

# A sequence analysis of the $\beta$ -glucosidase sub-family B

A. Rojas, A. Romeu\*

Universitat Rovira i Virgili, Department of Biochemistry and Biotechnology, E-43005 Tarragona, Spain

Received 24 October 1995

**Abstract** This computational study is a summary of structural properties of the  $\beta$ -glucosidase subfamily B. Computations were carried out using GCG package programs. All sequences used in this analysis were taken from the protein data bank. The multialignment and the phylogenetic tree of the  $\beta$ -glucosidase subfamily B are shown. The conserved patterns: DGP, GRNFE, DPYL, KHF, SDW, GLD, VLLKN in the N-terminal region and FGYGLSY in the C-terminal part should be pointed out. C-terminal parts of the *Butyrivibrio fibrisolvens* and *Ruminococcus albus*  $\beta$ -glucosidase sequences can be aligned to the N-terminal region of the other members of the subfamily. A crossed homology model in sub-family B  $\beta$ -glucosidases is described.

**Key words:**  $\beta$ -Glucosidase; Amino acid sequence analysis; Phylogenetic tree; Amino acid multialignment; Structural signature pattern; Computer analysis

## 1. Introduction

$\beta$ -Glucosidases (1,4- $\beta$ -D-glucoside glucohydrolases; EC 3.2.1.21) are defined as enzymes that hydrolyze compounds containing  $\beta$ -glucosidic linkages by splitting off the terminal, non-reducing  $\beta$ -D-glucose residues and releasing  $\beta$ -D-glucose [1]. His, Glu and Asp are the amino acids thought to be involved in enzymatic hydrolysis using an acid catalytic mechanism [2]. Because of their biotechnological importance, a number of  $\beta$ -glucosidase genes have been cloned and attempts at organizing  $\beta$ -glucosidase families and domains have been reported [3,4]. In a previous study we described a  $\beta$ -glucosidase family chart which was revealed by computer analysis of protein sequences [5]. Two sub-families A and B clearly exist. Sub-family A includes vegetal  $\beta$ -glucosidases and prokaryotic enzymes. While sub-family B groups rumen bacteria and fungal  $\beta$ -glucosidases. The aim of this work is to make an in depth study of the well conserved amino acid motifs and the protein identity of the amino acid sequences which are shared by the  $\beta$ -glucosidases of sub-family B. The phylogenetic arrangement of the  $\beta$ -glucosidases sequenced to date within the sub-family is also shown. This paper reports data that is part of a program for the study of enzymes which are involved in the degradation of plant material.

## 2. Materials and methods

All sequences used in this analysis were taken from Swiss Protein Resources (Swiss-Prot), Protein International Resources (PIR), Gen-

Bank and EMBL sequence data bank. Table 1 shows the organisms, gene, size (number of amino acids) and the individual access numbers of the cloned sub-family B  $\beta$ -glucosidases taken from the data bank. It should be noted that the *Schizophyllum commune* and *Aspergillus wentii* data correspond to only a part of their amino acid sequences. Multiple alignment was obtained using the PILEUP program that is available in the GCG software package version 8.0 (Genetic Computer Group, Inc. University of Wisconsin, Madison). Phylogenies of the aligned  $\beta$ -glucosidase sequences were estimated using the DISTANCES program. The phylogenetic tree was obtained by the GROWTREE program using the UPGMA method. A protein search for similarity between the query sequences was made by the FASTA program. Structural signature patterns were given by the PROSITE dictionary incorporated into GCG (Amos Bairoch, University of Geneva).

