

Characterization and Engineering of Thermostable Glycoside Hydrolases

Johan F.T. van Lieshout

Promotoren:

Prof. dr. J. van der Oost

Persoonlijk hoogleraar bij het Laboratorium voor Microbiologie
Wageningen Universiteit

Prof. dr. W.M. de Vos

Hoogleraar Microbiologie
Wageningen Universiteit

Leden van de
promotiecommissie:

Prof. dr. M. Rossi

University of Naples

Prof. dr. L. Dijkhuizen

Rijksuniversiteit Groningen

Prof. dr. ir. F. Voragen

Wageningen Universiteit

Prof. dr. ir. W. Norde

Wageningen Universiteit

Dit onderzoek is uitgevoerd binnen de onderzoekschool VLAG

Characterization and Engineering of Thermostable Glycoside Hydrolases

Johan F.T. van Lieshout

Proefschrift

Ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
Prof. dr. M.J. Kropff,
in het openbaar te verdedigen
op dinsdag 6 februari 2007
des namiddags te vier uur in de Aula

Johan F.T. van Lieshout – Characterization and engineering of thermostable glycoside hydrolases –
2007

Ph.D. thesis Wageningen University, Wageningen, The Netherlands (2007) 152 p, With summary
in Dutch

ISBN: 90-8504-586-x

Pap en Mam, dit is voor jullie

ABSTRACT

Van Lieshout, J.F.T. (2006) Characterization and Engineering of Thermostable Glycoside Hydrolases. PhD thesis. Laboratory of Microbiology, Wageningen University, The Netherlands.

Glycoside hydrolases form a class of enzymes that play an important role in sugar-converting processes. They are applied as biocatalyst in both the hydrolysis of natural polymers to mono- and oligo-saccharides, and the reverse hydrolysis or transglycosylation, by which oligosaccharides are synthesized from mono- and disaccharides. The latter synthesis process requires high substrate concentrations; to avoid technical problems (e.g. poor solubility, low diffusion rate, microbial contamination) this process is best performed at higher temperatures. Specific oligosaccharides (prebiotics) can stimulate the growth of so-called beneficial microorganisms in the gastro-intestinal tract. The complexity of this ecological system requires a specific combination of oligosaccharides which appears not to be present in currently available commercial preparations. Therefore, the production of highly specific oligosaccharides is desired. Biocatalysts for such a synthesis preferably need to have high catalytic specificity as well as high thermo-activity, and -stability. In this context, the work described in this thesis is aimed to gain knowledge on (i) catalytic mechanisms as well as stability strategies of thermostable glycoside hydrolases, (ii) state-of-the-art engineering methods to optimize these features, and (iii) novel molecular strategies to enhance protein production.

Three classes of glycoside hydrolases have been selected for detailed analysis: α -galactosidase, β -glucosidase and β -glucanase. The α -galactosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* has been cloned, functionally produced and characterized. Successful identification of its catalytic nucleophile allows for the future application of a synthase. The β -glucanase from *P. furiosus* (laminarinase, LamA) has been converted into a glycosynthase by a similar nucleophile mutation: glycosylation was observed with yields of up to 30% of oligosaccharides. In addition, analysis of the molecular basis for the extremely high chemical and thermal stability of LamA revealed an important role for calcium. The stability of LamA has also been analyzed after immobilization. Remarkably, despite a slight loss in secondary structure, LamA remained active upon adsorption to Silica and Teflon up to 130 °C. A mesophilic β -glucanase from *Bacillus* (lichenase LicA) has been successfully stabilized by an innovative method, in which its polypeptide backbone was circularized by intein-driven protein splicing. Finally, we designed a special cloning strategy to generate covalently closed circular messenger RNA (mRNA) that encodes a β -glucosidase from *P. furiosus*. Both *in vivo* and *in vitro* translation of the circular mRNA resulted in functional enzyme. In cases where mRNA stability is limiting the efficiency of protein production, the described engineering approach of transcript-cyclization may provide a solution. In conclusion, the work described in this thesis contributes to establishing a toolbox that may be instrumental for protein engineering in general, and for optimizing enzyme for oligo-saccharide synthesis in particular.

Table of Contents

	Preface and Outline of the Thesis	1
Chapter 1	Isolation and Characterization of Glycoside Hydrolases from <i>Pyrococcus furiosus</i>	3
Chapter 2	Identification and Molecular Characterization of a Novel Type of α -galactosidase from <i>Pyrococcus furiosus</i>	21
Chapter 3	Hydrolase and Glycosynthase Activity of Endo-1,3- β -glucanase from the Thermophile <i>Pyrococcus furiosus</i>	39
Chapter 4	Calcium-induced Tertiary Structure Modifications of Endo- β -1,3-glucanase from <i>Pyrococcus furiosus</i> in 7.9 M Guanidinium Chloride	53
Chapter 5	In Situ Structure and Activity Studies of an Enzyme Adsorbed on Spectroscopically Undetectable Particles	73
Chapter 6	Stabilization of an Endo- β -1,3-1,4-Glucanase by Cyclization	91
Chapter 7	Ribozyme-mediated Engineering of Circular mRNA and its Functional <i>in vivo</i> and <i>in vitro</i> Translation	107
Chapter 8	Summary and Concluding Remarks	125
	Samenvatting en Conclusies	129
	Dankwoord/Acknowledgments	133
	<i>Curriculum vitae</i>	135
	List of Publications	137

Preface and Outline of the Thesis

This thesis describes the research conducted in the course of the STW project “Biocatalysis of specific non-digestible oligosaccharides”. Oligosaccharides are substrates for many intestinal microorganisms, notably *Bifidobacteria* that are believed to be beneficial for human health. Specific non-digestible oligosaccharides can therefore be used as so called prebiotics to promote the growth of these potentially beneficial bacteria in the gastro-intestinal tract. However, the present commercial preparations contain mixtures of different oligosaccharides, which are not very selective in their stimulation of preferential growth of beneficial micro-organisms. Therefore, there is a need to produce more specific oligosaccharides with the highest possible efficiency. The synthesis of these oligosaccharides can be catalyzed by glycoside hydrolases. This class of enzymes can not only be instrumental in the hydrolysis of natural polymers generating specific oligosaccharides, but also catalyze the reverse hydrolysis or transglycosylation, using mono- and di-saccharides as a starting substrate. In the latter synthesis process, product formation is highly enhanced by high initial substrate concentrations. This can be achieved at elevated reaction temperatures. Therefore, thermostable enzymes are most suitable for this process.

The goal of the research described in this thesis was threefold: (i) to isolate and characterize thermostable glycoside hydrolases and their ability to be used as efficient biocatalysts in oligosaccharide synthesis, (ii) to improve their properties with respect to catalysis and stability and (iii) to enhance the production of the biocatalysts.

Chapter 1 gives a general overview on glycoside hydrolases, including classification, reaction mechanism, and protein stability, with special focus on thermostable variants and the methods that are used to study this class of enzymes, like isolation, cloning and heterologous expression, production, activity assays and thermal inactivation experiments.

Chapter 2 describes the isolation and characterization of a novel α -galactosidase (GalA) from the hyperthermophilic archaeon *Pyrococcus furiosus*. A comparison is made with other glycoside hydrolases at a molecular level, and a molecular as well as a biochemical study is performed to identify the catalytic nucleophile residue.

Chapter 3 reports on the engineering of a glycosynthase of the *Pyrococcus furiosus* laminarinase (LamA) and its subsequent characterization. The catalytic nucleophile (Glu 170) was changed for an alanine residue to suppress hydrolase activity. The synthesis of oligosaccharides by the mutated enzyme was studied with respect to the specificity of both the substrates and the glycosidic bond that was created.

Chapter 4 provides detailed insights into the stability aspects of LamA, which maintains residual tertiary interactions and intact secondary structure elements under extreme denaturing conditions. The effect of calcium in 7.9 M guanidinium chloride (GdmCl) was analyzed by circular dichroism and fluorescence spectroscopy. Furthermore, structural interactions that might be responsible for the stability of LamA are discussed by means of a three-dimensional model that was generated based on homology.

Chapter 5 sheds light on the effect of immobilizing LamA on Teflon and silica. Immobilization of an enzyme makes it a more efficient biocatalyst, because it facilitates the development of a continuous reactor system and it leads to ease of removal from the reaction mixture and thus re-use of the enzyme. The structural characteristics and enzymatic activity of the adsorbed endoglucanase were studied and compared with those of the enzyme free in solution.

Chapter 6 focuses on the use of an unconventional method to stabilize enzymes. A covalent linkage was generated between the N-terminus and the C-terminus of the polypeptide chain of the lichenase (LicA) from the bacterium *Bacillus licheniformis*, to create circular variants with increased thermostability.

Chapter 7 presents an engineering exercise to stabilize mRNA. Through a specially designed cloning strategy, the self-splicing activity of the *Tetrahymena thermophila* ribozyme was applied for the production of circular mRNA of the *celB* gene of *Pyrococcus furiosus* that codes for the β -glucosidase CelB.

Chapter 8 summarizes the results from the previous chapters and briefly discusses the importance of the work described in this thesis.

Chapter 1

Isolation and Characterization of Glycoside Hydrolases from *Pyrococcus furiosus*

Parts of this chapter were adapted from:

Kaper, T., Verhees C.H., Lebbink, J.H., van Lieshout, J.F.T., Kluskens, L.D., Ward, D.E., Kengen, S.W., Beerthuyzen, M.M., de Vos, W.M., and van der Oost, J. (2001) Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus*.

Methods in Enzymology **330**, 329-346.

Introduction

Enzymes from bacterial and archaeal hyperthermophiles have been studied extensively for their catalytic properties and stability at extremely high temperatures, as well as for their biotechnological potential as biocatalysts at elevated temperatures [1]. Considerable attention has been given to sugar-converting enzymes from heterotrophic hyperthermophilic archaea, mainly in members of the archaeal orders Sulfolobales (e.g. *Sulfolobus solfataricus*) and Thermococcales (e.g. *Pyrococcus furiosus*). The aim here is to summarize relevant characteristics of enzymes that catalyze the hydrolysis of sugars, the glycoside hydrolases (GHs). In addition, methods are described that have been instrumental in the molecular and biochemical characterization of such enzymes, with selected examples. The reviewed studies form an essential basis for a next generation of experiments, the engineering of GHs for optimal exploitation of their potential.

Glycoside Hydrolases

Many organisms utilize carbohydrates as carbon and energy sources. Consequently, enzymatic systems to hydrolyze these glycosides are of physiological importance. These systems comprise the class of glycoside hydrolases (GHs). The diversity and complexity of carbohydrate stereochemistry is reflected by a vast variety of GHs present in nature. Enzymes are highly specific for hydrolysis of either the α - or the β -glycosidic bond. They can catalyze the hydrolysis of the glycosidic bond in the middle of a poly- or oligosaccharide (endo-acting) or at the end of a saccharide molecule (exo-acting). However, some glycoside hydrolases, particularly the exo-acting β -specific ones, show broad specificities (CelB; BmnA;) [2]. This complicates a classification according to the Enzyme Commission (EC) designations, which are based on specificity. To facilitate distinguishing, Henrissat [3] proposed an alternative classification system that is based on amino acid sequence and folding similarities. At present this system comprises 107 families that have been organized in 14 clans (http://afmb.cnrs-mrs.fr/CAZY/fam/acc_GH.html) [4]. This classification has the advantage that it puts (new) enzymes not only in a mechanistic, but also in a structural context. Enzymes in each family share similarity of both polypeptide fold and reaction mechanism. Per definition the catalytic residues (see below) are conserved within each family [5].

Pyrococcus furiosus is a strictly anaerobic Euryarchaeon that grows optimally at 100 °C. Like many heterotrophic hyperthermophiles, *P. furiosus* can grow on several carbohydrates like starch, maltose, cellobiose and laminarin because it possesses the enzymes required for their degradation [5-7]. In addition, genome sequence analysis has revealed new genes encoding glycoside hydrolases whose specificity and (physiological) function yet have to be determined. Table 1.1 summarizes the glycoside hydrolases found in *Pyrococcus* species.

Table 1.1. Glycoside hydrolases from *Pyrococcus* sp.

Family	Organism	Enzyme	UniProt Accession number
1	<i>P. furiosus</i>	β -glucosidase (CelB)	Q51723
		β -mannosidase (BmnA)	PF1208*
		β -glucosidase (BglA)	Q9HHB3
		β -mannosidase (BglB)	Q8U3U9
	<i>P. horikoshii</i>	alkyl β -glycosidase (BgPh)	O58104
		β -mannosidase (BglB)	O58237
	<i>P. abyssi</i>	β -mannosidase	Q9UYG5
		β -mannosidase	Q9V058
	<i>P. woesei</i>	β -galactosidase	O52629
5	<i>P. horikoshii</i>	Endo-glucanase (EglB)	O58925
	<i>P. abyssi</i>	Endo-glucanase (EglB)	Q9V052
12	<i>P. furiosus</i>	Endo-glucanase (EglA)	Q9V2T0
13	<i>P. furiosus</i>	α -amylase/ cyclomaltodextrinase	Q8TZP8
		α -amylase (AmyA)	Q8U3I9
	<i>P. woesei</i>	α -amylase (AmyA)	Q9P9L0
16	<i>P. furiosus</i>	Laminarinase (LamA)	O73951
18	<i>P. furiosus</i>	ChitinaseA	Q8U1H4
		ChitinaseB	Q8U1H5
35	<i>P. furiosus</i>	β -galactosidase (BglC)	Q8U3U2
	<i>P. horikoshii</i>	β -galactosidase (BglC)	O58247
	<i>P. abyssi</i>	β -galactosidase (BglC)	Q9UYH2
38	<i>P. horikoshii</i>	α -mannosidase	O58565
57	<i>P. furiosus</i>	4- α -glucanotransferase (AmyA)	P49067
		Amylopullulanase (Apu)	Q8TZQ1
		α -galactosidase (GalA)	Q9HHB5
		α -amylase	PF1393*
	<i>P. horikoshii</i>	α -amylase (AmyA)	O57932
		α -galactosidase	O58106
		α -amylase	PH1386*
	<i>P. abyssi</i>	α -amylase (AmyA)	Q9V298
		amylopullulanase	Q9V294
		α -amylase	PAB1857*
65	<i>P. horikoshii</i>	Trehalase	O58512
?	<i>P. furiosus</i>	α -glucosidase/invertase	PF0132*

*gene number

Gene Cloning

A variety of methods have been used for the isolation of GH-encoding genes from *P. furiosus* and related species. In the absence of either orthologs or paralogs, a classical reversed genetics approach may be used. This is realized through biochemical enrichment of the activity, eventually resulting in purification and the N-terminal sequencing of the enzyme of interest. Subsequently,

oligonucleotide primers can be designed for screening phage- or plasmid-based genomic libraries, as have been used for the identification of *P. furiosus* *celB* and *lamA* genes [6, 8]. Alternatively, a genomic library is constructed in a heterologous host (usually an *Escherichia coli* strain) that is plated on solid media and screened for selected activity. This approach has been applied for the isolation of CelB- and BmnA-encoding genes [2]. When potential homologous sequences are available, degenerate primers can be designed for cloning the counterpart by performing polymerase chain reaction (PCR) on genomic DNA from the organism of interest. The latter method has been used to screen 21 isolates that belong to the order Thermococcales, using primers that were based on conserved regions of thermostable family 1 β GHs. As a result, a number of homologous fragments were obtained that appear to form a subclass of GH family 1 (BglB) [9].

At present, the genome sequences of 29 archaea have been completed and 56 archaeal genome projects are ongoing (<http://www.genomesonline.org/gold.cgi>); three genomes are available of the genus *Pyrococcus*: *P. horikoshii*, *P. abyssi*, and *P. furiosus*.⁶ This genomics era has led to a spectacular increase in the release of sequence information in public databases and has subsequently resulted in the identification of many more GHs in *Pyrococcus* spp. (GalA; Table 1.1) [10, 11].

Enzyme Overproduction

Many of the genes that encode the *P. furiosus* GHs listed in Table 1.1 have been isolated, and attempts have been made to overproduce them in *E. coli* using either pUC-derived vectors [12] or the T7 expression system [13]. Stable overproduction of the *P. furiosus* β -glucosidase (CelB), β -mannosidase (BmnA), and laminarinase (LamA) is obtained when their genes are amplified by PCR using *Pfu* DNA polymerase and as *NcoI*-*Bam*HI (*celB*, *lamA*) or *Bsp*HI-*Blp*I (*bmnA*) fragments fused translationally to the T7 promoter of the pET9d vector (Novagen), which carries a kanamycin resistance (Kan^R) gene. The resulting plasmids are transformed to *E. coli* BL21(DE3), in which the gene coding for T7-RNA polymerase has been integrated in the chromosome under control of the *lacUV5* promoter [13]. In contrast to relatively instable derivatives of plasmids with ampicillin resistance as the selective marker, enzyme production of the three pyrococcal β GHs encoded by pET9d-derived vectors is stable in culture volumes of 1-8 liter. Optimal expression (50 mg/liter in rotating Erlenmeyer or 90 mg/liter in fermentor) is observed when cultures are grown overnight, without the need of isopropyl- β -D-thiogalactopyranoside (IPTG) induction [9]. A drawback of an expression system based on the strong T7 promoter may be the formation of inclusion bodies, as was observed in the expression of the *P. furiosus* genes *eglA*, *bglA*, and *P. horikoshii* *bglB*, of which the majority of the protein (75-95%) is inactive and insoluble. In the case of the *P. furiosus* *bglA* gene (coding for a β -glycosidase) and *P. horikoshii* *bglB* gene (coding for a β -mannosidase), a high level of expression is observed as well, but only 5% of the recombinant protein is found in the soluble fraction of the cell-free extract. Optimization might be accomplished by (i) cultivation at suboptimal conditions (e.g., lower temperature; lower aeration rate; minimal

medium), [14] (ii) use of a low-copy T7 expression variant, [15] (iii) a different vector, such as high copy (pUC-derivatives) or low copy (pBR322 derivatives), with the gene of interest under control of a variety of inducible promoters (lactose, tryptophan/lactose, arabinose, rhamnose), [16-19] and (iv) a different host/vector system.

For the application of enzymes in food, it is necessary to produce them in a food-grade manner, i.e., nontoxic and safe. The lactic acid bacterium *Lactococcus lactis* has a long history of safe use as a starter for industrial dairy fermentation and can be employed for the food-grade production of enzymes. A nisin-controlled expression (NICE) system has been developed for this organism [20-22]. In a pilot experiment CelB was cloned in *L. lactis*, and heterologous expression was achieved. β -glucosidase activity was detected in cell extracts and the protein was revealed as a major protein band by SDS-PAGE analysis. Under optimal conditions, CelB made up 2% of the total cell protein (not shown).

Protein Purification

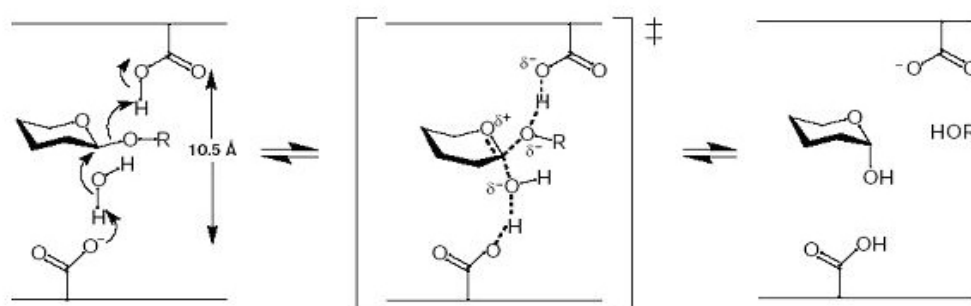
When the enzymes are produced in mesophilic hosts such as *Escherichia coli*, they can be purified easily by employing their superior thermostability: heat incubation followed by centrifugation usually results in at least 90% pure protein. In general, subsequent purification is performed by ion-exchange chromatography and by either gel filtration or hydrophobic interaction chromatography. LamA has been produced in a 1-liter culture of *E. coli* BL21(DE3) cells harboring pLUW532 (*lamA* cloned in pET9d) in LB medium (16-20 hr). After harvesting, the LamA protein has been purified from *E. coli* by resuspending the cell paste in 20 ml 150 mM sodium-citrate buffer (pH 5.0) and a threefold passage through a French Press (100 MPa). The cell-free extract is subjected to a heat incubation (45 min at 80°C), after which the precipitated proteins are removed by centrifugation (10,000g, 10 min). A heat incubation in citrate buffer is significantly more efficient than one in a Tris buffer due to stabilizing properties of the latter (see below, Thermal Inactivation). The resulting heat-stable, cell-free extract is dialyzed against 20 mM Tris-HCl (pH 8.0) and is loaded on an anion-exchange column (Q-Sepharose Fast Flow; Äkta-FPLC; Amersham Biosciences) equilibrated with the same buffer. During a linear gradient of NaCl (0.0-1.0 M, in the same Tris buffer), the active fractions (as determined by the DNS Method, see later) eluted around 0.5 M NaCl. Pooled LamA-containing fractions are concentrated (Centricon concentrator, 10-kDa cut off; Amicon), and 0.5 ml is loaded on a gel filtration column (Superdex 200, Amersham Biosciences), which removes traces of contaminating DNA. On this column, LamA migrates as a 31-kDa monomer. The *P. furiosus* enzymes CelB and BmnA could be produced in the same T7 expression system and are purified similarly. In a larger scale CelB or LamA production (8-liter fermentor), the gel filtration step in the purification method is replaced by phenyl-Sepharose chromatography. The contaminating DNA appears in the flow through, whereas CelB elutes from the column at the end of a $(\text{NH}_4)_2\text{SO}_4$ gradient (1.0-0.0 M) [23].

Reaction Mechanism of Glycoside Hydrolases

Even though the variety of glycoside hydrolases is enormous, and also the three-dimensional structures reveal a remarkable structural diversity (14 clans with at least 7 different folds; http://afmb.cnrs-mrs.fr/CAZY/fam/acc_GH.html), hydrolysis of the glycosidic bond occurs only by one of two mechanisms, either with retention or inversion of the anomeric configuration [24, 25].

Both mechanisms employ a pair of carboxylic acids in the active site. Inverting enzymes use one residue as a general acid and the other as a general base catalyst, and are suitably placed, about 10.5 Å apart, to allow both a substrate and a water molecule between them (Fig. 1.1A). In retaining enzymes, on the other hand, one residue acts as a nucleophile and the other as a general acid/base catalyst and are only 5.5 Å apart (Fig. 1.1B).

A



B

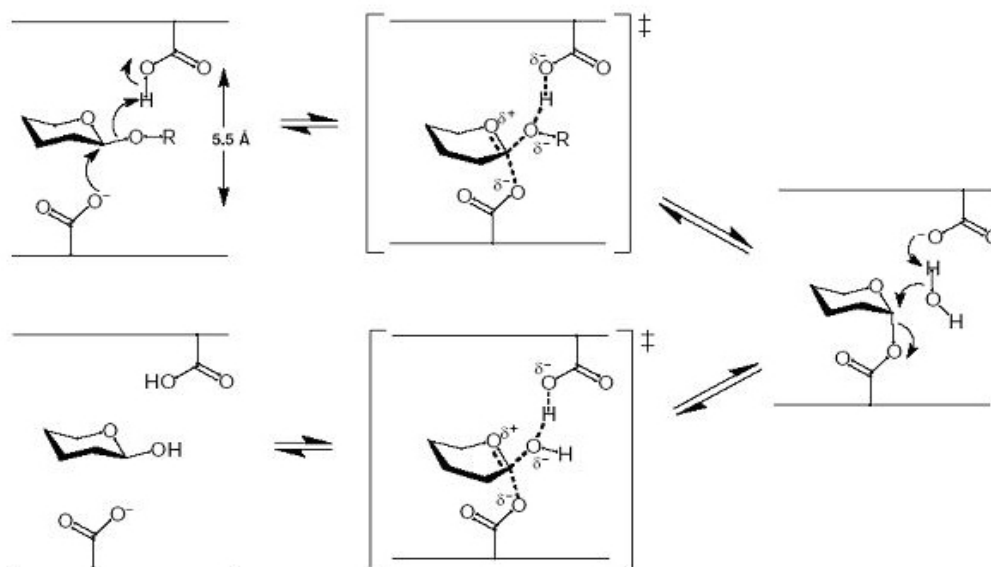


Figure 1.1. Mechanism of hydrolysis by (A) inverting and (B) retaining glycoside hydrolases after http://afmb.cnrs-mrs.fr/CAZY/fam/acc_GH.html

The inverting mechanism proceeds via a single displacement involving an oxocarbenium ion-like transition state. By contrast, the reaction by retaining enzymes proceeds via a double displacement mechanism involving a covalent glycosyl-enzyme intermediate. First, one of the carboxylic residues functions as a general acid catalyst by protonating the glycosidic oxygen. The other carboxylic residue acts as a nucleophile, by attacking the C1 atom of the glycosidic bond and forming a covalent glycosyl-enzyme intermediate. In the next step, the carboxyl group that first acted as an acid catalyst, now acts as a general base catalyst by deprotonating an incoming water molecule which attacks at the anomeric center resulting in the release of the sugar molecule. For a chitinase it has been demonstrated that the acetamido group at C-2 of the substrate can take over the role of catalytic nucleophile, and thus shows substrate assisted catalysis [26]. Furthermore, a distinct catalytic mechanism has been shown for a family of glycoside hydrolases in which NAD⁺ is used as a cofactor [27].

Whereas in thermodynamically controlled reactions reverse hydrolysis is possible with both types of enzymes, only with retaining enzymes transglycosylation can be achieved in kinetically controlled reactions, because of formation of the glycosyl-enzyme intermediate [25].

Activity Assay

A variety of methods are available for GH activity measurements, depending on the substrate specificity as well as the level of purification [colony on agar plate, lysate on polyacrylamide gel, or (partially) purified in solution] (Table 1.2). This section describes different techniques based on either artificial chromogenic substrates or "natural" sugar compounds.

Table 1.2. Activity assays of exo- and endo-glycoside hydrolases. X-gly, 5-bromo-4-chloro-3-indole-glycoside; NP-gly, ortho/para-nitrophenol-glycopyranoside; Mb-gly, methylumbelliferyl-glycoside; glu-ox., glucose-oxidase; DNS, dinitrosalicylic acid.

<i>medium</i>	<i>chromogenic substrates</i>	<i>oligo/di-saccharides</i>	<i>poly-saccharides</i>
agar plates	X-gly		congo red
solution	NP-gly	HPLC, glu-ox.	DNS, HPLC
native PAGE	Mb-gly		congo red

X-Glycosides

The dye 5-bromo-4-chloro-3-indole oxidizes and precipitates as a blue complex when free in solution. This dye is the coloring agent in all the X-linked substrates commonly used for the detection of exoglycosidase activity in bacterial colonies on agar plates. The well-known blue-white screening system for *E. coli*, based on cloning vectors with "α-complementation" of the

truncated β -galactosidase LacZ, makes use of the hydrolysis of X- β -Dgalactopyranoside (X-Gal) to discriminate between transformant colonies containing a plasmid with an insert (often white phenotype) or without an insert (blue phenotype). Several derivatives of the X dye with different colors are available, enabling simultaneous detection of different glycosidase activities (Biosynth, Staad, Switzerland), e.g., in the case of *P. furiosus* β -glucosidase CelB and *E. coli* β -glucuronidase GusA [28]. The *P. furiosus* β -glucosidase (CelB) has its optimum temperature for activity at 100 °C. Nevertheless, colonies of the LacZ-deficient host *E. coli* JM109(DE3) that are transformed with a *celB* expression vector (see earlier discussion) turn blue when grown at 37 °C on agar plates containing X- β -D-glucopyranoside [X-Glu, 0.16% (w/v)]. This feature has been used to construct a library of *E. coli* cells that contains expression vectors with randomly mutagenized *celB* genes [23]. In contrast to *E. coli* LacZ (GH family 2), the *P. furiosus* CelB has a broad substrate specificity, capable of hydrolyzing X-Glu, X-Gal, and, to a lesser extent, X- β -xyloside. BmnA has been characterized as a β -mannosidase, but when produced in *E. coli*, colonies turn blue on X-Glu- or X-Gal-containing agar plates, which is in agreement with the generally broad substrate specificity of family 1 GHs [2, 9]. Because LacZ specifically hydrolyzes X-Gal, even *E. coli* strains that have an intact *lacZ* gene, such as BL21(DE3), can be screened for functional β -glucosidase activity by culturing on agar plates containing X-Glu. When selecting for mutant enzymes with altered hydrolytic characteristics, however, the screening of glycosidase activity on X substrate-containing agar plates by analysis of the color intensity should not be regarded as a quantitative measure of mutant enzyme activity. In the case of *celB* variants, blue color formation after overnight growth in general corresponds to initial enzyme activity, but discrepancies have been noted for a significant number of clones [23].

Nitrophenol Glycosides

Nitrophenol is a commonly used chromogen in kinetic measurements with hydrolytic enzymes such as glycosidases. Many different mono- and oligoglycosides are available as *o*-nitrophenyl (*o*Np) and *p*-nitrophenyl (*p*Np) derivatives (Sigma). The substrates can be used in either discontinuous or continuous assays at the elevated temperatures at which thermostable glycosidases operate. The *P. furiosus* β -glucosidase (CelB) activity has been analyzed at 90 °C using both assays. Discontinuous and continuous assays are performed in a 150 mM sodium citrate buffer of pH 5.0 (set at room temperature; theoretically, the change in pH of the citrate buffer at 90 °C is expected to be less than 0.2 units [29]), the pH at which CelB has its optimum for activity. In the discontinuous assay, 495 μ l citrate buffer with 10 mM *p*Np- β -D-glucopyranoside (*p*Np-glu) in a 1.5-ml Eppendorf reaction vial is pre-heated for at least 2 min at 90 °C, after which the reaction is started by the addition of 5 μ l enzyme solution. After a defined period of time, usually 15 min, the reaction is stopped by the addition of 1.0 ml ice-cold 0.5 M Na₂CO₃. This causes the pH to rise to about 9-10, which terminates the reaction and enhances the specific absorption coefficient of the liberated nitrophenol (see below). The absorption of the reaction mixture is measured at 405 nm. The measured activity

should always be corrected for nonenzymatic hydrolysis of the substrate. The discontinuous assay has been mainly used in assays with a fixed substrate concentration, such as pH optima determination, qualitative assays in protein purification, and thermostability experiments. The continuous assay has been used for the determination of kinetic parameters of *P. furiosus* CelB. The liberation of nitrophenol in the reaction mixture is directly followed at 405 nm using a UV/VIS spectrophotometer equipped with a temperature controller. After at least 2 min of preheating 995 μ l buffer with *p*Np-Glu (0.05–10 mM) at 90 °C in a quartz cuvette, the nonenzymatic hydrolysis of the substrate is recorded for about 2 min. Then 5 μ l of enzyme is added and the increase of absorption is monitored for about 3 min. The activity is determined from the initial increase in absorbance and is expressed as units per milligram, in which 1 unit (U) is defined as 1 μ mol nitrophenol released per minute. It should be emphasized that the specific absorption coefficient (ϵ) of nitrophenol depends on both pH and temperature and that ϵ is different for *o*Np and *p*Np. Therefore, ϵ should be determined for each nitrophenol derivative under the assay condition used. As an example, the specific absorption coefficients for *p*Np at 90 °C have been determined at different pH values (Fig. 1.2A). When the pH is raised from 4 to 8, the specific absorption coefficient for *p*Np increases more than 100-fold. This should be taken into account when determining pH optima for activity. For example, the pH optimum for the *P. furiosus* β -mannosidase BmnA has been determined with and without correction of the specific absorption coefficient. Without correction the pH optimum appeared to be at a pH value that is about 2 units higher than the actual optimum (Fig. 1.2B).

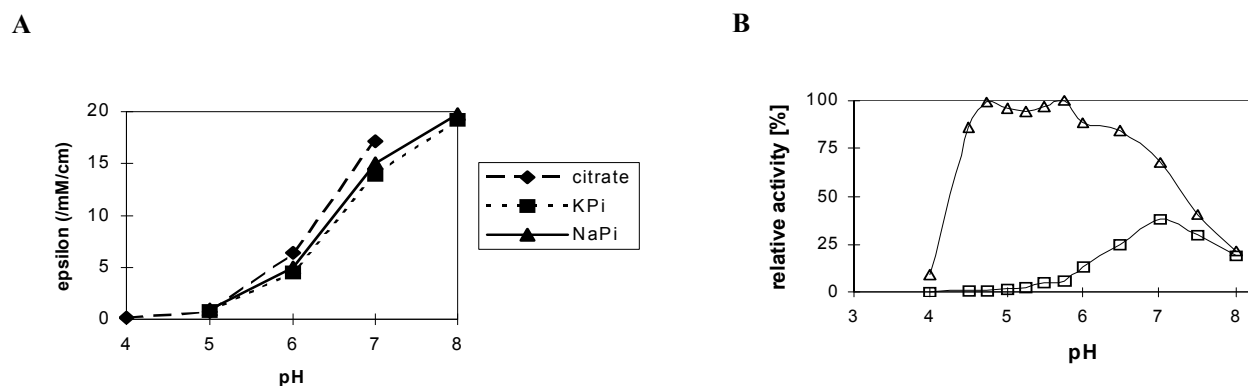


Figure 1.2. (A) Specific absorption coefficients (ϵ) for para-nitrophenol (pNp) at 405 nm in sodium-citrate, KPi and NaPi buffer (100 mM each) at different pH values at 90°C. (B) pH optimum curve for BmnA for the hydrolysis of pNp-man with correction (triangles) and without correction (squares) for pH.

Methylumbelliferylglycosides

Whereas 5-bromo-4-chloro-3-indole (X) and nitrophenol (Np) develop a color on hydrolysis, free 4-methylumbelliferone emits light with an optimum at 460 nm (fluorescent blue) after excitation with UV light. This makes the dye more sensitive in activity staining in native PAGE gels compared to X substrates and nitrophenol substrates. Again, mono- and oligoglycosides are

available as conjugates with 4-methylumbelliferone (Biosynth). To detect β -glucosidase activity in *P. furiosus* cell free extracts, a sample is loaded and run on a native PAGE gel. The gel (1.5 ml) is soaked in 10 ml (1.7 g/liter) 4-methylumbelliferyl- β -D-glucopyranoside, preheated at 90 °C. After 15 min of incubation at 90 °C, the gel is analyzed under UV light (standard DNA visualizing equipment), and activity is visible as bright spots on the gel.

Disaccharides

For reasons of convenience, the activity of exo-glycosidases is often measured using several artificial substrates, but it is important to keep in mind that the hydrolysis of physiological substrates (di- and oligosaccharides) can make quite a difference (see below). The activities of β -glucosidase CelB and β -mannosidase BmnA for the disaccharides cellobiose and lactose are determined by incubation at 90 °C in 150 mM sodium citrate buffer (pH 5.0) with different concentrations of lactose or cellobiose. Lactose and cellobiose could be dissolved to at least 250 and 500 mM, respectively. For the detection of free glucose in solution, several glucose oxidase kits are available commercially, which can be used to measure the activity on glucose- containing disaccharides, such as maltose, cellobiose, laminaribiose, lactose, and sucrose, or exoglucosidase activity on glucose polymers, such as starch, pullulan, cellulose, laminarin, or lichenan. In Eppendorf tubes, 495 μ l of 0-300 mM lactose in 150 mM sodium citrate buffer (pH 5.0) is pre-heated for 5 min at 90 °C. Next, 5 μ l of a CelB or BmnA dilution is added and after 15 min the reaction is stopped by cooling on ice. A 10- μ l sample of the assay mix is transferred to microtiter plate wells, which each contain 200 μ l of the Peridochrom glucose GOD-GAP kit (Boehringer). After 30 min, the developed color is analyzed at 492 nm in an Elisa reader (e.g. ATTC, SLT lab instruments). A standard series of known glucose concentrations (typically 0-10 mM) is always included and used to calculate activities. The signals of lactose, cellobiose, and galactose are negligible in this assay. The kinetic parameters for the hydrolysis of cellobiose and lactose are compared to those of their chromogenic counterparts *p*Np-Glu and *p*Np-Gal. In both cases, the efficiencies of CelB for the hydrolysis of the disaccharides are about 40 to 80 times lower compared to the *p*Np substrates. This was mainly because the affinity of CelB for cellobiose and lactose is lower (high *K_m*, Table 1.3). For BmnA, the difference in efficiency is smaller, as the affinity of BmnA for *p*Np-Glu and *p*Np-Gal is just a little higher than for cellobiose and lactose (2- to 15-fold difference). Hence, the activity on chromogenic substrates gives only an approximate indication for the enzyme's hydrolytic efficiency of natural substrates.

Table 1.3. Kinetic parameters of CelB and BmnA at 90°C on pNp-glu, pNp-gal, pNp-man, cellobiose and lactose [9, 47]

		CelB	BmnA
pNp-glu	v_{\max} (U/mg)	1800	486
	K_m (mM)	0.42	69
	k_{cat}/K_m ($\text{mM}^{-1}\text{s}^{-1}$)	3900	6.9
pNp-gal	v_{\max} (U/mg)	2600	496
	K_m (mM)	5.0	29
	k_{cat}/K_m ($\text{mM}^{-1}\text{s}^{-1}$)	480	16.8
pNp-man	v_{\max} (U/mg)	72.4	190
	K_m (mM)	1.3	0.99
	k_{cat}/K_m ($\text{mM}^{-1}\text{s}^{-1}$)	54.7	198
cellobiose	v_{\max} (U/mg)	720	260
	K_m (mM)	14	100
	k_{cat}/K_m ($\text{mM}^{-1}\text{s}^{-1}$)	48	2.6
lactose	v_{\max} (U/mg)	1500	101
	K_m (mM)	120	89
	k_{cat}/K_m ($\text{mM}^{-1}\text{s}^{-1}$)	11	1.1

Congo Red

Hydrolytic enzyme activity toward β -linked sugar polymers such as laminarin (β -1,3), lichenan (β -1,3-1,4), and carboxymethyl-cellulose (CMC; β -1,4) can be visualized qualitatively using Congo Red. This red dye will attach to the polysaccharide and enzyme activity can be detected by the appearance of clearing zones ("halo"). For the detection of *P. furiosus* endo- β -1,3-glucanase (laminarinase, LamA), the assay is carried out on (granulated) agar plates that, just before solidifying (50°C), are mixed with a sugar polymer solution (final concentration 3 g/liter) in 50 mM phosphate buffer (pH 6.0). CMC can be added in concentrations up to 10 g/liter (final concentration) due to a higher solubility. Subsequently, 5 μ l of enzyme extract is applied onto the plate and incubated at 55 °C for 2 hrs. Enzyme activity is visualized by staining the plates with 1 g/liter Congo Red in deionized water for 30 min and destaining twice with 2 M NaCl for 10 min. For a sharper contrast of the clearing zones, 1 M HCl is added after destaining, which causes a color shift from deep red to dark blue. Because of differences in substrate specificity, it is possible to detect independently endo- β -1,3-glucanase LamA activity and endo- β -1,4-glucanase EglA activity in concentrated culture supernatant or resuspended cell debris of *P. furiosus* cultures. Laminarin is only hydrolyzed by LamA, whereas EglA activity could be detected using CMC [10].

DNS Method

The action of endo-glucanases on sugar polymers results in an increase of sugar reducing ends. The accumulation of sugars with reducing ends can be determined colorimetrically by the addition of dinitrosalicylic acid (DNS), which results in a change of color from yellow toward brown and can be detected spectrophotometrically [30]. The DNS reagent (per 100 ml: 1.6 g NaOH, 1 g 3,5-dinitrosalicylic acid, 30g potassium sodium tartrate, 0.2 g phenol, 0.05 NaSO₃) should be stored at room temperature and protected from light. For *P. furiosus* LamA, a 5-mg/ml substrate solution of laminarin and lichenan, respectively, is made in a 50 mM phosphate buffer of pH 6.0. Enzyme is added and the solution is incubated at 80 °C for 30 min and subsequently transferred to 0 °C. A 100- μ l aliquot of the enzyme reaction is added to 100 μ l DNS reagent. As a standard, 100 μ l of a glucose dilution series is taken and assayed similarly. The samples are boiled for 5 min, after which 1 ml of MQ water is added. Color development is determined spectrophotometrically at 575 nm.

HPLC

High-performance liquid chromatography (HPLC) can be used to monitor reactions of glycoside hydrolases both qualitatively and quantitatively. Depending on the column that is used, saccharides with a different degree of polymerization (*DP* 1-20) are resolved. Because glycoside hydrolases change the degree of polymerization of saccharides, HPLC is well suited to observe not only changes in the amount of substrates and products, but also the range/variation of products. The β -glucosidase CelB and β -glucanase LamA were tested for their single and combined ability to hydrolyze the β -1,3-linked glucose polymer laminarin. A 0.5% (w/v) solution of laminarin is incubated without enzyme, with 5 μ g/ml CelB, with 5 μ g/ml LamA, and with both 5 μ g/ml CelB and 5 μ g/ml LamA for 4 hr at 80 °C. The reactions are stopped by adding H₂SO₄ up to a final concentration of 0.1 M. This causes the pH to lower to approximately pH 2, which terminates the reactions and makes the mixtures suitable for loading on the HPLC column. Samples are run on an Aminex HPX-87-H column (Bio-Rad) at a temperature of 30 °C with a flow of 0.6 ml/min in 0.01 M H₂SO₄. Eluates are analyzed by a refractive index detector. By calculating the relative peak areas in the chromatogram, the amount of product formed and substrate converted are determined (Fig. 1.3) [6].

Protein stability

To be functionally active at high temperatures glycoside hydrolases from *P. furiosus* have to maintain their native folded form. Generally they display tremendous stability towards extreme thermal and chemical conditions (CelB; LamA; GalA) [7, 11, 31]. Even when they are heterologously produced in mesophilic hosts this stability is demonstrated. This implies that

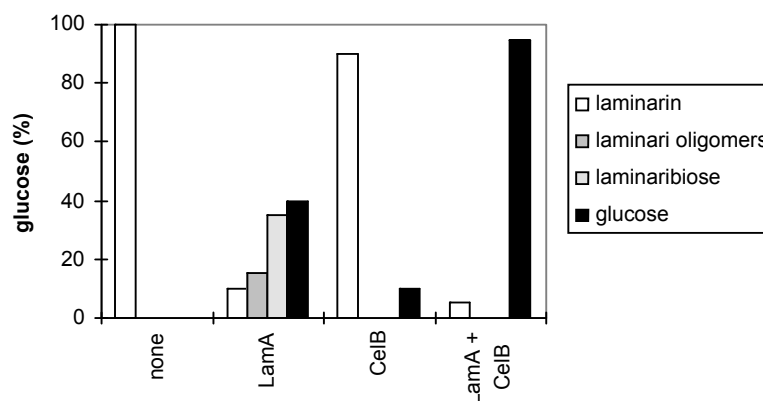


Figure 1.3. Action of LamA and CelB on laminarin (0.5% (w/v)). Laminarin was incubated for 4 hrs. at 80°C with 5 µg/ml LamA and/or 5 µg/ml CelB. Samples were analysed by HPLC (modified after [6]).

stability is determined by the amino acid sequence. It has been shown that the residues Ala, Arg and Tyr occur more frequently in thermostable proteins than in their mesophilic homologs, whereas the appearance of the residues Asn, Gln and Cys is lower in the sequences of enzymes from thermophiles than in those from mesophiles [32-34]. Despite this difference in amino acid composition, high sequence similarities can be found between homologs from mesophiles and thermophiles, and in essence even three-dimensional structures are virtually identical. However, several small structural factors have been observed that contribute to a higher intrinsic thermostability, and usually a combination of some of them can be found in thermostable enzymes. Stabilizing features include reduced surface area, improved packing, hydrophobic interactions, improved subunit interactions, (networks of) ion pairs and incorporation of metal ions like Ca^{2+} , Co^{2+} , Mg^{2+} , Mn^{2+} [31, 35-39]. Strategies to increase the stability of less-stable proteins are based on the introduction of one or more of these stabilizing features [40-44].

Thermal Inactivation

Generally, thermostability of enzymes is indicated by a half life time of inactivation at a given temperature. In this section, a method is described to determine this parameter of thermostability of glycoside hydrolases from *P. furiosus*. Routinely, 200 µl of a 50-µg/ml CelB or LamA solution is transferred to glass vials closed with a screw cap with a Teflon inlay. The vials are then submerged in a silicon oil bath, set at 95 to 105 °C. At constant time intervals, a vial is transferred from the oil bath and cooled down. Residual activity is subsequently determined, e.g., with *p*Np-Glu or laminarin as a substrate using either the continuous or the discontinuous activity assay (see earlier discussion). This kind of inactivation usually displays a first-order relationship with time. Half-life values for thermal inactivation are obtained from data fits.

The chemical environment, such as buffer components, pH, and protein concentration, has a significant influence on the thermostability of enzymes. The buffers Tris

[tris(hydroxymethyl)aminomethane], PIPES (1,4-piperazine- diethanesulfonic acid), and imidazole have a stabilizing effect on β -glucosidase CelB [45]. The effect of pH on the thermostability of enzymes has been investigated with *P. furiosus* β -glucosidase BglA. The enzyme is incubated at 100 °C in McIlvain's buffer with different pH values (pH 4-8). Optimal stability is observed at pH 6 (Fig. 1.4), which is close to the optimal pH for activity (pH 5.0) [14]. The irreversible loss of activity due to prolonged incubation at high temperatures is most likely caused by the irreversible unfolding of protein; however, inactivation might also be the result of a limited, local effect. Thermal inactivation experiments are, therefore, only an indication of the conformational stability of a protein. Usually, the unfolding of peptide chains by high temperature or incubation in guanidinium hydrochloride or urea is evaluated by analysis of the fluorescence spectrum or circular dichroism absorption [46].

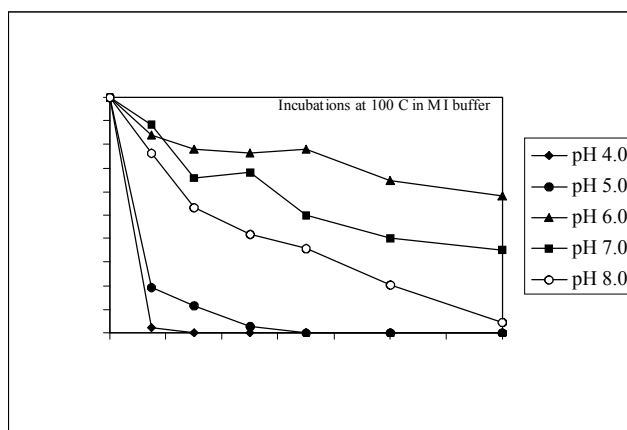


Figure 1.4. Thermal stability of BglA at 100°C in McIlvain-buffer at different pH values. Residual activity was determined with the continuous assay at 90°C using pNp-glu.

Concluding Remarks

At present, many sugar-converting enzymes involved in the hydrolysis of either α - or β -linked sugar substrates have been characterized genetically and biochemically, as well as biophysically. An obvious next step is to gain knowledge concerning the molecular basis that determines the substrate specificity as well as extreme thermal and chemical stability of this class of enzymes. Two distinct approaches might give rise to a higher level of understanding the different features of GHs: (i) screening for natural variants and (ii) optimization of currently available enzymes. The latter approach can be divided into directed and random engineering. Currently, several *P. furiosus* GHs are being engineered with a directed mutagenesis approach, a random method of "molecular evolution," [23, 47] or a combination thereof. The obvious goal is to optimize the GHs in such a way that they efficiently catalyze the conversion of a specific substrate at desired physical conditions.

References

1. Niehaus F., Bertoldo C., Kähler M., and Antranikian G. (1999) Extremophiles as a source of novel enzymes for industrial application. *Appl. Microbiol. Biotechnol.* **51**, 711-729.
2. Bauer M.W., Bylina E.J., Swanson R.V., and Kelly R.M. (1996) Comparison of a beta-glucosidase and a beta-mannosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Purification, characterization, gene cloning, and sequence analysis. *J. Biol. Chem.* **271**, 23749-23755.
3. Henrissat B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309-316.
4. Coutinho P.M. and Henrissat B. (1999) Carbohydrate-active enzymes: an integrated database approach. In "Recent Advances in Carbohydrate Bioengineering", H.J. Gilbert, G. Davies, B. Henrissat and B. Svensson eds., The Royal Society of Chemistry, Cambridge, pp. 3-12.
5. Fiala G. and Stetter K. O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* **161**, 168-175.
6. Gueguen Y., Voorhorst W. G. B., van der Oost J., and de Vos W. M. (1997) Molecular and biochemical characterization of an endo-beta-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **272**, 31258-31264.
7. Kengen S. W., Luesink E. J., Stams A. J., and Zehnder A. J. (1993) Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur. J. Biochem.* **213**, 305-312.
8. Voorhorst W.G., Eggen R.I., Luesink E.J., and de Vos W.M. (1995) Characterization of the *celB* gene coding for beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*. *J. Bacteriol.* **177**, 7105-7111.
9. Kaper T. and van der Oost J. Unpublished results.
10. Driskill L.E., Kusy K., Bauer M.W., Kelly R.M. (1999) Relationship between glycosyl hydrolase inventory and growth physiology of the hyperthermophile *Pyrococcus furiosus* on carbohydrate-based media. *Appl. Environ. Microbiol.* **65**, 893-7.
11. Van Lieshout J.F.T., Verhees C.H., Ettema T.J., van der Sar S., Imamura H., Matsuzawa H., van der Oost J., de Vos W.M. (2003) Identification and molecular characterization of a novel type of alpha-galactosidase from *Pyrococcus furiosus*. *Biocatalysis and Biotransformation* **21**, 243-252.
12. Yanisch-Perron C., Vieira J., and Messing J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.
13. Studier F.W., Rosenberg A.H., Dunn J.J., and Dubendorff J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185**, 60-89.
14. Verhees C. and van der Oost J. Unpublished results.
15. Ward D.E., Ross R.P., Van der Weijden C.C., Snoep J.L., and Claiborne A. (1999) Catabolism of branched-chain alpha-keto acids in *Enterococcus faecalis*: the *bkd* gene cluster, enzymes, and metabolic route. *J. Bact.* **181**, 5433-5442.
16. Fuller F. (1982) A family of cloning vectors containing the lacUV5 promoter. *Gene* **19**, 43-54.
17. Amann E., Brosius J., and Ptashne M. (1983) Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*. *Gene* **25**, 167-178.
18. Guzman L.M., Belin D., Carson M.J., and Beckwith J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121-4130.
19. Krebsfänger N., Zocher F., Altenbuchner J., and Bornscheuer U.T. (1998) characterization and enantioselectivity of a recombinant esterase from *Pseudomonas fluorescens*. *Enzyme Microbial. Technol.* **22**, 641-646.
20. de Ruyter P.G., Kuipers O.P., and de Vos, W.M. (1996) Controlled gene expression systems for *Lactococcus lactis* with the food-grade inducer nisin. *Appl. Environ. Microbiol.* **62**, 3662-3667.

