University of Groningen

Prevention of retrogradation of starch

Maarel, van der, Marc Jos Elise Cornelis; Dijkhuizen, Lubbert; Binnema, Doeve Jacob; Veen ,van der, Bartele Andries; Vos, Arnold

Publication date:
2003

Citation for published version (APA):
Maarel, van der, M. J. E. C., Dijkhuizen, L., Binnema, D. J., Veen ,van der, B. A., & Vos, A. (2003). Prevention of retrogradation of starch. (Patent No. *WO03002728*).

(51) International Patent Classification⁷: C12N 9/10

(21) International Application Number: PCT/NL02/00427

(22) International Filing Date: 1 July 2002 (01.07.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1018426 29 June 2001 (29.06.2001) NL

(71) Applicant (for all designated States except US): NED-ERLANDSE ORGANISATIE VOOR TOEGEPAST-NATUURWETENSHAPPELIJK ONDERZOEK TNO [NL/NL]; Schoemakerstraat 97, NL-2628 VK Delft (NL).

(72) Inventors; and
(75) Inventors/Applicants (for US only): VAN DER MAAREL, Marc, Jos, Elise, Cornelis [NL/NL]; Laagveld 17, NL-9753 KG Haren (NL). DIJKHUIZEN, Lubbert [NL/NL]; Ter Borch 28, NL-9472 RB Zuid-laren (NL). BINNEMA, Doeve, Jacob [NL/NL]; Helene Swarthlaan 86, NL-9721 TX Groningen (NL). VAN DER VEEN, Bartele, Andries [NL/NL]; Zwitserlaan 8, NL-9231 HK Surhuisterveen (NL). VOS, Arnold [NL/NL]; Oliemuldersweg 134A, NL-9713 VH Groningen (NL).

(74) Agent: PRINS, A., W.; Nieuwe Parklaan 97, NL-2587 BN Den Haag (NL).

(81) Designated States (national): AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PREVENTION OF RETROGRADATION OF STARCH

(57) Abstract: The invention provides an isolated or recombinant nucleic acid derived from a nucleic acid encoding a polypeptide essentially having alpha-glucanotransferase activity but having essentially no hydrolysing activity, said isolated or recombinant nucleic acid encoding a polypeptide with hydrolytic activity.

Title: Prevention of retrogradation of starch.

Starch containing crops form an important constituent of the human diet and a large proportion of the food consumed by the world's population originates from them. Besides the use of the starch-containing plant parts directly as a food source, starch is harvested and used as such or chemically or enzymatically processed into a variety of different products such as starch hydrolysates, glucose syrups, fructose, starch or maltodextrin derivatives, or cyclodextrins. In spite of the large number of plants able to produce starch, only a few plants are important for industrial starch processing. The major industrial sources are maize, tapioca, potato, and wheat. In the European Union, 3.6 million tons of maize starch, 2 million tons of wheat starch, and 1.8 million tons of potato starch were produced in 1998 (DeBaere, 1999).

Starch is found as granules containing polymers of glucose linked to one another through the C1 oxygen, known as the glycosidic bond. This glycosidic bond is stable at high pH but hydrolyses at low pH. At the end of the polymeric chain, a latent aldehyde group is present. This group is known as the reducing end. Two types of glucose polymers are present in starch granules: (i) amylose and (ii) amylopectin. Amylose is a linear polymer consisting of up to 6,000 glucose units with alpha,1-4 glycosidic bonds. The number of glucose residues, also indicated with the term DP (degree of polymerization), varies with the origin. Amylose from e.g. potato or tapioca starch has a DP of 1,000 – 6,000 while amylose from maize or wheat amylose has a DP varying between 200 and 1,200. The avarage amylose content in starches can vary between almost 0 to 75%, but a typical value is 20-25%. Amylopectin consists of short alpha,1-4 linked linear chains of 10-60 glucose units and alpha,1-6 linked side chains with 15-45 glucose units. The average number of branching points in amylopectin is 5%, but varies with the botanical origin. The complete amylopectin molecule contains on avarge about 2,000,000 glucose units, thereby being one of the largest molecules in nature. The most commonly accepted model of the structure of amylopectin is the cluster model, in which the side chains are ordered in clusters on the longer backbone chains (see Buléon et al., 1998; Myers et al., 2000).

While amylopectin is soluble in water, amylose and the starch granule itself are insoluble in cold water. This makes it relatively easy to extract starch granules from

their plant source. When a water-starch slurry is heated, the granules first swell untill a point is reached at which the swelling is irreversible. This swelling processes is termed gelatinization. During this process, amylose leaches out of the granule and causes an increase in the viscocity of the slurry. Further increase in temperature then leads to maximum swelling of the granules and increased viscosity. Finally, the granules break apart resulting in a complete viscous colloidal dispersion. Subsequent cooling results in association of the amylose chains, resulting in the formation of insoluble aggregates. In diluted starch suspensions these aggregates precipitate, cooling of a concentrated colloidal starch dispersion results in the formation of an elastic gel. This retrogradation is primarily caused by the amylose; amylopectin, due to its highly branched organization, is less prone to retrogradation.

A large variety of bacteria employ extracellular or intracellular enzymes able to convert starch or glycogen, that thus can serve as energy and carbon sources (Fig. 2).

There are basically four groups of starch-converting enzymes: (i) endoamylases; (ii) exoamylases; (iii) debranching enzymes; and (iv) transferases.

Endoamylases are able to cleave alpha,1-4 glycosidic bonds present in the inner part (endo-) of the amylose or amylopectin chain. alpha-Amylase (EC 3.2.1.1) is a well known endoamylase. It is found in a wide variety of microorganisms, belonging to the Archaea as well as the Bacteria (Pandey et al., 2000). The end products of alpha-amylase action are oligosaccharides with varying length with an alpha-configuration and alpha-limit dextrins, which constitute branched oligosaccharides.

Enzymes belonging to the second group, the exoamylases, either exclusively cleave alpha,1-4 glycosidic bonds such as b-amylase (EC 3.2.1.2) or cleave both alpha,1-4 and alpha,1-6 glycosidic bonds like amyloglucosidase or glucoamylase (EC 3.2.1.3) and alpha-glucosidase (EC 3.2.1.20). Exoamylases act on the external glucose residues of amylose or amylopectin and thus produce only glucose (glucoamylase and alpha-glucosidase), or maltose and b-limit dextrin (b-amylase).

The third group of starch-converting enzymes are the debranching enzymes that exclusively hydrolyse alpha,1-6 glycosidic bonds: isoamylase (EC 3.2.1.68) and pullanase type I (EC 3.2.1.41). Pullulanases hydrolyse the alpha,1-6 glycosidic bond in pullulan and amylopectin, while isoamylase can only hydrolyse the alpha,1-6 bond in amylopectin. These enzymes exclusively degrade amylopectin, thus leaving long linear polysaccharides.

There are also a number of pullulanase type enzymes that hydrolyse both alpha,1-4 and alpha,1-6 glycosidic bonds. These belong to the group II pullulanase and are also referred to as alpha-amylase-pullulanase or amylopullulanase. The main degradation products are maltose and maltotriose.

The fourth group of starch-converting enzymes are transferases that cleave an alpha,1-4 glycosidic bond of the donor molecule and transfer part of the donor to a glycosidic acceptor with the formation of a new glycosidic bond. Enzymes such as amylomaltase (EC 2.4.1.25) and cyclodextrin glycosyltransferase (EC 2.4.1.19) form a new alpha,1-4 glycosidic bond while branching enzyme (EC 2.4.1.18) forms a new alpha,1,6 glycosidic bond.

Cyclodextrin glycosyltransferases have a very low intrinsic hydrolytic activity and make cyclic oligosaccharides with 6,7,or 8 glucose residues and highly branched high molecular weight dextrins, the cyclodextrin glycosyl-transferase limit dextrins. Cyclodextrins are produced via an intramolecular transglycosylation reaction in which the enzyme cleaves an alpha,1-4 glycosidic bond and concomitantly links the reducing to the non-reducing end (Takaha and Smith, 1999; van der Veen et al., 2000a).

Amylomaltases are very similar to cyclodextrin glycosyltransferases with respect to the type of enzymatic reaction. The major difference is that amylomaltase performs a transglycosylation reaction resulting in a linear product while cyclodextrin glycosyl-transferase gives a cyclic product. Another difference is that they essentially do not hydrolyse starch. Amylomaltases have been found in different microorganisms in which it is involved in the utilization of maltose or the degradation of glycogen (Takaha and Smith, 1999).

Glucan branching enzymes are invloved in the synthesis of glycogen in many microorganisms. They are responsible for the formation of alpha,1-6 glycosidic bonds in the side chains of glycogen and in general do not hydrolyse either. Although glycogen has been found in a large number of microorganisms (Preiss, 1984), only a limited number of microbial glucan branching enzymes have been characterized (Kiel et al., 1991; Kiel et al., 1992; Takata et al., 1994; Binderup and Preiss, 1998).

Most of the enzymes mentioned above belong to one family based on amino acid sequence homology: the alpha-amylase family or family 13 glycosyl hydrolases according to the classification of Henrissat (1991). This group comprises those enzymes that have the following features: (i) they possess a (bete/alpha)8 or TIM

barrel (Fig. 3) structure containing the catalytic residues; (ii) they have four highly conserved regions in their primary sequence (Table 1) which contain the amino acids that form the catalytic site, as well as some amino acids that are essential for the stability of the conserved TIM barrel topology (Kuriki and Imanaka, 1999); (iii) they act on alpha-glycosidic bonds and hydrolyse or transglycosylate this bond with retention of the alpha-anomeric configuration through a double displacement mechanism.

The enzymes that match the above mentioned criteria and belong to the alpha-amylase family are listed in Table 2.

During the last three decades, alpha-amylases have been exploited by the starch-processing industry as a replacement of acid hydrolysis in the production of starch hydrolysates. This enzyme is also used for removal of starch in beer, fruit juices, or from clothes and porcelain. Another starch-hydrolysing enzyme that is used on a large scale is thermostable pullulanase for the debranching of amylopectin. A recent application is directed at the use of maltogenic amylases as an anti-staling agents to prevent the retrogradation of starch in bakery products.

The baking industry is a large consumer of starch and starch modifying enzymes. Bread baking starts with dough preparation by mixing flour, water, yeast and salt and possibly additives. Flour consists mainly of gluten, starch, non-starch polysaccharides and lipids. Immediately after dough preparation, the yeast starts to ferment the available sugars into alcohols and carbon dioxide, which causes rising of the dough. Amylases can be added to the dough to degrade the damaged starch in the flour into smaller dextrins, which are subsequently fermented by the yeast. The addition of malt or fungal α-amylase to the dough results in increased loaf volume and improved texture of the baked product

After rising, the dough is baked. When the bread is removed from the oven, a series of changes start which eventually leads to deterioration of quality. These changes include increase of crumb firmness, loss of crispness of the crust, decrease in moisture content of the crumb and loss of bread flavor. All undesirable changes that do occur upon storage together are called staling. Retrogradation of the starch fraction in bread is considered to be very important in staling (Kulp and Ponte, 1981). Especially the extent of amylopectin retrogradation strongly correlates with the firming rate of bread (Champenois et al., 1999). Staling is of considerable economic

importance for the baking industry since it limits the shelf life of baked products. In the USA, for instance, bread worth more than 1 billion US$ is discarded annually.

To delay staling, to improve texture, volume and flavor of bakery products, several additives may be used in bread baking. These include chemicals, small

5   sugars, enzymes or combinations of these. Well known additives are: milk powder, gluten, emulsifiers (mono- or diglycerides, sugar esters, lecithin etc.), granulated fat, oxidant (ascorbic acid or potassium bromate), cysteine, sugars or salts (Spendler and Jørgensen, 1997). Rapid advances in biotechnology have made "new" enzymes available for the industry. Since enzymes are produced from natural ingredients, they

10   will find greater acceptance by the consumers as they demand for products without chemicals. Several enzymes have been suggested to act as dough and/or bread improvers, by modifying one of the major dough components. Examples are glucose oxidase, hemicellulase, lipase, protease and xylanase. These enzymes, however, do not act on the starch fraction itself. Enzymes active on starch have been suggested to

15   act as anti-staling agents. Examples are: α-amylases (De Stefanis and Turner, 1981; Cole, 1982), branching (Okada et al., 1984) and debranching (Carroll et al., 1987) enzymes, maltogenic amylases (Olesen, 1991), β-amylases (Würsch and Gumy, 1994), and amyloglucosidases (Vidal and Gerrity, 1979). Present nti-staling agents, however, often act to fast.

20   Originally, α-amylases were added during dough preparation to generate fermentable compounds. Besides generating fermentable compounds, α-amylases also have an anti-staling effect in bread baking, and they improve the softness retention of baked goods (De Stefanis and Turner, 1981; Cole, 1982; Sahlström and Bråthen, 1997). Despite a possible anti-staling effect, the use of α-amylases as anti-staling

25   agent is not widespread because even a slight overdose of α-amylase results in a sticky bread. Positive effects of delayed staling, on the contrary, are measured only after 3 to 4 days (Olesen,1991). The increased gummyness of α-amylase treated bread is associated with the production of branched maltodextrins of DP20-100 (De Stefanis and Turner, 1981). Debranching enzymes are claimed to strongly decrease the

30   problems associated with the use of α-amylases as anti-staling agents in baking. In this method a thermostable pullulanase, and an α-amylase are used together. The pullulanase rapidly hydrolyzes the branched maltodextrins of DP20-100 produced by the α-amylase, while they have little effect upon the amylopectin itself (Carroll et al.,

1987). Pullulanase thus specifically removes the compound responsible for the gummyness associated with α-amylase treated bakery products.

Branching enzyme is claimed to increase shelf life and loaf volume of baked goods (Okada et al., 1984; Spendler and Jorgensen, 1997). These effects are achieved by

5   modifying the starch material in the dough during baking. Improved quality of baked products is also obtained when the branching enzyme is used in combination with other enzymes, such as α-amylase, maltogenic amylase, cyclodextrin glycosyltransferase, β-amylase, cellulase, oxidase and/or lipase (Spendler and Jorgensen, 1997).

10   The use of cyclodextrin glycosyltransferase as dough additive is claimed to increase loaf volume of the backed product (Van Eijk and Mutsaers, 1995). The effect is suggested to result from the gradual formation of cyclodextrins in the dough after mixing.

Exo-amylases, such as β-amylase and amyloglucosidase, shorten the external side

15   chains of amylopectin by cleaving of maltose or glucose molecules, respectively. Both enzymes are suggested to delay bread staling by reducing the tendency of the amylopectin compound in bakery products to retrograde (Würsch and Gumy, 1994). Anti-staling effects of amylo-glucosidase in baking are claimed in a few patents (Van Eijk, 1991; Vidal and Gerrity, 1979). The synergetic use of an α- and a β-amylase is

20   also claimed to increase the shelf life of baked goods (Van Eijk, 1991).

Since α-amylases cause stickiness of backed goods, especially when overdosed, it was suggested that these problems could be solved using an exo-amylase, since they do not produce the branched maltooligosaccharides of DP20-100. Such enzymes, called maltogenic amylases, produce linear oligosaccharides of 2 to 6 glucose residues.

25   Maltogenic amylases producing maltose (Olesen, 1991), maltotriose (Tanaka et al., 1997) and maltotetraose (Shigeji et al., 1999a; Shigeji et al., 1999b) are claimed to increase the shelf life of bakery products by delaying retrogradation of the starch compound. Currently, a thermostable maltogenic amylase of *Bacillus stearothermophilus* (Diderichsen and Christansen, 1988) is used commercially in the

30   bakery industry. Although this enzyme has some endo-activity (Christophersen et al., 1998), it does act as an exo-acting enzyme during baking, modifying starch at a temperature when most of the starch starts to gelatinize (Olesen, 1991).

Cherry et al. (1999) described in detail the 3D-structure of the maltogenic alpha-amylase and used this to suggest specific amino acid modifications to obtain variants of the enzyme with improved product specificity, altered pH optimum, improved thermostability, increased specific activity, altered cleavage pattern and thus have an increased ability to reduce retrogradation of starch or staling of bread.

Cyclodextrins are cyclic alpha,1-4 linked oligosaccharides mainly consisting of 6, 7,or 8 glucose residues. The glucose residues in the rings are arranged in such a manner that the inside is hydrophobic thus resulting in an apolar cavity while the outside is hydrophilic. This enables cyclodextrins to form inclusion complexes with a variety of hydrophobic guest molecules. Specific cyclodextrins are required for complexation of guest molecules of specific sizes. The formation of inclusion complexes leads to changes in the chemical and physical properties of the guest molecules, such as stabilization of light- or oxygen sensitive compounds, stabilization of volatile compounds, improvement of solubility, improvement of smell or tast, or modification of liquid compounds to powders. These altered characteristics of the encapsulated compounds have led to various applications of cyclodextrins in analytical chemistry (Armstrong, 1988; Loung et al., 1995), agriculture (Saenger, 1980; Oakes et al., 1991), biotechnology (Allegre and Deratani, 1994; Szejtli, 1994), pharmacy (Albers and Muller, 1995; Thompson, 1997), food (Allegre and Detrani, 1994; Bicchi et al., 1999) and cosmetics (Allegre and Detrani, 1994).

A major drawback for the application of cyclodextrins on a large scale is that all enzymes used today produce a mixture of cyclodextrins. Two different industrial approaches are used to purify the cyclodextrin mixtures: selective crystallization of beta-cyclodextrin, which is relatively poorly water-soluble, and selective complexation with organic solvents. Major disadvantages of the latter method is the toxitiy, flammability, and need for solvent recovery (Pedersen et al., 1995). This makes the production of cyclodextrins too costly for many applications. Additionally, the use of organic solvents limits applications involving human consumption.

For the industrial production of cyclodextrins, starch is first liquefied by a heat-stable alpha-amylase and then the cyclization occurs with a cyclodextrin glycosyltransferase from *Bacillus macerans* (Riisgaard, 1990) sp. A major drawback of this process is that the cyclization reaction has to be performed at lower temperatures than the initial liquefaction because of the low thermostability of the bacillus cyclodextrin glycosyltransferase. The use of cyclodextrin glycosyltransferase

from thermophilic microoganisms can solve this problem. Thermostable cyclodextrin glycosyltransferases have been found in a *Thermoanaerobacter* species (Starnes, 1990; Norman and Jørgensen, 1992; Pedersen et al., 1995), *Thermoanaero-bacterium thermosulfurogenes* (Wind et al., 1995), and *Anaerobranca bogoriae* (Prowe et al.,

5   1996).

Cyclodextrin glycosyltransferases can also be used for the production of novel glycosylated compounds, making use of the transglycosylation activity. A commercial application is the glycosylation of the intense sweetener stevioside, isolated from the leaves of the plant *Stevia rebaudania*, thereby increasing solubility and decreasing

10   the bitterness (Pedersen et al. 1995).

Other cyclic products that can be generated from starch are cycloamyloses. These large cyclic glucans (DP >20) contain antiparalel helices, providing long cavities with a diameter similar to that of alpha-cyclodextrin. Unlike cyclodextrins, cycloamylose is formed by all transglycosylating enzymes of the alpha-amylase family

15   (Takata et al., 1996; Terada et al., 1997; Terada et al., 1999). Formation of cyclodextrins occurs by an intramolecular transglycosylation reaction whereas the formation of large cycloamylose molecules is the result of an intramolecular transglycosylation. To form cycloamylose, low concentrations of high molecular weight amylose are incubated with a relatively high amount of enzyme. This reaction

20   is therefore not based on a novel catalytic mechanism but is a direct effect of the limited availability of acceptor molecules. Production of cycloamylose is currently not done on an industrial scale.

alpha-Amylase, pullulanase, cyclodextrin glycosyltransferase, and maltogenic amylase are nowadays widely used by industry in various applications (Table 3).

25   alpha-Amylase probably has the most wide-spread use. Besides their use in hydrolysis leading to the saccharification or liquefaction of starch, these enzymes are also used for the preparation of viscous, stable starch solutions used for the warp sizing of textile fibers, the clarification of haze formed in beer or fruit juices, or for the pretreatment of animal feed to improve the digestibility. A growing new area of

30   application of alpha-amylases is in the fields of laundry and dish-washing detergents. A modern trend among consumers is to use colder temperatures for doing the laundry or dish-washing. At these lower temperatures the removal of starch from cloth and porcelain becomes more problematic. Detergents with alpha-amylases optimally working at moderate temperatures and alkaline pH can help to solve this problem.

Two starch-modifying enzymes of the alpha-amylase family that do not find large scale application yet are amylomaltase and branching enzyme. Application of branching enzymes is limited by the lack of commercially available enzymes that are sufficiently thermostable. A potentially interesting industrial application of

5   amylomaltase is the production of thermoreversible starch gels. As already indicated above, a normal untreated starch gel cannot be dissolved in water after it has retrograded. However, starch that has been treated with amylomaltase has obtained thermoreversible gelling characteristics: it can be dissolved numerous times upon heating. This behaviour is very similar to gelatine. Van der Maarel et al. (2000)

10   described this process using the amylomaltase from the hyperthermophilic bacterium *Thermus thermophilus*. Currently, no amylomaltases are commercially available and the thermoreversible starch gel is not produced on an industrial scale.

Table 1. The four conserved regions and the corresponding b-sheets found in the

amino acid sequence of amylomaltase and alpha-amylase family enyzmes.

Highlighted are the conserved active site amino acid residues. The following enzymes

were used for the alignment: amylomaltase of *Thermus aquaticus* (Terada et al.

5 1999); amylosucrase of *Neisseria polysaccharea* (Büttcher et al. 1997); CGTase:

cyclodextrin glucosyltranferase of *Bacillus circulans* 251 (Lawson et al. 1994);

CMDase: cyclomaltodextrinase of *Clostridium thermohydrosulfuricim* 39E

(Podkovyrov & Zeikus 1992); BE: branching enzyme of *Bacillus stearothermophilus*

(Kiel et al. 1991); isoamylase of *Pseudomonas amyloderamosa* (Amemura et al. 1988);

10 M. amylase: maltogenic alpha-amylase of *Bacillus stearothermophilus* (Cha et al.

1998); pullulanase of *Bacillus flavocaldarius* KP 1228 (Kashiwabara et al. 1999);

sucrose Pase: sucrose phosphorylase of *Escherichia coli* K12 (Aiba et al. 1996); Taka-

amylase: alpha-amylase of *Aspergillus oryzae* (Matsuura et al. 1980). b2, b4, b5, and

b7 indicate the beta-sheet in which this region is present.

