doi:10.1016/j.jmb.2003.10.042 J. Mol. Biol. (2004) 335, 261–274

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

X-ray Structures of the Maltose–Maltodextrin-binding Protein of the Thermoacidophilic Bacterium *Alicyclobacillus acidocaldarius* Provide Insight into Acid Stability of Proteins

Karsten Schäfer[1], Ulrika Magnusson[2], Frank Scheffel[3] André Schiefner[1], Mats O. J. Sandgren[2], Kay Diederichs[1] Wolfram Welte[1], Anja Hülsmann[3], Erwin Schneider[3] and Sherry L. Mowbray[4]*

[1]*Fachbereich Biologie Universität Konstanz Universitätsstraße 10, 78457 Konstanz, Germany*

[2]*Department of Cell and Molecular Biology, Uppsala University, Biomedical Center Box 596, S-751 24 Uppsala Sweden*

[3]*Humboldt Universität zu Berlin, Institut für Biologie/Bakterienphysiologie Chausseestraße 117, D-10115 Berlin, Germany*

[4]*Department of Molecular Biosciences, Division of Structural Biology, Swedish University of Agricultural Sciences, Biomedical Center Box 590, S-751 24 Uppsala Sweden*

**Corresponding author*

Maltose-binding proteins act as primary receptors in bacterial transport and chemotaxis systems. We report here crystal structures of the thermoacidostable maltose-binding protein from *Alicyclobacillus acidocaldarius*, and explore its modes of binding to maltose and maltotriose. Further, comparison with the structures of related proteins from *Escherichia coli* (a mesophile), and two hyperthermophiles (*Pyrococcus furiosus* and *Thermococcus litoralis*) allows an investigation of the basis of thermo- and acidostability in this family of proteins.

The thermoacidophilic protein has fewer charged residues than the other three structures, which is compensated by an increase in the number of polar residues. Although the content of acidic and basic residues is approximately equal, more basic residues are exposed on its surface whereas most acidic residues are buried in the interior. As a consequence, this protein has a highly positive surface charge. Fewer salt bridges are buried than in the other MBP structures, but the number exposed on its surface does not appear to be unusual. These features appear to be correlated with the acidostability of the *A. acidocaldarius* protein rather than its thermostability.

An analysis of cavities within the proteins shows that the extremophile proteins are more closely packed than the mesophilic one. Proline content is slightly higher in the hyperthermophiles and thermoacidophiles than in mesophiles, and this amino acid is more common at the second position of β-turns, properties that are also probably related to thermostability. Secondary structural content does not vary greatly in the different structures, and so is not a contributing factor.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: maltose-binding protein; maltodextrin-binding protein; acidophile; thermoacidophile; hyperthermophile

Introduction

Starch is one of the major sources of carbon and energy available to heterotrophic bacteria and archaea. Since polysaccharides cannot penetrate the cell membrane, many microorganisms secrete amylases that produce maltose and maltodextrins (oligosaccharides with up to seven α-1,4 linked glucose units) as major degradation products of starch. The subsequent uptake of maltodextrins is

Abbreviations used: MBP, maltose–maltodextrin-binding protein; *Acy*MBP, *Alicyclobacillus acidocaldarius* MBP; *Eco*MBP, *Escherichia coli* MBP; *Pfu*MBP, *Pyrococcus furiosus* MBP; *Tli*MBP, the trehalose–maltose-binding protein of *Thermococcus litoralis*; r.m.s, root-mean-square.

E-mail address of the corresponding author: mowbray@xray.bmc.uu.se

0022-2836/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.

usually mediated by a member of the superfamily of ATP-binding-cassette (ABC) transport systems. Maltose ABC transporters are composed of a membrane-bound complex comprising the two hydrophobic permease subunits (MalF and MalG) and two copies of the ATPase subunit (MalK). In addition, a cognate receptor, the maltose-binding protein (MBP) is essential for this activity.[1] The binding proteins interact with the membrane permeases that actually transport the solute into the cell. The ATPase component provides the energy both for transport itself, and for overcoming the binding protein's tight interaction with the ligand. In Gram-negative bacteria, the binding proteins are situated in the periplasm between the outer and inner membrane, where they bind their ligands with high affinity. Gram-positive bacteria and archaea, which do not have a periplasm, instead anchor the binding proteins to the outer surface of the cell membrane *via* an N-terminal lipid moiety,[2] or alternatively, as observed for some archaea, an N-terminal hydrophobic helix.[3]

Like other essential systems, binding-protein dependent ABC transporters are widespread in organisms that thrive in extreme environmental niches. The principles that allow some proteins to remain stable and active under harsh conditions, such as high temperature or extreme pH, are of great interest for academic reasons as well as for potential industrial applications. For thermophilic organisms, there is no escape from their hostile environment, and all proteins in the cell must be able to function at high temperature. Therefore, many structures of thermostable proteins have been solved, and a number of features have been identified that are correlated with this property. In contrast, acidophilic microorganisms maintain their cytoplasmic pH close to neutrality by the means of respiratory chain activity or ATP-driven proton pumps.[4] Only proteins that are secreted or otherwise exposed to an acidic environment actually need to be stable under these conditions, and so there are currently very few subjects suitable for studies of the molecular basis of acidostability. Slightly more is known about proteins from thermoacidophilic organisms such as *Alicyclobacillus acidocaldarius*. This Gram-positive bacterium has a pH optimum of 3.6 and a temperature optimum of 57 °C.[5,6] Comparison of the core $(\beta\alpha)_8$-barrel domain of the secreted α-amylase (amylopullulanase)[7] from *A. acidocaldarius* with the available structures of homologous proteins from mesophilic organisms indicated that charged amino acid residues were often replaced with polar but uncharged ones in the thermoacidophile's protein, especially on the molecular surface.[8] This trend had not been identified in thermostable proteins, and so was believed to reflect the enzymes's acidostability. At the time, it was not clear whether this reflected a more general phenomenon.

We report here the ligand-bound structure of MBP of *A. acidocaldarius* (*Acy*MBP) at high resolution, and in multiple crystal packing environments. This protein binds maltose with high affinity (K_d 1.5 μM) over a wide pH range (2.5–7) and at temperatures up to 80 °C, making it an excellent candidate for studies of thermoacidostability.[9] Furthermore, the availability of the crystal structures for *Escherichia coli* MBP (*Eco*MBP)[10] as well as two hyperthermostable MBPs (from *Pyrococcus furiosus*, *Pfu*MBP[11] and *Thermococcus litoralis*, *Tli*MBP)[12] allow an excellent opportunity to investigate the molecular adaptations related to acidostability.

Results

Protein expression and purification

A truncated variant of *Acy*MBP was isolated from the culture supernatant of maltose-grown cells of *A. acidocaldarius* and purified by affinity chromatography on Sepharose-linked amylose.[9,13] N-terminal sequencing (SALPKGQTIT) and mass spectrometry (M_r 41,250 Da) confirmed earlier conclusions[13] that the truncated protein begins at residue 24 of the mature sequence. Cleavage of the protein during the preparation is believed to be due to the action of an extracellular protease.[8] Dynamic light scattering experiments showed this sample represented a monomer of molecular weight 42 kDa in the temperature range +5 °C to +50 °C.

A full-length version of the protein was purified from *E. coli* strain JM109 harboring plasmid pRF1 as an N-terminal fusion to six consecutive histidine residues by Ni-NTA chromatography (see Materials and Methods for details).

Crystallization, data collection, structure solution and refinement

The structure of one molecule of closed, ligand-bound *Acy*MBP (full-length protein, referred to as *Acy*MBP-1 in further discussion) was obtained by molecular replacement using the structure of ligand-bound *Eco*MBP,[14] and refined to 1.80 Å resolution. The first 28 amino acid residues of the mature protein (out of a total of 402) were not visible in the electron density and so were not included in the model. Molecular replacement with this structure located two molecules of the full-length protein in a different unit cell; these structures (*Acy*MBP-2a and 2b; residues 30–395 visible in both molecules) were refined to 1.45 Å resolution. An additional data set represented two copies of the truncated protein in a very similar crystal form. These were refined to 1.53 Å resolution (residues 30–399, and 31–399, respectively), and will be referred to as *Acy*MBP-3a and 3b. Statistics for data collection and final refined structures are presented in Table 1.

