





Crystal Structure at 2.3 Å Resolution and Revised Nucleotide Sequence of the Thermostable Cyclodextrin Glycosyltransferase from Thermoanaerobacterium thermosulfurigenes EM1

Knegtel, RMA; Wind, RD; Rozeboom, Henriette; Kalk, KH; Buitelaar, RM; Dijkhuizen, Lubbert; Dijkstra, BW

Published in: Journal of Molecular Biology

DOI: 10.1006/jmbi.1996.0113

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version Publisher's PDF, also known as Version of record

Publication date: 1996

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA): Knegtel, R. M. A., Wind, R. D., Rozeboom, H. J., Kalk, K. H., Buitelaar, R. M., Dijkhuizen, L., & Dijkstra, B. W. (1996). Crystal Structure at 2.3 Å Resolution and Revised Nucleotide Sequence of the Thermostable Cyclodextrin Glycosyltransferase from Thermoanaerobacterium thermosulfurigenes EM1. Journal of Molecular Biology, 256(3), 611-622. DOI: 10.1006/jmbi.1996.0113

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

JMB



Crystal Structure at 2.3 Å Resolution and Revised Nucleotide Sequence of the Thermostable Cyclodextrin Glycosyltransferase from *Thermoanaerobacterium thermosulfurigenes* EM1

Ronald M. A. Knegtel¹, Richèle D. Wind², Henriëtte J. Rozeboom¹ Kor H. Kalk¹, Reinetta M. Buitelaar², Lubbert Dijkhuizen³ and Bauke W. Dijkstra^{1*}

¹BIOSON Research Institute and Laboratory of Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Nijenborgh 4 9747 AG Groningen The Netherlands

²Agrotechnological Research Institute, P.O. Box 17, 6700 AA Wageningen The Netherlands

³Department of Microbiology Groningen Biomolecular Sciences and Biotechnology Institute (GBB), University of Groningen, Kerklaan 30 9751 NN Haren The Netherlands The crystal structure of the cyclodextrin glycosyltransferase (CGTase) from the thermophilic microorganism *Thermoanaerobacterium thermosulfurigenes* EM1 has been elucidated at 2.3 Å resolution. The final model consists of all 683 amino acid residues, two calcium ions and 343 water molecules, and has a crystallographic *R*-factor of 17.9% ($R_{\rm free}$ 24.9%) with excellent stereochemistry.

The overall fold of the enzyme is highly similar to that reported for mesophilic CGTases and differences are observed only at surface loop regions. Closer inspection of these loop regions and comparison with other CGTase structures reveals that especially loops 88-95, 335-339 and 534-539 possibly contribute with novel hydrogen bonds and apolar contacts to the stabilization of the enzyme. Other structural features that might confer thermostability to the T. thermosulfurigenes EM1 CGTase are the introduction of five new salt-bridges and three Gly to Ala/Pro substitutions. The abundance of Ser, Thr and Tyr residues near the active site and oligosaccharide binding sites might explain the increased thermostability of CGTase in the presence of starch, by allowing amylose chains to bind non-specifically to the protein. Additional stabilization of the A/E domain interface through apolar contacts involves residues Phe273 and Tyr187. No additional or improved calcium binding is observed in the structure, suggesting that the observed stabilization in the presence of calcium ions is caused by the reduced exchange of calcium from the protein to the solvent, rendering it less susceptible to unfolding.

The 50% decrease in cyclization activity of the *T. thermosulfurigenes* EM1 CGTase compared with that of *B. circulans* strain 251 appears to be caused by the changes in the conformation and amino acid composition of the 88–95 loop. In the *T. thermosulfurigenes* EM1 CGTase there is no residue homologous to Tyr89, which was observed to take part in stacking interactions with bound substrate in the case of the *B. circulans* strain 251 CGTase. The lack of this interaction in the enzyme-substrate complex is expected to destabilize bound substrates prior to cyclization. Apparently, some catalytic functionality of CGTase has been sacrificed for the sake of structural stability by modifying loop regions near the active site.

© 1996 Academic Press Limited

*Corresponding author

Keywords: X-ray structure; cyclodextrin glucanotransferase; cyclodextrin glycosyl transferase; thermostability; nucleotide sequence

Abbreviations used: BC8, *Bacillus circulans* strain 8; BC251, *Bacillus circulans* strain 251; CGTase, cyclodextrin glycosyltransferase; Hepes, *N*-2-hydroxyethylpiperazine-*N*-2-ethanesulfonic acid; r.m.s., root-mean-square; α , β , γ -CD, α , β , γ -cyclodextrin.

Introduction

The commercial production of cyclodextrins (CDs) for use as complexing agents in food, pharmaceutical and cosmetic applications (Szejtli, 1982), makes use of bacterial enzymes called cyclodextrin glycosyltransferases (CGTases; EC 2.4.11.9). Three different types of CDs, consisting of six, seven or eight glucopyranose units, respectively, find applications in industry and these are commonly referred to as α -, β - or γ -cyclodextrins. Depending on the major product of the cyclization reaction, CGTases are classified as α -, β - or γ -CGTases (Schmid, 1989). Besides producing mixtures of differently sized cyclodextrins from starch, CGTases are capable of performing coupling, disproportionation and hydrolytic reactions (Penninga et al., 1995). The limited product specificity, stability and the inhibition of the enzyme by its products have instigated us to perform protein engineering and crystallographic studies of CGTases with the aim of rationally designing enzymes with improved properties. Recently, we have reported the three-dimensional structure of the CGTase from Bacillus circulans strain 251 (Lawson et al., 1994) and its complexes with inhibitors (Strokopytov et al., 1995; B. Strokopytov et al., unpublished results) and natural substrates and products (Knegtel et al., 1995). The enzyme was found to consist of five domains, designated A through E with the catalytic carboxylates Asp229, Glu257 and Asp328 being located at the C-terminal ends of the β -strands of the $(\beta/\alpha)_8$ -barrel A-domain. Three oligosaccharide binding sites located in the E and C domains were observed to bind three maltose molecules by means of hydrogen bonds and stacking interactions with aromatic residues (Lawson et al., 1994). Maltose binding sites 1 (involving residues near Trp662 and Trp616) and 3 (involving residues near Trp413) are thought to attach the enzyme to raw starch granules, while maltose binding site 2 (located near Tyr633) is thought to guide the starch chain towards the active site (unpublished results). These studies have provided a structural explanation for the catalytic mechanism and product inhibition, and have provided insight into the factors determining product specificity.

Besides improving the catalytic properties of CGTase we are interested in increasing the stability of this enzyme under conditions of the industrial production of CDs. In order to gain more insight into the factors determining structural stability in this class of enzymes we have studied the thermostable CGTase from the thermophilic anaerobic microorganism Clostridium thermosulfurigenes EM1, lately reclassified as Thermoanaerobacterium thermosulfurigenes EM1 (Lee et al., 1993). This enzyme was recently shown to be a CGTase rather than an α -amylase, as it was originally characterized (Wind et al., 1995). The enzyme has a temperature optimum for the cyclization reaction at 85°C and a broad pH optimum around pH 6. Its thermostability is reflected by the retention of cyclization activity greater than 90% when kept at 80°C for five hours in the presence of starch. Besides starch, calcium has been shown to improve the enzyme's thermostability. Compared with the CGTase from *B. circulans* strain 251, however, the *T. thermosulfurigenes* EM1 CGTase has a twofold decreased cyclization activity and a sixfold increased hydrolytic activity at 60°C (Wind *et al.*, 1995).

The structural basis for the thermostability of proteins isolated from thermophilic microorganisms has been the subject of an increasing number of studies due to the importance of stable enzymes for use in industrial biotechnological applications. A number of factors have been suggested to play a role in protein thermostability. On the basis of sequence and structure comparisons between mesophilic enzymes and their thermophilic counterparts, it was concluded that an increase of the Ala content of α -helices at the expense of Gly, Ser, Val and Lys residues as well as Gly to Ala or Pro mutations could improve thermostability of proteins (Argos et al., 1979). Replacement of Lys by Arg has also been implicated in this respect (Mrabet et al., 1992). Structural features such as an increased number of disulphide bridges and extension of the hydrophobic core have also been suggested to improve thermostability (Fontana, 1991). Introducing disulphide bridges, however, has not always proven to be successful. For instance, although the introduction of additional disulphide bridges in bacteriophage T4 lysozyme increased its thermostability (Matsumura et al., 1989), the same strategy failed for a neutral protease (van den Burg et al., 1993). The recent analysis of the structures of aldehyde ferredoxin oxidoreductase and rubredoxin from Pyrococcus furiosus (Chan et al., 1995; Day et al., 1992) suggested that an increase of salt-bridges located at the protein surface, the shortening of loop regions and a reduction of the total protein surface contribute to the extreme stability of these proteins. Tight binding of calcium ions has been implicated in stabilizing protein structure (Fontana, 1991), as was, amongst others, observed in CGTase where the presence of calcium increases thermostability (Wind et al., 1995). In general, thermostable enzymes appear to be more rigid than mesophilic homologues as indicated by hydrogen exchange measurements, resistance against proteolytic degradation (Fontana, 1991) and a correlation between structural stability and reduced crystallographic B-factors (Vihinen, 1987). This rigidity has been suggested to explain the low catalytic efficiency of thermostable enzymes at room temperature, since they achieve the flexibility required for catalysis only at elevated temperatures. Although many different factors have been proposed to contribute to the thermostability of proteins, it generally cannot be explained by a single factor. Thermostability seems rather to be caused by a complex and subtle interplay of many different factors, and it is often dependent on the function and environment of the protein (Fontana, 1991; Jaenicke, 1991).

Besides allowing the rationalization of the observed thermostability in the presence of starch and calcium, the elucidation of the threedimensional structure of the *T. thermosulfurigenes* CGTase could provide a structural rationale for the differences in product ratio and hydrolytic activity compared with the CGTase from *B. circulans* strain 251 (Wind *et al.*, 1995). Here, we present the corrected nucleotide sequence of the *T. thermosulfurigenes* EM1 CGTase and its three-dimensional crystal structure at 2.3 Å resolution.

Results and Discussion

DNA sequencing

Figure 1 lists the revised amino acid sequence of the CGTase from T. thermosulfurigenes EM1. When compared with the sequence reported by Bahl et al. (1991) differences are observed at residues 101 (Tyr instead of a deletion), 106 to 111 (FKRTNP instead of L-REQS) and 363 to 371 (EQYMTGNGD instead of V-Y-DRQWR). Figure 2 depicts the electron density in two regions in the T. thermosulfurigenes EM1 CGTase structure where the corrected amino acid sequence deviates from that reported by Bahl et al. (1991). It is clear that the corrected amino acid sequence fits well in the electron density derived from the X-ray diffraction data. Only the side-chain of Arg108 has a reduced electron density. This residue sticks out into the bulk solvent and is likely to undergo motional averaging. Besides the corrected regions shown in Figure 2, no other discrepancy was observed between sequence and electron density during refinement.

Quality and general features of the structure

After refinement the final model had a crystallographic *R*-factor of 17.9 % (for all data in the 6.0 to 2.3 Å range with $|F_0| > 0$), a free *R*-factor of 24.9% and excellent stereochemistry as is shown in Table 2. It contains all 683 amino acid residues, 2 calcium ions and 343 water molecules. The Ramachandran plot shows no outliers except for Phe196, which is located in a generously allowed region of the Ramachandran plot. Phe196 is placed centrally in the active site, analogously to residue Tyr195 in the CGTase from Bacillus circulans strain 251 (Lawson et al., 1994), which has similar phi-psi torsion angles. Gly654 deviates from the phi-psi torsion angle distribution for glycine residues as given by PROCHECK. This residue is located in the Lys652 to Ile657 surface loop region, which displays poor quality electron density, probably due to motional averaging.

Crystal packing

In contrast to the CGTase from *B. circulans* strain 251, the CGTase from *T. thermosulfurigenes* EM1

crystallizes without the requirement for maltose or α -CD being present in the crystallization medium, similar to the CGTase from *B. circulans* strain 8 (Klein & Schulz, 1991). The intermolecular contacts observed in the T. thermosulfurigenes EM1 crystals differ, however, from those observed for the *B. circulans* strain 8 CGTase where only two of the surrounding four protein molecules of the $P2_12_12_1$ space group were contacted. As shown in Table 1 the T. thermosulfurigenes EM1 CGTase is involved in direct hydrogen bonds with three surrounding protein molecules, one of which is its own copy in the neighbouring unit cell in the direction of the a axis. There is no significant hydrophobic interaction involved in crystal packing.

Comparison with other CGTase structures

The availability of the crystal structures at 2.0 Å resolution of the CGTases of the mesophilic B. circulans strains 251 (BC251) (Lawson et al., 1994) and 8 (Klein & Schulz, 1991) (BC8, PDB entry codes 1CGD and 1CGT, respectively), and at 2.5 Å resolution of the thermophilic Bacillus stearothermophilus CGTase (Kubota et al., 1991: PDB entry code 1CYG) allows for a detailed structural comparison of the thermolabile and thermostable CGTase molecules. The overall fold of the CGTase of T. thermosulfurigenes EM1 is highly similar to that of the other CGTase structures. When the B. circulans and B. stearothermophilus CGTase structures are superimposed on the C^{α} atoms of the structurally homologous regions of the T. thermosulfurigenes EM1 CGTase, the r.m.s. deviations are 0.60 Å and 0.63 Å, for the CGTases from BC251 and BC8, and 0.59 Å for the *B. stearothermophilus* CGTase, respectively. Significant differences occur only at loop regions near residues (using *T. thermosulfurigenes* EM1 numbering) 88 to 95, 335 to 339, 494 to 499, 534 to 540 and 654 to 657. The B. circulans strain 8 CGTase structure has an additional deletion and insertion in loops 472 to 476 and 616 to 618, respectively. Figure 3 shows the superposition of the C^{α} backbones of the B. circulans and B. stearothermophilus CGTases onto that of T. thermosulfurigenes EM1. As can be seen in Figure 3, the non-superimposable loops in the thermostable CGTase are all located at the surface of the protein. The role of these loop regions in the thermostability of CGTase will be discussed in more detail in the sections below.