15

Region Ib2   Region IIb4 Region IIIb5 Region IVb7

Amylomaltase EALGIRIIGDMPIFVAED LFHLVRIDHFRG VPVLAEDLGVI
VVYTGTHDNDT

20 Amylosucrase HEAGISAVVDFIFNHTSN GVDILRMDAVAF VFFKSEAIVHP
VNYVRSHDDIG

CGTase   HAKNIKVIIDFAPNHTSP GIDGIRMDAVKH VFTFGEWFLGV
VTFIDNHDMER

CMDase   HDNGIKVIFDAVFNHCGY DIDGWRLDVANE AIIVGEVWHDA
25 FNLIGSHDTER

BE   HQAGIGVILDWVPGHFCK HVDGFRVDAVAN ILMIAEDSTDW
FILPFSHDEVV

Isoamylase HNAGIKVYMDVVYNHTAE GVDGFRFDLASV LDLFAEPWAIG
INFIDVHDGMT

30 M.amylase   HQKAIRVMLDAVFNHSGY DIDGWRLDVANE AYILGEIWHDA
FNLLGSHDTPR

Pullulanase HAHGVRVILDGVFNHTGR GVDGWRLDVPNE AYIVGEIWEEA
MNLLTSHDTPR

Sucrose Pase LGECSHLMFDFVCNHMSA GAEYVRLDAVGF TVIITETNVPH
FNFLASHDGIG
Taka-amylase HERGMYLMVDVVANHMGY SIDGLRIDTVKH VYCIGEVLDGD
GTFVENHDNPR

5

Table 2. Enzymes of the alpha-amylase family that act on glucose-containing substrates, their corresponding E.C. number, the domain organization as far as it has been described, and main substrates.

Enzyme	E.C. number	Domains	Main substrate
amylosucrase	2.4.1.4		sucrose
sucrose phosphorylase	2.4.1.7		sucrose
glucan branching enzyme	2.4.1.18	A, B, F	starch, glycogen
cyclodextrin glycosyltransferase	2.4.1.19	A, B, C, D, E	starch
amylomaltase	2.4.1.25	A, B1, B2	starch, glycogen
maltopentaose-forming amylase	3.2.1.-	A, B, I	starch
alpha-amylase	3.2.1.1	A, B, C	starch
oligo-1,6-glucosidase	3.2.1.10	A, B	amylopectin
alpha-glucosidase	3.2.1.20		starch
amylopullulanase	3.2.1.41 or	A, B, H, G, 1	pullulan
cyclomaltodextrinase	3.2.1.54	A, B	cyclodextrins
isopullulanase	3.2.1.57		pullulan
isoamylase	3.2.1.68	A, B, F, 7	amylopectin
maltotetraose-forming amylase	3.2.1.60	A, B, C, E	starch
glucodextranase	3.2.1.70		starch
trehalose-6-phosphate	3.2.1.93		trehalose
maltohexaose-forming amylase	3.2.1.98		starch
maltogenic amylase	3.2.1.133	A, B, C, D, E	starch
neopullulanase	3.2.1.135	A, B, G	pullulan
malto-oligosyl trehalose	3.2.1.141		trehalose
malto-oligosyl threhalose	5.4.99.15		maltose

5

Table 3. Different fields of application of enzymes belonging to the alphalpha-amylase family

Application	Enzyme
Starch liquefaction	alpha-amylase
Starch saccharification	amyloglucosidase, pullulanase, maltogenic alpha-amylase, alpha-amylase, isoamylase
Laundry detergent and cleaners; reduction of haze formation in juices, baking, brewing, digestibility of animal feed, fiber and cotton desizing, sanitary waste treatment	alpha-amylase
Cyclodextrin production	cyclodextrin glycosyltransferase
Thermoreversible starch gels	amylomaltase
Cycloamylose	amylomaltase, branching enzyme, cyclodextrin glycosyltransferase

The invention provides an isolated or recombinant nucleic acid encoding a 4-alpha- or 6-alpha-glucanotransferase, which, in a preferred embodiment, is provided with hydrolytic activity, a or functional fragment thereof. In one embodiment, the invention provides such a nucleic acid encoding an amylomaltase, the wild types of which are generaly not known for any hydrolysing activity. 4-α-Glucanotransferase (e.g. EC 2.4.1.25, amylomaltase (AMase) or D-enzyme) forms a separate family (77) of glycosyl hydrolases. However, it is closely related to the alpha-amylase family or family 13 of glycosyl hydrolases. Unlike most members of this family of enzymes 4-α-glucanotransferase is not directly involved in starch degradation, but promotes metabolism of starch degradation products inside the cell (AMase), or is involved in starch biosynthesis (D-enzyme). Recently, however, the action of amylomaltase from *Thermus thermophilus* on starch has been described, resulting in the production of a thermoreversible gel. To investigate the enzymatic properties responsible for this action the *T. thermophilus malQ* gene has been cloned and expressed in *E. coli*, and

14

its sequence as here been provided, allowing purification of large amounts of enzyme, and manipulation of the gene.

In order to determine the AMase reaction specificity its action on maltooligosaccharides and soluble starch was analyzed. Although the enzyme is closely related to the α-amylase family, of the wild type enzyme no hydrolyzing activity could be detected. In the disproportionation reaction the enzyme was found to prefer longer oligosaccharides as donor substrates, while shorter oligosaccharides are efficiently used as acceptors. As observed for other amylomaltases, maltose is not cleaved off and hardly used as acceptor by the enzyme.

The complete lack of hydrolyzing activity of wild type AMase and its specificity for donor and acceptor substrates makes it a very interesting enzyme to be studied regarding reaction and product specificity.

In another embodiment, the invention provides a nucleic acid encoding a enzyme or polypeptide derived from said non-hydrolysing enzyme, now provided with hydrolysing acitivity. For example, interaction with hydrophobic amino acids, such as F366, which is highly conserved in amylomaltases, is involved in the reaction specificity of the enzyme. Hydrolyzing activity can be introduced by mutating this residue, or other hydrophobic residues such as F251 or Y54. This hydrolyzing activity has significant effects on product profiles of the enzyme, indicating the necessity of complete absence of hydrolysis for the function of the wild type enzyme (the production of longer oligosaccharides from short substrates). Now that the enzyme has been provided with hydrolysing activity, it can be used in preventing retrogradation of starch. Especially useful in such prevention is the use of a newly hydrolising enzyme as provided herein that is derived from thermostable transferase, which can be found in a thermophilic micro-organism Particulary provided is such an enzyme wherein said micro-organism comprises *Thermus thermophilus, Thermus aquaticus* or *Aquifex aeolicus.*

Also, the branching enzyme (BE) gene from *Aquifex aeolicus* (BE Aae) was cloned, sequenced (for he amino acid sequence see fig 4) and overexpressed in *E. coli.* The thermostable branching enzyme was purified to homogeneity, and biochemically characterized. The temperature optimum for activity was 80 ⁰C, which is the highest optimum known for branching enzymes as compared with the other known thermostable branching enzyme from *Bacillus stearothermophilus* (BE Bst) which

has a temperature optimum of 50 °C. This higher temperature optimum is very

useful in hydrolysing starch. Furthermore, BE Aae was determined to be

thermostable up to 90 °C compared with approximately 60 °C for BE Bst. Branching

enzymes (BE) catalyze the formation of alpha- 1,6-glucosidic linkages in two steps

5   (pres. via cov. interm.). The first step is the cleavage of an alpha-1,4-glucosidic

linkage followed by a transfer of the oligosaccharide to the 6-position of another

glucose present within an alpha-1,4 glucosidic chain. This results in the branching

points present in starch and glycogen. It has been shown that a lot of organisms are

capable of producing starch or glycogen and express BE in order to do so. From

10   various sources the BE has been cloned and characterized. It has been shown that

BE's belong to the alpha-amylase family and that they posses the four conserved

regions present within the family. A 3D model of the BE from *Aquifex aeolicus* has

been designed. The crystal structure of isoamylase from *Pseudomonas* was used for

modelling using the program Swiss-Pdb viewer. All amino acids that are conserved in

15   the catalytic center within thealpha-amylase family were present in the active site of

the 3D-model of BE from *Aquifex aeolicus*. The most striking feature was the present

of hydrophobic residues (see fig 5) at the putative acceptor site. Alignment of

branching enzymes showed that these residues are highly conserved (see fig 4) .

These residues are mutated to more hydrophilic residues, for example according to

20   the table below

Table 4 Active site residue mutagenesis of BE.

function	mutant
acceptor site	W276Q W367Q
	W385Q M387S F458S
	Y460S
donor site	Y512S
catalytic site	D311N E362Q D430N

Now that a branching enzyme has been provided with hydrolysing activity, it can be

25   used in preventing retrogradation of starch. Especially useful in such prevention is

the use of a newly hydrolising enzyme as provided herein that is derived from

thermostable transferase, which can be found in a thermophilic micro-organism

Particulary provided is such an enzyme wherein said micro-organism comprises

Thermus thermophilus, Thermus aquaticus or *Aquifex aeolicus.*

In overview, the invention provides modified a transferase that is derived from or has an activity of an enzyme known under EC number 2.4.1.25 or 2.4.1.18, with added hydrolysing activity. These are derivable from   a nucleic acid according to the invention provided with a mutation leading to an alteration or loss of a codon originally encoding a hydrophobic amino acid located in or around a acceptor, a donor or a catalytic site extending from a TIM barrel structure of said transferase. Such mutation is preferably provided by site-directed mutagenesis, wherein said codon originally encoding a hydrophobic amino acid is altered into a codon encoding an amino acid which is substantially less hydrophobic. Preferably, the hydrophobic amino acid to be changed comprises phenylalanine, tryptophan or tyrosine, and is located at aor around the positions as indicated herein in the (beta/alpha)8 or TIM barel structure of the enzyme. For example, a nucleic acid is provided wherein said change in hydrophobic amino acid is located at or around an amino acid position essentially corresponding to amino acid position 54, 251, 258 or 366 of amylomaltase of *Thermus thermophilus* HB8. Furthermore, the invention provides a vector comprising a nucleic acid according to the invention and a host cell comprising a vector or a nucleic acid according to the invention.

As said, and further explained in the detailed descriptuion herein, the invention provides a method for providing a polypeptide or fragment thereof essentially having alpha glucanotransferase acitivity but having essentially no hydrolysing activity with specific hydrolysing activity said method comprising providing a nucleic acid encoding such a transferase with a mutation leading to an alteration or loss of a codon originally encoding a hydrophobic amino acid located in or around a acceptor, a donor or a catalytic site extending from a TIM barrel structure of said transferase, and provides a polypeptide obtainable therewith.

The invention also provided use a polypeptide or fragment according to the invention in reducing retrogradation of starch, such as in reducing retrogradation of amylopectine, particularly in reducing long-term retrogradation of amylpectine. The invention provides specific enzymes provided with one of more specific amino acid modifications to obtain variants of the enzyme with hydrlolysing activity, and thus with improved product specificity, altered pH optimum, improved thermostability when strating with a thermostable enzyme as provided herein, increased specific

activity, altered cleavage pattern. An enzyme as provided herein has increased ability to reduce retrogradation of starch or staling of bread.

Also, the invention provides use such a polypeptide or fragment in hydrolysing starch, said uses for example applied in the prevention or at least temporarily

5   avoiding of staling of bakery products such as bread, or as a replacement of acid hydrolysis in the production of starch hydrolysates. Such prevention of staling comprises use of a method for reducing retrogradation of starch comprising treating said starch with a polypeptide or fragment, such as a amylomaltase or branching enzyme provide with hydrolysing activity according to the invention. Improved

10   quality of baked products is further obtained when the alpha-glucanotransferase (e.g. amylomaltase or branching enzyme) provided with hydrolysing activity according to the invention is used in combination with other enzymes, such as α-amylase, maltogenic amylase, cyclodextrin glycosyltransferase, β-amylase, cellulase, oxidase and/or lipase Furthermore, the invention provides a bakery ingredient comprising a

15   polypeptide according to the invention and a bakery product such as bread comprising a polypeptide according to the invention. The invention is further explained in the detailed description provided herewith.

Figure legends

Figure 1: Overall secondary structure of the amylomaltase from *Thermus*
thermophilus. The central (b/a)₈ barrel is shown; this barrel consists of 8 β-sheets,
depicted as arrows, surrounded by 8 α-helices, depicted as spirals. The amino acid
residues constituting the catalytic site extend from this barrel into the active site
surrounded by subdomains B1, B2 and B3 respectively. Amino acid residues involved
in binding of the donor and acceptor substrates are located in and extending from
subdomain B1 and loops protruding from the (b/a)₈ barrel.

Figure 2: A model showing the binding of a maltoheptaose substrate in the active
site of the *T. thermophilus* amylomaltase. The sugar residues are numbered
according to the general subsite labeling scheme proposed for all glycosyl hydrolases
by Davies et al. (Biochem. J. 1997, 321: 557-559), in which the glycosidic bond
between -1 and +1 is the bond which is cleaved, and the substrate reducing end is at
position +3. The positively numbered subsites constitute the acceptor binding site.
The following amino acid residues are shown: (i) The catalytic residues Asp293 and
Glu340; (ii) those involved in interactions with the substrate by hydrogen bonds,
which are indicated by dotted lines; (iii) the aromatic amino acids involved in
hydrophobic stacking interactions, being Tyr54, Trp258, Phe251, and Phe366. The
model was constructed manually with the program O (Jones et al. 1991 Acta
Crystallogr. D55, 849-861) on basis of the 3D structures of a porcine pancreatic a-
amylase-hexasaccharide complex (Machius et al. 1996, J. Mol. Biol. 260, 409-421) and
a cyclodextrin glycosyltransferase-maltononaose complex (Uitdehaag et al. 1999,
Nature Struct. Biol. 6, 432-436). For clarity the model does not show the conserved
catalytic site residues Tyr 59, Arg 291, His 294 and Asp 395.

Figure 3: The amino acid sequence alignment of *Aquifex aeolicus* branching enzyme
(glgB Aqu) with *Pseudomonas amyloderamosa* isoamylase (isoamyla) used for
constructing the 3-D model of the *Aquifex aeolicus* branching enzyme. Symbols
represent the following: dots, functionally similar amino acids; *, identical amino
acids; s, amino acids present in a β-sheet; amino acids present in an α-helix. b1-b8

and a1-a8 represent the alternating β-sheets and α-helices, respectively, comprising the (β/α)8 barrel.

Figure 4: Detailed overview of the active site of *Aquifex aeolicus* branching enzyme, showing the catalityc amino acid residues Asp311 (D311 cat.res.), Glu362 (E362 cat.res.), and Asp430 (D430 cat.res.) and the hydrophobic amino acid residues surrounding the catalytic site Trp276 (W276), Trp367 (W367), Trp385 (W385), Met387 (M387), Phe458 (F458), Tyr460 (Y460), and Tyr512 (Y512).

Detailed description

Kinetic analysis of amylomaltase from *Thermus thermophilus* HB8: donor and acceptor specificities

5

Family 77 of glycosyl hydrolases consists of a single group of enzymes; 4-α-glucanotransferases (EC 2.4.1.25, amylomaltase (AMase) or D-enzyme). AMase is found in prokaryots and promotes metabolism of starch degradation products inside the cell as shown for *Escherichia coli*. In other organisms, lacking other enzymes

10   required for growth on oligosaccharides (p.e. maltodextrin phosphorylase), it may be involved in glycogen metabolism as suggested for *Aquifex aeolicus*. D-enzyme is found in plants and is reported to be involved in in starch metabolism. Recent studies on *Chlamydomonas rheinhardtii* show that D-enzyme is essential for biosynthesis of starch.

15   Sequence comparisons and 3-D structure similarities show that AMase is closely related to the á-amylase family or family 13 of glycosyl hydrolases. The a-amylase family is a very diverse group of enzymes that have the ability to modify and degrade starch. In the past, many 3D structures of enzymes from the a-amylase family have been elucidated, showing that all members share an (alpha/beta)8-barrel

20   architecture of the catalytic domain, containing a conserved active site that comprises seven amino acid residues. For this reason, it is thought that all members of the a-amylase family catalyze the same reaction cycle. This is suggested to proceed according to a two-step a-retaining mechanism. In the first step an a-glycosidic bond is cleaved in the substrate and a covalently bound enzyme-glycosyl intermediate is

25   formed. In the second step, the leaving group is exchanged for an acceptor molecule, which is then linked via a new a-glycosidic bond to the intermediate.

Recently, amylomaltases from thermophile organisms like *Thermus aquaticus* and *Thermus thermophilus* HB8 have been isolated. These enzymes have a high thermostability, which makes them suitable for industrial applications, such as the

30   production of large cyclic glucans and the production of thermoreversible gels from starch. A 2.0 Å 3D structure of the amylomaltase from *Thermus aquaticus* shows that the enzyme consists of a compact (alpha/beta)8-barrel catalytic domain with three loop excursions that are probably responsible for part of the enzyme's specificity. In the catalytic site, 6 out of the 7 conserved residues of the a-amylase family are

21

present, showing the close relatedness between amylomaltase and the alpha-amylase family.

Here we describe the cloning and characterization of the *T. thermophilus* AMase. Further glycosyl hydrolase families 13 and 77 are compared regarding

5   reaction (mechanism and) specificity.

EXPERIMENTAL PROCEDURES

Escherichia coli TOP10 was used for recombinant DNA manipulations. AMase

10   (mutant) proteins were produced with *E. coli* BL21 (DE3).

DNA manipulations - Restriction endonucleases and DNA polymerase were purchased from Pharmacia LKB Biotechnology, Sweden, and used according to the manufacturer's instructions. DNA manipulations and calcium chloride transformation of *E. coli* strains were as described {350}.

15   *Cloning and expression of the T. thermophilus MalQ gene* - A *T. thermophilus* gene library was constructed by inserting the 4-8 kb fragments of a partial *Sau*3A digest of genomic DNA in the *Bam*HI site of pZerO. This construct was transformed to E. coli TOP10 cells and plated on LB agar plates. After replicaplating the transformants were screened for amylomaltase activity by overlaying the

20   motherplate with 5 ml of a 0.5 % soluble starch solution, incubating for 24 h at 70°C, and staining with 4 ml Lugol solution. Positive colonies showed a shift from blue to red staining due to the disproportionation of the starch chains by amylomaltase. The DNA sequence of one of these clones was determined using the dideoxynuleotide chain termination method on a cycle sequencer (Pharmacia)

25   The *malQ* gene was amplified with PCR using the following primers:

Forward: GGCAGC<u>CATATG</u>GAGCTTCCCCGCGCTTTCGG

Reverse: GCAGCC<u>AGATCT</u>AGAGCCGTTCCGTGGCCTCGGC

The PCR product was digested with *Nde*I (<u>CATATG</u>) and *Bgl*II (<u>AGATCT</u>, overhang compatible with *Bam*HI) and ligated with either plasmid pET9c or plasmid pET15b

30   digested with *Nde*I and *Bam*HI, resulting in pGJ6002 or pCCBmalQ, respectively. Transformation of these plasmids resulted in expression of the native enzyme (pGJ6002) or of the amylomaltase with an N-terminal His6-tag (pCCBmalQ).

Production and purification of AMase - For the production of AMase protein *E. coli* BL21(DE3), containing the pCCBmalQ vector, was grown overnight in a 1 l

flask with 250 ml LB medium containing ampicilin.

Protein determination - Protein concentrations were determined with the

Bradford method {63} using the Bio-Rad reagent and bovine serum albumin as a

standard (Bio-Rad Laboratories, Richmond, CA, USA).

5   *Enzyme assays* - All assays were performed in a 25 mM sodium maleate buffer

(pH 6.5) at 70 °C.

Disproportionation reaction - Disproportionation activities were determineded

using the ability of AMase to release glucose from oligosaccharides. Various

concentrations (upto 50 mM) of (mixtures of) oligosaccharides (G2-G7) were

10   incubated with appropriatly diluted enzyme. For the determination of donor

specificity different concentrations of maltooligosaccharides as donor and methyl-α-

D-glucose as acceptor.At regular time intervals 50 il samples were taken and added

to 200 il GOD-PAP reagent (Roche) to measure the amount of glucose released.

Hydrolyzing activities were measured as described earlier using 1% soluble

15   starch (Lamers & Pleuger, Belgium) as substrate and dinitrosalicylic acid to

determine the number of reducing ends {680}.

In above assays 1 U of activity is defined as the amount of enzyme required for

the processing of 1 :mole of donor substrate per minute. Kinetic parameters were

fitted using the computer program Sigma Plot (Jandel Scientific).

20   *Product formation from oligosaccharides* was analyzed by HPLC. For this

purpose 1 ml of a 25 mM G3, G5, or G7 solution was incubated with 0.1 U AMase at

70 °C for 8 h. Samples were taken at regular time intervals and the products formed

were applied to a 25 cm Econosphere-NH2 5 micron column (Alltech Associates Inc.

USA) eluted with acetonitrile/water (60/40, v/v) at a flow rate of 1 ml per min.

25

In the assay for the disproportionation reaction various oligosaccharides (G2-

G7) were used as single (donor and acceptor) substrate. The K_M and Vmax values for

the formation of glucose varied with the different oligosaccharides. The highest Vmax

is observed for G4, which also shows the highest affinity. No activity on G2 was

30   observed. Adding G3 to the G5 reaction mixture resulted in a further increase in

activity (Fig. 2), whereas the addition of G2 had no effect (not shown). At high G3

concentrations a decrease in activity is observed, indicating competition between G3

and G5.

The donor specificity of AMase was further investigated using the various

oligosaccharides as donor and M-α-DG as acceptor substrates. Fig 3 shows that the addition of this monosaccharide clearly affects disproportionation activities, especially with the lower concentrations of oligosaccharide (donor) substrates. At higher M-α-DG concentrations the monosaccharide has an inhibitory effect. At lower

5   concentrations, however, it can efficiently be used as acceptor, allowing a clearer determination of affinities of the different oligosaccharides for the donor binding site. G5, showing the lowest KM value, clearly is the best donor substrate. Combined with the above observation of the stimulating effect of G3 on disproportionation of G5, this suggests that G3 is a better acceptor substrate.

10   *Hydrolyzing activity* on soluble starch was investigated, but even overnight incubation did not result in an increase of reducing ends, thus no hydrolyzing reaction is performed by the enzyme. Furthermore the enzyme was incubated with 4-nitrophenyl-α-D-maltoheptaoside-4-6-O-ethylidene (EPS; Boehringer Mannheim) (a maltoheptasaccharide which is blocked at the non-reducing end and with a para-

15   nitrophenyl group at its reducing end). This compound is generally used for the detection of α-amylase activity. However also with this substrate no hydrolyzing activity was observed. Furthermore it reacted very weakly when accepting oligosaccharides were added, suggesting that amylomaltase is an exo-acting enzyme, requiring the presence of a non-reducing end glucose.

20   *The oligosaccharide formation of AMase* was analyzed with HPLC (Fig. 4). With G3 as substrate the initial products were G1 and G5 (see Fig 4.a). After an initial lagfase, the production of G1 increased, while G3 decreased, however with a less significant increase of G5. Various larger oligosaccharides are produced, indicating that the initial product (G5) is used as donor and G3 is mainly used as

25   acceptor. With G5 as substrate the predominant initial products were G3 and G7, although also considerable amounts of G1, G4, G6, and G9 were formed. In both cases little maltose is produced initially, as observed previously (T. aquaticus, potato). The final production of maltose is probably caused by the tranfer of glucose from the donor to a glucose acceptor. This supported by the early formation of G4 from G5,

30   which indicates transfer of a glucose moyety, and by the above results with MαDG, which indicate that glucose can indeed be used as acceptor.