21. Kuipers O.P., de Ruyter P.G., Kleerebezem M., and de Vos, W.M. (1997) Controlled overproduction of proteins by lactic acid bacteria. *Trends Biotechnol.* **15**, 135-40.
22. de Vos W.M., Hols P., van Kranenburg R., Luesink E., Kuipers O.P., van der Oost J., Kleerebezem M., and Hugenholtz, J. (1998) Making More of Milk Sugar by Engineering Lactic Acid Bacteria. *Int. Dairy Journal* **8**, 227-233.
23. Lebbink J.H., Kaper T., Kengen S.W., van der Oost J., and de Vos W.M. (2001) beta-Glucosidase CelB from *Pyrococcus furiosus*: production by *Escherichia coli*, purification, and in vitro evolution. *Methods Enzymol.* **330**, 364-379.
24. McCarter J. D. and Withers S. G. (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885-892.
25. Sinnott M. L. (1990) Catalytic mechanisms of glycosyl transfer. *Chem. Rev.* **90**, 1171-1202.
26. Terwisscha van Scheltinga A., Armand S., Kalk K.H., Isogai A., Henrissat B., and Dijkstra B.W. (1995) Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and x-ray structure of a complex with allosamidin. Evidence for substrate assisted catalysis. *Biochemistry* **34**, 15619-15623.
27. Rajan S.S., Yang X., Collart F., Yip V.L., Withers S.G., Varrot A., Thompson J., Davies G.J., and Anderson W.F. (2004) Novel catalytic mechanism of glycoside hydrolysis based on the structure of an NAD⁺/Mn²⁺ - dependent phospho-alpha-glucosidase from *Bacillus subtilis*. *Structure.* **12**, 1619-1629.
28. Sessitsch A., Wilson K.J., Akkermans A.D., and de Vos W.M. (1996) Simultaneous detection of different *Rhizobium* strains marked with either the *Escherichia coli gusA* gene or the *Pyrococcus furiosus celB* gene. *Appl. Environ. Microbiol.* **62**, 4191-4194.
29. Bates R.G. and Paabo M. (1970) In "Handbook of Biochemistry: Selected Data for Molecular Biology" (H.A. Sober, ed.) The Chemical Rubber Company, Cleveland, Ohio.
30. Sumner J. B. and Somers G. F. (1949) Dinitrosalicylic method for glucose. In *Laboratory experiments in biological chemistry* (Sumner, J. B. & Somers, G. F., eds), pp. 38-39. Academic Press, New York.
31. Chiaraluce R., van der Oost J., Lebbink J.H., Kaper T., and Consalvi V. (2002) Persistence of tertiary structure in 7.9 M guanidinium chloride : the case of endo-beta-1,3-glucanase from *Pyrococcus furiosus*. *Biochemistry* **41**, 14624-14632.
32. Jaenicke R. (2000) Stability and stabilization of globular proteins in solution. *J. Biotechnol.* **79**, 193-203.
33. Kumar S., Tsai C.J., and Nussinov R. (2000) Factors enhancing protein thermostability. *Protein Eng.* **13**, 179-191.
34. Vieille C., and Zeikus G.J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1-43.
35. Yip K. S., Britton K. L., Stillman T. J., Lebbink J., de Vos W. M., Robb F. T., Vetriani C., Maeder D., and Rice D. W. (1998) Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family. *Eur. J. Biochem.* **255**, 336-46.
36. Aguilar C.F., Sanderson I., Moracci M., Ciaramella M., Nucci R., Rossi M., and Pearl L.H. (1997) Crystal structure of the beta-glycosidase from the hyperthermophilic archeon *Sulfolobus solfataricus*: resilience as a key factor in thermostability. *J. Mol. Biol.* **271**, 789-802.
37. Lebbink J.H.G. (1999) Molecular characterization of the thermostability and catalytic properties of enzymes from hyperthermophiles. PhD Thesis. Wageningen University, Wageningen, The Netherlands
38. Chang C., Park B.C., Lee D.S., and Suh S.W. (1999) Crystal structures of thermostable xylose isomerases from *Thermus caldophilus* and *Thermus thermophilus*: possible structural determinants of thermostability. *J. Mol. Biol.* **288**, 623-34.
39. Vieille C., Epting K.L., Kelly R.M., and Zeikus J.G. (2001) Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase. *Eur. J. Biochem.* **268**, 6291-6301.
40. Van den Burg B., Vriend G., Veltman O.R., Venema G., and Eijsink V. G. (1998) Engineering an enzyme to resist boiling. *Proc. Natl. Acad. Sci. USA.* **95**, 2056-60.
41. Van den Burg B. and Eijsink V.G. (2002) Selection of mutations for increased protein stability. *Curr. Opin. Biotechnol.* **13**, 333-7.

42. Machius M., Declerck N., Huber R., and Wiegand G. (2003) Kinetic stabilization of *Bacillus licheniformis* alpha-amylase through introduction of hydrophobic residues at the surface. *J. Biol. Chem.* **278**, 11546-53.
43. Eijsink V.G., Bjork A., Gaseidnes S., Sirevag R., Synstad B., van den Burg B., and Vriend G. (2004) Rational engineering of enzyme stability. *J. Biotechnol.* **113**, 105-20.
44. Eijsink V.G., Gaseidnes S., Borchert T.V., and van den Burg B. (2005) Directed evolution of enzyme stability. *Biomol. Eng.* **22**, 21-30.
45. Kengen S.W.M., and Stams A.J.M. (1994) An extremely thermostable β -glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* : a comparison with other glycosidases. *Biocatalysis* **11**, 79-88.
46. Pouwels J., Moracci M., Cobucci-Ponzano B., Perugino G., van der Oost J., Kaper T., Lebbink J.H., de Vos W.M., Ciaramella M., and Rossi M. (2000) Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal beta-glycosidases. *Extremophiles*, **4**, 157-164.
47. Lebbink J.H., Kaper T., Bron P., Van der Oost J., and Vos W.M. (2000) Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* beta-glucosidase CelB by directed evolution. *Biochemistry* **39**, 3656-3665.

Chapter 2

Identification and Molecular Characterization of a Novel Type of α -galactosidase from *Pyrococcus furiosus*

Van Lieshout, J.F.T., Verhees, C.H., Ettema, T.J., Van der Sar, S., Imamura, H., Matsuzawa, H.,
Van der Oost, J., De Vos, W.M.

Biocatalysis and Biotransformation **21**, 243-252

Abstract

An α -galactosidase gene from *Pyrococcus furiosus* was identified, cloned and functionally expressed in *Escherichia coli*. It is the first α -galactosidase from a hyperthermophilic archaeon described to date. The gene encodes a unique amino acid sequence compared to other α -galactosidases. Highest homology was found with α -amylases classified in family 57 of glycoside hydrolases. The 364 amino acid protein had a calculated mass of 41.6 kDa. The recombinant α -galactosidase specifically catalyzed the hydrolysis of *para*-nitrophenyl- α -1,4-D-galactopyranoside, and to some extent that of melibiose and raffinose. The enzyme showed to be an extremely thermo-active and thermostable α -galactosidase with a temperature optimum of 115 °C and a half-life time of 15 hours at 100 °C. The pH optimum is between 5.0 and 5.5. Sequence analysis showed four conserved carboxylic residues. Site-directed mutagenesis was applied to identify the potential catalytic residues. Glu117Ala showed decreased enzyme activity, which could be rescued by the addition of azide or formate. It is concluded that glutamate 117 is the catalytic nucleophile, whereas the acid/base catalyst remains to be identified.

Introduction

Pyrococcus furiosus is a heterotrophic hyperthermophilic archaeon that can utilize carbohydrates as a carbon and energy source. *P. furiosus* has initially been shown to grow on α -linked glucosidic substrates like starch and maltose (Fiala and Stetter, 1986). Later it was demonstrated that growth also occurred on β -glucosides like cellobiose, laminarin and lichenan (Kengen *et al.*, 1993, Gueguen *et al.*, 1997). To date, no information is available on the capacity of *P. furiosus* to ferment non-glucose (poly)saccharides. The saccharolytic capacity of an organism is reflected by its production of glycoside hydrolases. The enzymes described to date comprise two α -amylases (Laderman *et al.*, 1993a, Laderman *et al.*, 1993b, Jorgensen *et al.*, 1997, Dong *et al.*, 1997a), an amylopullulanase (Dong *et al.*, 1997b), an α -glucosidase (Costantino *et al.*, 1990), two endo- β -glucanases (Gueguen *et al.*, 1997, Bauer *et al.*, 1999) and four β -glycosidases with diverse/broad substrate specificities (Bauer *et al.*, 1996, Kaper *et al.*, 2001, Kengen *et al.*, 1993). Furthermore, genomic analysis reveals two other putative α -amylase-like genes as well as two chitinases and a β -galactosidase.

In addition to glucosides, *P. furiosus*' natural environment, a geothermally heated marine sediment, most likely also contains α -linked galactosides, although an analysis of its chemical composition has never been reported. However, a hyperthermophilic bacterium isolated from the same site, *Thermotoga maritima*, has been shown to contain an α -galactosidase that liberates galactose from 4-nitrophenyl- α -galactopyranoside (Gal α pNp) as well as from the disaccharide melibiose (α -galactopyranosyl-1,6- α -glucopyranoside) and the trisaccharide raffinose (α -galactopyranosyl-1,6- α -glucopyranosyl-1,2- β -fructofuranoside) (Liebl *et al.*, 1998). Similar

enzymes have been found in other hyperthermophilic bacteria, such as *Thermus* sp. (Koyama *et al.*, 1990, Fridjonsson *et al.*, 1999, Berger *et al.*, 1995) and *Thermotoga neapolitana* (King *et al.*, 1998). α -Galactosidases catalyze the hydrolysis of α -1,6-linked galactose residues from oligosaccharides and polymeric galactomannans. Catalysis of transglycosylation has also been observed, which makes these enzymes interesting for the production of specific oligosaccharides that may have applications in the pharmaceutical and food industries, for instance as potential prebiotic food additives (Van Laere *et al.*, 1999). Moreover, α -galactosidases have been used to eliminate D-raffinose from sugar beet molasses in order to increase the yield of sucrose (α -glucopyranosyl-1,2- β -fructofuranoside) by facilitating its crystallization in the sugar industry (Ganter *et al.*, 1988). Also an application has been demonstrated in the treatment of soybean milk to alleviate flatulence caused by raffinose and stachyose (α -galactopyranosyl-1,6- α -galactopyranosyl-1,6- α -glucopyranosyl-1,2- β -fructofuranoside) (Sugimoto and van Buren, 1970, Thananunkul *et al.*, 1976). The application of thermostable α -galactosidases is desired to withstand the elevated temperatures required in the respective biotechnological processes.

α -Galactosidases known to date have been isolated from a variety of eukaryotes and bacteria and are classified in three different families of glycoside hydrolases according to Henrissat (Henrissat, 1991). Most eukaryal α -galactosidases share a substantial amino acid sequence homology and are classified in family 27. Bacterial and some fungal α -galactosidases are classified in family 36 and have limited homology among them. This family comprises the aforementioned thermophilic enzymes. Family 4 contains bacterial α -galactosidases along with β -phosphoglucosidases and α -glucosidases. So far, no archaeal α -galactosidase has been described. However, genomic analysis of *Sulfolobus solfataricus* predicted a putative family 36 α -galactosidase from this Crenarchaeon (She *et al.*, 2001). Analysis of the *Pyrococcus furiosus* genome revealed a cluster of genes involved in the utilization of galactosides, of which the galactokinase and the β -glucosidase have been described in detail (Ettema *et al.*, 2001, Kaper *et al.*, 2001, Verhees *et al.*, 2002). A second glycoside hydrolase, with unknown specificity, was present in this cluster. In this paper, the unknown glycosidase will be characterized as the first archaeal α -galactosidase. A comparison with functional as well as with sequential/structural homologs is made, and phylogenetic and physiological implications are discussed.

Materials and Methods

Bacterial hosts and vectors - The T7 expression vectors pET9d and pET24d were obtained from Novagen. *Escherichia coli* XL-1 blue (Stratagene) was used as an initial host for cloning while *E. coli* BL21 (DE3) (Stratagene) was used as an expression host for the pET-derivatives. *E. coli* was grown in TY medium in a rotary shaker at 37°C. Kanamycin was added to a final concentration of 30 μ g/ml.

Cloning and expression of the gene - *P. furiosus* genomic DNA was isolated as described previously (Sambrook *et al.*, 1989) and used as template in a PCR reaction. A primer set with *Nco*I and *Bam*HI restriction sites, introducing an extra alanine after the initial methionine, was designed (Table 2.1) to amplify *galA* by PCR on a Thermal Cycler (Perkin Elmer Cetus). In addition to 10 ng of *P. furiosus* chromosomal DNA and 100 ng of each primer, the 100 µl mixture contained 0.2 mM dNTP's, *Pfu* buffer and 5 U *Pfu* DNA polymerase. The PCR product with the expected size was digested and cloned into *Nco*I/*Bam*HI-digested vector pET24d, resulting in pLUW581. *E. coli* XL1-blue and BL21(DE3) were transformed with this plasmid. Sequence analysis of pLUW581 was done by the dideoxynucleotide chain termination method with a Li-Cor automatic sequencing system (model 4000L).

Overexpression of the *galA* gene and purification of recombinant protein - A 5 ml culture of freshly transformed *E. coli* BL21(DE3) cells with pLUW581 was used to inoculate 1 liter of TY medium containing 30 µg/ml kanamycin. When the OD₆₀₀ reached 0.5 the cells were induced with 50 µM IPTG and subsequently grown at 37 °C for approximately 16 hrs. Cells were harvested by centrifugation (4000 x g for 10 min.) and resuspended in a 100 mM NaP_i buffer containing 150 mM KCl (pH 7.0). Cells were lysed by passing 2-3 times through a French Pressure cell at 1000 Psi. After removal of the cell debris by centrifugation (48.000 x g for 30 min.) the supernatant was incubated for 30 min. at 80 °C. The supernatant obtained after another centrifugation (48.000 x g for 30 min.) was used for the enzymatic assays. For kinetic data a further purified enzyme was used. Ammonium sulfate was added to the cell free extract up to 1 M. The sample was applied to a Phenyl Sepharose column (Amersham Pharmacia). In a linear gradient towards 0 M (NH₄)₂SO₄ GalA eluted when the salt concentration reached zero. After dialysis against 20 mM Tris pH 8.0, the sample was applied to a Q-Sepharose column (Amersham Pharmacia). GalA did not bind to the column in contrast to most of the contaminating proteins. The flowthrough was concentrated and passed over a Superdex 200 column (Amersham Pharmacia) in 20 mM Tris pH 8.0 containing 100 mM NaCl. GalA eluted in the void volume. As judged by SDS-PAAGE analysis, the protein appeared to be more than 95% pure.

Table 2.1 Constructs that were used and primers that were used to make them.

Plasmid	Description	Primer ^a	Primer Sequence ^b	Introduced Restriction Site
pLUW581	GalA wildtype	BG378 s	GCGCGCCATGGCAAGAGCACTAGTCTTTCATGGG	<i>Nco</i> I
		BG379 a	GCGCGGGATCCCTAATGTTTCCCATTTTCACTCC	<i>Bam</i> HI
pWUR14	E72A	BG894 s	AGTGGCCTCATCGCGATTCTTGAA	<i>Nru</i> I
		BG895 a	TTCCAAGATTCGCGATGAGGCCACT	
pWUR15	E100A	BG896 s	AAAGAGATAGAGCAGTGAAAGAGAAC	<i>Bts</i> I
		BG897 a	GTTCTCTTTCAGTCTCTATCTCTT	
pWUR16	E117A	BG898 s	GATTCTGGCTACCTGCATTGGCATAT	<i>Bsp</i> MI
		BG899 a	ATATGCCAATGCAGGTAGCCAGAATC	
pWUR17	D139A	BG900 s	CTCTTCGCTGCAGGTGAGGCTATGC	<i>Bsp</i> MI
		BG901 a	GCATAGCCTCACCTGCAGCGAAGAG	
pWUR99	E193A	BG1122 s	CATAAATCTGTCTTTGCAAGGAAAAGTCACAC	<i>Cvi</i> RI
		BG1123 a	GTGTGACTTTTCCTGCAAGACAAGATTTATG	

^as = sense, a = antisense.

^bBases introducing the desired aminoacid mutations are in bold, introduced restriction sites are underlined with silent mutations in italic.

Mutagenesis - The introduction of mutations was performed by overlap extension PCR (Ho *et al.*, 1989) using plasmid pLUW581 as a template in the first step. For each mutation two PCR reactions were done each with the use of a primer introducing the mutation (Table 2.1) and one of the flanking primers. The purified overlapping PCR products were used in the second step both as a template and as a primer. With the flanking primers this reaction yielded the *galA* gene with the mutation. In order to facilitate the detection of the mutants, a restriction site was created as indicated in Table 2.1. The amplified DNA was cloned in pET24d as described above and checked by restriction analysis as well as by DNA sequence analysis. *E. coli* BL21(DE3) cells were transformed with the obtained plasmids (Table 2.1) for overproduction of the mutant α -galactosidases.

Enzymatic assays - Standard enzymatic assays were performed at 90°C in McIlvaine buffer (pH 5.0) containing 1 mM of pNp derivatives (ϵ_{405} pNp = 0.6 mM⁻¹cm⁻¹) as described elsewhere (Kengen *et al.*, 1993). Standard assays for polysaccharide hydrolysis were carried out at 90°C, for 30 min, using 0.2-1% (w/v) substrate in McIlvaine buffer (pH 5.0). The reducing sugars were detected by the dinitrosalicylic acid (DNS) method, with glucose as standard (Gueguen *et al.*, 1997). One unit is defined as the amount of enzyme required to release 1 μ mol of reducing sugars per min. Standard assay towards di/tri saccharides was determined by measuring the increase of NADPH in an indirect coupled assay with glucose dehydrogenase with significant galactose dehydrogenase activity from *Thermoplasma acidophilum* (Smith *et al.*, 1989). First the pyrococcal glycoside hydrolase was incubated in McIlvaine buffer (pH 5.0) containing 10 mM substrate at 90°C for 30 min. A 30- μ l aliquot was rapidly cooled down on ice and added to NaPi (pH 7.0) containing 0.4 mM NADP⁺ and 2 U *T. acidophilum* glucose dehydrogenase. The following assay was performed at 50°C. The absorbance of NADPH was followed at 340 nm (ϵ = 6.3 mM⁻¹cm⁻¹). One unit is defined as the amount of enzyme required to convert 1 μ mol of galactose/glucose per min.

The pH dependence was determined in McIlvaine buffer over pH 4-8 at 90°C. Buffers were adjusted at room temperature. Gal α pNp was used as substrate. The reaction was stopped after 5 min at 90°C by placing the vials on ice followed by the addition of Na₂CO₃ (ϵ_{405} = 5.3 mM⁻¹cm⁻¹ at room temperature). For the determination of the temperature optimum a 20- μ l sample of the produced pyrococcal glycoside hydrolases was injected into a crimp-sealed vial, containing 1 mM Gal α pNp in McIlvaine buffer (pH 5.5), which was incubated at the desired temperature. For temperatures above 90°C, an oil-bath was used. After 5 min of incubation, the reaction was stopped by placing the vials on ice. A 500- μ l aliquot was added to 1 ml 0.5 M Na₂CO₃ and the absorbance at 405 nm was measured.

Thermostability was determined in 20 mM Tris-HCl (pH 8.5) by heating the enzyme in small crimp-sealed vials, submerged in an oil bath. During a time series (0-20 h), 20- μ l aliquots were withdrawn and tested for remaining activity as described before.

Rescue of activity by azide and formate was performed by a discontinuous assay at 90 °C with 4 mM pNp- α -galactopyranoside in a 100 mM sodium phosphate buffer pH 5.5 with the

addition of 1 or 2 molar of sodium azide or 4 molar of sodium formate (final concentrations). After the addition of 1 ml 0.5 M Na₂CO₃ to the 500 µl sample the absorbance at 405 nm was measured.

All activities were corrected for spontaneous hydrolysis and for hydrolysis by cell free extracts of *E. coli* cells harboring pET24d without a gene insert.

Results and Discussion

Isolation of the *P. furiosus* gal gene cluster

In the course of the *P. furiosus* genome project (www.genome.utah.edu), a genomic clone was deposited in the NCBI database, potentially encoding a glycoside hydrolase with significant homology to family 1 glycoside hydrolases (accession number AAG28457). Molecular analysis of the locus (AF195244) revealed a clustering of five open reading frames (ORFs) (Fig. 2.1). BLAST searches of these ORFs revealed significant homologs in the GenBank/EMBL databases. The complete cluster was found to exist in the closely related *P. horikoshii* as well, but not in *P. abyssi* (Ettema *et al.*, 2001). The average amino acid identity of the five gene products between the two species is 85%. The first ORF (PF0445) was found to share high similarity with galactokinases (39% identity with a galactokinase from *Thermotoga neapolitana* (AAC24222), and was named *galK*. Heterologous expression in *Escherichia coli* resulted in a 40 kDa gene product showing ATP-dependent galactokinase activity (Verhees *et al.*, 2002). The second ORF (PF0444) appeared to resemble family 57 glycosidases (24% identity with an α -amylase from *Dictyoglomus thermophilum* (P09961), and was named *galA*. The third ORF (PF0443) showed high similarity with several hydrophobic sequences encoding secondary transport proteins (symporter/antiporter) of the major facilitator super family (25% identity with a transporter (major facilitator superfamily) from *Aquifex aeolicus* (AAC07232), and was named *mfsA*. From overlapping coding regions, TATA-box and transcription termination sequences (van der Oost *et al.*, 1998), it is concluded that *galK*, *galA* and *mfsA* are most likely organized in an operon structure (Fig. 2.1). The fourth ORF (PF0442) of the gal cluster closely resembled family 1 glycoside hydrolases (63% identity with β -glucosidase from *Thermococcus* sp. (Z70242), and was named *bglA*. BglA showed only 34% identity with β -glucosidase (CelB) from *P. furiosus*. Recombinant BglA appeared to have highest activity towards para-nitrophenyl- β -glucoside (Kaper *et al.*, 2001). Unlike CelB (Kengen *et al.*, 1993), BglA did not hydrolyze lactose or cellobiose. A relatively high activity was found with the hydrophobic substrates β -octyl-D-glucose and salicine. The study by Matsui *et al.* (Matsui *et al.*, 2000), claiming a novel substrate specificity of the ortholog BGPh from *P. horikoshii*, agrees well with our findings. The fifth ORF (PF0441) showed high similarity with uridylyltransferases (42% identity with a galactose-1-phosphate uridylyltransferase from *Thermotoga neapolitana* (AAC24221), and was named *galT*.

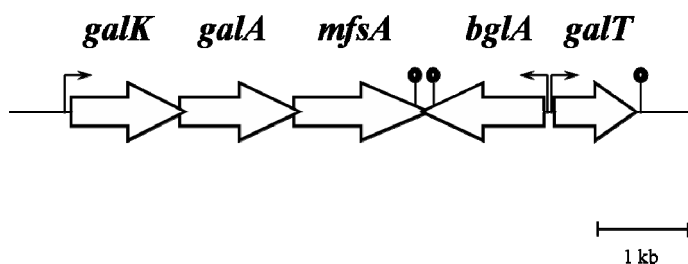


Figure 2.1 *GalA* gene cluster. Genetic organization of the *gal* locus in the genome of *Pyrococcus furiosus*. Putative TATA-boxes (∗) and transcription termination sites (↑) are indicated. The abbreviations used are as follows (*Pyrococcus furiosus* gene numbers are in straight brackets, primary accession numbers are in parentheses): *galK*: galactokinase [PF0445] (Q9HHB6), *galA*: α -galactosidase [PF0444] (Q9HHB5), *mfsA*: putative sugar transporter [PF0443] (Q9HHB4), *bglA*: β -glucosidase [PF0442] (Q9HHB3), *galT*: galactose-1-P uridylyltransferase [PF0441] (Q9HHB2).

GalA versus family 57 enzymes

The *galA* gene encodes a polypeptide of 364 amino acids. BLASTP analysis showed that GalA shares highest homology (up to 24%) with enzymes classified in family 57 of glycoside hydrolases (<http://afmb.cnrs-mrs.fr/CAZY/index.html>), whereas negligible homology (10%) was found on family 36 α -galactosidases. The dendrogram in Fig. 2.2A clearly shows the distance of GalA to family 36 and its close relationship to family 57. Enzymes from the latter family originate from hyperthermophilic bacteria and archaea. They comprise α -amylases (Fukusumi *et al.*, 1988, Laderman *et al.*, 1993b) and amylopullulanases (Erra-Pujada *et al.*, 1999, Dong *et al.*, 1997b). The α -amylases have been shown to have transferase activity as well. In both the cases of *Thermococcus litoralis* and of *T. kodakaraensis* transferase activity is even higher than the hydrolytic activity and for that reason these enzymes have been named α -glucanotransferases (Jeon *et al.*, 1997, Tachibana *et al.*, 1997). All enzymes from this family that have been characterized so far act on α -1,6 and/or on α -1,4 linkages in glucose-polymers like pullulan, starch and amylose. This, however, does not imply that GalA is active on similar substrates. The same feature is seen for instance in family 4, which comprises not only α -glucosidases but also α -galactosidases and even β -glucosidases.

The highest homology to family 57 enzymes was found to be located on the N-terminal part of the protein. Closer inspection of the amino acid sequences of all these enzymes reveals that this homologous domain characterizes the family (Fig. 2.2B, C). Most family 57 enzymes consist of one or more additional domains as well. Strikingly, this conserved domain was also found to share homology to parts of class 1 α -mannosidases classified in family 38. Alignment of the domain sequences results in 25%-93% homology within family 57 members, 58%-77% within family 38 and 20%-32% between family 57 and family 38. A dendrogram representing these homologous sequences is shown in Fig. 2.2B. The size of the domain is arguable, as homology gradually decreases between residues 200 and 260 of the GalA sequence (see also Fig. 2.3). It is clear in Fig. 2.3 that strongest conservation is found between residue 50 and 140 of the GalA sequence. In the COG database the domain has been defined as COG1543, with unknown function (www.ncbi.nlm.nih.gov/COG/). Recently a 3-dimensional structure of the 4- α -glucanotransferase from *Thermococcus litoralis* was published (Imamura *et al.*, 2003). The conserved sequence appeared to be part of the catalytic domain, suggesting a role in the mechanism of catalysis.

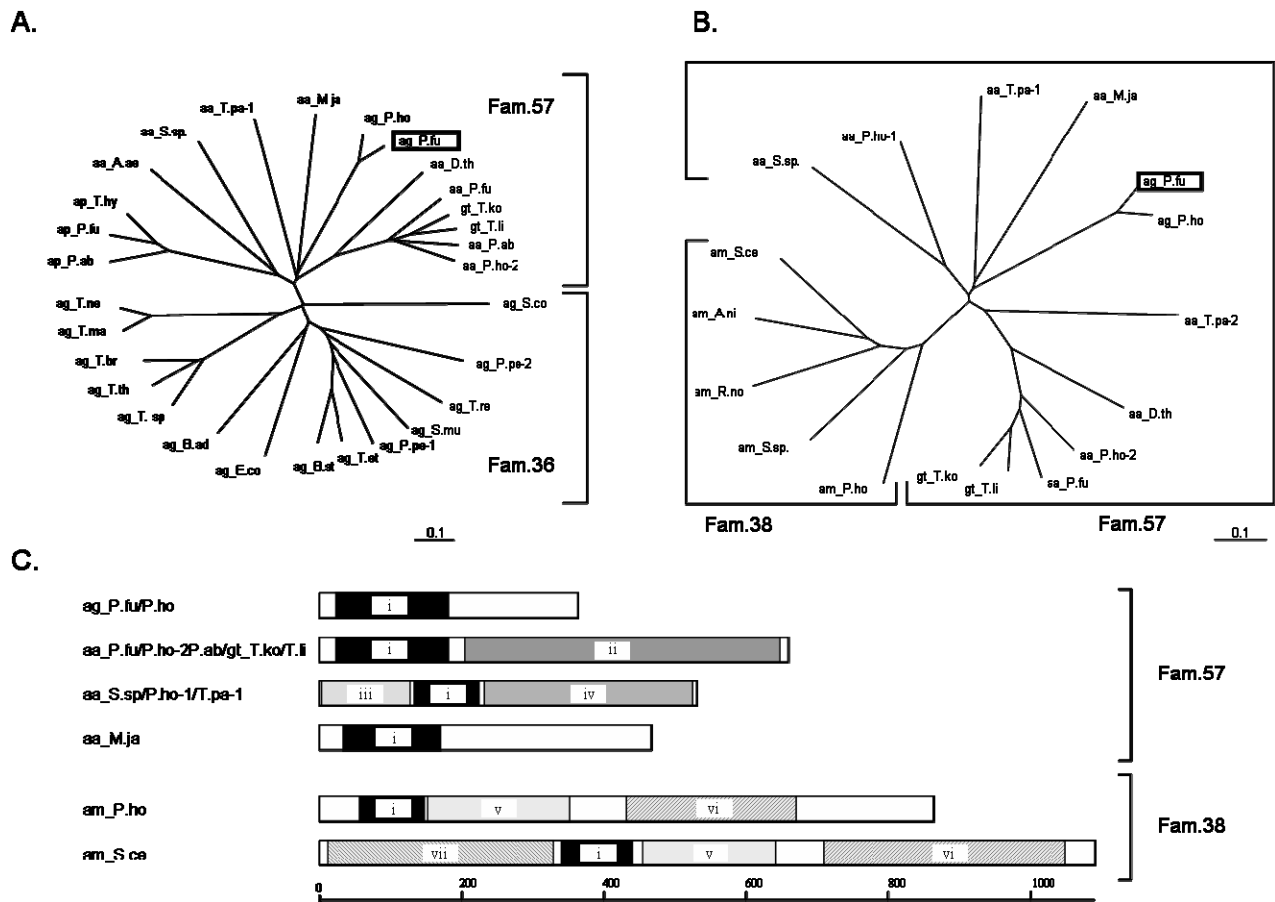


Figure 2.2 *Pyrococcus furiosus* α-galactosidase and related glycoside hydrolases. Dendrograms of selected glycoside hydrolases classified in family 57 and 36 (A) and of the conserved domain in family 57 and 38 glycoside hydrolases (B). Dendrograms are based on full length (A) and partial (B) multiple sequence alignments created by ClustalX and Genedoc. Dendrograms are made with Treeview. C is the schematic representation of domain arrangement in various members of family 57 and 38. Domains of conserved sequences are depicted as shaded boxes. Same numbers and shades represent homologous domains. The conserved domain shown in black boxes (i) corresponds to the domain represented by dendrogram B. Length of the protein sequences in C is indicated in amino acid residues by the bar underneath. The abbreviations used are as follows (primary accession numbers are in parentheses): ag = α-galactosidase; aa = α-amylase; ap = amylopullullanase; am = α-mannosidase; gt = 4-α-glucanotransferase; A.ae = *Aquifex aeolicus* VF5 (O66934); A.ni = *Aspergillus nidulans* (O13344); B.ad = *Bifidobacterium adolescentis* DSM 20083 (Q9XCX2); B.st = *Bacillus stearothermophilus* (Q9X624); D.th = *Dictyoglomus thermophilum* (P09961); E.co = *Escherichia coli* K12 (P16551); M.ja = *Methanococcus jannaschii* (Q59006); P.ab = *Pyrococcus abyssi* (aa: Q9V298, ap: Q9V294); P.fu = *Pyrococcus furiosus* DSM3638 (aa: P49067, ag: Q9HHB5, ap: Q30772); P.ho = *Pyrococcus horikoshii* OT3 (aa-1: O50094, aa-2: O57932, ag: O58106, am: O58565); P.pe = *Pediococcus pentosaceus* (ag-1: P43467, ag-2: P43469); R.no = *Rattus norvegicus* (P21139); S.ce = *Saccharomyces cerevisiae* S288C (P22855); S.co = *Streptococcus coelicor* A3(2) cosmid 9B10 (O50520); S.mu = *Streptococcus mutans* (P27756); S.sp. = *Synechosystis sp. strain PCC6803* (aa: Q55545, am: Q55528); T.al = *Thermococcus alcaliphilus* (US patent 5958751); T.br = *Thermus brockianus* IT1360 (Q9X6C5); T.et = *Thermoanaerobacter ethanolicus* (P77988); T.hy = *Thermococcus hydrothermalis* (Q9Y818); T.ko = *Thermococcus kodakaraensis* (O32450); T.li = *Thermococcus litoralis* (O32462); T.ma = *Thermotoga maritima* (O33835); T.ne = *Thermotoga neapolitana* (Q9R7H1); T.pa = *Treponoma pallidum* (aa-1: O83377, aa-2: O83182); T.re = *Trichoderma reesei* (Q92457); T.sp = *Thermus sp.T2* (Q9WXC1); T.th = *Thermus thermophilus* (Q9X6D2).

Expression and characterization of GalA

To determine the functionality of GalA, its gene was amplified from *P. furiosus* genomic DNA and cloned in the vector pET24d. The resulting plasmid, pLUW581, was used to transform *E. coli* BL21(DE3) cells. An overproduction of GalA of 15 - 30 % of total protein was achieved after induction with 50 μ M IPTG. The majority of GalA was present as insoluble, but at least partly active, aggregates. Extensive variation in the growth temperature (10-37°C), the induction conditions (0.0-0.4 mM IPTG), as well as the vector/host (pET9d/pET24d in BL21(DE3), JM109(DE3), C41(DE3) and C43(DE3) (Miroux and Walker, 1996) and pGEX2TK (GST-fusion) (Amersham Pharmacia) in BL21(DE3) did not improve the fraction of soluble enzyme. However, the amount of GalA in the soluble fraction was doubled by adding 150 mM KCl to the resuspension buffer (data not shown). Extraction of soluble protein by hydrophobic solvents or solutes from these aggregates, similar to what has been done with the pyrococcal α -amylase (Linden *et al.*, 2000), was unsuccessful. However, addition of 50% glycerol resulted in the extraction of 10-15% of the enzyme activity from the insoluble fraction.

Based on the deduced amino acid sequence, the molecular weight of GalA was calculated as 41.6 kDa. SDS-PAGE analysis of the purified GalA resulted in a single band of 44 kDa. Highest activity of GalA was detected with Gal α pNp (1.61 U/mg), and only little activity with melibiose (0.145 U/mg) and raffinose (0.006 U/mg). No GalA activity was measured with the polysaccharides carrageenan, ara-galactan, galactomannan and stractan. Neither on Glu α pNp nor on starch any activity could be detected, proving the non-amylolytic identity of GalA. Like the α -galactosidase of *Thermotoga maritima* (Liebl *et al.*, 1998), no stabilizing effect was observed upon addition of NAD⁺, DTT, and various metals/cations. Highest activity of GalA was observed at a pH of 5.2. The protein showed to be a very thermo-active and thermostable enzyme (Table 2.2). An extremely high optimal temperature was found of 115 °C. At 100°C in a 20 mM Tris-HCl (pH 8.5) buffer, GalA showed a half-life time of 15 hours. The thermostability was dependent on pH and decreased dramatically below pH 5.

Table 2.2 Biochemical characteristics (A) and substrate specificity (B) of *P. furiosus* GalA. Kinetic parameters have been determined at 90 °C and pH 5.5 with pNp- α -galactopyranoside as a substrate.

A	T _{opt}	115 °C
	pH _{opt}	5.0 - 5.5
	t _{1/2} at 100 °C	15 hrs
	K _m	0.25 mM
	V _{max}	880 U/mg
B	Substrate	GalA relative activity in %
	Gal α pNp	100
	Glc α pNp, Man α pNp, Rha α pNp, Gal β pNp	nd
	Melibiose	9
	Raffinose	4
	Carrageenan, ara-Galactan, Galactomannan, Stractan,	nd
	Starch	

Gal α pNp = para-nitrophenyl- α -galactopyranoside, Glc α pNp = para-nitrophenyl- α -glucopyranoside, Man α pNp = para-nitrophenyl- α -mannopyranoside, Rha α pNp = para-nitrophenyl- α -rhamnopyranoside, Gal β pNp = para-nitrophenyl- β -galactopyranoside. nd = not detected (activity > 0.3 %)

Regarding hydrolysis of Gal α pNp, GalA showed Michaelis-Menten kinetics at 90 °C under further optimal conditions. The affinity for the substrate is high, reflected by a K_m of 0.25 mM, which is very common among (thermophilic) α -galactosidases (Liebl *et al.*, 1998, Leder *et al.*, 1999). The calculated V_{max} of 800 U/mg is comparable to that of related glycoside hydrolases as well (Kaper *et al.*, 2001).

Identification of catalytic residues

Glycoside hydrolases act either via a double or single displacement mechanism releasing products with net retention or inversion of the anomeric carbon respectively (Sinnott, 1990, McCarter and Withers, 1994). Both mechanisms employ a pair of carboxylic acids in the active site. In retaining enzymes one residue acts as a nucleophile and the other as an acid/base catalyst. Inverting enzymes, on the other hand, uses one residue as a general acid and the other as a general base catalyst. In contrast to the inverting mechanism, the retaining mechanism involves a glycosyl-enzyme intermediate, allowing transglycosylation.

When GalA was incubated for 2 hrs at 80°C with 40% melibiose or 20% raffinose some transgalactosylation products (besides the hydrolysis products) were detected by HPLC analysis (results not shown). This indicates that the enzyme must use a retaining mechanism, which has previously been reported for the other members of family 57 as well as for α -galactosidases from family 27, 36 and 4.

To get further insight into the catalytic mechanism of GalA, a study was performed to identify the catalytic residues. In the classification system introduced by Henrissat (Henrissat, 1991) all enzymes in the same family share completely conserved catalytic amino-acid residues. Sometimes conservation is also observed between different families, within a clan, with ambiguity towards the nature of the catalytic residue itself. A conserved catalytic glutamate in one family can be homologous to a conserved catalytic aspartate in another. To identify the catalytic nucleophile and acid/base we made a multiple alignment of the amino acid sequences of various members of family 57 as well as of family 38 (Fig. 2.3). The alignment shows four more or less conserved carboxylic acids within family 57 enzymes. Glutamate 72 is well conserved within the enzymes of family 57. On the other hand, no carboxylic acid appears to be present at the corresponding site in family 38. At position 100 a glutamate is present in some family 57 α -amylases, which is also conserved in the amylopullulanases (not shown). An invariant glutamate at position 117 is in a region of the sequence that has highly conserved residues among family 57 enzymes. This glutamate has recently been identified as the catalytic nucleophile in 4- α -glucanotransferase from *Thermococcus litoralis* (Imamura *et al.*, 2001). In family 38 α -mannosidases an invariant aspartate is found at the corresponding position instead of a glutamate. It is this aspartate that has been identified as the catalytic nucleophile in the α -mannosidases from Jack Bean (Howard *et al.*, 1998) and bovine kidney lysosomes (Numao *et al.*, 2000) as well. The aspartate at position 139 is also consistent within family 57 and variable within family 38.

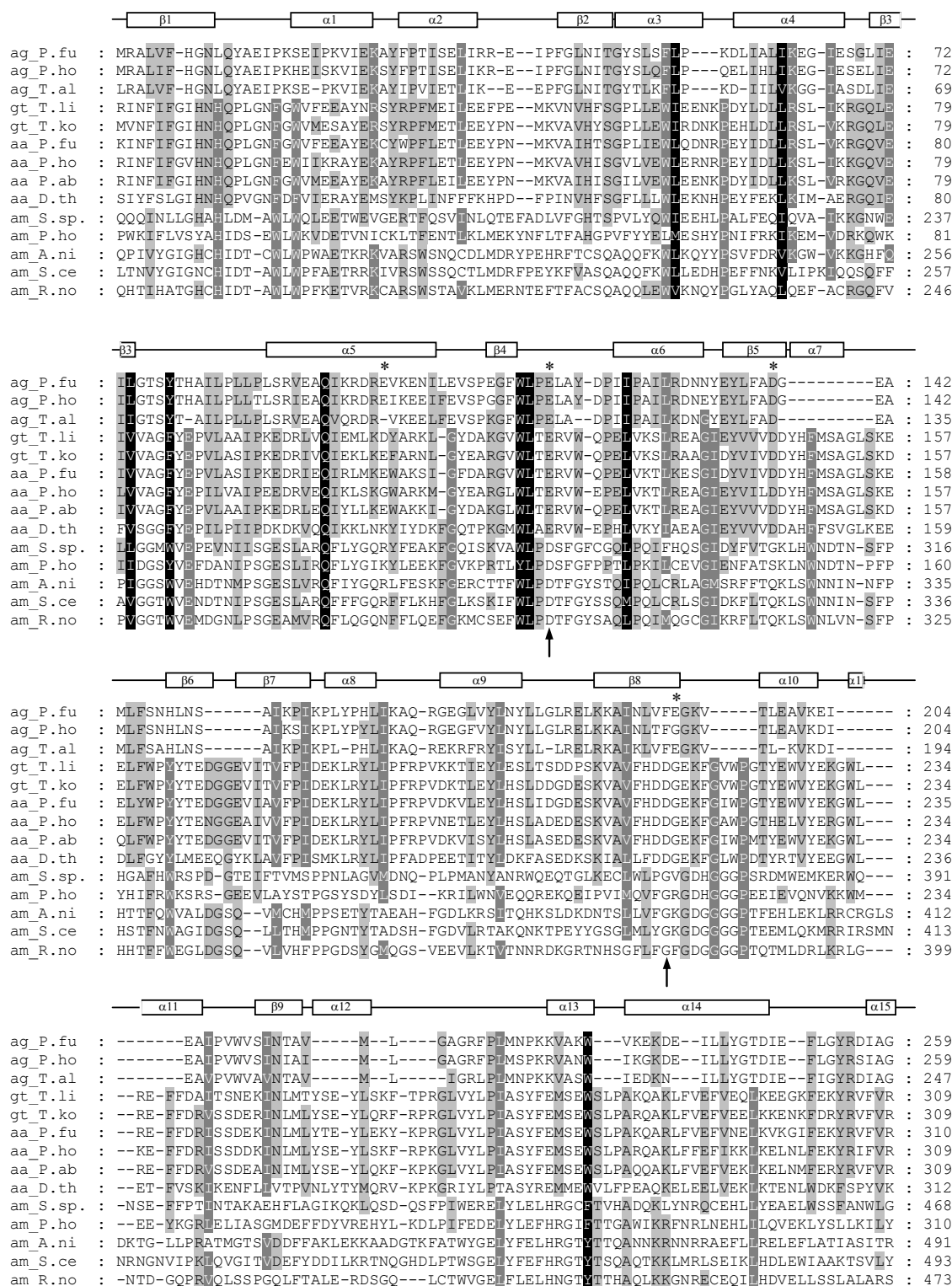


Figure 2.3 Multiple sequence alignment of selected glycoside hydrolases classified in family 57 and 38. Alignment shows the part of the sequences representing the conserved domain and was created using ClustalX and GeneDoc. The secondary structure of the *Thermococcus litoralis* 4- α -glucanotransferase is shown above the sequences. Conserved residues are shaded. Arrows indicate the catalytic residues identified in previous studies (see text). Residues of GalA that have been studied by site-directed mutagenesis are marked by asterisks. The abbreviations used are as follows (primary accession numbers are in parentheses) or identical to those used in Figure 2: ag_T.al = *Thermococcus alcaliphilus* α -galactosidase (US patent 5958751); aa_P.ho = *Pyrococcus horikoshii* OT3 putative α -amylase (O57932).

These four putative catalytic residues were changed into alanine residues. Hydrolytic activity of the mutants E72A, E100A and D139A was reduced ten- to forty-fold (Table 2.3). The remaining activity of the E117A variant was only 0.5 % compared to the wild-type enzyme. This suggests that glutamate 117 could be the catalytic nucleophile or acid/base although the reduction is less severe than might be expected for a mutation of one of the catalytic residues (Wang *et al.*, 1994, Viladot *et al.*, 1998, Moracci *et al.*, 1998). However, the remaining activity might be due to background activity of contaminating wild-type enzyme as has been reported before (Wang *et al.*, 1994, Moracci *et al.*, 1998). Moreover, most studies in which a 10^5 to 10^8 fold reduction has been reported, were performed with β -glycoside hydrolases. To date only a few studies have described the identification of the catalytic residues of α -glycosidases by means of mutagenesis. The nucleophile or acid/base mutants of these family 51 α -L-arabinofuranosidases and a family 67 α -glucuronidase showed reduced activities ranging between 5×10^{-2} and 10^{-5} (Zaide *et al.*, 2001, Debeche *et al.*, 2002, Shallom *et al.*, 2002a). Only the nucleophile mutant of the *Geobacillus stearothermophilus* enzyme appeared to have a 10^{-8} reduced activity (Shallom *et al.*, 2002b). It is not unlikely that the behaviour of α -glycosidases is different from β -glycosidases when the catalytic residues are changed. Therefore, care must be taken when activities are compared between the two types of enzymes. Chemical rescue by small external nucleophiles is considered to be a more important method to identify the catalytic residues, and has been successfully applied to many glycoside hydrolases (Ly and Withers, 1999). With this strategy, again, α -glycosidases showed a different tendency than β -glycosidases. Enzymes belonging to the former group could be reactivated 1.2 to 67 fold whereas those belonging to the latter are reported to show a reactivation of up to 10^5 fold, almost back to wildtype level (Debeche *et al.*, 2002, Shallom *et al.*, 2002a, Shallom *et al.*, 2002b, Wang *et al.*, 1994, Viladot *et al.*, 1998, Moracci *et al.*, 1998).

Upon addition of azide or formate the activity of E117A was recovered up to 6.8 %, whereas the wild-type activity was reduced with at least 20%. On the other hand, the activity of E72A, E100A and D139A did not change in the presence of azide or formate. This 13.6 fold rescue of activity by the addition of external nucleophiles provides further support that glutamate 117 is a catalytic residue. Since no other carboxylic acid residue was completely conserved according to our alignment, a prediction of the other catalytic residue was only possible after the elucidation of the 3-dimensional structure of the 4- α -glucanotransferase from *Thermococcus litoralis* (Imamura *et al.*, 2003). In the structure of the enzyme-acarbose complex it became evident that aspartate 214 of this enzyme is the acid/base catalyst, whereas glutamate 123 acts as the nucleophile. Based on our sequence alignment, 219 N-terminal residues of the GalA sequence were modelled against the catalytic domain of the structure, using the program SWISS-MODEL (<http://swissmodel.expasy.org/>). The sequence identity in that part is 18%, whereas homology is 34%. The model showed two carboxylic residues in close proximity of glutamate 117 (not shown). At 8.3 Å and 9.5 Å distance from the proposed nucleophile aspartate 139 and glutamate 193 were proposed as putative acid/base catalyst respectively. As the first proposed residue was not supported by our experimental results the latter was studied by mutagenesis as well. However, the

alignment shows that the direct region of glutamate 193 is not highly conserved between the α -amylases and the α -galactosidases. Moreover, there is a shift of one residue between position 188 and 195. This has been observed in a more obvious fashion in for instance family 16, where a missing methionine marks the difference between endo- β -1,3-1,4-glucanases and endo- β -1,3-glucanases (Gueguen *et al.*, 1997). In the tertiary structure this might be overcome by a twist of the β -strand or a beta-bulge (Hahn *et al.*, 1995, Richardson *et al.*, 1978).

E193A resulted in an enzyme with 12.6 fold decreased activity (Table 2.3). The addition of azide or formate caused no increase of activity. This suggests that glutamate 193 is probably not the acid/base catalyst in GalA. Moreover, this would imply that of two different glycoside hydrolases from the same family the role of acid/base catalyst is played by a different carboxylic residue. This is not in agreement with the classification criteria as defined by Henrissat (Henrissat, 1991). On the other hand, it might be considered to classify GalA, and its orthologs from *P. horikoshii* and *Thermococcus alcaliphicus*, in a new family that forms a clan with families 57 and 38.

Table 2.3 Relative activities of wildtype *P. furiosus* GalA, and mutants under standard conditions and upon addition of azide or formate as external nucleophiles. Activities are given in percentages, with 100 % activity for the wildtype GalA under standard conditions as stated in the materials and methods section. Assays are performed with heat stable cell free extracts. Activities are corrected for spontaneous hydrolysis and hydrolysis by cell free extracts of *E. coli* cells harboring a pET24d vector without gene insert.

	standard	+ 1M azide	+ 2M azide	+ 4M formate
GalA Wildtype	100	51.3	79.14	74.28
E72A	9	9.5	ND	ND
E100A	2.5	2.75	ND	ND
E117A	0.52	4.74	4.84	6.78
D139A	7	6.75	ND	ND
E193A	7.91	ND	2.33	4.30

ND = not determined.

In conclusion, the GalA from *Pyrococcus furiosus* functions as a bacterial-like α -galactosidase, however, without the capacity to hydrolyze polysaccharides. Since it is the first study of an α -galactosidase from a hyperthermophilic archaeon it is not surprising that GalA is the most thermo-active and thermostable α -galactosidase described to date. This is reflected by an optimum temperature of 115 °C and a half-life time of activity of 15 hours at 100°C. The uniqueness of GalA is its amino acid sequence. It shows no homology to other α -galactosidases, but rather to α -amylases from family 57. Based on our results and recent work from Imamura and coworkers (Imamura *et al.*, 2001) we conclude that glutamate 117 is the catalytic nucleophile. Furthermore, we could not support the hypothesis that acid/base catalysis is performed by glutamate 193, which makes classification within family 57 arguable. Finally, our results show that the glutamates at position 72, 100 and 193 and the aspartate at 139 are not one of the catalytic residues, but play an important role in catalysis.

Acknowledgments

This research was supported by the Earth and Life Sciences Foundation (ALW) and the Technology Foundation (STW), applied science division of NWO and the technology programme of the ministry of Economic Affairs.

References

- Bauer, M. W., Bylina, E. J., Swanson, R. V. and Kelly, R. M. (1996) Comparison of a beta-glucosidase and a beta-mannosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. Purification, characterization, gene cloning, and sequence analysis. *J. Biol. Chem.* **271**, 23749-23755.
- Bauer, M. W., Driskill, L. E., Callen, W., Snead, M. A., Mathur, E. J. and Kelly, R. M. (1999) An endoglucanase, EglA, from the hyperthermophilic archaeon *Pyrococcus furiosus* hydrolyzes beta-1,4 bonds in mixed-linkage (1-->3),(1-->4)- beta-D-glucans and cellulose. *J. Bacteriol.* **181**, 284-290.
- Berger, J. L., Lee, B. H. and Lacroix, C. (1995) Identification of new enzyme activities of several strains of *Thermus* species. *Appl. Microbiol. Biotechnol.* **44**, 81-87.
- Costantino, H. R., Brown, S. H. and Kelly, R. M. (1990) Purification and characterization of an alpha-glucosidase from a hyperthermophilic archaeobacterium, *Pyrococcus furiosus*, exhibiting a temperature optimum of 105 to 115 degrees C. *J. Bacteriol.* **172**, 3654-3660.
- Debeche, T., Bliard, C., Debeire, P. and O'Donohue, M. J. (2002) Probing the catalytically essential residues of the alpha-L- arabinofuranosidase from *Thermobacillus xylanilyticus*. *Protein Eng.* **15**, 21-28.
- Dong, G., Vieille, C., Savchenko, A. and Zeikus, J. G. (1997a) Cloning, sequencing, and expression of the gene encoding extracellular alpha-amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* **63**, 3569-3576.
- Dong, G., Vieille, C. and Zeikus, J. G. (1997b) Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* **63**, 3577-3584.
- Erra-Pujada, M., Debeire, P., Duchiron, F. and O'Donohue, M. J. (1999) The type II pullulanase of *Thermococcus hydrothermalis*: molecular characterization of the gene and expression of the catalytic domain. *J. Bacteriol.* **181**, 3284-3287.
- Ettema, T., van der Oost, J. and Huynen, M. (2001) Modularity in the gain and loss of genes: applications for function prediction. *Trends Genet.* **17**, 485-487.
- Fiala, G. and Stetter, K. O. (1986) *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. *Arch. Microbiol.* **161**, 168-175.
- Fridjonsson, O., Watzlawick, H., Gehweiler, A., Rohrhirsch, T. and Mattes, R. (1999) Cloning of the gene encoding a novel thermostable alpha-galactosidase from *Thermus brockianus* ITI360. *Appl. Environ. Microbiol.* **65**, 3955-3963.
- Fukusumi, S., Kamizono, A., Horinouchi, S. and Beppu, T. (1988) Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*. *Eur. J. Biochem.* **174**, 15-21.
- Ganter, C., Bock, A., Buckel, P. and Mattes, R. (1988) Production of thermostable, recombinant [alpha]-galactosidase suitable for raffinose elimination from sugar beet syrup. *J. Biotechnol.* **8**, 301-310.
- Gueguen, Y., Voorhorst, W. G. B., van der Oost, J. and de Vos, W. M. (1997) Molecular and biochemical characterization of an endo-beta-1,3-glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **272**, 31258-31264.
- Hahn, M., Olsen, O., Politz, O., Borriss, R. and Heinemann, U. (1995) Crystal structure and site-directed mutagenesis of *Bacillus macerans* endo-1,3-1,4-beta-glucanase. *J. Biol. Chem.* **270**, 3081-3088.