The hydrophobic core is well conserved in all four proteins and the two calcium binding sites are practically identical in all four cases. No additional calcium binding site is observed. The stabilizing effect of the presence of calcium ions on the *T. thermosulfurigenes* EM1 CGTase reported by Wind *et al.* (1995) could be due to a shift in the equilibrium between the apo and calcium-bound state of the enzyme towards the latter, rendering it less susceptible to unfolding. Also, long-range electro-

					10	20	30	40	50 6	50 70
TRACTERTIM	MKKTF-K	LTLV LMLSLT	UVEG LTAP	TOA ASDTAVS	NVV NYSTDVI	YOI VTDRFVD	GNT SNNPTGD	LYD PTHTSLK	YF GGDWQGIIN	K INDGYLTGMG
DCT2	MPRWIS	LVLS MSEVES	ATET VSDTOK	VTV EAAGNL-	NKV NFTSDVV	YÕI VVDRFVD	GNT SNNPSGA	LFS SGCTNLR	KYC GGDWQGIIN	IK INDGYLTDMG
B312 BC251	M VVELV	STAN INCLS	UTEG LE-SP	AOA APDTSVS	NKO NESTDVI	YOI FTDRFSD	GNP ANNPTGA	AFD GTCTNLRI	LYC GGDWOGIIN	K INDGYLTGMG
BC251	MEONAKDAFIC	TAR DADODO	CALD FLDACA	WVA DEDTAVT	NKO SESTOVI	YOV FTDRFLD	GNP SNNPTGA	AYD ATCSNLKI	LYC GGDWOGLIN	K INDNYFSDLG
BC8	MF QMAKKAF 15	NIMU LIVOID	SALF FURASA	TOT AAOADUT	NKV NYTROVI	YOT VTDRESD	GDP SNNPTGA	TYS ODCSDLH	KYC GGDWOGIII	K INDGYLTDLG
BOBH	LR	NEIV EEKIIP	UMIC IISD	VUN ADDTEVE	NKO NESTDVI	YOI FTDRESD	GNP ANNPTGA	AFD GSCTNLR	LYC GGDWOGIIN	K INDGYLTGMG
BSP382	MKRFMK	LTAV WILWLS	LILG LLSP	VIA APDISVS	MAQ MISIDVI		0111 12111 1012		ore copingern	
	0.0		1.0	0 11	0 12	0 13	0 14	0 15	0 160	170
	80	DNTVAULDDC	MECC CTCVU	C VWARDEKET	N PVFGSFTDF	O NI.TNTAHAH	N TRVTTDEAP	N HTSPASETD	P TYAENGRLYD	NGTLLGGYTN
TBACTERIUM	VTAIWISQPV	ENTIAVEPDS	C CONCYU	C VWARDERED	N DEFCTISOF	O RIVDAAHAK	G TKVITDFAP	N HTSPASETN	PSYMENGRLYD	NGTLLGGYTN
BST2	VTAIWISQPV	ENVESVMNDA	SGSASIN	G IWARDERKE	N PPPGIESDI	O NUTAAAHAK	N TEVITOFAP	N HTSPASSOO	P SFAENGRLYD	NGTLLGGYTN
BC251	VTAIWISQPV	ENITSII	NISGVINIAIA	C WWARDFRRI	N DVECTMADE	O NLITTAHAK	G TETVIDEAP	N HTSPAMETD	T SFAENGRLYD	NGTLVGGYTN
BC8	VTALWISQPV	ENIFATI	NYSGVINIAIA	G IWARDFARI	N PIPOIPADE	D DIMOTALICA	C TEVIMORTO	N HESPALETD	D SVAFNGAVYN	DGVLTGNYSN
вонв	ITAIWISQPV	ENVYA-LHPS	GITSIR	G IWARDIKKI	N PFIGDISDI	V NI TOTAHAH	N TEVITOFAP	N HTSPASSOD	P SFAENGRLYD	NGNLLGGYTN
BSP382	ITAIWISQPV	ENIYSVI	NISGVHNTAIH	G IWARDFREI	N PAIGINGDI	K NDIDIANAN	IN INVIIDIMI		Dimenoneire	
		100	200	210	220	230	240	250	260	270
	180	190	200	L NOONCETDC	VINCATVINI	DMGIDGIRLD	AVKHMPEGWO	KNEMDSTLSV	REVETEGEWE L	TNETDVNN
TBACTERIUM	DTNGYFHHYG	GTDFSSYEDG	IYRNLFDLAD	LNUONDUIDD	VINDAURAMUT	DMGIDGIRDD	AVIGHNITOWQ	KSLMDEIDNY	REVETEGEWE L	SENEVDANN
BST2	DANMYFHHNG	GTTFSSLEDG	I YRNLFDLAD	LNHQNPVIDK	VINDATIM	DICIDCIEMD	AVKUMPEGWO	KSEMAAUNNV	KPVETEGEWE L	WNEVSPEN
BC251	DTQNLFHHNG	GTDFSTTENG	TYKNLYDLAD	ENHINDS I VDV	VEVENTRIME	DEGIDGIRUD	AVKHMPLGWO	KSWMSSTVAH	KPVFTFGEWF L	SAASDADN
BC8	DTNGYFHHNG	GSDFSSLENG	TYKNLYDLAD	VDI NUMATION	VINECTNIMI	DECTOCIEVO	AVKUMSEGWO	TSLMSDIVAH	EDVETEGEWE L	SGEVDPON
BOHB	DPNNLFHHNG	GTDFSSYEDS	TIRNLIDLAD	I DENNI I VHDQ	VINDATKIMI	DICUDCIEVD	AVKHMPEGWO	KSEMSTINNY	KPVENEGEWE L	SVNETSPEY
BSP382	DTQNLFHHYG	GTDFSTIENG	TAKNPADPAD	LINHININSSVDV	IPKDATKIMP	DEGVEGIKVE	AVIANITIONQ			5111210121
		200	200	210	320	330	340	350	360	370
	280	290	DOVEDDNED	MYCIDEMIOS	TACOVNETNO	MUTETONHOM	DREV-NGGSTR	PVEOALAFTL	TSEGVPALYY	GTEOYMTGNG
TBACTERIUM	TYFANESGMS	LLDFRFSQKV	ROVERDNIDI	MIGLDSMIQ3	TASIMITIND	OWTEIDNHDM	DREMINGGOPR	KVDMALAVI.I.	TSEGVENING	TEOYMTGNG
BST2	HYFANESGMS	LEDFRFGQKL	ROVERNNSDN	MYCLVAMLEC	CANDANDER	OVERTONHOM	FREHASNANRR	KLEOALAFTL	TSEGVPAIYY	GTEOYMSGGT
BC251	HKFANESGMS	LLDFRFAQKV	ROVERDNIDN	MUNIDEMINE	TATINATINA	OVTEIDNHDM	DREKTSAVNNE	RLEOALAFTL	TSRGVPAIYY	GTEOYLTGNG
BC8	TDFANKSGMS	LLDFRFNSAV	RIVFRONTSN	MUNDENEMTAS	TRIDINGVIND	OVTETONHOM	SEESEEOSSNE	HTDIALAVIL	TSRGVPTIYY	GTEOYLTGGN
BOHB	HHFANESGMS	LEDFQFGQTI	RDVEMDGSSN	WIDENLEITAS	SEVDYAOUND	OVTETONHOM	FREHTSNGDRR	KLEOALAFTL	TSRGVPAIYY	GSEOYMSGGN
BSP382	HQFANESGMS	LUDEPEAQKA	ROVERDIVIDIN	MIGLINAMLEG	SEVDINGVIND	QVII IDMIDII	BIGHIBIODIG			
										450
	200	200	100	410	420	430	440	450	460	4/0
	380	390	400	410	420	430 VERKEGNNVA	440	450 SYNITGLYTA	460 LPAGTYTDVL G	470 GLLNGNSIS
TBACTERIUM	380 DPYNRAMMTS	390 FNTSTTAYNV	400 IKKLAPLRKS	410 NPAIAYGTTQ	420 QRWINNDVYI ORWINGDVYV	430 YERKFGNNVA YEROFGKDVV	440 LVAINRNLST LVAVNRSSSS	450 SYNITGLYTA NYSITGLFTA	460 LPAGTYTDVL G LPAGTYTDOL G	470 GLLNGNSIS GLLDGNTIQ
TBACTERIUM BST2	380 DPYNRAMMTS DPNNRKMMSS	390 FNTSTTAYNV FNKNTRAYQV	400 IKKLAPLRKS IQKLSSLRRN	410 NPAIAYGTTQ NPALAYGDTE	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA	440 LVAINRNLST LVAVNRSSSS	450 SYNITGLYTA NYSITGLFTA PASISGLVTS	460 LPAGTYTDVL G LPAGTYTDQL G LPOGSYNDVL G	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS
TBACTERIUM BST2 BC251	380 DPYNRAMMTS DPNNRKMMSS DPDNRARIPS	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA YERKFGKSVA	440 LVAINRNLST LVAVNRSSSS VVAVNRNLNA VVAVNRNLST	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS	460 LPAGTYTDVL G LPAGTYTDQL G LPQGSYNDVL G LPTGSYTDVL G	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT
TBACTERIUM BST2 BC251 BC8	380 DPYNRAMMTS DPNNRKMMSS DPDNRARIPS DPDNRAKMPS	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI QRWINNDVYV ERWINSDVYI	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA YERKFGKSVA	440 LVAINRNLST LVAVNRSSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS	460 LPAGTYTDVL G LPAGTYTDQL G LPQGSYNDVL G LPTGSYTDVL G LPOGOYTDEL 0	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT OLLDGNEIT
TBACTERIUM BST2 BC251 BC8 BOHB	380 DPYNRAMMTS DPNNRKMMSS DPDNRARIPS DPDNRAKMPS DPENRKPMSD	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTTNSYQI	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN	410 NPALAYGTTQ NPALAYGDTE NPALAYGSTQ NPALAYGSTQ NPALGYGNTS	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI QRWINNDVYV ERWINSDVYI ERWINNDVII	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA YERKFGKSVA YERSFGDSVV YERKFGNNVA	440 LVAINRNLST LVAVNRSSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT VVAINRMNT	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS	460 LPAGTYTDVL G LPAGTYTDQL G LPQGSYNDVL G LPTGSYTDVL G LPQGQYTDEL Q LPOGSYNDVL G	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT QLLDGNEIT GILNGNTLT.
TBACTERIUM BST2 BC251 BC8 BOHB BSP382	380 DPYNRAMMTS DPNNRKMMSS DPDNRARIPS DPDNRAKMPS DPENRKPMSD DPDNRARIPS	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPALGYGNTS NPAIAYGSTQ	420 QRWINNDVYI QRWINGDVYV ERWINNDVVI ERWINNDVYV ERWINSDVYI ERWINNDVII	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA YERKFGKSVA YERSFGDSVV YERKFGNNVA	440 LVAINRNLST LVAVNRSSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT VVAINRNMNT	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS	460 LPAGTYTDVL G LPAGTYTDQL G LPQGSYNDVL G LPTGSYTDVL G LPQGQYTDEL Q LPQGSYNDVL G	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNTIT QLLDGNEIT GILNGNTLT-
TBACTERIUM BST2 BC251 BC8 BOHB BSP382	380 DPYNRAMMTS DPNNRKMNSS DPDNRARIPS DPDNRAKMPS DPENRKPMSD DPDNRARIPS	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPALGYGNTS NPAIAYGSTQ	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI QRWINNDVYV ERWINSDVYI ERWINNDVII	430 YERKFGNNVA YERQFGKDVV YERKFGSNVA YERKFGKSVA YERSFGDSVV YERKFGNNVA	440 LVAINRNLST LVAVNRSSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT VVAINRNMNT	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS	460 LPAGTYTDVL G LPAGTYTDVL G LPQGSYNDVL G LPTGSYTDVL G LPQGQYTDEL Q LPQGSYNDVL G 50 56	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT QLLDGNEIT GILNGNTLT 0 570
TBACTERIUM BST2 BC251 BC8 BOHB BSP382	380 DPYNRAMMTS DPNNRKMMSS DPDNRARIPS DPDNRAKMPS DPENRKPMSD DPDNRARIPS 480	390 FNTSTTAYNV FNKNTRAYQV FSKSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 490	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS IQKLAPLRKS 500	410 NPALAYGTTQ NPALAYGDTE NPALAYGSTQ NPALAYGSTQ NPALAYGSTQ NPALAYGSTQ 510	420 QRWINNDVYI QRWINDDVYI ERWINNDVLI QRWINNDVYV ERWINNDVVI ERWINNDVII 5200	430 YERKFGNNVA YERKFGSNVA YERKFGSNVA YERKFGSVV YERKFGNVA 530 CECTTSCOVI	440 LVAINRNLST LVAVNRSSS VVAVNRLNA VVAVNRLST LTAVNSG-DT VVAINRNMNT	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 US WDDTEVKV	460 LPAGTYTDVL G LPAGTYTDQL G LPQGSYNDVL G LPTGSYTDVL G LPQGQYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT QLLDGNEIT GILNGNTLT- 0 570 I SLKTSSGATS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM	380 DPYNRAMMTS DPDNRARIPS DPDNRARIPS DPENRKPMSD DPDNRARIPS 480 VASDGSVTPF	390 FNTSTTAYNV FNKNTRAYQV FSKSTSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 490 TLSAGEVAVW	400 IKKLAPLRKS IQKLAPLRKC ISKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL	410 NPALAYGTTQ NPALAYGDTE NPALAYGSTQ NPALAYGSTQ NPALGYGNTS NPALAYGSTQ 510 IGHVGPTMTK	420 QRWINNDVYI ERWINNDVLI QRWINNDVLI QRWINNDVYV ERWINNDVYI ERWINNDVII 520 AGQTITIDGR	430 YERKFGNNVA YERKFGNVA YERKFGKSVA YERKFGNVA 530 GFGTTSGQVL GEGTNTGQVL	440 LVAINRNLST LVAVNRSSSS VVAVNRLST LTAVNSG-DT VVAINRNMNT FGSTAGTI	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV VS WDDTEVKV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGSYNDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN XV PNVSPGKYN	470 GLLNGNSIS GLLDGNTIQ GLLNGNTIS GVLNGNNIT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVOSSSCOTS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2	380 DPYNRAMMTS DPDNRAKMMSS DPDNRAKMPS DPDNRAKMPS DPDNRAKPSD DPDNRAKIPS 480 VASDGSVTPF VGSNGSVNAF	390 FNTSTTAYNV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 490 TLSAGEVAVW DLGPGEVGVW	400 IKKLAPLRKS IQKLAPLRKC ISKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AXSATESTPI	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPTMTK	420 QRWINNDVYI QRWINGDVYV ERWINNDVLI QRWINNDVVI ERWINNDVVI ERWINNDVII 520 AGQTITIDGR VGHQVTIDGE	430 YERKFGNNVA YERKFGSNVA YERKFGSNVA YERKFGSNVA YERKFGNVA 530 GFGTTSGQVL GFGTTTSGQVL GFGTTTGTVK	440 LVAINRNLST LVAVNRSSS VVAVNRNLST LTAVNSG-DT VVAINRNMNT FGSTAGTJ FGTTAATJ FGTTAATJ	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINILMTS PASITGLVTS 40 5 VS WDDTEVKV VVS WSNNQUY UTS WEDTOLKV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPGGSYNDVL G LPQGSYNDVL G 50 56 KV PSVTPGKYN KV PSVTPGKYN KV PNVSPGKYN	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNTLT GLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSSCQTS I KVANAGTAS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251	380 DPYNRAMMTS DPDNRARIPS DPDNRARIPS DPDNRARIPS DPDNRARIPS 480 VASDGSVTPF VGSNGSVNAF VGSGGAASNF	390 FNTSTTAYNV FSKSTTAYQV FSKSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 490 TLSAGEVAVW DLGPGEVGVW TLAAGGTAVW	400 IKKLAPLRKS IQKLSPLRKC ISKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTAATATPT	410 NPAIAYGTTQ NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPTMTK IGHVGPTMAGQ IGHVGPTMAGQ	420 QRWINDVYI QRWINGDVYV ERWINNDVII QRWINNDVYI ERWINNDVYI SCOUNTIDGE VGHQVTIDGE PGVTITIDGR PGVTITIDGR	430 YERKFGNNVA YERKFGKDVV YERKFGSNVA YERKFGSVVA YERKFGNVA 530 GFGTTSGQVL GFGTTSGQVL GFGTNTGTVK GFGSSKGTVY	440 LVAINRNLST LVAVNRSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT VVAINRNMNT FGSTAGT FGSTAGT FGTTAAN FGTTAVSGAD	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 US WDDTEVKV VVS WSNNQIVV TS WEDTQIKV	460 LPAGTYTDVL G LPAGTYTDVL G LPQGSYNDVL G LPQGSYNDVL G LPQGSYNDVL G 50 56 KV PSVTPGKYN VAV PNVSPGKYN VAV PNVSPGKYN KL PAVAGGNYN	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNTLT GILNGNTLT 0 570 I SLKTSSGATS I TVQSSSCQTS I KVANAAGTAS V KVAAAGTAS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8	380 DPYNRAMMTS DPNNRAMMSS DPDNRARIPS DPDNRAKMPS DPENRKPMSD DPDNRARIPS 480 VASDGSVTPF VGSNGSVNAF VGSGGAASNF -STNGSINNF	390 FNTSTTAYDV FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 1USAGEVAVW DLGPGEVGVW TLLAAGETAVW TLLAAGATAVW	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTAATATPT QYTAATATPT	410 NPAIAYGTQ NPAIAYGTQ NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPVMKQ IGHVGPVMKA IGHVGPVMKA IGHVGPVMKA	420 QRWINNDVYI QRWINNGDVYV ERWINNDVLI QRWINNDVYV ERWINNDVII ERWINNDVII S20 AGQTITIDGR VGHQVTIDGR PGVVTIDGR PGVVTIDGR	430 YERKFGNVA YEROFGKDVV YERKFGSNVA YERKFGSNVA YERKFGNVA 530 GFGTTSGQVL GFGTTSGQVL GFGTTSGVK GFGSSKGTVY GFGSSKGTVY GFGSSKGTVY GFGSSKGTVY	440 LVAINRNLST LVAVNRSSS VVAVNRNLNA VVAVNRNLST LTAVNSG-DT VVAINRNMNT FGSTAGT FGTTAAN FGTTAVSGADJ FGTTAVTGAAJ	450 SYNITGLYTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV VS WSNNQIVV TS WEDTQIKV TS WEDTQIKV US WEDTQIKV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN XI PAVAGGNYN XI PAVAGGNYN TI PSVAAGNYA	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSSQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSOS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB	380 DPYNRAMMSS DPNNRAMSS DPDNRAKMSS DPDNRAKPS DPDNRARIPS 480 VASDGSVTPF VGSNGSVNAF VGSGGASNFF VGSGGASNFF VNSNGAVDSF	390 FNTSTTAYINV FNINTRAYQV FSTSTTAYQV FSTSTTAYQV FDRTTNSYQI FSTTTAYQV ULSAGEVAVW DLGPGEVGVW TLAAGGTAVW QLSANGVSVW	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISTLASLRQN IQKLAPLRKS 500 QVVSSSNSPL AYSATESTPI QYTAATAPTPI QTITEEHASEL	410 NPAIAYGTYQ NPAIAYGDYE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPMMKK IGHVGPMMKK IGHVGPMMKK IGHVGPMMKK	420 QRWINDDVYU QRWINDDVYU ERWINNDVLI QRWINNDVVI ERWINNDVII ERWINNDVII 520 AGQTITIDGR VGHQVTIDGE PGVVITIDGR HGNTVTITGE	430 YERKFGNVA YERKFGNVA YERKFGSVV YERKFGSVV YERKFGSVV YERKFGNVA 530 GFGTTSGQVL GFGTNGTVK GFGSKGTVY CFGSTKGTVY CFGCTGTV	440 LVAINENLST LVAVNENLSSS VVAVNENLSA VVAVNENLST LTAVNSG-DT VVAINENMMT FGSTAAN FGSTAAN FGTTAVSGADT FGSTAVTGAAT FDSDFSD	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINULNTS PASITGLVTS 40 5 SVS WDDTEVKV VVS WSNNQIV VVS WSNNQIV TIS WEDTQIKV TIS WEDTQIKV VLS WSDTKIEV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPGGSYNDVL G LPQGSYNDVL G 50 56 KV PSVTPGKYN NAV PNVSPGKYN KI PAVAGGNYA TI PSVAAGNYA SV PDVTAGHYD KI LPVGGYV	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNTLT GILNGNTLT 0 570 I SLKTSSGATS I TVQSSSCQTS I KVANAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGCAS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 RC251 BC8 BOHB BSP382	380 DPYNRRAMTSS DPDNRRAMTSS DPDNRAKMSS DPDNRAKMSD DPDNRARIPS 480 VASDGSVTPF VGSNGSVNAF VGSNGSVNAF VGSGGAASNF VNSNGAVDSF VNSNGAVDSF	390 FNTSTTAYDV FNTSTTAYDV FSTSTTAYOV FSKSTTAFNV FDRTTNSYQI FSTTTTAYOV ULSAGEVAVW TLAAGATAW TLAAGATAW TLAPGGTAVW	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKS ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTAATAPTT QYTATATTPT QTTEHASPL QTTEHASPL	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPMMKQ IGHVGPMMKK IGHVGPMMKK NGNVGPMMKK	420 QRWINDDVYI QRWINDDVYI ERWINDDVIU ERWINDDVYU ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDGE PGVVTIDGE AGVTITIDGR	430 YERKFGNVA YERKFGNVA YERKFGSNVA YERKFGSVA YERKFGNVA 530 GFGTNTGTVK GFGSTGTVGTVK GFGSTGTVY GFGSTKGTVY GFGSTKGTVY A-SARQGTVY	440 UVAINRILST LVANNRLST LVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGT FGSTAGT FGTTA-SAD FGTTAVGAD FGTTAVGAD FGSTAVGAD	450 SYNITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS VS WDDTEVKV VVS WSNNQIVV TTS WEDTQIKV LIS WEDTQIKV LIS WEDTQIKV VA WEDTQIQV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGQYTDEL Q LPQGQYTDEL Q LPQGSYNDVL G S0 50 FXV PSVSPGKYN XAV PNVSPGKYN XAV PNVSPGKYN TI PSVAAGNYN TI PSVAAGNYN XSV PDVTAGHYD KL LRVPGGIYD	470 GLLNGN51S GLLDGNTIQ GLLNGNTIG GVLNGNNIT QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSSGATS I TVQSCSCQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I RVANAAGAS
