DOI: 10.1002/cbic.200700545

# Characterization of a New Glycosynthase Cloned by Using Chemical Complementation

Haiyan Tao,[a, b] Pamela Peralta-Yahya,[a] John Decatur,[a] and Virginia W. Cornish*[a]

Despite their fundamental role in biological processes and potential use as therapeutics, it still remains difficult to synthesize carbohydrates. In the past two decades, there has been tremendous progress in the chemical synthesis of complex carbohydrates.[1–7] However, chemical synthesis is still limited by the need for differentially protected intermediates and reactant-dependent coupling yields and stereocontrol. Enzymes, with their control of both regio- and stereochemistry, provide an obvious alternative to traditional small-molecule chemistry for the synthesis of oligosaccharides.[8,9]

Recently, Withers and co-workers demonstrated that retaining glycosidases can be engineered to glycosynthases simply by mutating the nucleophilic Glu residue at the base of the active site to a small hydrophobic residue and using an α-glycosyl fluoride as the donor substrate.[10] This strategy is based on extensive characterization of the mechanism of retaining glycosidases.[10,11] Mutation of the active-site nucleophile to a small residue both accommodates the glycosyl fluoride donor and inactivates the hydrolytic activity of the enzyme, allowing the reaction to proceed in the reverse direction. This approach was first demonstrated with the *Agrobacterium* sp. β-glucosidase/galactosidase (Abg).[10] The active-site nucleophile Glu358 was mutated to Ala. This Abg:E358A variant was shown to accept both galactosyl fluoride and glucosyl fluoride as donors to form glycosidic bonds with several mono- and disaccharides. This result opened a new route for carbohydrate synthesis, and already several retaining glycosidases have been successfully converted to glycosynthases with this strategy.[12–23] Directed evolution would offer an obvious route to improve the activity and alter the substrate selectivity of these enzymes, except that there is no intrinsic way to screen or select for glycosynthase activity. Mayer et al. developed a coupled enzyme assay using an *endo*-cellullase that can be used to screen for glycosynthase mutants with improved activity.[24] This screen, however, can only be used for glycosynthases that synthesize products that are substrates of the *endo*-cellulase. Screens only allow relatively small libraries, about $10^4$ variants, to be assayed.

Thus, our laboratory applied "chemical complementation", a general, high-throughput assay for enzyme catalysis of bond formation and cleavage reactions, to the directed evolution of glycosynthases.[25–27] In chemical complementation, glycosynthase activity is linked to reporter gene transcription and hence cell survival through covalent coupling of a methotrexate(Mtx)-disaccharide-fluoride donor and a dexamethasone-(Dex)-disaccharide acceptor, such that Dex-tetrasaccharide-Mtx effectively reconstitutes the transcriptional activator and increases transcription of a downstream reporter gene.[26] Use of the reporter gene *LEU2* allows for a growth selection in the absence of leucine (Figure 1). Using the *LEU2* selection, we previously demonstrated that chemical complementation can be used to read-out glycosynthase activity, and a *Humicola insolens* Cel7B:E197S variant with a fivefold increase in glycosynthase activity was selected from a Glu197 saturation library.[26]

Having established chemical complementation as a selection for glycosynthase activity, we then sought a glycosynthase that would provide a robust scaffold for the directed evolution of glycosynthase variants with altered substrate specificities. In our initial publication, the *H. insolens* CeL7B:E197A glycosynthase was employed because it was the only reported *endo*-glycosynthase at that time. However, this enzyme has poor expression properties and does not present an obvious scaffold for protein engineering.[26] Thus, we sought to develop an *endo*-glycosynthase derived from a family 5 glycosidase.[28,29] The in vitro activities and substrate specificities of many family 5 glycosidases have been extensively characterized, and several of these enzymes have shown good expression in *E. coli*. Moreover, family 5 retaining glycosidases are monomeric triose-phosphate isomerase (TIM) barrel enzymes, an appealing scaffold for enzyme engineering given that TIM barrels arguably are a "privileged" scaffold for enzyme catalysis of diverse chemical transformations.[30–34] In this paper, we report the cloning and characterization of a new glycosynthase from a family 5 glycosidase using a chemical complementation *LEU2* enrichment assay.