Table 1. Data collection and refinement statistics

	*Acy*MBP-1	*Acy*MBP-2	*Acy*MBP-3
Data collection			
Environment	Rotating anode, MAR 345 detector (T = 291 K)	Synchrotron, MAR CCD (T = 100 K)	Synchrotron, MAR 345 detector (T = 100 K)
Wavelength (Å)	1.54179	0.91964	1.0201
Space group	$P2_1$	$P2_1$	$P2_1$
Cell (Å and °)	50.11, 72.16, 57.78, β = 109.63	49.18, 70.53, 104.06 β = 96.98	49.23, 70.78, 104.67 β = 96.58
Resolution range (Å, highest shell)	50.0–1.80 (1.90–1.80)	40.14–1.45 (1.50–1.45)	40.0–1.53 (1.56–1.53)
No. unique reflections	35,989	137,804	99,996
Average multiplicity	3.9 (3.3)	6.2 (3.7)	5.2 (2.8)
R_{merge} (%)	15.5 (52.4)	8.4 (50.9)	8.3 (18.7)
Completeness	99.5 (91.4)	99.6 (98.7)	98.8 (88.3)
$I/\sigma I$	7.5 (2.1)	12.7 (2.4)	11.9 (5.8)
Refinement			
Resolution range (Å)	50–1.8	40.0–1.45	40.0–1.53
No. reflections total	34,421	137,798	99,996
R, R_{free} values (%)	19.4, 21.7	21.0, 23.2	20.6, 23.1
No. reflections for R_{free} calculation	1722	3450	3965
No. non-hydrogen atoms	3065	5705	5755
No. solvent water molecules	171	501	549
Mean B-factor, protein atoms (Å^2)	23.3	17.3, 19.0	11.6, 12.9
Mean B-factor, ligand atoms (Å^2)	12.8	15.6	10.1
Mean B-factor, solvent atoms (Å^2)	32.0	27.3	21.2
Ramachandran plot outliers (%)[a]	0.6	1.2	1.5
r.m.s. deviation bond length (Å)	0.005	0.005	0.005
r.m.s. deviation bond angle (°)	1.23	1.21	1.19

Numbers in parenthesis refer to data in the highest resolution shell.
[a] Using a stringent boundary Ramachandran plot.[54]

Overall structures

As typically found in periplasmic-binding proteins, the structure of *Acy*MBP is comprised of two domains, each containing a central β-sheet flanked by α-helices (Figure 1). The two domains are linked by three hinge segments; rotations of bonds in the hinge region allow the opening and closing of the protein required for entry and exit of ligand during its acquisition and transport.[10,15] The five *Acy*MBP structures are very similar, with small differences in the relationship between the two domains (~2°, calculated in FIT†) accounting for much of the observed r.m.s. differences (0.21–0.43 Å when comparing all C^{α} atoms). The different conformations of the closed forms are probably the result of crystal packing effects, as they are correlated with differences in the environment of the various molecules. Changes in side-chain conformations are restricted to residues exposed on the surface; those inside the protein and close to the

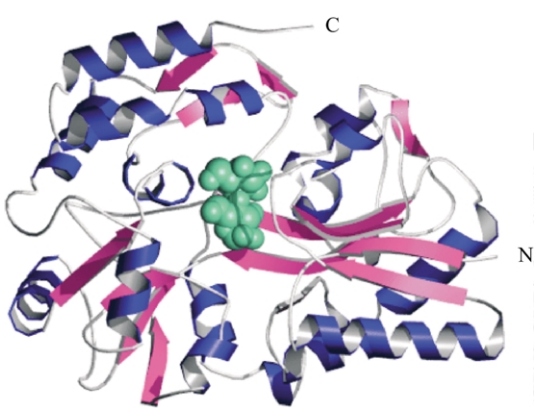

Figure 1. Overall structure. The structure of *Acy*MBP is shown as a ribbon cartoon, with β-strands colored magenta, and α-helices in blue. The N-terminal and C-terminal ends are marked. The N-terminal domain consists of residues 30–140 and 287–340, while the C-terminal domain consists of residues 141–286 and 341–399. Bound maltose is shown as a space-filling model (green).

† Lu, G. (1998). FIT. http://bioinfo1.mbfys.lu.se/~guoguang/fit.html

Table 2. Structural comparisons

Model	*Acy*MBP	*Eco*MBP	*Pfu*MBP	*Tli*MBP
*Acy*MBP	**374**	1.60	1.74	1.81
*Eco*MBP	321 (31.2)	**370**	1.65	1.83
*Pfu*MBP	299 (32.3)	309 (29.8)	**381**	1.91
*Tli*MBP	288 (25.0)	311 (23.8)	310 (26.6)	**407**

Numbers at top right (blue) are r.m.s. differences obtained using a cut-off of 3.5 Å in the superposition in LSQMAN (fast-force algorithm). Numbers at bottom left (green) are the number of residues matching (with % amino acid sequence identity in parentheses), and those on the diagonal (boldface) are the number of residues in the structures being compared.

ligand-binding site are very similar in all structures. Although three of the models (*Acy*MBP-1, 2a and 2b) represent a construct that originally included the N-terminal extension found *in vivo*,[13] no electron density is visible for this segment in the structures. Western blots using antibody directed to the His-tag confirmed that some or all of this region had been lost either during storage or crystallization. The construct used to obtain the *Acy*MBP-3 structures similarly showed no ordered electron density prior to residue 30 of the mature sequence.

Structures are currently available for MBPs from *E. coli*, *P. furiosus* and *T. litoralis*. These structures and *Acy*MBP are compared in Table 2. The most similar to *Acy*MBP is *Eco*MBP, in which 86% of the residues match with an r.m.s. difference of 1.6 Å using a 3.5 Å cut-off (referring to the *Acy*MBP-1 structure). Among the residues in this core structure, 31% of the amino acid residues are identical. A comparable result was obtained with *Pfu*MBP, with 80% of its residues matching to equivalents in *Acy*MBP. Each of these three proteins is more distantly related to the trehalose/maltose receptor *TliMBP*(which is also present as a second transport receptor in *P. furiosus*, presumably as the result of horizontal gene transfer).[16]

Ligand binding

Electron density for maltose (α-D-glucosyl-1,4-α-D-glucose) was visible in the cleft between the two domains, with full occupancy in all structures (Figure 2A). In addition, both $F_o - F_c$ and $2F_o - F_c$ maps showed a partial occupancy of an additional glucose ring at the reducing end of the sugar in *Acy*MBP-2 and 3. The electron density for this third ring was stronger in both *Acy*MBP-2 structures than in those of *Acy*MBP-3; it was completely absent in *Acy*MBP-1. This third ring presumably represents a maltotriose impurity that varies in quantity among the different batches of maltose used for purification and co-crystallization. It is unlikely that the electron density for the third ring results from the binding of maltose in two different modes, as the density for the first two rings is in all cases as strong as that of the surrounding protein atoms. To investigate the ratio of maltose:maltotriose, refinement was carried out with different occupancies for the third ring. The best fit to both $2F_o - F_c$ and $F_o - F_c$ maps was used to determine the closest match. The fractional occupancy of maltotriose in *Acy*MBP-2 was in this way estimated to be 0.65 in both molecules, and in *Acy*MBP-3a and 3b, the corresponding values are 0.30 and 0.45, respectively.

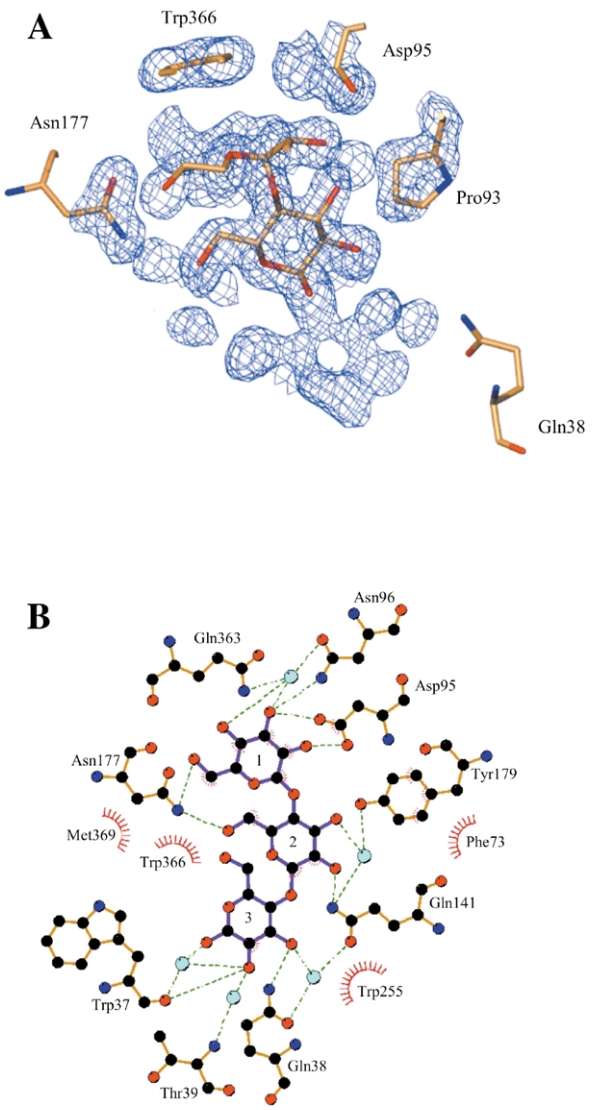

Figure 2. Ligand binding in *Acy*MBP. A, Maltose (and maltotriose) density in the last electron density map prior to addition of ligand in the *Acy*MBP-2a model. The structure of maltose from the final refined structure is illustrated with a ball-and-stick model. B, LIGPLOT cartoon illustrating maltotriose binding in *Acy*MBP-1.

