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

Karsten Schäfer¹, Ulrika Magnusson², Frank Scheffel³, André Schiefner¹, Mats O. J. Sandgren², Kay Diederichs¹, Wolfram Welte¹, Anja Hülsmann³, Erwin Schneider³ and Sherry L. Mowbray⁴*

1 Fachbereich Biologie
Universität Konstanz
Universitätsstraße 10, 78457 Konstanz, Germany

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

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

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

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
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. 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, or alternatively, as observed for some archaea, an N-terminal hydrophobic helix.

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. 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 *Acidocataldium acidocaldarius*. This Gram-positive bacterium has a pH optimum of 3.6 and a temperature optimum of 57 °C. Comparison of the core (βα)8-barrel domain of the secreted α-amylase (amylopullulanase) 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. This trend had not been identified in thermostable proteins, and so was believed to reflect the enzymes' 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* (AcyMBP) at high resolution, and in multiple crystal packing environments. This protein binds maltose with high affinity (Kₐ 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 thermostability. Furthermore, the availability of the crystal structures for *Escherichia coli* MBP (EcoMBP) as well as two hyperthermostable MBPs (from *Pyrococcus furiosus*, PfuMBP and *Thermococcus litoralis*, TliMBP) allow an excellent opportunity to investigate the molecular adaptations related to acidostability.

**Results**

**Protein expression and purification**

A truncated variant of AcyMBP was isolated from the culture supernatant of maltose-grown cells of *A. acidocaldarius* and purified by affinity chromatography on Sepharose-linked amylase. N-terminal sequencing (SALPKGQTTT) and mass spectrometry (M, 41,250 Da) confirmed earlier conclusions 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.

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 AcyMBP (full-length protein, referred to as AcyMBP-1 in further discussion) was obtained by molecular replacement using the structure of ligand-bound EcoMBP, 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 (AcyMBP-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 AcyMBP-3a and 3b. Statistics for data collection and final refined structures are presented in Table 1.
Overall structures

As typically found in periplasmic-binding proteins, the structure of AcyMBP 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 AcyMBP 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 Cα 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

---

Table 1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th></th>
<th>AcyMBP-1</th>
<th>AcyMBP-2</th>
<th>AcyMBP-3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Environment</td>
<td>Rotating anode, MAR 345 detector ($T = 291$ K)</td>
<td>Synchrotron, MAR CCD ($T = 100$ K)</td>
<td>Synchrotron, MAR 345 detector ($T = 100$ K)</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.54179</td>
<td>0.91964</td>
<td>1.0201</td>
</tr>
<tr>
<td>Space group</td>
<td>$P2_1$</td>
<td>$P2_1$</td>
<td>$P2_1$</td>
</tr>
<tr>
<td>Cell (Å and °)</td>
<td>50.11, 72.16, 57.78, $\beta = 109.63$</td>
<td>49.18, 70.53, 104.06, $\beta = 96.98$</td>
<td>49.23, 70.78, 104.67, $\beta = 96.58$</td>
</tr>
<tr>
<td>Resolution range (Å, highest shell)</td>
<td>50.0–1.80 (1.90–1.80)</td>
<td>40.14–1.45 (1.50–1.45)</td>
<td>40.0–1.53 (1.56–1.53)</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>35,989</td>
<td>137,804</td>
<td>99,996</td>
</tr>
<tr>
<td>No. reflections total</td>
<td>34,421</td>
<td>137,798</td>
<td>99,996</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>15.5 (32.4)</td>
<td>8.4 (50.9)</td>
<td>8.3 (18.7)</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.5 (91.4)</td>
<td>99.6 (98.7)</td>
<td>98.8 (88.3)</td>
</tr>
<tr>
<td>$I/\sigma I$</td>
<td>7.5 (2.1)</td>
<td>12.7 (2.4)</td>
<td>11.9 (5.3)</td>
</tr>
</tbody>
</table>

| **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 β-factor, protein atoms (Å²) | 23.3                          | 17.3, 19.0                    | 11.6, 12.9                    |
| Mean β-factor, ligand atoms (Å²) | 12.8                          | 15.6                          | 10.1                          |
| Mean β-factor, solvent atoms (Å²) | 32.0                          | 27.3                          | 21.2                          |
| Ramachandran plot outliers (%) | 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.
* Using a stringent boundary Ramachandran plot.34