Given that not all retaining glycosidases provide efficient glycosynthases upon mutation of the active-site nucleophile, we adapted our *LEU2* selection as an enrichment assay to clone the new glycosynthase.[27] Using this assay, the family 5 TIM-barrel glycosynthase was cloned by screening the active-site E:G and E:S variants of known family 5 glycosidases. Specifically, three family 5 glycosidases, *Clostridium cellulolyticum* Cel5A, *Clostridium thermocellum* CelG and *Clostridium cellulolyticum* Cel5N, were tested.[35–38] Among these genes, Cel5A and CelG have been overexpressed and purified from *E. coli*, while the catalytic domains of CelG and Cel5N share high sequence identity. The high-resolution structure of the Cel5A catalytic domain shows a classic TIM barrel fold, with Glu170 as the

[a] *Dr. H. Tao, P. Peralta-Yahya, Dr. J. Decatur, Prof. V. W. Cornish*
*Department of Chemistry, Columbia University*
*New York, NY 10027 (USA)*
*Fax: (+1) 212-932-1289*
*E-mail: vc114@columbia.edu*

[b] *Dr. H. Tao*
*Current address: Genomics Institute of the Novartis Research Foundation*
*San Diego, CA 92121 (USA)*

*Supporting information for this article is available on the WWW under http://www.chembiochem.org or from the author: Experimental procedures and NMR data.*

  

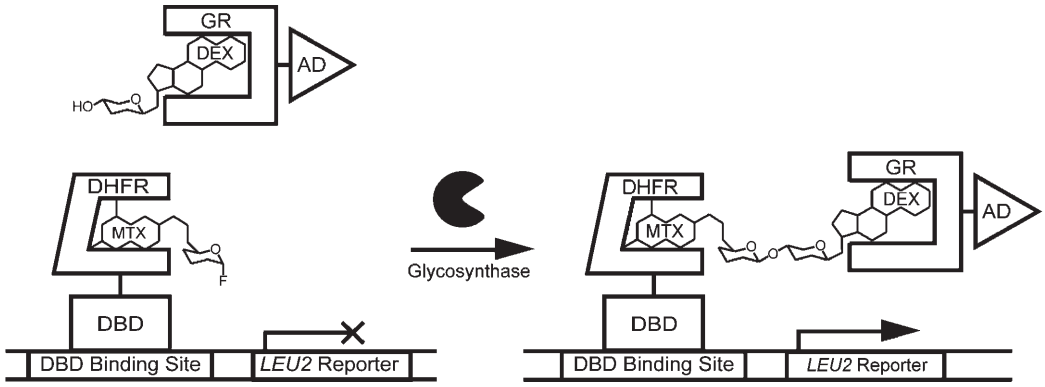


**Figure 1.** Chemical complementation provides a high-throughput assay for glycosynthase activity. The heterodimeric small molecule reconstitutes a transcriptional activator, turning on the transcription of a downstream reporter gene. Here, a Dex-Mtx yeast three-hybrid system is used. Glycosynthase activity is detected as formation of a glycosidic linkage between a Mtx fluoride donor and a Dex acceptor. Dex = dexamethasone; Mtx = methotrexate; DBD = DNA binding domain; DHFR = dihydrofolate reductase; GR = glucocorticoid receptor; AD = activation domain.

general acid/base and Glu307 as the active-site nucleophile.[35,39] These two Glu residues are separated by 5.5 Å, as expected for a glycosidase following a retaining mechanism. Since there are no high-resolution structures available for CelG and Cel5N, the active-site residues of these enzymes were determined using sequence alignment with well-characterized enzymes in the same family.[34,40] For both enzymes, Glu190 was assigned as the general acid/base, and Glu345 was assigned as the nucleophile, when signal peptides were excluded. Previous literature showed that some retaining glycosidases can be converted to glycosynthases by mutation of the active-site nucleophile to a Gly, Ser or Ala residue, with higher glycosynthase activities typically observed for Gly or Ser mutants.[24,41] Therefore, we chose to screen the Gly and/or Ser variants of the three family 5 glycosidases for glycosynthase activity, and these variants were expressed in yeast three-hybrid strain V1019Y under either a *MET25* or a *GAL1* promoter.[42] While these variants also could have been purified and assayed in vitro, we chose to use the chemical complementation selection for the cloning both because it could be carried out with slightly more ease and because it allowed us to assess the feasibility of using chemical complementation to clone enzymes based on function.

