Engineering of β-Glycosidases from Hyperthermophilic Archaea

Thijs Kaper

Promotor: **Prof. Dr. W.M. de Vos**
Hoogleraar in de Microbiologie
Wageningen Universiteit

Co-promotor: **Dr. J. van der Oost**
Universitair docent bij de leerstoelgroep Microbiologie
Wageningen Universiteit

Leden van de **Prof. Dr. L. Dijkhuizen**
promotie Rijksuniversiteit Groningen
commissie:

Dr. M. Moracci
Istituto di Biochimica delle Proteine ed Enzimologia, Consiglio Nazionale
delle Ricerche, Napoli, Italia

Prof. Dr. R.J. Siezen
Katholieke Universiteit Nijmegen

Prof. Dr. A.G. Voragen
Wageningen Universiteit

Engineering of β-Glycosidases from Hyperthermophilic Archaea

Thijs Kaper

Proefschrift

ter verkrijging van de graad van doctor

op gezag van de rector magnificus

van Wageningen Universiteit,

prof.dr.ir. L. Speelman,

in het openbaar te verdedigen op

vrijdag 19 oktober 2001

des namiddags te vier uur in de Aula.

16 2 9 3 0 2

voor mijn ouders

P ropositions

1. Proteins from hyperthermophilic origin can be shuffled to thermostable hybrid enzymes.
    This thesis.

2. The inhibition by glucose of *P. furiosus* β-glucosidase CelB and *S. solfataricus* β-glycosidase LacS is comparable.
    This thesis.
    Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.

3. A complete genome sequence should be regarded as a momentarily image of the genetic information of an organism.

4. Protein inclusion bodies can be regarded as an intermediate step in *in vivo* protein folding.
    Carrió, M.M., and Villaverde, A. (2001) *FEBS Lett 489*, 29-33

5. Scientific information on the internet will be more widely accessible when authors make their articles available on personal websites.
    http://www.nature.com/nature/debates/e-access/

6. Science would be served by presentation of failed experiments alongside successful ones, instead of mentioning them in informal discussions only ('tears in beers').
    C. Schmidt-Dannert and F.H. Arnold (1999) *Trends Biotechnol 17*, 135-6

7. By selling monumental buildings, Wageningen University deprives itself of charm, charisma and historical awareness.

8. Images in cooking books have too often nothing to do with reality.

Propositions belonging to the thesis: '*Engineering of β-glycosidases from hyperthermophilic Archaea*'
Thijs Kaper
Wageningen, 19 oktober 2001

S tellingen

1. Eiwitten van hyperthermofiele oorsprong kunnen 'geshuffled' worden tot thermostabiele hybride enzymen.
Dit proefschrift.

2. De glucose-inhibitie van *P. furiosus* β-glucosidase CelB en *S. solfataricus* β-glycosidase LacS is vergelijkbaar.
Dit proefschrift.
Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.

3. Een complete genoom sequentie moet gezien worden als een momentopname van de genetische informatie van een organisme.

4. Eiwit inclusion bodies kunnen beschouwd worden als een intermediar in *in vivo* eiwitvouwing.
Carrió, M.M., and Villaverde, A. (2001) *FEBS Lett 489*, 29-33

5. Wetenschappelijke informatie op het internet zal breder toegankelijk zijn wanneer auteurs hun artikelen beschikbaar stellen op persoonlijke websites.
http://www.nature.com/nature/debates/e-access/

6. De wetenschap zou gediend zijn bij het presenteren van mislukte experimenten tegelijkertijd met succesvolle, in plaats van deze slechts tijdens informele discussies te noemen ('tears in beers').
C. Schmidt-Dannert and F.H. Arnold (1999) *Trends Biotechnol 17*, 135-6

7. Met de verkoop van monumentale panden ontneemt de Wageningen Universiteit zichzelf charme, uitstraling en historisch besef.

8. Afbeeldingen in een kookboek hebben maar al te vaak niets met de realiteit te maken.

Stellingen behorend bij het proefschrift: *'Engineering of β-glycosidases from hyperthermophilic Archaea'*
Thijs Kaper
Wageningen, 19 oktober 2001

B edankt...

Ja, mijn naam staat dan wel als enige voorop het boekje, maar ik heb het toch echt niet allemaal zelf gedaan. Veel mensen, meer dan ik hier kan noemen, hebben al dan niet bewust bijgedragen aan dit werk. Voorop staat dat de afgelopen 5 jaren een geweldige tijd zijn geweest.

John, jou wil ik in de eerste plaats bedanken voor de begeleiding van de afgelopen 5 jaar: nimmer aflatend enthousiasme, positieve instelling, relativisme en een grote vrijheid om zelf het onderzoek in te vullen. Het is tenslotte allemaal maar een spelletje, niet?

Willem, meer op de achtergrond met het overzicht, bedankt voor de ontspannen begeleiding en de stimulerende discussies.

Jurre, Jeroen, Bert, Hester en Stan, zonder jullie was dit proefschrift een stuk dunner geweest en het werk op het lab een stuk saaier. Bedankt voor jullie inzet.

(Ex-)BacGenners bedankt: Leon, niet alleen in maar vooral buiten het lab, vanaf die eerste Chouffes met Jägermeister tot de petzelbauers, substlate speficificity en de colhoes; Joyce, gezellig om weer samen te werken in Zweden; Ans, voor het strakke koffie regime, de thee en de verse tuinkruiden; Arjen, voor de uurtjes jazz in de collegezaal; Corné voor de introductie in het Brabantse carnavalsleven; Ana, voor de culturele uitjes; Don, thanks for initiating some good habits, like fast food on Friday and the work-outs in the gym; Wilfried, Ineke, Thijs (the real one??), Johan, Servé, Hauke, Pino, Stan, Valerie, Nina, Ken, Yannick, Gaël, voor de gezelligheid tijdens de koffiepauzes, kartmiddagen, tot de verbeelding sprekende lunch discussies, kerstdiners, zeiltochten, Horizon lunches, barbecues, kroegmiddagen, -avonden en –nachten. Verder alle collega's op Microbiologie bedankt voor de goede werksfeer en speciaal Ria, die onmisbaar was voor de cafeïne elke morgen, Nees, voor het wijzen van de weg in bureaucratische en financiële doolhoven, en Wim, voor alle hulp met dwarsliggende rekenapparatuur.

Binnen de Wageningen Universiteit heb ik met veel plezier samengewerkt met collega's aan de overkant van de straat. Willem van Berkel, Carlo van Mierlo, Maurice Franssen en Matthew de Roode, bedankt de prettige samenwerking, overleg, adviezen en gebruik van apparatuur.

T able of contents

A bbreviations

BglB	β-mannosidase BglB from *Pyrococcus horikoshii*
CelB	β-glucosidase CelB from *Pyrococcus furiosus*
LacG	6-phospho-β-galactosidase LacG from *Lactococcus lactis*
LacS	β-glycosidase LacS from *Sulfolobus solfataricus*
Bgly	β-glycosidase from *Thermosphaera aggregans*
BglA	β-glucosidase BglA from *Bacillus polymyxa*
oNp	*ortho*-nitrophenol/yl
pNp	*para*-nitrophenol/yl
oNp-Gal	oNp-β-D-galactopyranoside
oNp-Gal-6-P	oNp-β-D-galactopyranoside-6-phosphate
oNp-Glc	oNp-β-D- glucopyranoside
pNp-Gal	pNp-β-D-galactopyranoside
pNp-Glc	pNp-β-D-glucopyranoside
pNp-Man	pNp-β-D-mannopyranoside
PheImGlc	(5R,6R,7S,8S)-5-(hydroxymethyl)-2-phenyl-5,6,7,8-tetrahydroimidazol[1,2-a]pyridine-6,7,8-triol
IPTG	isopropyl-β-D-thiogalactopyranoside
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
X-Glc	5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside

1

General introduction

The first chapter of this thesis gives background information on the separate parts of its title "Engineering of β-glycosidases from hyperthermophilic Archaea". It starts with a description of hyperthermophilic, or heat-loving organisms. Then, β-glycosidases, the enzymes that have been the topic of the studies described in this thesis, are described. Next, thermostabilizing mechanisms in proteins are discussed. Thereafter, different approaches for enzyme engineering are described, which is followed by possible applications for β-glycosidases in industry. Finally, the outline of the thesis is described.

1.1 Hyperthermophiles

The proteins that have been studied in this thesis originate from hyperthermophilic, or "heat-loving", organisms. This section will cover the natural habitats of these organisms, which are exclusively microorganisms, their biological diversity, the relevance of recently elucidated hyperthermophile genome sequences, as well as the possible biotechnological implications. In addition, the carbohydrate metabolism of the hyperthermophiles *Pyrococcus furiosus* and *Sulfolobus solfataricus* is described in more detail, since the enzymes that have been topic of the studies described in this thesis, play important roles in carbohydrate degradation in these microorganisms.

Hyperthermophilic habitats. Life can be classified by the temperature, at which an organism reaches its optimal growth rate. Hyperthermophiles have been defined as organisms that grow optimally above 80°C. Heat-adapted life has been mainly found in aqueous systems, usually in areas of volcanic activity in both marine and terrestrial ecosystems. Marine ecosystems are characterized by high concentrations of salt and a moderate pH. The solidifying magma in undersea volcanic areas vaporizes water to steam with a temperature of 400°C (*1*). Upon contacting the cold ocean water, the – often sulfur-containing - minerals in the steam precipitate to form chimney-like structures or 'black smokers', of which the porous walls are habitats for microbial life (*2*). Members from the order of Thermococcales, to which *Pyrococcus furiosus* belongs, are abundantly found at these sites (*3*). In addition to these and other microbes, undersea volcanic areas are full of Eukaryal life that has evolved to withstand the high concentrations of sulfur (*4*). Terrestrial habitats for hyperthermophiles include hot springs, mud holes and solfataric fields. Often found in areas of volcanic activity, the habitats are diverse in temperature, pH and chemical composition. In solfataric soils, an oxidized, ferric-iron-containing zone is often found above an anoxic, ferrous sulfide-containing zone (*2*). Furthermore, solfataric soils are characterized by steep temperature and pH gradients. Several Sulfolobales, including *Sulfolobus solfataricus*, have been isolated from hot, acidic pools in these places (Figure 1.1) (*5*).

Hyperthermophile diversity. Presently more than 70 hyperthermophilic organisms have been described, which belong to 32 different genera in 10 orders (*2*). The majority of these orders are Archaea, with the only bacterial orders being the Thermotagales and Aquificales (*2, 6*). The hyperthermophiles branch deeply in a universal phylogenetic tree, which fuels speculations about the hyperthermophilic character of the proposed last universal common ancestor of earthly life (Figure 1.2). However, the matter of hyperthermophilic Archaea being the first living organisms on a young, still hot earth is still under debate (*7, 8*). Hyperthermophilic Archaea are found in all three archaeal kingdoms, the Crenarchaeota, Euryarchaeota and Korarchaeota and show considerable phylogenetic diversity, which is reflected by their diverse morphology, physiology, and biochemistry (*9*).

Hyperthermophilic Euryarcheaota have found to be sulfur-reducing and sulfur-independent heterotrophs, CO_2-reducing methanogens, and sulfate reducing hydrogen oxidizers, while representatives of the Crenarchaeota include autotrophs and heterotrophs, which can be strictly or facultatively anaerobic or microaerophilic (*10*). Whereas the former two kingdoms have several isolated members, the presence of Korarchaeotal rRNA sequences in hydrothermal sites indicates that this kingdom includes hyperthermophiles as well (*11, 12*).

Figure 1.1 Hot acidic mud pool in the Solfatara crater, Pozzuoli (NA), Italy, habitat of *Sulfolobus* (courtesy of Leon Kluskens).

Genomes of hyperthermophilic microorganisms. At present, 12 genome sequences of hyperthermophiles have been listed as completed (www.tigr.org). Genomic sequences of hyperthermophilic microorganisms are of evolutionary and biochemical interest. The genome sequences can be used for establishing a comprehensive view of evolutionary relationships between organisms, in addition to the rRNA based classification of life (*13*). Also, such an analysis could be of help in the description of the last universal common ancestor for all earthly life, which might have had a hyperthermophilic character (*7, 9*). However, the hypothesis that the deeply rooted position of hyperthermophiles is an indication of a slow evolution pace (*14*), seems contradictory to recent analysis of the genome sequences. The genomes of hyperthermophilic *P. furiosus*, *P. horikoshii* and *P. abyssi* revealed a relatively high genome plasticity within one genus (*15-17*). Further analysis of bacterial and archaeal genomes might challenge the current view on evolution with the finding of domain-crossing characteristics in single species (*13, 18, 19*). Whereas a stable core of household genes can be identified, there is a considerable number of gains and losses of so called lifestyle genes, such as carbohydrate degrading pathways (*20*). Evidence has been presented that these genes may be laterally transferred from one organism to another (*21*). Furthermore, analysis of the genomes of hyperthermophiles provides information on the hyperthermophilic way of life, including its metabolic aspects, the molecular basis for protein thermostability, and might result in the identification of genes coding for enzymes of industrial interest (*14*).

Biotechnological implications. Enzymes from hyperthermophiles often have an extreme stability, which allows their application in industrial processes at elevated temperature, high pressure, and in organic solutes. A high process temperature may have a beneficial effect on substrate solubility, viscosity, diffusion and reaction rates (*22*). Furthermore, above 60 °C the risk for reactor contamination is greatly reduced. At present, enzymes from hyperthermophiles find mainly applications in PCR and other research tools in molecular biology as well as the starch and paper processing industry (*6*). Despite their great potential, however, the applications for hyperthermophiles and their enzymes are currently rather limited (*23*). Still, applications of enzymes or products from hyperthermophiles in food and non-food industry are anticipated to result from recently developed initiatives in functional genomics as well as from development of hyperthermophilic expression systems (*24*).

Sugar metabolism in Pyrococcus furiosus and Sulfolobus solfataricus. Pyrococcus furiosus is a strictly anaerobic Euryarchaeon that grows optimally at 100°C and was originally isolated from shallow, thermal marine sediments (*25*). Recently, its complete genome sequence (1.8 Mb) has been determined (*26*). *P. furiosus* is capable of heterotrophic growth on organic acids, proteinaceous substrates, α- and β-linked saccharides and monomeric glucose (*25, 27-29*). Starch, pullulanan and glycogen are degraded by extracellular α-amylase B (*30, 31*) and amylopullulanase (*32, 33*) to malto-oligosaccharides, which are converted to monomeric glucose by an intracellular α-
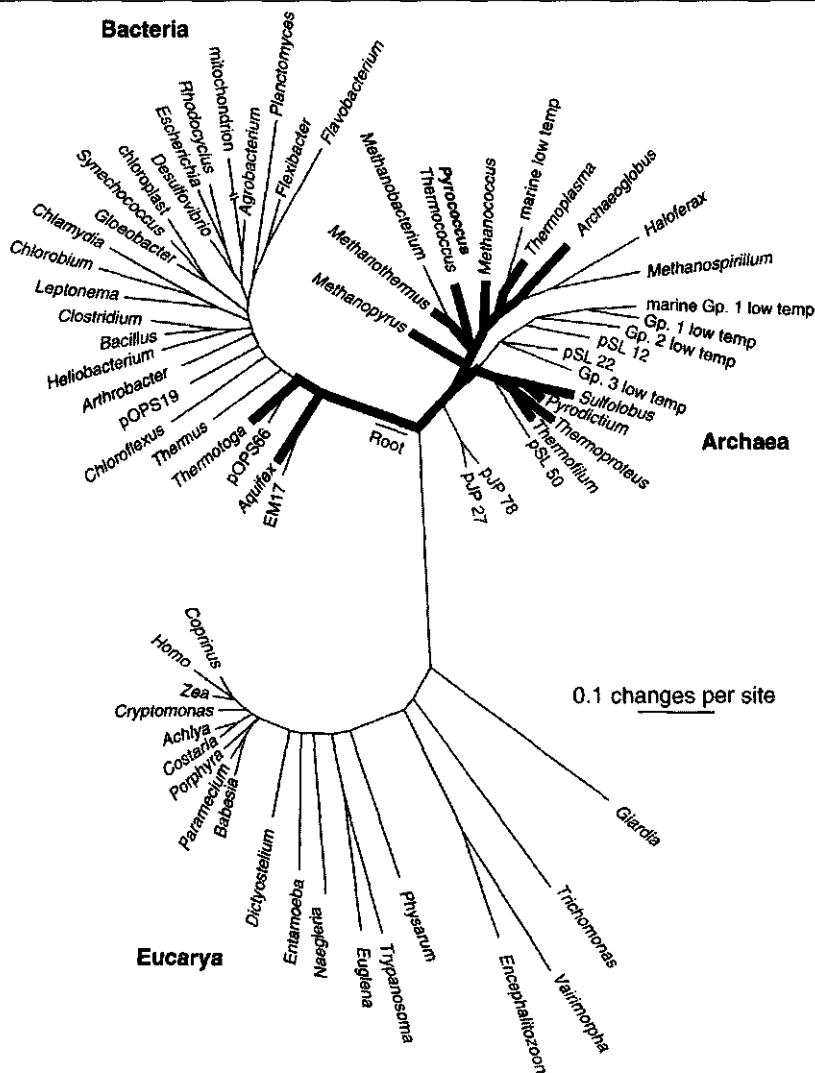
Figure 1.2 Universal phylogenetic tree based on SSU rRNA sequences, modified after (http://cas.bellarmine.edu/tietjen/RootWeb/ a_molecular_view_of_microbial_di.htm). Thick lines represent hyperthermophilic lineages. The scale bar corresponds to 0.1 changes per nucleotide.

glucosidase *(34)*. A homologous gene cluster that codes for enzymes involved in uptake of maltose and trehalose in *Thermococcus literalis* *(35)* has been identified in *P. furiosus* *(36)*. Intracellularglycogen is degraded by cytosolic α-amylase A *(37, 38)*. In addition to α-linked glucosides, *P. furiosus* is able to grow on β-1,4-linked cellobiose *(27)*, β-1,3-linked laminarin *(28)* and mixed linkage barley lichenan *(29)* (Figure 1.3). For their degradation *P. furiosus* produces an extracellular laminarinase LamA *(28)* and an endo-β-1,4-glucanase EglA *(39)*. The β-linked oligo-

saccharides are actively transported into the cell for degradation to glucose by the β-glucosidase CelB (*27, 36, 40*). The *celB* and *lamA* genes are divergently located in a single operon and their expression is induced in the presence of β-linked glucosides (*41*). The β-glucosidase and endo-glucanases have been found to synergistically degrade β-linked glucose polymers to glucose *in vitro* (*28, 42*). In the cell, monomeric glucose enters a modified Embden-Meyerhof glycolysis pathway with novel ADP-dependent kinases (*43-45*) and a single step conversion from glyceraldehyde-3-phosphate to 3-phospho-glycerate by the GAPOR enzyme, in which no energy seems to be conserved (*46*). Furthermore, *P. furiosus* uses alanine production as an electron sink in the presence of a high hydrogen partial pressure in combination with the absence of elementary sulfur (*47*).

The Crenarchaeon *Sulfolobus solfataricus* was originally isolated from solfatara fields and is a micro-aerophilic acidophile, which grows optimally at 85°C and pH 2-4 (*48*). The genome of *S. solfataricus* P2 is close to 3 Mb large and has recently been sequenced (*49*). While there exist some differences between a number of *S. solfataricus* strains in the capacity to utilize monomeric sugars, they are able to grow on several α- and β-linked saccharides (*5, 50*). These are taken up by binding protein-dependent ABC-type transporters (*50*). Growth on polysaccharides seems restricted to starch, dextrin, and xyloglucan (*5, 51*). Intracellularly, *S. solfataricus* produces an α-glucosidase, a β-glycosidase LacS with broad substrate specificity, and an α-xylosidase XylA (*52-55*). LacS and XylA seem to be involved in xyloglucan degradation (*51*). Monosaccharides are oxidized to CO_2, using O_2 as the terminal electron acceptor, which involves a partially non-phosphorylated type of Entner-Douderoff pathway (*45*), a TCA cycle and an unique respiratory chain (*49*).
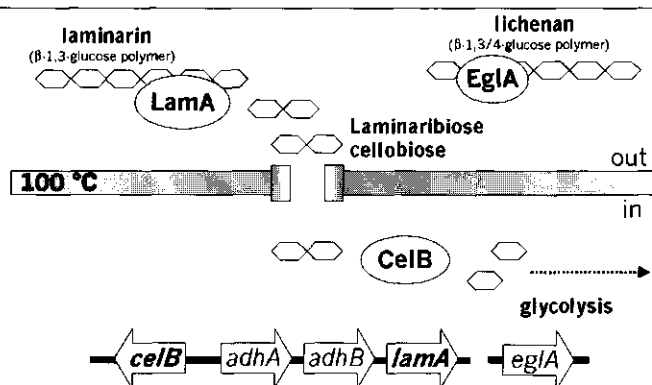
Figure 1.3 Schematic model of β-glucan degradation in *P. furiosus*. CelB: β-glucosidase, LamA: endo-β-1,3-glucanase, EglA: endo-β-1,3-1,4-glucanase (*27, 28, 39, 41*).

1.2 Family 1 glycosyl hydrolases from hyperthermophiles

Glycosyl hydrolases are enzymes that are capable of hydrolyzing glycosidic bonds. Given the extensive variety in carbohydrate stereochemistry, there are many different enzymes present in nature that are involved in their processing. This section summarizes general features of glycosyl hydrolases found in hyperthermophiles. Subsequently, family 1 of glycosyl hydrolases will be introduced with a description of enzyme structure, active site structure and mechanism of catalysis.

Glycosyl hydrolases from hyperthermophiles. Many hyperthermophiles can use carbohydrates as an energy source for growth and posses the necessary enzyme sets for their degradation (*45*). The produced enzymes are highly specific for either cleavage of the α- or the β-glycosidic bond in the middle (endo-acting) or at the end (exo-acting) of a poly-, oligo, or disaccharide molecule. Initially, α-glycosyl hydrolases were isolated from starch grown cultures, while, more recently, hyperthermophiles were found to be a source for β-glycosyl hydrolases as well (*56, 57*). In addition to the experimentally verified hyperthermophilic glycosidases, the recent advance in genome sequences has supplied a large number of novel glycosidase encoding genes, whose functions remain to be established. Among hyperthermophiles, the genome of the eubacterium *Thermotoga maritima* encodes a record number of 39 annotated glycosyl hydrolases, of which more than half have not been characterized yet (*58*). The isolated enzymes have their optimum for catalysis at high temperatures and are highly resistant against thermal and chemical inactivation, which makes them attractive biocatalysts for biotechnological applications (*6*).

Glycosyl hydrolases have been classified in over 80 families that have been organized in 11 clans based on protein sequence and folding similarities (*59*). The advantage of such classification is that the enzymes in each family share a similar peptide fold and reaction mechanism (*60*). Despite the large variety in glycosidases, there are only two mechanisms by which hydrolysis of the glycosidic bond occurs: either with retention or inversion of the stereochemistry at the anomeric carbon atom (*61*). Glycosyl hydrolases from hyperthermophiles are similar to those found in organisms with lower optimal growth temperatures and in 33 glycosidase families hyperthermophilic representatives can be found (*58*). Family 1 of glycosyl hydrolases is made up by exo-acting β-glycosidases that can be found in all three domains of life. Extreme thermostable representatives of this family originating from hyperthermophilic Archaea have been studied in the work described in this thesis.

Family 1 β-glycosidases from hyperthermophilic Archaea. The family 1 glycosidases of hyperthermophilic Archaea can be divided in three separate groups (Figure 1.4.), which have each one or more characterized members. The *P. furiosus* β-glucosidase CelB and *S. solfataricus* β-glycosidase LacS fall in group A (Figure 1.4) and have been the subject of numerous structure-function and structure-stability studies (*62-66*). Both enzymes are very thermostable and highly thermoactive, and have *in vivo* roles in the degradation of β-linked sugars, such as laminarin in *P. furiosus* (*41*) and xyloglucan in *S. solfataricus* (*51*). *P. furiosus* BmnA and *P. horikoshii* BglB of

group B are β-mannosidases with low and very low hydrolytic activity, respectively (*67*). The physiological role of these enzymes is unknown, but their specificity and low activity could suggest a role in biosynthesis or turnover of mannose containing compatible solutes (*68*). The *P. horikoshii* BglA of group C is membrane-associated and has a high specificity for alkalic glucosides (*69*).
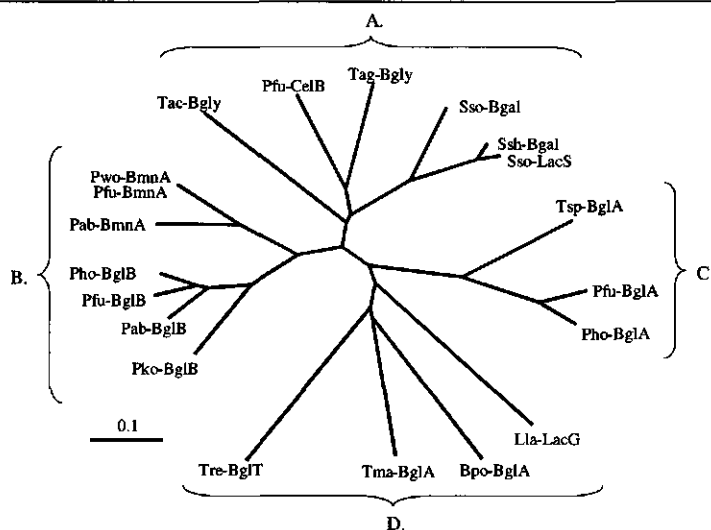
Figure 1.4 Phylogenetic tree of Archaeal family 1 β-glycosyl hydrolase protein sequences. A. (putative) β-glycosidases: Pfu_CelB: *P. furiosus* β-glucosidase CelB (AF013169), Tag-Bgly: *T. aggregans* β-glycosidase (AF053078), Sso-LacS: *S. solfataricus* β-glycosidase LacS (M34696), Sso-Bgal: *S. solfataricus* β-galactosidase Bgal (X15950), Ssh-Bgal: *S. shibatae* β-galactosidase Bgal (L47841). B. (putative) β-mannosidases: Pho-BglB: *P. horikoshii* β-mannosidase (AP000002), Pab-BglB: *P. abyssi* putative β-mannosidase BglB (AJ248288), Pfu-BglB: *P. furiosus* putative β-mannosidase BglB (*P. furiosus* genome ORF Pf_368506), Tko-BglB: *Thermococcus kodakaraensis* putative β-mannosidase BglB (AB028601), Pfu_BmnA: *P. furiosus* β-mannosidase BmnA (U60214), Pab-BmnA: *P. abyssi* putative β-mannosidase BmnA (AJ248285). C: (putative) membrane associated β-glucosidases: Pfu-BglA: *P. furiosus* putative β-glucosidase BglA (AF195244), Pho-BglA: *P. horikoshii* β-glucosidase BglA (C71144), Tsp-BglA: *Thermococcus* sp. putative β-glucosidase BglA (Z70242), D. Bacterial and eukaryal β-glycosidases: Tre-BglT: *Trifolium repens* cyanogenic β-glucosidase (X56733), Tma-BglA: *Thermotoga maritima* β-glucosidase BglA (X74163), Bpo-BglA: *Bacillus polymyxa* β-glucosidase BglA (M60210), Lla-LacG: *Lactococcus lactis* 6-phospho-β-galactosidase LacG (M28357). Tree was produced in ClustalX (*157*) and visualized using Treeview (*158*).

The hyperthermophilic origin of the β-glycosidases is reflected by their extreme thermostability. This stability is intrinsic to the amino acid sequence and independent of the producing organism, since functional expression of the thermostable β-glycosidases in *Escherichia coli* and *Lactococcus lactis* (*40, 70*), and *Saccharomyces cerevisiae* (*71*) yields enzymes with wild-type properties. The use of mesophilic hosts offer simple down-stream processing approaches for purification of these hyperthermostable ennzymes: virtually pure protein can be obtained after a heat incubation and subsequent centrifugation of the cell-free extract (*72*).

Mechanism of catalysis by family 1 glycosyl hydrolases. Family 1 glycosyl hydrolases hydrolyze their substrates with overall retention of the stereochemistry at the C1 atom. This proceeds through a double displacement mechanism, in which a carboxylic residue undertakes a nucleophilic attack on the C1 atom of the non-reducing saccharide residue and forms a covalent glycosyl-enzyme intermediate (Figure 1.5). The glycosidic bond is cleaved and the leaving group abstracts a proton from a second carboxylic residue that acts as a general acid base. Next, an incoming acceptor, water in the case of hydrolysis, attacks the covalent bond between the enzyme and the glycoside. The kinetic scheme for the reaction is as follows (*73*):

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\longrightarrow} E\text{-}S \overset{k_3}{\longrightarrow} EP$$
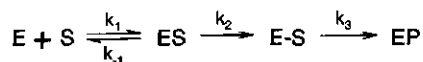
The rate-limiting step in the reaction is either k_2 or k_3, depending on the pK_a-value of the leaving group. Below a pK_a-value of 8-9, k_3 is rate-limiting, while at higher pKa-values k_2 becomes
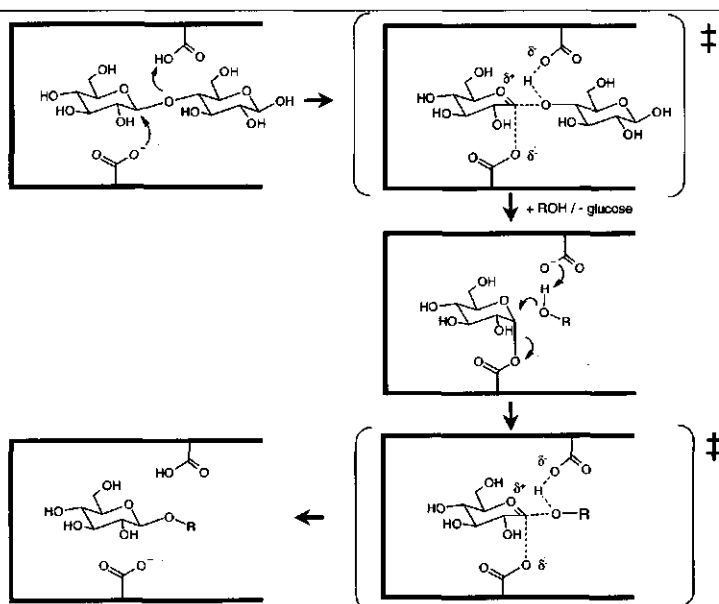
Figure 1.5. Mechanism of hydrolysis by retaining β-glycosyl hydrolases after ref. (*159*).

Figure 1.6 Ribbon presenatation of tetrameric CelB (*62*). Image was generated using Swiss-PDB viewer (*160*) and visualized using Povwin (*161*).

the rate limiting step (*73-75*). This implicates that the hydrolysis of lactose or cellobiose, is rate-limited by k_2, since the pK_a of a leaving 4-OH group of glucose is about 12.4 (*76*). In the hydrolysis of chromogenic substrates with leaving groups as *ortho-* or *para*-nitrophenol (pK_a = 7.2 (*73*)), k_3 determines the reaction rate. A glucose-β-1,4-glucose bond is the most stable covalent bond in biopolymers with an estimated half-life of 4.7 million years at 25°C (*76*).

The enhancement in hydrolysis of a factor of 10^{17}, makes glycosidases exceptionally efficient biocatalysts (*76*). This illustrates that glycosidases have been extremely well optimized for the lowering of the activation energy of the reaction. Retaining β-glycosidases have been found to distort the glycoside residue in the −1 subsite upon binding of the substrate in the active site, thus putting strain on the glycosidic bond (*77*). Furthermore, the active sites of β-glycosidases have been optimized for stabilization of the transition-state of the reaction, which is believed to have a trigonal geometry at the C1-atom and substantial oxo-carbenium ion character (Figure 1.5) (*73, 78*). This is clearly illustrated by the finding that glucoside hydrolysis by family 1 enzymes is strongly and competitively inhibited by transition state analogues with K_i-values that are at least 3 orders of magnitude smaller than the K_m for the corresponding substrates (*73, 74, 79*). The experimentally deduced double-displacement mechanism has been confirmed by the presence and position of two

fully conserved glutamate residues in family 1 protein sequences and determined 3D structures which serve as the nucleophile and general acid/base in the reaction *(80, 81)*. Substitution of the general acid/base residue results in about 10^3-fold reduced reaction rate which is pH-independent *(65, 81)*. In contrast, removal of the nucleophile of the reaction results in virtually complete inactivation *(40, 65, 80, 82)*, which can be partly rescued by another nucleophile- acting amino acid residue *(82)* or by addition of external nucleophiles, such as azide and formate *(83, 84)*.

The central fold of family 1 glycosyl hydrolases are $(\beta\alpha)_8$-barrels *(85)*, which in the case of the determined structures of *S. solfataricus* LacS *(64)*, *T. aggregans* Bgly *(86)* and *P. furiosus* CelB *(62)* have been arranged in a homotetrameric configuration (Figure 1.6). Organisation in oligomers has been observed in other thermostable enzymes as well and is believed to be a thermostablizing mechanism *(6, 87)*. The active sites of determined crystal structures of family 1 enzymes have been very well conserved with at most minor structural changes, which must determine the slight differences in substrate specificity *(62, 64, 86, 88-91)*. The 10 amino acid residues, which have been shown to interact with the substrate in the −1 subsite in crystal structures co-crystallized with ligands *(89, 92)*, have been highly conserved in the archaeal family 1 protein sequences (Figure 1.7).
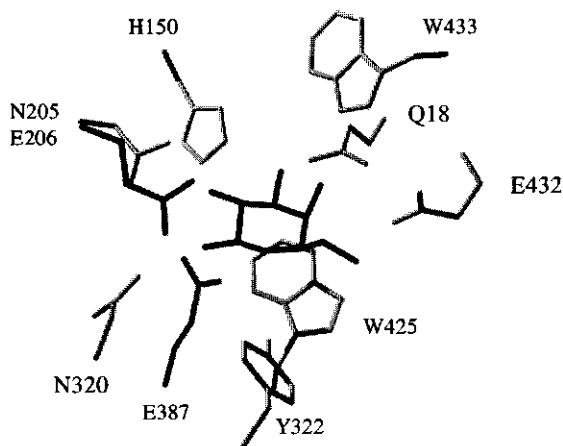
Figure 1.7 Conserved residues in −1 subsite of the active site of family 1 β-glycosidase LacS from *S. solfataricus* with a modeled galactose molecule *(64)*. Catalytic Glu 207 (acid/base) and Glu 387 (nucleophile) and galactose have been shaded black. Image was generated using Swiss-PDB viewer *(160)* and visualized using Povwin *(161)*.

1.3 Protein stability

In general, enzymes from hyperthermophiles display extreme thermal and chemical stabilities. Below, the initial folding of a protein is described, followed by a general description of conformational stability of proteins. Thereafter, an overview of currently identified thermostabilizing mechanisms is given.

Protein folding. In their native form all proteins adopt a fold that is determined by their amino acid sequences. At the moment, the way from an unorganized peptide sequence to a folded protein is described most accurately by a 3D folding landscape or folding funnel, in which different routes can lead to the native folded protein structure of minimal energy (*93*). Clustering of hydrophobic groups to prevent energetically-unfavorable hydration, also called the hydrophobic effect, is the major driving force to a collapsed structure that after rearrangement reactions results in the native protein (*94*). The conformational stability of the folded peptide structure is the difference between entropy, which strives for optimal conformational freedom of the peptide chain, and a large number of weak, stabilizing forces. The main stabilizing forces for the folded protein are hydrophobic interactions and hydrogen bonding (*95*). Each effect contributes about equally and results in a stability, which equals only a few weak intermolecular interactions (*96*). Interestingly, the conformational stability of hyperthermostable enzymes can be similar to that of less stable counterparts (*97*).

Thermostabilizing mechanisms. Proteins from hyperthermophilic organisms can endure much higher temperatures than their mesophilic counterparts. Since heterologously produced enzymes from hyperthermophilic origin often display wild-type stability, the molecular determinants for thermostability lay encoded solely by the amino acid composition and peptide sequence. The differences between proteins from mesophiles and hyperthermophiles are marginal since their proteins are all composed of the same 20 natural amino acids and can have amino acid sequences that are highly homologous. This results in near-identical structures that have been optimized for catalysis employing identical reaction mechanisms (*6, 87, 98*). Moreover, no general rules for thermostabilization have been identified so far. Rather, a number of factors have been observed in extreme thermostable proteins that seem to be instrumental in protein stabilization in hyperthermophiles (*6, 87, 96, 99*). From genome sequencing data it has been deduced that proteins from hyperthermophiles contain relatively more charged residues as well as hydrophobic residues (*6*). Asparagine and glutamine are sensitive for deamination at higher temperatures and probably for that reason occur less than in the average open reading frame from mesophiles (*96, 100*). Since there are large differences between hyperthermophile genomes, however, these data cannot be generalized (*6*). The number of cysteines is slightly lower in proteins from hyperthermophiles compared to mesophiles (*6*). However, disulfide bridges have also been found in hyperthermophilic proteins (*86*). Some enzymes have required thermostability by organization in higher oligomeric states compared to homologous enzymes from mesophiles, which reduces the surface area (*62, 64, 86*). Optimized subunit interactions furthermore, enhance thermostability, and for instance may

include reduction of subunit cavities (*100*) or optimized packing density (*101*). In general, but especially at subunit interfaces, a high number of ion-pair interactions between charged amino acid residues, has been observed in many hyperthermophilic protein structures (*64, 86, 100, 102*). Several studies have indicated that these ion-pairs are involved in protein stabilization, since disruption of surface ion-pair networks reduced protein stability (*103-107*). This has been confirmed by improvement of protein stability by introduction of ion-pairs, which, however, in some cases emphasized the importance of the exact positioning of the charged residues to obtain a stability increase (*108-110*). More surface area reduction results from shortening of loops (*100, 107*). Furthermore, an optimized packing efficiency has been observed in thermostable proteins, as well as modified hydrogen bonding patterns (*102*). In proteins from hyperthermophilic origin, buried solvent molecules have been observed, which could stabilize by providing missing van der Waals interactions or hydrogen bonds (*86*). However, for some proteins from hyperthermophiles it has not yet been possible to explain their extreme stability (*111*).

1.4 Approaches for enzyme design

By changing the amino acid sequence of a protein, the role of single amino acid residues in protein structure of function can be probed. When the gene of a protein is known and it can be expressed in a convenient heterologous host, a wide array of molecular biological techniques are at the scientist's disposal to evaluate structure-function relationships. The choice, which technique to use, is guided by the resources available, the purpose of the study and the available information on the enzyme of interest. For the studies described in this thesis, two fundamentally different approaches have been used, site-directed mutagenesis and directed evolution, which both will be described below. Each approach will be illustrated by recent studies that have relevance to the work described in this thesis.

Site-directed mutagenesis. Site-directed mutagenesis is the technique of choice for assessing the role of single amino acid residues in enzyme catalysis mechanism or stability. By this technique, a defined amino acid residue is deleted from a protein sequence, introduced into it, or substituted by a residue of choice. Analysis of the biochemical properties of the enzyme variant will give information about role of the designed residue in the protein. The obvious basis of this approach is the design of the mutation. Ideally, a high-resolution 3D structure is used for the identification of mutagenesis sites. When a 3D structure of the target protein is not available, one may use the structure of a related enzyme, and identify target residues in multiple sequence alignments. Alternatively, the structure of a related protein can be used for the construction of a 3D-model of the target protein. In the absence of any structural information, chemical labeling, followed by analysis of peptide hydrolysates, might identify, for instance, mechanistically active residues.

Site-directed mutagenesis has been an important tool for elucidation of protein structure-function relationships (*40, 84, 112*) and protein stabilizing mechanisms (*103, 108, 109, 113, 114*). Substitition of catalytic residues has been instrumental in the elucidation of the catalytic mechanism

of family 1 glycosyl hydrolases (*40, 65, 82, 115, 116*), while in the same enzyme family the substrate specificity has been probed by replacement of active site residues after structural comparison of β-glycosidases with different specificities (*62, 117*).

Directed evolution. Where site-directed mutagenesis represents a top-down approach towards the design of a small number of enzyme variants, directed evolution is a bottom-up approach, in which the creation of a large numbers of enzyme variants is followed by screening for variants that exhibit the desired properties. Any protein of N amino acids can be composed in 20^N ways, which represents the sequence space for the protein. Directed evolution is the exploration of this sequence space (*118*). Mutants that show the desired phenotype can serve as a starting material for new mutagenesis rounds. It has been argued that the steps per evolutionary cycle should be small to avoid accumulation of mutations with neutral or negative effects (*118*). In combining random mutagenesis and selective pressure, followed by proliferation of the most successful variants, the process is similar to that of Darwinian evolution. At the moment it is the approach of choice for engineering of enzyme properties. An advantage of this method is the limited amount of knowledge that is required about the enzyme's structure and reaction mechanism. Alternatively, determining factors are the generation of a large number of enzyme variants with randomly distributed mutations, a suitable expression system and, most important, an efficient and reliable screening that specifically selects for the desired properties (Figure 1.8).

Several methods for the creation of a large variety in gene sequences have been developed recently. In the case of a single gene, the genetic diversity of a pool of randomly mutated variants can be enlarged by DNA shuffling, a procedure which includes recombination of the sequences after fragmentation and subsequent reassembly in which new variants are created (Figure 1.9A) (*119, 120*). The use of multiple genes with substantial homology in a DNA family shuffling
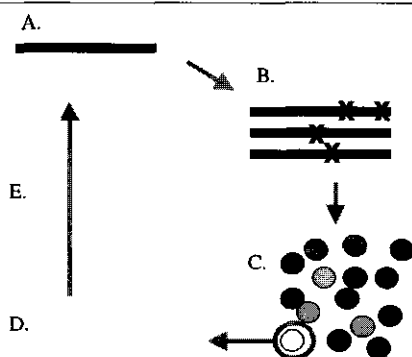
Figure 1.8 General approach for directed evolution of enzymes. A: gene coding for enzyme of interest, B: generation of genetic variants, C: production of enzyme variants in suitable host according to one cell, one sequence, D: isolation of genes coding for variants with desired features, E: improved genes serve as starting material for a new cycle (after http://www.che.caltech.edu/groups/fha/Enzyme/ directed.html).
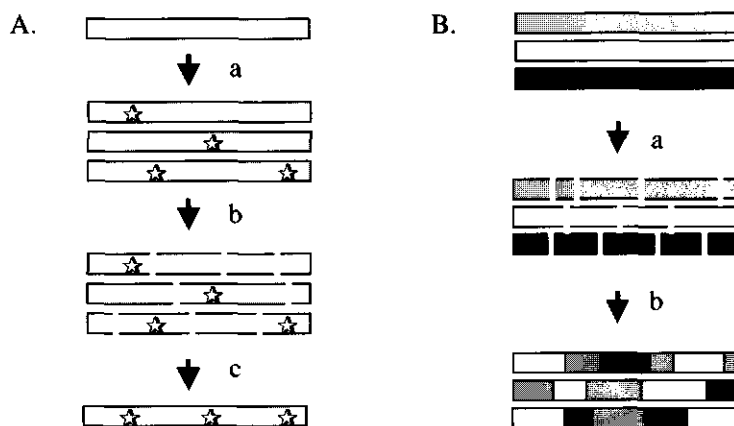
Figure 1.9 Generation of genetic diversity by DNA shuffling. A: Single gene shuffling: (a) introduction of random mutations (error-prone PCR), (b) gene fragmentation (DNAse treatment) and (c) reassembly (*119*). B: DNA family shuffling (homology >70%): (a) gene fragmentation (DNAse treatment) and (b) reassembly (*121, 162*).

approach (Figure 1.9B), accelerates the evolution of enzyme function (*121*), because a larger amount of sequence space can be explored (*118*). For shuffling or chimeragenesis of genes with low or no homology, various methods have been developed (*122-124*). Directed evolution approaches have proven themself in tailoring enzyme properties including thermostability (*125, 126*), and activity in organic solutes (*127*). Relatively few enzymes from hyperthermophilic organism have been subject in directed evolution experiments. In the reported cases, single genes were optimized for increased activity at room-temperature (*63, 128, 129*). In the family 1 of glycosyl hydrolases, the β-glucosidase from mesophilic *Bacillus polymyxa* was thermostabilized in a laboratorium evolution approach (*130*). Recently, a screening for *Agrobacterium faecalis* β-glucosidase mutants was reported, which allowed for a quantitative detection of the synthesis of β-1,4-oligo-glucosides (*131*).

1.5 Industrial applications of β-glycosidases

In quantity and price, lactose, galactose-β-1,4-glucopyranoside, is the most important β-linked industrial sugar (Figure 1.10). Lactose is found only in mammalian milk and is produced in large amounts by the cheese industry in the form of cheese whey and whey permeate (*132*). Pure lactose has several applications in the food and pharmaceutical industry, but could be used for the manufacturing of derivatives with an increased economical value. Isomerization of lactose gives lactulose with laxative properties, while reduction or oxidation of the free aldehyde groups yield lactitol or lactobionic acid, which can be used as a low-calorie sweetener or preservator of transplant organs, respectively. Through the action of β-glycosidases, value can be added to lactose by hydrolysis to monomeric glucose and galactose, synthesis of galacto-oligosaccharides and
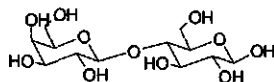
Figure 1.10 Molecular structure of lactose (galactosyl-β-1,4-glucose).

glycoconjugates (*132*) (Figure 1.11). The advantages of thermostable glycosidases are their longer operational stability that allows for higher process temperatures, which reduces risks of bacterial contamination and, increases substrate solubility and reaction rates (*22*).

Hydrolysis of lactose. Hydrolysis of lactose to glucose and galactose can be applied for the production of low-lactose milk and the valorization of whey. Lactose in milk can be the cause of digestive problems for persons that lack the gene coding for the enzyme lactase that hydrolyzes lactose to monomeric galactose and glucose. Treatment of milk with β-galactosidase can replace the lactase in the human intestine. Expression of lactase in milk by recombinant techniques is an elegant solution (*133*), but not yet feasible. Other techniques involve treatment of milk with immobilized β-galactosidase (*134*). Although hydrolysis of the lactose present in whey increases the sweetness and has beneficial effects for the use of whey in the food-industry, it is economically not interesting due to the competition of other sweeteners (*132*). The β-glycosidases CelB and LacS, from *P. furiosus* and *S. solfataricus* respectively, have recently been shown to be suitable for such processes at temperatures above 70 °C (*135*). By directed evolution, it has been demonstrated that the *P. furiosus* CelB and *S. solfataricus* LacS can be used for the creation of hybrid enzymes with enhanced lactose hydrolysis capacities at 70 °C (*136*).

Oligosaccharide synthesis. Short chain non-digestible carbohydrates have potential use as prebiotics in functional foods (*137*) by selectively stimulating bacteria in the intestine and, as such, may have a beneficial effect on the health of the host (*138*). Oligo-saccharides have reported effects on colon cancer protection (*139*), mineral absorption (*140*) and lipogenic enzyme expression (*141*). These studies, however, concentrate mainly on a limited number of fructose-containing oligosaccharides (*137*), while the variety in saccharides allows for an extended number of different short chain carbohydrates. Glycosidases can be employed for the synthesis of novel oligo-saccharides, which occurs as a side reaction during disaccharide hydrolysis through the retaining double-displacement mechanism (Figure 1.11) (*142*). A promising source for oligo-saccharides is provided by the milk disaccharide lactose (*132*), which is reflected by several studies on enzymatic galactooligosaccharide (GOS) synthesis from lactose using retaining β-glycosidases (*143-146*). The limited solubility of high amounts of lactose at room temperature initiated synthesis studies at elevated temperatures with the extreme thermostable β-glycosidases CelB and LacS as catalysts. Both enzymes produce GOS in relatively high yields during hydrolysis of lactose (*135, 147, 148*). The elevated temperatures resulted in higher lactose concentrations and subsequent higher GOS yields up to 43% w/w (*148*). The optimum temperature for the GOS synthesis was well below the
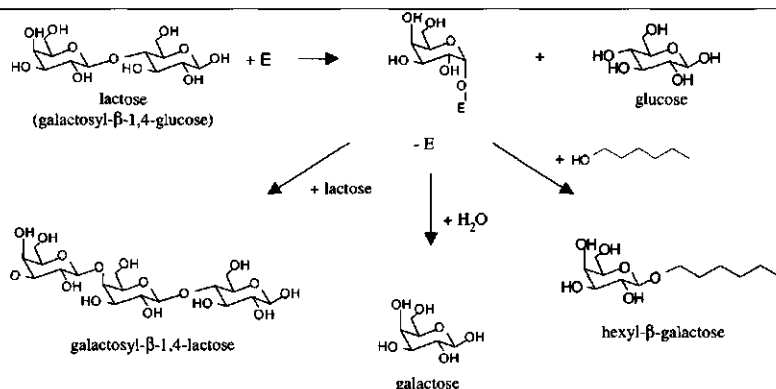
Figure 1.11 Reactions catalyzed by family 1 β-glycosidases

optimum temperatures of both enzymes, since above 75 °C the enzymes were rapidly inactivated due to Maillard reactions (*147*).

Both CelB and LacS have a preference for the formation of galactose-β-1,3- and 1,6-glucose bonds (*149*), which for CelB correlates with its apparent involvement in growth on β-1,3 glucosides in *P. furiosus* (*41*). Besides the use of elevated temperatures, the high yields for CelB in GOS synthesis (*147-149*) are determined by the structure of their catalytic centers, since the yields could be manipulated by active site variations (*148*). The *P. furiosus* β-mannosidase BmnA and *P. furiosus* β-glucosidase CelB were found to perform equally well in GOS synthesis. Taking into account that the two enzymes differ considerably in hydrolysis rates, the yield and pH-optimum of GOS synthesis by thermostable family 1 β-glycosidases was found to be independent of their catalytic parameters for hydrolysis (*148*).

A drawback of oligosaccharide synthesis by retaining wild-type enzymes is the activity of the β-glycosidases towards their synthesis products, which can be hydrolyzed again. By removal of the amino acid residue, which acts as the nucleophile in the synthesis and hydrolysis reactions, hydrolysis of the synthesis products is prevented. In this case however, the activity of the enzyme has to be restored by use of α-glycosylfluorides (*116, 150*) or external nucleophiles (*83, 84, 151*). When the catalytic nucleophile Glu387 was replaced by Ala or Gly, *S. solfataricus* LacS was completely inactivated. Activity was restored by addition of azide or formate, which in the case of formate yielded a saccharide synthesizing enzyme (*84*). The LacS "glycosynthase" can be used for the synthesis of linear and branched oligosaccharides with β-1,3-, 1,4- and 1,6 linkages, depending on the used external nucleophile (*152*).

Glycoconjugate synthesis. The coupling of sugars to aglycons yields products that find various applications in the food, personal care or farmaceutical industry (*153*). By using family 1 β-glycosidases, the non-reducing moiety of a disaccharide can be coupled to alkylic chains (Figure 1.11) (*153*). Alkylic galactosides can be produced by *S. solfataricus* β-glycosidase and have

potential as surfactants or components for sugar base polymers (*154*). In contrast to mesophilic counterparts, the β-glucosidase CelB from *P. furiosus* is able to glucosylate tertiary alcohols by transglycoslylation (*155*). Moreover, the CelB enzyme has been shown to glucosylate hexanol via direct glucosylation (*156*).

1.6 Aims and outline of the thesis

This thesis describes studies on the substrate specificity and activity of β-glycosidases from hyperthermophilic Archaea. With the β-glucosidase CelB from *P. furiosus* as starting point, the aims of the studies were (i) to elucidate molecular determinants of substrate recognition and catalysis in β-glycosidases, (ii) to identify and characterize novel β-glycosidases, (iii) to change their substrate specificity and activity by enzyme engineering, and (iv) test the enzymes in applied reactions.

Chapter 1 introduces various aspects of hyperthermophilic life and its enzymes, and focuses on β-glycosidases of family 1 with their applications for lactose conversion.

A crystal structure of the model enzyme CelB, based on 3.3 Å diffraction data, is described in Chapter 2. Within family 1 of glycosyl hydrolases, the 6-phospho-β-glycosidases form a distinct subgroup with a unique substrate specificity that was believed to result from a structural adaptation of the active site. The structure of the CelB active site was compared to that of the related 6-phospho-β-galactosidase LacG of *Lactococcus lactis* (*92*) and 3 mutations were introduced in CelB for enhancement of hydrolysis of 6-phospho-galactosides. A detailed kinetic analysis of the CelB variants is presented.

In several members of the Thermococcales, to which the genus of *Pyrococcus* belongs, genes coding for family 1 enzymes with an apparent unusual active site structure were identified. The gene from *P. horikoshii*, denoted *bglB*, was chosen as a representative and its gene and its product BglB is described in Chapter 3. In addition, the unique substrate specificity of BglB was studied by introduction of the two unique active site residues in CelB from *P. furiosus* and *vice versa*. The effects of the mutations on catalysis and active site structure are discussed in detail.

Chapter 4 introduces the β-glycosidase LacS from *Sulfolobus solfataricus* by the first direct biochemical comparison with the β-glucosidase CelB from *P. furiosus*. The comparison covered the thermoactivity, substrate specificity and thermostability and provided insight in the similarities in activity and differences in the thermostabilization of the two enzymes.

The next two studies were focused on the laboratory evolution of β-glycosidases for improved catalytic activity at suboptimal temperatures. In the study described in Chapter 5, the β-glucosidase CelB was optimized for catalysis of the chromogenic substrate *para*-nitrophenol-β-D-glucoside at room temperature by error-prone PCR combined with DNA shuffling. The results of two rounds of evolution and characterization of isolated high-performance hybrids are presented.

The creation of β-glycosidase hybrids by shuffling of the CelB and LacS-encoding genes has been described in Chapter 6. Hybrid enzymes were specifically selected for hydrolysis of the natural disaccharide lactose at 70 °C, the temperature of choice for lactose conversion (*135*). Structural and biochemical interpretation of high-performance hybrids is presented.

CelB and several CelB variants, which have been described in Chapters 2 and 3, were tested for their transgalactosylation ability. The enzymes were tested for the formation of galacto-oligosaccharides from lactose at high temperatures. This is described in Chapter 7.

Chapter 8 summarizes the obtained results and discusses them in the context of other studies in related research fields.

References.

1. Jupp, T., and Schultz, A. (2000) *Nature 403*, 880-3.
2. Huber, R., Huber, H., and Stetter, K. O. (2000) *FEMS Microbiol Rev 24*, 615-23.
3. Reysenbach, A. L., Longnecker, K., and Kirshtein, J. (2000) *Appl Environ Mirobiol 66*, 3798-806.
4. Zierenberg, R. A., Adams, M. W. W., and Arp, A. J. (2000) *Proc Nat Acad Sci 97*, 12961-2.
5. Grogan, D. W. (1989) *J Bacteriol 171*, 6710-9.
6. Vieille, C., and Zeikus, G. J. (2001) *Microbiol Mol Biol Rev 65*, 1-43.
7. Wiegel, J., and Adams, M. W. W. (1998) pp 346, Taylor & Francis, London, UK.
8. Nisbet, E. G., and Sleep, N. H. (2001) *Nature 409*, 1083-91.
9. DeLong, E. F. (2001) *Methods Enzymol 330*, 3-11.
10. Stetter, K. O. (1999) *FEBS letters 452*, 22-5.
11. Barns, S. M., Delwiche, C. F., Palmer, J. D., and Pace, N. R. P. (1996) *Proc Nat Acad Sci 93*, 9188-93.
12. Marteinsson, V. T., Kristiansson, J. K., Kristmansdottir, H., Dahlkvist, M., Saemmundsson, K., Hannington, M., Petursdoittir, S. K., Geptner, A., and Stoffers, P. (2001) *Appl Environ Microbiol 67*, 827-33.
13. Doolittle, W. F. (1999) *Trends Cell Biol 9*, M5-8.
14. Nelson, K. E., and *al, e.* (1999) *Nature 399*, 323-38.
15. Maeder, D. L., Weiss, R. B., Dunn, D. M., Cherry, J. L., González, J. M., DiRuggerio, J., and Robb, F. T. (1999) *Genetics 152*, 1299-1305.
16. Ettema, T. J. G., Van der Oost, J., and Huynen, M. A. (2001) *Trends Genet 17*, in press.
17. Lecompte, O., Ripp, R., Puzos-Barbe, V., Duprat, S., Heilig, R., Dietrich, J., Thierry, J. C., and Poch, O. (2001) *Genome Res 11*, 981-93.
18. Kyrpides, N. C., and Olsen, G. J. (1999) *Trends in Gen 15*, 298-9.
19. Aravind, L., Tatusov, R. L., Wolf, Y. I., Walker, D. R., and Koonin, E. V. (1998) *Trends Genet 14*, 442-4.
20. Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I., and Koonin, E. V. (1999) *Genome Res 9*, 608-28.
21. DiRuggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, H., Chatard, J., Horlacher, R., Robb, F. T., Boos, W., and Weiss, R. B. (2000) *Mol Microbiol 38*, 684-93.
22. Niehaus, F., Bertoldo, C., Kähler, M., and Antraninkian, G. (1999) *Appl Microbiol Biotechnol 51*, 711-729.
23. Adams, M. W., and Kelly, R. M. (1998) *Trends Biotechnol 16*, 329-32.
24. Jarrell, K. F. (1999) in *http://www.findarticles.com?cf_0/m1042/7_49/55294941/print.jhtml*.
25. Fiala, G., and Stetter, K. O. (1986) *Arch Microbiol 145*, 56-61.
26. Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B., and Dunn, D. M. (2001) *Methods Enzymol 330*, 134-57.
27. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.
28. Gueguen, Y., Voorhorst, W. G., van der Oost, J., and de Vos, W. M. (1997) *J Biol Chem 272*, 31258-64.
29. Driskill, L., Kusy, K., Bauer, M., and Kelly, R. (1999) *Appl Environ Microbiol 65*, 893-7.