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382	380 DPYNRAMMSS DPDNRAKMSS DPDNRAKMSS DPENRAKMSD DPENRAKMSD DPENRAKMSD DPENRAKMSD DPENRAKMSS DPDNRARIPS 480 VASDGSVTPF VGSGGAASNF -STNGSINNF VNSNGAVDSF VGSGGAASNF	390 FNTSTTAYIN FNKNTRAYQV FSTSTTAYQV FSTSTTAYQV FSTSTTAYQV TSTTTAYQV 490 TLSAGEVAVW DLGPGEVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 00 QVVSSSNSPL AYSATESTPI QVTAATTPT QVTTAETTPT QVTTAETTPT QVTTAETAFT	410 NPAIAYGTTQ NPAIAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPTMGK IGHVGPVMGK IGHVGPVMGK	420 QRWINDDVYI QRWINDDVYI ERWINDDVJI ERWINDDVJI ERWINDDVJI 520 AGQTITIDGR VGHQVTIDGE HGNTVTITGE AGVTITIDGR	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTTSGQVL GFGSTKGTVY GFGSKGTVY GFGSKGTVY GFGDNEGSVL A-SARQGTVY	440 LVAINRNLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGTT FGSTAGTT FGTTAVTGAD FGTTAVTGAD FGTTAVTGAD	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VIS WDDTEVKV VIS WSNNQIVV VIS WSNNQIVV VIS WSDTQIKV VLS WDTQIKV VLS WDTKIEV VLA WEDTQIQV	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDUL G LPTGSYTDVL G LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN XI PAVAGGNYN TI PSVAAGNYN TI PSVAAGNYN TI PSVAAGNYN TI PSVAAGNYN TI PSVAAGNYN KI LRVPGGIYE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSGATS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVVNAGDSQS 670
TBACTERIUM BST2 BC251 BC8 BOHB BSF382 TBACTERIUM BST2 RC251 BC8 BOHB BSP382	380 DPYNRAMMSS DPNNRAMSS DPDNRAKMSS DPDNRAKMSS DPDNRARTPS 480 VASDGSVTPF VGSNGSVNAF VGSNGSVNAF SSINSF VNSNGAVDSF VGAGGAASNF	390 FNTSTTAYINV FNNTRAYQV FSTSTTAYQV FSTSTTAYQV FDRTTNSYQI FSTTTAYQV TLSAGEVAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW 590	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 007VSSSNSPL AYSATESTPI QYTATATAPT QYTATATAPT QUTEEHASPL QYTTDATAPI 000000000000000000000000000000000000	410 NPAIAYGTYQ NPAIAYGDYE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPMMGK IGHVGPMMGK NGNVGPMMAK 610	420 QRWINDDVYI QRWINDDVYI ERWINDDVLI QRWINDDVU ERWINDDVII ERWINDDVII 520 AGQTITIDGR VGHQVTIDGE PGVVITIDGR HGNTVTIDGR HGNTVTIGE AGVTITIDGR 622	430 YERKFGNVA YERKFGNVA YERFGSVV YERKFGSVV YERKFGSVV YERKFGNVA 530 GFGTTSGQVL GFGTNGTVK GFGSKGTVY CFGDNEGSVL A-SARQGTVY 0 63 SHENOLOGYO	440 LVAINENLST LVAVNENLSS VVAVNENLSA VVAVNENLSA LTAVNSG-DT VVAINENMNT FGSTAGTJ FGSTAGTJ FGTTAVSGADJ FGTTAVTGADJ FGTTAVTGADJ 0 641	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINULNTS PASITGLVTS 640 5 SVS WDDTEVKU VIS WDDTEVKU VIS WDDTEVKU TIS WEDTQIKU TIS WEDTQIKU VIS WDTQIKU VIS WDTQIQ O 650 ACTTIORETI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPGGSYNDVL G LPQGSYNDVL G 50 56 KV PSVTPGKYN VAV PNVSPGKYN KI PAVAGGNYN TI PSVAAGNYA SV PDVTAGHYD KI LRVPGGIYE KKNG-NTITWF	470 GLLNGNSS GLLDGNTIQ GLLNGNTLS GVLNGNTLT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSSCQTS I KVANAGTAS V KVA-ASGVNS I SVVNAGDSQS I RVANAGAAS 670 GGSNHTYTVP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 RC251 BC8 BOHB BSP382 TBACTERIUM	380 DPYNRRAMTS DPDNRRAMTS DPDNRRAMSS DPDNRARTPS DPENRKPMSD DPENRKPMSD DPDNRARTPS 480 VASDGSVTPF VGSNGSVNFF VGSNGANSF VNSNGAVDSF VNSNGAVDSF VSAGAASNF 580 NTYNNINILT	390 FNTSTTAYDV FNTSTTAYDV FSKSTTAFDV FSKSTTAFNV FDRTINSVQI FSTTTTAYQV 1LAAGGTAVW TLAAGGTAVW ULSANGVSVW TLAPGGTAVW S90 GNQICVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKC ISKLAPLRKS 500 QVVSSNSPL QVSSSNSPL QVTSATAPT QVTTAETTPT QTTAETTPT QTTAETAPT QUTTAATAPT QUTTAATAPT QUTTAATAPT QUTTAATAPT QUTTAATAPT	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPMMSQ IGHVGPMMSQ IGHVGPMMSK IGHVGPMMSK IGHVGPMMSK GHUGPMMSK 610 VYLITNVAEL	420 QRWINDDVYI QRWINDDVYI ERWINDDVI ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGVVTIDGE PGVVTIDGE AGVTITIDGR 62' GNWDTSK-AI	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERFGDSVV YERKFGNVA 530 GFGTTSGVL GFGTTSGVL GFGSTKGTVY GFGSTKGTVY CFGDNEGSVL A-SARQGTVY 0 63 9 PMFNOVVQ D 63	440 UVAINRNLST LVAINRNLST LVAVNRSSS VVAVNRLNA LTAVNSG-DT VVAINRNMNT SGSTAGTT VVAINRNMNT FGTTAVSGAD FGTTAVSGAD FGTTAVSGAD FGTTAVSGAD GGTAVTGAD	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 SVS WDDTEVKU VIS WSDNQIVU TS WEDTQIKU US WEDTQIKU US WEDTQIKU US WEDTQIKU O 650 PAGTIOFKFJ PAGTIOFKFJ	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN KI PAVAGGYYN KI PAVAGGYYN KI LRVPGGIYE KKNG-NTITWE KKDSGNVTME	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNIT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I TVQSSCQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS GGSNHTTTVP SGGNHVYTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC8 BOHB BSP382 TBACTERIUM BST2	380 DPYNRAMMSS DPNNRAMMSS DPDNRARMSS DPDNRARMSD DPENRARMSD DPENRARMSD DPENRARMSS DPENRARMSS DPENRAMSS VGSGGAASNF -STNGSINNF VMSNGAVDSF VGSGGAASNF 580 STYNNINLT AAYDNFEVLT	390 FNTSTTAYINV FNKNTRAYQV FSKSTTAYQV FSKSTTAFNV FDRTTNSYQI FSTTTTAYQV 1LSAGEVAVW DLGPGEVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW S90 GNQICVRFVV NDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS OQVOSSSNSPL AYSATESTPI QYTAATTSTPI QYTAATTSTPI QYTAATTSTPI QYTAATATPT QYTTAATATPI QYTTAATATPI QYTTAATATPI 0 OCO 0 NNASTVYGEN NNATTNLGQN	410 NPAIAYGTYQ NPAIAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPTMTK IGHVGPTMAK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMAL IYIVGNVYEI IYIVGNVYEI	420 QRWINDVII QRWINDDVYI ERWINDVII ERWINDVII ERWINDVVI ERWINDVVII 520 AGQTITIDGR VGHQVTIDGE HGNTVTITGE AGVTITIDGR GNWDTSK-AI(GNWDTSK-AI(GNWDTSK-AI	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTTSGQVL GFGSTKGTVY GFGSKGTVY GFGSKGTVY GFGDNEGSVL A - SARQGTVY 0 63 9 PMFNQVVYS 5 PMFNQVVYS	440 LVAINRNLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGTV FGSTAGTV FGTTAVGADJ FGTTAVGADJ GTTAVGADJ 0 640 Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VIS WDDTEVKV VIS WSDNQIVV VIS WSDNQIVV VIS WSDTKIEV VIS WSDTKIEV VIS WSDTKIEV O 650 PAGTTIQFKFI P GGTTIQFKFI P GGTTIFFFFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN XI PAVAGGNYN XI PAVAGGNYN XI PAVAGGNYN XI LRVPGGIYE XKKOG-NTITWE KKKOGSYTWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GULNGNTLT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAGDSQS GGSNHTYTVP : SGSNHYTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC8 BOHB BSP382 TBACTERIUM BST2 BST2 BC51	380 DPYNRKAMSS DPDNRKAMSS DPDNRAKMSS DPDNRAKMSS DPDNRAKMSD DPENRKKMSD DPENRKMSD VASDGSVTNF VGSNGSVNAF VGSNGSVNAF VGSGGAASNF VMSNGAVDSF VMSNGAVDSF S80 MYYNNINILT NYYDNFEVLT NYYDNFEVLT	390 FNTSTTAYDV FNTSTTAYDV FSTSTTAYOV FSKSTTAFNV FDRTINSVQI FSTTTTAYOV ULSAGEVAVW TLAAGGTAVW TLAAGATAVW ULSANGVSVW TLAPGGTAVW S90 GNQICVFFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTATATTPT QYTAATATPT QYTAATATPT QTTEHASPL QTTEHASPL QTTEHASPL MNATVLGON NNATTALGON	410 NPAIAYGTYQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPMMGQ IGHVGPMMGQ IGHVGPMMGK NGNVGPMMGK NGNVGPMMGK 0610 VYLTGNVÆL YYLTGNVÆL	420 QRWINDDVYI QRWINDDVYI ERWINDDVYI ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDGR PGNVTIDGR GNVDTSK-AII GNWDTSK-AII GNWDTSK-AII	430 YERKFGNVA YERKFGNVA YERFGSVV YERKFGSVV YERKFGSVV YERKFGNVA 530 GFGTTSGQVL GFGTNGTVK GFGSKGTVY CFGDNEGSVL A-SARQGTVY 6 S PMFNQVVYQ S PMFNQVVYQ D DENEVVVIQ	440 UVAINRILST LVAINRILST LVAVNRILST LTAVNSG-DT VVAINRILST LTAVNSG-DT VVAINRNMNT FGTTAGTD FGTTAGTD FGTTAVTGAD FGTTAVTGAD FGTTAVTGAD 0 644 Y PTWYLDVSU Y PTWYLDVSU Y PTWYLDVSU	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINLINTS PASITGLVTS SVS WDDTEVKU VS WSDNQIVU TS WEDTQIKU VIS WSDTKIEV VA WEDTQIQU D 6651 P GGTIEFKFI P AGKTIEFKFI P AGKTIEFKFI P AGKTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPQGSYNDVL G LPQGQYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN KI PAVAGGNYN KI PAVAGGNYN KI LRVPGGIYE 0 660 KKKG-NTITWE KKKGGSTVTWE KKKGG-STTTWE	470 GLLNGN51S GLLDGNT1Q GLLNGNT1Q GULNGN1T QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSGATS 1 TVQSSSCQTS 1 KVANAAGTAS 1 SVVNAGDSQS 1 RVANAAGAAS 670 6 GGSNHTYTVP 2 SGSNHTTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BC8	380 DPYNRRAMTS DPDNRRAMTS DPDNRRAMSS DPDNRARTPS DPENRKMSD DPENRKMSD DPDNRARTPS 480 VASGGAASNF VGSNGAVDSF VGSGAASNF VGSGAASNF S80 NTYNNINILT AYDNFEVLS NAYNNFFULT	390 FNTSTTAYIN FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTINSYQI FSTTTAYQV LQGQEVGVW ULAPGGTAVW ULAPGGTAVW ULAPGGTAVW TLAPGGTAVW S90 GNQICVRFVV DQVSVRFVV GDQVSVRFVV GDQVTVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKC ISKLAPLRKS 500 QVSSNSPL QVSSNSPL QVSSNSPL QVTAATAPT QVTAATAPT QUTEEHASPL QUTEATAPT QUTTAATAPT QUTAATAPT 0 600 NNASTVYGEN NNASTVYGEN NNASTLGQN	410 NPAIAYGTTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPMMKK IGHVGPMMKK IGHVGPMMKK IGHVGPMMKK 610 VYLTGNVAEL IYIVGNVYEL IYIVGNVYEL LYLTNVAEL	420 QRWINDDVYI QRWINDDVYI ERWINDDVI ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDGE PGNVTIDGE GNVDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG	430 YERKFGNVA YERKFGNVA YERFGSDVV YERKFGSVA YERFGDSVV YERFGDSVV YERFGDSVV YERFGNVA GFGTTSGVL GFGTTSGVL GFGTTSGVL GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY 0 63 9 MFNQVVYQ 9 MFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ	440 UVAINRNLST LVAVNRSSS VVAVNRLNA VVAVNRLST LTAVNSG-DT VVAINRNMNT FGSTAGTT FGSTAGTTAVGAD FGTTAVGAD FGTTAVGAD GTTAVGAD 0 64(Y PTWYDVSV) Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI	450 SYNITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV VS WSNNQIVV UTS WEDTQIKV VS WSNNQIVV UTS WEDTQIKV US WEDTQIKV US WEDTQIKV D 650 PAGTIOFKFI P AGKLIFKFI P AGKLIFKFI P AGKLIFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPTGSYTDVL G LPQQYTDEL Q LPQGYTDEL Q LPQGYTDEL Q S0 50 56 KV PSVTPGKYN KI PAVAGGNYN KI PAVAGGNYN KI LRVPGGIYE KKNG-NTITWE KKNG-STVTWE KKNG-STITWE KKNG-STITWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNETT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I TVQAAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAAGSQS GGSNHTYTPP GGSNHTFTAP SGSNHTFTAP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC7 BC7 BC7 BC7 BC7 BC7 BC7 BC7 BC7 BC7	380 DPYNRAMMSS DPNNRAMMSS DPDNRARMSS DPDNRARMSS DPENRARMSS DPENRARMSS DPENRARMSS DPENRARMSS DPENRAMSS A80 VASDGSVTPF VGSGGAASNF -STNGSINNF VGSGGAASNF -STNGSINNF VGSGGASNF SINS VGSGGASNF SINS SINS VGSGGASNF TISNINILT AAYDNFEVLT NYYDNFEVLT	390 FNTSTTAYINV FNKNTRAYQV FSKSTTAYQV FSKSTTAFNV FDRTINSYQI FSTTTTAYQV ULSAGEVAVW DLGPGEVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAPGGTAVW TLAPGGTAVW GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 0QVVSSSNSPL AYSATESTPI QVTAATTSTP QVTAATTPT QVTAATTPT QVTAATTPT QVTAATAPT OCCOMPANIE NNASTVGEN NNATTNLGQN NNASTLGQN NNATTSLGGN	410 NPAIAYGTYQ NPAIAYGDTE NPAIAYGDTE NPAIAYGSTQ NPAIGYGNTS NPAIAYGSTQ 510 IGHVGPTMTK IGHVGPTMGK IGHVGPTMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK 1GHVGPVMGK LYLTGNVAEL LYLTGNVAEL LYLTGNVAEL	420 QRWINDUYI QRWINDUYI ERWINDUVI ERWINDUVI ERWINDUVI ERWINDUVI S20 AGQTITIDGR YGHQVTIDGE HGNTVTITGE AGVTITIDGR GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTNGTVK GFGSTKGTVY GFGSKGTVY GFGSKGTVY GFGDNGSVL A-SARQGTVY 0 63 PMFNQVVYQ 5 PMFNQVVYQ 5 PMFNQVVYQ 5 PMFNQVVYQ 6 PMFNQVHQ 5 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 6 PMFNQVHQ 7 PMFNQVHQ 6 PMFNQVHQ 7 PMFNQ 7 PMFNQ 7 PMFNQVHQ 7 PMFNQ 7 PMFNQ	440 LVAINENLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGTV FGSTAGTV FGTTAVGSAD FGTTAVTGAD FGTTAVTGAD 0 640 Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 340 5 US WDDTEVKV VS WSNNQIVV VS WSNNQIV VS WSDTQIKV US WSDTQIKV JLS WSDTQIKV JLS WSDTQIKV D 650 PAGTTIQFKFI P AGKTIEFKFI P AGKLEFKFI P AGKLEFKFI P AGKLEFKFI P AGKLEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGYTDVL G 50 56 KV PSVTPGKYN KI PAVAGGNYN KI PAVAGGNYN KI PAVAGGYY KI LRVPGGIYI 0 660 KKNG-NTITWE KKNG-STITWE KKNG-STITWE KKNG-STITWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GULNGNTLT QLLDGNEIT GILNGNTLT 0 570 I SLKTSSGATS I KVANAAGTAS V KVA-ASGVNS I KVANAAGTAS V KVA-ASGVNS I SVNAGDSQS I RVANAAGAS GGSNHTYTVP SGSNHTYTTP SGSNHTTTP SGANNHTTTP SGANNHTTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BSP382 TBACTERIUM BST2 BC251 BC251 BC251 BC251 BC251 BC251 BC251 BC8 BOHB BSP382	380 DPYNRKAMSS DPDNRKAMSS DPDNRKMSS DPDNRAKMSS DPDNRAKMSD DPENRKMSD DPDNRARIPS 480 VASDGSVTPF VGSNGSVNAF VGSGGAASNF VGSGGAASNF VSNGGAVDSF VSNGGAVDSF VSNGGVDFEVLT AAYDNFEVLT NYYDMFEVLT PYYDKFEVLT	390 FNTSTTAYDV FNTSTTAYDV FSKSTTAFVV FSKSTTAFVV FSKSTTAFVV DLSPGEVGVW TLSAGEVAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW OLSANGVSVW TLAPGGTAVW S90 GNQ1CVFFVU GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKS ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTATATTPT QYTAATAPT QYTAATAPT QTTEHASPL QTTEHASPL QTTEHASPL MNATTNLGQN NNATTNLGQN NNATTSLGTN NNATTALGQN	410 NPAIAYGTYQ NPAIAYGTYQ NPAIAYGSTQ NPAIAYGSTQ NPAIGYGNTS NPAIAYGSTQ IGHVGPMMGQ IGHVGPMMGQ IGHVGPMMGK NGNVGPMMAK 610 VYLTGNVAEL IYIUGNVYEL LYMYCNVNEL VYLTGNVAEL	420 QRWINDDVYI QRWINDDVYI ERWINDDVI ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDGE PGVTITIDGR GNVTIGE AGVTITIDGR 62/ GNWDTSK-AI(GNWDTSK-AI(GNWDFAK-AIG GNWTGSTAI- GNWDPAK-AIG	430 YERKFGNVA YERKFGNVA YERKFGSNVA YERKFGSNVA YERKFGNVA GFGTSGVU GFGTTSGVL GFGSTGTVG GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY 3 PMFNQVVYQ 3 PMFNQVVYQ 3 PMFNQVVYQ 3 PMFNQVVYQ	440 UVAINRILST LVAINRILST LVAVNRLST LTAVNSG-DT VVAINRINT FGTTAGTD FGTTAGTD FGTTAVGAD FGTTAVGAD FGTTAVGAD 0 640 Y PTWYDVSU Y PTWYDVSU Y PTWYDVSU Y PTWYDUSU Y PTWYDLSU Y PTWYDLSU Y PTWYDLSU	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINLINTS PASITGLVTS SVS WDDTEVKU VS WSDNQIVU TS WEDTQIKU VIS WSDTKIEV VIS WSDTKIEV VIS WSDTKIEV D 65(PAGTIGFKFI PAGKQLEFKFI PAGKQLEFKFI PAGKQLEFKFI PAGQTIEFKFI PAGQTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN SV PNVSPGKYN KI PAVAGGNYN KI PAVAGGNYN KI LRVPGGIYE C KKNG-STYTWE KKNG-STYTWE L KKDSGNVVWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTIQ GULNGNIT QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSSGATS I TVQSCSCQTS I TVQSCSCQTS I KVANAAGAAS V KVA-ASGVNS I SVVNAGDSQS I RVANAAGAAS 670 CGGSNHTYTP CGGSNHTYTP SGSNHTFTP SGNNHTYTP CGGANRTFTP CGGANRTFTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 PC251 BC5 BC55 BC8 BOHB BSP382	380 DPYNRRAMTSS DPDNRRAMTSS DPDNRRAMSS DPENRKMSD DPENRKMSD DPENRKMSD DPDNRARIPS 480 VASDGSVTPF VGSNGSVNFF VGSNGANSF VGSGAASNF VGSGAASNF S80 NTYNNINILT AYDNFEVLS NAYNNFFULT PYDKFEVLT NIYDNFEVLS	390 FNTSTTAYIN FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTINSYQI FSTTTTAYQV LQGQEVGVW ULAAGGTAVW ULAAGGTAVW ULAAGGTAVW ULAAGGTAVW GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISKLAPLRKS 500 QVXSSNSPL AYSATESTPI QYTAATATPT QYTAATATPT QITEEHASPL QITEDATAPI QUTTAATNEGON NNASTVYGEN NNASTNLGON NNASTLLGON NNASTLLGON	