Maltose makes direct hydrogen bonds with residues Asp95, Asn96, Gln141 and Asn177 (Figure 2B). Several indirect hydrogen bonds are mediated by water molecules, and link the disaccharide to residues Tyr179 and Gln363. As is typical for sugar-binding proteins,[17] non-polar interactions largely involve contacts with aromatic residues, in this case Tyr179, Trp255, Met356, Trp366 and Met369 of the C-terminal domain, and

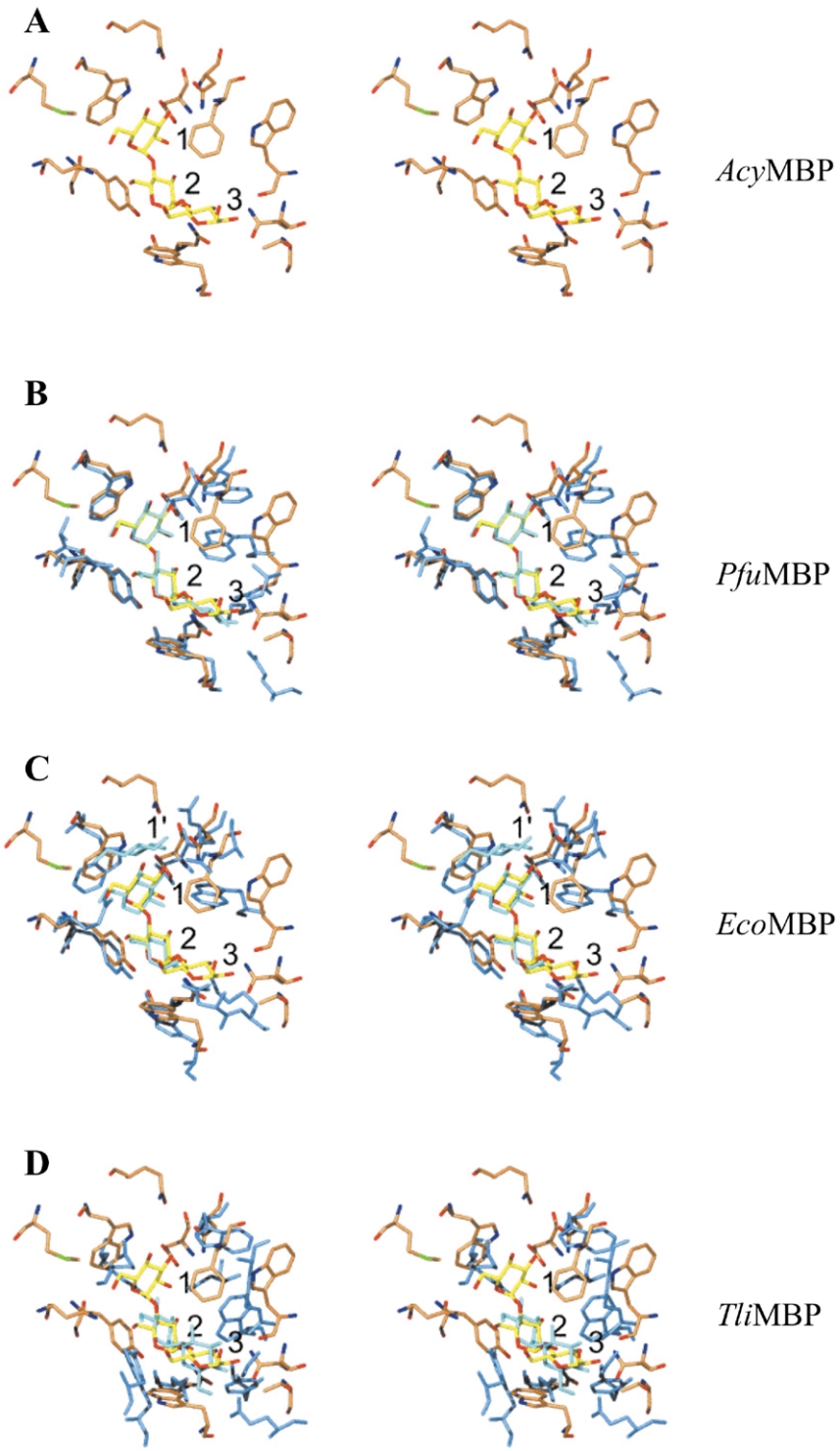

Figure 3. Comparison of ligand binding in available MBP structures. In each case, *Acy*MBP is shown as a stick model with normal atomic coloring, and the protein to be compared is shown with cyan (darker for protein atoms). The view is similar to that in Figure 1. A, *Acy*MBP. The aromatic residues interacting with the ligand and the residues involved in hydrogen bonding are shown, and the three sugar sites are labeled, beginning from the non-reducing end. B, Equivalent view of *Pfu*MBP, showing maltotriose bound in the same three subsites observed for *Acy*MBP. Although the residues mediating hydrogen bonds are sometimes different, the same total number of interactions is formed. Most interacting aromatic residues are also conserved. C, Comparison with *Eco*MBP. Two of the subsites used by *Acy*MBP and *Pfu*MBP are used in binding the same sugar in *Eco*MBP. The third sugar unit is, however, placed differently in *Eco*MBP (labeled as 1′); the loop at residues 12–17 has moved and therefore it is not possible to accommodate a glucose ring at site 3. D, Comparison with *Tli*MBP. In this case, sugar is positioned in sites 2 and 3, as defined above, but with different ring orientations, and no conserved interactions.

to a lesser extent Phe73 of the N-terminal domain. Binding of the β-anomer of the disaccharide would be precluded by steric clashes with Trp255. The third α-D-glycosyl ring of maltotriose forms additional hydrogen bonds to Trp37 (main-chain oxygen), Gln38, and Gln141, and through water molecules to Trp37 (main-chain oxygen), Thr39 (main-chain nitrogen) and Gln141. There are also further stacking interactions with Trp255. Assuming the same sugar conformation, extension of the trisaccharide at the reducing end would lead to clashes near residue 67 in the closed form. Proceeding from the non-reducing end, intra-molecular hydrogen bonds link O2 of the first sugar unit to O3 of the second, and O2 of the second sugar to O3 of the third; the result is a curved trisaccharide (Figure 3A), consistent with the form expected for the free sugar in solution.

The two fully occupied sites in the *Acy*MBP structures also have full occupancy in the available *Pfu*MBP structure[11] (Figure 3B). When maltose is bound to *Eco*MBP, the same two sites are preferred[18] (Figure 3C). Many of the interactions in these two sites are conserved in the three proteins. Counterparts to *Acy*MBP's Asp95, Tyr179, Phe180, Trp255 and Trp366 are present in both *Pfu*MBP and *Eco*MBP. A third glycosyl ring was also observed with partial occupancy in *Pfu*MBP at a position very similar to that seen in *Acy*MBP (site 3 in Figure 3B); this was likewise believed to represent maltotriose, a common contaminant of maltose. When maltotriose is bound to *Eco*MBP, however, the third ring is placed at the other end of the binding site, i.e. at the non-reducing end of the maltose (site 1′ in Figure 3C). In *Acy*MBP, residues 37–41 form hydrogen bonds to the glycosyl unit in site 3, and corresponding interactions are found in *Pfu*MBP. This site is blocked in the *E. coli* protein; the loop at residues 12–17 is different and therefore it is not possible to accommodate a glycosyl ring at this position. At the same time, local changes in the 1′-site of the *Acy*MBP and *Pfu*MBP structures prevent them from binding a third glycosyl unit in the same way as *Eco*MBP. Binding of trehalose (α-D-glucosyl-1,1-α-D-glucose) to *Tli*MBP positions the two glycosyl units in sites 2 and 3 of *Acy*MBP and *Pfu*MBP (Figure 3D). The plane of the sugar ring in site 2 has a very similar orientation in *Tli*MBP, although no interactions are conserved. The plane of the sugar ring in site 3 is almost anti-parallel with that in *Acy*MBP and *Pfu*MBP, and so it is not surprising that no interactions are conserved.