---


---

Figure 1. Overall structure. The structure of AcyMBP 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).
ligand-binding site are very similar in all structures. Although three of the models (AcyMBP-1, 2a and 2b) represent a construct that originally included the N-terminal extension found in vivo, 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 AcyMBP-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 AcyMBP are compared in Table 2. The most similar to AcyMBP is EcoMBP, 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 AcyMBP-1 structure). Among the residues in this core structure, 31% of the amino acid residues are identical. A comparable result was obtained with PfuMBP, with 80% of its residues matching to equivalents in AcyMBP. 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).

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 AcyMBP-2 and 3. The electron density for this third ring was stronger in both AcyMBP-2 structures than in those of AcyMBP-3; it was completely absent in AcyMBP-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 AcyMBP-2 was in this way estimated to be 0.65 in both molecules, and in AcyMBP-3a and 3b, the corresponding values are 0.30 and 0.45, respectively.

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, non-polar interactions largely involve contacts with aromatic residues, in this case Tyr179, Trp255, Met356, Trp366 and Met369 of the C-terminal domain, and

<table>
<thead>
<tr>
<th>Model</th>
<th>AcyMBP</th>
<th>EcoMBP</th>
<th>PfuMBP</th>
<th>TliMBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcyMBP</td>
<td>374</td>
<td>1.60</td>
<td>1.74</td>
<td>1.81</td>
</tr>
<tr>
<td>EcoMBP</td>
<td>321 (31.2)</td>
<td>370</td>
<td>1.65</td>
<td>1.83</td>
</tr>
<tr>
<td>PfuMBP</td>
<td>299 (32.3)</td>
<td>309 (29.8)</td>
<td>381</td>
<td>1.91</td>
</tr>
<tr>
<td>TliMBP</td>
<td>288 (25.0)</td>
<td>311 (23.8)</td>
<td>310 (26.6)</td>
<td>407</td>
</tr>
</tbody>
</table>

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.

Figure 2. Ligand binding in AcyMBP. A, Maltose (and maltotriose) density in the last electron density map prior to addition of ligand in the AcyMBP-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 AcyMBP-1.
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 AcyMBP structures also have full occupancy in the available PfuMBP structure11 (Figure 3B). When maltose is bound to EcoMBP, the same two sites are preferred18 (Figure 3C). Many of the interactions in these two sites are conserved in the three proteins. Counterparts to AcyMBP’s Asp95, Tyr179, Phe180, Trp255 and Trp366 are present in both PfuMBP and EcoMBP. A third glycosyl ring was also observed with partial occupancy in PfuMBP at a position very similar to that seen in AcyMBP (site 3 in Figure 3B); this was likewise believed to represent maltotriose, a common contaminant of maltose. When maltotriose is bound to EcoMBP, 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 AcyMBP, residues 37–41 form hydrogen bonds to the glycosyl unit in site 3, and corresponding interactions are found in PfuMBP. 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 AcyMBP and PfuMBP structures prevent them from binding a third glycosyl unit in the same way as EcoMBP. Binding of trehalose (α-D-glucosyl-1,1-α-D-glucose) to TliMBP positions the two glycosyl units in sites 2 and 3 of AcyMBP and PfuMBP (Figure 3D). The plane of the sugar ring in site 2 has a very similar orientation in TliMBP, although no interactions are conserved. The plane of the sugar ring in site 3 is almost anti-parallel with that in AcyMBP and PfuMBP, 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 AcyMBP that the experimentally measured isoelectric point (~10) was very different from that calculated based on its amino acid sequence (6.5). The AcyMBP 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 AcyMBP, EcoMBP and TliMBP have a similar number of basic residues on their surfaces, while PfuMBP 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 AcyMBP 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 AcyMBP 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 AcyMBP.

