



DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

DEPARTAMENTO DE QUÍMICA

- ÁREA DE BIOFISICOQUÍMICA -

**“DISEÑO DE MUTANTES TERMORRESISTENTES DE LA β -GLUCOSIDASA B DE
Paenibacillus polymyxa, EMPLEANDO HERRAMIENTAS DE BIOINFORMÁTICA”**

T E S I S

QUE PRESENTA

I.B.Q.I. MIRIAM ORDOÑEZ DIAZ

PARA OBTENER EL GRADO DE

MAESTRA EN CIENCIAS (QUÍMICA)

MÉXICO D.F.

PRIMAVERA 2008

El presente trabajo fue realizado en el Área de Biofísicoquímica del Departamento de Química, División de Ciencias Básicas e Ingeniería, de la Universidad Autónoma Metropolitana, Unidad Iztapalapa, bajo la asesoría del Dr. Rafael Arturo Zubillaga Luna.

Agradecemos el apoyo económico brindado por CONACyt a través del proyecto con No. de registro 46166, para la realización del presente trabajo.

Dedico el presente trabajo a mis Papás *Delfina* y *Santiago*.

Gracias por apoyarme en la realización de mis sueños.

Agradezco al Dr. Rafael A. Zubillaga, por haberme permitido desarrollar este proyecto y a su vez aprender cosas nuevas e interesantes.

Agradezco a la Dra Jaqueline A. Padilla Zúñiga por compartir conmigo sus conocimientos y dedicarme su tiempo, sin restricción.

ÍNDICE

Pág.

Capítulo 1. Introducción, Antecedentes, Justificación y Objetivos del Proyecto . . .	1
1.1 Introducción	2
1.2 Antecedentes sobre el estudio de β -glucosidasas.	13
1.3 Justificación	16
1.4 Objetivos del proyecto	17
Capítulo 2. Materiales y Métodos	18
2.1 Búsqueda de la secuencia de aminoácidos de la β gl-B de <i>Paenibacillus polymyxa</i>	20
2.2 Búsqueda de las secuencias de proteínas homólogas a β gl-B	20
2.3 Clasificación de las proteínas homólogas a β gl-B con base en la temperatura óptima de crecimiento de los microorganismos que las contienen	21
2.4 Alineamientos múltiples de las secuencias de proteínas homólogas a β gl-B	22
2.4.1 Obtención de las secuencias consenso mesófila y termófila	23
2.4.2 Alineamiento múltiple de las secuencias consenso mesófila y termófila, β gl-B-PAEPO y β gl-THETH	24
2.4.3 Variabilidad porcentual de la conservación de aminoácidos en las secuencias consenso mesófila y termófila	25
2.5 Análisis comparativo del alineamiento múltiple de las dos secuencias consenso vs un alineamiento tridimensional	25
2.6 Identificación de posibles mutaciones que confieran termoestabilidad a β gl-B	26
2.7 Análisis tridimensional y creación de las posibles mutaciones identificadas en las dos secuencias consenso	27

2.8	Evaluación energética de las mutantes realizadas	30
2.8.1	Generación de los archivos de entrada de las mutantes seleccionadas para la ejecución del programa APBS	30
2.8.2	Evaluación energética de las interacciones electrostáticas en las mutantes individuales generadas <i>in silico</i>	31
2.8.3	Búsqueda de posibles mutaciones múltiples en β gl-B	35
2.9	Comparación molecular <i>in silico</i> del sitio activo de las mutaciones generadas en β gl-B	36

Capítulo 3. Resultados 39

3.1	La secuencia de aminoácidos de la β gl-B de <i>Paenibacillus polymyxa</i>	40
3.2	Secuencias de las proteínas homólogas a β gl-B encontradas	41
3.3	Clasificación de las proteínas homólogas a β gl-B con base en la temperatura óptima de crecimiento de los microorganismos que las contienen	43
3.4	Alineamientos múltiples de las secuencias de proteínas homólogas a β gl-B	44
3.4.1	Obtención de las secuencias consenso mesófila y termófila	46
3.4.2	Alineamiento múltiple de las secuencias consenso mesófila y termófila, β gl-B-PAEPO y β gl-THETH	47
3.4.3	Variabilidad porcentual de la conservación de aminoácidos en las secuencias consenso mesófila y termófila	47
3.5	Análisis comparativo del alineamiento múltiple de las dos secuencias consenso, vs un alineamiento tridimensional	50

3.6	Identificación de posibles mutaciones que confieran termoestabilidad a β gl-B	50
3.7	Análisis tridimensional y creación de las posibles mutaciones identificadas en las dos secuencias consenso	52
3.8	Determinación de las energías de solvatación de las mutantes generadas	61
3.9	Comparación molecular <i>in silico</i> del sitio activo de las mutantes generadas y de la β gl-B	65
Capítulo 4. Análisis y discusión de resultados		70
Conclusiones		77
Perspectivas		78
ANEXOS		79
Anexo A.	Secuencias de aminoácidos de las proteínas homólogas a β gl-B	80
Anexo B.	Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a β gl-B	92
Anexo C.	Formato de archivos .in y .pqr obtenidos por el servidor PDB2PQR	102
REFERENCIAS		104

Capítulo 1. Introducción, Antecedentes, Justificación y Objetivos del Proyecto

1.1 INTRODUCCIÓN

Las proteínas son las macromoléculas más abundantes en las células que constituyen todos los organismos. Existe una gran variedad de ellas en una sola célula, que varían en tamaño desde péptidos relativamente pequeños hasta polímeros enormes de masas moleculares del orden de millones de daltones. Las proteínas constituyen más del 50% del peso seco de las células, y desempeñan una gran variedad y cantidad de funciones vitales que derivan principalmente de su estructura precisa en el espacio y de la estabilidad de ésta en el entorno celular. EL vocablo *proteína* deriva del griego *proteos*, que significa “primero o fundamental”.

La gran diversidad de funciones biológicas que desempeñan las proteínas incluye el papel de hormonas, neurotransmisores, anticuerpos, constituyentes de fibras musculares y nerviosas, algunas actúan como transportadores, otras son proteínas que se ingieren en la dieta diaria. Algunas otras son antibióticos, venenos de hongos, y en un lugar particularmente relevante se encuentra la función de catálisis que realizan las enzimas, las cuales son proteínas especializadas que aceleran considerablemente las reacciones químicas, incluso millones de veces. Prácticamente todas las reacciones que sustentan la vida celular están catalizadas por enzimas.

Otra función importante que desempeñan las proteínas es el ser los instrumentos moleculares mediante los cuales se expresa la información genética, para preservar las características de los miembros de una especie que les hayan permitido adaptarse a entornos ambientales muy diversos y en algunos casos extremos, como aquéllos con temperaturas bajas o elevadas, con alta salinidad o bajo pH. La base estructural de todas las proteínas se encuentra en unas subunidades relativamente simples llamadas *aminoácidos*.

Todas las proteínas, tanto si provienen de los linajes más antiguos como las arqueas, o bien de organismos recientes y complejos (bacterias, hongos, plantas y animales), están construidas a partir del mismo conjunto de 20 aminoácidos unidos de forma covalente a través de un enlace amida, comúnmente denominado *enlace peptídico*; por medio de él, dos aminoácidos pueden unirse formando un dipéptido; de la misma manera se pueden unir tres aminoácidos mediante dos enlaces peptídicos para formar un tripéptido y así sucesivamente hasta formar un *polipéptido*. Se considera que se tiene una proteína cuando se encuentran unidos más de 50 aminoácidos en una cadena polipeptídica, y puede estar constituida por uno o más de estos polipéptidos.

Debido a que cada uno de estos aminoácidos tiene una cadena lateral propia que determina sus características químicas y físicas particulares, se puede considerar a este grupo de 20 moléculas, como el alfabeto en el que está escrito el lenguaje de la estructura proteica.

Existen varios niveles de organización estructural de las proteínas, normalmente se definen cuatro niveles:

Estructura primaria: describe la secuencia de los residuos aminoácidos unidos de forma covalente por enlaces peptídicos y puentes disulfuro.

Estructura secundaria: se refiere a disposiciones particularmente estables de los aminoácidos que dan lugar a patrones estructurales repetitivos, como las hélices α y las hojas β .

Estructura terciaria: representa la manera en cómo la cadena polipeptídica se curva para formar la estructura estrechamente plegada y compacta de una proteína globular.

Estructura cuaternaria: es la disposición espacial de dos o más cadenas polipeptídicas que poseen algunas proteínas llamadas oligoméricas.

La evolución molecular de las proteínas a través del tiempo, se centra generalmente en familias de proteínas estrechamente relacionadas, las cuales comparten un ancestro en común, es decir, son *homólogas*. En este tipo de proteínas, los aminoácidos esenciales para su función se conservan a lo largo del tiempo; mientras que aquellos residuos de aminoácido menos importantes para la función pueden ser susceptibles a cambios debidos quizá, al proceso adaptativo que hayan sufrido los diversos organismos en su hábitat. La localización de estos residuos en un conjunto de proteínas homólogas que comparten una misma función biológica, podría proporcionarnos información valiosa para trazar la historia evolutiva que condujo a éstas proteínas hasta su actual composición y secuencia.

Actualmente en Internet se encuentran disponibles bancos gigantescos de datos de proteínas, los cuales proporcionan información no solamente de la secuencia de aminoácidos que constituyen la proteína de interés, sino que además aportan datos útiles acerca del número de cadenas o unidades que la componen, el tipo de organismo o microorganismo del cual fue obtenida; el hábitat de dichos organismos, la temperatura a la cuál su función se lleva a cabo de manera óptima; esta última característica es dependiente de la clasificación general que ubica a los microorganismos de acuerdo a su temperatura óptima crecimiento (T_{opt}) en¹:

Psicrófilos ($0 < T_{opt} < 20^{\circ}\text{C}$), *Mesófilos* ($20 < T_{opt} < 50^{\circ}\text{C}$)
Termófilos ($50 < T_{opt} < 80^{\circ}\text{C}$), *Hipertermófilos* ($80 < T_{opt} < 120^{\circ}\text{C}$)

Haciendo uso de esta información, y de la gran cantidad de herramientas bioinformáticas actualmente disponibles en Internet de manera gratuita, es posible obtener la secuencia de aminoácidos de todas las proteínas conocidas que sean homólogas a una proteína de interés. Esto permite generar arreglos de secuencias de proteínas en código de una letra, que constituyen verdaderos alineamientos múltiples, los cuales permiten identificar zonas altamente

conservadas que sin duda están íntimamente relacionadas con la estructura y función, o zonas variables.

Basándose en esta información, es posible obtener la **secuencia consenso**, la cual está constituida por los aminoácidos más frecuentemente conservados en cada posición a lo largo de la secuencia de aminoácidos en un grupo de proteínas homólogas. Lehmann M. et al², demostraron exitosamente la importancia del uso de secuencias consenso cuando estudiaron un conjunto de fitasas homólogas de organismos mesófilos extraídas de un cierto tipo de hongos. Su intención era obtener variantes termoestables mediante el empleo de técnicas de mutagénesis dirigida. Las mutaciones potencialmente útiles fueron identificadas con la información derivada de un alineamiento múltiple de proteínas homólogas mesófilas, en él se identificaron a los residuos cargados consensos, los cuales estaban ausentes en las secuencias de aminoácidos de las fitasas, y al generar experimentalmente estas mutaciones observaron que las mutantes obtenidas resultaron ser termorresistentes. Ellos consideraron que en este alineamiento múltiple, se encontraba la información suficiente para obtener enzimas más termoestables, sin considerar a los residuos consenso derivados de un alineamiento múltiple de proteínas homólogas termófilas².

La ingeniería de proteínas es una herramienta poderosa para obtener proteínas modificadas de tal forma que se logre un desempeño mejorado de sus funciones, o para lograr que trabajen en entornos muy diferentes a los que naturalmente requieren. La obtención de propiedades termorresistentes en una proteína proveniente de un organismo mesófilo puede conseguirse a través tres estrategias³:

Diseño racional:

Involucra una serie de principios a través de los cuales la termoestabilización de una proteína puede ser alcanzada o incrementada al aplicar varios de ellos, otorgando cada uno un efecto pequeño pero aditivo. Tales principios consisten en disminuir la entropía del estado desnaturalizado por la introducción de puentes

de hidrógeno adicionales o mediante mutaciones del tipo X→Pro; incrementar la propensión de las α -hélices mediante sustituciones Gly→Ala o por estabilización de los macrodipolos de las hélices; mejorar las interacciones electrostáticas entre residuos cargados superficiales por la introducción de puentes salinos o redes de puentes salinos y prediciendo mutaciones termoestabilizantes en base a cálculos de potenciales electrostáticos.

Evolución dirigida:

Comprende una serie de técnicas experimentales, mediante las cuales se trata de imitar en una escala de tiempo acelerada, el proceso de selección natural de una proteína. Esto se puede lograr a través de la combinación de mutagénesis al azar seguida de un proceso de selección de fácil aplicación. Con frecuencia se emplea también la técnica de intercambio de fragmentos de DNA (DNA shuffling).

La secuencia consenso:

Anteriormente descrita y cuya utilización es considerada como una aproximación semiracional, está basado en la hipótesis de que los aminoácidos consenso que poseen un porcentaje de conservación alto en la familia de proteínas homólogas de interés, contribuyen en mayor medida a la estabilidad y a la función de las mismas; no así, los aminoácidos consenso poco conservados.

Los alineamientos múltiples de secuencias de proteínas homólogas, son una poderosa herramienta que se puede utilizar para diseñar proteínas con alguna característica en particular modificada, por ejemplo conferir termoestabilidad. A partir de ellos es posible observar como han cambiado dichas secuencias a lo largo de la evolución, algunas diferencias estarán allí porque la mutación no ha alterado la función, han sido más o menos neutrales; alternativamente, otras diferencias tendrán que ver con las condiciones ambientales del organismo, por ejemplo las bacterias que viven en fuentes termales a temperaturas muy altas tienen proteínas con una T_m (temperatura media de desnaturalización) también muy alta.

Por otra parte, el hecho de que determinadas posiciones de las secuencias permanezcan *conservadas*, nos indica que éstas tienen una especial importancia en el mantenimiento de la estructura o la función de la proteína y su modificación no ha sido tolerada a lo largo de la evolución.

Al tratar de explicar las bases estructurales responsables de la termoestabilidad molecular, se han realizado comparaciones de secuencias y estructuras de proteínas provenientes de microorganismos termófilos e hipertermófilos, encontrando diferencias importantes respecto a los mesófilos.

Estas diferencias comprenden: un número mayor, tanto de residuos cargados en la superficie molecular, como de enlaces de hidrógeno presentes en la cadena principal y lateral (formado por residuos polares o cargados)^{5-9,11-15,18}; un número mayor de interacciones hidrofóbicas^{4,7,9,13}; mayor presencia de residuos de Pro en asas^{4,7,9,15}; mayor cantidad de estructuras secundarias principalmente hélices α ^{4,7,9,10,12,14,18}; mayor cantidad de residuos cargados en hélices α que las estabilizan por puentes salinos intrahelicoidales y más residuos con carga negativa en su extremo N-terminal^{4,10,15,18}, observando una importante contribución a la estabilidad de las hélices en muchos casos por Arg, debido a su capacidad para mantener su carga positiva y proveer un enlace de hidrógeno adicional en comparación con Lys^{8,15}.

También se ha observado en termófilos un mayor apilamiento de los anillos de residuos aromáticos (fenilalanina, tirosina y triptófano)¹²; un incremento en el grado de oligomerización de las cadenas polipeptídicas^{7,9,12,14}; al igual que un mayor nivel de compactación que reduce las cavidades en el centro hidrofóbico de la proteína^{5-9,12-15}. Así mismo, la cantidad total de superficie polar y/o cargada accesible al solvente^{7-9,14}, y de las interfaces de proteínas multiméricas termorresistentes¹⁴, son más amplias que aquéllas observadas en las proteínas homólogas provenientes de mesófilos^{7,8}. Se ha observado además que la sustitución de residuos en estructuras secundarias es factible para incrementar la termoestabilidad de una proteína^{7,9}.

Estudios recientes han revelado que en diversas familias de termófilos, otro factor importante que contribuye a la estabilidad térmica de sus proteínas, es la presencia de una mayor cantidad de interacciones electrostáticas bastante fuertes pero menos importantes que los pares iónicos, conocidas como “interacciones catión- π ”, formadas entre el aminoácido Lys (catión) y el enlace π del anillo aromático de la Tyr^{10, 12, 14}.

Por otra parte, en las proteínas de termófilos el tamaño de las asas es más corto^{5-7,9,12,14}, poseen menor cantidad de aminoácidos en su secuencia¹⁴, contienen una menor cantidad de Gly la cual está relacionada con la flexibilidad molecular, sobre todo en la proximidad del sitio activo⁸ y se encuentran en menor proporción los residuos lábiles a temperaturas elevadas (Asn, Gln, y Cys)^{7, 9, 12, 16}.

A pesar de esta gran cantidad de diferencias estructurales observadas en los estudios comparativos descritos anteriormente, un factor que parece ser primordial para conferir termoestabilidad a una proteína, es la formación de nuevos *pares iónicos* conocidos también como *puentes salinos*, en la superficie molecular (expuestos al solvente). La creación de *redes de interacciones electrostáticas* entre éstos, puede otorgar un efecto aditivo a la estabilidad de una proteína a altas temperaturas.

Normalmente los grupos cargados que conforman los pares iónicos, se encuentran distantes en la secuencia de una proteína, lo cual permite la generación de enlaces cruzados que estabilizan notablemente la estructura terciaria; tales redes de interacciones se optimizan por medio de la disminución de los efectos repulsivos ocasionados por residuos vecinos que poseen la misma carga eléctrica^{1,3-16,18}.

A través de un estudio comparativo sobre el patrón de aminoácidos localizados en la superficie versus residuos internos, para proteínas homólogas pertenecientes a organismos mesófilos y termófilos, se observó que la diferencia

en la composición de aminoácidos, está principalmente localizada en la superficie de la proteína, y radica en la presencia de un gran número de residuos cargados potencialmente formadores de pares iónicos y de residuos polares; está marcada diferencia sin duda, se encuentra estrechamente relacionada con diferencias entre los hábitats en los que viven tales organismos¹³.

Se ha observado que la energía de las interacciones electrostáticas está determinada básicamente por^{4,17,18}:

- (1) La distancia de separación entre los residuos con carga eléctrica que interaccionan.
- (2) El grado de exposición al solvente de los mismos.
- (3) El efecto que ocasionan los residuos cargados vecinos en el entorno de estas interacciones, ante la posibilidad de formar redes de pares iónicos.

La distancia a la cual se considera la existencia de una interacción electrostática de magnitud considerable, aun no ha sido estandarizada; Barlow y Thornton¹⁸ sugieren que existe una interacción electrostática de magnitud fuerte, si la distancia entre los grupos con carga opuesta es $\leq 4 \text{ \AA}$; mientras que Chakravarty y Varadarajan¹⁰, proponen un criterio en el cual dos residuos con carga opuesta forman un puente salino si, tanto la distancia entre los centros de las dos cadenas laterales como la distancia más corta entre los átomos que soportan la carga eléctrica es $< 6 \text{ \AA}$.

En general, las interacciones electrostáticas se forman entre los residuos cargados a pH 7: Glu (E), Asp (D), Arg (R), Lys (K) (ver **Fig.1**). Las interacciones de esta naturaleza más comunes se observan entre los pares de residuos: E-R, E-K, D-R, D-K. El residuo de His (H) está presente en estas interacciones en menor proporción¹⁰.

Sin embargo, estudios recientes han mostrado que la estabilidad térmica de una proteína mesófila puede ser aumentada, mutando un residuo no cargado localizado en la superficie de la proteína (expuesto al solvente), por uno con carga, el cual establezca una nueva interacción electrostática con otro residuo que posea carga opuesta ó fortalezca las ya existentes a través de la generación de redes de éste tipo de interacciones^{11,19,20}.

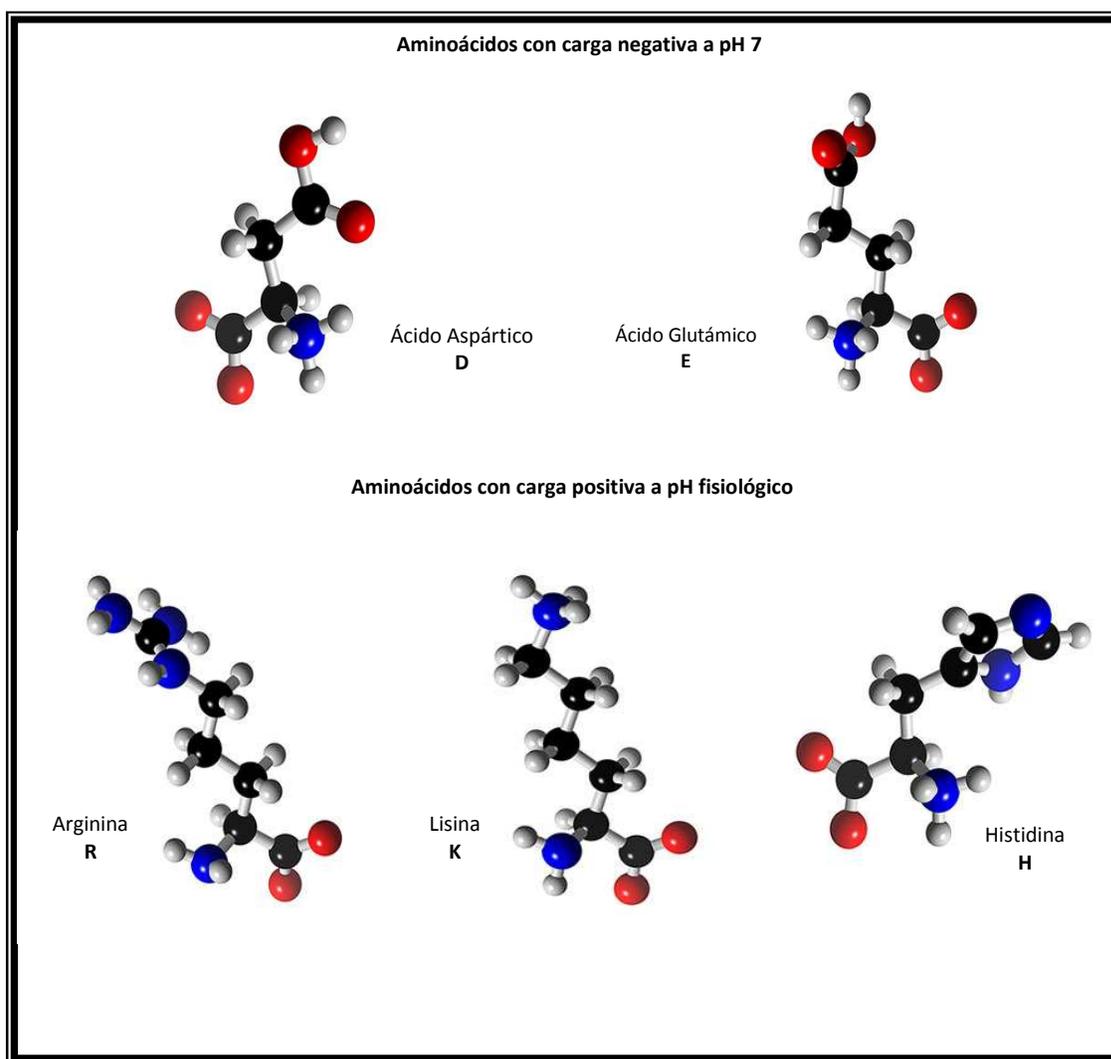


Figura 1. Representación tridimensional de los aminoácidos que adquieren carga eléctrica (positiva o negativa) a pH 7. Se encuentran representados sus diferentes átomos constituyentes por colores: en rojo los átomos de oxígeno, en azul el nitrógeno, en negro el carbono y en gris oscuro el hidrógeno.

Las enzimas son utilizadas ampliamente en diferentes procesos industriales como: fermentaciones, producción de medicamentos, elaboración de alimentos para la alimentación humana y animal, en la fabricación de detergentes, la degradación de agentes contaminantes del medio ambiente, como componentes de biosensores, etc., siendo éstas afines a un grupo muy específico de sustratos; por ejemplo las β -glucosidasas que son ampliamente utilizadas en la elaboración de productos lácteos bajos en contenido de lactosa, en el tratamiento de aguas residuales provenientes de aserraderos y en la fabricación de papel.

Las β -glucosidasas se conocen también como glicosil hidrolasas y se caracterizan por hidrolizar los enlaces β -glicosídicos. Estas enzimas han sido clasificadas en familias extensas y bien caracterizadas, como por ejemplo la clasificación de Henrissat²¹. Esta clasificación establece un catálogo de enzimas que asigna a cada una de ellas un código numérico basado en el tipo de reacción catalizada y en la especificidad por el sustrato. En esta clasificación se encuentran agrupadas todas las β -glucosidasas en familias 1,2 y 3. La familia 1 incluye a enzimas presentes en todo tipo de seres vivos: arqueas, bacterias y eucariotas (hongos, plantas y animales) y se caracterizan por hidrolizar enlaces β -glicosídicos de disacáridos como la celobiosa y la lactosa. A este grupo también pertenece una enzima de gran importancia como lo es la lactasa intestinal humana²².

En la actualidad debido al incremento en la demanda de la utilización de enzimas en diversos procesos industriales, se ha tratado de modificar las enzimas naturales con ayuda de la ingeniería de proteínas, para que mantengan su actividad biológica aún bajo condiciones de trabajo extremas requeridas, por ejemplo: altas temperaturas, pH diferente al óptimo para la adecuada funcionalidad de la enzima, condiciones de salinidad elevada, etc.

Se ha observado que algunas enzimas disminuyen su actividad como consecuencia de una desnaturalización irreversible ocasionada por altas temperaturas.

En el caso de una enzima mesófila de importancia biotecnológica, a la cual se desee conferir propiedades termorresistentes, es necesario conservar su actividad a temperaturas mayores a la óptima de operación. Por lo que la permanencia del estado nativo de una proteína, en determinadas condiciones ambientales de trabajo y en una escala de tiempo finita, está condicionada no sólo por su estabilidad termodinámica, esto es, un valor positivo para la energía libre de desnaturalización, $\Delta G^{N \rightleftharpoons D}$ (ver **Fig. 2**), sino también por su estabilidad cinética, es decir, por una barrera de activación para la desnaturalización suficientemente alta, ΔG^\ddagger .

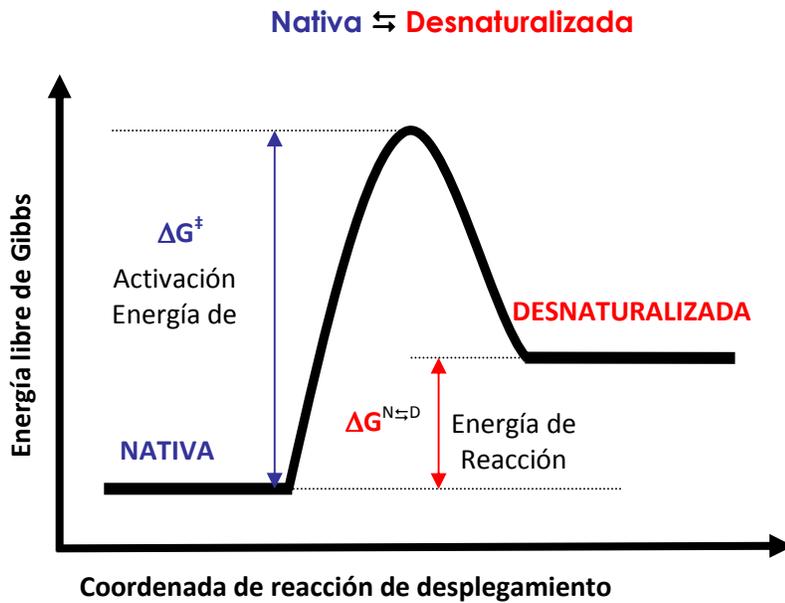


Figura 2. Representación esquemática de la barrera energética que se debe vencer para pasar del estado nativo de una proteína al desplegado.

Lo que se desea en la creación de proteínas con propiedades termorresistentes justamente, es incrementar el tamaño de la barrera energética a vencer para pasar del estado nativo al desplegado, la cual nos garantice que la enzima permanezca nativa a temperaturas elevadas durante periodos de tiempo más prolongados.

1.2 Antecedentes sobre el estudio de β -glucosidasas

En *Paenibacillus polymyxa*, existen dos β -glucosidasas homólogas que pertenecen a la familia 1 de la clasificación de Henrissat; la β -glucosidasa A (β glA) y la β -glucosidasa B (β gl-B), ambas proteínas están integradas por 448 residuos de aminoácidos, comparten un 47 % de identidad en secuencia, presentan una estructura de barril (β/a)₈ y poseen diferente estructura cuaternaria. La β glA es un octámero de alrededor de 400 kDa, su estructura tridimensional ha sido determinada experimentalmente, cuyo código pdb es 1BGA²³. Por su parte la β gl-B es una enzima monomérica que posee una masa molecular de 51.5 kDa. Recientemente ha sido obtenida su estructura cristalográfica por Isorna *et al*²⁴, la cual se presenta en la **Fig. 3**, su código pdb es 2O9P.

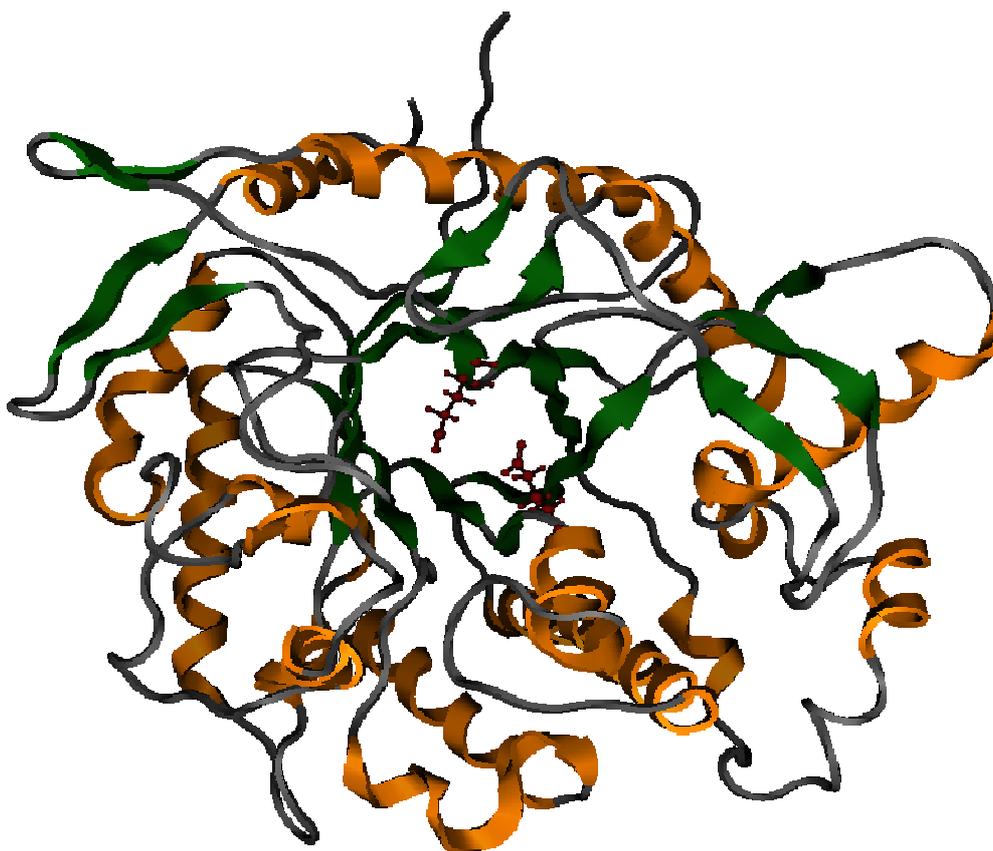


Figura 3. Estructura tridimensional de la enzima β -glucosidasa B de *Paenibacillus polymyxa*. En verde se representan las hojas β que conforman el sitio activo, en naranja las hélices α y en gris las estructuras irregulares. Los residuos catalíticos se muestran en rojo.

Esta enzima realiza la hidrólisis del enlace β -glucosídico a través de dos residuos catalíticos, los cuales se encuentran inmersos en el barril (sitio activo): el Glu 167 que actúa como donador de protones (par ácido/base) y el Glu 356 que actúa como nucleófilo o base. El mecanismo de reacción de esta enzima se presenta en la **Fig. 4**²⁴.

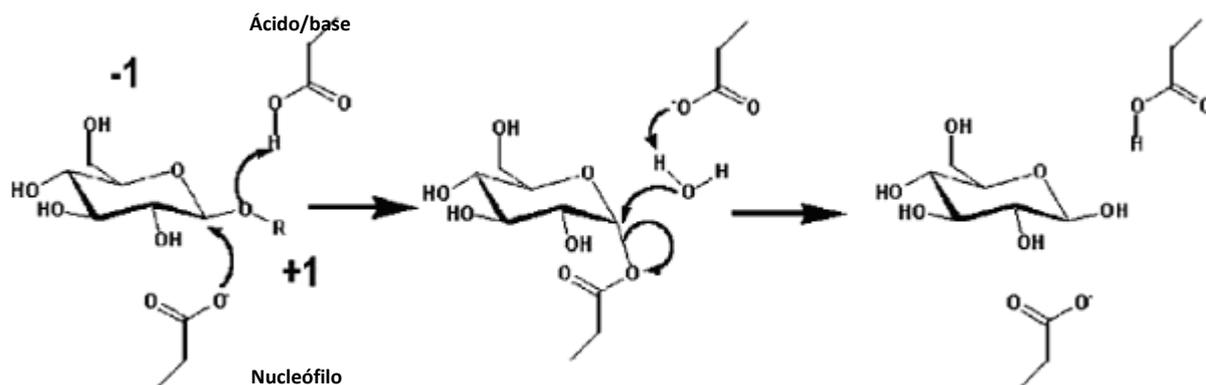


Figura 4. Representación del mecanismo de reacción de la β gl-B sobre su sustrato.

Las dos variantes conocidas de β -glucosidasas de *Paenibacillus polymyxa*, han sido sometidas a procesos mutagénicos con el propósito de obtener mutantes termorresistentes por el grupo del Dr. Julio Polaina (Instituto de Agroquímica y Tecnología de los Alimentos, IATA, Valencia, España). En el caso de la β glA se han llevado a cabo experimentos de mutagénesis aleatoria, encontrando 6 mutantes individuales⁶. Por otra parte, la β gl-B ha sido sometida a experimentos de mutagénesis al azar, encontrando 5 mutantes individuales²⁵.

En los resultados obtenidos de estabilidad térmica para ambas enzimas, se observó que en las mutantes en las que el residuo cambiado tenía carga eléctrica, hubo un incremento más marcado en la estabilidad térmica en comparación con las mutantes en las que el residuo cambiado fue de otra índole (polar, sin carga o hidrofóbico).

Una de las mutantes obtenidas de β gl-B que presentó mayor resistencia térmica, fue la H62R, en la cual, el cambio de un residuo de His por otro de Arg, generó un nuevo par iónico al reducirse la distancia entre la carga negativa del Glu429 y la carga positiva del residuo mutado²⁵.

Esta mutante fue estudiada posteriormente por Zubillaga *et al* ²⁶, quienes realizaron cinéticas de desnaturalización térmica de esta mutante y la β gl-B, con la finalidad de evaluar el efecto ocasionado por el nuevo par iónico sobre la barrera de activación del desplegamiento. Ellos encontraron que dicha barrera fue mayor para la mutante en alrededor de 1.0 Kcal/mol con respecto a la enzima silvestre, como se esquematiza en la **Fig. 5**.