30. Koch, R., Zablowski, P., Spreinat, A., and Antranikian, G. (1990) *FEMS Microbiol Lett 71*, 21-6.
31. Jorgensen, S., Vorgias, C. E., and Antranikian, G. (1997) *J Biol Chem 272*, 16335-42.
32. Brown, S. H., and Kelly, R. M. (1993) *Appl Environ Microb 59*, 2614-21.
33. Dong, G., Vieille, C., and Zeikus, G. (1997) *Appl Environ Microb 63*, 3577-84.
34. Costantino, H. R., Brown, S. H., and Kelly, R. M. (1990) *J Bacteriol 172*, 3654-60.
35. Horlacher, R., Xavier, K. B., Santos, H., DiRuggiero, J., Kossmann, M., and Boos, W. (1998) *J Bacteriol 180*, 680-9.
36. Koning, S., Elferink, M. G. L., Konings, W. N., and Driessen, A. J. M. (2001) *in press*.
37. Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. S., Griko, Y. V., Privalov, P. L., and Anfinsen, C. B. (1993) *J Biol Chem 268*, 24394-401.
38. Laderman, K. A., Asada, K., Uemori, T., Mukai, H., Taguchi, Y., and Anfinsen, C. B. (1993) *J Biol Chem 268*, 24402-7.
39. Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J., and Kelly, R. M. (1999) *J Bacteriol 181*, 284-90.
40. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J Bacteriol 177*, 7105-11.
41. Voorhorst, W., Gueguen, Y., Geerling, A., Schut, G., Dahlke, I., Thomm, M., Van der Oost, J., and De Vos, W. (1999) *J Bacteriol 181*, 3777-3783.
42. Driskill, L. E., Bauer, M. W., and Kelly, R. M. (1999) *Biotechnol Bioeng 66*, 51-60.
43. Kengen, S. W., de Bok, F. A., van Loo, N. D., Dijkema, C., Stams, A. J., and De Vos, W. M. (1994) *J Biol Chem 269*, 17537-41.
44. Kengen, S. W., Tuininga, J. E., de Bok, F. A., Stams, A. J., and De Vos, W. M. (1995) *J Biol Chem 270*, 30453-7.
45. Kengen, S. W. M., Stams, A. J. M., and De Vos, W. M. (1996) *FEMS Microbiol Rev 18*, 119-37.
46. van der Oost, J., Schut, G., Kengen, S. W., Hagen, W. R., Thomm, M., and De Vos, W. M. (1998) *J Biol Chem 273*, 28149-54.
47. Kengen, S. W. M., and Stams, A. J. M. (1994) *Archiv Microbiol 161*, 168-75.
48. Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., and Scholz, I. (1980) *Arch Microbiol 125*, 259-69.
49. She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C. Y., Groth Clausen, I., and al., e. (2001) *Proc Natl Acad Sci 98*, 7835-40.
50. Elferink, M. G., Albers, S. V., Konings, W. N., and Driessen, A. J. M. (2001) *Mol Microbiol 39*, 1494-503.
51. Cobucci Ponzano, B., Trincone, A., Rossi, M., and Moracci, M. (2001) *4th Carbohydrate Bioengineering Meeting Stockholm, Sweden*.
52. Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.
53. Nucci, R., Moracci, M., Vaccaro, C., Vespa, N., and Rossi, M. (1993) *Biotechnol Appl Biochem 17*, 239-50.
54. Rolfsmeier, M., and Blum, P. (1995) *J Bacteriol 177*, 482-5.
55. Moracci, M., Cobucci Ponzano, B., Trincone, A., Fusco, S., De Rosa, M., Van der Oost, J., Sensen, C. W., Charlebois, R. L., and Rossi, M. (2000) *J Biol Chem 275*, 22082-9.
56. Bauer, M. W., Driskill, L. E., and Kelly, R. M. (1998) *Curr Opin Biotechnol 9*, 141-5.
57. Sunna, A., Moracci, M., Rossi, M., and Antranikian, G. (1997) *Extremophiles 1*, 2-13.
58. Henrissat, B. (2001) *Methods enzymol 330*, 183-201.
59. Henrissat, B. (2001) *http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html*.
60. Henrissat, B., and Davies, G. (1997) *Curr Opin Struct Biol 7*, 637-644.
61. Davies, G., and Henrissat, B. (1995) *Structure 3*, 853-9.
62. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.
63. Lebbink, J. H. G., Kaper, T., Bron, P., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 3656-65.

64. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.

65. Moracci, M., Capalbo, L., Ciaramella, M., and Rossi, M. (1996) *Protein Eng 9*, 1191-5.

66. D'Auria, S., Moracci, M., Febbraio, F., Tanfani, F., Nucci, R., and Rossi, M. (1998) *Biochimie 80*, 949-57.

67. Kaper, T., Verhees, C. H., Lebbink, J. H. G., van Lieshout, J. F. T., Kluskens, L. D., Ward, D. E., Kengen, S. W. M., Beerthuyzen, M. M., De Vos, W. M., and van der Oost, J. (2001) *Methods Enzymol 330*, 329-46.

68. Da Costa, M. S., Santos, H., and Galinski, E. A. (1998) in *Advances in Biochemical Engineering Biotechnology* (Atranikian, G., Ed.) pp 117-54, Springer-Verlag, Berlin.

69. Matsui, I., Sakai, Y., Matsui, E., Kikuchi, H., Kawarabayasi, Y., and Honda, K. (2000) *FEBS Lett 467*, 195-200.

70. Moracci, M., Nucci, R., Febbraio, F., Vaccaro, C., Vespa, N., La Cara, F., and Rossi, M. (1995) *Enzyme Microb Technol 17*, 992-7.

71. Moracci, M., La Volpe, A., Pulitzer, J. F., Rossi, M., and Ciaramella, M. (1992) *J Bacteriol 174*, 873-82.

72. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2001) *Methods Enzymol 330*, 364-79.

73. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.

74. Bauer, M. W., and Kelly, R. M. (1998) *Biochemistry 37*, 17170-8.

75. Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausín, C., Querol, E., Planas, A., and Pérez-Pons, J. A. (2001) *Biochemistry 40*, 5975-82.

76. Wolfenden, R., Lu, X., and Young, G. (1998) *J Am Chem Soc 120*, 6814-5.

77. Davies, J. D., Mackenzie, L., Varrot, A., Dauter, M., Brzozowoski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707-13.

78. White, A., and Rose, D. R. (1997) *Curr Opin Struct Biol 7*, 645-51.

79. Kaper, T., Van Heusden, H. H., Van Loo, B., Vasella, A., Van der Oost, J., and De Vos, W. M. (2001) *submitted*.

80. Trimbur, D. E., Warren, R. A., and Withers, S. G. (1992) *J Biol Chem 267*, 10248-51.

81. Wang, Q., Trimbur, D., Graham, R., Warren, R. A., and Withers, S. G. (1995) *Biochemistry 34*, 14554-62.

82. Gebler, J. C., Trimbur, D. E., Warren, A. J., Aebersold, R., Namchuk, M., and Withers, S. G. (1995) *Biochemistry 34*, 14547-53.

83. Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J., and Withers, S. G. (1994) *J Am Chem Soc 116*, 11594-95.

84. Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., and Rossi, M. (1998) *Biochemistry 37*, 17262-70.

85. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) *Nature 255*, 609-14.

86. Chi, Y. I., Martinez-Cruz, L. A., Jancarik, J., Swanson, R. V., Robertson, D. E., and Kim, S. H. (1999) *FEBS Letters 445*, 375-383.

87. Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L., and Engel, P. C. (1998) *Biochimie 80*, 933-41.

88. Wiesmann, C., Beste, G., Hengstenberg, W., and Schulz, G. E. (1995) *Structure 3*, 961-8.

89. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.

90. Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995) *Structure 3*, 951-60.

91. Hakulinen, N., Paavilainen, S., Korpela, T., and J., R. (2000) *J Struct Biol 129*, 69-79.

92. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.

93. Radford, S. E. (2000) *Trends Biochem Sci 25*, 611-8.

94. Jaenicke, R. (1996) *Naturwissenschaften 83*, 44-54.

95. Pace, C. N., Shirley, B. A., McNutt, M., and Gajiwala, K. (1996) *FASEB J 10*, 75-83.

96. Jaenicke, R. (2000) *J Biotechnol 79*, 193-203.

97. Catanzano, F., Graziano, G., De Paola, B., and Barone, G. (1998) *Biochemistry 37*, 14484-90.

98. Jaenicke, R. (1998) *Biochemistry (Mosc) 63*, 312-22.

99.   Jaenicke, R., and Bohm, G. (1998) *Curr Opin Struc Biol 8 (6)*, 738-748.

100. Chang, C., Park, B. C., Lee, D. S., and Suh, S. W. (1999) *J Mol Biol 288*, 623-34.

101. Dams, T., Auerbach, G., Bader, G., Jacob, U., Ploom, T., Huber, R., and Jaenicke, R. (2000) *J Mol Biol 297*, 659-72.

102. Britton, K. L., Yip, K. S. P., Sedelnikova, S. E., Stillman, T. J., Adams, M. W. W., Ma, K., Maeder, D. L., Robb, F. T., Tolliday, N., Vetriani, C., Rice, D. W., and Baler, P. J. (1999) *J Mol Biol 293*.

103. Pappenberger, G., Schurig, H., and Jaenicke, R. (1997) *J Mol Biol 274*, 676-83.

104. Moracci, M., Ciaramella, M., Cobucci-Ponzano, B., Perugino, G., Rossi, M. (1998) *International Conference on Thermophiles 1998, Brest, France*.

105. Lebbink, J. H. G. (1999) PhD Thesis pp 200, Wageningen University, Wageningen, The Netherlands.

106. Arnott, M. A., Michael, R. A., Thompson, C. R., Hough, D. W., and Danson, M. J. (2000) *J Mol Biol 304*, 657-68.

107. Walden, H., Bell, G. S., Russell, R. J., Siebers, B., Hensel, R., and Taylor, G. L. (2001) *J Mol Biol 306*, 745-57.

108. Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S. P., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H., and Robb, F. T. (1998) *Proc Natl Acad Sci USA 95*, 12300-5.

109. Lebbink, J. H. G., Knapp, S., Van der Oost, J., Rice, D. W., Ladenstein, R., and De Vos, W. M. (1999) *J Mol Biol 289*, 357-69.

110. Rahman, R. N. Z. A., Fujiwara, S., Nakamura, H., Takagi, M., and Imanaka, T. (1998) *Biochem Biophys Res Commun 248*, 920-6.

111. Leonard, P. M., Smits, S. H., Sedelnikova, S. E., Brinkman, A. B., De Vos, W. M., Van der Oost, J., Rice, D. W., and Rafferty, J. B. (2001) *EMBO J 20*, 990-996.

112. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) *Nat Struct Biol 5*, 812-818.

113. Tomschy, A., Bohm, G., and Jaenicke, R. (1994) *Protein Eng 7*, 1471-8.

114. Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R., and de Vos, W. M. (1998) *J Mol Biol 280*, 287-96.

115. Lawson, S. L., Warren, R. A., and Withers, S. G. (1998) *Biochem J 330*, 203-9.

116. Mayer, C., Zechel, D. L., Reid, S. P., Warren, A. J., and Withers, S. G. (2000) *FEBS Letters 466*, 40-4.

117. Schulte, D., and Hengstenberg, W. (2000) *Protein Eng 13*, 515-8.

118. Arnold, F. H. (2001) *http://www.che.caltech.edu/groups/fha/Enzyme/directed.html*.

119. Stemmer, W. P. C. (1994) *Nature 370*, 389-91.

120. Stemmer, W. P. C. (1994) *Proc Natl Acad Sci 91*, 10747-51.

121. Crameri, A., Raillard, S.-A., Bermudez, E., and Stemmer, W. P. C. (1998) *Nature 391*, 288-291.

122. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) *Proc Natl Ac Sci USA 96*, 3562-7.

123. Sieber, V., Martinez, C. A., and Arnold, F. H. (2001) *Nat Biotechnol 19*, 456-60.

124. Coco, W. M., Levinson, W. E., Crist, M. J., Hektor, H. J., Darzins, A., Pienkos, P. T., Squires, C. H., and Monticello, D. J. (2001) *Nat Biotechnol 19*, 354-9.

125. Giver, L., Gershenson, A., Freskgard, P. O., and Arnold, F. H. (1998) *Proc Natl Acad Sci USA 95*, 12809-13.

126. Zhao, H., and Arnold, F. H. (1999) *Protein Eng 12*, 47-53.

127. You, L., and Arnold, F. H. (1996) *Protein Eng 9*, 315-19.

128. Merz, A., Yee, M. C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W. P., Yanofsky, C., and Kirschner, K. (2000) *Biochemistry 39*, 880-9.

129. Roovers, M., Sanchez, R., Legrain, C., and Glansdorff, N. (2001) *J Bacteriol 183*, 1101-5.

130. Arrizubieta, M. J., and Polaina, J. (2000) *J Biol Chem 275*, 28843-8.

131. Mayer, C., Jakeman, D. L., Mah, M., Karjala, G., Gal, L., Warren, R. A., and Withers, S. G. (2001) *Chem Biol 8*, 437-43.

132. Yang, S. T., and Silva, E. M. (1995) *J Dairy Sci 78*, 2541-62.

133. Jost, B., Vilotte, J. L., Duluc, I., Rodeau, J. L., and Freund, J. N. (1999) *Nat Biotechnol 17*, 160-4.

134.   Sun, S., Xuyan, L., Nu, S., and You, X. (1999) *J Agric Food Chem 47*, 819-23.

135.   Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.

136.   Kaper, T., Brouns, S. J. J., Geerling, A. C. M., De Vos, W. M., and Van der Oost, J. (2001) *submitted*.

137.   Roberfroid, M. B. (2000) *Am J Clin Nutr 71(suppl)*, 1682S-7S.

138.   Schrezenmeir, J., and De Vrese, M. (2000) *Am J Clin Nutr 73(suppl)*, 361S-4S.

139.   Wollowski, I., Rechkemmer, G., and Pool-Zobel, B. L. (2000) *Am J Clin Nutr 73(suppl)*, 451S-5S.

140.   Scholz-Ahrens, K. E., Schaafsma, G., E.G.H.M., V. d. H., and Schrezenmeir, J. (2000) *Am J Clin Nutr 73(suppl)*, 459S-64S.

141.   Delzenne, N. M., and Kok, N. (2000) *Am J Clin Nutr 73(suppl)*, 456S-8S.

142.   Watt, G. M., Lowden, P. A. S., and Flitsch, S. L. (1997) *Curr Opin Struct Biol 7*, 652-60.

143.   Prenosil, J. E., Stuker, E., and Bourne, J. R. (1987) *Biotechnol Bioeng 30*, 1026-31.

144.   Boon, M. A., Janssen, A. E. M., and Van 't Riet, K. (2000) *Enzyme Microb Technol 26*, 271-81.

145.   Yanahira, S., Kobayashi, T., Suguri, T., Nakakoshi, M., Miura, S., Ishikawa, H., and Nakajima, I. (1995) *Biosci Biotechn Biochem 59*, 1021-6.

146.   Nakao, M., Harada, M., Kodama, Y., Nakayama, T., Shibano, Y., and Amachi, T. (1994) *Appl Microbiol Biotechnol 40*, 657-63.

147.   Boon, M. A., Van der Oost, J., De Vos, W. M., Jansen, A. E. M., and Van 't Riet, K. (1998) *Appl Biochem Biotechnol 75*, 269-78.

148.   Hansson, T., Kaper, T., Van der Oost, J., De Vos, W. M., and Adlercreutz, P. (2000) *Biotechnol Bioeng 73*, 203-10.

149.   Petzelbauer, I., Zeleny, R., Reiter, A., Kulbe, K. D., and Nidetzky, B. (2000) *Biotechnol Bioeng 69*, 140-9.

150.   Malet, C., and Planas, A. (1998) *FEBS Letters 440*, 208-12.

151.   MacLeod, A. M., Tull, D., Rupitz, K., Warren, R. A., and Withers, S. G. (1996) *Biochemistry 35*, 13165-72.

152.   Trincone, A., Perugino, G., Rossi, M., and Moracci, M. (2000) *Bioorg Med Chem Lett 10*, 365-8.

153.   De Roode, B. M. (2001) PhD Thesis, pp. 165, *Wageningen University, Wageningen, The Netherlands*.

154.   Riva, S. (1999) pp 18, Instituto di Biocatalisi e Riconoscimento Molecolare, C.N.R., Milano.

155.   Fischer, L., Bromann, R., Kengen, S. W., de Vos, W. M., and Wagner, F. (1996) *Biotechnology N Y 14*, 88-91.

156.   De Roode, B. M., Van der Meer, T. D., Kaper, T., Franssen, M. C. R., Van der Padt, A., Van der Oost, J., Boom, R. M., and De Groot, A. (2001) *Enzyme Microb Tech*, in press.

157.   Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res 24*, 4876-82.

158.   Page, R. D. M. (1996) *Comput Appl Biosci 12*, 357-8.

159.   McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958-9966.

160.   Guex, N., Diemand, A., and Peitsch, M. C. (1999) *TiBS 24*, 364-7.

161.   Cason, C. (1999), The Persisitant of Vision Development Team, Indianapolis, IN, USA.

162.   Volkov, A. A., and Arnold, F. H. (2000) *Methods Enzymol 328*, 447-56.

2

Comparative structural analysis and substrate specificity engineering of the hyperthermostable β-glucosidase CelB from *Pyrococcus furiosus*

Authors
Thijs Kaper
Joyce H.G. Lebbink
Jeroen Pouwels
Jürgen Kopp
Georg E. Schulz
John van der Oost
Willem M. de Vos

Abstract

The substrate specificity of the β-glucosidase (CelB) from the hyperthermophilic archaeon Pyrococcus furiosus, a family 1 glycosyl hydrolase, has been studied at a molecular level. Following crystallization and X-ray diffraction of this enzyme, a 3.3Å resolution structural model has been obtained by molecular replacement. CelB shows a homo-tetramer configuration, with subunits having a typical $(\beta\alpha)_8$-barrel fold. Its active site has been compared to the one of the previously determined 6-phospho-β-glycosidase (LacG) from the mesophilic bacterium Lactococcus lactis. The overall design of the substrate binding pocket is very well conserved, with the exception of three residues that have been identified as a phosphate binding site in LacG. To verify the structural model and alter its substrate specificity, these three residues have been introduced at the corresponding positions in CelB (E417S, M424K, F426Y) in different combinations: single, double, and triple mutants. Characterization of the purified mutant CelB enzyme revealed that F426Y resulted in an increased affinity for galactosides, whereas M424K gave rise to a shifted pH optimum (from 5.0 to 6.0). Analysis of E417S revealed a 5-fold and a 3-fold increase of the efficiency of hydrolyzing o-nitrophenol-β-D-galactopyranoside-6-phosphate, in the single and triple mutants, respectively. In contrast, their activity on nonphosphorylated sugars was largely reduced (30-300-fold). The residue at position E417 in CelB seems to be the determining factor for the difference in substrate specificity between the two types of family 1 glycosidases.

2.1 Introduction

Glycosyl hydrolases are widely spread among living organisms. Members of this enzyme superfamily are involved in a broad range of physiological processes, such as food supply and storage, cell wall synthesis and degradation, defense systems, and signaling events (*1*). A large number of glycosyl hydrolases have been isolated and characterized, and the corresponding amino acid sequences have been determined. At the moment, comparison of these sequences has led to the identification of 77 families of glycosyl hydrolases (*2*). Family 1 consists of β-glycosidases, 6-phospho-β-glycosidases, lactase-phlorizin hydrolases, and myrosinases originating from all three domains of life: Bacteria, Archaea, and Eukarya. These enzymes hydrolyze their substrate with retention of configuration at the anomeric carbon atom, which occurs via a double displacement mechanism, in which a covalent intermediate is believed to be formed (*3-5*). Two glutamate residues serve as a general acid/base and a nucleophile in this reaction. Although family 1 glycohydrolases in general display a broad substrate specificity, the glycosidases have been optimized for the hydrolysis of nonphosphorylated glycosides and have hardly any activity on phosphorylated sugars, while the opposite is true for the family 1 6-P-glycosidases.

In the last years, the 3D structures have been solved of various family 1 glycosyl hydrolases from Eukarya (*6, 7*), Bacteria (*8-10*), and Archaea (*11, 12*). The enzymes display a common $(\beta\alpha)_8$-barrel motif, and, being members of the glycosyl hydrolase 4/7-superfamily, all have the two catalytic glutamates located at the C-terminal end of β-strands 4 and 7 (*13*). The involvement of the two glutamates in catalysis has been well established by mutagenesis and inhibitor studies (*14-18*). Recently, the 3D structure of an inactive mutant of the *Lactococcus lactis* 6-P-β-galactosidase LacG has been elucidated with a galactose-6-P bound in the active site (*10*).

Previously, a β-glucosidase has been purified from the hyperthermophilic archaeon *Pyrococcus furiosus* (*19*). The enzyme was active as a homo-tetramer with 58 kDa subunits and displayed a high thermostability, with a half-life of thermal inactivation of 85 h at 100 °C. The corresponding *celB* gene has been cloned and functionally expressed in *E. coli*, and its sequence characterization identified the *P. furiosus* β-glucosidase as a family 1 glycosyl hydrolase, in which Glu372 was identified as the nucleophilic catalytic residue by site-directed mutagenesis (*16*). A translational fusion of the *celB* gene to the T7 promoter of the pET9d vector, in which an alanine residue was inserted after the N-terminal methionine, allowed for stable overproduction in *E. coli* of the *P. furiosus* β-glucosidase (*20*). The activity, stability, and kinetic parameters of the resulting enzyme, referred to here as CelB, were comparable to those of the β-glucosidase isolated from *P. furiosus* (*21*). CelB serves as a model system to study the substrate specificity of glycosyl hydrolase family 1, since it has been well characterized, and can be produced in *E. coli* in wild-type and mutant forms and displays extreme stability, broad specificity, and high catalytic activity (*19, 20, 22-24*). In addition, CelB has developed into a model enzyme that may have wide applications, such

as disaccharide hydrolysis and synthesis of glycoconjugates and oligosaccharide production (*25-27*).

Here, we report the investigation and mutagenesis of the active site of CelB. Using X-ray diffraction data, a 3D model of the *P. furiosus* β-glucosidase was built by the molecular replacement method, in which the structures of *L. lactis* 6-P-β-galactosidase LacG (*8*) and *Sulfolobus solfataricus* β-glycosidase LacS (*11*) served as search models. Based on the structural model, we designed three substitutions at three positions to adapt the enzyme to the hydrolysis of phosphorylated galactosides. CelB mutants have been purified, characterized, and compared to the wild-type enzyme. The results show that a single substitution is mainly responsible for a significant increase of the catalytic activity on phosphorylated galactose.

2.2 Materials and methods

Strains, Plasmids, and Chemicals. Escherichia coli BL21-(DE3) *(hsdS gal λcIts* 857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene 1) or *E. coli* MC1061 was used for heterologous expression; *E. coli* TG1 *(supE hsdΔ5 thi Δ(lac-proAB)* F'[*traD36 proAB⁺ lacI^q lacZΔM15*]) was used in the construction of derivatives of pET9d (T7 promoter, kan^R) (Novagen) following established procedures (*28*). Plasmid pLUW511 is a pET9d derivative carrying the *P. furiosus celB* gene (Gen-Bank accession no. AF013169), derived from pLUW510 (*16*), translationally fused to the bacteriophage T7 φ10 gene start codon (*20*). All chemicals used were of analytical grade. O-Nitrophenyl-β-D-glucopyranoside (oNp-Glc), o-nitrophenyl-β-D-galactopyranoside (oNp-Gal), and o-nitrophenyl-β-D-galactopyranoside-6-phosphate (oNp-Gal-6-P) were purchased from Sigma (Zwijndrecht, The Netherlands). *Pfu* DNA polymerase was obtained from Stratagene.

Construction of Mutants. Mutations were introduced in *celB* using *Pfu* polymerase in the PCR-based overlap extension method (*29*). For each mutation, a sense/antisense primer pair was designed. BG266 (5'-AACAGACAACTACTCCTG-3')/BG267 (5'-GGGCCCAGGAGTAGT TGTCTGTT-3') intro-duced mutation E417S (introduced mutations in boldface, introduced restriction sites underlined, and restriction enzymes in parentheses). BG268 (5'-CAAGGGTTCAGGAAAAGATACGGATTGG TTT-3')/BG269 (5'-AAACCAATCCGTA TCTTTCCTGAACCCTTG-3') introduced M424K/F426Y. BG333 (5'-GGGTTCAGGAAA AGATTCGGATTGGTT-3')/BG334 (5'-AACCAATCCGAATCTTTCCTGAACCC-3')(*Xmn*I) introduced M424K. F426Y was introduced by BG335 (5'-GGTTCAGGATGAGGTACGG ATTGGTT-3')/BG336 (5'-AACCAATCCGTACCTCATCCTGAACC-3')(*Rsa*I). In each case the sense primer BG238 (5'-GCGCGC*CAT*GGCAAAGTTCCCAAAAAACTTCATGT TTG-3', *celB* start codon in italics) *(Nco*I) and antisense primer BG239 (5'-CGCGCGGATC*CCC TA*CTTTCTTGTAACAAATTTGAGG-3', *celB* stop codon in italics) *(Bam*HI) were used as flanking primers for the amplification of the *celB* gene. The resulting PCR products were digested with *Nco*I and *Bam*HI and ligated in *Nco*I/*Bam*HI-digested pET9d. Plasmid pLUW511 was used as

a template in PCR reactions for the construction of pLUW513 (E417S), 515 (M424K), and 522 (F426Y). Plasmid pLUW513 was used as a template in the construction of pLUW519 (E417S/M424K/F426Y). Plasmid pLUW518 (M424K/F426Y) was constructed by replacing a pLUW519 *Pst*I-*Afl*II fragment with a pLUW511 *Pst*I-*Afl*II fragment. Mutations were verified by DNA sequence analysis using the Thermosequenase cycle sequencing kit with infrared dye labeled primers. Reactions were analyzed on a Licor 4000L automated sequencer (Licor, Lincoln, NE) (data not shown).

Enzyme production and purification. CelB was routinely produced in *E. coli* BL21(DE3) harboring pET9d derivatives and purified from heat-incubated lysates following anion exchange and gel filtration chromatography (20). This resulted in pure CelB, as was confirmed by SDS-PAGE 1 electrophoresis, and was stored at 4 °C with 0.02% sodium azide. Protein concentrations were determined spectropho-tometrically at 280 nm with a molar extinction coefficient ε_{280} of 128 280 M^{-1} cm^{-1}, calculated according to Gill and Von Hippel (*30*).

Enzyme preparation, crystallization, and data collection. P. furiosus β-glucosidase was purified from *E. coli* MC1061 cells harboring pLUW510 as essentially described previously except for the heat incubation of 15 min at 60 °C (*16*). Initial crystallization conditions were determined after screening according to Jancarik and Kim (*31*). After fine-tuning, the enzyme was found to crystallize optimally by vapor diffusion using the hanging-drop method at 20 °C. Drops of 10 μL contained 10 mg/ml protein, 8% PEG-400, 150 mM CaCl2, and 10 mM Tris, pH 8.0, and the reservoir contained 20% PEG-400. Within a few days, crystals grew to an average size of 600 x 600 x 600 μm^3. The crystals diffracted merely to medium resolution. A native data set to 3.3 Å could be collected at RT with a multiwire area detector (X1000, Nicolet/Siemens) on a rotating anode with exposure times up to 240 s per frame. The space group was I_132 with a unit cell of *a* = *b* = *c* = 219.3 Å. The structures of 6-phospho-β-galactosidase LacG of *L. lactis* (PDB entry 1PBG) (8) and β-glycosidase LacS of *S. solfataricus* (PDB entry 1GOW) (*11*) were used as search models for molecular replacement and subsequent construction of the structural model of *P. furiosus* β-glucosidase. Solutions for *L. lactis* LacG and *S. solfataricus* LacS found with the programs AMORE (*32*) and AUTOAMORE (*32*), respectively, were consistent with each other. A subunit of LacS was used for an initial rigid body refinement with X-plor (*33*). After introduction of all CelB residues, the structure was refined with X-plor (*33*) using bulk solvent correction. The R-factor dropped to 32.5% (Rfree = 39.0%), and the common B-factor was 31 Å 2 . All residues were in the energetically most favorable or allowed regions of the Ramachandran plot of backbone dihedral angles.

Kinetic analyses. (A) pH optima. The pH optima of wild-type CelB and CelB mutants were determined at 90 °C in combined citrate/phosphate buffer (80 mM/80 mM) in a pH range of 3.5-8.0 (set at 25 °C). As substrates, oNp-Glc, oNp-Gal, and oNp-Gal-6-P were used. In the case of oNp-Gal-6P, reaction vials (1.5 ml) containing 49 μl of 40 mM oNp-Gal-6-P in buffer were preincubated

in a water bath at 90 °C for 2 min. The reaction was started by the addition of 1 µl of enzyme solution to the vials. After 5 min, the reaction was stopped by the addition of 100 µl of ice-cold 0.5 M Na2CO3, which terminated the reaction and augmented the molar extinction coefficient of the released nitrophenol; 125 µl samples were transferred to a 96-well microtiter plate, and the exinction of the reaction mixtures as determined at 405 nm in a microtiter plate reader (SLT Lab Instruments 340 ATTC). Maximum extinction was set at 100%. When oNp-Glc and oNp-Gal were used as substrates, reaction vials containing 195 µl of 10 mM substrate in buffer were treated as for oNp-Gal-6-P; only the reactions were initiated by the addition of 5 µl of enzyme and stopped with 0.4 ml of carbonate, and 150 µl samples were transferred to a microtiter plate for analysis.

(B) K_M and V_{max} determinations. The K_M and V_{max} of wild-type CelB and CelB mutants were measured for the hydrolysis of the three substrates at 90 °C in 100 mM citrate/phosphate buffer. We followed the release of nitrophenol at 405 nm (pH 5.0) or 470 nm (pH 6.0) in quartz cuvettes using a spectrophotometer equipped with a temperature controller (Hitachi, San Jose, CA). For each substrate, the activities were measured at differerent concentrations (oNp-Glc: 0.05-50 mM; oNp-Gal: 0.10-50 mM; oNp-Gal-6-P: 0.4-24 mM). The reaction volume was 1.00 ml for oNp-Glc and oNp-Gal and 0.50 ml for oNp-Gal-6-P. In each case, an enzyme volume of 5.0 µl was added to start the reaction. Molar extinction coefficients of o-nitrophenol at 90 °C in assay mixtures were the following: $\varepsilon_{405;pH5.0}$ 0.6243 cm^{-1}mM^{-1}; $\varepsilon_{470;pH6.0}$, 0.6339 cm^{-1}mM^{-1}; $\varepsilon_{405,pH4.5}$, 0.5614 cm^{-1}mM^{-1}; $\varepsilon_{405,pH4.0}$, 0.5398 cm^{-1}mM^{-1} . K_M and V_{max} values were calculated by fitting the activities according to Michaelis-Menten kinetics using the nonlinear regression program TableCurve (Jandel Scientific, AISN Software).

2.3 Results

CelB structure. The *P. furiosus* β-glucosidase CelB is the most thermostable and thermoactive representative of family 1 glycosyl hydrolases and shows a broad substrate specificity. To characterize its structure-function relation, the *P. furiosus* β-glucosidase was purified from an overproducing *E. coli* strain to homogeneity, crystallized by the hanging-drop method, and subjected to X-ray diffraction. A dataset was collected at 3 Å resolution, which allowed for the analysis of its tertiary structure and active site. The tetrameric structure of the enzyme was established in the molecular replacement procedure (Figure 2.1). The search models in this procedure were the 3D structures of the *L. lactis* 6-phospho-β-galactosidase LacG and the *S. solfataricus* β-glycosidase LacS, which share 16% and 53% amino acid identity with the pyrococcal enzyme, respectively. Attempts were made to obtain more regular crystals from CelB variants with altered surface residues, but crystals that diffracted at a higher resolution were not obtained (data not shown). Therefore, the initial structural model was used in the analysis of the CelB tertiary structure, the active site, and for the design of specific substitutions.
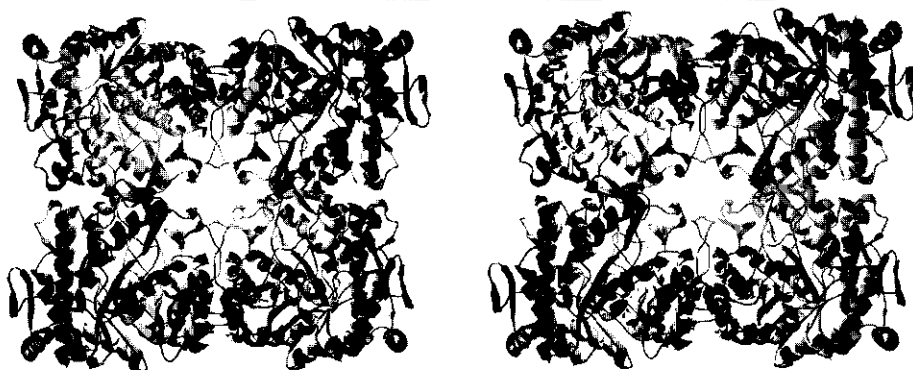
Figure 2.1 Ribbon model of *P. furiosus* β-glucosidase, viewed along one of the 2-fold axes of the tetramer.

The CelB homo-tetramer has point symmetry 222, with one subunit in the crystallographically asymmetric unit. The enzyme consists of two dimers, of which the two barrels are placed next to each other. In the final tetramer, two dimers appear to be placed on top of each other with the barrel openings facing each other (Figure 2.1). The overall size of the tetramer is 110 Å x 90 Å x 55 Å, and its appearance is that of a slightly twisted square. The axes of the $(\beta\alpha)_8$-barrels are almost parallel to one of the tetramer axes (Figure 2.1). Each subunit of CelB consists of a single domain of 472 amino acids, with 18 helices and 16 β-strands. The center of the monomer is formed by an $(\beta\alpha)_8$-barrel, with the catalytic residues Glu207 and Glu372 at the C-terminal end of the 4th and 7th β-strand of the barrel, respectively.

Architecture of the CelB Active Site. The positions of the active site residues in LacG and CelB are very well conserved, as illustrated by a superposition (Figure 2.2). Asn17, Arg77, His150, Asn206, Tyr307, and Trp410 are conserved residues and likely to be involved in substrate binding (*10*). Glu207 and Glu372 in CelB are the equivalents of the catalytic glutamate residues in LacG. The importance of Glu372 in catalysis at 100 °C has been shown by site-directed mutagenesis (*16*). The average distance between the oxygen atoms of these glutamate carboxylic groups is 4.3 Å (± 1Å) in CelB, which is in the range of the general observed distance in retaining glycosyl hydrolases (~ 5 Å)(*34*). The active site is located in the center of a subunit and can be reached from the outside by a 20 Å long channel, which is located between the fifth and sixth βα-motives and is opened by a 50° bending of the helix of the fifth βα-unit.
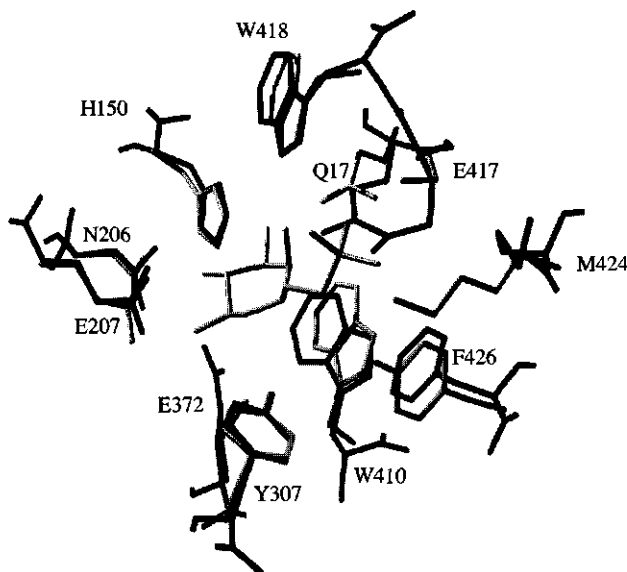
Figure 2.2 Superposition of *P. furiosus* CelB (black) with *L. lactis* LacG (gray) that has galactose-6-phosphate bound in the active site. The residues that are involved in the glucose or galactose binding in both enzymes have been represented. The residues have been indicated according to CelB numbering. The calculated RMS deviation of equivalent atoms of these residues is about 1 Å, which corresponds to the expected accuracy of the CelB model. The general superposition was based on the backbone atoms of both proteins and was performed using the Swiss PDB Viewer (*42*). The figure was generated using the program POV-Ray (Hallam Oaks Pty. Ltd.).

Design and production of wild-type and mutant P. furiosus CelB. In 6-P-β-galactosidase LacG of *Lactococcus lactis*, Ser428, Lys435, and Tyr437 form a binding site for the phosphate group of galactose-6-phosphate (*10*). These three residues are highly conserved among family 1 6-P-β-galactosidases (Figure 2.3). In the *P. furiosus* CelB, the corresponding residues at these positions are Glu417, Met424, and Phe426, residues which, in turn, are highly conserved among family 1 β-glycosidases that hydrolyze nonphosphorylated substrates. To study their contribution to substrate specificity, the residues that create the phosphate-binding site in LacG were introduced in CelB by a PCR-based method. This resulted in three CelB single mutants (E417S, M424K, and F426Y), one double mutant (M424K/F426Y), and one triple mutant (E417S/M424K/F426Y). The wild-type and mutated *celB* genes were overexpressed in *E. coli*, all resulting in significant overproduction of active enzyme, which amounted up to more than 20% of the total cell protein. Subsequently, the mutant CelB enzymes were purified to apparent homogeneity as judged by SDS-PAGE and UV spectroscopy (not shown) and used for further characterization and comparison with the wild-type CelB.

pH Optima of mutant and wild-type CelB. Wild-type CelB hydrolyzes the chromogenic substrate pNp-Glc optimally at pH 5.0 *(19)*. The glycosidic bond is cleaved by a common double displacement mechanism *(24)*. The pH optimum for this mechanism has been proposed to be determined by the pK_a values of the two catalytic residues *(35)*. Since the substitutions E417S and M424K introduce charge differences in the active site, these might influence the pK_a values of the catalytic glutamate residues and, as such, change the pH optimum for hydrolysis. Therefore, the pH optima for the hydrolysis of o-nitrophenol-β-D-glucopyranoside (oNp-Glc), o-nitrophenol-β-D-galactopyranoside (oNp-Gal), and o-nitrophenol-β-D-galactopyranoside-6-phosphate (oNp-Gal-6-P) were determined for different mutants (Table 2.1). The stability of the mutants at the different pH values was found to be equal to that of wild-type CelB, with the exception of CelB M424K and CelB M424K/F426Y, which were more sensitive to inactivation below pH 4.5 (data not shown).

The pH optima of wild-type and mutant CelB enzymes for the hydrolysis of oNp-Glc and oNp-Gal were essentially the same (Table 2.1). However, large differences were observed in the shape of the pH curves. Three groups could be distinguished: (i) wild-type CelB, M424K, and

```
Pfu CelB   NAYELP--MIITENGMADAADR----------YRPHYLVSHLKA   : 393
Sso LacS   R-YHLY--MYVTENGIADDADY----------QRPYYLVSHVYQ   : 408

Lla LacG   D-FPNYKKIYITENGLGYKDEFVDN--TVYDDGRIDYVKQHLEV   : 405
Sau LacG   D-FPNYHKIYITENGLGYKDEFIESEKTVHDDARIDYVRQHLNV   : 407
Eco BglB   R-YQ--KPLFIVENGLGAKDS-VEADGSIQDDYRIAYLNDHLVQ   : 391
                        #

Pfu CelB   VYNAMKEGADVRGYLHWSLTDNYEWAQG-FRMRFGLVYVDFETK   : 436
Sso LacS   VHRAINSGADVRGYLHWSLADNYEWASG-FSMRFGLLKVDYNTK   : 451
Lla LacG   LSDAIADGANVKGYFIWSLMDVFRWSNG-YERRGLFYVDFDTQ   : 448
Sau LacG   IADAIIDGANVKGYFIWSLMDVFRWSNG-YERRGLFYVDFETQ   : 450
Eco BglB   VNEAIADGVDIMGYTSWGPIDLVRASHSQMSRRFIYVDRDDN   : 435
                                      *        *  *

Pfu CelB   K-----RYLRPSALVFREIATQKEIPEELAHLADLKFVT--RK.   : 472
Sso LacS   R-----LYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH.   : 489

Lla LacG   E-----RYPKKSAHWYKKLAETQVIE.                    : 468
Sau LacG   E-----RYPKKSAYWYKELAETKEIK.                    : 470
Eco BglB   GEGSLTRTRKKSFGWYAEVIKTRGLSLKKITIKAP.           : 470
```

Figure 2.3 Alignment of partial amino acid sequences. *Pfu* CelB, *P. furiosus* β-glucosidase CelB (Swiss Prot Q51763); *Sso* LacS, *S. solfataricus* β-glycosidase LacS (P22498); *Lla* LacG, *L. lactis* 6-phospho-β-galactosidase LacG (P11546); *Sau* LacG, *Staphylococcus aureus* 6-phospho-β-galactosidase LacG (P11175); *Eco* BglB, *Escherichia coli* 6-phospho-β-glucosidase BglB (P11988). Conserved residues in all proteins have been shaded light gray; residues that interact with the phosphate group of galactose-6-phosphate in *L. lactis* LacG have been indicated (*) and shaded black in the 6-phospho-β-glycosidases and dark gray at the corresponding positions in the β-glycosidases. The conserved catalytic nucleophile in all enzymes has been indicated (#).

Table 2.1 pH Optima of wild-type CelB and CelB mutants for hydrolysis at 90 °C in 0.1 M Sodium citrate/NaPi

	oNp-Glc	oNp-Gal	oNp-Gal-6-P
CelB wt	5.0	5.0	4.0
CelB E417S	4.5	4.5	5.0
CelB M424K	6.0	5.75	4.5
CelB F426Y	5.0	5.0	4.0
CelB double mutant	6.0	6.25	4.5
CelB triple mutant	5.0	5.0	6.0

F426Y that showed an optimum activity at a broad pH range for oNp-Glc and oNp-Gal, but a narrow range for oNp-Gal-6-P with a lower optimal pH than with oNp-Glc and oNp-Gal as substrates (Figure 2.4A,B); (ii) E417S and the triple mutant that showed a pH dependency of activity for the three substrates that was opposite of that of wild-type CelB (Figure 2.4C,D); (iii) the double mutant M424K/F426Y hydrolyzed both oNp-Glc and oNp-Gal as well as oNp-Gal-6-P optimally at a relatively wide pH range, the optimal pH for hydrolysis with oNp-Gal-6-P being lower than with oNp-Glc and oNp-Gal as substrates (Figure 2.4C,D). The mutation F426Y did not have any effect on the pH optimum of hydrolysis, compared to the wild type. However, M424K hydrolyzed oNp-Glc and oNp-Gal optimally at pH 6.0, 1 unit higher than the wild-type. This mutant also showed the fastest hydrolysis of oNp-Gal-6-P at a higher pH than the wild-type.

Kinetic parameters of mutant and wild-type CelB. The kinetic constants at optimal pH for the hydrolysis of oNp-Glc and oNp-Gal were determined (Table 2.2), as well as those for the hydrolysis of oNp-Gal-6-P (Table 2.3). The wild-type CelB shows high affinity for oNp-Glc (K_M value of 0.28 mM) and a lower affinity for oNp-Gal (K_M value of 2.3 mM). All introduced mutations decreased the affinity for oNp-Glc compared to the wild type. Moreover, the turnover number decreased for most mutants, resulting in lower efficiencies for the hydrolysis of oNp-Glc. The affinity for oNp-Gal decreased in E417S, M424K, and the triple mutant, was unaffected in the double mutant, but was significantly increased in F426Y. This made F426Y slightly more efficient on oNp-Gal than the wild type. The K_M values for oNp-Glc and oNp-Gal in the mutants that contain the mutation E417S (E417S and the triple mutant) increased 20-fold to 200-fold, but due to the limited solubility of oNp-Glc and oNp-Gal, it was not possible to perform assays at saturating conditions for these mutants.

Although wild-type CelB is able to hydrolyze oNp-Gal-6-P, its efficiency is low with a K_M of 31 mM and a k_{cat} of 49 s^{-1} under optimal conditions (Table 2.3). Remarkably, mutant F426Y has a slightly higher KM value and a reduced k_{cat} value with respect to wild-type CelB, making it even less efficient in hydrolysis of the phosphorylated substrate. The CelB mutants that contain the substitution E417S had lost their affinity for nonphosphorylated glycosides, but these variants (mutant E417S and the triple mutant) have increased affinities and turnover numbers for the hydro-
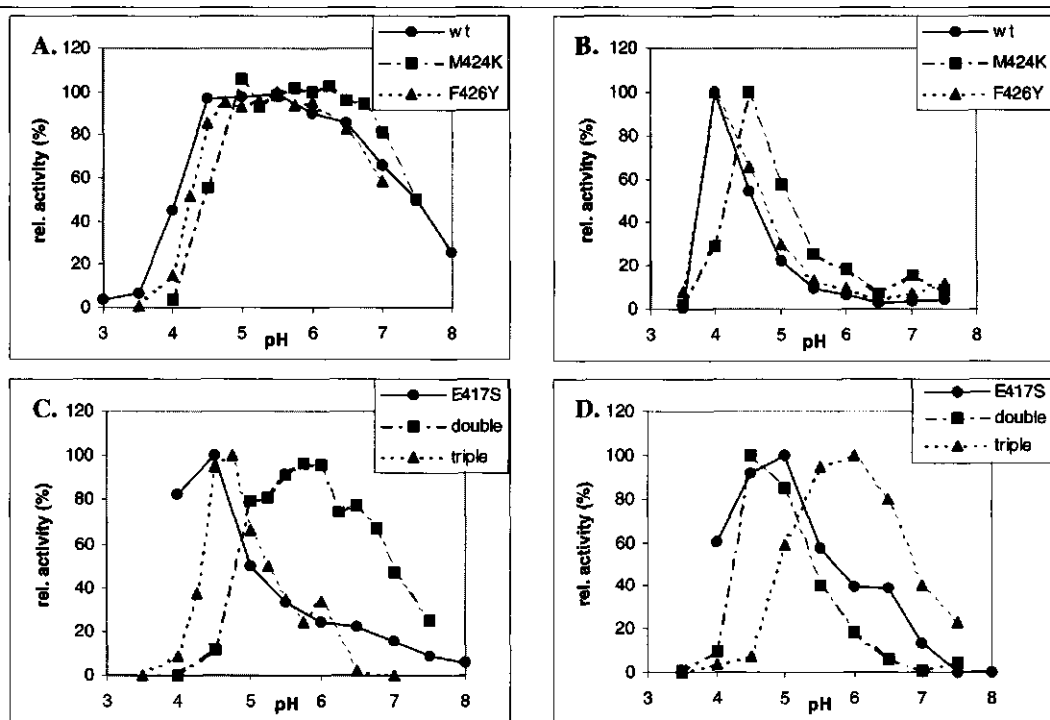
Figure 2.4 Typical effects of pH on enzyme activity at 90 °C in combined citrate-phosphate buffer. Assays were performed with 8.3 ng to 3.3 μg of enzyme. Maximum activity in each case was set at 100%. (A) Activity on oNp-Glc of wild-type CelB (max. 920 units/mg), CelB M424K (max. 805 units/mg), and CelB F426Y (max. 910 units/mg); (B) activity on oNp-Gal-6-P of wild-type CelB (max. 31 units/mg), CelB M424K (max. 5 units/mg), and CelB F426Y (max. 14 units/mg); (C) activity on oNp-Glc of CelB E417S (max. 29 units/mg), CelB double mutant (max. 1240 units/mg), and CelB triple mutant (max. 124 units/mg); (D) activity on oNp-Gal-6-P of CelB E417S (max. 89 units/mg), CelB double mutant (max. 27 units/mg), and CelB triple mutant (max. 54 units/mg). 1 Unit corresponds to the release of 1 μmol oNp per minute.

lysis of oNp-Gal-6-P. As a consequence, E417S and the triple mutant are 4.9-fold and 2.7-fold more efficient, respectively, in hydrolyzing oNp-Gal-6-P than wild-type CelB.

2.4 Discussion

CelB structural model and active site. A structural model of *P. furiosus* β-glucosidase with a resolution of 3.3 Å has been constructed based on X-ray diffraction of its crystals, followed by molecular replacement of the 3D structures of *L. lactis* LacG and *S. solfataricus* LacS. Although this resolution did not allow a detailed analysis of the entire protein, the structure of the active site was analyzed in more detail, since it could be established with a high accuracy (Figures 2.1 and 2.2) and confirms earlier predictions that CelB is a retaining enzyme (*22, 24, 25*).

The central fold of CelB is a (βα)$_8$-barrel (Figure 2.1), first found in chicken triosephosphate isomerase (*36*). This fold has been predicted for all family 1 glycosyl hydrolases and has indeed

been observed for all established structures of family 1 glycosyl hydrolases (*6-8, 10, 11*). A tetrameric subunit organization has been observed for all hyperthermostable representatives of family 1 β-glycosidases, including *Thermosphaera aggregans* and *Sulfolobus solfataricus* as well (*11, 12*). This oligomerization may contribute to stability, since mesophilic and thermophilic family 1 enzymes are mainly active as monomers or dimers (*6, 8*). However, higher degrees of oligomerization have been observed (*9*). The positions of the active site residues in CelB, LacG, and LacS are highly conserved (Figure 2.2). The glutamate residues are located at the C-terminal ends of the fourth and seventh β-strands of the barrel and identify the enzymes as a member of the glycosyl hydrolase 4/7-superfamily (*13*). The opening at the entrance of the substrate channel toward the active site, which results from a bend in the helix of the fifth βα-unit, is similar to the one found in *S. solfataricus* LacS. However, the unusual amino acid sequence, WWFF, at the corresponding positions (287-290) in *S. solfataricus* LacS is not present in CelB.

The family 1 6-P-β-galactosidases have evolved a binding site for the phosphate group of galactose-6-phosphate. In *L. lactis* LacG, Ser428, Lys435, and Tyr437 were found to interact with the phosphate group (*10*). The corresponding residues in CelB (Glu417, Met424, Phe426) are highly conserved in the β-glycosidases (Figure 2.3). The negatively charged Glu417 would have a repulsive action on the phosphate group of an incoming 6-phosphoglycoside. The mutations E417S, M424K, and F426Y were therefore introduced to create a phosphate-binding site in CelB that may affect the kinetics of hydrolysis of 6-phosphoglycosides.

Characterization of mutants. The structure of the active site confirms that CelB hydrolyzes its substrates by a retaining mechanism in which a covalent intermediate is formed (*24*). In the catalysis cycle, the C1 atom of an incoming glycoside is attacked by the nucleophilic glutamate (Glu372 in CelB). The glycosidic bond is broken, and a covalent enzyme-substrate intermediate is formed. An incoming water molecule hydrolyzes the covalent bond, and the enzyme returns to the initial state (*37*). In wild-type CelB, the hydrolysis of the glycosyl-enzyme intermediate is the rate-

Table 2.2 Kinetic constants of wild-type and mutant CelB with oNp-Glc and oNp-Gal as substrates at 90 °C in 0.1 M sodium citrate/NaPi at determined pH optimum[a]

	oNp-Glc			oNp-Gal		
	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M ($mM^{-1}s^{-1}$)
CelB wt	0.28 ± 0.03	860 ± 25	4048	2.3 ± 0.4	2050 ± 86	899
CelB E417S	29 ± 10	3300 ± 630	113	42 ± 15	1060 ± 250	25.3
CelB M424K	0.51 ± 0.12	772 ± 47	1510	4.8 ± 0.3	1880 ± 50	396
CelB F426Y	0.57 ± 0.09	840 ± 37	1484	0.89 ± 0.05	892 ± 13	1000
CelB double mutant	3.0 ± 0.5	1400 ± 89	422	3.0 ± 0.6	969 ± 65	325
CelB triple mutant	59 ± 7	777 ± 65	13	40 ± 8	175 ± 21	4.4

[a] Amounts of 83 ng of CelB wt, 0.11 μg of CelB E417S, 0.13 μg of CelB M424K, 0.23 μg of CelB F426Y, 0.14 μg of CelB double mutant, and 1.3 μg of CelB triple mutant were used in the assays.

Table 2.3 Kinetic constants of wild-type and mutant CelB at 90 °C with oNp-Gal-6-P as substrate in 0.1 M sodium citrate/NaPi at pH optimum [a]

	oNp-Gal-6-P		
	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})
CelB wt	31 ± 6	49 ± 6	1.6
CelB E417S	22 ± 2	171 ± 11	7.7
CelB M424K	-	(11)[b]	-
CelB F426Y	38 ± 10	26 ± 5	0.7
CelB double mutant	-	(58)[b]	-
CelB triple mutant	16 ± 2	69 ± 5	4.2
LacG[c]	0.4	83	208

[a] Amounts of 3.3 íg of CelB wt, 2.1 μg of CelB E417S, 26 μg of CelB M424K, 47 μg of CelB F426Y, 2.9 μg of CelB double mutant, and 3.8 μg of CelB triple mutant were used in the assays. [b] Data could not be fitted according to Michaelis-Menten kinetics; highest measured kcat shown. [c] LacG activity measured at 25 °C; data obtained from (*41*).

limiting step in the reaction with oNp-Glc. The rate is dependent on the ionization state of the general acid/base catalytic residue that abstracts a proton of an incoming water, which, in turn, attacks the covalent bond between the enzyme and the sugar moiety (*3*).

The substitutions M424K and F426Y had relatively small effects on the hydrolysis of oNp-Glc and oNp-Gal. The pH profile of F426Y was identical to wild-type CelB, as was expected since this substitution is charge-neutral. The substitution M424K did introduce a positive charge into the active site, and in M424K and the double mutant M424K/F426Y, the pH optimum had shifted to a higher value (Figure 2.4). Probably, the introduced positive charge at this position results in an increase of the pK_a value of the nucleophilic glutamate, as can be deduced from the shape of the curve at low pH values (*35*). The distance between Glu372 and Lys424 is too large (~15 Å) for a direct interaction, but there could be an influence of the charge or size on the local conformation of the enzyme. Residue Lys424 is also present at the corresponding position in the wild-type β-glucosidases from *Thermotoga maritima* and *Caldocellum saccharolyticum*. Remarkably, the pH optimum of CelB M424K was similar to the one determined for the β-glucosidase of *T. maritima* (*38*). In general, the substitutions M424K and F426Y caused a decrease in catalytic efficiencies on oNp-Glc and oNp-Gal, except for F426Y on oNp-Gal. For both substrates, the double mutant showed the combined effects of the single mutants, an indication for the proper folding of the mutant's active sites (Table 2.2).