- Henrissat, B.** (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309-316.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. and Pease, L. R.** (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.
- Howard, S., He, S. and Withers, S. G.** (1998) Identification of the active site nucleophile in jack bean α -mannosidase using 5-fluoro-beta-L-gulosyl fluoride. *J. Biol. Chem.* **273**, 2067-2072.
- Imamura, H., Fushinobu, S., Jeon, B. S., Wakagi, T. and Matsuzawa, H.** (2001) Identification of the catalytic residue of *Thermococcus litoralis* 4- α -glucanotransferase through mechanism-based labeling. *Biochemistry* **40**, 12400-12406.
- Imamura, H., Fushinobu, S., Yamamoto, M., Kumasaka, T., Jeon, B. S., Wakagi, T. and Matsuzawa, H.** (2003) Crystal structures of 4- α -glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor. *J. Biol. Chem.* **278**, 19378-19386.
- Jeon, B. S., Taguchi, H., Sakai, H., Ohshima, T., Wakagi, T. and Matsuzawa, H.** (1997) 4- α -glucanotransferase from the hyperthermophilic archaeon *Thermococcus litoralis*--enzyme purification and characterization, and gene cloning, sequencing and expression in *Escherichia coli*. *Eur. J. Biochem.* **248**, 171-178.
- Jorgensen, S., Vorgias, C. E. and Antranikian, G.** (1997) Cloning, sequencing, characterization, and expression of an extracellular α -amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. *J. Biol. Chem.* **272**, 16335-16342.
- Kaper, T., Verhees, C. H., Lebbink, J. H., van Lieshout, J. F., Kluskens, L. D., Ward, D. E., Kengen, S. W., Beerthuizen, M. M., de Vos, W. M. and van der Oost, J.** (2001) Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus*. *Meth. Enzymol.* **330**, 329-346.
- Kengen, S. W., Luesink, E. J., Stams, A. J. and Zehnder, A. J.** (1993) Purification and characterization of an extremely thermostable beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Eur. J. Biochem.* **213**, 305-312.
- King, M. R., Yernool, D. A., Eveleigh, D. E. and Chassy, B. M.** (1998) Thermostable α -galactosidase from *Thermotoga neapolitana*: cloning, sequencing and expression. *FEMS Microbiol. Lett.* **163**, 37-42.
- Koyama, Y., Okamoto, S. and Furukawa, K.** (1990) Cloning of α - and β -galactosidase genes from an extreme thermophile, *Thermus* strain T2, and their expression in *Thermus thermophilus* HB27. *Appl. Environ. Microbiol.* **56**, 2251-2254.
- Laderman, K. A., Asada, K., Uemori, T., Mukai, H., Taguchi, Y., Kato, I. and Anfinsen, C. B.** (1993a) α -amylase from the hyperthermophilic archaeobacterium *Pyrococcus furiosus*. Cloning and sequencing of the gene and expression in *Escherichia coli*. *J. Biol. Chem.* **268**, 24402-24407.
- Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. S., Griko, Y. V., Privalov, P. L. and Anfinsen, C. B.** (1993b) The purification and characterization of an extremely thermostable α -amylase from the hyperthermophilic archaeobacterium *Pyrococcus furiosus*. *J. Biol. Chem.* **268**, 24394-24401.
- Leder, S., Hartmeier, W. and Marx, S. P.** (1999) α -galactosidase of *Bifidobacterium adolescentis* DSM 20083. *Curr. Microbiol.* **38**, 101-106.
- Liebl, W., Wagner, B. and Schellhase, J.** (1998) Properties of an α -galactosidase, and structure of its gene galA, within an α - and β -galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. *Syst. Appl. Microbiol.* **21**, 1-11.
- Linden, A., Niehaus, F. and Antranikian, G.** (2000) Single-step purification of a recombinant thermostable α -amylase after solubilization of the enzyme from insoluble aggregates. *J. Chromatogr. B Biomed. Sci. Appl.* **737**, 253-259.
- Ly, H. D. and Withers, S. G.** (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem.* **68**, 487-522.
- Matsui, I., Sakai, Y., Matsui, E., Kikuchi, H., Kawarabayasi, Y. and Honda, K.** (2000) Novel substrate specificity of a membrane-bound β -glycosidase from the hyperthermophilic archaeon *Pyrococcus horikoshii*. *FEBS Lett.* **467**, 195-200.
- McCarter, J. D. and Withers, S. G.** (1994) Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**, 885-892.

- Miroux, B. and Walker, J. E. (1996) Overproduction of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289-298.
- Moracci, M., Trincone, A., Perugino, G., Ciaramella, M. and Rossi, M. (1998) Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. *Biochemistry* **37**, 17262-17270.
- Numao, S., He, S., Evjen, G., Howard, S., Tollersrud, O. K. and Withers, S. G. (2000) Identification of Asp197 as the catalytic nucleophile in the family 38 alpha-mannosidase from bovine kidney lysosomes. *FEBS Lett.* **484**, 175-178.
- Richardson, J. S., Getzoff, E. D. and Richardson, D. C. (1978) The beta bulge: a common small unit of nonrepetitive protein structure. *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2574-2578.
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular cloning: A laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Shallom, D., Belakhov, V., Solomon, D., Gilead-Gropper, S., Baasov, T., Shoham, G. and Shoham, Y. (2002a) The identification of the acid-base catalyst of alpha-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *FEBS Lett.* **514**, 163-167.
- Shallom, D., Belakhov, V., Solomon, D., Shoham, G., Baasov, T. and Shoham, Y. (2002b) Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from *Geobacillus stearothermophilus* T-6, a family 51 glycoside hydrolase. *J. Biol. Chem.* **277**, 43667-43673.
- She, Q., Singh, R. K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M. J., Chan-Weiher, C. C., Clausen, I. G., Curtis, B. A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P. M., Heikamp-de Jong, I., Jeffries, A. C., Kozera, C. J., Medina, N., Peng, X., Thi-Ngoc, H. P., Redder, P., Schenk, M. E., Theriault, C., Tolstrup, N., Charlebois, R. L., Doolittle, W. F., Duguet, M., Gaasterland, T., Garrett, R. A., Ragan, M. A., Sensen, C. W. and Van der Oost, J. (2001) The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7835-7840.
- Sinnott, M. L. (1990) Catalytic mechanisms of glycosyl transfer. *Chem. Rev.* **90**, 1171-1202.
- Smith, L. D., Budgen, N., Bungard, S. J., Danson, M. J. and Hough, D. W. (1989) Purification and characterization of glucose dehydrogenase from the thermoacidophilic archaebacterium *Thermoplasma acidophilum*. *Biochem. J.* **261**, 973-977.
- Sugimoto, H. and van Buren, J. P. (1970) Removal of oligosaccharides from soy milk by an enzyme from *Aspergillus saitoi*. *J. Food Sci.* **35**, 655-660.
- Tachibana, Y., Fujiwara, S., Takagi, M. and Imanaka, T. (1997) Cloning and Expression of the 4 Alpha Glucanotransferase Gene From the Hyperthermophilic Archaeon *Pyrococcus* Sp Kod1, and Characterization of the Enzyme. *J. Ferment. Bioeng.* **83**, 540-548.
- Thananunkul, D., Tanaka, M., Chichester, C. O. and Lee, T. (1976) Degradation of raffinose and stachyose in soybean milk by α -galactosidase from *Mortierella vinacea*. *J. Food Sci.* **41**, 173-175.
- van der Oost, J., Ciaramella, M., Moracci, M., Pisani, F. M., Rossi, M. and de Vos, W. M. (1998) Molecular biology of hyperthermophilic Archaea. *Adv. Biochem. Eng. Biotechnol.* **61**, 87-115.
- Van Laere, K. M., Hartemink, R., Beldman, G., Pitson, S., Dijkema, C., Schols, H. A. and Voragen, A. G. (1999) Transglycosidase activity of *Bifidobacterium adolescentis* DSM 20083 alpha-galactosidase. *Appl. Microbiol. Biotechnol.* **52**, 681-688.
- Verhees, C. H., Koot, D. G., Ettema, T. J., Dijkema, C., de Vos, W. M. and van der Oost, J. (2002) Biochemical adaptations of two sugar kinases from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Biochem. J.* **366**, 121-127.
- Viladot, J. L., de Ramon, E., Durany, O. and Planas, A. (1998) Probing the mechanism of *Bacillus* 1,3-1,4-beta-D-glucan 4-glucanohydrolases by chemical rescue of inactive mutants at catalytically essential residues. *Biochemistry* **37**, 11332-11342.
- Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J. and Withers, S. G. (1994) Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* **116**, 11594-11595.

- Zaide, G., Shallom, D., Shulami, S., Zolotnitsky, G., Golan, G., Baasov, T., Shoham, G. and Shoham, Y. (2001)**
Biochemical characterization and identification of catalytic residues in alpha-glucuronidase from *Bacillus stearothermophilus* T-6. *Eur. J. Biochem.* **268**, 3006-3016.

Chapter 3

Hydrolase and Glycosynthase Activity of Endo-1,3- β -glucanase from the Thermophile *Pyrococcus furiosus*

Van Lieshout, J.F.T., Faijes, M., Nieto, J., Van der Oost, J., Planas, A.

Archaea **1**, 285-292

Abstract

Pyrococcus furiosus laminarinase (LamA, PF0076) is an endo-glycosidase that hydrolyzes β -1,3-gluco-oligosaccharides, but not β -1,4-gluco-oligosaccharides. We studied the specificity of LamA towards small saccharides by using 4-methylumbelliferyl β -glucosides with different linkages. Besides endo-activity, wild-type LamA has some exo-activity, and catalyzes the hydrolysis of mixed-linked oligosaccharides (Glc β 4Glc β 3Glc β -MU (Glc = glucosyl, MU = 4-methylumbelliferyl)) with both β -1,4 and β -1,3 specificities. The LamA mutant E170A had severely reduced hydrolytic activity, which is consistent with Glu170 being the catalytic nucleophile. The E170A mutant was active as a glycosynthase, catalyzing the condensation of α -laminaribiosyl fluoride to different acceptors. The best condensation yields were found at pH 6.5 and 50 °C, but did not exceed 30%. Depending on the acceptor, the synthase generated either a β -1,3 or a β -1,4 linkage.

Introduction

For the biocatalytic production of oligosaccharides, glycoside hydrolases are of interest because of their generally simple heterologous production and the availability of low-cost substrates. Synthesis of oligosaccharides is achieved by thermodynamically controlled reverse hydrolysis and kinetically controlled transglycosylation, which both require high substrate concentrations and low water activity, often achieved by elevated temperatures and the addition of organic solvents. Final yields, however, do not exceed 50% because the product is a substrate of the enzyme's hydrolase activity. Increased yields have been obtained at higher temperatures with thermophilic enzymes (Boon et al. 1998, Kaper et al. 2001). Currently, a maximal yield can be accomplished only by using glycosynthases, glycoside hydrolases in which the catalytic nucleophile has been replaced by a small non-catalytic amino acid residue (Mackenzie et al. 1998). Glycosynthases couple activated glycosyl donors, which have the opposite anomeric configuration to that of the normal substrate, to a range of acceptors. Because these enzymes cannot hydrolyze the product, they carry out transglycosylation efficiently (Fig. 3.1). Typical product yields are 65-80%, but some studies have reported yields of up to 100% (for reviews, see Moracci et al. 2001, Planas and Faijes 2002, Williams and Withers 2002). The glycosynthase strategy has now been applied to five exo- β -glycosidases (Mackenzie et al. 1998, Trincone et al. 2000, Nashiru et al. 2001, Jakeman and Withers 2002, Perugino et al. 2003), four endo- β -glucanases (Malet and Planas 1998, Fort et al. 2000, Hrmova et al. 2002, Jahn et al. 2003) and an exo- α -glucosidase (Okuyama et al. 2002). The combination of two endo-glycosynthases resulted in the efficient synthesis of hexasaccharides through the sequential transglycosylation of different disaccharides (Faijes et al. 2001).

The β -glucansynthase from *Bacillus licheniformis* is an efficient enzyme, mainly because of its very strict regio- and stereospecificity (Malet and Planas 1998, Faijes et al. 2003), traits that are also observed in the original wild-type hydrolase (Malet and Planas 1997, Planas 2000). The endo-

β -1,3-1,4-glucanase hydrolyzes the β -1,4 glycosidic bond in a 3-O-substituted glucopyranosyl unit of mixed-linked glucans (Planas 2000). The enzyme is classified in Family 16 of glycoside hydrolases (Henrissat 1991), which also comprises endo- β -1,3-glucanases. Both types of Family 16 enzymes have an amino acid homology of 15-30%. The endo- β -1,3-1,4-glucanases, also called laminarinases, hydrolyze mixed-linked glucans such as lichenan, but prefer β -1,3-glucans such as laminarin, in contrast to the endo- β -1,3-1,4-glucanases, or lichenases. Therefore, we were interested in the behavior of a laminarinase, not only as a hydrolase, but also as a glycosynthase, particularly with respect to its specificity and efficiency.

The laminarinase (LamA, PF0076) from *Pyrococcus furiosus* is an extremely thermostable ($t_{1/2}$ 100 °C = 16 h) and thermoactive (T_{opt} = 100-105 °C) endo- β -1,3-glucanase (Gueguen et al. 1997, Chiaraluce et al. 2002) that has been reported to hydrolyze laminarin-oligosaccharides, but not cello-oligosaccharides and polysaccharides (Driskill et al. 1999). Hydrolysis of mixed linked oligosaccharides has never been reported. However, polysaccharides containing both β -1,3 and β -1,4 linkages are hydrolyzed by LamA to a certain extent, but the glycosidic linkages of the products have not been identified (Gueguen et al. 1997). The same holds true for products generated in glycosylation reactions with laminarin-oligosaccharides (Driskill et al. 1999). Therefore, it is still unclear whether the enzyme hydrolyzes or synthesizes β -1,3 glycosidic linkages only or also β -1,4 bonds. Here we report on the hydrolytic specificity of LamA towards a series of small synthetic 4-methylumbelliferyl (MU) glucosides. In addition, the glycosynthase activity of a LamA nucleophile mutant was studied and compared with the activity of 1,3-1,4- β -glucansynthase from *B. licheniformis*.

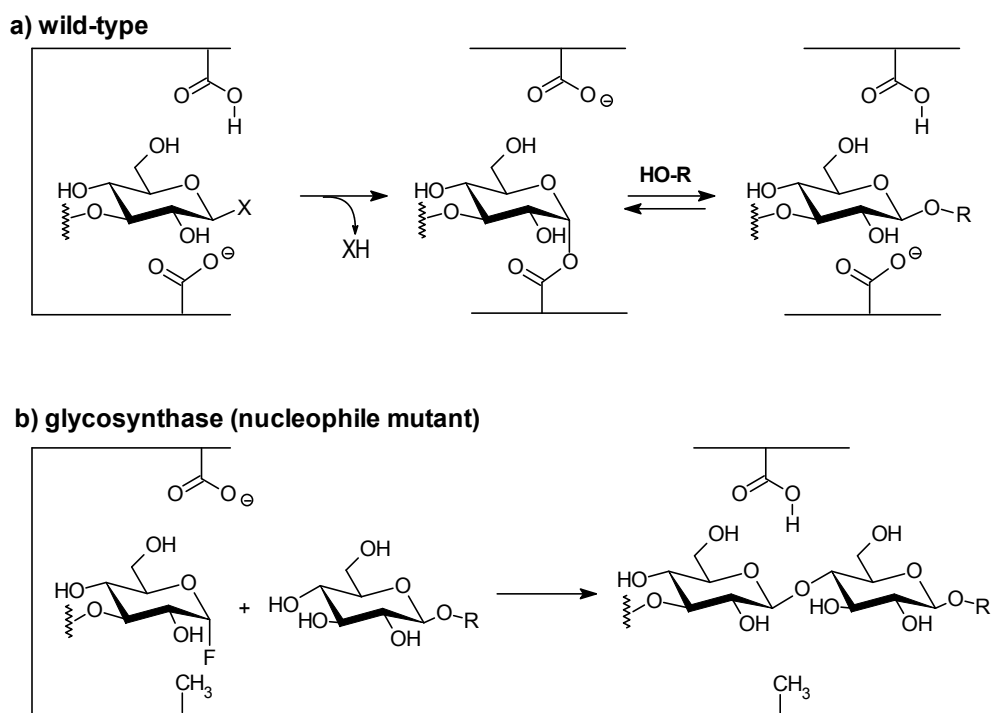


Figure 3.1 Transglycosylation and glycosynthase reactions catalyzed by retaining glycosidases. **(A)** Transglycosylation or hydrolysis (R = hydrogen) catalyzed by a wild-type glycosidase. **(B)** Glycosynthase reaction catalyzed by a nucleophile mutant glycosidase.

Material and Methods

General - NMR spectra were recorded on a Varian Gemini-300 spectrometer (Varian, Palo Alto, CA) in deuterium oxide (D_2O). Chemical shifts (δ in ppm) were referenced to residual DMSO (δ_H 2.49) or d_6 DMSO (δ_C 39.7) as internal standards for 1H or ^{13}C NMR, respectively. Mass spectra were recorded on a VG Autospec spectrometer (Waters-Micromass Technologies, Milford, MA) by the fast atomic bombardment (FAB) technique with an NaCl matrix.

Enzymes - The LamA mutant, E170A, was prepared by overlap extension polymerase chain reaction (PCR) (Ho et al. 1989) with the primers BG892, 5'-AAATTGTGGAGCGATCGACATATGGAGT-3', (sense) and BG893, 5'-ACTCCATTATGTCTATCGCTCCACAATTT-3', (antisense). Both recombinant wild-type and E170A mutant LamA proteins were expressed in *Escherichia coli* BL21 (DE3) (Stratagene, La Jolla, CA) and purified as previously reported (Gueguen et al. 1997, Kaper et al. 2001). The purity of the enzyme preparations was greater than 95% as judged by SDS-PAGE, according to Laemmli (Laemmli 1970). Enzyme concentrations were determined by UV spectrometry based on an $\epsilon_{280} = 83,070 \text{ M}^{-1}\text{cm}^{-1}$ and a molecular weight of $30,085 \text{ g mol}^{-1}$, as calculated by ProtParam (Gill and Von Hippel 1989).

Substrates - Substrates used in this study are shown in Fig. 3.2. 4-Methylumbelliferyl β -laminaribioside (**1**), 4-methylumbelliferyl 3-*O*- β -cellobiosyl- β -D-glucopyranoside (**2**), and 4-methylumbelliferyl 3-*O*- β -cellotriosyl- β -D-glucopyranoside (**3**) were synthesized as reported previously (Malet et al. 1995). 4-Methylumbelliferyl β -cellobioside (**4**) was obtained as described by Van Tilbeurgh et al. (1982), and 4-methylumbelliferyl β -D-glucoside (**5**) was obtained from Fluka (Buchs, Switzerland). α -Laminaribiosyl fluoride (**6**) was prepared as described by Malet and Planas (Malet and Planas 1998) by reacting laminaribiose peracetate with hydrogen fluoride in pyridine, followed by de-*O*-acetylation with sodium methoxide in methanol.

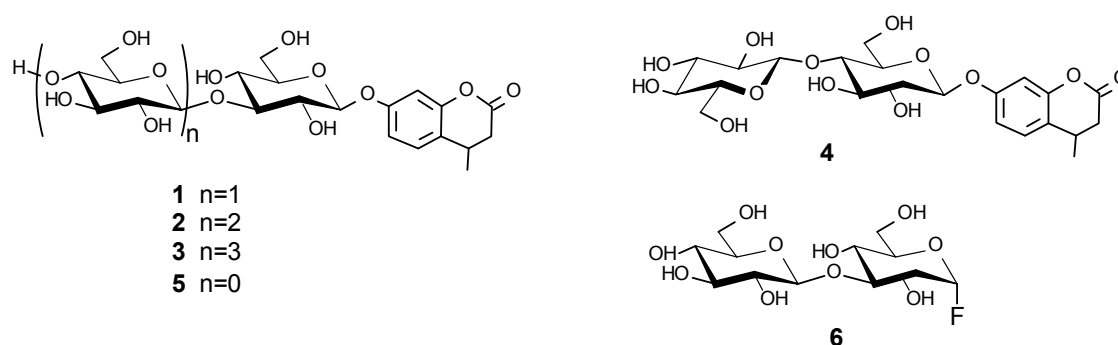


Figure 3.2 Substrates for wild-type and E170A LamA used in this study: 1 = 4-methylumbelliferyl β -laminaribioside (Glc β 3Glc β -MU); 2 = 3-*O*- β -cellobiosyl- β -D-glucopyranoside (Glc β 4Glc β 3Glc β -MU); 3 = 3-*O*- β -cellotriosyl- β -D-glucopyranoside (Glc β 4Glc β 4Glc β 3Glc β -MU); 4 = 4-methylumbelliferyl β -cellobioside (Glc β 4Glc β -MU); 5 = 4-methylumbelliferyl β -D-glucoside (Glc β -MU); 6 = α -laminaribiosyl fluoride (Glc β 3Glc α F).

Hydrolytic activity of wt and E170A enzymes

Kinetics with laminarin as substrate - Reactions were performed with 0.5% laminarin in 100 mM sodium phosphate buffer, pH 6.5, and 0.02 μM wild-type enzyme or 1 μM E170A mutant, at 80°C. Aliquots of the reaction were withdrawn at 10-min intervals, and the increase in reducing power was determined by the dinitrosalicylic (DNS) method (Miller 1959), with glucose as the standard.

Kinetics with 4-methylumbelliferyl glycoside substrates - Kinetics with Substrates **1** and **2** were performed by measuring changes in UV absorbance at 365 nm due to the release of 4-methylumbelliferone with a Varian Cary 4 spectrophotometer with a Peltier temperature control system. Rates of enzyme-catalyzed hydrolysis were determined by incubating the substrate (0 - 8 mM) in phosphate buffer (100 mM, pH 6.5) for 5 min in a thermostated cell holder at 80°C, followed by addition of the corresponding enzyme (0.02 μM for wild-type or 1 μM for E170A) and monitoring the absorbance change at $\lambda = 365$ nm (molar extinction coefficient, $\Delta\epsilon = 4452 \text{ M}^{-1}\text{cm}^{-1}$).

High performance liquid chromatography (HPLC) - Reaction mixtures containing the substrate (4 mM) and wild-type LamA (20-200 nM) in phosphate buffer (100 mM, pH 6.5) were incubated at 80°C. Aliquots were taken at different time intervals and analyzed by HPLC in a Nova-Pak C18 column (4μm, 3.9x150 mm) (Waters, Milford, MA) The flow rate was 1 mL min⁻¹, with 16% MeOH in 50 mM phosphate buffer (pH 6.5) as the mobile phase, and detection at 316 nm. Only products containing the 4-methylumbelliferyl chromophore were detected at this wavelength. Yields were calculated from the peak areas based on the corresponding response factor determined for free 4-methylumbelliferone (8770 area mM⁻¹) and for 4-methylumbelliferyl glycosides (7603 area mM⁻¹), respectively. Chromatographic peaks were identified by co-injection with independent standards.

Glycosynthase activity of E170A LamA - α-Laminaribiosyl fluoride (**6**) (1 mM) and different glycosyl acceptors (5 mM) were dissolved in phosphate buffer (100 mM, pH 6.5) and pre-incubated for 5 min. The E170A mutant (2-7.5 μM) was added and the reactions were incubated at 35, 50 or 76°C). Aliquots were taken at different time intervals, diluted with H₂O and analyzed by HPLC as described previously (mobile phase 16-18%MeOH). New chromatographic peaks were first tentatively identified by co-injection with independent standards. Control reactions without enzyme were performed to assess the absence of any uncatalyzed condensation. The thermal stability of α-laminaribiosyl fluoride (**6**) was analyzed in phosphate buffer (100 mM, pH 6.5) by monitoring fluoride anion release with a fluoride selective electrode (Sentek, Essex, U.K.), interfaced with a CiberScan Bench pH/ion meter (Eutech Instruments, Nijkerk, The Netherlands) as previously reported (Faijes et al. 2003).

Preparative glycosynthase reactions - α-Laminaribiosyl fluoride donor (**6**) (10.6 mg, 0.031 mmol, 1 equivalent), acceptor **1** or **2** (5 equivalents), and E170A LamA mutant (0.6 mg) were dissolved in phosphate buffer (100mM, pH 6.5, 5.5 ml) and incubated at 50 °C for 2 days. Samples

(10 μ L) were withdrawn, diluted 1:10 with H₂O and analyzed by HPLC (Nova-Pak C18, MeOH 18%). During the reaction, more donor was added (final conditions: 6.9 ml, 40 mg donor). After centrifugation to remove traces of precipitate, the reaction mixture was loaded directly onto a reverse-phase Lichoprep RP-18 Lobar-A column (Merck, Whitehouse Station, NJ) and eluted with an H₂O:MeOH gradient from 0 to 25%. Fractions were freeze-dried.

a) Glc β 3Glc α F (6) + Glc β -MU (5) glycosynthase reaction - Unreacted acceptor was eluted first. Then, the trisaccharide Glc β 3Glc β 4Glc β -MU was obtained: m/z (FAB) 685 [M+Na]⁺. ¹H NMR (D₂O, 30 °C) δ 2.42 (s, 3 H, CH₃), 3.34-4.08 (m, 18H, H-2^{I-III}, H-3^{I-III}, H-4^{I-III}, H-5^{I-III}, H-6a^{I-III}, H-6b^{I-III}), 4.60 (d, $J_{1,2}$ = 8.1 Hz, 1H), H-1^{II}, 4.80 (d, 1H, H-1^{III}), 5.23 (d, $J_{1,2}$ = 7.5 Hz, 1H, H-1^I), 6.22 (s, 1H, H-3'), 7.04 (s, 1H, H-8'), 7.10 (d, $J_{5',6'}$ = 9.6 Hz, 1H, H-6'), 7.69 (d, $J_{5',6'}$ = 9.0 Hz, 1H, H-5'). ¹³C NMR (D₂O, 30 °C) δ 18.6 (CH₃), 60.4-61.4 (C-6^{I-III}), 68.7, 70.2 (C-4^{II,III}), 73.2-76.6 (C-2^{I-III}, C-3^{I, III}, C-5^{I-III}), 78.8 (C-4^I), 84.7 (C-3^{II}), 100.2-104.2 (C-1^{I-III}, C-8'), 111.9 (C-3'), 114.5 (C-6'), 116.0 (C-4a'), 127.3 (C-5'), 154.5 (C-4'), 156.8 (C-1a'), 160.1 (C-7'), 165.2 (C-2').

The third fraction was a minor compound tentatively assigned as the trisaccharide Glc3Glc3Glc β -MU. It was treated with wild-type LamA and the disaccharide Glc β 3Glc β -MU was detected by HPLC.

(b) Glc β 3Glc α F (6) + Glc β 3Glc β -MU (1) glycosynthase reaction - First, unreacted acceptor was eluted, followed by the tetrasaccharide Glc β 3Glc β 3Glc β 3Glc β -MU: m/z (FAB): 847.251 [M+Na]⁺ (calculated 847.248). ¹H NMR (D₂O, 30 °C) δ 2.42 (s, 3 H, CH₃), 3.34-4.01 (m, 24H, H-2^{I-IV}, H-3^{I-IV}, H-4^{I-IV}, H-5^{I-IV}, H-6a^{I-IV}, H-6b^{I-IV}), 4.76-4.87 (3H, H-1^{I-III}), 5.23 (d, $J_{1,2}$ = 7.5 Hz, 1H, H-1^I), 6.23 (s, 1H, H-3'), 7.06 (s, 1H, H-8'), 7.10 (d, $J_{5',6'}$ = 9.6 Hz, 1H, H-6'), 7.69 (d, $J_{5',6'}$ = 9.0 Hz, 1H, H-5'). ¹³C NMR (D₂O, 30 °C) δ 18.7 (CH₃), 61.0, 61.3 (C-6^{I-IV}), 68.4-70.5, (C-4^{I-IV}), 73.2-76.6 (C-2^{I-IV}, C-3^{IV}, C-5^{I-IV}), 84.4-84.8 (C-3^{I-III}), 100.0-104.2 (C-1^{I-III}, C-8'), 111.8 (C-3'), 114.4 (C-6'), 115.8 (C-4a'), 127.2 (C-5'), 154.3 (C-4'), 156.6 (C-1a'), 159.9 (C-7'), 164.9 (C-2').

A second condensation product corresponding to the tetrasaccharide Glc β 3Glc β 4Glc β 3Glc β -MU was isolated: m/z : calculated for C₃₄H₄₈O₂₃Na: 847.250; found: 847.248. ¹H and ¹³C NMR are identical to those reported in Faijes et al. (2003).

Results and discussion

Hydrolase activity and specificity of wild type LamA

The glucoside **5**, and the cellobioside **4** were not hydrolyzed by wild-type LamA, whereas the 4-methylumbelliferyl laminaribioside **1** and the mixed-linked oligosaccharides Glc β 4Glc β 3Glc β -MU (**2**) and Glc β 4Glc β 4Glc β 3Glc β -MU (**3**) were substrates. Cleavage specificity was determined by HPLC (Figures 3A-C). Only chromophoric products (free MU and MU-glycosides) were analyzed (UV detection at 316 nm), and identified by co-injection with independent standards. Reactions

were performed at low enzyme concentrations to extend the time course and allow the detection of transient intermediates.

Hydrolysis of 4-methylumbelliferyl laminaribioside (Glc β 3Glc β -MU) (**1**) by LamA produced an MU concentration of up to 70% of the initial substrate concentration, and 15% Glc β -MU (**5**) was formed after 400 min (10% Glc β 3Glc β -MU substrate was not hydrolyzed). Therefore, we concluded that the laminaribioside **1** was hydrolyzed at the glycosidic bond with the aglycon and at the β -1,3 bond with a cleavage ratio of 4.5 to 1. The disaccharide isomer, Glc β 4Glc β -MU (**4**), was not hydrolyzed by LamA, indicating that the enzyme does not accept β -1,4 linkages between the glucosyl units in subsites -1 and -2, but accepts β -1,3 linkages in this position. (Enzyme subsites are defined as the set of amino acid residues in the binding site (cleft) that interact with a monosaccharyl unit of the oligomeric substrate, and are numbered as -1, -2, ..., -n, for subsites on the non-reducing end, and +1, +2, ..., +n, on the reducing end of the saccharide ligand from the site of cleavage (scissile glycosidic bond).)

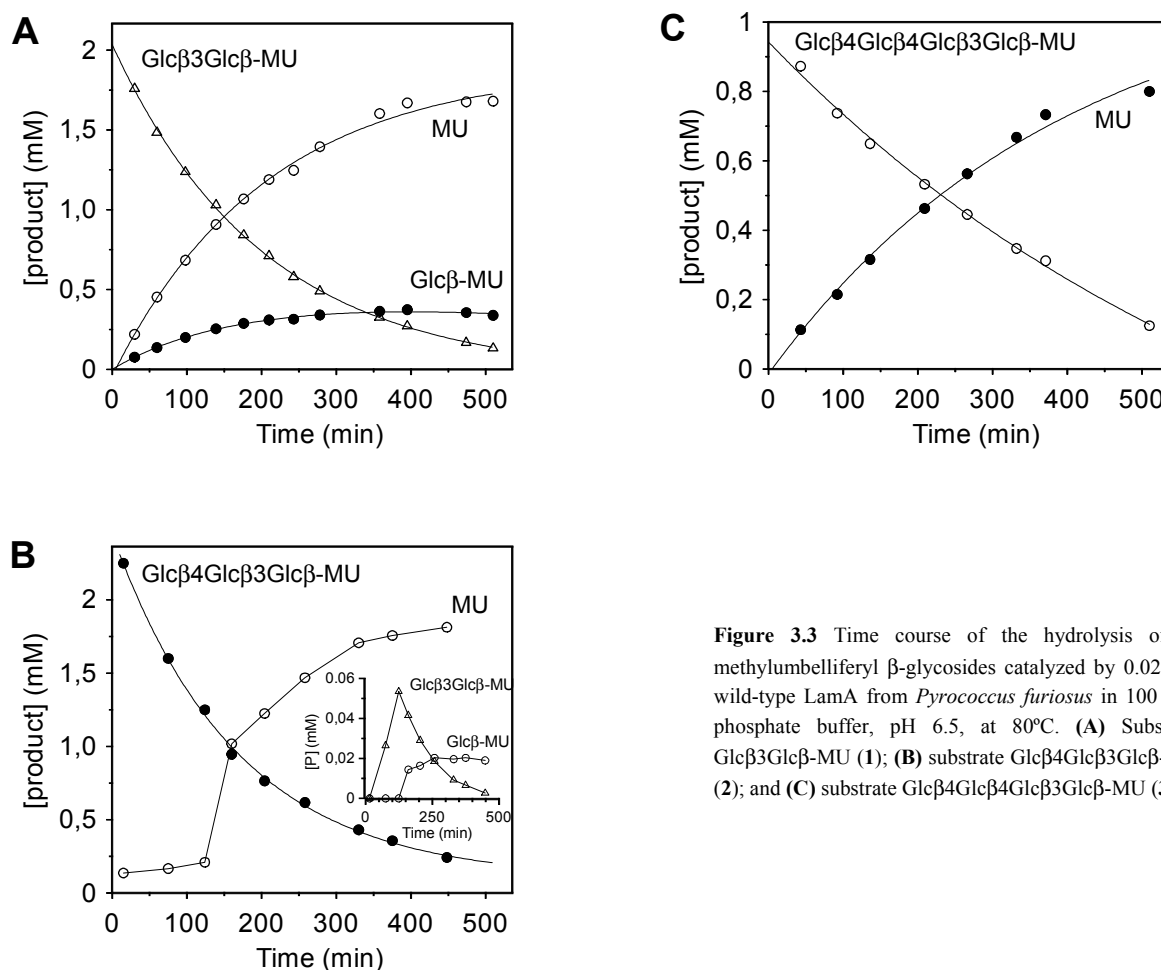


Figure 3.3 Time course of the hydrolysis of 4-methylumbelliferyl β -glycosides catalyzed by 0.02 μ M wild-type LamA from *Pyrococcus furiosus* in 100 mM phosphate buffer, pH 6.5, at 80°C. **(A)** Substrate Glc β 3Glc β -MU (**1**); **(B)** substrate Glc β 4Glc β 3Glc β -MU (**2**); and **(C)** substrate Glc β 4Glc β 4Glc β 3Glc β -MU (**3**).

With the Glc β 4Glc β 3Glc β -MU (**2**) substrate (Fig. 3.3B), Glc β 3Glc β -MU (**1**) was initially formed (up to 3%), but was then hydrolyzed, and Glc-MU (1%) and MU (85%) were the final hydrolysis products after 400 min of reaction (10% of the substrate was not hydrolyzed). Because Glc β 4Glc β -

MU (**4**) was not hydrolyzed, the Glc β -MU formed as the final product was unlikely to have been formed by cleavage of the β -1,3 bond of the trisaccharide, but was likely formed by hydrolysis of the initial disaccharide product **1**. Formation of this transient disaccharide indicates that the enzyme can also hydrolyze the β -1,4 glycosidic bond of the β 1,4- β 1,3-trisaccharide substrate. Finally, the tetrasaccharide Glc β 4Glc β 4Glc β 3Glc β -MU (**3**) was hydrolyzed with the release of MU (Fig. 3.3C), and no other shorter chromophoric oligosaccharides were detected.

Even though LamA is an endo-glycosidase, as previously established for the enzyme-catalyzed hydrolysis of β -1,3-glucans (Gueguen et al. 1997), the cleavage pattern for Glc β 3Glc β -MU (**1**) and Glc β 4Glc β 3Glc β -MU (**2**) shows that it has some detectable exo-activity on simple oligosaccharides, with both β -1,4 and β -1,3 specificity. Comparison with other laminarinases indicates that this behaviour is unique. However, it should be noted that studies on substrate specificity of laminarinases are not as extensive as those on lichenases from Family 16. LamR from *Rhodothermus marinus* has endo-activity on laminarin, but shows some exo-activity on oligosaccharides (Petersen et al. 2000, Borriss et al. 2003). Additionally, both β -1,3 and β -1,4 specificities have been reported for LamR; the enzyme hydrolyzes mixed-linked β -1,3-1,4 glucans with β -1,4 glycosidic bonds with higher specific activity than β -1,3 glucans. Moreover, activity on cello-oligosaccharides, but not on polysaccharides, has also been reported (Petersen et al. 2000). Hydrolysis of β -1,4 bonds of carboxymethyl cellulose has been observed with the putative laminarinases GluA and GluC from *Lysobacter enzymogenes* (Palumbo et al. 2003). However, the lichenase Lic16A from *Clostridium thermocellum* hydrolyzes lichenan and barley β -glucan, but with β -1,3 cleavage specificity (Fuchs et al. 2003). These examples illustrate that the specificities of laminarinases and lichenases in Family 16 of glycoside hydrolases cover a broad spectrum and require reclassification based on sequence similarity and substrate specificity criteria.

The substrate Glc β 3Glc β -MU (**1**) was selected as the reference substrate for the wild-type LamA glycosidase activity because it is preferentially hydrolyzed with release of the chromophoric aglycon, MU. The reaction follows Michaelis-Menten kinetics (Fig. 3.4) with $k_{\text{cat}} = 4.22 \pm 1.69 \text{ s}^{-1}$ and $K_{\text{M}} = 0.93 \pm 0.79 \text{ mM}$.

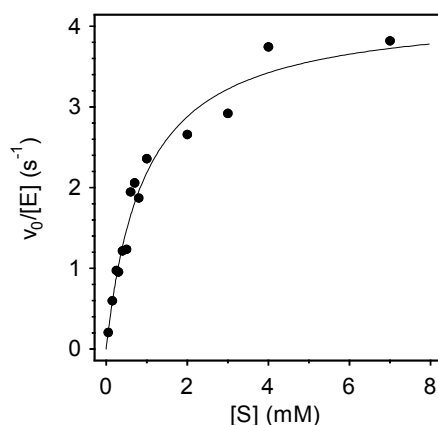


Figure 3.4 Kinetics of hydrolysis of Glc β 3Glc β -MU (**1**) catalyzed by 0.02 μM wild-type LamA from *Pyrococcus furiosus* in 100 mM phosphate buffer, pH 6.5, at 80°C. Abbreviations: v_0 = initial velocity; [E] = enzyme concentration; and [S] = substrate concentration.

Hydrolase and Glycosynthase activity of E170A

Residue Glu170 of *Pyrococcus furiosus* laminarinase is proposed to act as the catalytic nucleophile (Gueguen et al. 1997), based on sequence similarity with other family 16 glycosyl hydrolases for which the catalytic residues have been experimentally identified (Juncosa et al. 1994, Viladot et al. 1998, Planas 2000). The Glu170 residue was replaced by alanine to produce the mutant LamA, E170A, to confirm the role of Glu170 as an essential catalytic residue in the hydrolytic mechanism, and to evaluate its potential as a glycosynthase.

Table 3.1. Hydrolase activity of wild-type and E170A LamA on laminarin and 4-methylumbelliferyl β -glycosides. Reactions took place at 80°C in 100 mM phosphate buffer, pH 6.5, containing 0.5% laminarin, or 4 mM MU-glycoside. Activity was determined by the DNS method or as release of the chromophoric aglycon (MU = 4-methylumbelliferone).

Substrate	Wild-type	E170A
Laminarin	832 U mg ⁻¹	0.46 U mg ⁻¹
Glc β 3Glc β -MU (1)	3.75 s ⁻¹	4·10 ⁻⁴ s ⁻¹
Glc β 4Glc β 3Glc β -MU (2)	0.41 s ⁻¹	1.5·10 ⁻⁴ s ⁻¹
Glc β 4Glc β -MU (4)	< 10 ⁻⁴ s ⁻¹	< 10 ⁻⁴ s ⁻¹
Glc β -MU (5)	< 10 ⁻⁴ s ⁻¹	< 10 ⁻⁴ s ⁻¹

The hydrolase activity of the E170A mutant was measured with laminarin and several 4-methylumbelliferyl β -glycosides as substrates, and compared with the hydrolase activity of the wild-type enzyme (Table 3.1). With laminarin, the E170A mutant retained only 0.05% activity relative to the wild-type enzyme, whereas a 10⁴-fold reduction was obtained with the laminaribioside substrate **1**. This effect is consistent with Glu170 being the catalytic nucleophile.

Table 3.2. Glycosynthase-catalyzed condensations of α -laminaribiosyl fluoride donor (**6**) with acceptors **5** and **1**. Conditions: 100 mM phosphate buffer, pH 6.5, 50 °C, [donor] = 1 mM, [acceptor] = 5 mM, [enzyme] = 7.5 μ M. Yields were determined by HPLC after 24 h reaction. Boldface numbers indicate the linkage formed in the condensation reaction.

Acceptor	Product	Yield (%)
Glc β -MU (5)	Glc β 3Glc β 4 Glc β -MU (7)	20
	Glc β 3Glc β 3 Glc β -MU (8)	3
Glc β 3Glc β -MU (1)	Glc β 3Glc β 4 Glc β 3Glc β -MU (9)	2
	Glc β 3Glc β 3 Glc β 3Glc β -MU (10)	25

Because a laminaribiosyl unit binds productively in subsite -2/-1 (as deduced from hydrolysis of Glc β 3Glc β -MU (**1**) by wild-type LamA), α -laminaribiosyl fluoride (**6**) was used as the glycosyl donor to determine whether the E170A mutant has glycosynthase activity with different acceptors. Reactions were monitored by HPLC to detect MU-glycoside products. With Glc β -MU (**5**) as the acceptor and a reaction temperature of 35°C, a trisaccharide product (yield \approx 2% after 3h reaction) was obtained. When the reaction was performed at 50°C, the same trisaccharide was produced with a maximum yield of 20% at 24 h (4.5% after 3h reaction). The yield did not increase with longer incubation or with an increase in temperature to 76°C, at which the enzyme is probably more active because of its thermophilic nature. The initial rate of condensation was higher at 76°C ($17.7 \times 10^{-4} \text{ s}^{-1}$) than at 50 °C ($5.7 \times 10^{-4} \text{ s}^{-1}$), but the yield of condensation product did not exceed 10% at the higher temperature, most likely because of spontaneous hydrolysis of the α -glycosyl fluoride donor (Faijes et al. 2003). The reaction at 50°C was performed on a semi-preparative scale (10 mg of donor, 0.6 mg enzyme, two equivalents of acceptor) (Table 3.2). Two main products were isolated after chromatographic separation of the reaction mixture. The first (20% yield, as determined by HPLC) was identified as Glc β 3Glc β 4Glc β -MU (**7**) by NMR spectroscopy: a new signal at δ 4.60 ppm ($J = 8.1 \text{ Hz}$), characteristic of H-1 for a β -1,4 glycosidic bond, was observed in the ^1H NMR spectrum, whereas the ^{13}C NMR spectrum showed the 4 I and 3 II carbons involved in glycosidic linkages at δ 78.8 and 84.7 ppm, respectively. This product was identical to that reported for the glycosynthase reaction of the E134A mutant of *Bacillus licheniformis* endo-1,3-1,4- β -glucanase with the same donor and acceptor molecules (Malet and Planas 1998). A second product (3% yield, as determined by HPLC) was isolated; it was also a trisaccharide, tentatively assigned as Glc β 3Glc β 3Glc β -MU (**8**), because addition of wild-type LamA to the glycosynthase reaction mixture yielded Glc β 3Glc β -MU (**1**) as well as Glc β 4Glc β -MU (**4**), resulting from hydrolysis of **8** and **7**, respectively.

The reaction with E170A and Glc β 3Glc β -MU (**1**) acceptor was also analyzed at 50°C (Table 3.2). Two tetrasaccharide products were detected by HPLC with an overall initial rate ($v_0/[E]$) of $16.1 \times 10^{-2} \text{ s}^{-1}$ and 27% yield after 10 h (Fig. 3.5A). To elucidate the regioselectivity of the glycosynthase reaction, the mixture was digested with two glycosidases: LamA (1,3- β -glucanase) and LicA (endo-1,3-1,4- β -glucanase from *B. licheniformis*), which only hydrolyzes β -1,4 glycosidic bonds on 3-*O*-substituted glucosyl units (Planas 2000). Analysis by HPLC revealed that LicA degraded only the minor tetrasaccharide product of the glycosynthase reaction (Fig. 3.5C), indicating that the newly formed glycosidic bond was β -1,4 (Glc β 3Glc β 4Glc β 3Glc β -MU (**9**)). In contrast, LamA degraded the major tetrasaccharide product (Fig. 3.5D), which was then assigned to Glc β 3Glc β 3Glc β 3Glc β -MU (**10**). The structure of both tetrasaccharides was confirmed by ^1H and ^{13}C NMR spectroscopy. Tetrasaccharide **9** was identical to the product reported for the glycosynthase reaction of the E134A mutant 1,3-1,4- β -glucanase (Faijes et al. 2003). The major tetrasaccharide **10** (25% yield) was unambiguously identified: the three signals corresponding to H-1 in inter-glucosidic bonds were in the δ 4.76-4.87 ppm region in the ^1H NMR spectrum (characteristic of β -1,3 bonds), whereas the ^{13}C NMR spectrum showed that three C-3 carbons were involved in glycosidic bonds (δ 84.4-84.8 ppm).

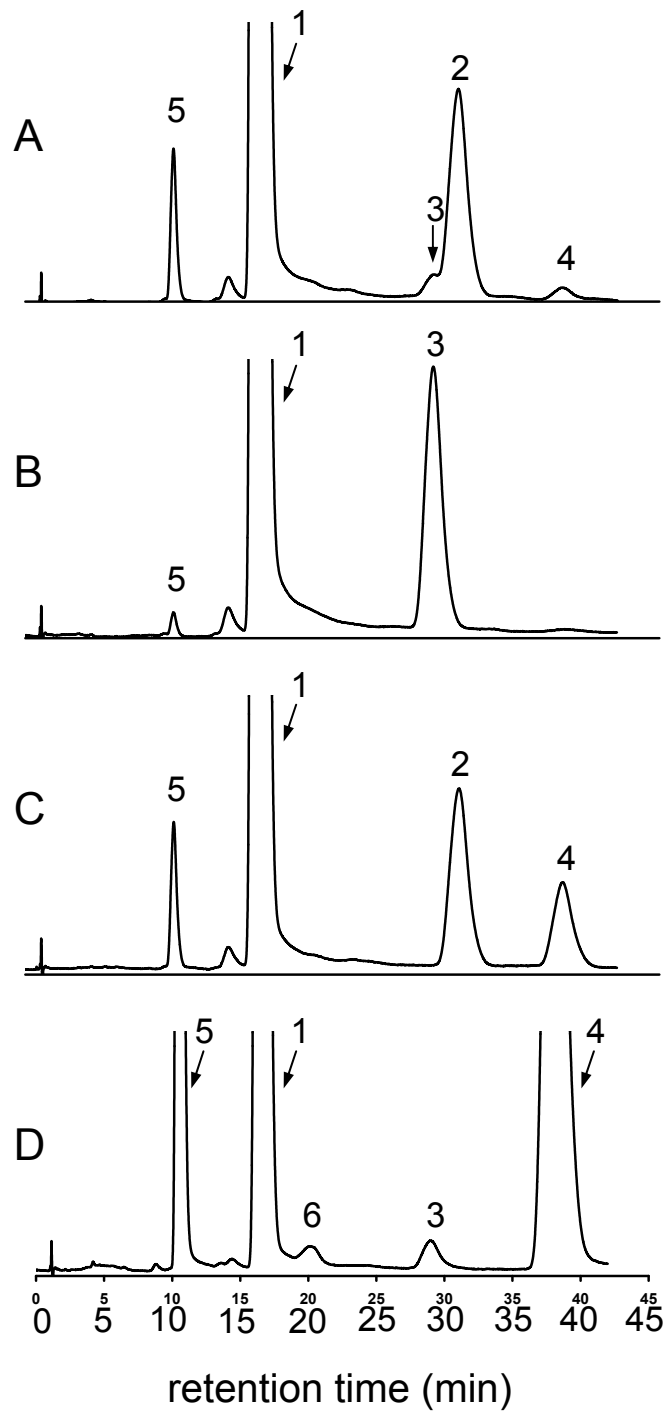


Figure 3.5 The HPLC analysis of reaction products. **(A)** glycosynthase reaction of E170A LamA with α -laminaribiosyl fluoride (**6**) and Glc β 3Glc β -MU (**1**); **(B)** glycosynthase reaction of E134A (endo- β -1,3-1,4-glucanase from *B. licheniformis*) with α -laminaribiosyl fluoride (**6**) and Glc β 3Glc β -MU (**1**); **(C)** digestion of the reaction mixture A with wild-type endo- β -1,3-1,4-glucanase from *B. licheniformis*; **(D)** digestion of the reaction mixture A with wild-type LamA. Peaks in the chromatogram were identified by coinjection with independent standards. Peak labels: 1. Glc β 3Glc-MU (**1**) with retention time, t_R = 17.2 min; 2. Glc β 3Glc β 3Glc β 3Glc-MU (**10**), t_R = 31.1 min; 3. Glc β 3Glc β 4Glc β 3Glc-MU (**9**), t_R = 29.5 min; 4. MU, t_R = 38.8 min; 5. Glc β -MU (**5**), t_R = 10.5 min; 6. unidentified hydrolysis product, t_R = 20.1 min.

In conclusion, LamA hydrolyzed mixed-linked oligosaccharides, confirming that the laminaribiose moiety is the preferred substrate unit; however, the enzyme also showed exo-activity, hydrolysing both β -1,3 and β -1,4 linkages. The E170A LamA mutant possessed glycosynthase activity, catalyzing the condensation of α -laminaribiosyl fluoride (**6**) donor with Glc β -MU (**5**) and Glc β 3Glc β -MU (**1**) as acceptors. The reaction was not regiospecific because both β -1,3 and β -1,4 glycosidic bonds were formed; with the monosaccharide acceptor **5**, a 7:1 ratio of β -1,4 to β -1,3 was obtained, whereas a 1:17 ratio was observed with the disaccharide acceptor **1**. These results contrast those reported for a Family 17 1,3- β -glucanase from barley (Hrmova et al. 2002), in which the corresponding glycosynthase mutant (E231A) is regiospecific for β -1,3 glycosidic linkages. Family 16 laminarinases seem to include a wider range of specificities. Transglycosylation catalyzed by wild-type laminarinases from *Oerskovia* sp. and *Spisula sachalinensis* exclusively produced β -1,3 linkages (Borriss et al. 2003), whereas LamR from *R. marinus* catalyzed the synthesis of both β -1,3 and β -1,4 bonds (Krah et al. 1998). The efficiency of E170A LamA as a glycosynthase is rather poor, at least with α -glycosyl fluoride as the activated donor, with preparative yields not exceeding 30% in condensation products (yields not optimized). This is probably a consequence of the hyperthermophilic nature of LamA from *Pyrococcus furiosus*. The maximum temperature for hydrolase activity of wild-type LamA is 100-105 °C, and activity drops to 10% at 50°C (Gueguen et al. 1997). The α -glycosyl fluoride donor is labile at high temperature and is readily hydrolyzed above 50°C. Therefore, the temperature used in this study is a compromise between donor stability and enzyme activity that renders conditions far from optimal.

Acknowledgments

This research was supported by the Technology Foundation (STW), applied science division of NWO and the technology program of the Ministry of Economic Affairs, The Netherlands, and grant BIO2001-2064-C02-02 from the Ministry of Science and Technology, Spain.

References

- Boon, M.A., J. Van Der Oost, W.M. De Vos, A.E.M. Jansen and K. Van 'T Riet. (1998) Synthesis of Oligosaccharides Catalyzed by Thermostable β -Glucosidase from *Pyrococcus furiosus*. *Appl. Biochem. Biotechnol.* **75**, 269-278.
- Borriss, R., M. Krah, H. Brumer III, M.A. Kerzhner, D.R. Ivanen, E.V. Eneyskaya, L.A. Elyakova, S.M. Shishlyannikov, K.A. Shabalin and K.N. Neustroev. (2003) Enzymatic Synthesis of 4-Methylumbelliferyl (1 \rightarrow 3)- β -D-Glucooligosaccharides. New Substrates for β -1,3-1,4-D-Glucanase. *Carbohydr. Res.* **338**, 1455-1467.
- Chiaraluce, R., J. Van Der Oost, J.H. Lebbink, T. Kaper and V. Consalvi. (2002) Persistence of Tertiary Structure in 7.9 M Guanidinium Chloride: The Case of Endo- β -1,3-Glucanase from *Pyrococcus furiosus*. *Biochemistry* **41**, 14624-14632.