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 EcoMBP (six buried), 13 in PfuMBP (nine buried) and nine in TliMBP (seven buried). In each of the AcyMBP structures, only four salt bridges were observed, two of which were buried. Thus, fewer salt bridges are buried in AcyMBP 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 AcyMBP, both on the surface and in the interior of the protein (Figure 4C and D). An increased number of polar residues in AcyMBP 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 search as being relatives of AcyMBP. Only a limited number of hyperthermoadophilic 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
determine whether any trends were general within the structural family. Results are summarized in Table 3. It immediately becomes apparent that, like AcyMBP, 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 AcyMBP.

Other features

The various MBP structures were assessed using
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, 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, 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 EcoMBP 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. 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)thermophilic structures (AcyMBP, four cases; PfuMBP, five cases; TliMBP, seven cases) than in the mesophilic one (EcoMBP, one case). The distribution of proline residues at the N and C-terminal ends of helices does not appear to differ significantly, while only PfuMBP shows an appreciable decrease in the frequency of proline residues within elements of secondary structure (Table 4).

It was noted earlier that PfuMBP has several dramatic clusters of isoleucine and aromatic residues, features that were suggested to increase its thermal stability relative to EcoMBP. However, this phenomenon is not apparent in either AcyMBP or TliMBP.

Discussion

In vivo, the mature form of AcyMBP 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). In the present study, two of the crystals were prepared using full-length AcyMBP. 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