Substitution E417S influenced the catalytic properties of CelB significantly. Although wild-type CelB shows the fastest hydrolysis of oNp-Glc and oNp-Gal at pH 5.0, it retains 90% of its activity between pH 4.5 and 6.0. E417S and the triple mutant showed narrow curves with sharp

optima at pH 4.5 and 5.0, respectively (Figure 2.4). It is likely that Glu417, like Lys424, is too far away from the catalytic glutamates Glu207 and Glu372 for a direct interaction.

The glutamate residue in *Bacillus polymyxa* β-glucosidase BglA that corresponds to Glu417 in CelB was found to interact with the hydroxyl group of C4 and C6 of a gluconate molecule bound in the active site. Since glucose and galactose are identical molecules except for the position of the hydroxyl group on the C4 position, equatorial in glucose and axial in galactose, it is believed that the position of this glutamate residue is partly responsible for the difference in substrate specificity between a glucosidase and a galactosidase (*9*). Furthermore, Glu417 is important for the stabilization of the glycosyl-enzyme intermediate in the hydrolysis cycle (*39*). The effects of the disturbed interaction with the C4 hydroxyl group become clear from the kinetic parameters on oNp-Glc and oNp-Gal in E417S and the triple mutant. The K_M values for oNp-Glc and oNp-Gal were found to increase 100-fold and 20-fold, respectively. In *S. solfataricus* LacS, Glu432, corresponding to residue Glu417 in CelB, was postulated to interfere with the carboxylate group of substrates such as galuronic acid or glucuronic acid (*11*). However, the replacement of Glu417 by a serine residue did not enable CelB to hydrolyze p-nitrophenol-β-D-glucuronate (data not shown).

The substitutions M424K and F426Y did not improve the hydrolysis of oNp-Gal-6-P. M424K increased the optimum pH for catalysis to pH 4.5, but, like F426Y, did not result in a higher catalytic efficiency. Substitution E417S has the largest effect on the hydrolysis of phosphorylated galactosides. On one hand, this substitution enables the hydrolysis of oNp-Gal-6-P at a higher pH value compared to oNp-Glc or oNp-Gal. On the other hand, it lowers the Michaelis constant for the phosphorylated substrate and increases the k_{cat}. As a consequence, mutant E417S and the triple mutant have 4.9 and 2.7 times improved efficiencies for the hydrolysis of phosphorylated galactose, respectively. The combination of the three mutations in CelB triple results in the lowest Km, and the highest pH optimum. Apparently, the lysine contributes to the pH optimum for the hydrolysis of phosphorylated galactose, as can be concluded from the difference in the pH optima of CelB E417S and the CelB triple mutant with oNp-Gal-6-P as a substrate.

When compared to the 6-phospho-β-galactosidase LacG, there is a striking difference in K_M values (Table 2.3). At 90 °C, the Michaelis constant K_M of the triple mutant is 40-fold higher than the K_M of LacG for oNp-Gal-6-P at 25 °C. This could be due to a difference in temperature, although this is not predictable, as has been found in other studies (*40*). In addition to the substitution of residues in the active site, additional replacements of second shell residues are probably needed to further optimize CelB for the hydrolysis of phosphorylated galactosides. However, it can be concluded that the residue at position 417 in CelB has a major determining role in the specificity of family 1 glycosyl hydrolases. A glutamate at this position will make the enzyme efficient in the hydrolysis of β-glycosides, while a serine at this position is mainly responsible for 6-phospho-β-galactosidase activity.

Based on a comparison of the 3D structure of *L. lactis* 6-phospho-β-galactosidase LacG and the model of the *P. furiosus* β-glucosidase, three substitutions were introduced in the active site of CelB, with the aim to increase its activity toward phosphorylated galactose. The roles of the three residues in hydrolysis have been evaluated and were found to be of importance for the enzyme's substrate affinity, pH-optimum, and hydrolysis rate. In conclusion, a molecular basis has been provided for the fundamental difference in substrate specificity between members of family 1 glycosyl hydrolases that hydrolyze specifically phosphorylated and nonphosphorylated substrates. Moreover, the CelB structural model has been verified and may provide a basis for further understanding of catalysis and stability at extreme high temperatures.

2.5 Acknowledgments

References

1. Davies, G., and Henrissat, B. (1995) *Structure* 3, 853-859.
2. Henrissat, B. (1999) http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html.
3. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9969.
4. Sinnott, M. L. (1990) *Chem. Rev. 90*, 1171-1202.
5. Wang, Q., Trimbur, D., Graham, R., Warren, R. A., and Withers, S. G. (1995) *Biochemistry 34*, 14554-14562.
6. Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995) *Structure 3*, 951-960.
7. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663-675.
8. Wiesmann, C., Beste, G., Hengstenberg, W., and Schulz, G. E. (1995) *Structure 3*, 961-968.
9. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J. Mol. Biol. 275*, 491-502.
10. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J. Mol. Biol. 269*, 851-860.
11. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J. Mol. Biol. 271*, 789-802.
12. Chi, Y. I., Martinez-Cruz, L. A., Jancarik, J., Swanson, R. V., Robertson, D. E., and Kim, S. H. (1999) *FEBS Lett. 445*, 375-383.
13. Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett. 362*, 281-285.
14. Street, I. P., Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9970-9978.
15. Trimbur, D. E., Warren, R. A., and Withers, S. G. (1992) *J. Biol. Chem. 267*, 10248-10251.
16. Voorhorst, W. G. B., Eggen, R. I. L., Luesink, E. J., and De Vos, W. M. (1995) *J. Bacteriol. 177*, 7105-11.
17. Moracci, M., Capalbo, L., Ciaramella, M., and Rossi, M. (1996) *Protein Eng. 9*, 1191-1195.
18. Febbraio, F., Barone, R., D'Auria, S., Rossi, M., Nucci, R., Piccialli, G., De Napoli, L., Orru, S., and Pucci, P. (1997) *Biochemistry 36*, 3068-3075.
19. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur. J. Biochem. 213*, 305-312.
20. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2000) *Methods Enzymol. 330*, 364-79.

21. Lebbink, J. H. G. (1999) Ph.D. Thesis, pp. 192, Wageningen University, Wageningen, The Netherlands.

22. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J. Bacteriol. 177*, 7105-7111.

23. Bauer, M. W., Bylina, E. J., Swanson, R. V., and Kelly, R. M. (1996) *J. Biol. Chem. 271*, 23749-23755.

24. Bauer, M. W., and Kelly, R. M. (1998) *Biochemistry 37*, 17170-17178.

25. Fischer, L., Bromann, R., Kengen, S. W., de Vos, W. M., and Wagner, F. (1996) *Biotechnology 14*, 88-91.

26. Boon, M. A., Van der Oost, J., De Vos, W. M., Jansen, A. E. M., and Van't Riet, K. (1998) *Appl. Biochem. Biotechnol. 75*, 269-278.

27. Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol. Bioeng. 64*, 322-332.

28. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) (Nolan, C., Ed.) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

29. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res. 16*, 7351-7367.

30. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319-326.

31. Jancarik, J., Scott, W. G., Milligan, D. L., Koshland, D. E., Jr., and Kim, S. H. (1991) *J. Mol. Biol. 221*, 31-34.

32. Navaza, J. (1994) *Acta Crystallogr. A50*, 157-163.

33. Brünger, A. T., Kuriyan, J., and Karplus, M. (1987) *Science 235*, 458-460.

34. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol. 4*, 885-892.

35. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958-9966.

36. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) *Nature 255*, 609-614.

37. White, A., and Rose, D. R. (1997) *Curr. Opin. Struct. Biol. 7*, 645-651.

38. Gabelsberger, J., Liebl, W., and Schleifer, K. (1993) *Appl. Microbiol. Biotechnol. 40*, 44-52.

39. Namchuk, M. N., and Withers, S. G. (1995) *Biochemistry 34*, 16194-16202.

40. Lebbink, J. H. G., Kaper, T., Bron, P., Van der Oost, J., and Vos, W. M. (2000) *Biochemistry 39*, 3656-3665.

41. Hengstenberg, W., Kohlbrecher, D., Witt, E., Kruse, R., Christiansen, I., Peters, D., Pogge von Strandman, R., Stadtler, P., Koch, B., and Kalbitzer, H. R. (1993) *FEMS Microbiol. Rev. 12*, 149-164.

42. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714-2723.

3

Substrate specificity engineering of a β-mannosidase and a β-glucosidase from *Pyrococcus* by exchange of unique active site residues

Authors
Thijs Kaper
Hester H.
   van Heusden
Bert van Loo
Andrea Vasella
John van der Oost
Willem M. de Vos

Abstract

A β-mannosidase gene (PH0501) was identified in the *Pyrococcus horikoshii* genome and cloned and expressed in *E. coli*. The purified enzyme (BglB) was most specific for the hydrolysis of *para*-nitrophenol-β-D-mannopyranoside (pNp-Man) (K_M: 0.44 mM) with a low turnover rate (k_{cat}: 4.3 s^{-1} at 90°C). The β-mannosidase was classified as a member of family 1 of glycosyl hydrolases with an apparent conservation of its active site region, which was confirmed by homology modeling. However, BglB has two unique active site residues, Gln77 and Asp206, which are an arginine and asparagine residues in all other known family 1 enzymes. The arginine and asparagine residues appear to interact with the catalytic nucleophile and equatorial C2-hydroxyl group of substrates, respectively. To study the role of these residues in substrate specificity and catalysis, the unique residues of *P. horikoshii* BglB were introduced in the highly active β-glucosidase CelB of *Pyrococcus furiosus* and *vice versa*, yielding 2 single and 1 double mutant for each enzyme. Analysis of the mutants confirmed the involvement of the introduced residues in substrate binding and catalysis. In CelB, both the R77Q and N206D substitutions resulted in an increased specificity for mannosides, coupled to 10-fold reduced hydrolysis rates. In the BglB Q77R mutant all substrate affinities and hydrolysis rates were reduced, most likely due to steric limitations to accommodate the introduced residue in the active site. In contrast, BglB D206N showed 10-fold increased hydrolysis rates and 35-fold increased affinity for the hydrolysis of glucosides. In combination with inhibitor studies it was concluded that the substituted residues participate in the ground-state binding of substrates with a equatorial C2-hydroxyl group, but contribute most to transition state stabilization. The unique activity profile of BglB seems to be caused by an altered interaction between the enzyme and C2-hydroxyl of the substrate and a specifically increased affinity for mannose that results from Asp206.

3.1 Introduction

Family 1 glycosidases are found in all domains of life and have been classified as β-glucosidases, β-galactosidases, 6-phospho-β-glycosidases, lactase-phlorizin hydrolases, β-mannosidases and myrosinases (*1*). While many members of this family have been demonstrated to be involved in energy supply or cell defense (*2-6*) not all characterized enzymes have clearly established metabolic roles (*7*). Family 1 enzymes have been shown to hydrolyze their substrates via a general acid/base catalyzed mechanism with overall retention of the configuration at the C1 atom of the non-reducing sugar residue (*8, 9*). In this reaction a covalent substrate-enzyme intermediate is formed, which has been trapped, allowing to establish 3D enzyme structures (*10, 11*). The transition state in the retaining hydrolysis is generally believed to have a substantial oxycarbenium-cation-like character, as supported by the high sensitivity of family 1 glycosidases for inhibition by glycoside derivatives with a trigonal geometry at the anomeric carbon atom, such as glucono-lactones (*8, 9, 12*). Two glutamate residues are essential for the hydrolysis (*13-15*). Their pK_a-value is tightly regulated and cycles during catalysis (*16*). Hydrogen bonds between the enzyme and substrate are crucial in the hydrolysis reaction and the interaction with the C2-hydroxyl group of the glycoside contributes most to the stabilization of the transition state (*17, 18*). All 3D structures revealed a $(\beta/\alpha)_8$-barrel as fold, although different quartenary organization forms have been observed, ranging from monomeric to octameric (*10, 19-24*). The enzyme-substrate interactions have been evaluated from enzymes that have been co-crystallized with ligands (*21, 25*). Nine residues in the active site pocket have been found to interact directly with the substrate, and appear to be highly conserved throughout family 1 enzymes (*21*). Only the 6-phospho-β-glycosidases make up a separate group, since they have developed a phosphate binding site that enables the hydrolysis of 6-phospho-glycosides (*24, 25*).

The β-glucosidase CelB the hyperthemophilic archaeon *Pyrococcus furiosus* represents the most thermostable family 1 enzyme described to date (*13, 26*). The enzyme is active as a tetramer and displays a high hydrolytic activity with a half life of 85 hours at 100°C (*26*). In agreement with its involvement in the degradation of glucans (*4, 26, 27*), CelB displays the highest hydrolytic efficiency on glucosides and fucosides, but the enzyme is also capable of hydrolyzing galactosides and, to a lower extent, mannosides as well (*26, 28*). The mechanism of glycoside hydrolysis by *P. furiosus* CelB at extremely high temperatures was found to be similar to that of the related mesophilic β-glucosidase from *Agrobacterium faecalis* (*28*). CelB has been crystallized and a structural model of the enzyme was constructed based on a 3.3 Å resolution X-diffraction data set (*24*). Analysis of CelB's active site revealed a high structural similarity to that of the previously determined 6-phospho-β-galactosidase LacG of *Lactococcus lactis* (*24*).

In the genome sequence of the hyperthermophilic archaeon *Pyrococcus horikoshii* two open reading frames have been annotated as family 1 glycosyl hydrolases (*29, 30*). PH0366 encodes a membrane-associated protein with high efficiency for the hydrolysis of long-chain alkyl-glycosides

(31). The hydrolase encoded by PH0501 has been annotated as a β-mannosidase *(29)* based on amino acid homology with the characterized β-mannosidase BmnA from *P. furiosus (7)*. However, this has not yet been verified experimentally.

This study describes the cloning and expression of PH0501, and the characterization of its gene product, the β-mannosidase BglB. This new thermostable member of glycosyl hydrolase family 1 has a low hydrolytic activity. Multiple sequence alignment and 3D modeling reveal that BglB represents a new subfamily of family 1 glycosyl hydrolases, characterized by two unique active-site residues. Their role was studied by substituting them for the corresponding residues of the well-characterized β-glucosidase CelB of *P. furiosus* by site-directed mutagenesis, and *vice versa*. The catalytic activity of the mutant enzymes has been compared to wild-type variants and the effects of the substitutions on substrate binding and catalysis are discussed in detail.

3.2 Materials & methods

Strains, plasmids and chemicals. Escherichia coli BL21(DE3) was used as a host for the cloning and production of wild-type and mutant enzymes. All genes were cloned in the expression vector pET9d (Novagen). Expression plasmid pLUW511 was used for the production of wild-type CelB, as described *(32)*. All chemicals were of analytical grade. The chromogenic substrates *para*-nitrophenol-β-D-glucopyranoside (pNp-Glc), *para*-nitrophenol-β-D-galactopyranoside (pNp-Gal), *para*-nitrophenol-β-D-mannopyranoside (pNp-Man) were obtained from Sigma. 5-Bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-glu) was purchased from Biosynth (Staad, Switserland). The thermostable transition state analogue (5R,6R,7S,8S)-5-(hydroxymethyl)-2-phenyl-5,6,7,8-tetrahydroimidazol[1,2-a]pyridine-6,7,8-triol (PheImGlc) was synthesized as described *(12)*. Chromosomal DNA of *P. horikoshii* OT3 was kindly provided by Ir. C. Verhees (Wageningen University, Wageningen, Netherlands).

Cloning of P. horikoshii bglB gene. The open reading frame PHO0501 coding for BglB (Genbank AP000002) was identified by a BLAST search of the *P. horikoshii* OT3 genome sequence using the CelB amino acid sequence. Based on the presence of a putative Shine-Dalgarno sequence, the translation start site of the *bglB* gene was predicted, which resulted in a protein that corresponded in size and composition to other family 1 enzymes. The gene was amplified by PCR using the 5'-oligo-nucleotide BG461 and the 3'-oligo-nucleotide BG462. BG461 (5'-gcgcgccATGgcaAAGTTTTACTGGGGC GTCGTT-3'(*bglB* sequence in capitals)) introduced an alanine residue following the initiator methionine in order to create a *Nco*I site (underlined) that allowed a translational fusion of *bglB* to the T7 promoter of pET9d. BG462 (5'-gcgcgcgctcagc TTATTTTCTTCCAGGTAGTTTCAT-3') introduced a *Blp*I site (underlined) after the stop codon of *bglB*. The alanine at position 2 will be referred to as Ala 2, which results in a identical residue numbering for a large part of both proteins. PCR reactions (50 μl) contained 2.5 U *Pfu* DNA polymerase (Stratagene) in the supplied buffer, 0.2 mM of each dNTP, 20 ng of chromosomal *P.*

horikoshii DNA and 10 pmol of each primer. The reactions were subjected to a denaturation step of 5 min at 95°C followed by 35 repeats of 1 min at 95°C, 45 sec at 50°C and 90 sec at 72°C, followed by 5 min at 72°C. PCR products were recuperated in 2 mM Tris HCl pH 8.0 after purification using a PCR purification kit (QIAgen). After digestion by *Nco*I and *Blp*I, the PCR product and pET9d vector were isolated from agarose gel using the QiaexII gel extraction kit (QIAgen) and ligated together. *E. coli* BL21(DE3) was transformed with the ligation mix by electroporation and plated on selective TY plates containing X-glu . Colonies expressing functional β-glycosidase activity were selected by their blue phenotype, since X-glu cannot be hydrolyzed by *E. coli* LacZ β-galactosidase. The nucleotide sequence of the *bglB* gene in plasmids from positive colonies was verified by DNA sequence analysis using the Thermosequenase cycle sequencing kit (Amersham-Pharmacia-Biotech) with infrared dye labeled primers. This provided the same sequence as deposited in the database (*29*). Reactions were analysed on a LiCor 4000L automated sequencer (data not shown). The resulting construct was denoted pLUW527.

Homology modeling. The BglB primary sequence was submitted to the Swiss model homology server (*33*). The structures of the β-glycosidases from *Sulfolobus solfataricus* (LacS, PDB: 1GOW) and *Thermosphaera aggregans* (Bgly, PDB: 1QVB) and the 6-phospho-β-galactosidase LacG form *Lactococcus lactis* (PDB: 2PBG) were used as templates in the modeling procedure. The quality of the model was verified using the programs Procheck (*34*) and Prosa (*35*). A galactose molecule was modeled in the active site of BglB by superimposition of the active sites of BglB and *L. lactis* LacG, crystallized with the ligand in the active site.

Construction of mutants. Mutations were introduced in *celB* and *bglB* using *Pfu* DNA polymerase using the PCR-based overlap extension method (*36*). For each mutation a sense/antisense primer pair was designed. R77Q was introduced in CelB by BG652/BG653 (*Acc*I). BG530 (5'-GACATGTG<u>GTCGAC</u>AATGGACGAA CCAAAC-3')/BG531 (5'-GTTTGGTT CGT CCATT<u>GTCGAC</u>CACATGTC-3') (*Sal*I) introduced N206D in CelB (introduced mutations in bold face, introduced restrictions sites underlined and restriction enzymes bracketed). BG654/BG655 (*Cla*I) introduced Q77R in BglB. BG528 (5'-GATTACTG<u>GTCGAC</u>ATTTAATGAACCAATG-3')/BG529 (5'-CATTGGTTCATTAAAT<u>GTCGAC</u>CA-GTAATC-3') (*Sal*I) resulted in D206N in BglB. These primers were used in combination with BG238/BG239 in the case of *celB* (*32*) and with BG461/462 in the case of *bglB* (see above). Mutated *bglB* and *celB* genes were cloned and isolated as described above. Plasmid pLUW511 (CelB wild-type) was used for the construction of pLUW524 (CelB R77Q) and pLUW525 (CelB N206D). Plasmid pLUW525 served as template for the construction of pLUW526 (CelB R77QlN206). Likewise, pLUW527 (wild-type BglB) was used to construct pLUW528 (BglB Q77R) and pLUW529 (BglB D206N). Plasmid pLUW533 (BglB Q77RlD206N) was constructed using pLUW529. Introduced mutations were confirmed by restriction analysis and DNA sequence analysis (see above).

Enzyme production and purification. Enzymes were produced and purified essentially as described before (*24*). The cell-free extracts of wild-type and mutant BglB-producing BL21 (DE3) strains were subjected to an extra centrifugation step of 10 min at 6000 rpm followed by a heat incubation of 40 min at 80 °C. Heat-stable cell free extracts were applied on an anion exchange column (Q-sepharose, Amersham-Pharmacia-Biotech) equilibrated with 20 mM Tris-HCl (pH 8.0) and eluted with a linear NaCl gradient (0-1.0 M). Pure protein was obtained after applying the pooled active fractions on a hydrophobic interaction column (phenyl-sepharose, Amersham-Pharmacia-Biotech), equilibrated with 20 mM Tris HCl (pH 8.0) with 1 M $(NH_4)_2SO_4$ followed by elution during a linear decreasing $(NH_4)_2SO_4$ gradient (1.0-0 M). Active fractions were pure as judged by SDS-PAGE analysis, and pooled fractions were dialyzed against 20 mM NaP_i (pH 7.5). The protein solutions were stored at 4°C and contained 0.02% NaN_3 to prevent microbial growth. Protein concentrations were measured at 280 nm according to Gill and Hippel (*37*), in which calculated ε_{280}-values of 128,280 $M^{-1}cm^{-1}$ and 120,740 $M^{-1}cm^{-1}$ were used for CelB and BglB, respectively.

Molecular weight determination by gel filtration. Samples of 100 µg purified BglB and CelB were loaded on a Superdex 200 HR16/60 column (Amersham-Parmacia-Biotech) equilibrated with 100 mM NaCl in 20 mM Tris HCl (pH 8.0) and eluted at a rate of 0.5 ml/min using an Äkta FPLC (Amersham-Pharmacia-Biotech). Besides the CelB protein with a previously determined size of 230 kDa (*26*), blue dextran (2000 kDa), thyroglobin (669 kDa), ferritin (440 kDa) catalase (232 kDa) chymotrypsin (25 kDa) and ribonuclease (13,7 kDa) were used as standard weight markers.

Kinetic stability. Kinetic stability was measured as described (*38*). Aliquots of wild-type BglB and CelB were diluted to 0.05 mg/ml in 150 mM sodium citrate (pH 5.0). Samples of 100 µl were dispensed in glass vials closed with a teflon cap and submersed in silicon oil of 106 °C. At regular intervals, vials were removed from the oil and immediately chilled on ice. Residual activity was determined using a routine activity assay (*38*). Half-lives of inactivation were calculated from data fits according to first order kinetics using the non-linear regression program Tablecurve 2D (Jandel Scientific).

Differential scanning calorimetry. Protein solutions of wild-type CelB and BglB were dialyzed extensively against 20 mM Na P_i (pH 7.5) and diluted to 4.6 and 4.4 µM in dialysis buffer, respectively. After 10 min degassing and equilibration, samples were analyzed in a VP-DSC differential scanning micro-calorimeter (MicroCal) between 50-125 °C at 0.5°C/min against the dialysis buffer. Enzyme scans were corrected using a buffer-buffer baseline.

Kinetic analyses. The optimal pH for hydrolysis and kinetic constants by wild-type and mutant enzymes at 90 °C were determined as described in earlier studies (*24*). The optimal pH for hydrolysis was determined for 15 mM pNp-Glc. To determine the kinetic parameters of the enzyme variants at their respective optimal pH at 90 °C, enzyme-catalyzed hydrolysis was determined at 10 to 15 different concentrations of pNp-Glc (0-20 mM), pNp-Gal (0-25 mM) and pNp-Man (0-12

A.

B.

```
Pho BglB   40 : KNDLVSGDLPEEGINNYELYEIDHRLAKELGLNAYQLTIEWSRIF :  85
Pfu BglB   40 : KNELVSGDLPEEGINNYDLYEIDHRLAKDLGLNAYQLTIEWSRIF :  85
Pab BglB   40 : KNELVSGDLPEEGINNYELYEIDHRLAKELGLNAYQLTIEWSRIF :  85
Tko BglB   41 : KNDLVSGHLPEDGINNYGLYEIDHQLAKDMGLNAYQITVEWSRIF :  86
Pfu BmnA   43 : EKGLVSGDLPEEGINNYELYEKDHEIARKLGLNAYRIGIEWSRIF :  88
Pab BmnA   43 : EKGLVSGDLPEDGINSYELYEVDHNLAKDLGLNMYRIGIEWSRIF :  88
Pfu CelB   41 : ASGLVSGDLPENGPAYWHLYKQDHDIAEKLGMDCIRGGIEWARIF :  86
Tag Bgly   42 : AAGLVSGDFPENGPGYWNLNQNDHDLAEKLGVNTIRVGVBWSRIF :  87
Sso LacS   43 : AAGLVSGDLPENGPGYWGNYKTFHDNAQKMGLKIARLNVEWSRIF :  88
                                                      ▲
                                                      1

Pho BglB  174 : VDERTIIEFAKYAAYVAWKFDNYVDYWSTFDEPMVTAELGYLAPY : 219
Pfu BglB  174 : VDEKTIVEFVKYVAYVAWKFDQYVDYWATFDEPMVTVELGYLAPY : 219
Pab BglB  174 : VDERAIVEFSKFAAYVAWKFDKYVDFWATFDEPMVTAELGYLAPY : 219
Tko BglB  175 : VDERAILEFAKFAAFVAWKRWDLVDYWATFDEPMVTVELGYLAPY : 220
Pfu BmnA  178 : VNPRTVIEFAKYAAYIAYKFGDIVDMWSTFNEPMVVVELGYLAPY : 223
Pab BmnA  178 : INKRTVIEFAKFAAYIAYKFGDIVDMWSTFNEPMVVVELGYLAPY : 223
Pfu CelB  175 : LDEKTVVEFVKFAAFVAYHLDDLVDMWSTMNEPNVVYNQGYINLR : 220
Tag Bgly  176 : LNEESVVEFAKYAAYIAWKMGELPVMWSTMNEPNVVYEQGYMFVK : 221
Sso LacS  174 : LSTRTVYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVK : 219
                                                      ▲#
                                                      2

Pho BglB  463 : RIPRRRSVEIYKKIVM-EGIE------------------------ : 483
Pfu BglB  463 : RIPRRKSVETYKMIVA-EGIE------------------------ : 483
Pab BglB  463 : RIPRRKSVETYKKVVK-EGIE------------------------ : 483
Tko BglB  463 : RIPRPRSVETYKKIVR-EGLP------------------------ : 483
Pfu BmnA  478 : RIPREKSVSIFREIVANNGVTKKIEBEL---LRGM---------- : 510
Pab BmnA  479 : RKPRKRSVEVFSRIIS-EGTTKGLQDEFRLKVRGERYENDSG--- : 520
Pfu CelB  437 : RYLR-PSALVFREIATQKEIPEELAHLADLKFVTRK--------- : 472
Tag Bgly  451 : RYLR-PSALVFREIATHNGIPDELQHLTLIQ-------------- : 481
Sso LacS  452 : LYWR-PSALVYREIATNGAITDEIEHLNSVPPVKPLRH------- : 489
```

Figure 3.1 A: Phylogenetic tree of thermostable (putative) family 1 glycosyl hydrolases: Pho_BglB: *P. horikoshii* β-mannosidase (Genbank AP000002), Pab_BglB: *P. abyssi* putative β-mannosidase BglB (AJ248288), Pfu_BglB: *P. furiosus* putative β-mannosidase BglB (*P. furiosus* genome ORF Pf_368506), Tko_BglB: *Thermococcus kodakaraensis* putative β-mannosidase BglB (Genbank AB028601), Pfu_BmnA: *P. furiosus* β-mannosidase BmnA (U60214), Pab_BmnA: *P. abyssi* putative β-mannosidase BmnA (AJ248285), Pfu_CelB: *P. furiosus* β-glucosidase CelB (AF013169), Tag_Bgly: *T. aggregans* β-glycosidase (AF053078), Sso-LacS: *S. solfataricus* β-glycosidase LacS (M34696). Alignment and tree were generated in ClustalX (*53*) and the tree was visualized using Treeview (*54*). B: Alignment of partial protein sequences of above mentioned thermostable family 1 glycosyl hydrolases, "1" and "2" indicate positions of unique glutamine and aspartic acid residues in BglB, respectively, "#" indicates active site catalytic acid/base residue.

mM) bracketing the K_m, when possible. Liberated pNp was measured at 405 nm using a U-2010 spectrophotometer (Hitachi) equipped with a temperature controller. Specific absorption coefficients of pNp were determined at 90 °C: $\varepsilon_{405,\,pH\,4.8}$: 556 $M^{-1}cm^{-1}$, $\varepsilon_{405,\,pH\,5.0}$: 905 $M^{-1}cm^{-1}$, $\varepsilon_{405,\,pH\,5.2}$:1200 $M^{-1}cm^{-1}$. Data were fitted according to Michaelis-Menten kinetics using the non-linear regression program Tablecurve 2D (Jandel Scientific). The turnover number, k_{cat}, has been defined as the amount of substrate molecules converted per second per active site in which molecular masses of 54,664 Da and 56,467 Da were used for subunits of wild-type CelB and wild-type BglB, respectively. The K_i value for the inhibition by PheImGlc was first estimated by measuring the enzyme-catalyzed hydrolysis at a fixed concentration pNp-Glc close to the K_m-value in the presence of varying inhibitor concentrations and plotting the data in a Dixon plot. Activities of enzyme-catalyzed hydrolysis were determined at 5 different substrate concentrations, bracketing the K_m and 7 different inhibitor concentrations, around the estimated K_i-value. The final K_i was calculated from fitting the data using the non-linear regression program Tablecurve 3D (Jandel Scientific) using a formula for competitive inhibition (*39*).

3.3 Results

Primary sequence analysis of P. horikoshii BglB. Open reading frame PH0501 of *P. horikoshii* is predicted to code for a 483 amino acid protein with a calculated size of 56,457 Da, designated here as BglB. Comparison of the predicted amino acid sequence of *P. horikoshii* BglB with those of the most significant hits in a BLAST search revealed that BglB belongs to glycosyl hydrolase family 1. It shows the highest protein sequence identity (> 70%) and clusters together in a phylogenetic tree with putative BglB-like proteins from other species from the order of *Thermococcales* (Figure 1A). The well-characterized *P. furiosus* β-glucosidase CelB (37% amino acid identity) (*13*) and *S. solfataricus* β-glycosidase LacS (38%) (*40*) are relatively distantly related to BglB. Since crystal structures of CelB and LacS and the *Thermosphaera aggregans* β-glycosidase Bgly (40%) are available (*22-24*), these enzymes can be used for structural interpretation of the BglB sequence as well as for the construction of a 3D model. The β-mannosidase BmnA (56%) from *P. furiosus* is the characterized glycosidase most closely related to *P. horikoshii* BglB (*7*).

A remarkable characteristic of family 1 glycosyl hydrolases is the high degree of conservation of the amino acid residues that are involved in catalysis and substrate binding. Surprisingly, two BglB active site residues were identified that are different from those present in other family 1 glycosyl hydrolases. These residues are the glutamine at position 77 and the aspartic acid at position 206 (Figure 1B, indicated by 1 and 2). In all other family 1 enzymes these residues are an arginine and an asparagine, respectively.
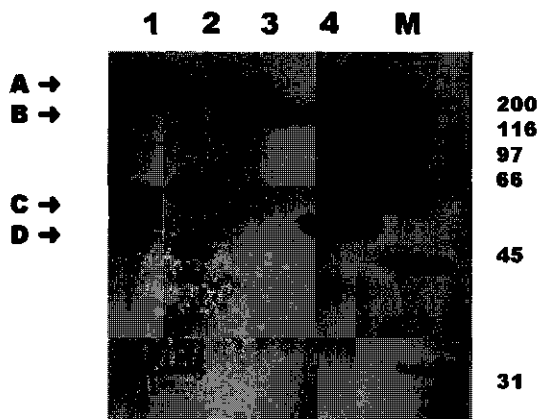
Figure 3.2 SDS-PAGE gel of purified BglB and CelB. 1 and 2 are BglB (15 μg), 3 and 4 are CelB (21 μg). Before loading, 1 and 3 were not incubated, 2 and 4 were incubated for 10 min at 100 °C in SDS-sample buffer. M: Biorad broad range marker, sizes in kDa. Arrows indicate positions of BglB multimer complex (A), CelB tetramer (B), BglB monomer (C), CelB monomer (D).

Characterization of P. horikoshii BglB. P. horikoshii ORF PHO501 was overexpressed in *E. coli* BL21(DE3). Although the majority of the protein was expressed as non-soluble, inactive inclusion bodies, a small fraction (~5%) was soluble and active and could be purified to homogeneity (Figure 2, lane 1,2). Attempts to improve the functional expression of the *bglB* gene by varying the addition and amount of IPTG induction, or growth temperature did not result in increased production yields (data not shown). The biochemical properties of the purified BglB were compared to those of *P. furiosus* β-glucosidase CelB. The native conformation and thermostability of BglB and CelB was studied by boiling of protein samples. Unboiled, BglB appears as a complex of high molecular mass on a SDS-PAGE gel (Figure 2, lane 1,A). After 10 min boiling in SDS-sample buffer, virtually all BglB has been denatured, although a small portion is still present as a multimer with an apparent size of about 160 kDa (Figure 2, lane 2,B). On the other hand, the CelB protein is mainly present in what is generally accepted as the tetrameric conformation, which appears as a protein band of about 150 kDa (Figure 2, lane 3). When CelB is boiled for 10 min, only a fraction of the protein is fully denatured and appears at the position of the size of a single subunit (Figure 2, lane 4). The position of denatured monomeric BglB (Figure 2, C) correlates with the calculated subunit size (56.5 kDa), which is slightly larger than that of monomeric CelB (52.1 kDa) (Figure 2, D). In gel filtration studies, the active form of BglB eluted at a position that corresponded to a molecular size of about 250 kDa. This suggests that the active form of BglB is a tetramer.

Table 3.1 Half-lives of thermal inactivation and apparent melting temperatures (T_m) of wild-type BglB and CelB.

	Half-life at 102°C	T_m (°C)
BglB	27 min	100.7
CelB	> 10 h	106.8

The thermostability of BglB was compared to CelB by heat incubation at 102 °C and was found to be significantly less than that of CelB (Table 3.1). For BglB and CelB apparent melting temperatures of 100.7 °C and 106.0 °C were determined, respectively, which correlated with the determined half-lives at 102°C.

Optimal hydrolytic activity of BglB at 90°C was observed at pH 4.75 (data not shown). The kinetic parameters of the enzyme for the hydrolysis of pNp-Glc, pNp-Gal and pNp-Man were determined at 90°C (Table 3.2). BglB shows highest turnover rates for pNp-glucose and pNp-galactose, but, because of a higher affinity for mannose, is most efficient in the hydrolysis of pNp-mannose.

3D structural model of BglB. The 3D model of BglB as obtained from the PDB homology server was analyzed with Procheck (*34*) and Prosa (*35*). The majority of the residues (96%) was located in the energetically most favorable or allowed regions of the Ramachandran plot of dihedral angles, with few in the generously allowed (2%) or disallowed (2%) regions. The Z-score of the *P. horikoshii* BglB model was −10.14, which makes it a reliable model (*41*). As expected, the overall structure of BglB shows a $(\beta/\alpha)_8$ barrel that resembles that of CelB to a high extent. Compared to CelB, BglB has 2 large insertions that run from residue 300 to 316 and residue 338 to 345. The first loop is not present in CelB, the second loop is considerably shorter in CelB. Another significant difference between the two proteins lies at the C-terminal end, where CelB is 16 residues longer than BglB. An overlay of the active sites of BglB and CelB shows the conservation of the positions of the active site residues (Figure 3A). A galactose molecule modeled in the active site visualizes the position of the two unique active site residues Gln 77 and Asp 206 with respect to the catalytic glutamate residues and the substrate (Figure 3B). These residues are located close to the two catalytic glutamate residues, which act as a nucleophile (BglB Glu 399, CelB Glu 372) and a catalytic acid/base (both Glu 207) in the hydrolysis reaction. The positions of backbone atoms of the unique glutamine 77 and aspartic acid 206 in BglB are identical to the corresponding residues of CelB (Figure 3B). Due to the shorter side chain, therefore, Gln 77 is not able to reach into the active site of BglB as far as Arg 77 in CelB. Arg 77 forms a salt bridge with the catalytic residue Glu 372 in CelB, while in BglB there is no interaction between Gln 77 and the corresponding catalytic residue Glu 399. Asp 206, the other unique active site residue, occupies the same space as its isosteric CelB counterpart Asn 206. The corresponding residue of Asp 206 in BglB has been found to interact with the hydroxyl group at carbon atom 2 of glucosides and galactosides in resolved

A.

B.
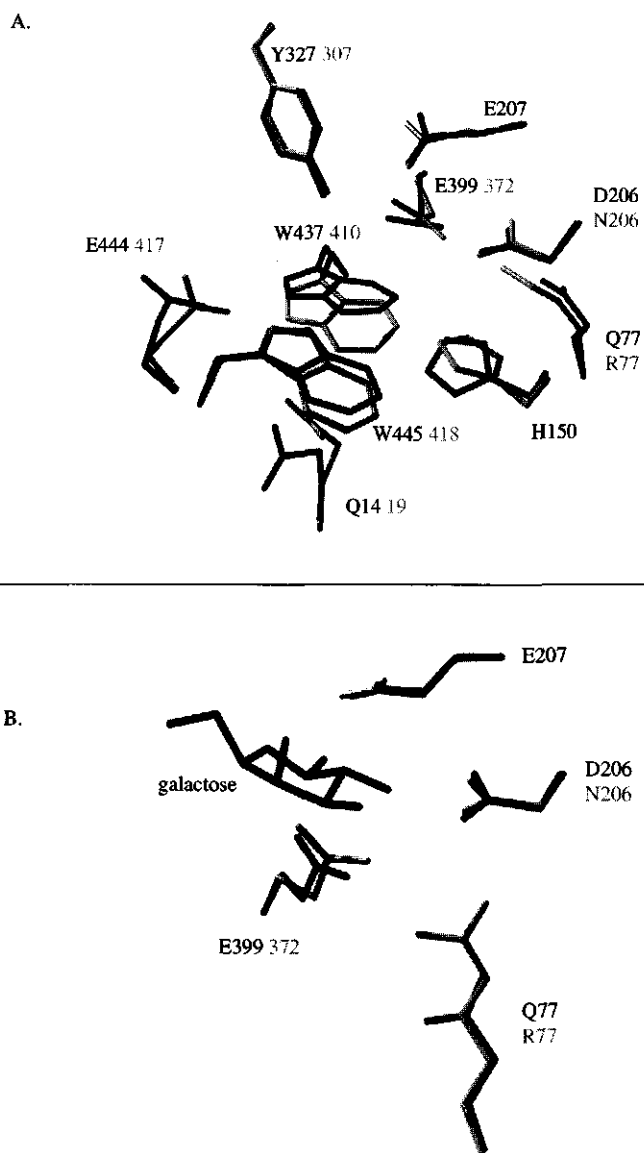
Figure 3.3 Superposition of the active site residues BglB residues (black) and CelB residues (where numbering is different from BglB: grey) (*24*) that align with substrate binding residues of *L. lactis* LacG (*25*) and *B. polymyxa* BglA (*21*) (A), and close up of unique BglB residues and catalytic glutamic acid residues (black) superimposed with corresponding CelB residues (grey), and a modeled galactose molecule (B). Images were made in PDBviewer (*55*) and visualized using Povwin (*56*).

Crystal structures of other family 1 glycosidases with bound ligands in the active site *(21, 25)*. Indeed, Asp 206 seems close enough (2.8 Å) to the hydroxyl at position 2 of the galactose molecule to have an interaction (Figure 3B). Interestingly, this is the hydroxyl group, which is axial in mannose and equatorial in glucose.

Characterization of mutants. Enzyme production. To analyze the role of the unique residues in the BglB active site, their CelB counterparts were introduced in BglB and vice versa. This resulted in two sets of mutants. The first set consisted of BglB variants that posses one or two CelB residues in the active site, two single mutants (BglB Q77R and BglB D206N), and one double mutant (BglB Q77R|D206N). The second set was made up by three CelB counter mutants, two single mutants (CelB R77Q and CelB N206D) and one double mutant (CelB R77Q|N206D). The enzymes were produced in *E. coli* BL21(DE3). The mutant enzymes resembled the corresponding wild-type enzymes in production and purification, evidencing proper folding. Similar as for the wild-type protein, about 5% of the BglB mutants was present in the soluble fraction and up to 2 mg of each BglB variant protein could be purified per liter cell culture.

pH optima. The optimal pH for hydrolysis of pNp-Glc at 90°C was determined (between pH 3-8). It was found that both the wild-type and mutant BglB and CelB variants show optimal catalysis at pH 4.75-5.0 (data not shown). Above pH 5.0, however, the k_{cat} of BglB Q77R decreased more rapidly than the one of the wild-type. In contrast, BglB D206N was relatively insensitive to a pH change above pH 5.0. The CelB mutants showed pH profiles identical to wild-type CelB. These findings are remarkable, since the known cases of removal or introduction of charges in the active site of CelB and other β-glycosyl hydrolases have resulted in significant pH-shifts for optimal hydrolysis *(24, 42)*.

Kinetic parameters. The catalytic parameters of wild-type CelB and CelB mutants for the hydrolysis of pNp-Glc, pNp-Gal and pNp-Man have been determined (Table 3.3). Wild-type CelB may be regarded as a β-glucosidase; it has the highest catalytic efficiency for the hydrolysis of pNp-Glc. However, the broad substrate specificity commonly found for family 1 enzymes was illustrated by CelB's considerable galactosidase activity and, although modest, mannosidase activity. When the unique residues of BglB were introduced in CelB, the hydrolysis of pNp-Glc and pNp-Gal were affected similarly: both the rate of hydrolysis and substrate affinity decreased (Table 3.3). Compared to N206D, the substitution R77Q had more effect on the substrate affinity for pNp-Glc, which decreased 85-fold, than on the hydrolysis rate, which was reduced only 5-fold. The second substitution N206D resulted in a 10-fold reduction of pNp-Glc hydrolysis, while the affinity was diminished by a factor 30. The hydrolysis-rate of pNp-Man was affected to a larger extent by both R77Q and N206D: 60 and 30-fold reduction, respectively. However, the substitutions resulted in a 2 to 8.7-fold increased affinity for mannose, the C2-epimer of glucose. The combination of both substitutions showed an additive effect and gave an enzyme with the combined activities of both separate mutants: the substrate affinity of R77Q combined with the slightly higher activity of N206D.

Table 3.2 Kinetic parameters for the hydrolysis of pNp-Glc, pNp-Gal, pNp-Man and PheImGlc inhibition of wild-type BglB and BglB mutants at 90°C in 0.1 M Na-citrate/0.1 M NaP$_i$. In the assays between 0.48 and 1.1 µg protein was used.

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}mM^{-1})	%	$K_{i, PheImGlc}$ (nM)
BglB wild-type	pNp-Glc	13.5 ± 3.3	34.4 ± 4.3	2.58	26	$1.0 \cdot 10^5$
	pNp-Gal	7.3 ± 1.1	13.5 ± 1.0	1.9	20	
	pNp-Man	0.44 ± 0.10	4.2 ± 0.3	9.8	100	
BglB Q77R	pNp-Glc	16.2 ± 2.3	25.5 ± 2.2	1.6	100	$1.1 \cdot 10^5$
	pNp-Gal	16.8 ± 2.9	14.2 ± 1.4	0.85	54	
	pNp-Man	1.1 ± 0.2	1.6 ± 0.1	1.4	91	
BglB D206N	pNp-Glc	0.30 ± 0.04	252 ± 9	870	100	14
	pNp-Gal	7.8 ± 1.0	571 ± 32	75.8	8.7	
	pNp-Man	3.2 ± 1.1	45.0 ± 5.0	14.7	1.7	
BglB Q77R|D206N	pNp-Glc	30.4 ± 3.5	439 ± 36	14.4	100.0	$2.0 \cdot 10^4$
	pNp-Gal	36.1 ± 5.7	66.5 ± 7.3	1.8	13	
	pNp-Man	0.68 ± 0.4	6.8 ± 0.4	9.9	69	

The combined substitutions resulted in an increased efficiency for the hydrolysis of mannosides. The mutants containing the substitution R77Q were even more efficient in the hydrolysis of pNp-Man, than of pNp-Gal. The Lineweaver-Burke plot for the hydrolysis of pNp-Man by these mutants was biphasic (Figure 4), which is an indication for oligosaccharide synthesis, as observed previously (*8, 28*). Indeed, products corresponding in size to saccharide oligomers were detected by FPLC analysis in samples containing pNp-Man and CelB R77Q or CelB R77Q\|N206D after prolonged incubation at 90 °C (data not shown). Although the β-mannosidase activity was increased relative to the β-glucosidase and β-galactosidase activity, β-glucosidase activity was still the most dominant activity in all CelB mutants.

In BglB, the introduction of an arginine at position 77 (Q77R) and an asparagine at position 206 (D206N) (as in CelB) has different effects (Table 3.2). When glutamine 77 was replaced by an arginine, none of the catalytic parameters on all three substrates was improved. However, upon the introduction of D206N the k_{cat} on all substrates increased about 10-fold and the K_m for the hydrolysis of pNp-Glc was reduced 45-fold. This made BglB D206N effectively a β-glucosidase. As in the CelB double mutant, BglB Q77R\|D206N showed the combined characteristics of the respective single mutants.

Inhibition by PheImGlc. Many glucose derivatives with a more or less trigonal C(1) atom, such as gluconolacton inhibit family 1 glycosyl hydrolases, but are insufficiently stable under hydrolytic or thermal conditions to be suitable for inhibition experiments at elevated temperatures. The inhibitor PheImGlc has a trigonal geometry at the center corresponding to C(1) and is

thermostable, as expected on the basis of the aromatic character of the imidazole ring (Figure 5). PheImGlc was used to test the inhibition of pNp-Glc hydrolysis at 90 °C by CelB, BglB and mutants (Table 3.2 & 3.3). At this temperature the inhibitor PheImGlc was stable within the time span of an activity assay (6 min). All enzymes were inhibited competitively by PheImGlc, although there were large differences in sensitivity towards the thermostable inhibitor. Wild-type CelB was inhibited by PheImGlc with a low inhibition constant of 6.5 nM. The inhibition constant of CelB for PheImGlc was found to be 6.5 nM. This value is lower than found with PheImGlc for the β-glucosidases from almond (100 nM) and Caldocellum saccharolyticum (18 nM) (*12*). Assuming, by analogy (*43*), that this inhibitor is at least a partial transition state analogue, this means that the active site of CelB (at 90 °C) is slightly more adapted to stabilize the reaction transition-state than the β-glucosidases of almond and *C. saccharolyticum*. Also, the reported inhibitor constants for the *bona fide* transition state analogues glucono-lactone and 1-deoxynojirimycin for pNp-Glc hydrolysis by CelB were at least 1000-fold higher than for PheImGlc (*28*). On the one hand, the PheImGlc inhibition constant for mutant CelB R77Q was 15-fold higher than for the wild-type enzyme, but was still relatively low compared to the 1500-fold increase of the inhibition constant for CelB N206D. The CelB double mutant was slightly more sensitive to inhibition by PheImGlc than CelB N206D. On the other hand, wild-type BglB was 15,000-fold less strongly inhibited by PheImGlc than wild-type CelB. BglB Q77R displayed a similar sensitivity. Strikingly, the inhibition constant for BglB D206N was comparable to that of CelB wild-type. The BglB double mutant was slightly more sensitive than wild-type BglB and BglB Q77R.
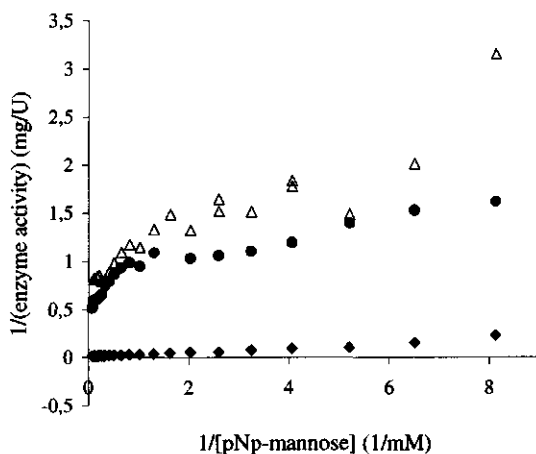
Figure 3.4 Lineweaver-Burke plot for the hydrolysis of pNp-man at 90 °C by wild-type CelB (◆), CelB R77Q (Δ), CelB R77Q/N206D (●).

Table 3.3 Kinetic parameters for the hydrolysis of pNp-Glc, pNp-Gal, pNp-Man and PheImGlc inhibition of wild-type CelB and CelB mutants at 90°C in 0.1 M Na-citrate/0.1 M NaP$_i$. In the assays between 85 ng and 2.0 µg protein was used. Mutants CelB R77Q and R77Q/N206D showed a bi-phasic behavior for the hydrolysis of pNp-Man (Fig. 4) with separate kinetic parameters above (I) and below (II) 1 mM pNp-Man

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}mM^{-1})	%	$K_{i, PheImGlc}$ (nM)
CelB WT	pNp-Glc	0.19 ± 0.06	1140 ± 86	7337	100	6.5
	pNp-Gal	5.0 ± 0.25	2827 ± 46	561	7.6	
	pNp-Man	1.3 ± 0.1	65.9 ± 1.6	49.8	0.7	
CelB R77Q	pNp-Glc	13.8 ± 3.0	207 ± 25	14.9	100	97
	pNp-Gal	21.7 ± 3.0	69.2 ± 5.8	3.2	21	
	pNp-Man I	0.64 ± 0.12	1.22 ± 0.05	1.9	13	
	II	0.15 ± 0.04	0.85 ± 0.07	5.7	38	
CelB N206D	pNp-Glc	4.6 ± 0.06	116 ± 7	24.9	100	9.8·10^3
	pNp-Gal	15.7 ± 2.0	180 ± 13	11.5	46	
	pNp-Man	0.63 ± 0.07	1.97 ± 0.07	3.1	13	
CelBR77Q|N206D	pNp-Glc	13.9 ± 1.2	172 ± 8	12.3	100	2.7·10^3
	pNp-Gal	22.0 ± 2.6	29.2 ± 2.1	1.3	11	
	pNp-Man I	1.5 ± 0.2	1.9 ± 0.1	1.3	11	
	II	0.10 ± 0.01	1.02 ± 0.03	10.2	83	
CelB N206S[a]	pNp-Glc	12	309	26	100	
	pNp-Gal	200[b]	1090[b]	5.5	21	
	pNp-Man	4.7	15	3.2	12	

[a] values obtained from (*32*)

[b] estimated values, due to the limited solubility of pNp-Gal

Plotting log k_{cat}/K_m versus log $1/K_i$ for PheImGlc for the enzyme variants visualized a linear trend with slope 0.65 between the free energy for activation and binding energy for PheImGlc (Figure 6). Comparison of log k_{cat}/K_m versus log $1/K_m$ for the hydrolysis of pNp-Glc and pNp-Gal by the wild-type and mutant enzymes revealed a linear trend for both substrates with slopes of 1.4 and 2.5, respectively (data not shown). For pNp-Man, such a linear trend was not observed (data not shown).

3.4 Discussion

Primary sequence analysis and characterization of P. horikoshii BglB. Gene PH0501 of the *P. horikoshii* genome codes for a β-mannosidase, denoted BglB, which can be placed in family 1 of glycosyl hydrolases. Uncharacterized homologues of BglB are found in two other *Pyrococcus* and a *Thermococcus* species, but the *P. furiosus* β-mannosidase BmnA is the closest related characterized hydrolase (Figure 1A.) (*7*). The *bglB* gene was cloned in the pET9d vector and over-expressed in *E. coli* BL21 (DE3). Purified BglB showed an apparent size that was slightly larger than the CelB tetramer in SDS-PAGE (Figure 2) and gel filtration. The most closely related and characterized enzymes, the *P. furiosus* BmnA and CelB, *S. solfataricus* LacS, and *T. aggregans* Bgly, are all active as tetramers (*23, 26, 44*). BglB does not show any large insertions or deletions compared to aforementioned enzymes. As expected, the active form of BglB was found to be tetrameric as well. In a direct comparison, the apparent melting temperature of BglB was determined to be 6 °C lower than the one determined for CelB, which is in agreement with the shorter half-life for BglB activity, as compared to that of CelB (Table 3.1). The lower stability of BglB can be partly explained by the relatively short C-terminus of BglB. The longer C-terminus of *S. solfataricus* LacS is involved in an intricate ion-pair network that bridges the four subunits at the center of the protein (*22*). A similar, though less extensive, interaction is postulated to for *P. furiosus* CelB, where the penultimate arginine residue is reaching towards an aspartic acid residue of an adjacent subunit to form a salt bridge (*45*). These interactions in LacS and CelB contribute to their stability (*45, 46*)). Such stabilizing interactions are ruled out for BglB, where the C-terminus is 16 residues shorter compared to *P. furiosus* CelB.

The most remarkable feature of BglB is the presence of two active site residues that differ from those in other family 1 enzyme sequences (Figure 1B.). The homologous BglB sequences in *P. furiosus*, *P. abyssi* and *P. kadokaraensis* display identical active site residue organization, which identifies them as a new subfamily of family 1 glycosyl hydrolases. If *P. horikoshii* BglB can be taken as a representative, the members of this subfamily display very low hydrolytic activities, as compared to the related β-mannosidase BmnA and β-glucosidase CelB from *P. furiosus* (*7, 26, 38*) (Tables 3.2 & 3.3). Although BglB has the highest turnover rates for the hydrolysis of pNp-Glc and pNp-Gal, it can be classified as β-mannosidase, because it has the highest catalytic efficiency for the hydrolysis of pNp-Man.
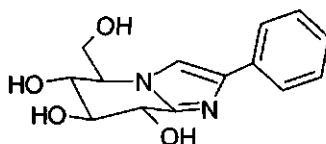

Figure 3.5 Molecular structure of the thermostable transition state inhibitor (5R,6R,7S,8S)-5-(hydroxymethyl)-2-phenyl-5,6,7,8-tetrahydroimidazol[1,2-a]pyridine-6,7,8-triol (PheImGlc).
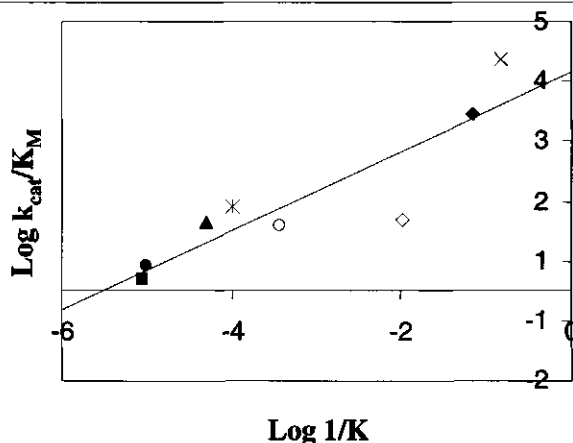
Figure 3.6 Correlation of activation free energy for turnover of pNp-Glc (log k_{cat}/K_m) by the CelB and BglB variants with the corresponding free energy of binding of transition state analogue PheImGlc (log $1/K_i$). Symbols indicate: CelB wt (×), CelB R77Q (◊), CelB N206D (✳), CelB R77Q/N206D (○), BglB wt (●), BglB Q77R (■), BglB D206N (◆), BglB Q77R/D206N (▲).