- Driskill, L.E., M.W. Bauer and R.M. Kelly.** (1999) Synergistic Interactions among β -Laminarinase, β -1,4-Glucanase, and β -Glucosidase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* During Hydrolysis of β -1,4-, β -1,3-, and Mixed-Linked Polysaccharides. *Biotechnol. Bioeng.* **66**, 51-60.
- Faijes, M., J.K. Fairweather, H. Driguez and A. Planas.** (2001) Oligosaccharide Synthesis by Coupled Endo-Glycosynthases of Different Specificity: A Straightforward Preparation of Two Mixed-Linkage Hexasaccharide Substrates of 1,3/1,4- β -Glucanases. *Chemistry* **7**, 4651-4655.
- Faijes, M., X. Perez, O. Perez and A. Planas.** (2003) Glycosynthase Activity of *Bacillus licheniformis* 1,3-1,4- β -Glucanase Mutants: Specificity, Kinetics, and Mechanism. *Biochemistry* **42**, 13304-13318.
- Fort, S., V. Boyer, L. Greffe, G. Davies, O. Moroz, L. Christiansen, M. Schulein, S. Cottaz and H. Driguez.** (2000) Highly Efficient Synthesis of β (1 \rightarrow 4)-Oligo- and -Polysaccharides Using a Mutant Cellulase. *J. Am. Chem. Soc.* **122**, 5429-5437.
- Fuchs, K.P., V.V. Zverlov, G.A. Velikodvorskaya, F. Lottspeich and W.H. Schwarz.** (2003) Lic16a of *Clostridium Thermocellum*, a Non-Cellulosomal, Highly Complex Endo- β -1,3-Glucanase Bound to the Outer Cell Surface. *Microbiology* **149**, 1021-1031.
- Gill, S.C. and P.H. Von Hippel.** (1989) Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal. Biochem.* **182**, 319-326.
- Gueguen, Y., W.G. Voorhorst, J. Van Der Oost and W.M. De Vos.** (1997) Molecular and Biochemical Characterization of an Endo- β -1,3-Glucanase of the Hyperthermophilic Archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **272**, 31258-31264.
- Henrissat, B.** (1991) A Classification of Glycosyl Hydrolases Based on Amino Acid Sequence Similarities. *Biochem. J.* **280**, 309-316.
- Ho, S.N., H.D. Hunt, R.M. Horton, J.K. Pullen and L.R. Pease.** (1989) Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction. *Gene* **77**, 151-59.
- Hrmova, M., T. Imai, S.J. Ruten, J.K. Fairweather, L. Pelosi, V. Bulone, H. Driguez and G.B. Fincher.** (2002) Mutated Barley (1,3)- β -D-Glucan Endohydrolases Synthesize Crystalline (1,3)- β -D-Glucans. *J. Biol. Chem.* **277**, 30102-30111.
- Jahn, M., D. Stoll, R.A.J. Warren, L. Szabo, P. Singh, H.J. Gilbert, V.M. Ducros, G.J. Davies, and S.G. Withers.** (2003) Expansion of the glycosynthase repertoire to produce defined manno-oligosaccharides. *Chem. Comm.* **2003**, 1327-1329.
- Jakeman, D.L. and S.G. Withers.** (2002) On expanding the repertoire of glycosynthases: mutant β -galactosidases forming β -(1,6)-linkages. *Can. J. Chem.* **80**, 866-860.
- Juncosa, M., J. Pons, T. Dot, E. Querol and A. Planas.** (1994) Identification of active site carboxylic residues in *Bacillus licheniformis* endo-1,3-1,4- β -D-glucan 4-glucanohydrolase by site-directed mutagenesis. *J. Biol. Chem.* **269**, 14530-14535.
- Kaper, T., C.H. Verhees, J.H. Lebbink, J.F. Van Lieshout, L.D. Kluskens, D.E. Ward, S.W. Kengen, M.M. Beerthuyzen, W.M. De Vos and J. Van Der Oost.** (2001) Characterization of β -Glycosyl hydrolases from *Pyrococcus furiosus*. *Methods Enzymol.* **330**, 329-346.
- Krah, M., R. Misselwitz, O. Politz, K.K. Thomsen, H. Welfle and R. Borriss.** (1998) The Laminarinase from Thermophilic Eubacterium *Rhodothermus marinus*. Conformation, Stability, and Identification of Active Site Carboxylic Residues by Site-Directed Mutagenesis. *Eur. J. Biochem.* **257**, 101-111.
- Laemmli, U.K.** (1970) Cleavage of Structural Proteins During the Assembly of the Head of Bacteriophage T4. *Nature* **227**, 680-685.
- Mackenzie, L.F., Q. Wang, R.A.J. Warren and S.G. Withers.** (1998) Glycosynthases: Mutant Glycosidases for Oligosaccharide Synthesis. *J. Am. Chem. Soc.* **120**, 5583-5584.
- Malet, C. and A. Planas.** (1997) Mechanism of *Bacillus* 1,3-1,4- β -D-Glucan 4-Glucanohydrolases: Kinetics and pH Studies with 4-Methylumbelliferyl β -D-Glucan Oligosaccharides. *Biochemistry* **36**, 13838-13848.
- Malet, C. and A. Planas.** (1998) From β -Glucanase to β -Glucansynthase: Glycosyl Transfer to α -Glycosyl Fluorides Catalyzed by a Mutant Endoglucanase Lacking Its Catalytic Nucleophile. *FEBS Lett.* **440**, 208-212.

- Malet, C., J.L. Viladot, A. Ochoa, B. Gallego, C. Brosa and A. Planas.** (1995) Synthesis of 4-Methylumbelliferyl- β -D-Glucan Oligosaccharides as Specific Chromophoric Substrates of (1 \rightarrow 3), (1 \rightarrow 4)- β -D-Glucan 4-Glucanohydrolases. *Carbohydr. Res.* **274**, 285-301.
- Miller, G.L.** (1959) Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Anal. Chem.* **31**, 426-428.
- Moracci, M., A. Trincone and M. Rossi.** (2001) Glycosynthases: New Enzymes for Oligosaccharide Synthesis. *J. Mol. Catal. B Enzym.* **11**, 155-163.
- Nashiru, O., D.L. Zechel, D. Stoll, T. Mohammadzadeh, R.A. Warren and S.G. Withers.** (2001) β -Mannosynthase: Synthesis of β -Mannosides with a Mutant β -Mannosidase. *Angew. Chem. Int. Ed. Engl.* **40**, 417-420.
- Okuyama, M., H. Mori, K. Watanabe, A. Kimura and S. Chiba.** (2002) α -Glucosidase Mutant Catalyzes "Alpha-Glycosynthase"-Type Reaction. *Biosci. Biotechnol. Biochem.* **66**, 928-933.
- Palumbo, J.D., R.F. Sullivan and D.Y. Kobayashi.** (2003) Molecular Characterization and Expression in *Escherichia coli* of Three β -1,3-Glucanase Genes from *Lysobacter enzymogenes* Strain N4-7. *J. Bacteriol.* **185**, 4362-4370.
- Perugino, G., A. Trincone, A. Giordano, J. Van Der Oost, T. Kaper, M. Rossi and M. Moracci.** (2003) Activity of Hyperthermophilic Glycosynthases Is Significantly Enhanced at Acidic pH. *Biochemistry* **42**, 8484-8493.
- Petersen, B.O., M. Krah, J.O. Duus and K.K. Thomsen.** (2000) A Transglycosylating 1,3(4)- β -Glucanase from *Rhodothermus marinus*. NMR Analysis of Enzyme Reactions. *Eur. J. Biochem.* **267**, 361-369.
- Planas, A.** (2000) Bacterial 1,3-1,4- β -Glucanases: Structure, Function and Protein Engineering. *Biochim. Biophys. Acta* **1543**, 361-382.
- Planas, A. and M. Faijes.** (2002) Glycosidases and Glycosynthases in Enzymatic Synthesis of Oligosaccharides. An Overview. *Afinidad* **59**, 295-313.
- Trincone, A., G. Perugino, M. Rossi and M. Moracci.** (2000) A Novel Thermophilic Glycosynthase That Effects Branching Glycosylation. *Bioorg. Med. Chem. Lett.* **10**, 365-368.
- Van Tilbeurgh, H., M. Claeysens and C.K. De Bruyne.** (1982) The Use of 4-Methylumbelliferyl and Other Chromophoric Glycosides in the Study of Cellulolytic Enzymes. *FEBS Lett.* **149**, 152-156.
- Viladot, J.L., E. De Ramon, O. Durany and A. Planas.** (1998) Probing the Mechanism of *Bacillus* 1,3-1,4- β -D-Glucan 4-Glucanohydrolases by Chemical Rescue of Inactive Mutants at Catalytically Essential Residues. *Biochemistry* **37**, 11332-11342.
- Williams, S.J. and S.G. Withers.** (2002) Glycosynthases: Mutant Glycosidases for Glycoside Synthesis. *Aust. J. Chem.* **55**, 3-12.

Chapter 4

Calcium-induced Tertiary Structure Modifications of Endo- β -1,3-glucanase from *Pyrococcus furiosus* in 7.9 M Guanidinium Chloride

Chiaraluce, R., Gianese, G., Angelaccio, S., Florio, R., Van Lieshout, J.F.T., Van der Oost, J., Consalvi, V.

Biochemical Journal **386**, 515-524

Abstract

The family 16 endo- β -1,3 glucanase from the extremophilic archaeon *Pyrococcus furiosus* is a laminarinase which in 7.9 M guanidinium chloride (GdmCl) maintains a significant amount of tertiary structure without any change of secondary structure. The addition of calcium to the enzyme in 7.9 M GdmCl causes significant changes to the near-UV CD and fluorescence spectra, suggesting a notable increase in the tertiary structure which leads to a state comparable, but not identical, to the native state. The capability to interact with calcium in 7.9 M GdmCl with a consistent recovery of native tertiary structure is a unique property of this extremely stable endo- β -1,3 glucanase. The effect of calcium on the thermodynamic parameters relative to the GdmCl-induced equilibrium unfolding has been analyzed by circular dichroism and fluorescence spectroscopy. The interaction of calcium with the native form of the enzyme is studied by Fourier transformed infrared spectroscopy in the absorption region of carboxylate groups and by titration in the presence of a chromophoric chelator. A homology-based model of the enzyme is generated and used to predict the putative binding site(s) for calcium and the structural interactions potentially responsible for the unusual stability of this protein, in comparison with other family 16 glycoside hydrolases.

Introduction

Laminarinase endo- β -1,3 glucanase (E.C. 3.2.1.39) from the hyperthermophilic archaeon *Pyrococcus furiosus* (*pfl*LamA) maintains residual tertiary structure and intact secondary structure elements in 7.9 M guanidinium chloride (GdmCl) [1]. The presence of residual tertiary interactions under such extreme denaturing conditions is a peculiar property of this protein which makes it an attractive model to investigate the role played by non-covalent interactions as structural determinants of protein stability.

*pfl*LamA belongs to family 16 glycoside hydrolases [2] (see URL <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>). The enzymes in this family show a number of catalytic activities and are characterized by a similar β -jelly roll fold, which is better conserved than primary structure. A similar fold has also been demonstrated in proteins that have been classified in different families of glycoside hydrolases [3]. A considerable number of primary structures of family 16 are available, however among the few crystal structures solved, detailed structural information on the laminarinase subfamily is lacking [4].

The presence of at least one metal binding site is one of the common features in the crystal structures of family 16 glycoside hydrolases [2, 4] and calcium has been reported to protect *pfl*LamA from heat-induced inactivation [5]. Indeed, previous sequence comparisons of the deduced *pfl*LamA primary structure have revealed that those residues presumably involved in metal binding are conserved, in particular Asp-287 which is present in most of the family 16 enzymes, as well as residues Glu-170 and Glu-175 which are involved in catalysis (*pfl*LamA amino acid numbering

used throughout manuscript) [5]. However, despite the sequence similarities with all the enzymes of family 16, the unusual resistance to structural loss in 7.9 M GdmCl is a peculiar property of *pfl*LamA. To gain more information about the possible structural determinants of the unusual stability of *pfl*LamA, we performed multiple sequence alignments by applying Hidden Markov Model using the SAM-T02 web pages [6]. *pfl*LamA was found to be closely related to the mesophilic κ -carrageenase from *Pseudoalteromonas carrageenovora* (*pcCar*) (EC 3.2.1.83), a family 16 glycoside hydrolases enzyme which cleaves the internal β (1 \rightarrow 4) linkages of carrageenans, a linear polymer of galactopyranose residues linked by alternating α (1 \rightarrow 3) and β (1 \rightarrow 4) linkages, and whose crystal structure has been solved at 1.54 Å resolution [2, 7]. The homology to *pcCar* allowed us to use it as a template for molecular modeling of the *pfl*LamA structure. Interestingly, κ -carrageenase in crystals binds seven cadmium ions, indicating that this fold has several potential binding sites for metals [7]. Here we describe that in 7.9 M GdmCl *pfl*LamA interacts with calcium with a consistent regain of tertiary structure, as indicated by near-UV CD and fluorescence spectra, without any change in the far-UV CD spectrum. The interaction of calcium with the native form of *pfl*LamA becomes evident by Fourier-transformed infrared (FTIR) spectroscopy in the absorption region of carboxylate groups. We analyzed the binding of the metal to the native enzyme in the presence of a chromophoric chelator, and present a structural model of the protein that reveals potential binding sites for metals and that may provide clues about the structural interactions potentially responsible for the extreme stability of *pfl*LamA.

Materials and methods

Data collection - The three-dimensional crystal structures of a lichenase, the hybrid 1,3-1,4- β -D-glucanase from *Bacillus amyloliquefaciens* and *Bacillus macerans* (*bl*1,3-1,4Glc; PDB code: 2AYH), and of the catalytic domain of *pcCar* (PDB code: 1DYP) were taken from the Brookhaven Protein Data Bank (PDB) [8]. The two structures were superimposed with the program CE [9] and a sequence alignment was derived from the structural consensus. The alignment was manually corrected to optimize the arrangement of insertions and deletions. The sequence of the 1,3- β -glucanase from *pfl*LamA was then lined up with the two sequences of the structural alignment, by the program CLUSTALW [10]. The alignment obtained was manually modified to take into account the structural information such as the observed and predicted secondary structural elements, the structurally and functionally conserved amino acids, and the position of insertions and deletions in the structures. The program HOMOLOGY in the package InsightII [11] was used for the processing of alignments and structures.

A search for β -glucanase structures displaying homology with *pfl*LamA for a comparative analysis was conducted in the PDB databank using PSI-BLAST [12] with two iterations. All the hybrid and circularly permuted structures were rejected. The final selection consisted of two lichenases: the 1,3-1,4- β -glucanases from *Bacillus licheniformis* (*bl*1,3-1,4Glc; PDB code: 1GBG) and from *B. macerans* (*bm*1,3-1,4Glc; PDB code: 1MAC).

Model building - The three-dimensional structures of *b1,3-1,4Glc* and *pcCar* were used as templates for the construction of a homology-derived model of *pfLamA*. Homology modelling was based on the multiple sequence alignment constructed as described in the “Data collection” section. Protein models were built with the MODELLER-4 package [13]. Ten different models at the highest optimization level were built for the target protein. The model displaying the lowest ‘objective function’ value, which measures the violation of constraints from the template structures, was selected [14]. Construction of slightly different models of the same protein structure can be used as an indicator of the most variable and, therefore, less reliable regions in the folding. The calcium ion was taken from the *b1,3-1,4Glc* structure and included in the model as a rigid body. The quality of the final model was assessed with the programs ProsaII [15] and PROCHECK [16].

Contact surface area for apolar atoms was calculated with the program PDB_NP_CONT [17] that computes pair wise atom contact areas between apolar atoms from structural data. The program is based on classification of points located on an interaction sphere around each atom and use a method similar to that described by Connolly [18]. The total contact surface was then normalized by the number of residues of the protein considered.

Ion pairs - Ion pair interactions were determined using the program WHAT IF [19]. Two atoms of opposite charge separated by a distance less than a defined threshold are defined as an ion pair. The distance limits of 4.0 Å, the value usually accepted [20], and 6.0 Å were selected to identify, respectively, “strong” and “weak” ion pairs. Atoms with positive charge were considered the side-chain nitrogens in Arg and Lys. Atoms with negative charge were the side-chain oxygens of Asp and Glu. Results with and without His residues are reported for completeness, however the assignment of protonation state of such residue in a protein may be difficult.

Chemicals and buffers - GdmCl, 8-anilidonaphthalene-1-sulfonic acid ammonium salt (ANS), 1,4-dithio-DL-threitol (DTT), EDTA and laminarin were from Fluka. 3',5'-dinitrosalicylic acid was purchased from Sigma. 5,5'-Br₂-BAPTA was from Molecular Probes Europe BV (The Netherlands). Buffer solutions were filtered (0.22 µm) and carefully degassed. All buffers and solutions were prepared with ultra-high quality water (ELGA UHQ, U.K.). Buffers for calcium titrations were prepared as described in [21].

Enzyme preparation and assay - *pfLamA* was functionally produced in *Escherichia coli* BL21(DE3) with pLUW532 and purified according to Kaper *et al.* [22]. The protein concentration was determined at 280 nm using a $\epsilon_{280} = 83070 \text{ M}^{-1} \times \text{cm}^{-1}$ calculated according to Gill and von Hippel [23]. Enzyme activity was determined by measuring the amount of reducing sugars released upon incubation in 0.1 M sodium phosphate buffer, pH 6.5, containing 5 mg/ml of laminarin, at 60 or 80 °C for 10 min, as described by Kaper *et al.* [22]. Calcium-depleted protein was obtained by extensive dialysis with 100 µM EDTA and 100 µM EGTA in 10 mM Tris/HCl, pH 7.4. All the precautions required to prevent Ca²⁺ contamination were followed during the preparation and storage of protein and buffer solutions [21]. Calcium-loaded protein refers to the protein in the presence of 40 mM CaCl₂.

Spectroscopic techniques - Intrinsic fluorescence emission and 90° light scattering measurements were carried out with a LS50B Perkin Elmer spectrofluorimeter using a 1.0-cm pathlength quartz cuvette. Fluorescence emission spectra were recorded at 300 - 450 nm (1 nm sampling interval) at 20 °C with the excitation wavelength set at 290 nm. 90° light scattering was measured at 20 °C with both excitation and emission wavelength set at 480 nm to check for the presence of aggregated particles.

Far-UV (180 - 250 nm) and near-UV (250 - 320 nm) CD measurements were performed at 20 °C in a 0.1-0.2 cm and 1.0-cm pathlength quartz cuvette, respectively. CD spectra were recorded on a Jasco J-720 spectropolarimeter. The results are expressed as the mean residue ellipticity ($[\Theta]$) assuming a mean residue weight of 110 per amino acid residue. In all the spectroscopic measurements at pH 7.4, 100-250 μ M EDTA was always present unless otherwise stated.

FTIR spectra were recorded on a Nicolet Magna 760 spectrometer (Nicolet) equipped with a liquid nitrogen-cooled mercury-cadmium-telluride solid-state detector. Attenuated total reflectance (ATR) spectra in solution were measured in a CIRCLE cell (Spectra Tech, Madison, WI) at 20 °C. Protein samples (130 μ l) of a 8 mg/ml protein solution in 20 mM Tris/HCl at pH 7.4, in the absence or in the presence of 40 mM CaCl_2 , were placed in the CIRCLE cell with a ZnSe crystal rod. For exchange of amide protons with deuterons, *pfl*LamA was lyophilized and dissolved twice in D_2O buffer at 20 °C before measurements. A total of 512 interferograms at 2 cm^{-1} resolution were collected for each spectrum, with Mertz apodization and two levels of zero filling. The sample chamber of the spectrometer was continuously purged with dry air to avoid water vapor interference on the bands of interest. The background spectra were collected immediately before the sample measurements and under the same conditions with the cell filled with everything but protein. At the end of the measurements, after prolonged washing, a spectrum of the cell was recorded to check for protein adsorption on the crystal rod [24]. Water vapor spectra were collected by reduction of the dry-air purge of the clean cell. Difference spectra were normalized to an equal protein content by normalizing spectra measured in H_2O to an amide II absorbance of 0.04 (difference in absorbance between 1545 and 1490 cm^{-1}) and spectra measured in D_2O to an amide I' absorbance of 0.06 (difference in absorbance between 1706 and 1636 cm^{-1}). The analysis of raw spectra was performed with GRAMS.

Experiments with the fluorescent dye anilinonaphtalene-8-sulfonic acid (ANS) were performed at 20 °C by incubating the protein and ANS at 1:10 molar ratio. After 5 min, fluorescence emission spectra were recorded at 400 - 600 nm with the excitation wavelength set at 390 nm. The maximum fluorescence emission wavelength and the intensity of the hydrophobic probe ANS depend on the environmental polarity, e.g. on the hydrophobicity of the accessible surface of the protein [25]. Fluorescence quenching was carried out by adding increasing amount of acrylamide (0-100 mM) to *pfl*LamA solution (40 μ g/ml) in 7.9 M GdmCl in the absence or in the presence of 40 mM CaCl_2 . Emission spectra (300-450) were recorded at 20 °C, 10 min after each acrylamide addition with the excitation wavelength set at 290 nm. The effective quenching constants were obtained from modified Stern-Vollmer plots by analyzing $F_0/\Delta F$ vs $1/[\text{acrylamide}]$ (25 data points) [26].

GdmCl induced unfolding and refolding - For equilibrium transition studies, *pf*LamA (final concentration 40-50 µg/ml) was incubated at 20 °C at increasing concentrations of GdmCl (0 - 8 M) in 20 mM Tris/HCl, pH 7.4, containing 100 µM DTT and 100 µM EDTA and, when indicated, 40 mM CaCl₂. After 24 h equilibrium was reached and intrinsic fluorescence emission and far-UV CD spectra (0.2-cm cuvette) were recorded in parallel at 20 °C. To test the reversibility of the unfolding, *pf*LamA was unfolded at 20 °C in 7.8 M GdmCl at 0.8 mg/ml protein concentration in 25 mM Tris/HCl, pH 7.4, containing 100 µM DTT and 100 µM EDTA, in the presence and absence of 40 mM CaCl₂. After 24 h, refolding was started by 20-fold dilution of the unfolding mixture, at 20 °C, into solutions of the same buffer used for unfolding containing decreasing GdmCl concentrations. The final enzyme concentration was 40 µg/ml. After 24 h, a time that was established to be sufficient to reach equilibrium, intrinsic fluorescence emission and far-UV CD spectra were recorded at 20°C.

Data analysis - Far-UV CD and near-UV CD spectra from GdmCl and Ca²⁺ titrations were analyzed by the singular value decomposition algorithm (SVD) [1, 27] using the software MATLAB (MathWorks, South Natick, MA). SVD is useful to find the number of independent components in a set of spectra and to remove the high-frequency noise and the low-frequency random error. CD spectra in the 210-250 nm region or in the 250-310 region (0.2 nm sampling interval) were placed in a rectangular matrix **A** of *n* columns, one column for each spectrum collected in the titration. The **A** matrix is decomposed by SVD into the product of three matrices: **A** = **U*****S*****V**^T where **U** and **V** are orthogonal matrices and **S** is a diagonal matrix. The columns of **U** matrix contain the basis spectra and the columns of the **V** matrix contain the denaturant or the Ca²⁺ dependence of each basis spectrum. Both **U** and **V** columns are arranged in terms of their decreasing order of the relative weight of information, as indicated by the magnitude of the singular values in **S**. The diagonal **S** matrix contains the singular values that quantify the relative importance of each vector in **U** and **V**. The signal-to-noise ratio is very high in the earliest columns of **U** and **V** and the random noise is mainly accumulated in the latest **U** and **V** columns. The wavelength averaged spectral changes induced by increasing denaturant or Ca²⁺ concentrations are represented by the columns of matrix **V**, hence the plot of the columns of **V** versus the denaturant or Ca²⁺ concentrations provides information about the observed transition.

GdmCl-induced equilibrium unfolding was analyzed by fitting baseline and transition region data to a two-state linear extrapolation model [28] according to

$$\Delta G_{\text{unfolding}} = \Delta G^{\text{H}_2\text{O}} + m_g [\text{GdmCl}] = - RT \ln K_{\text{unfolding}} \quad (\text{Eq. 1})$$

where $\Delta G_{\text{unfolding}}$ is the free energy change for unfolding for a given denaturant concentration, $\Delta G^{\text{H}_2\text{O}}$ is the free energy change for unfolding in the absence of denaturant and m_g is a slope term which quantitates the change in $\Delta G_{\text{unfolding}}$ per unit concentration of denaturant, R is the gas constant, T is the temperature and $K_{\text{unfolding}}$ is the equilibrium constant for unfolding. The model expresses the signal as a function of denaturant concentration:

$$y_i = \frac{y_N + m_N[X]_i + (y_D + m_D[X]_i) \cdot \exp[-(\Delta G^{\text{H}_2\text{O}} - m_g[X]_i)/RT]}{1 + \exp[-(\Delta G^{\text{H}_2\text{O}} - m_g[X]_i)/RT]} \quad (\text{Eq. 2})$$

where y_i is the observed signal, y_N and y_D are the native and denatured baseline intercepts, m_N and m_D are the native and denatured baseline slopes, $[X]_i$ is the denaturant concentration after the i th addition, $\Delta G^{\text{H}_2\text{O}}$ is the extrapolated free energy of unfolding in the absence of denaturant, m_g is the slope of a G unfolding versus $[X]$ plot, R is the gas constant and T is the temperature. The $[\text{GdmCl}]_{0.5}$ is the denaturant concentration at the midpoint of the transition and, according to Equation 1, is calculated as:

$$[\text{GdmCl}]_{0.5} = \Delta G^{\text{H}_2\text{O}} / m_g \quad (\text{Eq. 3})$$

Calcium titrations and determination of binding constant - Calcium-depleted native *p*fLamA (9-16 μM) was titrated with CaCl_2 in the presence of 24 μM of the chromophoric chelator 5,5'-Br₂-BAPTA [29]. 5,5'-Br₂-BAPTA concentration was determined by measuring the absorbance at 263 nm using $\epsilon_{263,5} = 1.6 \times 10^4 \text{ M}^{-1} \times \text{cm}^{-1}$ [21]. Titrations were performed at 20°C in 10 mM Tris/HCl pH 7.5 by addition of 1-2 μl of CaCl_2 solutions ranging from 0.015 to 20.0 mM to a 1 ml protein solution containing 24 μM 5,5'-Br₂-BAPTA. Absorbance spectra were monitored between 200 and 450 nm after each Ca^{2+} addition. To determine protein binding constants and number of binding sites, the variation of 5,5'-Br₂-BAPTA absorbance as a function of calcium addition was fitted by non-linear analysis using the CaLigator software [21]. The quantity χ^2 calculated by the program was used as the measure of the goodness-of-fit. All the precautions required to prevent Ca^{2+} contamination were followed during the preparation and storage of protein and buffer solutions [21].

Calcium titration of calcium-depleted *p*fLamA in 7.9 M GdmCl (25 mM Tris/HCl, pH 7.4 containing 200 μM DTT and 250 μM EDTA) was performed by addition of increasing CaCl_2 concentrations (0-40 mM) under continuous stirring. 5 min after each CaCl_2 addition, near-UV CD (240-320 nm, 22 μM protein concentration) and fluorescence (300-450 nm, 1.2 μM protein concentration) spectra were recorded at 20°C. The spectral changes observed after each CaCl_2 addition were not affected by longer incubation time. The concentration of unchelated Ca^{2+} was calculated by using the program WinMaxc Version 2.40 [30] (see URL <http://www.stanford.edu/~cpatton/maxc.html>).

Results

Effect of calcium on *pf*LamA in 7.9 M GdmCl

Addition of CaCl_2 to calcium-depleted *pf*LamA in 7.9 M GdmCl, 25 mM Tris/HCl, pH 7.4, containing 200 μM DTT, 250 μM EDTA, at 20 °C, causes significant changes in the enzyme tertiary structure, as indicated by near-UV CD and intrinsic fluorescence emission spectra (Fig. 4.1A, B). At 40 mM CaCl_2 the 295 nm band of Trp, which is significantly reduced in 7.9 M GdmCl, is restored with a concomitant regain of the negative ellipticity signal and of the fine structure in the 260-270 nm region (Fig. 4.1A). The regain in aromatic chirality is accompanied by a 1.5-fold decrease in fluorescence intensity at 342 nm and a blue shift from 357 to 343 nm in the maximum fluorescence emission wavelength, a value close to that measured at 342 nm for the native state (Fig. 4.1B). The far-UV CD spectrum in 7.9 M GdmCl, which is the same as that measured in the absence of denaturant [1], is not affected by the addition of CaCl_2 (data not shown). A titration of the enzyme in 7.9 M GdmCl with increasing amount of CaCl_2 , from 0.2 nM to 35 mM unchelated Ca^{2+} , and the analysis of the aromatic CD spectral changes at 295 nm indicates that, above 120 μM of unchelated Ca^{2+} concentration, no further changes are observed (Fig. 4.2). The near-UV CD ellipticity changes at 295 nm induced by increasing CaCl_2 concentration (Fig. 4.2) were analyzed after removal of the high-frequency noise and the low-frequency random error by SVD. The global changes in the spectral region 250-310 nm were analyzed by SVD that indicates that only two spectral components contribute to the near-UV CD spectra. The most signif-

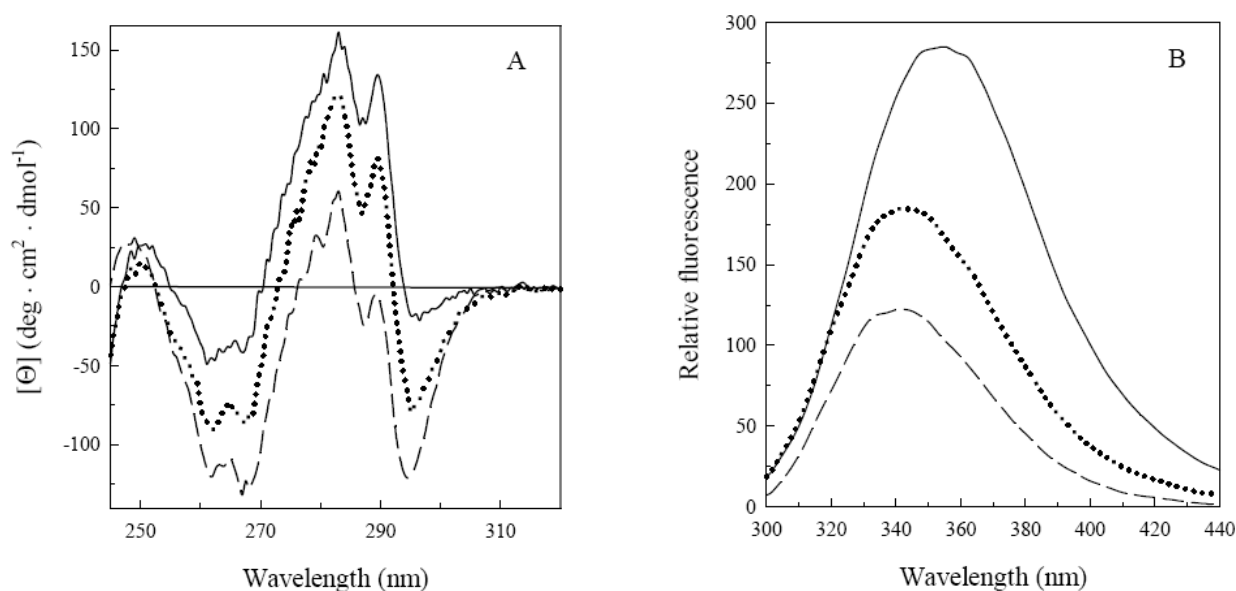


Figure 4.1 Effect of calcium on the spectral properties of *pf*LamA in 7.9 M GdmCl. **(A)** Near-UV CD spectra were recorded in a 1 cm quartz cuvette at 0.6 mg/ml protein concentration and **(B)** fluorescence spectra were recorded at 40 $\mu\text{g}/\text{ml}$ protein concentration (290 nm excitation wavelength). All the spectra were recorded at 20 °C after 24 h incubation of the protein in 20 mM Tris/HCl (pH 7.4) (---) and in 7.9 M GdmCl at pH 7.4 in the absence (—) and presence of 40 mM CaCl_2 (·····).

icant singular values are 4.7×10^3 , 1.0×10^3 , and 0.1×10^3 . All the other singular values are well below 10% of the largest singular value and progressively decrease approaching to zero. The first and the second columns of the V matrix (V_1 and V_2) show a similar dependence upon increasing Ca^{2+} concentration and both confirm that saturation occurs above 120 μM (data not shown). The $[\Theta]_{295}$ data were analyzed by nonlinear regression analysis to define two limiting slopes, intersecting at a value which suggests that essentially 2 moles of Ca^{2+} per mole of enzyme are necessary to reach an apparent saturation effect (Fig. 4.2). The intrinsic fluorescence emission quenching at 342 nm measured upon unchelated Ca^{2+} addition from 0.2 nM to 35.0 mM remains unchanged above 120 μM Ca^{2+} , similarly to what observed for the near-UV CD ellipticity changes (Fig. 4.2). The dependence of the spectral changes upon increasing Ca^{2+} concentration is hyperbolic, suggesting that binding of two Ca^{2+} essentially produce the same signal changes either on the decrease of the intrinsic fluorescence emission at 342 nm or on the regain of negative ellipticity at 295 nm. A comparable result is obtained from the plot of the changes of the maximum fluorescence emission wavelength (data not shown).

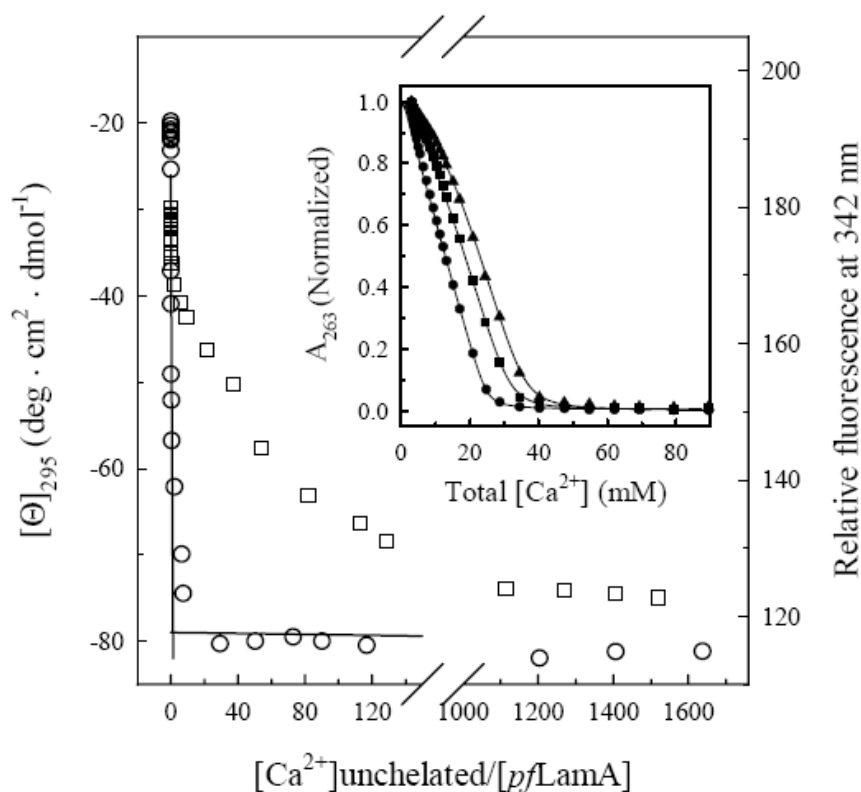


Figure 4.2 Interaction of calcium with *pfLamA*. $[\Theta]_{295}$ (○, 22 μM *pfLamA* in 7.9 M GdmCl) and fluorescence intensity at 342 nm (□, 1.2 μM *pfLamA* in 7.9 M GdmCl) were measured from near-UV CD and fluorescence spectra (290 nm excitation wavelength) recorded at 20 °C, 5 min after each Ca^{2+} addition. $[\Theta]_{295}$ is reported after removal of the high-frequency noise and the low-frequency random error by SVD. The two limiting slopes, calculated by non-linear regression analysis to the $[\Theta]_{295}$ data at 22 μM *pfLamA*, intersect at a point corresponding to $[\text{Ca}^{2+} \text{ unchelated}] / [\text{pfLamA}] = 2$. The reported unchelated Ca^{2+} concentrations, calculated according to [30], are 0.2 nM–35.0 mM and 0.2 nM–1.82 mM for $[\Theta]_{295}$ and fluorescence changes, respectively. The inset shows a plot of the decrease in 5,5'-Br₂-BAPTA absorbance at 263 nm towards $[\text{Ca}^{2+}]$ in the absence (●) and presence of 10 μM (■) and 20 μM (▲) native *pfLamA* in 10 mM Tris/HCl (pH 7.5), at 20 °C. The solid lines represent fitting of the data by non linear analysis using the Caligator software [21].

Effect of calcium on the native *pf*LamA

The near and far-UV CD spectra, as well as the intrinsic fluorescence emission spectrum of the calcium-depleted native *pf*LamA, in 25 mM Tris/HCl, pH 7.4, containing 200 μ M DTT and 250 μ M EDTA, are not affected by addition of CaCl_2 up to 0.5 M (data not shown). The interaction of calcium with the native enzyme was therefore studied by titration of *pf*LamA with CaCl_2 in the presence of the chromophoric chelator 5,5'-Br₂-BAPTA. The effect of calcium chelation on the absorbance spectra of 5,5'-Br₂-BAPTA is characterized by a decrease of the absorbance at 263 nm accompanied by an increase in absorbance around 239 nm, with an isosbetic point at 247 nm [29]. A plot of the decrease in 5,5'-Br₂-BAPTA absorbance at 263 nm towards $[\text{Ca}^{2+}]$ gives a titration profile, progressively shifted to the right in the presence of increasing *pf*LamA concentration (Fig. 4.2, inset). This indicates the affinity of the native protein for the cation that was quantitatively analyzed by CaLigator software [21]. At all protein concentrations tested, best fits of the absorbance data ($\chi^2 = 9.8 \times 10^{-5}$) were obtained with two calcium binding sites and two corresponding binding constants of $5.0 \times 10^7 \text{ M}^{-1}$ and $2.6 \times 10^5 \text{ M}^{-1}$, respectively, were found. Fitting the data according to only one calcium binding site gave a single binding constant of $3.3 \times 10^7 \text{ M}^{-1}$ and a 3-fold higher χ^2 value.

The effect of calcium on native *pf*LamA was also investigated by FTIR spectroscopy, in H_2O and in D_2O , in the amide I, amide II and in the spectral region where the side chains of Asp and Glu are known to exert their contributions [31, 32] (Fig. 4.3). In the presence of 40 mM Ca^{2+} , FTIR absorbance spectra in H_2O show an amide I region almost identical to that observed in the absence of the cation, with the same maximum at 1634 cm^{-1} , and a shift of the amide II maximum from 1542 to 1544 cm^{-1} which is accompanied by an increase in the amide II intensity (Fig. 4.3A). In D_2O , FTIR absorbance spectra of *pf*LamA in the presence and absence of calcium show a maximum centred at 1632 cm^{-1} and a band at 1560 cm^{-1} (Fig. 4.3B). Similarly to what observed in H_2O , the presence of the cation does not significantly affect the amide I' region, in agreement with the lack of any measurable change induced by the cation on the far UV CD spectrum. In D_2O , a decrease of the intensity in the amide II region and a shift of the amide II to 1455 cm^{-1} are also observed (Fig. 4.3B), similarly to what reported for other proteins upon deuteration [33, 34]. However, in the presence of Ca^{2+} the decrease in intensity of the amide II' region at 1544 cm^{-1} observed in D_2O is less pronounced (Fig. 4.3B, C), thus suggesting a reduced hydrogen-deuterium exchange indicative of a more compact structure [35].

The region of antisymmetric stretching vibration (ν_{as}) COO^- mode of the unprotonated carboxyl groups ($1595\text{-}1550 \text{ cm}^{-1}$) and the region $1430\text{-}1390 \text{ cm}^{-1}$ corresponding to the symmetric stretching vibration (ν_{s}) COO^- mode of the unprotonated carboxyl groups were explored [32, 36]. FTIR difference absorption spectra were obtained by subtracting the absorption spectrum of the calcium-depleted from that of the calcium-loaded form of *pf*LamA. Hence, positive peaks in the difference spectrum correspond to increase in the intensities of the absorption spectrum upon calcium binding, and vice versa (Fig. 4.3C). In D_2O , the binding of calcium gives rise to a main positive band centred at 1547 cm^{-1} and a minor peak at 1594 cm^{-1} (Fig. 4.3C). In H_2O , the difference spectrum is comparable to that obtained in D_2O , with a main positive peak centred at 1550 cm^{-1} , accompanied

by a shoulder at about 1575 cm^{-1} (Fig. 4.3C). The regions of the $\nu_s\text{ COO}^-$ mode in H_2O and in D_2O are also very similar, with a main positive peak at 1408 and at 1410 cm^{-1} , respectively. Besides the ν_{as} and $\nu_s\text{ COO}^-$ (1595 - 1550 and 1430 - 1390 cm^{-1}), in H_2O two positive peaks are evident at 1466 and at 1385 cm^{-1} similarly to the large shoulder at about 1468 cm^{-1} and to the positive peak at 1380 cm^{-1} observed in D_2O (Fig. 4.3C). Above 1610 cm^{-1} , in the amide I and amide I' regions, both the difference spectra are similar, without any significant change in the presence of calcium, thus confirming the lack of changes in the secondary structure elements induced by the cation (data not shown). The 1800 - 1700 cm^{-1} region, corresponding to the C=O mode of protonated carboxyl groups, is not affected by the presence of calcium (data not shown).

Resolution enhancement by second derivative analysis of the FTIR spectra confirms the changes observed in the difference spectra (Fig. 4.3D, E). Below the Tyr ring mode, which in D_2O is downshifted from 1516 to 1514 cm^{-1} , the effect of calcium is evident, particularly in H_2O , as an upshift from 1396 to 1403 cm^{-1} which is comparable to the upshift from 1400 to 1403 cm^{-1} observed in D_2O . In the ν_{as} (1595 - 1550 cm^{-1}), the frequencies of the bands are not shifted by the presence of calcium, neither in H_2O nor in D_2O , and an increased resolution of the band in the 1575 - 1550 cm^{-1} interval is observed in H_2O . In the presence of calcium, changes are observed also in frequencies regions apparently not directly related to the ν_{as} and $\nu_s\text{ COO}^-$, such as a downshift from 1475 to 1468 cm^{-1} and an upshift from 1377 to 1386 cm^{-1} in H_2O , and a downshift from 1385 to 1382 cm^{-1} in D_2O .

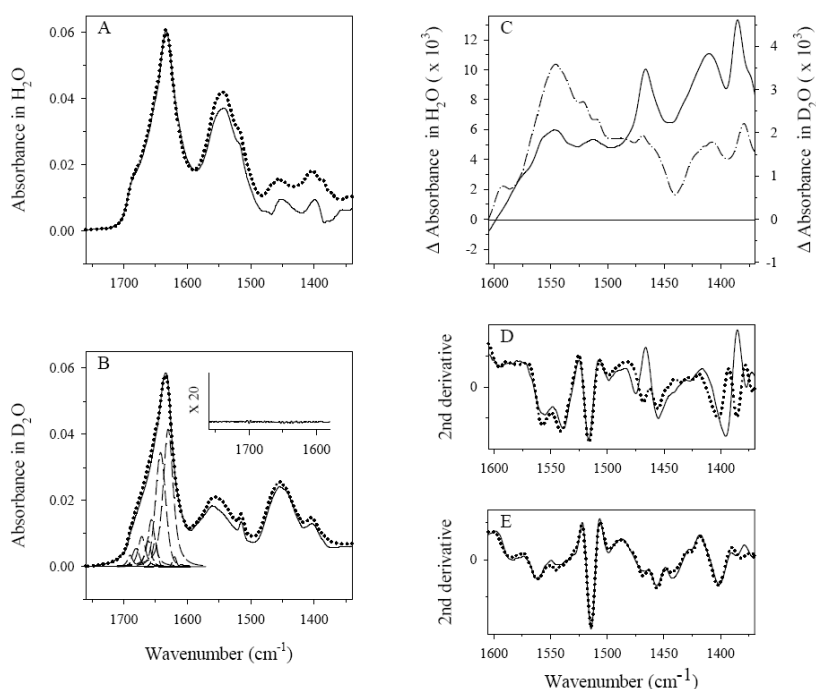


Figure 4.3 Effect of calcium on the IR spectra of *p/LamA*. Solution ATR- FTIR spectra of 8 mg/ml *p/LamA* were measured at 20°C in 20 mM Tris/HCl (pH 7.4) in a CIRCLE cell with a ZnSe crystal rod. A total of 512 interferograms at 2 cm^{-1} resolution were collected for each spectrum. Absorbance spectra of amide I and II in H_2O (A) and in D_2O (B) in the absence (—) and in the presence of 40 mM CaCl_2 (·····). Deconvolution of the amide I' spectrum was used to identify the individual components in the amide I' region by curve-fitting of the raw spectrum (B). Inset in (B) shows the difference between the fitted curve and the original expanded 20 times. (C) Difference spectra in H_2O (—, left axis) and in D_2O (·····, right axis) were obtained by subtracting the spectrum recorded in the absence of calcium from that in the presence of 40 mM CaCl_2 . (D, E) Second derivative of the raw spectra in the amide II region measured in H_2O (D) and in D_2O (E) in the absence (—) and in the presence of 40 mM CaCl_2 (·····).

Effect of calcium on the equilibrium transition in GdmCl

Incubation of *p/LamA* at increasing GdmCl concentrations (0-8 M) in 20 mM Tris/HCl, pH 7.4, containing 100 μ M DTT, 100 μ M EDTA, and 40 mM CaCl_2 for 20 h at 20 °C results in a progressive increase in the intrinsic fluorescence emission intensity at 342 nm (Fig. 4.4), with a fluorescence maximum emission wavelength still centered around 342 nm (Fig. 4.4, inset). The process follows a two-state mechanism without any detectable intermediates, as indicated by the sigmoidal transition profiles shown in Fig. 4.4, similarly to what observed in the absence of Ca^{2+} . A plot of the relative fluorescence intensity at 342 nm as a function of GdmCl concentration shows a transition midpoint at 6.0 M GdmCl in the presence of Ca^{2+} and at 6.7 M GdmCl in the absence of Ca^{2+} . The enzyme, which is inactive in 7.9 M GdmCl in the presence and absence of CaCl_2 , was fully reactivated upon dilution of the denaturant. The fluorescence changes were reversible and the $\Delta G^{\text{H}_2\text{O}}$ and m_g values were calculated by nonlinear regression fitting of the data reported in Fig. 4.4, according to Equation 2 and Equation 3. The $\Delta G^{\text{H}_2\text{O}}$ and m_g values were 41.8 kJ/mol and 6.99 kJ/mol/M and 61.5 kJ/mol and 9.20 kJ/mol/M in the presence and absence of Ca^{2+} , respectively, suggesting an apparent decrease in protein stability in the presence of the cation. In the presence of Ca^{2+} , the analysis of *p/LamA* far UV CD spectra in the 210-250 nm region revealed no changes upon increasing GdmCl concentration (data not shown), similarly to what has been reported for the enzyme in the absence of Ca^{2+} [1]. This result was confirmed by the reconstitution of the spectra after SVD data analysis and by the random variation in magnitude and sign of the two most significant columns of the **V** matrix as a function of denaturant concentration.

The fluorescence emission spectrum measured upon incubation in 7.9 M GdmCl and 40 mM CaCl_2 is comparable to that resulting from the progressive addition of CaCl_2 to the protein in 7.9 M GdmCl (Fig. 4.4, inset).

ANS fluorescence and Acrylamide quenching

The accessibility of hydrophobic residues upon incubation of *p/LamA* in 7.9 M GdmCl in the presence and absence of Ca^{2+} was compared by the analysis with the fluorescent probe ANS. The fluorescence emission spectrum of ANS shows a modest, two-fold increase in intensity in the presence of the protein in 7.9 M GdmCl, either in the presence and absence of 40 mM CaCl_2 , without any change in the maximum fluorescence emission wavelength (data not shown). This suggests that in 7.9 M GdmCl the hydrophobic surface area of the protein is not significantly exposed.

The uncharged fluorescence quencher acrylamide was used to probe the accessibility of the hydrophobic core and the dynamic properties of *p/LamA* in 7.9 M GdmCl in the presence and absence of Ca^{2+} in comparison with that of the native enzyme. Effective acrylamide quenching constants from the modified Stern-Vollmer plots for the protein in the absence and presence of Ca^{2+} were 13.0 and 11.4 M^{-1} in 7.9 M GdmCl and 7.9 and 6.9 M^{-1} for the native enzyme, respectively. These results suggest that Ca^{2+} causes a decreased accessibility of protein fluorophores to the quencher both in the native and in the 7.9 M GdmCl state.

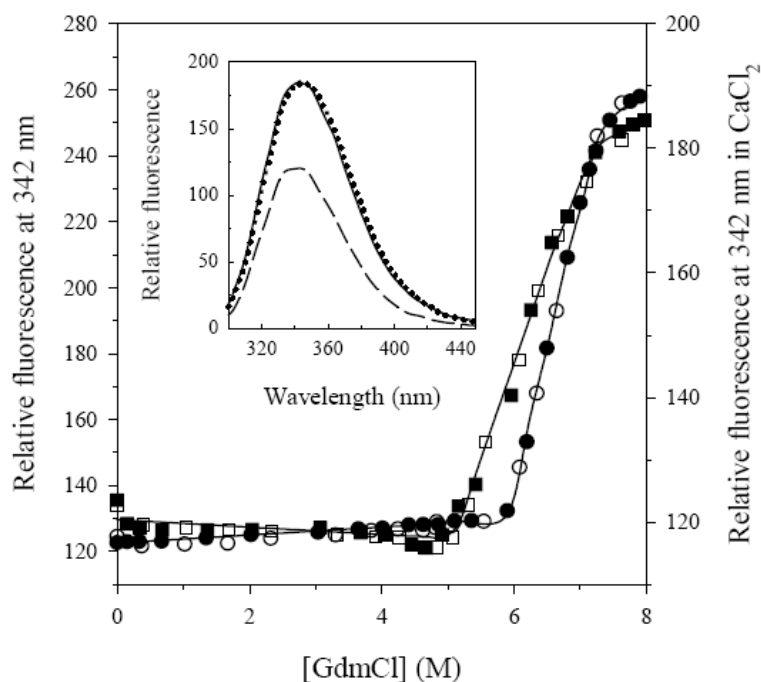


Figure 4.4 Effect of calcium on the GdmCl-induced fluorescence changes of *pfLamA*. Relative fluorescence emission intensity at 342 nm (290 nm excitation wavelength) were recorded at 20 °C after 24 h incubation at varying denaturant concentration, in the absence (●, left axis) and presence of 40 mM CaCl_2 (■, right axis), as described in the text (see Materials and Methods). Continuous lines are the non-linear regressions to Equation 2 of the fluorescence data. The reversibility points (empty symbols) were not included in the non-linear regression analysis. **(Inset)** the intrinsic fluorescence emission spectra of *pfLamA* in 7.9 M GdmCl and 40 mM CaCl_2 : —, the spectrum measured after 24 h incubation in 7.9 M GdmCl in the presence of 40 mM CaCl_2 ; ·····, the spectrum resulting from the progressive addition of CaCl_2 to the protein after 24 h incubation in 7.9 M GdmCl; ---, the spectrum of the native *pfLamA* in 20 mM Tris/HCl (pH 7.4). All the spectra were recorded at 20 °C.

Structural analysis

Estimation of secondary structure

A structural model for *pfLamA* has been generated on the basis of an alignment of its amino acid sequence with that of two family 16 enzymes that have established crystal structures and that share the highest sequence similarity with *pfLamA*: *b1,3-1,4Glc* and *pcCar* (Fig. 4.5). The resulting *pfLamA* model has been analyzed in an attempt to unravel details of its extreme stability, and of its calcium binding site(s). The relative contribution of secondary structure elements on *pfLamA* model was calculated by the program DSSP [37] and corresponds to a 42.4 % of β -sheets, 2.6 % of α -helices, 12.4 % turns (including residues in isolated β -bridges) and 42.6 % loops. This result was compared with the deconvoluted FTIR amide I' region in D_2O (Fig. 4.3B) which reveals nine individual peaks centered at 1620, 1630, 1642, 1651, 1658, 1662, 1672, 1682 and 1690 cm^{-1} . The main peak at 1630 cm^{-1} and the minor peak at 1620 cm^{-1} can be assigned to β -sheets and correspond to 40.0 % and 1.0 % of the total amide I' area, respectively, which is in good agreement with the model and with the secondary structure prediction given by the server SSpro [38]. The peak at 1642 cm^{-1} , 29.8 %, can be attributed to unordered structures [39] as well as to flexible loops [40]. The assignment to flexible loops seems to be more satisfactory since *pfLamA* folds like the

concanavalinA-like lectins [41] and a comparable band has been reported upon deconvolution of the amide I' of lectins [42]. Furthermore, the comparative analysis of the ten generated *pfl*LamA models reveals that the most variable regions, potentially reflecting the most flexible elements of the protein structure, correspond to loop regions for 29.7 % of the total sequence. The band at 1651 cm^{-1} , 3.0 %, may be assigned to α -helices. The peaks at 1658 cm^{-1} is 10.2 % of the amide I' area and can be assigned to loops with dihedral angles close to α -helix [42]. Noteworthy, the Ramachandran plot of *pfl*LamA model calculated by PROCHECK [16] shows 12.4 % of residues within the region corresponding to right-handed α -helix; hence, the exclusion of residues really involved in α -helix structures (2.6 %) results in a remaining 9.8 % of residues which may correspond to loop regions with α -helix dihedral angles. The peaks at 1662, 1672 and 1682 cm^{-1} are assigned to β -turns [31, 34, 42] and correspond to 4.5, 6.1 and 3.1% of the total amide I' area. The peak centered at 1690 cm^{-1} can be referred to antiparallel β -sheets [42] and correspond to 2.3 % of total amide I' (Fig. 4.3B). The relative content of secondary structure elements in amide I' is in agreement with that determined by FTIR in H_2O and far UV CD spectroscopy [1] and corresponds to that determined from the homology model of *pfl*LamA.