<table>
<thead>
<tr>
<th>Organism</th>
<th>Optimum t of growth (deg.C)</th>
<th>No. Res.</th>
<th>BLAST E value</th>
<th>BLAST id (overlap)</th>
<th>Asp</th>
<th>Asn</th>
<th>Glu</th>
<th>Gln</th>
<th>Arg</th>
<th>Lys</th>
<th>His</th>
<th>Total Asn + Gln</th>
<th>Total Asp + Glu</th>
<th>Total Arg + Lys + His</th>
<th>Pro</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>AcylMBP</em></td>
<td>57</td>
<td>374</td>
<td>0.0</td>
<td>–</td>
<td>5.1</td>
<td>5.1</td>
<td>2.4</td>
<td>8.0</td>
<td>1.3</td>
<td>5.9</td>
<td>1.3</td>
<td>49</td>
<td>28</td>
<td>32</td>
<td>6.7</td>
</tr>
<tr>
<td><em>Thermoplasma acidophilum</em></td>
<td>55–60</td>
<td>437</td>
<td>1 x 10^−9</td>
<td>21% (363)</td>
<td>3.0</td>
<td>7.8</td>
<td>4.1</td>
<td>4.8</td>
<td>0.2</td>
<td>3.9</td>
<td>1.1</td>
<td>55</td>
<td>51</td>
<td>23</td>
<td>5.0</td>
</tr>
<tr>
<td><em>Sulfolobus solfataricus</em></td>
<td>80</td>
<td>450</td>
<td>5 x 10^{−23}</td>
<td>27% (407)</td>
<td>2.0</td>
<td>6.7</td>
<td>3.6</td>
<td>7.8</td>
<td>1.3</td>
<td>2.4</td>
<td>0.9</td>
<td>65</td>
<td>25</td>
<td>21</td>
<td>6.4</td>
</tr>
<tr>
<td>Acid-tolerant mesophiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Lactobacillus plantarum</em></td>
<td>37</td>
<td>370</td>
<td>2 x 10^{−32}</td>
<td>29% (334)</td>
<td>5.4</td>
<td>7.6</td>
<td>2.7</td>
<td>4.1</td>
<td>0.8</td>
<td>13.5</td>
<td>0.3</td>
<td>43</td>
<td>40</td>
<td>30</td>
<td>5.4</td>
</tr>
<tr>
<td><em>Lactococcus lactis</em></td>
<td>25–35</td>
<td>369</td>
<td>7 x 10^{−21}</td>
<td>27% (380)</td>
<td>8.1</td>
<td>6.5</td>
<td>2.4</td>
<td>5.7</td>
<td>0.0</td>
<td>9.8</td>
<td>0.0</td>
<td>45</td>
<td>39</td>
<td>36</td>
<td>5.1</td>
</tr>
<tr>
<td><em>Lactobacillus gasseri</em></td>
<td>37</td>
<td>381</td>
<td>9 x 10^{−14}</td>
<td>25% (377)</td>
<td>4.7</td>
<td>7.6</td>
<td>4.7</td>
<td>5.2</td>
<td>0.3</td>
<td>12.9</td>
<td>1.0</td>
<td>49</td>
<td>36</td>
<td>54</td>
<td>5.2</td>
</tr>
<tr>
<td><em>Thermoactinomyces vulgaris</em></td>
<td>40–50</td>
<td>397</td>
<td>3 x 10^{−39}</td>
<td>29% (381)</td>
<td>5.0</td>
<td>5.8</td>
<td>6.0</td>
<td>6.0</td>
<td>1.0</td>
<td>11.3</td>
<td>0.5</td>
<td>47</td>
<td>44</td>
<td>51</td>
<td>6.5</td>
</tr>
<tr>
<td><em>Thermoanaerobacter tengcongensis</em></td>
<td>75</td>
<td>372</td>
<td>1 x 10^{−75}</td>
<td>43% (369)</td>
<td>7.0</td>
<td>5.4</td>
<td>4.6</td>
<td>5.1</td>
<td>0.3</td>
<td>9.7</td>
<td>1.1</td>
<td>39</td>
<td>43</td>
<td>41</td>
<td>7.0</td>
</tr>
<tr>
<td><em>Thermotoga maritima</em></td>
<td>70–80</td>
<td>373</td>
<td>4 x 10^{−40}</td>
<td>33% (337)</td>
<td>5.9</td>
<td>5.1</td>
<td>7.5</td>
<td>3.8</td>
<td>2.1</td>
<td>7.5</td>
<td>0.3</td>
<td>33</td>
<td>50</td>
<td>40</td>
<td>6.2</td>
</tr>
<tr>
<td><em>TliMBP</em></td>
<td>85</td>
<td>409</td>
<td>2 x 10^{−32}</td>
<td>30% (334)</td>
<td>5.4</td>
<td>4.4</td>
<td>8.1</td>
<td>4.9</td>
<td>2.7</td>
<td>8.1</td>
<td>1.5</td>
<td>38</td>
<td>55</td>
<td>50</td>
<td>6.4</td>
</tr>
<tr>
<td><em>PfuMBP</em></td>
<td>100</td>
<td>381</td>
<td>1 x 10^{−31}</td>
<td>30% (350)</td>
<td>5.5</td>
<td>4.2</td>
<td>8.9</td>
<td>4.5</td>
<td>0.5</td>
<td>7.9</td>
<td>1.0</td>
<td>33</td>
<td>55</td>
<td>56</td>
<td>7.6</td>
</tr>
<tr>
<td>Mesophiles</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clostridium perfringens</em></td>
<td>40</td>
<td>379</td>
<td>2 x 10^{−76}</td>
<td>41% (375)</td>
<td>7.1</td>
<td>5.0</td>
<td>5.8</td>
<td>5.0</td>
<td>0.5</td>
<td>9.2</td>
<td>0.8</td>
<td>38</td>
<td>49</td>
<td>40</td>
<td>5.0</td>
</tr>
<tr>
<td><em>Deinococcus radiouarus</em></td>
<td>26–30</td>
<td>375</td>
<td>6 x 10^{−48}</td>
<td>34% (354)</td>
<td>4.5</td>
<td>5.9</td>
<td>2.1</td>
<td>4.5</td>
<td>2.1</td>
<td>8.0</td>
<td>0.5</td>
<td>39</td>
<td>25</td>
<td>40</td>
<td>5.9</td>
</tr>
<tr>
<td><em>EcoMBP</em></td>
<td>37</td>
<td>370</td>
<td>9 x 10^{−41}</td>
<td>31% (383)</td>
<td>6.5</td>
<td>5.7</td>
<td>7.3</td>
<td>2.4</td>
<td>1.6</td>
<td>10.0</td>
<td>0.8</td>
<td>30</td>
<td>51</td>
<td>46</td>
<td>5.7</td>
</tr>
<tr>
<td><em>Enterobacter aerogenes</em></td>
<td>25–35</td>
<td>407</td>
<td>3 x 10^{−41}</td>
<td>30% (382)</td>
<td>6.2</td>
<td>5.9</td>
<td>5.9</td>
<td>2.4</td>
<td>1.6</td>
<td>10.5</td>
<td>0.8</td>
<td>31</td>
<td>45</td>
<td>48</td>
<td>5.9</td>
</tr>
<tr>
<td><em>Shigella flexneri</em></td>
<td>30–37</td>
<td>370</td>
<td>3 x 10^{−40}</td>
<td>30% (383)</td>
<td>6.5</td>
<td>5.7</td>
<td>7.3</td>
<td>2.4</td>
<td>1.6</td>
<td>10.0</td>
<td>0.8</td>
<td>30</td>
<td>51</td>
<td>46</td>
<td>5.7</td>
</tr>
<tr>
<td><em>Salmonella enterica</em></td>
<td>37</td>
<td>370</td>
<td>3 x 10^{−40}</td>
<td>32% (394)</td>
<td>6.8</td>
<td>5.7</td>
<td>5.9</td>
<td>3.2</td>
<td>1.6</td>
<td>10.3</td>
<td>0.5</td>
<td>33</td>
<td>47</td>
<td>46</td>
<td>5.7</td>
</tr>
</tbody>
</table>