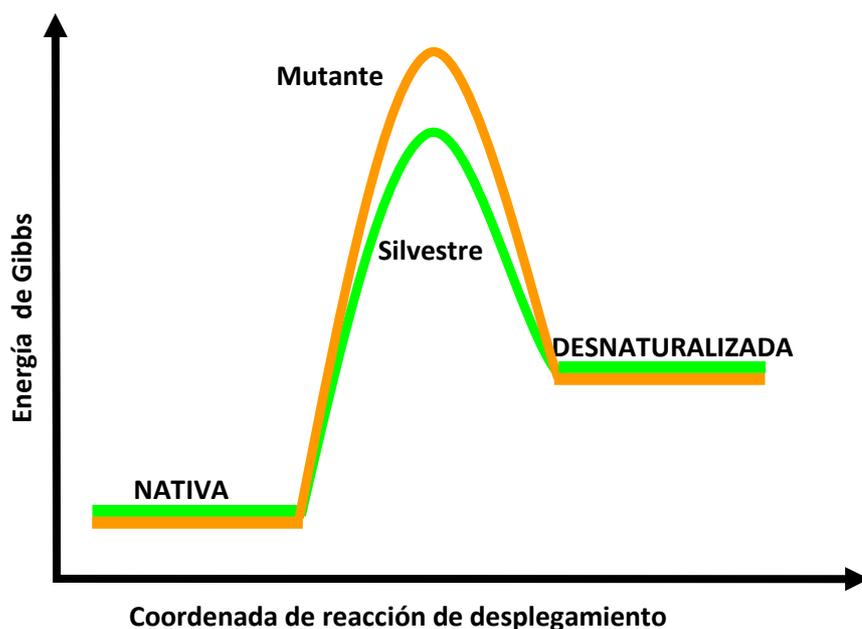


Figura 5. Representación gráfica del efecto ocasionado por una mutación estabilizante sobre la altura de la barrera energética de activación en el proceso de desnaturalización térmica de una enzima.

1.3 JUSTIFICACIÓN

La importancia biotecnológica de obtener una β -glucosidasa termorresistente, radica en la elaboración de productos lácteos bajos en contenido de lactosa, que son indispensables en nuestra alimentación. Debido a que una gran cantidad de seres humanos son intolerantes a la lactosa, la ingesta de estos productos puede proporcionar los nutrientes propios de la leche que son indispensables en nuestra alimentación, sin ocasionar molestias digestivas.

Al incrementar la termorresistencia de esta enzima, ésta podría realizar su actividad en un amplio intervalo de temperatura esto es, entre 20 y 80 °C o quizás aun mayor; tolerando el proceso de pasteurización de la leche sin desnaturalizarse^{22,27}. Haciendo más eficiente este proceso, lo cual repercute tanto en los costos de producción de la leche pasteurizada y de sus derivados lácteos, como en los costos de venta de estos productos, que sin duda se ven reflejados en la economía de las personas que consumen estos productos de canasta básica.

Como se ha descrito en párrafos anteriores, dado que el favorecimiento de la formación de interacciones electrostáticas superficiales, podrían contribuir para conferir termorresistencia a proteínas provenientes de microorganismos mesófilos⁵; en el presente trabajo se pretende diseñar bajo un criterio racional, mutaciones *in silico* que pudieran conferir estabilidad térmica a la enzima β -glucosidasa B (β gl-B) del microorganismo mesófilo *Paenibacillus polymyxa*, con ayuda de herramientas de bioinformática.

La idea central es que al generar nuevas interacciones electrostáticas estabilizantes en la estructura de β gl-B, las variantes resultantes tendrán una termorresistencia mayor que la molécula original.

1.4 **OBJETIVOS DEL PROYECTO**

Objetivo General

- Diseñar, construir y evaluar energéticamente mutantes *in silico* que podrían conferir termorresistencia a la β -glucosidasa B (β gl-B) de *Paenibacillus polymyxa*, utilizando herramientas de bioinformática.

Objetivos Particulares

- Obtención de las secuencias consenso a partir de grupos de proteínas homólogas a β gl-B provenientes de organismos mesófilos y termófilos.
- Localización de las diferencias en residuos cargados entre las secuencias consenso y la secuencia de β gl-B.
- Construcción *in silico* de mutantes de β gl-B que involucren cambios en la carga eléctrica acordes con las secuencias consenso.
- Evaluación de la energía libre de Gibbs electrostática total de las mutantes diseñadas *in silico* y de β gl-B.
- Comparación molecular *in silico* del sitio activo de la β gl-B y de las mutantes generadas.

Capítulo 2. Materiales y Métodos

En el presente capítulo se detallan los aspectos teóricos involucrados así como la metodología empleada y la información que se requiere para ejecutar cada programa utilizado con la finalidad de alcanzar los objetivos planteados en el presente trabajo.

Actualmente es posible emplear programas de simulación molecular para estudiar biomoléculas como el ADN y las proteínas, en diferentes conformaciones y condiciones de trabajo (en presencia de disolvente acuoso y cosolutos salinos, o en ausencia de estos), e incluso con mutaciones generadas en su cadena. Entre los programas disponibles, los hay tanto comerciales como gratuitos en Internet.

Diferentes servidores en Internet realizan búsquedas o análisis de las características de grupos de proteínas que comparten un ancestro en común, en bases gigantescas de secuencias de proteínas caracterizadas experimentalmente en todo el mundo. También es posible generar una comparación simultánea de estas proteínas a través de un *alineamiento múltiple* de las secuencias homólogas, con el fin de ubicar zonas de mayor conservación en sus secuencias a lo largo de la evolución; las cuales seguramente estarán vinculadas con la sustentación de la estructura, de la función molecular o con ambas.

Empleando programas de simulación molecular y las estructuras tridimensionales de proteínas homólogas cristalizadas, es posible realizar un alineamiento tridimensional de ellas con la finalidad de visualizar las diferencias estructurales entre ellas, y a su vez poder comparar dicho alineamiento contra un alineamiento múltiple de secuencias.

Con ayuda de estas herramientas bioinformáticas, también es posible generar mutaciones en la secuencia de una proteína y determinar los efectos ocasionados por estas, por ejemplo: la modificación de las diferentes zonas cargadas positiva o negativamente en la estructura de la proteína, el cambio energético provocado evaluado a través de la energía de interacción proteína-ligando, la modificación del sitio catalítico, cálculos electrostáticos, entre otras cosas.

2.1 Búsqueda de la secuencia de aminoácidos de la β gl-B de *Paenibacillus polymyxa*.

Para iniciar el desarrollo del presente trabajo, era necesario tener la secuencia de aminoácidos de la proteína bajo estudio, la β -glucosidasa B (β gl-B) de *Paenibacillus polymyxa*; la búsqueda de tal secuencia de aminoácidos, se realizó empleando la página disponible en Internet, de herramientas bioinformáticas muy diversas, **ExpASY Proteomics Server (Expert Protein Analysis System)** del [Swiss Institute of Bioinformatics](http://ca.expasy.org), cuya dirección electrónica es <http://ca.expasy.org>; en la cual se tiene acceso directo a la base de datos **Swiss-Prot and TrEMBL** no redundante que contiene miles secuencias de proteínas; traducidas del genoma de diversos organismos, las cuales fueron obtenidas por secuenciación experimental en diversas partes del mundo.

La búsqueda se realizó anotando el nombre de *Paenibacillus polymyxa* en la ventana de diálogo de la base de datos **Swiss-Prot and TrEMBL**, la cual se encuentra disponible al inicio de la página de **ExpASY**, obteniendo como resultado una lista de los archivos existentes en esta base de datos que contienen la palabra escrita.

De los archivos listados se seleccionó la opción β gl-B de *Paenibacillus polymyxa*, de donde puede obtenerse la secuencia de la proteína bajo estudio en código de una letra, agrupada en segmentos de 10 aminoácidos (**Fig. 8**) y en formato FASTA (**Fig. 9**), el cual es requerido en diversos programas de análisis bioinformático.

2.2 Búsqueda de las secuencias de proteínas homólogas a β gl-B.

La búsqueda de las secuencias de las proteínas homólogas a β gl-B se realizó empleando el servidor disponible en Internet del programa **BLAST2** “**Basic Local Alignment Search Tool**”²⁸. A este servidor puede accederse desde la página de **Expasy Proteomics Server**, en el apartado **Tools and software packages**. Este programa realiza la búsqueda de las secuencias de las proteínas homólogas a una proteína de interés, a través de alineamientos por pares de secuencias, es decir entre la secuencia de trabajo y cada una de las

secuencias homólogas existentes en su base de datos (**Swiss-Prot y TrEMBL**). Únicamente solicita ingresar la secuencia de aminoácidos de la proteína bajo estudio, en este caso la β gl-B en la ventana de diálogo del servidor del programa **BLAST2**; esta búsqueda se realizó con los parámetros preestablecidos en el servidor. A este programa puede accederse alternativamente, a partir de la página de **NCBI (National Center For Biotechnology Information)** cuya dirección electrónica es www.ncbi.nlm.nih.gov/. Las secuencias de aminoácidos de las proteínas homólogas a β gl-B obtenidas de ésta manera, se solicitaron en formato Fasta y todas ellas se encuentran en el **Anexo A** del presente trabajo. En la presente búsqueda se encontraron algunas proteínas con estructura tridimensional conocida, las cuales se emplearon en etapas posteriores durante el desarrollo del presente trabajo.

2.3 Clasificación de las proteínas homólogas a β gl-B con base en la temperatura óptima de crecimiento de los microorganismos que las contienen.

A través de un análisis de la lista que contiene toda la información de las proteínas homólogas a β gl-B (secuencia de aminoácidos, el nombre del microorganismo del que fue obtenida dicha proteína, el código PDB en el caso de que haya sido cristalizada, etc.), obtenida por el servidor **BLAST2**; se advirtió la presencia de microorganismos cuyos nombres sistemáticos no permitían una fácil clasificación de acuerdo a sus temperaturas óptimas de crecimiento (psicrófilos, mesófilos, termófilos e hipertermófilos).

Para buscar la información taxonómica de los microorganismos a partir de los cuales fueron obtenidas las proteínas homólogas a β gl-B, se empleó el servidor **ENTREZ Genome Project**, al que puede accederse desde la página de **Pub Med** del **NCBI**. Este servidor contiene una vasta base de datos que incluye a todos los microorganismos cuyo genoma es conocido. Para realizar la búsqueda es suficiente con ingresar el nombre sistemático del microorganismo de interés.

La información desplegada contiene los siguientes datos: hábitat del microorganismo del que fue aislado, sus características celulares (tipo de tinción Gram, forma celular, arreglo celular, tendencia a la formación de endosporas, movilidad). De igual manera se presentan las características ambientales del microorganismo (condiciones de salinidad (halotolerantes), requerimiento de oxígeno (aerobio o anaerobio), condiciones de temperatura (temperatura óptima y su clasificación con base en ella (psicrófilos, mesófilos, termófilos e hipertermófilos)).

Con la información obtenida para cada microorganismo, se realizó posteriormente una clasificación de ellos en dos grupos, mesófilos y termófilos.

Basándose en ambas clasificaciones, se procedió a agrupar dichas secuencias de proteínas homólogas. Dentro de esta clasificación se observó en algunas secuencias la presencia de más de una cadena, las cuales se diferenciaban por la variación en promedio de entre 3-10 residuos, y/o la ausencia de algunos residuos al final de estas secuencias. En estos casos se eligió a una sola cadena la cual tuviera una longitud aproximada a la secuencia de β gl-B (había cadenas bastante largas), o bien la que contuviera más residuos cargados. En otros casos había secuencias que tenían la insignia de "probable beta-glucosidasa" y no había rastro del microorganismo del que fueron aisladas, por lo que estas secuencias no se tomaron en cuenta.

2.4 Alineamientos múltiples de las secuencias de proteínas homólogas a β gl-B.

Para realizar el alineamiento múltiple de las secuencias de las proteínas homólogas a β gl-B, se utilizó el servidor del programa **MultAlin**²⁹, cuya dirección electrónica es: <http://bioinfo.genopole-toulouse.prd.fr/multalin/multalin.html>.

Con las secuencias en formato FASTA de las proteínas homólogas a β gl-B clasificadas en mesófilas y termófilas, se realizó un alineamiento múltiple de ellas, ingresando el conjunto de secuencias FASTA de cada grupo de proteínas

homologas por separado en este servidor, efectuando tales alineamientos múltiples bajo los parámetros preestablecidos en este servidor.

Este programa construye alineamientos múltiples de proteínas, generados a partir de cúmulos de secuencias que resultan del alineamiento de la proteína de interés con las homólogas restantes por pares posibles; para lograrlo, abre espacios (inserciones) en cada secuencia polipeptídica, con la finalidad de emparejar fragmentos de conservación similar en dicho grupo de proteínas, cuyo puntaje sea el más alto. Estos arreglos múltiples permiten advertir rápidamente las zonas totalmente conservadas y las zonas de conservación variable, entre las diferentes secuencias polipeptídicas homólogas estudiadas, como se observa en los alineamientos completos en el **Anexo B**.

El programa colorea la salida del alineamiento múltiple, en base a la conservación de aminoácidos, en rojo se presentan los residuos altamente conservados, en azul las posiciones conservadas, y en color negro se anotan los residuos poco conservados.

2.4.1 Obtención de las secuencias consenso mesófila y termófila.

MultAlin genera la secuencia consenso de un grupo de secuencias de proteínas homólogas, encontrando en cada posición, un carácter que está en función de los caracteres en todas las secuencias en la misma posición del alineamiento múltiple. Si más del 90% de los caracteres en la columna correspondiente, tiene la misma letra, el carácter consenso es esa letra en mayúsculas rojas. Si más de un 90% de los caracteres de la columna, están conservados e integran el mismo grupo de conservación representado por un símbolo: ! (IV), \$ (LM), % (FY), # (NDQE); el carácter colocado en esa posición de la secuencia consenso es el correspondiente símbolo en rojo. Si el porcentaje de conservación es inferior al 90%, el carácter consenso es representado en minúsculas azules. Una posición que no contiene residuos conservados se representa por un punto, y las zonas de inserciones se representan por un espacio en blanco.

Para obtener las secuencias consenso, se consideró una conservación en secuencia que oscilaba entre el 5% y el 95%, los demás parámetros involucrados para el proceso de alineamiento no se variaron. En algunas zonas de ellas, se advirtió la presencia de símbolos que colocó este programa al no poder tomar una decisión para colocar un aminoácido en esa posición; por lo que se tuvo que editar manualmente con ayuda de un procesador de textos (Word), dando prioridad en importancia consecutiva a los residuos con carga, polares, o a los que no causaran impedimento estérico de acuerdo a la estructura química que presentan estos residuos. Ambas secuencias se utilizaron ampliamente en etapas subsecuentes del proyecto y pueden observarse a detalle en la **Fig. 13**.

2.4.2 Alineamiento múltiple de las secuencias consenso mesófila y termófila, β gl-B-PAEPO y β gl-THETH.

Al realizar la búsqueda de las proteínas homólogas a β gl-B, se encontró una β -glucosidasa, la β gl-THETH, obtenida del microorganismo termófilo *Thermus thermophilus*, cuya estructura se encuentra depositada en la base de datos del PDB con la clave 1UG6, la cual se consideró como representativa de su grupo de proteínas homólogas y se empleó en las siguientes etapas de desarrollo del presente trabajo. Es una enzima monomérica integrada por 431 aminoácidos, fue cristalizada con una resolución de 0.99 Å, su estructura tridimensional se muestra en la **Fig. 6**.

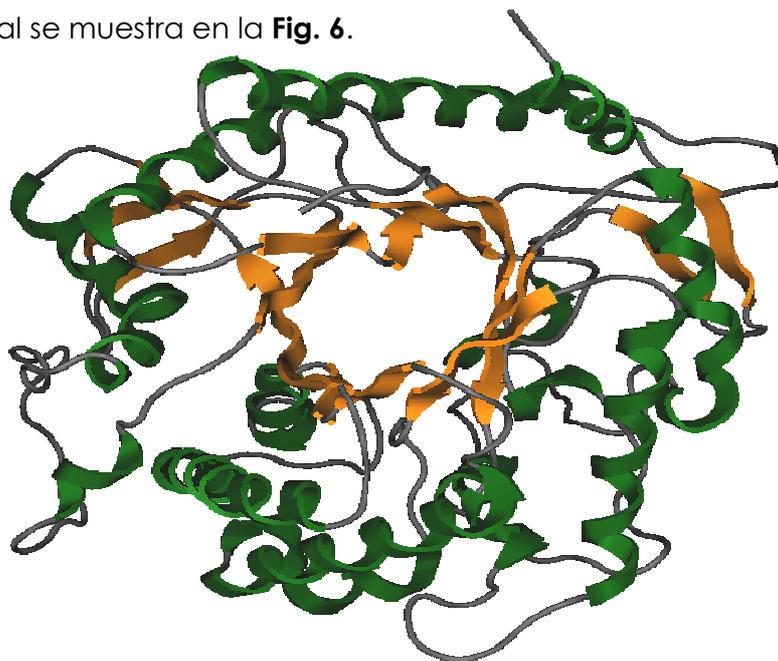


Figura 6. Estructura tridimensional de la β gl-THETH de *Thermus thermophilus*.

Se realizó un alineamiento múltiple de las secuencias consenso obtenidas junto con la β gl-B y la β gl-THETH, de igual manera utilizando el servidor **MultAlin**; con la finalidad de poder comparar y ubicar diferencias en residuos cargados.

2.4.3 Variabilidad porcentual de la conservación de aminoácidos en las secuencias consenso mesófila y termófila.

Para poder observar las diferencias evolutivas en ambas familias de proteínas homólogas (mesófilas y termófilas) de una manera cuantitativa, se calculó el porcentaje de conservación de cada aminoácido en cada posición del alineamiento múltiple realizado para los dos grupos de proteínas bajo estudio; se calcularon los porcentajes de conservación con ayuda de la hoja de cálculo Excel de Windows y los alineamientos múltiples de ambas familias.

A continuación se agruparon éstos porcentajes en intervalos, asignándole un color a cada intervalo y se procedió a colorear cada aminoácido de ambas secuencias consenso basándose en estos porcentajes calculados. Esta información se detalla en la **Fig. 14**.

2.5 Análisis comparativo del alineamiento múltiple de las dos secuencias consenso vs una superposición estructural.

Una superposición de estructuras de proteínas homólogas, es de gran utilidad para comparar visualmente la compatibilidad o similitud de ésta con respecto a un alineamiento múltiple de secuencias generado por servidores como **MultAlin**. Para superponer las estructuras tridimensionales de β gl-B y β gl-THET (termófila cristalizada) en este trabajo, se empleó el programa **MOE** (**M**olecular **O**perating **E**nvironmet) versión 2005.06 creado por el [CHEMICAL COMPUTING GROUP INC.](http://www.chemcomp.com/) (<http://www.chemcomp.com/>). Este programa requiere los archivos en formato pdb de las proteínas de interés, en este caso la β gl-B y la β gl-THET, los cuales poseen las coordenadas tridimensionales para cada átomo presente en la molécula de proteína, que son utilizadas por el programa **MOE** para representar tridimensionalmente a estas estructuras proteicas (MOE Tutorials versión 2005.06)³⁰.

En una primera etapa se superpusieron las estructuras de β gl-B-PAEPO y β gl-THET, y en una ventana alterna llamada **Sequence Editor**, en donde se encuentran las secuencias de ambas proteínas, y que está ligada a la ventana gráfica de **MOE**, se procedió a editar este alineamiento colocando las inserciones necesarias para convertirlo en uno idéntico al generado por **MultAlin**, que se presenta en la **Fig. 15**.

Posteriormente se seleccionaron los residuos laterales a cada inserción en el **Sequence Editor** y en la ventana gráfica, observando que los residuos seleccionados se encontraban en la misma posición estructural y paralelamente orientados, como se puede observar en la **Fig. 15**. Basándose en lo antes mencionado, se pudo afirmar que el alineamiento generado por **MultAlin** es confiable y compatible estructuralmente.

2.6 Identificación de posibles mutaciones que confieran termoestabilidad a β gl-B.

Una vez corroborada la confiabilidad del alineamiento secuencial a partir de la superposición estructural, como se mencionó en la sección anterior, la identificación de las posibles mutaciones termoestabilizantes de la β gl-B, se realizó localizando a los residuos cargados (E, D, R, K) en el alineamiento múltiple de la **Fig. 13**; en el que se encuentran las dos secuencias consenso, la β gl-B-PAEPO y la β gl-THETH; haciendo un análisis manual, posición a posición a lo largo del mismo, con la finalidad de localizar aquellos sitios que presentaran residuos cargados en al menos una de las secuencias consenso y los cuales estuviesen ausentes en esa misma posición, en la secuencia de β gl-B.

De esta forma se identificaron los **55** sitios potencialmente viables para mutarse que se anotan en la **Tabla 3**. Estos sitios no involucraron mutaciones en regiones altamente conservadas simultáneamente en ambos grupos de proteínas empleados en este trabajo, como se observa en la **Fig. 14**. Este análisis se realizó con la finalidad de evitar cualquier posible alteración en la estabilidad, la estructura y/o la función, de la β gl-B.

2.7 Análisis tridimensional y creación de las posibles mutaciones identificadas en las dos secuencias consenso.

Con las 55 posibles mutaciones identificadas, se procedió a crear cada una de ellas a partir del archivo pdb de la β gl-B, empleando el programa **MOE**.

En cada caso se mutó el residuo propuesto por al menos una de las secuencias consenso, en aquellos sitios en las que la mutación propuesta difería en carga, se probaron ambas. Se revisaron los vecinos cercanos a la mutación generada, con la finalidad de localizar a los residuos con carga opuesta potencialmente formadores de nuevas interacciones electrostáticas estabilizantes. Posteriormente, se realizó un proceso de minimización de energía de cada una de las variantes de β gl-B.

La minimización de energía de una molécula consiste en encontrar el conjunto de coordenadas atómicas que corresponden al mínimo local de la función de energía potencial empleada por el campo de fuerzas.

Este proceso se realiza aplicando técnicas de optimización no lineal a gran escala, que determinan la conformación de la molécula en la cual, la resultante de las fuerzas que se ejercen sobre los átomos sea cero, con esto se intenta encontrar una geometría similar a la que presenta la proteína en el medio celular³⁰.

La energía potencial de un sistema molecular está en función de las coordenadas atómicas, y puede calcularse de acuerdo a la contribución aditiva de varias interacciones, como se describe en la siguiente ecuación:

$$E(x) = E_{str} + E_{ang} + E_{stb} + E_{oop} + E_{tor} + E_{vdw} + E_{ele} + E_{sol} + E_{res} \quad [1]$$

Cada uno de los términos de la energía, es a su vez una suma que involucra el tipo de interacciones atómicas para cada par, tercia o cuarteto de átomos de los que está comprendida la molécula completa. En esta expresión:

E_{str} es la energía de estiramiento y compresión del enlace entre dos átomos.

E_{ang} es la energía debida a la magnitud del ángulo de enlace entre tres átomos enlazados covalentemente.

E_{stb} es la energía debida a la combinación tanto del estiramiento del enlace como de la magnitud del ángulo.

E_{oop} es la energía que debe remontarse para alterar la configuración de un isómero.

E_{tor} es la energía de torsión debida al ángulo diedro.

E_{vdw} es la energía de van der Waals.

E_{ele} es la energía electrostática.

E_{sol} y E_{rest} son parámetros que evalúan la energía de solvatación y de restricción geométrica, de acuerdo con las condiciones particulares de uso.

Para realizar la minimización de la energía de cada mutante realizada en β gl-B, se empleó el campo de fuerzas CHARMM 27, el cual está parametrizado para biomoléculas como proteínas, ADN y ARN.

Con la finalidad de alcanzar el estado más relajado para todos los átomos presentes en cada una de las moléculas de β gl-B mutadas, se diseñó la minimización por etapas partiendo del residuo mutado, permitiendo que éste se moviera libremente hasta encontrar una posición estructural adecuada con respecto al resto de la estructura proteica, la cual permaneció fija.

Posteriormente se permitió la movilidad tanto del residuo mutado como de sus vecinos cercanos a 4.5 Å; finalmente, se permitió la movilidad de toda la molécula de β gl-B para encontrar el estado más relajado de todos los átomos que la conforman.

Antes de iniciar la minimización, y con la mutación ya generada, se midió la distancia entre los átomos con carga de las cadenas laterales de los residuos involucrados en la nueva interacción electrostática.

Esta distancia fue monitorizada a lo largo del proceso de minimización, la mayoría de estas mutantes presentaron grupos con carga opuesta ubicados a una distancia menor de 6.5 Å de distancia al final de este proceso.

De las 55 mutaciones propuestas con base en los alineamientos de las secuencias consenso, hubo una gran cantidad de ellas que al construirse no mostraron vecinos de carga opuesta con los cuales pudieran formar una nueva interacción electrostática. En otros casos, al revisar el entorno inmediato a la mutación, se observó que al generarlas se destruirían las interacciones electrostáticas ya existentes en esa zona de la estructura de β gl-B; por tanto, estas mutantes no fueron empleadas en el resto del trabajo.

Por otra parte, de acuerdo con los valores reportados por Barlow y Thornton¹⁸, Chakravarty y Varadarajan¹⁰, la distancia máxima entre los átomos con carga opuesta en un par iónico de magnitud considerable está entre 4 y 6 Å. En el presente trabajo la mayor parte de las interacciones electrostáticas propuestas se mantuvo dentro del intervalo de distancias mencionado, al concluir la minimización, por lo que estas se pueden considerar como pares iónicos nuevos (ver **Fig. 17**).

De esta manera se obtuvieron 17 mutantes individuales a partir de las 55 seleccionadas originalmente. Tales mutaciones se señalan con colores sobre el alineamiento múltiple de las dos secuencias consenso con β gl-B y β gl-THETH (**Fig. 16**).

A continuación se generaron los archivos pdb de cada mutante, utilizando el programa **MOE**, los cuales se utilizaron más adelante en el desarrollo del presente trabajo.

2.8 Evaluación energética de las mutantes realizadas.

2.8.1 Generación de los archivos de entrada de las mutantes seleccionadas para la ejecución del programa APBS.

Para poder evaluar energéticamente cada una de las mutantes, se partió de sus archivos pdb eliminando todas las moléculas de agua presentes. Cada uno de estos archivos se transfirió al servidor del programa **PDB2PQR** (<http://pdb2pqr.sourceforge.net/>); el cual transforma los archivos pdb de las proteínas a formato .pqr, que contiene información adicional sobre las cargas y los radios de cada uno de los átomos presentes en la molécula proteica y también adiciona hidrógenos a la estructura molecular optimizándolos.

La salida del programa proporciona dos archivos, uno con extensión **“.in”** que contiene las dimensiones de la malla que ajusta a la superficie molecular de la proteína, por ejemplo para la β gl-B es de 161x129x161, el radio del solvente (1.40 Å), las constantes dieléctricas tanto del solvente (78.5400) como de la proteína (2.0), la fuerza iónica del amortiguador empleado (opcional), la temperatura considerada (298.15 °C), el tipo de algoritmo empleado para efectuar el cálculo, ya sea lineal (**lpbe**) o no lineal (**npbe**) de la Ecuación de Poisson-Boltzmann (**EPB**), etc. Todos estos parámetros son asignados con base en el campo de fuerzas elegido, para este trabajo se utilizó CHARMM. El segundo archivo con extensión **“.pqr”** contiene parte de la información del archivo pdb original además de la carga asignada a cada átomo y su radio. Ambos archivos de salida son requeridos por diversos programas que realizan cálculos electrostáticos como el programa **APBS (Adaptive Poisson-Boltzmann Solver)** empleado en este trabajo³¹.

Una vez que se tuvieron los archivos **.in** y **.pqr** para cada mutante, se les hizo una pequeña edición para después utilizarlos como entrada del programa **APBS**. En los archivos **.in** se asignó un valor de 4.0 para la constante dieléctrica de la proteína, y se adicionó el valor de la fuerza iónica (0.088 M)

del amortiguador de fosfatos ($\text{KH}_2\text{PO}_4 / \text{K}_2\text{HPO}_4$), en el cual se estudia a la β gl-B silvestre en nuestro laboratorio, además se eligió la forma no lineal (**npbe**) de la Ecuación de Poisson-Boltzmann para efectuar el cálculo electrostático. Los parámetros que quedan inalterados en el archivo **.in** son tanto el radio del solvente acuoso (1.4 Å), su constante dieléctrica (78.54) como la temperatura del medio (298.15 °C). En los archivos **.pqr** sólo se eliminaron las denominaciones de algunos átomos de la cadena tales como: el N terminal (NTER por TER), el C terminal (CTER por TER) y los residuos de GLU-P (GLU-P por GLU), asignados por el mismo servidor de **PDB2PQR**, simplemente para que el programa APBS identificara a estos residuos como parte de la proteína. Un ejemplo tanto del archivo **.in** como del **.pqr**, se presentan en el **Anexo C**.

2.8.2 Evaluación energética de las interacciones electrostáticas en las mutantes individuales generadas *in silico*.

Como se mencionó anteriormente, ambos archivos (**.in** y **.pqr**) son requeridos por el programa **APBS (Adaptive Poisson-Boltzmann Solver)**, para evaluar las propiedades electrostáticas de biomoléculas enormes como las proteínas y el ADN.

La energía libre electrostática total ($\Delta G_{ElsTot}^{\ddagger}$) de un sistema macromolecular puede ser calculada a partir de las contribuciones de las interacciones entre los pares de átomos cargados de la macromolécula, así como de las interacciones de las cargas individuales con el solvente, el cual puede contener electrolitos. La primera de estas contribuciones es conocida como la energía libre electrostática coulombica ($\Delta G_{ElsCoul}^{\ddagger}$), que puede calcularse a partir de las interacciones coulombicas existentes entre los grupos cargados. La segunda contribución puede conocerse a través de la energía libre electrostática de solvatación ($\Delta G_{ElsSolv}^{\ddagger}$) que considera tanto la interacción de las cargas de la molécula con un solvente polarizable, como la contribución causada por la interacción de esas cargas con la atmósfera iónica del solvente. De manera que la energía electrostática total de una macromolécula puede calcularse como³²

$$\Delta G_{ElsTot}^{\ddagger} = \Delta G_{ElsCoul}^{\ddagger} + \Delta G_{ElsSolv}^{\ddagger}$$

El programa **APBS** puede calcular ambos sumandos de la ecuación anterior. El (ΔG_{Ele}^{solv}) es calculado directamente sólo pidiendo que se realice el cálculo sobre el archivo **.in** en cuestión, el cual hace referencia al archivo **.pqr** de la misma molécula. Dentro del subdirectorio del programa **APBS** existe el programa **Coulomb**, el cual calcula el ΔG_{Ele}^{Coul} sobre el archivo **.pqr**

El programa **APBS** resuelve en forma numérica la **Ecuación de Poisson-Boltzmann (EPB)**, en la cual la presencia de iones en el sistema se encuentra descrita mediante una distribución de Boltzmann. La **EPB** se muestra a continuación:

$$-\nabla \cdot \varepsilon(x) \nabla \phi(x) + \kappa^2(x) \sinh \phi(x) = f(x) \quad [2]$$

Esta ecuación relaciona el potencial electrostático ($\phi(x)$), dependiente de la posición (x), con las propiedades dieléctricas de la proteína y el solvente ($\varepsilon(x)$), la fuerza iónica de la solución y la accesibilidad de los iones en el interior de la proteína (κ^2), y la distribución de las cargas parciales atómicas de la misma ($f(x)$). Para facilitar la resolución de esta ecuación no lineal, en ocasiones se aproxima a una forma linealizada asumiendo que el $\sinh \phi(x) \approx \phi(x)$; ambas formas de la ecuación pueden ser resueltas por algoritmos de elementos de frontera, elementos finitos y diferencias finitas³³.

La solución de la Ecuación de Poisson-Boltzmann (**EPB**) proporciona el potencial electrostático generado por la distribución de cargas atómicas de la macromolécula inmersa en un medio con una determinada constante dieléctrica. A partir de este potencial, el programa calcula el valor de la energía libre de solvatación electrostática (ΔG_{Ele}^{solv}) para la macromolécula (APBS Tutorial 2007)³⁴. La contribución electrostática a la energía libre de solvatación está dada por:

$$\Delta G_{solv}^{ele} = \frac{1}{2} \sum_i q_i [\phi(x_i)_{soluta} - \phi(x_i)_{vacio}] \quad [3]$$

Donde: q_i es la carga parcial de cada átomo, $\phi(x_i)_{\text{solute}}$ y $\phi(x_i)_{\text{vacio}}$ son los potenciales electrostáticos calculados para la distribución de cargas de la proteína en solución y de la proteína en el vacío, donde no hay límites dieléctricos, respectivamente³⁴.

El programa **APBS**, realiza estos cálculos considerando un modelo de solvente continuo conocido también como, de solvente implícito o de dieléctrica continua. En este modelo el soluto, que corresponde a la proteína, se detalla a nivel atómico, mientras que las moléculas de solvente y los posibles electrolitos disueltos en él, se tratan como un continuo sin estructura caracterizado por la dieléctrica del solvente³⁴⁻⁴⁰.

Comúnmente, tanto la proteína como la fase líquida (agua) son representados como materiales homogéneos. En estudios de dinámica molecular de proteínas y en cálculos electrostáticos, los valores de la constante dieléctrica más frecuentemente utilizados son de 4.0 para el soluto y 80 para el solvente³⁷.

El límite dieléctrico entre la proteína y el solvente, es sin duda la superficie molecular de ella; este depende tanto del radio atómico como del tamaño del radio esférico del solvente de prueba empleado^{33,37,38}.

Como se mencionó anteriormente, el programa **APBS** resuelve la **EPB** de dos formas, la linealizada (**LPBE**) y la no linealizada (**NLPBE**). La **EPB** puede expresarse de 3 maneras diferentes dependiendo de la fuerza iónica empleada^{33,35}.

- (1) En ausencia de atmósfera iónica, la ecuación de Poisson-Boltzmann, se transforma en la ecuación de Poisson.
- (2) Para fuerzas iónicas pequeñas, se utiliza la expresión lineal de la ecuación de Poisson-Boltzmann.
- (3) Para fuerzas iónicas grandes se requiere la expresión **PB** no lineal.

Se ha demostrado que ambas soluciones (**NLPBE** y **LPBE**) son cercanas para fuerzas iónicas pequeñas, debido a que ambos tratamientos están ligados básicamente a la densidad de carga en la interface proteína-agua^{33,35}.

El programa **APBS** resuelve ambas formas de la ecuación **PB** mediante el método numérico iterativo de diferencias finitas (**FDPB**). Esta aproximación se lleva a cabo utilizando una malla fina cúbica, que cubre toda la superficie de la proteína y un cierto volumen de solvente circundante a ella (ver **Fig. 7**).

El potencial electrostático se determina para cada punto de la malla como función de los potenciales en puntos contiguos de la misma, repitiéndose este proceso hasta la convergencia. Los resultados son muy sensibles al tamaño de malla utilizada, a menudo se emplean mallas muy finas, lo cual incrementa el costo computacional del cálculo.