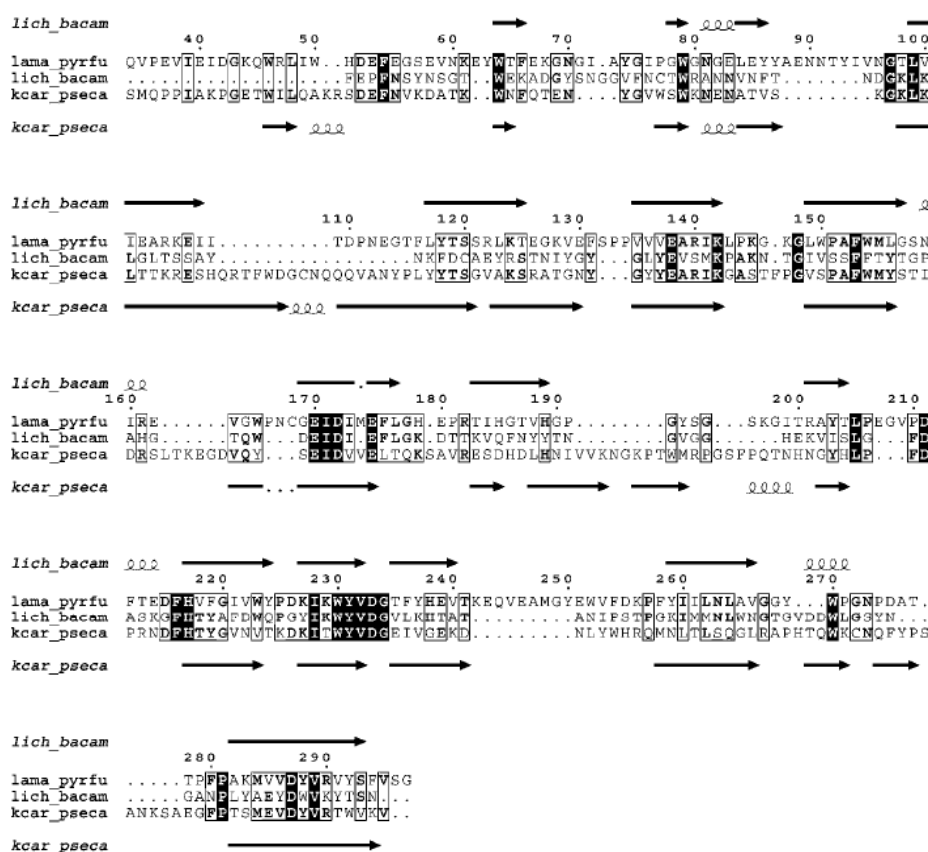


Figure 4.5 Multiple sequence alignment of *pfl*LamA (lama_pyrfu) with structural templates of *b1,3-1,4Glc* (lich_bacam) and *pcCar* (kcar_pseca). Single-letter code has been used for the amino acids. Dots represent deletions. Numbers above the sequences represent sequence numbering of *pfl*LamA. Invariant positions are boxed in black; aligned columns displaying an amino acid identity in two of the three sequences are boxed, and the most conserved residue is shown in boldface. Secondary structures of lich_bacam and kcar_pseca are reported in the first and the last line of each block respectively; α -helices and β -strands are shown as squiggles and arrows respectively. The figure was generated by using the program ESPript 2.2 [43].

Ion pairs

The number of ion pairs responsible for strong interactions (distance threshold of 4 Å) in the *pfLamA* model is 11, slightly higher than that observed in the lichenases *bl*1,3-1,4Glc and *bm*1,3-1,4Glc, where the ion pairs are 8 and 9, respectively, but lower than that observed in *pcCar* which displays 15 ion pairs. The inclusion of weak ion pairs (distance threshold of 6 Å) leads to a significantly higher number of potential ion pairs in the hyperthermophilic *pfLamA*, some of which are involved in the formation of salt bridge networks. With the 6 Å threshold, the number of ion pairs in *pfLamA* increases to 45, whereas that of *bl*1,3-1,4Glc and *bm*1,3-1,4Glc reaches 21 and 30, respectively. Instead, the number of ion pairs in *pcCar* increases more than those in the two lichenases and arrives at 44. The inclusion of His in the calculation of ion pair interactions increases the difference between the *pfLamA* and the two lichenases. The strong ion pairs increase from 8 to 9 and from 9 to 10 in *bl*1,3-1,4Glc and *bm*1,3-1,4Glc, respectively, whereas they rise from 11 to 18 in *pfLamA*, and from 15 to 22 in *pcCar*. If weak ion pairs are also considered, the number of total interactions reaches 63 in *pfLamA* and 33, 39 and 72 in *bl*1,3-1,4Glc, *bm*1,3-1,4Glc and *pcCar*, respectively.

Calcium ions

Calcium ion in the model is bound in the convex face of the molecule, to the backbone carbonyl oxygens of Glu-53, Gly-97 and Asp-287 and to side-chain carboxylate oxygens of Glu-53 and Asp-287 (*pfLamA* amino acid numbering) (Fig. 4.6). Residues Glu-53, Gly-97 and Asp-287 correspond, respectively, to Pro-9, Gly-45 and Asp-207 in *bl*1,3-1,4Glc. Gly-97 (45) and Asp-287 (207) are conserved among most of the family 16 β -glucanases whereas Glu-53 is apparently conserved only in β -glucanases from thermophilic organisms.

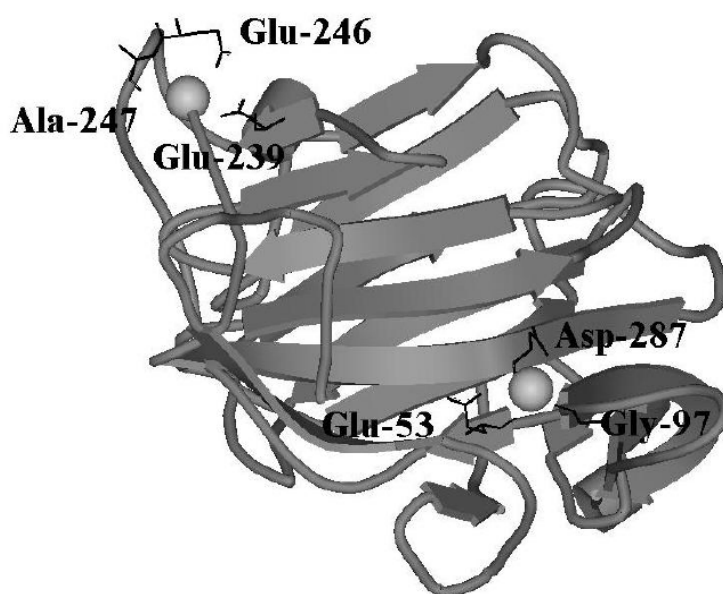


Figure 4.6 Schematic representation of *pfLamA* model. Calcium ions are represented as grey spheres. Amino acid residues putatively interacting with calcium are labelled and represented as sticks. The figure was prepared with DS ViewerPro 5.0 (Accelrys, San Diego, CA, U.S.A.).

The cadmium ions of *pcCar* were used to investigate further potential binding sites for the calcium ions in *pfLamA*. The cadmium ion bound in the active site most likely reflects a crystallization artefact [7]. Another cadmium ion is bound on the convex face of the enzyme in a position corresponding to the binding site of the calcium ion, above mentioned, in *pfLamA*. The remaining five cadmium ions are bound to carboxylate oxygens of surface-exposed acidic amino acid residues. The residues of *pfLamA* at the corresponding positions of the alignment that was used to build the model do not show acidic residues except for Glu-239 (corresponding to Glu-241 in *pcCar*). This Glu residue does not seem to be conserved in the β -glucanases of the family 16, not even in those from other thermophiles. However, residues Glu-246 and Ala-247 in *pfLamA* are in a suitable position to interact through, respectively, one carboxylate and the carbonyl oxygens with the same calcium ion potentially bound to Glu-239 (Fig. 4.6). Moreover, Glu-246 and Ala-247 are located within a loop region that is the most variable in the model (see Experimental). It is possible that such loop is provided of lower conformational restrictions so as to arrange Glu-246 and Ala-247 in the right position for the interaction with a calcium ion. Noteworthy, the loop conformation with the effective orientation of these two residues is that belonging to the model with the lowest ‘objective function’ value, *i.e.* the most reliable model.

Apolar contact surface

The extension of the apolar contact surface reveals some differences. *bl*1,3-1,4Glc, *bm*1,3-1,4Glc and *pcCar* display a total hydrophobic surface of about 11959, 12301 and 14980 Å², respectively, corresponding to a surface per residue of 55.9, 58.0 and 55.3 Å². The apolar contact surface of the *LamA* model reaches an area of about 15907 Å², corresponding to a surface per residue of 60.2 Å².

Discussion

In 7.9 M GdmCl, *pfLamA* tertiary structure can be consistently regained upon calcium addition. The spectral properties of calcium-depleted *pfLamA* in 7.9 M GdmCl indicate the integrity of secondary structure elements [1] with the persistence of residual tertiary structure which shows significant alterations, such as the solvent exposure of some Trp residues, indicated by the red-shift of the fluorescence emission maximum wavelength, and the loss of the asymmetric environment for the Trp residue(s) contributing to the 295 nm band of the near-UV CD spectrum. The changed spectral properties of the protein in 7.9 M GdmCl induced by calcium addition, *i.e.* the blue-shift of the intrinsic fluorescence emission wavelength to 343 nm and the recovery of the Trp 295 nm band in the near-UV CD spectrum, suggest a notable increase in the tertiary structure which leads to a state very close to the native one. In the presence of Ca²⁺ in 7.9 M GdmCl, the structure of *pfLamA* is almost completely recovered, as judged from the spectral properties and the accessibility of hydrophobic core is comparable to that of the native state, as suggested by ANS fluorescence and acrylamide quenching. However the lack of catalytic activity confirms that this state is comparable, but not identical, to the native state.

Amongst the 380 carbohydrate active enzymes of family 16 glycoside hydrolases [2] (see URL <http://afmb.cnrs-mrs.fr/~cazy/CAZY/index.html>), 6 out of the 16 3D-structures available in PDB presents calcium ion(s) bound to the crystal structures. The interaction with Ca^{2+} is also common amongst the lectin-like proteins [44, 45] that share a similar fold with family 16 glycoside hydrolases. The stabilizing effect of calcium on proteins is well known, however the capability of *pfLamA* in 7.9 M GdmCl to interact with calcium with a consistent recovery of native tertiary structure is, to the best of our knowledge, unprecedented. In 7.9 M GdmCl, metal binding should be, at least in principle, highly unfavoured since it requires the presence of several lateral chains locked in the proper orientation and within the correct relative distance. In this report we suggest that *pfLamA* in 7.9 M GdmCl may still preserve the correct position of the residues involved in metal binding, despite the perturbation of its tertiary structure induced by the denaturant. Indeed, in 7.9 M GdmCl the interaction between *pfLamA* and Ca^{2+} displays the specificity typical for a protein-ligand interaction, e.g. the saturability. Interestingly, this interaction shows the same hyperbolic dependence upon increasing calcium concentration, independent of the spectral probe used to follow the tertiary structure changes, suggesting that the process of structural regain of *pfLamA* does not proceed via any detectable structural intermediate.

The two-state GdmCl equilibrium transition of calcium-loaded *pfLamA* leads to a state indistinguishable from that obtained upon the progressive addition of calcium to the calcium-depleted protein in 7.9 M GdmCl; this suggests that in the presence of calcium the state of the enzyme in 7.9 M GdmCl is the same, independently from the way it was formed. The spectral properties of this state are very similar to those of the *pfLamA* native state. Interestingly, the $\Delta G^{\text{H}_2\text{O}}$ value relative to the GdmCl induced fluorescence changes of the calcium-loaded protein is 41.8 kJ/mol, a value relatively high if one considers the close structural similarity between the native state and the calcium-loaded state in 7.9 M GdmCl. This observation points out the remarkably high thermodynamic stability of *pfLamA*, as also supported by the $\Delta G^{\text{H}_2\text{O}}$ value of 61.5 kJ/mol relative to the partial unfolding of the calcium-depleted *pfLamA*, and suggests that much higher $\Delta G^{\text{H}_2\text{O}}$ values may reflect the complete unfolding of this 31 kDa monomeric protein. High values of free energy of stabilization are not unprecedented, and have been reported for some oligomeric [46] and for one large monomeric protein [47], however the thermodynamic parameters determined for *pfLamA* are associated with a state which is only partially denatured. The difference in thermodynamic parameters for the GdmCl-induced fluorescence changes in the presence and absence of calcium could lead to consider the calcium-loaded protein less stable than the calcium-depleted protein. The decrease in thermodynamic stability is only apparent, since in 7.9 M GdmCl the spectral properties of the calcium-loaded state are more similar to those of the native state compared to those of the calcium-depleted state.

The relevant stability of *pfLamA*, as well as the residual tertiary structure in 7.9 M GdmCl capable to interact with calcium, can be related to some peculiar features deduced from the comparison of its homology-based model with the structures of the mesophilic counterparts. The *pfLamA* model indicates that the number of ion pairs is almost doubled in comparison with the closely related mesophilic lichenases [7] and that the extension of the hydrophobic contact surface is significantly increased. Interestingly, these structural features are reported to be important for

increasing the stability of thermophilic proteins [48, 49] and may also contribute in maintaining *pfl*LamA's residual structure in 7.9 M GdmCl [50].

The native state of *pfl*LamA does not show any fluorescence or dichroic signal changes upon Ca^{2+} addition, however the decreased accessibility of the native protein fluorophores to acrylamide in the presence of calcium suggests a more compact structure as supported by the reduced hydrogen/deuterium exchange observed in the FTIR amide II region in the presence of the cation [35]. The result of Ca^{2+} binding to the native state can be directly monitored from the changes in intensity of the FTIR signal in the region of the antisymmetric COO^- stretching vibration of the carboxylate moiety of the amino acid side chains of Glu and Asp residues accompanied by an upshift of the corresponding symmetric bands. This evidence is in line with the possible involvement of Asp and Glu predicted by the homology modelling of *pfl*LamA. The quantitative analysis of calcium binding by 5,5'-Br₂-BAPTA titration indicates that two Ca^{2+} interact with native *pfl*LamA with high affinity. A correlation between protein stabilization and *in vivo* calcium concentration in *P. furiosus* is difficult to establish at the moment, also in consideration of the variability of Ca^{2+} concentration in the marine hydrothermal environment [51]. The effect of different metals on the residual tertiary structure of *pfl*LamA is in progress to make a comparative analysis with the structural changes induced by calcium and to gain information about the interactions of metals with non-native states of proteins.

Acknowledgments

We thank Dr. Roberto Contestabile for critical reading of the manuscript. This work was partially supported by a grant 'Progetti Strategici MIUR Legge 499/97' Project GENEFUN.

References

1. Chiaraluce, R., Van Der Oost, J., Lebbink, J.H., Kaper, T. and Consalvi, V. (2002) Persistence of tertiary structure in 7.9 M guanidinium chloride: the case of endo-beta-1,3-glucanase from *Pyrococcus furiosus*. *Biochemistry* **41**, 14624-14632.
2. Coutinho, P.M. and Henrissat, B. (1999) in Recent Advances in Carbohydrate Bioengineering (Gilbert, H.J., Davies, G.J., Henrissat, B. and Svensson, B., eds.) pp. 3-12, Royal Society of Chemistry, Cambridge.
3. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C. and Eddy, S.R. (2004) The Pfam protein families database. *Nucleic Acids Res.* **32**, Database issue: D138-141.
4. Allouch, J., Jam, M., Helbert, W., Barbeyron, T., Kloareg, B., Henrissat, B. and Czjzek, M. (2003) The three-dimensional structures of two beta-agarases. *J. Biol. Chem.* **278**, 47171-47180.
5. Gueguen, Y., Voorhorst, W.G., van der Oost J. and de Vos W.M. (1997) Molecular and biochemical characterization of an endo- β -1,3- glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* **272**, 31258-31264.

6. Karplus, K., Karchin, R., Draper, J., Casper, J., Mandel-Gutfreund, Y., Diekhans, M. and Hughey, R. (2003) Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. *Proteins* **53**, Suppl. 6, 491-496.
7. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. *Structure* **9**, 513-525.
8. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.* **28**, 235-242.
9. Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**, 739-747.
10. Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTALW: improving the sensitivity of progressive multiple alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
11. InsightII, Version 2000 (2000) Molecular Modeling System, User Guide. Accelrys 2000, San Diego, CA.
12. Friedberg, I., Kaplan, T. and Margalit, H. (2000) Evaluation of PSI-BLAST alignment accuracy in comparison to structural alignments. *Protein Sci.* **9**, 2278-2284.
13. Šali, A., Potterton, L., Yuan, F., Van Vlijmen, H. and Karplus, M. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins* **23**, 318-326.
14. Burke, D.F., Deane, C.M., Nagarajaram, H.A., Campillo, N., Martin-Martinez, M., Mendez, J. and Molina, F. (1999) An iterative structure-assisted approach to sequence alignment and comparative modeling. *Proteins* **3**, 55-60.
15. Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* **17**, 355-362.
16. Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-291.
17. Drablos, F. (1999) Clustering of non-polar contacts in proteins. *Bioinformatics* **15**, 501-509.
18. Connolly, M.L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**, 709-713.
19. Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56.
20. Barlow, D.J. and Thornton, J.M. (1983) Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867-885.
21. Andre, I. and Linse, S. (2002) Measurement of Ca²⁺-binding constants of proteins and presentation of the CaLigator software. *Anal. Biochem.* **305**, 195-205.
22. Kaper, T., Verhees, C.H., Lebbink, J.H., van Lieshout, J.F., Kluskens, L.D., Ward, D.E., Kengen, S.W., Beerthuyzen, M.M., de Vos, W.M. and van der Oost, J. (2001) Characterization of beta-glycosylhydrolases from *Pyrococcus furiosus*. *Methods Enzymol.* **330**, 329-346.
23. Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319-326.
24. Oberg, K. A. and Fink, A. L. (1998) A new attenuated total reflectance Fourier transform infrared spectroscopy method for the study of proteins in solution. *Anal. Biochem.* **256**, 92-106.
25. Semisotnov, G.V., Rodionova, N.A., Razgulyaev, O.I., Uversky, V.N., Gripas', A.F. and Gilmanshin, R.I. (1991) Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe. *Biopolymers* **31**, 119-128.
26. Lehrer, S.S. (1971) Solute perturbation of protein fluorescence. The quenching of the tryptophyl fluorescence of model compounds and of lysozyme by iodide ion. *Biochemistry* **10**, 3254-3263.
27. Ionescu, R., Smith, V. F., O'Neill, J. C. Jr. and Matthews, C. R. (2000) Multistate equilibrium unfolding of *Escherichia coli* dihydrofolate reductase: thermodynamic and spectroscopic description of the native, intermediate, and unfolded ensembles. *Biochemistry* **39**, 9540-9550.
28. Santoro, M. M. and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**, 8063-8068.

29. **Tsien, R.Y.** (1980) New calcium indicators and buffers with high selectivity against magnesium and protons: design, synthesis, and properties of prototype structures. *Biochemistry* **19**, 2396-2404.
30. **Patton, C., Thompson, S. and Epel, D.** (2004) Some precautions in using chelators to buffer metals in biological solutions. *Cell Calcium* **35**, 427-431.
31. **Jackson, M. and Mantsch, H.H.** (1995) The use and misuse of FTIR spectroscopy in the determination of protein structure. *Crit. Rev. Biochem. Mol. Biol.* **30**, 95-120.
32. **Barth, A.** (2000) The infrared absorption of amino acid side chains. *Prog. Biophys. Mol. Biol.* **74**, 141-173.
33. **Susi, H.** (1972) Infrared spectroscopy conformation. *Methods Enzymol.* **26**, 455-472.
34. **Barth, A. and Zscherp, C.** (2002) What vibrations tell us about proteins. *Q. Rev. Biophys.* **35**, 369-430.
35. **Jackson, M., Haris, P.I. and Chapman, D.** (1991) Fourier transform infrared spectroscopic studies of Ca^{2+} -binding proteins. *Biochemistry* **30**, 9681-9686.
36. **Fabian, H. and Vogel, H.J.** (2002) Fourier transform infrared spectroscopy of calcium-binding proteins. *Methods Mol. Biol.* **173**, 57-74.
37. **Kabsch, W., and Sander, C.** (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577-2637.
38. **Pollastri, G., Przybylski, D., Rost, B. and Baldi, P.** (2002) Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles. *Proteins* **47**, 228-235.
39. **Heredia, P. and De Las Rivas, J.** (2003) Calcium-dependent conformational change and thermal stability of the isolated PsbO protein detected by FTIR spectroscopy. *Biochemistry* **42**, 11831-11838.
40. **Fabian, H., Naumann, D., Misselwitz, R., Ristau, O., Gerlach, D. and Welfle, H.** (1992) Secondary structure of streptokinase in aqueous solution: a Fourier transform infrared spectroscopic study. *Biochemistry* **31**, 6532-6538.
41. **Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C.** (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536-540.
42. **Chehin, R., Iloro, I., Marcos, M.J., Villar, E., Shnyrov, V.L. and Arrondo, J.L.** (1999) Thermal and pH-induced conformational changes of a beta-sheet protein monitored by infrared spectroscopy. *Biochemistry* **38**, 1525-1530.
43. **Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F.** (1999) ESPript: analysis of multiple sequence alignments using PostScript. *Bioinformatics* **15**, 305-308.
44. **Cioci G., Mitchell E.P., Gautier C., Wimmerova M., Sudakevitz D, Perez S, Gilboa-Garber, N. and Imberty A.** (2003) Structural basis of calcium and galactose recognition by the lectin PA-IL of *Pseudomonas aeruginosa*. *FEBS Lett.* **555**, 297-301.
45. **Loris, R.** (2002) Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* **1572**, 198-208.
46. **Mitra N., Srinivas, V.R., Ramya, T.N., Ahmad, N., Reddy, G.B. and Surolia, A.** (2002) Conformational stability of legume lectins reflect their different modes of quaternary association: solvent denaturation studies on concanavalin A and winged bean acidic agglutinin. *Biochemistry* **41**, 9256-9263.
47. **Schoeffler, A.J., Joubert, A.M., Peng, F., Khan, F., Liu, C.C. and LiCata, V.J.** (2004) Extreme free energy of stabilization of Taq DNA polymerase. *Proteins* **54**, 616-621.
48. **Yano, J.K. and Poulos, T.L.** (2003) New understandings of thermostable and peizostable enzymes. *Curr. Opin. Biotechnol.* **14**, 360-365.
49. **Kumar, S. and Nussinov, R.** (2001) How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* **58**, 1216-1233.
50. **Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J.M. and Marqusee, S.** (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. USA.* **100**, 11345-11349.
51. **Holden, J.F. and Adams, M.W.** (2003) Microbe-metal interactions in marine hydrothermal environments. *Curr. Opin. Chem. Biol.* **7**, 160-165.

Chapter 5

In Situ Structure and Activity Studies of an Enzyme Adsorbed on Spectroscopically Undetectable Particles

Koutsopoulos, S., Tjeerdsma, A.-M., Van Lieshout, J.F.T., Van der Oost, J., Norde, W.

Biomacromolecules **6**, 1176-1184

Abstract

The structural characteristics and the activity of a hyperthermophilic endoglucanase were investigated upon adsorption. Silica (hydrophilic) and Teflon (hydrophobic) surfaces were selected for the study. The materials were specially designed so that the interaction of the particles with light was negligible and the enzyme conformation in the adsorbed state was monitored *in-situ*. The adsorption isotherms were determined, and the adsorbed endoglucanase was studied using a number of spectroscopic techniques, enzymatic activity tests, and dynamic light scattering. Experiments were performed at pH values below, at, and above the isoelectric point of the enzyme. It was shown that the enzyme adsorbed on the hydrophobic surface of Teflon with higher affinity as compared to the hydrophilic silica nanoparticles. In all cases, adsorption was followed by (slight) changes in the secondary structure resulting in decreased β -structural content. The changes were more profound upon adsorption on Teflon. The adsorbed enzyme remained active in the adsorbed state in spite of the structural changes induced when interacting with the surfaces.

Introduction

Hyperthermophiles thrive in natural superheated environments above 100°C in subterranean geothermally heated biotopes and hydrothermally heated rocks in the deep sea and in artificial settings such as hot outflows from power plants [1-4]. The stability of hyperthermophilic microorganisms and of their biomolecules under such extreme conditions opened a challenging research field with very promising applications [1]. In many biocatalytical and biotechnological processes proper yields are reached only under extreme conditions of temperature. Most microorganisms and enzymes lose activity in such immoderate environments, but hyperthermophiles and hyperthermostable enzymes form an exception to this rule [4]. Increasing the temperature in the production line results in higher rate constants and higher diffusion coefficients that lead to more efficient processes. Moreover, working at high temperatures minimizes the risk of microbial contamination.

Pyrococcus furiosus flourishes in the low depth marine sand surroundings of sulfurous marine volcanic areas at temperatures ranging from 70 to 103°C and pH 5 – 9 [5]. A monomeric, extracellular endo- β -1,3-glucanase (LamA) from the hyperthermophilic *P. furiosus* was isolated and characterized [6]. LamA hydrolyzes β -branched polysaccharides such as laminarin, which is the main storage product of various sea organisms such as algae [4-7]. It shows maximum enzymatic activity at 104°C with a relatively long half-life of thermal inactivation. Calorimetric studies have shown that LamA maintains its active conformation up to 135°C depending on pH and interaction with surfaces [8,9]. Hence, LamA can be used in protein microarray technology to probe various biochemical activities and in biotechnology as a biosensor and as a biocatalyst. Moreover, LamA is a natural protein from a well-studied microorganism with a known metabolism, and, therefore, its use can be easily certified. This is very important for the food industry. The high-temperature enzymatic activity, efficiency, and selectivity of LamA can be used to improve the

filterability and decrease the viscosity of the wort in the brewing industry or remove β -glucans during wine processing and beverage production [10]. Furthermore, because microbial cell walls contain β -glucans, LamA can aid in cell lysing and provide further sterilization in a high-temperature environment used to recover specific cell products and pharmaceuticals [10,11].

In various industrial applications, the immobilization of the enzyme is required. Covalent binding is such a means, but this method often induces structural perturbations that may result in loss of the enzymatic activity. Physical adsorption, on the other hand, is a relatively mild method to immobilize proteins on a carrier. Nevertheless, in most cases, proteins that are biologically active in solution become inactive in the adsorbed state. Hence, for practical and economical reasons, it is important to investigate the physical, chemical, and biological properties of surface-bound proteins for process optimization. For this study two types of surfaces were selected: Teflon latex (hydrophobic) and silica nanoparticles (hydrophilic) are well-characterized materials. Hydrophobic surfaces have little or no tendency to adsorb water, and, hence, they are poorly wetted. Hydrophilic surfaces exhibit an affinity for water which readily adsorbs and wets the surface, forming a film. Both types of particles were specially designed so that they do not interact with light. Using these particle suspensions made it possible to investigate the protein features in the adsorbed state applying standard spectrophotometric techniques without interference from the particles.

We here report on LamA adsorption, emphasizing the relation between conformational characteristics and enzymatic activity in the adsorbed state. Structural features of LamA adsorbed at the solid/liquid interface are compared with those of the enzyme in solution. The effect of the sorbent's surface properties on the enzyme characteristics is discussed as well.

Materials and Methods

Purification of LamA - LamA was produced heterologously in *Escherichia coli* BL21(DE3) using the T7 expression system and subsequently purified by fast protein liquid chromatography (FPLC) as described in detail elsewhere [6-13]. The purity of the enzyme was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectroscopy (PerSeptive Biosystems Voyager DE-RP mass spectrometer, using sinapinic acid crystallized on a gold-coated well plate; spectra were calibrated with protein standards). The isoelectric point (IEP) of LamA in solution is at pH 4.4, as determined by isoelectric focusing. After purification the isolated protein was stored at 4°C in 0.01 M sodium phosphate buffer at pH 7.0 without the addition of azide or other preservatives which might alter the physicochemical characteristics of the protein.

Sorbents - Negatively charged hydrophobic Teflon particles (copolymer of tetrafluoroethylene and perfluorovinyl ether) were prepared by emulsion polymerization. The colloidal suspension was stable in water owing to the sulphate groups on the surface of the particles originating from potassium persulphate which was used as a polymerization initiator. The particles were free of nonionic stabilizers and other contaminants. The surface charge provides colloidal stability to the

particles. Electrokinetic measurements showed that the particles are negatively charged at the pH values investigated. Monodispersity was confirmed by dynamic light scattering, DLS, and transmission electron microscopy, TEM, (JEOL 1200 EX). From this data the average diameter of the particles was calculated to be 215 nm. The specific surface area of the smooth particles is 12.5 m²/g as determined by N₂ adsorption triple-point BET method. The charged groups, providing colloidal stability to the Teflon suspension, cover only a slight fraction of the surface area of the particles. Consequently, Teflon is hydrophobic as shown from contact angle measurements of a sessile droplet of 0.01 M sodium phosphate solution at pH 7.0 on pelletized particles. Microscopic analysis revealed a contact angle of 96°.

Negatively charged silica particles (Ludox HS-40, Aldrich) are nonporous, smooth spheres with a diameter of 13 nm, as accurately determined by TEM analysis. The suspension is monodisperse as confirmed by DLS measurements. The specific surface area was estimated using the triple-point BET method based on N₂ adsorption (220 m²/g). At the experimental conditions employed the surface is negatively charged (point of zero charge, pzc, is at pH 2.7) and highly hydrophilic as shown from water contact angle measurements on pelletized particles; water droplets spread on the surface and the contact angle was below 5°.

The particles used are suitable to study *in-situ* the structural characteristics of the protein in the adsorbed state using standard spectroscopic techniques. The refractive index of the Teflon particles in suspension is 1.35, which is close to the respective value of water (1.33), and moreover, this preparation does not contain UV-absorbing groups. Hence, a low concentration of protein-covered Teflon particles can be directly used for circular dichroism (CD) and fluorescence measurements on the adsorbed protein without interference of light with the Teflon particles. The extremely small size of the ultrafine silica nanoparticles allows for negligible scattering of light by a highly dispersed particle suspension, which moreover, is characterized by large area/volume ratio enabling high protein concentrations in the adsorbed state.

Adsorption experiments and desorption tests - The adsorption experiments were performed in Eppendorf tubes each containing 1.2 ml of the sample. Each tube, which was previously tested and did not give significant adsorption of LamA, contained different LamA concentrations in 0.01 M sodium phosphate buffer at the desired pH and the same total sorbent surface area (i.e., 0.1 m² per ml of solution), to give a series of samples with the same total volume. The adsorption experiments were performed at different pH values below, at and above the IEP of LamA to investigate the effect of pH on the adsorption process. The protein solution and the Teflon and silica suspensions were first brought to the desired pH before mixing. The samples were incubated end-over-end overnight at room temperature. This period of time was sufficient for adsorption to reach a steady-state. The enzyme-covered particles were separated from unbound protein in the solution by centrifugation. The concentration of LamA in the supernatant was measured with a spectrophotometer from the absorption peak at 280 nm (the extinction coefficient $\epsilon_{280} = 83,190 \text{ M}^{-1} \text{ cm}^{-1}$ was calculated according to Gill and von Hippel) [14]. The adsorbed amount per unit surface area, Γ , was calculated from the difference of the protein concentration in solution before and after adsorption, c_{eq} , using mass balance considerations and the sorbent's specific surface area. All data

points in the adsorption isotherm represent an average of three experiments. After separation from the supernatant the enzyme-covered particles were redispersed in the original buffer and used for further analyses. Immobilization of LamA in the adsorbed state was tested by overnight incubation of the redispersed enzyme-covered particles at room temperature and for 6 h at 90°C. The enzyme-covered particles were redispersed in 2 mL of the original sodium phosphate buffer and after heat-treatment were centrifuged again, and the supernatant was examined for desorbed protein with a spectrophotometer at 280 nm.

Enzymatic activity measurements - The enzymatic activity of LamA free in solution and in the adsorbed state was measured for 10 min at 80°C using the 3,5-dinitrosalicylic acid (DNS) assay [15]. The method is based on the spectrophotometric determination of the amount of hydrolyzed ends carried by the oligosaccharides after degradation of the substrate (i.e., laminarin, Sigma). Glucose was used for the calibration. Comparison with a standard curve gave the specific enzymatic activity expressed in moles of mono- and oligo-saccharides per gram of enzyme. The specific enzymatic activities of LamA in solution and in the adsorbed state were then compared to assess the effect of adsorption on the biological activity. Samples of adsorbed LamA were selected along the adsorption isotherm. The Michaelis-Menten constant, K_m , of LamA for the hydrolysis of laminarin is 2.8 mg ml⁻¹ [6]. To test whether adsorption affected K_m , the enzymatic activities were also measured using a range of 0.5 – 100 mg of substrate/ml.

Dynamic light scattering - The 5 mg ml⁻¹ solution of LamA in 0.01 M phosphate buffer at pH 3.5, 4.4 and 7.0 was filtered through 0.45 µm pore size Acrodisc filters prior to measuring light scattering. The DLS measurements were carried out on a home-built setup equipped with a ALV-5000/E multiple tau digital correlator, an ALV-125 argon laser light scattering spectrometer/goniometer and an ALV/SO SPID detector. The laser wavelength was set to 488 nm, at a fixed scattering angle of 90°. Each sample was measured 20 times for 20 s. The solvent viscosity and the refractive index of the buffer were assumed to be 0.8935 cP and 1.3325, respectively, at 20°C. The data were processed using the CONTIN algorithm [16].

Circular Dichroism - Far-UV CD spectra (190 – 260 nm) of LamA before and after adsorption were recorded in a J-715 (JASCO) spectrophotometer equipped with a Jasco PTC 348 WI temperature controller set at 20°C. Samples were selected from different domains of the adsorption isotherm. The enzyme-covered particles were resuspended in 0.01 M phosphate buffer so as to have an enzyme concentration of 0.3 mg ml⁻¹. Solutions of LamA with the same concentration were also measured and compared with those of adsorbed LamA at the same pH. Spectra were recorded in 0.1-cm path length quartz cuvettes with a scan rate of 100 nm/min, and 0.2 nm resolution. The spectra resulted from accumulation of 32 scans which were subsequently averaged. Blank spectra of the buffer and of the particles without protein were obtained at identical conditions and subtracted from the spectra containing the protein. Suspensions of LamA-covered Teflon particles at pH 4.4 and at pH 3.5 were turbid and slowly precipitated. Therefore, CD spectra at these conditions were averaged over a lower number of spectra and were acquired at higher scan rates.

Data analysis was performed by fitting the acquired spectra with reference spectra using the CONTIN software, which is based on nonlinear regression fitting algorithms without constraints (ridge-regression analysis) [17,18]. This program gives a much better estimate of β -sheets and turns than simple multiple linear regression [19]. An average molar mass of 110 Da per amino acid residue of the protein was used for calculating the ellipticity, θ .

Fluorescence emission spectroscopy - Emission steady-state fluorescence was measured by a Varian Cary Eclipse spectrophotometer. All measurements were carried out at 20°C using quartz cuvettes of 1-cm path length. Emission spectra of 0.03 mg ml⁻¹ LamA at pH 3.5, 4.4, and 7.0 adsorbed on Teflon or silica and free in solution were recorded in the range 300 – 400 nm on excitation at 300 nm. Different samples along the isotherm were selected to see the influence of surface coverage on the tertiary structure of the adsorbed enzyme. The excitation and emission slit widths were set at 5.0 and 2.5 nm, respectively. All spectra were corrected for the background emission peak of water.

Results and Discussion

LamA is a single domain protein with a molar mass of 30,085 Da. For the graphical representation in Fig. 5.1 molecular modeling software was utilized [20], assuming structural similarity of LamA with the highly homologous 1,3-1,4- β -glucanase from *Bacillus licheniformis* and with a κ -carrageenase fragment from *Pseudoalteromonas carrageenovora* whose crystal structures are known (PDB entries 1GBG and 1DYP, respectively) [21,22]. Preliminary NMR analysis of the solution structure of LamA and CD and fluorescence emission spectroscopic data from this work were also taken into consideration in the modeling process and the selection of the best model. The secondary structure is dominated by β -sheets and turns. The sheets are slightly bent and stacked on top of each other. The shape of the protein is globular-ellipsoid and the calculated dimensions are 4.6 nm \times 3.2 nm \times 3.4 nm. Experimental data from DLS, mass spectroscopy (MALDI TOF), and size exclusion chromatography showed that native LamA in solution is a monomer and, hence, higher complexation clusters of the enzyme molecules will not interfere with the interpretation of the data from the adsorption experiments.

Dynamic light scattering

The effect of pH on the size of LamA in solution was investigated by DLS. Analysis of the data showed that the hydrodynamic radius of the negatively charged LamA at pH 7.0 is 3.1 ± 0.2 nm (Fig. 5.2). At the IEP (pH 4.4) and below (pH 3.5) the LamA molecules appear slightly bigger with radii of 4.0 ± 0.2 nm and 3.6 ± 0.2 nm, respectively. These values are in fairly good agreement with those calculated from time-resolved anisotropy experiments (data to be published) and the molecular modeling if a hydration layer of 0.23 nm surrounding the protein in solution is taken into account. The small difference between the hydrodynamic radius and the dimensions deduced by modeling are probably due to the assumption made in the DLS calculations that the protein is spher-

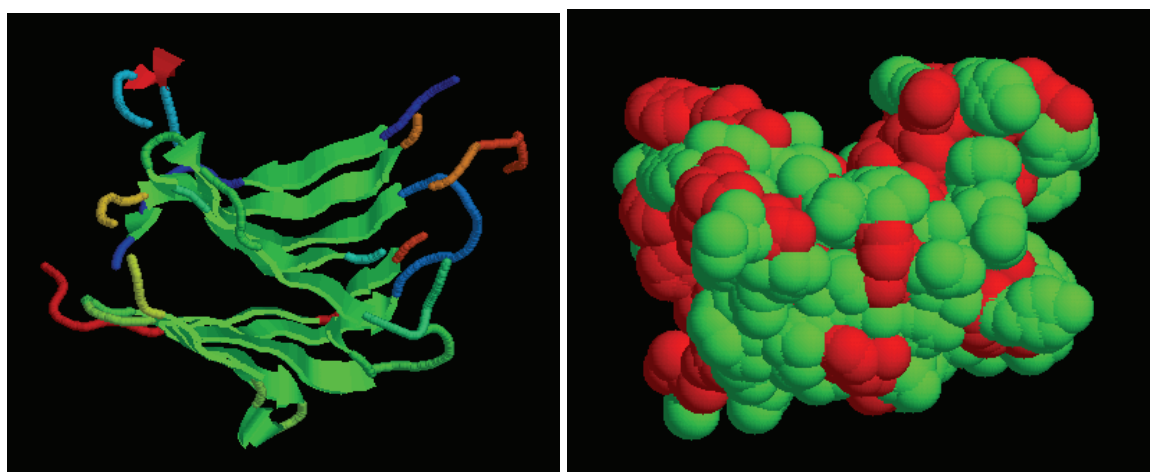


Figure 5.1 Graphic display of the structure of LamA. The red balls represent hydrophobic amino-acids, and the green represent polar amino acids. The enzymatic cleft may be seen on the top of the structural representation. Left and right panels represent the backbone shown as ribbon and the entire molecule in a space-filling format, respectively (graphics created with Rasmol).

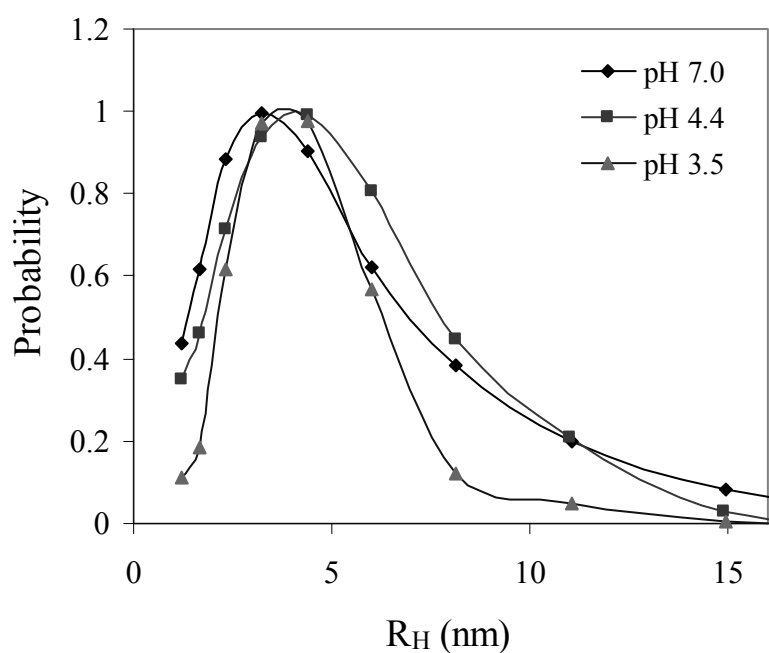


Figure 5.2 Size distribution of LamA at pH 3.5, 4.4, and 7.0 from DLS measurements. For the data analysis the CONTIN algorithm was used.

rical instead of elliptical. The results show that at the IEP LamA is bigger, while at pH 7.0, which is close to the pH of maximum activity (i.e., pH 6.5), the enzyme acquires a more compact form. The differences in the calculated size in solution of different pH values may be due to variations in the efficiency of the protein to form hydrogen bonds at different pH values.

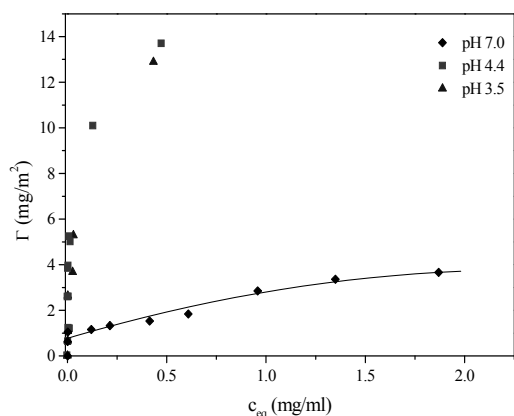


Figure 5.3 Adsorption isotherms for LamA on negatively charged hydrophobic Teflon particles in 0.01 M sodium phosphate buffer at pH 3.5, 4.4, and 7.0.

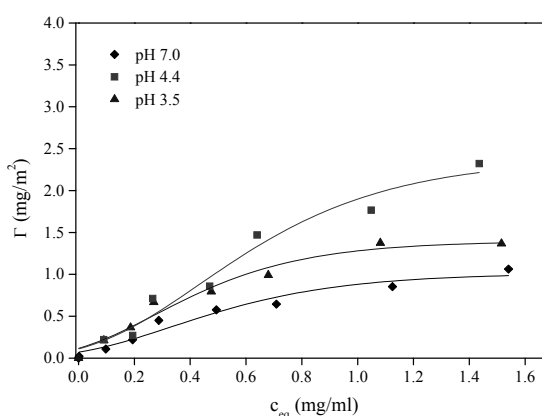


Figure 5.4 Adsorption isotherms for LamA on negatively charged hydrophilic silica particles in 0.01 M sodium phosphate buffer at pH 3.5, 4.4, and 7.0.

Adsorption isotherms

Fig. 5.3 and 5.4 show the adsorption isotherms for LamA on the hydrophobic Teflon particles and on the hydrophilic silica nanoparticles, respectively, at different pH values. The surface concentration of LamA (Γ) is plotted as a function of the enzyme concentration in solution (c_{eq}) after a steady-state was reached. At the pH values investigated, that is, 3.5, 4.4 and 7.0, both Teflon and silica (pzc at pH 2.7) surfaces are negatively charged but the net surface charge of LamA varies from positive to neutral to negative. Hence, clearly, because at pH 7.0 the negatively charged protein adsorbs on both types of negatively charged surfaces, the adsorption of LamA is not dominated by electrostatic interactions [23].

At all pHs the initial part of the adsorption isotherms of LamA on Teflon coincide with the Γ -axis, suggesting high affinity of the enzyme for the sorbent possibly owing to dehydration of the hydrophobic surface. At pH 7.0, where LamA is negatively charged, the adsorption isotherm reaches a plateau at about 3.5 mg/m^2 . Using the molecular dimensions of LamA, this surface concentration is compatible with complete monolayer formation of native LamA molecules. This also suggests that LamA is not structurally collapsed upon adsorption. If adsorption-induced spreading of the protein molecules was involved the isotherms would have revealed significantly smaller amounts of adsorbed LamA than that required for monolayer coverage with native molecules. At pH values below and at the IEP adsorption of LamA leads to multilayer formation (Fig. 5.3). Also at these pH values adsorption of LamA resulted in particle aggregation. This implies that the repulsive forces which maintain colloidal stability of the suspended Teflon particles are weakened or screened owing to the adsorbed enzyme. Turbidity was more evident when Teflon particles were allowed to interact with solutions of low LamA concentration. More concentrated samples of LamA resulted in less turbid suspensions possibly as a result of the stabilizing electrostatic interactions between enzyme-covered-particles. Aggregation may also have resulted in enzyme encapsulated between the particles, and this phenomenon can be falsely interpreted as increased adsorption.

As shown in Fig. 5.4, the initial slopes of the adsorption isotherms of LamA on silica do not coincide with the vertical axis, which implies low affinity for the surface. The adsorption profiles were rather similar for all three pH values even though LamA's net charge is different at each one of them. Inspection of the isotherms reveals two very interesting features. The first is that electrostatic interactions between LamA and the silica surface affect adsorption but they are not critical. Negatively charged LamA was readily adsorbed on the like-charged silica surface at pH 7.0. The highest adsorbed amount of LamA on silica is observed at the IEP of the enzyme. Saturation of the surface with enzyme was found to be at 1.4, 2.5 and 1.1 mg/m² for adsorption at pH 3.5, 4.4 and 7.0, respectively. Lower adsorption at either side of the IEP may be explained by lateral repulsive interactions between adjacent protein molecules in the adsorbed state. The second observation is the sigmoidal shape of the isotherm which is more profound in the cases of adsorption of LamA at pH 7.0 and 4.4. This pattern implies a cooperative adsorption mechanism employed by LamA to overcome the low affinity for the surface. Adsorption of the first LamA molecules facilitates further adsorption by a mechanism that promotes protein-protein interactions at the silica surface. Furthermore, it has been shown that adsorption of a charged species on a surface with the same polarity induces opposite charges to the surface to meet electroneutrality conditions. This implies that the surface of silica could be less negative or even positive upon adsorption of negatively charged LamA molecules at pH 7.0 and 4.4. In this regard further adsorption is facilitated as previously observed in other adsorption systems and predicted theoretically [8,9,24,25].

Comparison of the adsorption isotherms for the same pH shows that the hydrophobic surface of Teflon attracts more enzyme molecules as compared to the hydrophilic silica. Molecular modeling showed that the external surface of LamA is covered by hydrophobic groups up to about 30%. Dehydration of the hydrophobic Teflon surface in favor of adsorbing protein molecules is entropically highly favorable and, hence, strongly contributes to the Gibbs energy of adsorption, whereas at the hydrophilic silica surface LamA and water compete for adsorption. Adsorption-induced structural rearrangements of LamA are not likely, as will be shown in detail in a following paragraph (except in the case of low surface coverage at pH 3.5).

In addition to hydrophobic dehydration, the adsorption of LamA on Teflon at pH 3.5 is assisted by the attractive electrostatic forces between the surfaces of the enzyme and the sorbent. At zero net charge on the protein (pH 4.4) the electrostatic contribution is minimal, and at pH 7.0, where the protein and the Teflon surface repel each other electrostatically, adsorption still occurs because of the prevailing entropic contribution from hydrophobic dehydration.

The situation is different for adsorption of LamA on the negatively charged hydrophilic silica. At the IEP, electrostatic interactions do not oppose adsorption, and below the IEP, the overall charge of the enzyme is positive and adsorption is aided by electrostatic attraction. At pH 7.0 spontaneous adsorption still occurs despite adverse hydration effects and opposing electrostatic interactions. Adsorption under such conditions could result from increased conformational entropy associated with structural rearrangements in the adsorbing protein. Conformational entropy-driven adsorption is often observed for "soft" proteins, that is, proteins with relatively low structural stability [26]. Nevertheless, as will be demonstrated in a following section, LamA on silica retains

structural integrity and biological functioning in the adsorbed state and, therefore, it can be categorized as a “hard” protein. However, data from literature have shown that the interaction of “hard” proteins with hydrophilic surfaces leads to adsorption only under electrostatically attractive conditions [27-29]. The adsorption of LamA on silica is an exemption to the rule and probably originates from favorable interactions between the protein molecules at the surface, which leads to a cooperative adsorption mode (reflected by the sigmoidal shape of the isotherms).

The charge on the protein-covered sorbent surface is compensated by ions in solution, together forming an electrical double layer. By using standard formulation from interfacial thermodynamics, it was shown that under the experimental conditions employed in the adsorption experiments the thickness of the electrical double layer is about 1.5 nm. During adsorption the electrical double layer exerts an orientation effect on small proteins according to their electric dipole [30]. However, LamA is larger than this layer and, therefore, only partly immersed in it. Hence, the electrical double layer around the particle does not sense uneven charge distribution on the LamA molecule, which suggests that LamA molecules will not be electrostatically directed by the surface to a preferred orientation.

Heat treatment up to 90°C for 6 h did not result in desorption of LamA from Teflon and silica. LamA is immobilized in the adsorbed state even at electrostatically adverse conditions on both hydrophobic and hydrophilic surfaces. This information is important for the consistency of the upcoming analyses (e.g., enzymatic activity tests and spectroscopic studies). Immobilization can be due to multiple adsorption contact points of LamA with the surface which requires a high activation energy of desorption. In the case of adsorption on Teflon, hydrophobic interactions also opposes desorption while immobilization on silica may be due to the cooperative character of the isotherm as protein-protein interactions result in clusters of adsorbed LamA molecules at the silica surface. This would make desorption more difficult to occur because it requires detachment of the complete cluster rather than of single LamA molecules.

Circular Dichroism

Typical far-UV CD spectra of LamA in solution and adsorbed on Teflon and silica particles are shown in Fig. 5.5. The spectral profiles are characterized by a broad negative peak at 217 nm and a positive absorption band from about 207 nm, which indicate the predominant presence of β -structures (sheets and turns) and negligible contribution from α -helices. Spectral analysis in Table 5.1 presents the relative content of LamA’s secondary structural elements. At pH 3.5, 4.4 and 7.0 the CD spectra are virtually identical suggesting that the secondary structure of LamA is not significantly affected by pH. This is in line with data reported in the literature [31].

Upon adsorption on Teflon the far-UV CD spectra of LamA slightly deviate from that of LamA in solution. The main difference is the broadening of the negative absorption band and the absence of the shoulder at about 230 nm (Fig. 5.5, left panel). As compared to LamA in solution the spectral analysis showed that LamA adsorbed on Teflon at pH 7.0 has a lower content of β -sheets and turns and an increase of non-resolved domains (Table 5.1), along with a notable increase of the α -helical content as has been previously observed for the adsorption of amyloid β -peptide [32]. This phenomenon was observed at low and high monolayer coverage at pH 7.0 and at low surface

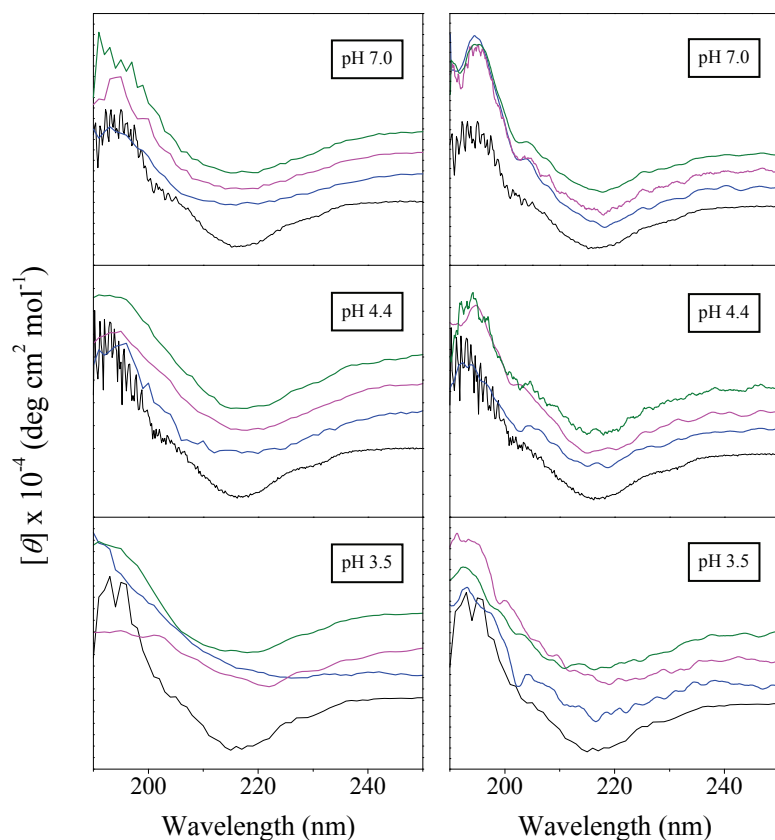


Figure 5.5 Far-UV CD spectra of LamA in solution (black lines), adsorbed on Teflon latex (left panel) and adsorbed on silica nanoparticles (right panel) at different surface coverages and pH. Lines represent low (blue), medium (magenta), and high (green) surface coverages, as shown in Table 5.1. Graphs were vertically shifted for comparison.