Electrostatic properties

It had been noted during the biochemical analysis of the truncated form of *Acy*MBP that the experimentally measured isoelectric point (~10) was very different from that calculated based on its amino acid sequence (6.5). The *Acy*MBP structures provided a clear explanation for this unusually large discrepancy: the protein has a striking excess of basic residues on its surface. This observation prompted us to undertake a closer investigation of charged residues in all of the available MBP structures. The comparisons showed that *Acy*MBP, *Eco*MBP and *Tli*MBP have a similar number of basic residues on their surfaces, while *Pfu*MBP has slightly fewer (Figure 4A). Since our main interest lies in understanding the adaptation of proteins to environmental conditions, we have been consistent in considering histidine with the positively charged residues, although some of these will not be charged in the normal surroundings of the mesophile. Fewer acidic residues are exposed in *Acy*MBP than in any of the other MBPs, leading to an overall positive charge. Furthermore, the total number of both positively and negatively charged residues was seen to be much smaller for *Acy*MBP than for the other proteins (Figure 4B). While the number of acidic and basic residues in the thermoacidophile's protein is approximately equal, most of its basic residues are exposed on the surface whereas most acidic residues are buried in the structure. The differences are particularly striking on the face of the protein near the carbohydrate-binding site, where acidic residues are totally absent in *Acy*MBP.

As an increase in the number of salt bridges has been correlated with thermostability of proteins, we manually searched for ion pairs in the various proteins at the graphics-terminal, using a cut-off distance of 3.3 Å. This inspection identified eight ion pairs in *Eco*MBP (six buried), 13 in *Pfu*MBP (nine buried) and nine in *Tli*MBP (seven buried). In each of the *Acy*MBP structures, only four salt bridges were observed, two of which were buried. Thus, fewer salt bridges are buried in *Acy*MBP than in the other MBP structures, but the number exposed on its surface does not appear to be outside the normal range.

Charged residues in the other MBPs are frequently substituted for uncharged or non-polar residues in *Acy*MBP, both on the surface and in the interior of the protein (Figure 4C and D). An increased number of polar residues in *Acy*MBP is particularly evident on the side facing away from the binding cleft (Figure 4C). Since the surface near the binding cleft has fewer charged residues as well, it will have a more non-polar character than the equivalent face in the other MBPs. The same pattern can be seen in more detail using structure-based sequence alignments (Figure 5). To further investigate these patterns, we looked at the sequences of a number of other MBPs that were identified by a BLAST[19] search as being relatives of *Acy*MBP. Only a limited number of hyperthermoacidophilic MBPs appeared in this list, although large numbers of thermophilic and mesophilic sequences were located, as well as ones from the mildly acid-tolerant lactobacilli. The most similar of these sequences were selected for further analysis, since the sequence/structural relationships were expected to be easiest to interpret in this set; some more distant relatives were also included, to

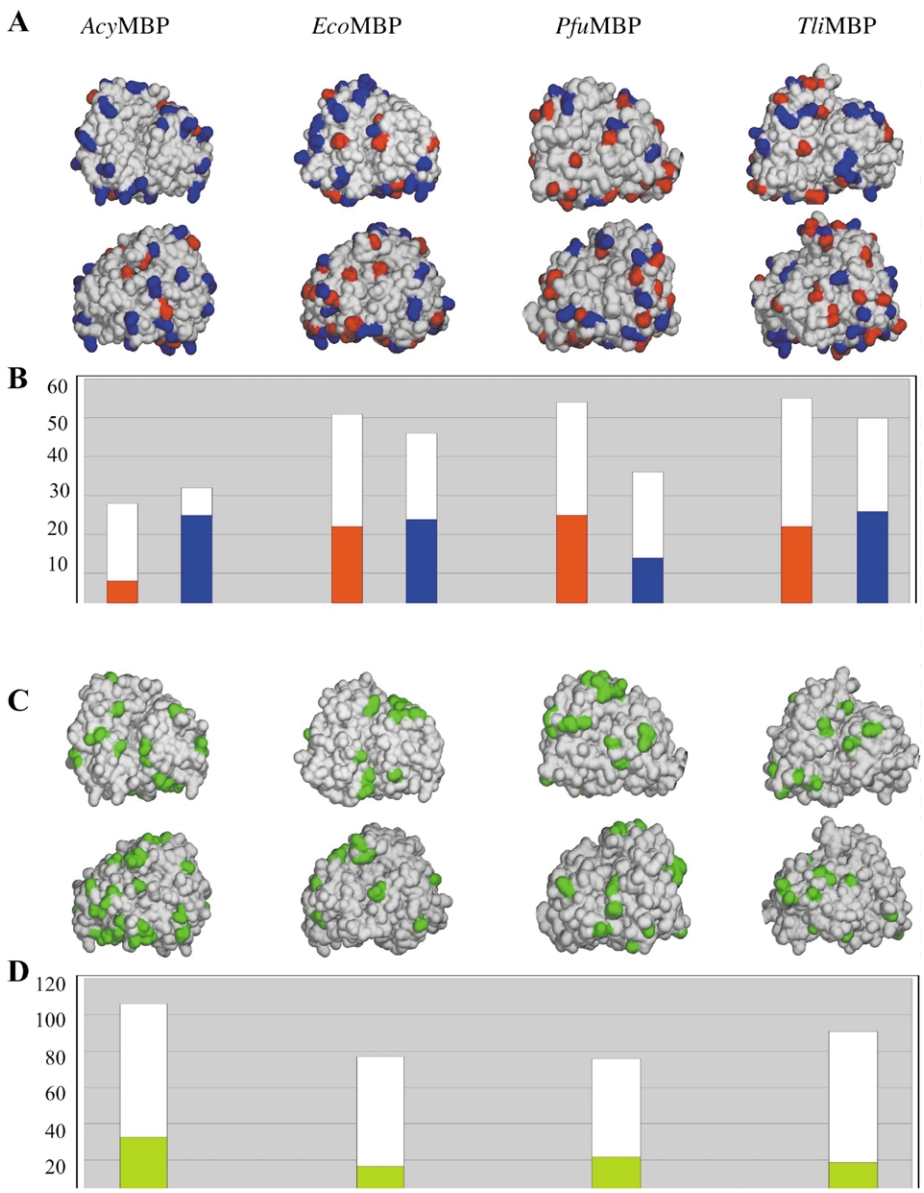

Figure 4. Distribution of charged and polar residues. A, Acidic (red) and basic (blue) residues on the surfaces of the various proteins are shown. In the top views, the orientation is the same as that in Figure 1, while those at the bottom are rotated 180° around a vertical axis. Surfaces were prepared in the program O. B, The absolute numbers of charged residues for each case in A are presented in a histogram, with exposed residues shown with the corresponding color, and buried residues in white. C, Polar residues (in this case, asparagine, glutamine, serine and threonine) on the surfaces are colored green in the same proteins. D, The absolute numbers of polar residues on the surface and buried in each instance in C are presented in a histogram, with exposed residues shown in the corresponding color.

determine whether any trends were general within the structural family. Results are summarized in Table 3. It immediately becomes apparent that, like *Acy*MBP, the hyperthermoacidophilic MBPs have fewer charged residues than average. Hyperthermophiles, thermophiles and mesophiles all have higher contents of both acidic and basic residues than the (hyper)thermoacidophiles. Among the acidic residues, the content of both aspartate and glutamate is reduced; a similar trend is apparent for the mildly acid-tolerant lactic acid bacteria as well. For basic residues, lysine accounts for most of the differences, primarily because it is much more common than either arginine or histidine. In contrast, the contents of polar but uncharged residues (most consistently asparagine and glutamine, the structural homologues of aspartate and glutamate) are generally increased. Sequence alignments (not shown) confirmed that the patterns observed for the hyperthermoacidophilic proteins were the same as those for the thermoacidophilic *Acy*MBP.