Model of P. horikoshii BglB and design of mutations. The modeled structure of *P. horikoshii* BglB shows the classical $(β/α)_8$-barrel fold, typical for all 3D structures of family 1 β-glycosidases (*19, 20, 22-24*). Compared to CelB, the structure of BglB is slightly less compact. Throughout family 1 the structure of the active site has been highly conserved and the active site of BglB is no exception (Figure 3A). The two unique active site residues in BglB are located near the two glutamic acid residues that are essential for catalysis (Figure 3B). Moreover, the two residues are close enough to a modeled galactose molecule to have an interaction. However, the interactions of a mannoside with the enzyme can be expected to be different from those in Figure 3B. Upon binding in the active site, the sugar ring is likely to be distorted, as found previously (*47*). This will bring the glycosidic oxygen atom of the mannoside in plane with catalytic acid Glu 207, whereupon the glycosyl-enzym intermediate is formed (*47*). The β-1,4-glucanase Cex from *Cellulomonas fimi* hydrolyzes β-glucosides by the retaining mechanism, like CelB. The C2-hydroxyl group of a covalently bound glucose residue in the −1 subsite of in the active site of Cex has been found to have a strong interaction with the carbonyl oxygen of the carboxyl group of the nucleophilic glutamate (*48*). This interaction is highly stabilizing in hydrolysis, hence the reduction in activity of β-glycosidases upon removal of the equatorial C2-hydroxyl group of a glycosidic substrate (*17, 49*). Due to the axial C2-hydroxyl, a mannose molecule is not capable of a similar interaction. However, *P. horikoshii* BglB has two residues in the active site that are unique among other family 1 enzymes, such as *P. furiosus CelB*, which are likely to have an effect on catalysis. The first residue is the glutamine at position 77 in BglB, which corresponds to arginine 77 in CelB (Figure 3B). Arg

77 has been postulated to position (by an ionic interaction) the nucleophilic glutamate for catalysis (*25*). Replacement of Arg 77 by Gln can result in: (i) lost stabilization of the nucleophile, due to the significantly shorter amino acid side chain of glutamine and (ii) a different electronic balance, influencing the pK_a-values of the catalytic residues. The effect on the electronic balance could be further influenced by the second unique active site residue, which involves the aspartic acid at position 206 next to the catalytic glutamic acid. At this position CelB has an asparagine (206). While the aspartic and asparagine side chains are isosteric, this change may introduce an extra negative charge, when the aspartate is dissociated. Asp 206 can be expected to be protonated during glycosylation and dissociated during deglycosylation, due to the pK_a-cycling of the neighboring Glu 207, the catalytic acid/base (*50*). The net charge difference between the two active sites is thus up to 2 charges. In the direct vicinity of both residues (within a radius of 6 Å) there are no amino acids that could compensate for this charge difference. The role of the unique components in the active site of BglB in catalysis of pNp-glycosides was evaluated by exchange of the unique active site residues between the two enzymes, yielding two single mutants and a double mutant for each enzyme.

Characterization of mutants. Kinetic parameters and inhibition by PheImGlc. The two wild-type enzymes have very different kinetic natures. Whereas CelB has a high affinity for pNp glucosides, BglB has very low activity with a high affinity for mannosides. This difference is mainly caused by the active site differences between the two enzymes. As retaining enzymes, the hydrolysis of saccharides by BglB and CelB occurs by formation of a glycosyl- enzyme intermediate in which the glycosidic bond is hydrolyzed (*9*). Next, the glycosyl-enzyme bond is hydrolyzed. Since nitrophenol is an excellent leaving group, the parameters obtained in this study provide information about the latter, deglycosylating step of the reaction (*8*).

The replacement of arginine 77 by glutamine and asparagine 206 by aspartic acid in CelB resulted in significantly increased affinities for mannosides, coupled to a reduced turnover rate. All CelB mutants, however, are most efficient in the hydrolysis of glucosides. This indicates that besides the unique active site residues, the position of the other conserved active site, or second-shell residues contributes to the specificity differences between the two wild-type enzymes, similar to what has been concluded in other studies (*24, 51*). In BglB Q77R, the unfavorable catalytic effects might be the result of limited space in the BglB active site to accommodate the bulky arginine side-chain. The residue at position 206 is crucial for the deglycosylating step of the hydrolysis. Glu 207 positions and deprotonates the water molecule that attacks the anomeric centre of the glycosyl-enzyme. A deprotonated Asp 206 could remove the water molecule from its optimal position and thus decrease the catalysis rate. When an asparagine residue was introduced at the position of aspartic acid 206 in BglB, the hydrolysis rate increased ten-fold and the affinity for glucosides increased 30-fold. Clearly, CelB and BglB represent two different classes of family 1 glycosyl hydrolases. Whereas CelB has been optimized for high-turnover saccharide hydrolysis,

BglB is adapted for the specific, low rate catalysis of mannoside conversion, mostly likely because of its unique active site structure.

Inhibition by PheImGlc. Since the sensitivity of glycosyl hydrolases for transition state analogues is several orders of magnitude higher than for ground state-mimicking inhibitors, one can conclude that the enzymes have been primarily optimized to lower the activation energy to reach the transition state (*8, 28*). With a trigonal geometry at the center corresponding to C1, the inhibitor PheImGlc resembles the transition states (*43*) (Figure 5). Conceivably, differences in accommodating the phenyl substituent of the inhibitor between the mutants may also contribute to differences ion the inhibition. However, this option is not supported by the observed correlation of k_{cat} and $1/K_i$, which indicates that the mutated residues are directly involved in the stabilization of the transition states (Figure 6). The mutated residues seem to be involved in binding of equatorial C2-hydroxyls. The slope of the regression line (0.66) indicates that the mutated residues are involved in ground state substrate binding as well. This is supported by the data for activation free energy for turnover of pNp-Glc versus the free energy of binding for pNp-Glc (slope 1.4). Like pNp-Glc, hydrolysis of pNp-Gal gives a log k_{cat}/K_m versus log $1/K_m$ plot with a linear character, however with a slope of 2.5. Glucose and galactose are epimers with an equatorial and axial orientation of the C4-hydroxyl, respectively. The larger slope for galactose (2.5) than for glucose (1.4) indicates that in an active site, which has been optimized for an equatorial C4-hydroxyl on glucose (CelB) or mannose (BglB), an altered C4-hydroxyl interaction can be expected to put more weight on the C2-hydroxyl interaction during catalysis.

Interestingly, log k_{cat}/K_m versus log K_m for pNp-Man does not show such apparent correlation, which indicates that substrate binding and transition state stabilization are less coupled compared to glucose and galactose. Wild-type CelB has the highest activation free energy for the reaction, like for pNp-Glc and pNp-Gal. Interestingly, the activation free energy for wild-type BglB, BglB D206N, BglB Q77R|D206N and CelB R77Q|N206D is comparable, although they display different K_m-values for pNp-Man hydrolysis. This could indicate that the stabilization of the transition state is independent of the binding of the ground-state of the substrate, which can be expected for a substrate with an axial C2-hydroxyl.

Substrate binding in CelB. The catalytic roles of the arginine 77 and asparagine 206 residues in CelB-like family 1 hydrolases were studied in the CelB mutants. The arginine is highly conserved and has no interaction with the substrate, but has been postulated to aid in catalysis by positioning the nucleophilic glutamate (*25*). When Arg77 is replaced by a glutamine in CelB, the K_m values for glucosides and galactosides increase, and the mutant becomes more specific for mannosides. Besides the removal of a positive charge, the substitution of Arg 77 by the smaller glutamine also creates more space. This may entail minor local rearrangements, which could impair substrate binding and transition state stabilization. Indeed, the hydrolysis rates decreased about 15-fold, which is in agreement with the observed increased K_i for PheImGlc in this mutant. The second substitution in CelB, N206D, had a smaller effect on the K_m. Since aspartic acid and asparagine are

isosteric residues, differences in substrate binding and catalysis must result from the negative charge of the aspartic acid. Remarkably, N206D results in a catalytic profile, which is similar to R77Q. However, CelB N206D is much less inhibited by PheImGlc, which is taken to mean that the transition-state is less stabilized in this mutant. This interpretation is confirmed by the BglB mutant D206N, which showed 10-fold increased hydrolysis rates and a 5 orders of magnitude increased sensitivity towards PheImGlc compared to wild-type BglB.

The role of CelB Asn 206 in substrate binding can be deduced from an analysis of the CelB active site, which was found to be highly homologous to the one of 6-phospho-β-galactosidase LacG from *Lactococcus lactis* (*24*). In LacG the interactions between the enzyme and substrate have been identified from a LacG mutant co-crystalized with the hydrolysis product galactose-6-phosphate in the active site (*25*). The LacG residue Asn 159 corresponding to Asn 206 was found to interact with the C2 OH group of galactose-6-phosphate. Glucose and galactose have both an equatorial C2 hydroxyl group, whereas the C2 hydroxyl group in mannose is axial. Such an axial C2 hydroxyl group would be too distant from the Asp 206 residue for an interaction in CelB. Any Asn206 substitution can, therefore, be expected to have a greater impact on the hydrolysis of glucosides and galactosides, relative to mannoside hydrolysis. Indeed, we observed for the CelB N206D mutant that the Michaelis constants for hydrolysis of pNp-Glc and pNp-Gal increased 30-fold and 3-fold, respectively, while the K_m-value for pNp-Man conversion was halved. Thus, the N206D destabilizes binding of glucose and galactose, and stabilizes the one one of mannose, resulting in a more specific binding of mannose. However, the hydrolysis rates in this mutant were reduced about 10-fold for all substrates tested, most likely due to a decreased stabilization of the reaction intermediate. Previously, the contribution to catalysis of non-covalent interactions between enzyme and substrate have been analyzed in the family 1 β-glucosidase of *A. feacalis* through the use of deoxy glucosides and fluorine substituted glucosides (*17*). Of all interactions between glucose hydroxyl-groups and the enzyme, the equatorial 2-hydroxyl groupwas found to contribute most to the stabilization of the reaction intermediate. Additionally, the hydrolysis rate of a 2-deoxy-glucoside was comparable to that of the corresponding mannoside, supporting the structural observation that there should be no interaction between the enzyme and the 2-hydroxyl of a mannoside. The increased affinity for mannose in CelB N206D seems to result directly from the charged aspartic acid residue, since a previously studied CelB N206S mutant was found to have hydrolysis rates comparable to CelB N206D, but a significantly reduced affinity for pNp-Man (Table 3.3) (*32*). The higher affinity for pNp-Man in CelB N206D is caused, therefore, by the properties of the introduced aspartic acid. The Michaelis constants for the hydrolysis of pNp-Glc and pNp-Gal in CelB N206S were higher than in CelB N206D, indicating that the glycoside equatorial 2-hydroxyl groups and the aspartic acid interact with each other, albeit significantly less compared to the interaction with Asn206 in wild-type CelB.

The β-mannosidase BglB of *P. horikoshii* has several homologues in related *Pyrococcus* species. This might indicate that the low hydrolytic activity has an important function in the pyrococcal metabolism. The specificity for mannosides of this subfamily of family 1 glycosidases suggests a role in the synthesis or turnover of specific mannose containing biomolecules, such as the abundant mannosyl-glycerate (52). This specificity results directly from the unique active site residues as demonstrated by exchanging these residues with those present in the β-glucosidase CelB of *P. furiosus*. We have shown that in *P. furiosus* CelB the affinity for mannosides was increased at the cost of the turnover rate, while *P. horikoshii* BglB could be turned into an efficient β-glucosidase. An altered interaction between the glycoside-2-hydroxyl and the enzyme upon substrate binding could explain the differences in activity between these two enzymes.

3.5 Acknowledgements

References

1. Henrissat, B. (2001) *http://afmb.cnrs-mrs.fr/~pedro/CAZY/ghf.html*.
2. Vetter, J. (2000) *Toxicon 38*, 11-36.
3. Rask, L., Andreasson, E., Ekbom, B., Eriksson, S., Pontoppidan, B., and Meijer, J. (2000) *Plant Mol Biol 42*, 93-113.
4. Voorhorst, W., Gueguen, Y., Geerling, A., Schut, G., Dahlke, I., Thomm, M., Van der Oost, J., and De Vos, W. (1999) *J Bacteriol 181*, 3777-3783.
5. de Vos, W. M., Boerrigter, I., van Rooyen, R. J., Reiche, B., and Hengstenberg, W. (1990) *J Biol Chem 265*, 22554-60.
6. Fernandez, P., Canada, F. J., Jiminez-Barbero, J., and Martin-Lomas, M. (1995) *Carbohydr Res 271*, 31-42.
7. Bauer, M. W., Bylina, E. J., Swanson, R. V., and Kelly, R. M. (1996) *J Biol Chem 271*, 23749-55.
8. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.
9. White, A., and Rose, D. R. (1997) *Curr Opin Struct Biol 7*, 645-51.
10. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663-75.
11. Sidhu, G., Withers, S. G., Nguyen, N. T., McIntosh, L. P., Ziser, L., and Brayer, G. D. (1999) *Biochemistry 38*, 5346-54.
12. Panday, N., Canac, Y., and Vasella, A. (2000) *Helv Chim Acta 83*, 58-79.
13. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J Bacteriol 177*, 7105-11.
14. Wang, Q., Trimbur, D., Graham, R., Warren, R. A., and Withers, S. G. (1995) *Biochemistry 34*, 14554-62.
15. Moracci, M., Capalbo, L., Ciaramella, M., and Rossi, M. (1996) *Protein Eng 9*, 1191-5.
16. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958-66.
17. Namchuk, M. N., and Withers, S. G. (1995) *Biochemistry 34*, 16194-202.
18. Rivera-Sagredo, A., Canada, F. J., Nieto, O., Jimenez-Barbero, J., and Martin-Lomas, M. (1992) *Eur J Biochem 209*, 415-22.
19. Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., and Hughes, M. A. (1995) *Structure 3*, 951-60.
20. Wiesmann, C., Beste, G., Hengstenberg, W., and Schulz, G. E. (1995) *Structure 3*, 961-8.

21. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.
22. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.
23. Chi, Y. I., Martinez-Cruz, L. A., Jancarik, J., Swanson, R. V., Robertson, D. E., and Kim, S. H. (1999) *FEBS Letters 445*, 375-383.
24. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.
25. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.
26. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.
27. Gueguen, Y., Voorhorst, W. G., van der Oost, J., and de Vos, W. M. (1997) *J Biol Chem 272*, 31258-64.
28. Bauer, M. W., and Kelly, R. M. (1998) *Biochemistry 37*, 17170-8.
29. Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., and Kikuchi, H. (1998) *DNA Res 5*, 147-55.
30. Kawarabayasi, Y. (2001) *Methods Enzymol 330*, 124-34.
31. Matsui, I., Sakai, Y., Matsui, E., Kikuchi, H., Kawarabayasi, Y., and Honda, K. (2000) *FEBS Lett 467*, 195-200.
32. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2001) *Methods Enzymol 330*, 364-79.
33. Guex, N., Diemand, A., and Peitsch, M. C. (1999) *Trends Biochem Sci 24*, 364-7.
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thronton, J. M. (1993) *J Appl Crystal 26*, 283-91.
35. Sippl, M. J. (1993) *proteins 17*, 355-62.
36. Higuchi, R., Krummel, B., and Saiki, R. K. (1988) *Nucleic Acids Res 16*, 7351-67.
37. Gill, S. C., and von Hippel, P. H. (1989) *Anal Biochem 182*, 319-26.
38. Kaper, T., Verhees, C. H., Lebbink, J. H. G., van Lieshout, J. F. T., Kluskens, L. D., Ward, D. E., Kengen, S. W. M., Beerthuyzen, M. M., de Vos, W. M., and van der Oost, J. (2001) *Methods Enzymol 330*, 329-46.
39. Wong, J. T. (1977) *Kinetics of enzyme mechanisms*, 2nd ed., Academic press, London.
40. Cubellis, M. V., Rozzo, C., Montecucchi, P., and Rossi, M. (1990) *Gene 94*, 89-94.
41. Sali, A. (2000), http://guitar.rockefeller.edu/pg/.
42. Joshi, M. D., Sidhu, G., Pot, I., Brayer, G. D., Withers, S. G., and McIntosh, L. P. (2000) *J Mol Biol 299*, 255-79.
43. Ermert, P., Vasella, A., Weber, M., Rupitz, K., and Withers, S. G. (1993) *Carbohydr Res 250*, 113-128.
44. Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.
45. Lebbink, J. H. G. (1999) PhD Thesis pp 200, Wageningen University, Wageningen.
46. Moracci, M., Ciaramella, M., Cobucci-Ponzano, B., Perugino, G., Rossi, M. (1998) *International Conference on Thermophiles 1998, Brest, France*.
47. Davies, J. D., Mackenzie, L., Varrot, A., Dauter, M., Brzozowoski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707-13.
48. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) *Nat Struct Biol 5*, 812-818.
49. McCarter, J. D., Adam, M. J., and Withers, S. G. (1992) *Biochem J 286*, 721-727.
50. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Körner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958-9966.
51. Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausín, C., Querol, E., Planas, A., and Pérez-Pons, J. A. (2001) *Biochemistry 40*, 5975-82.

52.   Da Costa, M. S., Santos, H., and Galinski, E. A. (1998) in *Advances in Biochemical Engineering Biotechnology* (Atranikian, G., Ed.) pp 117-54, Springer-Verlag, Berlin.
53.   Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res 22*, 4673-80.
54.   Page, R. D. C. (2000 Treeview 1.64b),, Glasgow, UK.
55.   Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714-2723.
56.   Cason, C. (1999), The Persisitant of Vision Development Team, Indianapolis, IN, USA.

4

Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal β-glycosidases

Authors
Jeroen Pouwels
Marco Moracci
Beatrice
  Cobucci Ponzano
Giuseppe Perugino
John van der Oost
Thijs Kaper
Joyce H.G. Lebbink
Willem M. de Vos
Maria Ciaramella
Mosé Rossi

Abstract
Sβgly and CelB are well-studied hyperthermophilic glycosyl hydrolases, isolated from the archaea *Sulfolobus solfataricus* and *Pyrococcus furiosus*, respectively. Previous studies revealed that the two enzymes are phylogenetically related, they are very active and stable at high temperatures, and their overall 3-D structure is very well conserved. In order to get insight in the molecular determinants of thermostability and thermoactivity of these enzymes, we have performed a detailed comparison, under identical conditions, of enzymological and biochemical parameters of Sβgly and CelB, and we have probed the basis of their stability by perturbations induced by temperature, pH, ionic strength and detergents. The major result of the present study is that, whereas the two enzymes are remarkably similar with respect to kinetic parameters, substrate specificity and reaction mechanism, they are strikingly different in the stability to the different physical or chemical perturbations induced. These results provide useful information for the design of further experiments aimed at understanding the structure-function relationships in these enzymes.

4.1 Introduction

Hyperthermophilic enzymes show a high thermal stability as well as an increased resistance to common protein denaturants such as organic solvents, detergents, extremes of pH. These enzymes are not only of potential interest in many biotechnological applications, they also are ideal models for the study of the structural basis of protein stability (*1*).

Despite extensive research on this topic, no general strategies for protein (thermo)-stabilization have been found up to now. Thermal stability is currently believed to be due to cumulative effects of small, local modifications within the protein. Several controversial features have been proposed to be responsible for high stability, such as increased rigidity, structural resilience, substitution of labile amino acids, hydrophobic interactions, hydrogen bonds, increased packing density, buried solvent molecules, oligomerization, and others (*2*) and references therein).

Recently the resolution of several 3-D structures has shown that a high number of salt bridges, mainly at subunit interface and often arranged in large networks, is a common feature of hyperthermophilic proteins (*3-16*) with only few exceptions (*17, 18*). However, functional studies on the role of ionic networks in hyperthermophilic enzymes have produced contradictory results (*19-21*) showing that their contribution to protein stabilization cannot be easily predicted. Moreover, data available for different thermophilic proteins have often been obtained under very different experimental conditions, therefore it is difficult to extrapolate generally applicable rules.

In the present study, we have focused our attention on two glycosyl hydrolases, Sβgly and CelB, isolated from the hyperthermophilic archaea *S. solfataricus* and *P. furiosus*, respectively (*22, 23*). *S. Solfataricus* is an extremely thermo-acidophilic aerotolerant *Crenarchaeon* which grows optimally at 87 °C, and pH 3.0; *P. furiosus* belongs to the *Euryarchaeota*, is strictly anaerobic, grows optimally at 100 °C, and pH 7.0.

CelB and Sβgly share a number of features: they both belong to the family 1 of glycosyl-hydrolases, their aminoacid sequence are 55% identical, and they have comparable masses (220 and 240 kDa, respectively) and about the same isoelectric point (4.4 and 4.5, respectively). Both are intracellular homo-tetramers, are only active in the tetrameric form, and do not contain cysteine bridges. They catalyze the hydrolysis of β-O-glycosidic bonds following a retaining mechanism, and show a broad substrate specificity, being active, although with different efficiency, on β-Glcco-, β-fuco-, β-galacto- and β-xylo-sides, as well as on Glccose oligmers of moderate length (*24-29*). between the two enzymes, concerning mainly optimal temperatures (102-105 °C for CelB and 95 °C for Sβgly), optimal pH's (5.0 for CelB and 6.5 for Sβgly), and thermostability (half lives of 85 hours at 100 °C for CelB and 15 hours at 85 °C for Sβgly).

The 3-D structure of Sβgly, solved at 2.6 Å (*12*), showed the $(\beta\alpha)_8$-barrel typical of many glycosyl-hydrolases; however two peculiar features compared to mesophilic counterparts concern a high amount of buried solvent molecules, and a high number of salt bridges. In particular, 58% of the charged residues are involved in ion-pairs, and 60% of these ion-pairs are part of a multiple

network. The contribution of these interactions to thermal stability of Sβgly has not been tested experimentally.

Unfortunately at present the 3-D structure of CelB is only known at a resolution of 3.3 Å (*30*). The overall folding of the molecule is similar to that of Sβgly, but at this low resolution it is not possible to pinpoint the presence and distribution of ion-pairs. Preliminary modeling of the structure of CelB on that of Sβgly suggests that the most peculiar difference concerns an extensive ion-pair network occuring at the center of the Sβgly tetramer, which involves six amino acids at the C-terminal end (*12*); such residues are not conserved in CelB, whose C-terminal region forms a different structure (*29*).

Due to their structure/function similarities, and because of the considerable amount of enzymological, chemical-physical, and structural data available, CelB and Sβgly are suitable models to elucidate which factors are important in their stability and activity. However, a reliable comparison of the two enzymes is hampered by the fact that most of the data available have not been obtained under similar conditions. Here we present a detailed parallel comparison of CelB and Sβgly activity and stability. The implications of the results for the understanding of the similarities and differences of the two enzymes are discussed.

4.2 Material and methods

Chemicals and solutions. All chemicals were of analytical grade, purchased from Sigma (St. Louis USA).

Enzyme purifications. For production of Sβgly a new expression vector was constructed. For this purpose NdeI and NcoI sites were introduced by PCR amplification to the 5' and 3' end, respectively, of the *lacS* coding sequence (*31*). The resulting fragment was cloned in the NcoI-NdeI sites of pET29 (Novagen), producing plasmid pETGly29. *Escherichia coli* BL21 (DE3) cultures transformed with pETGly29 were grown in LB medium supplemented with 30 μg/ml Kanamycin at 1.0 OD_{600}, induced with 0.5 mM IPTG and grown overnight. Cells were broken with French Press; the lysate was clarified by centrifugation at 15.000 rpm for 30', and was heated for 30' at 70 °C. Heat-denatured host proteins were eliminated by centrifugation at 15.000 rpm for 30', supernatant was loaded on a FPLC phenyl-sepharose column in 50 mM phosphate buffer, pH 6.8, 1 M $(NH_4)_2SO_4$, and eluted with 50 mM phosphate buffer, pH 6.8.

For CelB a comparable construct was generated in a pET9d vector (Novagen) (*32*). After transformation into *E. coli* BL21 (DE3), cells were grown in LB medium supplemented with 30 μg/ml Kanamycin (without IPTG induction), and CelB was purified as described by Voorhorst et al. (*28*), with the exception that 20 mM sodium citrate buffer, pH 5.4, instead of phosphate buffer was used. Both enzymes were homogeneous as judged by SDS-PAGE (not shown). Protein concentrations were determined by the method of Bradford (*33*).

Determination of the pH optima. Buffers used were: 50 mM sodium citrate at pH 3.4, 3.9, 4.4, 5.0, and 5.3; 50 mM phosphate at pH 5.8, 6.3, 6.5, 6.9, and 7.9; and 50 mM KCl Borate at pH 8.0, 8.5, 9.1, 9.5, and 10.0. Reaction mixtures (final volume 1.0 ml), containing 50 mM of one of the buffers reported above and 5 mM para-nitrophenyl-β-D-galactopyranoside (pNp-Gal) were pre-heated for 2' at the desired temperature, then 0.05-5 µg of enzyme was added and the reaction was allowed to proceed for 10'. Reactions were stopped by adding 150 µl of the reaction mixture to 1350 µl of a cold 0.33 M Na_2CO_3 solution. The $\varepsilon_{mM,405}$ for para-nitrophenol (pNp) was 18,300 $M^{-1}cm^{-1}$. The OD_{405} was measured at room temperature in a Varian DMS 2000 spectrophotometer (Varian, Australia); one unit of enzyme activity was defined as the amount of enzyme catalyzing the hydrolysis of 1 µmol of substrate in 1 minute at the indicated temperature. Data were plotted and refined using the program Grafit (*34*).

Kinetic parameters. Reaction mixtures (total volume 1 ml) contained: 50 mM sodium citrate buffer (pH 5.4), varying concentrations (0.1-20 mM) of substrate: para-nitrophenyl-β-D-galacto- and gluco-pyranoside (pNp-Gal and pNp-Glc); orthonitrophenyl-β-D-galacto- and gluco-pyranoside (oNp-Gal and oNp-Glc). They were preheated for 2', then the desired amount of enzyme (0.1-5 µg) was added and the change in absorbance at 405 nm was followed for 2' with a Cary 1E UV-Visible spectrophotometer (Varian, Australia), equipped with a circulating waterbath. A blank mixture containing all the reactants except the enzyme was used to correct for spontaneous degradation. To calculate the kinetic parameters, the molar extinction coefficients of pNp were measured at 405 nm in 50 mM sodium citrate buffer (pH 5.4) at all temperatures used in this experiment. The resulting molecular extinction coefficients were: 720 M^{-1} cm^{-1} (30 °C), 920 M^{-1} cm^{-1} (40 °C), 1150 M^{-1} cm^{-1} (50 °C), 1450 M^{-1} cm^{-1} (60 °C), 1650 M^{-1} cm^{-1} (65 °C), 2020 M^{-1} cm^{-1} (75 °C) and 2640 M^{-1} cm^{-1} (85 °C). The molar extinction coefficient of ortho-nitrophenol (oNp) at 75 °C in 50 mM sodium citrate buffer (pH 5.4) was 704 M^{-1} cm^{-1}. The program Grafit was used to plot the results and determine the K_M and k_{cat}.

Inactivation measurements. Inactivation by temperature was determined by incubating enzymes at 0.05 mg/ml in 50 mM sodium citrate buffer, pH 5.4 (final volume 500 µl) for 30' at different temperatures in a Delphi 1000 thermocycler (MJ Research, Watertown, USA). 5 µl of the incubation mixture were assayed for activity in 50 mM sodium citrate buffer, pH 5.4, saturating concentrations of pNp-Gal (10 mM for CelB and 5 mM for Sβgly, respectively); final volume was 1 ml; activity was measured following the change in absorption for 2' at 65 °C (standard activity assay). Protein aggregation was determined by measuring the light scattering with excitation and emission wavelength of 480 nm in a Jasco FP-777 spectrofluorometer (Jasco, Tokyo, Japan). Quartz cuvettes with a path length of 10 mm were used (Starna Ltd, Romford, UK).

For inactivation by salts, the enzymes were incubated in 50 mM sodium citrate buffer, pH 5.4 containing different concentrations of either NaCl or Na_2SO_4 (final volume 500 µl); samples

were incubated for 30' at 75 °C, centrifuged for 10' at 13000 rpm, and 5 µl of the supernatant were assayed for activity and aggregation as described above.

For inactivation by SDS, the enzymes were incubated in in 50 mM sodium citrate, pH 5.4 containing different SDS concentrations at 50 °C for 30'; for inactivation by extreme pH, the enzymes were incubated either in 50 mM sodium citrate, pH 3.4 or in 50 mM HCl borate, pH 10.0 at 75 °C for increasing time spans. After incubations in the presence of denaturants, 5 µl aliquots were were transferred to no-denaturing conditions and assayed for activity and aggregation as described above.

4.3 Results and discussion

Determination of the pH optima. Previously reported pH optima are 6.5 for Sβgly, determined at 75 °C *(22)* and 5.0 for CelB, determined at 80 °C *(23)*. We have now determined the activity dependence on pH of the two enzymes at 65 °C and 75 °C on 5 mM pNp-Gal, using a modification of the previously reported procedure: since the $\varepsilon_{mM,405}$ of pNp significantly varies with pH and temperature, after the end of the reaction the pH of all samples was adjusted to 10.0 and the absorption was determined at room temperature. Although the specific activity of CelB was about three times that of Sβgly at both temperatures tested, the two enzymes showed similar bell-shaped curves, and the same optimum at pH 5.4, which was temperature-independent (Figure 4.1A, B).

The discrepancy with previously reported data could be attributed to the different assay procedure used in the present study, since other conditions, such as buffer composition and concentration were not changed.
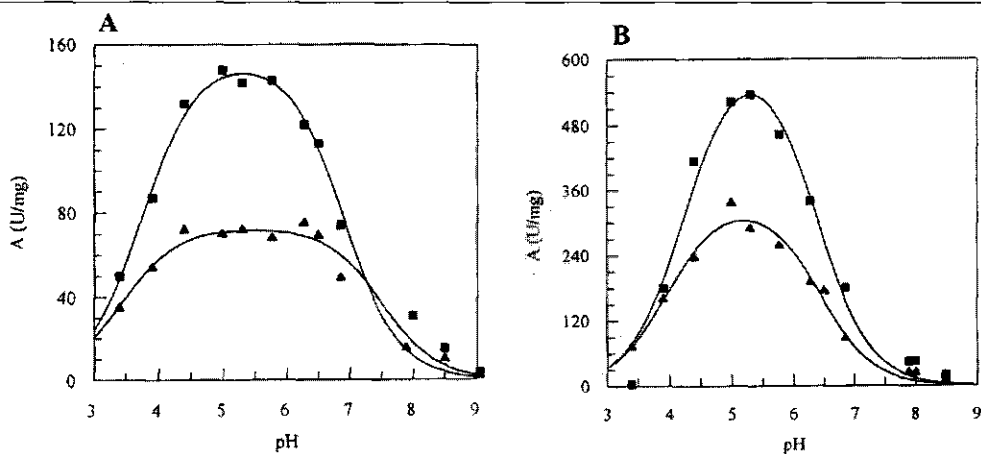
Figure 4.1 The pH-optima of Sβgly (A) and CelB (B), determined at 75 °C (closed squares) and 65 °C (triangles).

Kinetic studies. We have determined K_M and k_{cat} values for the hydrolysis of pNp-Gal by Sβgly and CelB at different temperatures (Table 4.1). For both enzymes the K_M increases with temperature: the increase is moderate at lower temperatures (from 30 °C to 65 °C) and more evident between 65 °C and 85 °C. For both enzymes the k_{cat} increases exponentially with temperature confirming the thermophilicity of the two enzymes. Arrhenius plots were derived from the above reaction, and used to calculate thermodynamic parameters. The energy of activation was constant over the whole range of temperatures and was comparable for both enzymes (65 kJ/mol for CelB and 71 kJ/mol for Sβgly). Also ΔG, ΔH and ΔS were very similar for the two enzymes and were temperature-independent (not shown). The efficiency (k_{cat}/K_M) of the reaction catalized by the two enzymes increases with temperature (Figure 4.2). However, whereas at lower temperatures (30 °C to 50 °C) the behaviour is the same, between 50 °C and 65 °C the efficiency of Sβgly increases more than that of CelB; Sβgly reaches its maximal efficiency at 65 °C, while the k_{cat}/K_M of CelB is still increasing at 85 °C.

Both enzymes have the ability to hydrolyze different β-glycosides; in order to compare their substrate specificity we have determined the K_M and k_{cat} on different substrates at 75 °C and pH 5.4 (Table 4.2). The reaction of Sβgly on oNp-Glc and pNp-Glc showed a biphasic behavior in Lineawever-Burk plots (not shown), suggesting that transglycosylation reactions occur at higher substrate concentrations (*35*); kinetic constants obtained at lower (I) and higher (II) substrate concentrations were calculated (Table 4.2). No such evidence was found in all other reactions under the conditions used. Data in Table 4.2 show that the substate specificity of the two enzymes is similar, and that they are slighltly more active on aryl-glucosides than on aryl-galactosides.

Table 4.1 The K_M and k_{cat} values of CelB and Sβgly on pNp-Gal at different temperatures. Assays were performed in 50 mM sodium citrate buffer, pH 5.4

| T | CelB | | Sβgly | |
(°C)	K_M (mM)	k_{cat} (s^{-1})	K_M (mM)	k_{cat} (s^{-1})
30	1.11	90	0.59	56
40	1.31	236	0.82	143
50	2.05	580	1.09	293
60	2.09	1222	0.72	648
65	1.97	1486	0.77	1000
75	2.80	2907	1.61	2288
85	4.10	4863	3.16	4380
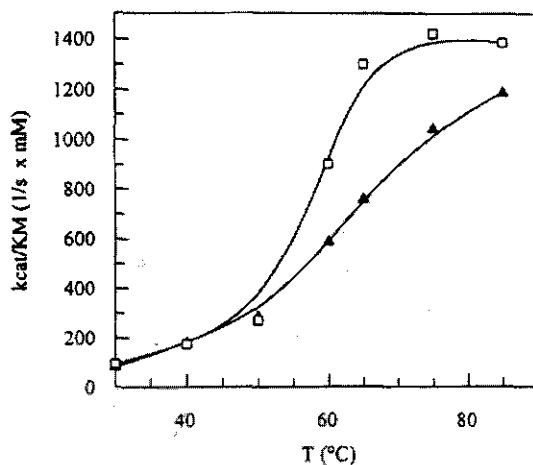
Figure 4.2 The temperature dependence of K_M/k_{cat} for Sβgly (squares) and CelB (triangles).

Thermostability. In order to compare the intrinsic stability of the two enzymes we followed enzyme inactivation after incubations for 30' at different temperatures at pH 5.4 (Figure 4.3). As expected Sβgly was far less stable: at 88 °C it showed 50% activity, while CelB was completely stable up to the highest temperature used (97.5 °C). Inactivation of Sβgly was coupled to irreversible aggregation. Aggregation of CelB was negligible (Figure 4.3). Again, the lower stability of Sβgly reflects the lower growth temperature of *S. solfataricus* with respect to *P. furiosus*; marginal stability at physiological conditions is of great importance for biological function (*36*). Data in Figure 4.3 also show that Sβgly and CelB incubated for 30' at 75 °C are stable enough (90% and 100% residual activity, respectively) for following inactivation/denaturation experiments.

Perturbation by salts. Salts may be used to test the type of interactions involved in the stabilization of model proteins since they affect protein stability in different ways (*37, 38*). Chaotropic salts, big ions with low charge density (like I^-, ClO_4^-, $NO3^-$, and Cs^+), destabilize proteins. On the other hand, cosmotropic salts, small ions of high charge density (like F^-, $SO4^{2-}$, and acetate) are known to increase protein stability and decrease their solubility by strengthening hydrophobic interactions. Cl^- has neither a stabilizing nor a destabilizing effect on hydrophobic interactions, but might affect ionic interactions. In preliminar experiments we followed inactivation and aggregation in time in the presence of NaI, $NaNO_3$, NaCl and Na_2SO_4, at a concentration of 4 I (ionic strenght equivalents). As expected, for both enzymes inactivation and aggregation with chaotropes was much faster than with other salts (data not shown).

Table 4.2 Kinetic constants for CelB and Sβgly on different substrates, measured in 50 mM sodium citrate buffer, pH 5.4, at 75 °C. For the reaction of Sβgly with aryl-glucosides kinetic constants obtained at lower (I) and higher (II) substrate concentrations are reported.

Substrate	CelB			Sβgly					
	K_M (mM)	k_{cat} (s^{-1})	k_{cat}/K_M (mM^{-1}s^{-1})		K_M (mM)		k_{cat} (s^{-1})		k_{cat}/K_M (mM^{-1}s^{-1})
pNp-Gal	2.80	2900	1000		1.61		2300		1400
oNp-Gal	2.57	9600	3700		0.87		2400		2700
pNp-Glc	0.37	2600	7100	(I)	0.026	(I)	520	(I)	20000
				(II)	1.32	(II)	1000	(I)	7600
oNp-Glc	0.25	3300	12900	(I)	0.053	(I)	702	(I)	13250
				(II)	2.76	(II)	1600	(I)	580

Then CelB and Sβgly were incubated for 30' at 75 °C and pH 5.4 at several NaCl concentrations, and both activity and aggregation were followed (Figure 4.4A); both enzymes were inactivated by NaCl, but Sβgly was more sensitive than CelB. Inactivation was coupled to aggregation, which was more evident for Sβgly. Aggregation was at least partially reversible since the aggregates could be resuspended and partially reactivated (data not shown), suggesting that some active protein is aggregated and then inactivated. In 4 M NaCl aggregation occurred at every protein concentration starting from 0.001 mg/ml (data not shown). We performed a similar experiment using Na_2SO_4 (Figure 4.4B); in this case the two enzymes showed more remarkable differences. Sβgly was inactivated and aggregated, although aggregation was less pronounced than with NaCl at the same ionic strenghts; in contrast, CelB only showed a slight inactivation (80% residual activity at the highest ionic strenght used) and very little aggregation.

Taken together, our results are consistent with the different nature of the salts: chaotropes have strong destabilizing effect on the proteins by perturbing hydrophobic interactions. On the other hand, the denaturing effect of NaCl and Na_2SO_4 suggests that ionic interactions are also involved in maintaining the native conformation of both enzymes at high temperature, and in particular in Sβgly, which is more sensitive to inactivation by these salts. A behaviour similar to that of Sβgly has been reported for a *S. solfataricus* carboxypeptidase (*39*). Unfortunately, the occurrence of aggregation precludes from quantitative conclusions and from CD and fluorescence analysis to test if inactivation by salts is caused by unfolding.

Perturbation by extreme pH. Extreme pH values are strong protein denaturants since charge perturbations normally displace stabilizing interactions; as such, pH may be useful to test the importance of ion-pairs in proteins.

In order to test the effect of low pH on protein activity and stability, CelB and Sβgly were incubated at 75 °C and pH 3.4, and inactivation was determined after different time periods. Both enzymes were very sensitive to inactivation by acidic pH: their half life was less than 5', and they aggregated irreversibly (data not shown). When analyzed by SDS-PAGE after incubations, both

enzymes showed several bands migrating faster than the monomers, suggesting that inactivation by acidic pH is due to hydrolysis of the polypeptide chains (data not shown). The occurrence of hydrolysis, which is often observed in acidic pH at high temperatures (*1*) impaired the analysis of the effect of such pH on the stability of CelB and Sβgly.

CelB and Sβgly were then incubated at 75 °C at pH 10 and inactivation was determined after different time spans. The two enzymes showed striking differences under these conditions: Sβgly was very sensitive to inactivation (half life less than 5'), whereas CelB was completely active after 90' of incubation at pH 10.0 (Figure 4.5). CD analysis showed that inactivation of Sβgly is not due to loss of secondary structure (data not shown); D'Auria et al have previously reported that at alkaline pH Sβgly is highly destabilized and shows remarkable changes in tertiary structure (*40, 41*).
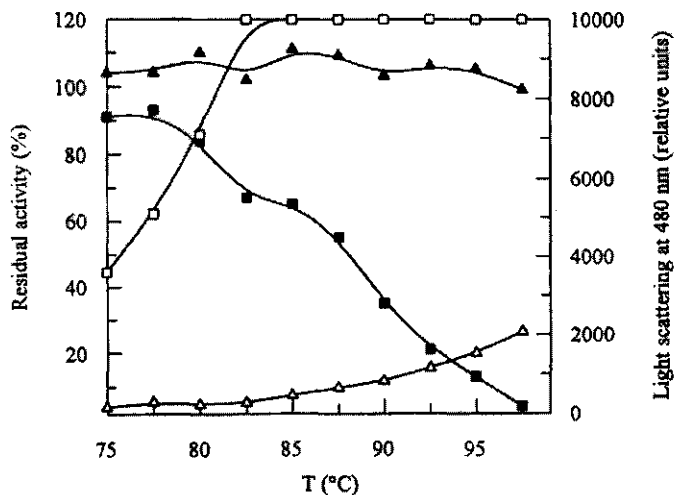
Figure 4.3 Inactivation and aggregation of CelB and Sβgly after incubation at different temperatures. Incubations were performed for 30' in 50 mM sodium citrate buffer, pH 5.4. Protein concentration was 0.05 mg/ml. Standard activity assays were performed as described in: Materials and Methods. Residual activities are shown as fractions of the activity of a non-incubated sample; samples were not centrifuged before the assay, therefore they may contain some aggregated protein. Closed squares: Sβgly activity; open squares Sβgly aggregation; closed triangles, CelB activity; open triangles, CelB aggregation. The highest value of aggregation reached the limit of sensitivity of the spectrofluorometer under the conditions used, therefore does not necessarily indicate saturation.
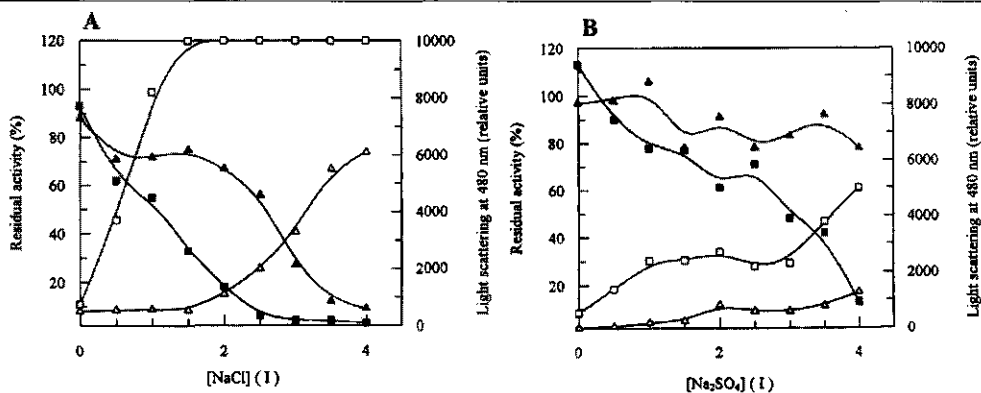
Figure 4.4 Inactivation and aggregation of CelB and Sβgly at different concentrations of NaCl (A) and Na$_2$SO$_4$ (B). Incubations were performed for 30' at 75 °C in 50 mM sodium citrate buffer, pH 5.4. Protein concentration was 0.05 mg/ml. Activity of each sample was expressed as fraction of the activity of the same sample measured at t=0, taken as 100%; residual activities of the supernatants after centrifugation are reported. Closed squares: Sβgly activity; open squares Sβgly aggregation; closed triangles, CelB activity; open triangles, CelB aggregation. See also the legend to Figure 4.3 for comments to the aggregation values.

Perturbation by SDS. The effects of SDS on the structure and stability of Sβgly have been extensively studied (*40, 42*). SDS caused remarkable changes in the tertiary, but not secondary structure; in addition, at moderate concentrations (0.05- 0.1 %), SDS induced enzyme activation. For these reasons it has been suggested that SDS increases the protein flexibility. In contrast CelB has been reported to be completely and irreversibly inactivated by 0.05% SDS in 50 mM citrate buffer, pH 5.0; SDS-PAGE showed that the enzyme dissociated into the 58 kDa subunits (*43*).

We have determined inactivation of Sβgly and CelB by incubating the enzymes for 30' at 50 °C at different concentrations of SDS in 50 mM sodium citrate buffer pH 5.4 (Figure 4.6). Strikingly, the two enzymes showed a very different behaviour: CelB was very sensitive to SDS (50% residual activity in 0.03%, less then 10% in 0.04%), whereas Sβgly showed a slight activation in 0.03%, and then a moderate inactivation, but residual activity was still 60% in 0.3% SDS, the highest concentration used. Values were unchanged after up to 4.5 hours of incubation, indicating that the equilibrium was already reached in 30' (not shown). No aggregation occurred in these conditions.

When the enzymes were incubated at 50 °C in 0.1% SDS for one hour, CelB had 2% residual activity and Sβgly about 60%, yet secondary structure tested by far-UV CD spectra was unchanged for both enzymes, indicating that inactivation was not due to complete unfolding of the monomers (not shown). Moreover, analysis by SDS-PAGE showed that after incubation for 10' in

0.1% SDS at 50 °C CelB was completely dissociated into monomers, while Sβgly was about half in the tetrameric and half in the monomeric form (not shown). We conclude that in the presence of SDS both enzymes dissociate into monomers, which maintain significant secondary structure, but are inactive. The fact that Sβgly is much more resistant than CelB to the inactivation and dissociation suggests that interactions less affected by SDS play an essential role in the maintainance of the quaternary structure of the former enzyme. In particular, we note that a large ion-pair network in Sβgly, which occurs at the interface of the monomers (*12*), could stabilize the tetramer even when other interactions are broken. Preliminary observations from molecular modelling showed that the corresponding region of CelB could form a different structure (Kaper at al. unpublished results); this hypothesis should be confirmed by high resolution of the 3-D structure of CelB, which is currently lacking.
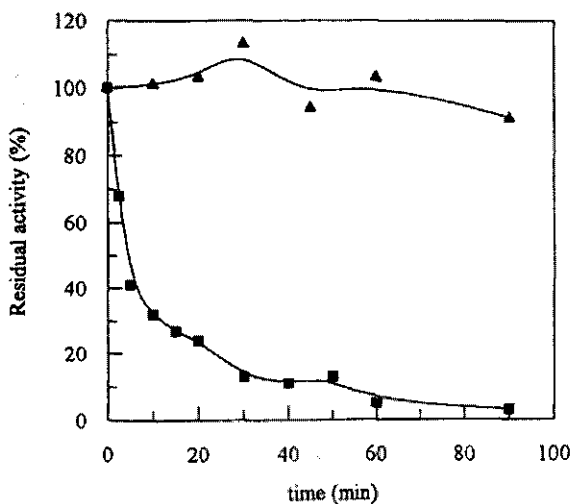
Figure 4.5 Inactivation of CelB (triangles) and Sβgly (squares), when incubated for different time periods at pH 10.0. Incubations were performed at 75 °C in KCl borate buffer, pH 10.0. Activity of each sample was expressed as fraction of the activity of the same sample measured at t=0, taken as 100%. Protein concentration was 0.05 mg/ml.
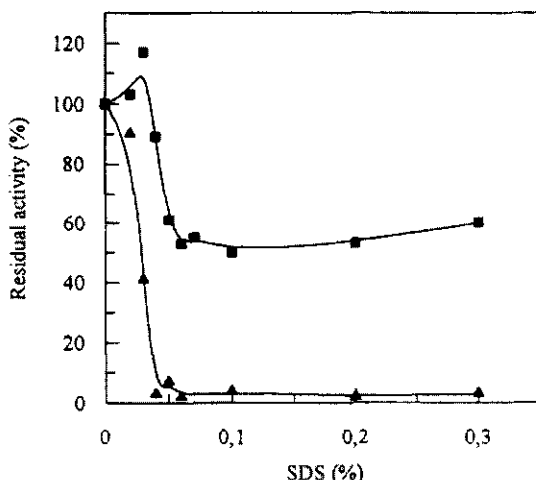
Figure 4.6 Inactivation of CelB (triangles) and Sβgly (squares) by different concentrations of SDS. Samples were incubated for 30' at 50 °C in 50 mM sodium citrate buffer, pH 5.4. Protein concentration was 0.05 mg/ml. The activity of a sample incubated in the same conditions but without SDS was taken as 100%.

4.4 Conclusions

From the comparison of Sβgly and CelB we can conclude that they show only limited differences in enzyme kinetics, substrate specificity and pH optima. Previous data indicated that the reaction mechanism is the same in the two enzymes, and that catalytic residues are conserved (*27, 28*); present results suggest that there is a wider structural and functional conservation in all components implicated in catalysis.

In contrast, the two enzymes showed remarkable differences in their stability to different denaturing agents. CelB is far more stable than Sβgly to temperature, salts and high pH; on the other hand, CelB is more sensitive to inactivation by SDS. Taken together, the results of the present work suggest that the importance of various stabilizing interactions might be different in the two enzymes. Testing this hypothesis requires additional experiments including site-directed mutagenesis as well as the comparison of high resolution 3-D structures.

4.5 Acknowledgments

References

1.   Vieille, C., Burdette, D. S., and Zeikus, J. G. (1996) *Biotechnol Annu Rev 2*, 1-83.
2.   Jaenicke, R., and Bohm, G. (1998) *Curr Opin Struc Biol 8 (6)*, 738-748.
3.   Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J. B., Z.H., Z., and Adams, M. W. W. (1992) *Protein Sci 1*, 1494-1507.
4.   Russell, R. J. M., Hough, D. W., Danson, M. J., and Taylor, G. (1994) *Structure 2*, 1157-67.
5.   Tomschy, A., Bohm, G., and Jaenicke, R. (1994) *Protein Eng 7*, 1471-8.
6.   Chan, M. K., Mukund, S., Kletzin, A., and Adams, M. W. W. (1995) *Science 267*, 1463-69.
7.   Goldman, A. (1995) *Structure 3*, 1277-9.
8.   Hennig, M., Dairmont, B., Kirschner, K., and Jansonius, J. N. (1995) *Structure 3*, 1295-1306.
9.   Korndorfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995) *J Mol Biol 246*, 511-21.
10. Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., Consalvi, V., and et al. (1995) *Structure 3*, 1147-58.
11. Starich, M. R., Sandman, K., Reeve, J. N., and Summers, M. F. (1996) *J Mol Biol 255*, 187-203.
12. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.
13. Hennig, M., Sterner, R., Kirschner, K., and Jansonius, J. N. (1997) *Biochemistry 36*, 6009-16.
14. Knapp, S., De Vos, W. M., Rice, D., and Ladenstein, R. (1997) *J Mol Biol 267*, 916-32.
15. Lim, J. H., Yu, Y. G., Han, Y. S., Cho, S., Ahn, B. Y., Kim, S. H., and Cho, Y. (1997) *J Mol Biol 270*, 259-74.
16. Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J., and Taylor, G. (1997) *Biochemistry 36*, 9983-94.
17. Robinson, H., Gao, Y. G., McCrary, B. S., Edmonson, S. P., Shriver, J. W., and Wang, A. H. J. (1998) *Nature 392*, 392:202-5.
18. Usher, K. C., De La Cruz, A. A., Dahlquist, F. W., Swanson, R. V., Simon, M. I., and Remington, S. J. (1998) *Protein Sci 7*, 403-12.
19. Pappenberger, G., Schurig, H., and Jaenicke, R. (1997) *J Mol Biol 274*, 676-83.
20. Lebbink, J. H., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R., and de Vos, W. M. (1998) *J Mol Biol 280*, 287-96.
21. Merz, A., Knochel, T., Jansonius, J. N., and Kirschner, K. (1999) *J Mol Biol 288*, 753-63.
22. Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.
23. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.
24. Nucci, R., Moracci, M., Vaccaro, C., Vespa, N., and Rossi, M. (1993) *Biotechnol Appl Biochem 17*, 239-50.
25. Moracci, M., Ciaramella, M., Nucci, R., Pearl, L. H., Sanderson, I., Trincone, A., and Rossi, M. (1994) *Biocatalysis 11*, 89-103.
26. Moracci, M., Nucci, R., Febbraio, F., Vaccaro, C., Vespa, N., La Cara, F., and Rossi, M. (1995) *Enzyme Microb Technol 17*, 992-7.
27. Moracci, M., Capalbo, L., Ciaramella, M., and Rossi, M. (1996) *Protein Eng 9*, 1191-5.
28. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J Bacteriol 177*, 7105-11.
29. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.
30. Kopp, J. (1998).
31. Cubellis, M. V., Rozzo, C., Montecucchi, P., and Rossi, M. (1990) *Gene 94*, 89-94.
32. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2001) *Methods Enzymol 330*, 364-79.
33. Bradford, M. M. (1976) *Anal Biochem 72*.
34. Leatherbarrow, R. J. (1992), Grafit version 3.0, Erithacus Software Ltd, Staines, U.K.
35. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.
36. Somero, G. N. (1995) *Annu Rev Physiol 57*, 43-68.
37. Hatefi, Y., Hanstein, W.G. (1969) *Proc. Natl. Acad. Sci.USA 62*, 1129-36.
38. Von Hippel, P. H., Peticolas, V., Schack, L., and Karlson, L. (1973) *Biochemistry 12*, 1256-64.
39. Villa, A., Zecca, L., Fusi, P., Colombo, S., Tedeschi, G., and Tortora, P. (1993) *Biochem J 295*, 827-31.
40. D'Auria, S., Rossi, M., Nucci, R., Irace, G., and Bismuto, E. (1997) *Proteins 27*, 71-9.
41. D'Auria, S., Moracci, M., Febbraio, F., Tanfani, F., Nucci, R., and Rossi, M. (1998) *Biochimie 80*, 949-57.
42. D'Auria, S., Barone, R., Rossi, M., Nucci, R., Barone, G., Fessas, D., Bertoli, E., and Tanfani, F. (1997) *Biochem J 323*, 833-40.
43. Kengen, S., and Stams, A. (1994) *Biocatalysis 11*, 79-88.

5

Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* β-glucosidase CelB by directed evolution

Authors

Joyce H.G. Lebbink
Thijs Kaper
Peter Bron
John van der Oost
Willem de Vos

Abstract

The β-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* (CelB) is the most thermostable and thermoactive family 1 glycosyl hydrolase described to date. To obtain more insight in the molecular determinants of adaptations to high temperatures and study the possibility of optimizing low-temperature activity of a hyperthermostable enzyme, we generated a library of random CelB mutants in *Escherichia coli*. This library was screened for increased activity on *p*-nitrophenol-β-D-glucopyranoside at room temperature. Multiple CelB variants were identified with up to three-fold increased rates of hydrolysis of this aryl-glucoside, and ten of them were characterized in detail. Amino acid substitutions were identified in the active site region, at subunit interfaces, at the enzyme surface, and buried in the interior of the monomers. Characterization of the mutants revealed that the increase in low-temperature activity was achieved in different ways, including altered substrate specificity and increased flexibility by an apparent overall destabilisation of the enzyme. Kinetic characterization of the active site mutants showed that in all cases the catalytic efficiency at 20°C on *p*-nitrophenol-β-D-glucose as well as on the disaccharide cellobiose, was increased up to two-fold. In most cases, this was achieved at the expense of β-galactosidase activity at 20°C and total catalytic efficiency at 90°C. Substrate specificity was found to be affected by many of the observed amino acid substitutions, of which only some are located in the vicinity of the active site. The largest effect on substrate specificity was observed with the CelB variant N415S that showed a 7.5-fold increase in the ratio of *p*-nitrophenol-β-D-glucopyranoside/*p*-nitrophenol-β-D-galactopyranoside hydrolysis. This asparagine at position 415 is predicted to interact with active site residues that stabilize the hydroxyl group at the C4 position of the substrate, the conformation of which is equatorial in glucose-containing, and axial in galactose-containing substrates.

5.1 Introduction

Carbohydrate polymers can be degraded by a wide variety of glycosyl hydrolases that have been classified into more than 70 different families based on amino acid sequence comparisons (http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html) *(1,2)*. In recent years, considerable progress has been achieved in the determination and characterization of primary and three-dimensional structures from family 1 glycosyl hydrolases. From among these enzymes isolated from mesophiles, the crystal structures have been solved of the 6-phospho-β-galactosidase (LacG) from *Lactococcus lactis* and the β-glucosidase from *Bacillus polymyxa (3,4,5)*. Both of these enzymes have been cocrystallized with either the substrate or an inhibitor in the active site and therefore allow analysis of the interactions between the enzyme and its substrate. Biochemical characterization and primary structure determination have been reported for several family 1 glycosyl hydrolases from hyperthermophilic organisms, that optimally grow at and around the normal boiling point of water. These include the β-glucosidase from *Thermotoga maritima (6)*, the β-glycosidase LacS from *Sulfolobus solfataricus (7,8)* and the β-glucosidase CelB and β-mannosidase BmnA from *Pyrococcus furiosus (9,10,11)*. The crystal structure of LacS as well as the structure of the β-glycosidase from *Thermosphaera aggregans* have recently been reported *(12,13)*. A three-dimensional model for CelB, based on 3.3 Å X-ray diffraction data, is available *(39, 40)*. The comparison of these structures with the homologous structures from the mesophiles, allows for studying of the molecular adaptations of family 1 enzymes to high temperatures.