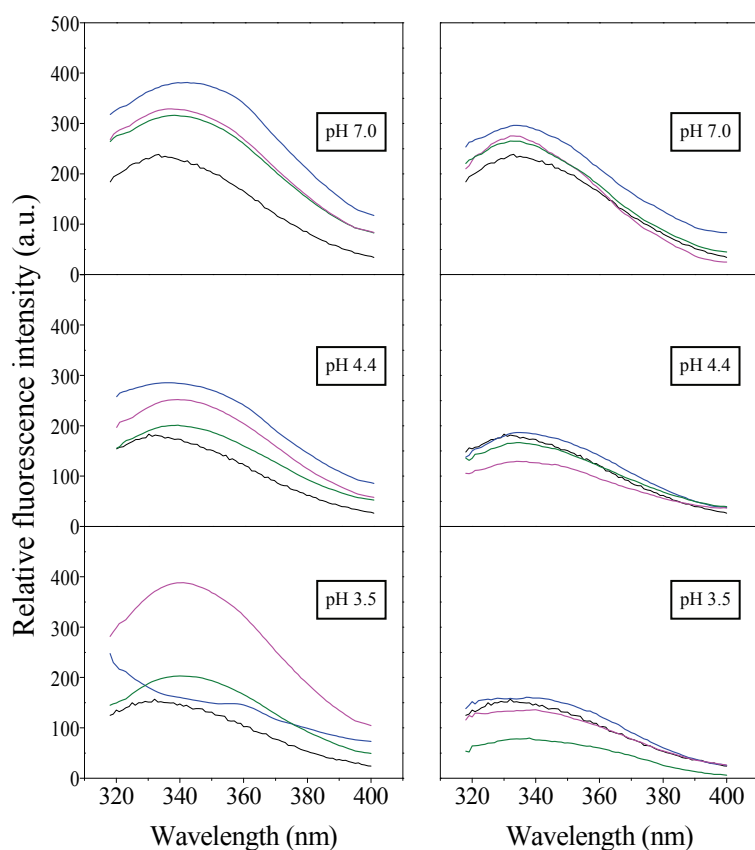


Figure 5.6 Fluorescence emission spectra of 0.03 mg ml⁻¹ LamA obtained on excitation at 300 nm in solution (black lines), adsorbed on Teflon latex (left panel), and adsorbed on silica nanoparticles (right panel) at different surface coverages and pH. Lines represent low (blue), medium (magenta) and high (green) surface coverages, as shown in Table 5.1.

Table 5.1 Secondary structure of LamA in 0.01 M sodium phosphate buffer solution at pH 7.0, 4.4, and 3.5 and in the adsorbed state at different surface coverages as calculated from the CD spectra using the CONTIN program

			adsorbed LamA (mg/m ²)	α -helix (%)	β -sheet (%)	β -turn (%)	unknown (%)
LamA in solution pH 7.0				1 \pm 0	79 \pm 3	17 \pm 2	3 \pm 1
LamA on Teflon pH 7.0	low	1		7 \pm 2	58 \pm 4	23 \pm 3	12 \pm 6
	medium	2		7 \pm 2	64 \pm 3	15 \pm 2	14 \pm 3
	high	3		8 \pm 4	68 \pm 4	18 \pm 3	6 \pm 4
LamA on silica pH 7.0	low	0.6		1 \pm 1	75 \pm 4	16 \pm 3	9 \pm 4
	medium	0.9		0 \pm 0	76 \pm 3	15 \pm 2	9 \pm 3
	high	1.1		0 \pm 0	77 \pm 5	16 \pm 4	7 \pm 6
LamA in solution pH 4.4				1 \pm 1	79 \pm 2	16 \pm 2	4 \pm 2
LamA on Teflon pH 4.4	low	1		7 \pm 2	56 \pm 3	15 \pm 2	22 \pm 2
	medium	4		3 \pm 1	63 \pm 4	18 \pm 2	16 \pm 4
	high	10		3 \pm 1	62 \pm 3	20 \pm 1	15 \pm 2
LamA on silica pH 4.4	low	0.9		0 \pm 0	75 \pm 2	21 \pm 2	4 \pm 3
	medium	1.5		0 \pm 0	68 \pm 1	21 \pm 1	11 \pm 3
	high	3.2		0 \pm 0	78 \pm 2	19 \pm 2	3 \pm 1
LamA in solution pH 3.5				1 \pm 1	81 \pm 3	15 \pm 3	3 \pm 1
LamA on Teflon pH 3.5	low	1		3 \pm 2	51 \pm 4	21 \pm 2	25 \pm 5
	medium	4		10 \pm 6	59 \pm 8	16 \pm 7	15 \pm 8
	high	13		5 \pm 2	61 \pm 4	22 \pm 2	12 \pm 5
LamA on silica pH 3.5	low	1.0		2 \pm 1	63 \pm 1	21 \pm 1	14 \pm 2
	medium	1.4		0 \pm 0	67 \pm 2	27 \pm 2	6 \pm 3
	high	3.0		2 \pm 1	63 \pm 2	28 \pm 2	7 \pm 2

coverage at pH 3.5 and 4.4. At these pH values higher surface concentrations did not involve helix formation as they probably represent multilayer adsorption where the effect of the Teflon surface on the secondary structure of LamA is less pronounced.

Irrespective of surface coverage, the adsorption of LamA on the silica nanoparticles at pH 7.0 and pH 4.4 did not significant affect the CD spectra and, hence, the secondary structure of LamA (Fig. 5.5, left panel; Table 5.1). The effect of adsorption on the secondary structure of LamA is notable at pH 3.5 probably owing to an additional destabilizing effect of the relatively strong acidic environment on the structure of LamA.

Overall, the results from the CD analysis indicate that the secondary structure of LamA is moderately affected upon adsorption on the hydrophobic Teflon while the interaction with the hydrophilic silica nanoparticles induces only slight changes, if significant at all. In both cases of LamA adsorption on Teflon and silica, the CD spectra and the data analysis showed that when adsorption exceeds monolayer coverage, the spectral characteristics resemble those of native LamA in solution. This is not surprising as the protein molecules in the second and subsequent adsorbed layers outnumber the ones in the first layer and, also, interact less intimately with the underlying surface.

Fluorescence emission spectroscopy

On excitation at 300 nm, tryptophan emission dominates the fluorescence spectra of proteins. LamA contains 11 tryptophans homogeneously distributed over the sequence and, thus, local changes in their microenvironment can be used to detect changes in the tertiary structure. At pH 7.0 LamA in solution shows maximum intensity at 335 nm, which is characteristic of partially but not completely exposed tryptophans [33]. At the IEP (pH 4.4) and at pH 3.5 the emission maxima were observed at 332 nm and the intensity decreased by 57% and 50%, respectively (Fig. 5.6). Relative to that in LamA at pH 7.0, the small blue shift of the emission maxima implies less tryptophan exposure to the polar solvent, and the decreased intensity indicates a different energy transfer mechanism possibly due to specific tryptophans structural relocation as a function of pH [33,34]. Apart from the profound effect of protonation of side groups, especially in the vicinity of or at the active site, these conformational changes provide an additional qualitative explanation of LamA's enzymatic activity drop at low pH (Fig. 5.7).

At pH 7.0 the fluorescence emission maximum of LamA adsorbed on Teflon is red-shifted, and the intensity increased as compared to that of LamA in solution, that is, 335 nm (for the same enzyme concentration). At low surface coverage the emission maximum is at 342 nm, and at higher LamA surface concentrations it is at 338 nm. The red shift of the maximum is due to increased tryptophan exposure to the polar solvent. This, however, does not mean that all of the tryptophans have increased contact with water. The fact that the peak is broadened suggests that some tryptophans are more shielded from the solvent upon adsorption (as a result of association with the surface), while other tryptophans are more exposed to the solvent. The change in the emission intensity also suggests altered tertiary structure upon adsorption, as mentioned before. The more pronounced conformational changes in the case of low surface coverage suggest a surface-induced structural relaxation effect. This means that the enzyme optimizes its contacts with the hydrophobic surface of Teflon when there is enough space available for doing so. It is likely that at low surface concentration LamA is more spread over the surface of Teflon relative to LamA adsorbed at higher surface coverage [26,30].

The fluorescence spectra of LamA adsorbed on Teflon could not be easily determined at pH 4.4 and 3.5 due to aggregation of the enzyme covered particles. The presence of zero- and positively charged LamA at the IEP and at pH 3.5, respectively, resulted in screening of the stabilizing negative charge of the Teflon particles in the suspension. The enzyme-covered particles do not sufficiently repel each other, leading to aggregation and precipitation. Apart from affecting the stability of the colloidal suspension, this also resulted in a decreased intensity of the fluorescence emission from the protein. Nevertheless, the information obtained from the wavelength of the emission unambiguously showed that upon adsorption of LamA on Teflon at pH 4.4 and 3.5 the emission maximum shifted to 339 nm and 342 nm, respectively (LamA in solution at both pH 4.4 and 3.5 shows maximum emission at 332 nm). The red shift and the peak broadening observed upon adsorption are indicative of changes in the tertiary structure of LamA. Furthermore, at pH 3.5 and at a surface concentration of 1 mg m^{-2} of LamA on Teflon the emission profile showed an unusual shoulder at about 360 nm, which clearly suggests significant changes in the tertiary structure of LamA at these adsorbing conditions.

In the case of adsorption on the silica nanoparticles, the fluorescence emission spectra of adsorbed LamA resembled those of LamA in solution when the pH of the solution was 7.0 and 4.4. This indicates that at these conditions LamA's tertiary structure did not significantly change upon adsorption. In the case of adsorption at pH 3.5 and especially at low surface coverage the changes observed in the tertiary structure of LamA were pronounced. This probably resulted from the strong destabilizing effect of the relatively low pH and the interaction between protein and surface.

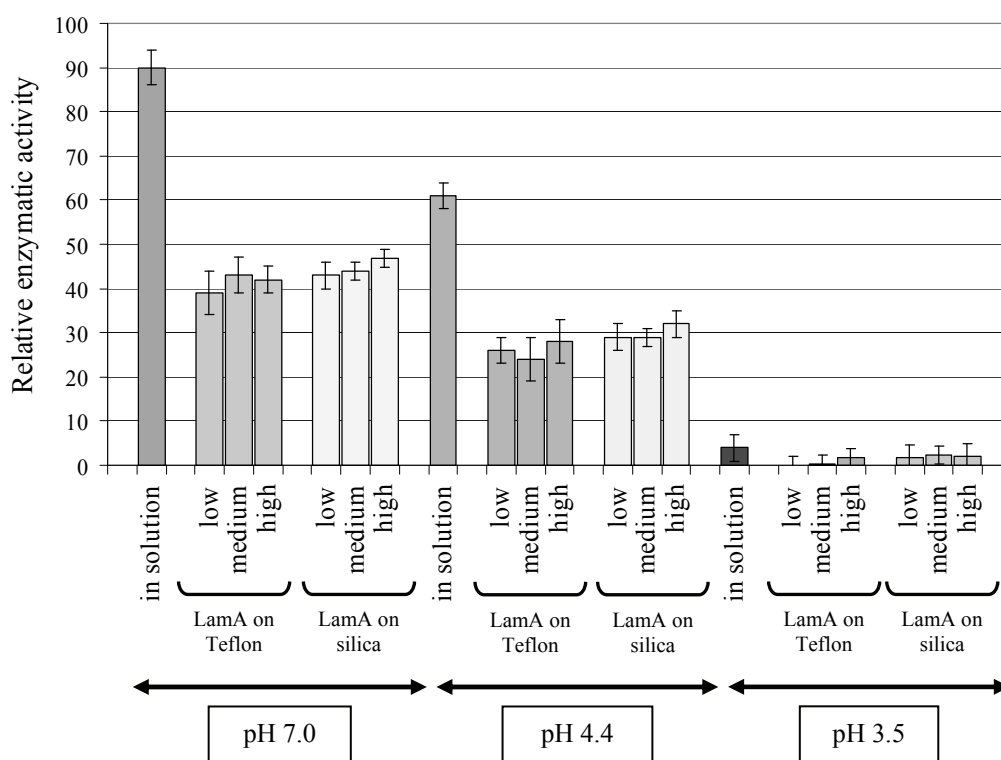


Figure 5.7 Relative specific enzymatic activity at 80°C of LamA free in solution and adsorbed on hydrophobic Teflon and hydrophilic silica particles in 0.01 M sodium phosphate buffer at pH 7.0, 4.4, and 3.5 (the maximum enzymatic activity at pH 6.5 was taken as 100%).

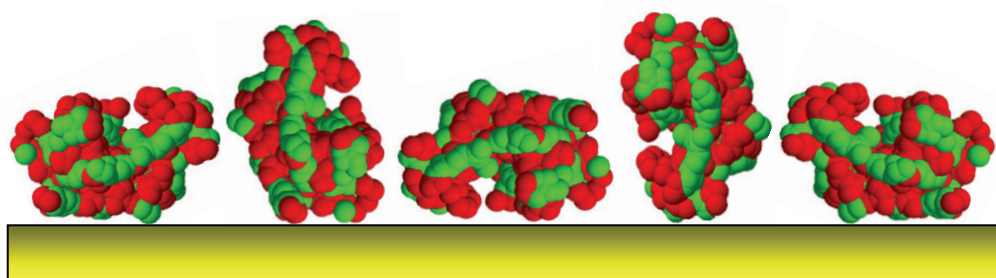


Figure 5.8 Adsorption model schematically. The structure of LamA was obtained from molecular modeling.

Enzymatic activity tests

The specific enzymatic activity of LamA in solution at 80°C is 1073 units/mg. The results presented in Fig. 5.7 resemble the activity profile reported before [6,35] and show that LamA's activity is pH-dependent. The maximum activity was observed at pH 6.0 – 6.5. At pH 7.0 the relative specific activity is 10% lower. At the IEP the activity dropped by 62% relative to the maximum, and at pH 3.5 the enzyme was practically inactive.

For the enzymatic activity tests of adsorbed LamA samples enzyme-covered particles were selected from the rising part of the adsorption isotherm and from the plateau. The observation that LamA remains biologically active in the adsorbed state primarily features the properties of “hard” proteins [25]. For all surface occupancies the specific activity of immobilized LamA was decreased about 50% relative to that of LamA in solution at the same pH (except in the case of low and medium adsorption on Teflon at pH 3.5 where the most severe structural changes were observed in the secondary and tertiary structures). In all other cases, the K_m constants of LamA in solution and in the adsorbed state were the same, and, therefore, it is safe to conclude that the decreased activity is not due to structural changes. Hence, the activity drop is attributed to random orientation of LamA in the adsorbed state where a fraction of the molecules adsorbs with their active site oriented toward the surface while others expose their cleft to the solution where it is accessible for the substrate (Fig. 5.8).

Conclusions

In this work LamA's structural characteristics and biological activity were investigated in solution and adsorbed at the solid-liquid interface of hydrophobic Teflon and dispersions of hydrophilic silica nanoparticles, at different pH values. The results of this study complete the previously published experimental data and confirmed the suggestions therein [9]. Experiments were performed at pH 7.0, 4.4 and 3.5, that is, above, at, and below the IEP of the enzyme, respectively. The adsorption isotherms were determined, and the specific enzymatic activity was measured. Using the specially designed Teflon and silica particles, we were able to monitor the enzyme's structural features in the adsorbed state by standard spectroscopic techniques. We showed that, regardless, the minor or, in the case of adsorption on the hydrophobic Teflon particles, more significant changes observed in the secondary and tertiary structure of LamA, biological activity was maintained in the adsorbed state. With an exception of LamA adsorbed on Teflon at low and medium surface coverage at very low pH, the adsorption-induced structural changes were not significant in terms of biological functioning or the protein domains involved with the structural changes were not associated with the structure of the enzymatic cleft.

This work lends further support to the previously suggested mechanism employed by nature to preserve life at extreme environmental conditions. Apart from the internal structural features, which predominantly ensure conformational stability, the extracellular hyperthermophilic LamA may use adsorption as an essential complementary mode to maintain biological functionality at temperatures higher than those the enzyme can tolerate when dissolved. The enzyme may adsorb on

the surface of hydrophobic or hydrophilic soil minerals, rocks, and sediments in the neighborhood of the microorganism's colony to cope with temperature changes without loss of activity.

Acknowledgement

We are grateful to R. Fokkink for performing the DLS experiments. Professor Ton Visser is kindly acknowledged for helpful discussions. Dupont de Nemours is acknowledged for donating the Teflon particles. This work was supported by a fellowship from the Graduate School VLAG, Wageningen (to S.K.)

References

1. Niehaus, F., Bertoldo, C., Kähler, M., and Antranikian, G. (1999) *Appl. Microbiol. Biot.* **51**, 711-729.
2. Adams, M. W. W. and Kelly, R. M. (1998) *Trends Biotechnol.* **16**, 329-332.
3. Andrade, C. M. M. C., Pereira Jr. N., and Antranikian, G. (1999) *Rev. Microbiol.* **30**, 287-298.
4. Stetter, K. O. (1999) *FEBS Lett.* **452**, 22-25.
5. Fiala, G. and Stetter, K.O. (1986) *Arch. Microbiol.* **145**, 56-61.
6. Gueguen, Y., Voorhorst, W. G. B., van der Oost, J., and de Vos, W. M. (1997) *J. Biol. Chem.* **272**, 31258-31264.
7. de Vos, W. M., Kengen, S. W. M., Voorhorst, W. G. B., and van der Oost, J. (1998) *Extremophiles* **2**, 201-205.
8. Koutsopoulos, S., van der Oost, J., and Norde, W. (2004) *Langmuir* **20**, 6401-6406.
9. Koutsopoulos, S., van der Oost, J., and Norde, W. (2005) *Biophys. J.* **88**, 467-474.
10. Linko, Y-Y. and Linko, P. (1979) *Biotechnology Lett.* **1**, 489-494.
11. Bang, M. L., Villadsen, I., and Sandal, T. (1999) *Appl. Microbiol. Biot.* **51**, 215-222.
12. Rink, R., Fennema, M., Smids, M., Dehmelt, U., and Janssen, D. B. (1997) *J. Biol. Chem.* **272**, 14650-14657.
13. Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60-89.
14. Gill, S. C. and von Hippel, P. H. (1989) *Analytical Biochemistry* **182**, 319-326.
15. Miller, G. L. (1959) *Anal. Chem.* **31**, 426-428.
16. Provencher, S. W. (1982) *Computer Physics Communication* **27**, 213-227.
17. Venyaminov, S. Yu., Baikalov, I. A., Shen, Z. M., Wu, C-S. C., and Yang, J. T. (1993) *Anal. Biochem.* **214**, 17-24.
18. Provencher, S.W. and Glöckner, J. (1981) *Biochemistry* **20**, 33-37.
19. Greenfield, N. J. (1996) *Anal. Biochem.* **235**, 1-10.
20. Combet, C., Jambon, M., Deleage, G., and Geourjon, C. (2002) *Bioinformatics* **18**, 213-214.
21. Hahn, M., Pons, J., Planas, A., Querol, E., and Heinemann, U. (1995) *FEBS Letters* **374**, 221-224.
22. Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) *Structure* **9**, 513-525..
23. Haynes, C. A. and Norde, W. (1995) *J. Coll. Interface Sci.* **169**, 313-328.
24. Norde, W. and Zoungrana, T. (1998) *Biotechnol. Appl. Biochem.* **28**, 133-143.
25. Bremer, M., Duval, J., Norde, W., and Lyyklema, J. (2004) *Colloids Surfaces A* **250**, 29-42.
26. Haynes, C. A. and Norde, W. (1994) *Colloids Surf. B: Biointerfaces* **2**, 517-566.
27. Norde, W. and Anusiem, A. C. I. (1992) *Colloids and Surfaces* **66**, 73-80.
28. Norde, W., Arai, T., and Shirahama, H. (1991) *Biofouling* **4**, 37-51.
29. Shirahama, H., Lyklema, J., and Norde, W. (1990) *J. Coll. Interf. Sci.* **139**, 177-187.
30. Norde, W. (1995) *Cells and Mater.* **5**, 97-112.

31. **Chiaraluce, R., van der Oost, J., Lebbink, J. H. G., Kaper, T., and Consalvi, V.** (2002) *Biochemistry* **41**, 14624-14632.
32. **Giacomelli CE and Norde W.** (2003) *Biomacromolecules*. **4**, 1719-1726.
33. **Lakowicz, J. R.** (1999) *Principles of Fluorescence Spectroscopy*; 2nd Ed.; Kluwer Academic/Plenum Press: New York, NY,.
34. **Eftink, M. R.** (1991) *Method Biochem. Anal.* **35**, 127-205
35. **Driskill, L. E., Bauer, M. W., and Kelly, R. M.** (1999) *Biotechnol. Bioengineer.* **66**, 51-60.

Chapter 6

Stabilization of an Endo- β -1,3-1,4-Glucanase by Cyclization

Van Lieshout, J.F.T., Pérez Gutiérrez, O.N., Vroom, W., Koutsopoulos, S., Planas, A., De Vos, W.M., Van der Oost, J.

submitted

Abstract

A novel protein-engineering approach has recently been developed on the basis of intein-driven protein splicing. A modification of this process allows for the generation of a covalent linkage between the N-terminus and the C-terminus of a polypeptide chain. This method has been applied to create circular variants of endo- β -1,3-1,4-glucanase (LicA) from *Bacillus licheniformis*. Two cyclic variants were selected for further analysis: LicA-C1 has a linking loop of 20 amino acid residues, and LicA-C2 of 14 residues. Both have a higher thermal stability than the linear variant. Upon increasing temperatures, fluorescence spectroscopy as well as differential scanning calorimetry demonstrated that the cyclic enzymes start to unfold significantly later than the linear one. Both cyclic variants show catalytic activities comparable to that of the linear variant. Moreover, whereas the linear glucanase (LicA-L1) has lost half of its activity after 3 minutes at 65 °C, the two cyclic variants have 6-fold (LicA-C1) and 16-fold (LicA-C2) increased half-life times of inactivation. The most stable enzyme is the cyclic variant with the shortest linking loop.

Introduction

The application of enzymes as biocatalysts in industry has increased considerably over the last few decades [1-3]. Desired features for optimal performance in a bioreactor include high specificity, high activity and high stability. Because natural enzymes have evolved to function optimally in the context of a living cell, it is not likely that such proteins meet all chemical, physical and biological requirements for being perfect biocatalysts in a specific biotechnological application. Limited shelf life and low intrinsic stability of enzymes often are major drawbacks for an enzyme's successful application. Both random and directed mutagenesis strategies have been used to create variants with optimized features, either by improving specificity or activity of stable thermophilic enzymes [4-6], or by increasing stability of labile mesophilic enzymes with desired catalytic features [7, 8]. Intrinsic factors that contribute to thermostability include ionic or hydrophobic interactions, improved subunit interactions, reduced surface area, improved packing, or a combination thereof [9, 10]. To stabilize a protein, these factors have to be introduced by random or directed mutagenesis approaches in the corresponding gene (reviewed by [11]).

A rather poorly explored feature is stabilization of a protein by a covalent cyclization of its backbone structure. Several peptides (or small proteins) have been identified in bacteria, plants and mammals that have a covalently-linked N- and C- terminus [12, 13]. Not only are they more stable at a broad range of chemical and physical conditions, they often also show an increased resistance to proteases [14]. Although details on the biosynthesis of most of these naturally occurring circular proteins are still unknown, several studies have proven that it is possible to engineer a covalent linkage of the N-terminal and C-terminal amino acids of a polypeptide chain, thereby generating a cyclic enzyme. Distinct cyclization strategies have been developed: (i) a chemical approach that applies solid phase synthesis combined with chemical ligation or crosslinking [15, 16], (ii) a biochemical strategy based on intein-driven self-ligation *in vitro* [17, 18], and (iii) a biological

method that is based on the ability of inteins to perform self-splicing *in vivo* (Fig. 6.1) [18-20]. Little is known about the effect of cyclization on protein stability since limited information has been reported in the few previous studies [17-20].

In the present study *in vivo* cyclization has been applied to the endo- β -1,3-1,4-glucanase (LicA) from *Bacillus licheniformis* in order to study the effect of covalently linking the N- and C-termini on catalysis and stability. Bacterial 1,3-1,4- β -glucanases (or lichenases, EC 3.2.1.73) hydrolyze β -1,4-glycosidic bonds on 3-*O*-substituted glycosyl residues in linear mixed-linked glucans, like barley β -glucan and lichenan, and have been studied extensively over the years (reviewed by Planas 2000 [21]). These enzymes are classified in family 16 of glycoside hydrolases [22] and have a monomeric jelly-roll β -sandwich structure [23-25] (Fig. 6.2). The monomeric structure of LicA, with the N- and C-termini in close proximity, makes the *in vivo* intein-based method a particularly suitable strategy for this enzyme.

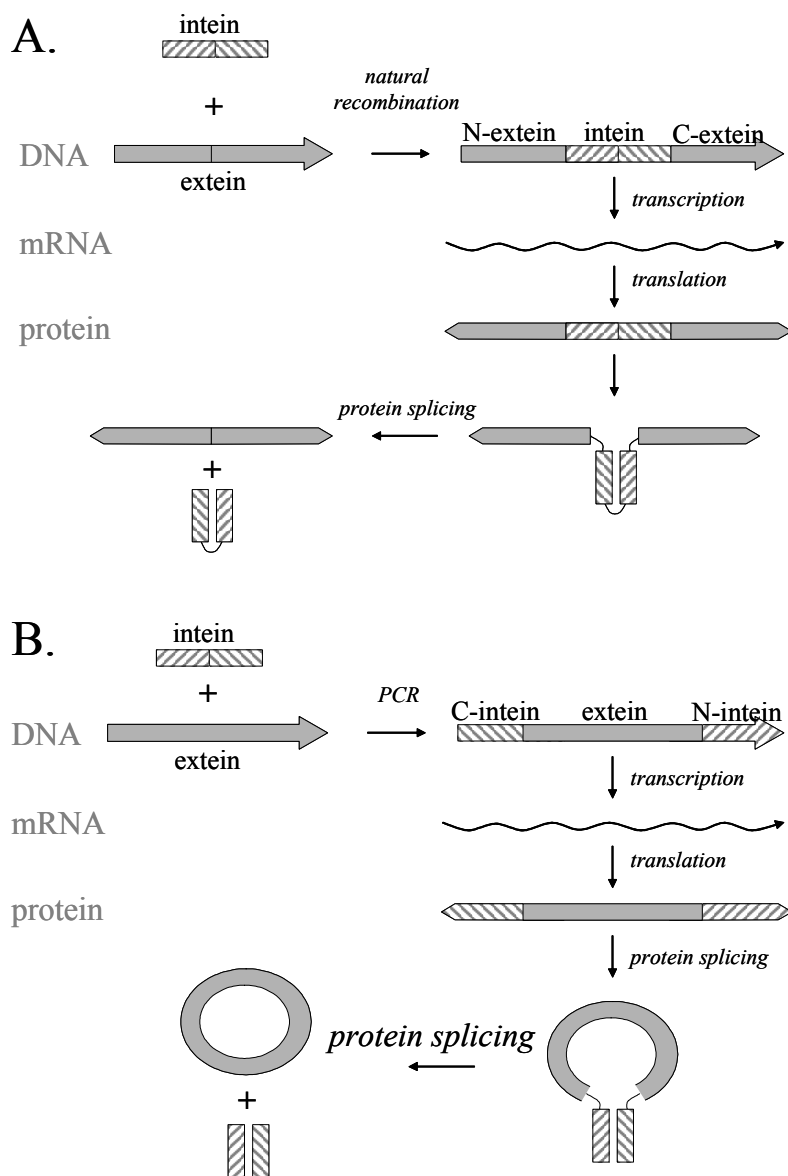


Figure 6.1. Inteins in nature and as engineering tool. **(A)** Protein splicing by a natural intein, resulting in ligation of the two extein fragments. **(B)** The engineered split intein domains do fold, interact and form an active intein complex that catalyzes splicing and ligation of the extein fragment as does the aforementioned natural system. The result is a complex of the two intein domains, and the extein domain with a circular peptide backbone.

Table 6.1. Primers used in this study

* s = sense, a = antisense

† sequences are given from 5'-3'

Primer*	primer sequence†	Description	LicA-variant
BG1260 s	GCGCGCCATGGGACATGAGTACATCTATGACAGA	intein C + <i>Nco</i> I-site	C1(a), C1-C6
BG1261 a	GTGGTGGTGGTGGTGTCCGGTGTGGACGAAAATCATTC	intein C + His-tag	C1(a), C1, C2, C6
BG1262 s	GGACACCAACACCAACCAACCAACGGGCGGGTCGTTTATGAAC	glucanase + His-tag	C1(a), C1, C2, C6
BG1263 a	CCCGGTTCCCTCGTGTACTAGTCTTTTGTGTAAACGCAACCAATG	glucanase + thrombin-site	C1(a), C1, C5
BG1264 s	CTAGTACCACGAGGAACCGGGTGCATAGACGGAAGGCCAAG	intein N + thrombin-site	C1(a), C1, C5
BG1265 a	GCGCGCTCGAGCTTAACATGTGAGTGGTATTATC	intein N + <i>Xho</i> I-site	C1(a)
BG1306 s	GCGCGCCATGGGGCAACGGGCGGGTCGTTTT	glucanase + <i>Nco</i> I-site + Δ signalsequence + SA \rightarrow MG	L1
BG1307 a	GCGCGCTCGAGTCTTTTGTGTAAACGCACCCA	glucanase + <i>Xho</i> I-site + Δ stopcodon	L1
BG1351 s	GGACACCAACCAACCAACCAACGGGTCGTTTATGAACCGTTCAAC	as BG1262 + Δ QTG	C3, C4, C5
BG1352 a	TGGCCTTTCCGTCTATGCACCCTGTGTAAACGCACCAATGTAATGAG	as BG1263 + Δ thrombin-site + Δ KR	C3, C6
BG1353 s	CTCATTACATTGGGTGCGTTACACAGGGTGCTAGACGGAAAGGCCA	as BG1264 + Δ thrombin-site + Δ KR	C3, C6
BG1354 a	GCGCGCTCGAGTTACTTAACATGTGAGTGGTATTATCAAA	as BG1265 + stopcodon	C1-C6
BG1429 a	TCTATGCACCCCTCTTTTGTGTAAACGCACCCCAATG	as BG1263 + Δ thrombin-site	C2
BG1430 s	GTTACACAAAAAGAGGGTGCATAGACGGAAAGGCC	as BG1264 + Δ thrombin-site	C2
BG1431 a	CCCGGTTCCCTCGTGTACTAGTGTGTAAACGCACCCCAATGTAATG	as BG1263 + Δ KR	C4
BG1432 s	CTAGTACCACGAGGAACCGGGTGCATAGACGGAAAGGCC	as BG1264 + Δ KR	C4
Intein-f a	CGAGCCGAGGACGTTCTACGATC	forward sequence primer, annealing to intein	-
Intein-r s	GCTTGTATCTCTCGTACATCTCCTC	reverse sequence primer, annealing to intein	-

Materials and methods

Bacterial hosts and vectors - The T7 expression vector pET24d was obtained from Novagen. *Escherichia coli* XL-1 Blue (Stratagene) was used as an initial host for cloning, while either one of the *E. coli* strains BL21(DE3) or JM109 (DE3) (Stratagene) was used as expression host for the pET-derivatives. *E. coli* was grown in TY medium [26] in a rotary shaker at 37°C. Kanamycin was added to a final concentration of 30 μ g/ml.

Cloning and expression - The gene coding for *Bacillus licheniformis* 1,3-1,4- β -glucanase previously cloned in the pUC119-derived pD6-2 [27] was used as a template for PCR amplification. For expression of the linear enzyme without the signal sequence the gene was subcloned in pET24d using primers BG1306 and BG1307 (Table 6.1), introducing a C-terminal His-tag, resulting in pWUR146.

P. furiosus genomic DNA was isolated as described previously [26] and used as template in PCR reactions to amplify the intein PI-*PfuI* [20, 28]. In a first series of PCR reactions the two parts of the intein and the glucanase gene were amplified separately (Fig. 6.3A). In PCR-2, an overlap extension PCR [29], the three overlapping fragments were fused, and the full-length hybrid molecule was amplified using the flanking primers BG1260 and BG1265/1354 (2.1 kb; Fig. 6.3A). The 2.1 kb PCR product was then ligated into pET24d at the *NcoI* and *XhoI* restriction sites, and transformed to *E. coli* XL1-Blue. Sequence analysis of this construct (and of all variant constructs described below, generated with a similar approach using variant primers (Table 6.1)) was done by the dideoxynucleotide chain termination method with a Li-Cor automatic sequencing system (model 4000L). Two sets of sense/antisense sequence primers were used that anneal either to the promoter and terminator sequence of the pET24d vector, or to the sequences of the two parts of the intein flanking the glucanase gene (Intein-f and Intein-r in Table 6.1).

Overexpression of the gene and purification of recombinant protein – Both the *licA-L1* gene and the permuted genes were expressed in freshly transformed BL21(DE3) or JM109 (DE3) cells. A 5 ml overnight culture was used to inoculate 500 ml of TY medium containing 30 μ g/ml kanamycin. When the OD₆₀₀ reached 0.5, the cells were induced with 100 μ M IPTG and subsequently grown at 37 °C for another 4 h. Cells were harvested by centrifugation (10 min, 4000 x g) and resuspended in 50 mM sodium phosphate buffer (pH 7.7) containing 300 mM NaCl and 10 mM imidazole. After addition of lysozyme to 1 mg/ml, the cells were incubated on ice for 30 min before sonication (six times 15 sec). After removal of the cell debris by centrifugation (30 min, 10,000 x g) the glucanase variants containing a Histidine hexa-peptide (His-tag) were further purified from the supernatant using Ni-NTA spin columns according to the provided protocol for native conditions (Qiagen). The active fractions were collected and purified to homogeneity on a Superdex200 column (Amersham Biosciences). Depending on subsequent analysis of the enzyme, elution was performed with either a 50 mM sodium phosphate buffer (pH 7.7) or a 20 mM PIPES buffer (pH 7.0).

Enzymatic assays - Standard enzymatic assays were performed at 55°C in 30 mM sodium phosphate buffer (pH 7.0) containing 0.1 mM of CaCl₂ with barley β -glucan (final concentration 0.4 %) as a substrate. The reducing sugars were detected by the dinitrosalicylic acid (DNS) method, with glucose as standard [30, 31]. One unit is defined as the amount of enzyme required to release 1 μ mol of reducing sugars per min. Temperature induced inactivation was determined in 50 mM sodium phosphate (pH 7.7) and 1 mM CaCl₂ by heating the purified enzyme (30 μ g/ml) in small crimp-sealed vials, submerged in a water bath. During a time series (0-20 h), 20- μ l aliquots were tested for remaining activity as described above. All activities were corrected for spontaneous hydrolysis in the absence of enzyme.

Thrombin digestion - To linearize the circular proteins containing a thrombin recognition site, incubation with thrombin (Sigma) was performed overnight at 22 °C. A ratio of 1 unit of thrombin was used per 100 μ g of protein.

Fluorescence emission spectroscopy - For fluorescence experiments proteins were purified with 20 mM PIPES buffer (pH 7.0) as the eluent in the final step on the Superdex200 column. In 10 mm Quartz SUPRASIL precision cells (Hellma), 20 mM PIPES (pH 7.0) was mixed with the protein solution to a final concentration of 15 μ g/ml with a final volume of 3.0 ml. When indicated, 1.0 mM CaCl₂ or 1.0 mM EDTA was added. All solutions were degassed prior to use. The fluorescence emission was measured in the temperature range 30 – 95°C, and with a scan rate of 0.5 °C/min by a Varian Cary Eclipse spectrophotometer. Emission spectra were recorded in the range 300 – 400 nm upon excitation of the tryptophans at 295 nm, with the excitation and emission slit widths set at 10 nm, and the photomultiplier at 610 V. All spectra were corrected for the background emission peak of water. The emission spectra of samples containing 1 mM of either CaCl₂ or EDTA were corrected using buffer baselines acquired at the same conditions.

Differential Scanning Calorimetry – Temperature-controlled calorimetric studies of 0.3 mg/ml LicA-C1 were carried out in a VP-DSC calorimeter (MicroCal Inc., Northampton, MA) using as reference the respective buffer solution. All samples were degassed for 15 min prior to loading the cells and the enzyme solution was kept under 1.5 bar pressure to avoid boiling of the sample. The temperature increased with a heat rate of 0.5 °C/min.

Results and Discussion

Protein stability can be described as the Gibbs free-energy of folding (ΔG), i.e. the difference in free energy between the protein's denatured and native state. The actual net free energy difference is the sum of a large number of stabilizing and destabilizing interactions. This difference is usually very small, typically 5-17 kcal/mol, which is comparable to the energy of only a few hydrophobic interactions, ion pairs, or hydrogen bonds (reviewed by [11]). According to Gibbs' equation ($\Delta G =$

$\Delta H - T\Delta S$), protein stability (ΔG) can be increased either by increasing the enthalpy change (ΔH) or by decreasing the entropy difference (ΔS). Several examples have been described of “entropic stabilization” (rigidification), i.e. mutations that result in increased ΔG because of a decreased ΔS (lowering the entropy of the denatured state). Such engineering of protein stability includes the engineering of a more rigid protein backbone by specific substitutions (e.g. Gly > Ala in an α -helix, or Xxx > Pro in a β -turn), or by the introduction of disulfide bridges [11]. An alternative approach to reach entropic stabilization is described in the present study: the covalent linkage of the polypeptide’s N- and C-termini. Apart from enhanced rigidity, such a cyclic polypeptide would also be resistant towards exo-proteases.

The goal of the present study was to introduce a covalent link between N- and C-terminus of an active enzyme, and subsequently analyze potential stabilization of the generated circular backbone structure. The three-dimensional structure of the *Bacillus licheniformis* 1,3-1,4- β -glucanase (LicA) [24] indicates that this approach could be feasible since the termini of this enzyme are in close proximity, as they reside on adjacent anti-parallel beta strands (Fig. 6.2). Hence, this enzyme was selected for further engineering experiments.

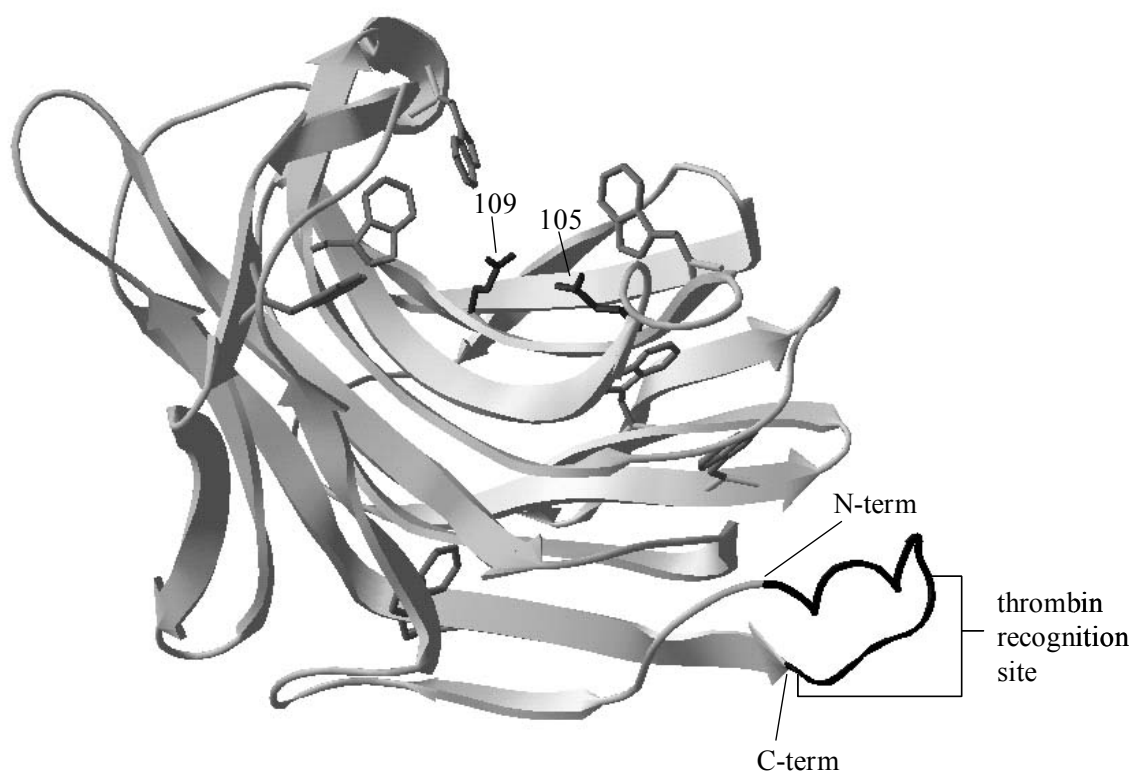
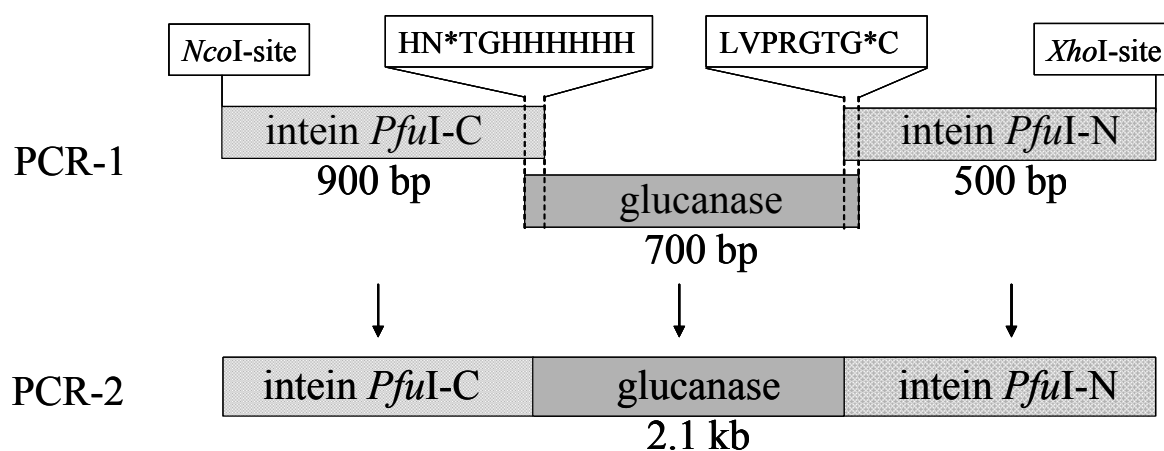


Figure 6.2. Circular LicA structural model. The model shown is one possible structure of construct LicA-C1. 15 extra residues were added to the x-ray structure (Protein Data Bank code 1GBG) and manually turned to form a closed peptide backbone with the program Swiss-PdbViewer. The linking loop is shown as a black coil. The original N- and C-terminal locations (Gln-1 and Arg-214 respectively) are indicated, as well as the thrombin recognition site. The two catalytic residues (Glu-105 and Glu-109; numbering according to 3D-structure in Protein Data Bank) are displayed in black. Tryptophan residues are depicted in dark grey.

A. PCR design of circular LicA



B. Linear LicA-construct

	[C-term]	[N-term]	length loop	activity (U/mg)
L1	: <u>...WVRYTKRLEHHHHHH</u>	MGQTGGSFYE...	-	2.7

C. Circular LicA-constructs

	["C-term"] * ["N-term"]	length loop	activity (U/mg)
C1	: <u>...WVRYTKRLVPRGTG*TGHHHHHHQTGGSFYE...</u>	20	4.9
C1a	: <u>...WVRYTKRLVPRGTG*TGHHHHHHQTGGSFYE...</u>	20	4.7
C2	: <u>...WVRYTKR-----G*TGHHHHHHQTGGSFYE...</u>	14	7.7
C3	: <u>...WVRYT-----G*TGHHHHHH---GSFYE...</u>	9	0.0
C4	: <u>...WVRYT--LVPRGTG*TGHHHHHH---GSFYE...</u>	15	0.0
C5	: <u>...WVRYTKRLVPRGTG*TGHHHHHH---GSFYE...</u>	17	7.9
C6	: <u>...WVRYT-----G*TGHHHHHHQTGGSFYE...</u>	12	0.0

linking loop

Figure 6.3. Circular constructs made in this study. (A) Schematic representation of the PCR-based engineering of the constructs used for the intein-based circularization of LicA; for details see text; *NcoI* and *XhoI* are the introduced restriction sites. The amino acid sequence of the overlap of the PCR-1 fragments is shown. * indicates the connection point of the sequence of the intein and the glucanase gene; (B/C) Amino acid sequences of the N- and C-terminal regions of (B) linear and (C) circular variants of LicA. The extein sequences, corresponding to the wild-type sequence, are underlined; note that in all constructs the signal sequence has been deleted, resulting in intracellular production of the corresponding proteins. The length of the various loops is indicated, as well as the activity of the purified proteins (U/mg). LicA-C1 and LicA-C1a are identical, but the precursor protein of LicA-C1a has a His-tag at the N-terminal part of the intein, in contrast to that of LicA-C1 (see text). * indicates the connection point of the N- and C-terminal sequence.

To produce a circular enzyme we used a self-splicing intein to ligate protein backbones. The intein PI-*PfuI* from *Pyrococcus furiosus* has recently been demonstrated to perform this cyclization in *E. coli in vivo* [20], and has therefore been selected for the present study. In a two-step PCR reaction we connected the C-terminal part of the intein (residues 161-454) to the N-terminus of the 1,3-1,4- β -glucanase, and the N-terminal part (residues 1-160) to the C-terminus of the 1,3-1,4- β -glucanase (Fig. 6.3A). After cloning in pET24d, this resulted in a chimeric gene of 2058 basepairs coding for a precursor protein with a total length of 686 amino acids (79.4 kDa). At the same time, a His-tag was introduced at the N-terminus of the glucanase to facilitate purification. Furthermore, the connecting loop contained a thrombin recognition site (LVPRGT) to enable subsequent linearization by thrombin cleavage. After processing, the cyclic protein LicA-C1 has a length of 229 amino acids (Fig. 6.3C), whereas the linear LicA-L1 consists of 224 amino acids (Fig. 6.3B).

The glucanase gene, without the signal sequence but with His-tag (Fig. 6.3B), was expressed in *E. coli* BL21(DE3) cytoplasm as a soluble protein (LicA-L1). After purification LicA-L1 has a specific activity (2.7 U/mg), comparable to that of the wild-type LicA, expressed with its signal sequence but without His-tag, the excreted mature product of which was purified from the *E. coli* medium (A. Planas, personal communication). This indicates that both the loss of the signal peptide (29 amino acids), and the addition of six C-terminal histidines (Fig. 6.3B), does not have a significant effect on the specific activity of the 1,3-1,4- β -glucanase.

Next, cyclic enzymes LicA-C1 and LicA-C1a (Fig. 6.4) were expressed as soluble proteins in *E. coli* cells. Construct LicA-C1 generates, upon *in vivo* processing, a circular glucanase with the His-tag and the thrombin cleavage site in the linking loop that connects the N- and C-termini (Fig. 6.3C). On the other hand, LicA-C1a is a control construct that will produce the same cyclic β -glucanase, but has a second His-tag at the N-terminal part of the intein. Cell-free extracts of *E. coli* BL21(DE3) cells overproducing either LicA-C1 or LicA-C1a showed three prominent additional bands on a SDS-gel, with sizes of 35, 27 and 19 kDa (Fig. 6.4). After application of the LicA-C1a cell-free extract on a Ni-NTA column, the proteins corresponding to these three bands were separated from the *E. coli* proteins (Fig. 6.4). The sizes corresponded to the expected molecular weights of the glucanase and of the C- and N-terminal parts of the intein. This indicates that the splicing reaction has taken place in the cells, leading to three separate polypeptide-fragments. Moreover, the co-purification on the Ni-NTA column shows that the C-terminal part of the intein, lacking a histidine tag, is associated with its N-terminal part, which is necessary for the splicing reaction to take place. A similar phenomenon was observed by Iwai et al. [20], however, they also found the presence of an unprocessed 80 kDa product. The absence of such an unprocessed form in our system indicates the efficiency of the present approach.

In the case of construct LicA-C1, the intein fragments were not affinity-purified on the Ni-NTA column (Fig. 6.4) because the N-terminal intein fragment does not contain a His-tag. In all other variant constructs (see below), this His-tag at the N-terminal intein was omitted as well, resulting in the presence of only a single His-tag, which is located in the linking loop of the circular glucanase, like in LicA-C1, allowing for a convenient 1-step purification of the circular enzymes. Thrombin was added to the purified spliced LicA-C1 protein and incubated overnight. SDS-PAGE analysis (Fig. 6.4) showed a different migration behavior between the thrombin diges-

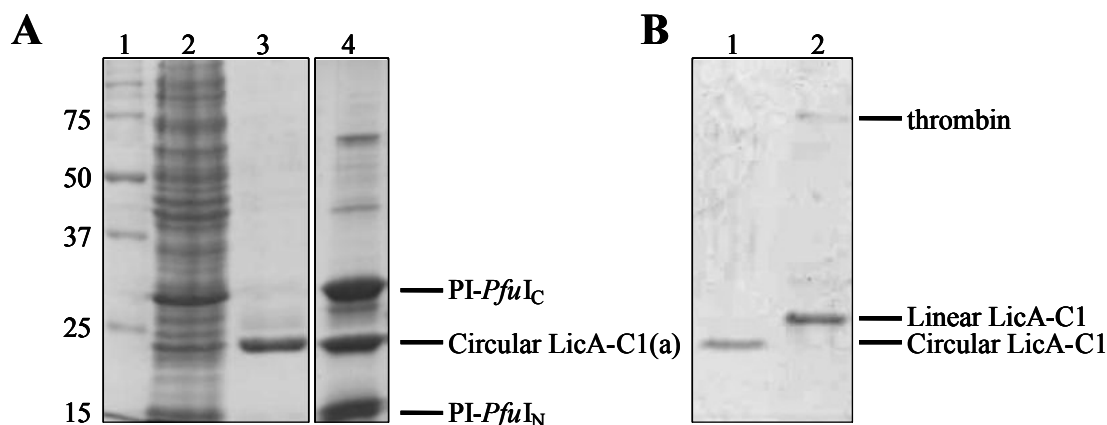


Figure 6.4. SDS-PAGE analysis of expressed and purified circular and linearized LicA-C1. **(A)** Lane 1, molecular weight marker, with corresponding sizes shown on the left in kDa. Lane 2, cell free extract of *E. coli* BL21(DE3) expressing construct 1. Lane 3, circular LicA-C1 after purification on a NI-NTA column. Lane 4, circular LicA-C1a after purification on a NI-NTA column. PI-PfuI_C and PI-PfuI_N denote the C- and N-terminal part respectively of the intein, that are visible in lane 2 and 4. **(B)** LicA-C1 after purification on Superdex200, before (Lane 1) and after (Lane 2), treatment with thrombin.

ted sample and the untreated protein. The non-digested LicA-C1 protein migrates faster (Fig. 6.4), as would have been expected for a circular form because of its slightly different denatured conformation. This phenomenon has been reported before with circular and linearized polypeptides [17, 19, 20].