Other features

The various MBP structures were assessed using

```
AcyMBP    --GQTITVWSWQT--GPELQDVKQIAAQWAKAHG-DKVIVVDQSSNPKGFQFYA-TAARTG
EcoMBP    -KIEEGKLVIWIN-GDKGYNGLAEVGKKFEKDTG-IKVTVEHPD---KLEEKFP-QVAATG
PfuMBP    MKIEEGKVVIWHAMQPNELEVFQSLAEEYMALCPE--VEIVFEQKP-NLEDALK-AAIPTG
TliMBP    IEEGKIVFAVGGA--PNEIEYWKGVIAEFEKKYPGVTVELKRQATDTEQRRLDLVNALRGK

AcyMBP    KGPDVVFGMPHDNNGVFAEEGLMAPVPSGVLNT-----GLYAPNTIDAIKVNGTMYSVPVS
EcoMBP    DGPDIIFWA-HDRFGGYAQSGLLAEITPDKAFQ-----DKLYPFTWDAVRYNGKLIAYPIA
PfuMBP    QGPDLFIWA-HDWIGKFAEAGLLEPIDEYVTEDLL---NEFAPMAQDAMQYKGHYYALPFA
TliMBP    SSDPDVFLMDVAWLGQFIASGWLEPLDDYVQKDNYDLSVFFQSVINLADKQGGKLYALPVY

AcyMBP    VQVAAIYYNKKLV-----PQPPQTWAEFVKDA------------NAHGFM----YDQANLY
EcoMBP    VEALSLIYNKDLL-----PNPPKTWEEIPALDKEL---------KAKGKS--ALMFNLEPY
PfuMBP    AETVAIIYNKEMV-----SEPPKTFDEMKAIMEKYYD-------PANEKY--GIAWPINAY
TliMBP    IDAGLLYYRKDLLEKYGYSKPPETWQELVEMAQKIQSGERETNPNFWGFVWQGKQYEGLVC

AcyMBP    FDYAIIGGYGGYVFKDNNGTLDPNNIGLDTPGAVQAYTLMRDMVSKYHWMTPS----TNGS
EcoMBP    FTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDL-IKNKHMNAD----TDYS
PfuMBP    FISAIAQAFGGYYFDDK-----TEQPGLDKPETIEGFKFFFTE-IWPYMAP-----TGDYN
TliMBP    DFVEYVYSNGGSLGEFKDGKWV---PTLNKPENVEALQFMVDLIHKYKISPPNTYTEMTEE

AcyMBP    IAKAEFLAGKIGMYVSGPWDTADIEKAKID----FGVTPWPTL---PNGKHATPFLGVITA
EcoMBP    IAEAAFNKGETAMTINGPWAWSNIDTSKVN----YGVTVLPTF----KGQPSKPFVGVLSA
PfuMBP    TQQSIFLEGRAPMMVNGPWSINDVKKAGIN----FGVVPLPPIIKDGKEYWPRPYGGVKLI
TliMBP    PVRLMFQQGNAAFERNWPYAWGLHNADDSPVKGKVGVAPLPHF----PGHKSAATLGGWHI

AcyMBP    FVNKESKTQAADWSLVQA-LTSAQAQQMYFRD-SQQIPALLSVQR-SSAVQSSPTFKAFVE
EcoMBP    GINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEE---ELAKDPRIAATME
PfuMBP    YFAAGIKNKDAAWKFAKWLTTSEESIKTLALE-LGYIPVLTKVLDDP-EIKNDPVIYGFGQ
TliMBP    GISKYSDNKALAWEFVKF-VESYSVQKGFAMN-LGWNPGRVDVYDDPAVVSKSPHLKELRA

AcyMBP    QLRYAVPMPNIPQMQAVWQAMSI-LQ-NIIAGKVSPEQGAKDFVQNIQKGIMAQGS
EcoMBP    NAQKGEIMPNIPQMSAFWYAVRT-AVINAASGRQTVDEALKDAQTRITK-------
PfuMBP    AVQHAYLMPKSPKMSAVWGGVDG-AINEILQDPQNAD--IEGILKKYQQEILNNMQ
TliMBP    VFENAVPRPIVPYYPQLSEIIQKYVN-SALAGKISPQEALDKAQKEAEELVKQYS-
```

Figure 5. Structure-based sequence alignments. Alignment of the structures of *Acy*MBP, *Eco*MBP, *Pfu*MBP and *Tli*MBP was carried out using the default parameters in LSQMAN. Residues were then colored according to whether they are acidic (Asp, Glu; red), basic (Arg, Lys, His; blue) and polar (in this case, Asn, Gln, Ser, Thr; green).

a number of criteria that have been associated with thermostability, i.e. higher secondary structural content, fewer/smaller internal cavities, clusters of hydrophobic residues, proline disposition, etc.

Secondary structural content was evaluated using DSSP;[20] the results are summarized in Table 4. The content of helix and sheet does not vary substantially among the various proteins, remaining relatively constant in terms of the fraction of total structure.

Using the program VOIDOO,[21] we searched for internal cavities in the various proteins; the results are summarized in Table 5. Both the number of cavities, and their total volume, is greater in *Eco*MBP than in any of the other proteins. Thus all of the extremophile MBPs are more closely packed than that of the mesophile.

It has been reported that proper placement of proline residues with respect to secondary structural elements can enhance protein stability.[22–24] The total proline content is in general slightly higher in the (hyper)thermophilic MBPs than in the mesophiles (Table 3). Proline residues are more frequently found at the second position of β-turns in the (hyper)thermophile structures (*Acy*MBP, four cases; *Pfu*MBP, five cases; *Tli*MBP, seven cases) than in the mesophilic one (*Eco*MBP, one case). The distribution of proline residues at the N and C-terminal ends of helices does not appear to differ significantly, while only *Pfu*MBP shows an appreciable decrease in the frequency of proline residues within elements of secondary structure (Table 4).

It was noted earlier that *Pfu*MBP has several dramatic clusters of isoleucine and aromatic residues, features that were suggested to increase its thermal stability relative to *Eco*MBP.[11] However, this phenomenon is not apparent in either *Acy*MBP or *Tli*MBP.

Discussion

In vivo, the mature form of *Acy*MBP includes an ~30-residue N-terminal sequence that is believed to act as a linker between the binding domains and the lipid anchor (a palmitate covalently attached to Cys1).[13] In the present study, two of the crystals were prepared using full-length *Acy*MBP. Another crystal represented a truncated version in which much of the N-terminal sequence was removed by the action of a protease during protein preparation. No electron density corresponding to the N-terminal segment was visible in any of the maps. Tests of the supposedly full-length protein indicated that it, too, had lost the

Table 3. Amino acid composition of MBPs

Organism	Optimum t of growth (deg.C)	No. Res.	BLAST E value	BLAST id (overlap)	Asp	Asn	Glu	Gln	Arg	Lys	His	Total Asn + Gln	Total Asp + Glu	Total Arg + Lys + His	Pro
(*Hyper)thermoacidophiles															
*Acy*MBP	57	374	0.0	–	5.1	5.1	2.4	8.0	1.3	5.9	1.3	49	28	32	6.7
Thermoplasma acidophilum	55–60	437	1×10^{-9}	21% (363)	3.0	7.8	4.1	4.8	0.2	3.9	1.1	55	31	23	5.0
*Sulfolobus solfataricus**	80	450	5×10^{-23}	27% (407)	2.0	6.7	3.6	7.8	1.3	2.4	0.9	65	25	21	6.4
Acid-tolerant mesophiles															
Lactobacillus plantarum	37	370	2×10^{-32}	29% (334)	5.4	7.6	2.7	4.1	0.8	13.5	0.3	43	30	54	3.9
Lactococcus lactis	25–35	369	7×10^{-21}	27% (380)	8.1	6.5	2.4	5.7	0.0	9.8	0.0	45	39	36	5.1
Lactobacillus gasseri	37	381	9×10^{-16}	25% (377)	4.7	7.6	4.7	5.2	0.3	12.9	1.0	49	36	54	5.2
(*Hyper)thermophiles															
Thermoactinomyces vulgaris	40–50	397	3×10^{-39}	29% (381)	5.0	5.8	6.0	6.0	1.0	11.3	0.5	47	44	51	6.5
Thermoanaerobacter tengcongensis	75	372	1×10^{-79}	43% (369)	7.0	5.4	4.6	5.1	0.3	9.7	1.1	39	43	41	7.0
Thermotoga maritime	70–80	373	4×10^{-40}	33% (337)	5.9	5.1	7.5	3.8	2.1	7.5	0.3	33	50	37	6.2
*Tli*MBP*	85	409	2×10^{-32}	30% (334)	5.4	4.4	8.1	4.9	2.7	8.1	1.5	38	55	50	6.4
*Pfu*MBP*	100	381	1×10^{-31}	30% (350)	5.5	4.2	8.9	4.5	0.5	7.9	1.0	33	55	36	7.6
Mesophiles															
Clostridium perfringens	40	379	2×10^{-76}	41% (375)	7.1	5.0	5.8	5.0	0.5	9.2	0.8	38	49	40	5.0
Deinococcus radiodurans	26–30	375	6×10^{-48}	34% (354)	4.5	5.9	2.1	4.5	2.1	8.0	0.5	39	25	40	5.9
*Eco*MBP	37	370	9×10^{-41}	31% (383)	6.5	5.7	7.3	2.4	1.6	10.0	0.8	30	51	46	5.7
Enterobacter aerogenes	25–35	407	3×10^{-41}	30% (382)	6.2	5.9	5.9	2.4	1.6	10.5	0.8	31	45	48	5.9
Shigella flexneri	30–37	370	3×10^{-40}	30% (383)	6.5	5.7	7.3	2.4	1.6	10.0	0.8	30	51	46	5.7
Salmonella enterica	37	370	3×10^{-40}	32% (384)	6.8	5.7	5.9	3.2	1.6	10.3	0.5	33	47	46	5.7

Optimal growth of the organism (American Type Culture Collection, http://www.atcc.org/SearchCatalogs/Bacteria.cfm) is not necessarily the same as the temperature optimum for protein activity. Related proteins were identified using a BLAST search with the folded region of *Acy*MBP as a probe. Sequences selected as described in the text were obtained from GenBank as follows: *Acy*MBP, gi|6686562|; *Sulfolobus solfataricus*, gi|15898024|; *Thermoplasma acidophilum*, gi|16081289|; *Lactobacillus plantarum*, gi|28377115|; *Lactococcus lactis*, gi|15673665|; *Lactobacillus gasseri*, gi|23003417|; *Thermoanaerobacter tengcongensis*, gi|20808247|; *Thermotoga maritima*, gi|15643960|; *Thermoactinomyces vulgaris*, gi|13537290|; *Tli*MBP, gi|13787046|; *Pfu*MBP, gi|13096471|; *Clostridium perfringens*, gi|18311325|; *Deinococcus radiodurans*, gi|15805588|; *Eco*MBP, gi|2781044|. The mature regions corresponding to the structures compared here were calculated with the aid of SignalP, and then analyzed using ProtParam.