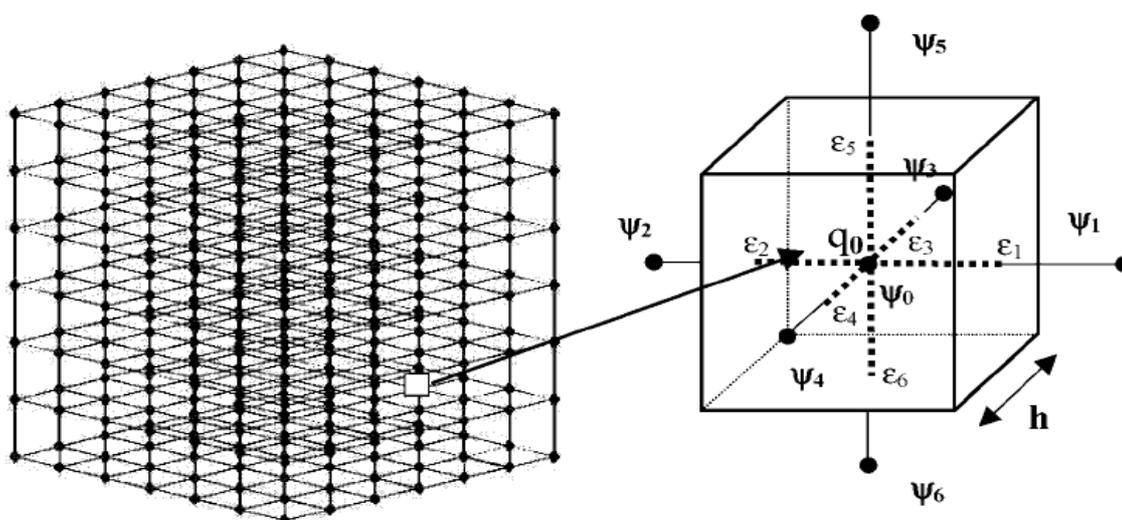


Figura 7. Esquema de la malla utilizada en el proceso de solución de la ecuación de Poisson-Boltzmann mediante el método de diferencias finitas (**FDPB**). En la parte derecha se encuentra una ampliación de un punto de la malla, en donde se indica las variables distribuidas en el espacio que intervienen en el cálculo; Ψ_i es el potencial electrostático, q es la carga, ϵ_i es la constante dieléctrica, h es la arista del cubo³⁵.

Alternativamente a la solución de la ecuación de **PB** para la obtención del potencial electrostático y la energía libre de solvatación electrostática (ΔG_{solv}^{ele}), el programa **APBS** tiene la capacidad de graficar este potencial sobre

la superficie de la proteína, mostrando en color rojo las zonas de potencial electrostático negativo, en azul las zonas positivas, y en tonos de gris las zonas con potencial electrostático intermedio^{34,35}.

En el presente trabajo se evaluaron energéticamente las 17 mutantes individuales creadas *in silico* en la β gl-B, empleando la forma no linealizada de la ecuación de Poisson-Boltzmann (**NLPBE**) en el programa **APBS** que está diseñado para trabajar en ambiente **LINUX**. Las especificaciones para tal evaluación, se detallaron en el apartado anterior (2.8.1).

2.8.3 Búsqueda de posibles mutaciones múltiples en β gl-B.

Para la búsqueda de posibles mutaciones múltiples que garantizaran aún más el incremento de la termorresistencia molecular en β gl-B, se utilizaron los archivos de cada una de las 17 mutantes individuales generadas en **MOE**. La creación de las mutantes múltiples se realizó empleando tres criterios de diseño independientes:

1. A partir de un primer residuo mutado, se seleccionaron los residuos comprendidos en una esfera de 10 Å, buscando al menos una de las restantes mutaciones individuales realizadas. Así se encontraron 10 mutaciones dobles, 2 triples, e incluso una mutante cuádruple, las cuales se muestran en las **Tablas 5 y 6**. A continuación se procedió a generar cada una de estas mutaciones de forma similar a las anteriores en el programa **MOE**, haciendo los cambios de los residuos de forma simultánea. El proceso de minimización de energía, se realizó de forma análoga por etapas empleando el campo de fuerzas CHARMM 27; permitiendo en un principio solo la movilidad de los residuos mutados, en la siguiente etapa se permitió la movilidad de estos junto con sus vecinos a 4.5 Å, posteriormente se permitió la movilidad de toda la molécula. A continuación se generaron también los archivos pdb para cada mutante múltiple creada, se convirtieron los archivos a formato **pqr**, de la forma descrita anteriormente empleando el servidor **PDB2PQR** y se evaluaron energéticamente con el programa **APBS**.

2. Basándose en los resultados energéticos obtenidos por el programa **APBS** para las mutantes individuales, se seleccionaron las que presentaron menor energía electrostática (3 mutantes), y se realizó la combinación entre ellas para diseñar 3 dobles mutantes y una triple (**Tabla 7**), las cuales se generaron también con el programa **MOE**, cambiando los residuos involucrados simultáneamente, y minimizando cada molécula de la manera descrita anteriormente. Se generaron los archivos pdb, y se sometieron al mismo proceso de edición que crea la entrada del programa **APBS** para su evaluación energética.

3. En la revisión bibliográfica que se realizó al inicio de este proyecto, se encontró que es factible generar mutaciones en asas externas de la proteína, las cuales podrían estabilizarla térmicamente⁷. Con base en esta información se eligió a la mutación A91R, que se encuentra en un asa gigante de β gl-B, para generar todas las mutantes dobles posibles, haciendo la combinatoria con las 16 mutaciones individuales restantes, empleando para ello el programa **MOE**. Los archivos se evaluaron energéticamente con el programa **APBS**, bajo los mismos parámetros indicados en párrafos precedentes. Los resultados de esta evaluación se muestran en la **Tabla 8**.

2.9 Comparación molecular *in silico* del sitio activo de las mutaciones generadas en β gl-B.

Para determinar si las mutaciones construidas *in silico* habían modificado la estructura del sitio activo de una manera significativa, se propuso evaluar y comparar, la unión de un mismo ligando común en el sitio activo, de cada una de las 17 mutaciones generadas junto con la β gl-B silvestre. Para ello se utilizó la estructura cristalográfica de la β gl-B que contiene una molécula de glucosa en el sitio activo. Esta enzima se encuentra depositada en el PDB con código **2O9T**. El cálculo involucró la evaluación de la energía de interacción de los dos residuos catalíticos (**E167** y **E356**) con la glucosa, a partir de la superposición por pares de la estructura cristalográfica (**2O9T**) con cada una de las mutantes creadas, siguiendo el procedimiento que se describe a continuación.

En una primera etapa, se realizó la superposición de ambas estructuras, eliminando todas las moléculas de agua presentes en los archivos. Posteriormente se seleccionaron los residuos catalíticos, tanto de la **2O9T** como de la mutante en turno, y se identificaron los vecinos ubicados a 7Å para superponer las dos estructuras, incluyendo a la glucosa. Se observó que esta distancia garantizaba la cobertura tanto del barril de hojas β , donde se encuentra el sitio activo, como un número considerable de residuos perimetrales a él.

Una vez realizada esta superposición, se suprimió la estructura de la **2O9T**, quedando en la pantalla gráfica de **MOE**, la mutante bajo estudio junto con la molécula de glucosa. A continuación se procedió a minimizar energéticamente a estas moléculas superpuestas por etapas, con la finalidad de alcanzar el estado más relajado para todos sus átomos constituyentes.

Inicialmente se asignaron cargas eléctricas empleando el campo de fuerzas **CHARMM 27**, posteriormente se permitió la movilidad de la glucosa, manteniendo rígida la estructura de variante de β gl-B mutada individualmente. Enseguida se seleccionaron los dos residuos catalíticos, la glucosa y sus vecinos próximos a 4.5 Å de distancia. Nuevamente se realizó la minimización de la energía permitiendo la movilidad de los átomos correspondientes a esta selección, mientras el resto de la estructura se mantenía rígida. Por último se permitió la movilidad de todos los átomos de ambas moléculas minimizando una vez más.

Finalmente se determinaron los enlaces de hidrógeno formados entre la glucosa y sus vecinos inmediatos a una distancia esférica de 4.5 Å; para la β gl-B, la mutante termorresistente H62R, cada variante individual y aquellas variantes múltiples que presentaron una menor energía libre de solvatación; empleando el programa **MOE**. En la **Tabla 13**, se encuentran reportadas tanto las distancias de los enlaces de hidrógeno formados, como los residuos involucrados en ellos.

Cabe mencionar que esta evaluación energética de interacción no es propiamente un *docking*. Sin embargo, este proceso de comparación nos permite visualizar el acomodo más relajado del ligando con respecto a los residuos catalíticos, así como la determinación cuantitativa de la energía de interacción entre la glucosa y los residuos catalíticos de la mutante. Los resultados de esta evaluación aplicada tanto a β gl-B como a las mutantes individuales, mutantes múltiples de menor energía de acuerdo a los cálculos del programa APBS, y de las mutantes obtenidas mediante mutagénesis al azar²⁵, se muestran en las **Tablas 9-11**.

El método que **MOE** utiliza para realizar la superposición estructural de proteínas se fundamenta en la similitud de las cadenas polipeptídicas de las moléculas comparadas, dado que el primer paso en este proceso es el alineamiento de las estructuras correspondientes. La superposición tridimensional de estructuras de proteínas permite la identificación de las regiones de mayor cambio conformacional que sufre la proteína mutada respecto a la silvestre.

En el caso en el que dos moléculas no enlazadas covalentemente entre sí se encuentren interactuando estrechamente, como ocurre con una proteína y un ligando, la energía potencial de interacción (E_{int}) calculada por **MOE**, solo considera a la energía de van der Waals (E_{vdw}) y a la energía electrostática (E_{ele}). Lo anterior es debido a que los términos energéticos restantes que integran la Ec. 1 (sección 2.7), involucran la información detallada del tipo de interacciones atómicas que se encuentran unidas covalentemente en la proteína. Por lo tanto, la energía potencial de interacción entre una proteína su un ligando, está dada por la Ec. 4.

$$E_{int} = E_{vdw} + E_{ele} \quad [4]$$

O bien:

$$E_{int} = \sum_{i=1}^{ligando} \sum_{j=1}^{receptor} \left[\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \left(\frac{q_i q_j}{\epsilon r_{ij}} \right) \right] \quad [5]$$

Donde: A_{ij} y B_{ij} son los parámetros de van der Waals para cada par de átomos interactuantes, r_{ij}^{12} y r_{ij}^6 son las distancias entre átomos, q_i y q_j son las cargas de los átomos correspondientes, ϵ es la constante dieléctrica del sistema y r_{ij} es la distancia entre cargas.

Capítulo 3. Resultados.

En este capítulo se presentan los archivos de salida obtenidas de todos los servidores y programas utilizados en el presente trabajo. La discusión de los resultados se realiza en el siguiente capítulo.

3.1 La secuencia de aminoácidos de la β gl-B de *Paenibacillus polymyxa*.

A continuación se presenta en las **Figs. 8 y 9**, la secuencia de aminoácidos de la β gl-B de *Paenibacillus polymyxa* en código de una letra, agrupada en segmentos de 10 aminoácidos y en formato FASTA respectivamente. Ambas representaciones se encuentran en la base de datos **Swiss-Prot y TrEMBL**.

```
      10      20      30      40      50      60
MSENTFIFPÄ TFMWGTSTSS YQIEGGTDEG GRTPSIWDTF CQIPGKVIGG DCGDVACDHF

      70      80      90     100     110     120
HHFKEDVQLM KQLGFLHYRF SVAWPRIMPA AGIINEEGLL FYEHLLEIE LAGLIPMLTL

     130     140     150     160     170     180
YHWDLPQWIE DEGGWTQRET IQHFKTYASV IMDRFGERIN WWNTINEPYC ASILGYGTGE

     190     200     210     220     230     240
HAPGHENWRĒ AFTAHHILM CHGIASNLHK EKGLTGKIGĪ TLNMEHVDAĀ SERPEDVAAA

     250     260     270     280     290     300
IRRDGFINRĀ FAEPLFNGKY PEDMVEWYGT YLNGLDVQPF GDMELIQQPG DFLGINYYTR

     310     320     330     340     350     360
SIIRSTNDAS LLQVEQVHME EPVTDMGWEI HPESFYKLLT RIEKDFSKGL PILITENGAA

     370     380     390     400     410     420
MRDELVNGQĪ EDTGRHGYIĒ EHLKACHRFĪ EGGQLKGYF VVSFLDNFEW AWGYSKRFGĪ

     430     440
VHINYETQER TPKQSALWFK QMMAKNGF
```

Figura 8. Secuencia de aminoácidos de la β gl-B de *Paenibacillus polymyxa*, agrupada en segmentos.

```
>2o9p_A mol:protein length:454 Beta-Glucosidase B From Paenibacillus  
Poly__MHHHHHSENTFIFPATFMWGTSTSSYQIEGGTDEGGRTPSIWDTFCQIPGKVIGGDCGDVA  
CDHFHFKEDVQLMKQLGFLHYRFSVAWPRIMPAAGIINEEGLLFYEHLLEIAGLIPMLTLYHWDL  
PQWIEDEGGWTQRETIQHFKTYASVIMDRFGERINWWNTINEPYCASILGYGTGEHAPGHENWREA  
FTAAHHILMCHGIASNLHKEKGLTGKIGITLNMEHVDAASERPEDVAAAIRRDGFINRWFAPLFGK  
YPEDMVEWYGTYLNGLDVQPGDMELIQQPGDFLGINYYTRSIIRSTNDASLLQVEQVHMEEPVTD  
MGWEIHPEFVKLLTRIEKDFSKGLPILITENGAAMRDELVNGQIEDTGRQRYIEEHLKACHRFIEEGGQ  
LKG YFVWSFLDNFEWAWGYSKRFGIVHINYETQERTPKQSALWFKQMMAKNGF
```

Figura 9. Secuencia de aminoácidos en formato FASTA, de la β gl-B de *Paenibacillus polymyxa*. Los residuos de histidina marcados en rojo se insertaron para facilitar la purificación de esta enzima, pero no aparecen en su estructura tridimensional.

3.2 Secuencias de las proteínas homólogas a β gl-B encontradas.

La Lista de las proteínas homólogas a β gl-B, obtenidas con el servidor del programa **BLAST2**, se presenta en la **Fig. 10**. Cada renglón amarillo corresponde a una proteína homóloga alineada con la β gl-B (línea verde). La tonalidad del color de cada renglón corresponde al porcentaje de identidad entre ambas secuencias, es decir entre la β gl-B y su respectiva homóloga; de acuerdo a la barra de colores que se ubica en la parte inferior de la figura. Los códigos que aparecen en el margen izquierdo de la lista de proteínas homólogas antes del guión, pertenecen a la clave de acceso a la base de datos **Swiss-Prot y TrEMBL** de cada proteína; mientras que, en algunos casos, los códigos que aparecen después del guión corresponden a la abreviatura del nombre del microorganismo del cual fue aislada dicha proteína. A la derecha se muestra en gris la porción de la cadena de aminoácidos que no presenta homología con la proteína de referencia, en este estudio, la β gl-B de *Paenibacillus polymyxa*.

En el **Anexo A**, se encuentran en formato FASTA todas las secuencias de las proteínas homólogas a β gl-B, mesófilas y termófilas junto con su clave de acceso a la base de datos de proteínas **Swiss-Prot y TrEMBL**.

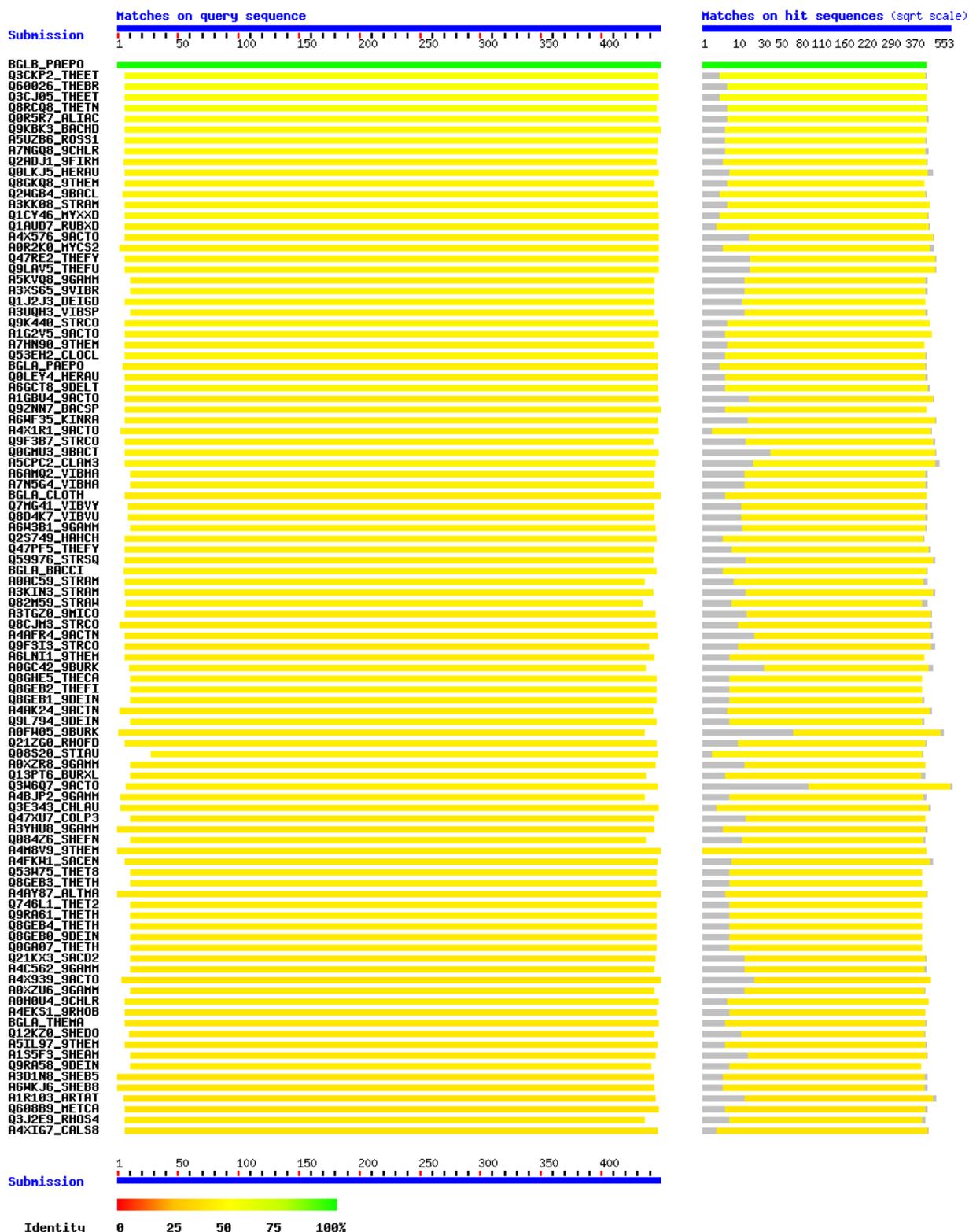


Figura 10. Lista de las proteínas homólogas a β gl-B con alrededor de un 50% de identidad, obtenidas con el servidor del programa **BLAST2**.

3.3 Clasificación de las proteínas homólogas a β gl-B con base en la temperatura óptima de crecimiento de los microorganismos que las contienen.

La clasificación de las proteínas homólogas a β gl-B, se realizó con base en la temperatura óptima ambiental de los microorganismos a partir de los cuales fueron aisladas (mesófilos y termófilos). La clasificación de estos microorganismos se realizó con base en la información contenida en el servidor **ENTREZ Genome Project**, obteniendo los resultados que se muestran a continuación en la **Tabla 1** y **Tabla 2**; en las cuales se muestra el hábitat y la temperatura óptima de crecimiento en °C, respectivamente.

Tabla 1. Microorganismos Mesófilos.

	MICROORGANISMO	HABITAT	TEMPERATURA ÓPTIMA DE CRECIMIENTO (°C)
1	<i>Paenibacillus polymyxa</i>	Terrestre	
2	<i>Agrobacterium tumefaciens</i>	Múltiple	25-28
3	<i>Bacillus halodurans</i>	Múltiple	
4	<i>Deinococcus geothermalis</i>	Acuático	47
5	<i>Herpetosiphon aurantiacus</i>	Múltiple	32
6	<i>Kineococcus radiotolerans</i>	Múltiple	32
7	<i>Magnetospirillum magneticum</i>	Acuático	
8	<i>Mycobacterium smegmatis str</i>	Asociado a un huésped	37
9	<i>Myxococcus xanthus</i>	Terrestre	20-30
10	<i>Sinorhizobium meliloti</i>	Múltiple	25-30
11	<i>Rhodobacter sphaeroides</i>	Múltiple	25-35
12	<i>Salinospira tropica</i>	Acuático	28
13	<i>Streptomyces sp</i>		
14	<i>Streptomyces ambofaciens</i>	Múltiple	25-35
15	<i>Streptomyces avermitilis</i>	Múltiple	25-35
16	<i>Burkholderia phytofirmans</i>	Terrestre	30
17	<i>Burkholderia phymatum</i>	Asociado a un huésped	
18	<i>Saccharophagus degradans</i>	Acuático	4-37
19	<i>Shewanella baltica</i>	Acuático	
20	<i>Streptomyces coelicolor</i>	Múltiple	25-35
21	<i>Vibrio vulnificus</i>	Acuático	20-30
22	<i>Hahella chejuensis</i>	Acuático	
23	<i>Stigmatella aurantica</i>	Terrestre	
24	<i>Rhodoferax ferrireducens</i>	Múltiple	25
25	<i>Frankia sp.</i>	Múltiple	
26	<i>Shewanella frigidimarina</i>	Múltiple	20-22
27	<i>Shewanella denitrificans</i>	Acuático	20-25
28	<i>Shewanella amazonensis</i>	Múltiple	37
29	<i>Bifidobacterium breve</i>	Asociado a un huésped	
30	<i>Rhizobium leguminosarum</i>	Asociado a un huésped	25-30
31	<i>Nocardioiodes sp.</i>	Terrestre	30
32	<i>Roseovarius sp.</i>	Acuático	
33	<i>Shewanella woodyi</i>	Múltiple	25

Tabla 2. Microorganismos Termófilos.

	MICROORGANISMO	HABITAT	TEMPERATURA ÓPTIMA DE CRECIMIENTO (°C)
1	<i>Acidothermus cellulolyticus</i>	Acuático	58
2	<i>Caldicellulosiruptor saccharolyticus</i>	Especializado	
3	<i>Chloroflexus aggregans</i>	Especializado	
4	<i>Chloroflexus aurantiacus</i>	Especializado	52-60
5	<i>Fervidobacterium nodosum</i>	Especializado	70
6	<i>Thermobifida fusca</i>	Múltiple	50-55
7	<i>Halothermothrix orenii</i>	Acuático	60
8	<i>Methylococcus capsulatus</i>	Múltiple	45
9	<i>Roseiflexus castenholziistrain</i>	Acuático	50
10	<i>Roseiflexus sp.</i>	Especializado	
11	<i>Rubrobacter xylanophilus</i>	Especializado	60
12	<i>Thermoanaerobacter ethanolicus</i>	Acuático	
13	<i>Thermoanaerobacter tengcongensis</i>	Especializado	75
14	<i>Thermoanaerobacter brockii</i>	Especializado	
15	<i>Thermomonospora fusca</i>	Especializado	
16	<i>Thermus nonproteolyticus</i>	Especializado	
17	<i>Thermus caldophilus</i>	Especializado	
18	<i>Thermus filiformis</i>	Especializado	
18	<i>Thermus thermophilus</i>	Especializado	
19	<i>Thermus sp.</i>	Especializado	
20	<i>Alicyclobacillus acidocaldarius</i>	Especializado	
21	<i>Bacillus circulans</i>	Especializado	
22	<i>Fervidobacterium sp.</i>	Especializado	
23	<i>Paenibacillus sp.</i>	Especializado	

NOTA: los microorganismos que no tienen temperatura óptima de crecimiento son organismos geotérmicos.

3.4 Alineamientos múltiples de las secuencias de proteínas homólogas a β gl-B.

El alineamiento múltiple de las secuencias de las proteínas homólogas a β gl-B y la obtención de las secuencias consenso para cada grupo (mesófilas y termófilas), se realizó empleando el servidor del programa **MultAlin**; ingresando sus secuencias en formato Fasta. Una parte de ambos alineamientos se muestra en las **Figs. 11** y **12**. La salida aparece marcada en colores; en rojo se presentan los residuos completamente conservados, en azul las posiciones conservadas, y en color negro se anotan los residuos poco conservados.

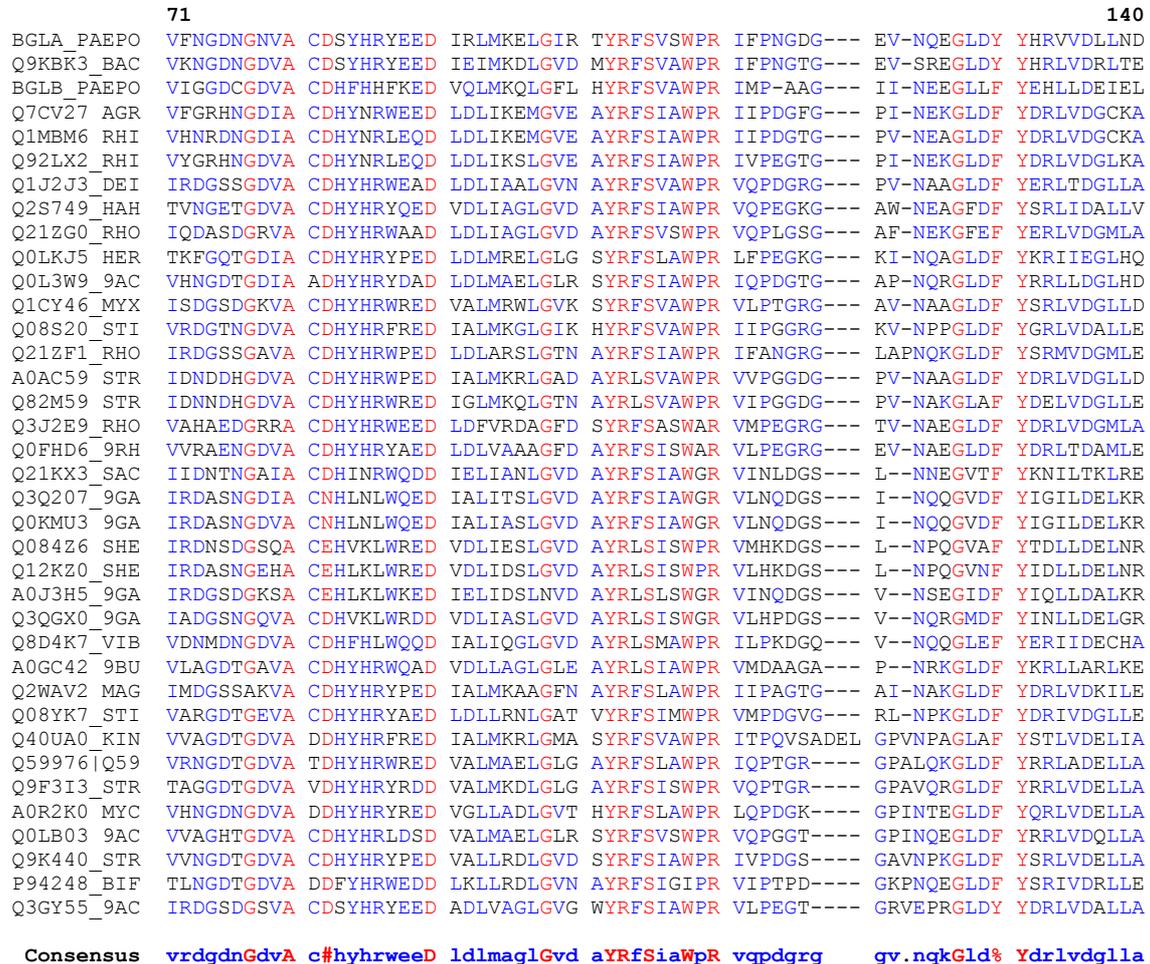


Figura 11. Parte del alineamiento múltiple de proteínas mesófilas homólogas a βgl-B generado por **MultAlin**.

NOTA: El Alineamiento múltiple completo de las proteínas mesófilas homólogas a βgl-B generado por **MultAlin** se muestra en el **ANEXO B**.

	71						140
A0LR48_ACI	ADHYHRYVGD	VRLMADLGV	SYRFSVAWPR	ILPSGSGAVN	RAGLDFYSRL	VDELLNHGIT	FALTTYHWDL
Q47RE2_THE	CDHYNRYRDD	VALMRELGVG	AYRFSIAWPR	IQPEGKGTVP	EAGLDFYDRL	VDCLLEAGIE	PWPTTYHWDL
Q9LAV5_THE	CDHYNRYRDD	VALMRELGVG	AYRFSIAWPR	IQPEGKGTVP	EAGLDFYDRL	VDCLLEAGIE	PWPTTYHWDL
Q1AUD7_RUB	CDHYHRLIED	LDLMARFGLG	AYRFSVAWPR	IQPEGRGPAN	RSGLDFYRRL	VEGLGERGIE	PVLTLYHWDL
Q2ADJ1_9FI	CDHYHLYRED	IELMKEIGIR	SYRFSVSWPR	ILPEGKGRVN	QKGLDFYKRL	VDNLLKANIR	PMITLYHWDL
Q3CKP2_THE	CDHYHRYKED	VELLKEIGVK	AYRFSIAWPR	IFPE-EGKYN	PKGMDFYKRL	IDELLKDIM	PTATLYHWDL
Q60026_THE	CDHYHRYKED	VELLKEIGVK	AYRFSIAWPR	IFPE-EGKYN	PKGMDFYKRL	IDELQKRDIV	PAATLYHWDL
Q8RCQ8_THE	CDHYHRYKED	VQIMKEIGVK	AYRFSIAWPR	IFPE-EGKYN	PKGMDFYKRL	VDELLKREII	PVATLYHWDL
Q0R5R7_ALI	CDHYHRYQDD	VRLMKELGIS	SYRFSIAWPR	VMPE-KGRVW	VKGLDFYKRL	ATALLEHGIR	PAATMYHWDL
A0ICG5_9CH	CDHYHRWRDD	IALMKSLGFP	AYRFSIAWPR	IMPKGRGAVN	PAGLDFYDRL	VDGLLAAGIR	PVVTLYHWDL
Q1FAN6_9CH	CDHYHRWRDD	ITLMKSLGLQ	AYRFSIAWPR	IIPQGRGQVN	PAGLDFYDRL	VDGLLDAGIR	PVVTLYHWDL
Q2WGB4_9BA	CDSYHRYEED	IELMKKLGIN	TYRFSIAWPR	IIPDGDGEIN	REGLDFYHRF	VDKLLAAGIE	PFCTLYHWDL
BGLA_BACCI	CDSYHRVEED	VQLKDLGVK	VYRFSISWPR	VLPQGTGEVN	RAGLDYHRL	VDELLANGIE	PFCTLYHWDL
A0HPV3_9TH	CDHYHRFKED	VALMKEIGLN	AYRFSISWPR	VMQDGN-IN	QKGLDFYDRL	VDELLKNDII	PFVTLYHWDL
Q8GKQ8_9TH	CDHYHRYKED	IQLMKEIGLD	AYRFSISWPR	IMPDGN-IN	QKGLDFYDRL	VDELLKNDII	PFVTLYHWDL
BGLA_THEME	CDHYNRWKED	IEIEKLGVK	AYRFSISWPR	ILPEGTGRVN	QKGLDFYDRL	IDTLLKAGIT	PFVTLYHWDL
BGLA_THENE	CDHYNRWKED	IEIEKIGAK	AYRFSISWPR	ILPEGTGKVN	QKGLDFYDRL	IDTLLKKNIT	PFVTLYHWDL
Q608B9_MET	CDHYRRFRED	VALMKALGLS	AYRFSIAWPR	IFPEKGRIN	WRGIAHYQAL	VETLLEHGIR	PMATLHHWDL
A0H0U4_9CH	CDHYHRYEED	LEHMAAMGLK	AYRFSIAWPR	IFPDGTGQPN	QRGLDFYRRL	IDGLHRRRIL	PVATLYHWDL
Q3E343_CHL	CDHYHRYEED	LDHMAALGLK	AYRFSIAWPR	IFPDGTGQPN	QRGLDFYRRL	IDGLHRRRIL	PVATLYHWDL
BGLS_CALSA	CDHYHRFEED	VSLMKELGLK	AYRFSIAWTR	IFPDGFGTVN	QKGLDFYDRL	INKLVENGIE	PVVTLYHWDL
Q2ZJP6_CAL	CDHYHRFEED	VSLMKELGLK	AYRFSIAWTR	IFPDGFGTVN	QKGLDFYDRL	INKLVENGIE	PVVTLYHWDL
Q8GEB4_THE	CDHYRRYEED	IALMQSLGVR	AYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLASGIT	PFLTLYHWDL
Q8GEB0_9DE	CDHYRRYEED	IALMQSLGVR	AYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLASGIT	PFLTLYHWDL
Q8GHE5_THE	CDHYHRYEED	IALMQSLGVG	VYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLAAGIT	PFLTLYHWDL
Q8GEB2_THE	CDHYHRYEED	IALMQSLGVG	VYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLAAGIT	PFLTLYHWDL
Q8GEB1_9DE	CDHYHRYEED	IALMQSLGVG	AYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLAAGIT	PFLTLYHWDL
Q9L794_9DE	CDHYHRYEED	IALMQSLGVG	VYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLAAGIT	PFLTLYHWDL
Q9RA58_9DE	CDHYRRYEED	IALMQSLGVG	AYRFSVAWPR	ILPEGRGRIN	PKGLAFYDRL	VDRLLASGIT	PFLTLYHWDL
Consensus CDHyHrye#D valmkelGvk aYRFS!aWpR !lPegkgrvn qkGldfYdrl !#eLleagIe PfLTlyHWDL							

Figura 12. Parte del alineamiento múltiple de proteínas termófilas homólogas a βgl-B generado por **MultAlin**.

NOTA: El Alineamiento múltiple completo de las proteínas termófilas homólogas a βgl-B generado por **MultAlin** se muestra en el **ANEXO B**.

3.4.1 Obtención de las secuencias consenso mesófila y termófila.

En la parte final de la salida arrojada por **MultAlin** de ambos alineamientos múltiples, se encuentra la *secuencia consenso* de los correspondientes grupos de proteínas homólogas empleadas en este trabajo. En la línea de la *secuencia consenso*, un residuo que es altamente conservado (más del 90%) se presenta en mayúsculas y en color rojo. Si más de un 90% de los caracteres de la columna, están conservados e integran el mismo grupo de conservación representado por un símbolo de homología: ! (IV), \$ (LM), % (FY), # (NDQE); el carácter colocado en esa posición de la *secuencia consenso* es el correspondiente símbolo en rojo.

Si el porcentaje de conservación es inferior al 90%, los caracteres son representados en minúsculas azules. Una posición que no contiene residuos conservados se representa por un punto, y las zonas de inserciones se representan por un espacio en blanco.

Ambas secuencias consenso se editaron manualmente, en algunas zonas en donde **MultAlin** colocó símbolos de homología; tomando decisiones objetivas preferenciales para colocar en esas posiciones aminoácidos cargados, polares, o bien a los que causaran menor impedimento estérico, según fuera el caso de los residuos representados por estos símbolos. Lo anterior debido a que los grupos con carga favorecen la formación de interacciones electrostáticas fundamentalmente, los polares la formación un mayor número de puentes de hidrógeno, y los residuos pequeños permiten un arreglo estructural más compacto. Estas secuencias consenso completas pueden observarse a detalle en la **Fig. 13**.