We have focused on the most thermostable member of the family 1 glycosyl hydrolases, CelB from the archaeon *P.furiosus*. CelB was purified from *P.furiosus*, characterized, and found to be a tetrameric enzyme of 230 kD with identical 58 kD subunits *(10)*. The enzyme is very well adapted to the high growth optimum of the organism (100°C), with a half-life for thermal inactivation of 85 hours at 100°C and an optimum temperature for activity of 102-105°C. CelB shows high activity on the aryl-glucosides *p*-nitrophenol-β-1,4-D-glucopyranoside (pNp-Glc) and pNp-β-1,4-D-galactose (pNp-Gal), as well as on β-(1,4)-linked disaccharides cellobiose and lactose and on the β-(1,3)-linked disaccharide laminaribiose *(10,14)*. Low activity was detected on pNp-mannose and pNp-xylose. The *celB* gene was cloned, sequenced and functionally overexpressed in *Escherichia coli (9)*. Using site-directed mutagenesis, Glu 372 was identified as the active site nucleophile, being the equivalent of the nucleophile in mesophilic homologues according to a multiple sequence alignment *(9)*. Moreover, a recent study indicated that the enzymes from *P.furiosus* and the mesophilic bacterium *A.faecalis* share a common catalytic mechanism *(15)*. The 3-dimensional model for CelB shows that the conformation of the active site is identical to that of other family 1 enzymes *(39, 40)*.

An alternative approach to study structure-function relations in enzymes, is that of directed evolution by combining random mutagenesis, *in vitro* recombination and rapid screening procedures *(16,17,18)*. Subsequent biochemical and structural analysis of selected mutant enzymes has been

shown to contribute to understanding the molecular basis of the observed phenotypic changes and as such to our knowledge of enzyme catalysis and stability. Directed evolution approaches have hitherto been restricted to enzymes with moderate thermostability and have not yet been applied to enzymes from hyperthermophiles. CelB is extremely suitable to be used as hyperthermostable model enzyme in the development of such a procedure, since it is efficiently produced in *E.coli*, is not post-translationally modified, and is capable of hydrolyzing chromogenic substrates, both at low and high temperatures. In this study we describe the construction of a random CelB library and the screening of this library for increased activity on pNp-Glc at room temperature. Multiple random mutants were selected and characterized, with the aim to gain insight in substrate recognition and catalysis both at low and at high temperatures, and into the relation of these features with thermoactivity and thermostability.

5.2 Materials and methods

Construction of a random mutant CelB library. Random mutations were introduced into the *celB* gene by PCR amplification using primers BG238 (5'-GCGCG<u>CCATG G</u>CAAAGTTCCCAAAAAACTTCATGTTTG; *Nco*I restriction site underlined) and BG309 (5'-GTTAGCAGCC<u>GGATCC</u>CTA; *Bam*HI restriction site underlined) and proof-reading deficient *Taq* DNA polymerase (Pharmacia, Uppsala, Sweden). DNA shuffling of the *celB* gene was essentially performed according to the protocol of Stemmer (*16*) with the optimization described by Lorimer & Pastan (*19*). The resulting shuffled DNA fragments were digested with *Nco*I and *Bam*HI, the restriction sites of which overlapped the *celB* start- and stop-codon, respectively. The fragments were subsequently cloned into expression vector pET9d as translational fusion with phage φ10 translation initiation and termination signals (*20*). The resulting plasmids were used to transform *E.coli* JM109(DE3) (*21,22*). Transformation mixtures were plated onto selective TY agar plates (1% trypton, 0.5% yeast extract, 0.5% NaCl; 1.5% granulated agar and 30 μg/ml kanamycin) supplemented with 16 μg/ml of the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-Glc, Biosynth, Switzerland). Transformants harboring only pET9d or non-functional *celB* variants will remain white after overnight growth at 37°C, while transformants harboring functional *celB* genes will develop a blue color due to hydrolysis of the X-Glc. Single blue colonies were transferred to microtiter plates containing 200 μl/well of TY broth supplemented with 30 μg/ml kanamycin.

Screening for increased activity on pNp-Glc at room temperature. A replica of the random CelB library was prepared in microtiter plates and bacterial growth was quantified by determination of the optical density at 560 nm using a Thermomax microplate reader (Molecular Devices). Data was exported into a spreadsheet (Excel, Microsoft, USA) and corrected for blank absorption. *E.coli* cells were lysed by a combination of freezing/thawing and addition of 10 μl chloroform and 5 μl of

0.1% SDS. No negative effect of these additives on CelB activity was detected in control experiments. Viscosity was reduced by adding 10 µg/ml DNaseI and 10 mM MgCl$_2$. CelB activity at 20°C was assayed by transfer and mixing of 5 µl of lysed cells to a second set of microtiter plates containing 100 µl/well of 150 mM sodium citrate (pH 4.8) and 1 mM pNp-Glc. After 30 minutes reactions were stopped by addition of 200 µl 0.5 M Na$_2$CO$_3$. Free pNp formation was quantified at 405 nm using the Thermomax microplate reader. Potential high-performance mutants were selected on the basis of an increased ratio of OD$^{405\,nm}$/OD$^{560\,nm}$ relative to wild-type controls. These mutants were regrown and rescreened according to the same procedure. Plasmid DNA of confirmed high-performance mutants was subjected to a second round of mutation and selection.

Characterization of high-performance mutants. High performance mutants were grown overnight in 10 ml selective TY medium at 200 rpm. Cells were spun down, resuspended in 1 ml of a 20 mM sodium citrate buffer (pH 4.8) and lysed by sonication (Branson sonifier). Cell debris was spun down and cell free extract was incubated for 10 min at 80°C. Denatured *E.coli* proteins were precipitated by centrifugation. Protein concentration of the heat-stable cell free extract was determined according to Bradford *(23)* using the Bio-Rad protein assay (Biorad, Veenendaal, The Netherlands). SDS-PAGE analysis revealed that CelB was at least 90% pure in heat-stable cell-free extract (not shown). CelB activity at 20°C and 90°C was determined relative to wild-type CelB in 0.5 ml 150 mM sodium citrate (pH 4.8), containing 3 mM pNp-Glc or 40 mM pNp-Gal. The pNp-Glc concentration was chosen at 3 mM because at higher concentrations, the occuring substrate inhibition will cause an underestimation of the enzyme activity. pNp-Gal could be used at 40 mM since there is no indication at all for substrate inhibition for the wild-type CelB and any of the characterized mutant enzymes. Reactions were stopped by addition of 1 ml 0.5 M Na$_2$CO$_3$ and the free pNp concentration was quantified at 405 nm on a Hitachi U-1100 spectrophotometer. Temperature optima were determined in 150 mM sodium citrate buffer (pH 4.8) containing 3 mM pNp-Glc at desired temperatures up to 98°C in a water bath. Thermostability was analysed in an oilbath at 106°C. Heat-stable cell free extract was diluted to 50 µg/ml protein in a 20 mM sodium citrate buffer pH 4.8. 100 µl samples were incubated in glass vials closed by screw caps with teflon inlay and incubated at 106°C. Residual CelB activity after 1 h incubation was determined at 20°C as described above and compared to the activity of a sample kept at room temperature.

DNA sequencing and three-dimensional structure analysis. Plasmid DNA was isolated from 3 ml cultures using the Qiaprep spin plasmid kit (Qiagen, Westburg, NL) according to the included protocol. Sequencing was performed using infrared-labeled oligos (MWG, Germany) complementary to pET9d sequences immediately flanking the cloning site on the plasmid, and the Thermo Sequenase kit (Amersham Life Science) on the Li-Cor 4000L automated sequencer. Amino acid substitutions were deduced from identified mutations in the DNA sequence of the *celB* gene and their position and possible interactions in the three-dimensional structure of CelB was analysed using the 3D-structures

of BglA from *B.polymyxa* (*5*; pdb accession code 1bgg.pdb), LacS from *S.solfataricus* (*12*; pdb accession code 1gow.pdb) and the three-dimensional model for CelB (*39, 40*), with the use of the molecular visualization software Rasmol (Raswin Molecular Graphics V 2.6; Sayle, Glaxo Research and Development, U.K.) on a personal computer, and InsightII (V 97.0, Molecular Simulations Inc.) on a Silicon Graphics Indy workstation.

Purification of CelB and determination of biochemical parameters. One liter of selective TY medium in 2 liter baffled erlenmeyer flasks, was inoculated with *E.coli* harboring expression plasmids containing either wild-type or mutant *celB* genes and cultured overnight at 37°C while vigorously shaking. Cell lysate was prepared and CelB was purified using heat-precipitation and anion exchange chromatography essentially as described by (*9*) with an additional gel filtration in 20 mM Tris-HCl buffer (pH 8.0) on a 300 ml Superdex 75 column (Pharmacia, Sweden). Pure CelB fractions were pooled and dialysed against a 20 mM sodium citrate buffer (pH 4.8). Protein concentrations were determined at 280 nm using an extinction coefficient for one subunit of $\varepsilon_m^{280nm}=1.28*10^5$ $M^{-1}cm^{-1}$ according to Gill & von Hippel (*24*). Initial CelB activity of pure fractions was monitored continuously in a total volume of 1.0 ml 150 mM sodium citrate buffer (pH 4.8, set at 20°C) at 20 and 90°C on a Hitachi U-2010 spectrophotometer equipped with SPR-10 temperature controller (Hitachi, Tokyo, Japan).

The amount of wild-type and mutant CelB in the activity assays on pNp-Glc/Gal and cellobiose/lactose was in the range of 5-20 µg at 20°C, and 90-240 ng at 90°C. Formation of hydrolyzed pNp was measured at 405 nm. Catalytic activities were calculated using extinction coefficients for pNp in the assay buffer of $\varepsilon_m^{405\ nm} = 0.178$ mM^{-1} cm^{-1} at 20°C and $\varepsilon_m^{405\ nm} = 0.561$ mM^{-1} cm^{-1} at 90°C. Turnover rates (k_{cat} [s^{-1}]) were calculated using the subunit molecular weight of CelB $M_r = 54,580$ Da. A turnover rate of 0.967 s^{-1} corresponds to a catalytic activity of 1 U/mg, in which 1 U is defined as the amount of β-glucosidase activity needed to catalyze the liberation of 1 µmol of *p*-nitrophenol per minute at the given temperature. Kinetic parameters were determined by measuring initial velocity at varying substrate concentrations, ranging from 0 to 8 mM pNp-Glc (up to 40 mM for N415S at 90°C) and 0.2-40 mM pNp-Gal. Hydrolysis of disaccharides was performed in 150 mM citrate buffer (pH 4.8) with 0-150 mM cellobiose or 0-300 mM lactose at 20 and 90°C. Glucose formation was quantified using the Peridochrom Glucose kit (Boehringer Mannheim, Germany) according to the included protocol. The signals of the reaction substrates cellobiose and lactose and the product galactose, were found to be negligible. For each K_m and k_{cat} determination, at least 14 substrate/velocity data pairs were determined. Kinetic data were fitted according to functions describing Michaelis-Menten kinetics or, when necessary, according to a second-degree function describing substrate inhibition using the non-linear regression program Tablecurve (Tablecurve 2D for Windows, version 2.03, Jandel Scientific). In the second-degree inhibitor function, a $[S]^2$ term has been added to the denominator of the Michaelis-Menten equation.

Table 5.1 Characterization of high performance CelB mutants.[a]

Plasmid pLUW	Amino acid substitution(s)	Class	% Act. on pNp-Glc at 20°C	% Act.on pNp-Gal at 20°C	% Act. on pNp-Glc at 90°C	% Act. on pNp-Gal at 90°C	Temperature optimum (°C)	Thermostability (% residual act. after 1 h at 106°C)
pLUW511	wild-type		100	100	100	100	≥ 98	100
pLUW839	N415S	I	259	64	35	35	≥ 98	100
pLUW842	M424V	I	199	140	53	49	85	24
pLUW843	T371A	I	148	139	101	91	85-90	14
pLUW846	A419T	I	159	109	64	98	85	53
pLUW844	K285R	III	155	103	55	81	85-90	58
pLUW841	V211A	I	260	151	141	120	85-90	<1
	V163I	II						
	F447S	II						
pLUW847*	N415S	I	308	27	24	63	85	<1
	V211A	I						
	Y227H	II						
	E26G	III						
pLUW845	I161V	II	110	37	86	125	98	84
	E119G	III						
pLUW840	K70R	II	175	147	42	<1	70-80	<1
	L45P	II						
	F344L	III						
pLUW838	I67T	III	110	97	70	70	≥ 98	100
	D159N	III						
	A341T	III						

[a] For each mutant the amino acid substitutions are listed, as well as their classification. Class I involves amino acid substitutions in the active site region, Class II at or very near to subunit interface, and Class III at surface of the protein or buried in interior and not in one of the other categories. * pLUW847 was selected from a second generation library. Specific activities on 3 mM pNp-Glc and 40 mM pNp-Gal at 20 and 90°C were determined and are listed in percentages relative to wild-type CelB; absolute values of specific activities of wild-type CelB are:13.4 U/mg with pNp-Glc at 20°C; 9.4 U/mg with pNp-Gal at 20°C; 1655 U/mg with pNp-Glc at 90°C; and 2482 U/mg with pNp-Gal at 90°C (see also Tables 2 and 3). Temperature optima were determined up to 98°C and thermostability relative to the wild-type was determined at 106°C.

The formation of reaction products after 30 minutes incubation at 20 and 90°C of 20 µg wild-type CelB with 20 and 150 mM cellobiose was analysed by HPLC using a Polyspher OAHY column (Merck, Darmstadt, Germany). 20 mM solutions of glucose, cellobiose and a mixture of cellulodextrines were used as standards. Thermal inactivation of pure enzyme samples was studied by incubation of 50 µg/ml samples of wild-type and mutant CelB at 106°C in 20 mM sodium citrate buffer (pH 4.8). Remaining activity as a function of time was assayed relative to a standard kept at room temperature. First-order plots of inactivation yielded half-life values for thermal inactivation.

5.3 Results and discussion

Construction of a random mutant CelB library. The hyperthermostable β-glucosidase CelB from *P.furiosus* is optimally active at 102-105°C, and its activity at room temperature does not exceed 1% of its optimal activity. In order to study the possibility of increasing this low activity at room temperature and in order to gain more insight into the factors determining this temperature dependance of activity, we subjected the *celB* gene to random mutagenesis and low-temperature activity screening. Random mutations were introduced into the *celB* gene using error-prone PCR and *in vitro* recombination by DNA shuffling according to Stemmer (*16*). Upon cloning the mutated *celB* genes in *E.coli*, transformants harboring functional CelB enzymes were selected by their ability to hydrolyse the chromogenic substrate analogue X-Glc at low temperature. An ordered library consisting of 6160 random *E.coli* mutants was constructed. Complete DNA sequence analysis of 9 randomly picked mutants revealed an average mutation frequency of 2.3 basepair per *celB* gene, indicating that on average 1 or 2 amino acids will be changed in each CelB enzyme (data not shown).

Screening for increased activity on pNp-Glc at room temperature. The mutant CelB library was screened for increased activity on pNp-Glc at room temperature. This resulted in the identification of about 400 mutants with significantly higher activity than the wild-type controls. These mutants were transferred to new microtiter plates and rescreened. Heat-stable cell-free extracts of the 42 most active mutants were further analysed for low-temperature β-glucosidase activity. Eventually, the nine most active mutants (harboring plasmids pLUW838-pLUW846) were selected, analysed in detail and found to contain heat-stable β-glucosidase activities that were up to three-fold higher at room temperature than that of *E.coli* producing wild-type CelB (Table 5.1). Variation in enzyme production levels was ruled out as an explanation for this increased activity since protein analysis using SDS-PAGE, showed that CelB expression levels were similar in all mutants (data not shown). Plasmid inserts were sequenced and amino acid substitutions were deduced from the changes in the DNA sequence (Table 5.1).

Amino acid substitutions in high performance mutants. The 3-dimensional model of CelB (*39, 40*), and the crystal structures of LacS (*12*) and BglA from *Bacillus polymyxa* complexed with gluconate in the active site (*5*), were used to analyze the position and possible interactions of each of the amino acid substitutions found in the high-performance mutants. LacS is a close relative of CelB (53% amino acid identity), with an identical arrangement of subunits into a tetrameric conformation. BglA is distantly related, with only 17% amino acid identity but 100% conservation of the active site residues (all residues depicted in Figure 5.3 are identical in both CelB and LacS as well as in BglA). Because of the low resolution of the present CelB model, we have been careful in evaluating interactions and mainly restricted this to the active site region were high structural homology between CelB, LacS and BglA is found. In other regions we restricted the analysis to an evaluation of the position of the mutations and expected general effects.

```
                      *             G        . . . . . . P                      T  R
PF_B-GLU  : MKFPKNFMFGYSWSGFQFEMG--LPGSE-VESDWWVWVHDKENIASGLVSGDLPENGPAYWHLYKQDHDIAEKL  :  71
SS_B-GLY  : MYSFPNSFRFGWSQAGFQSEMG--TPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPGYWGNYKTFHDNAQKM :  73
PF-B-MAN  : MFPEKFLWGVAQSGFQFEMGDKLRRNIDTNTDWWHWVRDKTNIBKGLVSGDLPEEGINNYELYEKDHEIARKL   :  73

TM_B-GLU  : MNVKKFPEGFLWGVATASYQIEGS--PLADGAGMSIWHTFSHTPGNVKNGDTG----DVACDHYNRWKEDIEIIEKL  :  71
BP_B-GLU  : MTIFQFPQDFMWGTATAAYQIEGA--YQEDGRGLSIWDTFAHTPGKVFNGDNG----NVACDSYHRYEEDIRLMKEL  :  71
AF_B-GLU  : MTDPNTLAARFPGDFLFGVATASPQIEGS--TKADGRKPSIWDAFCNMPGHVFGRHNG----DIACDHYNRWEEDLDLIKEM :  76
LL_6P-GAL : MTKTLPKDFIFGGATAAYQAEGA--THTDGKGPVAWDKYLEDN----YWYTA----BPASDFYHKYPVDLELAEEY  :  66
                  20          40          60          80


                               G                                               *
PF_B-GLU  : GMDCIRGGIEWARIFPKPTFDVKVDVEKDEEG-NIISVDVPESTIKELEKIANMEALEHYRKIYSDWKERGKTFILNLYHWP : 152
SS_B-GLY  : GLKIARLNVEWSRIFFNP-LPRPQNFDESKQ--DVTKVEINENELKRLDEYANKDALNHYREIFPKDLKSRGLYFILNMYHWP : 152
PF-B-MAN  : GLNAYRIGIEWSRIFWPTTFIDVDYSYNESYNLIEDVKITKDTLEELDEIANKREVAYYRSVINSLRSKGFKVIVNLNHFT  : 155

TM_B-GLU  : GVKAYRFSISWPRILPEG------------------------------TGRVNQKGLDFYNRIIDTLLEKGITPFVTIYHWD : 123
BP_B-GLU  : GIRTYRFSVSWPRIFPNG------------------------------DGEVNQEGLDYYHRVVDLLNDNGIEPFCTLYHWD : 123
AF_B-GLU  : GVEAYRFSLAWPRIIPDG------------------------------FGPINEKGLDFYDRLVDGCKARGIKTYATLYHWD : 128
LL_6P-GAL : GVNGIRISIAWSRIFPTG------------------------------YGEVNEKGVEFYHKLFAECHKRHVEPFVTLHHFD : 118
                 100         120         140         160


                N V I.....                             *        A      ...       H...    .
PF_B-GLU  : LPLWIHDPIAVRKLGPDRAPAGWLDEKTVVEFVKPAAFVAYHLDDLVDMWSTMNEPNVVYNQGYINLRSGFPPGYLSFEAAE : 234
SS_B-GLY  : LPLWLHDPIRVRR-GDFTGPSGWLSTRTVYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGYLSFELSR : 233
PF-B-MAN  : LPYWLHDPIEARERALTNKRNGWVNPRTVIEFAKYAAYIAYKFGDIVDMWSTFNEPMVVVELGYLAPYSGFPPGVLNPEAAK : 237

TM_B-GLU  : LPFAIQL-----------KGGWANREIADWFAEYSRVLFENFGDRVKNWITLNEPWVVAIVGHLYG-VHAPGMR-DIYVAF  : 191
BP_B-GLU  : LPQALQD-----------AGGWGNRRTIQAFVQPAETMFREFHGKIQHWLTFNEPWCIAPLSNMLG-VHAPGLT-NLQTAI  : 191
AF_B-GLU  : LPLTLMG-----------DGGWASRSTAHAFQRYAKTVMARLGDRLDAVATFNEPWCAVWLSHLYG-VHAPGER-NMEAAL  : 196
LL_6P-GAL : TPEALHS-----------NGDFLNRENIEHFIDYAAFCFEEFP-EVNYWTTFNEIGPIGDGQYLVG-KFPPGIKYDLAKVF  : 186
                 180         200           A          240


                                                         R
PF_B-GLU  : KAKFNLIQAHIGAYDAIKEYS-EKS---------VGVIYAFA-WHDPLAEEYKDEVEEIRKKD------YEFVTILH----- : 294
SS_B-GLY  : RHMYNIIQAHARAYDGIKSVS-KKP---------VGIIYANS-SFQPLTDKDMEAVEMAENDNRW----WFFDAIIRGEI-- : 298
PF-B-MAN  : LAILHMINAHALAYRQIKKFDTEKADKDSKEPAEVGIIYNNIGVAYPKDPNDSKDVKAAENDNFFHS-GLFFEAIHKGKLNI : 318

TM_B-GLU  : RAVHNLLRAHARAVKVFRETVKDGK---------IGIVFNNG-YFEPAS-EKEEDIRAVRFMHQFNNYPLFLNPIYRGDYPE : 262
BP_B-GLU  : DVGHHLLVAHGLSVRRFRELGTSGQ---------IGIAPNVS-WAVPYS-TSEEDKAACARTISLHS-DWFLQPIYQGSYPQ : 261
AF_B-GLU  : AAMHHINLAHGFGVEASRHVAPKVP---------VGLVLNAH-SAIPAS-DGEADLKAAERAFQFHN-GAFFDPVFKGEYPA : 266
LL_6P-GAL : QSHHNMMVSHARAVKLYKDKGYKGE---------IGVVHALP-TKYPYDPENPADVRAAELEDIIHN-KFILDATYLGHYSD : 257
                 260         280         300         320


                                          * *                                    ...
PF_B-GLU  : ----------------------------SKGKLDWIGVNYYSRLVYGAKDGHLVPLP--------GYGFMSERGGFAKSGR : 339
SS_B-GLY  : -TRGNE----------------KIVRDDLKGRLDWIGVNYYTRTVVKRTEKGYVSLG-------GYGHGCERNSVSLAGL  : 354
PF-B-MAN  : EFDGET---------------FIDAPYLKGN-DWIGVNYYTREVVTYQEPMFPSIPLITFKGVQGYGYACRPGTLSKDDR  : 382

TM_B-GLU  : LVLEFAREYLPEN--------YKDDMSEIQEKIDFVGLNYYS-GHLVKFDPDAPAKVS-----------FVERDLP--    : 318
BP_B-GLU  : FLVDWFAEQGATVP-------IQDGDMDIIGEPIDMIGINYYS-MSVNRFNPEAGFLQSE-----------EINMGLP--  : 320
AF_B-GLU  : EMMEALGDRMPVV---------EAEDLGIISQKLDWWGLNYYT-PMRVADDATPGVEPPA-----------TMPAPAVSD  : 325
LL_6P-GAL : KTMEGVNHILAENGGELDLRDEDFQALDAAKDLNDFLGINYYMSDWMQAFDGETEIIHNG-----------KGEKGSSKY  : 326
                 340         360         380         400


                T  .L                             A              .......
PF_B-GLU  : PASDPG-------------WEMYPEGLENLLKYLNNAYELP--MIITENGMAD--------AADRYRPHYLVSHLKAVYN  : 396
SS_B-GLY  : PTSDFG-------------WEFFPEGLYDVLTKYWNRYHLY--MYVTENGIAD--------DADYQRPYYLVSHVYQVHR  : 411
PF-B-MAN  : PVSDIG-------------WELYPEGMYDSIVEAH-KYGVP--VYVTENGIAD--------SKDILRPYYIASHIKMIEK  : 438

TM_B-GLU  : -KTAMG-------------WEIVPEGIWILKKVK-EEYNPPEVYITENGAAFDDVVSEDGRVHDQNRIDYLKAHIGQAWK  : 384
BP_B-GLU  : -VTDIG-------------WPVESRGLYEVLHYLQ-KYGN-IDIYITENGACINDEVVN-GKVQDRRISYMQQHLVQVHR  : 384
AF_B-GLU  : VKTDIG-------------WEVYAPALHTLVETLY-ERYDLPECYITENGACYNMGVEN-GQVNDQPRLDYYAEHLGIVAD : 391
LL_6P-GAL : QIKGVGRRVAPDYVPRTDWDWIIYPEGLYDQIMRVKNDYPNYKKIYITENGLGYKDEFVD-NTVYDDGRIDYVKQHLEVLSD : 407
                 420         440            N          480


                *      S **T     V.*              ...          S.     .....  .
PF_B-GLU  : AMKEGADVRGYLHWSLTDNYEWAQGFRMRPGLVYVDFETKKKRYLR-PSALVFREIATQKEIPEELAHLADLKFVT  RK  : 472
SS_B-GLY  : AINSGADVRGYLHWSLADNYEWASGFSMRPGLLKVDYNTKRLYWR-PSALVYREIATNGAITDEIEHLNSVPPVKPLRH   : 489
PF-B-MAN  : APEDGYEVKGYFHWALTDNFEWALGFRMRPGLYEVNLITKERIPREKSVSIFREIVANNGVTKKIEEELLRGM         : 511

TM_B-GLU  : AIQEGVPLKGYFVWSLLDNFEWAEGYSKRFGIVYVDYSTQKRIVK-DSGYWYSNVVK-----NNGLED             : 446
BP_B-GLU  : TIHDGLHVKGYMAWSLLDNFEWAEGYNMRFGMIHVDPRTQVRTPK-ESYYWYRNVVS-----NNWLETRR           : 448
AF_B-GLU  : LIRDGYPMRGYFAWSLMDNFEWAEGYRMRFGLVHVDYQTQVRTVK-NSGKWYSALASGFPKGNHGVAKG            : 459
LL_6P-GAL : AIADGANVKGYFIWSLMDVFSWSNGYEKRYGLFYVDFDTQERYPK-KSAHWYKKLAET-----QVIE              : 468
                 500         520         540         560
```

The position of the identified amino acid substitutions is marked in the CelB sequence in a multiple amino acid sequence alignment, which furthermore highlights the active site residues and residues involved in intersubunit contacts (Figure 5.1). Analysis of the position of the substitutions in the CelB model revealed a more or less uniform distribution over the enzyme (Figure 5.2). Based on the position of the substitutions in the enzyme structure, three classes of CelB mutants were distinguished (Table 5.1, Figure 5.2). The first class (Class I) involves substitutions of residues in the immediate vicinity of the active site, that often directly interact with active site residues. The second class (Class II) involves substitutions that are located at or near subunit interfaces. The remainder of the mutations was assigned to Class III, and involves substitutions at the protein surface or buried in the interior of the monomers. Five of the high-performance mutants contained single amino acid substitutions. While in these mutants the genotype-phenotype relation is straightforward, this is less clear in the remaining mutants containing multiple amino acid substitutions.

The majority of the amino acid substitutions in the mutants containing only a single substitution, belong to Class I (Figure 5.3). These mutations predominantly affect the C-terminal part of the enzyme. Plasmid pLUW839 encodes amino acid substitution N415S, which involves an asparagine residue that is hydrogen bonded to active site residues glutamine 17 and glutamate 417. These two residues directly interact with the hydroxyl group at the C4-position of the substrate. The serine side chain in the mutant is much shorter than the original asparagine. Therefore, interactions will be significantly weakened or destroyed and a cavity may be introduced. In homologous family 1 glycosyl hydrolases this residue is conserved except for *L.lactis* LacG (and other 6-phospho-β-galactosidases), which contains a valine (or a leucine or glutamine) at this position (Figure 5.1) (*25*). Mutant N415S displays the highest increase in β-glucosidase activity and is as thermostable as the wild-type enzyme. DNA of pLUW839, coding for this mutant, was therefore used to start a second generation evolution, resulting in a derivative (harboring pLUW847) with considerably higher activity on pNp-Glc than the N415S mutant CelB (Table 5.1).

Figure 5.1 (Opposite page) Alignment of family 1 glycosyl hydrolases. Sequences are obtained from GenBank and Swiss Prot and include : PF_B-MAN = β-glucosidase (CelB) from *P.furiosus* (AF013169), SS_B-GAL = β-galactosidase (LacS) from *Sulfolobus solfataricus* (P22498), PF_B-MAN = β-mannosidase from *P.furiosus* (U60214), TM_B-GLU= β-glucosidase from *T.maritima* (Q08638), BP_B-GLU = β-glucosidase (BglA) from *Bacillus polymyxa* (P22505), AF_B-GLU = β-glucosidase (Abg) from *Agrobacterium faecalis* (P12614), LL_P-GAL = 6-phospho-β-galactosidase (LacG) from *L.lactis* (P11546). The active site nucleophile and acid-base catalyst are marked with N and A below the sequences, respectively, other active site residues are marked with an asterisk. Residues involved in intersubunit interactions, are indicated with dots. Identified amino-acid substitutions in high-performance mutants are indicated in bold and underscored.
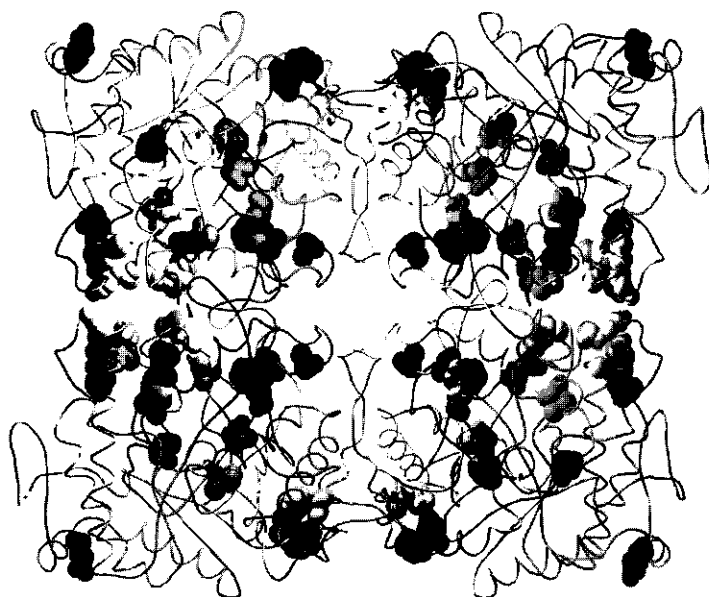
Figure 5.2 Model of the CelB tetramer based on 3.3 Å X-ray diffraction data (*39; 40*). Subunits are depicted in light-gray ribbon presentation. Side chains of the amino acids that have been substituted in the high-perfomance CelB mutants, are shown in space-filling representation with Class I substitutions in dark gray, Class II substitutions in light gray and Class III substitutions in black. The figure has been generated with Swiss PDB viewer (Glaxo Wellcome Experimental research) in combination with the ray-tracing program POV-RAY.

Plasmid pLUW842 encodes the single amino acid substitution M424V, which involves a methionine that, like the asparagine at position 415, contacts active site glutamate 417. The valine in the resulting mutant CelB introduces a shorter but more bulky side chain which may influence the conformation of this active site glutamate. Residues at position 424 in homologous enzymes are either methionine or lysine, which both have long, unbranched side chains (Figure 5.1). In LacG this lysine, together with a serine instead of E417, and a third conservative mutation (a tyrosine at position 426 instead of phenylalanine), serves to stabilize binding of the negatively charged phosphate group on the substrate (*4, 40*).

Plasmid pLUW843 encodes T371A CelB, with a substitution that results in the removal of the polar side chain of a threonine residue that is located adjacent to the catalytic nucleophile E372. Although this threonine does not seem to interact directly with the glutamate side chain, the fact that this residue and several adjacent residues are completely conserved in all family 1 glycosyl hydrolases, strongly suggests that interactions in this region are critical for correct catalysis (Figure 5.1).

Plasmid pLUW846 encodes A419T CelB which involves a buried alanine at the top of the active site barrel next to W418 that interacts with the hydroxyl group at the C3 position of the substrate. Because the alanine is buried, the larger threonine side chain will probably change the side chain packing in this area. Remarkably, the same substitution into threonine or into the similar polar residue serine has also occured in 6-phospho-β-galactosidases (Figure 5.1) (*25*).

Another plasmid that contains only a single base change in the *celB* gene, is pLUW844 coding for the amino acid substitution K285R. This mutation is located at the kink in the α-helix that creates a tunnel in the side of the $(\alpha\beta)_8$-barrel, which probably forms the substrate entrance to the active site (*12*). K285R is, however, a very conservative mutation and the residue is not conserved at all in homologous enzymes. The model of CelB predicts an ion-pair interaction between K285 and E281, which may still be formed in the mutant K285R CelB.

The fifth substitution in the active site region, although not as close to the active site itself as the other substitutions, is V211A that involves a completely buried valine in the same α-helix as active site residues N206 and the acid/base catalyst E207. Residues at this position are always hydrophobic and the mutation will introduce a cavity. This substitution is found in combination with two class II substitutions, V163I and F447S, at the subunit interface in the CelB mutant encoded by
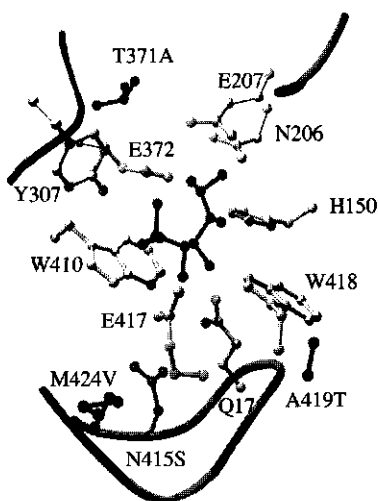
Figure 5.3 Close-up of the active site of *B.polymyxa* BglA with inhibitor gluconate (*5*). Gluconate is depicted in ball-and-stick representation. Active site residues are shown in stick representation. Corresponding residues that are changed in the high performance CelB mutants (T371A, M424V, N415S and A419T) are shown in ball-and-stick representation. Numbering of the residues is according to the homologous CelB residues in order to allow comparison with the text. The figure has been generated using Molscript (*38*).

pLUW841. The V163I substitution is located in a loop that is only present in the tetrameric family 1 enzymes. The phenylalanine at position 447 is completely buried close to the subunit interface and its substitution to serine will therefore result in a cavity. The aromatic ring at this position in the enzyme seems to be important, since in all family 1 enzymes either a phenylalanine or a tyrosine residue is present.

Two of the above mentioned active site substitutions, N415S and V211A, are also present in the enzyme encoded by plasmid pLUW847 that was selected from a second-generation library. The N415S substitution is derived from the template plasmid in the DNA shuffling procedure, the V211A substitution is introduced independently in this second round of evolution. Two additional mutations are present in pLUW847, resulting in E26G in which a surface-exposed charge is removed, and Y227H, in which intersubunit interactions are changed. This second generation CelB variant is significantly more active at room temperature than N415S CelB and exceeds the wild-type CelB activity more than three-fold (Table 5.1).

Two mutants contain multiple substitutions belonging to Class II and III. pLUW845 codes for substitution I161V which is also located in the subunit interface loop that appears to be restricted to tetrameric members of family 1 enzymes. Furthermore, the surface-exposed E119 has been exchanged for a glycine. This may result in the loss of electrostatic interactions and may have significant effects on the local flexibility of the polypeptide because of the large conformational freedom that is introduced with the glycine residue. pLUW840 codes for three substitutions: K70R, L45P and F344L. This triple mutant is highly destabilized, possibly mainly by the L45P substitution in which the conserved leucine is changed to a proline, that may introduce severe strain into the polypeptide chain.

A   B   C

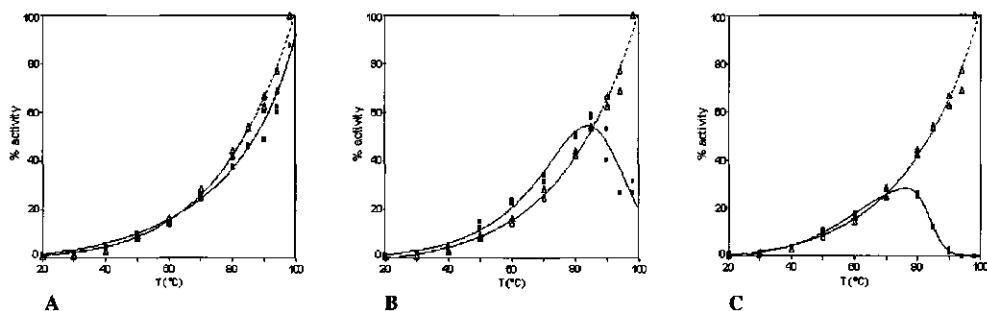**Figure 5.4** Optimum temperature curve for wild-type CelB (Δ) compared to mutant CelB N415S (■) (A), mutant CelB M424V (■) (B), and mutant CelB K70R/L45P/F344L (■) (C). The inset in a) zooms in on the increased activity of N415S CelB below 50°C.

Plasmid pLUW838 codes for three class III substitutions, viz. I67T, D159N and A341T. This CelB mutant displays an unchanged temperature optimum curve, is as thermostable as the wild-type enzyme, and shows relatively minor effects on activity (Table 5.1). These properties indicate that even a highly optimized, hyperthermostable enzyme is able to incorporate changes without loosing stability and thermoactivity. None of these residues is completely conserved among close relatives, and moreover, LacS also contains a threonine at position 341 (Figure 5.1).

Characterization of high-performance CelB variants. The increase in the rate of β-glucosidase activity on pNp-Glc that is found in the heat stable extract of the first generation mutants differs from 1.1- to 2.6- fold (Table 5.1). A further increase to more than 3-fold higher activity was obtained in the second generation mutant. However, β-galactosidase activity is increased in only four of the mutant CelB enzymes, is unchanged in three other mutants, and has decreased in the remaining three mutant CelB enzymes (Table 5.1). The largest increase in the ratio of pNp-Glc/pNp-Gal hydrolysis is observed in the mutants containing the active site-associated mutation N415S (pLUW839 and pLUW847) as well as in the enzyme encoded by pLUW845. This latter mutant CelB is remarkable because its amino acid substitutions are at the enzyme surface (E119G) and on the subunit interface (I161V), indicating that substrate specificity is not solely determined by interactions in the active site region. Large reductions in both β-glucosidase and β-galactosidase activity at 90°C are found in the mutant CelB mutants, except for the enzyme encoded by plasmid pLUW841 that has increased activity on both substrates, the active site mutant T371A that has unchanged activity, and, again remarkably, the CelB mutant encoded by pLUW845 that shows increased β-galactosidase activity, although this was severely reduced at room temperature.

To study whether the low temperature adaptation of the mutant enzymes affected the activities at other temperatures, their temperature optima were determined. Changes that were detected in the optimum curves, could be divided in three categories (Fig 4); (i) flattening of the curve, as is found for mutant N415S (Fig 4a) that shows below 50°C higher activity than the wild-type CelB, while the reverse was observed above 60°C, (ii) shifting of the curve to lower temperature values as observed for mutant M424V (Fig 4b), and (iii) a much earlier inactivation as found for the mutant encoded by pLUW840 (Fig 4c). All mutants with lower temperature optima for activity also have a lower resistance towards thermal inactivation (Table 5.1).

Purification and thermal inactivation of wild-type and active site mutant CelB. In order to gain more insight in the effects realized by the amino acid changes in the vicinity of the active site, we decided to purify and characterize mutant CelB enzymes containing a single substitution in this region, namely N415S, M424V, T371A and A419T. In order to obtain an accurate measure of the persistence of catalytic function at high temperature, thermal inactivation of purified wild-type and
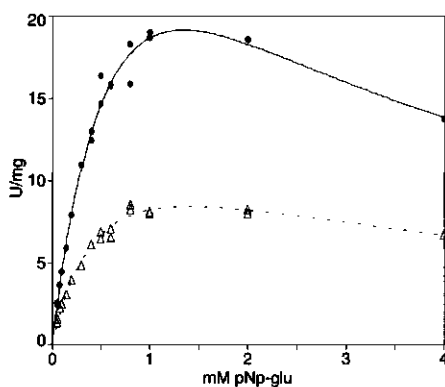
Figure 5.5 Comparison of rate of pNp-Glc hydrolysis by wild-type (Δ) and mutant CelB N415S (●) as a function of substrate concentration at 20°C.

CelB mutants was followed in time (Table 5.2). Mutant N415S CelB is as thermostable as the wild-type enzyme, while the other mutants are considerably destabilized; compared to all other family 1 glycosyl hydrolases they are, however, still extremely thermostable enzymes because they all retain a considerable amount of activity after incubation at 106°C.

Kinetic characterization of wild type CelB at 20 and 90°C. The hydrolysis of pNp-Glc by wild-type CelB revealed an apparent inhibition in a hyperbolic Michaelis-Menten plot, starting at substrate concentrations above 1 mM at 20°C (Figure 5.5) and above 2-4 mM pNp-Glc at 90°C (not shown). A similar phenomenon was observed during cellobiose hydrolysis (Figure 5.6). Plotting the data according to a Lineweaver-Burk equation, resulted in a curve typically found when substrate inhibition occurs (inset in Figure 5.6). This process has not been reported in the previous kinetic characterization of CelB, presumably because employed maximum substrate concentrations did not exceed 1.5 mM pNp-Glc at 90°C (*15*). Apparent substrate inhibition was, however, also observed with the β-glucosidase from *Thermotoga maritima* (*6*). Fitting of the data according to Michaelis-Menten kinetics, corrected for substrate inhibition, resulted in a k_{cat} of 13 s^{-1} and a K_m of 0.53 mM for pNp-Glc at 20 °C (Table 5.2). The enzyme is two-fold less efficient on pNp-Gal than on pNp-Glc, due to a lower hydrolysis rate and a lower affinity for the galactosyl substrate. On cellobiose and lactose, similar K_{cat} values are found, but affinity for the β-(1,4)-linked disaccharides is relatively low. K_{cat}-values for the aryl-glycosides at 90°C are 1.5- and 4-fold higher than reported earlier for the enzyme purified from *E.coli* by Voorhorst and coworkers and Bauer and Kelly, respectively (*9,15*). This difference may be related to the observed substrate inhibition or may reflect the variability between independent enzyme preparations which have been purified by different methods.

Calculated K_m values at 90°C are more or less similar to those already reported. Interestingly, while K_m values of wild-type CelB for pNp-Glc and cellobiose at 90°C *decrease* compared to those at 20°C, these values *increase* for pNp-Gal and lactose. Substrate affinity is, apparently, highly influenced by the operation temperature. It should be noted that, although k_{cat} values for pNp-Gal and lactose are considerably higher than for pNp-Glc and cellobiose, the enzyme shows, based on catalytic efficiency, a higher β-glucosidase than β-galactosidase activity.

Transglycosylation. Transglycosylation is common among family 1 glycosyl hydrolases. It has been described before for *P.furiosus* CelB and *A.faecalis* Abg in the case of pNp-xylose and pNp-arabinose substrates and for CelB as well using cellobiose and lactose *(15,26,27,28).* Cellotriose formation from cellobiose has been reported for BglA from *B.polymyxa* and for the β-glucosidase from *T.maritima (29,6).* Analysis of reaction products of CelB incubated with 150 mM cellobiose by HPLC revealed, apart from glucose and cellobiose, the presence of an additional product with the same retention time as cellotriose, strongly suggesting that transglycosylation is indeed occuring. Kinetic characterization of this process was not possible because its effect on product formation was completely masked by the large observed substrate inhibition (see above).

Kinetic characterization of active site mutant CelB enzymes at 20°C. The catalytic efficiencies of mutated CelB on pNp-Glc and cellobiose are increased up to two-fold in comparison with wild-type CelB (Table 5.2). This has either been achieved by large increases in K_{cat} values, as in the case of mutant N415S CelB on pNp-Glc (Figure 5.5a), or by a decrease in K_m values as observed for T371A CelB on cellobiose (Figure 5.5b). The increase in efficiency of pNp-Glc hydrolysis is at the expense of catalytic efficiency on pNp-Gal. However, a similar effect is not observed on β-(1,4)-
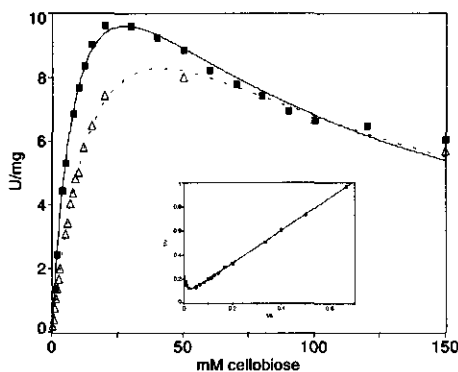

Figure 5.6 Comparison of the rate of cellobiose hydrolysis by wild-type (Δ) and mutant T371A CelB (■) as a function of substrate concentration at 20°C. Inset is Lineweaver-Burk plot for wild type CelB.

Table 5.2 Kinetic parameters of wild-type and mutant CelB enzymes at 20°C on pNp-Glc, pNp-Gal and on β-(1,4)-linked disaccharides cellobiose and lactose. In the last column the half-life of thermal inactivation at 106 °C is listed

CelB variant	pNp-Glc			cellobiose			pNp-Gal			lactose			Thermo-stability
	k_{cat} (s⁻¹)	K_m (mM)	Cat.eff.[1] (mM⁻¹ s⁻¹)	k_{cat} (s⁻¹)	K_m (mM)	Cat.eff.[1] (mM⁻¹ s⁻¹)	k_{cat} (s⁻¹)	K_m (mM)	Cat.eff.[1] (mM⁻¹ s⁻¹)	k_{cat} (s⁻¹)	K_m (mM)	Cat.eff.[1] (mM⁻¹ s⁻¹)	$t_{1/2}$ at 106°C (h)
wt	13	0.53	25	16	23	0.69	9.1	0.70	13	13	52	0.25	3.3
N415S	47	1.1	43	20	20	1.0	4.2	1.2	3.5	21	95	0.22	3.2
M424V	40	0.92	43	18	16	1.1	12	3.4	3.4	21	97	0.22	1.1
T371A	26	0.57	45	15	11	1.4	11	1.3	8.2	17	52	0.33	0.65
A419T	40	1.3	31	19	25	0.75	14	4.5	3.1	15	83	0.18	0.82

[1]Cat. eff. = k_{cat}/K_m

linked disaccharides, moreover, a significant increase in efficiency is reported for T371A CelB on lactose. Evidently, the effect of the leaving group (either pNp or glucose for the aryl-glycosides and the disaccharides, respectively) on catalytic efficiency is not comparable for glucosyl and galactosyl substrates. This is most clearly illustrated for mutant N415S CelB, for which a large reduction in maximum activity on pNp-Gal is found, that is now more than ten-fold lower than maximum activity on pNp-Glc. In contrast, cellobiose is hydrolysed with the same maximum velocity as lactose. The mutant enzymes show, in general, a higher substrate inhibition than the wild-type CelB, because the substrate concentration at which the activity is reduced to ½ V_{max}, is lower for all the mutants than for the wild-type CelB (results not shown).

Kinetic characterization of active site mutant CelB N415S at 90°C. Because N415S CelB is as thermoactive and thermostable as the wild-type CelB, we decided to analyse its kinetic parameters at 90°C in order to see the relation between these characteristics at low and high temperatures (Table 5.3). The maximum activity of mutant N415S CelB at 90°C was lower than for the wild-type enzyme for pNp-Glc and especially pNp-Gal. Because also K_m values increase, N415S CelB is less efficient than wild-type CelB on pNp-Glc and pNp-Gal. The same is found for the disaccharides. Higher performance on pNp-Glc at room temperature is therefore not only at the expense of galactosidase activity at room temperature, but also at the expense of high-temperature catalysis. Because of the apparent unchanged temperature optimum and thermostability of this mutant, this effect is not caused by an overall destabilisation of the enzyme, but must be a direct result of the effects of the single mutation on substrate binding and turnover rates.

Table 5.3 Kinetic parameters of wild-type CelB and mutant N415S at 90°C on pNp-Glc, pNp-Gal and on β-(1,4)-linked disaccharides cellobiose and lactose.

CelB variant	pNp-Glc			cellobiose			pNp-Gal			lactose		
	k_{cat} (s^{-1})	K_m (mM)	Cat.eff.[1] (mM^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	Cat.eff.[1] (mM^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	Cat.eff.[1] (mM^{-1} s^{-1})	k_{cat} (s^{-1})	K_m (mM)	Cat.eff.[1] (mM^{-1} s^{-1})
wt	1600	0.42	3900	670	14	48	2400	5.0	480	1300	120	11
N415S	1400	1.4	1000	240	38	6.4	1000	8.5	120	530	220	2.4

[1]Cat. eff. $= k_{cat}/K_m$

5.4 Conclusions

Screening of a random CelB library resulted in the selection of CelB variants with significantly higher activity on pNp-Glc at room temperature than the wild-type enzyme. Amino acid substitutions were located in the active site region (Class I), at or close to subunit interfaces (Class II), or either at the protein surface or buried in the subunit interior (Class III). All mutants containing subunit interface substitutions were less thermostable and had lower temperature optima than the wild-type CelB, suggesting that subunit interfaces play an important role in thermoadaptation. Hyperthermostable enzymes are most probably highly optimized with respect to packing efficiency(*30,31*). However, a recent database survey showed that proteins from mesophiles and thermophiles essentially do not differ in packing (*32*). The present study indicates that CelB is able to accommodate amino acids in its interior with larger or smaller side chains and different properties, without affecting its thermostability or temperature optimum for catalysis. Interestingly, none of the mutants displayed intact thermostability and increased activity at both low and high temperatures. This observation might indicate that CelB is highly optimized regarding both stability and flexibility and that severe restrictions exist on the simultaneous optimization of these intimately related parameters.

Remarkably, substrate specificity of CelB seems to be a characteristic which is determined by the enzyme as a whole, since changes in the ratio on pNp-Glc and pNp-Gal are not restricted to substitutions in the active site region. However, the most drastic effect on the pNp-Glc/pNp-Gal ratio (a 7.5-fold increase) is caused by the disruption of interactions between asparagine at position 415 and two active site residues that directly contact the hydroxyl group on the substrate at the C4 position. This is exactly the (only) position at which glucose and galactose differ in orientation of this hydroxyl group. The loss of hydrogen bonds between the substituted asparagine and active site residues Q17 and E417 may influence the position and the flexibility of these side chains, thereby changing their interactions with the substrate and the ratio between binding and turnover on substrates with either glucose- or galactose moieties at this position. This is indeed found for mutant N415S on aryl-glycosides and, to a lesser extent, on β-(1,4)-linked disaccharides. In fact, a key role in determining glucose/galactose specificity for residues interacting with the active site Q17 and E417 was proposed recently on the basis of the interactions between the inhibitor gluconate and active site residues of *B.polymyxa* BglA (*5*). Further indications that slight changes in conformational

freedom of active site residues may have large effects on catalysis and substrate recognition, is provided by the large and opposing effect of operational temperature on affinities for different substrates. While wild-type CelB has a 8-fold higher efficiency on pNp-Glc than on pNp-Gal at 90°C, this ratio is reduced to 2-fold at 20°C. Until recently it was assumed, based on lactate dehydrogenase from *T.maritima* and *T.neapolitana* xylose isomerase, that an increased temperature causes the decrease of one enzyme's affinity for its substrate (*33,34,35*). However, the results presented here indicate that this assumption does not hold. Also for glutamate dehydrogenase from *P.furiosus* and *T.maritima* we have reported different changes in substrate affinity with increasing temperature, depending on the substrate (*36,37*).

Our results show that the quality of the random CelB library is good and that the sensitivity of the screening procedure is sufficient to isolate mutants with only 10% increase in activity. Mutants displaying evolved properties, like N415S CelB with two-fold increased catalytic efficiency on pNp-Glc but unchanged stability, and mutant T371A CelB with two-fold higher affinity for cellobiose, could be interesting candidates for industrial or diagnostic applications.

In conclusion, this random screening approach using a hyperthermostable enzyme resulted in the identification of residues that are critical in determining thermostability, low-temperature activity, and substrate recognition. It was shown that low-temperature activity can be engineered into a hyperthermostable enzyme without affecting its extreme stability. Not only is directed evolution a powerful approach to introduce a desired property into an industrially relevant biocatalyst, but, as we have shown here, it may facilitate our understanding of the mechanisms determining enzyme catalysis, stability, and substrate recognition at physiological or extreme conditions.

5.5 Acknowledgements

References

1.   Davies, G. & Henrissat, B. (1995) *Structure 3*, 853-9.

2.   Henrissat, B. & Davies, G. (1997) *Curr. Opinion Struct. Biol. 7*, 637-644.

3.   Wiesmann, C., Beste, G., Hengstenberg, W. and Schulz, G. E. (1995) *Structure 3*, 961-8.

4.   Wiesmann, C., Hengstenberg, W. and Schulz, G. E. (1997) *J. Mol. Biol. 269*, 851-860.

5.   Sanz-Aparicio, J., Hermoso, J. A., Martinez-Ripoll, M., Lequerica, J. L. and Polaina, J. (1998) *J. Mol. Biol. 275*, 491-502.

6.   Gabelsberger, J., Liebl, W. and Schleifer, K.-H. (1993) *Appl. Microb. Biotechnol. 40*, 44-52.

7.   Cubellis, M. V., Rozzo, C., Montecucchi, P. and Rossi, M. (1990) *Gene 94*, 89-94.

8.   Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M. and Rossi, M. (1990) *Eur. J. Biochem. 187*, 321-8.

9.   Voorhorst, W. G., Eggen, R. I., Luesink, E. J. and de Vos, W. M. (1995) *J. Bacteriol. 177*, 7105-11.

10.   Kengen, S. W., Luesink, E. J., Stams, A. J. and Zehnder, A. J. (1993) *Eur. J. Biochem. 213*, 305-12.

11.   Bauer, M. W., Bylina, E. J., Swanson, R. V. and Kelly, R. M. (1996) *J. Biol. Chem. 271*, 23749-55.

12.   Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M. and Pearl, L. H. (1997) *J. Mol. Biol. 271*, 789-802.

13.   Chi, Y.-I., Martinez-Cruz, L.A., Jancarik, J., Swanson, R.V., Robertson, D.E. and Kim, S.-H. (1999) *FEBS Letters 445*, 375-383.

14.   Driskill L.E., Bauer M.W. and Kelly R.M. (1999) *Appl. Environ. Microbiol. 65*, 893-7.

15.   Bauer, M. W. and Kelly, R. M. (1998) *Biochemistry 37*, 17170-8.

16.   Stemmer, W. P. (1994) *Proc. Natl. Acad. Sci. U S A 91*, 10747-51.

17.   Kuchner, O. and Arnold, F. H. (1997) *Trends in Biotechnol. 15*, 523-530.

18.   Arnold, F. H. (1998) *Acc. Chem. Res. 31*, 125-131.

19.   Lorimer, I. A. and Pastan, I. (1995) *Nucleic Acids Res. 23*, 3067-8.

20.   Rosenberg, A. H., Lade, B. N., Chui, D. S., Lin, S. W., Dunn, J. J. and Studier, F. W. (1987) *Gene 56*, 125-35.

21.   Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60-89.

22.   Yanisch-Perron, C. and Messing V. J. (1985) *Gene 33*, 103-19.

23.   Bradford, M. M. (1976) *Anal. Biochem. 72*, 248-54.

24.   Gill, S. C. and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319-26.

25.   Witt, E., Frank, R. and Hengstenberg, W. (1993) *Protein Eng. 6*, 913-20.

26.   Kempton, J. B. and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.

27.   Fischer, L., Bromann, R., Kengen, S. W. M., Devos, W. M. and Wagner, F. (1996) *Biotechnology 14*, 88-91.

28.   Boon, M. A., Van der Oost, J., De Vos, W.M., Janssen, A.E.M. and Van `t Riet, K. (1999) *Appl. Biochem. Biotechnol. 75*, 269-278.