AMase is the ultimate disproportionating enzyme, producing a variety of (long) oligosaccharides from short substrates. In the disproportionation reaction the enzyme has a preference for longer oligosaccharides to be used as donor while shorter

oligosaccharides except maltose are efficiently used as acceptor. One of the requirements of doing this efficiently is a low hydrolyzing activity, which is extremely well met in AMase. The complete lack of hydrolyzing activity of this enzyme makes it a very interesting enzyme to be studied regarding reaction specificity in the α-

5   amylase family.

Reaction kinetics of the disportionation of oligosaccharides

DP	Km	Vmax
2	nd	nd
3	10	90
4	3.4	281
5	4.6	235
6	3.5	143
7	4.5	108

nd = not detectable

Affinity constants for oligosaccharides using MαDG as acceptor

DP	0 mM	4 mM	10 mM
3	8.0	4.7	12.2
4	3.4	3.0	2.9
5	4.6	1.5	2.7
6	3.5	2.5	2.8

Kinetic analysis of amylomaltase from *Thermus thermophilus* HB8: donor and acceptor specificities

Family 77 of glycosyl hydrolases consists of a single group of enzymes; 4-α-

glucanotransferases (EC 2.4.1.25, amylomaltase (AMase) or D-enzyme). AMase is found in prokaryots and promotes metabolism of starch degradation products inside the cell as shown for *Escherichia coli*. In other organisms, lacking other enzymes required for growth on oligosaccharides (p.e. maltodextrin phosphorylase), it may be involved in glycogen metabolism as suggested for *Aquifex aeolicus*. D-enzyme is found in plants and is reported to be involved in in starch metabolism. Recent studies on *Chlamydomonas rheinhardtii* show that D-enzyme is essential for biosynthesis of starch.

Sequence comparisons and 3-D structure similarities show that AMase is closely related to the á-amylase family or family 13 of glycosyl hydrolases. The a-amylase family is a very diverse group of enzymes that have the ability to modify and degrade starch. In the past, many 3D structures of enzymes from the a-amylase family have been elucidated, showing that all members share an (alpha/beta)8-barrel architecture of the catalytic domain, containing a conserved active site that comprises seven amino acid residues. For this reason, it is thought that all members of the a-amylase family catalyze the same reaction cycle. This is suggested to proceed according to a two-step a-retaining mechanism. In the first step an a-glycosidic bond is cleaved in the substrate and a covalently bound enzyme-glycosyl intermediate is formed. In the second step, the leaving group is exchanged for an acceptor molecule, which is then linked via a new a-glycosidic bond to the intermediate.

Recently, amylomaltases from thermophile organisms like *Thermus aquaticus* and *Thermus thermophilus* HB8 have been isolated. These enzymes have a high thermostability, which makes them suitable for industrial applications, such as the production of large cyclic glucans and the production of thermoreversible gels from starch. A 2.0 Å 3D structure of the amylomaltase from *Thermus aquaticus* shows that the enzyme consists of a compact (alpha/beta)8-barrel catalytic domain with three loop excursions that are probably responsible for part of the enzyme's specificity. In the catalytic site, 6 out of the 7 conserved residues of the a-amylase family are present, showing the close relatedness between amylomaltase and the alpha-amylase family.

Here we describe the cloning and characterization of the *T. thermophilus* AMase. Further glycosyl hydrolase families 13 and 77 are compared regarding reaction (mechanism and) specificity.

EXPERIMENTAL PROCEDURES

5   *Escherichia coli* TOP10 was used for recombinant DNA manipulations. AMase
(mutant) proteins were produced with *E. coli* BL21 (DE3).

DNA *manipulations* - Restriction endonucleases and DNA polymerase were
purchased from Pharmacia LKB Biotechnology, Sweden, and used according to the
manufacturer's instructions. DNA manipulations and calcium chloride
10   transformation of *E. coli* strains were as described {350}.

Cloning and expression of the T. thermophilus MalQ gene - A *T. thermophilus*
gene library was constructed by inserting the 4-8 kb fragments of a partial *Sau*3A
digest of genomic DNA in the *Bam*HI site of pZerO. This construct was transformed
to E. coli TOP10 cells and plated on LB agar plates. After replicaplating the
15   transformants were screened for amylomaltase activity by overlaying the
motherplate with 5 ml of a 0.5 % soluble starch solution, incubating for 24 h at 70ºC,
and staining with 4 ml Lugol solution. Positive colonies showed a shift from blue to
red staining due to the disproportionation of the starch chains by amylomaltase. The
DNA sequence of one of these clones was determined using the dideoxynuleotide
20   chain termination method on a cycle sequencer (Pharmacia)

The *malQ* gene was amplified with PCR using the following primers:

Forward: GGCAGC<u>CATATG</u>GAGCTTCCCCGCGCTTTCGG

Reverse: GCAGCC<u>AGATCT</u>AGAGCCGTTCCGTGGCCTCGGC

The PCR product was digested with *Nde*I (<u>CATATG</u>) and *Bgl*II (<u>AGATCT</u>, overhang
25   compatible with *Bam*HI) and ligated with either plasmid pET9c or plasmid pET15b
digested with *Nde*I and *Bam*HI, resulting in pGJ6002 or pCCBmalQ, respectively.
Transformation of these plasmids resulted in expression of the native enzyme
(pGJ6002) or of the amylomaltase with an N-terminal His6-tag (pCCBmalQ).

Production and purification of AMase - For the production of AMase protein
30   *E. coli* BL21(DE3), containing the pCCBmalQ vector, was grown overnight in a 1 l
flask with 250 ml LB medium containing ampicilin.

Protein determination - Protein concentrations were determined with the
Bradford method {63} using the Bio-Rad reagent and bovine serum albumin as a
standard (Bio-Rad Laboratories, Richmond, CA, USA).

Enzyme assays - All assays were performed in a 25 mM sodium maleate buffer (pH 6.5) at 70 °C.

Disproportionation reaction - Disproportionation activities were determineded using the ability of AMase to release glucose from oligosaccharides. Various

5   concentrations (upto 50 mM) of (mixtures of) oligosaccharides (G2-G7) were incubated with appropriatly diluted enzyme. For the determination of donor specificity different concentrations of maltooligosaccharides as donor and methyl-α-D-glucose as acceptor.At regular time intervals 50 il samples were taken and added to 200 il GOD-PAP reagent (Roche) to measure the amount of glucose released.

10   *Hydrolyzing activities* were measured as described earlier using 1% soluble starch (Lamers & Pleuger, Belgium) as substrate and dinitrosalicylic acid to determine the number of reducing ends {680}.

In above assays 1 U of activity is defined as the amount of enzyme required for the processing of 1 :mole of donor substrate per minute. Kinetic parameters were

15   fitted using the computer program Sigma Plot (Jandel Scientific).

Product formation from oligosaccharides was analyzed by HPLC. For this purpose 1 ml of a 25 mM G3, G5, or G7 solution was incubated with 0.1 U AMase at 70 °C for 8 h. Samples were taken at regular time intervals and the products formed were applied to a 25 cm Econosphere-NH_2 5 micron column (Alltech Associates Inc.

20   USA) eluted with acetonitrile/water (60/40, v/v) at a flow rate of 1 ml per min.

In the assay for the disproportionation reaction various oligosaccharides (G2-G7) were used as single (donor and acceptor) substrate. The K_M and Vmax values for the formation of glucose varied with the different oligosaccharides. The highest Vmax

25   is observed for G4, which also shows the highest affinity. No activity on G2 was observed. Adding G3 to the G5 reaction mixture resulted in a further increase in activity (Fig. 2), whereas the addition of G2 had no effect (not shown). At high G3 concentrations a decrease in activity is observed, indicating competition between G3 and G5.

30   *The donor specificity of AMase* was further investigated using the various oligosaccharides as donor and M-α-DG as acceptor substrates. Fig 3 shows that the addition of this monosaccharide clearly affects disproportionation activities, especially with the lower concentrations of oligosaccharide (donor) substrates. At higher M-α-DG concentrations the monosaccharide has an inhibitory effect. At lower

concentrations, however, it can efficiently be used as acceptor, allowing a clearer determination of affinities of the different oligosaccharides for the donor binding site. G5, showing the lowest KM value, clearly is the best donor substrate. Combined with the above observation of the stimulating effect of G3 on disproportionation of G5, this suggests that G3 is a better acceptor substrate.

Hydrolyzing activity on soluble starch was investigated, but even overnight incubation did not result in an increase of reducing ends, thus no hydrolyzing reaction is performed by the enzyme. Furthermore the enzyme was incubated with 4-nitrophenyl-α-D-maltoheptaoside-4-6-O-ethylidene (EPS; Boehringer Mannheim) (a maltoheptasaccharide which is blocked at the non-reducing end and with a para-nitrophenyl group at its reducing end). This compound is generally used for the detection of α-amylase activity. However also with this substrate no hydrolyzing activity was observed. Furthermore it reacted very weakly when accepting oligosaccharides were added, suggesting that amylomaltase is an exo-acting enzyme, requiring the presence of a non-reducing end glucose.

The oligosaccharide formation of AMase was analyzed with HPLC. With G3 as substrate the initial products were G1 and G5 After an initial lagfase, the production of G1 increased, while G3 decreased, however with a less significant increase of G5. Various larger oligosaccharides are produced, indicating that the initial product (G5) is used as donor and G3 is mainly used as acceptor. With G5 as substrate the predominant initial products were G3 and G7, although also considerable amounts of G1, G4, G6, and G9 were formed. In both cases little maltose is produced initially, as observed previously (T. aquaticus, potato). The final production of maltose is probably caused by the tranfer of glucose from the donor to a glucose acceptor. This supported by the early formation of G4 from G5, which indicates transfer of a glucose moyety, and by the above results with MαDG, which indicate that glucose can indeed be used as acceptor.

AMase is the ultimate disproportionating enzyme, producing a variety of (long) oligosaccharides from short substrates. In the disproportionation reaction the enzyme has a preference for longer oligosaccharides to be used as donor while shorter oligosaccharides except maltose are efficiently used as acceptor. One of the requirements of doing this efficiently is a low hydrolyzing activity, which is extremely well met in AMase. The complete or near complete lack of hydrolyzing activity of this enzyme makes it a very interesting enzyme to be studied regarding reaction

specificity in the α-amylase family.

Reaction kinetics of the disportionation of oligosaccharides

DP	Km	Vmax
2	nd	nd
3	10	90
4	3.4	281
5	4.6	235
6	3.5	143
7	4.5	108

nd = not detectable

Affinity constants for oligosaccharides using MαDG as acceptor

DP	0 mM	4 mM	10 mM
3	8.0	4.7	12.2
4	3.4	3.0	2.9
5	4.6	1.5	2.7
6	3.5	2.5	2.8

STRUCTURES OF THE THERMOSTABLE AMYLOMALTASE FROM THERMUS
THERMOPHILUS HB8 IN TWO DIFFERENT SPACE GROUPS

Enzymes from the alpha-amylase family, or glycosyl hydrolase family 13, are a
very diverse group of starch-converting enzymes, which have a common architecture
of their catalytic site. Many enzymes from the alpha-amylase family are used in
industrial starch processing, and many have been structurally characterized with the
aim of improving them for specific applications. Because of a lack of sufficient
homology to allow Molecular Replacement, the phase problem for most of these
structures has been solved by using multiple isomorphous replacement (MIR) or
multi wavelength anomalous dispersion (MAD) approaches.

Here we show that the phase problem in the alpha-amylase family can be solved
by using six 'high potential' heavy atom compounds that bind to conserved elements
in the family. The effectiveness of this strategy was demonstrated by the elucidaton
of the structure of the amylomaltase from *Thermus thermophilus* HB8, which is the
most divergent member of the alpha-amylase family.

The structure of the amylomaltase from *Thermus thermophilus* HB8 was solved in
space groups $P2_12_12$ and C2, whereas the highly (99.8%) identical amylomaltase from
Thermus aquaticus was solved earlier in space group $P6_4$. A comparison of these
three structures shows that the core of the enzyme is highly rigid, whereas some
loops around the catalytic site can vary in conformation.

1. Introduction

The alpha-amylase family is a very diverse group of enzymes that have the ability
to modify and degrade starch. Some well-known members of this family, such as
bacterial alpha-amylases, cyclodextrin glycosyltransferase, and iso-amylase are used
in industrial starch processing. Other enzymes, such as human salivary and
pancreatic alpha-amylases are therapeutic targets in the treatment of diabetes,
whereas insect alpha-amylases are useful targets in the development of crop
protectants. In the past, many 3D structures of enzymes from the alpha-amylase
family have been elucidated (Table 1), showing that all members share an
(alpha/beta)8-barrel architecture of the catalytic domain, in which a conserved active
site is that comprises seven amino acid residues. For this reason, it is thought that all
members of the alpha-amylase family catalyze the same reaction cycle. This is

suggested to proceed according to a two-step a-retaining mechanism. In the first step an a-glycosidic bond is cleaved in the substrate and a covalently bound enzyme-glycosyl intermediate is formed. In the second step, the leaving group is exchanged for an acceptor molecule, which is then linked via a new a-glycosidic bond to the

5   intermediate.

The most divergent member of the alpha-amylase family is, on basis of sequence comparisons, the enzyme amylomaltase. Amylomaltase is a 57 kDa intracellular enzyme that is also known as 4a-glucanotransferase in bacteria and D-enzyme in plants. Investigations with *Escherichia coli* have established that amylomaltase is

10   the product of the MalQ gene and is essential for the growth on maltose. Presumably, the function of the enzyme is to synthesize long amylose-like oligosaccharides from shorter oligosaccharides, which can then be further catabolized. This synthesizing capacity of amylomaltase is probably related to the enzyme's high transglycosylation activity. This forms an interesting contrast with the activity of 'classical' alpha-

15   amylases that degrade starch and mainly perform hydrolysis.

Recently, amylomaltases from thermophile organisms like *Thermus aquaticus* and *Thermus thermophilus* HB8 have been isolated. These enzymes have a high thermostability, which makes them suitable for industrial applications, such as the production of large cyclic glucans and the production of thermoreversible gels from

20   starch. A 2.0 Å 3D structure of the amylomaltase from *Thermus aquaticus* shows that the enzyme consists of a compact (alphalpha/betaeta)$_8$-barrel catalytic domain with three loop excursions that are responsible for part of the enzyme's specificity. In the catalytic site, 6 out of the 7 conserved residues of the alpha-amylase family are present, establishing amylomaltase as a member of the alpha-amylase family.

25   We have investigated two 3D structures of the amylomaltase from *Thermus thermophilus* HB8 (TTHB8), determined from data in space groups P$_{2_1 2_1}$2 and C2 to 2.3 Å and 3.1 Å, respectively. The TTHB8 enzyme has a 99.8% sequence identity to the amylomaltase from *Thermus aquaticus*, which crystallizes in space group P6$_4$. However, the structures of the TTHB8 enzyme were solved independently using a

30   MIRAS strategy with general applicability for alpha-amylase enzymes.

32

2. Materials and methods

2.1 Crystallization and data collection

The amylomaltase from *Thermus thermophilus* HB8 was cloned and expressed in *E. coli*, and purified by a series of standard chromatographic steps until all heterogeneities had dissappeared as judged from silver-stained SDS page. The isolation and characterization of the enzyme will be described in detail in another publication. For crystallization an enzyme preparation in 25 mM Tris-HCl, pH 7.5 was used, which was concentrated to 2.5 mg/ml using a Filtron 30K system. The TTHB8 amylomaltase was crystallized at 293 K with the hanging drop vapor diffusion method, using a reservoir solution of 12% (w/v) PEG 20000 and 100 mM MES (2-[N-morpholino] ethanesulfonic acid) buffer at pH 6.8. Crystals appeared after five days, in the form of very thin plates with dimensions 0.15 x 0.15 x 0.04 mm^3. Prior to data collection, they were frozen in a cryo-mother liquor consisting of 25% (v/v) glycerol, 10% (w/v) PEG 20000 and 100 mM MES buffer at pH 6.8.

Due to the small dimensions of these crystals, the diffraction of the amylomaltase crystals on a rotating anode source was limited to 8 Å resolution. However, by using synchrotron radiation, a complete dataset could be obtained to 2.3 Å resolution. The intensity distribution of the data was very anisotropic, most likely because of the non-uniform dimensions of the crystals. The space group of the crystals was P2$_1$2$_1$2, with cell dimensions a=115.2 Å, b=93.7 Å, c=53.5 Å.

In addition to this crystal form, small microneedles (0.04 x 0.04 x 0.20 mm^3) were found growing in hanging drops at 12% (w/v) PEG 20000 and 100 mM maleate at pH 6.8 and 0.1% (w/v) maltotriose. These crystals were frozen by transferring them to an identical solution to which 20% (v/v) glycerol was added, and subsequently dipping them in liquid nitrogen. The frozen crystals were exposed to synchrotron radiation and belong to space group C2, with cell dimensions a=104.9 Å, b=52.5 Å, c=104.9 Å, and a=90°,b= 96.4°, and g=90°. Unfortunately, their diffraction was limited to 3.1 Å, therefore we performed further soaking experiments with the better-diffracting P2$_1$2$_1$2 crystal form.

2.2 Phasing

Because sequence comparisons suggest that amylomaltase is a member of the alpha-amylase family, we initially attempted to solve the phase problem for

33

amylomaltase by Molecular Replacement using a poly-alanine TIM barrel domain as search model. Several models were tried, originating from cyclodextrin glycosyltransferase (CGTase) and *Aspergillus oryzae* (Taka) alpha-amylase, but all attempts failed. This is not surprising since similar Molecular Replacement attempts were also problematic in cases in the alpha-amylase family where model and target had much more structural homology. Therefore, as a next strategy, we decided to use ab initio phasing with multiple isomorphous replacement combined with anomalous scattering (MIRAS).

To determine a suitable MIRAS strategy, an overview was made of the compounds that were used in the past to solve structures of enzymes from the alpha-amylase family. It appears that many structures have been solved using the same heavy atom compounds. Out of 14 cases, a $HgCl_2$ derivative was useful 8 times, a K_2PtCl_4 derivative 7 times, a UO_2^{2-} derivative 6 times and a Sm^{3+} derivative 3 times. This suggests that these compounds bind to conserved features in alpha-amylase family enzymes and thus would have general applicability within the family.

To check this hypothesis and to solve the structure of the TTHB8 amylomaltase, we used these compounds to soak crystals and collected data at the EMBL beamline BW7B at DESY, Hamburg and the EMBL beamline ID14-3 of the ESRF, Grenoble (Table 3). Despite non-isomorphism in the length of the longest cell axis, all the four above-mentioned compounds turned out to be useful derivatives. In addition we found an ethylmercury phosphate derivative.

From these data, heavy atom sites were located using the program Solve and subsequently refined with the program Sharp. Solvent flattening resulted in an experimental electron density map in which secondary structure elements were well discernible. Model building was was performed with the program O. To facilitate model building, we used the sequence and structure of amylomaltase from *Thermus aquaticus* as a template.

2.3 Refinement of the $P2_12_12$ crystal form

Our initial model was refined against our best data, those from a $HgCl_2$ soak which diffracted to 2.3 Å. Refinement was performed using the program CNS version 1.0. After initial rigid body refinement, full coordinate refinement, grouped B-factor refinement and individual atomic B-factor refinement against the CNS maximum likelihood target were applied. Solvent molecules were placed at peaks of at least 3.0

s in F_o-F_c difference electron density maps, at positions where they could form at least one hydrogen bond. This was done using in combination with refinement using the iterative procedure implemented in CNS. Manual rebuilding was done in s_A-weighted F_o-F_c, $2F_o$-F_c and OMIT F_o-F_c, and $2F_o$-F_c maps, calculated with CNS.

5   During rebuilding, a very strong peak close to a small peak in an F_o-F_c difference electron density was observed in the active site. A peak in a similar position was observed in an anomalous difference map (F_o^+-F_o^-, where + and - reflections are Bijvoet mates) from the HgCl$_2$ data. Moreover, the program Sharp had interpreted the position of this peak as a heavy atom binding site. From this we concluded that a

10   HgCl$^-$ ion bound in the active site should be included in our model.

The stereochemistry of the final model was checked with the programs Procheck and Whatcheck. The final model contains no residues in disallowed regions of the Ramachandran plot, in contrast to the structure of the *Thermus aquaticus* enzyme. The atomic coordinates and structure factors have been deposited at the Protein Data

15   Bank (www.rcsb.org, code 1FP8).

2.4 Refinement of the C2 crystal form

In order to study the influence of crystal contacts on the conformation of the enzyme, we also determined the 3.1 Å structure of TTHB8 amylomaltase in the

20   maltotriose-dependent C2 crystal form. An initial model was obtained from the structure in $P2_12_12$, by Molecular Replacement with the program AMoRe. This model was refined using CNS as outlined above. The final refinement step consisted of a few rounds of individual B-factor refinement, which was stopped after the free R-factor started to increase. No solvent molecules were incorporated. Although the crystals

25   were grown in the presence of sugars, we found no evidence for the presence of maltotriose or any other oligosaccharide in the electron density maps. Final model statistics, coordinates and structure factors have been deposited at the Protein Data Bank (www.rcsb.org, code 1FP9).

30   2.5 Binding locations of heavy atom ligands

To solve the phase problem for amylomaltase we used heavy atom compound with a high success rate in the alpha-amylase family, under the assumption that they bind to conserved features within the family. To check whether this is true, we investigated their location using anomalous difference electron density maps

computed with phases from refined models. As indicated above, the HgCl₂ soak
resulted in a HgCl⁻ ion bound in the conserved catalytic site of the alpha-amylase
family. At that position, the Cl⁻ atom binds to Tyr 59 with a typical halide-aryl
interaction, whereas the Hg²⁺ atom is bound by the conserved acidic residues Asp

5   395, Glu 340 and Asp 293.

In addition to HgCl₂, the ethylmercury phosphate soak also resulted in an active
site complex in which an Hg²⁺ moiety is bound by acidic residues. However, the other
soaks (UO₂Ac₂, K₂PtCl₄, SmCl₃) resulted in heavy atoms bound in non-conserved
regions (Table 3). This contrasts with other reports. UO₂Ac₂ was observed to bind in

10   the active site of the CGTase from *Bacillus circulans* strain 251, and UO₂(NO₃)₂ was
observed in the active site of Taka alpha-amylase. K₂PtCl₄ was observed to bind close
to the catalytic site in the CGTase from *Bacillus circulans* strain 8, near residue His
233 in the sugar binding subsite +1. In Taka alpha-amylase, K₂PtCl₄ was observed to
bind close to the catalytic site. To explain this discrepancy, we suggest that the

15   binding of UO₂Ac₂ and K₂PtCl₄ in the active site of amylomaltase is hindered by the
presence of a low concentration of HgCl₂ that was applied to stabilize the crystals.
The HgCl₂ might compete with the other compounds for binding.