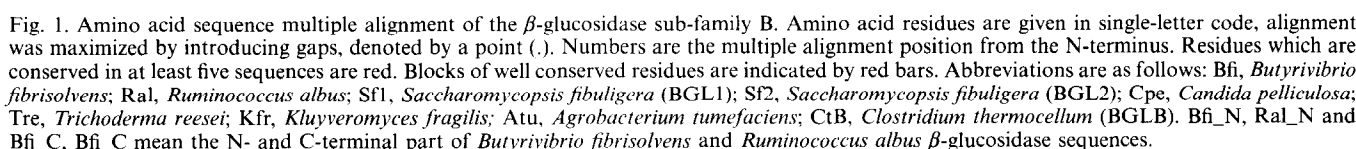
## 3. Results

Long tails were observed in the C-terminal region of the whole subfamily multialignment corresponding to the *Butyrivibrio fibrisolvens* and *Ruminococcus albus*  $\beta$ -glucosidases (results not shown). In order to characterize the structural basis of these bacterial enzymes, first, the C-tails were computationally cut, and two segments were obtained for each protein. Second, the two tails were individually submitted to the protein data base using the FASTA program. All the members of the sub-family B  $\beta$ -glucosidases were the only proteins which showed significantly high similarity scores. The C-tails matched the N-regions of these  $\beta$ -glucosidases. Third, as FASTA provides only individual comparisons, in an attempt to reveal new aspects of the global homology, both the C-tails and the N-terminal fragments of *Butyrivibrio fibrisolvens* and *Ruminococcus albus* enzymes were multiple aligned with the rest of the  $\beta$ -glucosidases using the Pileup program. Fig. 1 shows the amino acid alignment of seven of the more representative  $\beta$ -glucosidases of sub-family B (*S. fibuligera*, *T. reesei*, *C. pelliculosa*, *K. fragilis*, *A. tumefaciens*, and *C. thermocellum*) plus the N- and C-terminal fragments of *B. fibrisolvens* and *R. albus*. In this case the C-fragments also matched the N-terminal regions of the whole sub-family. A crossed homology model, based on a block-inversion of the *Butyrivibrio fibrisolvens* and *Ruminococcus albus* sequences is suggested for the sub-family B  $\beta$ -glucosidases. The conserved patterns: DGP, GRNFE, DPYL, KHF, SDW, GLD, VLLKN in the N-terminal region and FGYGLSY in the C-terminal part should be pointed out. Fig. 2 shows the phylogenetic tree of the  $\beta$ -glucosidase sub-family B. The fungal  $\beta$ -glucosidases such as *Schizophyllum commune*, *Trichoderma reesei* and *Clostridium thermocellum* are a branch of the phylogenetic tree.

## 4. Discussion

In the present work, a set of fourteen sub-family B  $\beta$ -glucosidases sequences were compared. On the basis of the sequence comparison using the iterative, pairwise multiple alignment

\*Corresponding author. Universitat Rovira i Virgili, Facultat de Química, Dept. Bioquímica i Biotecnologia, Pl. Imperial Tàrraco 1, E-43005 Tarragona, Spain. Fax: (34) (77) 559528.  
E-mail: romeu@quimica.urv.es