410 NPAIAYGTQ NPALAYGDTE NPAIAYGSTQ NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPTMTK IGHVGPTMKK IGHVGPTMKK IGHVGPTMKK IGHVGPTMKK 610 VYLTGNVAEL LYLTGNVAEL LYLTGNVAEL LYLTGNVAEL	420 QRWINDDVYI QRWINDDVYI ERWINDDVYV ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDGE PGNVTIDGE GNVDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDPQ-AI GNWDPQ-AI	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTTSGVL GFGTTSGVL GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY O 63 PMFNQVVYQ S PMFNQVVYQ S PMFNQVVYQ S PMFNQVVYQ S PMFNQVVYQ S PMFNQVVYQ	440 LVAINRNLST LVAVNRSSS VVAVNRLNA LTAVNSG-DT VVAINRNMT FGSTAGTT FGSTAGTTAVTGAD FGTTAVTGAD FGTTAVTGAD 0 644 Y PTWYYDVSU Y PTWYYDVSU Y PTWYYDVSU Y PTWYYDVSU Y PTWYYDVSU Y PTWYYDVSU	450 SYNITGLYTA NYSITGLFTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV VS WSNNQIVV UTS WEDTQIKV VS WSNNQIVV UTS WEDTQIKV VA WEDTQIQV 0 650 PAGTIOFKFI P AGQLEFKFI P AGQLEFKFI P AGQTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQQYTDEL Q LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q Source Contemporation (K PSVTPGKYN KI PAVAGGNYN KI PAVAGGNYN KI LRVPGGIYE (KKNG-NTITWE KKNG-STVTWE KKNG-STVTWE KKNG-STVTWE KKQG-STVTWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNETT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I TVQSSCQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAAGSAS 670 CGGSNHTYTPP CGGSNHTFTAP SGSNHTFTTP SGNHTFTTP CGGANRTFTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382	380 DPYNRAMMSS DPDNRRAMMSS DPDNRARMSS DPDNRARMSD DPENRKPMSD DPENRKPMSD DPDNRARIPS 480 VASDGSVTPF VGSGGAASNF -STNGSINNF VGSGGAASNF -STNGSINNF VMSNGAVDSF VGAGGASNF 580 NTYNNINILT AAYDNFEVLT NYYDNFEVLT NIYDNFEVLT NIYDNFEVLT	390 FNTSTTAYDV FSKSTTAYDV FSKSTTAFNVV FSKSTTAFNV FDRTTNSV0I FSTTTTAYOV ULSAGEVAVW DLGPGEVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAPGGTAVW TLAPGGTAVW SOQUSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 0QVVSSSNSPL AYSATESTPI QVTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT 000 NNASTVGEN NNATTALGQN NNATTALGQN NNATTALGQN NNATTALGQN	410 NPALAYGTQ NPALAYGDTE NPALAYGSTQ NPALAYGSTQ NPALAYGSTQ S10 IGHVGPTMTK IGHVGPTMTK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK CHUGPTGSVEL LYTCGNVEL LYTCGNVEL Similarity	420 QRWINDUYI QRWINDUYI ERWINDUJI QRWINDUYI ERWINDUYI ERWINDUYI S20 AGQTITIDGR VGHQVTIDGE HGNTVTITGE AGVVTIDGR GNWDTSK-AI GNWDTSK-AI GNWDTSK-AI GNWDPAK-AI GNWDPAK-AI GNWDPAK-AI Scores (%)	430 YERKFGNVA YERKFGNVA YERKFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTNGTVK GFGSTKGTVY GFGSKGTVY GFGSKGTVY GFGDNEGSVL A-SARQGTVY 0 63 PMFNQVVYQ PMFNQVVYQ 3 PMFNQVVYQ 5 PMYNQVVYQ 3 PMFNQVVYQ 5 PMYNQVVYQ	440 LVAINENLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGTJ FGSTAGTJ FGTTAVTGADJ FGTTAVTGADJ FGTTAVTGADJ 0 640 Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINLINTS PASITGLVTS 640 5 US WDDTEVKV VTS WEDTQIKV VTS WEDTQIKV VTS WEDTQIKV VLS WSDTKIEV VLS WSDTKIEV PAGTTIQFKFI PAGTTIQFKFI PAGKLEFKFI PAGKLEFKFI PAGQTIEFKFI PAGQTIEFKFI	460 LPAGTYTDUL G LPAGTYTDUL G LPQGSYNDUL G LPTGSYTDUL G LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q S0 56 KV PSVTPGKYN XI PAVAGGNYN XI PAVAGGNYN XI PAVAGGNYN XI PAVAGGYY XI LRVPGGIYL 0 660 KKNG-NTITWE KKDG-STITWE KKNG-STITWE KKNG-STVTWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GULNGNTIT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSSQOTS I TVQSSSQOTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAGDSQS GGSNHTYTVP SGSNHTYTTP SGSNHTTTP CGGNNTFTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM	3800 DPYNRKAMSS DPDNRKAMSS DPDNRKMSS DPDNRAKMSS DPDNRAKPS DPDNRARIPS 480 VASDGSVTPF VGSGGAASMF -STNGSINNF VSSGGAASMF S80 NTYNNINILT AAYDNFEULT NYYDNFEULS NAYMFFULT NIYDNFEULT 680 SSSTGTVIVM	390 FNTSTTAYDV FSKSTTAYDV FSKSTTAFVV FSKSTTAFVV FSKSTTAFVV USSKSTAFVV USSKSVV TLAAGGTAW TLAAGGTAW TLAAGGTAW USSKSVV ODQVSVRFVV ODQVSVRFVV GDQVTVRFVI GDQVSVRFVV GDQVSVRFVV GDQVSVRFVV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKS ISKLAPLRKS ISTLASLRQN IQKLAPLRKS 500 QYVSSSNSPL AYSATESTPI QYTATATTPT QYTAETTPT QYTATATTPT QYTATATTPT QTTEHASPL QTTEHASPL QTTEHASPL MNATTNLGQN NNASTTLGQN NNASTTLGQN NNASTTLGQN NNATTALQQN NNATTALQQN al aa 83	410 NPAIAYGTTQ NPAIAYGTTQ NPAIAYGTTQ NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPMMSQ IGHVGPMMSQ IGHVGPMMSK IGHVGPMMSK NGNVGPMMSK NGNVGPMMSK ICHVGPMSEL IYIIVGNVFEL IYIIVGNVFEL LYMVGNVNEL Similarity	420 QRWINDDVYI QRWINDDVYI ERWINDDVYI ERWINDDVYI ERWINDDVYI ERWINDDVYI S20 AGQTITIDGR PGNVTIDE PGVTITIDGR GNVDTSK-AIG GNWDPSK-AIG GNWDPSK-AIG GNWDPSK-AIG GNWDPSK-AIG GNWDPAK-AIS GNWDPAK-AIS GNWDPAK-AIS GNWDPAK-AIS Scores (%)	430 YERKFGNVA YERKFGNVA YERKFGSNVA YERKFGSVA YERKFGNVA 530 GFGTTSGVL GFGTTSGVL GFGSTGTVY GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY 0 63 9 MFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ	440 UVAINRILST LVAINRILST LVAVNRLST LTAVNSG-DT VVAINRINT FGSTAGT FGSTAGT FGTTAVSGAD FGTTAVSGAD FGTTAVTGAD 0 64(Y PTWYDVSV) Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDLSVI Y PTWYDLSVI Y PTWYDLSVI	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS VS WDDTEVKU VVS WSNNQIVU TS WEDTQIKU UVS WSDTKIEV UVA WEDTQIQU D 650 PAGTIGFKFI P AGKTIEFKFI P AGKTIEFKFI P AGKTIEFKFI P AGQTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGQYTDEL Q LPQGQYTDEL Q LPQGSYNDVL G 50 56 KV PSVTPGKYN XV PNVSPGKYN XV PNVSPGKYN XV PNVSPGKYN XV PDVTAGHYD KI LRVPGGIYE 0 660 KKNG-NTITWE KKDSGONVTWE KKDSGONVTWE KKDGSTVTWE	470 GLLNGN51S GLLDGNTIQ GLLNGNTIQ GULNGNIT QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSSGATS I TVQSCSCQTS I TVQSCSCQTS I TVQSCSCQTS I SVVNAGDSQS I SVVNAGDSQS I SVVNAGDSQS GGSNHTYTVP SGGSNHTYTVP SGSNHTYTTP SGGNHTYTPP SGANRTFTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC51 BC8 BOHB BSP382	380 DPYNRAMTS DPDNRRAMTS DPDNRRAMSS DPDNRARTPS DPENRKMSD DPENRKMSD DPENRKMSD DPENRATPS 480 VASDGSVTPF VGSNGSVNFF VGSNGANSF VGSGAASNF VGSGAASNF S80 NTYNNINII AYDNFEVLS NAYNNFFILT PYDKFEVLT NIYDNFEVLS SSSTGTVIVM TNTTGKIIVD	390 FNTSTTAYIN FNKNTRAYQV FSTSTTAYQV FSKSTTAFNV FDRTINSYQI FSTTTAYQV ULSAGEVGVW ULSAGEVGVW ULAAGGTAVW ULSAGGTAVW ULSANGVSVRFVU GDQVSVRFVV FV FV FV FV FV FV FV FV FV FV FV FV	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS 500 QVXSSNSPL AYSATESTPI QYTAATATPT QYTAATATPT QTTEEHASPL QTTAETTPT QTTAETTPT QTTAETTPT QTTAATATPA NAATTLGON NNATTLGON NNATTLGON NNATTLGON NNATTLGON ZAL aa 83 30	410 NPAIAYGTTQ NPAIAYGTTQ NPAIAYGTTC NPAIAYGTTQ NPAIAYGSTQ S10 IGHVGPTMTK IGHVGPTMTK IGHVGPTMTK IGHVGPTMMK IGHVGPTMMK IGHVGPTMMK G10 VYLTGNVAEL LYLTGNVAEL LYLTGNVAEL LYLTGNVAEL Similarity 55.9	420 QRWINDUYI QRWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI S20 AGQTITIDGR PGNVTIDGE PGNVTIDGE PGNVTIDGE GNWDTSK-AIG GNWDTSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDPQ-AI GNWDPQ-AI Scores (%)	430 YERKFGNVA YERKFGNVA YERFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTTSGVL GFGSTSGVL GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY O 63 PMFNQVVYQ O 63 PMFNQVVYQ O 94 PMFNQVVYQ O PAFNQVHQ O PMFNQVVYQ O PMFNQVVYQ O PMFNQVVYQ O PMFNQVVYQ O 0 0 O 0 O 0 O 0 O 0 O 0 O 0 O	440 LVAINRNLST LVAVNRSSS VVAVNRLNA LTAVNSG-DT VVAINRNMT FGSTAGTT FGSTAGTTAVTGAD FGTTAVTGAD FGTTAVTGAD 0 644 Y PTWYIDVSVI Y PTWYIDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI	450 SYNITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV VS WSNNQIVV TS WEDTQIKV VS WSNQIVV US WSDTVIK VA WEDTQIQV 0 650 PAGTIOFKFI P AGKQLEFKFI P AGKQLEFKFI P AGGTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q Source Contemporation (KV PSVTPGKYN KI PAVAGGNYN KI PAVAGGNYN KI LRVPGGIYE (KKNG-NTITWE KKNG-STVTWE KKNG-STVTWE KKNG-STVTWE KKQG-STVTWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNETT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I TVQSSCQTS I KVANAAGTAS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAAGSQS GGSNHTYTPP SGSNHTFTAP SGSNHTFTTP CGANRTFTTP
TBACTERIUM BST2 BC251 BC8 BOHB BSF382 TBACTERIUM BST2 BC251 BC8 BOHB BST22 BC251 BC8 BOHB BST282 TBACTERIUM BST2 BC8 BOHB BSF382	380 DPYNRAMMSS DPNNRKMSS DPDNRAKMSS DPDNRAKMSS DPENRKMSD DPENRKMSD DPENRKMSD DPENRKMSD DPENRKMSD DPENRKMSD STOGSUNF VGSGGASNF -STNGSINNF VGSGGASNF -STNGSINNF VGSGGASNF SSNGAUDSF VGSGGASNF -STNGSINNF VNSNGAUDSF VGSGGASNF SSTTVIVN NYDNFEULT NYDNFEULT NYDNFEULT NYDNFEULT SSTTTVIVN SSGTATINVN	390 FNTSTTAYINV FNKNTRAYQV FSKSTTAYQV FSKSTTAYQV FSKSTTAYQV TSKSTTAYQV TLSAGEVAVW DLGPCeVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW GDQVSVRFVV F	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS S500 QVVSSSNSPL AYSATESTPI QVTAATTST QVTTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT QVTTAETTPT ONASTVGEN NNATTNLGQN NNATTNLGQN NNATTLGQN NNATTLGQN NNATTLGQN NNATTLGQN NNATTLGQN S80 S80	410 NPAIAYGTQ NPAIAYGTE NPAIAYGTC NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPTMTK IGHVGPTMTK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK IGHVGPTMGK S10 VYLTGNVAEL LYTCGNVAEL LYTCGNVAEL Similarity 55.9 59.2 2022	420 QRWINDV1 QRWINDV1 ERWINDV1 ERWINDV1 520 AGQTITIDGR VGHQVTIDGE HGNTVTITGE AGVVTIDGR HGNTVTITGC GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDPAK-AIG GNWDPAK-AIG GNWDPAK-AIG Scores (%)	430 YERKFGNVA YERKFGNVA YERKFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTNGTVK GFGTNGTVK GFGSKGTVY GFGSKGTVY GFGDNEGSVL A-SARQGTVY 0 63 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ 9 PMFNQVVYQ	440 LVAINENLST LVAINENLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINENMT FGSTAGTJ FGSTAGTJ FGTTAVTGADJ FGTTAVTGADJ FGTTAVTGADJ 0 640 Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ Y PTWYDVSVJ	450 SYNITGLYTA NYSITGLYTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 340 5 US WDDTEVKV VS WSNNQIVV VS WSNNQIVV VS WSNDTEVKV VS WSDTQIKV VLS WSDTQIKV VLS WSDTQIKV VLS WSDTQIKV P AGTTIQFKFI P AGTTIQFKFI P AGQTIEFKFI P AGQTIEFKFI P AGQTIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPQGSYNDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q S0 50 56 KV PSVTPGKYN XI PAVAGGNYN XI PAVAGGNYN XI PAVAGGNYN XI LRVPGGIYL 0 660 KKNG-NTITWE KKDSGONYTWE KKNG-STITWE KKNG-STITWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTIQ GLLNGNIT QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSGATS 1 TVQSSSCQTS 1 KVANAAGTAS 1 TVQSSCQTS 1 KVANAAGTAS 670 670 CGGSNHTYTP 2 GGSNHTYTP 2 GGANRTFTP 2 GGANRTFTP 2 GGANRTFTP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 RC251 BC8 BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382	3800 DPYNRKAMSS DPDNRKAMSS DPDNRKAMSS DPDNRAKMSS DPDNRAKPS DPDNRARIPS 480 VASDGSVTPF VGSGGASNF VGSGGASNF VGSGGASNF VNSNGAVDSF VGSGGASNF STOGINDF VNSNGAVDSF VNSNGAVDSF SST NAYNNTILT PTYDKFEVLT NYDNFEVLS SSSTTVIVN TNTTGKIIUD SSGTATINVN ASGTATVTVN	390 FNTSTTAYDV FNKNTRAYQV FSKSTTAFNV FSKSTTAFNV FSKSTTAFVV USSURFVV ULAAGGTAVW ULAAGGTAVW ULAAGATAVW ULAAGATAVW GDQVSVRFVV F	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS S00 QVVSSSNSPL QYXSSNSPL QYTAATATPT QYTAATATPT QYTAATATPT QUTTEHASPL QUTTEHASPL QUTTAATAPI NAATTALGON NNATSLIGON NNATSLIGON NNATSLIGON NNATSLIGON NNATSLIGON NNATSLIGON S83 S80 S86 S84	410 NPAIAYGTTQ NPAIAYGTTQ NPAIAYGTTQ NPAIAYGTTQ NPAIAYGSTQ NPAIGYGNTS NPAIAYGSTQ IGHVGPMMGQ IGHVGPMMGQ IGHVGPMMGK NGNVGPMMGK NGNVGPMMAK 610 VYLTGSVSEL LYLVGNVAEL LYLVGNVAEL LYMVGNVEL Similarity 55.9 59.2 57.3	420 QRWINDUYI QRWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI S20 AGQTITIDGR PGNVTIDGE PGVTITIDGR GNUTISK-AIG GNWDTSK-AIG GNWDTSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG SCOPES (%)	430 YERKFGNVA YERKFGNVA YERKFGSNVA YERKFGSVA YERKFGSVV YERKFGNVVA GFGTTSGVL GFGTTSGVL GFGSTKGTVY GFGSTKGTVY GFGSTKGTVY OFGSTKGTVY OFGSTKGTVY OFGSNEGSVL A-SARQGTVY OFGSNEGSVL A-SARQGTVY OFGSNEGSVL A-SARQGTVY OFGSNEGSVL A-SARQGTVY OFGSNEGSVL OFANQVYQ OFMNQVVYQ OFMNQVVYQ OFMNQVVYQ	440 UVAINRNLST LVAINRNLST LVAVNRSSS VVAVNRLST LTAVNSG-DT VVAINRNMT FGSTAGTT FGSTAGTT FGTTAVSGAD FGTTAVSGAD FGTTAVTGAD 0 64(Y PTWYDVSV) Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI Y PTWYDVSVI	450 SYNITGLYTA NYSITGLFTA PASISGLVTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 SVS WDDTEVKU VVS WSNNQIVU TS WEDTQIKU UTS WEDTQIKU UTS WEDTQIKU US WEDTQIKU D 650 PAGTIFEKFI P AGKTIEFKFI P AGKLEFKFI P AGQLIEFKFI	460 LPAGTYTDVL G LPAGTYTDVL G LPAGTYTDVL G LPTGSYTDVL G LPTGSYTDVL G LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q SO 50 50 KV PSVTPGKYN TI PSVAGGYTN TI PSVAGGYTN TI PSVAGGYTN KI LRVPGGIYE MAGGYTAGHYE KKDSGONTME KKDSGONTME KKDSGONTME KKDSGSNVWE KKQG-STVTWE	470 GLLNGN51S GLLDGNTIQ GLLNGNTIG GULNGNIT QLLDGNEIT GILNGNTLT. 0 570 1 SLKTSSGATS 1 TVQSSCQTS 1 KVANAAGTAS 1 SVVNAGDSQS 1 RVANAAGAAS 670 CGGSNHTTTVP CGGSNHTFTAP 2 SGSNHTFTAP 2 SGSNHTFTAP
TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 RC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB BSP382 TBACTERIUM BST2 BC251 BC8 BOHB	380 DPYNRAMTS DPDNRRAMTS DPDNRRAMSS DPDNRARTPS DPDNRARTPS 480 VASDGSVTPF VGSGGASNF -STNGSINNF VNSGGASNF STNGSINNF VNSGGASNF S80 NTYNNINILT AXDNFEVLS NYDNFEVLS NAYNNFTLLT PTYDKFEVLT NIYDNFEVLS SSSTGTVIVN ASGTATVTVN ANGTDTVLUZ	390 FNTSTTAYINV FNKNTRAYQV FSKSTTAYQV FSKSTTAYQV FSKSTTAYQV TSKSTTAYQV UGPCEVGVW ULGPCEVGVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW TLAAGGTAVW GDQVSVRFVV GDQV GDQVSVRFVV GDQVSVRFVV GDQV GDQV GDQV GDQV GDQV GDQV GDQV GD	400 IKKLAPLRKS IQKLSSLRRN IQKLAPLRKC ISKLAPLRKS ISKLAPLRKS 500 QVXSSNSPL AYSATESTPI QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT QVTAETTPT OCCUP NNATTNLGON NNASTVLGON NNASTNLGON NNATTSLGON NNATTSLGON ZNASTLGON Z	410 NPAIAYGTQ NPAIAYGTC NPAIAYGTC NPAIAYGSTQ NPAIAYGSTQ S10 IGHVGPTMTK IGHVGPTMTK IGHVGPTMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK IGHVGPVMGK S10 VYLTGNVAEL LYLTGNVAEL LYLTGNVAEL LYLTGNVAEL Similarity 55.9 55.2 57.3 43.1 77.1	420 QRWINDUYI QRWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI ERWINDUYI S20 AGQTITIDGR PGNVTIDGE PGVVITIGE AGVTITIGE GNWDTSK-AIG GNWDTSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG GNWDFSK-AIG SCOTES (%)	430 YERKFGNVA YERKFGNVA YERKFGSVA YERKFGSVA YERKFGSVA YERKFGNVA 530 GFGTTSGVL GFGTTSGVL GFGSKGTVY GFGSKGTVY GFGSKGTVY GFGSKGTVY A-SARQGTVY 0 63 PMFNQVVYQ PMFNQVVYQ PAFNQVHQ PAFNQVHQ PAFNQVHQ PMFNQVVYQ	440 LVAINRNLST LVAINRNLST LTAVNSG-DT VVAINRNMT FGSTAGTT VVAINRNMT FGTTAVTGAD FGTTAVTGAD FGTTAVTGAD 0 644 Y PTWYYDVSVI Y PTWYYDVSVI Y PTWYYDVSVI Y PTWYYDVSVI Y PTWYYDVSVI Y PTWYYDVSVI Y PTWYYDVSVI	450 SYNITGLYTA NYSITGLFTA PASISGLYTS SASITGLSTS SYTINNLNTS PASITGLVTS 40 5 VS WDDTEVKV TS WEDTQIKV TS WEDTQIKV US WSDTKIEV US WEDTQIKV LS WEDTQIKV A WEDTQIQV 0 650 PAGTIOFKFI PAGKQLEFKFI PAGKQLEFKFI PAGKQLEFKFI PAGQTIEFKFI	460 LPAGTYTDUL G LPAGTYTDUL G LPAGTYTDUL G LPTGSYTDUL G LPTGSYTDUL G LPQGYTDEL Q LPQGYTDEL Q LPQGYTDEL Q S50 56 KV PSVTPGYN KI PAVAGGNYN KI PAVAGGNYN KI PAVAGGYY KI LRVPGGIYE C KKDG-STYTWE KKDG-STYTWE KKDG-STYTWE KKDG-STYTWE	470 GLLNGNSIS GLLDGNTIQ GLLNGNTLS GVLNGNNTT QLLDGNEIT GILNGNTLT. 0 570 I SLKTSSGATS I TVQSSCQTS I KVANAAGTAS V KVA-ASGVNS V KVA-ASGVNS V KVA-ASGVNS I SVVNAGDSQS I SVVNAGDSQS I SVNAAGSQN GGSNHTYTVP SGSNHTFTAP SGSNHTFTTP GGANRTFTTP