Thus, we show that there exist 'high potential' compounds, which are much more
succesful than average in forming heavy atom derivatives of a crystallized alpha-

20   amylase family enzymes. Most of these compounds were reported to bind in the
conserved catalytic site, though this could not always be reproduced for TTHB8
amylomaltase. Nevertheless, it was shown that with these compounds, the phase
problem for alpha-amylase-family enzymes can be quickly and efficiently solved.

25

3. Results

3.1 Secondary structure

The three-dimensional structure of the amylomaltase from *Thermus thermophilus*

30   HB8 in its P2₁2₁2 crystal form is depicted in Figure 1. It is similar to the *Thermus
aquaticus* amylomaltase and consists of a central (alpha/beta)₈ or TIM-barrel domain
from which three other small domains protrude. Although the (alpha/beta)₈-barrel
domain is a feature that is shared by all enzymes from the alpha-amylase, a
superposition of the (alpha/beta)₈-barrel domain in amylomaltase with those from

cyclodextrin glycosyltransferase (CGTase) and Taka alpha-amylase shows large differences in the position, length and orientation of the a-helices that surround the central b-barrel. These differences explain the difficulty of solving the phase problem by using TIM-barrels from alpha-amylase family enzymes as templates for a

5   Molecular Replacement search. Moreover, they shows that the folding pattern of the (alpha/beta)8-barrel is more conserved than the precise three-dimensional orientation of its constituent secondary structure elements.

From the central (alpha/beta)8-barrel domain in amylomaltase three subdomains protrude that are labelled B1, B2 and B3. Subdomain B2 comprises residues 68 to

10   179 and protrudes at the third beta-strand of the TIM barrel, which makes this subdomain the structural homolog of domain B in CGTases and alpha-amylases. Subdomain B1 comprises residues 222 to 272 and 294 to 320, and subdomain B3 comprises residues 398 to 427. Both these latter domains are unique to amylomaltase.

15

3.2 The active site of amylomaltase

Another determinant of alpha-amylase family membership is the presence of seven conserved residues in the catalytic site in a characteristic orientation. The catalytic site of the TTHB8 amylomaltase is compared with the catalytic site of CGTase, a

20   representative member of the alpha-amylase family. It appears that the nucleophilic catalytic residue Asp 229 in CGTase, and the acid/base catalyst Glu 257 have amylomaltase equivalents in Asp 293 and Glu 340. Residues Arg 227, His 327 and Tyr 100, which are important in strabilization of the transition state and the covalent intermediate have equivalents in amylomaltase in Arg 291, His 394 and Tyr 59,

25   respectively. Interestingly however, of two residues in CGTase that are important for distortion of a bound substrate, Asp 328 and His 140, only Asp 328 has an equivalent in amylomaltase in Asp 293, whereas the position of His 140 is taken by Asn 260. In this respect, amylomaltase is different from all other members of the alpha-amylase family. Interestingly, when His 140 is replaced in CGTase or alpha-amylase, the

30   activity decreases 50-100x times.However, amylomaltase has an optimal enzymatic rate that is comparable to that of other alpha-amylases. This might indicate that amylomaltase has found a way of compensating for the absence of a His 140 equivalent by an unknown mechanism.

3.3 Putative sugar binding sites

In addition to the catalytic site, amylomaltase possesses at least seven sugar binding subsites that assist in substrate processing. We attempted to identify these sugar binding subsites by a crystal-soaking procedure, in which $P2_12_12$ crystals of TTHB8 amylomaltase were subjected to a stabilizing solution containing the oligosaccharide inhibitor acarbose. This inhibitor is known to bind strongly in the catalytic site of alpha-amylase-family enzymes and in adjacent sugar binding subsites. Unfortunately, after subsequent data collection on these crystals, this inhibitor could not be observed in the electron density, and therefore had not bound inside the crystals. Probably, the active site of amylomaltase in its $P2_12_12$-crystalline form is not accessible to oligosaccharide binding.

To nevertheless estimate the location of extra sugar binding subsites, we constructed a model of sugar binding. We superimposed the 3D structure of a maltohexaose inhibitor in complex with Porcine pancreatic alpha-amylase on amylomaltase on basis of the conserved active site in both enzymes. The torsion angles of the glycosidic bonds in the maltohexaose inhibitor were subsequently adjusted to improve its fit in the active site of amylomaltase. This remodelling was aided by comparisons with the conformations of other oligo-saccharides in complex with alpha-amylase family enzymes, such as maltononaose bound to CGTase. The final model is schematically drawn in Figure 3, and is the first detailed model of how amylomaltase might bind an oligosaccharide, and is provides the guidance needed for site-directed mutagenesis experiments that alter the properties of amylomaltase in a desired fashion.

3.4 Crystal contacts in the $P2_12_12$ and C2 crystal forms

The structure of TTHB8 amylomaltase was determined to high resolution in a $P2_12_12$ crystal form, and to lower resolution in a C2 crystal form. This allows us to establish whether the conformation of amylomaltase is influenced by the crystalline packing of the molecules.

In the $P2_12_12$ form, crystal contacts are formed in three regions. In the first a loop of residues Gly 149-Gly 153 is grabbed by residues Gly 422-Arg 426 and the C-terminus Ala 492-Leu 500. A second, weaker contact is formed between residues Gln 27-Glu 38 and the two stretches Glu 313-:Lys 318, Gly 343-Val 349.

Interestingly, a third contact is formed by the only cysteine residue in amylomaltase. Cys 308 is at a distance of 5.2 Å of a Cys 308 of another amylomaltase molecule inside the crystal. This suggests that one crystal contact is formed by an intramolecular disulphide bond. The possibility of a disulphide-linked crystal

5   contact is corroborated by the electron density at this location, which suggest the (partial) presence of a disulphide bond. This suggests that the crystal lattice consists of a mixture of disulphide-bonded dimers and monomeric units. Dynamic light scattering experiments with our sample support the presence of a small amount of dimers (results not shown) mixed with monomers. It is not unlikely

10   that the dimeric impurities enforce the presence of a disulphide-bonded crystal contact, inhibiting the formation of other (stronger) contacts, such as present for example in the P6$_4$ crystal form of the *Thermus aquaticus* amylomaltase.

The crystals in space group C2 show a similar disulphide-linked crystal contact. However, at the other crystal contacts, there are significant differences. The

15   stretch of residues Lys 148-Glu 173 binds to Pro 378-Gly 385, and the residues Gly 26-Asp 31 and Leu 74- Gly 89 bind to Gly 26-Asp 31 and Leu 74- Gly 89 in another molecule. Due to these differences, the C2 crystal form can be regarded as independent from the P2$_1$2$_1$2 crystal form.

20   3.5 Comparison of amylomaltases from *T. aquaticus* and *T. thermophilus* HB8

We have determined two structures of the amylomaltase from *Thermus thermophilus* HB8. Earlier, the 3D structure of the amylomaltase from *Thermus aquaticus* was determined, which has a sequence identity to the TTHB8 amylomaltase of 99.8%. Only Gln 27 and Leu 154 in the TTHB8 enzyme have been substituted by Arg 27 and

25   Pro 154 in the *Thermus aquaticus* enzyme. Strangely, the published amino acid sequence of the *Thermus aquaticus* enzyme, which gives Pro 154, does not correspond to the sequence derived from the 3D structure of the *Thermus aquaticus* amylomaltase, which gives Leu 154. If this is interpreted as a correction on a sequencing error, both amylomaltases only differ in amino acid sequence at position

30   27. Therefore, for all practical purposes these structures can be regarded as independently solved structures of the same enzyme in different space groups. A comparison could reveal interesting areas of flexibility.

3.6 Conformational differences between the three structures of amylomaltase

Since the structures of *T. thermophilus* HB8 amylomaltase in space groups C2 and
$P2_12_12$ and the structure of *T. aquaticus* amylomaltase in space group $P6_4$ can be
regarded as three structures of the same enzyme in different crystal packing
environments, differences between these structures can show how crystal contacts
influence the conformation of the enzyme, and in which areas it is very flexible or
very rigid.

If we take the '$P2_12_12$-structure' as basis, and superimpose the 'C2-structure', we
observe that the position of most amino-acids is identical (r.m.s.d. 0.5 Å). However,
two loops in the active site cleft have a significantly different conformation. First, the
loop that comprises residues Tyr 141 to Ala 170 has shifted in the C2-form towards
the active site (maximally 1.5 Å). Secondly, the loop of residues Val 242-Leu 262 (and
its adjacent loop Tyr 301-Val 317), which cover the active site cleft, have shifted ~0.5
Å outwards in the C2 form, thereby opening the cleft a little.

When the structure of *T. aquaticus* amylomaltase is superimposed on the '$P2_12_12$-
structure', this shows that they have an almost identical conformation (r.m.s.d.
0.4 Å). Interestingly, also in the *T. aquaticus* enzyme the loop of residues 141-170
has a position that is oriented more toward the active site (maximum difference
1.3 Å). This position resembles the conformation of this loop in the C2-crystal
form.

The flexibility of amylomaltase was further studied through the atomic temperature
factors. In general, all three structures show a similar temperature factor
distribution, indicating only a marginal influence of crystal packing contacts. In
all cases amylomaltase appears to be rigid, with specific areas having higher
temperature factors, and thus higher flexibility. These include four loop stretches
near the catalytic site comprising residues 80-93, 114-125, 342-348 and most
strongly 249-253.

Thus, in general amylomaltase appears to have a rigid, well-determined
conformation, which might in part explain the enzyme's thermostability.

However, when information on conformational variability and temperature factor
distributions is combined, it appears that there are two interesting regions in the
enzyme. The first is the loop 242-262 (comprising 249-253) that can have
different conformations and is also very flexible (high B-factors). This loop
incorporates residues Tyr 250 and Phe 251, which might be involved in substrate

binding (Figure 3). The second is the loop 141-170, which is conformationally variable but has a very low temperature factor. Therefore, this loop is not flexible, but can 'switch' between two rigid conformations. As was observed for other alpha-amylase family enzymes, such conformational variations could play an important role in promoting catalysis.

Overview of heavy atom compounds used to solve 3D structures of alpha-amylase-family proteins

Enzyme	method[c]	heavy atom compounds used
Animal alpha-amylases		
porcine pancreas	MIR (2x)[a]	$OCMP^b/K_2PtCl_4/K_2HgI_4/PbNO_3/HgAc_2/U_2O_7$
human salivary	MR	
human pancreas	MR	
yellow meal worm	MR	
Fungal and plant alpha-amylases		
Taka (*Aspergillus oryzae*)	MIR	$HgCl_2/UO_2(NO_3)_2/AgNO_3/K_2PdCl_4/K_2PtCl_4/K_2Pt(CN)_4/KAu(CN)_2$
Aspergillus niger	MIR/MR	$HgCl_2/SmAc_3/K_2PtCl_6/PbAc_2$
Barley	MIR	$HgCl_2/Eu(NO_3)_3/K_2PtCl_4$
bacterial alpha-amylases		
B. licheniformis	MIR (2x)	$UO_2Ac_2/Pb(CH_3)_3Ac/HgCl_2/K_2PtCl_4/K_2PtCl_6{}^a$
B. subtilis	MIR	$K_2PtCl_4/HgCl_2$
P. stutzeri	MIR	$K_3UO_2F_5/SmCl_3$
Alteromonas haloplanctis	MR	
cyclodextrin glycosyltransferases		
B. circulans strain 8	MIR	$K_2PtCl_4/cis-(NH_3)_2PtCl_2/UO_2C_2O_4$
B. stearothermophilus		
B. sp. 1011	MR	
B. circulans strain 251	SIRAS	UO_2Ac_2
other enzymes		
B. cereus oligo-1,6-glucosidase		$HgCl_2/UO_2(NO_3)_2/Sm(NO_3)_3$
P. amyloderamosa iso-amylase	MIR	$NaAuCl_4/HgCl_2$
B. stearothermophilus maltogenic alpha-amylase	MR	
Thermoactinomyces vulgaris A47 alpha-amylase II	MIRAS	$PbAc_2/C_2H_5Hg^-$
Thermus strain maltogenic alpha-amylase	MIR/MR	$Se-Met/PtCl2(NH_3)_2/HoCl_2$
Thermus aquaticus amylomaltase	MIR	$PCMBS/HgCl_2/K_2PtCl_4/KAu(CN)_2/K_2Pt(SCN)_6/Pb(CH_3)_3Ac$

[a]used in the most recent report. [b]OCMP means ortho-chloromercuriphenol. PCMBS means para-chloromercuriphenylsulfonic acid. [c]MIR(AS) means Multiple isomorphous replacement (with anomalous scattering), MR means Molecular Replacement, SIR(AS) means single isomorphous replacement (with anomalous scattering).

Data collection and refinement statistics

Data collection	Native $P2_12_12$	Native C2
X-ray source	EMBL beamline ID14-3 ESRF, Grenoble	
Temperature (K)	120	120
Wavelength (Å)	0.933	0.931
Space group	$P2_12_12$	C2
Cell axis (Å)	115.2, 93.7, 53.5	104.9, 52.4, 104.9
	90.0, 90.0, 90.0	90.0, 96.4, 90.0
Resolution (Å)	53.4-2.30	52.2-3.13
No. of unique reflections	26405	10132
Completeness (%)	99.7	96.3
R_{merge}[a] and $<I/s>$	0.076 / 16.2	0.082 / 8.7
Statistics of the last resolution shell	(2.37 Å-2.30 Å)	(3.17 Å-3.10 Å)
Completeness (%)	99.5	96.3
R_{merge} and $<I/s>$	0.37 / 3.5	0.20 / 2.8
Refinement statistics		
No. of amino acids	500 (all)	500 (all)
No. of solvent sites	270	0
Average B-factor (Å²)	35.6	34.5
Final R-factor[b] (incl. bulk solvent correction)	0.194	0.239
Final free R-factor[c] (incl. bulk solvent correction)	0.232	0.292
R.m.s. deviation from ideal geometry		
bond lengths (Å)	0.006	0.004

bond angles (deg.)	1.204	1.031
B-factor correlations between neighbouring main chain atoms ($\mathrm{\mathring{A}}^2$)	1.37	2.46
percentage of residues in allowed regions of the Ramachandran plot	100.0	100.0
[a]R_{merge}=$S_h S_i$	I(h)-I_i(h)	/$S_h S_i$ I_i(h) where reflection h has intensity I_i(h) on occurrence i and mean intensity I(h). [b]R factor=S_h	F_o-F_c	/$S_h F_o$ where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively. [c]The free R factor is calculated as the R factor, using F_o that were excluded from the refinement (5% of the data).		

Summary of phasing statistics

	unsoaked	HgCl₂[a]	UO₂Ac₂[b]	(C₂H₅Hg)₃PO₄	SmCl₃[b]	K₂PtCl₄
X-ray source	BW7B	ID14-3	BW7B	BW7B	BW7B	BW7B
Soaking conditions (days/mM)	-	3/3	3/3	3/5	3/10	3/3
Temperature (K)	100	120	100	100	100	100
Wavelength (Å)	0.8439	0.9330	0.8439	0.8439	0.8439	0.8439
Resolution (Å)	36.7-3.0	53.4-2.30	50-2.57	50-2.73	50-2.94	50-3.0
No. of unique reflections	11853	26436	19162	16205	13139	12372
Longest cell axis (Å) (in P2₁2₁2)	117.8	115.2	116.3	116.3	116.0	117.6
Completeness	95.0	99.7	98.2[e]	99.0[e]	98.6	99.6[e]
R$_{merge}$ and <I/s>	0.22/5.0	0.076/16.2	0.11/12.1[e]	0.13/8.4[e]	0.14/9.1	0.15/7.6[e]
Phasing to 3.3 Å (centric/acentric)						
Number of sites	-	1	3	4	3	5
PP[c]	-	1.1/1.1	1.7/2.3	1.2/1.4	0.76/0.80	0.95/0.92
Anomalous PP[c]	-	-/-	-/1.2	-/1.0	-/-	-/0.92
Nearest amino acid residue	-	Asp 293 Asp 395 Glu 340	Asp 293[b] Asp 52 Tyr 54	His 228 His 304 His 409 Glu 460	Asp 96 Asp 369 Glu 122	Met 145 Met 372 His 409 Arg 209 Arg 194
Combined FOM[d]	0.61/0.53					

[a]Due to their high resolution, the final model was refined against these data.

[b]Because HgCl₂ appeared to stabilize the crystals, soakings with this compound was preceded by a soaking in HgCl₂. The site labelled 'b' is therefore probably a Hg²⁺ ion.

5   [c]PP: Phasing Power. [d]Before solvent flattening. [e]Regarding Bijvoet mates as separate reflections.

Generation of mutants

Family 77 of glycosyl hydrolases consists of a single group of enzymes; 4-α-glucanotransferases (EC 2.4.1.25, amylomaltase (AMase) or D-enzyme). AMase is

5   found in prokaryots and promotes metabolism of starch degradation products inside the cell as shown for *Escherichia coli*. In other organisms, lacking other enzymes required for growth on oligosaccharides (p.e. maltodextrin phosphorylase), it may be involved in glycogen metabolism as suggested for *Aquifex aeolicus*. D-enzyme is found in plants and is reported to be involved in in starch metabolism. Recent studies on

10   *Chlamydomonas rheinhardtii* show that D-enzyme is essential for biosynthesis of starch. In each case the role of AMase is based on its transglycosylating activity, which enables the enzyme to produce long oligosaccharides from short chained substrates or transfer oligosaccharides to branched polymers (glycogen, amylopectin). The synthesizing capacity of wid type amylomaltase is probably related to the

15   enzyme's high transglycosylation activity and lack of hydrolyzing activity. This forms an interesting contrast with the activity of 'classical' a-amylases that degrade starch and mainly perform hydrolysis.

         Sequence comparisons and 3-D structure similarities show that AMase is closely related to the alpha-amylase family or family 13 of glycosyl hydrolases. The a-

20   amylase family is a very diverse group of enzymes that have the ability to modify and degrade starch. In the past, many 3D structures of enzymes from the a-amylase family have been elucidated, showing that all members share an (alpha/beta)$_8$-barrel architecture of the catalytic domain, containing a conserved active site that comprises seven amino acid residues. For this reason, it is thought that all members of the a-

25   amylase family catalyze the same reaction cycle. This is suggested to proceed according to a two-step a-retaining mechanism. In the first step an a-glycosidic bond is cleaved in the substrate and a covalently bound enzyme-glycosyl intermediate is formed. In the second step, the leaving group is exchanged for an acceptor molecule, which is then linked via a new a-glycosidic bond to the intermediate.

30   Recently, amylomaltases from thermophile organisms like *Thermus aquaticus* and *Thermus thermophilus* HB8 have been isolated. These enzymes have a high thermostability, which makes them suitable for industrial applications, such as the production of large cyclic glucans and the production of thermoreversible gels from starch. A 2.0 Å 3D structure of the amylomaltase from *Thermus aquaticus* shows that

the enzyme consists of a compact (alpha/beta)8-barrel catalytic domain with three

loop excursions that are probably responsible for part of the enzyme's specificity. In

the catalytic site, 6 out of the 7 conserved residues of the a-amylase family are

present, showing the close relatedness between amylomaltase and the a-amylase

5   family.

EXPERIMENTAL PROCEDURES

10   *Bacterial strains and plasmids* - *Escherichia coli* TOP10 (Invitrogen) was used

for recombinant DNA manipulations. AMase (mutant) proteins were produced with

E. coli BL21(DE3) (Stratagene). The *malQ* gene was amplified with PCR using the

following primers:

15   *Thermus thermophilus*:

Forward: GGCAGCCATATGGAGCTTCCCCGCGCTTTCGG

Reverse: GCAGCCAGATCTAGAGCCGTTCCGTGGCCTCGGC

        Aquifex aeolicus:

Forward: GGCAGCCATATGAGATTGGCAGGTATTTTAC

20   Reverse: GCAGCCGGATCCTTAAACTTCTCTTCCG

The PCR product was digested with *Nde*I (CATATG), and *Bgl*II (AGATCT, overhang

compatible with *Bam*HI, *T. thermophilus*) or *Bam*HI (GGATCC, *A. aeolicus*) and

ligated with plasmid pET15b (Novagen), digested with *Nde*I and *Bam*HI. The

25   resulting construct (pCCBmalQ) encodes the amylomaltase with an N-terminal His6-

tag.

        Site-directed mutagenesis - For site-directed mutagenesis a method based

upon PCR reactions using PWO-DNA polymerase was used. In a first PCR reaction a

mutagenesis primer together with the reverse primer was used. The product of this

30   reaction was used as a primer in a second PCR reaction together with the forward

primer. This PCR product was cloned in pET15b using the same strategy as for the

wild type. The following mutagenesis primers were used to produce the mutations:

        Thermus thermophilus:

F251L/S: 5'-CCC'CCC'GAC'TAC'TYG'AGC'GAG'ACC'GGT'CAG'CGC'TGG' GGC-3',

F366L/S: 5'-AAG'GTC'CTG'<u>CAA'TYG</u>'GCC'TTT'GAC'GAC-3'

Aquifex aeolicus:

F244L/S: 5'-CCT'CCT'GAT'TTC'<u>TYG'AGT</u>'AAA'ACG'GG-3'

F359L/S:5'-GTT'ATT'GAG'TYG'GCC'TTC'TAC'G-3'

In these primers Y= T (F-L) of C (F-S). Successful mutagenesis resulted in appearance of the underlined restriction sites, allowing rapid screening of potential mutants. For F251L/S this restriction site was *Age*I (ACCGGT); for F251S an additional *Xho*I site (CTCGAG) was introduced. Mutation F366L/S caused deletion of a *Pst*I site; for F366L an additional *Mun*I site (CAATTG) was introduced. For F244S an *Xho*I site (CTCGAG) was introduced. All mutations were confirmed by restriction analysis and DNA sequencing.

DNA manipulations - Restriction endonucleases were purchased from Pharmacia LKB Biotechnology, Sweden; NEB; or Boehringer, and used according to the manufacturer's instructions. DNA manipulations and calcium chloride transformation of *E. coli* strains were as described.

Growth conditions - Plasmid carrying bacterial strains were grown on LB medium containing 50 ìg/ml ampicilin (*E. coli* TOP10) or 50 ìg/ml ampicilin and 50 ìg/ml chloramphenicol (*E. coli* DE3(RP)). For the production of (mutant) AMase proteins *E. coli* DE3(RP), containing the pCCBmalQ vector, was grown in a 1 l flask with 250 ml LB medium containing 50 ìg/ml ampicilin.

Protein determination - Protein concentrations were determined with the Bradford method {63} using the Bio-Rad reagent and bovine serum albumin as a standard (Bio-Rad Laboratories, Richmond, CA, USA).