The cell-free extracts containing the circular protein LicA-C1 showed hydrolytic activity on barley- β -glucan, reflecting its proper folding. After purification, the specific activity of the circular enzyme is comparable to that of the linear one, apparently slightly more active (Fig. 6.3). Fluorescence spectroscopy after excitation at 295 nm demonstrated a similar emission spectrum of the linear LicA-L1 and the circular LicA-C1 (not shown), also suggesting correct folding of the protein. The optimal temperature for hydrolytic activity for the spliced protein was determined to be 56 °C, which is in the same range as the 55°C optimum of the wild-type enzyme [32].

From these results we conclude that endo-1,3-1,4- β -glucanase was expressed, excised and ligated successfully to acquire its designed circular backbone structure. Moreover, the comparable activities of the circular enzyme variant LicA-C1 and the linear LicA-L1 indicate that the introduced junction and loop do not result in inactivation of the enzyme, indicating that the enzyme's functionality has not been affected by the generated covalent link. Earlier studies showed that circular permutations in the compact jellyroll domain of the *Bacillus* glucanase is tolerated without severe change of enzymatic activity or tertiary structure [33, 34]. In the latter studies, it was concluded that the novel peptide bond linking the original N- and C-termini, without adding extra residues, apparently did not introduce strain into the molecule. However, in those studies new N- and C- termini were introduced in another loop of the glucanase structure, and as such there was still an open chain (linear) structure, which might enable release of strain. In our study, a covalently-closed circular polypeptide chain is generated in which a high specific activity indicates that no severe strain is introduced.

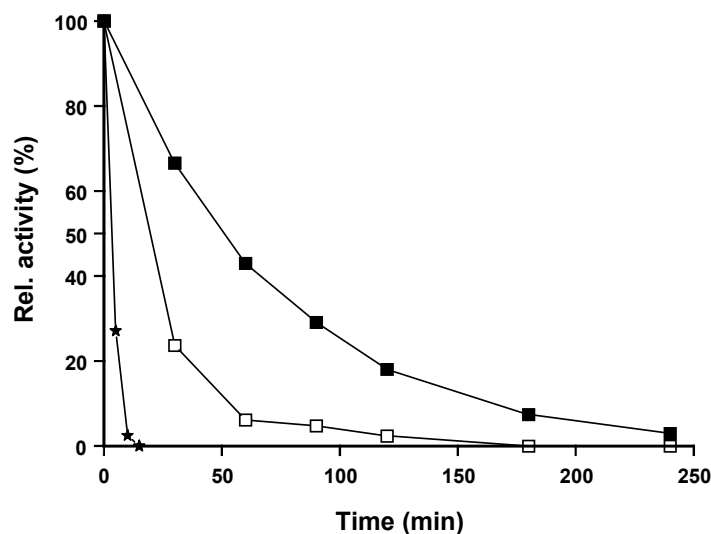


Figure 6.5. Temperature induced inactivation. Residual activity of linear LicA-L1 (asterisks), circular LicA-C1 (open squares) and circular LicA-C2 (closed squares), when incubated at 65 °C at a concentration of 30 μ g/ml in a 50 mM sodium phosphate buffer (pH 7.7) containing 1 mM CaCl_2 . Activity of each sample is expressed relative to the sample measured at $t=0$, taken as 100%.

The effect of the ligated backbone on thermostability was investigated by comparing the thermal inactivation of LicA-C1 with that of LicA-L1. Incubation at 65 °C caused the linear enzyme to inactivate very rapidly ($t_{1/2} = 3$ min; Fig. 6.5). Complete inactivation occurred within 15 min. The cyclic LicA-C1, however, still retained half of its initial activity after 20 min (Fig. 6.5), a 6-fold increase as compared to the linear LicA-L1.

Temperature-induced unfolding was monitored by fluorescence spectroscopy to determine the transition midpoint of the linear enzyme in comparison with its cyclic derivative, as an alternative means to study the effect of cyclization on protein stability. A titration of a protein solution with a chemical or physical stress factor (e.g. temperature gradient) generally results in a transient rearrangement of the polypeptide structure, and consequently in a transient decrease of the tryptophan fluorescence emission. The LicA enzyme contains 8 tryptophan residues that are randomly distributed over the protein backbone (Fig. 6.2), which were used as intrinsic fluorophores for fluorescence emission studies upon excitation at 295 nm. The fluorescence intensities of the linear and the circular LicA were monitored at 350 nm as a function of temperature (Fig. 6.6). The intensities measured at different wavelengths (360 and 375 nm) gave similar transition temperatures (not shown).

First the effect of calcium was examined. While addition of EDTA caused the linear enzyme to melt at 61.8 °C, the presence of calcium resulted in a melting temperature of 64.5 °C. This confirmed the stabilizing effect of calcium for endo-1,3-1,4- β -glucanase as reported before [35, 36], and calcium was added in all further experiments. In the presence of calcium, the circular LicA-C1 showed a melting temperature of 67.7, an increase of 3 °C as compared to the linear LicA-L1 (Fig. 6.6); a destabilizing effect of EDTA as described for the linear enzyme, was also observed with the circular variants (results not shown). The apparent melting temperatures obtained for linear and circular enzymes match very well with the trend observed in the above mentioned in-

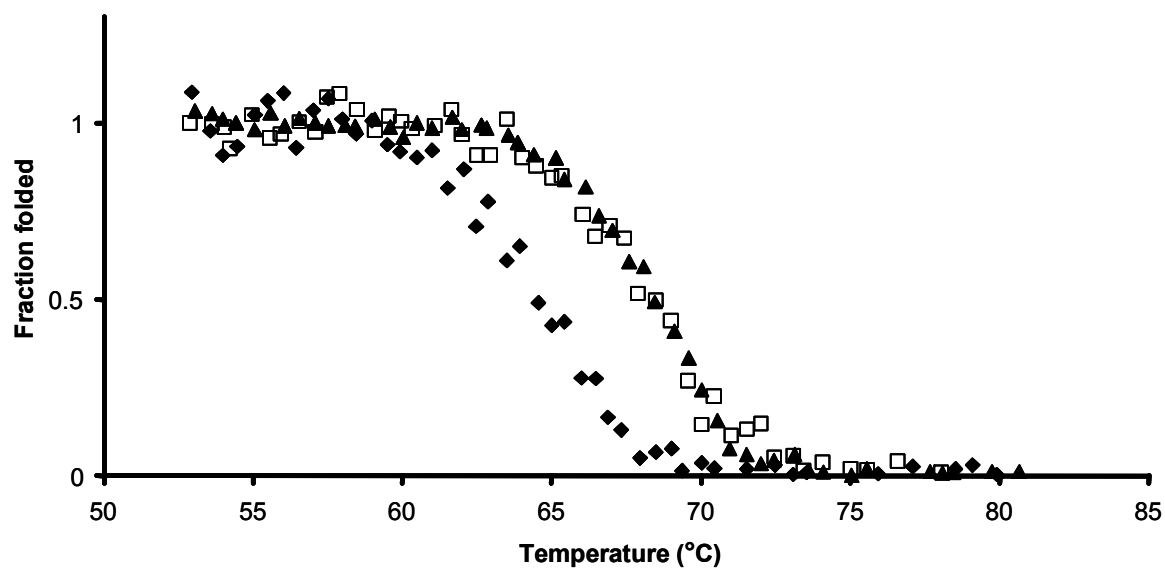


Figure 6.6. Temperature induced unfolding. Fraction of folded linear LicA-L1 (closed clubs), circular LicA-C1(open squares) and circular LicA-C2 (closed triangles), upon increasing temperature. Fraction was calculated by normalizing the emission data at 350 nm obtained by thermal denaturation fluorescence spectroscopy.

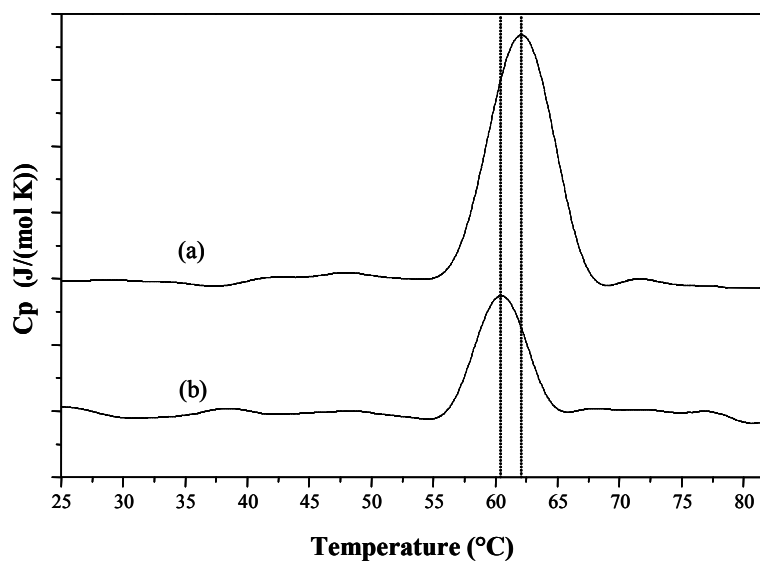


Figure 6.7. Temperature induced unfolding. Differential scanning thermograms of (A) LicA-C1 and (B) linearized LicA-C1 in the absence of Ca^{2+} at pH 7.00 at a protein concentration of 0.3 mg/ml. Graphs were vertically shifted for comparison. Vertical dotted lines emphasize the shift of the denaturation temperature.

activation experiment at 65 °C (Fig. 6.5). While the linear enzyme is beyond its melting temperature at this condition and thus starts to inactivate rapidly, the circular variant is only at the early stage of unfolding, resulting in a much slower inactivation process.

DSC studies of circular and linearized (circular variant after treatment with thrombin) LicA-C1 in 50 mM sodium phosphate buffer (pH 7.7), revealed a single denaturation transition peak. The denaturation temperature, T_d , was found to be 60.0 °C and 61.7 °C for the linearized and the circular LicA-C1, respectively (Fig. 6.7). These denaturation temperatures are in good agreement with the transition temperatures observed from monitoring the fluorescence intensity under the same conditions. Varying the scanning rate between 0.25 and 1.5 °C/min did not affect the T_d or the shape of the endothermal peak and the enthalpy associated with the transition (not shown). However, after cooling the samples to room temperature subsequent heating up did not reveal a denaturation peak; this indicates irreversible unfolding that prevents the calculation of changes in ΔS and ΔH . However, from the area of the transition peak, the calorimetrically-determined enthalpy associated with the denaturation of the circular enzyme was estimated to be two-fold higher than that calculated for the denaturation of the linearized enzyme (not shown). The enthalpy of denaturation is a measure of the intramolecular bonds, which maintain the active protein conformation. This strongly suggests that the backbone cyclization increases the internal stability of LicA-C1.

To investigate the effect of the introduced loop, a series of constructs with different lengths and composition were designed. Shorter loops were constructed either by deletion of three N-terminal residues (QTG), two C-terminal residues (KR), the thrombin recognition site (LVPRGT), or a combination thereof, resulting in linking loops consisting of 9 - 20 amino acid residues (Fig. 6.3C). The six histidine residues were always maintained to facilitate purification. Based on SDS-PAGE analysis of Ni-NTA purified CFEs of the complete series of clones, it is concluded that not all variants were produced as soluble proteins; the cases where no protein band was detectable corresponded to the ones in which no activity was measured (LicA-C3, LicA-C4 and LicA-C6) (Fig. 6.3C). The constructs for which proteins of the expected size were detected (not shown), were purified and activity was analyzed: LicA-C1, LicA-C2 and LicA-C5. The two variants with the shortest loop (LicA-C3 and LicA-C6) are inactive, whereas the ones with a longer loop (LicA-C1 and LicA-C5) show hydrolytic activity (Fig. 6.3C). On the other hand, the medium-sized loops result in both an active (LicA-C2) and an inactive (LicA-C4) cyclic variant, at least indicating that length is not the only factor affecting catalytic function and/or correct processing. Moreover, it can not be ruled out that the C-terminal residues (RK) are important; a mutant with only this deletion has not been included. Removal of the N-terminal QTG or the thrombin recognition site has hardly any effect. Overall, it is concluded that both the size of the loop and the nature of amino acid residues in the region of the original C-terminus, determines the efficiency of intein processing, which is required to generate active circular endo-1,3-1,4- β -glucanase.

The variant with the shortest loop that was still active, LicA-C2, was compared with the long loop variant, LicA-C1, with respect to temperature optimum and stability. Like variant LicA-C1, also LicA-C2 had an optimal temperature of 55 °C, identical to the linear LicA-L1 (not shown). Both cyclic variants are more stable than the linear enzyme. Fluorescence spectroscopy showed that

variant LicA-C2 unfolds in the same temperature range as variant LicA-C1 (Melting temperature of 68.2 °C, Fig. 6.6). As discussed earlier, LicA-C1 inactivated 6-fold slower than LicA-L1 upon incubation at 65 °C, yet LicA-C2 showed an inactivation half-life time of 50 min at this temperature: more than a 16-fold increase as compared to the linear enzyme (Fig. 6.5). The difference in stability between the two cyclic variants is significant. LicA-C2 differs from LicA-C1 by the deletion of 6 amino acid residues, the thrombin recognition site, in the loop (Fig. 6.3C). The shorter loop of LicA-C2 might result in a more stable protein because it is more compact, whereas the longer loop of LicA-C1 might be more flexible, what makes the protein more susceptible to unfolding at increasing temperatures. Furthermore, solving the three-dimensional structures of LicA-C1 and LicA-C2 would be required to reveal interactions between residues in the loop and other residues that might contribute to stability of the cyclic enzymes.

In summary, the endo-1,3-1,4- β -glucanase from *Bacillus licheniformis* has successfully been circularized by using the cyclization approach based on circular permutation of a precursor protein flanked by two intein domains. The circular variants showed catalytic properties comparable to the original linear enzyme, but were significantly more stable. Although the design of the connecting loop will differ per protein, we conclude that cyclization may be an effective tool to moderately stabilize polypeptides.

Acknowledgments

This research was supported by the Technology Foundation (STW), applied science division of NWO and the technology programme of the ministry of Economic Affairs. O. P. received a fellowship of the VLAG graduate school (Wageningen University).

References

1. van Beilen, J. B. & Li, Z. (2002) Enzyme technology: an overview, *Curr. Opin. Biotechnol.* **13**, 338-344.
2. Kirk, O., Borchert, T. V. & Fuglsang, C. C. (2002) Industrial enzyme applications, *Curr. Opin. Biotechnol.* **13**, 345-351.
3. Schmid, A., Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M. & Witholt, B. (2001) Industrial biocatalysis today and tomorrow, *Nature*. **409**, 258-268.
4. Lebbink, J. H., Kaper, T., Bron, P., van der Oost, J. & de Vos, W. M. (2000) Improving low-temperature catalysis in the hyperthermostable *Pyrococcus furiosus* beta-glucosidase CelB by directed evolution, *Biochemistry*. **39**, 3656-3665.
5. Kaper, T., Lebbink, J. H., Pouwels, J., Kopp, J., Schulz, G. E., van der Oost, J. & de Vos, W. M. (2000) Comparative structural analysis and substrate specificity engineering of the hyperthermostable beta-glucosidase CelB from *Pyrococcus furiosus*, *Biochemistry*. **39**, 4963-4970.
6. Kaper, T., Brouns, S. J., Geerling, A. C., De Vos, W. M. & Van der Oost, J. (2002) DNA family shuffling of hyperthermostable beta-glycosidases, *Biochem. J.* **368**, 461-470.
7. Machius, M., Declerck, N., Huber, R. & Wiegand, G. (2003) Kinetic stabilization of *Bacillus licheniformis* alpha-amylase through introduction of hydrophobic residues at the surface, *J. Biol. Chem.* **278**, 11546-11553.

8. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G. & Eijlsink, V. G. (1998) Engineering an enzyme to resist boiling, *Proc. Natl. Acad. Sci. U S A.* **95**, 2056-2060.
9. Yip, K. S., Britton, K. L., Stillman, T. J., Lebbink, J., de Vos, W. M., Robb, F. T., Vetriani, C., Maeder, D. & Rice, D. W. (1998) Insights into the molecular basis of thermal stability from the analysis of ion-pair networks in the glutamate dehydrogenase family, *Eur. J. Biochem.* **255**, 336-346.
10. van den Burg, B. & Eijlsink, V. G. (2002) Selection of mutations for increased protein stability, *Curr. Opin. Biotechnol.* **13**, 333-337.
11. Eijlsink, V. G., Bjork, A., Gaseidnes, S., Sirevag, R., Synstad, B., van den Burg, B. & Vriend, G. (2004) Rational engineering of enzyme stability, *J. Biotechnol.* **113**, 105-120.
12. Craik, D. J., Daly, N. L., Saska, I., Trabi, M. & Rosengren, K. J. (2003) Structures of naturally occurring circular proteins from bacteria, *J. Bacteriol.* **185**, 4011-4021.
13. Trabi, M. & Craik, D. J. (2002) Circular proteins--no end in sight, *Trends Biochem. Sci.* **27**, 132-138.
14. Craik, D. J., Cemazar, M., Wang, C. K. & Daly, N. L. (2006) The cyclotide family of circular miniproteins: nature's combinatorial peptide template, *Biopolymers.* **84**, 250-66.
15. Camarero, J. A., Fushman, D., Sato, S., Giriat, I., Cowburn, D., Raleigh, D. P. & Muir, T. W. (2001) Rescuing a destabilized protein fold through backbone cyclization, *J. Mol. Biol.* **308**, 1045-1062.
16. Goldenberg, D. P. & Creighton, T. E. (1983) Circular and circularly permuted forms of bovine pancreatic trypsin inhibitor, *J. Mol. Biol.* **165**, 407-413.
17. Iwai, H. & Pluckthun, A. (1999) Circular beta-lactamase: stability enhancement by cyclizing the backbone, *FEBS Lett.* **459**, 166-172.
18. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q. & Xu, M. Q. (2000) Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803, *J. Biol. Chem.* **275**, 9091-904.
19. Scott, C. P., Abel-Santos, E., Wall, M., Wahnnon, D. C. & Benkovic, S. J. (1999) Production of cyclic peptides and proteins in vivo, *Proc. Natl. Acad. Sci. U S A.* **96**, 13638-13643.
20. Iwai, H., Lingel, A. & Pluckthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfufI intein from *Pyrococcus furiosus*, *J. Biol. Chem.* **276**, 16548-16554.
21. Planas, A. (2000) Bacterial 1,3-1,4-beta-glucanases: structure, function and protein engineering, *Biochim. Biophys. Acta.* **1543**, 361-382.
22. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities, *Biochem. J.* **280**, 309-316.
23. Hahn, M., Olsen, O., Politz, O., Borriss, R. & Heinemann, U. (1995) Crystal structure and site-directed mutagenesis of *Bacillus macerans* endo-1,3-1,4-beta-glucanase, *J. Biol. Chem.* **270**, 3081-308.
24. Hahn, M., Pons, J., Planas, A., Querol, E. & Heinemann, U. (1995) Crystal structure of *Bacillus licheniformis* 1,3-1,4-beta-D-glucan 4-glucanohydrolase at 1.8 Å resolution, *FEBS Lett.* **374**, 221-224.
25. Keitel, T., Simon, O., Borriss, R. & Heinemann, U. (1993) Molecular and active-site structure of a *Bacillus* 1,3-1,4-beta-glucanase, *Proc. Natl. Acad. Sci. U S A.* **90**, 5287-5291.
26. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular cloning: A laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
27. Planas, A., Juncosa, M., Lloberas, J. & Querol, E. (1992) Essential catalytic role of Glu134 in endo-beta-1,3-1,4-D-glucan 4- glucanohydrolase from *B. licheniformis* as determined by site-directed mutagenesis, *FEBS Lett.* **308**, 141-145.
28. Ichianagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfufI, *J. Mol. Biol.* **300**, 889-901.
29. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene.* **77**, 51-59.
30. Sumner, J. B. & Somers, G. F. (1949) Dinitrosalicylic method for glucose in *Laboratory experiments in biological chemistry* (Sumner, J. B. & Somers, G. F., eds) pp. 38-39, Academic Press, New York.

31. **Gueguen, Y., Voorhorst, W. G., van der Oost, J. & de Vos, W. M.** (1997) Molecular and biochemical characterization of an endo-beta-1,3- glucanase of the hyperthermophilic archaeon *Pyrococcus furiosus*, *J. Biol. Chem.* **272**, 31258-31264.
32. **Lloberas, J., Querol, E. & Bernues, J.** (1988) Purification and Characterization of Endo-Beta-1 3-1 4-D Glucanase Activity from *Bacillus-Licheniformis*, *Applied Microbiology and Biotechnology.* **29**, 32-38.
33. **Hahn, M., Piotukh, K., Borriss, R. & Heinemann, U.** (1994) Native-like in vivo folding of a circularly permuted jellyroll protein shown by crystal structure analysis, *Proc. Natl. Acad. Sci. U S A.* **91**, 10417-10421.
34. **Ay, J., Hahn, M., Decanniere, K., Piotukh, K., Borriss, R. & Heinemann, U.** (1998) Crystal structures and properties of de novo circularly permuted 1,3-1,4-beta-glucanases, *Proteins.* **30**, 155-167.
35. **Keitel, T., Meldgaard, M. & Heinemann, U.** (1994) Cation binding to a *Bacillus* (1,3-1,4)-beta-glucanase. Geometry, affinity and effect on protein stability, *Eur. J. Biochem.* **222**, 203-214.
36. **Chiaraluce, R., Gianese, G., Angelaccio, S., Florio, R., van Lieshout, J. F., van der Oost, J. & Consalvi, V.** (2004) Calcium-induced tertiary structure modifications of endo-beta-1,3-glucanase from *Pyrococcus furiosus* in 7.9 M guanidinium chloride, *Biochem. J.* **386**, 515-524.

Chapter 7

Ribozyme-mediated engineering of circular mRNA and its functional *in vivo* and *in vitro* translation

Van Lieshout, J.F.T., Vroom, W., De Vos, W.M., Van der Oost, J.

submitted

Abstract

The instability of prokaryotic mRNA in some instances may result in a decreased overall yield of gene expression. Hence, better insight in the features that determine mRNA stability, and the ability to engineer stable messenger variants, might contribute to optimized production of certain proteins. The aim of the present study was to explore novel strategies for stabilization of mRNA by cyclization. A series of pET9d-derived plasmids has been generated, including a reporter (*Pyrococcus furiosus* β -glycosidase, CelB) and a ribozyme (*Tetrahymena thermophila* self-splicing intron), and used to transform the LacZ-deficient *Escherichia coli* JM109(DE3). The design of the constructs allowed for a simple *in vivo* selection for ribozyme-catalyzed mRNA processing, as the cyclization of the RNA molecule resulted in re-ligation of truncated *celB*-fragments. Subsequent translation of this cyclic mRNA resulted in a functional β -glycosidase enzyme, as was judged by blue staining after growth on LB agar supplemented with X-Gal. In addition, β -glycosidase activity was demonstrated in lysates of liquid cultures of the selected recombinants. Moreover, analysis of total RNA of these recombinants by RT-PCR confirmed the presence of covalently closed circular mRNA. Also transcription *in vitro* resulted in cyclic mRNA that showed resistance to degradation by exonuclease. Several variant constructs have been designed aiming at optimal splicing, ligation, and subsequent transformation of the cyclic transcript. Although the present cyclic variants did not give rise to an enhanced protein production compared to the linear construct, this engineering exercise showed potential applicability for its improvement of mRNA stability.

Introduction

The *Tetrahymena thermophila* self-splicing intron is inserted in the 26S ribosomal RNA (rRNA) gene of this unicellular eukaryote, where it autocatalyses its excision from the primary rRNA transcript. This ribozyme has been classified as a Group I intron, based on its conserved sequence elements that form distinctive secondary and tertiary structures by intramolecular Watson-Crick base-pairing (Fig. 7.1). In general, Group I introns catalyse self-splicing from their primary transcript in two sequential trans-esterification reactions, resulting in excision and circularization of the intron fragment, and ligation of the exon fragments (Fig. 7.2A) [1].

Thorough analysis has identified the essential sequence elements required for splicing and accurate recognition of the splice site by the intron. These elements include the 5' region of the intron, the internal guide sequence (IGS) which pairs with the 5' exon to form the Paired region 1 (P1) [2, 3]. This structure is needed for recognition of the conserved 5' exon sequence. A GU wobble base pair at the 5' splice site (Fig. 7.1B) determines the exact location of the cleavage site and is required for the initiation of the splicing reaction [4]. The sole requirements for splicing appear to be the presence of both monovalent and divalent cations (usually Mg^{2+}) and of guanosine or one of its 5'-phosphorylated forms (GMP, GDP, or GTP) [1]. These characteristics allow efficient splicing *in vivo* in *Escherichia coli* as well as *in vitro*.

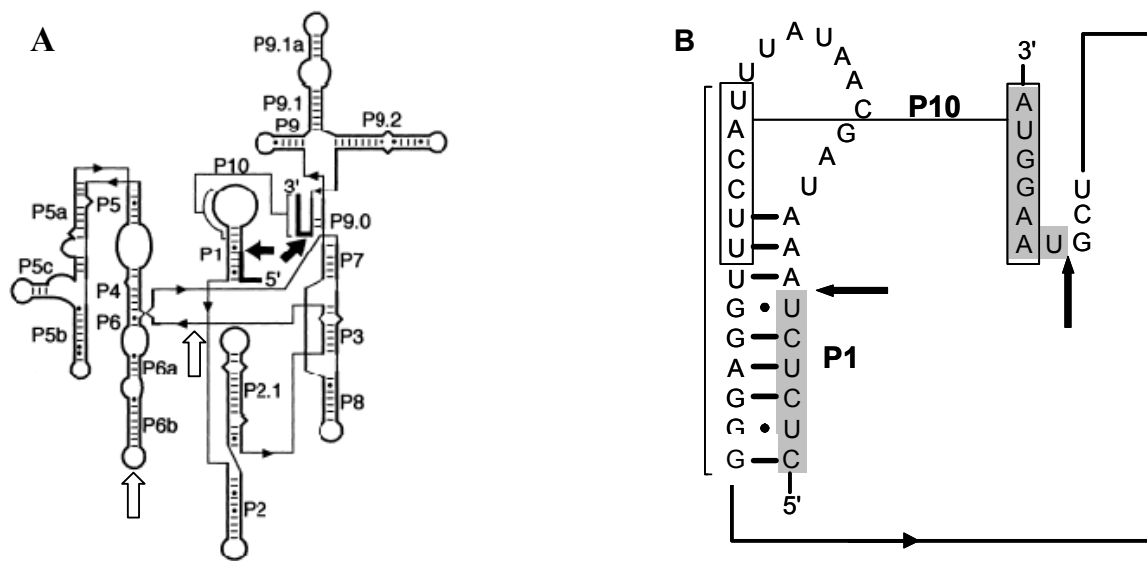


Figure 7.1 Secondary structure of the ribozyme from *T. thermophila*. (A) Open arrows indicate truncation sites for constructs pWUR96 and pWUR97. This figure is reprinted, with modifications, with permission from [5]. (B) A close-up of the P1 and P9 stem loops. Closed arrows indicate the 5' and 3' splice sites. Exon nucleotides are indicated in grey. The boxed nucleotides align to form the P10 helix, required in the second step of the splicing reaction. Adapted from [6], with permission.

Previous experiments with the Group I intron of the bacteriophage T4 (td intron) [7, 8] and the Group I introns of *Anabaena* [9, 10] have shown the possibility to divide the ribozyme into two parts and to rearrange the exon and intron sequences in such a way that the 3'-portions of both ribozyme and exon is followed by the 5' halves of exon and ribozyme (Fig. 7.2B). Such a DNA construct has been shown to yield a circular exon RNA after splicing both *in vitro* and *in vivo*. Furthermore, it has been demonstrated that a circular GFP-encoding transcript, including the sequential elements needed for translation initiation (RBS, ribosome binding site; DB, downstream box), can be effectively translated in *E. coli*, yielding the properly folded GFP [8]. An interesting observation was the apparent accumulation of circular mRNA, suggesting enhanced messenger stability, most likely due to increased exonuclease resistance [8]. In the same study, deletion of the stop-codon resulted in the "non-stop translation" of extremely long poly-proteins from circular messengers, proving the ability of the *E. coli* ribosomes to bind the circular RNA, and to continue translation for at least ten rounds. Unfortunately, translation efficiency of these circular mRNAs proved to be rather low, 20% compared to translation of their linear equivalent. Considering that circular messengers were found to accumulate *in vivo* [8, 11], the translational efficiency per RNA molecule may be even less.

In this study, the described techniques have been applied using the group I intron from *Tetrahymena thermophila* for generating circular transcripts of a β -glycosidase reporter protein. Several unprecedented designs have been made with the goal to establish an efficient system for *in vivo* and *in vitro* translation of circular mRNA. Furthermore, the increased stability of circular mRNA towards degradation by exonucleases has been analyzed.

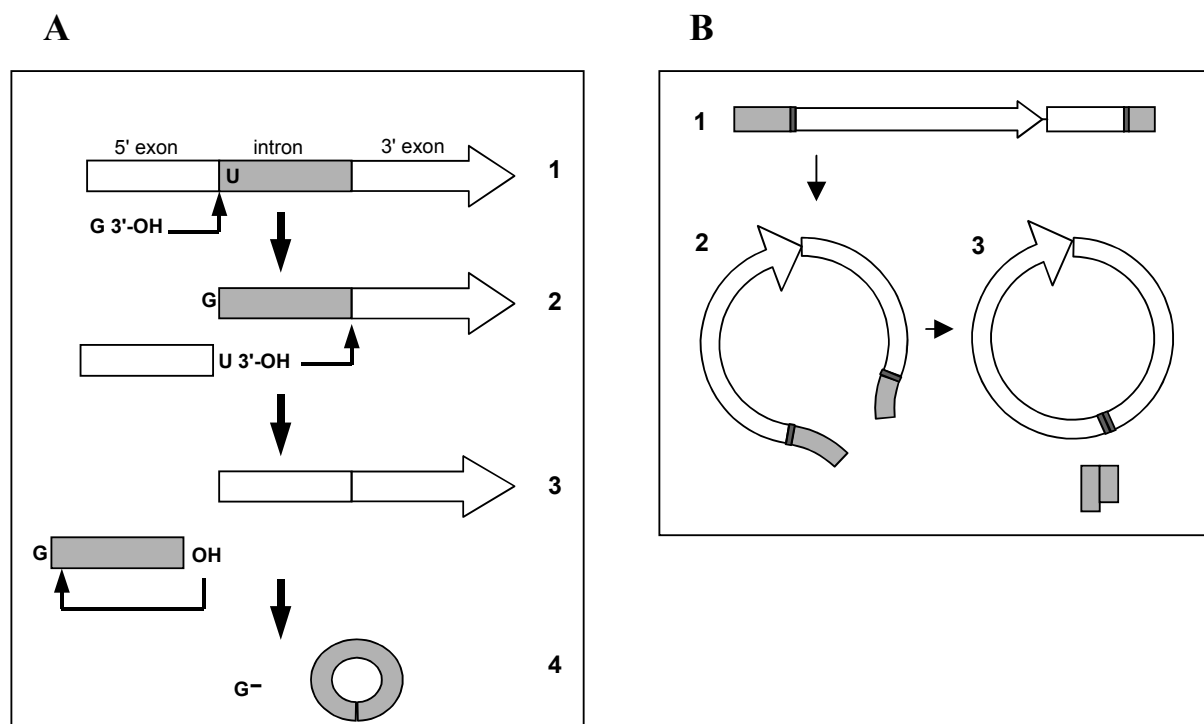


Figure 7.2 Schematic representation of the splicing mechanism of a Group I intron **(A)** and after engenderic rearrangement, yielding a circular exon **(B)**. Exon sequence is drawn white, intron sequence light grey. The dark grey fragments are the mutated exon boundaries in the *celB* gene. **(A)** The reaction is initiated by the nucleophilic attack of a 3' hydroxyl of a guanosine cofactor at the 5' splice site (1). The exon-intron phosphodiester bond is cleaved and the guanosine forms a 3',5'-phosphodiester bond at the 5' end of the intron. The liberated 3' hydroxyl of the 5' exon then makes a nucleophilic attack at the 3' splice site to form the ligated exons and release the intron with the non-encoded guanosine (2). Nucleotides close to the 5' end of the intron are realigned on the "internal guide sequence" (IGS) and the highly conserved 3' terminal guanosine of the intron makes a nucleophilic attack at a phospho-diester bond between nucleotides 15 and 16 or between nucleotides 19 and 20 within the intron in a reaction that is analogous to the first step of splicing (3). The intron is circularised (4) and a small fragment containing the non encoded guanosine is released (adapted from [5]). **(B)** 1 represents the transcribed, unspliced messenger. 2 shows both parts of the intron realigned *in vivo*. Finally, the splicing reaction yields a ligated, circular exon and a complex of two intron fragments (3).

Materials and Methods

Bacterial hosts and vectors - The T7 expression vector pET9d was obtained from Novagen. *Escherichia coli* JM109(DE3) (Stratagene) was used as an expression host for the pET-derivatives. *E. coli* was grown in TY medium in a rotary shaker at 37°C. Kanamycin was added to a final concentration of 30 µg/ml.

Cloning and expression - The *celB* gene coding for *Pyrococcus furiosus* β-glycosidase previously cloned in the pET9d-derived pLUW511 [12] was used as a template in PCR reactions. The intron has been PCR-amplified from *T. thermophila* genomic DNA (a kind gift from Dr. Hackstein, Radboud University Nijmegen). The primers that were used for the different constructs are listed in Table 7.1. All mutations and permutations were generated using overlap extension PCR [13]. The generated set of permuted genes was then ligated into pET9d after digestion with *Nco*I and *Bam*HI,

Table 1. Primers used in this study^as = sense, a = antisense^bsequences are given from 5'-3'

primer ^a	primer sequence ^b	Description
BG64 a	TACGGAAATTAATGTTGCC	anneals at position 235 in <i>ceIB</i>
BG238 s	GCAGCCATGGCAAGTCCCAAAATTCATGTTTG	forward primer <i>ceIB</i>
BG239 a	CGCGGATCCCTACTTTC TTGTAAACAAATTTGAGG	reverse primer <i>ceIB</i>
BG1372 s	CTCTCTAAATAGCAATA TTTACCTTTGG	forward primer <i>Tth</i> intron
BG1373 a	TACCTTACGAGTACTCCAAACTAA TCAA TATAC	reverse primer <i>Tth</i> intron
BG1374 s	CAGAGTCCTCTCTTAAGGTACTAGAGAAAA TTGCCAACA TGGAG	introducing mutation in <i>ceIB</i> , to create pWUR94
BG1375 a	CTAGTACCCTTAAGAGAGGACTCTGGACGCTACGGAAATTTATG	introducing mutation in <i>ceIB</i> , to create pWUR94
BG1397 s	GGAGTACTCGTAAGGTACTAGAGAAAA TTGCCAAC	connecting 3' half of <i>ceIB</i> with 5' end of <i>Tth</i> intron, to create pWUR95
BG1398 a	ATA TTGCTA TTTAGAGAGGACTCTGGAACGTCTAC	connecting 5' half of <i>ceIB</i> with 3' end of <i>Tth</i> intron, to create pWUR95
BG1399 s	CCAGATCCCTCTCTAAATAGCAATA TTTACCTTTG	connecting 3' end of <i>Tth</i> intron with 5' half of <i>ceIB</i> , to create pWUR95
BG1400 a	TTCTCTAGTACCTTACGAGTACTCCAAACTAA TC	connecting 5' end of <i>Tth</i> intron with 3' half of <i>ceIB</i> , to create pWUR95
BG1401 s	AAGAAAGTAGGAAGGAGATA TACCATGGCAAAAGTT	circular permutation pWUR96 and pWUR97, with rbs of pET9d
BG1402 a	ATGGTATA TCCTCTCCTACTTCTTTGTAAACAAA TTTGAGGTC	circular permutation pWUR96 and pWUR97, with rbs of pET9d
BG1403 s	CGCGCTCTAGAAA TTGCGGGAAGGGGTCAACAG	circular permutation pWUR96
BG1404 a	CGCGCTCAGCTGACGGCTTTGCC TTTTAAACCG	circular permutation pWUR96
BG1405 s	CGCGCTCTAGATTCTGTTGATATGGA TGCAGTTTAC	circular permutation pWUR97, pWUR115 and pWUR124
BG1406 a	CGCGCTCAGCGATCTGTTGACTTAGGACTTGGC	circular permutation pWUR97, pWUR115 and pWUR124
BG1436 s	ACCTCAAA TTAGGAGGTAGAAAATGAGGATCCGGCTGCTAACAAAGCC	introducing A TGA motif and rbs in pWUR94, to create pWUR114
BG1437 s	CTCAAA TTAGGAGGTAGAAAATGAAGTTCCCAAAAACCTTCATGTTGG	introducing A TGA motif and rbs in pWUR97, to create pWUR115
BG1438 a	TCATTTTCTACCTCTCTAA TTTGAGGTCTGCGAGGTGAGC	introducing A TGA motif and rbs in pWUR94 and pWUR97, to create pWUR114 and pWUR115
BG1486 a	CCTAACTACTTCTTGTAAACAAA TTTGAGGT	introducing rbs from pWUR115 into pWUR97, to create pWUR124
BG1495 s	CAAA TTTGTACAAAGAAAGTATTAGGAGGTAGACCA TGGCAAAAGTTCCCAAAAAA ACT	introducing rbs from pWUR115 into pWUR97, to create pWUR124
pET rev a	CCCGTCTCTGTGGATATCCGG	pET reverse primer, used with BG1436, to create pWUR114

or *Xba*I and *Blp*I (pWUR97), and constructs were used to transform to *E. coli* JM109(DE3). Sequence analysis of all constructs was done by the dideoxynucleotide chain termination method with a Li-Cor automatic sequencing system (model 4000L). The expression of the *celB* gene was monitored by blue-white-screening of colonies grown on plates containing 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-Gal) and in cell lysates by a discontinuous activity assay with pNp- β -D-glucopyranoside (pNp-Glu) as a substrate.

Enzymatic assays - Standard enzymatic assays were performed at 90°C in 60 mM citrate buffer (pH 7.0) with pNp-Glu (final concentration 3mM) as a substrate. An amount of 495 μ l was preheated in a 1.5 ml Eppendorf vial. The reaction was started by the addition of 5 μ l of heat-stable cell-free extract (soluble fraction of the cell-free extract after 30 minutes at 80°C). After exactly 7 minutes, the reaction was stopped by the addition of 1.0 ml of ice cold 0.5 M Na₂CO₃. This causes the pH to rise to about 9-10, terminating the reaction and enhancing the specific absorption coefficient of the liberated nitrophenol. ($\epsilon_{\text{pNp}} = 18.3 \text{ mM}^{-1}$ at pH = 9.8). The absorption of the reaction mixture was measured at 405 nm. Protein concentrations were determined according to Bradford [14]. All activities were corrected for spontaneous hydrolysis and for hydrolysis by cell free extracts of *E. coli* cells harboring pET9d without a gene insert.

In vitro expression - All *in vitro* reactions were performed using the Novagen EcoPro System, according to the corresponding protocol. Briefly, DNA-template samples were extracted twice with an equal amount of TE-buffered phenol:CIAA (1:1; CIAA is 24 parts chloroform, 1 part isoamyl alcohol) to remove possible RNase. After extracting once with CIAA NaOAc was added to a final concentration of 0.3 M. The DNA was precipitated with 2.5 volumes of 96% ice-cold ethanol and kept at -20 °C for 60 minutes. After centrifugation of the sample for 30 minutes at maximum speed the pellet was washed with 70% ethanol and resuspended in 20 μ l DEPC-treated sterile milliQ water. For *in vitro* expression, 5 μ l of this resuspended DNA was used per 50 μ l of total reaction volume.

RNA - For all RNA work, RNase free solutions and glassware were used. All plasticware was handled wearing gloves and sterilized in an autoclave. Solutions were treated with 0.1% DEPC for at least 12 hours at 37 °C, after which they were sterilized in an autoclave. DNA-samples are treated with a phenol-extraction to remove possible RNase. The isolation of total RNA was performed with a Qiagen RNeasy mini-kit. The RNA was eluted from the Qiagen RNA column with 30 μ l of sterile DEPC-treated milliQ water. Contaminating DNA was removed using the Qiagen RNase-Free DNase Set.

RT-PCR was performed with the Ambion RLM-Race kit, using MMLV reverse transcriptase. The Qiagen-isolated RNA was directly used in the reverse transcription experiment as described in the RLM-Race protocol. However, 4 μ l RNA was used in the reverse transcription reaction, instead of the recommended 2 μ l. Also, 2 μ l RT-product was used in 50 μ l amplification reaction, instead of the recommended 1 μ l. Two primer sets were used to prove the presence of

circular mRNA. Primer set 1 consisted of primers BG238 and BG239, primer set 2 consisted of primers BG64 and BG1374 (Table 7.1). Amplification of cDNA in the second part of the protocol was performed with an annealing temperature of 52 °C.

Exonuclease - The stability of circular mRNA towards degradation by exonuclease was examined by adding a specific exonuclease (phosphodiesterase II, Sigma) to the *in vitro* expression reaction. Samples were taken at different time points, and the reaction was stopped by placing the sample on ice. The amount of enzyme added to the *in vitro* reaction was 7×10^{-3} units per 30 µl reaction mix. This corresponds to 7 µl of a 10 mU/µl stock of phosphodiesterase in milliQ. The expression of *celB* was determined by a discontinuous activity assay as described earlier, but with an incubation period of 15 minutes at 90 °C.

Results

The self-splicing activity of the *Tetrahymena thermophila* ribozyme was applied for the production of circular RNA, i.e. by rearranging the sequences of the ribozyme intron and a reporter exon (Fig. 7.2B and 7.4). The *celB* gene of the hyperthermophilic archaeon *Pyrococcus furiosus* was used as a reporter gene to monitor efficient and accurate splicing of the ribozyme and the consecutive translation of the circularized *celB* messenger *in vivo*. The *celB* gene encodes a heat-stable β -glucosidase [15]. Functional *celB* expression in a β -glucosidase (LacZ) mutant of *E. coli* can easily be verified by blue-staining of recombinant colonies grown on X-Gal containing agar plates, and, in solution, by an activity assay with pNp-Glu as a substrate. The translated and properly folded CelB monomers assemble into a tetramer with four active subunits [15].

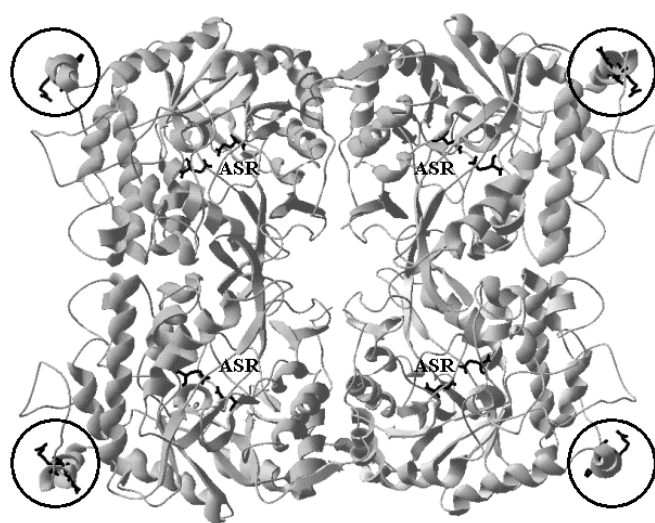


Figure 7.3 Model of the 3D-structure of CelB. Ribbon model of *Pyrococcus furiosus* β -glucosidase CelB, viewed along one of the 2-fold axes of the tetramer [12]. Active site residues (marked with ASR) and mutated residues (in encircled areas) are depicted in black.

Previous research [6, 16] has shown the importance of the flanking exon boundaries for an efficient and accurate recognition of the splice site. Therefore the *celB* gene was adapted for the insertion of the group I intron from *T. thermophila*. To ensure that such a mutation of the coding sequence of the gene would not disrupt the activity of the translated enzyme, a site at the surface of the tetramer and far from the active sites of the enzyme was chosen (Fig. 7.3). Using PCR, a 15 base pair substitution was introduced in the *celB* gene to create exon boundaries at the 5' and 3' flanking regions of the selected intron insertion site, yielding pWUR94 after cloning into *NcoI/BamHI* digested pET9d. The introduced flanking regions were similar to those in the 26S rRNA gene (Fig. 7.4A), and as such should perfectly pair with the internal guide sequence (IGS) of the intron to ensure proper folding at the splice sites. The intron was inserted in the 5' and 3' flanking regions of pWUR94, by the overlap extension method, resulting in pWUR95. Both constructs were used to transform *E. coli* JM109(DE3) cells, and colonies were screened for blue-staining on X-Gal supplemented agar plates. Cell-free extracts of cultures of the different recombinants were prepared and incubated at 80 °C for 30 minutes; these conditions result in enrichment of the heat-stable CelB because of denaturation of most *E. coli* proteins. The level of production of CelB was compared by discontinuous activity assays and analysis on SDS-PAGE (Fig. 7.5). Functional expression of active CelB was observed for both constructs, confirming i) that the introduced mutation did not impair enzyme activity, and ii) the capacity of the *Tth* ribozyme to perform self-splicing in *E. coli* JM109(DE3), with the introduced mutations in the exon boundaries in the *celB* gene. The increase in specific activity of the samples after heat incubation, indicated that it was in fact the heat stable CelB that was responsible for the β -glycosidase activity measured.

Next, a circular permutation of the gene construct was made subsequently to create a DNA sequence that would lead to the production of circular RNA, provided that the splicing reaction still occurs. Analysis of the known 2-dimensional structure of the *Tth* group I intron (Fig. 7.1A) showed that dividing the intron in the P6b stem at position 238 (indicated with an open arrow in Fig. 7.1A) would most probably result in an active ribozyme even when divided by a 1.5 kb gene. The bacteriophage T4 td intron has previously been truncated in the P6 region as well, yielding an active ribozyme [11]. Moreover, division of the intron in the P6b stem would result in large complementary strands with the potential to strongly interact. Using partially overlapping primers, both parts of the gene construct with the mutated *celB* and the ribozyme were PCR-amplified and ligated in a consecutive PCR reaction, finally resulting in pWUR97 (Fig. 7.4). This construct contains, from 5' to 3': the 3' portion of the ribozyme, the 3' exon with a stop codon (TAG), a ribosome binding site (AGGAGAT), a start-codon (ATG), the 5' exon and finally the 5' portion of the ribozyme.

A second truncation site in the intron was made in the joining region of stem 3 and 4 (at position 104 from the 5' end of the intron; indicated with an open arrow in Fig. 7.1A), yielding pWUR96. Motivation for this truncation site was the reported high-affinity self-assembly of separately expressed domains of the *Tth* ribozyme (P4-P6 and P3-P9), resulting in an active ribozyme [17].

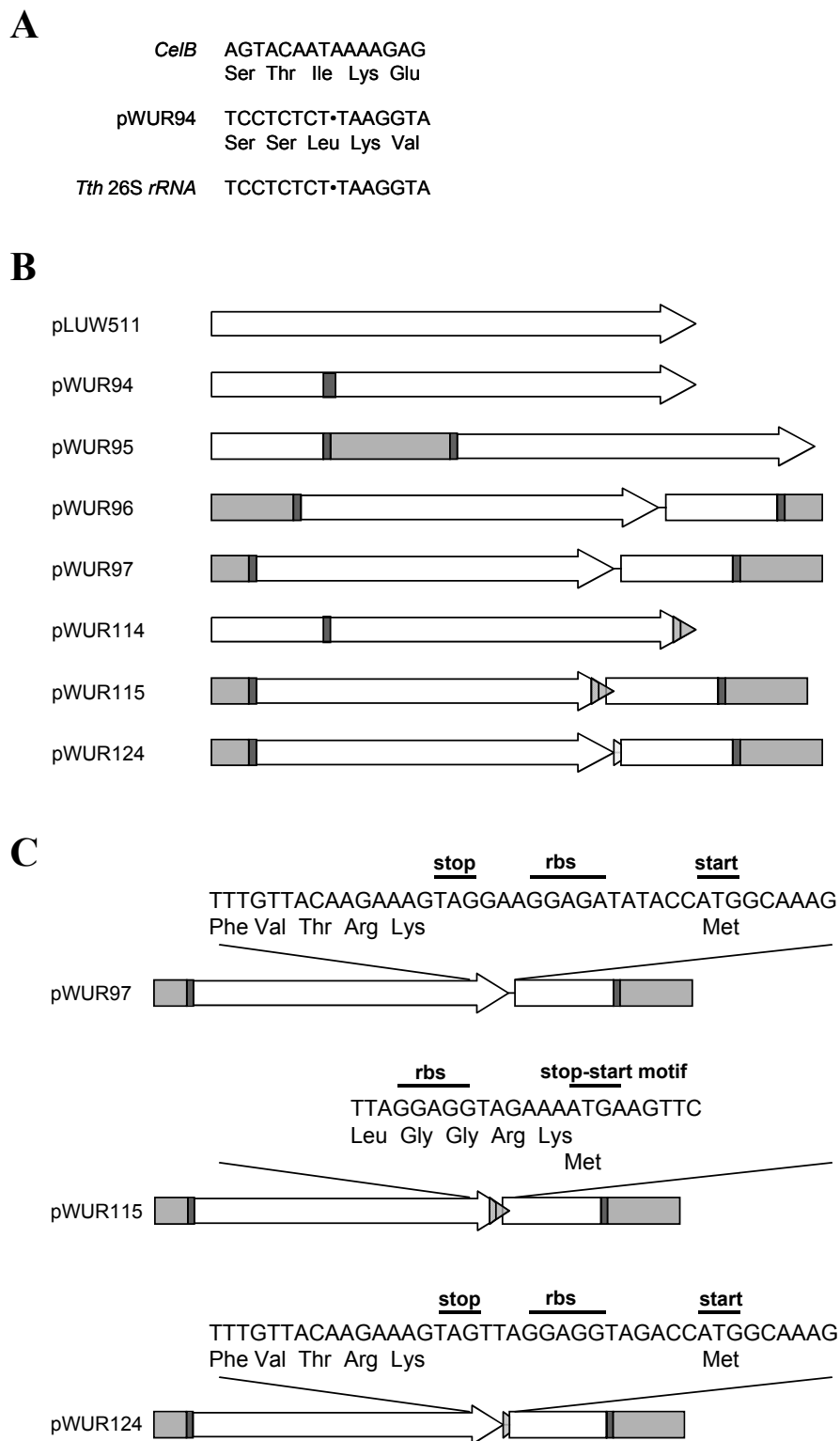


Figure 7.4 Overview of all DNA constructs. **(A)** sequence alignment of the flanking exon boundaries in the 26S rRNA of *Tetrahymena thermophila* (*Tth* 26S rRNA), the one that has been introduced in *celB* (pWUR94) and the original sequence of *celB* at that position. **(B)** and **(C)** Schematic presentation of the different constructs. Only the DNA insert of the vectors is shown. Colors as in Fig. 2. The introduced ribosome binding site (rbs) in pWUR114, 115 and 124 is depicted in outlined diamond pattern, the introduced ATGA motif is depicted in dark horizontal pattern. **(C)** detailed sequence view of the stop-start region of the 3 constructs with different designs of ribosome binding site (rbs) or stop-start motif. Stop- and start-codons are indicated.