3.4.2 Alineamiento múltiple de las secuencias consenso mesófila y termófila, β gl-B-PAEPO y β gl-THETH.

Posteriormente se realizó un alineamiento múltiple de igual manera con programa **MultAlin**, de las dos secuencias consenso obtenidas junto con la β gl-B-PAEPO (*Paenibacillus polymyxa*) y la β gl-THET (*Thermus thermophilus*), con la finalidad de comparar posición a posición estas secuencias para observar sus diferencias principalmente en residuos cargados: R, K, H, E y D. Tal alineamiento se muestra en la **Fig.13**.

3.4.3 Variabilidad porcentual de la conservación de aminoácidos en las secuencias consenso mesófila y termófila.

Para tener una descripción gráfica y cuantitativa de la variabilidad y/o conservación de aminoácidos, tanto en la familia de las proteínas mesófilas como en la de proteínas termófilas homólogas a β gl-B, se calcularon los porcentajes de conservación para cada aminoácido en cada posición de los alineamientos múltiples, los cuales se marcaron por colores de acuerdo a los porcentajes de conservación clasificados por intervalos en ambas secuencias consenso. La **Fig. 14** reúne toda la información citada.



Figura 13. Alineamiento generado por MultAlin de las secuencias consenso de proteínas homólogas a βgl-B de organismos mesófilos y termófilos, βgl-B-PAEPO y βgl-B-THETH.

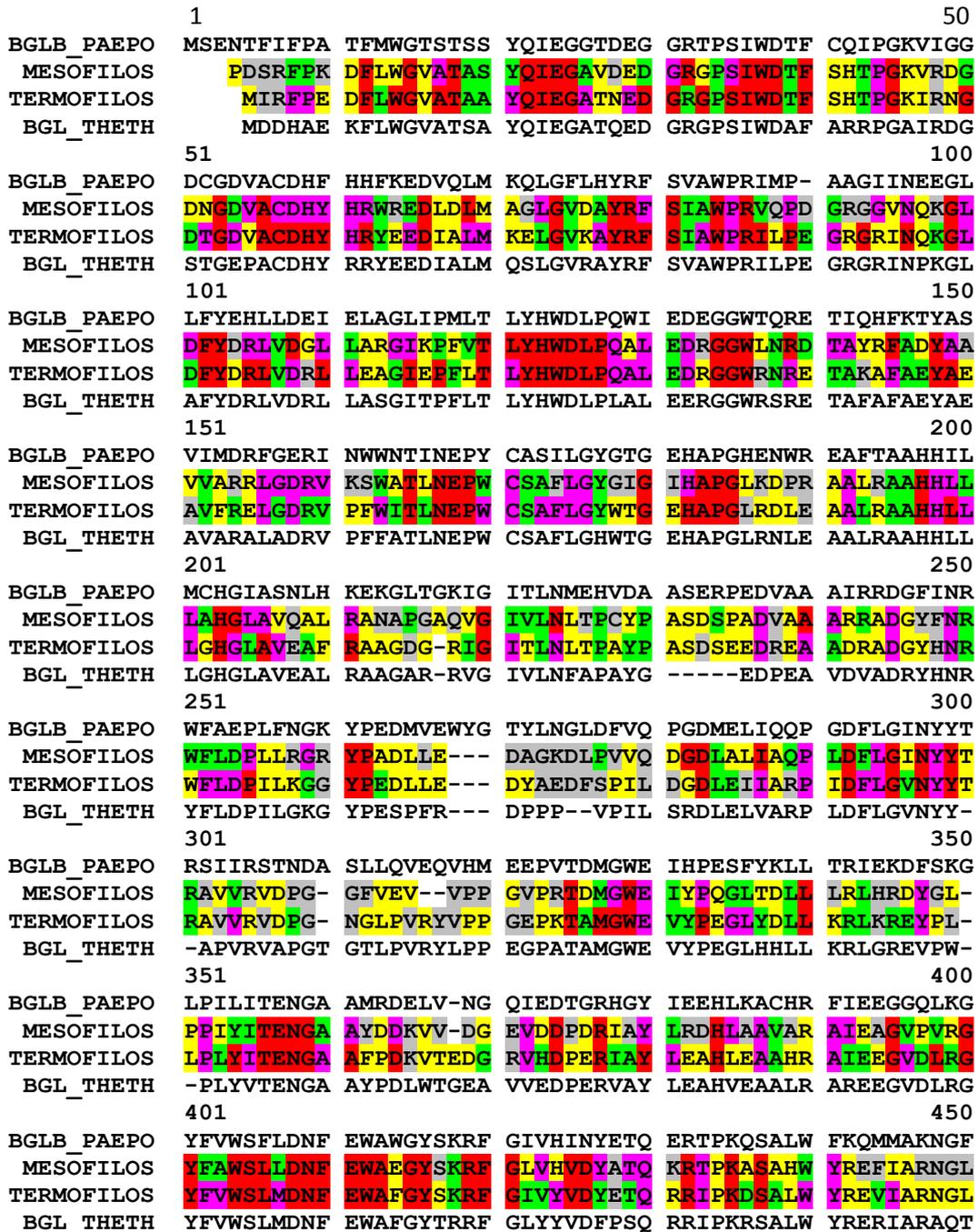


Figura 14. Alineamiento de las dos secuencias consenso de proteínas homólogas a β glB de organismos mesófilos y termófilos vs. la secuencia de β glB y β gl-THETH. En colores se representa el intervalo del % de conservación de cada aminoácido en las secuencias consenso:

0 <  < 30, 30 ≤  < 50, 50 ≤  < 70, 70 ≤  < 90, 90 ≤  ≤ 100.

3.5 Análisis comparativo del alineamiento múltiple de las dos secuencias consenso vs una superposición estructural.

Se realizó una superposición de las estructuras tridimensionales de la β gl-B-PAEPO y la β gl-THETH con la finalidad de verificar si el alineamiento de sus secuencias de aminoácidos era compatible tridimensionalmente. Para ello se utilizó el programa **MOE**, en el cual se editó el alineamiento de ambas secuencias (generado por **MultAlin**), colocando todas las inserciones asignadas por el servidor. Posteriormente se seleccionaron los residuos laterales a cada una de las zonas de inserción y se le pidió al programa que los resaltara en la pantalla gráfica de **MOE**, en ambas estructuras. Se observó que estas zonas se encontraban tanto en posiciones estructuralmente equivalentes, como en la misma orientación espacial. Lo antes mencionado puede observarse en la **Fig.15**.

3.6 Identificación de posibles mutaciones que confieran termoestabilidad a β gl-B.

Una vez corroborada la reproducibilidad del alineamiento de las secuencias consenso con la superposición estructural de la β gl-B-PAEPO y la β gl-THET. Se identificaron 55 posibles mutaciones termoestabilizantes de la β gl-B, con base en la localización de residuos cargados (E, D, R, K, H) presentes en al menos una de las secuencias consenso, que estuvieran ausentes en la β gl-B; haciendo un análisis posición a posición a lo largo de tal alineamiento. Todas estas mutaciones se listan en la **Tabla 3**.

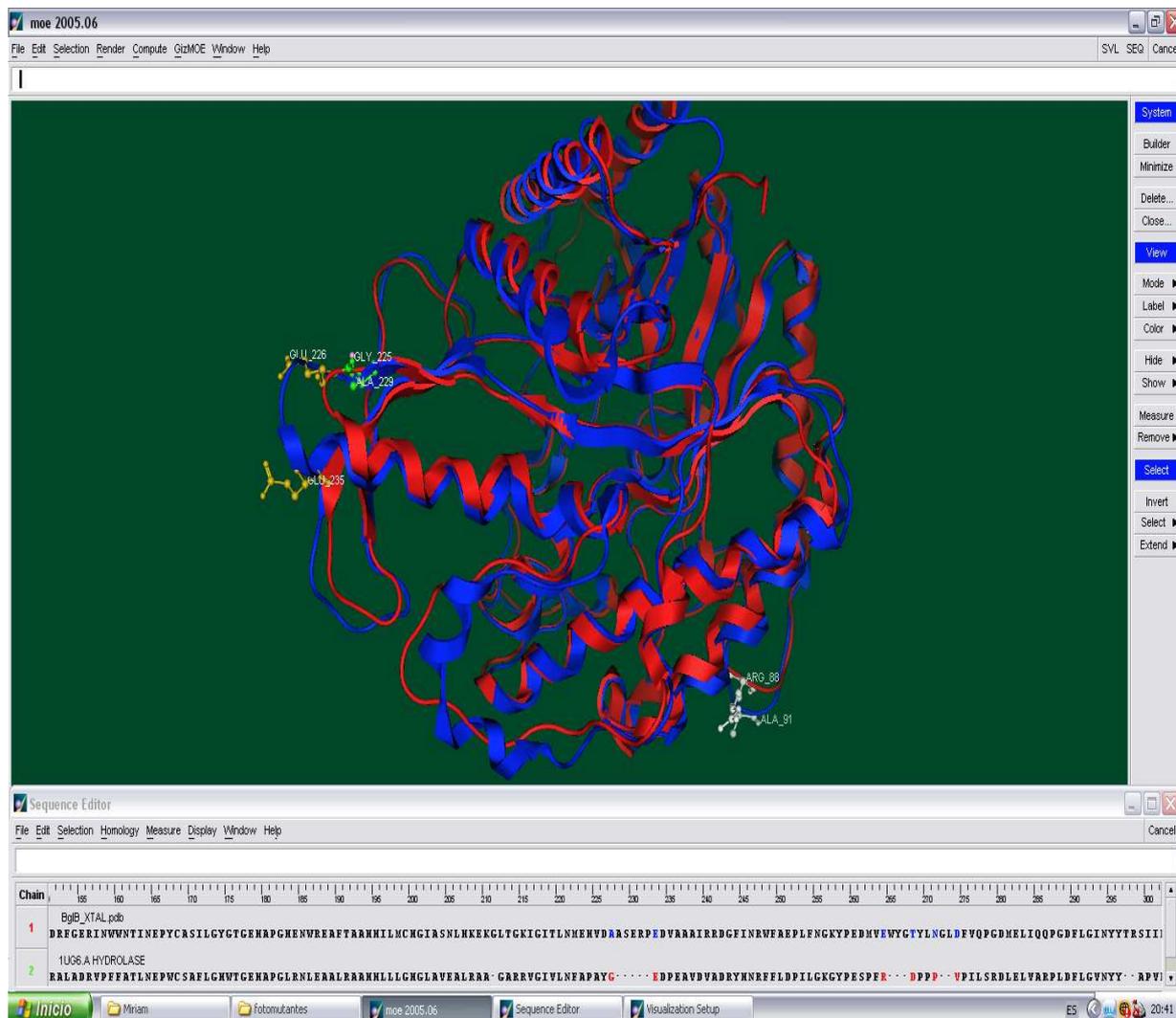


Figura 15. Superposición estructural generado por **MOE**, de la β gl-B-PAEPO (en azul) y la β gl-THET (en rojo). En la ventana inferior se muestra el alineamiento de ambas secuencias, la cadena 1 pertenece a β gl-B-PAEPO y la cadena 2 es la β gl-THET; se encuentran señalados en rojo a los residuos que flanquean cada inserción en la secuencia de la β gl-THET y en azul a los residuos que se encuentran en la misma posición en la secuencia de β gl-B.

3.7 Análisis tridimensional y creación de las posibles mutaciones identificadas en las dos secuencias consenso.

Las 55 mutaciones antes seleccionadas se construyeron sobre la estructura tridimensional de β gl-B empleando el programa **MOE**, para revisar el entorno inmediato a cada aminoácido nuevo y determinar su tendencia a formar interacciones iónicas importantes. Se encontró que, de los 55 sitios susceptibles de mutarse para conferir termorresistencia a β gl-B, que se presentan en la **Tabla 3**; sólo 17 de ellos podrían formar nuevas interacciones electrostáticas, al haber en su entorno inmediato residuos de carga opuesta, con los cuales se podían formar nuevas interacciones electrostáticas, a una distancia inferior a 7 Å entre los átomos cargados más cercanos de los residuos involucrados es estas.

Tabla 3. Mutaciones seleccionadas en las secuencias consenso revisadas tridimensionalmente

	Conservado en Mesófilos	Conservado en Termófilos	Conservado en ambas	Variable en ambas
1	T5D	Q72E	I7R	A10 K,E
2	G49D	I93R	T11D	L76D,K
3	Q68D	L111E	G30D	I115K,E
4	A112R	T136R	Q42H	R189R,E
5	N160K	Q142K	I48R	I241R,D
6	I247H	S149E	K64E	N273K,E
7	Q313E	R154E	A91R,K	Q369E,R
8	W412E	N207E	E97K	
9		L214E	L100D	
10		P234E	T146D,E	
11		V237R	D153R	
12		A238E	T193R	
13		Q288R	N257R,K	
14		T340K	T270D	
15		K384E	P280D	
16		Q395D	T306D	
17		Q334D	V323R,K	
18			E343H,K	
19			E364K	
20			N367D	
21			G374D,E	
22			N224D	
23			Q441E	

NOTA: Las mutaciones en color rojo, son las 17 que establecieron interacciones iónicas con cargas opuestas a menos de 7 Å. En la notación A10K,E de la última columna; la letra K corresponde al aminoácido consenso en mesófilos y la letra E al aminoácido consenso en termófilos. La numeración corresponde a la secuencia de β gl-B en el alineamiento de la **Fig. 12**.

En el alineamiento de la **Fig. 16**, se muestran las 17 posibles mutaciones encontradas y realizadas sobre la estructura de β gl-B, y en la **Fig. 17**, se muestran el o los residuos con los cuales interaccionan cada uno de los residuos mutados y sus distancias de interacción en Å, después de haber minimizado cada molécula de β gl-B mutada individualmente.

	1					50
BGLB_PAEPO	MSENTFIFPA	TFMWGTSTSS	YQIEGGTDEG	GRTPSIWDTF	CQIPGKVIIG	
MESOFILOS	PDSRFPK	DFLWGVATAS	YQIEGAVDED	GRGPSIWDTF	SHTPGKVRDG	
TERMOFILOS	MIRFPE	DFLWGVATAA	YQIEGATNED	GRGPSIWDTF	SHTPGKIRNG	
BGL_THETH	MDDHAE	KFLWGVATSA	YQIEGATQED	GRGPSIWDAF	ARRPGAIRDG	
	51					100
BGLB_PAEPO	DCGDVACDHF	HHFKEDVQLM	KOLGFLHYRF	SVAWPRIMP-	AAGIINEEGL	
MESOFILOS	DNGDVACDHY	HRWREDLDLM	AGLVDAYRF	SIAWPRVQPD	GRGGVNQKGL	
TERMOFILOS	DTGDVACDHY	HRYEEDIALM	KELGVKAYRF	SIAWPRILPE	GRGRINQKGL	
BGL_THETH	STGEPACDHY	RRYEEDIALM	QSLGVRAYRF	SVAWPRILPE	GRGRINPKGL	
	101					150
BGLB_PAEPO	LFYEHLLEI	ELAGLIPMLT	LYHWDLPQWI	EDEGGWTORE	TIOHFKTYAS	
MESOFILOS	DFYDRLVDGL	LARGIKPFVT	LYHWDLPQAL	EDRGGWLNRD	TAYRFADYAA	
TERMOFILOS	DFYDRLVDRL	LEAGIEPFLT	LYHWDLPQAL	EDRGGWNRRE	TAKAFAYAE	
BGL_THETH	AFYDRLVDRL	LASGITPFLT	LYHWDLPAL	EERGGWRSRE	TAFAFAYAE	
	151					200
BGLB_PAEPO	VIMDRFGERI	NWWNTINEPY	CASILGYGTG	EHAPGHENWR	EAFTHAAHIL	
MESOFILOS	VVARRLGDRV	KSWATLNEPW	CSAFLGYGIG	IHAPGLKDP	AALRAAHLL	
TERMOFILOS	AVFRELGDRV	PFWITLNEPW	CSAFLGYWTG	EHAPGLRDLE	AALRAAHLL	
BGL_THETH	AVARALADRV	PPFATLNEPW	CSAFLGHWTG	EHAPGLRNLE	AALRAAHLL	
	201					250
BGLB_PAEPO	MCHGIASNLH	KEKGLTGKIG	ITLNMHVDA	ASERPEDVAA	AARRDGFNIR	
MESOFILOS	LAHGLAVQAL	RANAPGAQVG	IVLNLTPCYP	ASDSPADVAA	ARRADGYFNR	
TERMOFILOS	LGHGLAVEAF	RAAGDG-RIG	ITLNLTPAYP	ASDSEEDREA	ADRADGYHNR	
BGL_THETH	LGHGLAVEAL	RAAGAR-RVG	IVLNFAPAYG	-----EDPEA	VDVADRYHNR	
	251					300
BGLB_PAEPO	WFAEPLFNGK	YPEDMVEWYG	TYLNGLDFVQ	PGDMELIQOP	GDFLGINYYT	
MESOFILOS	WFLDPLLGR	YPADLLE---	DAGKDLPVVQ	DGDLALIAQP	LDFLGINYYT	
TERMOFILOS	WFLDPILKGG	YPEDLLE---	DYAEDFSPII	DGDLEI IARP	IDFLGVNYYT	
BGL_THETH	YFLDPILKGG	YPESPFR---	DPPP--VPIL	SRDLEIVARP	LDFLGVNYY-	
	301					350
BGLB_PAEPO	RSIIRSTNDA	SLLQVEQVHM	EEPVTDMGWE	IHPESFYKLL	TRIEKDFSKG	
MESOFILOS	RAVVRVDPG-	GFVEV--VPP	GVPRTDMGWE	IYPQGLTDLL	LRLHRDYGL-	
TERMOFILOS	RAVVRVDPG-	NGLPVRYVPP	GEPKTAMGWE	VYPEGLYDLL	KRLKREYPL-	
BGL_THETH	-APVRVAPGT	GTLPVRYLPP	EGPATAMGWE	VYPEGLHLLL	KRLGREVPW-	
	351					400
BGLB_PAEPO	LPILITENGA	AMRDELV-NG	QIEDTGRHGY	IEEHLKACHR	FIEEGGOLKG	
MESOFILOS	PPIYITENGA	AYDDKVV-DG	EVDDPDRIAY	LRDHAAVAR	AIEAGVEVRG	
TERMOFILOS	LPLYITENGA	AFPDKVTEG	RVHDPERIAI	LEAHLEAAHR	AIEEGVDLRG	
BGL_THETH	-PLYVTENGA	AYPDLWTGEA	VVEDPERVAY	LEAHVEAALR	AREEGVDLRG	
	401					450
BGLB_PAEPO	YFVWSFLDNF	EWAWGYSKRF	GIVHINYETQ	ERTPKQSALW	FKOMMAKNGF	
MESOFILOS	YFAWSLLDNF	EWAEGYSKRF	GLVHVDYATQ	KRTPKASAHW	YREFIARNGL	
TERMOFILOS	YFVWSLMDNF	EWAFGYSKRF	GIVYVDYETQ	RRIPKDSALW	YREVIARNGL	
BGL_THETH	YFVWSLMDNF	EWAFGYTRRF	GLYYVDFPSQ	RRIPKRSALW	YRERIARAQT	

Figura 16. En este alineamiento se encuentran marcadas las diferencias de los residuos cargados conservados en: termófilos y la THET, en mesófilos y termófilos, en mesófilos, termófilos y la THET, los residuos conservados únicamente en termófilos.

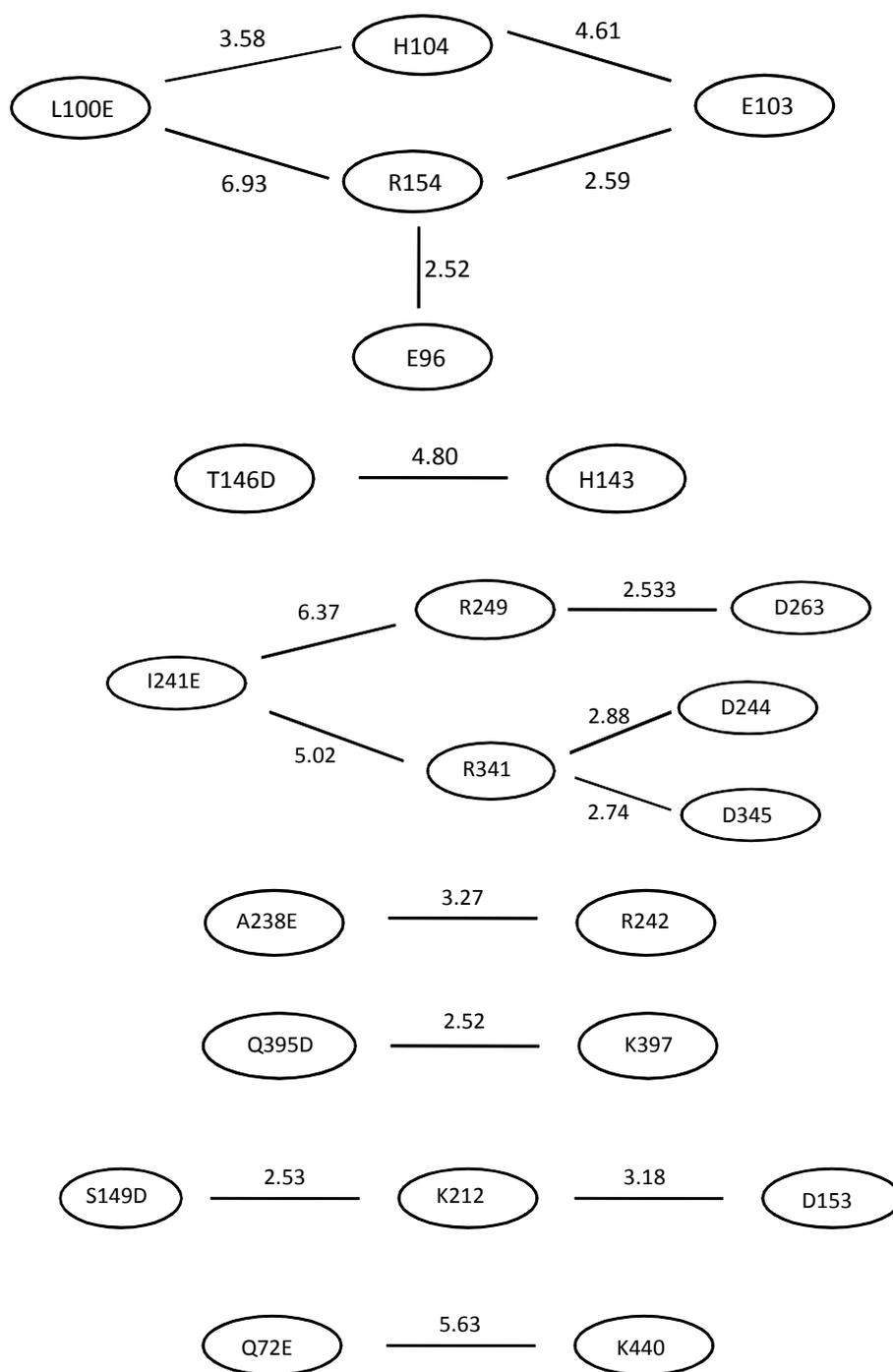


Figura 17. Representación esquemática de las 17 mutaciones realizadas en βgl-B y de los residuos involucrados en la formación de las nuevas interacciones electrostáticas formadas. El número que aparece arriba o abajo de las líneas, corresponde a la distancia de interacción en Å.

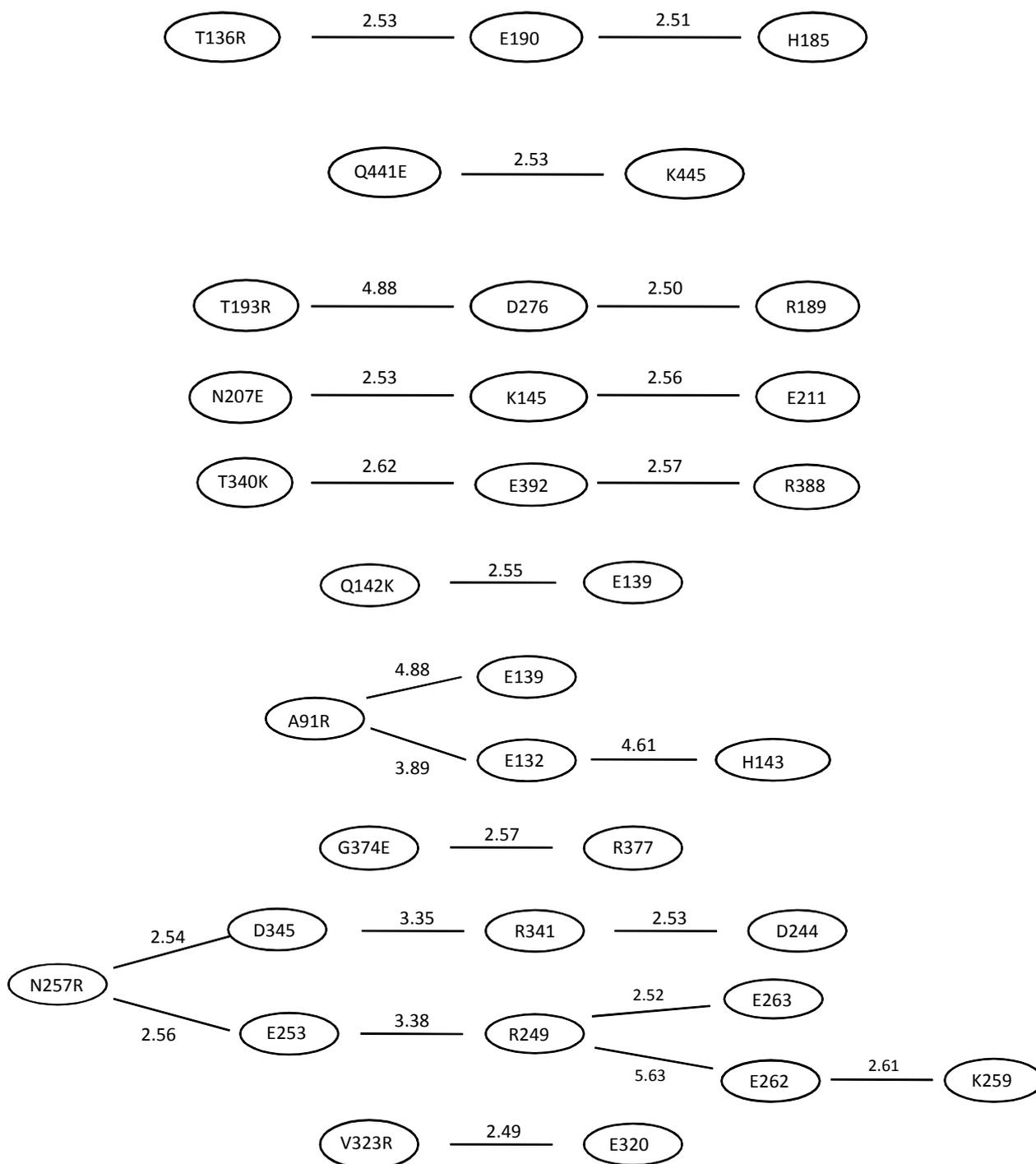


Figura 17. Representación esquemática de las 17 mutaciones realizadas en β gl-B y de los residuos involucrados en la formación de las nuevas interacciones electrostáticas, el número que aparece arriba o abajo de las líneas, corresponde a la distancia de interacción en Å.

Cabe mencionar que las 17 mutaciones individuales realizadas, se encuentran en zonas de baja conservación de aminoácidos, en ambas familias de proteínas homólogas mesófilas y termófilas, de acuerdo a la **Fig. 14**. La estructura secundaria de cada una de las variantes de β gl-B generadas en el programa **MOE**, se muestran a continuación en las **Figuras 18-34**, en ellas se muestra el residuo mutado en esferas y líneas, y los residuos con los que se establecieron interacciones electrostáticas en líneas; las distancias de interacción entre grupos con carga opuesta están en Å:

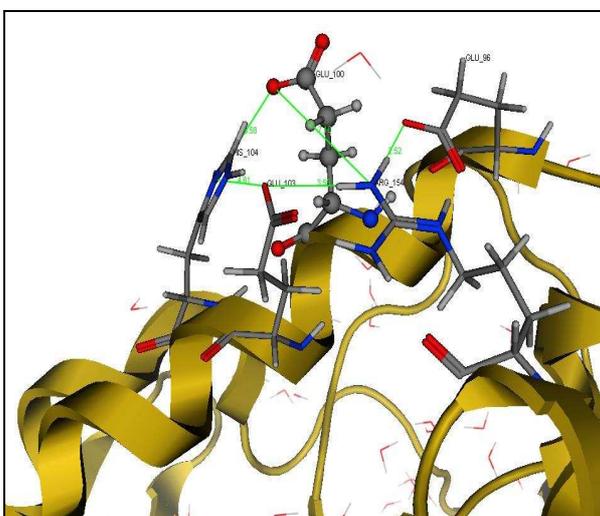


Figura 18. Mutación L100E

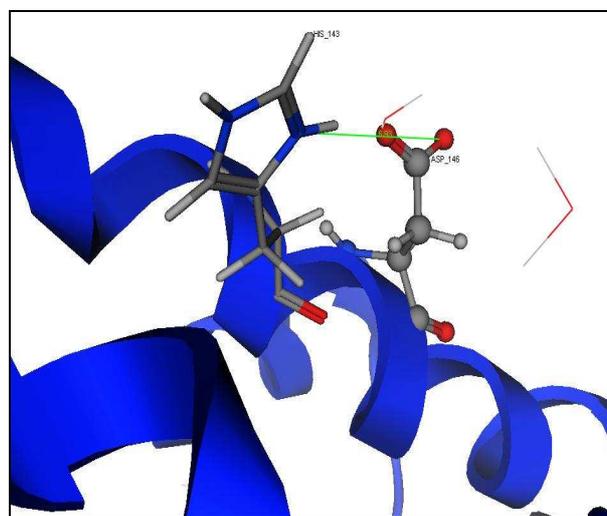


Figura 19. Mutación T146D

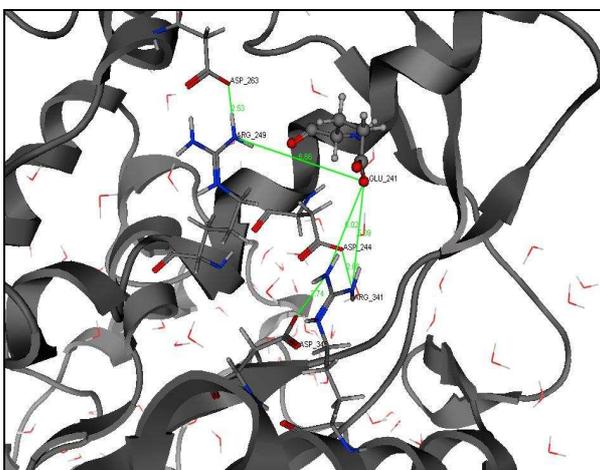


Figura 20. Mutación I241E

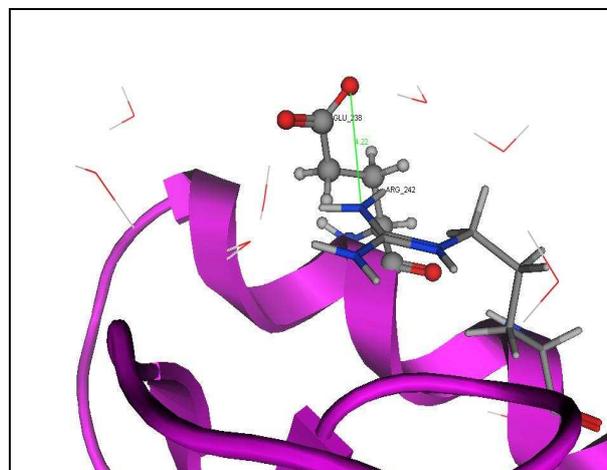


Figura 21. Mutación A238E

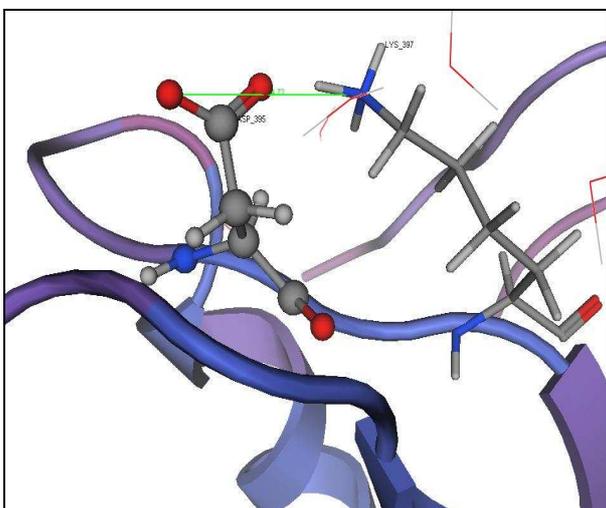


Figura 22. Mutación Q395D

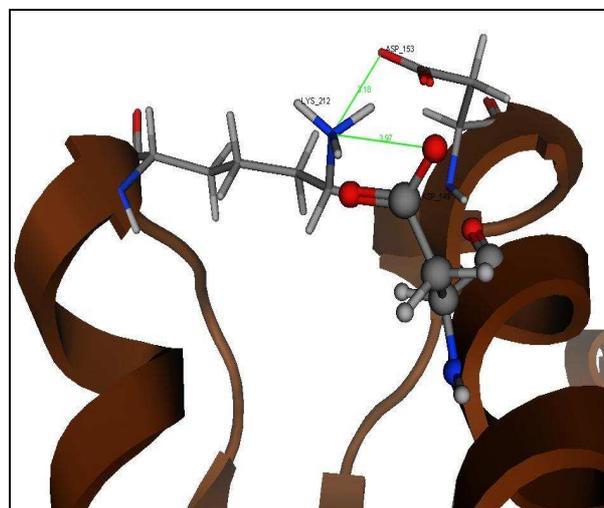


Figura 23. Mutación S149D

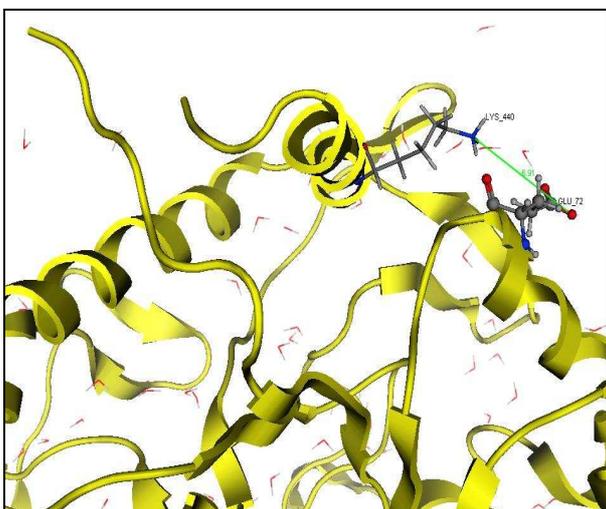


Figura 24. Mutación Q72E

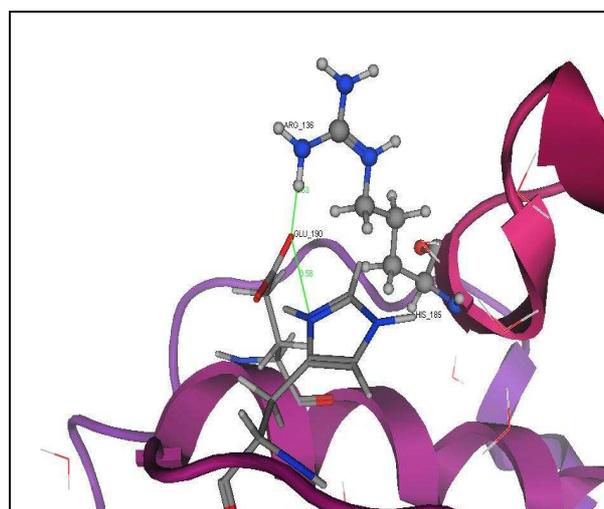


Figura 25. Mutación T136R

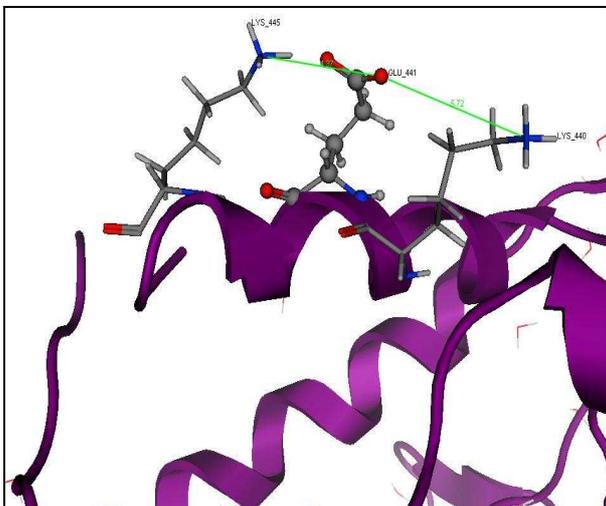


Figura 26. Mutación Q441E

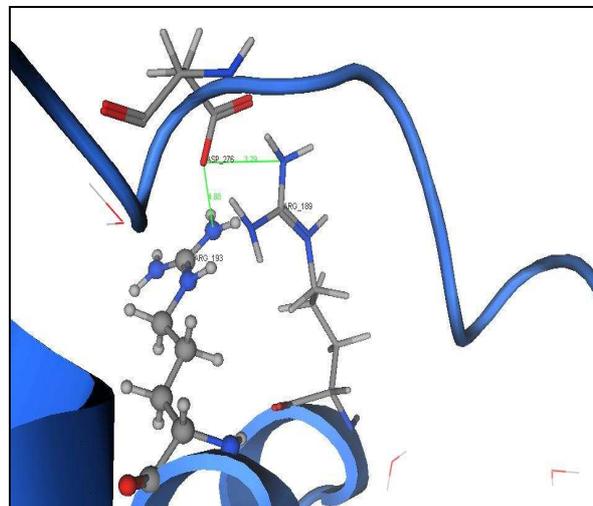


Figura 27. Mutación T193R

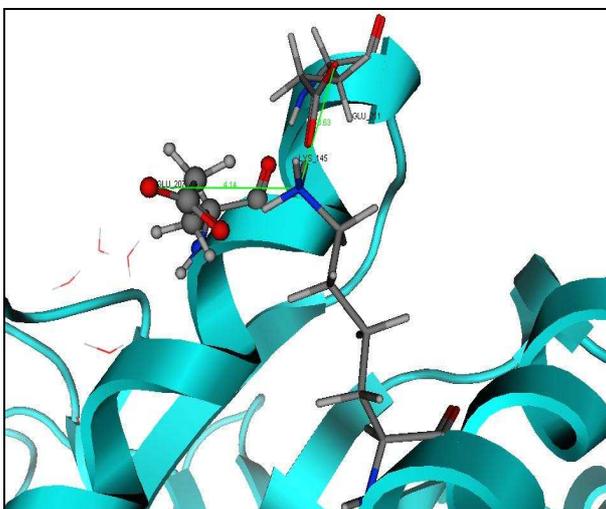


Figura 28. Mutación N207E

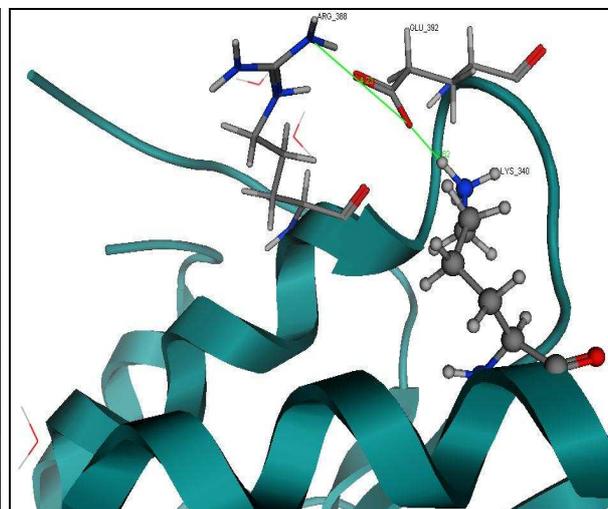


Figura 29. Mutación T340K

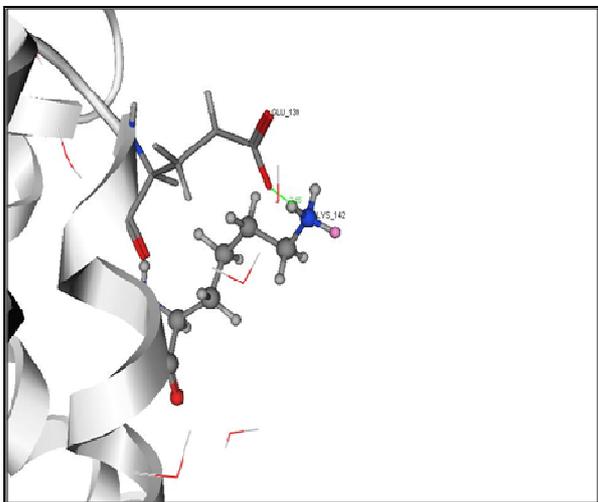


Figura 30. Mutación Q142K

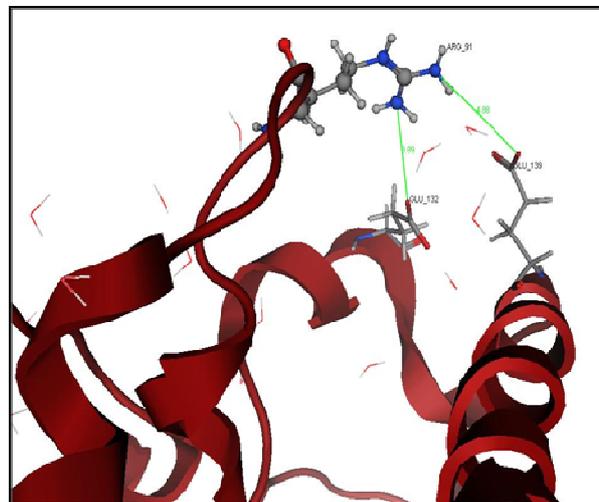


Figura 31. Mutación A91R

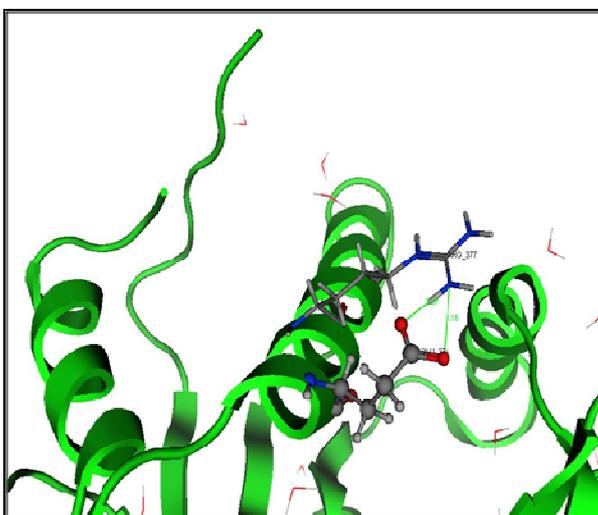


Figura 32. Mutación G374E

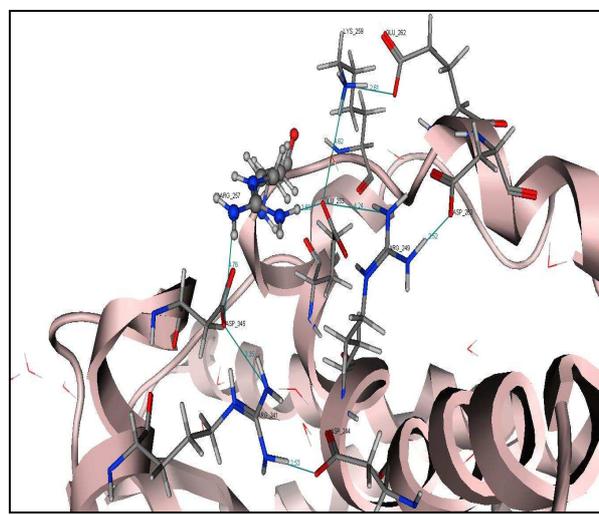


Figura 33. Mutación N257R

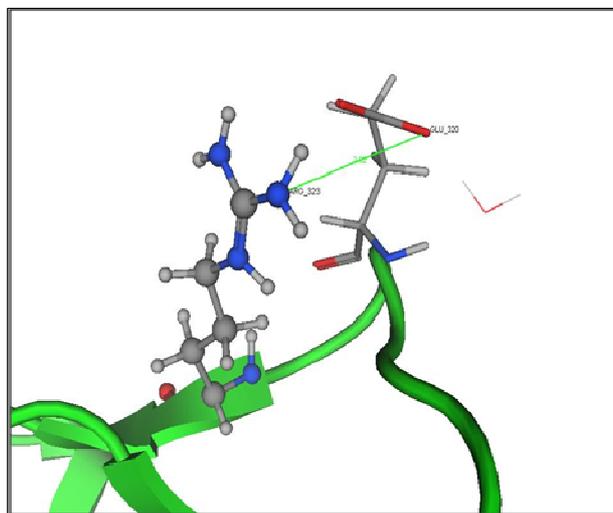


Figura 34. Mutación V323R

3.8 Determinación de las energías de solvatación de las mutantes generadas.

Para la evaluación energética de cada una de las 17 mutaciones individuales, así como de las mutantes dobles, triples y la cuádruple, obtenidas como se describió en el apartado 2.7 del presente trabajo; primeramente se transformaron cada uno de los archivos pdb a formato **.pqr**, estos archivos se editaron de la manera descrita en la sección 2.8, y se tomaron como entrada para evaluarlos en el programa **APBS**. El tiempo de evaluación para cada mutante osciló entre 7 y 8 min (en una computadora con procesador Intel (R) Pentium (R) M, Processor 2.00 GHZ, 797 MHZ, 1 GB RAM), se calcularon los valores de $\Delta G_{Ele\ Cou1}^2$ y ΔG_{soliv}^{ele} y con ambos él $\Delta G_{Ele\ Tot}^2$, además de la diferencia del mismo ($\Delta\Delta G_{Ele\ Tot}^2$) con respecto al valor obtenido para β gl-B para cada una de las variantes. Los resultados se muestran en las siguientes **Tablas 4 a 8**.

Tabla 4. Valores de la energía libre de Gibbs electrostática total de β gl-B, β gl-THET, la mutante termorresistente H62R, y de las mutantes individuales analizadas. En la última columna se comparan las energías de las diferentes variantes respecto a β gl-B ($\Delta\Delta G_{\text{Elec Total}} = \Delta G_{\text{Elec Total Variante } i} - \Delta G_{\text{Elec Total } \beta\text{gl-B}}$)

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
	β gl-B	-121357.6	0.00
	β gl-THET	-125533.5	-4175.90
	H62R	-121541.6	-184.00
1	A91R	-122208.15	-850.55
2	V323R	-122126.8	-769.20
3	T193R	-121934.85	-577.25
4	N257R	-121885.4	-527.80
5	T136R	-121850.7	-493.10
6	T340K	-121453.5	-95.90
7	Q142K	-121415.8	-58.20
8	G374E	-121286.45	71.15
9	A238E	-121260.85	96.75
10	Q395D	-121152	205.60
11	S149D	-121147.75	209.85
12	N207E	-121086.45	271.15
13	I241E	-120911.65	445.95
14	Q72E	-120902.85	454.75
15	Q441E	-120834.2	523.40
16	L100E	-120767.05	590.55
17	T146D	-120658.25	699.35

Tabla 5. Valores de la energía libre de Gibbs electrostática total y de la diferencia de la misma con respecto a β gl-B, de las mutantes dobles localizadas a menos de 10Å del primer residuo mutado.

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
1	T136R - T193R	-122699.75	-1342.15
2	A91R - Q142K	-122353.75	-996.15
3	V323R - G374E	-122185.65	-828.05
4	G258K - N257K	-121686.85	-329.25
5	N257R - I241E	-121684.64	-327.04
6	T340K - Q395D	-121614.55	-256.95
7	A238E - I241E	-121267.15	90.45
8	Q72E - Q441E	-121046.85	310.75
9	N257K - I241E	-121034.9	322.70
10	Q142K - T146D	-120876.3	481.30

Tabla 6. Valores de la energía libre de Gibbs electrostática total y de la diferencia de la misma con respecto a β gl-B, de las mutantes múltiples localizadas a menos de 10Å del primer residuo mutado.

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
1	S149D - Q142K - N207E	-121046.65	310.95
2	S149D - Q142K - T146D - N207E	-120901.2	456.4
3	S149D - T146D - N207E	-120347.1	1010.5

Tabla 7. Valores de la energía libre de Gibbs electrostática total y de la diferencia de la misma con respecto a β gl-B, de la combinatoria de las 3 mutantes individuales de más baja energía.

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
1	A91R-V323R	-123148.69	-1791.09
2	A91R-T193R	-122444.53	-1086.93
3	A91R-T193R-V323R	-122287.17	-929.57
4	V323R-T193R	-121371.94	-14.34

Tabla 8. Valores de la energía libre Gibbs electrostática total y de la diferencia de la misma con respecto a β gl-B, de la combinatoria de A91R con las 16 mutaciones individuales anteriores.

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
1	A91R-V323R	-123148.69	-1791.09
2	A91R-T136R	-122740.56	-1382.96
3	A91R-N257R	-122636.26	-1278.66
4	A91R-T193R	-122444.53	-1086.93
5	A91R-T340K	-122301.12	-943.52
6	A91R-A238E	-122277.61	-920.01
7	A91R-I241E	-122187.91	-830.31
8	A91R-Q72E	-122133.46	-775.86
9	A91R-S149D	-122071.79	-714.19
10	A91R-N207E	-122054.99	-697.39
11	A91R-Q142K	-121974.08	-616.48
12	A91R-Q395D	-121906.12	-548.52
13	A91R-T146D	-121805.66	-448.06
14	A91R-L100E	-121769.14	-411.54
15	A91R-G374E	-121672.84	-315.24
16	A91R-Q441E	-121644.13	-286.53

También se realizaron y se evaluaron, bajo los mismos parámetros que las demás variantes, las mutantes termorresistentes de β gl-B, obtenidas mediante mutagénesis al azar por Arrizubieta y Polaina²⁵. Todas estas mutaciones se reportan junto con sus valores de energía libre de solvatación, en la **Tabla 9**.

Tabla 9. Mutantes termorresistentes de β gl-B obtenidas mediante mutagénesis al azar por Arrizubieta y Polaina²⁵.

	VARIANTE	$\Delta G_{\text{Elec Total}}$ (kJ/mol)	$\Delta\Delta G_{\text{Elec Total}}$ (kJ/mol)
	β gl-B	-121357.6	0
1	H62R	-121541.6	-184
2	H62R-N223Y-M319I	-121306.75	50.85
3	M319I	-121085.65	271.95
4	N223Y	-120805.65	551.95

3.9 Comparación molecular *in silico* del sitio activo de las mutantes generadas y de la β gl-B.

Como se mencionó en la Introducción del presente trabajo, la β gl-B posee un barril de hojas β en el que se encuentran inmersos los residuos catalíticos (E167 y E356), una ilustración de ello se presenta en la **Fig. 35**.

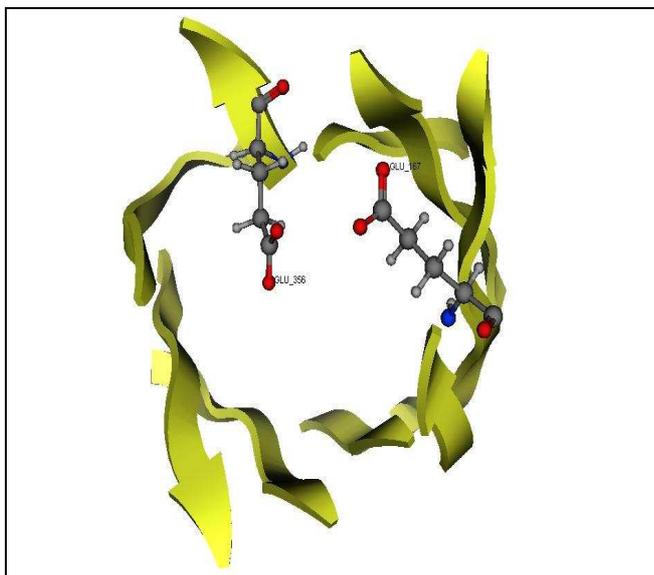


Figura 35. Representación del barril de hojas β y de los dos Glu catalíticos

La representación tridimensional de la β gl-B que fue cristalizada con una molécula de glucosa en el sitio activo (**2o9T**) se presenta en la **Fig. 36**, y en la **Fig. 37**, se presenta una ampliación del barril de hojas β con los dos residuos catalíticos y la glucosa en donde se puede apreciar la orientación de éstos.

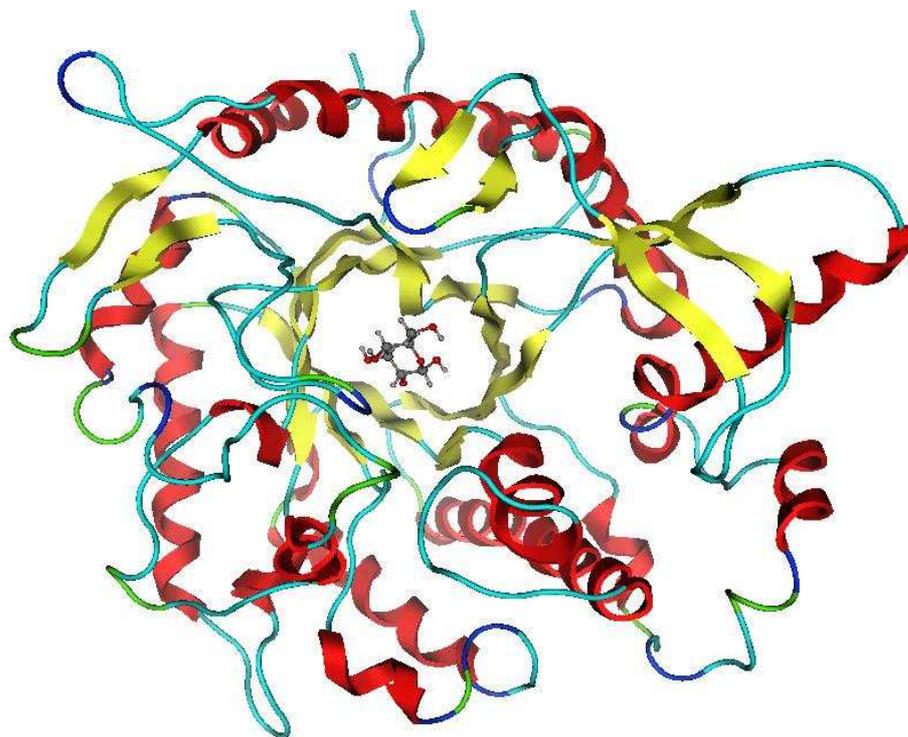


Figura 36. Representación tridimensional de la **2O9T** con una molécula de glucosa unida en el sitio activo.

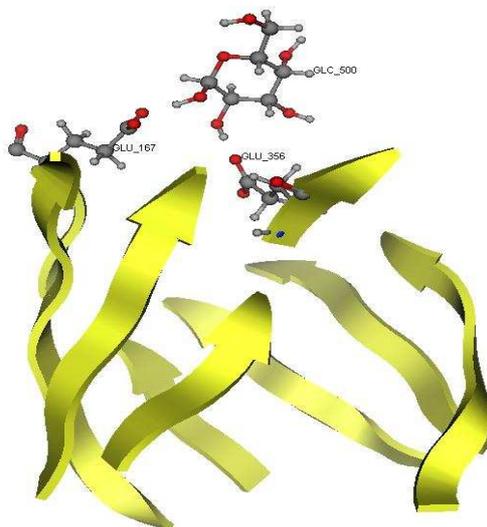


Figura 37. Representación del barril de hojas β con los dos residuos catalíticos y la glucosa en la **2O9T**.

La comparación *in silico* del sitio activo de cada una de las 17 mutaciones generadas y de la β gl-B, se realizó de igual manera en el programa **MOE**; con base en el cálculo de la energía de interacción de los dos residuos catalíticos (E167 y E356) con la glucosa. Para lo cual se realizó la superposición de la β gl-B **2O9T** con cada variante generada. Un ejemplo de tal superposición se muestra en la **Fig. 38**. Al final de este proceso, en la información arrojada por **MOE**, se encuentra la citada energía de interacción, estos datos se reportan para cada mutante individual y algunas mutantes múltiples que presentaron menor $\Delta\Delta G_{\text{Elec Total}}$, de acuerdo a los cálculos del programa **APBS**, y también para las mutantes obtenidas mediante mutagénesis al azar²⁵, en las **Tablas 10-12**, respectivamente.

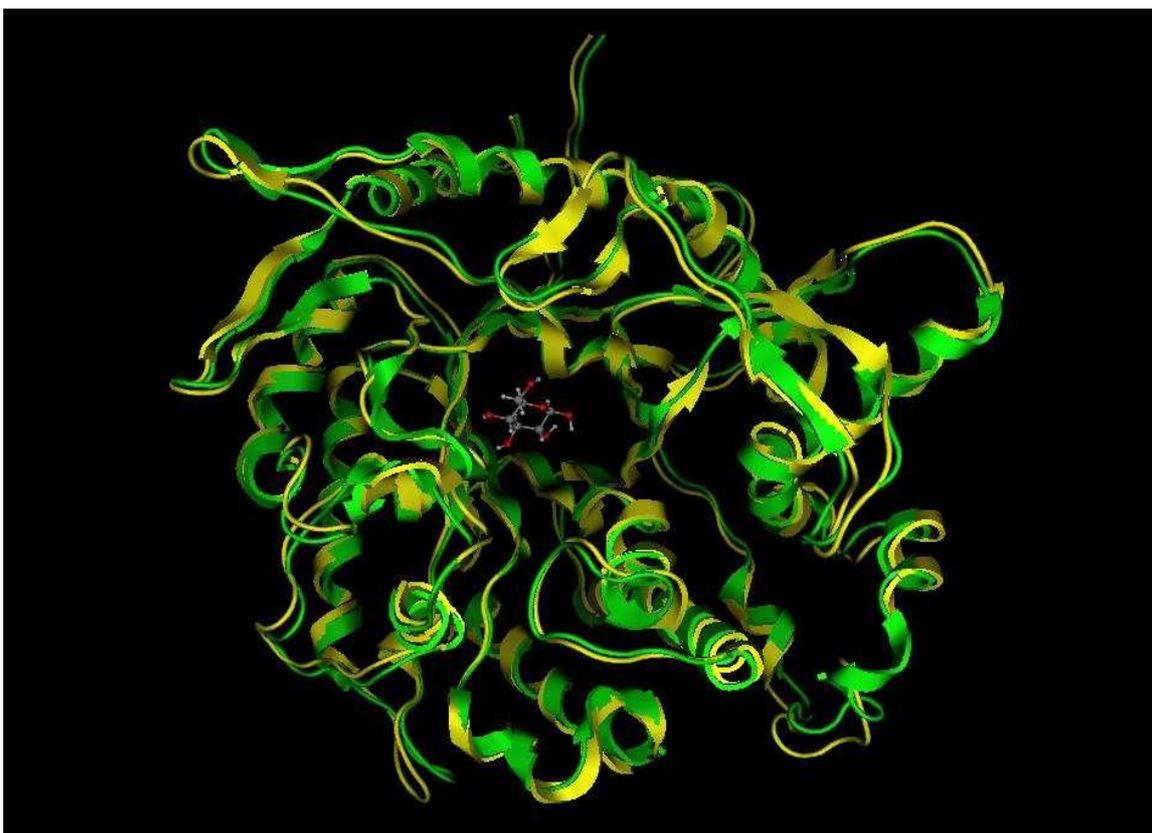


Figura 38. Superposición estructural de la β gl-B (verde) y 2O9T (amarillo).

Tabla 10. Cálculo de la energía de interacción con glucosa para la β gl-B, H62R, y mutantes individuales.

	VARIANTE	E_{int} (kcal/mol)
	β gl-B	-114.159
	H62R	-112.799
1	G374E	-123.586
2	Q72E	-122.314
3	T340K	-119.501
4	Q395D	-116.964
5	T136R	-116.587
6	Q142K	-115.826
7	S149D	-115.727
8	N207E	-115.286
9	V323R	-114.863
10	L100E	-114.829
11	I241E	-110.224
12	T146D	-110.099
13	N257R	-109.371
14	A91R	-108.416
15	T193R	-108.242
16	A238E	-106.269
17	Q441E	-106.227

NOTA: Las variantes que se encuentran en rojo son las que presentan valores lejanos de la energía de interacción con la glucosa a los presentados por la β gl-B y la H62R

Tabla 11. Cálculo de la energía de interacción con glucosa para las mutantes múltiples de menor $\Delta\Delta G_{Elec Total}$.

	VARIANTE	E_{int} (kcal/mol)
1	A91R-T193R-V323R	-115.07
2	A91R-T136R	-112.043
3	T136R - T193R	-111.478
4	A91R - Q142K	-110.409
5	A91R-N257R	-109.994
6	A91R-T193R	-109.894
7	T340K - Q395D	-109.813
8	V323R-T193R	-109.81
9	V323R - G374E	-109.508
10	A91R-V323R	-102.78
11	N257R - I241E	-97.805

Adicionalmente, se calculo la E_{int} con glucosa para las variantes termorresistentes probadas experimentalmente.

Tabla 12. Cálculo de la energía de interacción con glucosa para las mutantes obtenidas mediante mutagénesis al azar²⁵.

	Variante	E_{int} (kCal/mol)
	βgl-B	-114.159
1	H62R	-112.799
2	H62R-N223Y-M319I	-108.381
2	N223Y	-121.279
4	M319I	-112.905

A continuación se presentan los resultados de la exploración de los enlaces de hidrógeno formados después de la minimización de la energía de cada una de las variantes individuales superpuestas con glucosa.

Tabla 13. Distancias de los enlaces de hidrógeno formados entre la glucosa y sus residuos vecinos a 4.5 Å en cada variante.

8.5	Glu 167	Glu 356	Glu 409	Trp 402	Trp 410	Gln 22	Gln 22	Tyr298
βgl-B	2.52	2.5	2.47		2.85	2.88		
H62R	2.48	2.49	2.47	2.78	2.97	2.9		
L100E	2.53	2.49	2.47		2.79	2.85		
T146D	2.47	2.49	2.52		2.79	2.75	3.09	
I241E	2.51	2.49	2.47		2.86	2.79		
A238E	2.52	2.47	2.47	2.7	2.85	2.88		
Q359D	2.47	2.48	2.47	2.85	2.98	2.84		
S149D	2.5	2.47	2.48	2.91	2.84	2.89		
Q72E	2.51	2.45	2.46		2.86	3.29		
T136R	2.46	2.5	2.45	2.8	3.29	2.91		
Q441E	2.55	2.5	2.5		2.81	3.01	2.67	
T193R	2.52	2.51	2.51		2.84	3.01	2.68	
N207E	2.5	2.48	2.48	2.92	2.83	2.97		
T340K	2.46	2.49	2.46	2.79	2.98	3.1		
Q142K	2.47	2.48	2.48	2.82	2.81	2.89		
A91R	2.49	2.5	2.53		2.8	2.75	3	
G374R	2.49	2.48		3.31	3.28	3.14		
N257R	2.47	2.5	2.49		2.83	3.02	2.71	
V323R	2.49	2.5	2.49	3.12	2.86	2.81		

Capítulo 4. Análisis y discusión de resultados.

De las cinco mutantes termorresistentes de la β -glucosidasa B obtenidas por evolución dirigida en el laboratorio de J. Polaina²⁵, la variante H62R fue la que mostró mayor termoestabilidad y la única que involucró residuos con carga eléctrica, formándose además un nuevo puente salino (R62-E429) de acuerdo al modelo generado por este autor. Al calcular la energía libre electrostática total de esta mutante (-121541.6 kJ/mol), observamos que fue inferior a la de la β gl-B silvestre (-121357.6 kJ/mol), lo cual confirma nuestra suposición hecha al principio del desarrollo de este trabajo, de que las variantes de menor energía libre electrostática total, en las cuales se haya formado una nueva interacción electrostática, al generar mutaciones que involucren residuos cargados, deben ser termorresistentes.

Por lo anteriormente expuesto, para seleccionar mutantes individuales, y/o múltiples, de la proteína bajo estudio, en este trabajo se propone generar experimentalmente las variantes de β gl-B que presenten menor energía libre de solvatación en los cálculos teóricos.

El alineamiento múltiple de las dos secuencias consenso junto con la β gl-B y la β gl-THET, permitió identificar fácilmente la diferencia en el número de residuos con carga positiva o negativa presentes en ambas secuencias consenso.

Esta diferencia se ilustra en la **Fig. 39**, en la cual se encuentran marcados en color azul los residuos con carga positiva (R, K, H) y en rojo a los residuos con carga negativa (D y E). En esta figura puede observarse que la cantidad de residuos con carga positiva se mantiene conservada en las dos secuencias consenso, mientras que hay una marcada diferencia en residuos con carga negativa (15 residuos) en la secuencia consenso termófila, debido quizá a la adaptación ambiental a altas temperaturas que tuvieron que sufrir a lo largo de la evolución los diferentes microorganismos a partir de los cuales se aislaron estas proteínas termófilas.

En sólo 17, de las 55 posibles mutaciones generadas *in silico* y listadas en la **Tabla 3**, se establecieron nuevas interacciones electrostáticas con los residuos vecinos cargados a una distancia menor a 7 Å, como se detalla en la **Fig. 17**.

Cabe mencionar que los residuos mutados en estas 17 posiciones, no involucraron residuos altamente conservados en las secuencias de proteínas homólogas, lo cual puede corroborarse en la **Fig. 14**, que muestra en colores el grado de conservación de los residuos en las secuencias consenso.

Adicionalmente, de acuerdo con las variantes construidas *in silico*, todas estas mutaciones se encuentran en la superficie molecular, por lo que se esperaría que ellas no afectaran drásticamente la función y/o la estabilidad estructural de la enzima al ser generadas *in vivo*. Con la finalidad de ubicar redes de interacciones electrostáticas que confirieran una mayor estabilidad térmica a la molécula de β gl-B, se generaron variantes múltiples empleando como base a los 17 sitios antes mencionados (**Tablas 5-8**).

Para saber si la magnitud de la energía de libre de Gibbs electrostática total está correlacionada con la distancia de interacción entre los residuos de carga opuesta en la nueva interacción electrostática de cada variante (es decir, a menor distancia de interacción, menor energía libre electrostática total), se realizó la comparación que se muestra en la **Tabla 14**. En esta tabla se observa el hecho de que una menor energía libre electrostática total, como la registrada en las variantes A91R, V323R, T193R, N257R y T136R, está vinculada marginalmente con una distancia corta de interacción. Esto indica que la energía de interacción electrostática está influenciada en mayor medida por factores distintos a la formación de pares iónicos, entre los que podríamos mencionar: la interacción con otros residuos cargados a mayores distancias, y la ubicación de los átomos cargados respecto a la frontera dieléctrica proteína/agua.

Tabla 94. Comparación de la energía libre electrostática total con la distancia de interacción de las 17 variantes seleccionadas.

VARIANTE	Interacciona con	Distancia (Å)	$\Delta\Delta G_{\text{Elec Tot}}$ (kJ/mol)
L100E	H104	3.58	590.55
	R154	6.93	"
T146D	H143	4.8	699.35
I241E	R249	6.37	445.95
	R341	5.02	"
A238E	R242	3.27	96.75
Q395D	K397	2.52	205.6
S149D	K212	2.53	209.85
Q72E	K440	5.63	454.75
T136R	E190	2.53	-493.1
Q441E	K445	2.53	523.4
T193R	D276	4.88	-577.25
N207E	K145	2.53	271.15
T340K	E392	2.62	-95.9
Q142K	E139	2.55	-58.2
A91R	E139	4.88	-850.55
	E132	3.89	"
G374E	R377	2.57	71.15
N257R	D345	2.54	-527.8
	E253	2.56	"
V323R	E320	2.49	-769.2

Para seleccionar que mutantes individuales, y/o múltiples, podrían generarse experimentalmente por mutagénesis dirigida en la β gl-B de *Paenibacillus polymyxa*, para aumentar su termorresistencia, se consideraron las variantes con menor energía libre electrostática total respecto la reportada por la β gl-B silvestre ($\Delta\Delta G_{\text{Elec Tot}}$, **Tabla 14**).

Adicionalmente se determinó la energía de interacción de la glucosa (ligando presente en la estructura cristalográfica **2O9T**) con los dos residuos catalíticos de cada variante, y también se rastrearon los enlaces de hidrógeno nuevos formados o conservados entre la glucosa y los residuos del sitio activo.

A partir del estudio estructural sobre la mutante H62R, se encontró un valor de $\Delta\Delta G_{\text{Elec Tot}}$ negativo con respecto a $\beta\text{gl-B}$ (**Tabla 14**). Esto nos dio un indicio de que las variantes que presenten los valores más pequeños (negativos) de $\Delta\Delta G_{\text{Elec Tot}}$, podrían ser aquellas que manifestaran mayor termorresistencia, y por tanto, serían las mejores candidatas para ser construidas experimentalmente. Cabe mencionar que los residuos catalíticos de esta mutante poseen tanto una energía de interacción con la glucosa, como una red de enlaces de hidrógeno con este ligando, similares a la que presenta la $\beta\text{gl-B}$ silvestre.

Una vez realizado el análisis energético anteriormente descrito sobre las estructuras de las 17 variantes individuales generadas, se puede presentar la siguiente lista de mutantes propuestas para su construcción: A91R, V323R, T193R, N257R, T136R, T340K y Q142K. Todas estas variantes presentan un valor de $\Delta\Delta G_{\text{Elec Tot}}$ negativo, lo cual sugiere que ellas logren una mayor estabilidad electrostática que la $\beta\text{gl-B}$ silvestre. Además presentan una red de enlaces de hidrógeno entre los residuos del sitio activo y la glucosa inserta en el sitio activo, semejantes a la presentada en la $\beta\text{gl-B}$ silvestre y H62R.

En particular, las mutantes T136R, Q142K y V323R, mostraron una energía de interacción con la glucosa muy cercana a la que posee la proteína silvestre y su variante estable térmicamente. El hecho de mantener parámetros similares de interacción con el ligando, hace suponer que el sitio activo permaneció sin alteraciones importantes al efectuar estas 7 mutaciones, lo cual puede explicarse por la ubicación de todas ellas en exterior de la molécula.

Por otra parte las variantes dobles localizadas a menos de 10 Å del primer residuo mutado, que presentan valores de $\Delta\Delta G_{\text{Elec Tot}}$ negativo y que además podrían presentar alta termorresistencia, propuestas en este trabajo para generarse experimentalmente son: T136R-T193R, A91R-Q142K, V323R-G374E, G258K-N257R, N257R-I241E, T340K-Q395D. De este grupo destaca la primera variante con un valor de -1342.15 kJ/mol, que involucra dos cambios simultáneos detectados en el análisis citado en el párrafo anterior.