Enzyme assays - All assays were performed in a 25 mM sodium maleate buffer (pH 6.5) at 70 °C.

Disproportionation reaction - Disproportionation activities were determineded using the ability of AMase to release glucose from oligosaccharides. Various concentrations (upto 50 mM) of (mixtures of) oligosaccharides (G2-G7) were incubated with appropriatly diluted enzyme. For the determination of donor specificity different concentrations of maltooligosaccharides as donor and methyl-α-D-glucose as acceptor.At regular time intervals 50 ìl samples were taken and added to 200 microliter GOD-PAP reagent (Roche) to measure the amount of glucose released.

Hydrolyzing activities were measured as described earlier using 1% soluble starch (Lamers & Pleuger, Belgium) as substrate and dinitrosalicylic acid to determine the number of reducing ends .

In above assays 1 U of activity is defined as the amount of enzyme required for the processing of 1 mmole of donor substrate per minute. Kinetic parameters were fitted using the computer program Sigma Plot (Jandel Scientific).

Product formation from oligosaccharides was analyzed by HPLC. For this purpose 1 ml of a 25 mM G3, G5, or G7 solution was incubated with 0.1 U AMase at 70 °C for 8 h. Samples were taken at regular time intervals and the products formed were applied to a 25 cm Econosphere-NH$_2$ 5 micron column (Alltech Associates Inc. USA) eluted with acetonitrile/water (60/40, v/v) at a flow rate of 1 ml per min.

Sequence alignments - Sequence alignments of various amylomaltases indicate that the two phenylalanines selected based on the structure of amylomaltase and the model of the maltoheptaose bound in the active site are (functionally) conserved in all amylomaltases.

5

T.thermophilus   GVPPDYFSETGQRWGNP

T.aquaticus   {4}   GVPPDYFSETGQRWGNP

Synechocystis {101}   GVPPDYFSATGQLWGNP

A.aeolicus   {97}   GVPPDFFSKTGQLWGNP

10   S.tuberosum   {25}   GVPPDAFSETGQLWGSP

C.butyricum   {11}   GCPPDAFSETGQLWGNP

S.pneumoniae   {55}   GCPPDEFSVTGQLWGNP

M.tuberculosis {96}   GAPPDEFNQLGQDWSQP

H.influenzae {103}   GAPPDPLGPVGQNWNLP

15   E.coli   {40}   GAPPDILGPLGQNWGLP

C.pneumoniae   {94}   GAPPDLYNSEGQNWHLP

C.psittaci   {99}   GAPPDIYNTEGQNWHLP

C.trachomatis {95}   GAPPDLYNAEGQNWHLP

20   T.thermophilus   LAEDLGVITPEVEALRDRFGLPGMKVLQFAF

T.aquaticus   LAEDLGVITPEVEALRDRFGLPGMKVLQFAF

Synechocystis   VAEDLGVITPEVEALRDEFNFPGMKVLHFAF

A.aeolicus   IAEDLGFITDEVRYLRETFKIPGSRVIEFAF

S.tuberosum   IAEDLGVITEDVVQLRKSIEAPGMAVLQFAF

25   C.butyricum   IAEDLGYLTEETLEFKKRTGFPGMKIIQFAF

S.pneumoniae   IAEDLGFMTDEVIELRERTGFPGMKILQFAF

M.tuberculosis   VGEDLGTVEPWVRDYLLLRGLLGTSILWFEQ

H.influenzae   IGEDLGTVPDEVRWKLNEFQIFSYFVLYFAQ

E.coli   IGEDLGTVPVEIVGKLRSSGVYSYKVLYFEN

30   C.pneumoniae   IGEDLGIIPQDVKTTLTHLGICGTRIPRWER

C.psittaci   IGEDLGSVPTDVKETLVKLGICGTRIPRWER

C.trachomatis   IGEDLGTIPSDVKRMLESFAVCGTRIPRWER

49

Construction of mutant enzymes - One mutant (F366L, *Thermus*) has been constructed and confirmed by sequence analysis. Other mutants have been constructed (for example F366S (*Thermus*), F359L/S (*Aquifex*))

Disproportionation activity - Mutant F366L has been analyzed concerning the

5   disproportionation of maltotriose. The activity (25 U/mg) was four times lower than that of the wild type, whereas the affinity (Km = 3.5) was threefold higher than the wild type.

Hydrolyzing activity - As for the wild type, no hydrolyzing activity could be determined during incubation of soluble starch, even with large amounts of enzyme.

10   However, contrary to the wild type enzyme, an increase in reducing power of the reaction mixture after overnight incubation was detected, indicating that hydrolysis had taken place.

Product formation from maltotriose - HPLC analysis of the products formed during incubation of the enzyme with maltotriose clearly shows that hydrolysis takes

15   place. Whereas the wild type produces essentially no maltose, which cannot be cleaved of by the enzyme, the mutant produces maltose as one of the main compounds.

Interaction with hydrophobic amino acids, such as F366, which is highly conserved in amylomaltases, is involved in the reaction specificity of the enzyme.

20   Hydrolyzing activity can be introduced by mutating this residue or other hydrophobic residues. This hydrolyzing activity has significant effects on product profiles of the enzyme, indicating the necessity of essentially complete or practically complete absence of hydrolysis for the function of the wild type enzyme (the production of longer oligosaccharides from short substrates).

25

Alignement of Branching Enzymes

Bstearothermophilus   --

Bcaldolyticus   --

5   Bsubtilis   --

mycobacterium   --------------------------MSRSEKLTGEH-LAPEPA----

EMARLVAGT

Streptomyces

MSAARQPSPTVRDKAAPEPAAPAAPKGARAPRARRAAPPHGVRPAPALAAEERAR

10 LLEGR

E.   ----------------------------------MSDRIDRDVINALIAGH

H.influenzae   ---------------------------------MTTAVTQAIIDGFFDAS

Agro.tume   ----------------------------MKKPLNSAEEKKTGDITKAEIEAIKSGL

Aquifex_a.   ---

15 Synechococcus   ----------------------------TGTTPLPSSSLSVEQVNRIASNQ

Synechocystis   -------------------------------MTYTINADQVHQIVHNL

Butyrivibrio   ---

CHLAMYDIA   ---------------------------------MDPFFLNTQHVELLVSGK

20

Bstearothermophilus   --

Bcaldolyticus   --

Bsubtilis   --

mycobacterium   HHNPHGILGAHEYDDHTVIR-----AFRPHAVEVVALVGK---

25 DRFSLQHLD-SGLFAVA

Streptomyces   HHDPHAVLGARTQRGGVAFR-----VLRPYAKAVTVVAKG---

LRTELVDEG-DGLFSGL

E.   FADPFSVLGMHKTTAGLEVR-----

ALLPDATDVWVIEPKTGRKLAKLECLDSRGFFSGV

30 H.influenzae   NGDPFATLGMHETEQGIEIR-----

TLLPDANRMVVIERESGKEITELDCVDERGFFVGV

Agro.tume   HSNPFQIIPLHETPEGFSAR-----CFIPGAEEVSVLTLD-

GNFVGELKQIDPDGFFEGR

Aquifex_a.   --

Synechococcus   EQNPFDILGPHPYEHEGQAG-WVIRAYLPEAQEAAVICPAL-
RREFAMHPVHHPHFFETW

Synechocystis
HHDPFEVLGCHPLGDHGKVNQWVIRAYLPTAEAVTVLLPTD-

5 RREVIMTTVHHPNFFECV

Butyrivibrio   --

CHLAMYDIA   QSSPQDLLGIVS-ESLNQDR---IVLFRPGAETVFVELRG----
KIQQAESHHSGIFSLP

10

Bstearothermophilus   --------------------------
MIAANPTDLEVYLFHEGSLYKSYELFGAHV--

Bcaldolyticus   -------------------------
MIAANPTDLEVYLFHEGRLYQSYELFGAHV--

15 Bsubtilis   -------------------------
MAAASPTAHDVYLFHEGSLFKSYQLFGSHY--

mycobacterium   LPFVD-
LIDYRLQVTYEGCEPHTVADAYRFLPTLGEVDLHLFAEGRHERLWEVLGAHPRS

Streptomyces   LPLTG-VPDYRLLVTYDSDE-

20 IEVHDPYRFLPALGELDLHLIGEGRHEELWTALGSQP--

E.   IPRRKNFFRYQLAVVWHGQQ-
NLIDDPYRFGPLIQEMDAWLLSEGTHLRPYETLGAHA--

H.influenzae   IPNCRQFFAYQLQVFWGNEA-
QIIEDPYRFHPMIDDLEQWLLSEGSMLRPYEVLGAHF--

25 Agro.tume   IDLSK-RQPVRYRACRDDAE-
WAVTDPYSFGPVLGPMDDYFVREGSICGYSTGWARIP--

Aquifex_a.   -------------------------
MKKFSLISDYDVYLFKEGTHTRLYDKLGSHV--

Synechococcus   VPEET-

30 LEIYQLRITEGERERIIYDPYAFRSPLLTDYDIHLFAEGNHHRIYEKLGAHP--

Synechocystis   LELEE-
PKNYQLRITENGHERVIYDPYGFKTPKLTDFDLHVFGEGNHHRIYEKLGAHL--

Butyrivibrio   -------------------------
MSQKVFISEDDEYLFGQGTHYDIYDKLGAHP--

```
CHLAMYDIA            VMKGISPQDYRVYHQN-G---
LLAHDPYAFPLLWGEIDSFLFHEGTHQRIYERMGAIP--
                           : ..:*        .
```

Bstearothermophilus   -INEGG-

```
KVGTRFCVWAPHAREVRLVGSFNDWDGTDFRLEKVND-EGVWTIVVPENLEGH
```

Bcaldolyticus   -IRGGG-

```
AVGTRFCVWAPHAREVRLVGSFNDWNGTNSPLTKVND-EGVWTIVVPENLEGH
```

Bsubtilis   -RELNG-

10   KSGYEFCVWAPHASEVRVAGDFNSWSGEEHVMHRVND-NGIWTLFIPGIGEKE

mycobacterium

```
FTTADGVVSGVSFAVWAPNAKGVSLIGEFNGWNGHEAPMRVLGP-
SGVWELFWPDFPCDG
```

Streptomyces   -MEHQG-

15   VAGTRFTVWAPNALGVRVTGDFSYWDAVAYPMRSLGA-SGVWELFLPGVAEGA

E.   -DTMDG-

```
VTGTRFSVWAPNARRVSVVGQFNYWDGRRHPMRLRKE-SGIWELFIPGAHNGQ
```

H.influenzae   -MECDG-

```
VSGVNFRLWAPNARRVSIVGDFNYWDGRRHPMRFHSK-SGVWELFLPKASLGQ
```

20   Agro.tume   -LKLEG-

```
VEGFHFAVWAPNGRRVSVVGDFNNWDGRRHVMRFRKD-TGIWEIFAPDVYA-C
```

Aquifex_a.   -IELNG-

```
KRYTFFAVWAPHADYVSLIGDFNEWDKGSTPMVKREDGSGIWEVLLEGDLTGS
```

Synechococcus   -CELEN-

25   VAGVNFAVWAPSARNVSILGDFNSWDGRKHQMAR-RS-NGIWELFIPELTVGA

Synechocystis   -MTVDG-

```
VKGVYFAVWAPNARNVSILGDFNNWDGRLHQMRK-RN-NMVWELFIPELGVGT
```

Butyrivibrio   -SEEKG-

```
KKGFFFAVWAPNAADVHVVGDFNGWDENAHQMKRSKT-GNIWTLFIPGVAIGA
```

30   CHLAMYDIA   -CEIDG-

```
VPGVRFIVWAPHAQRVSVIGDFNGWHGLVNPLHKVSD-QGVWELFVPGLTAGA
         . *:***.*:*.*.*   :     :*:.
```

Bstearothermophilus   LYKYEIVTPDGQVL-
FKADPYAFYSELRPHTASIAYDLKGYQWNDQSWKRKKRRKRIYDQ

Bcaldolyticus   LYKYEIITPDGRVL-
LKADPYAFYSELRPHTASIVYDLKGYEWNDSPWQRKKRRKRIYDQ

5   Bsubtilis   RYKYEIVTNNGEIR-
LKADPYAIYSEVRPNTASLTYDLEGYSWQDQKWQKKQKAKTLYEK

mycobacterium   LYKFRVHGADGVVT-DRADPFAFGTEVPRQTASRVT-
SSDYTWGDDDWMAGRALRNPVNE

Streptomyces   LYKYEITRPDGGRT-LRADPMARYAEVPPANASIVT-
10   ASRYEWQDAEWMARRGALAPHQA

E.   LYKYEMIDANGNLR-LKSDPYAFEAQMRPETASLIC-
GLPEKVVQTEERKKANQFDA---

H.influenzae   LYKFELIDCHGNLR-LKADPFAFSSQLRPDTASQVS-
ALPNVVEMTEARKKANQGNQ---

15   Agro.tume
AYKFEILGANGELLPLKADPYARRGELRPKNASVTAPELTQKWEDQAHREHWAQ
VDQRRQ

Aquifex_a.   KYKYFIKNGNYEVD--KSDPFAFFCEQPPGNASVVW-
KLNYRWNDSEYMKKRKRVNSHDS

20   Synechococcus   AYKYEIKNYDGHIYE-
KSDPYGFQQEVRPKTASIVADLDRYTWGDADWLERRRHQEPLRQ

Synechocystis   SYKYEIKNWEGHIYE-
KTDPYGFYQEVRPKTASIVADLDGYQWHDEDWLEARRTSDPLSK

Butyrivibrio   LYKFLITAQDGRKLY-
25   KADPYANYAELRPGNASRTTDLSGFKWSDSKWYESLKGKDMNRQ

CHLAMYDIA   CYKWEMVTESGQVL-IKSDPYGKFFGPPPWSVSVVI-
DDSYEWTDSEWLEERIKKTEG--

        .:   ...   *..*

Bstearothermophilus   PMVIYELHFGSWKKK----------------------
30   DGRFYTYREMADELISYVLDH

Bcaldolyticus   PMVIYELHFGSWKKKP----------------------
DGRFYTYREMADELIPYVLER

Bsubtilis   PVFIYELHLGSWKKHS----------------------
DGRHYSYKELSQTLIPYIKKH

54

mycobacterium   AMSTYEVHLGSWRP-------------------------
GLSYRQLARELTDYIVDQ

Streptomyces   PMSVYELHLASWRP-------------------------
GLSYRQLAEQLPAYVKEL

5   E.   PISIYEVHLGSWRRH-----------TDNN----------
FWLSYRELADQLVPYAKWM

H.influenzae   PISIYEVHLGSWRRN-----------LENN----------
FWLDYDQIADELIPYVKEM

Agro.tume   PISIYEVHAGSWQR-----------SEDG----------
10   TFLSWDELEAQLIPYCTDM

Aquifex_a.   PISIYEVHVGSWRRVP----------EEGN----------
RFLSYRELAEYLPYYVKEM

Synechococcus
PISVYEVHLGSWMHASSDAIATDAQGKPLPPVPVADLKPGARFLTYRELADRLIPY
15   VLDL

Synechocystis   PVSVYELHLGSWLHTAYDEPVKTLHGEGVP-
VEVSEWNTGARFLTYYELVDKLIPYVKEL

Butyrivibrio   PIAIYECHIGSWMKHP--------DGTEDG-----------
FYTYRQFADRIVEYLKEM

20   CHLAMYDIA   PMNIYEVHVGSWRWQE----------------------
GQPLNYKELADQLALYCKQM
                        .: ** * .**   : :: : *

Bstearothermophilus
25   GFTHIELLPLVEHPLDRSWGYQGTGYYAVTSRYGTPHDFMYFVDRCHQAGIGVIM
DWVPG

Bcaldolyticus
GFTHIELLPLVEHPLDRSWGYQGTGYYSVTSRYGTPHDFMYFVDRCHQAGLGVII
DWVPG

30   Bsubtilis
GFTHIELLPVYEHPYDRSWGYQGTGYYSPTSRFGPPHDLMKFVDECHQQNIGVIL
DWVPG

mycobacterium

GFTHVELLPVAEHPFAGSWGYQVTSYYAPTSRFGTPDDFRALVDALHQAGIGVIVD

WVPA

Streptomyces

GFTHVELMPVAEHPFGGSWGYQVTGFYAPTSRMGTPDDFRFLVDALHRAGIGVIV

DWVPA

E.

GFTHLELLPINEHPFDGSWGYQPTGLYAPTRRFGTRDDFRYFIDAAHAAGLNVILD

WVPG

H.influenzae

GFTHIEFLPLSEFPFDGSWGYQPLGLYSPTSRFGSPEAFRRLVKRAHEAGINVILD

WVPG

Agro.tume

GFTHIEFLPITEHPYDPSWGYQTTGLYAPTARFGDPEGFARFVNGAHKVGIGVLLD

WVPA

Aquifex_a.

GFTHVEFLPVMEHPFYGSWGYQITGYFAPTSRYGTPQDFMYLIDKLHQEGIGVILD

WVPS

Synechococcus

GYSHIELLPIAEHPFDGSWGYQVTGYYAATSRYGSPEDFMYFVDRCHQNGIGVILD

WVPG

Synechocystis

GYTHIELLPIAEHPFDGSWGYQVTGYYAPTSRFGSPEDFMYFVDQCHLNGIGVIID

WVPG

Butyrivibrio

KYTHIELIGIAEHPFDGSWGYQVTGYYAPTARYGEPTDFMYLINQLHKHGIGVILD

WVPA

CHLAMYDIA

HYTHVELLPVTEHPLNESWGYQTTGYYAPTSRYGSFEDLQYFIDTMHQHGIGVIL

DWVPG

      ..*.*.. . *.* ***** ,.. * *   : .. * ,.*..****,

Bstearothermophilus

HFCKDAHGLYMFDGAPTYEYANEKDRENYVWGTANFDLGKPEVRSFLISNALFW

LEYYHI

Bcaldolyticus

5   HFCKDAHGLYMFDGAPTYEYANEKDRENYVWGTANFDLGKPEVRSFLISNALFW

LEYYHV

Bsubtilis

HFCKDAHGLYMFDGEPLYEYKEERDRENWLWGTANFDLGKPEVHSFLISNALY

WAEFYHI

10   mycobacterium

HFPKDAWALGRFDGTPLYEHSDPKRGEQLDWGTYVFDFGRPEVRNFLVANALY

WLQEFHI

Streptomyces

HFPRDDWALAEFDGRPLYEHQDPRRAAHPDWGTLEFDYGRKEVRNFLVANAVY

15   WCQEFHV

E.

HFPTDDFALAEFDGTNLYEHSDPREGYHQDWNTLIYNYGRREVSNFLVGNALYW

IERFGI

H.influenzae

20   HFPSDTHGLVAFDGTALYEHEDPREGYHQDWNTLIYNYGRNEVKNFLSSNALYW

LERFGV

Agro.tume

HFPTDEHGLRWFDGTALYEHADPRQGFHPDWNTAIYNFGRIEVMSYLINNALYW

AEKFHL

25   Aquifex_a.

HFPTDAHGLAYFDGTHLYEYEDWRKRWHPDWNSFVFDYGKPEVRSFLLSSAHF

WLDKYHA

Synechococcus

HFPKDGHGLAFFDGTHLYEHADSRQGEHREWGTLVFNYGRHEVRNFLAANALF

30   WFDKYHI

Synechocystis

HFPKDGHGLAFFDGTHLYEHGDPRKGEHKEWGTLIFNYGRNEVRNFLVANALF

WFDKYHI

Butyrivibrio

HFCPDEFGLACFDGTCIYEDPDPRKGEHPDWGTKIFNLAKPEVKNFLIANALYWI

RKFHI

CHLAMYDIA

5   HFPIDSFAMSGFDGTPLYEYTRNPSPLHPHWHTYTFDYAKPEVCNFLLGSVLFWI

DKMHV

           ** * .. *** **   : *: ::..** .* ..:*

Bstearothermophilus   DGFRVDAVANMLYWPNNDRL------YE----

10   NPYAVEFLRKLNEAVFAYDPNALMIAED

Bcaldolyticus   DGFRVDAVANMLYWPNNDRL------YE----

NPYAVEFLRQLNEAVFAYDPNVWMIAED

Bsubtilis   DGFRVDAVANILYWPNQDER------HT----

NPYAVDFLKKLNQTMREAYPHVMMIAED

15   mycobacterium

DGLRVDAVASMLYLDYSRPEGGWTPNVHGGRENLEAVQFLQEMNATAHKVAPGI

VTIAEE

Streptomyces

DGLRADAVASMLYLDYSRDEGDWSPNAHGGREDLDAVALLQEMNATVYRRFPGV

20   VTIAEE

E.

DALRVDAVASMIYRDYSRKEGEWIPNEFGGRENLEAIEFLRNTNRILGEQVSGAVT

MAEE

H.influenzae

25   DGIRVDAVASMIYRDYSRAEGEWIPNQYGGRENLEAIEFLKHTNWKIHSEMAGAI

SIAEE

Agro.tume

DGLRVDAVASMLYLDYSRKEGEWIPNEYGGRENLESVRFLQKMNSLVYGTHPGV

MTIAEE

30   Aquifex_a.   DGLRVDAVASMLYLDYSRKE--

WVPNIYGGKENLEAIEFLRKFNESVYRNFPDVQTIAEE

Synechococcus

DGIRVDAVASMLYLDYNRKEGEWIPNEYGGRENIEAADFLRQVNHLIFSYFPGALS

IAEE

Synechocystis

DGMRVDAVASMLYLDYCREEGEWVANEYGGRENLEAADFLRQVNSVVYSYFPGI
LSIAEE

Butyrivibrio

5   DGLRVDAVASMLYLDYGKKDGQWVPNKYGDNKNLDAIEFFKHFNSVVRGTYPNI
LTIAEE

CHLAMYDIA

DGIRVDAVSSMLYLDYGRYAGEWVPNRYGGRENLDAIRFLQQFNTVIHEKYPGVL
TFAEE

10   * .:* .***:...*   : : .::. *   .   .**:

Bstearothermophilus

STDWPKVTAPTYEGGLGFNYKWNMGWMNDMLKYMETPPYERRHVHNQVTFSL
LYAYSENF

15   Bcaldolyticus

STDWPRVTAPTYDGGLGFNYKWNMGWMNDMLKYMETPPHERKYAHNQVSFSL
LYAYSENF

Bsubtilis

STEWPQVTGAVEEGGLGFHYKWNMGWMNDVLKYMETPPEERRHCHQLISFSLL

20 YAFSEHF

mycobacterium

STPWSGVTRPTNIGGLGFSMKWNMGWMHDTLDYVSRDPVYRSYHHHEMTFSML
YAFSENY

Streptomyces

25 STAWDGVTRPTDSGGLGFGLKWNMGWMHDTLRYVSKEPVHRKYHHHDMTFGM
VYAFSENF

E.