Having prepared the Gly and Ser variants from Cel5A and CelG, and Ser variants from Cel5N, we next determined if they had significant glycosynthase activity in what we term a chemical complementation enrichment assay. If active, the glycosynthase variant should be able to couple the two substrates, methotrexate-lactose-fluoride (Mtx-Lac-F) donor and dexamethasone-cellobioside (Dex-Cel) acceptor, synthesizing Dex-Cel-Lac-Mtx. Dex-Cel-Lac-Mtx should then dimerize the transcriptional activator, activating transcription of the *LEU2* reporter gene and providing a growth advantage to yeast three-hybrid strain V1019Y in the absence of leucine (Figure 1).[26,27] In the enrichment assay, a mixture of plasmids encoding a roughly 1:4 ratio of the glycosidase nucleophile mutant to the wild-type glycosidase was transformed en masse into yeast three-hybrid strain V1019Y. The resulting transformants were incubated in selective media lacking the appropriate auxotrophs and leucine and containing both the Mtx-Lac-F donor and the Dex-Cel acceptor substrates. After six days of incubation, all assays showed different extent of cell growth. A portion of culture was then taken from each assay tube, the cells were recovered in nonselective medium, and the plasmids were extracted and subjected to restriction analysis. The rest of cells had been further incubated in selective medium, and assayed at the ninth day of the selection. Since only cells expressing active glycosynthase get enriched during selection, the glycosynthase activity was linked to the enrichment of the plasmids encoding glycosynthase variants. As shown in Figure 2, the fold of enrichment can be conveniently visualized by restriction analysis. Since we chose the 1:4 starting ratio of the glycosidase nucleophile mutant to the wild-type glycosidase, we can observe the inverse of their intensity after a relatively small fold of enrich-

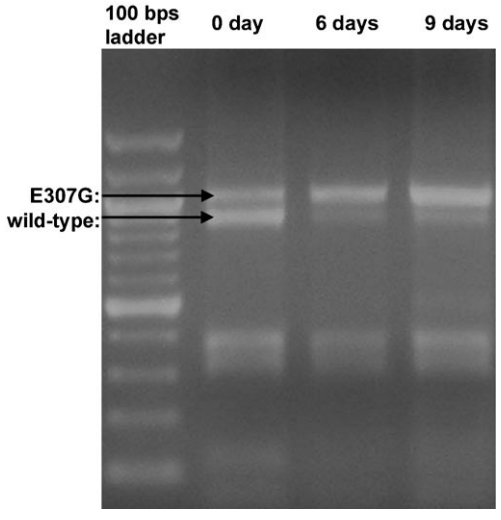


**Figure 2.** Cel5A:E307G was identified as a glycosynthase with the chemical complementation enrichment assay. Before and after enrichment assay, the plasmids were extracted from reaction, amplified by PCR, and digested using HphI and PstI restriction enzymes. Since the Cel5A wild-type has one more HphI site than the E307G mutant, the enrichment fold can be readily determined by restriction digestion.

  

ment. In a total of 10 enrichment assays the Cel5A:E307G mutant under the GAL1 promoter gave the best enrichment after six or nine days of incubation (Figure 2 and Figure S1 in the Supporting Informtaion).

Next the glycosynthase activity of Cel5A:E307G was confirmed in vitro using α-lactosyl fluoride (Lac-F) donor and *p*-nitrophenyl β-cellobioside (PNPC) acceptor substrates.[27,41] The Cel5A:E307G was overexpressed in *E. coli* from a P*tac* promoter and purified via a 6xHis tag.[37] The consumption of Lac-F was monitored with a fluoride ion-selective electrode. When all of the Lac-F was consumed, the reaction mixture was purified by reversed-phase HPLC, and the products were characterized by NMR and mass spectrometry. To our surprise, while Cel5A is a β-1,4-hydrolase, the Cel5A:E307G variant shows specificity for the synthesis of the β-1,3-Glu-Glu linkage, with no detectable β-1,4-Lac-Cel product based on HPLC purification and NMR characterization of the products (see the Supporting Information). The resulting tetrasaccharide can be further glycosylated by another α-lactosyl fluoride molecule, resulting in a hexasaccharide as a minor product. To understand whether or not the unexpected regioselectivity is a consequence of the 4′-axial hydroxyl group of the Lac-F donor substrate, we also assayed the Cel5A:E307G variant with α-cellobiose fluoride (Cel-F) as the donor substrate. Interestingly, the reaction of Cel-F and PNPC gave a mixture of tetrasaccharide regioisomers and other higher order oligosaccharide products (Figure 3). Further NMR analysis identified the presence of both Celβ1→3Celβ1→PNP and the Celβ1→4Celβ1→PNP tetrasaccharide products (see the Supporting Information). Therefore, the Cel5A:E307G glycosynthase does have different substrate specificity as its wild-type enzyme, and the specificity is related to the donor substrates used in the reaction. This phenomenon was also observed from previous glycosynthase references.[17,43] The Gly mutation increased the size of the active-site pocket and abolished the hydrogen-bond network derived from the Glu residue of the wild-type enzyme, thus allowed the substrates to bind in a more flexible manner.