En estas 6 variantes dobles se ven incrementadas las redes de interacciones electrostáticas que presentan tanto las variantes individuales como la β gl-B silvestre; además de que los parámetros de interacción geométricos y energéticos de los residuos catalíticos con la glucosa, fueron semejantes a los presentados en la proteína natural y su variante estable térmicamente.

Por otra parte las 3 mutantes dobles y la triple mutante, resultantes de la combinatoria de las 3 variantes individuales de más baja energía, presentaron valores de $\Delta\Delta G_{\text{Elec Tot}}$, negativo y por lo tanto también se proponen generarlas experimentalmente. Sin embargo las variantes: A91R-T193R, V323R-T193R, A91R-T193R-V323R, fueron las que presentaron valores de la energía de interacción de la glucosa con los residuos catalíticos cercanos a los valores presentados por β gl-B y la H62R. Por lo tanto se esperaría que en ellas no se hubiera modificado drásticamente el sitio activo. Por el contrario en la variante A91R-V323R parece haberse modificado su sitio activo al presentar una energía de interacción de los dos residuos catalíticos con la glucosa de -102.78 kCal/mol

Del resto de las variantes construidas tanto individuales como dobles, triples y una cuádruple, mostraron valores de $\Delta\Delta G_{\text{Elec Tot}}$ positivos, por lo que parecen ser poco útiles para incrementar la termorresistencia en la proteína silvestre bajo estudio en este trabajo.

CONCLUSIONES

- Con base en las secuencias consenso obtenidas de los alineamientos múltiples de las proteínas homólogas a la β -glucosidasa B, provenientes de microorganismos mesófilos y termófilos por separado, se identificaron 55 sitios en los cuales existían residuos cargados conservados en ellas, susceptibles a mutarse en β gl-B al haber en ella residuos no cargados en esas posiciones.
- En solo 17 posiciones de los 55 sitios identificados, se pudieron generar nuevas interacciones electrostáticas en β gl-B a menos de 7.0 Å de distancia, con residuos de carga opuesta.
- La energía de solvatación electrostática calculada para la mutante termorresistente H62R, que posee un nuevo par iónico, fue inferior a la de la β gl-B silvestre en 1.9%.
- De las mutantes individuales analizadas, la A91R mostró la menor energía libre electrostática total en un 0.7% inferior a la β gl-B silvestre.
- En el caso de las dobles mutantes, la T136R-T193R mostró la menor energía libre electrostática total, 1.1% menor que la β gl-B silvestre.
- A partir de la realización de la combinatoria de las variantes que presentaron menor $\Delta\Delta G_{\text{Elec Total}}$, se obtuvo la variante A91R-V323R que presentó el valor de $\Delta\Delta G_{\text{Elec Total}}$ más bajo de todas las variantes generadas en el presente trabajo.
- Finalmente, de acuerdo a los resultados obtenidos puede observarse que existe una ligera correlación entre la distancia de interacción de los grupos cargados y la energía libre electrostática total electrostática en las variantes construidas. Por lo que es probable que se requiera moléculas de solvente explícito en los cálculos y la simulación de la dinámica molecular de las proteínas estudiadas, modificadas o no, para tener una descripción más realista de la energética en las interacciones electrostáticas generadas *in silico* en el presente trabajo.

PERSPECTIVAS

- Generar *in vitro* las 8 variantes de β gl-B propuestas en el presente trabajo, en las cuales se establecieron nuevas interacciones electrostáticas, al mutar residuos sin carga por residuos cargados, dictados por al menos una de las secuencias consenso, obtenidas a partir de los alineamientos múltiples de proteínas homólogas a β gl-B mesófilas y termófilas.
- Extracción y purificación de las variantes generadas experimentalmente.
- Evaluación fisicoquímica de estas variantes.
- Comparación de los resultados de la evaluación fisicoquímica con lo predicho energéticamente a través de las herramientas bioinformáticas utilizadas en el presente trabajo.

ANEXOS

Anexo A. Secuencias de aminoácidos de las proteínas homólogas a β gl-B

En el presente, se encuentran las secuencias de aminoácidos de las proteínas mesófilas y termófilas homólogas a β gl-B empleadas en la realización de éste proyecto. Se encuentran en formato FASTA, el código de letras y números inicial (azul) corresponde a la clave de acceso para cada proteína en la base de datos **Swiss-Prot and TrEMBL**.

MESÓFILAS

1. [P22073](#)
>BGLA_PAEPO Beta-glucosidase A - Paenibacillus polymyxa (Bacillus polymyxa).
2. [P22505](#)
>BGLB_PAEPO Beta-glucosidase B - Paenibacillus polymyxa (Bacillus polymyxa).
3. [Q7CV27](#)
>Q7CV27_AGR_T5 AGR_L_770p - Agrobacterium tumefaciens (strain C58 / ATCC 33970).
4. [Q9KBK3](#)
>Q9KBK3_BACHD Beta-glucosidase - Bacillus halodurans.
5. [Q1J2J3](#)
>Q1J2J3_DEIGD Beta-glucosidase - Deinococcus geothermalis (strain DSM 11300).
6. [Q0LEY4](#)
>Q0LKJ5_HERAU Beta-glucosidase - Herpetosiphon aurantiacus ATCC 23779.
7. [Q40UA0](#)
>Q40UA0_KINRA Beta-glucosidase - Kineococcus radiotolerans SRS30216.
8. [Q2WAV2](#)
>Q2WAV2_MAGMM Beta-glucosidase A - Magnetospirillum magneticum (strain AMB-1 / ATCC 700264).
9. [A0R2K0](#)
>A0R2K0_MYCSM Beta-glucosidase A - Mycobacterium smegmatis str. MC2 155.
10. [Q1CY46](#)
>Q1CY46_MYXXD Beta-glucosidase - Myxococcus xanthus (strain DK 1622).
11. [Q92LX2](#)
>Q92LX2_RHIME PROBABLE BETA-GLUCOSIDASE PROTEIN - Rhizobium meliloti (Sinorhizobium meliloti).
12. [Q3J2E9](#)
>Q3J2E9_RHOS4 Putative Beta-glucosidase A - Rhodobacter sphaeroides (strain ATCC 17023 / 2.4.1 / NCIB 8253 / DSM 158).
13. [Q21ZF1](#)
>Q21ZF1_RHOFD Beta-glucosidase - Rhodoferax ferrireducens (strain DSM 15236 / ATCC BAA-621 / T118).
14. [Q21KX3](#)
>Q21KX3_SACD2 Beta-glucosidase - Saccharophagus degradans (strain 2-40 / ATCC 43961 / DSM 17024).
15. [Q0L8C7](#)
>Q0LB03_9ACTO Beta-glucosidase - Salinispora tropica CNB-440.
16. [Q0L3W9](#)
>Q0L3W9_9ACTO Beta-glucosidase - Salinispora tropica CNB-440.
17. [Q3Q207](#)
>Q3Q207_9GAMM Beta-glucosidase - Shewanella baltica OS155.
18. [Q59976](#)
>Q59976|Q59976_STRSQ Beta-glucosidase - Streptomyces sp.
19. [Q8CJM3](#)
>Q9K440_STRCO Putative beta-glucosidase - Streptomyces coelicolor.
20. [Q9F3I3](#)

Anexo A. Secuencias de aminoácidos de las proteínas homólogas a β gl-B

>Q9F3I3_STRCO Putative cellobiose hydrolase - Streptomyces coelicolor.
21. Q8D4K7

>Q8D4K7_VIBVU Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase - Vibrio vulnificus.
22. Q2S749

>Q2S749_HAHCH Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase - Hahella chejuensis (strain KCTC 2396).
23. A0AC59

>A0AC59_STRAM Putative beta-glucosidase - Streptomyces ambofaciens ATCC 23877.
24. Q82M59

>Q82M59_STRAW Putative beta-glucosidase - Streptomyces avermitilis.
25. A0GC42

>A0GC42_9BURK Glycoside hydrolase, family 1 - Burkholderia phytofirmans PsJN.
26. A0GC42

>A0GC42|A0GC42_9BURK Glycoside hydrolase, family 1 - Burkholderia phytofirmans PsJN.
27. Q08S20

>Q08S20_STIAU Beta-glucosidase A - Stigmatella aurantiaca DW4/3-1.
28. Q08YK7

>Q08YK7_STIAU Beta-glucosidase A - Stigmatella aurantiaca DW4/3-1.
29. Q21ZG0

>Q21ZG0_RHOFD Beta-glucosidase - Rhodoferax ferrireducens (strain DSM 15236 / ATCC BAA-621 / T118).
30. Q084Z6

>Q084Z6_SHEFN Beta-glucosidase - Shewanella frigidimarina (strain NCIMB 400).
31. Q12KZ0

>Q12KZ0_SHEDO Beta-glucosidase - Shewanella denitrificans (strain OS217 / ATCC BAA-1090 / DSM 15013).
32. Q3QGX0

>Q3QGX0_9GAMM Beta-glucosidase - Shewanella amazonensis SB2B.
33. Q0KMU3

>Q0KMU3_9GAMM Beta-glucosidase - Shewanella baltica OS195.
34. P94248

>P94248_BIFBR Beta-D-glucosidase - Bifidobacterium breve.
35. Q1MBM6

>Q1MBM6_RHIL3 Putative beta-glucosidase - Rhizobium leguminosarum bv. viciae (strain 3841).
36. Q3GY55

>Q3GY55_9ACTO Beta-glucosidase - Nocardioides sp. JS614.
37. Q0FHD6

>Q0FHD6_9RHOB Putative Beta-glucosidase A - Roseovarius sp. HTCC2601.
38. A0J3H5

>A0J3H5_9GAMM Glycoside hydrolase, family 1 - Shewanella woodyi ATCC 51908.

TERMÓFILAS

1. [A0LR48](#)
>A0LR48_ACICE Beta-glucosidase - Acidothermus cellulolyticus 11B.
2. [P10482](#)
>BGLS_CALSA Beta-glucosidase A - Caldocellum saccharolyticum (Caldicellulosiruptor saccharolyticus).
3. [Q2ZJP6](#)
>CALSA Beta-glucosidase - Caldicellulosiruptor saccharolyticus DSM 8903.
4. [A0H0U4](#)
>A0H0U4_9CHLR Beta-glucosidase - Chloroflexus aggregans DSM 9485.
5. [A0HPV3](#)
>A0HPV3_9THEM Beta-glucosidase - Fervidobacterium nodosum Rt17-B1.
6. [Q47PF5](#)
>Q47RE2_THEFY Beta-glucosidase - Thermobifida fusca (strain YX).
7. [Q2ADJ1](#)
>Q2ADJ1_9FIRM Glycoside hydrolase, family 1 - Halothermothrix orenii H 168.
8. [Q608B9](#)
>Q608B9_METCA Beta-glucosidase - Methylococcus capsulatus.
9. [A0ICG5](#)
>A0ICG5_9CHLR Beta-glucosidase - Roseiflexus castenholzii DSM 13941.
10. [Q1FAN6](#)
>Q1FAN6_9CHLR Beta-glucosidase - Roseiflexus sp. RS-1.
11. [Q1AUD7](#)
>Q1AUD7_RUBXD Beta-glucosidase - Rubrobacter xylanophilus (strain DSM 9941 / NBRC 16129).
12. [B0KCV1](#)
>Q3CKP2_THEET Beta-glucosidase - Thermoanaerobacter ethanolicus ATCC 33223.
13. [Q8RCQ8](#)
>Q8RCQ8_THETN Beta-glucosidase/6-phospho-beta-glucosidase/beta-galactosidase - Thermoanaerobacter tengcongensis.
14. [Q3E343](#)
>Q3E343_CHLAU Glycoside hydrolase, family 1 - Chloroflexus aurantiacus J-10-fl.
15. [Q8GEB4](#)
>Q8GEB4_THETH Beta-glycosidase - Thermus thermophilus.
15. [Q0R5R7](#)
>Q0R5R7_ALIAC Beta-glycosidase - Alicyclobacillus acidocaldarius (Bacillus acidocaldarius).
16. [Q60026](#)
>Q60026_THEBR Beta-glucosidase - Thermoanaerobacter brockii (Thermoanaerobium brockii).
17. [Q8GKQ8](#)
>Q8GKQ8_9THEM Beta-glucosidase - Fervidobacterium sp. YNP.
18. [Q9ZNN7](#)
>Q2WGB4_9BACL Beta-glucosidase - Paenibacillus sp. HCl.
19. [Q9LAV5](#)
>Q9LAV5_THEFU Beta-glucosidase BglC - Thermomonospora fusca.
20. [Q03506](#)
>BGLA_BACCI Beta-glucosidase - Bacillus circulans.
21. [Q8GHE5](#)
>Q8GHE5_THECA Beta-glycosidase - Thermus caldophilus.
22. [Q8GEB2](#)
>Q8GEB2_THEFI Beta-glycosidase - Thermus filiformis.
23. [Q8GEB1](#)
>Q8GEB1_9DEIN Beta-glycosidase - Thermus sp. IB-21.
24. [Q8GEB0](#)
>Q8GEB0_9DEIN Beta-glycosidase - Thermus sp. IB-21.

Anexo A. Secuencias de aminoácidos de las proteínas homólogas a β gl-B

25. [Q9L794](#)
>Q9L794_9DEIN Beta-glycosidase - Thermus nonproteolyticus.
26. [Q08638](#)
>BGLA_THEMA Beta-glucosidase A - Thermotoga maritima.
27. [Q9RA58](#)
>Q9RA58_9DEIN Beta-glucosidase - Thermus sp. Z-1.
28. [O33843](#)
>BGLA_THENE Beta-glucosidase A - Thermotoga neapolitana.

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B.

En rojo se presentan los residuos completamente conservados, en azul las posiciones conservadas, en color negro se anotan los residuos no conservados. En la línea de la secuencia consenso, un residuo que es altamente conservado se presenta en mayúsculas y en color rojo, un residuo que es poco conservado aparece en minúsculas y en azul. Una inserción es representada por un espacio en blanco. Los símbolos siguientes representa a cualquiera de los aminoácidos entre paréntesis : ! (IV), \$ (LM), % (FY), # (NDQE).

		FAMILIA	MESÓFILA					
1						70		
BGLA_PAEPO		MTIF	QFPQDFMWGT	ATAAYQIEGA	YQEDGRGLSI	WDTFAHTPGK		
Q9KKB3_BAC		MSII	QFPKEMKGV	ATASYQIEGA	INAGGRGASI	WDFVAKTPGK		
BGLB_PAEPO		MSENTF	IFPATFMWGT	STSSYQIEGG	TDEGGRTPSI	WDTFCQIPGK		
Q7CV27_AGR	MDRKAS	RMTDPQTLAA	RFPGDFLFGV	ATASFQIEGA	TKADGRKPSI	WDAFCNMPGH		
Q1MBM6_RHI		MIDPKKLAE	RFPGDFTFGV	ATAAFQIEGA	SKADGRKPSI	WDAFCNMPGR		
Q92LX2_RHI		MMIEAKKLA	RFPGDFVFGV	ATASFQIEGA	SKADGRKASI	WDAFCNMPGR		
Q1J2J3_DEI	MTQT	RPAPDTLRRS	DFFPAGFTFGI	ATSAYQIEGA	TSEDGRGPSI	WDTFCREPGR		
Q2S749_HAH		MNQ	KFAKDFTWGV	ATSAYQIEGA	ANEGGRGPSI	WDTFSHTPGK		
Q21ZG0_RHO	M	IALMTTTSNS	AFFPKNFVWGV	ATSAFQIEGA	SEADGKGPSI	WDDFCRIPGA		
Q0LKJ5_HER		MTTVEQ	HFPADFMWGT	ATSSYQIEGA	VHEDGRGESI	WDRFSHTPGK		
Q0L3W9_9AC	MSTHPRQ	QPLPHAASPD	SLERKPAGTL	RFPFGFWGA	AKEDGRGESI	WDTFSHTPGR		
Q1CY46_MYX		MR	QFPNDFLWGV	ATSAFQIEGA	TSADGRGESI	WDRFAATPGK		
Q08S20_STI					MEDGRGESI	WDRFSKTPGK		
Q21ZF1_RHO	MKNKK	TDDLPLLKRT	DFALDFRWGC	STSSYQIEGG	VLDLGRGESI	WDRFCATPGH		
A0AC59_STR		MSDSLDA	AFPADFTWGT	ATSAYQIEGA	VAEDGRAPSI	WDTFSHTPGK		
Q82M59_STR		MTIDLA	ALPHDFLWGT	ATAAYQIEGA	VAEDGRSPSI	WDTFSHTPGK		
Q3J2E9_RHO		MTFSRA	DFFPADFLFGV	ATSAYQIEG-	HGAGGAGRTH	WDDFAAPGN		
Q0FHD6_9RH		MLLTRK	DFPEGFFTGT	ATSSYQIEG-	HKFGGAGRTH	WDDFADTPGN		
Q21KX3_SAC	M	NRLTLPSSR	LRSKEFTFGV	ATSSYQIEGG	ID--SRLPCN	WDTFCEQPNT		
Q3Q207_9GA		MKISLPKNSR	LQSEAFTFGV	ATASFQIEGG	VD--SRQTCI	WDTFCATPDK		
Q0KMU3_9GA		MKISLPKNSR	LQSEAFTFGV	ATASFQIEGG	VD--SRQTCI	WDTFCATPDK		
Q084Z6_SHE		MKIMLPADSK	MHSAFTFGV	ATASFQIEGA	VD--YRLPCI	WDTFCATPGK		
Q12KZ0_SHE		MKISLPADSN	MHSTSFYIGV	ATASFQIEGG	VD--SRLPCI	WDTFCATEGN		
A0J3H5_9GA		MKISLPEASR	LNQKEFTYGV	ATASFQIEGG	YD--SRLPCI	WDTFCSTQGK		
Q3QGX0_9GA	MTTL	SHFTLPGDSV	MMQKDFLFGV	ATASFQIEGD	AE--HRQPCI	WDTFCDTPGK		
Q8D4K7_VIB	M	NKYQLPQDSQ	LRQADFLFGV	ATSSYQIEGG	AQLGGRTPSI	WDTFCNQPGA		
A0GC42_9BU	MANDASAVSD	ALSPRSPSLA	DFPTPADSS	LWRKSFLLGA	ATASYQIEGA	VNEDGRGPSI	WDTFSAKPGK	
Q2WAV2_MAG	MTVAK	AAQSAPQGAR	QFPKDFFWGA	STAAYQIEGA	YDTDGRGMTI	WDKFTAD-GK		
Q08YK7_STI		M	RFPFGFLWGV	STSSYQIEGG	APDDGRGRSI	WDTYCATPGK		
Q40UA0_KIN	MTTTEERS	TTPAPSAAR	TFFPGFVWGT	ATASYQVEGA	VAEDGRTPSI	WDTFSHTPGR		
Q59976 Q59	MVPAAQ	QTATAPDAAL	TFFPGFLWGS	ATASYQIEGA	AAEDGRTPSI	WDTYARTPGR		
Q9F3I3_STR	M	LESATVPVPV	TFFPAFLWGA	ATSAYQIEGA	VREDGRTPSI	WDTFSHTPGK		
A0R2K0_MYC		MSQPDRTL	GFPDGFVWGT	ATASYQIEGA	VTAGGRAPSI	WDTFSHKPGA		
Q0LB03_9AC	MSNPASPPA	VGVLAEERPL	AFFPGFLWGA	ATAAYQIEGA	ATAGGRTPSI	WDTFSHTPGR		
Q9K440_STR		MSEYP	GFPFGFVFGA	ATASYQIEGA	ATEDGRGPSI	WDTYSRTPGL		
P94248_BIF		MTM	IFPKGFMTGT	ATAAYQIEGA	VAEGGRTPSI	WDTFSHT-GH		
Q3GY55_9AC		MSLHL	PSBSTLAYGA	ATASYQIEGA	TAEDGRGASI	WDTFTTRPGA		
Consensus	p.dsr	lfpkdflwgv	aTas%Q!EGA	vdedgrgpsi	WDT%shtpgk	
71						140		
BGLA_PAEPO		VFNGDNGNVA	CDSYHRYEED	IRLMKELGIR	TYRFSVSWPR	IFPNGDG---	EV-NQEGLDY	YHRVVDLLND
Q9KKB3_BAC		VKNGDNGDVA	CDSYHRYEED	IEIMKDLGVD	MYRFSVAWPR	IFPNGTG---	EV-SREGLDY	YHRLVDRLTE
BGLB_PAEPO		VIGDCCGDVA	CDHFHFKED	VQLMKQLGFL	HYRFSVAWPR	IMP-AAG---	II-NEEGLLF	YEHLLDEIEL
Q7CV27_AGR		VFGRHNGDIA	CDHYNRWEED	LDLIKEMGVE	AYRFSIAWPR	IIPDGF---	PI-NEKGLDF	YDRLDVGCKA
Q1MBM6_RHI		VHNRDNGDIA	CDHYNRLEQD	LDLIKEMGVE	AYRFSIAWPR	IIPDGT---	PV-NEAGLDF	YDRLDVGCKA
Q92LX2_RHI		VYGRHNGDVA	CDHYNRLEQD	LDLIKSLGVE	AYRFSIAWPR	IVPEGT---	PI-NEKGLDF	YDRLDVGLKA
Q1J2J3_DEI		IRDGSSGDVA	CDHYHRWEAD	LDLIAALGVN	AYRFSVAWPR	VQPDGR---	PV-NAAGLDF	YERLTDGLLA
Q2S749_HAH		TVNGETGDVA	CDHYHRYQED	VDLIAGLGVD	AYRFSIAWPR	VQPEGK---	AW-NEAGDF	YSRLIDALLV
Q21ZG0_RHO		IQDASDGRVA	CDHYHRWAAD	LDLIAGLGVD	AYRFSVSWPR	VQPLGSG---	AF-NEKGFEF	YERLDVGLMA
Q0LKJ5_HER		TKFGQTGDIA	CDHYHRYPED	LDLMRELGLG	SYRFSIAWPR	LFPEGK---	KI-NQAGLDF	YKRIIEGLHQ
Q0L3W9_9AC		VHNGDTGDIA	ADHYHRYDAD	LDLMAELGLR	SYRFSIAWPR	IQPDGT---	AP-NQRGLDF	YRRLDGLHD
Q1CY46_MYX		ISDGS DGKVA	CDHYHRWRED	VALMRWLGVK	SYRFSVAWPR	VLPTGR---	AV-NAAGLDF	YSRLVDGLLD
Q08S20_STI		VRDGTNGDVA	CDHYHRFRED	IALMKGLGIK	HYRFSVAWPR	IIPGGR---	KV-NPPGLDF	YGRLDVADLE

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

Q21ZF1_RHO IRDGSSGAVA CDHYHRWPED LDLARSLGTN AYRFSIAWPR IFANGRG--- LAPNQKGLDF YSRMVDGML E
A0AC59_STR IDNDHDGVA CDHYHRWPED IALMKRLGAD AYRLSVAWPR VVPGGDG--- PV-NAAGLDF YDRLVDGLLD
Q82M59_STR IDNDHDGVA CDHYHRWRED IGLMKQLGTN AYRLSVAWPR VIPGGDG--- PV-NAKGLAF YDELVDGLLE
Q3J2E9_RHO VAHAEDGRRA CDHYHRWEED LDFVRDAGFD SYRFSASWAR VMPEGRG--- TV-NAEGLDF YDRLVDGMLA
Q0FHD6_9RH VVRAENGVA CDHYHRYAED LDLVAAAGFD AYRFSISWAR VLPEGRG--- EV-NAEGLDF YDRLTDAMLE
Q21KX3_SAC IIDNTNGAIA CDHYHRWQDD IELIANLQVD AYRFSIAWGR VINLDGS--- L--NNEGVTDF YKNILTKLRE
Q3Q207_9GA IRDASNGDIA CNHNLNLWQED IALITSLGVD AYRFSIAWGR VLNQDGS--- I--NQQGVDF YIGILDELKR
Q0KMU3_9GA IRDASNGDVA CNHNLNLWQED IALIASLGVD AYRFSIAWGR VLNQDGS--- I--NQQGVDF YIGILDELKR
Q084Z6_SHE IRDNSDGSQA CEHVKLWRED VDLIESLGVD AYRLSISWPR VMHKDGS--- L--NPQGVAF YTDLLDELNR
Q12KZ0_SHE IRDASNGEHA CEHLKLWRED VDLIDSLGVD AYRLSISWPR VLHKDGS--- L--NPQGVNF YIDLLDELNR
A0J3H5_9GA IRDGS DGKSA CEHLKLWRED IELIDSLNVD AYRLSLSWGR VINQDGS--- V--NSEGIDF YIQLLDALKR
Q3QGX0_9GA IADGSNGQVA CDHYKLWRDD VDLIASLGVD AYRLSISWGR VLHPDGS--- V--NQRGMDF YINLLDELGR
Q8D4K7_VIB VDNMDNGDVA CDHFHLWQDD IALIQLGVD AYRLSMWAPR ILPKDQG--- V--NQQGLEF YERIIDECHA
A0GC42_9BU VLAGDTGAVA CDHYHRWQAD VDLLAGLQLE AYRLSIAWPR VMDAAGA--- P--NRKGLDF YKRLDLARLKE
Q2WAV2_MAG IMDGSSAKVA CDHYHRYPED IALMKAAGFN AYRFSIAWPR IIPAGTG--- AI-NAKGLDF YDRLVDKILE
Q08YK7_STI VARGDTGEVA CDHYHRYAED LDLLRNLGAT VYRFSIMWPR VMPDGVG--- RL-NPKGLDF YDRIVDGLLE
Q40UA0_KIN VVAGDTGDVA DDHYHRFRED IALMKRGLMA SYRFSVWAPR ITPQVSADEL GPVNPAGLAF YSTLVDELIA
Q59976|Q59 VRNGDTGDVA TDHYHRWRED VALMAELGLG AYRFSIAWPR IQPTGR---- GPALQKGLDF YRRLADELLA
Q9F3I3_STR TAGGDTGDVA VDHYHRYRDD VALMKDLGLG AYRFSISWPR VQPTGR---- GPAVQRGLDF YRRLVDELLA
A0R2K0_MYC VHNGDNGDVA DDHYHRYRED VGLLADLVGT HYRFSIAWPR LQPDGK---- GPINTEGLDF YQRLVDELLA
Q0LB03_9AC VVAGTGDVA CDHYHRLDSD VALMAELGLR SYRFSVSWPR VQPGGT---- GPINQEGGLDF YRRLVDQLLA
Q9K440_STR VVNGDTGDVA CDHYHRYPED VALLRDLGVD SYRFSIAWPR IVPDGS---- GAVNPKGLDF YSRVLDELLA
P94248_BIF TLNGDTGDVA DDFYHRWEDD LKLLRDLGVN AYRFSIGIPR VIPTPD---- GKPNQEGGLDF YSRIVDRLLE
Q3GY55_9AC IRDGS DGKSA CDSYHRYEED ADLVAGLGVG WYRFSIAWPR VLPEGT---- GRVEPRGLDY YDRLVDALLA
Consensus vrdgdnGdVA c#hyhrweeD ldlmaglGvd aYRFSiaWpR vqpdgrg gv.nqkGld% Ydrlvdglla

141 210
BGLA_PAEPO NGIEPFCTLY HWDLPQALQD ----AGGWGN RRTIQAFVQF AETMFREFHG KIQHWLTFNE PWCIAFLSNM
Q9KKB3_BAC NGIQPMCTLY HWDLPQALQE ----KGGWDN RDTIDAFVRY AEVMFKEFGD KINHWITFNE LWCVSFLSNY
BGLB_PAEPO AGLIPMLTLY HWDLPQWIED ----EGGWTO RETIQHFKTY ASVIMDRFGE RINWNTINE PYCASILGYG
Q7CV27_AGR RGIKTYATLY HWDLPQLTMG ----DGGWAS RSTAHAFQRY AKTVMARLGD RLDAVATFNE PWCNAVLSHL
Q1MBM6_RHI RGIKTFATLY HWDLPPLL LAG ----EGGWTA RSTAYAFQRY AKTVMNRLGD LDRVATFNE PWCIVVLSHL
Q92LX2_RHI RGIKAFATLY HWDLP LALMG ----DGGWTA RTTAYAYQRY AKTVIARLGD RLDAVATFNE PWC SVVLGHL
Q1J2J3_DEI RGIEPHVTLY HWDLPQPLQD ----TGGWPN RDTAERFAEY AA AVAGRLGD RVRSYATLNE PWCSSILSYE
Q2S749_HAH KGLQPVYVTLY HWDLPQALE D ----EGGWRN RDTCHRFADY AA EFMRRYGD SVKTISTHNE PWCTATLGH A
Q21ZG0_RHO RGLKPYLTLN HWDLP SALQA ----TGGWEN RDTVQRFVDY ACEVARRLGD RVVSICTHNE PWWVAVFNQ
Q0LJK5_HER RHLTPMATLY HWDLPQALQD ----KGGWMN RDTALRFAEY AEAMYRQLGE SVPFWITHNE PWWAAFVGHF
Q0L3W9_9AC RGIQPVATLF HWDLPQALQD ----RGGWES REVTHRFADY ADHVFRALGD RVPTWLTINE PKTVVQNGYL
Q1CY46_MYX AGIEPFVTLY HWDLPQALQD ----LGGWPS RDTASAFVEY ADVMSRKLGD RVKRWITHNE PWCISVLGYG
Q08S20_STI AGIEPVYVTLY HWDLPQVLQD ----EGGWAK RSTAEAFVEY AGVVARSLGD RVKRWITHNE PWCASMSLYQ
Q21ZF1_RHO RGLPEWVTLY HWDLPQALQE ----QGGWAN RDTVDFAVEY TDVVSRLGD RIKHWITHNE PLCTAFHNGY
A0AC59_STR AGITPNVTLY HWDLPQALQD ----RGGWTA RETAEHLAEY ASVVAGRLGD RVTQWATLNE PLCSAWIGHL
Q82M59_STR AGITPSVTLY HWDLPQVLQD ----RGGWPE RATAEHFAAY AAVVAERLGD RVTHWATLNE PLCSAWIGHL
Q3J2E9_RHO RGLKPALTYL HWELPSALQD ----LGGWRN RDIAGWFADY AEVLLGRIGD RVWSTAPVNE PWCVAWLSHF
Q0FHD6_9RH RGLKPYATLY HWELPSALAD ----LGGWRN RDIQWFGDF TEVVM SKLGD R METVAPINE PWCVAWLSHF
Q21KX3_SAC KNLKAYITLY HWDLPQHLED ----AGGWLN RDTAYKFRDY VNLITQALDD DVFCYTTLNE PFC SAYLGYE
Q3Q207_9GA RNIKAFVTLY HWDLPQHIED ----QGGWLN RDTAYLFKDY ADKISQAFGD RVYSYATLNE PFCSSYLGYE
Q0KMU3_9GA RNIKAFVTLY HWDLPQHIED ----QGGWLN RDTAYLFKDY ADKISQAFGD RVYSYATLNE PFCSSYLGYE
Q084Z6_SHE RGIKTFVTLY HWDLPQHIED ----NGGWLN RETAYLFADY ADKITQAFGD RVYSYATFNE PFCSSYLGYE
Q12KZ0_SHE RGIKAFVTLY HWDLPQHIED ----NGGWLN RDTAYLFADY ADKISQAFGD RVYSYATFNE PFCSSYLGYE
A0J3H5_9GA KKIKAFTLY HWDLPQHIED ----NGGWLN RETAFLFQDY ADKVSQILGD RVYSYATLNE PFCSSYLGYE
Q3QGX0_9GA RGINVFVTLY HWDLPQHLED ----KGGWLN RDTAVAFANY AAIVANALGN RVYAYSTLNE PFCSAFLGYE
Q8D4K7_VIB RGLKVFTLY HWDLPQYLED ----KGGWLN RETAYKFAEY AKVVSIFYGN KIDSYATLNE PFC SAYLGYR
A0GC42_9BU KGITTFVTLY HWDLPQHLED ----CGGWLN RDTAYRFADY ADLMSREL AG SVDAWMTLNE PWC SAYLGYG
Q2WAV2_MAG AGIKPMACTLY HWDLPQPLED ----KGGWQG RDIVGPF AEY ARIATKRLGD RVKDWMYMLNE PNVVAIIGY G
Q08YK7_STI RGLRAWPCLY HWDLPQALQD ----RGGWAN RDIVGWFAEY TAVMARRLGD RVDQWVTFNE PWSAVWVGYE
Q40UA0_KIN AGITPNVTLY HWDLPQALE D ----AGGWTA RTTAERFAEY AEVVAGALGD RVRTFITLNE PWC SAYLGYA
Q59976|Q59 KGIQPVATLY HWDLPQELN ----PGGWPE RPTAERFAEY AAIADALGD RVKTWTTLNE PWC SAFLGYG
Q9F3I3_STR AGIKPAVTLY HWDLPQELD ----AGGWPE RDTAYRF AEY AQIVGEALGD RVEQWITLNE PWCA AFLGYA
A0R2K0_MYC HDIQPVVTLY HWDLPQTLED ----AGGWPA RDTAQRFADY AAMTHGK LHD RIRHWTTLNE PWC SAFLGYA
Q0LB03_9AC NGIEPWTLTY HWDLPQPLED ----AGGWPA RDTAARFADY AALVAGALGD RVRYWTTLNE PWC SAFLGYG
Q9K440_STR AGIEPAATLY HWDLPQALE D ----RGGWRV RETAERFAEY TAVVAEHLSD RVP RWITLNE PWC SSFLGYS
P94248_BIF YGIAPIVTLY HWDLPQYMAS GDGREGGWLE RETAYRIADY AGIVAKCLGD RVHTYTTLNE PWC SAHLSYG
Q3GY55_9AC RGVSPATLY HWDLPQALE D ----RGGWLE RSTAEAFADY AMVVHERLGD RVGWATHNE PWCAAYLGYA
Consensus rgikpfvtL% HW#LPqaled rGGWln Rdtayrfad% aavvarrlgd rvkswatlNE Pwcsaflgyg

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

Q2WAV2_MAG AVWNRVPLDG VMRGAI PDVL ----- AEKMAHIVKP GDMETIKFPI DMLGINYYSR MTKM-----
 Q08YK7_STI DKWNGVFLDP VYHGRYPASV ----- EAKFASHVQP GDLEEHQPI DFLGVNHYFP SYVQ-----
 Q40UA0_KIN GLQNRVFLDP VLRGTYPADV QADT----SS -----VTDWSFVRE GDALAIHQPL DVLGVNYYSP TVVRAWDGSA
 Q59976_Q59 ALANRVFTGP MLQGAYPEDL VKDT----AG -----LTDWSFVRD GDLRLAHQKL DFLGVNYYSP TLVSEADG-S
 Q9F3I3_STR DLSGGIFHGP ILHGAYPQTL LEAT----AP -----LTDWSFVQD GDLGQIRQPI DAICLNYYTP TVVSAADADS
 A0R2K0_MYC AVSNRFLFLDA LLRGRYPADL SADV----AM -----ISDFGFVQP GDEAVIAAPL DFLGVNYYFR ETVRFGTG--
 Q0LB03_9AC GLANRFFLDP LLRGSYPADL MSDL----RQ -----VSDFGHVRA GDLATIAAPL DLVGINYYSR HVVAAPTA--
 Q9K440_STR TQHNLVWTEP ILAGRYPATE EETW----GE LITGDFRRD GDELEISQPL DFLGVNYYRP LVVAD-----
 P94248_BIF LIGNRVFLDP MLRGRYPDEL FSIT----KG -----ICDWFVCD GDLDLIHQPI DVLGINYYST NLVKMSD---
 Q3GY55_9AC AIRNRLWLG P LVDGAYDDGL LAVA----PE -----LADPDVVEH GDELELVGSA DWIGINYYTP FRPTLAD---
Consensus glfnrwlfdp llrgr%Padl l.....al apkd!pfvqd gd!al!aqpl Dflgin%Ytr avvr.....