Optimal growth of the organism (American Type Culture Collection, [http://www.atcc.org/SearchCatalogs/Bacteria.cfm](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 *AcylMBP* as a probe. Sequences selected as described in the text were obtained from GenBank as follows: *AcylMBP*, gi|6686562|; *Sulfolobus solfataricus*, gi|158988024|; *Thermoplasma acidophilum*, gi|160812289|; *Lactobacillus plantarum*, gi|28377115|; *Lactococcus lactis*, gi|15673665|; *Lactobacillus gasseri*, gi|13537290|; *Thermoactinomyces vulgaris*, gi|13787046|; *TliMBP*, gi|13096471|; *Clostridium perfringens*, gi|18311325|; *Deinococcus radiouarus*, gi|15805888|; *EcoMBP*, gi|2781044|. The mature regions corresponding to the structures compared here were calculated with the aid of SignalP, and then analyzed using ProtParam.
Acy- and Heat Stability of a Maltose-Binding Protein

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, AcyMBP 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 (Kd for maltose is 1.5 μM). 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 amylopullulanase of A. acidocaldarius. The aromatic component in E. coli instead by residues from a different part of 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 AcyMBP’s site 1 (Tyr341)’s site 1 of 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 AcyMBP structure. It can be provided instead by residues from a different part of the sequence: Trp65 in PfuMBP and Trp62 in EcoMBP. The aromatic component in E. coli’s site (Tyr341) is missing in AcyMBP and PfuMBP, 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 TliMBP is substantially different from that of the other three MBPs, and resemblance is only in the approximate position of the bound ligand. TliMBP has two sites in common