Table 4. Comparison of absolute numbers (and percentage) of residues involved in secondary structure

	α-Helix	3_{10}-Helix	β-Strand	β-Bridge	Total α	Total β	Total sec. str.	Total res.	Proline residues in sec. str.	Proline residues outside sec. str.
*Acy*MBP	159 (42.5)	12 (3.2)	68 (18.2)	3 (0.8)	171 (45.7)	71 (19.0)	242 (64.7)	374	11 (44.0)	14 (56.0)
*Eco*MBP	157 (42.4)	6 (1.6)	72 (19.5)	5 (1.4)	163 (44.1)	77 (20.8)	240 (64.9)	370	9 (42.9)	12 (57.1)
*Pfu*MBP	158 (42.0)	17 (4.5)	65 (17.3)	8 (2.1)	175 (46.5)	73 (19.4)	248 (66.0)	381	9 (32.1)	19 (67.9)
*Tli*MBP	182 (44.8)	10 (2.5)	65 (16.0)	9 (2.2)	192 (47.3)	74 (18.2)	266 (65.5)	407	11 (42.3)	15 (57.7)

N-terminal segment during either storage or crystallization. The combined observations suggest that this segment is normally disordered, which agrees well with the suggested biological role: a linker should be flexible to allow MBP to interact with carbohydrates in the periplasm as well as with the membrane-bound transport proteins. Presumably the linker lacks sequences that are recognized by proteases in its natural environment.

Like other MBPs, *Acy*MBP is designed to bind small α(1,4) polymers of glucose of varying size. Both aromatic stacking and hydrogen-bonding interactions are important in binding the cognate ligand tightly (K_d for maltose is 1.5 μM).[9] The presence of three clear glycosyl-binding subsites in the structure suggests that maltotriose should be the optimal ligand, and indeed, maltotriose is seen in the electron density (occupancy 0.35–0.65 in the various final structures), although it can have been only a minor contaminant (<3%) of the sugar added during purification and crystallization. Theoretically, there is sufficient space for longer sugar moieties at the non-reducing end, but additional glycosyl units would protrude from the binding site and lack good contacts with the protein. A preference for maltose and maltotriose would make sense in biological terms, since these are the main products when starch is digested by the amylopullulanase of *A. acidocaldarius*.[8]

Comparison of the *Acy*MBP structures with related proteins from a mesophile (*E. coli*) and two hyperthermophiles (*P. furiosus* and *T. litoralis*) gave other insights into structure and function. For example, the overall structural differences provide a good basis for understanding the various approaches to ligand binding. Two subsites (sites 1 and 2 in Figure 3) are shared by three of the proteins (*Acy*MBP, *Pfu*MBP and *Eco*MBP), and represent their common mode of binding to maltose. Like *Acy*MBP, *Pfu*MBP shows a strong preference for maltotriose over maltose or trehalose, placing the third sugar ring in site 3. *Eco*MBP also binds more tightly to maltotriose, but for several reasons binds the third glycosyl unit in a different subsite (site 1′ in Figure 3). Therefore, *Acy*MBP and *Pfu*MBP will deliver a trisaccharide to their membrane permease partners in a different way than *Eco*MBP, which further suggests that differences in the permeases will exist in these organisms.

Although the contributing residues are not always conserved, the number and position of hydrogen bonds for the two central glycosyl units (sites 1 and 2) are very similar in these three structures. Regardless of the mode of binding the third glycosyl unit, each protein utilizes a similar number of hydrogen bonds in this site. Further, *Acy*MBP, *Eco*MBP and *Pfu*MBP have similar layers of aromatic residues interacting with the relatively non-polar faces of the sugar rings, which provide much of the energy of binding. As might be expected, these interactions can be provided in a fairly non-specific fashion, although the large flat surfaces of aromatic residues are a natural complement to the shape and character of a glycosyl ring.[17] In sites 1–3, the largest surface (at left in Figure 3) is contributed by the C-terminal domain, and the residues involved are quite well conserved. By contrast, the interactions supplied by Phe45 in the N-terminal domain of the *Acy*MBP structure (at right in Figure 3) can be provided instead by residues from a different part of the sequence: Trp65 in *Pfu*MBP and Trp62 in *Eco*MBP. The aromatic component in *E. coli*'s site 1′ (Tyr341) is missing in *Acy*MBP and *Pfu*MBP, and inspection of a number of MBP sequences suggests the presence/absence of this tyrosine will be the best predictor of binding mode for the trisaccharide; the effects of insertions/deletions near site 3 are more difficult to evaluate.

Despite the sequence/structural homology, ligand binding in *Tli*MBP is substantially different from that of the other three MBPs, and resemblance is only in the approximate position of the bound ligand. *Tli*MBP has two sites in common

Table 5. Comparison of cavities

Structure	No. cavities	Total volume (Å^3)
*Acy*MBP-1	5 (0.7)	104.8 (26.6)
*Eco*MBP	11	511.3
*Pfu*MBP	4	331.8
*Tli*MBP	4	124.4

Minimum volume for a cavity was set to 5 Å^3. Standard deviations for the multiple observations of the *Acy*MBP structure are shown in parentheses.

with *Acy*MBP and *Pfu*MBP (sites 2 and 3 in Figure 3), but it has a completely different approach to binding in each of those sites. In contrast to the other proteins, *Tli*MBP accepts trehalose as well as maltose, but does not bind to longer oligosaccharides. The *P. furiosus* genome contains a second binding protein corresponding very closely to *Tli*MBP that can presumably take on the role of disaccharide binding in that organism.

Since acidophilic organisms can regulate their internal pH,[4,25,26] only proteins exposed to the external environment (i.e. either located in the periplasm or secreted, such as the binding proteins) must actually be stable under acidic conditions. *Acy*MBP thus represents a very rare opportunity to study a protein that is both thermostable and acid-stable. However, the problems of thermostability and acidostability are by no means the same. Factors that are thought to be correlated with the ability of a protein to remain stable and active at high temperatures include an increased number of hydrogen bonds, additional or improved electrostatic interactions including those mediated by salt bridges, optimized hydrophobic interactions, increased compactness or packing density, increased secondary-structural content, truncation of solvent-exposed loops, greater number of proline residues in loops or other key positions, and increased polar compared with non-polar surface areas.[27–29] Much less is known about acidostability or thermoacidostability. The availability of a range of MBP structures and sequences allows us to assess their features in the light of differing physical settings.

Matzke *et al.*[30] analyzed the sequences of α-amylases using the crystal structure of the Taka enzyme from *Aspergillus oryzae* and predicted a reduced charge density on the surface of the secreted *A. acidocaldarius* enzyme. It was believed that this adaptation would prevent electrostatic repulsion of charged groups at low pH. Our analysis of *Acy*MBP showed that there are indeed fewer charged residues on its surface, and furthermore that most of these carry a positive charge. Although the numbers of basic and acidic residues are equal in *Acy*MBP (28 *versus* 29), most acidic residues are buried within the protein. At the pH optimal for the organism (~3.5), the few exposed acidic groups will in general be protonated, giving this protein a rather dramatic positive charge surplus. This is a different strategy to that observed in the acid-stable (but not thermostable) xylanase from *Aspergillus kawachii* in which many acidic residues are exposed (although even that protein is expected to be positively charged at the pH optimum of 2).[31] The preponderance of basic residues would seem to have the advantage that their high pK_a values are outside the pH range for optimum activity (pH 2.5–7), which would effectively protect the protein from destabilizing changes in surface charge if the pH in the environment varies. This is in fact the case with more "normal" proteins: the pK_a values of the vast majority of their surface groups (3–5 for acidic residues and 9–11 for basic residues) are outside the usual pH range of their environment. The only exception is histidine with a pK_a of 6.05; histidine, however, is relatively rare in the compared MBPs. Base-stable proteins might show a corresponding reduction in the number of basic residues. If so many surface residues carry the same charge, it is reasonable to postulate that charge repulsion will be a problem if their number is too large. Overall, the thermoacidophilic amylopullulanase was shown to contain approximately 30% fewer charged residues than its closest relatives.[8,30] Analysis of the amino acid sequences shows that (hyper)thermoacidophilic MBPs also have a smaller total number of charged residues than their (hyper)thermophilic or mesophilic counterparts. These results suggest that the trend is a general one, and linked to acidostability specifically.