STDFPGVSRPQDMGGLGFWYKWNLGWMHDTLDYMKLDPVYRQYHHDKLTFGI
LYNYTENF

30 H.influenzae

STSFAGVTHPSENGGLGFNFKWNMGWMNDTLAYMKLDPIYRQYHHNKMTFGM
VYQYSENF

Agro.tume

STSWPKVSQPVHEGGLGFGFKWNMGFMHDTLSYFSREPVHRKFHHQELTFGLL
YAFTENF

Aquifex_a.

5   STAWPMVSRPTYVGGLGFGMKWNMGWMNDTLFYFSKDPIYRKYHHEVLTFSIW
YAFSENF

Synechococcus

STSWPMVSWPTYVGGLGFNLKWNMGWMHDMLDYFSMDPWFRQFHQNNVTFSI
WYAFSENF

10 Synechocystis

STSWPMVSWPTYVGGLGFNLKWNMGWMHDMLDYFSMDPWFRQFHQNSITFSM
WYNHSENY

Butyrivibrio

STAWPKVTAPPEEDGLGFAFKWNMGWMHDFCEYMKLDPYFRQGAHYMMTFAM

15 SYNDSENY

CHLAMYDIA

STTFPKITVSVEEGGLGFDYKWNMGWMHDTLHYFEKDFPYRPYHQSDLTFPQW
YAFSERF

          ** . .. . .**** ***.*.*.* *..   * : ..* * .*.

20 Bstearothermophilus

ILPFSHDEVVHGKKSLLNKMPGSYEEKFAQLRLLYGYMMAHPGKKLLFMGNEFA
QFDEWK

Bcaldolyticus

ILPFSHDEVVHGKKSLLNKMPGSYEEKFAQLRLLYGYMMAHPGKKLLFMGSEFA

25 QFDEWK

Bsubtilis

VLPFSHDEVVYGKKSLLNKMPGDYWQKFAQYRLLLGYMTVHPGKKLIFMGSEFA
QFDEWK

mycobacterium

30 VLPLSHDEVVHGKGTLWGRMPGNNHVKAAGLRSLLAYQWAHPGKQLLFMGQEF
GQRAEWS

Streptomyces

VLPISHDEVVHGKRSLVSKMPGDWWQQRATHRAYLGFMWAHPGKQLLFMGQEF
AQGSEWS

E.

VLPLSHDEVVHGKKSILDRMPGDAWQKFANLRAYYGWMWAFPGKKLLFMGNEF

AQGREWN

H.influenzae

VLPLSHDEVVHGKYSLLGKMPGDTWQKFANLRAYYGYMWGYPGKKLLFMGNEF

AQGREWN

Agro.tume

VLPLSHDEVVHGKGSLIAKMSGDDWQKFANLRSYYGFMWGYPGKKLLFMGQEF

AQWSEWS

Aquifex_a.

VLPLSHDEVVHGKGSLIGKMPGDYWQKFANLRALFGYMWAHPGKKLLFMGGEF

GQFKEWD

Synechococcus

MLALSHDEVVHGKSNLIGKMPGDEWQKFANLRCLLGYMFTHPGKKTLFMGMEF

GQWAEWN

Synechocystis

MLALSHDEVVHGKSNMLGKMPGDEWQKYANVRALFTYMFTHPGKKTMFMSME

FGQWSEWN

Butyrivibrio

ILPLSHDEVVHLKCSMVEKMPGYKVDKYANLRVGYTYMFGHSGKKLLFMGQDF

GQEREWS

CHLAMYDIA

LLPFSHDEVVHGKRSLIGKMPGDAWRQFAQLRLLLGYQICQPGKKLLFMGGEFG

QGREWS

          .*..******. * .. .*.*   . * *   .   .**..**..*.* **.

Bstearothermophilus   FEDELDWVLFDF-----

ELHRKMNDYMKELIACYKRYKPFYELDHDPQGFEWIDVHNAEQ

Bcaldolyticus   FAEELDWVLFDF-----

ELHRKMDEYVKQLIACYKRYKPFYELDHDPRGFEWIDVHNAEQ

Bsubtilis   DTEQLDWFLDSF-----

PMHQKASVFTQDLLRFYQKSKILYEHDHRAQSFEWIDVHNDEQ

mycobacterium   EQRGLDWFQLDE----

NGFSNGIQRLVRDINDIYRCHPALWSLDTTPEGYSWIDANDSAN

```
Streptomyces
ETYGPDWWVLDSSYPAAGDHLGVRSLVRDLNRTYTASPALWERDSVPEGFAWVE
ADAADD
E.              HDASLDWHLLEG---
GDNWHHGVQRLVRDLNLTYRHHKAMHELDFDPYGFEWLVVDDKER
H.influenzae      YEESLDWFLLDENI-
GGGWHKGVLKLVKDLNQIYQKNRPLFELDNSPEGFDWLVVDDAAN
Agro.tume      EKGSLDWNLRQY-----
PMHEGMRRLVRDLNLTYRSKAALHARDCEPDGFRWLVVDDHEN
Aquifex_a.      HETSLDWHLLEY-----
PSHRGIQRLVKDLNEVYRREKALHETDFSPEGFEWVDFHDWEK
Synechococcus      VWGDLEWHLLQY-----
EPHQGLKQFVKDLNHLYRNAPALYSEDCNQAGFEWIDCSDNRH
Synechocystis      VWGDLEWHLLNF-----
PPHQQLKQFFTELNHLYKNEPALYSNDFDESGFQWIDCSDNRH
Butyrivibrio      EKRELDWFLLEN-----
DLNRGMKDYVGKLLEIYRKYPALYEVDNDWGGFEWINADDKER
CHLAMYDIA      PGRELDWELLDI-----      ·
SYHQGVHLCSQELNALYVQSPQLWQADHLPSSFRWVDFSDVRN
                :*  .         .: *   :  *  .:*:


Bstearothermophilus   SIFSFIRRGKKED-DVLVIVCNFTNQAYDDYKVGVP-
LLVPYREVLNSDAVTFGGSGHVN
Bcaldolyticus      SIFSFIRRGKKEG-DVLVIVCNFTNQAYDDYKVSVP-
LLAPYREVLNSDAAEFGGSGHVN
Bsubtilis      SIFSFIRYGQKHG-EALVIICNFTPVVYHQYDVGVP-
FFTQYIEVLNSDSETYGGSGQIN
mycobacterium      NVLSFMRYGSDG--SVLACVFNFAGAEHRDYRLGLP-
RAGRWREVLNTDATIYHGSGIGN
Streptomyces      NVFAFLRFARDG--
SPLLCVSNFSPVVRHGYRIGVPQEVGQWREVLNTDLEPYGGSGVHH
E.              SVLIFVRRDKEG--NEIIVASNFTPVPRHDYRFGIN-
QPGKWREILNTDSMHYHGSNAGN
```

```
        H.influenzae        SVLAFERRSSNG--ERIIVVSNFTPVPRHNYRIGVN-
VAGKYEEILNTDSMYYEGSNVGN
        Agro.tume           SVFAWLRTAPGE--KPVAVICNLTPVYRENYYVPLG-
VAGRWREILNTDAEIYGGSGKGN
5       Aquifex_a.          SVISFLRKDKSGK-EIILVVCNFTPVPRYDYRVGVP-
KGGYWREIMNTDAKEYWGSGMGN
        Synechococcus       SIVSFIRRAHESD-RFLVVVCNFTPQPHAHYRIGVP-
VAGFYREIFNSDARSYGGSNMGN
        Synechocystis       SVVSFIRRAKNSA-EFVVTICNFTPQPHSHYRVGVP-
10      VPGFYTELFNSDARQYGGSNMGN
        Butyrivibrio        STYSFYRRASNGK-DNILFVLNMTPMERKGFKVGVP-
FDGTYTKILDSAKECYGGSGSSV
        CHLAMYDIA           GVVAYLRFADADAKKALLCVHHFGVGYFPHYLLPIL-
PLESCDLLMNTDDTRFGGSGKGF
15              .         . :*        :  ::      : . :        ::::   : **.


        Bstearothermophilus    GKR-LSAFNEPFHGK------------P--
YHVRMTIPPFGISILRPVQKRGERKRNEK
        Bcaldolyticus          GKR-LPAFSEPFHGK------------P--
20      YHVRMTIPPFGISILRPVQKRGERKQNEE
        Bsubtilis              KKP-LSAKKGALHHK------------P--
CYITMTIPPYGISILRAVKKRGEIKR---
        mycobacterium          LGG-VDATDDPWHGR------------P--
ASAVLVLPPTSALWLTPA-----------
25      Streptomyces           ARA-LRPEPVPAQGR-------------A--VSLRMTLPPMATVWLRP----
---------
        E.                  GGT-VHSDEIASHGR------------Q--HSLSLTLPPLATIWLVREAE-------
--
        H.influenzae        FGC-VASEQIESHGR------------E--NSISVSIPPLATVYLRLKTK--
30      -------
        Agro.tume           GG----RVQAVDAGG------------E--IGAMLVLPPLATIMLEPEN-----
-----
        Aquifex_a.          LGG-KEADKIPWHGR------------K--FSLSLTLPPLSVIYLKHEG---
-------
```

```
Synechococcus        LGG-KWTDEWSCHNR------------P--
YSLDLCLPPLTTLVLELASGPES----LS
Synechocystis        LGG-KWTEEWSFHEQ------------P--
YSLDLCLPPLSVLVLKLSQNAEENTVPAE
```

⁵
```
Butyrivibrio         PDK-IKAVKGLCDYK------------D--
YSIEFDLPPYGAEVFVFQTKKTKN-----
CHLAMYDIA
REPEILTPEIARQEREAAGLIEADDESGPDCWGLDIELPPSATLIFSVTLQ--------
                          . :**   :
```

10

```
Bstearothermophilus    EMHRHVIGRRARKSASLADDKHR------------
Bcaldolyticus          EVHRHVIGRRARKPASLADEKHRETSRAVWGEVPDH
Bsubtilis              -----------------------------------
mycobacterium          -----------------------------------
Streptomyces           -----------------------------------
E.                     -----------------------------------
H.influenzae           -----------------------------------
Agro.tume              -----------------------------------
Aquifex_a.             -----------------------------------
Synechococcus          EAANSPL----------------------------
Synechocystis          EASNIA-----------------------------
Butyrivibrio           -----------------------------------
CHLAMYDIA              -----------------------------------
```

25

Alignement of BE and isoamylases.

Bstearothermophilus   --

5   Bcaldolyticus   --

Bsubtilis   --

mycobacterium   --------------------------MSRSEKLTGEH-LAPEPA----
EMARLVAGT

Streptomyces

10 MSAARQPSPTVRDKAAPEPAAPAAPKGARAPRARRAAPPHGVRPAPALAAEERAR
LLEGR

E.   ---------------------------------MSDRIDRDVINALIAGH

H.influenzae   -------------------------------------MTTAVTQAIIDGFFDAS

Agro.tume   ----------------------------MKKPLNSAEEKKTGDITKAEIEAIKSGL

15 Aquifex_a.   --

Synechococcus   ----------------------------------TGTTPLPSSSLSVEQVNRIASNQ

Synechocystis   ----------------------------------MTYTINADQVHQIVHNL

Butyrivibrio   ---

CHLAMYDIA   ------------------------------------MDPFFLNTQHVELLVSGK

20

Bstearothermophilus   --

Bcaldolyticus   --

Bsubtilis   --

25 mycobacterium   HHNPHGILGAHEYDDHTVIR-----AFRPHAVEVVALVGK---
DRFSLQHLD-SGLFAVA

Streptomyces   HHDPHAVLGARTQRGGVAFR-----VLRPYAKAVTVVAKG---
LRTELVDEG-DGLFSGL

E.   FADPFSVLGMHKTTAGLEVR-----

30 ALLPDATDVWVIEPKTGRKLAKLECLDSRGFFSGV

H.influenzae   NGDPFATLGMHETEQGIEIR-----
TLLPDANRMVVIERESGKEITELDCVDERGFFVGV

Agro.tume   HSNPFQIIPLHETPEGFSAR-----CFIPGAEEVSVLTLD-
GNFVGELKQIDPDGFFEGR

```
Aquifex_a.        ----------------------------------------------------

Synechococcus     EQNPFDILGPHPYEHEGQAG-WVIRAYLPEAQEAAVICPAL-

RREFAMHPVHHPHFFETW

Synechocystis

HHDPFEVLGCHPLGDHGKVNQWVIRAYLPTAEAVTVLLPTD-

RREVIMTTVHHPNFFECV

Butyrivibrio      --------------------------------------------------------

CHLAMYDIA         QSSPQDLLGIVS-ESLNQDR---IVLFRPGAETVFVELRG----

KIQQAESHHSGIFSLP
```

```
Bstearothermophilus     ----------------------------

MIAANPTDLEVYLFHEGSLYKSYELFGAHV--

Bcaldolyticus           ----------------------------

MIAANPTDLEVYLFHEGRLYQSYELFGAHV--

Bsubtilis               ----------------------------

MAAASPTAHDVYLFHEGSLFKSYQLFGSHY--

mycobacterium     LPFVD-

LIDYRLQVTYEGCEPHTVADAYRFLPTLGEVDLHLFAEGRHERLWEVLGAHPRS

Streptomyces      LPLTG-VPDYRLLVTYDSDE-

IEVHDPYRFLPALGELDLHLIGEGRHEELWTALGSQP--

E.                IPRRKNFFRYQLAVVWHGQQ-

NLIDDPYRFGPLIQEMDAWLLSEGTHLRPYETLGAHA--

H.influenzae      IPNCRQFFAYQLQVFWGNEA-

QIIEDPYRFHPMIDDLEQWLLSEGSMLRPYEVLGAHF--

Agro.tume         IDLSK-RQPVRYRACRDDAE-

WAVTDPYSFGPVLGPMDDYFVREGSICGYSTGWARIP--

Aquifex_a.        ----------------------------

MKKFSLISDYDVYLFKEGTHTRLYDKLGSHV--

Synechococcus     VPEET-

LEIYQLRITEGERERIIYDPYAFRSPLLTDYDIHLFAEGNHHRIYEKLGAHP--

Synechocystis     LELEE-

PKNYQLRITENGHERVIYDPYGFKTPKLTDFDLHVFGEGNHHRIYEKLGAHL--
```

Butyrivibrio   -----------------------

MSQKVFISEDDEYLFGQGTHYDIYDKLGAHP--

CHLAMYDIA   VMKGISPQDYRVYHQN-G---

LLAHDPYAFPLLWGEIDSFLFHEGTHQRIYERMGAIP--

5   : ..:*   .

Bstearothermophilus   -INEGG-

KVGTRFCVWAPHAREVRLVGSFNDWDGTDFRLEKVND-EGVWTIVVPENLEGH

Bcaldolyticus   -IRGGG-

10   AVGTRFCVWAPHAREVRLVGSFNDWNGTNSPLTKVND-EGVWTIVVPENLEGH

Bsubtilis   -RELNG-

KSGYEFCVWAPHASEVRVAGDFNSWSGEEHVMHRVND-NGIWTLFIPGIGEKE

mycobacterium

FTTADGVVSGVSFAVWAPNAKGVSLIGEFNGWNGHEAPMRVLGP-

15   SGVWELFWPDFPCDG

Streptomyces   -MEHQG-

VAGTRFTVWAPNALGVRVTGDFSYWDAVAYPMRSLGA-SGVWELFLPGVAEGA

E.   -DTMDG-

VTGTRFSVWAPNARRVSVVGQFNYWDGRRHPMRLRKE-SGIWELFIPGAHNGQ

20   H.influenzae   -MECDG-

VSGVNFRLWAPNARRVSIVGDFNYWDGRRHPMRFHSK-SGVWELFLPKASLGQ

Agro.tume   -LKLEG-

VEGFHFAVWAPNGRRVSVVGDFNNWDGRRHVMRFRKD-TGIWEIFAPDVYA-C

Aquifex_a.   -IELNG-

25   KRYTFFAVWAPHADYVSLIGDFNEWDKGSTPMVKREDGSGIWEVLLEGDLTGS

Synechococcus   -CELEN-

VAGVNFAVWAPSARNVSILGDFNSWDGRKHQMAR-RS-NGIWELFIPELTVGA

Synechocystis   -MTVDG-

VKGVYFAVWAPNARNVSILGDFNNWDGRLHQMRK-RN-NMVWELFIPELGVGT

30   Butyrivibrio   -SEEKG-

KKGFFFAVWAPNAADVHVVGDFNGWDENAHQMKRSKT-GNIWTLFIPGVAIGA

CHLAMYDIA   -CEIDG-

VPGVRFIVWAPHAQRVSVIGDFNGWHGLVNPLHKVSD-QGVWELFVPGLTAGA

.   *.*** . *.:.*.* :   .* ..

Bstearothermophilus   LYKYEIVTPDGQVL-

FKADPYAFYSELRPHTASIAYDLKGYQWNDQSWKRKKRRKRIYDQ

Bcaldolyticus   LYKYEIITPDGRVL-

5   LKADPYAFYSELRPHTASIVYDLKGYEWNDSPWQRKKRRKRIYDQ

Bsubtilis   RYKYEIVTNNGEIR-

LKADPYAIYSEVRPNTASLTYDLEGYSWQDQKWQKKQKAKTLYEK

mycobacterium   LYKFRVHGADGVVT-DRADPFAFGTEVPPQTASRVT-

SSDYTWGDDDWMAGRALRNPVNE

10   Streptomyces   LYKYEITRPDGGRT-LRADPMARYAEVPPANASIVT-

ASRYEWQDAEWMARRGALAPHQA

E.   LYKYEMIDANGNLR-LKSDPYAFEAQMRPETASLIC-

GLPEKVVQTEERKKANQFDA---

H.influenzae   LYKFELIDCHGNLR-LKADPFAFSSQLRPDTASQVS-

15   ALPNVVEMTEARKKANQGNQ---

Agro.tume

AYKFEILGANGELLPLKADPYARRGELRPKNASVTAPELTQKWEDQAHREHWAQ

VDQRRQ

Aquifex_a.   KYKYFIKNGNYEVD--KSDPFAFFCEQPPGNASVVW-

20   KLNYRWNDSEYMKKRKRVNSHDS

Synechococcus   AYKYEIKNYDGHIYE-

KSDPYGFQQEVRPKTASIVADLDRYTWGDADWLERRRHQEPLRQ

Synechocystis   SYKYEIKNWEGHIYE-

KTDPYGFYQEVRPKTASIVADLDGYQWHDEDWLEARRTSDPLSK

25   Butyrivibrio   LYKFLITAQDGRKLY-

KADPYANYAELRPGNASRTTDLSGFKWSDSKWYESLKGKDMNRQ

CHLAMYDIA   CYKWEMVTESGQVL-IKSDPYGKFFGPPPWSVSVVI-

DDSYEWTDSEWLEERIKKTEG--

                .:   ...   * ..*

30   Bstearothermophilus   PMVIYELHFGSWKKK----------------------

DGRFYTYREMADELISYVLDH

Bcaldolyticus   PMVIYELHFGSWKKKP----------------------

DGRFYTYREMADELIPYVLER

```
     Bsubtilis          PVFIYELHLGSWKKHS--------------------
DGRHYSYKELSQTLIPYIKKH
     mycobacterium      AMSTYEVHLGSWRP-------------------------
GLSYRQLARELTDYIVDQ
 5   Streptomyces       PMSVYELHLASWRP--------------------------
GLSYRQLAEQLPAYVKEL
     E.                 PISIYEVHLGSWRRH----------TDNN----------
FWLSYRELADQLVPYAKWM
     H.influenzae       PISIYEVHLGSWRRN----------LENN----------
10   FWLDYDQIADELIPYVKEM
     Agro.tume          PISIYEVHAGSWQR-----------SEDG----------
TFLSWDELEAQLIPYCTDM
     Aquifex_a.         PISIYEVHVGSWRRVP---------EEGN----------
RFLSYRELAEYLPYYVKEM
15   Synechococcus
PISVYEVHLGSWMHASSDAIATDAQGKPLPPVPVADLKPGARFLTYRELADRLIPY
VLDL
     Synechocystis      PVSVYELHLGSWLHTAYDEPVKTLHGEGVP-
VEVSEWNTGARFLTYYELVDKLIPYVKEL
20   Butyrivibrio       PIAIYECHIGSWMKHP--------DGTEDG-----------
FYTYRQFADRIVEYLKEM
     CHLAMYDIA          PMNIYEVHVGSWRWQE---------------------
GQPLNYKELADQLALYCKQM
            .:  ** * .**              : ::  :  *
25

     Bstearothermophilus
GFTHIELLPLVEHPLDRSWGYQGTGYYAVTSRYGTPHDFMYFVDRCHQAGIGVIM
DWVPG
     Bcaldolyticus
30   GFTHIELLPLVEHPLDRSWGYQGTGYYSVTSRYGTPHDFMYFVDRCHQAGLGVII
DWVPG
     Bsubtilis
GFTHIELLPVYEHPYDRSWGYQGTGYYSPTSRFGPPHDLMKFVDECHQQNIGVIL
DWVPG
```

mycobacterium

GFTHVELLPVAEHPFAGSWGYQVTSYYAPTSRFGTPDDFRALVDALHQAGIGVIVD
WVPA

Streptomyces

5   GFTHVELMPVAEHPFGGSWGYQVTGFYAPTSRMGTPDDFRFLVDALHRAGIGVIV
DWVPA

E.
GFTHLELLPINEHPFDGSWGYQPTGLYAPTRRFGTRDDFRYFIDAAHAAGLNVILD
WVPG

10   H.influenzae
GFTHIEFLPLSEFPFDGSWGYQPLGLYSPTSRFGSPEAFRRLVKRAHEAGINVILD
WVPG

Agro.tume
GFTHIEFLPITEHPYDPSWGYQTTGLYAPTARFGDPEGFARFVNGAHKVGIGVLLD
15   WVPA

Aquifex_a.
GFTHVEFLPVMEHPFYGSWGYQITGYFAPTSRYGTPQDFMYLIDKLHQEGIGVILD
WVPS

Synechococcus
20   GYSHIELLPIAEHPFDGSWGYQVTGYYAATSRYGSPEDFMYFVDRCHQNGIGVILD
WVPG

Synechocystis
GYTHIELLPIAEHPFDGSWGYQVTGYYAPTSRFGSPEDFMYFVDQCHLNGIGVIID
WVPG

25   Butyrivibrio
KYTHIELIGIAEHPFDGSWGYQVTGYYAPTARYGEPTDFMYLINQLHKHGIGVILD
WVPA

CHLAMYDIA
HYTHVELLPVTEHPLNESWGYQTTGYYAPTSRYGSFEDLQYFIDTMHQHGIGVIL
30   DWVPG

                ..*.*.. . *.* ***** .. * *   . .. * ..*..****.
                .:.:.::.:.*.   .:   .::   * :. ..:.:.:

Bstearothermophilus

HFCKDAHGLYMFDGAPTYEYANEKDRENYVWGTANFDLGKPEVRSFLISNALFW

LEYYHI

Bcaldolyticus

HFCKDAHGLYMFDGAPTYEYANEKDRENYVWGTANFDLGKPEVRSFLISNALFW

LEYYHV

Bsubtilis

HFCKDAHGLYMFDGEPLYEYKEERDRENWLWGTANFDLGKPEVHSFLISNALY

WAEFYHI

mycobacterium

HFPKDAWALGRFDGTPLYEHSDPKRGEQLDWGTYVFDFGRPEVRNFLVANALY

WLQEFHI

Streptomyces

HFPRDDWALAEFDGRPLYEHQDPRRAAHPDWGTLEFDYGRKEVRNFLVANAVY

WCQEFHV

E.