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AcyMBP</td>
<td>159</td>
<td>12</td>
<td>68</td>
<td>3</td>
<td>171</td>
<td>71</td>
<td>242 (64.7)</td>
<td>374</td>
<td>11 (44.0)</td>
<td>14 (56.0)</td>
</tr>
<tr>
<td>EcoMBP</td>
<td>(42.5)</td>
<td>(3.2)</td>
<td>(18.2)</td>
<td>(0.8)</td>
<td>(45.7)</td>
<td>(19.0)</td>
<td>240 (64.9)</td>
<td>370</td>
<td>9 (42.9)</td>
<td>12 (57.1)</td>
</tr>
<tr>
<td>PfuMBP</td>
<td>157</td>
<td>6</td>
<td>72</td>
<td>5</td>
<td>163</td>
<td>77</td>
<td>248 (66.0)</td>
<td>381</td>
<td>9 (32.1)</td>
<td>19 (67.9)</td>
</tr>
<tr>
<td>TliMBP</td>
<td>182</td>
<td>17</td>
<td>65</td>
<td>8</td>
<td>173</td>
<td>74</td>
<td>268 (65.5)</td>
<td>407</td>
<td>11 (42.3)</td>
<td>15 (57.7)</td>
</tr>
</tbody>
</table>

Minimum volume for a cavity was set to 5 Å³. Standard deviations for the multiple observations of the AcyMBP structure are shown in parentheses.

Table 5. Comparison of cavities

<table>
<thead>
<tr>
<th>Structure</th>
<th>No. cavities</th>
<th>Total volume (Å³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcyMBP-1</td>
<td>5 (0.7)</td>
<td>104.8 (26.6)</td>
</tr>
<tr>
<td>EcoMBP</td>
<td>11</td>
<td>511.3</td>
</tr>
<tr>
<td>PfuMBP</td>
<td>4</td>
<td>331.8</td>
</tr>
<tr>
<td>TliMBP</td>
<td>4</td>
<td>124.4</td>
</tr>
</tbody>
</table>

Minimum volume for a cavity was set to 5 Å³. Standard deviations for the multiple observations of the AcyMBP structure are shown in parentheses.

The aromatic component in E. coli’s site (Tyr341) is missing in AcyMBP and PfuMBP, 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 TliMBP is substantially different from that of the other three MBPs, and resemblance is only in the approximate position of the bound ligand. TliMBP has two sites in common.
with AcyMBP and PfuMBP (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, TliMBP 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 TliMBP that can presumably take on the role of disaccharide binding in that organism.

Since acidophilic organisms can regulate their internal \( \text{pH} \), \( 25,26 \) 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. AcyMBP 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.\(^{27-28}\) 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 \( \alpha \)-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 \( \text{pH} \).\(^{29}\) Our analysis of AcyMBP 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 AcyMBP (Table 4), most acidic residues are buried within the protein. At the \( \text{pH} \) optimal for the organism (\( \sim 3.5 \)), 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 \( A.\ kawachii \) in which many acidic residues are exposed (although even that protein is expected to be positively charged at the \( \text{pH} \) optimum of 2).\(^{31}\) The preponderance of basic residues would seem to have the advantage that their high \( \text{pK}_a \) values are outside the \( \text{pH} \) range for optimum activity (\( \text{pH} 2.5-7 \)), which would effectively protect the protein from destabilizing changes in surface charge if the \( \text{pH} \) in the environment varies. This is in fact the case with more “normal” proteins: the \( \text{pK}_a \) values of the vast majority of their surface groups (3–5 for acidic residues and 9–11 for basic residues) are outside the usual \( \text{pH} \) range of their environment. The only exception is histidine with a \( \text{pK}_a \) 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.\(^{32,33}\) 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 AcyMBP 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 AcyMBP 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 AcyMBP (6.7% compared to \( \sim 5.7% \) 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 \( \beta \)-turns is a recurring feature in the (hyper)thermophilic proteins. AcyMBP 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). AcyMBP 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 AcyMBP was prepared from the culture supernatant of maltose-grown cells of *A. acidocaldarius* as described. Isoelectric focusing experiments were carried out using Amphoteric PAG plate 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, 10 mM maltose (the normal storage buffer).