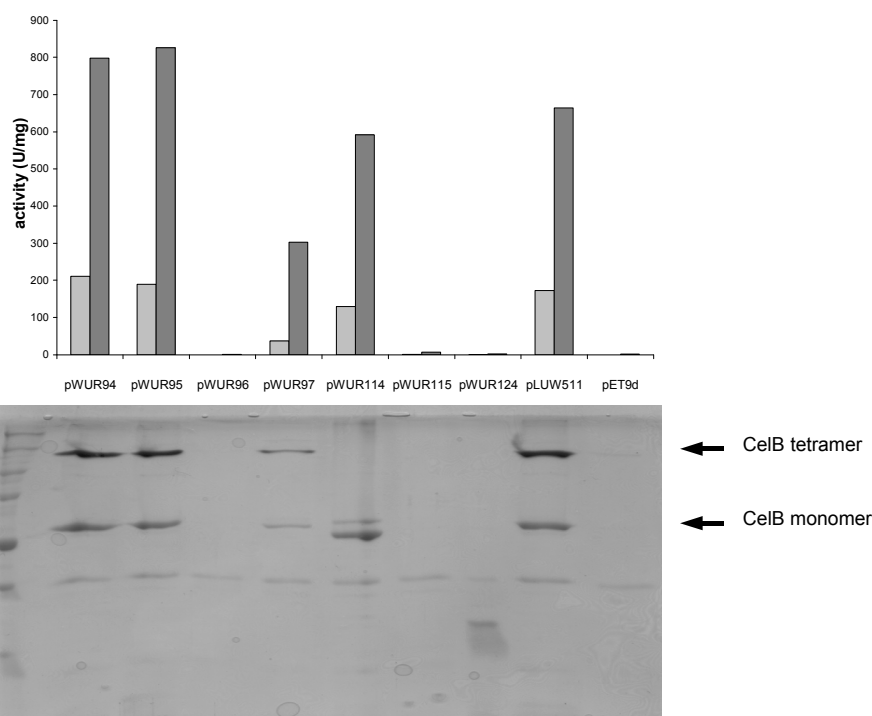


Figure 7.5 Expression and activity of CelB from different constructs. The upper panel shows the specific activity of CelB in Units per mg protein in cell-free extract, CFE (depicted in light grey) and heat-stable cell-free extract, HSCFE (depicted in dark grey). The lower panel shows the corresponding SDS-PAGE analysis of the heat stable cell-free extracts of the different constructs. Arrows indicate the tetramer (216 kDa) and monomer (54 kDa) of CelB. Lane 1, Biorad precision marker; Lane 2, pWUR94; Lane 3, pWUR95; Lane 4, pWUR96; Lane 5, pWUR97; Lane 6, pWUR114; Lane 7, pWUR115; Lane 8, pWUR124; Lane 9, pLUW511; Lane 10, pET9d;

A recent study by Ishida *et al.* (2002) [18] has shown how an upstream small open reading frame (encoding a leader peptide) may enhance the expression of an adjacent gene in *E. coli*, especially when the stop and start codons overlap (ATGA). Introducing an ATGA-motif with a 4-basepair overlap between the stop codon of a leader open reading frame and the start codon of the downstream coding region of a gene, increased protein expression more than 3-fold, compared to the expression from a construct with a 10 nt gap between the leader ORF and the start codon of the gene. The authors proposed that the enhanced expression of the downstream gene was a result of “efficient ribosome recruitment” [18]. In analogy, pWUR115 was designed with an ATGA motif and a RBS introduced at the 3' end of the *celB* coding region (Fig. 7.4). The design of construct pWUR115 with respect to the RBS and the spacing towards the start-codon was based on previous research on the role of the Shine-Dalgarno (SD) sequence in prokaryotes [19]. To introduce the RBS at the 3' end of *celB*, three amino acids had to be substituted. To evaluate the influence of this mutation on the activity of the enzyme, the same mutation was introduced in pWUR94, resulting in pWUR114. PCR using partially overlapping primers was used to create pWUR115 and pWUR114 with the described mutations in the 3' end of the sequence. The importance of the exact sequence of the SD region for ribosome recruitment, and therefore translation initiation, has been shown before [18]. To examine the influence of the RBS and its directly flanking region on the expression from pWUR115, this sequence was replaced in pWUR97, resulting in pWUR124 (Fig. 7.4).

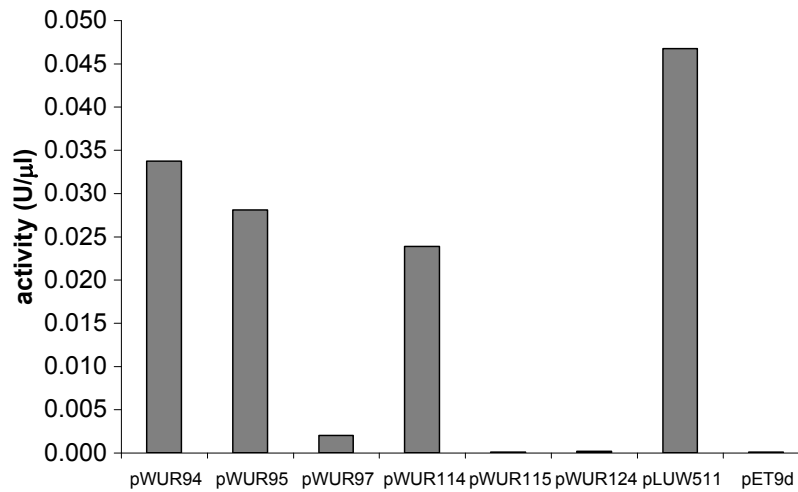


Figure 7.6 *In vitro* expression. The expression levels *in vitro* are indicated by the activity in Units per μ l of the *in vitro* transcription/translation reaction.

No functional expression of CelB was observed for constructs pWUR96, pWUR115 and pWUR124, while pWUR97 yielded lower levels of active protein compared to wild-type *celB* and pWUR94 and pWUR95. Analysis on SDS-PAGE revealed the presence of CelB tetramer and monomer in the different heat-stable cell-free extracts (Fig. 7.5), as has been observed before [20]. All samples were incubated at 99°C for 10 minutes, prior to loading on gel to dissociate the CelB tetramer. However, to some extent the tetramer was still present on gel, except for sample pWUR114 (Fig. 7.5C, lane 6).

In addition to *in vivo* heterologous expression, all constructs were tested for transcription, splicing and translation *in vitro*, using the Novagen Ecopro system, which is a prokaryotic *in vitro* expression system based on a bacterial extract. Expression levels of CelB were examined using a discontinuous activity assay, showing similar trends as observed *in vivo* (Fig. 7.6).

An alternative assay to prove the presence of circular RNA was performed by RT-PCR [21]. RNA was purified from cultures in the mid log phase, and used as a template in a reverse transcription reaction by MMLV reverse transcriptase to produce the corresponding cDNA. Two sets of primers for the RT-reaction were designed such that only covalently closed circular RNA (cccRNA) could give rise to cDNA with both primer sets. Linear, unspliced RNA would only give rise to amplification with one primer set (set 1, grey dashed line in Fig. 7.7A), and RNA from the wild-type gene would give rise to amplification from the other primer set (set 2, black dotted line in Fig. 7.7A). The cDNA produced in the RT-reaction was consecutively amplified in a normal PCR reaction with *Taq* polymerase to produce sufficient amounts of DNA to visualize by Ethidium Bromide staining after electrophoresis on an agarose gel. For pWUR97, this analysis indeed showed the presence of amplified DNA products of the expected size for both primer sets, thereby confirming the presence of circular mRNA (Fig. 7.7B). The same reaction with pWUR96-RNA showed only amplification from primer set 1, indicating the presence of unspliced mRNA.

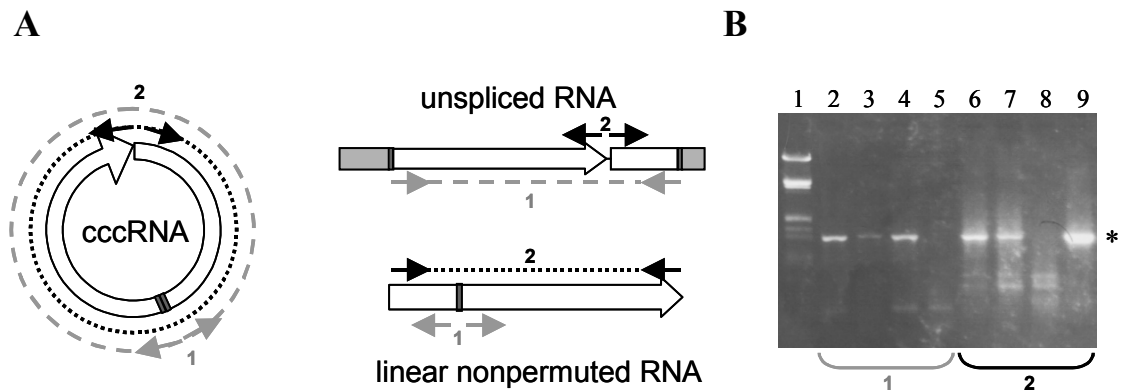


Figure 7.7 RT-PCR. **(A)** Two sets of primers can amplify different products, depending on the template present. Only cccRNA will give rise to a product with both primersets. Colors as in Fig. 2. **(B)** Gel electrophoresis after RT-PCR, the asterix indicates the expected product (1.4 kb). Lane 1, DNA marker; Lanes 2-5 represent the products with primer set 1, with in Lane 2, pWUR97; Lane 3, pWUR97; Lane 4, pWUR96; Lane 5, pLUW511; Lanes 6-9 represent the products obtained with primer set 2, with in Lane 6, pWUR97; Lane 7, pWUR97; Lane 8, pWUR96; Lane 9, pLUW511;

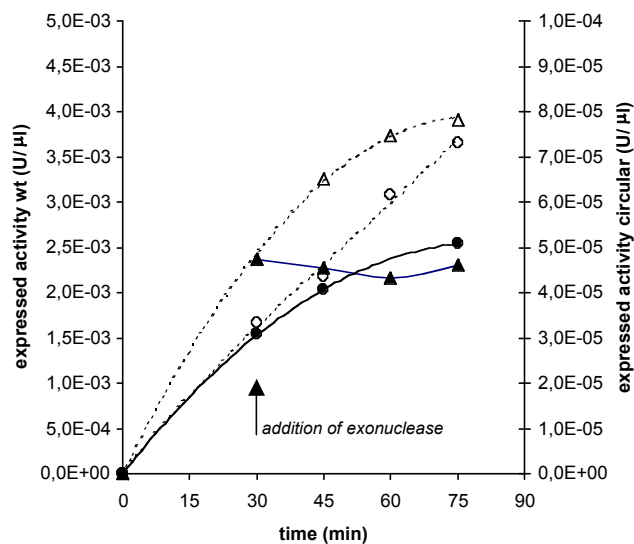


Figure 7.8 Stability of circular mRNA. *In vitro* expression of *celB* from linear and circular messenger RNA, after addition of exonuclease (phosphodiesterase II, Sigma). The lines with triangles show the results from the wild-type *celB* gene (left axis), the lines with closed circles show the results from circular RNA from construct pWUR97 (right axis). After 30 minutes of *in vitro* expression, the exonuclease was added to both samples. The dashed lines show the increase in expressed CelB activity in the absence of exonuclease.

pLUW511-RNA gave rise to a product with primer set 2, indicating the presence of normal, non-permuted messengers, but no circular messenger.

The stability of circular and linear mRNA against exonuclease degradation was compared. During *in vitro* expression, a specific exonuclease (phosphodiesterase II, Sigma) was added to the reaction and samples were taken at different time points. The increase in CelB-activity after addition of exonuclease was measured and compared to the CelB-activity in samples without added exonuclease (Fig. 7.8). The results indicated an enhanced yield in the case of the circular messenger, strongly suggesting a stabilization of the circular mRNA.

Discussion

In the present study we have shown that a rearranged group I intron, derived from the ciliate *T. thermophila*, can be used to synthesize large circular mRNAs in *E. coli*, and that this circular transcript is functionally translated to an active β -glycosidase. Additional analyses, using SDS-PAGE and *in vivo* and *in vitro* activity assays, were instrumental for assessing the influence of the constructs' design on the expression level of *celB*.

We could show that introduction of an intron in the *celB* gene, as in pWUR95, did not significantly decrease the expression level of the enzyme. This confirmed results of Guo and Cech [6], showing that the eukaryotic splicing mechanism of the intron can easily occur *in vivo* in *E. coli*. Unlike the latter study, however, the original "intron insertion site" introduced in the *celB* gene proved to be effective for an accurate and efficient splicing reaction, without decreasing the enzyme activity or stability.

Furthermore, we could demonstrate that the success of a circular permutation of construct pWUR95 strongly depends on the actual site of the permutation. Truncating the intron between stem P3 and P4, as in construct pWUR96, did not give rise to the production of CelB. This is most likely due to a disturbed splicing mechanism of the transcribed mRNA. The failure of this construct is unexpected given the results of Doudna and Cech [17], and probably indicates the effect of the additional sequence (*celB* fragments) that may affect the correct assembly of the ribozyme domains. On the other hand, a division in stem-loop P6, as in pWUR97, yielded functional CelB expression. This successful re-assembly of the ribozyme domains agrees well with analogous truncation experiments with T4 td intron [11]. The production of active CelB by cells harbouring pWUR97 confirmed that the transcribed messenger can indeed perform autocatalytic splicing and ligation *in vivo*, and that the created circular messenger can be translated by the prokaryotic translation machinery. Although translation of circular mRNAs has been shown once before with the T4 Group I intron [8], the here described circular messenger is slightly larger in size, it concerns an enzyme-encoding transcript, and the unprecedented design of shuffled gene fragments is a convenient tool for screening functional construct designs.

Comparison of the expression levels of the permuted construct and the wild type gene revealed a decrease in activity of approximately 90%; this is comparable with the GFP production from a circular messenger using the T4 td intron [8]. Since the introduction of an intron in the native *celB* gene does not affect expression levels, the truncation of the intron most likely results in a decreased splicing efficiency of the messenger because of sub-optimal assembly of the ribozyme domains. This may result in lower levels of accurately processed mRNAs, or may even slow down the overall splicing and ligation reactions, allowing nucleases to degrade the linear messenger. Alternatively, efficiency of translation initiation could be decreased in case of a circular messenger. Ribosome recruitment on a circular messenger might be less efficient than recruitment on a linear messenger. This could also negatively affect protein production levels.

In vitro expression experiments revealed expression profiles for the different constructs that agree very well with the trends observed *in vivo*. Comparison of the wild-type gene and the permuted construct showed an approximate 95% decrease in functional CelB production, which is

slightly worse than observed *in vivo*. The anticipated increased stability of the circular RNA does not appear to be an extra advantage *in vitro*.

Apart from the technical challenge, a motivation of generating circular messengers was to enhance *in vivo* expression, because of anticipated longer messenger lifetime, and maybe even further by increased translation efficiency via a “rolling circle mechanism”. However, these effects could neither be demonstrated in the case of CelB (this study) nor GFP [8]. A likely explanation for this observation is that the transcript stability in the case of these two reporters is not a limiting factor for protein production.

To accomplish rolling-circle translation of the *celB* messenger, the construct was engineered to introduce a 4-basepair overlap in the stop and start codon of the *celB* gene, as is observed in many natural poly-cistronic messengers in prokaryotes. This should allow the ribosome to continuously read and translate without dissociating from the circular mRNA. However, pWUR115, did not yield active CelB. To create the 4-basepair overlap in stop and start codon (ATGA-motif), radical changes had to be introduced at the C-terminus of CelB, and to some extent in the SD-sequence. Two constructs were designed to evaluate these changes introduced in the DNA-sequence of pWUR115: pWUR114 and pWUR124 (Fig. 7.7AB). In construct pWUR114, the mutation in the 3' end of the *celB* gene was introduced in the mutated *celB* gene, without intron sequence. Although some activity was measured in the cell-extracts of cells containing this construct, it was approximately 50% less than the expression from the wild-type gene. On SDS-PAGE, the overexpressed protein appeared to migrate slightly faster than the wild-type CelB. Moreover, the very stable tetramer of the enzyme was still present in the lanes with the wild-type CelB, but did not appear in the pWUR114 lane. This suggests that either the tetramer of the enzyme was not present at all, or it was less stable than the tetramer of the wild-type CelB and had therefore completely dissociated during the heat-incubation of the sample, prior to loading on gel. The C-terminal end of the CelB monomer has previously been reported to play a crucial role in tetramer formation and stability [20]. A mutation in this part of the protein might therefore indeed decrease the stability of the tetramer. Since only the tetramer of the protein is known to be active [20], this might also explain the decreased enzyme activity in the cell-extract.

The influence of the changed SD-sequence was evaluated with construct pWUR124, in which the SD-sequence and spacer between stop- and start-codon of construct pWUR97 was replaced by the SD-sequence as present in pWUR115 (Fig. 7.7). The lack of activity of this construct (which has a more perfect match with the anti-SD of *E. coli*) is puzzling, as also sequence analysis did not indicate mistakes. This might reflect the importance of the sequence directly upstream of the gene for protein expression [22]. However, the introduced ribosome binding site is known to be functional in *E. coli*. An alternative explanation might be that the introduced substitutions enable the formation of some α -specific interaction with the ribozyme domains that disrupt its splicing and/or ligation activity. Future research is required to draw final conclusions on the effect of the 4-basepair overlap in an ATGA-motif. Random mutagenesis of the SD-sequence might lead to variants with enhanced efficiency, as well as applying the proposed design on a gene with a less crucial 3' end.

Addition of exonuclease apparently resulted in an immediate stop of CelB production from a linear construct (Fig. 7.8). This is in agreement with the expected degradation of the linear messenger RNA by the enzyme, thus destroying any template for the translation reaction. Addition of exonuclease to a circular RNA-template decreased CelB production as well, but certainly not to the same extent. The decrease in production compared to the samples without exonuclease can be explained by assuming the degradation of primary, unspliced, and therefore linear, transcript. However, the already present pool of circular mRNA remains available as a template for translation, and as such for functional CelB production.

A method was developed to prove the presence of a specific circular RNA by means of RT-PCR. Although the presence of an active enzyme itself is a strong indication for a successful intron splicing and an effective translation of the messenger, the RT-PCR method provides elegant means to confirm the presence of circular RNA. Moreover, in case of disturbed messenger translation, RT-PCR might still reveal the presence of circular RNA. Finally, the increased stability of circular mRNA towards exonuclease degradation is demonstrated.

Instead of optimizing the translational efficiency of the circular messenger, another way of improving protein expression in future research could be to improve the splicing efficiency of the permuted messenger *in vivo*. Studies by Guo and Cech (2002) [6] indicated that the flanking regions are extremely important: base pairing should not be too strong, but proper base pairing of the 5' exon to the IGS to form the P1 stem loop, as well as base pairing of the 3' exon to the IGS to form the P10 stem loop, are essential. Analysis of the secondary structure of the *Tth* 26S rRNA, shows a highly organised structure which most likely is crucial for optimal intron splicing. Previous research has in fact revealed that the minimum length for optimal splicing of the *Tth* 26S rRNA, is 145 nucleotides upstream and 85 nucleotides downstream of the intron [23]. More research might reveal essential sequential elements in this region for an improved intron splicing, and might be useful to explain some of the efficiency-related matters that were encountered in the present study.

A rather interesting question that remains is whether the ribosome is capable of translating the circularized mRNA more than once without dissociating from its RNA template. Evaluating the expression of different constructs described in this study does not give a definite answer to this question. In the study of Perriman and Ares [8] a design was made in which the stop codon was deleted, resulting in a continuous circular open reading frame; this indeed resulted in long repetitive strings of proteins, indicating the occurrence of rolling circle translation. Alternative designs presented here did not lead to the anticipated result, for reasons that are not understood at present.

The system described in this study may be useful for detailed analysis of ribozyme-catalyzed splicing, as well as of ribosome-mediated translation. Apart from these fundamental processes, the system may have potential for improved expression of genes that have very instable messengers, or for certain *in vitro* applications that require mRNA templates with enhanced stability.

Acknowledgments

This research was supported by the Technology Foundation (STW), applied science division of NWO and the technology programme of the ministry of Economic Affairs.

References

1. Cech, T. R. (1990) Self-splicing of group I introns, *Annu. Rev. Biochem.* **59**, 543-568.
2. Been, M. D. & Cech, T. R. (1986) One binding site determines sequence specificity of *Tetrahymena* pre-rRNA self-splicing, trans-splicing, and RNA enzyme activity, *Cell*. **47**, 207-216.
3. Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. (1986) Splice-site selection by a self-splicing RNA of *Tetrahymena*., *Nature*, 133-139.
4. Strobel, S. A. & Cech, T. R. (1995) Minor groove recognition of the conserved G.U pair at the *Tetrahymena* ribozyme reaction site, *Science*. **267**, 675-679.
5. Tanner, N. K. (1999) Ribozymes: the characteristics and properties of catalytic RNAs, *FEMS Microbiol. Rev.* **23**, 257-275.
6. Guo, F. & Cech, T. R. (2002) *In vivo* selection of better self-splicing introns in *Escherichia coli*: the role of the P1 extension helix of the *Tetrahymena* intron, *RNA*. **8**, 647-658.
7. Perriman, R. (2002) Circular mRNA encoding for monomeric and polymeric green fluorescent protein, *Methods Mol. Biol.* **183**, 69-85.
8. Perriman, R. & Ares, M., Jr. (1998) Circular mRNA can direct translation of extremely long repeating-sequence proteins *in vivo*, *RNA*. **4**, 1047-1054.
9. Puttaraju, M. & Been, M. D. (1992) Group I permuted intron-exon (PIE) sequences self-splice to produce circular exons, *Nucleic Acids Res.* **20**, 5357-5364.
10. Puttaraju, M. & Been, M. D. (1996) Circular ribozymes generated in *Escherichia coli* using group I self-splicing permuted intron-exon sequences, *J. Biol. Chem.* **271**, 26081-26087.
11. Ford, E. & Ares, M., Jr. (1994) Synthesis of circular RNA in bacteria and yeast using RNA cyclase ribozymes derived from a group I intron of phage T4, *Proc. Natl. Acad. Sci. U S A.* **91**, 3117-3121.
12. Kaper, T., Lebbink, J. H., Pouwels, J., Kopp, J., Schulz, G. E., van der Oost, J. & de Vos, W. M. (2000) Comparative structural analysis and substrate specificity engineering of the hyperthermostable beta-glucosidase CelB from *Pyrococcus furiosus*, *Biochemistry*. **39**, 4963-4970.
13. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K. & Pease, L. R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene (Amsterdam)*. **77**, 51-59.
14. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, *Anal. Biochem.* **72**, 248-254.
15. Voorhorst, W. G., Eggen, R. I., Luesink, E. J. & de Vos, W. M. (1995) Characterization of the celB gene coding for beta-glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* and its expression and site-directed mutation in *Escherichia coli*, *J. Bacteriol.* **177**, 7105-7111.
16. Hagen, M. & Cech, T. R. (1999) Self-splicing of the *Tetrahymena* intron from mRNA in mammalian cells, *EMBO J.* **18**, 6491-6500.
17. Doudna, J. A. & Cech, T. R. (1995) Self-assembly of a group I intron active site from its component tertiary structural domains, *RNA*. **1**, 36-45.
18. Ishida, M. & Oshima, T. (2002) Effective structure of a leader open reading frame for enhancing the expression of GC-rich genes, *J. Biochem. (Tokyo)*. **132**, 63-70.
19. Ma, J., Campbell, A. & Karlin, S. (2002) Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures, *J. Bacteriol.* **184**, 5733-5745.

20. **Pouwels, J., Moracci, M., Cobucci-Ponzano, B., Perugino, G., van der Oost, J., Kaper, T., Lebbink, J. H., de Vos, W. M., Ciaramella, M. & Rossi, M.** (2000) Activity and stability of hyperthermophilic enzymes: a comparative study on two archaeal beta-glycosidases, *Extremophiles*. **4**, 157-164.
21. **Mackie, G. A.** (2000) Stabilization of circular rpsT mRNA demonstrates the 5'-end dependence of RNase E action in vivo, *J. Biol. Chem.* **275**, 25069-25072.
22. **Sprengart, M. L. & Porter, A. G.** (1997) Functional importance of RNA interactions in selection of translation initiation codons, *Mol. Microbiol.* **24**, 19-28.
23. **Woodson, S. A.** (1992) Exon sequences distant from the splice junction are required for efficient self-splicing of the *Tetrahymena* IVS, *Nucleic Acids Res.* **20**, 4027-4032.

Chapter 8

Summary and Concluding Remarks

The catalytic properties, thermostability, and production of thermostable glycoside hydrolases have been investigated. Moreover, attempts have been made to improve these parameters in order to efficiently generate biocatalysts that are better suited for application in sugar-converting processes, in which high temperatures are often desired. Various strategies and techniques have been applied, some of which unconventional. A novel α -galactosidase (GalA), a laminarinase (LamA) and a β -glucosidase (CelB) from the hyperthermophilic archaeon *Pyrococcus furiosus* served as models in this study, as well as a lichenase (LicA) from the bacterium *Bacillus licheniformis*.

As a general introduction to this thesis, **Chapter 1** provides a review on glycoside hydrolases, with specific attention those from hyperthermophilic origin. The classification and reaction mechanism of this class of enzymes is discussed as well as the methodology that is used to study these biocatalysts. An extended characterization of one such enzyme, an α -galactosidase (GalA) from *Pyrococcus furiosus*, is described in **Chapter 2**. The GalA gene (*galA*) was identified, cloned and functionally expressed in *Escherichia coli*. GalA is a novel type of α -galactosidase from a hyperthermophilic archaeon. The *galA* gene encodes a unique amino acid sequence compared to other α -galactosidases. Highest homology was found with α -amylases classified in family 57 of glycoside hydrolases. GalA specifically catalyzed the hydrolysis of *para*-nitrophenol- α -1,4-D-galactopyranoside, and to some extent that of melibiose and raffinose. The enzyme appeared to be an α -galactosidase with extremely high thermo-activity (temperature optimum of 115 °C) and thermal stability (half-life time of 15 hours at 100 °C). Sequence analysis showed four conserved carboxylic residues. Site-directed mutagenesis was applied to identify the potential catalytic residues. Glu117Ala showed decreased enzyme activity, which could be rescued by the addition of azide or formate. It is concluded that glutamate 117 is the catalytic nucleophile, whereas the acid/base catalyst remains to be identified. The importance of identifying the catalytic residues is illustrated in **Chapter 3** that describes the engineering of the *Pyrococcus furiosus* laminarinase (LamA), an endo-glycosidase known to hydrolyze β -1,3-gluco-oligosaccharides, but unable to degrade β -1,4-gluco-oligosaccharides. The nucleophile mutant Glu170Ala had a severely reduced hydrolytic activity, but was active as a glycosynthase: the condensation of α -laminaribiosyl fluoride and different acceptors resulted in oligosaccharides with a yield of up to 30%. Depending on the acceptor, the synthase generated either a β -1,3 or a β -1,4 linkage. In the same chapter, the specificity of the wild-type hydrolase towards small saccharides has been investigated in more detail, using 4-methylumbelliferyl β -glucosides with different linkages. Besides endo-activity, wild-type LamA also was demonstrated to have some exo-activity, and the potential to hydrolyze mixed-linked oligosaccharides with both β -1,4 and β -1,3 specificities.

In **Chapter 4** certain details of the stability and tertiary structure conformation of LamA are described. In 7.9 M GdmCl the addition of calcium to the enzyme causes significant changes to the near-UV circular dichroism (CD) and fluorescence spectra, suggesting a notable increase in the tertiary structure which leads to a state comparable, but not identical, to the native state. The capability to interact with calcium in 7.9 M GdmCl with a consistent recovery of native tertiary structure is a unique property of this extremely stable endo- β -1,3 glucanase. The effect of calcium on the thermodynamic parameters relative to the GdmCl-induced equilibrium unfolding has been

analyzed by circular dichroism and fluorescence spectroscopy. The interaction of calcium with the native form of the enzyme was studied by Fourier transformed infrared spectroscopy in the absorption region of carboxylate groups and by titration in the presence of a chromophoric chelator. A homology-based model of the enzyme was generated and used to predict the putative binding site(s) for calcium and the structural interactions potentially responsible for the unusual stability of this protein, in comparison with other family 16 glycoside hydrolases.

In various industrial applications the immobilization of the enzyme is desired. The structural characteristics and activity of LamA upon immobilization on Teflon and silica has been described in **Chapter 5**. The materials were designed such that the interaction of the particles with light was negligible and the enzyme conformation in the adsorbed state could be monitored *in situ*. The adsorption isotherms were determined and the adsorbed endoglucanase was studied using a number of spectroscopic techniques, enzymatic activity assays, and dynamic light scattering. It was shown that the enzyme adsorbed to the hydrophobic surface of Teflon with higher affinity than to the hydrophilic surface of Silica nanoparticles. In all cases, adsorption resulted in slight changes in the secondary structure. The changes were more profound upon adsorption on Teflon. The activity of the adsorbed enzyme showed the same pH-dependency as the enzyme free in the solution and it remained active in the adsorbed state in spite of the structural changes induced when interacting with the surfaces.

To explore innovative means for enhancing enzyme stability, a recently described method of protein engineering has been used, as presented in **Chapter 6**. The approach is based on the natural process of “intein-driven protein splicing”. A variation on this theme allows for the generation of a covalent linkage between the N-terminus and the C-terminus of a polypeptide chain. The method has been applied to create circular variants of LicA, a lichenase (endo- β -1,3-1,4-glucanase) from *Bacillus licheniformis*. Two cyclic variants showed a significantly higher thermal stability than the linear variant, whereas catalytic activity was sustained. After 3 minutes at 65 °C the linear glucanase had lost half of its activity, whereas the two cyclic variants showed a 6-fold and 16-fold increase of half-life times of activity at this temperature. Upon increasing temperatures, fluorescence spectroscopy demonstrated that the cyclic polypeptides started to unfold significantly later than the linear one, with an increase in apparent melting point by 2 °C.

Finally, in **Chapter 7**, attempts were made to stabilize messenger RNA (mRNA) by cyclization, and its effect on the efficiency of heterologous expression has been analyzed. In some instances the instability of prokaryotic mRNA may result in a decreased overall yield of gene expression. Better insight in the features that determine mRNA stability, and the ability to engineer stable messenger variants, might contribute to optimisation of protein production by microbial Cell Factories. A series of constructs has been generated including a reporter (β -glucosidase) and a ribozyme (self-splicing intron). They were used to transform *E. coli*. The constructs were designed in such a way that ribozyme-catalyzed mRNA processing resulted in cyclization of the RNA molecule with re-ligation of truncated β -glucosidase-fragments as a consequence. Subsequent translation of this cyclic mRNA resulted in a functional β -glucosidase enzyme, as was shown after growth both by activity screens on agar plates and in liquid cultures. Moreover, analysis of total RNA of these recombinants by RT-PCR confirmed the presence of covalently closed circular

mRNA. Also transcription *in vitro* resulted in cyclic mRNA that showed resistance to degradation by exonuclease. Several design variations have been tested aiming at optimal splicing, ligation, and subsequent transformation of the cyclic transcript. Although the present cyclic variants did not give rise to an enhanced protein production compared to the linear construct, this engineering exercise showed potential applicability for its improvement of mRNA stability.

This thesis provides tools to identify, isolate, and characterize new glycoside hydrolases, and to change them into more efficient and specific glycosynthases. Moreover, approaches are provided to study and increase thermostability of enzymes, and to increase the stability of mRNA in order to improve the expression of certain genes. In conclusion, the described toolbox includes different strategies for protein engineering in general, and for enzymes with potential for oligosaccharide synthesis in particular.

Samenvatting en Conclusies

De katalyse, thermostabiliteit, en productie van thermostabiele glycoside hydrolasen zijn onderzocht. Daarnaast is getracht bepaalde eigenschappen van deze enzymen te verbeteren waardoor ze beter geschikt zijn voor toepassing in processen waarin korte suiker ketens (oligosachariden) gesynthetiseerd worden, een proces dat efficiënter verloopt bij hogere temperaturen. Diverse strategieën en technieken zijn toegepast, waarvan enkele onconventioneel. Als model voor deze studie zijn een nieuw α -galactosidase (GalA), een laminarinase (LamA) en een β -glucosidase (CelB) van het hyperthermofiele archaeon *Pyrococcus furiosus* gebruikt, alsmede een lichenase (LicA) van de bacterie *Bacillus licheniformis*.

Als een algemene inleiding tot dit proefschrift wordt in **Hoofdstuk 1** een overzicht gegeven van glycoside hydrolasen, met specifieke aandacht voor die van hyperthermofiele origine. De classificatie en het reactiemechanisme van deze klasse van enzymen worden beschreven alsmede de methoden die worden gebruikt om deze biokatalysatoren te bestuderen. Een uitgebreide karakterisatie van één zo'n enzym, een α -galactosidase (GalA) uit *Pyrococcus furiosus*, wordt beschreven in **Hoofdstuk 2**. Het GalA gen (*galA*) werd geïdentificeerd, gekloneerd en functioneel tot expressie gebracht in *Escherichia coli*. GalA is een nieuw type α -galactosidase; de hoogste homologie werd gevonden met α -amylasen die geclassificeerd zijn in familie 57 van glycoside hydrolasen. GalA katalyseert specifiek de hydrolyse van *para*-nitrophenol- α -1,4-D-galactopyranoside, en tot op zekere hoogte die van melibiose en raffinose. Het enzym bleek een α -galactosidase te zijn met een extreem hoge thermo-activiteit (temperatuur optimum van 115 °C) en thermostabiliteit (halfwaardetijd van 15 uur bij 100 °C). Sequentie analyse heeft geleid tot de identificatie van vier geconserveerde carboxylaat residuen. Plaatsgerichte mutagenese werd toegepast om de mogelijke katalytische residuen te identificeren. Glu117Ala vertoonde een verlaagde enzymactiviteit, welke hersteld kon worden door toevoeging van azide of formiaat. Er wordt geconcludeerd dat glutamaat 117 het katalytische nucleofiel is, terwijl de zuur/base katalysator nog geïdentificeerd moet worden. Het belang van het identificeren van de katalytische residuen wordt geïllustreerd in **Hoofdstuk 3**, waarin de engineering beschreven wordt van het *Pyrococcus furiosus* laminarinase (LamA), een endo-glycosidase waarvan bekend is dat het β -1,3-gluco-oligosachariden hydrolyseert, maar niet in staat is tot afbraak van β -1,4-gluco-oligosachariden. De nucleofiel mutant Glu170Ala had een drastisch gereduceerde hydrolytische activiteit, maar was actief als een glycosynthase: de condensatie van α -laminaribiosyl fluoride en verschillende acceptoren resulteerde in oligosachariden met een opbrengst tot 30%. Afhankelijk van de acceptor, genereerde het synthase hetzij een β -1,3 danwel een β -1,4 binding. In hetzelfde hoofdstuk is de specificiteit van het wild-type hydrolase met betrekking tot kleine sachariden onderzocht, waarbij gebruik gemaakt werd van 4-methylumbelliferyl β -glucosiden met verschillende bindingen. Naast endo-activiteit is ook gedemonstreerd dat het wild-type LamA enige exo-activiteit heeft, en de potentie om oligosachariden te hydrolyseren met zowel β -1,4 als β -1,3 specificiteit.

In **Hoofdstuk 4** worden bepaalde details van de stabiliteit en tertiaire structuur conformatie van LamA beschreven. In 7.9 M GdmCl wordt aangetoond dat de toevoeging van calcium aan het enzym significante veranderingen veroorzaakt in circulair dichroïsme (CD) en fluorescentie

spectra, wat een kleine maar significante verandering suggereert van de tertiaire structuur t.o.v. de natieve toestand. Het vermogen tot interactie met calcium in 7.9 M GdmCl, met een consistent herstel van natieve tertiaire structuur is een unieke eigenschap van dit extreem stabiele endo- β -1,3-glucanase. Het effect van calcium op de thermodynamische parameters, relatief ten opzichte van het GdmCl-geïnduceerde evenwicht, is geanalyseerd door middel van circulair dichroïsme en fluorescentie spectroscopie. De interactie van calcium met de natieve vorm van het enzym werd bestudeerd aan de hand van Fourier getransformeerde infrarood (FTIR) spectra. Van het enzym werd een op homologie gebaseerd model gegenereerd welke gebruikt werd om de vermoedelijke bindingsplaats(en) van calcium te voorspellen, alsmede de structurele interacties die mogelijk verantwoordelijk zijn voor de ongewone stabiliteit van dit eiwit, vergeleken met andere familie 16 glycoside hydrolasen.

In verschillende industriële toepassingen is immobilisatie van het enzym gewenst. De structurele eigenschappen en activiteit van LamA na immobilisatie op Teflon en Silica zijn beschreven in **Hoofdstuk 5**. De materialen waren zo ontworpen dat de interactie van de deeltjes met licht te verwaarlozen was en de enzymconformatie in de geadsorbeerde toestand *in situ* gecontroleerd kon worden. De adsorptie isothermen werden bepaald en het geadsorbeerde endoglucanase werd bestudeerd door middel van diverse spectroscopische technieken, enzym activiteit assays, en dynamic light scattering. Er werd aangetoond dat het enzym met hogere affiniteit adsorbeerde aan het hydrofobe oppervlak van Teflon dan aan het hydrofiele oppervlak van Silica nanodeeltjes. In alle gevallen resulteerde de adsorptie in kleine veranderingen in de secundaire structuur. De veranderingen waren intenser bij de adsorptie aan Teflon. De activiteit van het geadsorbeerde enzym vertoonde dezelfde pH-afhankelijkheid als het enzym vrij in oplossing, en het bleef actief in de geadsorbeerde toestand ondanks de structurele veranderingen die geïnduceerd zijn wanneer interactie met de oppervlakken plaatsvindt.

Om vernieuwende manieren voor enzym stabilisatie te onderzoeken, is een recent beschreven methode gebruikt voor eiwit engineering, wat gepresenteerd wordt in **Hoofdstuk 6**. De aanpak is gebaseerd op het natuurlijke proces “intein-driven protein splicing”. Een variatie op dit thema maakt generatie mogelijk van een covalente binding tussen de N-terminus en de C-terminus van een polypeptide keten. De methode is toegepast om circulaire varianten te creëren van LicA, een lichenase (endo- β -1,3-1,4-glucanase) van *Bacillus licheniformis*. Twee cyclische varianten vertoonden een significant hogere thermische stabiliteit dan de lineaire variant, terwijl de catalytische activiteit behouden bleef. Na 3 minuten bij 65 °C had het lineaire glucanase de helft van zijn activiteit verloren, terwijl de twee cyclische varianten een 6-voudige en 16-voudige verhoging van de halfwaardetijden vertoonden voor de activiteit bij deze temperatuur. Bij verhoging van de temperatuur, liet fluorescentie spectroscopie zien dat de cyclische polypeptiden significant later begonnen te ontvouwen dan de lineaire, met een verhoging van het schijnbare smeltpunt van 2 °C.

Tot slot, in **Hoofdstuk 7**, zijn pogingen ondernomen om boodschapper RNA (mRNA) te stabiliseren door middel van cyclisatie, en is het effect hiervan op de efficiëntie van heterologe expressie geanalyseerd. In sommige gevallen kan de instabiliteit van prokaryoot mRNA resulteren in een verlaagde totale opbrengst van gen expressie. Een beter inzicht in de kenmerken die mRNA

stabiliteit bepalen, en de mogelijkheid om stabiele boodschapper varianten te construeren, zou kunnen bijdragen aan optimalisatie van eiwit productie door microbiële Cell Factories. Een serie van constructen is gegenereerd die een reporter (β -glucosidase) en een ribozym (self-splicing intron) bevatten. Deze werden gebruikt om *E. coli* te transformeren. De constructen werden zodanig ontworpen dat ribozym-gekatalyseerde mRNA processing resulteerde in cyclisatie van het RNA molecuul met als consequentie de re-ligatie van verbroken β -glucosidase fragmenten. De translatie van dit cyclische mRNA die hierop volgt resulteerde in een functioneel β -glucosidase, zoals werd aangetoond middels activiteits screenings na groei op zowel agar platen, als in vloeibare culturen. Bovendien werd de aanwezigheid van covalent gesloten circulair mRNA bevestigd door analyse van het totale RNA van deze recombinante culturen met behulp van RT-PCR. Ook transcriptie *in vitro* resulteerde in cyclisch mRNA, dat in tegenstelling tot lineair mRNA niet gevoelig bleek voor hydrolyse door een exonuclease. Verschillende variaties in het ontwerp zijn getest, strevend naar een optimale splicing, ligatie, en vervolgens transformatie van het cyclische transcript. Hoewel de huidige cyclische varianten geen verbeterde eiwit productie tot stand bracht vergeleken met het lineaire construct, liet deze uitvoering van engineering een potentiële toepassing zien vanwege zijn verbetering van mRNA stabiliteit.

Dit proefschrift verschaft gereedschappen voor het identificeren, isoleren, en karakteriseren van nieuwe glycoside hydrolasen, en voor het veranderen van deze enzymen in meer efficiënte en specifieke glycosynthasen. Verder worden een aantal manieren beschreven om de thermostabiliteit van enzymen te bestuderen en te verhogen, en om de stabiliteit van mRNA te verhogen om de expressie van sommige genen te verbeteren. Concluderend, de beschreven gereedschapskist bevat verschillende strategieën voor eiwit engineering in het algemeen, en voor enzymen met potentie voor oligosacharide synthese in het bijzonder.

Dankwoord/Acknowledgments

Voor velen leek het onmogelijk, maar het is nu dan toch echt waarheid geworden: een proefschrift in je handen van Johan van Lieshout. Ik kan het zelf ook nog moeilijk bevatten, zeker omdat op het moment dat ik dit schrijf het nog niet zover is. Het was een hele opgave, die nooit volbracht had kunnen worden door mij alleen. Een heleboel mensen om mij heen hebben me op één of andere manier geholpen om dit voor elkaar te krijgen. Op deze plaats wil ik al diegenen bedanken.

Ten eerste jou, John, voor je onophoudelijke enthousiasme. Alleen al daarom ben ik ooit bij je begonnen als AIO. Ik heb een erg plezierige tijd gehad in jouw groep. Geweldig hoe je ons allemaal leidt. Bovendien zit je altijd vol fantastische ideeën, zonder welke dit boekje een paar hoofdstukken had moeten missen. Naast je motivatie ben ik je ook nog dankbaar voor het geduld dat je voor mij hebt gehad in de laatste (lange) fase, het afronden van mijn proefschrift. Het gaat lukken!

Willem, naast John, bedankt dat je me de mogelijkheid hebt gegeven om bij je te promoveren. We hebben elkaar niet zo heel vaak gesproken, maar je was altijd volledig op de hoogte van mijn onderzoek, wat ik altijd heb weten te waarderen. Daarnaast zijn je interesse en je invloed op de achtergrond niet onopgemerkt gebleven.

Zoals eerder gezegd heb ik met veel plezier aan mijn promotie gewerkt in de groep Bacteriële Genetica. Ik ben hiervoor vele BacGenners dank verschuldigd. Het begon al met Joyce, het vervolgpacticum dat jij begeleidde bracht me op de BacGen weg. De weg die volgde was er een vol met hilarische congresbezoeken in het buitenland, saté-menu's in de Horizon, sportieve uitdagingen (Elfstedentocht fietsen, Veluweloop, Brug-tot-Brug), de Tour in de collegezaal, Quake, Sudden Death, en nog veel meer. En vooral veel lol en gezelligheid in het lab en de AIO kamers. Bedankt allemaal, Wilfried, Hauke, Joyce, Thijs K. (voor het delen van je kamer, zowel thuis als tijdens congressen (dit geldt overigens voor meerderen in dit lijstje)), Leon (voor alle hilariteiten, uitbreiding van mijn woordenschat (Tarrel), en niet te vergeten de bijnaam (De Beul van Bakel) die je me gaf), Don, Gael, Ineke, Pino, Ana (sjongejonge), Arjen (hoe heb je het kunnen uithouden met Corné en mij in dezelfde kamer?), Corné (wat een atmosfeer in de kamer!). Ans, wat moeten we zonder jou? Soms was je de enige man tussen al die jongens. Bedankt voor het in stand houden van BacGen. Evenals Servé (en voor de ritjes naar Nijmegen). Verder uiteraard de generatie die na mij is begonnen, Thijs E. en Krisztina (en vóór mij zijn geeindigd), Stan (gaan we nog een keer voor 'n piccolo in de Moonlight Bar?), Jasper (Status Quo op de SDS-gel schudder, geweldig!), Odette (nice fellowship!), Suzanne, Ronnie (bedankt voor je datum), Harmen, Hao, en alle nieuwe AIO's. Daarnaast zijn er nog vele fellows, guests en (PGO- en afstudeer-) studenten geweest die de sfeer bepaalden in het lab. Met name mijn studenten Sjaak, Bram en Wietse moet ik bedanken voor hun inzet en enthousiasme.

Naast BacGen ben ik het gehele Laboratorium voor Microbiologie dankbaar voor de gezelligheid in en om het gebouw. De Moleco's (nu nog steeds, maar dan anders...) en MicFysers. Veel dank ben ik ook verschuldigd aan de vaste staf, Francis, Nees, Wim, Renee, Jannie, en Ria. Jullie hielden de boel draaiende.

Dit proefschrift heeft ook tot stand kunnen komen door diverse samenwerkingsverbanden. In eerste instantie binnen het NWO-project zelf. Dank gaat uit naar Marieke, Anja, Remko, en Jolanda, Chantal, Ben, Henk, Jean-Paul en Fons. Ook wil ik alle leden van de begeleidingscommissie van het project bedanken voor hun interesse, feedback en input. Verder heb ik fijn samengewerkt met Fysische chemie en Kolloïdchemie, welke geleid heeft tot een aantal publikaties. Bedankt Sotiris en Willem. Thanks also to Roberta and Valerio, for a nice collaboration on LamA. Toni, thank you for having me in your lab for 4 weeks. It has been a wonderful experience. Thanks also to all the people there at the IQS for a pleasant stay! And, of course, Magda, thank you so much for everything! It wouldn't have been the same without you.

Inmiddels ben ik alweer een tijdje bezig op een nieuwe werkplek, maar de OLVM houden een goede traditie in stand met de OLVMTP. Thanks dudes! En bedankt Ron, voor die nieuwe werkplek, terwijl ik nog niet gepromoveerd was. En dat ik de mogelijkheid heb gekregen om het af te maken. (Oud-) collega's in Amsterdam wil ik ook graag bedanken: Teunie, Louis, Zulfı, Rokus, Tillman, Vic, Gerrit-Jan, Petra en Britta. En ook Martijn, Floris, Martin, Dick en Marcel. Het is fijn samenwerken met jullie.

Paranimfen Jasper en Corné, bedankt dat jullie me bij willen staan. Jasper, ontzettend bedankt voor de tijd en moeite die je hebt genomen in de laatste fase van het drukklaar maken van mijn boekje. Hauke, dank je wel voor je waardevolle adviezen en snelle commentaar op delen van mijn proefschrift.

Familie en vrienden. Zonder jullie interesse en motivatie was het me nooit gelukt en hadden jullie dit boekje nu niet in handen gehad. Ontzettend bedankt voor jullie steun en toeverlaat!! Irene en Bas, een tijdje uit het oog, maar niet uit het hart. Jullie zijn heel erg belangrijk voor mij geweest in al die jaren. Ik mis jullie. Evenals pa en ma Kövi en de rest van de familie. Bedankt voor jullie liefde. Hélène, Suzanne, Andries, Eddy, en de kinderen, en vooral papa en mama, bedankt voor de steun, interesse en liefde die jullie me hebben gegeven. Het is fijn om zo'n thuisfront te hebben. Eindelijk zien jullie nu het resultaat van al die jaren.

Odette, thanks a lot for all your love and support. And for the way you were in the stressfull periods of this thesis. It's great to have you by my side. I hope I can mean as much to you in two years from now, when you will be finishing your thesis, as you meant to me now. Huge kiss!

Johan
Januari 2007

Curriculum vitae

Johannes Franciscus Thomas werd geboren op 18 oktober 1972 te Helmond. Hij groeide op in Bakel, waar hij de peuterspeelzaal, kleuterschool en basisschool doorliep. Na zes jaar op en neer fietsen naar Deurne, behaalde hij aldaar in 1991 het VWO diploma aan het St. Willibrord Gymnasium. In september van datzelfde jaar begon hij aan de studie Bioprocestechnologie aan de Landbouw Universiteit Wageningen. Hij specialiseerde zich in de cellulaire en moleculaire richting middels twee afstudeervakken en een stage. Als eerste werd onderzoek verricht aan de afbraak van de kankerverwekkende stof benzo-(a)-pyreen in grond door de wit-rot schimmel *Bjerkandera*, aan de vakgroep Industriële Microbiologie, onder begeleiding van Ir. Michiel Kotterman, Prof. Dr. Ir. Jan de Bont, en Dr. Jim Field. Het vervolgpacticum Microbiologie wekte bij hem de interesse op in het onderzoek van de groep Bacteriële Genetica van Dr. John van der Oost in het Laboratorium van Microbiologie, onder leiding van Prof. Dr. Willem M. de Vos. Het tweede afstudeervak werd dan ook gedaan in die groep. Onder enthousiaste begeleiding van Ir. Joyce Lebbink werd door middel van mutagenese de thermostabiliteit bestudeerd van het enzym glutamate dehydrogenase uit *Thermotoga maritima*. Vervolgens ging hij op stage naar Heidelberg in Duitsland. Op het EMBL aldaar, in de groep van Prof. Dr. Matti Saraste, werkte hij, begeleid door Elena Baraldi, MSc., aan verschillende klonerings-en expressie-technieken om de interactie tussen twee eiwitdomeinen te onderzoeken. Vervolgens studeerde Johan in maart 1999 af en werd aansluitend aangesteld als assistent in opleiding aan het Laboratorium voor Microbiologie van Wageningen Universiteit. In de eerder genoemde groep van Dr. John van der Oost werd het promotieonderzoek uitgevoerd als onderdeel van NWO-project WLM.4479, getiteld: "Biocatalytic production of specific oligosaccharides". De resultaten van dit onderzoek zijn beschreven in dit proefschrift. Sinds september 2004 werkt Johan als Post-Doc voor de Universiteit van Amsterdam. In de Biokatalyse groep van Prof. Dr. Ron Wever verricht hij onderzoek voor het IBOS-project 053.63.003, getiteld: "Use of sulfatases in the production of sulfated carbohydrates and steroids".

List of publications

- Kaper, T., Verhees, C.H., Lebbink, J.H.G., Van Lieshout, J.F.T., Kluskens, L.D., Ward, D.E., Brinkman, A.B., Kengen, S.W.M., de Vos, W.M., Van der Oost, J. (2001) Characterization of beta-glycosyl hydrolases from *Pyrococcus furiosus*. Meth. Enzymol. 330, 329-346
- Van Lieshout, J.F.T., Verhees, C.H., Ettema, T.J., Van der Sar, S., Imamura, H., Matsuzawa, H., Van der Oost, J., De Vos, W.M. (2003) Identification and molecular characterization of a novel type of alpha-galactosidase from *Pyrococcus furiosus*. Biocatalysis and Biotransformation 21, 243-252
- Bruins, M.E., Strubel, M., van Lieshout, J.F.T., Janssen, A.E.M., Boom, R.M. (2003) Oligosaccharide synthesis by the hyperthermostable β -glucosidase from *Pyrococcus furiosus*: kinetics and modeling. Enzyme Microb. Technol. 33, 3-11
- Van Lieshout, J.F.T., Faijes, M., Nieto, J., Van der Oost, J., Planas, A. (2004) Hydrolase and glycosynthase activity by endo-1,3- β -glucanase from the thermophile *Pyrococcus furiosus*. Archaea 1, 285-292
- Chiaraluce, R., Gianese, G., Angelaccio, S., Florio, R., Van Lieshout, J.F.T., Van der Oost, J., Consalvi, V. (2005) Calcium-induced Tertiary Structure Modifications of Endo-beta-1,3-glucanase from *Pyrococcus furiosus* in 7.9 M Guanidinium Chloride. Biochem. J. 386, 515-524
- Koutsopoulos, S., Tjeerdsma, A.-M., Van Lieshout, J.F.T., Van der Oost, J., Norde, W. (2005) Adsorption induced changes in structure and activity of an endoglucanase from the hyperthermophilic Archaeon *Pyrococcus furiosus*. Biomacromolecules 6, 1176-1184
- Chiaraluce, R., Florio, R., Angelaccio, S., Gianese, G., Van Lieshout, J. F. T., Van der Oost, J., and Consalvi, V. (2006) Role of Glu-53 and Asp-287 in *Pyrococcus furiosus* endo- β -1,3-glucanase thermodynamic stability and in calcium binding in 7.9 M guanidinium chloride (submitted)
- Van Lieshout, J.F.T., N. Pérez Gutiérrez, O., Vroom, W., Koutsopoulos, S., Planas, A., De Vos, W.M., Van der Oost, J. (2006) Stabilization of an endo- β -1,3-1,4-glucanase by Cyclization. (submitted)
- Van Lieshout, J.F.T., Vroom, W., De Vos, W.M., Van der Oost, J. (2006) Ribozyme-mediated engineering of circular mRNA and its functional *in vivo* and *in vitro* translation (submitted)
- Koutsopoulos, S., Van Lieshout, J.F.T., Ippel, H., Van der Oost, J., Norde, W., Van Mierlo C.P.M. (2006) A dynamic NMR study of a hyperthermostable enzyme reveals unique structural features. (in prep.)

The work presented in this thesis was performed at the Laboratory of Microbiology of Wageningen University, Wageningen, The Netherlands, and was financially supported by the Technology Foundation (STW), applied science division of the Netherlands Organization for Scientific Research (NWO) and the technology program of the Ministry of Economic Affairs, The Netherlands, with project number WLM.4479.

Printing Grafisch bedrijf Ponsen & Looijen B.V., Wageningen, The Netherlands

Cover design Jasper Akerboom en Johan van Lieshout
Structural model of a circular endo- β -1,3-1,4-glucanase