This study	100 Gly GGT	Tyr TAT	Trp TGG	Ala GCT	Arg CGT	Asp GAT	Phe TTT	Lys AAG	Arg AGA	110 Thr ACA	Asn AAT	Pro CCA	Tyr TAC
Bahl et al. (1991)	Gly GGT	-	Trp TGG	Ala GCT	Arg CGT	Asp GAT	Leu TTA	-	Arg AGA	Glu GAA	Gln CAA	Ser TCA	Tyr TAC
This study	362 Thr ACA	Glu GAA	Gln CAG	Tyr TAT	Met ATG	Thr ACA	Gly GGC	Asn AAT	370 Gly GGA	Asp GAC	Pro CCT		
Bahl et al. (1991)	Thr ACA	Val GTA	-	Tyr TAT	-	Asp GAC	Arg AGG	Gln CAA	Trp TGG	Arg AGA	Pro CCT		

Figure 1. (a) Alignment of amino acid sequences of bacterial CGTases. BC251, *B. circulans* strain 251 (Lawson *et al.*, 1994); BC8, *B. circulans* strain 8 (Nitschke *et al.*, 1990); BOHB, *B. ohbensis* (Sin *et al.*, 1991); BSP382, *Bacillus* sp. strain 38-2 (Kaneko *et al.*, 1988); BST2, *B. stearothermophilus* strain 2 (Fujiwara *et al.*, 1992); TBACTERIUM, *T. thermosulfurigenes* EM1 (this study; Bahl *et al.*, 1991). Amino acid numbering is started after the signal peptide and according to TBACTERIUM. Similarity scores are given compared with the CGTase from *T. thermosulfurigenes* EM1. (b) Differences between the amino acid sequence reported by Bahl *et al.* (1991) and the revised sequence of the *T. thermosulfurigenes* EM1 CGTase.