HFPTDDFALAEFDGTNLYEHSDPREGYHQDWNTLIYNYGRREVSNFLVGNALYW

IERFGI

H.influenzae

HFPSDTHGLVAFDGTALYEHEDPREGYHQDWNTLIYNYGRNEVKNFLSSNALYW

LERFGV

Agro.tume

HFPTDEHGLRWFDGTALYEHADPRQGFHPDWNTAIYNFGRIEVMSYLINNALYW

AEKFHL

Aquifex_a.

HFPTDAHGLAYFDGTHLYEYEDWRKRWHPDWNSFVFDYGKPEVRSFLLSSAHF

WLDKYHA

Synechococcus

HFPKDGHGLAFFDGTHLYEHADSRQGEHREWGTLVFNYGRHEVRNFLAANALF

WFDKYHI

Synechocystis

HFPKDGHGLAFFDGTHLYEHGDPRKGEHKEWGTLIFNYGRNEVRNFLVANALF

WFDKYHI

Butyrivibrio

HFCPDEFGLACFDGTCIYEDPDPRKGEHPDWGTKIFNLAKPEVKNFLIANALYWI

RKFHI

CHLAMYDIA

5   HFPIDSFAMSGFDGTPLYEYTRNPSPLHPHWHTYTFDYAKPEVCNFLLGSVLFWI

DKMHV

              ** * .: *** **   : * : :: .: ** .* .. :*

Bstearothermophilus   DGFRVDAVANMLYWPNNDRL------YE----

10   NPYAVEFLRKLNEAVFAYDPNALMIAED

Bcaldolyticus   DGFRVDAVANMLYWPNNDRL------YE----

NPYAVEFLRQLNEAVFAYDPNVWMIAED

Bsubtilis   DGFRVDAVANILYWPNQDER------HT----

NPYAVDFLKKLNQTMREAYPHVMMIAED

15   mycobacterium

DGLRVDAVASMLYLDYSRPEGGWTPNVHGGRENLEAVQFLQEMNATAHKVAPGI

VTIAEE

Streptomyces

DGLRADAVASMLYLDYSRDEGDWSPNAHGGREDLDAVALLQEMNATVYRRFPGV

20   VTIAEE

E.

DALRVDAVASMIYRDYSRKEGEWIPNEFGGRENLEAIEFLRNTNRILGEQVSGAVT

MAEE

H.influenzae

25   DGIRVDAVASMIYRDYSRAEGEWIPNQYGGRENLEAIEFLKHTNWKIHSEMAGAI

SIAEE

Agro.tume

DGLRVDAVASMLYLDYSRKEGEWIPNEYGGRENLESVRFLQKMNSLVYGTHPGV

MTIAEE

30   Aquifex_a.   DGLRVDAVASMLYLDYSRKE--

WVPNIYGGKENLEAIEFLRKFNESVYRNFPDVQTIAEE

Synechococcus

DGIRVDAVASMLYLDYNRKEGEWIPNEYGGRENIEAADFLRQVNHLIFSYFPGALS

IAEE

Synechocystis

DGMRVDAVASMLYLDYCREEGEWVANEYGGRENLEAADFLRQVNSVVYSYFPGI

LSIAEE

Butyrivibrio

5 DGLRVDAVASMLYLDYGKKDGQWVPNKYGDNKNLDAIEFFKHFNSVVRGTYPNI

LTIAEE

CHLAMYDIA

DGIRVDAVSSMLYLDYGRYAGEWVPNRYGGRENLDAIRFLQQFNTVIHEKYPGVL

TFAEE

10   *..*.***...*   : : :::. *   . .**.

Bstearothermophilus

STDWPKVTAPTYEGGLGFNYKWNMGWMNDMLKYMETPPYERRHVHNQVTFSL

LYAYSENF

15 Bcaldolyticus

STDWPRVTAPTYDGGLGFNYKWNMGWMNDMLKYMETPPHERKYAHNQVSFSL

LYAYSENF

Bsubtilis

STEWPQVTGAVEEGGLGFHYKWNMGWMNDVLKYMETPPEERRHCHQLISFSLL

20 YAFSEHF

mycobacterium

STPWSGVTRPTNIGGLGFSMKWNMGWMHDTLDYVSRDPVYRSYHHHEMTFSML

YAFSENY

Streptomyces

25 STAWDGVTRPTDSGGLGFGLKWNMGWMHDTLRYVSKEPVHRKYHHHDMTFGM

VYAFSENF

E.

STDFPGVSRPQDMGGLGFWYKWNLGWMHDTLDYMKLDPVYRQYHHDKLTFGI

LYNYTENF

30 H.influenzae

STSFAGVTHPSENGGLGFNFKWNMGWMNDTLAYMKLDPIYRQYHHNKMTFGM

VYQYSENF

Agro.tume

STSWPKVSQPVHEGGLGFGFKWNMGFMHDTLSYFSREPVHRKFHHQELTFGLL

YAFTENF

Aquifex_a.

STAWPMVSRPTYVGGLGFGMKWNMGWMNDTLFYFSKDPIYRKYHHEVLTFSIW

YAFSENF

Synechococcus

STSWPMVSWPTYVGGLGFNLKWNMGWMHDMLDYFSMDPWFRQFHQNNVTFSI

WYAFSENF

Synechocystis

STSWPMVSWPTYVGGLGFNLKWNMGWMHDMLDYFSMDPWFRQFHQNSITFSM

WYNHSENY

Butyrivibrio

STAWPKVTAPPEEDGLGFAFKWNMGWMHDFCEYMKLDPYFRQGAHYMMTFAM

SYNDSENY

CHLAMYDIA

STTFPKITVSVEEGGLGFDYKWNMGWMHDTLHYFEKDFPYRPYHQSDLTFPQW

YAFSERF

      ** : :: . .**** ***.*.*.* *.. * : ..* * .*. .

Bstearothermophilus

ILPFSHDEVVHGKKSLLNKMPGSYEEKFAQLRLLYGYMMAHPGKKLLFMGNEFA

QFDEWK

Bcaldolyticus

ILPFSHDEVVHGKKSLLNKMPGSYEEKFAQLRLLYGYMMAHPGKKLLFMGSEFA

QFDEWK

Bsubtilis

VLPFSHDEVVYGKKSLLNKMPGDYWQKFAQYRLLLGYMTVHPGKKLIFMGSEFA

QFDEWK

mycobacterium

VLPLSHDEVVHGKGTLWGRMPGNNHVKAAGLRSLLAYQWAHPGKQLLFMGQEF

GQRAEWS

Streptomyces

VLPISHDEVVHGKRSLVSKMPGDWWQQRATHRAYLGFMWAHPGKQLLFMGQEF

AQGSEWS

E.

VLPLSHDEVVHGKKSILDRMPGDAWQKFANLRAYYGWMWAFPGKKLLFMGNEF
AQGREWN

H.influenzae

VLPLSHDEVVHGKYSLLGKMPGDTWQKFANLRAYYGYMWGYPGKKLLFMGNEF
AQGREWN

Agro.tume

VLPLSHDEVVHGKGSLIAKMSGDDWQKFANLRSYYGFMWGYPGKKLLFMGQEF
AQWSEWS

Aquifex_a.

VLPLSHDEVVHGKGSLIGKMPGDYWQKFANLRALFGYMWAHPGKKLLFMGGEF
GQFKEWD

Synechococcus

MLALSHDEVVHGKSNLIGKMPGDEWQKFANLRCLLGYMFTHPGKKTLFMGMEF
GQWAEWN

Synechocystis

MLALSHDEVVHGKSNMLGKMPGDEWQKYANVRALFTYMFTHPGKKTMFMSME
FGQWSEWN

Butyrivibrio

ILPLSHDEVVHLKCSMVEKMPGYKVDKYANLRVGYTYMFGHSGKKLLFMGQDF
GQEREWS

CHLAMYDIA

LLPFSHDEVVHGKRSLIGKMPGDAWRQFAQLRLLLGYQICQPGKKLLFMGGEFG
QGREWS

        .* .*******. * .   .* *   . * *   :   .**:.**..*.* **.

Bstearothermophilus   FEDELDWVLFDF-----
ELHRKMNDYMKELIACYKRYKPFYELDHDPQGFEWIDVHNAEQ
Bcaldolyticus   FAEELDWVLFDF-----
ELHRKMDEYVKQLIACYKRYKPFYELDHDPRGFEWIDVHNAEQ
Bsubtilis   DTEQLDWFLDSF-----
PMHQKASVFTQDLLRFYQKSKILYEHDHRAQSFEWIDVHNDEQ
mycobacterium   EQRGLDWFQLDE----
NGFSNGIQRLVRDINDIYRCHPALWSLDTTPEGYSWIDANDSAN

Streptomyces

ETYGPDWWVLDSSYPAAGDHLGVRSLVRDLNRTYTASPALWERDSVPEGFAWVE

ADAADD

E.   HDASLDWHLLEG---

GDNWHHGVQRLVRDLNLTYRHHKAMHELDFDPYGFEWLVVDDKER

H.influenzae   YEESLDWFLLDENI-

GGGWHKGVLKLVKDLNQIYQKNRPLFELDNSPEGFDWLVVDDAAN

Agro.tume   EKGSLDWNLRQY-----

PMHEGMRRLVRDLNLTYRSKAALHARDCEPDGFRWLVVDDHEN

Aquifex_a.   HETSLDWHLLEY-----

PSHRGIQRLVKDLNEVYRREKALHETDFSPEGFEWVDFHDWEK

Synechococcus   VWGDLEWHLLQY-----

EPHQGLKQFVKDLNHLYRNAPALYSEDCNQAGFEWIDCSDNRH

Synechocystis   VWGDLEWHLLNF-----

PPHQQLKQFFTELNHLYKNEPALYSNDFDESGFQWIDCSDNRH

Butyrivibrio   EKRELDWFLLEN-----

DLNRGMKDYVGKLLEIYRKYPALYEVDNDWGGFEWINADDKER

CHLAMYDIA   PGRELDWELLDI-----

SYHQGVHLCSQELNALYVQSPQLWQADHLPSSFRWVDFSDVRN

                        :*   .   .: *   : *   .: *:

Bstearothermophilus   SIFSFIRRGKKED-DVLVIVCNFTNQAYDDYKVGVP-

LLVPYREVLNSDAVTFGGSGHVN

Bcaldolyticus   SIFSFIRRGKKEG-DVLVIVCNFTNQAYDDYKVSVP-

LLAPYREVLNSDAAEFGGSGHVN

Bsubtilis   SIFSFIRYGQKHG-EALVIICNFTPVVYHQYDVGVP-

FFTQYIEVLNSDSETYGGSGQIN

mycobacterium   NVLSFMRYGSDG--SVLACVFNFAGAEHRDYRLGLP-

RAGRWREVLNTDATIYHGSGIGN

Streptomyces   NVFAFLRFARDG--

SPLLCVSNFSPVVRHGYRIGVPQEVGQWREVLNTDLEPYGGSGVHH

E.   SVLIFVRRDKEG--NEIIVASNFTPVPRHDYRFGIN-

QPGKWREILNTDSMHYHGSNAGN

H.influenzae   SVLAFERRSSNG--ERIIVVSNFTPVPRHNYRIGVN-VAGKYEEILNTDSMYYEGSNVGN

Agro.tume   SVFAWLRTAPGE--KPVAVICNLTPVYRENYYVPLG-VAGRWREILNTDAEIYGGSGKGN

Aquifex_a.   SVISFLRKDKSGK-EIILVVCNFTPVPRYDYRVGVP-KGGYWREIMNTDAKEYWGSGMGN

Synechococcus   SIVSFIRRAHESD-RFLVVVCNFTPQPHAHYRIGVP-VAGFYREIFNSDARSYGGSNMGN

Synechocystis   SVVSFIRRAKNSA-EFVVTICNFTPQPHSHYRVGVP-VPGFYTELFNSDARQYGGSNMGN

Butyrivibrio   STYSFYRRASNGK-DNILFVLNMTPMERKGFKVGVP-FDGTYTKILDSAKECYGGSGSSV

CHLAMYDIA   GVVAYLRFADADAKKALLCVHHFGVGYFPHYLLPIL-PLESCDLLMNTDDTRFGGSGKGF

      . :*   : ::   :.:   :::: .:**.

Bstearothermophilus   GKR-LSAFNEPFHGK------------P--YHVRMTIPPFGISILRPVQKRGERKRNEK

Bcaldolyticus   GKR-LPAFSEPFHGK------------P--YHVRMTIPPFGISILRPVQKRGERKQNEE

Bsubtilis   KKP-LSAKKGALHHK------------P--CYITMTIPPYGISILRAVKKRGEIKR---

mycobacterium   LGG-VDATDDPWHGR------------P--ASAVLVLPPTSALWLTPA-----------

Streptomyces   ARA-LRPEPVPAQGR------------A--VSLRMTLPPMATVWLRP------------

E.   GGT-VHSDEIASHGR------------Q--HSLSLTLPPLATIWLVREAE--------

H.influenzae   FGC-VASEQIESHGR------------E--NSISVSIPPLATVYLRLKTK---------

Agro.tume   GG----RVQAVDAGG------------E--IGAMLVLPPLATIMLEPEN----------

Aquifex_a.   LGG-KEADKIPWHGR------------K--FSLSLTLPPLSVIYLKHEG----------

```
     Synechococcus      LGG-KWTDEWSCHNR------------P--
     YSLDLCLPPLTTLVLELASGPES----LS
     Synechocystis      LGG-KWTEEWSFHEQ------------P--
     YSLDLCLPPLSVLVLKLSQNAEENTVPAE
5    Butyrivibrio       PDK-IKAVKGLCDYK------------D--
     YSIEFDLPPYGAEVFVFQTKKKTKN-----
     CHLAMYDIA
     REPEILTPEIARQEREAAGLIEADDESGPDCWGLDIELPPSATLIFSVTLQ--------
                              . :**   :

10

     Bstearothermophilus    EMHRHVIGRRARKSASLADDKHR------------
     Bcaldolyticus          EVHRHVIGRRARKPASLADEKHRETSRAVWGEVPDH
     Bsubtilis          ------------------------------
     mycobacterium      ------------------------------
15   Streptomyces       ------------------------------
     E.                 ------------------------------
     H.influenzae       ------------------------------
     Agro.tume          ------------------------------
     Aquifex_a.         ------------------------------
20   Synechococcus          EAANSPL---------------------------
     Synechocystis          EASNIA----------------------------
     Butyrivibrio       ------------------------------
     CHLAMYDIA          ------------------------------


25
```

Nucleotide sequence of *T. thermophilus* AMase

    1 ATGGAGCTTC CCCGCGCTTT CGGTCTGCTT CTCCACCCCA CGAGCCTCCC
CGGCCCCTAC
   61 GGCGTCGGCG TCCTGGGCCA GGAGGCCCGG GACTTCCTCC
GCTTCCTCAA GGAGGCAGGG
  121 GGGCGGTACT GGCAGGTCCT CCCCTTGGGC CCCACGGGCT
ATGGCGACTC CCCCTACCAG
  181 TCCTTCAGCG CCTTCGCCGG AAACCCCTAC CTCATAGACC TGAGGCCCCT
CGCGGAAAGG
  241 GGCTACGTGC GCCTGGAGGA CCCCGGCTTC CCCCAAGGCC
GGGTGGACTA CGGCCTCCTC
  301 TACGCCTGGA AGTGGCCCGC CCTGAAGGAG GCCTTCCGGG
GCTTCAAGGA AAAGGCCTCC
  361 CCGGAGGAGC GGGAGGCCTT CGCCGCCTTC CGGGAGAGGG
AGGCCTGGTG GCTCGAGGAC
  421 TACGCCCTCT TCATGGCCCT GAAGGGGGCG CACGGGGGGC
TTCCCTGGAA CCGGTGGCCC
  481 CTTCCCCTGC GGAAGCGGGA AGAGAAGGCC CTTAGGGAGG
CGAAAAGCGC CTTGGCCGAG
  541 GAGGTGGCCT TCCACGCCTT CACCCAGTGG CTCTTCTTCC
GCCAGTGGGG GGCCTTGAAG
  601 GCGGAGGCCG AGGCGTTGGG CATCCGGATC ATCGGGGACA
TGCCCATCTT CGTGGCCGAG
  661 GACTCCGCCG AGGTCTGGGC CCACCCCGAG TGGTTTCACC
TGGACGAGGA GGGCCGCCCC
  721 ACGGTGGTGG CGGGGGTGCC CCCCGACTAC TTCTCGGAGA
CGGGCCAGCG CTGGGGCAAC
  781 CCCCTTTACC GCTGGGACGT TTTGGAGCGG GAGGGGTTCT
CCTTCTGGAT CCGCCGTCTG
  841 GAGAAGGCCC TGGAGCTCTT CCACCTGGTG CGCATAGACC
ACTTCCGCGG CTTTGAGGCC
  901 TACTGGGAGA TCCCCGCAAG CTGCCCCACG GCGGTGGAGG
GGCGCTGGGT CAAGGCCCCG
  961 GGGGAGAAGC TCTTCCAGAA GATCCAGGAG GTCTTCGGCG
AGGTCCCCGT CCTCGCCGAG
 1021 GACCTGGGGG TCATCACCCC CGAGGTGGAG GCCCTGCGCG
ACCGCTTCGG CCTTCCCGGG
 1081 ATGAAGGTCC TGCAGTTCGC CTTTGACGAC GGGATGGAAA
ACCCCTTCCT CCCCCACAAC
 1141 TACCCTGCCC ACGGCCGGGT GGTGGTCTAC ACCGGCACCC
ACGACAACGA CACCACCCTG
 1201 GGCTGGTACC GCACGGCCAC CCCCCACGAG AAGGCCTTCA
TGGCGCGGTA CCTGGCGGAC
 1261 TGGGGGATCA CCTTCCGGGA AGAGGAGGAG GTGCCCTGGG
CCCTGATGCA CCTGGGGATG
 1321 AAGTCCGTGG CCCGGCTCGC CGTCTACCCG GTGCAGGACG
TCCTGGCCCT GGGCAGCGAG
 1381 GCCCGGATGA ACTACCCGGG AAGGCCCTCG GGGAACTGGG
CCTGGCGGCT CCTCCCGGGG

1441 GAGCTTTCCC CGGAGCACGG GGCGAGGCTT AGGGCCATGG
CCGAGGCCAC GGAACGGCTC
1501 TAG

Amino acid sequence of *T. thermophilus* AMase

   1 MELPRAFGLL LHPTSLPGPY GVGVLGQEAR DFLRFLKEAG GRYWQVLPLG
PTGYGDSPYQ
  61 SFSAFAGNPY LIDLRPLAER GYVRLEDPGF PQGRVDYGLL YAWKWPALKE
AFRGFKEKAS
 121 PEEREAFAAF REREAWWLED YALFMALKGA HGGLPWNRWP
LPLRKREEKA LREAKSALAE
 181 EVAFHAFTQW LFFRQWGALK AEAEALGIRI IGDMPIFVAE DSAEVWAHPE
WFHLDEEGRP
 241 TVVAGVPPDY FSETGQRWGN PLYRWDVLER EGFSFWIRRL EKALELFHLV
RIDHFRGFEA
 301 YWEIPASCPT AVEGRWVKAP GEKLFQKIQE VFGEVPVLAE DLGVITPEVE
ALRDRFGLPG
 361 MKVLQFAFDD GMENPFLPHN YPAHGRVVVY TGTHDNDTTL
GWYRTATPHE KAFMARYLAD
 421 WGITFREEEE VPWALMHLGM KSVARLAVYP VQDVLALGSE
ARMNYPGRPS GNWAWRLLPG
 481 ELSPEHGARL RAMAEATERL

Nucleotide sequence of *A. aeolicus* MTase

```
   1 ATGAGATTGG CAGGTATTTT ACTTCACGTA ACTTCACTTC CCTCTCCTTA
     CGGGATAGGG
  61 GATCTCGGAA AAGAAGCCTA CAGGTTTCTG GACTTCTTAA AGGAGTGCGG
     TTTTAGCCTT
 121 TGGCAGGTTC TACCTCTGAA CCCCACTTCA CTTGAGGCGG GAAACTCACC
     CTACAGTTCA
 181 AACTCCCTCT TCGCGGGCAA TTACGTACTA ATAGACCCTG AAGAATTATT
     GGAGGAGGAC
 241 TTAATAAAAG AAAGGGACTT AAAAAGATTT CCCTTGGGTG AAGCCCTTTA
     CGAAGTCGTG
 301 TACGAGTATA AAAAGAGTT GCTCGAAAAA GCCTTTAAAA ATTTCAGGAG
     ATTTGAACTG
 361 CTTGAAGATT TTCTGAAGGA ACACTCTTAC TGGCTCAGAG ATTACGCACT
     TTACATGGCT
 421 ATAAAAGAAG AAGAGGGAAA GGAGTGGTAT GAATGGGATG
     AAGAATTGAA GAGGAGAGAA
 481 AAAGAGGCTT TAAAAAGGGT GTTAAATAAG TTAAAGGGGA GGTTTTACTT
     CCACGTATTC
 541 GTCCAGTTTG TTTTCTTCAA GCAGTGGGAA AAACTGAGAA GATACGCAAG
     GGAAAGGGGG
 601 ATAAGCATAG TTGGAGATCT TCCAATGTAC CCCTCGTACT CAAGTGCGGA
     CGTGTGGACA
 661 AATCCTGAAC TTTTTAAACT GGACGGAGAT TTAAAACCCC TTTTTGTAGC
     GGGTGTTCCT
 721 CCTGATTTTT TCAGTAAAAC GGGACAGCTG TGGGGAAATC CCGTTTACAA
     CTGGAAGAA
 781 CACGAAAAGG AAGGCTTCAG ATGGTGGATA AGGAGAGTTC
     ATCACAACTT AAAACTCTTT
 841 GACTTTTTAA GACTTGACCA CTTCAGGGGA TTTGAGGCGT ACTGGGAGGT
     TCCTTACGGT
 901 GAAGAAACTG CGGTAAACGG AAGGTGGGTA AAGGCTCCCG
     GAAAGACACT ATTTAAAAAA
 961 CTCTTATCAT ACTTCCCGAA GAACCCATTC ATAGCGGAGG ACTTAGGTTT
     TATAACGGAC
1021 GAAGTGAGGT ACTTGAGGGA AACTTTTAAA ATCCCGGGAA
     GCAGAGTTAT TGAGTTTGCC
1081 TTCTACGATA AGGAAAGTGA GCACCTTCCC CACAACGTTG
     AAGAGAACAA CGTTTACTAC
1141 ACTTCAACTC ATGACCTTCC TCCGATAAGA GGATGGTTTG AGAATTTAGG
     AGAAGAATCA
1201 AGAAACGAT TATTTGAATA CTTGGGAAGG GAGATTAAAG
     AGGAAAAAGT TAACGAGGAG
1261 CTTATAAGAC TCGTTTTAAT CTCAAGGGCG AAGTTCGCAA TAATCCAGAT
     GCAGGACTTA
1321 CTCAATCTCG GCAATGAAGC GAGGATGAAT TACCCCGGAA
     GACCTTTCGG AAATTGGAGG
1381 TGGAGAATAA AGGAAGATTA CACACAAAAG AAGGAATTTA
     TTAAAAAACT CCTCGGAATT
1441 TACGGAAGAG AAGTTTAA
```