	601		700
Bfi_N	.....DKS VRFYQEYVAD QYRLGIAPGM IKEP.....	ALPEDILADA	
Ral_N	.....EDV LKAYKDYVAE H.PYDYGEGM GGEPWCQEEM PLDDSLVKRA		
Sf1	.....G SVGSPKYQVT PFEISYSLAR KNKMQFDYIR		
Sf2	.....G SVGYPKYQVT PFEISAMAR KNKMQFDYIR		
Cpe	.....G AGTFSYF.VT PADGIGARAQ QEKISYEFIG		
Tre	.....G AVNYPYF.VA PYDAINTRAS SQGTQVTLN		
Kfr	KHEKVDPKNP YFFVTLTGQY VPQEDGDYIF SLQVYGSGLF YLNDELIDQ KHNQERGSFC FGAGTKERTK KLTLLKGGQY NVRYVEYSGP TSLVGEFGA		
Atu	PSGDLDLAD .FSARMTATF VPQETGEHIF GNTNAGLARL FVDGELVVDG YDGTWKGENF FGTAENSEQR AVTLGAARRY RVVVEYEAPK AS...LDGINI		
Ctb	.....TRLDDI YEEIKKAGAD KVNLYVSEGY R.....		
Bfi_C	.....		
Ral_C	NFQSKLLRG RRAYR.....		
	701		800
Bfi_N	AAYADTAIAI ISRFSGEGND RKVAGVDREI KCEAKDLVEQ GNKIFDHGDF YLTNAEKKM KMVKENFSSV IVVMNVGGV DTTWFKKDDQ IS...SVLMAW		
Ral_N	AESSDTAICI IGRTAG.....EQ DNSC.KAGSY LLTDGEKAIL RKVRDNFSKM VILLNVGNII DMGFI...DE FSPDVMYVW		
Sf1	ESYDLAQVTK VASDAHLSIV VVSASGEGY ITVDGNQG...DRKNL TLWNGGKLI ETVAENCANT VVVTSTGQI NFEGFADHPN VT...AIVNAG		
Sf2	ESFDLTQVST VASDAHMSIV VVSASGEGY LIIDGNRG...DKNMV TLWHSNDNLI KAVAENCANT VVVTSTGQV DVESFADHPN VT...AIVNAG		
Cpe	DSWNQAAAMD SALYADAAIE VANSVAGEEI GDVDGNYG...DLNML TLWNNVAVPLI KNISSINNT IVIVTSQQI DLEPFIDNEN VT...AVIYSS		
Tre	TD.NTSGAS AARGKDAIV FITADSGDY ITVEGNAG...DRNML DPWHNGNALV QAVAGANSV IVVHVSVAI ILEQILALPQ VK...AVNAG		
Kfr	GGFQAGVKA IDDEEIRNA AELAAKHDKA VLIIGLNGEW ETEGYDRENM DLPRKTNELV RAVLKANPNT VIVNQSGTPV EFPWL...EE AN...ALVQAW		
Atu	CALRFGEVK L.GDAGIAEA VETARKSDIV LLLVGREGEW DTEGLDLPDM RLPGRQEELI EAVAETNPV VVVLQTGGPI EMPWL...GK VR...AVLQAW		
Ctb	.....LEND GIDEELINEA KKAASSDVA VVAGLPDEY ESEGFORTHM SIPENQRLI EAAVEQSNH VVLLNGSPV EMPWL...DK VK...SVLEAY		
Bfi_C	.....		
Ral_C	.....		
	801		900
Bfi_N	QGGIEGGLAA ARILLGKVN SGKLSDTFAA RLEDYPSLEG ....FHEDDD YVDYTEDIV GYRYFETIPG AKEKVNPFG YGLSYTFL EDYKAEPFVA		
Ral_N	QGGMTGGTG ARVLLGEVSP CGKLPDTIAY DITDYPDSKN ....FHNDDV DI.YAEDIFV GYRYFDTF... AKDRVRFPFG YGLSYTQF... ..		
Sf1	PLGDRSGTAI ANILFGKAMP SGHLPFTIAK TDDYIPIET YSPSSGEPED NHLVENLLV DYRYFEE... KNIEPRYAFG YGLSYNEYEV SNAKVSAAK		
Sf2	PLGDRSGTAI ANILFGKAMP SGHLPFTIAK SNDDYIPIVT YNPPNGEPED NTLAEDLLV DYRYFEE... KNIEPRYAFG YGLSYNEYEV SNAKVSAAK		
Cpe	YLQDFGTVL AKVLFGEENP SGKLPFTIAK DVNDYIPVIE ...KVDVDPD VDKFTESIYV DYRYFOK... YNKPVRFEFG YGLSYNFSL SDIEITQLP		
Tre	LPSQESGNAL VDVWGDVSP SGKLVYTIK SPNDYNTRI...VSGG SDSFSEGLFI DYKHFD... ANITPRYEFY YGL.YTKFNY SRLSVLSTAK		