Figure 2. Stereo views of the corrected portion of the amino acid sequence of the *T. thermosulfurigenes* EM1 CGTase as present in the refined model with corresponding electron density contoured at 1σ in σ_a -weighted $2F_o - F_c$ maps. Residues Tyr101, Lys107 to Pro111 and Glu363 to Asp371 are indicated.

static effects and stabilization of calcium binding loops by, for instance, the stacking of Tyr112 onto Pro34 at the second calcium binding site could contribute to tight calcium binding.

Table 1. Hydrogen bonds observed between symmetryrelated protein molecules in crystals of *Thermoanaerobacterium thermosulfurigenes* EM1 CGTase

		0			
Residue	Atom	Residue	Atom	Distance (Å)	Molecule
Asn443	$N^{\delta 1}$	Thr666	0	2.9	Ι
Pro479	0	Thr668	$O^{\gamma 1}$	2.7	Ι
Ser31	Ογ	Ser552	0	3.3	II
Asn62	0	Gln646	$N^{\epsilon 1}$	3.2	II
Asn62	$N^{\delta 2}$	Asn606	$O^{\delta 1}$	3.4	II
His129	$N^{\epsilon 2}$	Asn664	$N^{\delta 2}$	3.1	II
Asp183	$O^{\delta 2}$	Asn416	$N^{\delta 1}$	2.9	III
Asn264	$N^{\delta 2}$	Gln629	$O^{\epsilon 2}$	3.2	III
Asn297	0	Gln625	$N^{\epsilon 2}$	2.7	III
Asn335	0	Tyr597	Ν	2.8	III
Asn335	$N^{\delta 2}$	Thr595	0	2.7	III

Molecule I is the reference molecule that contacts its copy, which is shifted by one unit translation along the a axis.

Symmetry operations generating crystal contacts:

I = x + 1, y, z

II = -x + 1, y - 1/2, -z + 1/2III = x + 1/2 y + 1/2 z

III = x + 1/2, -y + 1/2, -z

The identity and conformation of active site residues near the catalytic carboxylates Asp230, Asp329 and Glu258 is highly similar to that observed in the other CGTase structures. At position 193 near the active site Lys is replaced by Arg with respect to the B. circulans CGTases, which may render the enzyme less sensitive to chemical modification by glucose (Mrabet et al., 1992). A second difference seen in the active site of the T. thermosulfurigenes EM1 CGTase is that the hydrogen bond between the side-chain carboxyl groups of Glu257 and Asp328, which is present in the B. circulans strain 251 and the other two CGTase structures, is lost. Instead the Glu258 side-chain has rotated about its χ_2 angle to assume a conformation similar to that observed in the complex between the B. circulans CGTase and the inhibitor acarbose (Strokopytov et al., 1995). The aromatic residues that take part in oligosaccharide binding at the maltose binding sites in the BC251 CGTase are all conserved in the T. thermosulfurigenes EM1 CGTase. Leu600, however, which was observed to be near the centre of the ring of an α -CD molecule bound at maltose binding site 2 of the BC251 CGTase (Knegtel et al., 1995), has been replaced by Tyr597.



Figure 3. Stereo view of the superimposed C^{α} backbone traces of the crystal structures of the CGTases from *B. circulans* strains 8 and 251, *B. stearothermophilus* and *T. thermosulfurigenes* EM1. The backbone of the *T. thermosulfurigenes* EM1 CGTase is drawn with thick lines. Maltose binding sites 1 to 3 (MBS1 to 3) and variable loop regions are indicated. The centre of the active site is indicated with an asterisk (*).

Structure and thermostability

The origins of thermostability in proteins have been subject of extensive debate in the literature (Fontana, 1991; Jaenicke, 1991; Rees & Adams, 1995). We have examined several proposed causes for the enhanced stability of thermostable proteins. An increase in the number of disulphide bridges has in some cases increased the stability of proteins. The two cysteine residues that formed a disulphide bridge in the *B. circulans* CGTases are replaced by His43 and Phe50, respectively, in the T. thermosulfurigenes EM1 CGTase. These two residues stack onto each other to form a small hydrophobic core together with Met377, while additional rigidity in this region is provided by the presence of Pro41. The CGTase from B. stearothermophilus still has the disulphide bridge present, although in this case the free cysteine residue at position 400 in the BC251 CGTase has been deleted from its sequence.