351

420

BGLA_PAEPO -----F N---PEAGFL QSEEinMGLP VTDIGWPVES RGLYEVLYHL -QKYG--ID IYITENGACI
 Q9KKB3_BAC -----Y K---ENEGLF DLEKVDAGYE KTDIGWNIYP EGFYKVLYYI TEQYQ--IP IYITENGSCY
 BGLB_PAEPO -----S T---NDASLL QVEQVHMEEP VTDMGWEIHP ESFYKLLTRI EKDFSKG-LP LLITENGAM
 Q7CV27_AGR -----D DATPGAFFPA TTPAPAVSEV KTDIGWEVYA PALKSLVETL YKRY-DL-PE CYITENGACY
 Q1MBM6_RHI -----D DAERKGDFFPW TVEAPPASDV KTDIGWEIYA PGLKLSVEDL YRRY-EL-PE CYITENGACD
 Q92LX2_RHI -----A DPAKGAEYPA TVNAKPVSNV KTDIGWEVYA PALGSLVETL NARY-RL-PP CYITENGACY
 Q1J2J3_DEI -----A S---GGG--- --V--PAGAP VTAMGWEIYP QGLTDLRLRL QADYPSL-PP LLITENGAAF
 Q2S749_HAH -----Q A---NPQ--- --IPAPAAMG LITDMGWEVYP QGLAQHLVRI TRDY-LP-PP IYITENGAAF
 Q21ZG0_RHO -----A G---EPW--- --DVKSSGRE ITDMGWEVYP QGLTELLRLR HHDY-TM-PP IYITENGAAF
 Q0LKJ5_HER -----A A---DESKFL ALVNSPAIGE TSFRGWEVVP AAFADLLKRV QRDY-GN-TP IYITENGSAF
 Q0L3W9_9AC -----A A---GTT--- --VRRWPTSE ASWQ--QIYP EGMVDILTRV TRDY-GP-IP LTVTENGLEPT
 Q1CY46_MYX -----S DRIPESQNA P RTVH--PEPE RTDMDWEVYA PALTRMLVHL HTDY-QP-GP LYITENGCA Y
 Q08S20_STI -----S DKVPEAQNEP RTVFAVEKE WTEMGWEVYP DGLREILMRV HLDY-RP-RK IYITENGAS Y
 Q21ZF1_RHO -----D A--PGVG VMS TQVIESDNVE RTAFGWEVSP EGMVTLLGRL AKDY-QP-AE IYLTENGSTY
 A0AC59_STR -----D D--PQGPVYP AREVRLDPV RTMGWQIDA GGLEALLRLR TDDY-GV-RK LYVTENGSAF
 Q82M59_STR -----D D--PTGPAPR VRSVRRLGVP RTGMWQIDA TGIETLLRLR THEY-GP-RK LYVTENGSAF
 Q3J2E9_RHO -----A A----PGPWP GLSEVEGPLP RTRMGWEIHP EGLSDILLRI HEGYTRG-LP LIVTENGMAA
 Q0FHD6_9RH -----A D----KSPWP SLKEVDGILP KTYMGWEIYP EGLLNFLERT AREYTG E-LP IFVTENGMAA
 Q21KX3_SAC -----A D----GNGG FTEQVPEGVE LITDMGWEVYP QGLTDLIDL NQRY-TL-PP LLITENGAM
 Q3Q207_9GA -----A D----AEQG FVQVDLPGVP KTDIGWEIHP QAFTDLLVSL NQTY-DL-PP IFITENGAM
 Q0KMU3_9GA -----A D----AEHG FVQVDLPGVT KTDIGWEIYP QAFTDLLVSL DHIY-DL-PP IFITENGAM
 Q084Z6_SHE -----A D----PATG FSQIDMVDKP KTDIGWEIYP QSFTDLLTSL HALY-PL-PP IYITENGAM
 Q12KZ0_SHE -----A D----EKTG FSQVDMQNV P KTQMGWEIYP QAFTDLLTSL NALY-PL-PP VYITENGAM
 A0J3H5_9GA -----H C----PEQR FKEITMPDVP RTDIGWEIYP QAFTDLLISL DKSY-QL-PP IYITENGAM
 Q3QGX0_9GA -----A G----GPLG FEEVRIDNVP RTAMWQICP QAFTDLLTGL AQEF-NL-PP IYITENGAAE
 Q8D4K7_VIB -----F D----ANGD LEEMPQDAE HTYIGWEIYP QALTDLLRLR KQRYPNL-PP VYITENGAA G
 A0GC42_9BU -----S D----GAHG FRDVPLEGVE RTQMGWEVYP DGLRDLLTGF KATYANL-PP IYITENGMA S
 Q2WAV2_MAG -----H E--EGHPFDV FWGDAHCD-R WTAMAWPVQP DGLYDLLREF KELY-GN-PA VFAENGAA Y
 Q08YK7_STI -----P A--ANGAWPF QHADPPLYFR RTETNWAIDG QAFYEALRN V QIRC-GN-PP VYVTENGGA F
 Q40UA0_KIN PRAENDGHGN SAHSPWVAST DVEFPPELPE KTAGMWSIDP RGMTELLRLR HREHPGL--E LLVTENGAA Y
 Q59976_Q59 GTHNSDGHGR SAHSPWPGAD RVAHFPPGGE TTAMGAVDVP SGLYELLRLR SSDFPAL--P LVITENGAA F
 Q9F3I3_STR RAPRADGHGA SDHSPWPAAE DVAFHQPPGE RTEMGWTVDP TGLHELMRY NREAPGM--P LYISENGAA Y
 A0R2K0_MYC --APSDNPHA RA---WVGST DVEPVRTGLP TTAMGWEIDP TGLYDVL SWV HREYGPV--P IYVTENGAA F
 Q0LB03_9AC --AAPPEPYW RTPSCWPGSE DVRFVARGMP VTDMDWEIDP SGLVETLQRV YEEYTDL--P LYVTENGSA F
 Q9K440_STR ---APHRES D PARRVATDNR YEEVHRHPGV HTAMNWPVVP DSFTDLLVAL KRQYGDALPP VHTENGSAE
 P94248_BIF ---RPQFPQS TEASTAPGAS DVWLPTAGP HTEMGWNIDP DALYETLVR LNDYPGM--P LVVTENGMA C
 Q3GY55_9AC ----PALETH PEVDAYPGAT PVSFVVRE-P RTDIGWEVEA RGLEELLVET HRRTGL--P LIVTENGAA Y
Consensusa d.....pgpg fvevppgvp rtdmgWe!yp qgldllrl hrdy.g! pp iy!tENGaa y

421

490

BGLA_PAEPO NDEV-----V NGKVQDDRRI SYMQQHLVQV HRTIHDLHV KGYMAWSLLD NFEWAEGYNM RFGMIHVDFR
 Q9KKB3_BAC NDEP-----V NGQVKDEGRI RYLSQH LTAL KRSMEG VNI KGYMAWSLLD NFEWAEGYSM RFGIVHVN YR
 BGLB_PAEPO RDEL-----V NGQIEDTGRH GYIEHLKAC HRFIEEGGQL KGYFVWSFLD NFEWAEGYSK RFGIVHIN YE
 Q7CV27_AGR NMGV-----E NGEVNDQPR L DYAEHLGIV ADLIRDGYPM RGYFAWSLMD NFEWAEGYRM RFGLVHVDY D
 Q1MBM6_RHI NTDV-----V DGEVDDTMR L DYVGDHLEIV AGLIKDGYPL RGYFAWSLMD NFEWAEGYRM RFGLVHVDY E
 Q92LX2_RHI NMGV-----E NGTVDDQPR L DYISDH LAVT ADLIAKGYP M RGYFAWSLMD NFEWAEGYRM RFGIVHVDY E
 Q1J2J3_DEI TDRL-----E DGRVHDPERV RYLQTHLAAL RRALDAGVDV RGYFAWSLMD NFEWAEGY EK RFGLVYVDY P
 Q2S749_HAH PDRM-----E KGEVQDPARI HYLETHLQAL RQAMELGADI RGYFYWSLAD NYEWNWGYSK RFGLT YVDY A
 Q21ZG0_RHO QDEV-----V DGRVHDLRRQ TYIANHIEAV AEAMRQGV RV NGYFVWSLLD NFEWASGYAK RFGIVRVDY D
 Q0LKJ5_HER ADLK---RAA DGSVNDGDRM SYLHTHLEAV ADAIAGVPV KGYFAWSMLD NYEWAEGYDE RFGIIEVDFA
 Q0L3W9_9AC PDTL---AA DGTVDDGERI QFLRDHLAAA HRADGVPL ESFHVWSLLD NFEWAEGYDQ RWGLVYVDY P
 Q1CY46_MYX ATGP---SE DGKVHDDKRV AYLRSHLEAS LEAIRQGVPL AGYFAWSLMD NFEWAFGYQK RFGMVYVDY D
 Q08S20_STI STAP---GE DGRVRDEKRL SFLRDHFIAA RRAMEQGAPV AGYFVWSLMD NFEWDRGYSQ RFGIVWVDY K
 Q21ZF1_RHO DDVL---SP EGNIDDVERR RYLVRHLQAT REIVAQQIPV KGYFAWSLLD NFEWAEGYIR RFGLVHVDY E
 A0AC59_STR PDAV---SP DGEVHDPRRA RYLEQHVGAC ARAVRKGAPL AGYFAWSLLD NFEWAEGYDK RFGLVHVDY A
 Q82M59_STR PDVV---RP DGTVDDPERQ DYLVRH LAAC ASAARKGAPL AGYFAWSLLD NFEWAEGYDK RFGLVHVDY R
 Q3J2E9_RHO ADRV-----Q AGEVQDPDR I AYLEGHLAAV QRRAIQGV PV RGYHVWSLLD NFEWAFGYDQ RFGLVHVDY Q

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

Q0FHD6_9RH HDDI-----V AGEVQDEHRI AYIEAHLADV RRAIARGVPV KGYFCWSLLD NYEWSFGYDK RFGLVHVDVD
 Q21KX3_SAC VDEL-----V NGEVNDIARI NYFQTHLQAV HNAIEQGVDV RGYFAWSLMD NFEWALGYSK RFGITYVDYQ
 Q3Q207_9GA DDKC-----I DGRVDDFDR L SYQHHLTAV DNAIVQGVNI QGYFAWSLMD NFEWAEGYLYK RFGIVYVDYA
 Q0KMU3_9GA DDKC-----I DGRVDDFDR L SYQHHLTAV DNAIVQGVNI QGYFAWSLMD NFEWAEGYLYK RFGIVYVDYA
 Q084Z6_SHE DDKL-----I EGKVVDDQDR L EYNAHLNAV NNAIEQGVNV VGYFAWSLMD NFEWAEGYLYK RFGIVYVDYE
 Q12KZ0_SHE DDKL-----Q NGNVDDQDR L QYNAHLNAV NNAIEQGVNV IGYFAWSLMD NFEWAEGYLYK RFGIVYVDYQ
 A0J3H5_9GA DDKL-----I AQS VNDLDRV QY YQGHLEAL DKAIEAGVNI QGYFAWSLMD NFEWAEGYLYK RFGIVYVDYN
 Q3QGX0_9GA DDAP-----F NGTVHDP MRL DY LQSHLLAV HQAIEAGVDI KGYFAWSLMD NFEWAEGYLYK RFGIVYVDYG
 Q8D4K7_VIB EDAC-----I NGEVNDQQRV RYFQSHLLAL DEAIRAGVNV QGYFAWSLMD NFEWAEGYLYK RFGIVYVDYA
 A0GC42_9BU DDKV-----I DGQVEDSQR I SFLKRHL SAV DQAIKAGVEI RGYFLWSLMD NFEWAFGYER RFGIVHVDYA
 Q2WAV2_MAG DDVV-----TP DGQVHDAERV AFIRDHVSEV ARAVKDGCNV KGYLVWSLLD NFEWAEGYLYK RFGIVYVDYE
 Q08YK7_STI IDVP----GP QGRVDDQDR I AYYREYLIGL QRAISEGADV RGFMPWSLLD NFEWALGYEK RFGIVHVDYR
 Q40UA0_KIN DDVVVA---G --QVH DVERT RYLRDH LAAV LDAEAGAPV TGYLWSFLD NFEWAGYGSK RFGVHVDYD
 Q59976|Q59 HDYADP--EG --NVNDPER I AYVRDHLAAV HRAIKDSDV RGYFLWSLLD NFEWAEGYLYK RFGAVYVDYP
 Q9F3I3_STR DDKPGA--DG --TVHDPDRV AYLNGLHTAV RQAIADGADV RGYFLWSLLD NFEWAEGYLYK RFGAVYVDYA
 A0R2K0_MYC DDDAEA--DG --SIRDDDR I TFLDSHFRA T LRAIDDGADI RGYFVWSLLD NFEWAEGYLYK RFGLIHVDYA
 Q0LB03_9AC VD-AVV--DG --KVDDPDRV AYFEAHLRAA HQAIAAGVPL RGYFAWSLMD NFEWAEGYLYK RFGMIHVDYR
 Q9K440_STR DDAAAA--DG --TVHDTDRV AYLRDH LTA L RQAIADGADV RGYFVWSLLD NFEWAEGYLYK RFGVHVDYD
 P94248_BIF PDKVEVGTDG VKMVHDNDRI DYLRHLEAV YRAIEEGTDV RGYFAWSLMD NFEWAEGYLYK RFGITYVDYE
 Q3GY55_9AC ADDTLR-EGA AGVIDDQDR I AYLRDHIAAT ERARSAGADV RAYIVWTL L D NFEWAEGYLYK TFGVVHVDPK
Consensus ddkv... .v dge!dDpdRi a%lrdHlaav araieaGvvp rg%faWSLSD N%EWaeG%sk RFGlvhV#%a

491

535

BGLA_PAEPO TQVRTPKESY YWYRNVSNN WLETRR
 Q9KKB3_BAC TLERTKKDSF YWYKQMIANQ FFEL
 BGLB_PAEPO TQERTPKQSA LWFKQMAKN GF
 Q7CV27_AGR TQVRTLKNSG KWYSALASRF PKGNHGVVKG
 Q1MBM6_RHI TQVRTVKKSG KWHLEAAQF PKGNHKPG
 Q92LX2_RHI TQVRTIKKSG RWYKDLAERF PSGNHKPG
 Q1J2J3_DEI TQTRVLKDSG HWYRQFLR
 Q2S749_HAH TQQRTLKASG HWYRDFIARQ RA
 Q21ZG0_RHO TQERTLKDSA LWYRAF LSEQ KRKGM
 Q0LKJ5_HER TQKRTPKRTA RWYQQIVANN GLPSLPADVQ ALAERYRNCP IGPQD
 Q0L3W9_9AC TQRRVPKRSA HWYREVIAAG GF
 Q1CY46_MYX SQRRIPKDSA HLYKALVEKN GLDVELAA
 Q08S20_STI TQQRIPKDSA LWYRGVIAEN ALSVS
 Q21ZF1_RHO TQQRRLKLSG EWYGAFLKGE L
 A0AC59_STR TQARTVKTSG RRYADIRAH RQSHA
 Q82M59_STR TQVRTIKGTG HRYADIVRTH RGRARRAA
 Q3J2E9_RHO NLQRTPKASY HALARALAR
 Q0FHD6_9RH TLARLPKASY HALKALTEG QS
 Q21KX3_SAC TQKRTLKASG HAFAEFVSSR S
 Q3Q207_9GA SQTRTIKASG QAYSDLIRSR AHFTNNNNK
 Q0KMU3_9GA SQTRTIKASG HAYSDLIRSR AHLTNNNNK
 Q084Z6_SHE TQKRTLKASA HAYRDFINAR K
 Q12KZ0_SHE TQQRTLKASA HGYRDFIQAR K
 A0J3H5_9GA TQVRTIKASG LAYRDFITAR -HTDN
 Q3QGX0_9GA TQQRILKSSA KAYQGMLAIR QEASQQ
 Q8D4K7_VIB TQKRTLKQSA IAYRNTLLAR ABEKQ
 A0GC42_9BU TQKRTIKRSA NLVSKFLKER KARTE
 Q2WAV2_MAG TLKRTPKDSY KWFAEVIRTG RVG
 Q08YK7_STI TQKRTPKASY GFMREVIATN ALPGV
 Q40UA0_KIN TQERTVKDSA LWYADVSRN ALPAADAD
 Q59976|Q59 TGTRIPKASA RWYAEVARTG VLPTA
 Q9F3I3_STR SQQRTPKSSA LWYGRAARTG TLPPVDAAE
 A0R2K0_MYC TQVRTPKASA RWFAEVTRRN GVPHVFRQND APADIR
 Q0LB03_9AC SQARTLKSSG RWYAETIRRN GLAAQ
 Q9K440_STR TQRRTPKDSY RWYREMIAAN RG
 P94248_BIF SQERVKDSF DWYRRFIADH SAR
 Q3GY55_9AC DQTRTPKASY HWLAEHVAEA H
Consensus tqkRtpKasa hwyrefiarn gl.....

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

	FAMILIA		TERMÓFILA				
	1				70		
A0LR48_ACI	MTQIEE	RDQVESRPTL	RF PD DFV WG V	AT S AYQ IEGA	VA ED GRG PSI	WD T FS HT PGK	VV GD TGD VA
Q47RE2_THE	MTSQSTTPLG	NLEETPKPDI	RF PS DFV WG V	AT AS FQ IEGS	TT AD DGR GPSI	WD T FC AT PGK	VE NG DTG DP A
Q9LAV5_THE	MTSQSTTPLG	NLEETPKPDI	RF PS DFV WG V	AT AS FQ IEGS	TT AD DGR GPSI	WD T FC AT PGK	VE NG DTG DP A
Q1AUD7_RUB		M	RF PE GFL WG A	AT AA YQ VEGA	VE GG GRG PSI	WD T FS HT PGR	VY RG DTG D V A
Q2ADJ1_9FI		MAKI	IF PE DFI WG A	AT SS YQ IEGA	FN ED GK GESI	WD R FS HT PGK	IE NG DTG D I A
Q3CKP2_THE		MI	KF PK DFL WG T	AT SS YQ IEGA	VN ED GR T PSI	WD T FS K TEG K	TY NG H TGD V A
Q60026_THE		MIKLA	KF PR DFV WG T	AT SS YQ IEGA	VN ED GR T PSI	WD T FS K TEG K	TY K G H TG D V A
Q8RCQ8_THE		MIELV	KF PK DFT WG V	AT SS YQ IEGA	VN ED GR T PSI	WD T FS K TEG K	TY Q G H TG D V A
Q0R5R7_ALI		MREMR	KF PE G F V WG T	AT AS YQ VEGA	AR EG GRG RSI	WD T FS HT PGK	VA EG H TGD V A
A0ICG5_9CH		MAIR	HF PD DFL WG A	AT AA FQ IEGA	TR ED DGR GESI	WD R FC AT PGK	VL NG DTG DP A
Q1FAN6_9CH		MTTR	RF PQ GFL WG S	AT AA FQ IEGA	TR ED DGR GESI	WD R FC AT PGK	VL NG DTG DP A
Q2WGB4_9BA		MTIF	QF PQ DFR WG A	AT AS YQ IEGA	AQ EG GRG VSI	WD T F AR TPG K	VF NG D NGD V A
BGLA_BACCI		MSIH	MF PS DFK WG V	AT AA YQ IEGA	YN ED DGR GMSI	WD T F AR TPG K	VF NG D NG V A
A0HPV3_9TH		MIKRS	DF PK DFL FG T	AT AA YQ IEGA	AK ED GK GPSI	WD V FS HT PGK	TF NG DTG D I A
Q8GKQ8_9TH		MIRRS	DF PK DFI FG T	AT AA YQ IEGA	AN ED DGR GPSI	WD V FS HT PGK	TL NG DTG D V A
BGLA_THEMA		MNVK	KF PE GFL WG V	AT AS YQ IEGS	PL AD GAG MSI	WH T FS HT PGN	VK NG DTG D V A
BGLA_THENE		MK	KF PE GFL WG V	AT AS YQ IEGS	PL AD GAG MSI	WH T FS HT PGN	VK NG DTG D V A
Q608B9_MET		MSRY	EF PE RFL WG A	AT S AYQ VEGS	PL AD GAG PSN	WH R FC RQ PGR	IL NG DTG D T A
A0H0U4_9CH		MNHAI	RF PT NFI WG A	AT AA YQ IEGA	WN ED GK GESI	WD R F VR RPG A	IA D G ST GD V A
Q3E343_CHL		MSNDI	RF PS DFL WG A	AT AA YQ IEGA	WN ED GK GESI	WD R F VR RPG V	IA NG DTG D I A
BGLS_CALSA		MDM	SF PK GFL WG A	AT AS YQ IEGA	WN ED GK GESI	WD R F TH Q KRN	IL Y G H NG D V A
Q2ZJP6_CAL		M	SF PK GFL WG A	AT AS YQ IEGA	WN ED GK GESI	WD R F TH Q KRN	IL Y G H NG D V A
Q8GEB4_THE		MD	DH AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD A F AR RPG A	IR D G ST GE P A
Q8GEB0_9DE		MD	DH AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD A F AR RPG A	IR D G ST GE P A
Q8GHE5_THE		MT	EN AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD T F AR RPG A	IR D G ST GE P A
Q8GEB2_THE		MA	EN AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD T F AR RPG A	IR D G ST GE P A
Q8GEB1_9DE		MT	EN AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD A F AR RPG T	IR D G ST GE P A
Q9L794_9DE		MT	EN AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD T F AR RPG A	IR D G ST GE P A
Q9RA58_9DE		MT	EN AE KFL WG V	AT S AYQ IEGA	TQ ED DGR GPSI	WD A F AQ R P G A	IR D G ST GE P A
Consensus	mi r fp edFl w Gv	ATaa % Q ! EGA	tnedGr gp Si	wdtF sh tpgk	vl n GdtG # V A
	71						140
A0LR48_ACI	AD H YH RY VGD	VRL M ADL G V T	S Y RF S V A W P R	IL P S G S G A V N	R A GL D F Y S R L	V D EL L N H G I T	P A L T L Y H W D L
Q47RE2_THE	CD H Y N R Y R D D	VAL M REL G V G	A Y RF S I A W P R	I Q PE G K G T P V	E A GL D F Y D R L	V D CL L E A G I E	P W P T L Y H W D L
Q9LAV5_THE	CD H Y N R Y R D D	VAL M REL G V G	A Y RF S I A W P R	I Q PE G K G T P V	E A GL D F Y D R L	V D CL L E A G I E	P W P T L Y H W D L
Q1AUD7_RUB	CD H Y H R L E E D	LDL M AR F L G	A Y RF S V A W P R	I Q PE G R G P A N	RS G L D F Y R R L	V E GL L E R G I E	P V L T L Y H W D L
Q2ADJ1_9FI	CD H Y H L Y R E D	I E L M K E I G I R	S Y RF S T S W P R	I L PE G R G R V N	Q K G L D F Y K R L	V D N L L K A N I R	P M I T L Y H W D L
Q3CKP2_THE	CD H Y H R Y K E D	V E I L K E I G V K	A Y RF S I A W P R	I F P E -E G K Y N	P K G M D F Y K R L	I D EL L L K D I M	P T A T I Y H W D L
Q60026_THE	CD H Y H R Y K E D	V E I L K E I G V K	A Y RF S I A W P R	I F P E -E G K Y N	P K G M D F Y K K L	I D EL Q K R D I V	P A A T I Y H W D L
Q8RCQ8_THE	CD H Y H R Y K E D	V Q I M K E I G V K	A Y RF S I A W P R	I F P E -E G K Y N	P K G M D F Y K R L	V D EL L L K R E I I	P V A T I Y H W D L
Q0R5R7_ALI	CD H Y H R Y Q D D	V R L M K E L G I S	S Y RF S I A W P R	V M P E -K G R V W	V K G L D F Y K R L	A T A L L E H G I R	P A A T M Y H W D L
A0ICG5_9CH	CD H Y H R W R D D	I A L M K S L G F P	A Y RF S I A W S R	I M P K R G A V N	P A GL D F Y D R L	V D GL L A A G I R	P F V T L Y H W D L
Q1FAN6_9CH	CD H Y H R W R D D	I T L M K S L G L Q	A Y RF S I A W P R	I P P Q R G Q V N	P A GL D F Y D R L	V D GL L D A G I R	P F V T L Y H W D L
Q2WGB4_9BA	C D S Y H R Y E E D	I E L M K K L G I N	T Y RF S I A W P R	I I P D G D E I N	R E G L D F Y H R F	V D K L L E A G I E	P F C T L Y H W D L
BGLA_BACCI	C D S Y H R V E E D	V Q L L K D L G V K	V Y RF S I S W P R	V L P Q T G E V N	R A GL D Y Y H R L	V D EL L A N G I E	P F C T L Y H W D L
A0HPV3_9TH	CD H Y H R F K E D	V A I M K E I G L N	A Y RF S I S W P R	V M Q D G K N - I N	Q K G I D F Y N R L	V D EL L E N D I	P F I T L Y H W D L
Q8GKQ8_9TH	CD H Y H R Y K E D	I Q L M K E I G L D	A Y RF S I S W P R	I M P D G K N - I N	Q K G V D F Y N R L	V D EL L K N D I	P F V T L Y H W D L
BGLA_THEMA	CD H Y N R W K E D	I E I I E K L G V K	A Y RF S I S W P R	I L P E G T R V N	Q K G L D F Y N R I	I D T L L E K G I T	P F V T I Y H W D L
BGLA_THENE	CD H Y N R W K E D	I E I I E K I G A K	A Y RF S I S W P R	I L P E G T G K V N	Q K G L D F Y N R I	I D T L L E K N I T	P F I T I Y H W D L
Q608B9_MET	CD H Y H R F R E D	V A L M K A L G L S	A Y RF S I A W S R	I F P E G K R I N	W R G I A H Y Q A L	V E T L L E H G I R	P M A T L H H W D L
A0H0U4_9CH	CD H Y H R Y E E D	L E H M A A M G L K	A Y RF S I A W P R	I F P D G T G Q P N	Q R G L D F Y R R L	I D G L H R R R I L	P V A T L Y H W D L
Q3E343_CHL	CD H Y H R Y E E D	L D H M A A L G L K	A Y RF S I S W P R	I F P T G S G Q P N	Q R G L D F Y R R L	I D G L H R R H I E	P V A T L Y H W D L
BGLS_CALSA	CD H Y H R F E E D	V S L M K E L G L K	A Y RF S I A W T R	I F P D G F G T V N	Q K G L E F Y D R L	I N K L V E N G I E	P V V T L Y H W D L
Q2ZJP6_CAL	CD H Y H R F E E D	V S L M K E L G L K	A Y RF S I A W T R	I F P D G F G T V N	Q K G L E F Y D R L	I N K L V E N G I E	P V V T L Y H W D L
Q8GEB4_THE	CD H Y H R Y E E D	I A L M Q S L G V R	A Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A S G I T	P F L T L Y H W D L
Q8GEB0_9DE	CD H Y H R Y E E D	I A L M Q S L G V R	A Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A S G I T	P F L T L Y H W D L
Q8GHE5_THE	CD H Y H R Y E E D	I A L M Q S L G V G	V Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A A G I T	P F L T L Y H W D L
Q8GEB2_THE	CD H Y H R Y E E D	I A L M Q S L G V G	V Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A A G I T	P F L T L Y H W D L
Q8GEB1_9DE	CD H Y H R Y E E D	I A L M Q S L G V G	A Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A A G I T	P F L T L Y H W D L
Q9L794_9DE	CD H Y H R Y E E D	I A L M Q S L G V G	V Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A A G I T	P F L T L Y H W D L
Q9RA58_9DE	CD H Y H R Y E E D	I A L M Q S L G V G	A Y RF S V A W P R	I L P E G R G R I N	P K G L A F Y D R L	V D R L L A S G I T	P F L T L Y H W D L
Consensus	CDhYhRye#D	valmkelGvk	ayRFS!aWpR	!lPegkgrvn	qkGldfYdrl	!#eLleagIe	PfLtlYHWDL

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

	141											210
A0LR48	ACI	PQALQDQ-GG	WTNRATAQRF	AEYAVVVARE	LGDRVNFWIT	LNEPWCAAFI	GYGAGVHAPG	HTDSAEALTA				
Q47RE2	THE	PQALEDA-GG	WPNRDTAKRF	ADYAEIVYRR	LGDRITNWNT	LNEPWCSAFI	GYASGVHAPG	RQEPAAALAA				
Q9LAV5	THE	PQALEDA-GG	WPNRDTAKRF	ADYAEIVYRR	LGDRITNWNT	LNEPWCSAFI	GYASGVHAPG	RQEPAAALAA				
Q1AUD7	RUB	PQALEDQ-GG	WTSRQTSERF	AEYAALVYEA	LGGSVRFWIT	LNEPWVSAWM	GYGLGVHAPG	RRSTADALAA				
Q2ADJ1	9FI	PQALQDK-GG	WTNRDTAKYF	AEYARLMFEE	FNGLVDLWVT	HNEPWVVAFE	GHAFGNHAPG	TKDFKTALQV				
Q3CKP2	THE	PQWAYDKGGG	WLNRRDSVKWY	VEYATKLFEE	LGDVIPLWIT	HNEPWCSSIL	SYGIGEHAPG	HKNYREALIA				
Q60026	THE	PQWAYDKGGG	WLNRESIKWY	VEYATKLFEE	LGDAIPLWIT	HNEPWCSSIL	SYGIGEHAPG	HKNYREALIA				
Q8RCQ8	THE	PQWAYEKNGG	WLNRESVKWY	VEYASKLFEE	LGDVIPLWIT	HNEPWCSSIL	SYGIGEHAPG	HKDWREALIA				
Q0R5R7	ALI	PQWMEDE-GG	WNSRETVSRF	LEYSEILFRE	LGDLVPMWIT	HNEPWCSAFI	GYGIGVHAPG	LKDWRRAYRA				
A0ICG5	9CH	PQALEDA-GG	WPARDTAAAF	ADYADVVARR	LGDRVKHWIT	LNEPWCSAFI	GYWTGDHAPG	RKE-GPALAA				
Q1FAN6	9CH	PQALEDA-GG	WPARDTASAF	ADYADVVRER	LGDRVKHWIT	LNEPWCSAFI	GYWTGDHAPG	VRE-GPVLAA				
Q2WGB4	9BA	PQTLEDI-GG	WGNRRTVDAF	VKYAEVIFKE	FSGKINFWLT	FNEPWCIAFI	SNLLGIHAPG	NKDLQTSINV				
BGLA_BACCI		PQALQDQ-GG	WGSRTIDAF	AEYAEIMFKE	LGDKIKQWIT	FNEPWCMAFI	SNYLGVHAPG	NKDLQLAIDV				
A0HPV3	9TH	PYALYEK-GG	WLNDDIAMYF	RAYATLMFNE	LGDRVKHWIT	LNEPWCSSFL	GYFTGEHAPG	HQNLQEALIA				
Q8GKQ8	9TH	PYALYEK-GG	WLNPDIALYF	RAYATFMFNE	LGDRVKHWIT	LNEPWCSSFL	GYTGEHAPG	HQNLQEALIA				
BGLA_THEMA		PFALQLK-GG	WANREIADWF	AEYSRVLFEN	FGDRVKNWIT	LNEPWVVAIV	GHLYGVHAPG	MRDIYVAFRA				
BGLA_THENE		PFSLQLK-GG	WANRDIADWF	AEYSRVLFEN	FGDRVKNWIT	LNEPWVVAIV	GHLYGVHAPG	MKDIYVAFHT				
Q608B9	MET	PALEDL-GG	WANRDSAGWF	ADYAHTVIRA	IGNEIDLWAT	LNEPWVIMDA	GYVSGVHPPG	HRSLKDPAPWV				
A0H0U4	9CH	PQALEDR-GG	WINRDTAFYF	AEYADYLFQ	LGNDVALWAT	HNEPFIQAFY	GYNGENAPG	KRVPPWRVLHV				
Q3E343	CHL	PQALEDR-GG	WTQRDTALRF	AEYADYLFRT	IGGEVALWAT	HNEPFIQAFY	GYNGENAPG	KRVPPWRVLAV				
BGLS_CALSA		PQKLQDI-GG	WANPEIVNYY	FDYAMLVINR	YKDKVKKWIT	FNEPYCIAFI	GYFHGIHAPG	IKDFKVAMDV				
Q2ZJP6	CAL	PQKLQDI-GG	WANPEIVNYY	FDYAMLVINR	YKDKVKKWIT	FNEPYCIAFI	GYFHGIHAPG	IKDFKVAMDV				
Q8GEB4	THE	PLALEER-GG	WRSRETAFAF	AEYAEAVARA	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q8GEB0	9DE	PLALEER-GG	WRSRETAFAF	AEYAEAVARA	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q8GHE5	THE	PQALEDR-GG	WRSRETAFAF	AEYAEAVART	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q8GEB2	THE	PQALEDR-GG	WRSRETAFAF	AEYAEAVARA	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q8GEB1	9DE	PQALEDR-GG	WRSRETAFAF	AEYAEAVARA	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q9L794	9DE	PQALEDR-GG	WRSRETAFAF	AEYAEAVARA	LADRVPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Q9RA58	9DE	PLAQEER-GG	WRSRETAFAF	AEYAEAVARA	LADRLPFAT	LNEPWCSAFI	GHWTGEHAPG	LRNLEAALRA				
Consensus		Pqaledk.GG	Wanrdtaka%	aeYaelvfre	lgdr!kfwit	LNEPwcsafI	gygtGvhaPG	hrdlaaala				