In addition to the glycosyl transfer reaction, the Cel5A:E307G glycosynthase also catalyzes the hydrolysis of the Lac-F donor substrate, since only a portion of the PNPC is consumed when Lac-F is 100% consumed (Figure 3). For reaction of 10 mM Lac-F and 10 mM PNPC with 10 μM glycosynthase, 34% of PNPC was converted to tetrasaccharide, 22% to hexasaccharide, and 44% remained unreacted. The hydrolysis activity of Cel5A:E307G was confirmed using 2:1 Lac-F:PNPC (Figure 3). Considering that the hydrolysis activity might result from Cel5A wild-type contamination, we prepared the Cel5A:E307G enzyme from another *E. coli* single colony and analyzed glycosynthase activity with the HPLC assay. However, no improvement in reaction yield was observed (data not shown). The extent of hydrolysis with Cel-F as the substrate is ambiguous because the higher-order oligomerization occurs. This hydrolysis reaction can be treated as a glycosyl transfer reaction between the donor substrate and water.[41] For Cel5A:E307G glycosynthase with Lac-F as the donor substrate, the $k_{water}$ is slightly smaller than $k_{PNPC}$. Presumably the reaction efficacy, $k_{PNPC}/k_{water}$, could be improved using a classic directed-evolution approach, and chemical complementation provides a convenient selection assay for this approach. At last, the specific activity of Cel5A:E307G was measured based on the rate of expulsion of fluoride ion from the Lac-F donor substrate, with 10 mM Lac-F, 10 mM PNPC, and 10 μM enzyme in sodium phosphate buffer (pH 7). The Cel5A:E307G has a specific activity of 1.2 μmol [F-] per min per mg enzyme.

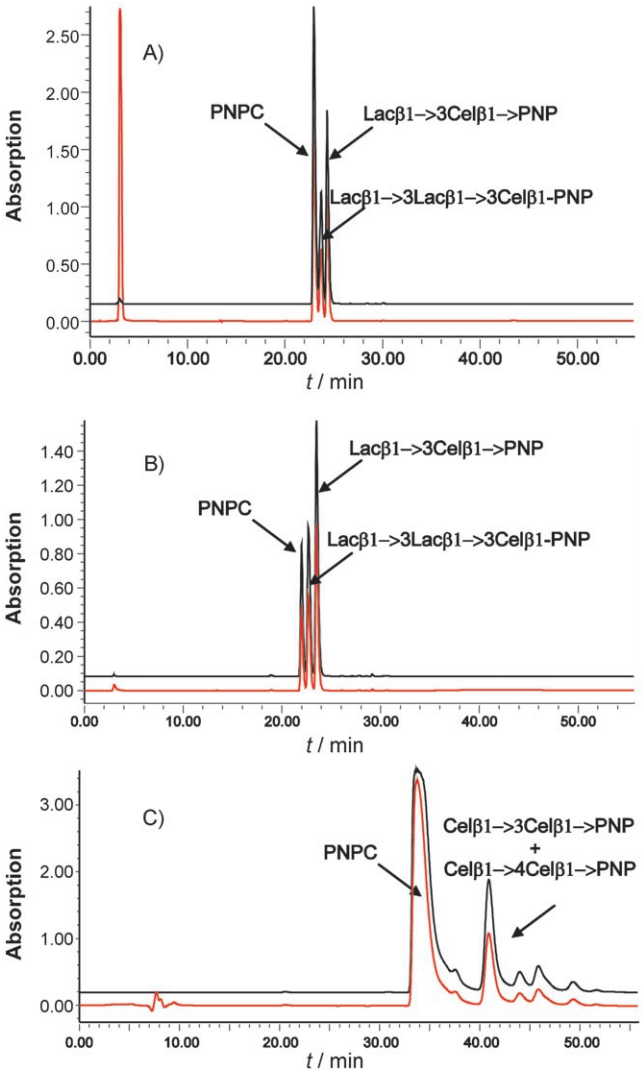


**Figure 3.** Analytical HPLC trace for the Cel5A:E307G catalyzed reactions. Reactions were monitored at 300 nm, as indicated by the black line, and at 220 nm, as indicated by the red line. HPLC was performed when donor substrates were completely consumed, as measured by the fluoride ion selective electrode. A) 10 mM Lac-F, 10 mM PNPC and 10 μM Cel5A:E307G; B) 10 mM Lac-F, 5 mM PNPC and 10 μM Cel5A:E307G; C) 10 mM Cel-F, 10 mM PNPC and 10 μM Cel5A:E307G.