Recent structural studies of thermophilic enzymes have suggested that an increase in surface salt-bridges could be responsible for structural stability of proteins at elevated temperatures (Chan et al., 1995; Day et al., 1992). The heat-labile CGTases from B. circulans strains 251 and 8 have a total of 27 and 16 salt-bridges, respectively (using a 3.5 Å distance limit for the Arg, Lys, Asp and Glu charged side-chain atoms), while the T. thermosulfurigenes EM1 CGTase has 20. Although the BC251 CGTase has a larger total number of ionic interactions, the T. thermosulfurigenes EM1 CGTase has five salt-bridges that have no homologues in the B. circulans CGTases. These salt-bridges involve residues 47/89, 189/193, 228/230, 245/510 and 276/557, all of which are located near or at the protein surface and possibly stabilize the B/E and A/D domain interfaces. The thermophilic B. stearothermophilus CGTase has a total of 18 salt-bridges, of which six unique salt-bridges are located at its surface, involving residues 113/116, 336/461, 99/149, 205/241, 212/241 and 406/408. Only the 99/149 salt-bridge appears to be involved

in stabilization of a domain interface. Thus, it is not only the number of surface salt-bridges that is important for conferring stability, but probably also the precise location of a salt-bridge might contribute to the overall stability of the enzyme (e.g. by providing interactions between domains or surface loops).

Other domain-interface stabilizing interactions are observed near Phe237, which assumes a conformation 120° rotated compared with that of the corresponding phenylalanine residue in the BC251 CGTase, as shown in Figure 4. This change in conformation results from a mutation of Thr to Tyr at position 187, and as a consequence a small hydrophobic core is formed involving Tyr272 and Phe273. Tyr272 is absent in both B. circulans CGTases and the observed contacts appear to improve the packing of two adjacent α -helices in the Å domain. A similar arrangement of aromatic residues is observed in the B. stearothermophilus CGTase. The space that is occupied by Phe237 in the BC251 CGTase structure is taken by Tyr187 in the T. thermosulfurigenes EM1 CGTase, where its side-chain makes van der Waals contacts with Phe623 and a water-mediated hydrogen bond with the hydroxyl group of Tyr635. The latter two residues are located in the E domain and these interactions may stabilize the domain interface.

An increased overall rigidity and compactness have been suggested to improve the stability of proteins under extreme conditions (Fontana, 1991; Vihinen, 1987). Thermostable proteins are assumed to be more rigid at moderate temperatures. A decreased flexibility is faintly reflected in the lower average *B*-factor for the protein atoms of the *T. thermosulfurigenes* EM1 (20.6 Å²) and the *B. stearothermophilus* CGTases (12.7 Å²), compared with those of the *B. circulans* strains 251 (26.1 Å²) and 8 (22.2 Å²) CGTases. The differences are quite small, however, and could equally well be caused by differences in crystal packing and solvent content. Of the loops with the largest conformational differences, three have clearly reduced



Figure 4. Stereo view of the rearrangement of aromatic residues at the A/E domain interface. The *T. thermosulfurigenes* EM1 structure is drawn in bold, the *B. circulans* strain 251 CGTase structure with thin lines. Aromatic residues of the thermostable protein are labelled. The side-chain of Phe237 is rotated such that it contacts Tyr272 and Phe273, while Tyr187 contacts Phe623 and Tyr635 from the E domain.

temperature factors (residues 42, 89 to 93 and 335 to 336) but, in contrast, the temperature factors of loop 655-656 are twice as high (58.3 Å² compared with 29.8 $Å^2$). Some rigidity may be conferred to the enzyme by the Gly to Ala or Pro substitutions at positions 41, 472 and 555. The compactness of thermostable protein structures or the reduction of the number and volume of cavities has been shown to affect the stability of aldehyde ferredoxin oxidoreductase from *P. furiosus* (Chan et al., 1995) and citrate synthase from T. acidophilum (Russell et al., 1994). We calculated cavity sizes and total volumes with VOIDOO, using the program's standard settings (Kleywegt & Jones, 1994). The results of these calculations do not agree with the expected trend where the thermostable protein has the smallest cavity and total volume. For instance, the thermolabile BC251 CGTase had the smallest total cavity volume of 45 Å³ compared with 117 Å³ for the T. thermosulfurigenes EM1 CGTase, while the latter had a slightly larger total volume of 69,060 Å³ $(68,290 \text{ Å}^3 \text{ for the BC251 CGTase}).$

As overall structural features conferring thermostability to the *T. thermosulfurigenes* EM1 CGTase, we considered loop regions that differ from those in the thermolabile CGTases. On the basis of the amino acid sequence no obvious trend in reduction or increase of the length of loops in the thermostable CGTases is observed. When the structures of the loops carrying insertions or deletions are examined in more detail, some stabilizing new contacts are observed. The loop region consisting of residues 88 to 95 is one residue longer compared with the same loop in the *B. circulans* CGTases (Figure 1). Therefore one might expect that this loop plays no significant role in the stabilization of the enzyme. The conformation of this loop, however, allows Phe92 to form a small hydrophobic core with Pro372, Tyr373, Leu46 and the methylene groups of the Lys47 side-chain. In the B. circulans CGTases there is no apolar contact attaching this loop to the core of the protein. The 335-339 loop region is one residue shorter than the equivalent *B. circulans* loop regions but the thermostable *B. stearothermophilus* CGTase has no deletion in this region. In the T. thermosulfurigenes EM1 CGTase additional stabilization of this loop is achieved by hydrogen bonds between Asn335 and Asn369/Arg295, and hydrophobic contacts between Tyr334 and Gly336, Thr339 and Tyr365. The loop region consisting of residues



Figure 5. Stereo view of the C^{α} trace of the loop region Phe531 Trp541 in four different to CGTase crystal structures. CGTases are labelled with their PDB entry code as follows: B. circulans strain 8 (1CGT), B. circulans strain 251 (1CGD), B. stearothermophilus (1CYG) and T. thermosulfurigenes EM1 (TtEM1). The thermolabile CGTases 1CGD and 1CGT carry a three residue insertion in this region, which disrupts a β -sheet that is completed in the thermostable enzymes.

Thr534 to Val539 in the thermostable CGTases contains the largest deletion (three residues) with respect to the thermolabile CGTases (Figure 5), thereby completing an anti-parallel β -sheet in the D-domain that is disrupted in the *B. circulans* CGTases. The thermolabile CGTase from *B. ohbensis* (Figure 1) has a similar deletion at this position but the remaining loop contains larger residues (Asp, Phe) than in the thermostable proteins (Ser, Gly, Ala, Thr) which could influence the conformation and stability of this loop. In conclusion, it appears that especially loops 88-95, 335-339 and 534-539 may contribute to the thermostability in CGTases.

Argos et al. (1979) suggested that the exchange of Ser by Thr, Gly or Ala, and of Gly and Val by Ala stabilizes α -helices and thus the overall protein structure. Inspection of the amino acid composition of the thermostable CGTases compared with that of the mesophilic enzymes shows that in fact a decrease of almost 45% of the alanine content is observed for the thermophilic CGTases. In terms of absolute numbers the alanine residues are replaced by Thr (Val in the case of the *B. stearothermophilus* CGTase) and to a lesser extent by Ser. At nine positions (residues 210, 214, 246, 249, 287, 306, 310, 313 and 400) Ser is introduced in α -helical regions in the T. thermosulfurigenes EM1 CGTase, suggesting that the replacement of Ala by Ser and/or Thr is important enough for overall thermostability to compensate possible destabilization of α -helices. A similar trend in amino acid composition has been observed for α -amylases (Ikai, 1980). Both α -amylases and CGTases are extracellular proteins that bind and convert long oligosaccharide chains and the preference for a higher Ser and Thr content might be related to their functionality. Most Ser and Thr residues in the T. thermosulfurigenes EM1 CGTase are positioned on one side of the protein surface, near the catalytic site and sugar binding domains. In general, reduction of the hydrophobicity of the protein exterior is thermodynamically favourable due to the increased solvent entropy. Alternatively, the increased number of hydrogen bonding surface residues such as Ser and Thr near the active site could provide tighter non-specific binding of the enzyme to raw starch granules at high temperatures. Indeed, a sharp increase in the thermostability of CGTase has been observed in the presence of soluble starch (Wind et al., 1995). It seems unlikely that this is caused by the binding of substrates to the maltose binding sites, since these are structurally conserved in the T. thermosulfurigenes EM1 CGTase, except for Tyr597 in maltose binding site 2, which is Leu in all other structures. Aromatic residues like Trp and Tyr have been observed previously to be involved in substrate binding at the maltose binding sites of CGTase (Lawson et al., 1994). In the T. thermosulfurigenes EM1 CGTase the number of tyrosine residues located at the protein surface is increased with respect to the B. circulans CGTases. A total of 11 tyrosine residues has been introduced at the protein surface compared with the BC251 CGTase. Only

two of them are, however, present in both the *B. stearothermophilus* and *T. thermosulfurigenes* EM1 CGTases; namely, Tyr272, which assists in stabilizing the A/E-domain interface and Tyr442, which is located near maltose binding site 3 (near Trp413). In conclusion, the observed increased stability in the presence of starch could be caused by increased non-specific binding of amylose chains at the surface of CGTase.

Increased hydrolytic activity of the *T. thermosulfurigenes* EM1 CGTase

Kinetic studies of the CGTase from T. thermosulfurigenes EM1 showed that it has a fivefold increased hydrolytic activity at 60°C compared with the CGTase from *B. circulans* strain 251 at 50°C, while its cyclization activity is reduced by almost 50% (Wind et al., 1995). The coupling and disproportionation activities appear to be very similar. Does the structure of the enzyme offer an explanation for the observed differences in activity? The absence of the hydrogen bond between the active site carboxylates Asp328 and Glu257 (BC251 numbering) could be due to the crystallization conditions. The pH range of 7.6 to 8.0 that was used for crystallization is beyond the pH optimum, and the other three CGTase structures were obtained at lower pH values (Klein & Schulz, 1991; Kubota et al., 1991; Lawson et al., 1994). On the other hand, this hydrogen bond is absent in BC251 CGTase complexed with acarbose (Strokopytov et al., 1995) or substrate (Knegtel et al., 1995) and in the complex with a G9 inhibitor (B. Strokopytov et al., unpublished results).

A possible explanation for the higher hydrolysis rate and lower cyclization activity of the CGTase from T. thermosulfurigenes EM1 compared with that of *B. circulans* strain 251 is the change of the size and conformation of the loop region consisting of residues 88 to 95. The structure of BC251 CGTase complexed with a maltononaose inhibitor was recently solved and revealed a stacking interaction between the aromatic side-chain of Tyr89 and glucose 3 and 4 of the inhibitor (B. Strokopytov et al., unpublished results). As is shown in Figure 6, where the structure of the maltononaose inhibitor is superimposed on the T. thermosulfurigenes EM1 CGTase structure, this interaction is no longer possible. In the thermostable enzyme the 88-95 loop is directed away from the inhibitor and no aromatic residue is at a position capable of interacting with the inhibitor. Bound oligosaccharides will thus lack an important stabilizing interaction and, certainly at elevated temperatures, cyclization will be reduced due to the increased flexibility of the accepting end of the amylose chain. A decrease in cyclization activity implies an increase in hydrolytic activity, which is in agreement with the experimental results (Wind et al., 1995). In the B. stearothermophilus CGTase the role of Tyr89 in the B. circulans CGTases appears to be taken over by Met84 (Met87 in Figure 1), of which the sulphur atom is at the same position as the centroid of the aromatic ring of Tyr89.



Figure 6. Stereo views of a close-up of the active site of the CGTases from (a) *B. circulans* strain 251 and (b) *T. thermosulfurigenes* EM1 with a maltononaose inhibitor in the active site (Strokopytov *et al.*, 1996). BC251 CGTase complexed with the maltononaose inhibitor was superimposed on the *T. thermosulfurigenes* EM1 CGTase and the maltononaose inhibitor was transferred to the *T. thermosulfurigenes* EM1 CGTase structure. Catalytic site residues and residues involved in substrate binding and stabilization of the 88-95 (*T. thermosulfurigenes* EM1 numbering) loop are labelled and the C^{α} trace of the loop region is drawn in bold. The *T. thermosulfurigenes* EM1 CGTase lacks the hydrophobic interaction of Tyr89 with the substrate resulting in a reduced cyclization activity.

Conclusions

The crystal structure of the cyclodextrin glycosyltransferase of the thermophilic *T. thermosulfurigenes* EM1 has been solved at 2.3 Å resolution. Its overall fold is highly similar to that reported for CGTases of mesophilic bacteria and structural differences are observed only at surface loops. The *T. thermosulfurigenes* EM1 CGTase structure provides another example of the subtle interplay of electrostatic and apolar interactions that causes thermostability in proteins. A number of structural features that could be responsible for the increased stability of this enzyme have been identified. Five or six saltbridges, absent from the thermolabile CGTase structures, are introduced at the protein surface near domain interfaces. Loop regions consisting of amino acid residues 88 to 95, 335 to 339 and 534 to 539 also appear to add to the general stability of the enzyme, either by introducing new hydrophobic interactions or hydrogen bonds or by removing flexibility existing in the less stable CGTases. Rearrangement of aromatic residues near Phe237 appears to stabilize the A domain and its interface with the E-domain. Finally, the increased incidence of Ser, Thr and Tyr residues at the protein surface near the active site and oligosaccharide binding sites could improve non-specific binding of the enzyme to raw starch granules. This might explain the dramatic increase of thermostability upon addition of soluble starch. Other factors that have been implicated in conferring thermostability to proteins, like reduced cavity volumes, disulphide bridges and an increase in the number of helix-stabilizing

amino acid residues appear not to be of major importance for the stability of CGTases.

The increased hydrolytic and decreased cyclization activity of the T. thermosulfurigenes EM1 compared with that of the CGTase from *B. circulans* strain 251 appears to be mediated mainly by the conformational changes in the loop region consisting of residues 88 to 95. A number of apolar interactions, not present in the thermolabile CGTase structures, fix this loop to the hydrophobic core of the enzyme. Apparently, stabilization of this region in order to increase the overall thermostability of the enzyme weighs more heavily than optimal cyclization activity. The loss of the hydrogen bond between Glu258 and Asp329 could also be of influence on the hydrolytic activity but is more likely to be an artefact of the high pH of the crystallization conditions.

The availability of the crystal structure of the *T. thermosulfurigenes* EM1 CGTase can serve as a starting point for protein engineering studies aimed at increasing the thermostability of industrially used CGTases and the manipulation of the ratio between cyclization and hydrolytic activity in this class of enzymes.

Materials and Methods

Enzyme preparation and activity assays

Production and purification of the CGTase from *T. thermosulfurigenes* EM1 was done according to Wind *et al.* (1995). Activity measurements were performed according to Penninga *et al.* (1995) at the temperature optimum of 50°C for the *B. circulans* 251 CGTase and at 60°C for the *T. thermosulfurigenes* EM1 enzyme, respectively. The temperature optimum of this latter enzyme is 85°C.