29.   Painbeni, E., Valles, S., Polaina, J. and Flors, A. (1992) *J. Bacteriol. 174*, 3087-91.

30.   Jaenicke, R. and Bohm, G. (1998) *Curr. Opinion Struct. Biol. 8*, 738-748.

31.   Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. and Engel, P. C. (1998) *Biochimie 80*, 933-941.

32.   Karshikoff A, L. R. (1998) *Protein Eng. 11*, 867-72.

33.   Hecht, K., Wrba, A. and Jaenicke, R. (1989) *Eur. J. Biochem. 183*, 69-74.

34.   Vieille, C., Hess, J. M., Kelly, R. M. and Zeikus, J. G. (1995) *Appl. Environ. Microbiol. 61*, 1867-75.

35.   Vieille, C. and Zeikus, J. G. (1996) *Trends in Biotechnol. 14*, 183-190.

36. Lebbink, J. H.G., Eggen, R. I. L., Geerling, A. C. M., Consalvi, V., Chiaraluce, R., Scandurra, R. and de Vos, W. M. (1995) *Protein Eng.8*, 1287-94.

37. Lebbink, J. H. G., Knapp, S., van der Oost, J., Rice, D., Ladenstein, R. and de Vos, W. M. (1998) *J. Mol. Biol. 280*, 287-96.

38. Kraulis, P.J. (1991) *J. Appl. Crystallog. 24*, 946-950.

39. Kopp, J. (1998) PhD thesis, Universität Freiburg, Germany

40. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.

6

Optimized catalytic features of extremely thermostable β-glycosidase hybrids created by DNA family shuffling

Authors

Thijs Kaper
Stan J.J. Brouns
Ans C.M. Geerling
Willem M. de Vos
John van der Oost

Abstract

A library of 2000 hybrid β-glycosidases was constructed by DNA family shuffling. The genes of two extremely stable β-glycosidases, CelB from *Pyrococcus furiosus* and LacS from *Sulfolobus solfataricus*, which share 56% nucleotide sequence identity, served as starting material. The hybrids were tested for their thermostability, ability to hydrolyze lactose and sensitivity towards inhibition by glucose. Three screening rounds at 70 °C led to the isolation of three high performance hybrid enzymes (hybrid 11, 18 and 20) that had 1.5 - 3.5 fold and 3.5 – 8.6 fold increased lactose hydrolysis rates compared to parental CelB and LacS, respectively. Sequence analysis revealed that all three variants are different, but resulted all from of a single crossover event resulting in hybrids with a LacS N-terminus and a CelB core.. Constructed 3D models of the hybrid enzymes visualized that the catalytic $(\beta\alpha)_8$-barrel was composed of both LacS and CelB elements. In addition, an extra intersubunit H-bond in hybrids 18 and 20 might explain their superior stability over hybrid 11. This study demonstrates that extremely thermostable enzymes with limited homology and different mechanisms of stabilization can be used to create stable hybrids with improved catalytic features.

6.1 Introduction

Elevated temperatures result in enhanced reaction rates for most enzyme-catalyzed conversions (*1*), which is advantageous for many biotechnological processes (*2*). However, the performance of enzymes under such conditions is often limited by their stability (*3, 4*). Although stabilization of proteins by rational design has been successful in a few cases (*5*), enzymes often have proven to be too complex to fully predict the effects of mutations on protein structure and catalysis (*6*). Recently, however, powerful strategies for the improvement of enzymatic properties have been developed that do not require information on protein structure or function. These strategies involve the generation of numerous enzyme variants by techniques in which wild-type or randomly mutated genes are recombined by DNA shuffling (*7, 8*), staggered extension process (StEP) (*9*) or incremental truncation for creation of hybrid enzymes (ITCHY) (*10*). Thus generated libraries are eventually screened for enzyme variants with desired features.

In order to enable biocatalysis at elevated temperatures, these techniques have recently been applied successfully for stabilization enzymes (*11-13*). Alternatively, one could start with enzymes from hyperthermophilic sources that generally have a superior stability compared to their mesophilic counterparts. Hence, their stability does not require optimization, but rather the catalytic features for improved performance at lower temperatures. Previously, we developed a random mutagenesis system for the β-glucosidase CelB from *Pyrococcus furiosus* and applied it for the generation of variants with high-performance at low-temperature (*14*). Since then, few studies have reported on the increased catalysis of thermostable enzymes, at temperatures well below their original optimum (*15, 16*). In these studies single genes were subjected to *in vitro* or *in vivo* random mutagenesis, followed by a suitable screening at ambient temperatures.

In the present study, we have focussed on the generation of functional hybrids by DNA shuffling of two genes coding for extreme thermostable β-glycosidases. The 'parental' enzymes are the β-glucosidase CelB from *Pyrococcus furiosus* and β-glycosidase LacS from *Sulfolobus solfataricus*, which are among the most thermo-active members of family 1 glycosyl hydrolases with optimal temperatures for hydrolysis of 105 and 95 °C, respectively (*17, 18*). The enzymes are 53% and 56% identical on the protein and DNA level, respectively, and have a similar catalytic mechanism and substrate specificity (*19*). However, the molecular basis of the high thermostability appears to be different in CelB and LacS. A biochemical comparison suggested that CelB is mainly stabilized by hydrophobic interactions, while salt-bridge interactions are crucial for the stability of LacS (*19*). Indeed, analysis of the crystal structures of CelB (*20*) and LacS (*21*) revealed a higher number of ion-pairs in the less stable LacS protein (*22*). Besides being model enzymes for the study of enzyme stability and high-temperature catalysis, CelB and LacS have been tested for their potential in the synthesis of oligosaccharides (*23-25*) and glycoconjugates (*26*). A recent study recognized CelB and LacS as potentially interesting biocatalysts for the hydrolysis of the milk sugar

lactose (galactose β-1,4-D-glucose) to monomeric glucose and galactose (*27*). However, for CelB a significant glucose inhibition was observed, while LacS displayed limited stability.

Here, we report the creation of functional β-glycosidase hybrids by shuffling of the genes coding for *P. furiosus* CelB and *S. solfataricus* LacS. In order to study catalysis at sub-optimal temperatures, a library of 2000 active variants has been screened for optimized hydrolysis of lactose at 70 °C in combination with a reduced inhibition by glucose as a model system. To our knowledge this is the first study on DNA family shuffling of extreme thermostable enzymes. After three screening rounds, three high-performance variants and one variant with reduced stability were isolated, purified, and characterized in detail on a biochemical and molecular level. Using crystal structures of CelB (*20*) and LacS (*21*), 3D models of the hybrids were constructed by homology modeling. The results show that it is possible to improve the catalytic features of hyperthermostable enzymes for catalysis at temperatures well below the optimum of the wild-type enzymes. Moreover, we found that hyperthermostable enzymes can tolerate a significant number of non-wild-type residues while retaining a considerable thermostability..

6.2 Experimental Procedures

Library construction. For the production of *P. furiosus* β-glucosidase CelB in *E. coli* an efficient expression system has been developed by cloning the *celB* gene in the pET9d vector, resulting in pLUW511 (*28*). Similarly, the *lacS* gene (*29*) was cloned in the pET9d vector (kan^R, p$_{T7}$, Novagen). The *lacS* gene was amplified in a PCR reaction using *Pfu* Turbo polymerase (Stratagene) and homologous primers BG745 (5'-GCGCGC<u>CCATGG</u>**C**ATAC CATTTCCAGATAGCTTT-3', introduced *Nco*I restriction site underlined, *lacS* start codon in bold face, introduced codon for alanine at position 1a in italics) and BG746 (5'-GCGCGC<u>GGATCC</u>TTAGTGCTTTAA TGGCTTTACTG-3', introduced *Bam*HI restriction site underlined, *lacS* stop anticodon in bold face). The PCR product was digested with *Nco*I and *Bam*HI and cloned in *Nco*I-*Bam*HI digested pET9d vector to form pWUR6.

Subsequently, the plasmids pLUW511 and pWUR6 were used as starting material for the DNA shuffling procedure in which *Pfu* turbo DNA polymerase (Stratagene) was used in all assembly and DNA amplification steps. The *celB* and *lacS* genes were amplified by PCR using the homologous primer sets BG238/BG239 (*28*) and BG745/BG746, respectively. Using DNAseI the genes were randomly digested as described before (*30*). The digestion mixture contained 4 µg DNA and 1 µg DNAseI per 50 µl DNAseI buffer (5 mM Tris-HCl pH 7.4, 1 mM MnCl$_2$) and was performed at room temperature in triplicate. Reactions were stopped at intervals by addition of 5 µl 0.5 M EDTA and stored on ice. DNA fragments were analyzed on a 1.5% agarose gel and fragments in the range of 50-300 bp (derived from 13-15 min incubation) were excised and purified using the QiaexII gel extraction kit (Qiagen). The fragmented genes were assembled by performing a PCR without primers. The reaction mix contained 0.5 µg of each fragmented gene, 0.2 mM of

each dNTP, 2.5 U *Pfu* turbo DNA polymerase (Stratagene) in the supplied buffer in a total volume of 50 μl. Reported assembly reactions have been carried out with annealing temperatures of 40 °C (*31*). Due to the limited homology of the celB and lacS genes, however, the use of lower annealing temperatures for re-assembly was explored. Attempts with annealing temperatures of 10 °C rendered unsuccessful. Hybrid genes were obtained by using a Tgradient thermocycler (Biometra), when the mixtures were subjected to 5 min 95°C, followed by two temperature programs of 21 cycles. The first 21 cycles consisted of 1 min 95°C, 1 min at 20°C with 1°C increase per cycle and 1 min 72°C with 5 sec increase per cycle. The second 21 cycles were 1 min 95°C, 1 min 40°C and 2 min 40 sec at 72°C with 2 sec increase per cycle. Subsequently, the shuffled constructs were enriched in PCR reactions using hybrid primer sets. In reaction mixtures containing the primer combination BG238/BG746 hybrid genes with N-terminal CelB and C-terminal LacS sequence were amplified. The primer combination BG745/BG238 selected genes with N-terminal LacS and C-terminal CelB sequence. This strategy introduced a bias towards hybrid genes that had termini derived from both gene parents. Sequences with termini derived from one parent with a middle section from the other were not amplified and therefore excluded from further screening.

The amplification mixtures contained 100 ng recombined DNA fragments, 5 pmol of each primer, 0.2 mM of each dNTP, and 2.5 U *Pfu* Turbo polymerase in the supplied buffer in a total volume of 50 μl. Temperature program was 2 min at 95°C followed by 25 repeats of 30 sec at 95°C, 1 min at 58°C, and 45 sec at 72°C with a 20 sec increase per cycle, which were followed by a final step of 5 min at 72°C. The PCR products were purified using the Qiaquick PCR purification kit (Qiagen) before digestion with *Nco*I and *Bam*HI. Digested hybrid genes were ligated in the similarly digested pET9d vector by T4 DNA ligase. *E. coli* JM109(DE3) (*lacZ* strain) electrocompetent cells were transformed with the ligation mixture using a Electroporator 2510 (Biorad) and plated on selective TY-agar (selective TY: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 30 μg/ml kanamycine, with 1.5% granulated agar) containing 1.6 μg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal)[1]. Colonies that expressed β-glycosidase activity showed a blue phenotype and were transferred to sterile flat bottom a 96-well microplate that contained selective TY-medium with 10% glycerol. Each plate contained negative (*E. coli* JM109(DE3)/pET9d) and positive controls (*E. coli* JM109(DE3)/pLUW511 or pWUR6). 2048 colonies were isolated and grown overnight at 37°C while shaking. These glycerol stocks were stored at −80°C.

Screening for lactose hydrolysis at 70°C. For the preparation of cell-free extract (CFE), the glycerol stock plates were replicated in 96-well plates containing selective TY-medium and grown for 72 hrs at 37°C while shaking. Induction of the expression system was not necessary, because leakage of the lacUV5 promoter results in a low level of constitutive T7 polymerase expression, as described previously (*28*). Cells were lyzed by freezing-thawing and the cell-free extract was subjected to a heat incubation of 60 min at 70°C, which denatured most of the *E. coli* host proteins.

After o/n incubation at 4°C the majority of the denatured proteins precipitated at the well bottom leaving a clear CFE. Microplates containing either 200 µl L-buffer (100 mM lactose in citrate buffer (150 mM sodium citrate, pH 5.0)) or 200 µl LG-buffer (L-buffer containing 10 mM glucose) per well were covered to prevent evaporation and preheated in a 70°C waterbath for 15 min. In an insulated microplate holder to prevent cooling, 5 µl CFE was transferred to the plates containing L-buffer and LG-buffer. Next, the plates were covered, manually inverted for mixing, and placed in a 70°C waterbath. After 60 min the hydrolysis reaction was stopped by placing the microplates in a water-ice mixture. The liberated glucose was measured by transferring 5 µl of the hydrolysis reaction to 200 µl GOD-PAP glucose detection kit (Roche) in 96 well microplates. After 30 min incubation at room temperature, the developed color was measured at 492 nm in a Thermomax microplate reader (Molecular Devices). The amount of liberated glucose was calculated from a calibration curve. The accuracy of the GOD-PAP kit allowed for a reliable calculation of the liberated glucose in the presence of 10 mM glucose.

Initial characterization of high performance hybrids. For a more detailed analysis, hybrids were grown o/n in 15 ml selective TY-medium at 37°C while shaking. At OD_{600} 1.0, isopropylthio-β-D-galactopyranoside (IPTG) was added to final concentration of 0.5 mM. Cells were spun down and resuspended in 0.75 ml citrate buffer pH 5.0. Cells were lysed by sonication and CFE was incubated at 70°C for 60 min. This yielded at least 90% pure hybrid β-glycosidase after centrifugation at 13 krpm for 30 min, as was judged by SDS-PAGE analysis. The hybrids were tested for their lactose-hydrolyzing activity and thermostability. For the lactose-hydrolyzing activity, samples of 0.25 µg enzym were added to 0.5 ml of preheated 100 mM lactose and 300 mM lactose both in citrate buffer and incubated at 70°C for 15 min min. Glucose was detected with the GOD-PAP kit. Thermostability of the hybrids was tested by incubating 50 µl of 50 µg/ml enzyme solution at 70°C for 1.5 hrs. The remaining activity was determined by incubating 1 µl enzyme solution with 0.5 ml 2 mM pNp-Gal in citrate buffer at 70°C and, after addition of 1 ml 0.5 M Na_2CO_3, measuring the liberated nitrophenol at 405 nm. The remaining activity was divided by the initial activity to give the inactivation rate.

Purification high performance hybrids. E. coli JM109(DE3) clones harboring pLUW511, pWUR6 or pET9d derivatives with hybrid genes, were grown o/n in 125 ml selective TY medium at 37°C while shaking. Prior to harvesting, the cultures were induced for 4 h by the addition of IPTG to a final concentration of 0.5 mM. Cells were collected by centrifugation at 5400 g for 10 min, resuspended in citrate buffer, and lyzed by sonication (Sonifier B12, Branson). E. coli proteins were denatured by a heat incubation at 70°C for 1.5 h and pelleted by centrifugation at 47.800 g for 30 min. Ammonium sulfate was added to the supernatant to a final concentration of 1.0 M. After filtering, the supernatant, containing 1.0 M $(NH_4)SO_4$, was loaded on a phenyl-superose column coupled to an Äkta FPLC (Amersham-Pharmacia-Biotech), equilibrated with 20 mM Tris-HCl (pH 8.0) with 1.0 M $(NH_4)SO_4$. Proteins were eluted by a linearly decreasing $(NH_4)_2SO_4$ gradient (1.0-0

M) and ran off the column around 0.2 M (NH$_4$)SO$_4$. Active fractions were pooled and dialyzed against 20 mM of sodium phosphate (pH 7.5). Absence of contaminating DNA was verified in wavelength scans in the range of 250-300 nm. Proteins were pure as judged by SDS-PAGE analysis. Protein concentrations were determined by measuring the absorption at 280 nm, in which calculated specific ε_{280} values of 128280 M^{-1}cm^{-1} and 140370 M^{-1}cm^{-1} were used for CelB and LacS, respectively (*32*).

Enzyme activity assays. Enzymes were tested for the hydrolysis of lactose, cellobiose, *para*-nitrophenyl-β-galactopyranoside (pNp-Gal) and *para*-nitrophenyl-β-glucopyranoside (pNp-Glc) at 70°C. In 1.5 ml microvials, 0.5 ml substrate in citrate buffer was preheated for 5 min. Reaction was started by the addition of 1 µg of enzyme in 5 µl and incubated for 15 minutes. In the case of lactose and cellobiose, the reaction was stopped by placing the tubes on ice. The liberated glucose was measured using the GOD-PAP glucose detection kit (Boehringer-Mannheim). In the case of pNp-Gal and pNp-Glc, the reaction was terminated by addition of 1.0 ml 0.5 M Na$_2$CO$_3$. The liberated nitrophenol was measured at 405 nm. The ε_{405} of pNp was determined as 15.8 mM^{-1}cm^{-1}. Kinetic data were obtained by fitting the data to the Michaelis-Menten equation using the non-linear regression program Tablecurve 2D (Jandel Scientific). Inhibition constants were determined for glucose and the thermostable transition state analogue (5R,6R,7S,8S)-5-(hydroxymethyl)-2-phenyl-5,6,7,8-tetrahydroimidazol[1,2-a]pyridine-6,7,8-triol (PheImGlc, generously supplied by Dr. Andrea Vasella, ETH Zürich, Switzerland) (*33*) with pNp-Gal as a substrate. An initial K$_i$ value was determined by incubation the enzyme in 3 mM pNp-Gal in the presence of 8 different concentrations of glucose (10-500 mM) or PheImGlc (1-100 nM) and plotting the data in a Dixon plot. Final inhibition constants were calculated by determining the enzyme's activity at 16 different pNp-Gal concentrations around the K$_m$ in the presence of 4 inhibitor concentrations bracketing the initial K$_i$ value. The apparent K$_m$ values were determined from Lineweaver-Burke plots and plotted against the inhibitor concentrations. A linear regression line was extrapolated to the x-intercept, which indicated the negative K$_i$ value.

Stability studies. Half-lives of thermal inactivation were determined at 92°C in 20 mM sodium phosphate pH 7.5 using 50 µg/ml enzyme. At time intervals, aliquots were removed and the remaining activity was measured using 3 mM pNp-Gal. Half-lives were calculated by fitting the data to first order inactivation kinetics. Optimal temperatures for catalysis were determined by measuring specific activities at 60, 70, 75, 80, 85, 90, 95, 100, 105 and 110 °C for the hydrolysis of 20 mM pNp-Gal in citrate buffer. Below 90 °C, 0.5 ml buffer with substrate were preheated for 5 min in a waterbath, before addition of 5 µl enzyme solution. After 5 min reactions were stopped by addition of 1 ml 0.5 M Na$_2$CO$_3$. At 90 °C and above, 250 µl buffer with substrate was preheated in 1 ml glass HPLC vials in a silicon oil bath for 5 min. Reactions were started by adding of 5 µl enzyme solution and terminated by addition of 0.5 ml 0.5 M Na$_2$CO$_3$. Liberated nitrophenol was measured at 405 nm using a spectrophotometer (Hitachi). Melting temperatures of wild-type and

Table 6.1. Characterization of high performance hybrids: lactose hydrolysis at 70°C and thermostability. ND: not determined

Hybrid	Lactose hydrolyzing activity (U/mg)		Residual activity after 1.5 h 70°C (%)
	100 mM	**300 mM**	
2	333	540	128
3	157	396	3
4	261	562	89
5	308	456	140
6	384	553	138
7	166	310	2
8	378	540	134
9	385	549	116
10	782	1119	107
11	554	926	117
12	604	959	128
13	545	960	130
14	804	941	139
15	588	949	125
16	619	996	114
17	624	948	121
18	728	1146	139
19	32	73	3
20	727	1089	110
21	637	902	132
CelB	400	581	ND
LacS	223	263	ND

hybrid enzymes were determined by differential scanning calorimetry in a VP-DCS micro calorimeter (MicroCal). Enzyme solutions were extensively dialyzed against 20 mM sodium phosphate buffer pH 7.5 and diluted to 4.6 µM. After 10 minutes degassing and equilibration, samples were analyzed between 50-125°C at 0.5°C/min against dialysis buffer. Enzyme scans were corrected using a buffer-buffer baseline.

Molecular analysis and homology modeling. The DNA sequence of hybrid genes was determined using the Thermo Sequenase kit (Amersham-Pharmacia-Biotech) with IR-labeled primers (MWG, Ebersberg, Germany) and subsequent analysis on an automated sequencer (LiCor, Lincoln, USA). Three-dimensional tetrameric models of the hybrid enzymes were obtained by

molecular replacement using the crystal structures of CelB (*20*) and LacS (PDB entry: 1GOW) as search models by SWISS-MODEL (*34*).

Construction of CelB S14A mutant. Using pLUW511 as a template, a S14A substitution was introduced in the *celB* gene via the QuickChange Site-Directed Mutagensis Kit (Stratagene) using primers BG996 (5'-TATTCTTGGgCTG GTTTCCAG-3', introduced mutation in lower case font) and BG997 (5'-CTGGAA ACCAGcCCAAGAATA-3'). The sequence with introduced mutation was verified by DNA sequence analysis (see above). CelB S14A protein was produced and purified as wild-type CelB (see above).

6.3 Results

Isolation of high performance hybrids. After the DNA shuffling procedure, 10,000 colonies were screened for X-Gal hydrolyzing activity on TY-agar plates at 37°C. About 20% of the colonies showed a blue phenotype indicating functional β-galactosidase expression. At this level the variety in functional hybrid enzymes was visible by the differences in blue color intensity between individual colonies. The enzymes were then tested for their (1) thermostability at 70 °C, (2) lactose hydrolyzing capacity and (3) inhibition by glucose. The heat incubation inactivated a significant part of the variants, indicating the presence of functional, but thermolabile hybrids. One of these instable variants, hybrid 1, was studied in more detail with respect to its stability and, after sequencing, a 3D model was constructed by homology modeling (see below). In each of the two subsequent screening rounds, 10% of the hybrids was selected for re-screening, which finally yielded 20 thermostable hybrid enzymes with improved lactose hydrolyzing activity and/or a reduced inhibition by glucose. These variants were partially purified and tested at 70°C for their activity on lactose, glucose inhibition, and stability. This more detailed characterization revealed that 11 hybrids, indeed showed significantly higher lactose-hydrolyzing activity at 70°C compared

Table 6.2. Kinetic parameters of wild-type and hybrid enzymes for the hydrolysis of lactose and cellobiose at 70°C. In the assays 0.1 μg enzyme was used in 150 mM sodiumcitrate pH 5.0.

Substrate	Lactose			Cellobiose[1]			
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1}s^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1}s^{-1})	$K_{m, lactose}/K_{m,cellobiose}$
CelB	59.8 ± 12.1	286 ± 15	4.78	3.2 ± 2.6	37.6 ± 3.2	11.8	18.7
LacS	37.9 ± 17.3	121 ± 13	3.19	13.5 ± 8.4	23.2 ± 2.9	1.72	2.81
11	71.8 ± 10.2	442 ± 18	6.16	11.8 ± 3.9	42.5 ± 2.7	3.60	6.08
18	88.6 ± 8.4	1022 ± 32	11.5	4.6 ± 1.8	264 ± 13	57.4	19.3
20	59.5 ± 5.5	447 ± 11	7.51	12.1 ± 2.4	60.1 ± 2.3	4.97	4.92

[1] CelB showed typical substrate inhibition on cellobiose.

Table 6.3. Kinetic parameters of wild-type and hybrid enzymes for the hydrolysis of *p*Np-Gal and *p*Np-Glc at 70°C. In the assays 0.1 μg enzyme was used in 150 mM sodiumcitrate pH 5.0.

Substrate	*p*Np-Gal			*p*Np-Glc[1]			
	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1}s^{-1})	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (mM^{-1}s^{-1})	$K_{m,pNp\text{-}Gal}/K_{m,pNp\text{-}Glc}$
CelB	2.60 ± 0.26	298 ± 9	115	0.80 ± 0.11	206 ± 11	258	3.25
LacS	1.57 ± 0.21	165 ± 6	105	0.12 ± 0.02	45 ± 2	375	13.1
11	3.99 ± 0.35	327 ± 9	82.0	0.71 ± 0.29	235 ± 28	331	5.62
18	5.63 ± 0.51	743 ± 13	132	0.58 ± 0.09	440 ± 21	759	9.71
20	6.58 ± 0.85	412 ± 20	62.6	0.75 ± 0.13	276 ± 17	368	8.77

[1] CelB, hybrids 18 and 20 showed typical substrate inhibition on *p*Np-Glc at concentrations over 3 mM (wt) and 4 mM (18 and 20), respectively. Parameters have been determined using activities obtained at lower substrate concentrations.

to the parental enzymes (Table 6.1). Remarkably, all of these hybrids were specifically selected with primers for an N-terminal LacS and a C-terminal CelB sequence. Hybrid 18 was the most active variant, while hybrid 11 and 20 represented high-performance hybrids with a relatively low (11) and average (20) activity. These three hybrids were selected for further analysis.

Kinetic parameters. To obtain more insight in the altered biochemical properties the hybrid enzymes 11, 18 and 20, they were purified to homogeneity together with the CelB and lacS enzymes, as described previously (*28*) and characterized. At 70 °C the kinetic parameters of the purified enzymes were determined for the hydrolysis of lactose, cellobiose and their respective chromogenic analogs pNp-Gal and pNp-Glc (Tables 6.2 and 6.3). At 70 °C, CelB is little moreactive on the galactosides lactose and pNp-Gal, than on pNp-Glc, whereas the efficiency of cellobiose hydrolysis is relatively low. With doubled activities on galactosides compared to glucosides, LacS displays a more pronounced β-galactosidase activity than CelB, as reported previously (*19*). On all tested substrates, LacS is less active than the other enzymes, but has the highest affinity for all substrates except cellobiose. Compared to LacS, CelB is about twice as active on all substrates, however with equally augmented Michaelis constants. The substrate affinities of the hybrids are similar to those of CelB. Hybrids 11, 18 and 20 have highest activity for the hydrolysis of lactose, which was the substrate used in the screening. This lactose-hydrolyzing activity is about 1.5-3.5 and 3.6-8.6 times higher than that of CelB and LacS, respectively (Figure 1A). Cellobiose hydrolysis by the hybrids proceeds with a 6.9- to 11.4-fold increased efficiency. For chromogenic substrates, the activities of the hybrids have been equally increased. CelB was inhibited at higher concentrations of the substrates cellobiose and pNp-Glc. Hybrid 20 showed similar inhibition, but only with pNp-Glc. All other substrates were hydrolyzed according to Michaelis-Menten kinetics.
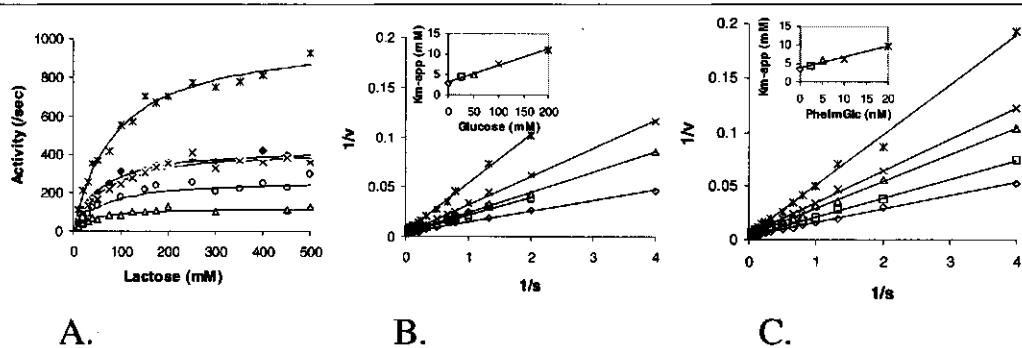
Figure 6.1: Michaelis-Menten and Dixon plots of hybrid enzymes. A: Michaelis-Menten kinetics of CelB (o), LacS (Δ), hydrids 11 (x), 18 ([) and 20 (◊) for the hydrolysis of lactose; B: inhibition of hybrid 11/20 by glucose; C: inhibition of hybrid 20 by transition state analogue PheImGlc; insets in (B) and (C) show the dependence of the K_m for glucose and PheImGlc, respectively.

Inhibition by glucose and PheImGlc. The inhibition constants of glucose and the thermostable transition state analogue PheImGlc on the hydrolytic activity of CelB, LacS and hybrids 11, 18 and 20 were determined (Table 6.4). pNp-Gal was chosen as a substrate, since no substrate inhibition was observed for the hydrolysis of this compound with any of the tested enzymes. The inhibiting effect of glucose on the activity of CelB and LacS is similar; the inhibition constants for glucose are comparable, whereas the three high-performance hybrids are about 2 times less sensitive for glucose inhibition (Fig. 1B). PheImGlc is a glucose derivative that has a flat geometry at the anomeric C1 saccharide atom and, as such, is believed to resemble the transition-state in the glycoside hydrolysis reaction in retaining β-glycosidases *(33)*. For all enzymes the strong inhibition by PheImGlc followed kinetics for competitive inhibition (Figure 1C). The inhibition constants for CelB and LacS were similar. The K_i value determined for hybrid 11 was slightly lower than that for CelB and LacS, while the K_i-value for hybrid 18 and 20 was about 2 times the value of hybrid 11.

Stability. The hybrids 11,18 and 20 were compared to CelB and LacS with respect to their thermal inactivation rate, optimum temperature for catalysis and melting temperature (Table 6.5). The different analysis methods roughly display similar trends for the enzyme variants. Hybrid 1 is the least stable variant, while the stability of hybrids 11, 18 and 20 are intermediate of that of CelB and LacS. Hybrid 20 is the most stable hybrid regarding the inactivation rate and melting temperature, while hybrid 18 has a slightly higher optimum temperature for hydrolysis (Figure 2). Remarkably, LacS is readily inactivated at 92 °C, while it displays a broad optimum temperature for catalysis, with a peak around 95 °C. For CelB and hybrid 1 and 18, the determined optimal temperature for catalysis is very close to the determined melting temperature. For hybrid 11 and 20,

Table 6.4. Inhibition of *p*Np-Gal hydrolysis by glucose and transition state analogue PheImGlc at 70 °C. In the assays 0.05 µg enzyme was used with 150 mM sodiumcitrate pH 5.0 as assay buffer.

	*p*Np-Gal		Glucose	PheImGlc
	K_m	k_{cat}	K_i	K_i
	(mM)	(s^{-1})	(mM)	(nM)
CelB wt	2.60 ± 0.26	298 ± 9	43	9.3
LacS	1.57 ± 0.21	165 ± 6	46	8.4
11	3.99 ± 0.35	327 ± 9	110	5.9
18	5.78 ± 0.51	725 ± 28	88	12
20	6.58 ± 0.85	412 ± 20	80	13

however, the T_{opt}-value is almost 10 degrees lower than the determined T_m-value. This might suggest different inactivation patterns between the hybrids. It was not possible to determine a T_m-value for LacS. This is most likely due to the fact that the unfolding of LacS does not involve two states, but rather occurs via a stable intermediate, which has been identified during guanidine hydrochloride-induced unfolding of LacS (Kaper and Van der Oost, unpublished results) (*35*). In previous differential scanning calorimetric experiments with LacS, a two-state unfolding was observed, but these experiments were performed in 10 mM CAPS buffer at pH 10 (*36*). Under those conditions, the melting temperature of LacS was 86.3 °C.

Structural analysis of hybrids. The primary structure of several hybrids was determined by DNA sequence analysis. All hybrids are the result of a single crossover event in different regions with 5 to 22 bases identical DNA sequence (Figure 3). All sequenced hybrids contain mainly the CelB amino acid sequence in which the C-terminus or N-terminus has been substituted with LacS sequence (Figure 3 & 4A). Starting at the N-terminus, hybrids 1 and 7 have 419 and 413 amino acids derived from CelB, followed by 55 and 61 residues of LacS until the C-terminus, respectively (Figure 3A). Hybrid 9 was identical to wild-type CelB. Hybrid 11 and 20 are very similar. Both proteins start with a LacS sequence of about 40 residues, followed by CelB sequence until the C-terminus. The only differences between hybrid 11 and 20 are the residues at position 43 and 45 (Figure 3B). Hybrid 11 has LacS residues in this position, respectively a methionine and an alanine, whereas hybrid 20, like CelB, has an isoleucine and a serine at these positions. Hybrid 14 was found to be identical to hybrid 20, while hybrid 18 was LacS until residue 33 with a point mutation that resulted in T23P, followed by 441 residues of CelB sequence (Figure 3B). The gene of hybrid 19 consisted of 68 nucleotides of LacS sequence followed by the complete wild-type *celB* gene sequence.

To obtain insight in the structural characteristics of the hybrids, 3D-models were constructed of hybrids 1, 11, 18 and 20 by homology modeling using the available crystal structures of CelB (*20*) and LacS (*21*). The structural models can be used for a global interpretation of the hybrid

Table 6.5. Half-lives at 92°C, optimal temperature for catalysis and melting temperatures of wild-type and hybrid enzymes.

	Half-life at 92°C (min)	T_{opt} (°C)	T_m (°C)
CelB wt	> > 100	104	106.0
LacS	< 3	95	ND [a]
SB1	30 [b]	80	80.7
SB11	7	85	94.0
SB18	8	96	94.5
SB20	100	93	101.8

[a] ND: not determined

[b] at 80°C

structures, since the resolutions of the CelB and LacS structures (3.3 and 2.6 Å resolution, respectively) do not allow for a detailed analysis of the exact positions of the amino acid side chains. Therefore, they provide a basis for the interpretation of the structural rearrangements in the chimeric enzymes. In hybrid 1, the LacS sequence replaces almost completely the CelB-CelB intersubunit contacts at the small subunit interface of the tetramer (Figure 4B). The 55 amino acid residues of the LacS C-terminus are 55% identical to the C-terminus of CelB. The LacS sequence is two residues longer than in CelB and fills up the cavity that is present in the center of CelB at the tetrameric subunit interface. The C-terminus is involved in an elaborate ion-pair network in LacS, where the penultimate arginine residue reaches across and participates in a 16-residue ion-pair network, that bridges the four subunits (*21*). In general, ion pairs have been shown to be involved in stabilization of proteins against high temperatures (*37*). In the C-terminus of CelB several charged residues are located that are possibly involved in an ion-pair, but align with non-charged residues in LacS and vice versa. Therefore, in hybrid 1 these possible ion pairs can not be formed. Interestingly, the ion partner of the next-to-last arginine residue in LacS, Glu345, aligns with Glu330 in CelB, which could indicate a LacS-like C-terminal organization in hybrid 1. However, the changes in hydrophobic interactions and H-bridges can not be evaluated with these models. Overall, the loss in thermostability can be interpreted as an incompatability of LacS and CelB in the structural stabilization of the C-terminus and the subunit interface. The active site in hybrid 1 is equal to that of CelB. This could explain the catalytic profile of hybrid 1, which is similar to that of CelB (data not shown). In contrast to hybrid 1, hybrids 11, 18 and 20 contain LacS sequences that completely traverse the subunit of the protein (Figure 4C). A close up of the catalytic $(\beta\alpha)_8$-barrel visualizes that in both variants the first β-strand of the CelB barrel has been replaced by its LacS counterpart (Figure 4D&E). CelB and LacS are 51% identical in the first 43 and 45 amino acid residues of the protein (Figure 3B.). In LacS, the N-terminus is tightly bound to the protein body by

an ion-ion interaction of the α-amino group with the side chain of Asp473 (*21*). Such an interaction is not possible in the hybrids, where an alanine (Ala 1a) has been introduced after the N-terminal methionine residue to enable a translational fusion of the *lacS* gene to the T7 promoter in a *Nco*I restriction site. In CelB, however, the α-amino group does not appear to make ionic interactions. The side chain of Lys2 appears to form a ion pair with Glu455 in CelB. The interactions of thecharged residues in the LacS stretch in hybrid 11, 18 and 20 appear to be similar to those in wild-type LacS, which could explain the relatively high stability of the hybrids.

Hybrids 11 and 20 differ only at two positions, Met43 and Ala45 for hybrid 11 versus Ile43 and Ser45 for hybrid 20. Since we measured a significant higher stability for hybrid 20, this must solely result from the effects of the two different residues. Interestingly, both residues are located at the larger subunit interface of the glycosidase tetramer. Earlier studies have shown that quarternary interactions contribute largely to the stability of proteins (*38*). These intersubunit interactions can be hydrophobic interactions, salt-bridges or hydrogen bonds. In modeling studies the interactions of Met43 and Ile43 in hybrid 11 and 20, respectively, appear to be very similar. However, an extra possible hydrogen bond was identified in hybrid 20, which involves the side chain of Ser45 and the backbone nitrogen atom of Leu221, located at the adjacent subunit (Figure5A). Since hybrid 11 has an alanine at position 45 a similar interaction is absent in this variant (Figure 5B). In hybrid 18 the residues at positions 43 and 45 are the same as in hybrid 20. Surprisingly, it is significantly less stable than hybrid 20. The point mutation in hybrid 18, which resulted in T23P, causes a small shift in the position of the N-terminal peptide chain in the 3D model, compared to that in hybrid 20. This shift could have a destabilizing effect on the interactions that 23 N-terminal residues have with the rest of the protein and could be an explanation for the lower stability of hybrid 18.
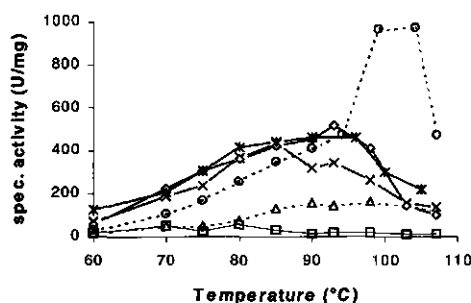

Figure 6.2: Typical influence of temperature on the activity of CelB, LacS and hybrids 1, 11, 18 and 20. Reactions were performed in 250 µl 20 mM pNp-Gal in 150 mM sodium citrate pH5.0 with 0.05 µg enzyme. At 60 and 70 °C, 75 to 85 °C and 90 to 107°C reactions were incubated for 10 min, 5 min and 3min, respectively. Dotted lines are CelB (o) and LacS (Δ). Solid lines are hybrid 1 ( ), 11 (x), 18 ([) and 20 (◊). 1 U equals the hydrolysis of 1 µmol of pNp-Gal per min.

A.

```
CelB    387   VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD   432
LacS    402   VSHVYQVHRAINSGADVRGYLHWSLADNYEWASGFSMRFGLLKVD   447
1       387   VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD   432
7       387   VSHLKAVYNAMKEGADVRGYLHWSLTDNYEWAQGFRMRFGLVYVD   432
              ***   *  *    ***********  ******  **  *****    **

CelB    433   PETKKRYLRPSALVFREIATQKEIPEELAHLADLKFVT--RK      472
LacS    448   YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH     489
1       433   YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH     474
7       433   YNTKRLYWRPSALVYREIATNGAITDEIEHLNSVPPVKPLRH     474
              **  *  ******  *****    *   *   **         *
```

B.

```
CelB    M-KFPKNFMFGYSWSGFQFEMGLPGSE-VESDWWVWVHDKENIASGLVSGDLPENGPA   56
LacS    MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPG   58
11      MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENMAAGLVSGDLPENGPA   58
18      MYSFPNSFRFGWSQAGFQSEMGPPGSEDPNTDWWVWVHDKENIASGLVSGDLPENGPA   58
14/20   MYSFPNSFRFGWSQAGFQSEMGTPGSEDPNTDWYKWVHDPENIASGLVSGDLPENGPA   58
        **  *  **  *    ***  ***  ****    **    ****   *  * ************
                                                          ⇑ ⇑
```

Figure 6.3. N- and C-terminal amino acid sequences of CelB, LacS and hybrids. CelB sequence has been underlined, while the corresponding regions of nucleotide sequence identity where crossovers took place, have been indicated in *italics*. C-terminal amino acid sequence of CelB, LacS and 1 and 7 (A). N-terminal sequences of CelB, LacS and hybrids 11,18 and 20 (B). The two residues in which hybrid 11 and 20 differ have been indicated by arrows. For reasons of clarity, the alanine residue at position 1a in the sequences is not shown.

Throughout family 1 of glycosyl hydrolases, the residues that line the active site have been highly conserved (*20*). Therefore, CelB and LacS share a nearly identical active site topology. However, in CelB the serine residue at position 14 is either a serine or an alanine residue in family 1 sequences (50/50). Ser 14 in CelB aligns with alanine at position 15 in LacS. The hydroxyl group of this serine side chain could possibly form a hydrogen bond with the residue at the non-reducing end of a hexose substrate in the active site. Overlay studies with the structure of the 6-phospho-β-galactosidase LacG from *Lactococcus lactis* with a bound galactose-6-phosphate in the active site (*39*) showed that Ser14 might interact with the 3-hydroxyl of galactose or glucose, which have an identical orientation at this position. To investigate whether this residue was involved in substrate inhibition, Ser14 was substituted by an alanine by site-directed mutagenesis resulting in mutant CelB S14A. However, this mutant was similarly inhibited by glucose as CelB and LacS and resembled CelB in substrate specificity and activity (data not shown).
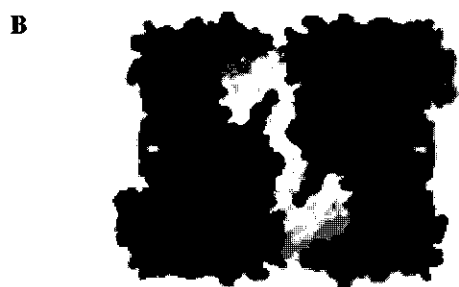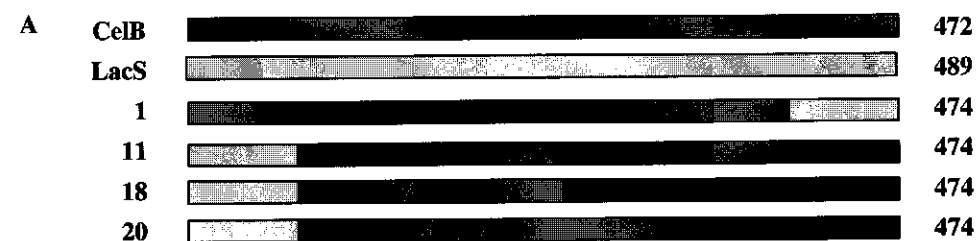
Figure 6.4 Structural representation of hybrids. Schematic represenation of primary structure of CelB, LacS, hybrids 1, 11, 18 and 20 (A). Surface plot of tetrameric hybrid 1 (B) and tetrameric hybrid SB20 (C). Side view (D) and top view (E) of hybrid 20 $(\beta\alpha)_8$-barrel. Images were generated using Swiss-PDB viewer (56) and visualized using Povwin (57).

6.4 Discussion

We here report on the construction of a library of hybrid β-glycosidases and the detailed characterization of isolated high-performance variants. The library was constructed by shuffling of the genes of the hyperthermostable family 1 β-glycosidases CelB from *P. furiosus* and LacS from *S. solfataricus*. This is the first example of improving extremely thermostable enzymes by a DNA family shuffling approach. The β-glycosidase-encoding genes used in this study are 56% identical, which is relatively low compared to other shuffling studies in which 60-90% identical genes were used (*8, 40-42*). The use of sequences with limited DNA identity introduced a risk of recombination of gene fragments back to wild-type sequences. However, the combination of extremely low annealing temperatures in the reassembly procedure and a subsequent PCR-reaction with hybrid primer sets, led to the formation and specific amplification of a variety in hybrid genes (Figure 3).

Previously, one of the parental enzymes in this study, the pyrococcal CelB, has been optimized for catalysis at room temperature by random mutagenesis in combination with DNA shuffling (*14*). It was demonstrated that variants with increased activity on pNp-Glc at room temperature, were not equally more active on the β-(1,4) linked glucose disaccharide cellobiose (*14*). Moreover, at higher temperatures the variants' activities approached wild-type activity, or were even less active. According to the first law in directed evolution "you get what you screen for" (*43*), the screening should be carried out with the substrate of interest at the conditions of interest. With this in mind, the screening described in this study was designed for the selection of hybrid β-glycosidases with improved hydrolytic properties for the natural sugar lactose at 70 °C. This screening proved to be reliable, since characterization of the purified enzymes confirmed the increased hydrolysis rates and decreased glucose sensitivities, observed in the screening rounds (Tables 1, 2, 3 and 4). Increased hydrolysis rates of three high performance hybrids were not restricted to hydrolysis of the galactose-β-(1,4)-glucose disaccharide. Increased turnover numbers were measured for hydrolysis of all tested substrates by hybrid 18 and, to a lesser extent, hybrid 11 and 20 (Table 2). In addition, the three hybrids were inhibited less by glucose, compared to the parental enzymes (Table 4). The inhibition of CelB, LacS and high-performance hybrids by PheImGlc is comparable and indicates that the stabilization of the reaction transition state in these enzymes is similar. Judging from the increased K_m-values for the hydrolysis of pNp-gal the affinity for the formation of Michaelis-complex has been reduced in the hybrids with respect to the parental enzymes, which is supported by the reduced glucose inhibition (Table 4). Therefore, the increase in k_{cat} might result from a less favorite binding of the ground state of the substrate, which would lower the activation energy to reach the transition state of the reaction (*44*).

The three high-performance hybrids resulted from single crossover events near the N-terminus of the protein. Based on the determined sequences of hybrids 1, 11, 18 and 20, 3D structural models were constructed using the available structures of CelB (*20*) and LacS (*21*). The hybrids are near-identical in sequence and consequently in structure, which means that structural
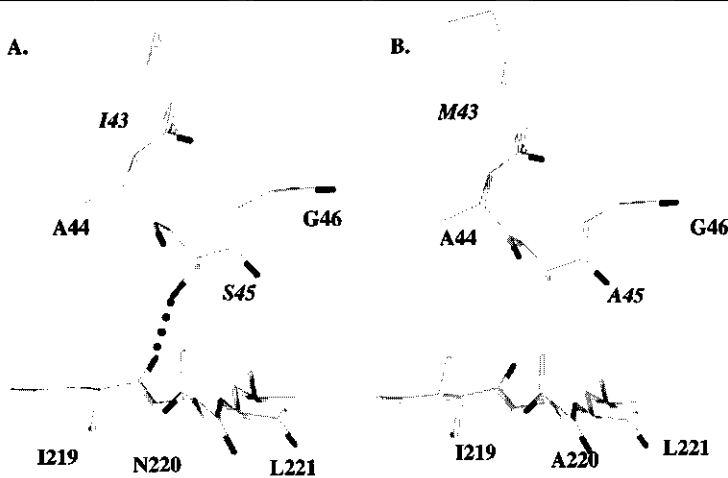
Figure 6.5 Modeled subunit interface interactions in hybrid 20 (A) and 11 (B) around residue 45. Residues that are different for hybrid 11 and 20 are in *italic* font. Images were generated using Swiss-PDB viewer (56) and visualized using Povwin (57).

explanations for substrate specificity have to result from differences between residue 34 to 45 in the hybrids (Figure 3). CelB has a more pronounced β-glucosidase character than LacS for the hydrolysis of disaccharides (Table 2). Interestingly, hybrid 18 has the shortest LacS sequence (residues 1-33) and resembles CelB more considering the ratio of the specificity for lactose and cellobiose (K_m), while hybrid 11 and 20 have more LacS-like specificities (Table 2). However, such specificity correlation can not be extended to the hydrolysis of chromogenic substrates, in which hybrid 18 and 20 resemble more LacS and hybrid 11 is CelB-like (Table 3). Therefore, there must be differences in binding of "natural" and chromogenic substrates, which can be rationalized when the differences in leaving group ability of the substrates are taken into account. For hydrolysis of disaccharides, the initial breaking of the glycosidic bond is most critical for the reaction rate due to the high pK_a-value of the leaving glucose (i.e. 12.4 (45)) (46). This requires considerable distortion of the non-reducing sugar residue (47). The second step of the reaction, hydrolysis of the covalent glycosyl-enzyme intermediate, is rate determining for hydrolysis of pNp-substrates, which have a good leaving group ($pK_a = 7.2$) (46). This indicates that initial binding of substrates is less critical for the hydrolysis of pNp-substrates. Trp420 is a conserved active site residue that has been found to interact with the 4-OH of substrates in the active site (39, 48), the hydroxyl that is axial in lactose and equatorial in cellobiose. Trp420 is positioned by Trp33, which is the first residue of a helix that contains Val37 in hybrid 18 and Lys 37 in hybrid 11 and 20. Probably, this difference has an effect on the positioning of Trp33 and Trp420 and, as such, the binding of substrates in the active site.

The gain in activity at lower temperatures is coupled to a decreased thermostability. The three hybrids have a stability that is between that of LacS and CelB, with hybrid 20 being the most

stable variant (Table 5, Figure 2). The loss in stability and increase in activity could be the result of increased flexibility, although this is not an absolute requirement for improved turnover numbers at lower temperatures, as has been found in other directed evolution studies (*11, 13*). The structural models were analyzed with respect to the hybrids' stability. The inferior stability of hybrid 1 most likely results from the loss of a relatively large number of stabilizing interactions at the C-terminus. The difference in stability between hybrid 11 and 20 is intriguing since the two hybrids differ only in two residues. The 3D model shows four extra intersubunit H-bridges in hybrid 20, which could explain the 7 °C difference in melting temperatures between the two hybrids. The forces that counteract conformational entropy loss and result in protein folding have a net energy gain for folded proteins that equals only a few weak intermolecular interactions (*49*). Therefore, the introduction of four H-bridges might result in such a large stability increase. Mutational studies of subunit interface interactions have mainly dealt with ion-pair interactions, but similar differences in melting temperatures have been reported upon altering single residues at subunit interfaces (*50, 51*).

High performance catalysis at 70 °C resulted from the incorporation of an N-terminal LacS sequence in CelB (Figure 3). In CelB, optimized low-temperature catalysis resulted from mutation of residues at various locations in the enzyme, ranging from second shell residues around the active site to residues on the surface or subunit interface (*14*). These substitutions could not have been predicted by rational design. Most mutations resulted in a reduced thermostability, an indication for the highly optimized protein structure of wild-type CelB (*14*). In the same enzyme family the β-glucosidase BglA of mesophilic *Bacillus polymyxa* has been the target of mutagenesis studies to increase its stability (*52*). The modeled structure of a single mutant BglA variant revealed the introduction of a non-wild-type ion-pair that reduced the flexibility of a large number of residues (*53*). Similar observations have been noted after detailed analysis of laboratory-evolved thermostable esterases (*54, 55*). Although the effects of most amino acid changes on conformation and flexibility could be rationalized, the authors state that it would have been impossible to design these mutations, even with the help of high-resolution protein structures. Without elaborated structural and conformational studies it is, therefore, difficult to rationalize the changed catalytic features of the hybrid β-glycosidases which have been evolved in this study.

This study has demonstrated that DNA family shuffling is also applicable for directed evolution of highly thermostable enzymes with different mechanisms of stabilization. We have successfully shuffled the genes coding for the extremely thermostable β-glycosidases CelB from *Pyrococcus furiosus* and LacS from *Sulfolobus solfataricus*. Hybrid enzymes were selected for thermostability and improved catalytic properties for the hydrolysis of lactose at 70°C. Composition of the hybrids showed that the highly optimized β-glycosidases can tolerate significant foreign peptide stretches, while remaining very thermostable. These results have encouraged us to continue directed evolution of enzymes from extreme thermophilic origin. At the moment the laboratory evolution of enzymes is the most promising strategy for improvement of enzymes, since the

necessary structural changes for altered catalytic or stability features are too subtle to be predicted by rational design. Analysis of high-performance hybrids is likely to increase our understanding of protein stability and catalysis at elevated temperatures.

6.5 Acknowledgements

References

1. Segel, I. H. (1993) *Enzyme kinetics : behavior and analysis of rapid equilibrium and steady-state enzyme systems*, Wiley classics library ed, New York.
2. Niehaus, F., Bertoldo, C., Kähler, M., and Antraninkian, G. (1999) *Appl Microbiol Biotechnol 51*, 711-729.
3. DeSantis, G., and Jones, J. B. (1999) *Curr Opin Biotechnol 10*, 324-30.
4. Govardhan, C. P. (1999) *Curr Opin Biotechnol 10*, 331-5.
5. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. H. (1998) *Proc Natl Acad Sci USA 95*, 2056-60.
6. Cedrone, F., Ménez, A., and Quéméneur, E. (2000) *Curr Opin Struct Biol 10*, 405-410.
7. Stemmer, W. P. C. (1994) *Nature 370*, 389-91.
8. Crameri, A., Raillard, S.-A., Bermudez, E., and Stemmer, W. P. C. (1998) *Nature 391*, 288-291.
9. Zhao, H., Giver, L., Affholter, J. A., and Arnold, F. H. (1998) *Nat Biotechnol 16*, 258-261.
10. Ostermeier, M., Nixon, A. E., Shim, J. H., and Benkovic, S. J. (1999) *Proc Natl Ac Sci USA 96*, 3562-7.
11. Giver, L., Gershenson, A., Freskgard, P. O., and Arnold, F. H. (1998) *Proc Natl Acad Sci USA 95*, 12809-13.
12. Zhao, H., and Arnold, F. H. (1999) *Protein Eng 12*, 47-53.
13. Arrizubieta, M. J., and Polaina, J. (2000) *J Biol Chem 275*, 28843-8.
14. Lebbink, J. H. G., Kaper, T., Bron, P., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 3656-65.
15. Merz, A., Yee, M. C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W. P., Yanofsky, C., and Kirschner, K. (2000) *Biochemistry 39*, 880-9.
16. Roovers, M., Sanchez, R., Legrain, C., and Glansdorff, N. (2001) *J Bacteriol 183*, 1101-5.
17. Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.
18. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.
19. Pouwels, J., Moracci, M., Cobuzzi-Ponzano, B., Perugino, G., Van der Oost, J., Kaper, T., Lebbink, J., De Vos, W., Ciaramella, M., and Rossi, M. (2000) *Extremophiles 4*, 157-64.
20. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.
21. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.
22. Lebbink, J. H. G. (1999) PhD Thesis pp 200, Wageningen University, Wageningen, The Netherlands.
23. Boon, M. A., Van der Oost, J., De Vos, W. M., Jansen, A. E. M., and Van 't Riet, K. (1998) *Appl Biochem Biotechnol 75*, 269-78.
24. Petzelbauer, I., Zeleny, R., Reiter, A., Kulbe, K. D., and Nidetzky, B. (2000) *Biotechnol Bioeng 69*, 140-9.
25. Petzelbauer, I., Reiter, A., Splechtna, B., Kosma, P., and Nidetzky, B. (2000) *Eur J Biochem 267*, 5055-66.