In conclusion, we cloned a new glycosynthase Cel5A:E307G which catalyzes the efficient synthesis of an *endo*-β-1,3-glycosidic linkage between Lac-F donor and PNPC acceptor substrates. Although the specific activity of Cel5A:E307G glycosynthase is relatively low, the chemical complementation can be used as a selection assay to improve its activity. Cel5A:E307G has a "privileged" TIM barrel scaffold, which should facilitate directed-evolution experiments to improve the enzyme activity

  

and alter substrate specificity through mutation of the catalytic loops. Cel5A:E307G can be readily expressed and purified, thus not only providing a robust scaffold for directed evolution but also ultimately a good protein for overproduction for commercial use. Moreover, we demonstrated that the chemical-complementation *LEU2* enrichment assay provides an efficient method for screening glycosynthase activity and thus cloning enzymes based on function. Currently, we are using the *LEU2* selection for the directed evolution of Cel5A:E307G variants with altered substrate specificities for use in carbohydrate synthesis.

## Acknowledgements

*The authors thank Dr. Henri-Pierre Fierobe from Scientifique Institut de Biologie Structurale et Microbiologie-Institut Fédératif de Recherche, France, for providing the Cel5A genes and the* C. cellulolyticum *genomic DNA, and Dr. Pierre Beguin from Institut Pasteur, France, for providing the CelG genes. We thank Prof. Stephen Withers, University of British Columbia, for insightful comments on the manuscript. We thank the New York Structural Biology Center for use of their 800 MHz NMR. The Center is a STAR center supported by the New York State Office of Science, Technology, and Academic Research. We are grateful for financial support for this work from the National Science Foundation (CHE-0350 183) and National Institutes of Health (GM62 867).*