The number of salt bridges buried within *Acy*MBP is smaller than for its hyperthermophilic as well as mesophilic counterparts, while the numbers exposed on the surface are roughly the same. This is in accordance with previous suggestions that buried salt bridges can actually be destabilizing,[32,33] and that an increased number of such salt bridges is tolerated in thermophilic proteins, rather than being an asset.[34] The results presented here suggest that reducing the number of buried salt bridges may be a factor in acidostability.

An increase in the number of polar but uncharged residues is found both on the surface and inside *Acy*MBP compared to the other proteins. Such substitutions on the surface would be expected to help maintain the polar-outside/non-polar-inside balance that is critical for a folded protein in an aqueous environment.[8]

The content of secondary structure does not appear to vary significantly in the various MBPs (Table 4). Proline content was slightly higher than average in *Acy*MBP (6.7% compared to ~5.7% for mesophiles, Table 3), but does not appear to be correlated with a strengthening of the existing secondary structure. Further, the slightly higher proline content is also observed for the hyperthermophilic MBPs, and so may be linked to thermo- rather than acidostability; an increased number of proline residues at the second position of β-turns is a recurring feature in the (hyper)thermophilic proteins.

*Acy*MBP and the hyperthermophilic MBPs have fewer and smaller cavities compared to their mesophilic counterpart, thus suggesting that more thermostable MBPs are generally more compact (Table 5). Other factors, such as the number of insertions and deletions in the sequences/structures, do not appear to be correlated with differences in stability (Figure 5). *Acy*MBP is a monomer (confirmed here by dynamic light scattering), and so improved interactions within a multimer[35] are not a viable route to its stabilization.

The results presented here provide a number of testable hypotheses that can be applied in further work using existing, as well as new, structures and sequences.

Materials and Methods

Protein preparation and analysis

The N-terminally truncated version of *Acy*MBP was prepared from the culture supernatant of maltose-grown cells of *A. acidocaldarius* as described.[13] Isoelectric focusing experiments were carried out using Ampholine PAGplate gels (pH range 3.5–9.5, Amersham Biosciences) on a PHAST electrophoresis system. Dynamic light scattering experiments were performed at a protein concentration of 6 mg/ml, in a solution of 20 mM sodium acetate (pH 3.5), 9.8 mM ammonium sulphate and 10 mM maltose (the normal storage buffer).

A full-length version of *Acy*MBP was purified from the cytosolic fraction of *E. coli* strain JM109(pRF1) by ion exchange chromatography and Ni-NTA affinity chromatography. Plasmid pFR1 is a derivative of expression vector pQE9 (Qiagen, Germany) that contains the *malE* gene lacking its signal sequence and with the wild-type initiation codon TGT (Cys) replaced by GCG (Ala). The resulting translated protein has the sequence $MRGSH_6$-GS fused to the N-terminal alanine residue. Cells were grown in Luria–Bertani broth at 30 °C; gene expression was induced with 0.5 mM IPTG at $A_{650} = 0.4$ and growth continued for four hours. Cell disruption was carried out with a French pressure cell in a solution containing 50 mM Tris–HCl (pH 7.5), 0.1 mM EDTA, 20% (v/v) glycerol, 2 mM DTT, 0.1 mM PMSF, 10 mM maltose, and followed by ultracentrifugation. The resulting supernatant was applied to an SP-Sepharose column equilibrated with 50 mM Tris–HCl (pH 7.2), 5% glycerol, 0.5 mM PMSF, 10 mM maltose (Buffer A). *Acy*MBP was eluted with Buffer A containing 100 mM NaCl and subsequently incubated with Ni-NTA matrix in the same buffer for 20 minutes at room temperature. The mixture was then transferred to a column and the resin was extensively washed with Buffer A containing 20 mM imidazole. *Acy*MBP was eluted with the same buffer containing 150 mM imidazole, concentrated, dialyzed against Buffer A without imidazole and stored at 4 °C until use.

Crystallization and data collection

The crystals for the *Acy*MBP-1 structure (intact mature protein) were obtained by the sitting-drop vapour diffusion method: 1 μl protein solution (9.6 mg/ml in 50 mM Tris–HCl (pH 7.2), 5% glycerol, 10 mM maltose) was mixed with 1 μl of reservoir solution containing 20% (w/v) PEG 8000 and 100 mM Ches buffer (pH 9.5), and equilibrated against the reservoir solution at 18 °C. Data were collected in a glass capillary at room temperature using a rotating anode X-ray source equipped with a Mar345 detector, and processed using XDS2002.[36] Data statistics are summarized in Table 1.

Crystals of the intact mature protein in a larger unit cell (*Acy*MBP-2) were obtained using the hanging-drop vapour diffusion method: 2 μl protein solution (15 mg/ml in 50 mM in 50 mM Tris–HCl (pH 7.2), 5% glycerol, 10 mM maltose) were mixed with 2 μl of reservoir solution (18% PEG 8000, 100 mM Ches (pH 9.5), 10% glycerol) and equilibrated against the latter. Prior to freezing, crystals were transferred to a cryo-protection solution containing 20% PEG 8000, the mixed Ches/Tris–HCl buffer, 20% glycerol, 10 mM maltose. Data were collected at the Swiss-Light-Source (beamline X06SA), as high and low-resolution sets that were processed and merged using XDS2002 (Table 1).

Crystals of the truncated protein in the larger unit cell (*Acy*MBP-3) grew from hanging drops after three weeks at 37 °C: 2 μl of protein solution (10 mg/ml in 10 mM sodium acetate (pH 3.8), 9.8 mM ammonium sulphate, 10 mM maltose) was mixed with 2 μl of reservoir solution (20% PEG 5000, 100 mM Tris–HCl (pH 8.0), 80 mM ammonium sulphate). Prior to freezing in liquid nitrogen, crystals were soaked in mother liquor containing 20% glycerol. Two diffraction data sets were collected, at I711 at MAX-lab, Lund, Sweden and ID14-4, Grenoble, France. Data were processed using DENZO and SCALEPACK[37] and merged using programs of the CCP4 package[38] (Table 1).

Structure solution and refinement

The structure of *Acy*MBP-1 was solved using molecular replacement with the program package CNS[39] including data in the resolution range 15–4 Å. The search model consisted of all protein atoms of the ligand-bound *E. coli* protein (PDB entry 1ANF).[14] Refinement was also carried out using CNS including all data to 1.8 Å. Combined energy minimization/simulated annealing, and alternating cycles of rebuilding with the graphics program O,[40] resulted in an R_{factor} of 19.4% and a R_{free} of 21.7% for the final model (summarized in Table 1).

The structures *Acy*MBP-2 and 3 were solved by molecular replacement using the final structure of *Acy*MBP-1, with all protein atoms as a search model. Two molecules were located in the asymmetric unit of both datasets using AMoRe[41] (resolution range 8–4 Å). This was followed by rigid body refinement (resolution range 40–3 Å) that allowed the two domains of each molecule to move independently (carried out using the program REFMAC[42] as implemented in CCP4).[38] Further cycles of restrained refinement using all data to the maximum observed resolution were alternated with manual rebuilding. Water molecules were added using CNS.[39] Statistics relating to the final refined models are summarized in Table 1.

Structural analysis, sequence and structural comparisons

Coordinate sets used for the comparisons with the new MBP structures were obtained from the Protein Data Bank[43,44] as follows: 1ANF (*Eco*MBP with maltose),[14] 3MBP (*Eco*MBP with maltotriose),[14] 4MBP (*Eco*MBP with maltotetraose),[14] 1ELJ (*Pfu*MBP)[11] and 1EU8 (*Tli*MBP).[12]

Structures were compared using O, Swiss-PdbViewer,[45] LSQMAN,[46] VOIDOO[21] and FIT†. Secondary structural content was assessed using DSSP.[20] Surface accessibility was assessed using a per-residue cut-off of 30 Å^2 in Swiss-PdbViewer[45] (accessibility calculated using a 1.5 Å probe). Structure-based and other sequence alignments were carried out using LSQMAN[47] and Indonesia (Madsen, Kleywegt and Johansson, Uppsala

† http://bioinfo1.mbfys.lu.se/~guoguang/fit.html

University). Similar sequences were located using BLAST.[19] Sequences were obtained from GenBank.[48] Signal sequence cleavage sites were predicted using SignalP version 1.1.[49] Calculation of isoelectric points and amino acid compositions was performed using the server in the web†. Figures were prepared with the programs O, Molray,[50] Molscript,[51] LIGPLOT[52] and Canvas (Deneba Systems, Inc.).

Accession numbers

Coordinates and structure factors have been deposited with the Protein Data Bank with entry codes 1urg, 1urs and 1urd.[44]

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SCHN 274/6-3;6-4), the Fonds der Chemischen Industrie (to E.S.), grants from the Swedish Research Foundation (VR) to S.L.M. as well as by Uppsala University and the Swedish University of Agricultural Sciences. The authors thank Rebecca Fleischer for constructing pRF1, C. Schulze-Briese and T. Tomizaki for support with data collection at SLS, and Gerard Kleywegt for helpful discussions.

