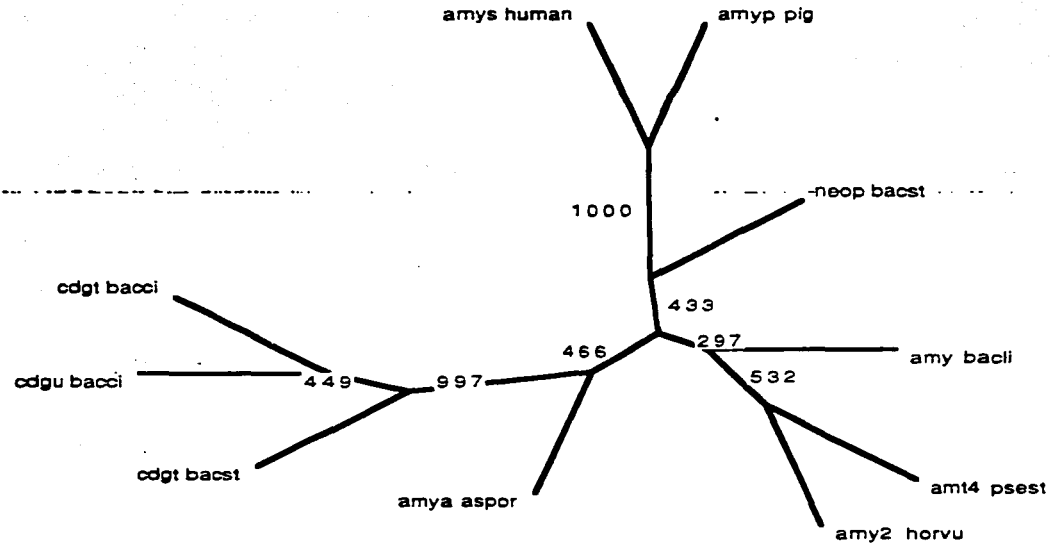
Figure 2.- Thin layer chromatography of starch depolymerization products. Six reaction times are presented for every enzymatic reaction: 0, 1, 2, 3, 4 and 5 hours. Panel A. Lanes 1-6, alpha-amylase hydrolysis products; lanes 7-12 CGTase transferase products. Panel B. Lanes 1-6 alpha-amylase(1):CGTase(1) depolymerization products; lanes 7-12 alpha-amylase(1):CGTase(2) depolymerization products. Panel C, lanes 1-6 alpha-amylase(2):CGTase(1) depolymerization products. For all of the panels, the last two are a ladder of glucose polymers from 1 to 9 (G1 to G9), and α-cyclodextin, respectively. See Methods for description of reaction condiction and nomenclature used.

FIG. 1A


amys human
amyp pig
neop bacst
1000
357
cdgt bacst
cdgu bacci
541
999
535
723
amy bacli
853
cdgt bacci
amya aspor
amy2 horvu
am14 psest

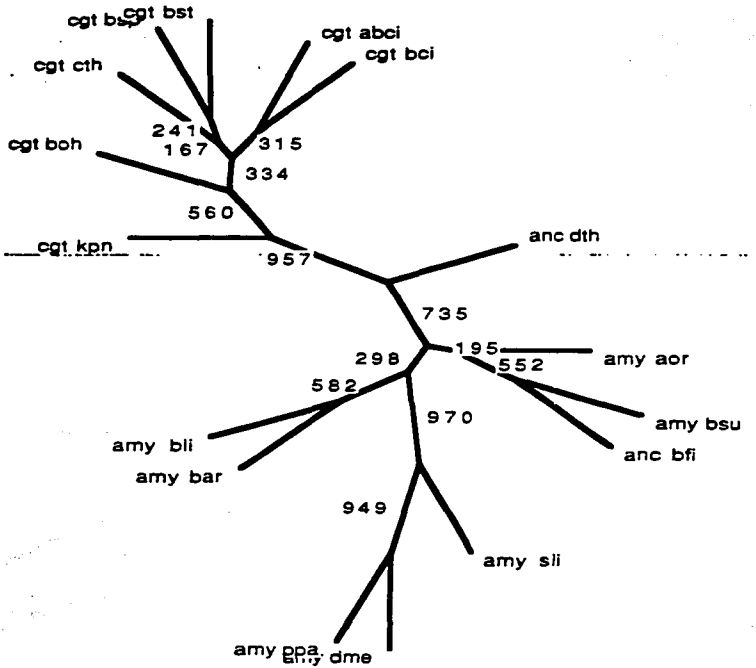
100

Fig. 1B


amys human
amyp pig
1000
neop bacst
433
cdgt bacci
466
297
amy bacli
cdgu bacci
449
997
532
cdgt bacst
amya aspor
amt4 psest
amy2 horvu

100

Fig. 1C


cgt bst
cgt bsp
cgt abci
cgt bci
cgt cth
241
167
315
cgt boh
334
560
cgt kpn
957
anc dth
735
298
195
552
amy aor
582
amy bsu
970
amy bli
anc bfi
amy bar
949
amy sli
amy ppa
amy dme

100