Kfr	YGGNELGNAI ADVLYGDVVP NGKLSLSWPF KLQDMPAFLN ....FKTEFG RVVYGEDIFV GYRYYEK... LQRKVAFPFG YGLSYTTFEL DISDFKYTD		
Atu	YPQELGNAL ADVLFGDVEP AGRLPQTFPK ALTONSAITD DPSIYPGQDG HVRYAEGIFV GYRHOT... REIEPLFPFG FGLGYTRFTW GAPQLSGTEM		
Ctb	LGGQALGGRW .RMCYSVSKI VGKLAETFPV KLSHNPSTLN ....FPGEDD RVEYKEGLFV GYRYYOT... KGIEPLFPFG HGLSYTKFEY SDISVDKDV		
Bfi_C	.....		
Ral_C	.....		
	901		1000
Bfi_N	SAAD.....	.....EV GKSDSLADA IVASVTVTNI	
Ral_N	.....EI SAEGRKTDG VVITAKVKNI		
Sf1	VDEELPEPAT YLSEFSYQNA KDSKNPSDAF APADLNRVNE YLYPYLDSNV TLK.DGNYEY PDGYSTEQRT TP.NQPGGGL GGNDALWEVA YNSTOKFVPQ		
Sf2	VDEELPQKL YLAEYSYNT EEINNPEDAF FPSNARRIQE FLYPYLDSNV TLK.DGNYEY PDGYSTEQRT TP.IQPGGGL GGNDALWEVA YKVEVDVQL		
Cpe	FSENAEPAN YSETYQYQS N...MDPSEYT VPEGFKELAN YTYPIHDAS SIKANSSYDY PEGYSTEQLD GPKSLAAGGL GGNT...CG MLVTLSSLKS		
Tre	SGPAT.....	.....GAVVP GGPSDLFQNV ATVTVDIANS	
Kfr	...DK.....	.....IDISVDVNT	
Atu	GADG.....	.....LTVTVDTNI	
Ctb	SDNSI.....	.....INVSVKYKV	
Bfi_C	.....		
Ral_C	.....		
	1001		1100
Bfi_N	G.KIPGKEVV QLYYSAPQKG LGKPAKVLGG YAKTRLLQPG ESQRVTIALY MEDMASYDOL GKVKKA.AHL LEKGEYHFFL GTSVRQTRLL DYTVELSKNI		
Ral_N	G.SAAGKEVV QVYLEAPNCK LGKAARVLGG FEKTKVLAPN EQTLTIEVT ERDIASYYDS GITGNAFAMV EEAGEYTFYA GSDVRSK.E CFAFTLDSTK		
Sf1	GNS.TDKFVP QLYLKHPEDG KFETPIQLRG FEKVE.LSPG EKKTVDLRL LRRDLSVNDT TRQS...MI VESGTYEALI QVAVNDIKTS VLFTI.....		
Sf2	GNS.TDKFVP QLYLKHPEDG KFETPIQLRG FEKVE.LSPG EKKTVFEFL LRRDLSVNDT TRQS...MI VESGTYEALI QVAVNDIKTS VLFTI.....		
Cpe	QIK.VLMLVG .LHLNCLDI QIMMNSQHLQ CNYVD.LKRC FWIKIILKL FLLN.....		
Tre	QGV.TGAEVA QLYITYPSSA PRTPPKQLRG FAKLN.LTPG QSGTATFNI LRRDLSVNDT ASQK...MV VPSGSGISV GASSRDILRT STLVA....		
Kfr	GDKFAGSEVV QVYFSALNSK VSRPVKELKG FEKVH.LEPG EKKTVNIELE LKDAISYFNE ELGK...MH VEAGEYLVSV GTSSDDILSV KEFKVKEDLY		
Atu	GDR.AGSDVV QLYYHSPNAR VERPFKELRA FAKLK.LAPG ATGTAVLKIA PRD.LAYFDV EAGR...FR ADAGKYELIV AASAI DIRAS VSIHLPVDHV		
Ctb	G.KMAGKEIV QLYYKDVKSS VRRPEKELKG FEKVF.LNPG EEKTVTFTLD .KRAFAYYNT QIKD...MH VESGEFLILI GRSSRDIVLK ESVRVNSTVK		
Bfi_C	.....		
Ral_C	.....		
	1101		1182
Bfi_N	IVEQVSNKLV PTLSPKRM..		
Ral_N	VIEQLEQALA PVTPEKRM..		
Sf1	.....		
Sf2	.....		
Cpe	.....		
Tre	.....		
Kfr	WKGL.....		
Atu	MEP.....		
Ctb	IRKRFTVNSA VEDVMSDSSA AAVLGPVLKE ITDALQIDMD NAHDMAANI KNMPLRSLVG YSQGRLEEM LEELVDKINN VE		
Bfi_C	.....		
Ral_C	.....		