References

1. Schneider, E. (2003). Import of solutes by ABC transporters: the maltose and other systems. In *ABC Proteins: From Bacteria to Man* (Cole, S., Higgins, C., Holland, E. B. & Kuchler, K., eds), pp. 157–185, Academic Press, London.
2. Sutcliffe, I. C. & Russell, R. R. (1995). Lipoproteins of gram-positive bacteria. *J. Bacteriol.* **177**, 1123–1128.
3. Albers, S. V., Elferink, M. G., Charlebois, R. L., Sensen, C. W., Driessen, A. J. & Konings, W. N. (1999). Glucose transport in the extremely thermoacidophilic *Sulfolobus solfataricus* involves a high-affinity membrane-integrated binding protein. *J. Bacteriol.* **181**, 4285–4291.
4. Bakker, E. P. (1990). The role of alkali-cation transport in energy coupling of neutrophilic and acidophilic bacteria: an assessment of methods and concepts. *FEMS Microbiol. Rev.* **75**, 319–334.
5. Darland, G. & Brock, T. D. (1971). *Bacillus acidocaldarius* sp. nov., an acidophilic thermophilic spore-forming bacterium. *J. Gen. Microbiol.* **67**, 9–15.
6. Wisotzkey, J. D., Jurtshuk, P., Jr, Fox, G. E., Deinhard, G. & Poralla, K. (1992). Comparative sequence analyses on the 16 S rRNA (rDNA) of *Bacillus acidocaldarius*, *Bacillus acidoterrestris*, and *Bacillus cycloheptanicus* and proposal for creation of a new genus, *Alicyclobacillus* gen. nov. *Int. J. Syst. Bacteriol.* **42**, 263–269.
7. Matzke, J., Herrmann, A., Schneider, E. & Bakker, E. P. (2000). Gene cloning, nucleotide sequence and biochemical properties of a cytoplasmic cyclomaltodextrinase (neopullulanase) from *Alicyclobacillus acidocaldarius*, reclassification of a group of enzymes [published erratum appears in FEMS Microbiol. Letters, 2000 Jul 1; 188(1): 107]. *FEMS Microbiol. Letters*, **183**, 55–61.
8. Schwermann, B., Pfau, K., Liliensiek, B., Schleyer, M., Fischer, T. & Bakker, E. P. (1994). Purification, properties and structural aspects of a thermoacidophilic alpha-amylase from *Alicyclobacillus acidocaldarius* ATCC 27009. Insight into acidostability of proteins. *Eur. J. Biochem.* **226**, 981–991.
9. Hülsmann, A., Lurz, R., Scheffel, F. & Schneider, E. (2000). Maltose and maltodextrin transport in the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius* is mediated by a high-affinity transport system that includes a maltose binding protein tolerant to low pH. *J. Bacteriol.* **182**, 6292–6301.
10. Sharff, A. J., Rodseth, L. E., Spurlino, J. C. & Quiocho, F. A. (1992). Crystallographic evidence of a large ligand-induced hinge-twist motion between the two domains of the maltodextrin binding protein involved in active transport and chemotaxis. *Biochemistry*, **31**, 10657–10663.
11. Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. (2001). Structural basis for oligosaccharide recognition by *Pyrococcus furiosus* maltodextrin-binding protein. *J. Mol. Biol.* **305**, 891–904.
12. Diez, J., Diederichs, K., Greller, G., Horlacher, R., Boos, W. & Welte, W. (2001). The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic archaeon *Thermococcus litoralis* at 1.85 Å. *J. Mol. Biol.* **305**, 905–915.
13. Herrmann, A., Schlösser, A., Schmid, R. & Schneider, E. (1996). Biochemical identification of a lipoprotein with maltose-binding activity in the thermoacidophilic Gram-positive bacterium *Alicyclobacillus acidocaldarius*. *Res. Microbiol.* **147**, 733–737.
14. Quiocho, F. A., Spurlino, J. C. & Rodseth, L. E. (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure*, **5**, 997–1015.
15. Shilton, B. H., Flocco, M. M., Nilsson, M. & Mowbray, S. L. (1996). Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins. *J. Mol. Biol.* **264**, 350–363.
16. Diruggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R. *et al.* (2000). Evidence of recent lateral gene transfer among hyperthermophilic archaea. *Mol. Microbiol.* **38**, 684–693.
17. Vyas, N. K. (1991). Atomic features of protein–carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1**, 732–740.
18. Spurlino, J. C., Lu, G.-Y. & Quiocho, F. A. (1991). The 2.3 Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266**, 5202–5219.
19. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
20. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.

† http://www.expasy.org/tools/protparam.html

21. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallog. sect. D*, **50**, 178–185.
22. Watanabe, K., Chishiro, K., Kitamura, K. & Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6-glucosidase from *Bacillus thermoglucosidasius* KP1006. *J. Biol. Chem.* **266**, 24287–24294.
23. Watanabe, K., Kitamura, K. & Suzuki, Y. (1996). Analysis of the critical sites for protein thermostabilization by proline substitution in oligo-1,6-glucosidase from *Bacillus coagulans* ATCC 7050 and the evolutionary consideration of proline residues. *Appl. Environ. Microbiol.* **62**, 2066–2073.
24. Prieto, J. & Serrano, L. (1997). C-capping and helix stability: the Pro C-capping motif. *J. Mol. Biol.* **274**, 276–288.
25. Cobley, J. G. & Cox, J. C. (1983). Energy conservation in acidophilic bacteria. *Microbiol. Rev.* **47**, 579–595.
26. Krulwich, T. A., Davidson, L. F., Filip, S. J., Jr, Zuckerman, R. S. & Guffanti, A. A. (1978). The protonmotive force and beta-galactoside transport in *Bacillus acidocaldarius*. *J. Biol. Chem.* **253**, 4599–4603.
27. Ladenstein, R. & Antranikian, G. (1998). Proteins from hyperthermophiles: stability and enzymatic catalysis close to the boiling point of water. *Advan. Biochem. Eng. Biotechnol.* **61**, 37–85.
28. Kumar, S. & Nussinov, R. (2001). How do thermophilic proteins deal with heat? *Cell. Mol. Life Sci.* **58**, 1216–1233.
29. Sanchez-Ruiz, J. M. & Makhatadze, G. I. (2001). To charge or not to charge? *Trends Biotechnol.* **19**, 132–135.
30. Matzke, J., Schwermann, B. & Bakker, E. P. (1997). Acidostable and acidophilic proteins: the example of the alpha-amylase from *Alicyclobacillus acidocaldarius*. *Comp. Biochem. Physiol. A Physiol.* **118**, 475–479.
31. Fushinobu, S., Ito, K., Konno, M., Wakagi, T. & Matsuzawa, H. (1998). Crystallographic and mutational analyses of an extremely acidophilic and acid-stable xylanase: biased distribution of acidic residues and importance of Asp37 for catalysis at low pH. *Protein Eng.* **11**, 1121–1128.
32. Waldburger, C. D., Jonsson, T. & Sauer, R. T. (1996). Barriers to protein folding: formation of buried polar interactions is a slow step in acquisition of structure. *Proc. Natl Acad. Sci. USA*, **93**, 2629–2634.
33. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nature Struct. Biol.* **2**, 122–128.
34. Elcock, A. H. (1998). The stability of salt bridges at high temperatures: implications for hyperthermophilic proteins. *J. Mol. Biol.* **284**, 489–502.
35. Russell, R. J. & Taylor, G. L. (1995). Engineering thermostability: lessons from thermophilic proteins. *Curr. Opin. Biotechnol.* **6**, 370–374.
36. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
37. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
38. Collaborative Computing Project Number 4 (1994). The CCP4 Suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
39. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system (CNS): a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
40. 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.
41. Navaza, J. & Saludjian, P. (1997). AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581–594.
42. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
43. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. T., Jr, Brice, M. D., Rodgers, J. R. *et al.* (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
44. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucl. Acids Res.* **28**, 235–242.
45. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
46. Kleywegt, G. J. & Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525–545.
47. Kleywegt, G. J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallog. sect. D*, **52**, 842–857.
48. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. (2003). GenBank. *Nucl. Acids Res.* **31**, 23–27.
49. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.
50. Harris, M. & Jones, T. A. (2001). Molray—a web interface between O and the POV-Ray ray tracer. *Acta Crystallog. sect. D*, **57**, 1201–1203.
51. Kraulis, P. (1991). Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
52. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein–ligand interactions. *Protein Eng.* **8**, 127–134.
53. Kleywegt, G. J. & Jones, T. A. (1996). Phi/Psi-cology: Ramachandran revisited. *Structure*, **4**, 1395–1400.

Edited by R. Huber

(Received 14 July 2003; received in revised form 15 October 2003; accepted 17 October 2003)