A full-length version of AcyMBP 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 MRGSFLGSG fused to the N-terminal alanine residue. Cells were grown in Luria-Bertani broth at 30 °C, and 20 mM sodium phosphate and 10 mM maltose (the protein concentration for all preparations was 20 mg/ml in 50 mM Tris–HCl (pH 7.2), 5% glycerol, 2 mM DTT, 0.1 mM PMSF, 10 mM maltose, and 0.5 mM PEG 8000 and 100 mM Ches buffer (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 (AcyMBP-3) grew from hanging drops within 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-IAB, Lund, Sweden and DI4-4, Grenoble, France. Data were processed using DENZO and SCALEPACK and merged using programs of the CCP4 package (Table 1).

**Structure solution and refinement**

The structure of AcyMBP-1 was solved using molecular replacement with the program package CNS, 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). 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, resulted in an Rfree of 19.4% and a Rsym of 21.7% for the final model (summarized in Table 1).

The structures AcyMBP-2 and 3 were solved by molecular replacement using the final structure of AcyMBP-1, with all protein atoms as a search model. Two molecules were located in the asymmetric unit of both datasets using AMoRe (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 as implemented in CCP4). Further cycles of restrained refinement using all data to the maximum observed resolution were alternated with manual rebuilding. Water molecules were added using CNS. Statistics relating to the final refined models are summarized in Table 1.

**Crystallization and data collection**

The crystals for the AcyMBP-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% (v/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. Data statistics are summarized in Table 1.

Crystals of the intact mature protein in a larger unit cell (AcyMBP-2) were obtained using the hanging-drop vapour diffusion method: 2 μl protein solution (15 mg/ml 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 (AcyMBP-3) grew from hanging drops within 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-IAB, Lund, Sweden and DI4-4, Grenoble, France. Data were processed using DENZO and SCALEPACK and merged using programs of the CCP4 package (Table 1).

**Structure solution and refinement**

The structure of AcyMBP-1 was solved using molecular replacement with the program package CNS, 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). 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, resulted in an Rfree of 19.4% and a Rsym of 21.7% for the final model (summarized in Table 1).

The structures AcyMBP-2 and 3 were solved by molecular replacement using the final structure of AcyMBP-1, with all protein atoms as a search model. Two molecules were located in the asymmetric unit of both datasets using AMoRe (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 as implemented in CCP4). Further cycles of restrained refinement using all data to the maximum observed resolution were alternated with manual rebuilding. Water molecules were added using CNS. 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 as follows: 1ANF (*Eco*MBP with maltose), 3MBP (*Eco*MBP with maltotriose), 4MBP (*Eco*MBP with maltotetraose), 1EIJ (*Pfu*MBP)1 and 1EUS (*Tih*MBP). Structures were compared using O, Swiss-PdbViewer, LSQMAN, VOIDOO and FIT. Secondary structural content was assessed using DSSP. Surface accessibility was assessed using a per-residue cut-off of 30 Å² in Swiss-PdbViewer (accessibility calculated using a 1.5 Å probe). Structure-based and other sequence alignments were carried out using LSQMAN and Indonesia (Madsen, Kleywegt and Johansson, Uppsala...
University). Similar sequences were located using BLAST.\(^9\) Sequences were obtained from GenBank.\(^8\) Signal sequence cleavage sites were predicted using SignalP 4.1.\(^6\) Calculation of isoelectric points and amino acid compositions was performed using the server in the web\(^7\). Figures were prepared with the programs O, Molray,\(^5\) Molscript,\(^5\) LIGPLOT\(^5\) 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.\(^4\)

---

**Acknowledgements**

This work was supported by the Deutsche Forschungsgemeinschaft (SCHN 274/6-3-6-4), the Fonds der Chemischen Industrie (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**


*Edited by R. Huber*

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