Amino acid sequence of *A. aeolicus* MTase

      1 MRLAGILLHV TSLPSPYGIG DLGKEAYRFL DFLKECGFSL WQVLPLNPTS
5 LEAGNSPYSS
     61 NSLFAGNYVL IDPEELLEED LIKERDLKRF PLGEALYEVV YEYKKELLEK
AFKNFRRFEL
    121 LEDFLKEHSY WLRDYALYMA IKEEEGKEWY EWDEELKRRE
KEALKRVLNK LKGRFYFHVF
10   181 VQFVFFKQWE KLRRYARERG ISIVGDLPMY PSYSSADVWT NPELFKLDGD
LKPLFVAGVP
    241 PDFFSKTGQL WGNPVYNWEE HEKEGFRWWI RRVHHNLKLF
DFLRLDHFRG FEAYWEVPYG
    301 EETAVNGRWV KAPGKTLFKK LLSYFPKNPF IAEDLGFITD EVRYLRETFK
15 IPGSRVIEFA
    361 FYDKESEHLP HNVEENNVYY TSTHDLPPIR GWFENLGEES RKRLFEYLGR
EIKEEKVNEE
    421 LIRLVLISRA KFAIIQMQDL LNLGNEARMN YPGRPFGNWR WRIKEDYTQK
KEFIKKLLGI
20   481 YGREV

Nucleotide sequence of *A. aeolicus* BE

```
  1 ATGAAGAAGT TCAGTCTCAT CAGTGATTAC GACGTTTACC TCTTTAAGGA
    GGGAACGCAC
 61 ACGAGACTTT ACGATAAACT TGGCTCCCAC GTTATAGAAC TAAACGGGAA
    AAGGTATACC
121 TTCTTTGCGG TTTGGGCACC CCACGCGGAT TACGTATCAC TTATAGGCGA
    TTTTAACGAA
181 TGGGATAAAG GTTCTACTCC CATGGTAAAG AGGGAGGACG
    GCTCCGGAAT ATGGGAGGTT
241 TTACTTGAAG GAGACCTGAC TGGTTCAAAG TACAAGTACT TTATAAAGAA
    CGGGAATTAC
301 GAAGTTGATA AGTCCGATCC CTTCGCATTT TTCTGTGAGC AACCCCCCGG
    AAACGCTTCC
361 GTAGTGTGGA AGCTCAATTA CAGGTGGAAC GACTCCGAAT
    ACATGAAAAA GAGGAAAGA
421 GTAAACTCAC ACGACTCGCC TATATCCATA TACGAAGTTC ACGTGGGTTC
    TTGGAGGAGA
481 GTTCCAGAAG AGGGAAACAG ATTTTTGAGC TATAGGGAAC TTGCCGAATA
    CCTCCCATAC
541 TACGTAAAAG AGATGGGATT TACTCACGTT GAGTTCTTAC CCGTTATGGA
    ACATCCCTTT
601 TACGGCTCTT GGGGCTACCA GATAACGGGC TACTTCGCTC CGACTTCCAG
    ATACGGAACT
661 CCTCAGGACT TTATGTACTT AATAGACAAA CTTCATCAAG AAGGGATAGG
    TGTGATACTA
721 GACTGGGTTC CCTCTCACTT TCCCACCGAT GCCCACGGGC TCGCATACTT
    TGACGGGACT
781 CACCTTTACG AGTACGAGGA CTGGAGAAAG AGGTGGCATC
    CCGACTGGAA CAGCTTTGTT
841 TTTGATTACG GAAAACCGGA AGTTCGCTCC TTTCTCCTGA GTTCTGCCCA
    CTTCTGGCTC
901 GACAAGTACC ACGCAGACGG TCTCAGAGTG GATGCAGTTG
    CTTCAATGCT TTACCTAGAT
961 TACTCTAGGA AAGAATGGGT TCCAAACATA TACGGAGGGA
    AAGAAAACCT CGAGGCTATA
1021 GAATTCCTCA GGAAGTTTAA CGAAAGCGTT TACAGAAATT TTCCAGACGT
    CCAGACAATA
1081 GCGGAGGAAT CAACAGCCTG GCCTATGGTG TCCAGACCTA
    CATACGTGGG GGGACTGGGA
1141 TTTGGAATGA AGTGGAATAT GGGTTGGATG AACGACACAC TCTTTTACTT
    TTCAAAGGAT
1201 CCCATCTACA GGAAGTACCA CCATGAAGTC CTCACTTTCA GTATATGGTA
    CGCTTTTTCC
1261 GAGAACTTCG TCCTTCCACT ATCCCACGAT GAAGTTGTTC
    ACGGAAAGGG TTCTCTGATA
1321 GGGAAGATGC CAGGAGATTA CTGGCAGAAG TTTGCAAACC
    TTAGAGCCCT TTTCGGATAC
1381 ATGTGGGCAC ACCCAGGGAA AAAACTCCTC TTTATGGGGG
    GAGAGTTCGG ACAGTTTAAG
```

1441 GAATGGGATC ACGAAACGAG TCTCGACTGG CACCTCTTGG
AATACCCTTC TCACAGAGGT
1501 ATTCAGAGAT TAGTTAAGGA CTTAAACGAA GTTTACAGGA
GGGAAAAGGC TTTGCACGAA
1561 ACGGATTTTT CACCTGAGGG CTTTGAGTGG GTAGACTTCC
ACGACTGGGA AAAGAGCGTT
1621 ATATCCTTCT TGAGAAAGGA CAAAAGCGGT AAGGAAATTA TACTCGTAGT
TTGCAACTTC
1681 ACACCCGTTC CGAGATACGA TTACAGGGTA GGTGTACCGA
AAGGCGGATA CTGGAGGGAG
1741 ATAATGAATA CCGATGCAAA GGAGTACTGG GGCTCCGGAA
TGGGAAATCT GGGTGGAAAA
1801 GAGGCTGATA AAATCCCGTG GCACGGAAGA AAATTCTCAC TTTCACTTAC
CCTGCCTCCC
1861 CTTTCCGTGA TCTATTTAAA GCACGAAGGA TGA

Amino acid sequence of *A. aeolicus* BE

1 MKKFSLISDY DVYLFKEGTH TRLYDKLGSH VIELNGKRYT FFAVWAPHAD
YVSLIGDFNE
61 WDKGSTPMVK REDGSGIWEV LLEGDLTGSK YKYFIKNGNY EVDKSDPFAF
FCEQPPGNAS
121 VVWKLNYRWN DSEYMKKRKR VNSHDSPISI YEVHVGSWRR
VPEEGNRFLS YRELAEYLPY
181 YVKEMGFTHV EFLPVMEHPF YGSWGYQITG YFAPTSRYGT PQDFMYLIDK
LHQEGIGVIL
241 DWVPSHFPTD AHGLAYFDGT HLYEYEDWRK RWHPDWNSFV
FDYGKPEVRS FLLSSAHFWL
301 DKYHADGLRV DAVASMLYLD YSRKEWVPNI YGGKENLEAI EFLRKFNESV
YRNFPDVQTI
361 AEESTAWPMV SRPTYVGGLG FGMKWNMGWM NDTLFYFSKD
PIYRKYHHEV LTFSIWYAFS
421 ENFVLPLSHD EVVHGKGSLI GKMPGDYWQK FANLRALFGY
MWAHPGKKLL FMGGEFGQFK
481 EWDHETSLDW HLLEYPSHRG IQRLVKDLNE VYRREKALHE
TDFSPEGFEW VDFHDWEKSV
541 ISFLRKDKSG KEIILVVCNF TPVPRYDYRV GVPKGGYWRE IMNTDAKEYW
GSGMGNLGGK
601 EADKIPWHGR KFSLSLTLPP LSVIYLKHEG

Claims

1.   An isolated or recombinant nucleic acid derived from a nucleic acid encoding a
polypeptide essentially having alpha-glucanotransferase activity but having
5   essentially no hydrolysing activity, said isolated or recombinant nucleic acid encoding
a polypeptide with hydrolytic activity.

2.   A nucleic acid according to claim 1 wherein said transferase comprises
amylomaltase or branching enzyme.
10

3.   A nucleic acid according to claim 1 or 2 wherein said transferase comprises a
thermostable transferase.

4.   A nucleic acid according to anyone of claims 1 to 3 wherein said transferase is
15   derived from a thermophilic micro-organism.

5.   A nucleic acid according to claim 4 wherein said micro-organism comprises
Thermus thermophilus, *Thermus aquaticus* or *Aquifex aeolicus*.

.

20

6.   A nucleic acid according to anyone of claims 1 to 5 wherein said transferase is
known under EC number 2.4.1.25 or 2.4.1.18

7.   A nucleic acid according to anyone of claims 1 to 6 provided with a mutation
25   leading to an alteration or loss of a codon originally encoding a hydrophobic amino
acid located in or around a acceptor, a donor or a catalytic site extending from a TIM
barrel structure of said transferase.

8.   A nucleic acid according to claim 7 wherein said codon originally encoding a
30   hydrophobic amino acid is altered into a codon encoding an amino acid which is
substantially less hydrophobic.

9.   A nucleic acid according to claim 7 or 8 wherein said hydrophobic amino acid
comprises phenylalanine, tryptophan or tyrosine.

10.   A nucleic acid according to anyone of claims 7 to 9 wherein said hydrophobic amino acid is located at or around an amino acid position essentially corresponding to amino acid position 54, 251, 258 or 366 of amylomaltase of *Thermus thermophilus*

5   HB8.

11.   A vector comprising a nucleic acid according to anyone of claims 1 to 10.

12.   A host cell comprising a vector according to claim 11 or a nucleic acid

10   according to anyone of claims 1 to 10.

13.   A method for providing a polypeptide or fragment thereof essentially having alpha glucanotransferase acitivity but having essentially no hydrolysing activity with hydrolysing activity said method comprising providing a nucleic acid encoding such a

15   transferase with a mutation leading to an alteration or loss of a codon originally encoding a hydrophobic amino acid located in or around a acceptor, a donor or a catalytic site extending from a (alphalpha/betaeta)8 barrel structure of said transferase.

20 14.   A polypeptide, or an enzymatically functional fragment thereof encoded by a nucleic acid according to anyone of claims 1 to 10 or obtainable by a method according to claim 13.

15.   Use a polypeptide or fragment according to claim 14 in reducing

25   retrogradation of starch.

16.   Use according to claim 15 in reducing retrogradation of amylopectine.

17.   Use according to claim 16 in reducing long-term retrogradation of

30   amylpectine.

18.   Use a polypeptide or fragment according to claim 14 in hydrolysing starch.

19.   A method for reducing retrogradation of starch comprising treating said starch

with a polypeptide or fragment according to claim 14.

20.   A method for hydrolysing starch comprising treating said starch with a polypeptide or fragment according to claim 14.

21.   A bakery ingredient comprising a polypeptide or fragment according to claim 14.

22.   A bakery product such as bread comprising a polypeptide or fragment according to claim 14.
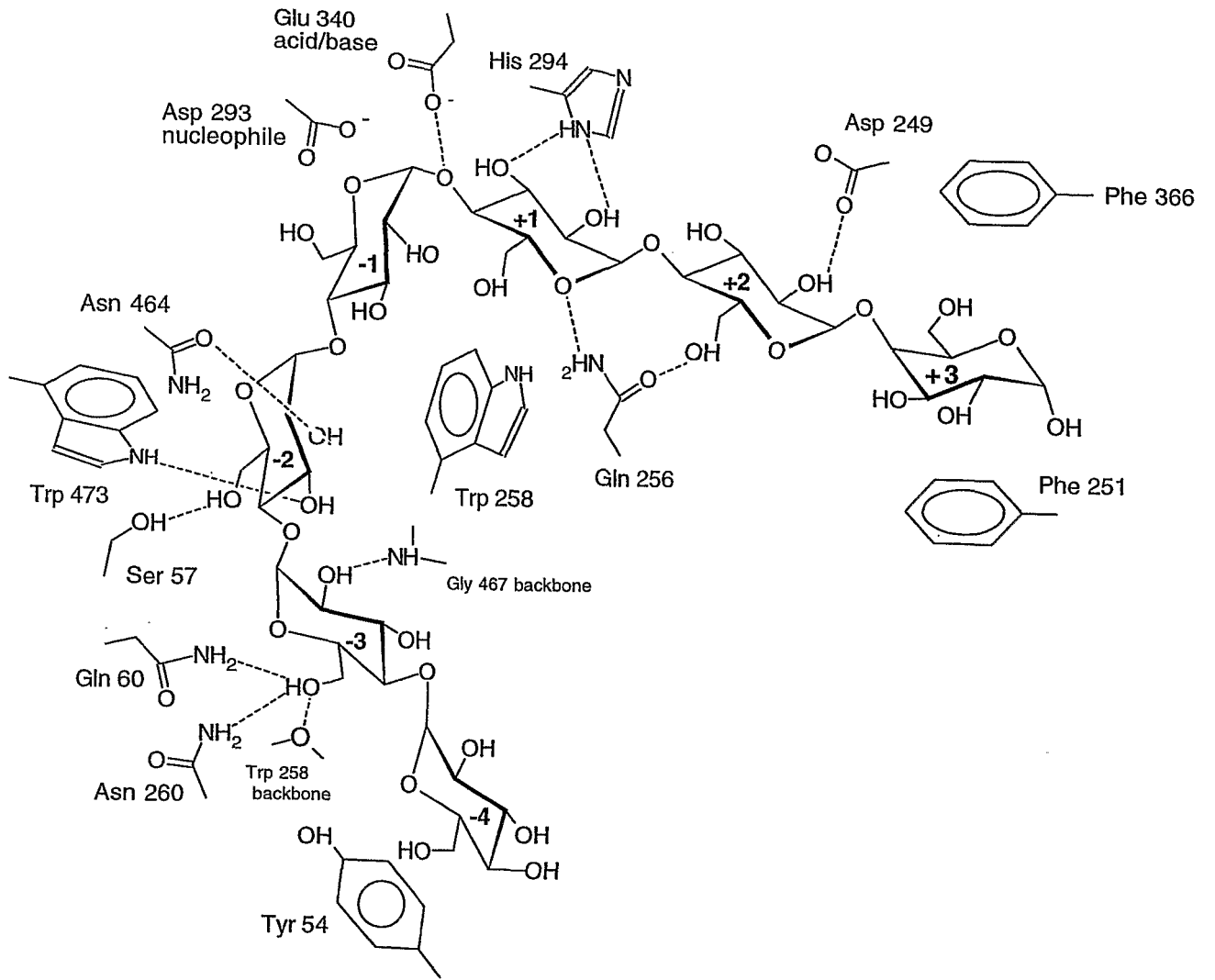
FIGURE 1

FIGURE 2

FIGURE 3

```
glgB Aqu      1    MKKFSLISD YDVY------ --LFKEGTHT RLYDKLGSHV IELNGKRYTF
isoamyla      1    AINSMSLGAS YDAQQANITF RVYSSQATRI VLYLYSAGYG VQESATYTLS
                   ..  ** .  **                   ..*.  **     ..    .. ..
glgB Aqu                                              ss sssss          ssssss
isoamyla              sss ss     sssss ssss    ss sssss          ssssss


glgB Aqu     42    FAVWAPHADY VSLIGDFNEW DKGSTPMVKR EDGSGIWEVL LEGDLTGSKY
isoamyla     51    PAGSGVWAVT VPVSSIKAAG ITGAVYYGYR AWGPNWPYAS NWGKGSQAGF
                   *   .  *  *...        *.     *   *.           *    ...
glgB Aqu           ss   sss      hhhhhh     ssssss ss
isoamyla           ss   ssssss ssshhhhhh    ssssss ss


glgB Aqu     92    KYFIK-NGNY EVDKS---DP FAFFCEQPPG NAS----VVW KLNYRWNDSE
isoamyla    101    VSDVDANGDR FNPNKLLLDP YAQEVSQDPL NPSNQNGNVF ASGASYRTTD
                   .  **.       .    ** .*   * *  *.*    *        ..
glgB Aqu                                   sss             s ss
isoamyla                                   sss           hhs ss


glgB Aqu    134    YMKKRKRVNS HDSPIS---- ---------- IYEVHV-GSW RRVPEEGNRF
isoamyla    151    SGIYAPKGVV LVPSTQSTGT KPTRAQKDDV IYEVHVRGFT EQDTSIPAQY
                                    .       ..        ****** *       .      ..
glgB Aqu           hhh    ssss                        ssss
isoamyla           hhh    ssss                        ssss hhh
                                                      b1


glgB Aqu    169    LSYRELAEYL PYYVKEMGFT HVEFLPVMEH PFYG------ ------SWGY
isoamyla    201    RGTYYGAGLK ASYLASLGVT AVEFLPVQET QNDANDVVPN SDANQNYWGY
                     *         . *. .* * ****** *        .          ***
glgB Aqu            hhhhh        hhhhhh  s sssss sss
isoamyla           hhhhh        hhhhhh  s sssss sss       hhh
                                a1              b2


glgB Aqu    207    QITGYFAPTS RY-------G TPQDFMYLID KLHQEGIGVI LDWVPSH---
isoamyla    251    MTENYFSPDR RYAYNKAAGG PTAEFQAMVQ AFHNAGIKVY MDVVYNHTAE
                     **.*      **         .*   ... .* ** *  .* * .*
glgB Aqu              sss                hhhhhhhhhh hhhhh   sss sss
isoamyla              sss              h hhhhhhhhhh hhhhh   sss sss
                                                              a2          b3


glgB Aqu    247    -------FPT DAHGLAY--F DGTHLYEYED WRKRWHPDWN S-FVFDYGKP
isoamyla    301    GGTWTSSDPT TATIYSWRGL DNATYYELTS GNQYFYDNTG IGANFNTYNT
                          **   *     .  . * .  **            .    *.  .
glgB Aqu                     sss         hhhssssss       sss       sss  h
isoamyla                     sss  hh     hhsssss         sss       sss  h


glgB Aqu    287    EVRSFLLSSA HFWLDKYHAD GLRVDAVASM LY--LDYSRK EWVPNIYGGK
isoamyla    351    VAQNLIVDSL AYWANTMGVD GFRFDLASVL GNSCLNGAYT ASAPNCPNGG
                   .....  *   .*  .    * *.* *   ..         *.  .    ** *
glgB Aqu           hhhhhhhhhh hhhhhh   s sssss   hh
isoamyla           hhhhhhhhhh hhhhhh   s sssss   hh  sss
                            a3              b4
```

FIGURE 3, Contd.


```
glgB Aqu    335  ENLEAIEFLR KFN----ESV YR---NFPDV QTIAEESTAW PMVSRPTYVG
isoamyla    401  YNFDAADSNV AINRILREFT VRPAAGGSGL DLFAEPWAIG G--NSYQLGG
                 *..*  .      .*     *    *      . . .**  .      .       *
glgB Aqu                                                  sssss
isoamyla             sss        hhhhh                     sssss
                                a4                          b5


glgB Aqu    378  GLGFGMKWNM GWMNDTL--- --FYFSKDPI YRKYHHEVLT FSIWYAFS--
isoamyla    449  FPQGWSEWNG LFR-DSLRQA QNELGSMTIY VTQDANDFSG SSNLFQSSGR
                          **         *.*        *.          ..      * . *
glgB Aqu              ssss   hh                   hhhhhhhh    hhhhhh
isoamyla              ssss   hhh hhhhhh hh        hhhhhhhh    hhhhhh
                      b6                                        a6


glgB Aqu    421  ---ENFVLPL SHD-EVVHGK GSLIGKMPGD YWQKFANLR- ALFGYMWAHP
isoamyla    498  SPWNSINFID VHDGMTLKDV YSCNGANNSQ AWPYGPSDGG TSTNYSWDQG
                 ... .      **    .     *   *   ..   *     ..      * * .
glgB Aqu         ssss       hhh                                       s
isoamyla         ssss       hhh                                       s
                 b7


glgB Aqu    466  GKKLLFMGGE FGQFKEWDHE TSLDWHLLEY PSHRG---IQ RLVKDLNEVY
isoamyla    548  MS--AGTGAA VDQRRAARTG MAFEMLSAGT PLMQGGDEYL RTLQCNNNAY
                 *.        *  .          ...       * .*     * .   *. *
glgB Aqu         ss   ssss h hhhhhhhhhh hhhhhh    s sssss
isoamyla         ss   sss h hhhhhhhhhh hhhhhh     s sssss
                                a7         b8


glgB Aqu    513  RREKALHETD FSPEGFEWVD FHDWEKSVIS FLRKDKSGKE IILVVCNFTP
isoamyla    596  NLDSSANWLT YS-WTTDQSN FYTFAQRLIA FRKAHPALRP SSWYSG--SQ
                 . . .      .*       . .  *     .*. *  .    . .          .
glgB Aqu                    hhhhh hhhhhhhhhh hhhh                      s
isoamyla                    hhhhh hhhhhhhhhh hhhh                      s
                                   a8


glgB Aqu    563  VPRYDYRVGV PKGGYWREIM N-TDAKEYWG SGMGNLGGKE ADKIPWHGRK
isoamyla    643  LTWYQPSGAV ADSNYWNNTS NYAIAYAING PSLGDSNSIY VAYNGWSSSV
                 . .*.     .*  . .** .   * . *    * ...*  .         * .
glgB Aqu         sssss    ss s hhhhh       ssssss hhh      ss sss
isoamyla         sssss    ss s hhhhh       sssssss hhh     sss sssss    s


glgB Aqu    612  FSLSLTLPPL SVIYLKHEG
isoamyla    693  TFTLPAPPSG TQWYRVTDTC DWNDGASTFV APGSETLIGG AGTTYGQCGQ
                  . *.  . *  .
glgB Aqu
isoamyla         ssss      sssssssss                   sss       sssss


glgB Aqu
isoamyla    743  SLLLLISK


glgB Aqu
isoamyla         sssssssss
```

FIGURE 4