Fig. 1. (Contd.).

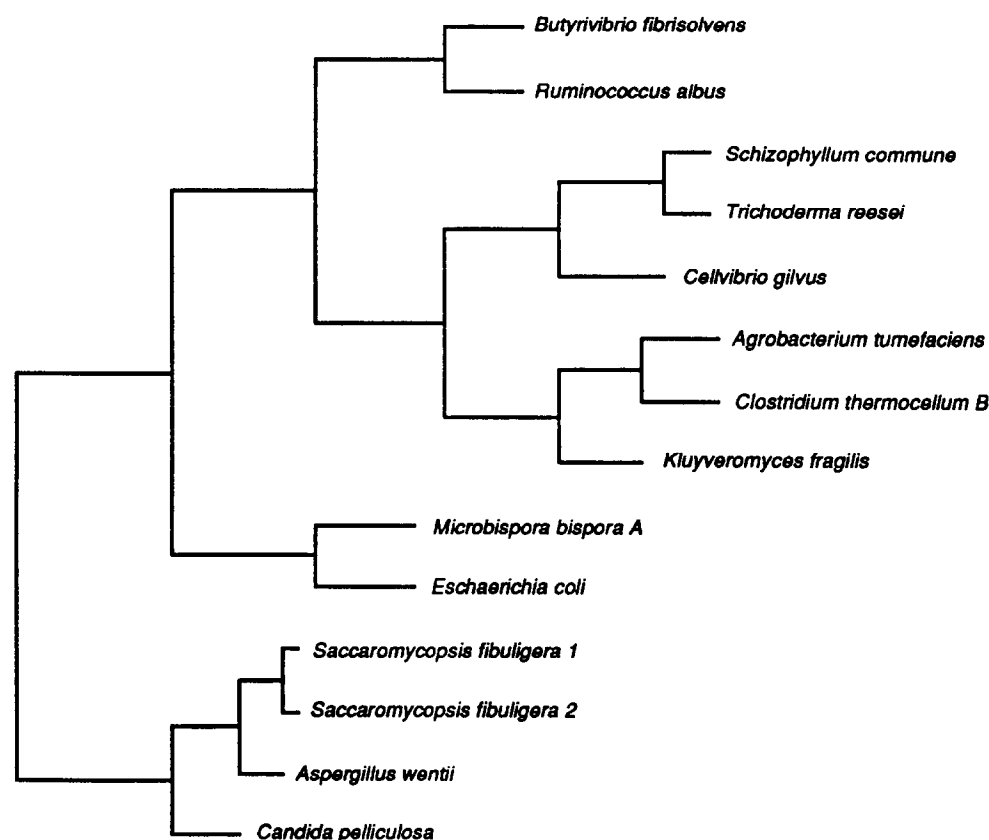


Fig. 2. The resulting phylogenetic tree of the  $\beta$ -glucosidase sub-family B according to the GROWTREE (GCG) program.

program PileUp (GCG), two domains showing similarity scores were observed in the N-terminal part, and in the C-terminal part, respectively. Consistently with the *B. fibrisolvens* and *R. albus* enzymes characteristics, it is remarkable that these anaerobic bacteria  $\beta$ -glucosidases are grouped together in the phylogenetic analysis. These results agree with the literature which reports similarities between *Butyrivibrio fibrisolvens* and *Kluyveromyces fragilis*, *Clostridium Thermocellum* and *Candida pelliculosa* sequences [6]. Of course the behaviour of *B. fibri-*

*solvens* and *R. albus*  $\beta$ -glucosidases requires attention. To the best of the author's knowledge, no attempt to address this problem has been made to date.

It is worth noting that the N-terminal region of the multiple alignment groups together the well known sequence motifs [2]. Legler et al. [21] observed that the conduritol B-epoxi inhibitor bound to the Asp residue of the SDW motif in *Aspergillus wentii*. The PROSITE data bank uses this region as a signature pattern. Based on mechanism studies with lysozyme [22], the

Table 1  
Summary of cloned  $\beta$ -glucosidases sub-family B

Organism	gene	Size (amino acids)	Protein data bank code			Reference
			SwissProt	GeneBank	PIR	
<i>Butyrivibrio fibrisolvens</i>	BGLA	830	P16084	M31120	A44768	6
<i>Ruminococcus albus</i>	BGLA	947	P15885	X15415	S08243	7
<i>Schizophyllum commune</i>	—	192*	P29091	—	A28571	8
<i>Trichoderma reesei</i>	BGL1	712	—	—	—	9
<i>Saccharomycopsis fibuligera</i>	BGL1	876	P22506	M22475	—	10
<i>Saccharomycopsis fibuligera</i>	BGL2	880	—	M22476	—	10
<i>Aspergillus wentii</i>	A3	63*	P29090	—	A29171	11
<i>Candida pelliculosa</i>	—	825	P06835	X02903	B23783	12
<i>Agrobacterium tumefaciens</i>	CBG1	818	P27034	M59852	A42292	13
<i>Kluyveromyces fragilis</i>	—	845	P07337	X05918	A29148	14
<i>Clostridium thermocellum</i>	BGLB	754	P14002	X15644	S04381	15
<i>Microbispora bispora</i>	BGLA	1020	—	L06134	—	16
<i>Cellvibrio gilvus</i>	—	854	—	—	—	17,18,19
<i>Escherichia coli</i>	BGLX	765	P33363	—	—	20

\*Partial sequence.

hydrolysis of  $\beta$ -glycosidic bonds is considered to proceed by general acid catalysis involving aspartic and glutamic acid residues. In the case of lysozyme, Glu-35 promotes catalysis by promoting the glycosyl oxygen, while the carboxylate form of Asp-52 stabilizes the glycosyl carbonium ion intermediate [22]. Conceivably, the aspartic amino acid residue of the SDW motif, as well as the conserved Glu residue nine amino acids upstream, could fulfill a function similar to Asp and Glu of lysozyme. Moreover, the protonated form of the strictly conserved His residue at position 249 within the KHF motif, might act as proton donor with Glu assisting in the stabilization of the cleavage intermediate. Another conserved sequence motif was found around asparagine which shows the putative glycosylation site consensus N(K,R)(S,T)N-X-S/T, where X is not Pro [9]. Alternatively, in view of their large size these enzymes may possess more than one active site. With respect to this, the Asp residue which falls in the VDYL motif has been described as a catalytic site in the *Schizophyllum commune*  $\beta$ -glucosidase [8].