26. Fischer, L., Bromann, R., Kengen, S. W., de Vos, W. M., and Wagner, F. (1996) *Biotechnology N Y 14*, 88-91.
27. Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.
28. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2001) *Methods Enzymol 330*, 364-79.
29. Cubellis, M. V., Rozzo, C., Montecucchi, P., and Rossi, M. (1990) *Gene 94*, 89-94.
30. Lorimer, I. A., and Pastan, I. (1995) *Nucleic Acids Res 23*, 3067-8.
31. Volkov, A. A., and Arnold, F. H. (2000) *Methods Enzymol 328*, 447-56.
32. Gill, S. C., and von Hippel, P. H. (1989) *Anal Biochem 182*, 319-26.
33. Panday, N., Canac, Y., and Vasella, A. (2000) *Helv Chim Acta 83*, 58-79.
34. Guex, N., Diemand, A., and Peitsch, M. C. (1999) *Trends Biochem Sci 24*, 364-7.
35. Catanzano, F., Graziano, G., De Paola, B., and Barone, G. (1998) *Biochemistry 37*, 14484-90.
36. D'Auria, S., Barone, R., Rossi, M., Nucci, R., Barone, G., Fessas, D., Bertoli, E., and Tanfani, F. (1997) *Biochem J 323*, 833-40.
37. Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S. P., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H., and Robb, F. T. (1998) *Proc Natl Acad Sci USA 95*, 12300-5.
38. Vieille, C., and Zeikus, G. J. (2001) *Microbiol Mol Biol Rev 65*, 1-43.
39. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.
40. Kikuchi, M., Ohnishi, K., and Harayama, S. (1999) *Gene 236*, 159-67.
41. Ness, J. E., Welch, M., Giver, L., Bueno, M., Cherry, J. R., Borchert, T. V., Stemmer, W. P. C., and Minshull, J. (1999) *Nat Biotechnol 17*, 893-6.
42. Chang, C. C. J., Chen, T. T., Cox, B. W., Dawes, G. N., Stemmer, W. P. C., Punnonen, J., and Patten, P. A. (1999) *Nat Biotechnol 17*, 793-7.
43. You, L., and Arnold, F. H. (1996) *Protein Eng 9*, 315-19.
44. Fersht, A. (1999) *Structure and mechanism in protein science*, W.H. Freeman and Company, New York.
45. Wolfenden, R., Lu, X., and Young, G. (1998) *J Am Chem Soc 120*, 6814-5.
46. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.
47. Davies, J. D., Mackenzie, L., Varrot, A., Dauter, M., Brzozowoski, A. M., Schüllein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707-13.
48. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.
49. Jaenicke, R. (2000) *J Biotechnol 79*, 193-203.
50. Lebbink, J. H. G., Knapp, S., Van der Oost, J., Rice, D. W., Ladenstein, R., and De Vos, W. M. (1999) *J Mol Biol 289*, 357-69.
51. Mandelman, D., Schwarz, F. P., Li, H., and Poulos, T. L. (1998) *Protein Sci 7*, 2089-98.
52. Lopez Camacho, C., Salgado, J., Lequerica, J. L., Madarro, A., Ballestar, E., Franco, L., and Polaina, J. (1996) *Biochem J 314*, 833-8.
53. Sanz-Aparicio, J., Hermoso, J. A., Martinez-Ripoll, M., Gonzalez, B., Lopez-Camacho, C., and Polaina, J. (1998) *Proteins 33*, 567-76.
54. Spiller, B., Gersherson, A., Arnold, F. H., and Stevens, R. C. (1999) *Proc Natl Acad Sci USA 96*, 12305-10.
55. Gerherson, A., Schauerte, J. A., Giver, L., and Arnold, F. H. (2000) *Biochemistry 39*, 4658-65.
56. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis 18*, 2714-2723.
57. Cason, C. (1999), The Persisitant of Vision Development Team, Indianapolis, IN, USA.

7

Improved oligosaccharide synthesis by protein engineering of β-glucosidase from hyperthermophilic *Pyrococcus furiosus*

Authors
Therese Hansson
Thijs Kaper
John van der Oost
Willem M. de Vos
Patrick Adlercreutz

Abstract
Enzymatic transglycosylation of lactose into oligosaccharides was studied using wild-type β-glucosidase (CelB) and active site mutants thereof (M424K, F426Y, M424K/F426Y) and wild-type β-mannosidase (BmnA) of the hyperthermophilic *Pyrococcus furiosus*. The effects of the mutations on kinetics, enzyme activity, and substrate specificity were determined. The oligosaccharide synthesis was carried out in aqueous solution at 95°C at different lactose concentrations and pH-values. The results showed enhanced synthetic properties of the CelB mutant enzymes. An exchange of one phenylalanine to tyrosine (F426Y) increased the oligosaccharide yield (45%) compared with the wild-type CelB (40%). Incorporation of a positively charged group in the active site (M424K) increased the pH optimum of transglycosylation reaction of CelB. The double mutant, M424K/F426Y, showed much better transglycosylation properties at low (10-20%) lactose concentrations compared to the wild-type. At a lactose concentration of 10%, the oligosaccharide yield for the mutant was 40% compared to 18% for the wild-type. At optimal reaction conditions, a higher ratio of tetrasaccharides to trisaccharides was obtained with the double mutant (0.42, 10% lactose) compared to the wild-type (0.19, 70% lactose). At a lactose concentration as low as 10%, only trisaccharides were synthesized by CelB wild-type. The β-mannosidase BmnA from *P. furiosus* showed both β-glucosidase and β-galactosidase activity and in the transglycosylation of lactose, the maximal oligosaccharide yield of BmnA was 44%. The oligosaccharide yields obtained in this study are high compared to those reported with other transglycosylating β-glycosidases in oligosaccharide synthesis from lactose.

7.1 Introduction

The increased interest in enzymatic synthesis of oligosaccharides has heightened the need for good catalysts. Although chemical synthesis of oligosaccharides is possible, it requires many reaction steps due to the necessary selective protection of the hydroxyl groups. Enzymatic synthesis, therefore, provides an attractive alternative, since protection/deprotection sequences are avoided and the specificity of a particular enzyme used allows complete control over newly generated anomeric centers. Another advantage of enzymatic synthesis is that milder reaction conditions can be used than with chemical synthesis, which is becoming more and more important because of the environmental impact of toxic reagents.

In enzymatic synthesis of oligosaccharides, either glycosyl hydrolases (EC 3.2) or glycosyl transferases (EC 2.4.) can be used (*1, 2*). Glycosyl hydrolases, e.g. glycosidases, acting on inexpensive simple substrates such as monosaccharides and disaccharides are the most common catalysts in oligosaccharide synthesis. These enzymes do not need any cofactors and they are generally quite stable and easy to handle. In contrast, the glycosyltransferases usually need activated substrates. Those can be synthesized *in situ* using other enzymes, but there is a need for many different enzymes acting simultaneously to achieve the continuous generation of the activated substrates. This has been accomplished successfully in a number of cases, but it is considerably more laborious than the use of glycosidases.

Although glycosidases belong to the hydrolase enzymes, some of them can also transfer glycosyl residues and form a glycosidic linkage in the presence of a suitable nucleophile other than water e.g. resulting in transglycosylation activity. Transglycosylation is a kinetically controlled reaction that usually gives higher yields than reverse hydrolysis, which is under thermodynamic control (*3, 4*). In a kinetically controlled reaction in contrast to a thermodynamically controlled one, the properties of the enzyme influence the maximal yield (*4*). Hence, an optimization of the enzyme structure might contribute to its efficiency in the transglycosylation reaction.

To optimize the transglycosylation activity of glycosidases in a rational way, the influence of the active site structure on the catalytic properties of the enzyme has to be understood. One way to gain this understanding, is to use protein engineering. Protein engineering by means of site-directed mutagenesis has become a powerful method in investigations of enzymatic structure/function relationship studies in glycosyl hydrolases. For some glycosidases mutational studies have been carried out to identify the catalytic residues (acid/base and nucleophile residues) in the active site (*5*) (*6*). The influence of active site structure on the transglycosylation activity has been studied for the following glycosidases (EC 3.2) (*7*): β-glycosidase (3.2.1.21/23) (*8*), α-amylase (3.2.1.1) (*9-11*), neopullulanase (3.2.1.135) (*12*), and lysozyme (3.2.1.17) (*13*). However, for β-glycosidases the knowledge about the relation of the active site structure and their transglycosylation activity is less than for the other ones. In one study it was possible to increase the tetrasaccharide yield from *para*-nitrophenyl-β-D-cellobioside by using protein engineering on the

β-glucosidase from *Cellulomonas fimi*, but the maximal yield was still below 13% (*8*). In a few studies (*14-16*) it has been shown possible to convert β-glycosidases to "glycosyl synthases" which means that the transglycosylation product is not a substrate of the enzyme and secondary hydrolyses is therefore avoided. This is achieved by mutation of the catalytic nucleophile of the enzyme and in order to restore activity an external nucleophile such as azide is added to the reaction mixture. Using this method oligosaccharide yields (sum of tri- and tetrasaccharides) between 29-54% were obtained using α-glucosyl fluoride as the donor in the presence of different monosaccharide acceptors (*15*).

In several studies of oligosaccharide synthesis from lactose using different β-glycosidases, maximal yields of 40-42% (*17-19*) have been reported. This could indicate that these yields represent the upper limit for what is possible to achieve by wild-type β-glycosidases. The aim of this study was to increase the synthetic yield of oligosaccharides from lactose further by protein engineering of the active site of the hyperthermostable *Pyrococcus furiosus* β-glucosidase (CelB). For comparison, β-mannosidase (BmnA) from the same organism was studied as well. Lactose acts as both donor and acceptor in this type of reactions and the product mixtures contain a large number of different trisaccharides, tetrasaccharides and also small amounts of larger oligosaccharides (*20*).

In the present study we have not aimed at the characterization of all the separate products. Instead the main aim was to compare the transglycosylation activity (resulting in oligosaccharide synthesis) and the hydrolysis activity of the enzymes, and thus only the total amounts of tri- and tetrasaccharides were determined. CelB is the most thermostable and thermoactive family 1 glycosyl hydrolase described to date. The use of hyperthermostable enzymes in the conversion of lactose is of interest because at higher temperatures, higher lactose concentrations can be used, favoring oligosaccharide synthesis. Another advantage of using thermostable enzymes is that using high temperatures can prevent microbial contamination of whey, which contains large quantities of lactose and is a byproduct of cheese manufacture. Recently, wild-type CelB has been tested for its capacity to hydrolyze lactose (*21*). Additionally, the transglycosylation properties of wild-type CelB in oligosaccharide and glycoside synthesis have been studied and optimized (*22-24*)Hansson and Adlercreutz, in preparation). The yields obtained with the CelB mutants in this study are high compared to those reported with other β-glycosidases in oligosaccharide synthesis from lactose.

7.2 Materials and methods

Materials. Solvents were obtained from Merck (Darmstadt, Germany). Lactose was obtained from BDH Chemicals Ltd (Poole, England) and other chemicals were obtained from Sigma-Aldrich, Sweden AB.

Enzyme preparation and purification. The wild-type β-glucosidase CelB and variants (EC 3.2.1.21), and β-mannosidase BmnA (EC 3.2.1.25) from *P. furiosus* were purified from *Escherichia coli* BL21 (DE3) as described before (*25*). Either the enzyme solution was just heat treated to

remove *E.coli* proteins or it was purified until it was free from other proteins. Wild-type CelB and F426Y were purified by subjecting the cell-free extracts to a heat incubation of 30 min at 80øC. This yielded protein with a purity of at least 90% after pelleting the denatured *E. coli* proteins by centrifugation. CelB variants M424K and M424K/F426Y, and wild-type BmnA were purified to homogeneity by applying the heat incubated cell-free extract on a Q-sepharose ion exchange column followed by gel filtration. This yielded pure protein as judged by SDS-PAGE analysis. To prevent bacterial growth the storage buffer contained 0.02% azide.

Enzyme structure. In the studies of the protein structure of *P. furiosus* β-glucosidase CelB, a three dimensional (3D) structural model was used. The model was calculated on the basis of homology modeling and energy minimization starting from known 3D homologous structures of high identity: *Sulfolobus solfataricus* β-glycosidase LacS and *Lactococcus lactis* 6-phospho-β-D-galactosidase LacG. On the amino acid level these enzymes are 54% and 16% identical to CelB, respectively. The model was obtained from the database SWISS-MODELL Repository, a database of automatically generated protein models, with the SWISS-PROT AC code Q51723. This database is located on the ExPASy Molecular Biology Server at the Swiss Institute of Bioinformatics (SIB), Genève, Switzerland.

Enzyme activity assay. The β-glucosidase, β-galactosidase, and β-mannosidase activities of the wild-type and the mutant enzymes were determined using *para*-nitrophenyl-β-D-glucopyranoside (pNp-Glc), *para*-nitrophenyl-β-D-galactopyranoside (pNp-Gal) and *para*-nitrophenyl-β-D-mannopyranoside (pNp-man) as the substrates. The hydrolysis reaction of the different nitrophenyl substrates was followed by spectrophotometric measurements (Shimadzu, UV-120-02) at 80°C by measuring the increase of absorbance at 405 nm due to the liberation of *para*-nitrophenol. The reaction mixture contained 50 mM sodium citrate buffer (pH 5.0), 2.8 mM substrate, and 40 μl enzyme (0.50-2.0 μg) in a final volume of 1.0 ml.

One unit of enzyme activity was defined as the amount of enzyme catalyzing the liberation of 1.0 æmol of *para*-nitrophenol per minute at 80°C under these conditions, using extinction coefficients of $\varepsilon_{405;pH\ 5.0}$: 0.290×10^3 M^{-1} cm^{-1} , at 405 nm for *para*-nitrophenol.

Protein concentration assay. Total protein concentration was determined by the Bradford method (26), using bovine serum albumin (BSA) as a standard.

Kinetic constants. The kinetic parameters (K_m, V_{max}, k_{cat}/K_m) for hydrolysis of pNp-Glc for each enzyme variant were determined in a spectrophotometric assay as described in the enzyme assay procedure, using initial pNp-Glc concentrations of 0.028-6.0 mM in 50mM sodium citrate buffer (pH 5.0). The K_m and V_{max} constants for wild-type and each mutant were calculated by non-linear regression in GraphPad Prism 3.0 (GraphPad Software) using rates determined at concentrations at which no substrate inhibition occurred. The specificity constant, k_{cat}/K_m, was calculated to determine the substrate specificity of each catalyst. To calculate the catalytic constant, k_{cat}, a molecular weight of 54665 Da for the wild-type enzyme was used.

Preparation of oligosaccharides. The optimal reaction conditions in oligosaccharide synthesis for the wild-type *P. furiosus* β-glucosidase CelB have been determined by Hansson and Adlercreutz (Hansson and Adlercreutz, in preparation), and were used as a basis of comparison between different CelB mutants and CelB wild-type. The optimal reaction conditions were: pH 5.0, 95°C, 70% (wt/vol) lactose concentration, and an enzyme amount of 30 U/ml β-glucosidase activity.

Dissolved lactose at high concentration (thermostat to 95°C) was added to a reaction mixture containing the catalyst in 50 mM citrate buffer (pH 5.0) to a final volume of 2 ml to give the final lactose concentration of 0.28-2.50 M (10-90% wt/vol). The lactose concentrations given in the study are expressed as weight percent lactose of the final volume (% wt/vol). The reaction mixture, stirred vigorously, was incubated in glass vials with tightly screwed teflon caps, in an oven, at 95°C. At regular time intervals of between 15 min and 24 hours, samples of 10 æl were removed and the reaction stopped by immersing the vials in ice for 20 minutes. The samples were diluted with cold (4°C) water before being injected into the HPLC. The oligosaccharide synthesis capacity of the enzymes was quantified in terms of maximum yield of oligosaccharides (sum of tri- and tetrasaccharides), which is expressed as percent by weight of product mixture (% wt/wt). In the studies of the influence of pH on oligosaccharide synthesis the buffers used were: pH 4.0-6.0, 50 mM citrate buffer; pH 7.0, 50 mM citrate-phosphate buffer.

Analytical method. The oligosaccharides were analyzed by HPLC with a Nucleosil NH_2 column. The HPLC system used was as described previously (Hansson and Adlercreutz, in preparation).

7.3 Results and discussion

The synthetic capacity of CelB active site mutants M424K, F426Y and M424K/F426Y was evaluated and compared to that of wild-type β-glucosidase CelB and wild-type β-mannosidase BmnA, both from *P. furiosus*. The M424 and F426 residues in CelB are conserved residues and are located ~15Å and ~10Å, respectively, from the catalytic acid/base Glu207. The catalytic glutamates (Glu207 and Glu372) have a distance of ~5Å between the oxygen atoms of the carboxylic groups.
Initially the M424K, F426Y and M424K/F426Y CelB variants were designed in an attempt to alter the substrate specificity of the enzyme to adapt it for the hydrolysis of phosphorylated sugars (25). The residues corresponding to M424 and F426 in 6-phospho-β-galactosidase LacG from *L. lactis*, K435 and Y437 respectively (Figure 7.1), were found to interact with the phosphate group of a galactose-6-phosphate molecule, bound in the active site (27). In addition, Y437 was found to have hydrogen bonding with the oxygen at carbon 6 of the galactose residue in the substrate.

Influence of the mutations on substrate specificity and enzymatic activity. Family 1 glycosidases in general shows broad substrate specificities. CelB wild-type hydrolyses β-glucosides most efficiently (100%), but has significant β-galactosidase activity (61%) and some β-manno-

```
                                                        207                        % identity
Pfu CelB    WLDEKTVVEFVKFAAFVAYHLDDLVDMWSTMNEPNVVYNQGYINLRSGFPPGY-LSFEAA 233      100
Sso LacS    WLSTRTVYEFARFSAYIAWKFDDLVDEYSTMNEPNVVGGLGYVGVKSGFPPGY-LSFELS 232       54
Pfu BmnA    WVNPRTVIEFAKYAAYIAYKFGDIVDMWSTFNEPMVVVELGYLAPYSGFPPGV-LNPEAA 236       47
Lla LacG    FLNRENIEHFIDYAAFCFEEFP-EVNYWTTFNEIGPIGDGQYLVG--KFPPGIKYDLAKV 185       16


                                   372
Pfu CelB    NLLKYLNNAYE--LPMIITENGMADA--------ADRYRPHYLVSHLKAVYNAMKEGADV 404
Sso LacS    DVLTKYWNRYH--LYMYVTKNGIADD--------ADYQRPYYLVSHVYQVHRAINSGADV 419
Pfu BmnA    DSIVEAH-KYG--VPVYVTKNGIADS--------KDILRPYYIASHIKMIEKAFEDGYEV 446
Lla LacG    DQIMRVKNDYPNYKKIYITENGLGYKDEFVDNTVYDDGRIDYVKQHLEVLSDAIADGANV 415


                              424 426
Pfu CelB    RGYLHWSLTDNYEWAQGFRSREGLVYVDFETKKRYLRP-SALVFREIATQKEIPEELAHL 463
Sso LacS    RGYLHWSLADNYEWASGFSREGLLKVDYNTKRLYWRP-SALVYREIATNGAITDEIEHL 478
Pfu BmnA    KGYFHWALTDNFEWALGFRREGLYEVNLITKERIPREKSVSIFREIVANNGVTKKIEEE 506
Lla LacG    KGYFIWSLMDVFSWSNGYERHGLFYVDFDTQERYPKK-SAHWYKKLAETQVIE------ 468
```

Figure 7.1 Sequence alignment of *Pfu* CelB, *Pyrococcus furiosus* β-glucosidase CelB (TrEmBL Q51723), with other enzymes with high identity on the amino acid level. Sequences were compared with the following enzymes / accession numbers: *Sso* LacS, *Sulfolobus solfataricus* β-galactosidase LacS (Swiss Prot P22498); *Pfu* BmnA, *Pyrococcus furiosus* β-mannosidase BmnA (TrEMBL Q51733); *Lla* LacG, *Lactococcus lactis* 6-phospho-β-galactosidase LacG (P11546). The amino acids shaded grey are the catalytic acid/base (*Pfu* CelB Glu207) and nucleophile (*Pfu* CelB Glu372) in all enzymes. The amino acids shaded black are the residues corresponding to the mutation positions in the different enzymes.

sidase activity (9.5%). The mutants had lower activities (Table 7.1) and substrate specificities similar to that of the wild-type, except for the M424K mutant, which expressed a higher ratio of β-galactosidase/β-glucosidase activity (0.77) and lower β-mannosidase activity (7.6%) than the wild-type. BmnA showed the highest activity on β-mannosides (139%), as expected, and was found capable of hydrolyzing β-glucosides (100%) and β-galactosides (59%) as well.

The enzyme kinetics (K_m, V_{max} and k_{cat}/K_m) of wild-type and mutant enzymes for the hydrolysis of pNp-Glc were studied in more detail (Table 7.1). In all cases, except for the β-mannosidase, some substrate inhibition could be seen at substrate concentrations higher than 3 mM. For M424K and M424K/F426Y this inhibition was larger than for the other ones. The kinetic constants were determined at substrate concentrations low enough to avoid the interference from substrate inhibition. The substitutions in the β-glucosidase CelB resulted in a decrease in the V_{max} values (Table 7.1). For the F426Y and the M424K mutants 65% and 61% activity remained, respectively. The substitution F426Y caused a decrease of 28% in the specificity constant (Table 7.1). A mutation of methionine to lysine in position 424 (M424K) increased the specificity constant with 56% due to a decrease in the K_m value. As expected, the β-mannosidase BmnA had a significantly lower catalytic efficiency for the hydrolysis of β-glucosidic bonds than the β-glucosidase CelB (Table 7.1).

In the oligosaccharide synthesis experiments (see below) the enzyme amount was standardized with respect to the β-glucosidase activity since this was the main activity of most of the enzymes. Since most enzymes have a β-galactosidase activity around 60% of the β-glucosidase activity, the β-galactosidase activity was almost the same in all those experiments as well. Due to lower specific activity of the mutants in comparison to the wild-type (Table 7.1), higher amounts (mg) of mutant enzyme were used to maintain equal amounts of β-glucosidase activity (U/ml) in the reactions. The amount of enzyme did not influence the maximum oligosaccharide yield, which agrees with earlier results (Hansson and Adlercreutz, in preparation).

Oligosaccharide synthesis. The maximum yield of oligosaccharides (sum of tri- and tetrasaccharides) from lactose was determined to investigate the transglycosylation activity of the wild-type and active site mutants of CelB and the wild-type BmnA. All the catalysts synthesized pentasaccharides, but the amounts were too small to be quantified. The time course of oligosaccharide formation for CelB wild-type at optimized reaction conditions is seen in Figure 7.2. The maximum was reached within 5 h. After the maximum was reached the synthesized oligosaccharides were hydrolyzed. A maximal oligosaccharide yield of 40% has been reached for this enzyme (Hansson and Adlercreutz, in preparation).

Mutational effects on oligosaccharide synthesis by β-glucosidase CelB of Pyrococcus furiosus (M424K, F426Y, M424K/F426Y). At optimal reaction conditions for the wild-type enzyme, the replacement of a methionine by a lysine at position 424 (M424K) resulted in a changed curve for maximal oligosaccharide yield as a function of lactose concentration (Figure 3A). The maximal synthetic yields by the M424K mutant were almost unaffected by the substrate concentration, but were lower than those for the wild-type with exception of the lowest lactose concentrations.

Increased maximal oligosaccharide yield (from 40 to 45%) was achieved by replacing the phenylalanine residue at position 426 by a tyrosine residue (F426Y) at optimized reaction conditions (Figure 2 and Table 7.2). The F426Y mutant produced higher maximum oligosaccharide yield than the wild-type enzyme in the range of lactose concentrations between 10-90% (Figure 7.3B).

Table 7.1 Kinetic parameters for wild-type (wt) and mutants of β-glucosidase CelB and wild-type β-mannosidase BmnA of Pyrococcus furiosus. Reaction conditions: 0.028-6 mM para-nitrophenyl-β-D-glucopyranoside, 50 mM citrate buffer (pH 5.0), 0.50-2.0 æg enzyme, 80°C.

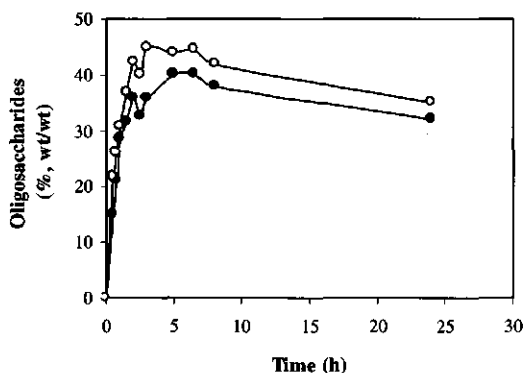| Enzyme | Kinetic parameters | | |
	K_m (mM)	V_{max} (U/mg)	k_{cat} / K_m (mM^{-1} s^{-1})
CelB wt	0.41 ± 0.03	833 ± 18	1958
CelB M424K	0.16 ± 0.02	507 ± 13	3051
CelB F426Y	0.37 ± 0.02	539 ± 10	1409
CelB M424K/F426Y	0.61 ± 0.04	227 ± 6	358
BmnA wt	0.92 ± 0.32	292 ± 35	306

Figure 7.2 Oligosaccharide yield as a function of time by the wild-type (●) and by the F426Y mutant (O) of *Pyrococcus furiosus* β-glucosidase CelB. Reaction conditions: 30 U/ml glucosidase activity, 95°C, 70% (wt/vol) lactose concentration and pH 5.0.

The maximum oligosaccharide yield for both the wild-type and the mutant increased with increasing lactose concentration up to 70% lactose. No further increase could be seen at higher lactose concentrations. This was probably due to substrate inhibition and/or inhibition by the large amounts of Maillard components (brown coloring) formed at high lactose concentration and high temperature (95°C). A higher yield in tetrasaccharide synthesis accounted mostly for the mutant's increased oligosaccharide yield at high lactose concentrations (70-90%). At 70% lactose the amount of trisaccharides was almost the same for the wild-type CelB and the mutant F426Y, 33% and 35%, respectively, but the tetrasaccharide yield was higher for the mutant (6% and 10%, respectively). At low lactose concentrations (10-20%) the higher yield by the mutant was seen in an increased trisaccharide yield.

For the M424K/F426Y double mutant the maximal oligosaccharide yield was influenced inversely by the substrate concentration compared to the wild-type (Figure 7.3C). Below lactose concentrations of 50% the mutant showed higher oligosaccharide yields than the wild-type. At a lactose concentration of 10%, the oligosaccharide yield for the mutant was 40% compared to 18% for the wild-type. The effects of the two single mutations: the small effect of the substrate concentration on the maximum yield observed for the single M424K mutation and the increased maximum yield caused by the single F426Y mutation, seem to be combined in the M424K/F426Y double mutant (Figure 7.3). The very good synthetic properties of the double mutant at low substrate concentrations can be of significant interest in an industrial process for oligosaccharides synthesis. Lactose is only present at a concentration of approximately 5% in milk or milk derived side products such as whey. It would be economically favorable if those relatively dilute solutions could be used without the need for concentration steps. Furthermore, high temperature reactions at low lactose concentrations are favorable because of the low extent of Maillard reactions.

The F426Y and M424K/F426Y mutants showed the same pH optima as the wild-type enzyme in oligosaccharide synthesis (pH 5.0). For the M424K mutant, with an extra positively charged amino acid in the active site, the pH optimum increased from 5.0 to 6.0. In Table 7.2, oligosaccharide yield can be seen for each mutant at optimal substrate concentration and pH. The same maximum yield as for the wild-type CelB enzyme (40%) was reached for the M424K and M424K/F426Y mutants. For the F426Y mutant, the maximum yield increased from 40 to 45% oligosaccharides.

Oligosaccharide synthesis using two different glycosyl hydrolase enzymes from Pyrococcus furiosus, CelB and BmnA.. The β-mannosidase BmnA and the β-glucosidase CelB were compared with respect to their conversion of lactose into oligosaccharides. The enzymes are closely related (46.5% identity of the amino acid sequences) (*28*). Differences in substrate specificity have been discussed above.

The curve of oligosaccharide yield as a function of lactose concentration by CelB and BmnA at the same reaction conditions and initial β-glucosidase activity showed almost the same shapes (Figure 7.3D). BmnA resulted in the same oligosaccharide yield (40%) at the optimal conditions for CelB and the same ratio between tri- and tetrasaccharide (Table 7.2). Both enzymes showed the same yields at high substrate concentrations, but at low substrate concentrations the oligosaccharide yields for BmnA were slightly higher. BmnA showed an optimal pH of 5.5 in oligosaccharide synthesis, and was more pH dependent than CelB.

Even if BmnA is not preferred as a catalyst in lactose conversion due to low specific β-glucosidase and β-galactosidase activities, these investigations show that if a high enzyme amount is used, the β-mannosidase BmnA of *P. furiosus* acts as a good transferase in the transglycosylation of lactose. These results also show that enzymes that have been differently classified according to their most prominent hydrolytic activities, can be used to produce the same kind of product.

Table 7.2 Maximum oligosaccharide yield at optimal reaction conditions for the wild-type (wt) β-glucosidase CelB enzyme (pH 5.0, 70% (wt/vol) lactose) and at the optimal reaction conditions for each catalyst (pH 4.0-7.0, 10-90% (wt/vol) lactose concentration). In all experiments 30 U/ml β-glucosidase activity and 95°C were used.

Enzyme	pH	Lactose (%wt/vol)	Maximum yield (%,wt/wt)[i]			[Tet] [Tri]
			% GO	% Tri	% Tet	
CelB wt	5.0	70	39.5 ± 1.1	33.2 ± 1.4	6.3 ± 0.4	0.19
CelB M424K	5.0	70	34.5 ± 1.1	29.2 ± 1.0	5.3 ± 0.1	0.18
CelB M424K	6.0	70	40.6 ± 0.8	33.7 ± 0.4	6.9 ± 0.5	0.20
CelB F426Y	5.0	70	44.6 ± 0.4	35.1 ± 0.7	9.5 ± 0.3	0.27
CelB M424K/F426Y	5.0	70	36.5 ± 1.5	31.8 ± 1.0	4.7 ± 0.5	0.15
CelB M424K/F426Y	5.0	10	40.1 ± 1.9	28.3 ± 1.0	11.8 ± 0.9	0.42
BmnA wt	5.0	70	40.4	33.9	6.5	0.19
BmnA wt	5.5	70	44.4	36.5	7.9	0.22

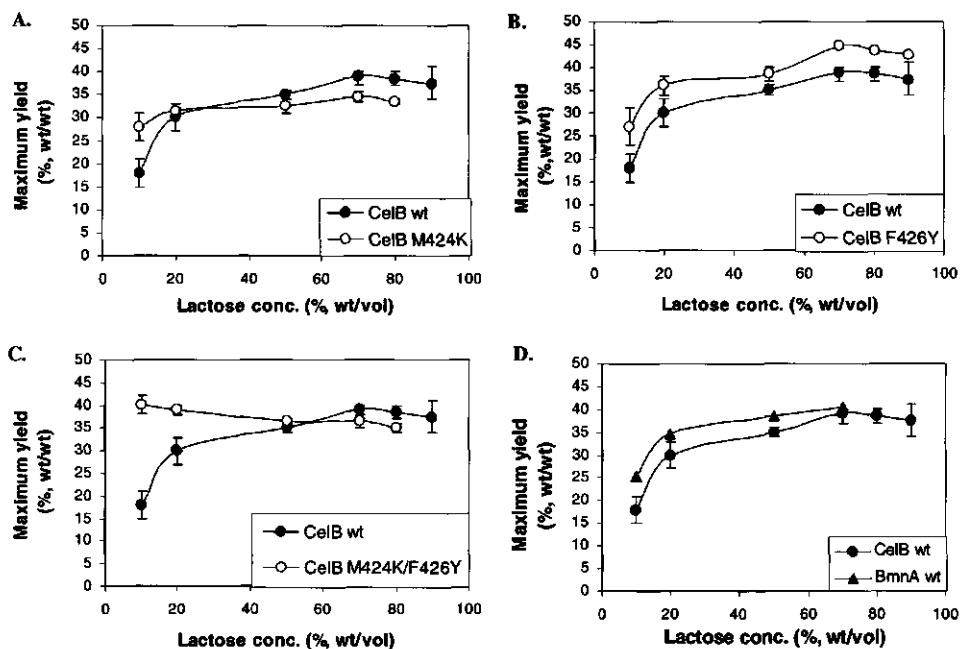[i] GO = Oligosaccharides, Tri = Trisaccharides, Tet = Tetrasaccharides

Figure 7.3 Maximal oligosaccharide yield as a function of lactose concentration. Comparison between (A) wild-type (wt) β-glucosidase CelB and CelB M424K (B) wild-type β-glucosidase CelB and CelB F426Y (yields at: 70% lactose, 44.6% ñ 0.4; 80% lactose, 43.5% ñ 0.5; 90% lactose, 42.7% ñ 0.2) (C) wild-type β-glucosidase CelB and CelB M424K/F426Y double mutant (D) wild-type β-glucosidase CelB and wild-type β-mannosidase BmnA. All from *Pyrococcus furiosus*. Reaction conditions: 30 U/ml β-glucosidase activity, 95°C and pH 5.0.

Background of the increased oligosaccharide yields. The increase in maximal oligosaccharide yield observed in this study can be due either to an increase in the transglycosylation/hydrolysis ratio of the mutant enzyme or to a lower rate of the breakdown of the oligosaccharide products compared to the wild-type enzyme. For the F426Y mutant, Figure 7.2 indicates that during the phase of reaction when the breakdown of oligosaccharides dominates (the curves after the maxima), the reactions occur at about the same rate for the mutant and the wild-type. The same was observed for the M424K/F426Y mutant at low substrate concentrations (data not shown). This suggests that the increased oligosaccharide yield by both these mutants (Figure 7.3) was due to an increase in the transglycosylation/hydrolysis ratio. This means that the mutations favor the reaction of the glycosyl-enzyme with carbohydrate acceptors (for example lactose) more than the hydrolysis of the glycosyl-enzyme. In hexyl-glucoside synthesis from pentyl-glucoside in hexanol, a 2.6 and 2.3 fold increase in the transglycosylation/hydrolysis ratio has been seen for CelB F426Y and CelB M424K/F426Y, respectively, compared to CelB wild-type (Hansson and Adlercreutz, unpublished results).

There is no simple way to model the interaction between enzyme and oligosaccharide, and to understand the influence of different amino acids in the enzymatic structure on product yield. Parameters influencing the formation of enzyme-carbohydrate complexes can be: the flexibility of the carbohydrates (conformational changes) and the numerous ways they can interact with the protein surface with different stabilizing interactions (such as hydrogen bonds, van der Waal contacts, nonpolar interactions etc.) (29-31).

It is very difficult to rationally design mutations that enhance the transglycosylation activity of these enzymes. A different approach would be the generation of a large number of randomly mutated CelB variants, followed by a selective assay for transglycosylating properties. Although CelB was randomly mutated and screened for optimal hydrolytic activity at room temperature (32), a transglycosylating assay for selection of optimized transglycosylation properties of CelB has not been described so far. The information available from this study is not sufficient to explain the increased oligosaccharide yields. Further studies with other substrates and different enzyme variants are currently carried out.

7.4 Conclusions
In this study we report on improvement from 40 to 45% of the oligosaccharide yield by a single mutation (F426Y) in the β-glucosidase CelB. Moreover, the 40% oligosaccharide yield obtained with the CelB double mutant (M424K/F426Y) at a lactose concentration of only 10% is promising for practical applications. At this lactose concentration, *Aspergillus oryzae* β-galactosidase and *Caldocellum saccharolyticum* β-glycosidase provided oligosaccharide yields of 10% and 7%, respectively (19). For both mutants, the results indicate that the improved yield is due to an increase in the ratio of transglycosylation to hydrolysis. This ratio is determined by the reaction kinetics and additional protein engineering of this or other glycosidases might increase it further.

7.5 Acknowledgements

References
1. Bucke, C. (1996) *J Chem Tech Biotechnol 67*, 217-20.
2. Crout, D. H. G., and Vic, G. (1998) *Curr Opin Chem Biol 2*, 98-111.
3. Faber, K. (1997) in *Biotransformations in organic chemistry* (Faber, K., Ed.) pp 275-87, Springer-Verlag, Berlin, Germany.
4. Kasche, V. (1986) *Enzyme Microb Tech 8*, 4-16.
5. Fujimoto, Z., Takase, K., Doui, N., Momma, M., Matsumoto, T., and Mizuno, H. (1998) *J Mol Biol 277*, 393-407.
6. Mackenzie, L. F., Brooke, G. S., Cutfield, J. F., Sullivan, P. A., and Withers, S. G. (1997) *J Biol Chem 272*, 3161-7.
7. Ly, H. D., and Withers, S. G. (1999) *Annu. Rev. Biochem. 68*, 487-522.
8. Nikolova, P. V., Duff, S., MacLeod, A., and Haynes, C. A. (1996) *Ann NY Acad Sci 799*, 19-25.
9. Matsui, I., Ishikawa, K., Miyairi, S., Fukui, S., and Honda, K. (1991) *Biochim Biophys Acta 1077*, 416-19.

10. Matsui, I., Yoneda, S., Ishikawa, K., Miyairi, S., Fukui, S., Umeyama, H., and Honda, K. (1994) *Biochemistry 33*, 451-8.
11. Saab-Rincon, G., del-Rio, G., Santamaria, R. I., Lopez-Munguia, A., and Soberon, X. (1999) *FEBS Lett 453*, 100-6.
12. Kuriki, T., Kaneko, H., Yanase, M., Takata H, S., J., Handa, S., Takada, T., Umeyama, H., and Okada, S. (1996) *J Biol Chem 271*, 17321-9.
13. Kuroki, R., Weaver, L. H., and Matthews, B. W. (1999) *Proc Natl Ac Sci USA 96*, 8949-54.
14. Malet, C., and Planas, A. (1998) *FEBS Letters 440*, 208-12.
15. Mackenzie, L. F., Wang, Q., Warren, R. A. J., and Withers, S. G. (1998) *J Am Chem Soc 120*, 5583-4.
16. Moracci, M., Trincone, A., Perugino, G., Ciaramella, M., and Rossi, M. (1998) *Biochemistry 37*, 17262-70.
17. Nakao, M., Harada, M., Kodama, Y., Nakayama, T., Shibano, Y., and Amachi, T. (1994) *Appl Microbiol Biotechnol 40*, 657-63.
18. Onishi, N., and Tanaka, T. (1995) *Appl Environ Microbiol 61*, 4026-30.
19. Stevenson, D. E., Stanley, A. R., and Furneaux, H. (1996) *Enzyme Microb Tech 18*, 544-9.
20. Toba, T., and Yokota, A. (1985) *Food Chem 16*, 147-56.
21. Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.
22. Boon, M. A., Van der Oost, J., De Vos, W. M., Jansen, A. E. M., and Van 't Riet, K. (1998) *Appl Biochem Biotechnol 75*, 269-78.
23. Fischer, L., Bromann, R., Kengen, S. W., de Vos, W. M., and Wagner, F. (1996) *Biotechnology N Y 14*, 88-91.
24. Petzelbauer, I., Zeleny, R., Reiter, A., Kulbe, K. D., and Nidetzky, B. (2000) *Biotechnol Bioeng 69*, 140-9.
25. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.
26. Bradford, M. M. (1976) *Anal Biochem 72*.
27. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.
28. Bauer, M. W., Bylina, E. J., Swanson, R. V., and Kelly, R. M. (1996) *J Biol Chem 271*, 23749-55.
29. Imberty, A., Bourne, Y., Cambillau, C., Roug, P., and Perez, S. (1993) *Adv Biophys Chem 3*, 71-117.
30. Johnson, L. N., Cheetham, J., McLaughlin, P. J., Acharya, K. R., Barford, D., and Phillips, D. C. (1988) *Curr Top Microbiol Immunol 139*, 81-134.
31. Quiocho, F. A. (1989) *Pure & Appl Chem 61*, 1293-1306.
32. Lebbink, J. H. G., Kaper, T., Bron, P., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 3656-65.

8

General discussion and concluding remarks

This chapter aims to give a comprehensive discussion of the most prominent results obtained in the studies described in the previous chapters. Results of the individual studies will be compared to each other, to other studies in the field, and to recent new data, which have not been included in previous chapters. First, an introduction of the properties of the *Pyrococcus furiosus* β-glucosidase CelB is given, including some of the methodologies used for its study and an overview of the distribution of family 1 glycosyl hydrolases in Thermococcales. In addition, the active site structure in family 1 enzymes will be discussed, followed by some comments on the reaction mechanism for hydrolysis of saccharides by these enzymes. Finally, the thermostability and thermoactivity of *P. furiosus* CelB are discussed together with an evaluation of the protein engineering approaches concluding with potential applications of thermostable β-glycosidases.

8.1 The family 1 β-glucosidase CelB of *P. furiosus*

Initially isolated from cell-free extracts of cellobiose-grown *P. furiosus* cells, the β-glucosidase CelB is the most thermostable and thermoactive family 1 glycosyl hydrolase described to date (*1, 2*). The tetrameric enzyme displays optimal activity at 100-102°C and is most specific for hydrolysis of the residue at the non-reducing ends of β-1,3 and 1,4-glucosides, but has significant β-fucosidase, β-galactosidase and β-xylosidase activity as well (*1, 3*). The β-glucosidase is able to release glucose residues from the β-1,3-glucose polymer laminarin, but not β-1,4-cellulose (*1*). Through the synergistic action of β-glucosidase CelB and the endo-β-1,3-glucanase LamA of *P. furiosus*, laminarin is efficiently degraded to monomeric glucose (*4*). The retaining mechanism by which β-glucosidase CelB hydrolyzes its substrates is the same as that in family 1 β-glucosidases from mesophiles (*3*). When the proposed catalytic glutamate that serves as a nucleophile was replaced by glutamine, hydrolytic activity was abolished, similar as has been found for family 1 glycosidases from mesophilic origin (*2*). The β-glucosidase is extremely resistant to thermal inactivation and shows a half-life of activity of 85 h at 100°C (*1*). This thermostability is believed to be principally the result of optimized hydrophobic interactions in CelB (Chapter 4). The *celB* gene was identified by a reverse genetics approach and cloned into the mesophilic host *E. coli*, where functional CelB made up over 20% of the total cell protein (*2*). The heterologously produced β-glucosidase CelB had the same biochemical properties as the wild-type enzyme isolated from *P. furiosus* cells.

8.2 *P. furiosus* β-glucosidase CelB in engineering studies

Enzyme studies in general benefit from a reliable and efficient expression system and enzyme purification protocol. In addition, site-directed mutagenesis studies require structural information son the target enzyme, preferably a high-resolution 3D structure (Chapter 1). On the other hand, directed evolution studies rely on, besides a suitable expression system, the generation of large amounts of genetic variants of the target gene, and efficient and reliable screening methods to select for the desired enzyme property (see Chapter 1). Each of these requirements has been optimized for the β-glucosidase CelB (Chapters 2, 5 & 6) (*5*).

The initial *E. coli*/pTTQ19 expression system for the production of β-glucosidase CelB, which carried the *celB* gene on a genomic DNA fragment (*2*), suffered from expressional instability in larger culture volumes. Hence, the large-scale production of β-glucosidase CelB in *E. coli* was optimized using a T7-derived expression system (*6*). Expression of the *celB* gene was placed under control of the T7-promoter located on the pET9d vector (Novagen) that carries the gene for kanamycin-resistancy (*5*). Using this system up to 90 mg of pure β-glucosdase CelB could be obtained per liter *E. coli* cell culture (*5*). This expression system has been used for the production of wild-type and mutated CelB in culture volumes ranging from microliters in microplates (Chapters 5 and 6) to liters (Chapters 2, 3, 4, 5, 6 & 7) in the described studies.

In Chapter 2, the elucidation of the 3D structure of *P. furiosus* β–glucosidase CelB to 3.3 Å resolution is described. This structure has been instrumental in the design of the site-directed CelB mutants (Chapters 2 & 3) as well as interpretation of CelB variants obtained in the laboratory evolution studies (Chapter 5) and the construction of hybrid enzyme 3D models (Chapter 6). The structure was established by the molecular replacement method, in which the 3D structures of the β-glycosidase LacS from *Sulfolobus solfataricus* (*7*) and 6-phospho-β-galactosidase LacG from *Lactococcus lactis* (*8*) served as search models. Due to the high structural homology between CelB, LacS and LacG, the active site structure of CelB could be established with greater precision than the rest of the protein. The β-glucosidase CelB has many characterized orthologs with known amino acid sequence structures in other organisms ranging from mesophiles to hyperthermophiles with several solved 3D structures. The CelB 3D structure has been valuable in comparison studies and interpretation of biochemical differences between orthologs (Chapters 2, 3, 5 & 6).

The evolution studies described in Chapters 5 and 6 were aimed at the optimization of catalytic features under suboptimal reaction conditions. The screening methods were therefore designed to detect activity differences between wild-type and mutant enzymes. Initially, activity screenings were performed with whole cells on agar plates at 37°C, which contained chromogenic substrates that released color after hydrolysis of the glycosidic bond (*5*). However, this screening approach was found to be less reliable than that based on cell-free extracts that were produced by treatment of cultures grown in microplates with chloroform-DNAse or freezing-thawing, as described in Chapters 5 and 6. The use of chromogenic substrates in the screening facilitated activity measurements by analyzing the amount of released chromophor in a microplate reader after termination of the hydrolysis reaction (Chapter 5). However, when lactose was used as a screening substrate, the introduction of an extra step was required in which the released glucose was determined enzymatically (Chapter 6).

The above described aspects of the studies on the β-glucosidase CelB provided the necessary framework for the engineering of CelB by site-directed mutagenesis or laboratory evolution approaches.

8.3 Family 1 β-glycosidases in Thermococcales

The currently known genes coding for family 1 β-glycosidases in Thermococcales can be divided in four groups of orthologous genes and have been identified and isolated by reverse genetics, use of expression libraries or analysis of pyrococcal and other genome sequences (*9*) (Table 8.1). Remarkably, the genes encoding these β-glycosidases are not present in the completely sequenced genomes of *P. furiosus*, *P. abyssi* and *P. horikoshii*. *P. furiosus* contains four family 1 enzymes. All of these enzymes have been characterized or have closely related characterized orthologs and were found to have a specific substrate specificity and activity (Chapter 3) (*1, 10, 11*) (Table 8.1). The gene coding for the first described β-glucosidase CelB in *P. furiosus* was initially isolated by

Table 8.1 Identified family 1 β-glycosidases homologues in Thermococcales (modified after ref. (*9*)) and hydrolytic activity of representative enzymes at 90°C. Confirmed presence (+), absence of genes (-) has been indicated. Not determined (ND).

Organism		CelB (*1*)[1]	BmnA (*10*)[2]	BglA (*11*)[3]	BglB (*65*)[4]
P. furiosus		+	+	+	+
P. abyssi		-	+	-	+
P. horikoshii		-	-	+	+
T. kodakaraensis KOD1		ND	ND	ND	+
Substrate					
cellobiose	K_m (mM)	720	260	1698	ND
	k_{cat} (s^{-1})	14	100	194	
	k_{cat}/K_m (s^{-1}mM^{-1})	48	2.6	0.11	
lactose	K_m (mM)	1500	101	ND	ND
	k_{cat} (s^{-1})	120	89		
	k_{cat}/K_m (s^{-1}mM^{-1})	11	1.1		
pNp-Glc	K_m (mM)	1800	486	0.35	13.5
	k_{cat} (s^{-1})	0.42	69	79	34.4
	k_{cat}/K_m (s^{-1}mM^{-1})	3900	6.9	226	2.6
pNp-Gal	K_m (mM)	2600	496	1.3	7.3
	k_{cat} (s^{-1})	5.0	29	123	13.5
	k_{cat}/K_m (s^{-1}mM^{-1})	480	16.8	94	1.9
pNp-Man	K_m (mM)	72.4	190	0.14	0.44
	k_{cat} (s^{-1})	1.3	0.99	2	4.2
	k_{cat}/K_m (s^{-1}mM^{-1})	54.7	198	15	9.8
n-d-Glc	K_m (mM)	ND	ND	0.03	ND
	k_{cat} (s^{-1})			36	
	k_{cat}/K_m (s^{-1}mM^{-1})			1153	

[1] kinetic data were taken from Lebbink *et al.* (*66*) and Kaper *et al.* (*65*)

[2] kinetic data were taken from Kaper *et al.* (*9*)

[3] kinetic data were taken from Matsui *et al.* (*11*)

[4] kinetic data were taken from Kaper *et al.* (*65*)

reverse genetics (*2*). CelB is likely to have a catabolic role in the high turnover of β-glucosides, since its gene expression in *P. furiosus* is induced by the presence of β-glucosides in the medium (*1, 12*). CelB has only been identified in *P. furiosus* (Table 8.1). This might be related to the habitat of the organism, since the *celB* gene is located in an operon together with the endo-β-1,3-glucanase LamA. This endo-glucanase degrades β-1,3-glucose polymer laminarin, a cell-wall component of red algae, to oligosaccharides which in turn can be degraded by *P. furiosus* (*4*). *P. furiosus* was

isolated from marine sediments at a beach on Vulcano Island, Italy (*13*), where algae can be expected to be present. In contrast, both *P. abyssi* and *P. horikoshii* were isolated from abyssal hydrothermal vents (*14, 15*), well out of range for photo-synthetic life, which might explain the absence of CelB homologues in these organisms. In support of this view is the finding that the hyperthermophilic Crenarchaeota *S. solfataricus* and *Thermosphaera aggregans* contain a CelB homologue and were both isolated from shallow depth (*16, 17*). Other CelB homologues might be present in *Pyrococcus* species, capable of saccharolytic growth, such as *P. glycovorans* (*18*). The gene coding for *P. furiosus* β-mannosidase BmnA was isolated from an expression library (*10*). A BmnA homologue has been found in *P. abyssi*, but not in *P. horikoshii*, which suggests that its activity is not crucial for the pyrococcal metabolism. This can not be easily reconciled with the proposed involvement of BmnA in compatible solute metabolism (*10*). The β-glucosidase BglA, the gene of which has been identified in *P. horikoshii* and *P. furiosus* genome sequences, has the highest specificity for alkyl glucosides with an implied function in glycosylation of cell components (*11*). However, a *bglA* gene is absent in *P. abyssi*. In conclusion, it is likely that CelB, BmnA and BglA can be considered as 'lifestyle' enzymes (*19*).

On the other hand, BglB homologues have been identified in all three sequenced *Pyrococcus* genomes, as described in Chapter 3. Further evidence of the presence of *bglB* homologues in Thermococcales resulted from the analysis of genomic DNA by PCR. In this approach to identify genes of family 1 glycosyl hydrolases, genomic DNA of Thermococcales species displaying β-glucosidase activity from the IFREMER collection (Brest, France) was screened for the presence of genes coding for β-glucosidase CelB homologues in a PCR based approach. Degenerated oligonucleotides were designed against conserved protein sequences in the tetrameric family 1 β-glycosidases from *P. furiosus* (CelB, BmnA) and *S. solfataricus* (LacS). The retrieved sequences mainly showed the typical active site signature of BglB and clustered with BglB in a phylogenetic tree based on the partial amino acid sequences (Figure 8.1). Due to the low hydrolytic activity of their characterized homologues, no attempts for further isolation of complete genes were undertaken. These findings, however, implied a basic functional role for BglB enzymes in the Thermococcal metabolism, which could possibly be the involvement in the degradation of compatible solutes, which would make it a 'household' enzyme (*19*). The specificity of BglB for mannosides agrees with the presence of the compatible solute mannosyl-β-glycerate in *P. furiosus* (*20*). If BglB is involved in compatible solute degradation, *bglB* transcription might be induced by a sudden decrease in salt concentration in the medium. However, Northern blot analysis did not show any *bglB* transcript in growing cells of *P. furiosus*, even when the salt concentration was lowered by dilution (data not shown). Alternatively, the expression of the *bglB* gene might be temperature-regulated.
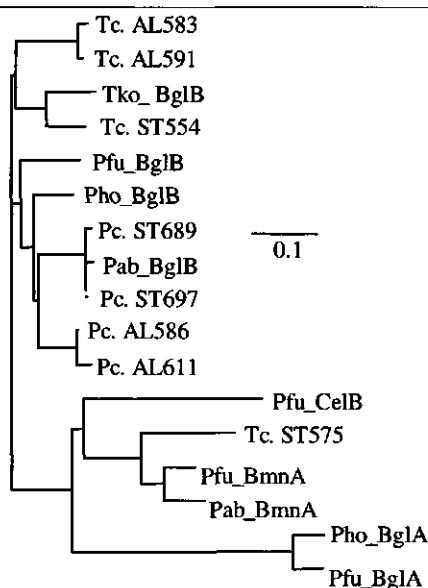
Figure 8.1 Phylogenetic tree of partial amino acid sequences of family 1 β-glycosidases from Thermococcales. Pc: *Pyrococcus*, Tc: *Thermococcus*, Pho_BglB: *P. horikoshii* β-mannosidase (AP000002), Pab_BglB: *P. abyssi* putative β-mannosidase BglB (AJ248288), Pfu_BglB: *P. furiosus* putative β-mannosidase BglB (*P. furiosus* genome ORF Pf_368506), Tko_BglB: *Thermococcus kodakaraensis* putative β-mannosidase BglB (AB028601), Pfu_BmnA: *P. furiosus* β-mannosidase BmnA (U60214), Pab_BmnA: *P. abyssi* putative β-mannosidase BmnA (AJ248285), Pfu_CelB: *P. furiosus* β-glucosidase CelB (AF013169). Alignment and tree were generated in ClustalX (*68*) and the tree was visualized using Treeview (*69*). DNA sequences were amplified by PCR on purified genomic DNA using primerS BG262 (GTNAGYGGNGAYYTNCCNGA) and BG265 (TARTARTTNACNCC DATCCARTC). Reactions containing ~25 ng DNA, 0.5 μg of each primer, 1 U *Taq* DNA polymerase (Pharmacia) in the supplied buffer with a final volume of 50 μl were subjected to a denaturation step of 5 min at 94°C. This was followed by 35 cycles of 45 sec at 94°C, 30 sec at 35°C, and 1 min at 72°C. After the last cycle, the reactions were incubated for 5 min at 72°C. Nucleotide sequences were determined as described (*70*).

8.4 The active site structure of 1 β-glycosidases

The active sites of β-glucosidase CelB and those of related family 1 enzymes have been intensively studied in this thesis. In each case, the results of the studies indicated that active site structure of family β-glycosidases has been conserved, despite the specificity differences that exist between the individual family members.

Chapter 2 describes the elucidation of the overall 3D structure of *P. furiosus* CelB and focuses on the active site structure of the β-glucosidase CelB from hyperthermophilic *P. furiosus*. It showed an organization of CelB monomers in a slightly twisted square to form a tetramer, highly

similar to those observed for *S. solfataricus* LacS (*21*) and *Thermosphaera aggregans* β-glycosidase (*7*). Furthermore, the active site structure of CelB was found to be highly homologous to those in all other determined 3D structures of family 1 β-glycosidases and was studied in relation to substrate specificity by structural comparison with related β-glycosidases in Chapters 2 and 3. In Chapter 2, the active site of the 6-phospho-β-galactosidase LacG of *Lactococcus lactis* was compared to that of CelB. The LacG enzyme is member of a subgroup of family 1 glycoside hydrolases that are specific for the hydrolysis of 6-phospho-glycosides and poorly capable of hydrolyzing non-phosphorylated glycosides (*22*). The study decribed in Chapter 2 demonstrates that this specificity is the result of the structural adaptations in the active site. In the active site of LacG three residues were identified that interact with the phosphate group of galactose-6-phosphate that has been cocrystallized with LacG (Ser 428, Lys 435 and Tyr 437) (*8*). Replacement of the corresponding residues in CelB (Glu 417, Met 424 and Phe 426) with LacG residues improved the 6-phospho-β-galactosidase activity of CelB. In a reverse approach, the substrate specificity of *L. lactis* 6-phospho β-galactosidase LacG was engineered by replacement of the phosphate interacting residues with the corresponding residues of the related cyanogenic β-glucosidase from *Trifolium repens* (*22*). However, the corresponding residue of Ser 428, a glutamate in cyanogenic β-glucosidase, was not introduced in LacG. Instead an aspartic acid residue was introduced which most likely led to the inactivation of the triple mutant LacG S428D/K435/Y437F (*22*). The double mutant LacG K4355/Y437F was more specific for the hydrolysis of non-phosphorylated galactosides over phosphorylated galactose (*22*).

In Chapter 3 the protein sequence of CelB was compared to that of the low-active β-mannosidase BglB of *P. horikoshii*. A constructed 3D model of BglB revealed a high structural similarity between the active sites of BglB and CelB, but highlighted the presence of 2 residues that were different between the enzymes (Gln 77 and Asp 206 in BglB, Arg 77 and Asn 206 in CelB). A modeled galactose monomer in the active site showed that these residues were located close to the catalytic glutamic acid residues. Asn 206 has an interaction with the C2-hydroxyl in 3D structures of family 1 enzymes cocrystallized with ligands (*8, 23*), and at this position glucose has an equatorial hydroxyl, while that in mannose is axial. The substrate specificities of the two enzymes could be exchanged by introduction of the two unique active site residues of BglB in CelB and *vice versa*. With the thermostable transition state analogue PheImGlc, the effect of the substitutions on the hydrolysis reaction at 90°C was evaluated. The results showed that the active site of CelB has been optimized to stabilize the transition state of the reaction. On the other hand, the unique residues in the active site of BglB result in a largely reduced affinity for the transition state. This is postulated to result from altered interaction with the substrate C2-hydroxyl group in BglB, which has an interaction with the asparagine residue at the corresponding position of Asp 206 in other family 1 enzymes (*8, 23*). This interaction has been shown to contribute most to stabilization of the

transition state in the family 1 β-glucosidase from *Agrobaterium faecalis* (*24*) and is most likely altered by the presence of Asp 206 in BglB.