Crystallization

Initial attempts at crystallizing the enzyme were performed with a Sparse Matrix (48 experiments) setup using hanging drops with a protein concentration of 3.4 mg/ml in 10 mM sodium acetate (pH 5.5) at room temperature. Intergrown crystals were obtained from either 0.1 M sodium citrate and 1.0 M ammonium phosphate in Hepes buffer at pH 7.6 or 0.8 M sodium potassium tartrate in Hepes buffer at pH 7.7. Refinement of the tartrate conditions showed that rod-like single crystals could be grown from 0.8 to 1.2 M sodium potassium tartrate at pH 7.3 to 7.7. Unfortunately these crystals did not diffract better than to 4 Å. Other experiments showed that good quality rod-like crystals could be grown from ammonium sulphate. Within one week crystals with dimensions of 0.5 mm $\times\,0.1\ mm\times0.1\ mm$ grew from 19 to 22% saturated ammonium sulphate at room temperature in a pH range from pH 7.6 to 8.0. These crystals belong to the orthorhombic space group $P2_12_12_1$.

Data collection and processing

A data set was recorded on an Enraf Nonius FAST area detector system and processed with MADNES (Messer-

schmidt & Pflugrath, 1987) with profile fitting of the intensities according to Kabsch (1988). The final data set contained 17,735 unique reflections with an $R_{\rm sym}$ of 19.8% and a completeness of 96.1% to 2.95 Å. This data set was used for molecular replacement, using the 2.0 Å resolution crystal structure of the *Bacillus circulans* strain 251 CGTase as the search model. After one month larger crystals had grown (0.5 mm × 0.2 mm × 0.2 mm). A new data set was recorded on a MacScience Dip2000K image plate system and processed with XDS (Kabsch, 1993). Cell dimensions and statistics for this data set are listed in Table 2. Reflections between 6 and 2.3 Å from this data set were used for refinement of the structure.

DNA sequencing

DNA sequencing was carried out by the dideoxy-chain termination method (Sanger *et al.*, 1977), using the fluoro-dATP⁺⁺ labelling mix in combination with the Autoread Sequencing kit and the Automated Laser Fluorescent (A.L.F.) DNA sequencer from Pharmacia. Plasmid pCT2 was used as template DNA (Bahl *et al.*, 1991). Primers were provided by Dr H. Bahl or synthesized by Pharmacia. Plasmid DNA for sequencing was isolated with the plasmid midi protocol of QIAGEN (Cathsworth, USA). The (corrected) *T. thermosulfurigenes* EM1 sequence has been deposited with the EMBL sequence database (accession number M57580).

Alignments of CGTase sequences were performed with the program DNASIS for Windows (Hitachi Software, San Bruno, USA). The same program was used for the calculation of similarity scores. Sequences were obtained from the EMBL sequence database. Accession numbers for CGTases: *T. thermosulfurigenes* EM1, M57580 (this

Table 2. Data collection statistics and quality of the final model

Cell parameters $(P2_12_12_1)$					
a (Å)	73.9				
b (Å)	97.4				
c (Å)	115.8				
Resolution range (Å)	34.3-2.3				
Total number of observations	119,291				
No. of discarded observations	5383				
No. of unique reflections	31,189				
R _{merge}	0.061				
Completeness of data (%)	82.4				
Completeness (%) of last	66.6				
resolution shell (Å)	(2.34 - 2.30)				
No. of protein atoms	5333				
No. of calcium atoms	2				
No. of solvent sites	343				
Average <i>B</i> -factor ($Å^2$)	19.24				
Final <i>R</i> -factor/ $R_{\rm free}$	17.9/24.9				
r.m.s. deviations from ideality for:					
Bond lengths (Å)	0.004				
Bond angles (deg)	0.9				
Torsion angles (deg)	18.6				
Trigonal planes (Å)	0.004				
Planar groups (Å)	0.009				
van der Waals contacts (Å)	0.019				
r.m.s. difference in <i>B</i> for					
neighbouring atoms (Å ²)	1.37				

 $R_{\rm merge}$ is defined as $R_{\rm merge} = (\Sigma\Sigma | I_j(hkl) - \langle I(hkl) \rangle |) / (\Sigma\Sigma \langle I(hkl) \rangle)$ and the crystallographic *R*-factor is defined as $R = \Sigma \| F_{\rm obs} | - |F_{\rm calc}\| / \Sigma | F_{\rm obs} |$. Free *R*-factors were calculated using 10% of the unique reflections. Observations were discarded when they deviated by more than 2.5 times the standard deviation from the average intensity.

study; Bahl et al., 1991); B. circulans strain 251, X78145 (Lawson et al., 1994); B. circulans strain 8, X68326 (Nitschke et al., 1990); Bacillus sp. 38-2, M19880 (Kaneko et al., 1988); B. ohbensis, D90243 (Sin et al., 1991); B. stearothermophilus no. 2, X59042 or S51853 (Fujiwara et al., 1992).

Molecular replacement

For molecular replacement a model was constructed based on the 2.0 Å structure of wild-type CGTase from Bacillus circulans strain 251. Based on the amino acid sequence derived from the original nucleotide sequence by Bahl et al. (1991), mutations to Ala or deletions were introduced with O (Jones et al., 1991) and all three maltose molecules and both calcium ions were removed from the starting model. Molecular replacement with the BIOMOL programs (Protein Crystallography Group, University of Groningen) FROTF0-3 (fast rotation function according to Crowther (1972)) and TRAFUN (translation function according to Crowther & Blow (1967)) using diffraction data in the 8.0 to 3.5 Å resolution range yielded a unique solution at 9σ that rotated the starting model over $\alpha = 7.5^{\circ}$, $\beta = 135^{\circ}$ and $\gamma = 265^{\circ}$ and translated it over -8.53 Å, -12.73 Å and -39.51 Å along the *a*, *b* and *c* axes, respectively.

Structure refinement and quality assessment

The resulting structure was submitted to rigid body refinement with TNT (Tronrud et al., 1987) followed by alternating cycles of manual rebuilding with O (Jones et al., 1991) and all-parameter refinement with TNT. Free R-factors (Brünger, 1993) were monitored using a randomly chosen subset of 10% of the reflections during all stages of the refinement. Strong restraints were applied to bond lengths and angles as well as planes and non-bonded contacts using the reference values reported by Engh & Huber (1991). No restraint was applied to dihedrals or hydrogen bonds. During the refinement it became clear that the published amino acid sequence did not fit with the observed electron density at several places. Some of these residues were highly conserved in other CGTases but according to the sequence should be absent in the CGTase from T. thermosulfurigenes EM1. To resolve these issues the protein was resequenced and a corrected sequence was obtained, which was used during further structure refinement.

Water molecules were placed at maxima larger than 3.5σ in σ_a -weighted $F_o - F_c$ electron density maps (Read, 1986) under the requirement that they were within 2.7 to 3.5 Å of hydrogen bond donating or accepting groups in the model. The final model was checked for stereochemical quality using the PROCHECK (Laskowski *et al.*, 1993) package. Total and cavity volumes of protein structures were calculated with VOIDOO (Kleywegt & Jones, 1994) using standard settings. Refinement statistics are listed in Table 2. The coordinates of the *T. thermosulfurigenes* EM1 CGTase (accession code 1CIU) have been deposited with the Protein Data Bank (Bernstein *et al.*, 1977).

Acknowledgements

R.D.W. acknowledges financial support by the European Community (AIR2-CT93-1023). We thank Dr H. Bahl (Department of Microbiology, Georg-August

References

- Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. & Tratschin, J. D. (1979). Thermal stability of protein structure. *Biochemistry*, 18, 5698–5703.
- Bahl, H., Burchhardt, G., Spreinat, A., Haeckel, K., Wienecke, A., Schmidt, B. & Antranikian, G. (1991). α -Amylase of *Clostridium thermosulfurigenes* EM1: nucleotide sequence of the gene, processing of the enzyme, and comparison to other α -amylases. *Appl. Environ. Microbiol.* **57(5)**, 1554–1559.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. 112, 535–542.
- Brünger, A. T. (1993). Assessment of the phase accuracy by cross validation: The free *R*-value. Methods and applications. *Acta Crystallog. sect. D*, **49**, 24–36.
- Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science*, **267**, 1463–1469.
- Crowther, R. A. (1972). Fast rotation function. In *The Molecular Replacement Method* (Rossmann, M. G., ed.), pp. 173–178, Gordon and Breach, New York.
- Crowther, R. A. & Blow, D. M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallog.* **23**, 544–548.
- Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J.-B., Zhou, Z. H., Adams, M. W. W. & Rees, D. C. (1992). X-ray crystal structures of the oxidized and reduced form of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Protein Sci.* 1, 1494–1507.
- Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.
- Fontana, A. (1991). How nature engineers protein (thermo) stability. In *Life Under Extreme Conditions* (di Prisco, G., ed.), pp. 89–113, Springer-Verlag, Berlin and Heidelberg.
- Fujiwara, S., Kakihara, H., Woo, K. B., Lejeune, A., Kanemoto, M., Sakaguchi, K. & Imanaka, T. (1992). Cyclization characteristics of cyclodextrin glucanotransferase are conferred by the NH2-terminal region of the enzyme. *Appl. Environ. Microbiol.* 58, 4016– 4025.
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.* **88**, 1895–1898.
- Jaenicke, R. (1991). Protein stability and molecular adaptation to extreme conditions. *Eur. J. Biochem.* 202, 715–728.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
- Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. J. Appl. Crystallog. 21, 916–924.
- Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown

symmetry and cell constants. J. Appl. Crystallog. 26, 795–800.

- Kaneko, T., Hamamoto, T. & Horikoshi, K. (1988). Molecular cloning and nucleotide sequence of the cyclomaltodextrin glucanotransferase from the alkalophilic *Bacillus* sp. strain no. 38-2. *J. Gen. Microbiol.* 134, 97–105.
- Klein, C. & Schulz, G. E. (1991). Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. J. Mol. Biol. 217, 737–750.
- Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallog. sect. D*, 50, 178–185.
- Knegtel, R. M. A., Strokopytov, B., Penninga, D., Faber, O. G., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. & Dijkstra, B. W. (1995). Crystallographic studies of the interaction of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 with natural substrates and products. *J. Biol. Chem.* 270, 29256– 29264.
- Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991). Molecular structure of *B. stearothermophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. *Denpun Kagaku*, **38**, 141–146.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallog. 26, 283–291.
- Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G., Penninga, D., Dijkhuizen, L. & Dijkstra, B. W. (1994). Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *J. Mol. Biol.* 236, 590–600.
- Lee, Y. E., Jain, M., Lee, C., Lowe, S. E. & Zeikus, J. G. (1993). Taxonomic distinction of saccharolytic thermophilic anaerobes. *Int. J. Syst. Bacteriol.* 43, 41–51.
- Matsumura, M., Signor, G. & Matthews, B. W. (1989). Substantial increase of protein stability by multiple disulphide bonds. *Nature*, **142**, 291–293.
- Messerschmidt, A. & Pflugrath, J. W. (1987). Crystal orientation and X-ray pattern prediction routines for area detector diffractometer systems in macromolecular crystallography. J. Appl. Crystallog. 20, 306–315.
- Mrabet, N. T., van den Broeck, A., van den Brande, I., Stanssens, P., Laroche, Y., Lambeir, A.-M., Matthijssens, G., Jenkins, J., Chiadmi, M., van Tilbeurgh, H., Rey, F., Janin, J., Quax, W. J., Lasters, I., De Maeyer, M. & Wodak, S. J. (1992). Arginine residues as stabilizing elements in proteins. *Biochemistry*, **31**, 2239–2253.
- Nitschke, L., Heeger, K., Bender, H. & Schulz, G. E. (1990). Molecular cloning, nucleotide sequence and

expression in *Escherichia coli* of the β -cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain no. 8. Appl. *Microbiol. Biotechnol.* **33**, 542–546.

- Penninga, D., Strokopytov, B., Rozeboom, H. J., Lawson, C. L., Dijkstra, B. W., Bergsma, J. & Dijkhuizen, L. (1995). Site directed mutations in Tyrosine 195 of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 affect activity and product specificity. *Biochemistry*, 34, 3368–3376.
- Read, R. J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. Acta Crystallog. sect. A, 42, 140–149.
- Rees, D. C. & Adams, M. W. W. (1995). Hyperthermophiles: taking the heat and loving it. *Structure*, **3**, 251–254.
- Russell, R. J. M., Hough, D. W., Danson, M. J. & Taylor, G. L. (1994). The crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma* acidophilum. Structure, 2, 1157–1167.
- Sanger, F., Nicklen, S. & Coulsen, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl Acad. Sci. USA, 74, 5463–5467.
- Schmid, G. (1989). Cyclodextrin glycosyltransferase production: yield enhancement by overexpression of cloned genes. *Trends Biotechnol.* 7, 224–248.
- Sin, K.-A., Nakamura, A., Kobayashi, K., Masaki, H. & Uozumi, T. (1991). Cloning and sequencing of a cyclodextrin glucanotransferase gene from *Bacillus* ohbensis and its expression in *Escherichia coli*. Appl. Microbiol. Biotechnol. 35, 600–605.
- Strokopytov, B., Penninga, D., Rozeboom, H. J., Kalk, K. H., Dijkhuizen, L. & Dijkstra, B. W. (1995). X-ray structure of cyclodextrin glycosyl transferase complexed with acarbose. Implications for the catalytic mechanism of glycosidases. *Biochemistry*, **34**, 2234– 2240.
- Szejtli, J. (1982). Cyclodextrins in food, cosmetics and toiletries. *Die Stärke*, **34**, 379–385.
- Tronrud, D. E., Ten Eyck, L. & Matthews, B. W. (1987). An efficient general-purpose least-squares refinement program for macromolecular structures. Acta Crystallog. sect. A, 43, 489–501.
- van den Burg, B., Dijkstra, B. W., van der Vinne, B., Stulp,
 B. K., Eijsink, V. G. H. & Venema, G. (1993). Introduction of disulfide bonds into *Bacillus subtilis* neutral protease. *Protein Eng.* 6, 521–527.
- Vihinen, M. (1987). Relationship of protein flexibility to thermostability. *Protein Eng.* **1**, 477–480.
- Wind, R. D., Liebľ, W., Buiteľaar, R. M., Penninga, D., Spreinat, A., Dijkhuizen, L. & Bahl, H. (1995). Cyclodextrin formation by the thermostable α -amylase of *Thermoanaerobacterium thermosulfurigenes* EM1 and reclassification of the enzyme as a cyclodextrin glycosyltransferase. *Appl. Environ. Microbiol.* **61**, 1257–1265.

Edited by I. A. Wilson

(Received 20 October 1995; accepted in revised form 5 December 1995)