**Keywords:** chemical complementation · directed evolution · enzymes · glycosidases · glycosynthases · screening

[1] S. J. Danishefsky, K. F. McClure, J. T. Randolph, R. B. Ruggeri, *Science* **1993**, *260*, 1307.
[2] B. Fraser-Reid, U. E. Udodong, Z. Wu, H. Ottosson, J. R. Merritt, C. S. Rao, C. Roberts, R. Madsen, *Synlett* **1992**, 927.
[3] O. J. Plante, R. B. Andrade, P. H. Seeberger, *Org. Lett.* **1999**, *1*, 211.
[4] R. R. Schmidt, *Angew. Chem.* **1986**, *98*, 213; *Angew. Chem. Int. Ed. Engl.* **1986**, *25*, 212.
[5] P. H. Seeberger, S. J. Danishefsky, *Acc. Chem. Res.* **1998**, *31*, 685.
[6] K. Toshima, K. Tatsuta, *Chem. Rev.* **1993**, *93*, 1503.
[7] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem.* **2006**, *118*, 4222; *Angew. Chem. Int. Ed.* **2006**, *45*, 4116.
[8] A. M. Daines, B. A. Maltman, S. L. Flitsch, *Curr. Opin. Chem. Biol.* **2004**, *8*, 106.
[9] K. M. Koeller, C. H. Wong, *Nature* **2001**, *409*, 232.
[10] L. F. Mackenzie, Q. Wang, R. A. J. Warren, S. G. Withers, *J. Am. Chem. Soc.* **1998**, *120*, 5583.
[11] C. S. Rye, S. G. Withers, *Curr. Opin. Chem. Biol.* **2000**, *4*, 573.
[12] M. Faijes, M. Saura-Valls, X. Perez, M. Conti, A. Planas, *Carbohydr. Res.* **2006**, *341*, 2055.
[13] S. M. Hancock, M. D. Vaughan, S. G. Withers, *Curr. Opin. Chem. Biol.* **2006**, *10*, 509.
[14] Y. Honda, M. Kitaoka, *J. Biol. Chem.* **2006**, *281*, 1426.
[15] Y. W. Kim, D. T. Fox, O. Hekmat, T. Kantner, L. P. McIntosh, R. A. Warren, S. G. Withers, *Org. Biomol. Chem.* **2006**, *4*, 2025.
[16] J. Müllegger, H. M. Chen, R. A. Warren, S. G. Withers, *Angew. Chem.* **2006**, *118*, 2647; *Angew. Chem. Int. Ed.* **2006**, *45*, 2585.
[17] H. M. Salleh, J. Müllegger, S. P. Reid, W. Y. Chan, J. Hwang, R. A. J. Warren, S. G. Withers, *Carbohydr. Res.* **2006**, *341*, 49.
[18] M. Sugimura, M. Nishimoto, M. Kitaoka, *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1210.
[19] A. Trincone, A. Giordano, G. Perugino, M. Rossi, M. Moracci, *ChemBioChem* **2005**, *6*, 1431.
[20] J. van Lieshout, M. Faijes, J. Nieto, J. van der Oost, A. Planas, *Archaea* **2004**, *1*, 285.
[21] M. D. Vaughan, K. Johnson, S. DeFrees, X. Tang, R. A. Warren, S. G. Withers, *J. Am. Chem. Soc.* **2006**, *128*, 6300.
[22] D. L. Zechel, S. P. Reid, D. Stoll, O. Nashiru, R. A. Warren, S. G. Withers, *Biochemistry* **2003**, *42*, 7195.
[23] O. Nashiru, D. L. Zechel, D. Stoll, T. Mohammadzadeh, R. A. Warren, S. G. Withers, *Angew. Chem.* **2001**, *113*, 431; *Angew. Chem. Int. Ed.* **2001**, *40*, 417.
[24] C. Mayer, D. L. Jakeman, M. Mah, G. Karjala, L. Gal, R. A. Warren, S. G. Withers, *Chem. Biol.* **2001**, *8*, 437.
[25] K. Baker, C. Bleczinski, H. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, V. W. Cornish, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16537.
[26] H. Lin, H. Tao, V. W. Cornish, *J. Am. Chem. Soc.* **2004**, *126*, 15051.
[27] H. Tao, P. Peralta-Yahya, H. Lin, V. W. Cornish, *Bioorg. Med. Chem.* **2006**, *14*, 6940.
[28] B. Henrissat, *Biochem. J.* **1991**, *280 ( Pt 2)*, 309.
[29] B. Henrissat, A. Bairoch, *Biochem. J.* **1993**, *293 ( Pt 3)*, 781.
[30] P. M. Coutinho, B. Henrissat, *Carbohydrate-Active Enzymes: An Integrated Database Approach*, The Royal Society of Chemistry, Cambridge, **1999**.
[31] J. A. Gerlt, P. C. Babbitt, *Annu. Rev. Biochem.* **2001**, *70*, 209.
[32] J. A. Gerlt, F. M. Raushel, *Curr. Opin. Chem. Biol.* **2003**, *7*, 252.
[33] R. Sterner, B. Hocker, *Chem. Rev.* **2005**, *105*, 4038.
[34] CAZY: http://www.cazy.org
[35] V. Ducros, M. Czjzek, A. Belaich, C. Gaudin, H. P. Fierobe, J. P. Belaich, G. J. Davies, R. Haser, *Structure* **1995**, *3*, 939.
[36] H. P. Fierobe, C. Gaudin, A. Belaich, M. Loutfi, E. Faure, C. Bagnara, D. Baty, J. P. Belaich, *J. Bacteriol.* **1991**, *173*, 7956.
[37] H. P. Fierobe, A. Mechaly, C. Tardif, A. Belaich, R. Lamed, Y. Shoham, J. P. Belaich, E. A. Bayer, *J. Biol. Chem.* **2001**, *276*, 21257.
[38] M. Lemaire, P. Beguin, *J. Bacteriol.* **1993**, *175*, 3353.
[39] A. Belaich, H. P. Fierobe, D. Baty, B. Busetta, C. Bagnara-Tardif, C. Gaudin, J. P. Belaich, *J. Bacteriol.* **1992**, *174*, 4677.
[40] B. Henrissat, G. Davies, *Curr. Opin. Struct. Biol.* **1997**, *7*, 637.
[41] V. M. Ducros, C. A. Tarling, D. L. Zechel, A. M. Brzozowski, T. P. Frandsen, I. von Ossowski, M. Schulein, S. G. Withers, G. J. Davies, *Chem. Biol.* **2003**, *10*, 619.
[42] D. Mumberg, R. Muller, M. Funk, *Nucleic Acids Res.* **1994**, *22*, 5767.
[43] S. M. Hancock, M. D. Vaughan, S. G. Withers, *Curr. Opin. Chem. Biol.* **2006**, *10*, 509.

Received: September 11, 2007

  