Further conservation of the active site structure is indicated by the biochemical comparison of CelB with the β-glycosidase LacS from *S. solfataricus* (Chapter 4). Whereas the hydrolytic characteristics of the two enzymes are highly similar, which indicates considerable active site homology, the differences are to be found within the mechanisms that have been used for thermostabilization of the proteins. Additionally, the active sites of LacS and CelB could be integrated together in the high-performance hybrids (Chapter 6). Experiments with the thermostable inhibitor PheImGlc resulted in similar inhibition constants for both proteins and the CelB-LacS high performance hybrids, which indicated a comparable structural organization of their active sites.

Because of the high degree of active site conservation in family 1 of glycosyl hydrolases, the vast amount of research done on substrate binding and hydrolysis mechanism of family 1 β-glycosidases is directly applicable to the thermostable family members. This provides an excellent starting point for activity engineering studies.

8.5 Mechanistic considerations

The hydrolysis of saccharides by retention of anomeric configuration through the double displacement mechanism has been well established for family 1 enzymes, including CelB of *P. furiosus* (*3*). Two catalytic glutamate residues are essential for the reaction. Substitution of these residues results in severely reduced hydrolytic activities (*2, 25-29*). The studies described in this thesis provide additional information on other residues that are located in or close to the active site of *P. furiosus* CelB (Table 8.2).

The substituted residues can be placed in two catagories: those that have an interaction with the substrate and those that have not. The substitutions in the latter category have various effects depending on the purpose of the study. All substitutions of the former category have a negative effect on the hydrolysis of pNp-Glc. The catalytic nucleophile binds C1 covalently during hydrolysis (*30*) and its substitution cripples the enzyme (*2*). Asn 206 and Glu 417 are postulated to interact with the glucose hydroxyl groups on C2 and C4, respectively. Substitution of Asn 206 by Ser (*5*) and Asp (Chapter 3) results in comparable increases in activation energy for the hydrolysis of pNp-Glc (Table 8.3). When Glu 417 is replaced by Ser, the increase in activation energy for hydrolysis of oNp-Glc is 2/3 of that for pNp-Glc hydrolysis by CelB N206S and CelB N206D. Assuming that the hydrolysis of oNp-Glc and pNp-Glc are comparable, this implies that the interaction of Asn 206 with the substrate contributes more to the hydrolysis of aryl glucosides than that of Glu 417. These observations are true for the deglycosylation step of the reaction, since the nitrophenol group is an excellent leaving group.

Table 8.2 Overview of substitutions in and near the active site of *P. furiosus* CelB and their effects on the activity of CelB. Interactions of enzyme residues have been postulated from 3D structures of *L. lactis* LacG and *B. polymyxa* BglA, which have been determined with ligands bound in their active sites(8, 23).

Residue	Description	Mutation	Effect	Ref.
Arg 77	salt-bridge with catalytic nucleophile Glu 372	Gln	reduced k_{cat}/K_m on pNp-Glc, pNp-Gal, reduced k_{cat} and K_m on pNp-Man at 90 C	Chapter 3
Asn 206	hydrogen bond with C2-hydroxyl	Asp	reduced k_{cat}/K_m on pNp-Glc, pNp-Gal, reduced k_{cat} and K_m on pNp-Man at 90 C	Chapter 3
		Ser	reduced k_{cat}/K_m on pNp-Glc, pNp-Gal, only reduced k_{cat} on pNp-Man at 90 C	(5)
Thr 371	adjacent to catalytic nucleophile	Ala	increased k_{cat} on pNp-Glc and pNp-Gal at RT, reduced thermostability	Chapter 5
Glu 372	catalytic nucleophile	Gln	1000-fold reduced k_{cat} on pNp-Glc at 90 C	(2)
		Asp	200-fold reduced k_{cat} on pNp-Glc at 90 C	(2)
Asn 415	hydrogen bond to active site residues Gln 17 and Glu 417	Ser	increased k_{cat} on pNp-Glc at RT	Chapter 5
Glu 417	hydrogen bond to C4-hydroxyl on glucose	Ser	increased k_{cat} on oNp-gal-6-P , reduced k_{cat}/K_m on oNp-Glc and oNp-Gal at 90 C	Chapter 2
Ala 419	next to Trp 418, which hydrogen bonds to C3-hydroxyl on glucose	Thr	increased k_{cat} on pNp-Glc at RT, reduced thermostability	Chapter 5
Met 424		Lys	pH optimum for hydrolysis from pH 5 to 6 at 90 C, in combination with F426Y increased yields in oligosaccharide synthesis at low lactose concentrations	Chapters 2 and 7
		Val	increased k_{cat} on pNp-Glc and pNp-Gal at RT, reduced thermostability	Chapter 5
Phe 426		Tyr	increased affinity for galactosides at 90 C, increased yields in oligosacharide synthesis	Chapters 2 and 7

Table 8.3 Kinetic parameters for hydrolysis of oNp-Glc and pNp-Glc at 90°C by wild-type CelB and CelB mutants and free energy for activation.

Enzyme	Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1}mM^{-1})	$\Delta\Delta G°$ (kJ/mol)[1]
CelB wt [2]	pNp-Glc	0.19 ± 0.06	1140 ± 86	7337	-
CelB N206D [2]	pNp-Glc	4.6 ± 0.06	116 ± 7	24.9	17.2
CelB N206S [3]	pNp-Glc	12	340	28.3	16.8
CelB wt [4]	oNp-Glc	0.28 ± 0.03	860 ± 25	4048	-
CelB E417S [4]	oNp-Glc	29 ± 10	3300 ± 630	113	10.8

[1] calculated by $\Delta\Delta G° = RT \ln((k_{cat}/K_m)_{CelB\,wt}/(k_{cat}/K_m)_{CelB\,mut})$ after ref. (*67*)

[2] data from Table 3.3

[3] data from (*5*)

[4] data from Table 2.2

The optimum pH for hydrolysis by this mechanism lies between the pK_a-values of the two catalytic glutamate residues (*31*). However, the control of the pK_a-values is less clear. The cycling of the pK_a-value of the general acid-base glutamate during the course of the reaction implies a tight electronic control in the active site. Interestingly, CelB M424K hydrolyzes glycosides optimally at a pH-value that is unit higher than wild-type CelB (Chapter 2). The responsible substitution (M424K) introduced a positive charge in the active site at considerable distance of the catalytic glutamates (~ 15 Å). On the other hand, introduced charge differences in the near vicinity of the catalytic residues, as described for the CelB and BglB variants in Chapter 3, did not affect the optimum pH for hydrolysis. This could not be explained, which indicates that the control of the pH optimum is not yet understood. Insight in the electronic balance in the active site is likely to follow from analysis of other enzyme variants with shifted pH optima for hydrolysis. Alternatively, the electronic organization of the active site could be computed in a molecular dynamics approach based on a protein crystal structure (*32*). Chapter 7 provides insight in the optimal pH of CelB variants for the kinetic transglycosylation reaction. For most variants the optimal yields in oligosaccharide synthesis were observed at the optimal pH for hydrolysis.

The present studies in which an enzyme was crystallized with a hydrolysis product or inhibitor in the active site (*8, 23*) allowed for analysis of the interactions of family 1 enzymes with the substrate in the −1 subsite. It appears that these interactions have been well preserved throughout family 1, since the substrate interacting residues have been almost fully conserved. The structure of family 1 *B. polymyxa* BglA with 2-deoxy-2-fluoro-glucose (2F-glucose) covalently bound in the active site has been deposited in the database (PDB entry: 1E4I). The 2F-glucose forms a glycosidic bond with Glu 352 with a remarkable bond angle of 86° around the oxygen atom, instead of the usually observed ~105° (Figure 8.2). This bond strain might facilitate hydrolysis in the consecutive reaction step. The 3D-structures of covalent glycosyl-enzyme intermediates of cellulases do not show this strained glycosidic bond (*33, 34*). The interactions of

the hydroxyl groups of the bound 2F-glucose with the enzyme are highly similar to those of the bound gluconate molecule in BglA (23). Furthermore, the positions of most residues are identical to those in the BglA structure without a covalently bound intermediate (PDB entry: 1TR1) (35). Only the nucleophilic Glu 352, covalently bound to 2F-glucose, has rotated 87° around C-atom B, pushing the side chain of Tyr 296 aside (Figure 8.2). Contrary to other studies, the active site of BglA has not been mutated in order to trap the covalent glycosyl-enzyme intermediate (33). However, the interactions of the fluorinated sugar can be expected to differ slightly from those of non-substituted glucose.

The glycosidic glucose-β-1,4-glucose bond is one of the most stable found in nature (36). In order to break this bond, the active sites of cellulases been found to put considerable strain on the bond by distorting the sugar moiety in the −1 subsite upon formation of the Michaelis-complex (33, 34). A similar distortion can be expected to occur in family 1 β-glycosidases. Such structure has not been determined up to now and would provide a more complete understanding on the catalysis by family 1 enzymes. Also, little is known about the enzyme-substrate interactions in the +1 subsite. However, the fact that family 1 enzymes can accept a variety of glycons and aglycons in the +1 subsite, indicates that these interactions are less critical than those in the −1 subsite.
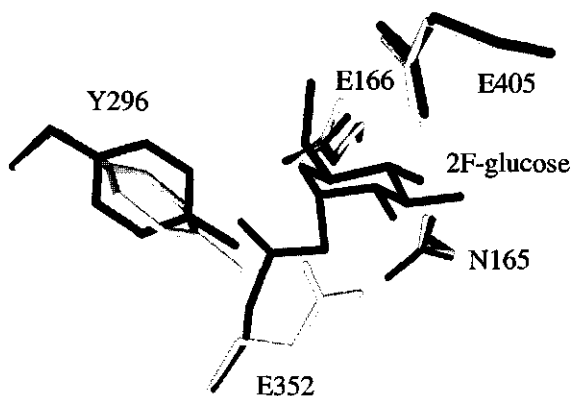

Figure 8.2 Overlay of active sites of *B. polymyxa* BglA β-glycosidase with 2-deoxy-2-fluoro-D-glucose (2F-glucose) covalently bound to Glu 352 (PDB: 1E4I, in black) and without ligand in the active site (PDB: 1TR1, in grey). Image generated in Swiss PDB viewer (71) and visualized using Pov-Ray (72).

```
Pfu_CelB : EIATQKEIPEELAHLADLKFVT--RK------- : 472
Sso_LacS : EIATNGAITDEIEHLNSVPPVKPLRH------- : 489
Pfu-BmnA : EIVANNGVTKKIEEEL---LRGM---------- : 511
Pab-BmnA : RIIS-EGTTKGLQDEFRLKVRGERYENDSG--- : 520
Tag_Bgly : EIATHNGIPDELQHLTLIQ-------------- : 481
Pho-BglB : KIVM-EGIE----------------------- : 483
Pfu-BglB : MIVA-EGIE----------------------- : 483
Pab-BglB : KVVK-EGIE----------------------- : 483
Tko-BglB : KIVR-EGLP----------------------- : 483
```

Figure 8.3 C-terminal amino acid sequences of archaeal β-glycosidases. Tag_Bgly: *T. aggregans* β-glycosidase (AF053078), Sso-LacS: *S. solfataricus* β-glycosidase LacS (M34696), for other sequence details see figure 8.1.

A covalent enzyme-substrate intermediate has been trapped in the family 1 myrosinase of *Sinapis alba* as well (*37*). However, the active site of myrosinase has been adapted for the hydrolysis of S-glycosidic bonds and differs from that of the β-glycosidases, because of the better leaving group ability of the S-aglycon. The effect of the leaving group ability becomes apparent from the analysis of the high performance LacS-CelB hybrids (Chapter 6). The ratio between the hydrolytic activities on the natural substrates cellobiose and lactose correlates with the length of the LacS fragment of the hybrid. Hybrids with a longer LacS fragment have a cellobiose/lactose ratio comparable to LacS, while hybrids with a shorter fragment, resemble CelB. However, this is not observed for the hydrolysis of pNp-substrates, presumably because of the less critical substrate distortion that is needed for the liberation of the nitrophenol, which is a much better leaving group than glucose.

8.6 Thermostability and thermoactivity

With a reported half-life for thermal inactivation of 85 h at 100°C, the β-glucosidase CelB of *P. furiosus* is the most stable family 1 member, described to date (*1*). In Chapter 1 an introduction on protein thermostability has been given. In many 3D structures of extremely stable proteins an increase of ion pairs, especially on the surface, has been identified compared to proteins with a lower stability. The determined structures of the family 1 β-glycosidases from *S. solfataricus* and *T. aggregans* (52% amino acid identity) show such an increase (*7, 21*). Interestingly, most residues involved in ion pairs do not seem to be conserved between the two enzymes or protein sequences of other related β-glycosidases from hyperthermophiles (*7*). The ion-pairs that have been conserved are located in the active site and also present in β-glycosidases from mesophiles, which suggests a role in catalysis. One of such ion-pairs is the interaction between an arginine and the nucleophilic glutamate as described in Chapter 3. A careful analysis of the 3D structure of CelB (*38*) revealed a lower number of possible ion-pairs than in *S. solfataricus* LacS, while CelB is more stable than LacS. This observation has been confirmed in the biochemical comparison of *P. furiosus* CelB and *S. solfataricus* LacS (Chapter 4), which indicated a strong stabilizing role for hydrophobic

interactions in CelB, instead of ion-pair interactions as found in LacS. Nevertheless, ion-pairs seem to be contributing to stability in CelB as well. At the C-terminus of LacS an intricate ion-pair network has been described, which involves the penultimate arginine residue at the fourfold subunit interface at the center of the protein (*21*). The arginine residue has been conserved between *P. furiosus* CelB, BmnA and *S. solfataricus* LacS. The arginine is possibly involved in anchoring of the C-terminus, a thermostabilizing mechanism (*39*). The formation of an ion-pair network at the C-terminus similar to that of LacS is not possible in CelB that is two residues shorter at the C-terminus (Figure 8.3). Instead, Arg 471 seems to form an ion-pair interaction with Asp 377 of the neighboring subunit, while Lys 472 seems to hydrogen bonds to the backbone oxygen of Arg 381 and Val 469 (Figure 8.4).

The effect of the interactions at the C-terminus of *P. furiosus* CelB to chemical and thermal inactivation has been evaluated by site-directed mutagenesis (*38*). In addition, the resulting C-terminal CelB mutants were tested for their chemical and thermal stability, together with 2 extra mutants with C-terminal substitutions (Table 8.4). Removal of the last and second last amino acid residue resulted in loss of kinetic stability (*38*). The results of the chemical and thermal stability studies are not so easily interpreted, since clearly different types of stability have been tested. The deletion mutants show reduced half-lives of thermal inactivation in conjunction with lowered
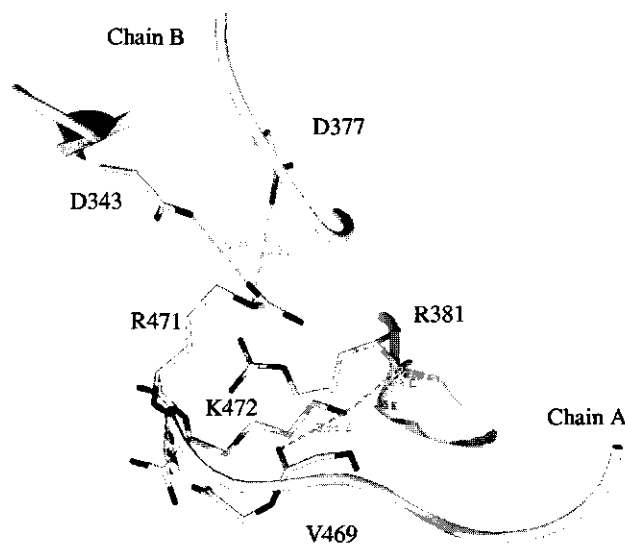

Figure 8.4 Proposed interactions of Arg 471 and Lys 472 in β-glucosidase of CelB. Image generated in Swiss PDB viewer (*71*) and visualized using Pov-Ray (*72*).

Table 8.4 Midpoint concentration (C_m) of guanidine hydrochloride-induced unfolding, apparent melting temperatures (T_m), and half-lives of activity at 106°C of C-terminal CelB mutants. ND = not determined

CelB	$C_m (M)$[1]	$T_m (°C)$[2]	$t_{1/2} (min)$[4]
wild-type	3.6	106.8	139 ± 3.5
R471A	3.5	106.5	91 ± 14
K472Δ	3.6	106.1	58 ± 4.8
R471ΔIK472Δ	3.4	102.0	137 ± ND[5]
K472A	3.1	107.3	43 ± 9.4
R471AIK472A	3.0	108.9	37 ± 7.4
D377A	3.1	99.5[3]	ND
D377AIR471A	3.0	100.2	ND

[1] Reaction conditions: 50 μg/ml protein in 50 mM NaP$_i$, pH 6.5, incubated for 48 h at 25 °C, all values ± 0.1 M

[2] Reaction conditions: 1.0 mg/ml protein at 20 mM NaP$_i$, pH 7.5. Denaturation was irreversible. Except for CelB D377A, in all cases two-state denaturation was observed, all values ± 0.1°C.

[3] Apparent three-state unfolding was observed, temperature of the first peak has been given.

[4] Data from Lebbink (*38*)

[5] At 90 °C

melting temperatures. However, the chemical stability of these mutants is sometimes comparable to that of wild-type CelB. Removal of the side chain of Arg 471 has in general a limited effect on the stability, since CelB R471A slightly less stable than wild-type CelB. Interestingly, the side chain of Lys 472 contributes more to the stability than that of its neighbor, while its side chain does not appear to be involved in a salt-bridge interaction in the CelB structure. Removal of the side chain of Asp 377 is highly destabilizing, which implicates additional interactions besides those with Arg 471. Finally, it can be concluded that the residues at the C-terminus appear to participate in protein stabilization, but the precise mechanism is remains unclear as so far.

From the results of the directed evolution study described in Chapter 5 it becomes clear that relatively few substitutions can have a considerable destabilizing effect. However, this coincided with an increased activity at room temperature. Similar results have been obtained in comparable studies, where thermozymes were optimized for low-temperature catalysis (*40, 41*). The current view is that proteins from hyperthermophiles have a limited flexibility at room temperature, compared to proteins from mesophiles, which would explain their relatively low activity at room temperature (*42*). A total increase in flexibility would increase the activity at low temperatures, but consequently compromise thermostability. Most likely, the CelB-LacS high-performance hybrids that have been described in Chapter 6 have acquired their increased activity at 70°C by a similar

mechanism, since their thermostability was considerably less than that of CelB. On the other hand, increased activity at lower temperatures together with wild-type stability could possibly result from a locally increased flexibility in or near the active site, while the structure around the active site remains intact. CelB variant N415S (Chapter 5) shows this behavior similar as has been found in comparable studies (*40, 41*). Reverse studies, in which proteins from mesophiles were thermostabilized by directed evolution, have yielded more stable variants in which the activity was similar to that of the wild-type proteins at low temperatures (*43-45*). A reduced surface loop flexibility was a major observation in the high performance variants (*43, 46*). Furthermore, a limited amount of engineered substitutions in a single loop area of *Bacillus steathermophilus* thermolysin-like protease increased the stability and shifted the optimum temperature for catalysis 21°C upward (*47*). The stabilized mutant displayed wild-type activity at low temperatures. In conclusion, it has been shown that low-temperature activity can be improved with retention of thermostability, and *vice versa*. The fact that the combination of these traits is less frequently observed in wild-type enzymes indicates that it is apparently not biologically relevant (*48*).

8.7 Protein engineering by rational design versus laboratory evolution

In the studies presented in this thesis, the *P. furiosus* β-glucosidase CelB has been engineered by rational design as well as by laboratory evolution, which allows for a comparison of the two approaches. The substrate specificity of β-glucosidase CelB was tested by structural comparison with the 6-phospho-β-galactosidase LacG from *L. lactis* (Chapter 2) and the β-mannosidase BglB from *P. horikoshii* (Chapter 3). Both LacG and BglB display different specificities that could be correlated to amino acid differences in their active sites. In each case, the substrate specificity of CelB was altered towards that of the other enzyme by site-directed mutagenesis, although LacG was still much more efficient in the hydrolysis of 6-phospho-galactosides than all CelB mutants (Chapter 2). Likewise, β-glucosidase activity was still the most dominant activity in the CelB variants with the inserted BglB active site residues (Chapter 3). This was attributed to differences in positioning of the active site residues by amino acids in the second shell, which must account for specificity differences between family 1 orthologs, as concluded in other studies (*49*). Extensive testing of enzyme variants might result in the discovery of unexpected catalytic features, such as increased oligosaccharide synthesis (Chapter 7).

The laboratory evolution approaches were successful in isolating β-glycosidases with the anticipated phenotypes (Chapters 5 & 6). A library with random mutated CelB variants was created by error-prone PCR combined with DNA shuffling, followed by efficient selection of variants with increased activity at room temperature (Chapter 5). One variant containing the single mutation N415S was over 3-fold more active than wild-type CelB at room temperature. CelB N415S displayed wild-type stability, while the increase in catalytic efficiency at room temperature was

highest for the screening substrate pNp-Glc. This exemplifies the first law in directed evolution: "You get what you screen for" *(50)*.

Creation of hybrid enzymes by DNA family shuffling of related genes requires substantial gene identity, usually 70% or higher *(51)*. The genes coding for *P. furiosus* CelB and *S. solfataricus* LacS are only 56% identical. Despite this low identity, the *celB* and *lacS* genes were successfully shuffled to result in thermostable hybrid β-glycosidases with over 8-fold increased lactose-hydrolytic capacity (Chapter 6). The high-performance hybrids had resulted from a single crossover event near the N-terminus and were composed of a 33 to 45 amino acids-long LacS N-terminal sequence and a CelB core. The hybrids displayed increased activities on all tested substrates, while their thermostability was between that of parental *P. furiosus* CelB and *S. solfataricus* LacS.

The activity and substrate specificity of the high-performance β-glycosidases of Chapter 5 and 6 might be correlated to their thermostability and, hence, flexibility (see above). CelB N415S is only more active on pNp-Glc and has wild-type stability. The increased activity is probably the result of a locally increased flexibility, which is only advantageous for the catalysis of the screening substrate. Enzymes with increased activity on additional substrates, for instance CelB T371A and the LacS-CelB hybrids, have a generally increased flexibility and are less thermostable.

In conclusion, by rational protein design it has been possible to improve catalytic features of β-glucosidase CelB, although further improvement of the resulting enzyme variants can most likely be achieved by a directed evolution approach. A combination of the two approaches, therefore, is likely to yield the best results in enzyme engineering *(52)*.

8.7 Applications of β-glycosidases from hyperthermophiles

The β-glucosidase CelB has been tested for hydrolysis of lactose in whey *(53)*, oligosaccharide synthesis (Chapter 7) *(54, 55)*, glycoconjugate synthesis *(56)*, and use as a reporter enzyme *(57)*. In other studies β-glycosidases from hyperthermophiles have been cloned in plants to facilitate downstream hydrolysis of plant carbohydrates *(58)* or tested for catalyzing of commercially relevant reactions *(59, 60)*. Inactive enzyme variants with the nucleophile residue removed, have potential in oligosaccharide synthesis *(61, 62)*. The studies described in Chapters 5, 6 and 7 have shown that it is possible to tailor β-glycosidases from hyperthermophiles for such applied reactions. Further studies might evaluate the use of these enzymes as biosensors, for instance in lactulose detection *(63)* or in production of low-lactose milk. During an ultra heat treatment the thermostable β-glycosidase would hydrolyze the lactose. Ultimately, the enzyme could be ectopically expressed in mammary glands of cows, analogous to studies in mice *(64)*.

8.8 Future perspectives

The β-glycosidases from hyperthermophilic sources have proven to be interesting and versatile catalysts from both fundamental and applied points of view. The studies described in this thesis show that β-glycosidases are quite abundantly present in hyperthermophiles and that the substrate specificity and thermoactivity of these enzymes can be tailored by protein engineering. Studies on hyperthermostable proteins will continue to elucidate the molecular mechanisms of high-temperature substrate recognition and catalysis, as well as protein stabilization.

References

1. Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.
2. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J Bacteriol 177*, 7105-11.
3. Bauer, M. W., and Kelly, R. M. (1998) *Biochemistry 37*, 17170-8.
4. Gueguen, Y., Voorhorst, W. G., van der Oost, J., and de Vos, W. M. (1997) *J Biol Chem 272*, 31258-64.
5. Lebbink, J. H. G., Kaper, T., Kengen, S. W. M., Van der Oost, J., and De Vos, W. M. (2001) *Methods Enzymol 330*, 364-79.
6. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol 185*, 60-89.
7. Chi, Y. I., Martinez-Cruz, L. A., Jancarik, J., Swanson, R. V., Robertson, D. E., and Kim, S. H. (1999) *FEBS Letters 445*, 375-383.
8. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.
9. Kaper, T., Verhees, C. H., Lebbink, J. H. G., van Lieshout, J. F. T., Kluskens, L. D., Ward, D. E., Kengen, S. W. M., Beerthuyzen, M. M., de Vos, W. M., and van der Oost, J. (2001) *Methods Enzymol 330*, 329-46.
10. Bauer, M. W., Bylina, E. J., Swanson, R. V., and Kelly, R. M. (1996) *J Biol Chem 271*, 23749-55.
11. Matsui, I., Sakai, Y., Matsui, E., Kikuchi, H., Kawarabayasi, Y., and Honda, K. (2000) *FEBS Lett 467*, 195-200.
12. Voorhorst, W., Gueguen, Y., Geerling, A., Schut, G., Dahlke, I., Thomm, M., Van der Oost, J., and De Vos, W. (1999) *J Bact 181*, 3777-3783.
13. Fiala, G., and Stetter, K. O. (1986) *Arch Microbiol 145*, 56-61.
14. Gonzalez, J. M., Masuchi, Y., Robb, F. T., Ammerman, J. W., Maeder, D. L., Yanagibayashi, M., Tamaoka, J., and Kato, C. (1998) *Extremophiles 2*, 123-30.
15. Marteinsson, V. T., Watrin, L., Prieur, D., Caprais, J. C., Raguenes, G., and Erauso, G. (1995) *Int J Syst Bacteriol 45*, 623-32.
16. Zillig, W., Stetter, K. O., Wunderl, S., Schulz, W., Priess, H., and Scholz, I. (1980) *Arch Microbiol 125*, 259-69.
17. Huber, R., Dyba, D., Huber, H., Burggraf, S., and Rachel, R. (1998) *Int J Syst Bacteriol 48*, 31-8.
18. Barbier, G., Godfroy, A., Meunier, J. R., Querellou, J., Cambon, M. A., Lesongeur, F., Grimont, P. A., and Raguenes, G. (1999) *Int J Syst Bacteriol 49*, 1829-37.
19. Makarova, K. S., Aravind, L., Galperin, M. Y., Grishin, N. V., Tatusov, R. L., Wolf, Y. I., and Koonin, E. V. (1999) *Genome Res 9*, 608-28.
20. Da Costa, M. S., Santos, H., and Galinski, E. A. (1998) in *Advances in Biochemical Engineering Biotechnology* (Atranikian, G., Ed.) pp 117-54, Springer-Verlag, Berlin.
21. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.
22. Schulte, D., and Hengstenberg, W. (2000) *Protein Eng 13*, 515-8.

23. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.

24. Namchuk, M. N., and Withers, S. G. (1995) *Biochemistry 34*, 16194-202.

25. Gebler, J. C., Trimbur, D. E., Warren, A. J., Aebersold, R., Namchuk, M., and Withers, S. G. (1995) *Biochemistry 34*, 14547-53.

26. MacLeod, A. M., Tull, D., Rupitz, K., Warren, R. A., and Withers, S. G. (1996) *Biochemistry 35*, 13165-72.

27. Wang, Q., Trimbur, D., Graham, R., Warren, R. A., and Withers, S. G. (1995) *Biochemistry 34*, 14554-62.

28. Withers, S. G., Rupitz, K., Trimbur, D., and Warren, R. A. (1992) *Biochemistry 31*, 9979-85.

29. Moracci, M., Capalbo, L., Ciaramella, M., and Rossi, M. (1996) *Protein Eng 9*, 1191-5.

30. White, A., and Rose, D. R. (1997) *Curr Opin Struct Biol 7*, 645-51.

31. McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., and Withers, S. G. (1996) *Biochemistry 35*, 9958-66.

32. Gogonea, V., Suárez, D., Van der Vaart, A., and Merz Jr, K. M. (2001) *Curr Opin Struct Biol 11*, 217-23.

33. Notenboom, V., Birsan, C., Nitz, M., Rose, D. R., Warren, R. A. J., and Withers, S. G. (1998) *Nature Struct Biol 5*, 812-818.

34. Davies, J. D., Mackenzie, L., Varrot, A., Dauter, M., Brzozowoski, A. M., Schülein, M., and Withers, S. G. (1998) *Biochemistry 37*, 11707-13.

35. Sanz-Aparicio, J., Hermoso, J. A., Martinez-Ripoll, M., Gonzalez, B., Lopez-Camacho, C., and Polaina, J. (1998) *Proteins 33*, 567-76.

36. Wolfenden, R., Lu, X., and Young, G. (1998) *J Am Chem Soc 120*, 6814-5.

37. Burmeister, W. P., Cottaz, S., Driguez, H., Iori, R., Palmieri, S., and Henrissat, B. (1997) *Structure 5*, 663-75.

38. Lebbink, J. H. G. (1999) PhD thesis pp. 200, Wageningen University, Wageningen, The Netherlands

39. Vieille, C., and Zeikus, G. J. (2001) *Microbiol Mol Biol Rev 65*, 1-43.

40. Merz, A., Yee, M. C., Szadkowski, H., Pappenberger, G., Crameri, A., Stemmer, W. P., Yanofsky, C., and Kirschner, K. (2000) *Biochemistry 39*, 880-9.

41. Roovers, M., Sanchez, R., Legrain, C., and Glansdorff, N. (2001) *J Bacteriol 183*, 1101-5.

42. Jaenicke, R. (2000) *Proc Natl Acad Sci 97*, 2962-4.

43. Arrizubieta, M. J., and Polaina, J. (2000) *J Biol Chem 275*, 28843-8.

44. Giver, L., Gershenson, A., Freskgard, P. O., and Arnold, F. H. (1998) *Proc Natl Acad Sci USA 95*, 12809-13.

45. Miyazaki, K., Wintrode, P. L., Grayling, R. A., Rubingh, D. N., and Arnold, F. H. (2000) *J Mol Biol 297*, 1015-26.

46. Spiller, B., Gersherson, A., Arnold, F. H., and Stevens, R. C. (1999) *Proc Natl Acad Sci USA 96*, 12305-10.

47. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. H. (1998) *Proc Natl Acad Sci USA 95*, 2056-60.

48. Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A. (2001) *Trends Biochem Sci 26*, 100-6.

49. Vallmitjana, M., Ferrer-Navarro, M., Planell, R., Abel, M., Ausín, C., Querol, E., Planas, A., and Pérez-Pons, J. A. (2001) *Biochemistry 40*, 5975-82.

50. You, L., and Arnold, F. H. (1996) *Protein Eng 9*, 315-19.

51. Volkov, A. A., and Arnold, F. H. (2000) *Methods Enzymol 328*, 447-56.

52. Cedrone, F., Ménez, A., and Quéméneur, E. (2000) *Curr Opin Struct Biol 10*, 405-410.

53. Petzelbauer, I., Nidetzky, B., Haltrich, D., and Kulbe, K. D. (1999) *Biotechnol Bioeng 64*, 322-32.

54. Boon, M. A., Van der Oost, J., De Vos, W. M., Jansen, A. E. M., and Van 't Riet, K. (1998) *Appl Biochem Biotechnol 75*, 269-78.

55. Petzelbauer, I., Zeleny, R., Reiter, A., Kulbe, K. D., and Nidetzky, B. (2000) *Biotechnol Bioeng 69*, 140-9.

56. Fischer, L., Bromann, R., Kengen, S. W., de Vos, W. M., and Wagner, F. (1996) *Biotechnology N Y 14*, 88-91.

57. Sessitsch, A., Wilson, K. J., Akkermans, A. D., and de Vos, W. M. (1996) *Appl Environ Microbiol 62*, 4191-4.

58. Montalvo-Rodriguez, R., Haseltine, C., Huess-LaRossa, K., Clemente, T., Soto, J., Staswick, P., and Blum, P. (2000) *Biotechnol Bioeng 70*, 151-9.

59. Briante, R., La Cara, F., Febbraio, F., Barone, R., Piccialli, G., Carolla, R., Mainolfi, P., De Napoli, L., Patumi, M., Fontanazza, G., and Nucci, R. (2000) *J Biotechnol 77*, 275-86.

60. Febbraio, F., Portaccio, M., Stellato, S., Rossi, S., Bencivenga, U., Nucci, R., Rossi, M., Gaeta, F. S., and Mita, D. G. (1998) *Biotechnol Bioeng 59*, 108-15.

61. Trincone, A., Perugino, G., Rossi, M., and Moracci, M. (2000) *Bioorg Med Chem Lett 10*, 365-8.

62. Mayer, C., Jakeman, D. L., Mah, M., Karjala, G., Gal, L., Warren, R. A., and Withers, S. G. (2001) *Chem Biol 8*, 437-43.

63. Moscone, D., Bernardo, R. A., Marconi, E., Amine, A., and Palleschi, G. (1999) *Analyst 124*, 325-9.

64. Jost, B., Vilotte, J. L., Duluc, I., Rodeau, J. L., and Freund, J. N. (1999) *Nat Biotechnol 17*, 160-4.

65. Kaper, T., Van Heusden, H. H., Van Loo, B., Vasella, A., Van der Oost, J., and De Vos, W. M. (2001) *submitted*.

66. Lebbink, J. H. G., Kaper, T., Bron, P., Van der Oost, J., and Vos, W. M. (2000) *Biochemistry 39*, 3656-65.

67. Kempton, J. B., and Withers, S. G. (1992) *Biochemistry 31*, 9961-9.

68. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res 24*, 4876-82.

69. Page, R. D. C. (2000) Treeview 1.64 b, Glasgow, UK.

70. Kaper, T., Lebbink, J. H. G., Pouwels, J., Kopp, J., Schulz, G. E., Van der Oost, J., and De Vos, W. M. (2000) *Biochemistry 39*, 4963-70.

71. Guex, N., Diemand, A., and Peitsch, M. C. (1999) *TiBS 24*, 364-7.

72. Cason, C. (1999) The Persisitant of Vision Development Team, Indianapolis, IN, USA.

S ummary

The substrate specificity, thermoactivity, and performance in industrially significant reactions of β-glycosidases from hyperthermophilic Archaea have been investigated. The following enzymes served as models: β-glycosidase CelB from *Pyrococcus furiosus* (*1*), β-glycosidase LacS from *Sulfolobus solfataricus* (*2*), and the novel β-mannosidase BglB of *P. horikoshii* (this thesis). In addition to the biochemical characterization of wild-type enzymes, the substrate specificity of these enzymes has been probed by site-directed mutagenesis following rational design, while thermoactivity has been addressed in laboratory evolution approaches.

Chapter 1
The introductory Chapter 1 provides a state-of-the-art review on hyperthermophilic organisms, family 1 of glycosyl hydrolases, protein stability, approaches for protein engineering, as well as industrially interesting conversions of the milk sugar lactose by family 1 β-glucosidases. Hyperthermophilic organisms dwell at the more exotic places found in nature and can vary considerably in genotype and phenotype. Several glycosyl hydrolases have been identified in these organisms, some of which belong to family 1 β-glycosidases. This family has been relatively well studied and several aspects of the reaction mechanism and substrate binding have been elucidated. In the absence of a general thermostabilizing mechanism, the high thermostability of proteins from hyperthermophiles is attributed to a combination of optimized interactions, which can vary between individual proteins. Rational design followed by site-directed mutagenesis and laboratory evolution represent two fundamentally different strategies for protein engineering. Family 1 β-glycosidases are able to perform several industrially interesting reactions with lactose as a substrate.

Chapter 2
Crystallization and structure determination of the β-glucosidase CelB from *P. furiosus* to 3.3 Å resolution by the molecular replacement method has been described in Chapter 2. In this procedure, the structures of the β-glycosidase LacS from *S. solfataricus* (*3*) and 6-phospho-β-galactosidase LacG from *Lactococcus lactis* (*4*) were used as search models. For verification of the model, the active site was adapted for the

hydrolysis of 6-phospho-galactosides by structural comparison with the structure of LacG from *L. lactis* with a galactose-6-phosphate bound in the active site (5). Three substitutions were introduced in CelB, which determined the specificity for phosphorylated substrates (E417S), increased the optimal pH for hydrolysis by 1 unit (M424K), and increased the specificity for galactosides (F426Y). The results indicate a high structural similarity between the active sites of CelB and LacG.

Chapter 3
The identification of a novel subfamily of family 1 glycosyl hydrolases and characterization of a representative member, BglB of *P. horikoshii*, has been described in Chapter 3. BglB was characterized as a β-mannosidase with low hydrolytic activity. All protein sequences in this subfamily, including BglB, showed a unique active site architecture with 2 residues differing from the family 1 consensus. These residues were introduced in *P. furiosus* CelB (R77Q and N206D) and shifted the activity profile towards that of *P. horkoshii* BglB, low activity with increased affinity for mannosides. The substitution D206N turned BglB into a β-glucosidase with 10-fold increased activities over wild-type BglB. As a result of the unique active site residues, the transition state of the reaction seems less stabilized in BglB, resulting in a low turnover rate.

Chapter 4
Chapter 4 describes the biochemical comparison of *P. furiosus* CelB and the homologous β-glycosidase LacS from *S. solfataricus* (amino acid identity 53%). The enzymes are very similar regarding their catalytic behavior, but differ considerably in stability. Physical and chemical perturbation experiments indicate that hydrophobic interactions contribute most to CelB stability, while ionic interactions seem the main contributors to LacS stability.

Chapter 5
In the first directed evolution study with CelB, two rounds of random mutants of the *celB* gene were generated and subjected to DNA shuffling, as described in Chapter 5. The functional CelB mutants that resulted from this procedure were screened for increased hydrolytic activity of pNp-β-D-glucose at room temperature. High-performance variants contained substitutions in various locations in the enzyme. Some of these mutants showed a reduced thermostability. The most active mutant contained a single substitution N415S, located near the active site, and was three-fold more active at room temperature compared to wild-type CelB. However, the activity on cellobiose was not equally increased and the activity on galactosides had even diminished. Asn 415 has been proposed to interact with active site residues that form a hydrogen bond with the equatorial C4-hydroxyl of glucosides (6). In galactosides this hydroxyl has been axially oriented.

Chapter 6
In Chapter 6, a second directed evolution study has been described, in which the genes coding for *P. furiosus* CelB and *S. solfataricus* LacS were combined in a DNA family shuffling approach. About 2000 active hybrids were screened at 70 °C for stability, increased lactose hydrolysis, and reduced glucose inhibition. Three high-performance hybrids with 1.5 to 8.6-fold increased activity over the parental enzymes were composed of an N-terminal LacS sequence and a CelB core. Besides a reduced glucose inhibition, the three hybrids showed increased hydrolysis rates on all tested

substrates as well. The stability of the hybrids was intermediate to that of CelB and LacS. This was quite remarkable considering the different mechanisms by which Celb and LacS have been stabilized to withstand high temperatures (see Chapter 4). Together with the generally increased catalysis rates, it seems likely that an increased overall flexibility gives rise to the observed activity optimization.

Chapter 7   Wild-type CelB and several site-directed CelB mutants, which have been described in Chapters 2 and 3, were tested for their oligo-saccharide synthetic capacity in the study described in Chapter 7. At 95 °C, CelB wild-type produces up to 40% (w/w) oligo-saccharides from lactose by trans-glycosylation in an all aqueous system. The substitution F426Y increased the maximal synthesis yield to 45%. CelB M424K|F426Y produced 40% oligo-saccharides versus 18% for wild-type CelB at low lactose concentration (10%). It is hard to rationalize these results due to the complexity of substrate, but they demonstrate that the synthesis yield in these enzymes results from the active site structure and is a trait that can be engineered.

Chapter 8   Finally, Chapter 8 presents a discussion of various aspects of family 1 β-glycosidases by combining the results obtained in the described studies, together with additional results. Specific attention is given to technical aspects of the studies, and the occurrence of family 1 enzymes in *Pyrococcus* and *Thermococcus* species. Furthermore, the active site structure in this enzyme family and its reaction mechanism are discussed. Additionally, the thermostability and thermoactivity of the β-glycosidases from hyperthermophiles, as well as the comparison of the outcomes from rational and *in vitro* evolutionary protein engineering are analyzed. Finally possible applications for β-glycosidases from hyperthermophiles are mentioned.

References   1.   Kengen, S. W., Luesink, E. J., Stams, A. J., and Zehnder, A. J. (1993) *Eur J Biochem 213*, 305-12.

2.   Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.

3.   Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.

4.   Wiesmann, C., Beste, G., Hengstenberg, W., and Schulz, G. E. (1995) *Structure 3*, 961-8.

5.   Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.

6.   Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.

Samenvatting

De substraatspecificiteit, thermoactiviteit, en toepassing in industrieel relevante reacties van β-glycosidases van hyperthermofiele Archaea zijn onderzocht. De volgende enzymen, behorende tot familie 1 van glycosyl hydrolasen, stonden hierbij model: β-glucosidase CelB uit *Pyrococcus furiosus* (*1*), β-glycosidase LacS uit *Sulfolobus solfataricus* (*2*) en het nieuwe β-mannosidase BglB uit *P. horikoshii* (dit proefschrift). Naast de biochemische karakterisering van de wild type enzymen, is de substraat specificiteit getest met behulp van plaatsgerichte mutagenese, terwijl de thermoaktiviteit is onderzocht met behulp van laboratorium evolutietechnieken.

Hoofdstuk 1   Het inleidende Hoofdstuk 1 geeft een actueel overzicht van hyperthermofiele organismen en glycosyl hydrolases behorende tot familie 1. Verder worden eiwitstabiliteit, technieken voor protein engineering, alsmede industrieel relevante omzettingen van het melksuiker lactose door familie 1 glycosyl hydrolasen besproken. Hyperthermofiele organismen groeien optimaal boven 80°C op de meer exotische plekken in de natuur en kunnen aanzienlijk verschillen in genotype en fenotype. Meerdere glycosyl hydrolasen zijn geïdentificeerd in deze organismen, waarvan enkelen tot familie 1 van glycosyl hydrolasen behoren. Deze familie enzymen is relatief goed bestudeerd en verschillende aspecten van hun reactie mechanisme en substraat binding zijn opgehelderd. Bij afwezigheid van een algemeen mechanisme voor thermostabilisering van eiwitten, wordt aangenomen dat thermostabiliteit het samengestelde resultaat is van geoptimaliseerde interacties, waarvan de combinatie kan variëren van enzym tot enzym. Rationeel ontwerp van plaatsgerichte mutaties en laboratoriumevolutie zijn twee fundamenteel verschillende benaderingen voor protein engineering (modificatie van eiwit structuren door mutatie van hun genen (*3*)). Verder kunnen familie 1 β-glycosidases verschillende industrieel interessante omzettingen uitvoeren met lactose als substraat.

Hoofdstuk 2   De kristallisatie en struktuur opheldering van het β-glucosidase CelB van *P. furiosus* tot een resolutie van 3.3 Å met behulp van de molecular replacement methode staan beschreven in Hoofdstuk 2. In deze procedure werden de strukturen van het β-glycosidase LacS uit *S. solfataricus* (*4*) en het 6-fosfo-β-galactosidase LacG uit *Lactococcus lactis*

(5) gebruikt als zoekmodellen. Om de struktuur te verifiëren werd het aktieve centrum van CelB aangepast voor de hydrolyse van 6-fosfogalactosiden. Dit werd gedaan aan de hand van vergelijking met de struktuur van het aktieve centrum van LacG uit *L. lactis*, die is opgehelderd met een gebonden galactose-6-fosfaat in het aktieve centrum. Drie aminozuursubstituties werden geïntroduceerd in CelB, die respectievelijk de specificiteit voor gefosforyleerde substraten bepaalden (E417S), de optimale pH voor hydrolyse 1 eenheid verhoogden (M424K) en de specificiteit voor galactosiden vergrootten (F426Y).

Hoofdstuk 3 De ontdekking van een nieuwe subgroep van familie 1 glycosyl hydrolasen en de karakterisatie van een representatief lid, BglB van *P. horikoshii*, staan beschreven in Hoofdstuk 3. BglB werd gekarakteriseerd als een β-mannosidase met een lage hydrolytische aktiviteit. Alle eiwit sequenties in deze subgroep, inclusief BglB, tonen een unieke struktuur in het aktief centrum met 2 residuen (Gln 77 en Asp 206) die verschillen van de consensus in familie 1 (resp. Arg en Asn). Deze residuen werden geïntroduceerd in *P. furiosus* CelB (R77Q en N206D) en verschoven het aktiviteitsprofiel in de richting van dat van *P. horikoshii* BglB: lage aktiviteit met een verhoogde affiniteit voor mannosiden. De tegenovergestelde substitutie D206N veranderde BglB in een β-glucosidase met 10-voudig verhoogde aktiviteit vergeleken met wild-type BglB. De unieke residuen in het aktieve centrum resulteren in een verminderde stabilisatie van de overgangs toestand van de reactie in BglB met als gevolg een lage omzettingssnelheid.

Hoofdstuk 4 Hoofdstuk 4 beschrijft de biochemische vergelijking van *P. furiosus* CelB en het homologe β-glycosidase LacS van *S. solfataricus* (aminozuur identiteit 35%). De twee enzymen lijken erg op elkaar wat betreft katalyse, maar verschillen aanzienlijk wat stabiliteit betreft. Fysische en chemische perturbatie experimenten geven aan dat hydrofobe interacties het meeste bijdragen aan de stabiliteit van CelB, terwijl ion interacties de grootste bijdrage leveren aan LacS' stabiliteit.

Hoofdstuk 5 In de eerste laboratoriumevolutiestudie met CelB zijn twee cycli van random mutanten van het *celB* gen gegenereerd en onderworpen aan DNA shuffling, zoals beschreven staat in Hoofdstuk 5. De functionele CelB mutanten die voortkwamen uit deze procedure werden geanalyseerd op verhoogde hydrolyse van para-nitrophenyl-β-D-glucopyranoside bij kamertemperatuur. Mutanten met verhoogde aktiviteit bevatten substituties op verschillende plekken in het enzym. Een aantal van deze mutanten vertoonden een verminderde thermostabiliteit. De meest actieve mutant bevatte de enkele substitutie N425S vlak bij het aktieve centrum en was driemaal aktiever dan wild type CelB. Echter, de aktiviteit op cellobiose, en galactosiden was niet evenredig hoger. Asn 415 wordt verondersteld te interacteren met de residuen in het aktieve centrum die een waterstofbrug vormen met de equatoriale C4-hydroxylgroep van glucosiden (6). In galactose is deze hydroxyl group axiaal gepositioneerd.

Hoofdstuk 6 In Hoofdstuk 6 is een tweede laboratoriumevolutiestudie beschreven, waarin de genen coderend voor *P. furiosus* CelB en *S. solfataricus* LacS werden gecombineerd in een DNA family shuffling experiment. Ruim 2000 aktieve hybriden werden bij 70°C geanalyseerd op

stabiliteit, verhoogde lactose hydrolyse aktiviteit en verminderde glucose inhibitie. Drie hybriden met 1.6 tot 8.6-maal hogere aktiviteit dan de uitgangsenzymen bestonden uit een N-terminale LacS sequentie and een CelB kern. Naast een gereduceerde glucose inhibitie vertoonden alle hybriden eveneens verhoogde hydrolyse aktiviteiten op alle geteste substraten. De stabiliteit van de hybriden lag tussen die van CelB en LacS in. Dit was opzienbarend met het oog op de verschillende mechanismen waardoor CelB en LacS zijn gestabiliseerd om hoge temperaturen te kunnen weerstaan (zie Hoofdstuk 4). Met het oog op de algemeen verminderde stabiliteit, lijkt het waarschijnlijk dat een verhoogde algehele flexibiliteit in de hybriden aanleiding geeft tot de verhoogde aktiviteit.

Hoofdstuk 7

Wild type CelB en een aantal plaats gerichte mutanten, die beschreven staan in Hoofdstuk 2 en 3, zijn getest op hun vermogen om oligosacchariden te synthetiseren, wat beschreven staat in Hoofdstuk 7. Bij 95°C produceerde CelB tot 40% (w/w) oligosacchariden uit lactose via transglycosylering in een geheel waterig systeem. De substitutie F426Y verhoogde de maximale synthese opbrengst naar 45%. CelB M424K|F426Y produceerde 40% oligosacchariden bij lage lactose concentratie (10%) tegen 18% voor wild-type CelB. Het is moeilijk om deze resultaten te verklaren vanwege de complexiteit van het substraat, maar ze tonen wel aan dat het synthetiserend vermogen een eigenschap is die aangepast kan worden.

Hoofdstuk 8

Tenslotte staat in Hoofdstuk 8 een discussie van verschillende aspecten van familie 1 β-glycosidasen aan de hand van de resultaten uit de beschreven studies, alsmede nieuwe gegevens. Speciale aandacht wordt gegeven aan de gebruikte methodologieën en de verspreiding van familie 1 enzymen in Thermococcales. Verder wordt het aktieve centrum en reaktie mechanisme van dit soort enzymen besproken. Daarnaast worden de thermostabiliteit en thermoaktiviteit van de β-glycosidasen van hyperthermofielen, alsmede de vergelijking van rationele en *in vitro* evolutionaire benaderingen voor protein engineering geanalyseerd. Uiteindelijk worden verschillende toepassingen voor β-glycosidasen van hyperthermofielen besproken.

Referenties

1. Voorhorst, W. G., Eggen, R. I., Luesink, E. J., and de Vos, W. M. (1995) *J Bacteriol 177*, 7105-11.
2. Pisani, F. M., Rella, R., Raia, C. A., Rozzo, C., Nucci, R., Gambacorta, A., De Rosa, M., and Rossi, M. (1990) *Eur J Biochem 187*, 321-8.
3. Fersht, A. (1999) *Structure and mechanism in protein science*, W.H. Freeman and Company, New York.
4. Aguilar, C. F., Sanderson, I., Moracci, M., Ciaramella, M., Nucci, R., Rossi, M., and Pearl, L. H. (1997) *J Mol Biol 271*, 789-802.
5. Wiesmann, C., Hengstenberg, W., and Schulz, G. E. (1997) *J Mol Biol 269*, 851-60.
6. Sanz Aparicio, J., Hermoso, J. A., Martinez Ripoll, M., Lequerica, J. L., and Polaina, J. (1998) *J Mol Biol 275*, 491-502.

C urriculum vitae

Thijs Kaper werd geboren op vrijdag 28 mei 1971 te Marken. Op zijn vijfde verhuisde hij naar Barneveld, waar hij de Openbare Lagere School I doorliep. Vanaf 1983 tot 1989 fietste hij elke dag naar Ede, om daar het Marnix College te bezoeken. Dit sloot hij in 1989 af met het behalen van zijn gymnasium diploma. Het volgend schooljaar, 1989-1990, ging hij er even tussenuit en deed hij het Senior year aan de Robert Service High School in Anchorage, Alaska, USA. Aangetrokken door de multidisciplinariteit van de studie, begon Thijs in 1990 met de studie Moleculaire Wetenschappen aan de Landbouwuniversiteit in Wageningen. Tijdens het PGO praktikum liep hij warm voor de moleculaire biologie aan hyperthermofiele micro-organismen. Na een milieu-gericht afstudeervak bij de werkgroep Xenobiotica (dr. Gosse Schraa), verdiepte hij zich onder begeleiding van prof. dr. Willem M. de Vos in de wereld van Archaea bij de werkgroep Bacteriële Genetica, beide bij de vakgroep Microbiologie. Vervolgens bestudeerde hij bij de vakgroep Experimentele Diermorfologie en Celbiologie (dr. Lidy Verburg-Van Kemenade en prof. dr. Wim van Muiswinkel) het immuunsysteem van de karper. In 1996 behaalde hij zijn bul. Een korte periode werkte Thijs vervolgens als onderzoeker op de Research, Development & Quality Division van Sara Lee/Douwe Egberts in Utrecht. Eind 1996 begon hij aan zijn promotie onderzoek bij de werkgroep Bacteriële Genetica van het Laboratorium voor Microbiologie van de Wageningen Universiteit, waar hij struktuur-funktie relaties in β-glycosidases van hyperthermofiele Archaea bestudeerde onder begeleiding van dr. John van der Oost en prof. dr. Willem M. de Vos. Dat werk heeft geleid tot dit proefschrift. Vanaf augustus 2001 leert Thijs de beginselen van de eiwitkristallografie en 3D-struktuuropheldering bij het Center for Structural Biology van het Karolinska Institutet in Zweden onder leiding van prof. dr. Rudolf Ladenstein. Begin 2002 wil hij zich verder gaan verdiepen in de directed evolution van enzymen als post-doctoral fellow bij de vakgroep Microbiologie van de Rijksuniversiteit Groningen onder begeleiding van prof. dr. Lubbert Dijkhuizen.

L ist of publications

1. Kaper, T., Lebbink, J.H.G., Pouwels, J., Kopp, J., Schulz, G.E., Van der Oost, J., and De Vos, W.M. (2000) "Engineering the substrate specificity of the hyperthermostable β-glucosidase CelB from *Pyrococcus furiosus.*" *Biochemistry*, 39(13):3656-65

2. Lebbink, H.J.G., Kaper, T., Peter Bron, Van der Oost, J., and De Vos W.M. (2000) Improving low temperature catalysis in the hyperthermostable *Pyrococcus furiosus* β-glucosidase CelB by directed evolution. *Biochemistry*, 39(17):4963-70

3. Lebbink, H.J.G., Kaper, T., Kengen, S.W.M., Van der Oost, J., and De Vos W.M. (2001) "The β-glucosidase CelB from *Pyrococcus furiosus*: Production by *Escherichia coli*, purification and *in vitro* evolution." *Methods Enzymol* 330:364-79

4. Kaper, T., Verhees, C., Lebbink, J.H.G., Van Lieshout, J., Kluskens, L., Ward, D.E., Kengen, S.W.M., De Vos, W.M., and Van der Oost, J. (2001) "Characterization of β-glycosyl hydrolases from *Pyrococcus furiosus.*" *Methods Enzymol* 330:329-46

5. Pouwels, J., Moracci, M., Cobucci-Ponzano, B., Perugino, G., Van der Oost, J., Kaper, T., Lebbink, J.H.G., De Vos, W.M., Ciaramella, M., and Rossi, M. (1999) "Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal β-glycosidases." *Extremophiles* 4, 157-64

6. Hansson, T., Kaper, T., Van der Oost, J., De Vos, W.M., and Adlercreutz, P. (2001) "Improved oligosaccharide synthesis by protein engineering of β-glucosidase from hyperthermophilc *Pyrococcus furiosus.*" *Biotechnol Bioeng* 73(3):203-10

7. Kaper, T., Van der Oost, J. and De Vos, W.M. (2001) "Engineering thermostable β-glycosidases for saccharide processing" In: Further advances in Carbohydrate Bioengineering, Eds. Teeri, T.T., Svensson, B., Gilbert, H.J., Feizi, T. *in press*

8. **De Roode, B.M., Van der Meer, T.D., Kaper, T., Franssen, M.C.R., Van der Padt, A., Van der Oost, J., Boom, R.J., and de Groot, A.** (2001) "Catalytic potency of the β-glucosidase CelB from *Pyrococcus furiosus* in direct glucosylation"*Enzyme Microb Tech, in press*

9. **Kaper, T, Van Heusden, H.H., Van Loo, B., Van der Oost, J., and De Vos, W.M.** "Substrate specificity engineering of a β-mannosidase and a β-glucosidase from *Pyrococcus* by exchange of unique active site residues." *Submitted*

10. **Kaper, T., Brouns, S., Van der Oost, J., and Willem M. De Vos** "Optimized catalytic features of β-glycosidase hybrids created by DNA family shuffling" *Submitted*