From the data in this paper, it is concluded that the C-terminal parts of both *B. fibrisolvens* and *R. albus*  $\beta$ -glucosidases are homologous with the N-terminal region of the others members of the sub-family B and show the signature patterns of these N-regions. This evidence suggests that *B. fibrisolvens* and *R. albus*  $\beta$ -glucosidases have a block-inversion in their amino acid sequences.

**Acknowledgements:** We thank John Bates for his help in writing this article. This research was supported by grants from the CICYT and the CIRIT of the Spanish and Catalan Governments, projects QFN92-4317 and GRQ94-7005.

## References

- [1] Free, S.N. (1993) *J. Biol. Chem.* 268, 9337–9342.
- [2] Gräbnitz, F., Seiss, M., Rücknagel, K.P. and Staudenbauer, W.L. (1991) *Eur. J. Biochem.* 200, 301–309.
- [3] Béguin, P. (1990) *Annu. Rev. Microbiol.* 44, 219–248.
- [4] Gilkes, N.R., Henrissat, B., Kilburn, D.G., Miller, R.C. and Warren, R.A.J. (1991) *Microbiol. Rev.* 55, 303–315.
- [5] Rojas, A., Arola, L.I. and Romeu, A. (1995) *Biochem. Mol. Biol.* 35, 1223–1231.
- [6] Ling, L.L., Rumbach, E., Zappe, H., Thomson, J.A. and Woods, D.R. (1990) *J. Gen. Microbiol.* 136, 1567–1576.
- [7] Ohmiya, K., Takano, M. and Shimizu, S. (1989) *Nucl. Acid Res.* 18, 671.
- [8] Moranelli, F., Barbier, J.R., Dove, M.J., MacKay, R.M., Seligy, V.L., Yaguchi, M. and Willick, G.E. (1986) *Biochem. Int.* 12, 905–912.
- [9] Barnett, C.C., Berka, R.M. and Fowler, T. (1991) *Biotechnology* 9, 562–565.
- [10] Machida, M., Ohtsuki, I., Fukui, S. and Yamashita, I., (1988) *Appl. Environ. Microbiol.* 54, 3147–3155.
- [11] Bause, E. and Legler, G. (1980) *Biochim. Biophys. Acta* 626, 459–465.
- [12] Kohchi, C. and Toh-e, A. (1985) *Nucl. Acids Res.* 13, 6273–6282.
- [13] Castle, L.A., Smith, K.D. and Morris, R.O. (1992) *J. Bacteriol.* 174, 1478–1486.
- [14] Raynal, A., Gerbaud, C., Francinques, M.C. and Guerincau, M. (1987) *Curr. Genet.* 12, 175–184.
- [15] Gräbnitz, F., Rücknagel, K.P., Seiss, M. and Staudenbauer, W.L. (1989) *Mol. Gen. Genet.* 217, 70–76.
- [16] Wright, R.M., Yablonsky, M.D., Shalita, Z.P., Goyal, A.K. and Eveleigh, D.E. (1992) *Appl. Environ. Microbiol.* 58, 3455–3465.
- [17] Kashiwagi, Y., Aoyagi, C., Sasaki, T. and Taniguchi, H. (1993) *J. Ferment. Bioengineer.* 75, 159–165.
- [18] Henrissat, B. (1993) *J. Ferment. Bioengineer.* 76, 240–242.
- [19] Kashiwagi, Y., Aoyagi, C., Sasaki, T. and Taniguchi, H. (1993) *J. Ferment. Bioengineer.* 75, 159–165.
- [20] Yang, M., Luoh, S., Goddard, A., Reilly, D., Henzel, W. and Bass, S. (1995) *SwissProt*, P33363.
- [21] Legler, G., Roeser, K.R. and Illig, H.K. (1979) *Eur. J. Biochem.* 101, 85–92.
- [22] Kirby, A.J. (1987) *CRC Crit. Rev. Biochem.* 22, 283–315.