	211											280
A0LR48	ACI	AHLLLLAHGL	A---VQALGS	VLP----PDC	QMAITLNPVAV	ARPASLA---	--EEDVAAAR	KVDGLQN-RL				
Q47RE2	THE	AHLLMLGHGL	AAAVMRDLAG	QAG----RSV	RIGVAHNQTT	VRPYTDS---	--EADRDAAR	RIDALRN-RI				
Q9LAV5	THE	AHLLMLGHGL	AAAVMRDLAG	QAG----RSV	RIGVAHNQTT	VRPYTDS---	--EADRDAAR	RIDALRN-RI				
Q1AUD7	RUB	THLLLLGHGL	ALEALRALGG	-----GG	RLGITLNLSP	VRAASAE---	--PADAEAR	RVNAN-RL				
Q2ADJ1	9FI	AHLLLLSHGM	AVDIFREEDL	-----PG-	EIGITLNLTP	AYPAGDS---	--EKDVKAAS	LLDDYIN-AW				
Q3CKP2	THE	AHLLLLSHGE	AVKAFREMNI	-----KGS	KIGITLNLTP	AYPASEK---	--EEDKLAQ	YADGFAN-RW				
Q60026	THE	AHLLLLSHGE	AVKAFREMNI	-----KGS	KIGITLNLTP	AYPASEK---	--EEDKLAQ	YADGFAN-RW				
Q8RCQ8	THE	AHLLLLSHGE	AVKVFDMNL	-----KGA	QIGITLNLTP	AYPASEK---	--EEDKLAQ	YADGFAN-RW				
Q0R5R7	ALI	AHLLLLSHGQ	AVRLYRELGL	-----PG-	EIGITLNLTP	VYAATPN---	--PEDLAAAD	RQDMFQN-RW				
A0ICG5	9CH	AHLLLLGHGL	ALAALRAAH	-----SDV	RAGITLNFSP	ADPASDS---	--DADRAAAW	RYDGFN-RW				
Q1FAN6	9CH	AHLLLLGHGL	ALAALRAAY	-----PDV	QAGITLNFSP	ADPASDS---	--DADRAAAW	RYDGFN-RW				
Q2WGB4	9BA	AHLLLVHAGK	AVQSFRLGT	-----TG	QIGIAPNVCW	AEPYKSK---	--PEDQAACD	RSIALNT-DW				
BGLA_BACCI		SHLLLVHAGR	AVTLFRELGI	-----SG	EIGIAPNTSW	AVPYRRT---	--KEDMEACL	RVNGWSG-DW				
A0HPV3	9TH	AHLLLRSHGH	AVQAFREEV-	-----RDG	KIGLTVNVTK	VEPGDSR---	--PESFFVAS	LVDKVIN-AW				
Q8GKQ8	9TH	AHLLLRSHGH	AVQAFREEV-	-----KDG	KVGLTVNVTK	IEPGDAK---	--PESFLVAS	LVDKFVN-AW				
BGLA_THEMA		VHLLLRSHAR	AVKVFRETIV	-----KDG	KIGIVFNNGY	FEPASEK---	--EEDIRAVR	FMHQFNYP				
BGLA_THENE		VHLLLRSHAK	SVKVFRETIV	-----KDG	KIGIVFNNGY	FEPASER---	--EEDIRAVR	FMHQFNYP				
Q608B9	MET	THLLLRSHAL	AVQAFRADG-	-----RG	QIGLTVNLEP	KYALTDSD---	--RDDRAAAE	RAHAYMN-RQ				
A0H0U4	9CH	VHLLLLSHGL	AVSAFRATKP	QPVRADLPSP	QIGIVLMIWP	QYPASDH---	--PADLKAQ	RIDGAMN-RL				
Q3E343	CHL	VHLLLLSHGL	AVSAFRAARP	QAVRSDMPS	QIGIVLMIWP	QYPASQR---	--IADQKAAQ	RIDGIMN-RL				
BGLS_CALSA		VHSLMLSHFK	VVKAVKENNI	-----DV	EVGITLNLTP	VYLQTERLGY	KVSEIEREMV	SLSSQLDNQL				
Q2ZJP6	CAL	VHSLMLSHFK	VVKAVKENNI	-----DV	EVGITLNLTP	VYLQTERLGY	KVSEIEREMV	SLSSQLDNQL				
Q8GEB4	THE	AHLLLLGHGL	AVEALRAAG-	-----AR	RVGIVLNFAP	AYG-----	---EDPEAVD	VADRYHN-RY				
Q8GEB0	9DE	AHLLLLGHGL	AVEALRAAG-	-----AR	RVGIVLNFAP	AYG-----	---EDPEAVD	VADRYHN-RY				
Q8GHE5	THE	AHLLLLGHGL	AVEALRAAG-	-----AK	RVGIVLNFAP	AYG-----	---EDPEAVD	VADRYHN-RY				
Q8GEB2	THE	AHLLLLGHGL	AVEALRAAG-	-----AK	RVGIVLNFAP	VYG-----	---EDPEAVD	VADRYHN-RY				
Q8GEB1	9DE	AHLLLLGHGL	AVEALRAAG-	-----TK	RVGIVLNFAP	VYG-----	---EDPEAVD	VADRYHN-RY				
Q9L794	9DE	AHLLLLGHGL	AVEALRAAG-	-----AR	RVGIVLNFAP	AYG-----	---EDPEAVD	VADRYHN-RY				
Q9RA58	9DE	AHLLLLGHGL	AVEALRAAG-	-----AK	RVGIVLNFAP	RYG-----	---EDPEAVD	VADRYHN-RF				
Consensus		AhHl\$lgHgl	aveafraIgl.dg	riGitlntltp	aypasds...	..eedraaar	radgfnh.rw				

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

281 350

A0LR48_ACI WLDPLFHGTY PQDVVNFTSK -VTDWSFVRD NDLAVIATPF ---DILGVNY YNPVIVGHYA GSGSRGRDGH
Q47RE2_THE FTEPLVKGRY PEDLIEDVAA -VDYSFVQD GDLKTI SANL ---DMMGVNF YNPSWVSGNR ENG--GSDRL
Q9LAV5_THE FTEPLVKGRY PEDLIEDVAA -VDYSFVQD GDLKTI SANL ---DMMGVNF YNPSWVSGNR ENG--GSDRL
Q1AUD7_RUB YLDPLFRGSY PQDMLHYRG -ASDFSFVRD GDLQRI SAPV ---DFLGVNY YMRHTVRAAP GGGVVRGPS--
Q2ADJ1_9FI FLSPVFKGSY PEELHHIYEQ NLGAF-TQP GDMDIISRDI ---DFLGINY YSRMVRHKKP GDN-----
Q3CKP2_THE FLDPFIKGN Y PEDMMELYSK IIGEFDFIKE GDLETISVPI ---DFLGVNY YTRSIKYN- EDS-----
Q60026_THE FLDPFIKGN Y PEDMMELYSK IIGEFDFIKE GDLETISVPI ---DFLGVNY YTRSIKYN- EDS-----
Q8RCQ8_THE FLDPFIKGN Y PEDMMELYSK IIGEFDFIRE GDLETISVPI ---DFLGVNY YTRSIKYN- EDS-----
Q0R5R7_ALI FLDPVLRGEY PEELLHRVDQ VVGGFDAVKP GDLDVIATPI ---DFLGVNY YTRAVVADDP SEP-----
A0ICG5_9CH YLDPLYRSAY PADMLALY AQ -MGQAPPVQD DDMRIIAAPL ---DFLGVNY YSRAVIRDDP QAG-----
Q1FAN6_9CH YLDPLYRGT Y PEDMLQLYAR -LGQTPPVQP DDMRIIAVPM ---DFLGVNY YSRAVIRDDP QAG-----
Q2WGB4_9BA FLDPFIKGN Y PQFMVDWFAE -AGATVPIQE GDMEIISQPI ---DLGINY YTMGINRFPN --E-----
BGLA_BACCI YLDPIYFGEY PKFMLDWYEN -LGYPPIVD GDMEIISQPI ---DFGINY YTSMMRNYNP GEA-----
A0HPV3_9TH FHDPIYFGEY PEEAVKNYVE ---MGLNVPD NDFDIISTPI ---DFGVNY YTRTLVVFDE TNP-----
Q8GKQ8_9TH FHDPIYFGEY PEEAVLYTE ---KGLQVPD SDMNIISTPI ---DFGVNY YTRTLVVFDM NNP-----
BGLA_THEME FLNPIYRGDY PELVLEFARE ---YLPENYK DDMSEIQEKI ---DFVGLNY YSGHLVKFDP DAP-----
BGLA_THENE FLNPIYRGDY PELVLEFARE ---YLPENYK DDMSEIQEKI ---DFVGLNY YSGHMVKYDP NSP-----
Q608B9_MET YLDPVLHGAY PDELAISAL ---HWPSFES EDLRVIQEP I ---DYLGINY YTRAVVRHDP SGG-----
A0H0U4_9CH FLEPLFRRY PADLVAHFAR RL-IFAPVKP GDMEIISQPI ---DFLGINY YTRTLVVFDM REP-----
Q3E343_CHL FLEPLFCRRY PADMIDHFAR RL-VFAPIQP GDFEIIQPI ---DFLGINY YTRTLVVFDM REP-----
BGLS_CALSA FLDPVLKGSY PQKLLDYLVQ KDLLDSQKAL SMQQEVKENF IFPDLGINY YTRAVRLYDE NSS-----
Q2ZJP6_CAL FLDPVLKGSY PQKLLDYLVQ KDLLDSQKAL SMQQEVKENF IFPDLGINY YTRAVRLYDE NSS-----
Q8GEB4_THE FLDPILKGY PESPFDRPPP -----VPILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GT-----
Q8GEB0_9DE FLDPILKGY PESPFDRPPP -----VPILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GT-----
Q8GHE5_THE FLDPILGRY PESPFQDPPP -----TPILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GP-----
Q8GEB2_THE FLDPILGRY PESPFQDPPP -----TPNLS RDLELVARPL ---DFLGVNY YAPVRVAPGT GP-----
Q8GEB1_9DE FLDPILGRY PESPFQDPPP -----VPILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GP-----
Q9L794_9DE FLDPILGRY PESPFQDPPP -----APILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GP-----
Q9RA58_9DE FLDPILKGY PESPFDRPPP -----VPILS RDLELVARPL ---DFLGVNY YAPVRVAPGT GT-----
Consensus %!#PilkgrY Redlledyaq ...dfspvqd gdleI!srpi ...DflGvNy Ytravvrydp gng.....

351 420

A0LR48_ACI GQGTGETWPG CPDIQFPEWP FRR TAMGWPI DPSGLYELLI RLNRDYPR-P IMITENGAAF DDVVT--DNN
Q47RE2_THE PDEGYSPSVG SEHVVEVDPG LPVTAMGWPI DPTGLYDTLT RLANDYPGLP LYITENGAAF EDKVV--DG-
Q9LAV5_THE PDEGYSPSVG SEHVVEVDPG LPVTAMGWPI DPTGLYDTLT RLANDYPGLP LYITENGAAF EDKVV--DG-
Q1AUD7_RUB -TGMRFGE LG AETVL--PEG VGT TAMGWPV EPDGLAEILV RVKEEYRDLV VFTTENGCAV HDYID--PEG
Q2ADJ1_9FI -----L FNAEVVKMED RPT TEMGW EI YPQGLYDILV RVNKEYTDKP LYITENGAAF DDKLT--EEG
Q3CKP2_THE -----M LKAENVPGPG -KR TEMGW EI SPESLYDLLK RLDREYTKLP MYITENGAAF KDEVT--EDG
Q60026_THE -----M LKAENVPGPG -KR TEMGW EI SPESLYDLLK RLDREYTKLP MYITENGAAF KDEVT--EDG
Q8RCQ8_THE -----M LKAENVPGPG -KR TEMGW EI SPESLYDLLK RLDREYTKLP MYITENGVAF KDEVT--EDG
Q0R5R7_ALI -----L LSVRHVPGE G -PR TEMDWEV YPDGLYD LLS RLRRDYGDIP IYITENGAAF DDRV---QDG
A0ICG5_9CH -----G LRYAHRPE G -EY TQMDWEV HPASLRLLLE RLHRDYAPT T LYITENGAAY PDEV S--SDG
Q1FAN6_9CH -----G LRYAHERPE G -EY THMDWEV HPDSLRLRLLLE RLHA EYAPGV LYITENGAAY PDEIA--PDG
Q2WGB4_9BA -----AG VLQSEEVDMG LTKTDIGWPV ESRGLYEFMH YL-QKYGNVD VYITENGACI NDDLE---NG
BGLA_BACCI -----GG MLSSEAI SMG APKTDIGWEI YAEGLYD LLR YTADKYGNPT LYITENGAAY NDGLS--LDG
A0HPV3_9TH ----- -MKFSYVSGD LPK TEMGW EI YPQGLFDMLI YLKERYR LP- LYITENGMAG PDKV---ENG
Q8GKQ8_9TH ----- -LGFSYVQGD LPK TEMGW EI YPQGLFDMLV YLKERYKLP- LYITENGMAG PDKL---ENG
BGLA_THEME ----- -AKVSFVERD LPK TAMGW EI VPEGIYWILK KVKEEYNPPE VYITENGAAF DDVVS--EDG
BGLA_THENE ----- -ARVSFVERD LPK TAMGW EI VPEGIYWILK GVKEEYNPPE VYITENGAAF DDVVS--EGG
Q608B9_MET -----P LEVTAVPQRG VEH TEMGW EV YPQGLKDVLA WVKARYGDIP LYITENGAAF ADPEG--ENG
A0H0U4_9CH ----- FLMTKQVPGP LPK TAMGW EI YPDCIVEALQ KA-REYTSLP LYITENGAAF DDP PPGPNDQ
Q3E343_CHL ----- FTMAKQVPGP LPK TAMGW EI YPDCIVEALQ KA-RS YTSLP LYITENGAAF DDP PPGPNDQ
BGLS_CALSA -----W IFPIRWEHPA GEY TEMGW EV FPQGLFDLLI WIKESYPQIP IYITENGAAY NDIVT--EDG
Q2ZJP6_CAL -----W IFPIRWEHPA GEY TEMGW EV FPQGLFDLLI WIKESYPQIP IYITENGAAY NDIVT--EDG
Q8GEB4_THE ----- -LPVRYLPPE GPATAMGW EV YPEGLHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q8GEB0_9DE ----- -LPVRYLPPE GPATAMGW EV YPEGLHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q8GHE5_THE ----- -LPVRYLPPE GPV TAMGW EV YPEGLYHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q8GEB2_THE ----- -LPVRYLPPE GPV TAMGW EV YPEGLYHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q8GEB1_9DE ----- -LPVRYLPPE GPV TAMGW EV YPEGLYHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q9L794_9DE ----- -LPVRYLPPE GPV TAMGW EV YPEGLYHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Q9RA58_9DE ----- -LPVRYLPPE GPATAMGW EV YPEGLYHLLK RLGREVP-WP LYITENGAAY PDLWT--GEA
Consensuslpvryvppg lpkTamgWe! ypeglydIlk rIkreydpIp l%!TENGAaf pdkvt..edg

Anexo B. Alineamientos de las secuencias de aminoácidos de las proteínas homólogas a βgl-B

421 490

A0LR48_ACI **R**VRDPARAAY **I**QEHLAALHQ **A**IADGVVDRG **Y**YLWSLIDNF **E**WAYGYSRRF **G**IVYVDFET- **Q**ERIKDSGY
Q47RE2_THE **A**VHDTERIAIY **L**DSLHRAAHA **A**IEAGVPLKG **Y**FAWSFMDNF **E**WALGYGKRF **G**IVHVDYES- **Q**TRTVKDSGW
Q9LAV5_THE **A**VHDTERIAIY **L**DSLHRAAHA **A**IEAGVPLKG **Y**FAWSFMDNF **E**WALGYGKRF **G**IVHVDYES- **Q**TRTVKDSGW
Q1AUD7_RUB **E**VNDVERVAY **L**DAHLRAAHA **A**LERGVDLRG **Y**MVWSLLDNF **E**WAEGYSKRF **G**LVYVEYGS- **Q**RRVPKRSAR
Q2ADJ1_9FI **K**IHDEKRINY **L**GDHFKQAYK **A**LKDGVPVLRG **Y**VVWSLMDNF **E**WAYGYSKRF **G**LIYVDYENG **N**RRFLKDSAL
Q3CKP2_THE **R**VHDDERIEY **I**KEHLKAAK **F**IGEGGNLKG **Y**FVWSLMDNF **E**WAHGYSKRF **G**IVYVDYET- **Q**KRILKDSAL
Q60026_THE **R**VHDDERIEY **I**KEHLKAAK **F**IGEGGNLKG **Y**FVWSLMDNF **E**WAHGYSKRF **G**IVYVDYTT- **Q**KRILKDSAL
Q8RCQ8_THE **R**VHDYERIEY **I**KEHLKAIAR **F**IEEGGNLKG **Y**FVWSLLDNF **E**WAHGYSKRF **G**IVYVDYET- **Q**KRILKDSAF
Q0R5R7_ALI **G**VHDADRVAIY **L**AGHFAAAHR **F**IEEGGNLKG **Y**VVWSLMDNF **E**WAFGYTKRF **G**IVYVDYDT- **L**ARIPKDSYF
A0ICG5_9CH **G**VHDPDRVRY **I**ARHLAACHD **A**IAAGVPLRG **Y**FVWSLMDNF **E**WAFGYSRRF **G**IVYVDYAT- **Q**RRIPKDSAL
Q1FAN6_9CH **G**VHDPDRIRY **I**ARHLVACHD **A**ITAGVPLRG **Y**FVWSLMDNF **E**WAFGYSRRF **G**IIYVDYAT- **Q**RRIPKDSAL
Q2WGB4_9BA **K**INDDRRIAY **Y**EQHLAQIHR **I**INDGINLKG **Y**MAWSLMDNF **E**WAEGYRMRF **G**LVHVDYRS- **L**VRTPKESFY
BGLA_BACCI **R**IHDQRIDY **L**AMHLIQASR **A**IEDGINLKG **Y**MEWSLMDNF **E**WAEGYGMRF **G**LVHVDYDT- **L**VRTPKESFY
A0HPV3_9TH **K**VIDDYRIEY **L**EKHFEKALE **A**INAGVNLKG **Y**FIWSLLDNF **E**WAYGYSKRF **G**IVYVDYNT- **Q**KRILKDSAQ
Q8GKQ8_9TH **R**VHDNYRIEY **L**EKHFEKALE **A**INAGVDLKG **Y**FIWSLMDNF **E**WAYGYSKRF **G**IIYVDYNT- **Q**KRILKDSAL
BGLA_THEMA **R**VHDQNRIDY **L**KAHIGQAWK **A**IQEGVPLKG **Y**FVWSLLDNF **E**WAEGYSKRF **G**IVYVDYST- **Q**KRIVKDSGY
BGLA_THENE **K**VHDQNRIDY **L**RAHIEQVWR **A**IQDGVPLKG **Y**FVWSLLDNF **E**WAEGYSKRF **G**IVYVDYNT- **Q**KRIKDSGY
Q608B9_MET **R**IDDTRRIAY **Y**RSHLRALHE **A**IAQGVVDRG **Y**FAWSLLDNF **E**WTYGYARRF **G**LVQVDPLT- **Q**RRIPKASAG
A0H0U4_9CH **I**VEDPDRVAIY **L**RSHIAACHR **A**LTAGIDLKG **Y**FVWTLMDNF **E**WAKGLSKRF **G**IIYTDYAT- **Q**RRVWKRSAH
Q3E343_CHL **V**VEDPQRVAIY **L**QSHIEACRR **A**IAAGIDLKG **Y**FVWTLMDNF **E**WAKGYSKRF **G**IIYTDYTT- **Q**RRVWKRSAF
BGLS_CALSA **K**VHDSKRIEY **L**KQHFEAARK **A**IENGVDLKG **Y**FVWSLMDNF **E**WAMGYTKRF **G**IIYVDYET- **Q**KRIKDSFY
Q2ZJP6_CAL **K**VHDSKRIEY **L**KQHFEAARK **A**IENGVDLKG **Y**FVWSLMDNF **E**WAMGYTKRF **G**IIYVDYET- **Q**KRIKDSFY
Q8GEB4_THE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q8GEB0_9DE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q8GHE5_THE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q8GEB2_THE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q8GEB1_9DE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q9L794_9DE **V**VEDPERVAIY **L**EAHVEAALR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Q9RA58_9DE **V**VEDPERVAIY **L**EAHVEGRLR **A**REEGVDLRG **Y**FVWSLMDNF **E**WAFGYTRRF **G**LYYVDFPS- **Q**RRIPKRSAL
Consensus **r**!hDpeR!aY **leaHleaahr** **aieeGvdIrg** **YfvwslsIDNF** **EwafGyskRF** **GivyvDyEt** **qrRipKdSal**

491 511

A0LR48_ACI **F**YSLVARTNT **I**AAP
Q47RE2_THE **W**YSRVMRNGG **I**FGQE
Q9LAV5_THE **W**YSRVMRNGG **I**FGQE
Q1AUD7_RUB **W**YAAVIRRG **P**EG
Q2ADJ1_9FI **W**YREVIEWKQ **V**EAN
Q3CKP2_THE **W**YKEVIQKNS **I**E
Q60026_THE **W**YKEVILDDG **I**ED
Q8RCQ8_THE **W**YKGVIEKGV **I**E
Q0R5R7_ALI **W**YQRVIREGG **L**VPAPPAETA **R**
A0ICG5_9CH **F**LRQVIAANA **L**TETQMFTFTR
Q1FAN6_9CH **F**MRQVIAANA **V**EEG
Q2WGB4_9BA **W**YQNVIKNNW **V**ETRI
BGLA_BACCI **W**YKGVISRGW **L**DL
A0HPV3_9TH **W**LKEFLRS
Q8GKQ8_9TH **W**LKEFLKS
BGLA_THEMA **W**YSNVVKNNG **L**ED
BGLA_THENE **W**YSNGIKNNG **L**TD
Q608B9_MET **F**YAEVAQTNG **A**VLDRDY
A0H0U4_9CH **V**YRDIIARNG **L**
Q3E343_CHL **V**YRDIIARNG **L**FGNESS
BGLS_CALSA **F**YQQYIKENS
Q2ZJP6_CAL **F**YQQYIKENS
Q8GEB4_THE **W**YRERIARAQ **T**
Q8GEB0_9DE **W**YRERIARAQ **T**
Q8GHE5_THE **W**YRERIARAQ **L**
Q8GEB2_THE **W**YRERIARAQ **L**
Q8GEB1_9DE **W**YRERIARAQ **T**GG SAR
Q9L794_9DE **W**YRERIARAQ **T**GG SAR
Q9RA58_9DE **W**YQ
Consensus **wyreviarng** **l**.....

ANEXO C. Formato de archivos **.in** y **.pqr** obtenidos por el servidor **PDB2PQR**.

Ambos archivos están editados con los parámetros bajo los cuales se evaluaron energéticamente β gl-B y todas las variantes generadas en el presente trabajo.

Archivo **.in** para β gl-B

```
read
  mol pqr BGLB.pqr
end
elec
  mg-auto
  dime 161 129 161
  cglen 111.2038 89.7974 106.3418
  fglen 85.4140 72.8220 82.5540
  cgcent mol 1
  fgcent mol 1
  mol 1
  npbe
  bcfl sdh
  ion 1 0.088 2.0
  ion -1 0.088 2.0
  pdie 4.0000
  sdie 78.5400
  srfm smol
  chgm spl2
  sdens 10.00
  srad 1.40
  swin 0.30
  temp 298.15
  calcenergy total
  calcforce no
end
elec
  mg-auto
  dime 161 129 161
  cglen 111.2038 89.7974 106.3418
  fglen 85.4140 72.8220 82.5540
  cgcent mol 1
  fgcent mol 1
  mol 1
  npbe
  bcfl sdh
  ion 1 0.088 2.0
  ion -1 0.088 2.0
  pdie 4.0000
  sdie 4.0000
  srfm smol
  chgm spl2
```

Anexo C. Formato de archivos .in y .pqr obtenidos por el servidor PDB2PQR

```
sdens 10.00
srad 1.40
swin 0.30
temp 298.15
calcenergy total
calcforce no
end
print elecEnergy 1 - 2 end
quit
```

Parte del archivo .pqr para β gl-B B

```
REMARK 1 PQR file generated by PDB2PQR (Version 1.2.1)
REMARK 1
REMARK 1 Forcefield Used: charmm
REMARK 1 Naming Scheme Used: charmm
REMARK 1
REMARK 1 pKas calculated by propka and assigned using pH 7.00
REMARK 1
REMARK 5 WARNING: Propka determined the following residues to be
REMARK 5 in a protonation state not supported by the
REMARK 5 charmm forcefield!
REMARK 5 All were reset to their standard pH 7.0 state.
REMARK 5
REMARK 5 CYS 41 A (negative)
REMARK 5
REMARK 5
REMARK 6 Total charge on this protein: -19.0000 e
REMARK 6
ATOM 1 N NTER 4 78.654 34.627 10.663 -0.3000 1.8500
ATOM 2 CA NTER 4 77.935 35.475 11.612 0.2100 2.2750
ATOM 3 C ASN 4 78.860 35.911 12.749 0.5100 2.0000
ATOM 4 O ASN 4 79.543 35.062 13.323 -0.5100 1.7000
ATOM 5 CB ASN 4 76.728 34.686 12.225 -0.1800 2.1750
ATOM 6 HB2 ASN 4 76.219 35.336 12.970 0.0900 1.3200
ATOM 7 HB1 ASN 4 77.085 33.781 12.757 0.0900 1.3200
ATOM 8 CG ASN 4 75.662 34.307 11.186 0.5500 2.0000
ATOM 9 HT1NTER 4 79.148 33.944 11.210 0.3300 0.2245
ATOM 10 HT2NTER 4 79.360 35.172 10.207 0.3300 0.2245
ATOM 11 HA NTER 4 77.565 36.351 11.093 0.1000 1.3200
ATOM 12 HT3NTER 4 77.958 34.211 10.070 0.3300 0.2245
ATOM 13 OD1 ASN 4 75.013 35.172 10.598 -0.5500 1.7000
ATOM 14 ND2 ASN 4 75.464 32.984 10.961 -0.6200 1.8500
ATOM 15 HD22 ASN 4 75.986 32.307 11.474 0.3000 0.2245
ATOM 16 HD21 ASN 4 74.749 32.730 10.314 0.3200 0.2245
ATOM 17 N THR 5 78.894 37.234 13.080 -0.4700 1.8500
ATOM 18 CA THR 5 79.609 37.811 14.229 0.0700 2.2750
ATOM 19 C THR 5 78.786 38.984 14.780 0.5100 2.0000
ATOM 20 O THR 5 78.291 39.831 14.027 -0.5100 1.7000
ATOM 21 HN THR 5 78.369 37.922 12.580 0.3100 0.2245
ATOM 22 HA THR 5 79.618 37.043 14.995 0.0900 1.3200
ATOM 23 CB THR 5 81.095 38.174 13.995 0.1400 2.2750
ATOM 24 HB THR 5 81.626 37.249 13.670 0.0900 1.3200
ATOM 25 OG1 THR 5 81.711 38.613 15.200 -0.6600 1.7700
```

Anexo C. Formato de archivos .in y .pqr obtenidos por el servidor PDB2PQR

ATOM	26	CG2	THR	5	81.327	39.259	12.916	-0.2700	2.0600
ATOM	27	HG21	THR	5	80.845	40.215	13.210	0.0900	1.3200
ATOM	28	HG22	THR	5	80.924	38.939	11.933	0.0900	1.3200
ATOM	29	HG23	THR	5	82.416	39.446	12.808	0.0900	1.3200
ATOM	30	HG1	THR	5	81.730	37.870	15.880	0.4300	0.2245
ATOM	31	N	PHE	6	78.602	39.013	16.123	-0.4700	1.8500
ATOM	32	CA	PHE	6	77.658	39.846	16.851	0.0700	2.2750
ATOM	33	C	PHE	6	78.395	40.779	17.825	0.5100	2.0000
ATOM	34	O	PHE	6	78.349	40.589	19.042	-0.5100	1.7000
ATOM	35	HN	PHE	6	79.086	38.317	16.707	0.3100	0.2245
ATOM	36	HA	PHE	6	77.122	40.470	16.147	0.0900	1.3200

Referencias

Referencias

1. Karshikoff, A. y Ladenstein, R. (2001), **Ion pairs and the thermotolerance of proteins from hyperthermophiles: a “traffic rule” for hot roads.** *TRENDS in Biochem. Sci.* 26: 550-556.
2. Lehmann M., Pasamontes L., Lassen S.F., Wyss M. (2000), Review. **The consensus concept for thermostability engineering of proteins.** *Biochimica et Biophysica Acta* 1543:408-415.
3. Lehmann M., and Wyss M. (2001). **Engineering proteins for thermostability: the use of sequence alignments versus racional design and directed evolution.** *Curr. Opin. Biotech.* 12:371-375.
4. Horovitz A., Serrano L., Avron B., Bycroft M. y Fersht A. (1990). **Strength and Cooperativity of Contributions of Surface Salt Bridges to Protein Stability.** *J. Mol. Biol.* 216:1031-1044.
5. Vetriani C., Maeder D., Tolliday N., Yip K., Stillman T., Britton K., Rice D., Klump H., and Robb F. (1998). **Protein thermostability above 100°C: A key role for ionic interactions.** *Biophysics* 95:12300–12305.
6. González G., Sanz J., González B., Hermoso J., y Polaina J. (2000). **Directed Evolution of β -Glucosidase A from *Paenibacillus polymyxa* to Thermal Resistance.** *The J. Biol. Chem.* 275:18:13708-13712.
7. Kumar S., Tsai C. and Nussinov R. (2000). **Factors enhancing protein thermostability.** *Protein Engineering* 13:3:179-191.
8. Panasik N., Brenchley J., Farber G. (2000). **Distributions of structural features contributing to thermostability in mesophilic and thermofilic α/β barrel glycosyl hidrolases.** *Biochimica et Biophysica* 1543:189-201.
9. Kumar S. and Nussinov R. (2001). **How do thermophilic proteins deal with heat?.** *CMLS, Cell. Mol. Life Sci.* 58:1216–1233.
10. Chakravarty S. and Varadarajan, R. (2002). **Elucidation of factors responsible for enhanced thermal stability of proteins: a structural genomics based study.** *Biochemistry* 41: 8152-8161.
11. Makhataдзе G., Loladze V., Ermolenko D., Chen X. and Thomas S. (2003). **Contribution of Surface Salt Bridges to Protein Stability: Guidelines for protein Engineering.** *J. Mol. Biol.* 327:1135-1148.
12. Yano J. and Poulos T. (2003). **New understanding of thermostable and peiozostable enzymes.** *Curr. Opin. Biotech.* 14:1-6.

Referencias

13. Bismuto E., Febbraio F., Limongelli S., Briante R. and Nucci R. (2003). **Dynamic Fluorescence Studies of β -Glycosidase Mutants From *Sulfolobus solfataricus*: Effects of Single Mutations on Protein Thermostability.** *Proteins* 51:10-20.
14. Robinson M., Alibés A. and Godzik A. (2006). **Contribution of Electrostatic Interactions, Compactness and Quaternary Structure to Protein Thermostability: Lessons from Structural Genomics of *Thermotoga maritima*.** *J. Mol. Biol.* 356:547-557.
15. Bogin O., Levin I., Hacham Y., Tel-or S., Peretz M., Frolow F., and Burstein Y. (2006). **Structural basis for the enhanced thermal stability of alcohol dehydrogenase mutants from the mesophilic bacterium *Clostridium beijerinckii*: contribution of salt bridging.** *Protein Science* 11:2561-2574.
16. Mizuguchi K., Sele M., and Cubellis M. (2007). **Environment specific substitution tables for thermophilic proteins.** *BMC Bioinformatics* 8(Suppl 1):S15.
17. Hendsch Z. and Tidor B. (1994). **Do salt bridges stabilize proteins? A continuum electrostatic analysis.** *Protein Science* 3:211-226.
18. Barlow D. J. and Thornton J. M. (1983). **Ion-pairs in proteins.** *J. Mol. Biol.* 168:867-885.
19. Loladze V., Ibarra-Molero B., Sanchez-Ruiz J.M. and Makhataдзе G. I. (1999). **Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface.** *Biochemistry*, 38:16419-16423.
20. Alsop E., Silver M. and Livesay D. (2003). **Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatics analysis.** *Protein Engineering*, 16:12:871-874.
21. Henrissat B. (1991). **A classification of glycosyl hidrolases base on amino acid sequence similarities.** *Biochem. J.*, 280:309-316.
22. Polaina J. (2004). **Estructura, función e ingeniería molecular de enzimas implicadas en la digestión de carbohidratos.** Mensaje Bioquímico, Vol XXVIII: 61-76. Universidad Nacional Autónoma de México.
23. Sanz-Aparicio J., Hermoso J. A., Martínez-Ripoll M., Lequerica J.L., Polaina J. (1998). **Crystal structure of beta-glucosidase A from *Bacillus polymyxa*: insights into the catalytic activity in family 1 glycosil hidrolases.** *J. Mol. Biol.* 275:491-502.
24. Isorna P., Polaina J., Latorre-García L., Cañada F., González B., Sanz-Aparicio J. (2007). **Crystal Structures of *Paenibacillus polymyxa* β -Glucosidase B Complexes Reveal the Molecular Basis of Substrate Specificity and Give New Insights into the Catalytic Machinery of Family I Glycosidases.** *J. Mol. Biol.*, 371:1204-1218
25. Arrizubieta M. and Polaina J., (2000). **Increased Thermal Resistance and Modification of the Catalytic Properties of a β -Glucosidase by Random Mutagenesis and *in Vitro* Recombination.** *The J. Biol. Chem.* 275:37:28843-28848

Referencias

26. Zubillaga R., García-Hernández E., Camarillo-Cadena M., León M., and Polaina J. (2006). **Effect of a New Ionic Pair on the Unfolding Activation Barrier of β -Glucosidase B. Protein & Peptide Letters**, 13:113-118
27. Kee S., Keun K., Kun j., Duck J., Ha S., Hee J., Gu H., Kwon S., Jaie Y., (2005). **Three forms of thermostable lactose-hydrolase from *Thermus* sp. IB-21: cloning, expression, and enzyme characterization.** *J. Biotech.* 116:337-346
28. Altschul S., Gish W., Miller W., Myers E. and Lipman D., (1990). **Basic Local Alignment Search Tool.** *J. Mol. Biol.* 215:403-410
29. Corpet F., (1988). **The method used is described in "Multiple sequence alignment with hierarchical clustering".** *Nucl. Acids Res.* 16:10881-10890.
30. **MOE 2005.06** Tutorials versión 2005.06 of CHEMICAL COMPUTING GROUP INC.
31. Dolinsky T., Nielsen J., McCammon A. and Baker N. (2004). **PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.** *Nucleic Acids Research*, Vol. 32:W665–W667
32. Gilson M. and Honig B. (1988). **Calculation of the Total Electrostatic Energy of a Macromolecular System: Solvation Energies, Binding Energies, and Conformational Analysis.** *Proteins: Structure, Function, and Genetics*, 4:7-18
33. Baker N., Sept D., Joseph S., Michael J. Holst M. and McCammon J. (2001). **Electrostatics of nanosystems: Application to microtubules and the ribosome.** *PNAS* 98:18:10037–10041
34. Baker N. **APBS Tutorial 2007.** Washington University in St. Louis, Department of Biochemistry and Molecular Biophysics, Center for Computational Biology.
35. Fogolari F., Brigo A., and Molinari H. (2002). **The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology.** *J. Mol. Recognit.* 15:377-392.
36. Wagoner J., and Baker N.A. (2004). **Salvation Forces on Biomolecular Structures: A Comparison of Explicit Solvent and Poisson-Boltzmann Models.** *J. Comput Chem* 25:1623-1629.
37. Miteva M. A., Tufféry P., and Villoutreix B. O. (2005). **PCE: web tools to compute protein continuum electrostatics.** *Nucl. Acids Res.* 33:W372-W375.
38. Lee M. S. and Olson M. A. (2005). **Evaluation of Poisson Solvation Models Using a Hybrid Explicit/Implicit Solvent Method.** *J. Phys. Chem. B.* 109:5223-5236.
39. Baker N.A. (2005). **Improving implicit solvent simulations: a Poisson-centric view.** *Curr Opin Struct Biol.* 15(2):137-143.
40. Cerutti D. S., Baker N.A., McCammon J. A. (2007). **Solvent reaction field potential inside an uncharged globular protein: A bridge between implicit and explicit solvent models?** *J. Chem. Phys.* 127:155101-155111.



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Iztapalapa

DIVISIÓN DE CIENCIAS BÁSICAS E INGENIERÍA

DEPARTAMENTO DE QUÍMICA

- ÁREA DE BIOFISICOQUÍMICA -

"DISEÑO DE MUTANTES TERMORRESISTENTES DE LA
 β -GLUCOSIDASA B DE *Paenibacillus polymyxa*, EMPLEANDO
HERRAMIENTAS DE BIOINFORMÁTICA"

T E S I S

QUE PRESENTA

I.B.Q.I MIRIAM ORDOÑEZ DIAZ

PARA OBTENER EL GRADO DE

MAESTRA EN CIENCIAS (QUÍMICA)

MÉXICO D.F.

PRIMAVERA 2008