the FEBS Journal

Crystal structure of the *Bc*ZBP, a zinc-binding protein from *Bacillus cereus*

Functional insights from structural data

Vasiliki E. Fadouloglou[1], Alexandra Deli[1], Nicholas M. Glykos[3], Emmanuel Psylinakis[1], Vassilis Bouriotis[1,2] and Michael Kokkinidis[1,2]

1 University of Crete, Department of Biology, Heraklion, Crete, Greece
2 Institute of Molecular Biology and Biotechnology, Heraklion, Crete, Greece
3 Democritus University of Thrace, Department of Molecular Biology and Genetics, Alexandroupolis, Greece

Keywords
Bacillus cereus; deacetylase; hydrolase; Rossmann fold; zinc-dependent enzyme

Correspondence
M. Kokkinidis, Institute of Molecular Biology and Biotechnology, PO Box 1527, Heraklion, Crete, Greece
Fax: +30 2810 394351
Tel: +30 2810 394351
E-mail: kokkinid@imbb.forth.gr

(Received 20 January 2007, revised 15 April 2007, accepted 17 April 2007)

doi:10.1111/j.1742-4658.2007.05834.x

Bacillus cereus is an opportunistic pathogenic bacterium closely related to *Bacillus anthracis*, the causative agent of anthrax in mammals. A significant portion of the *B. cereus* chromosomal genes are common to *B. anthracis*, including genes which in *B. anthracis* code for putative virulence and surface proteins. *B. cereus* thus provides a convenient model organism for studying proteins potentially associated with the pathogenicity of the highly infectious *B. anthracis*. The zinc-binding protein of *B. cereus*, *Bc*ZBP, is encoded from the *bc1534* gene which has three homologues to *B. anthracis*. The protein exhibits deacetylase activity with the *N*-acetyl moiety of the *N*-acetylglucosamine and the diacetylchitobiose and triacetylchitotriose. However, neither the specific substrate of the *Bc*ZBP nor the biochemical pathway have been conclusively identified. Here, we present the crystal structure of *Bc*ZBP at 1.8 Å resolution. The N-terminal part of the 234 amino acid protein adopts a Rossmann fold whereas the C-terminal part consists of two β-strands and two α-helices. In the crystal, the protein forms a compact hexamer, in agreement with solution data. A zinc binding site and a potential active site have been identified in each monomer. These sites have extensive similarities to those found in two known zinc-dependent hydrolases with deacetylase activity, MshB and LpxC, despite a low degree of amino acid sequence identity. The functional implications and a possible catalytic mechanism are discussed.

Bacillus cereus, an opportunistic pathogen that causes food poisoning and *Bacillus antracis*, the endospore-forming bacterium that causes inhalational anthrax, share a large number of homologous genes, as demonstrated by the recent genome sequencing and comparative analysis [1,2]. Given the laboratory safety precautions necessary for working with highly infectious agents and the recent concerns related to *B. anthracis* as a potential bioweapon, *B. cereus* offers an attractive alternative for studying the corresponding proteins of *B. anthracis* because it lacks infectiousness of the latter. The objective of the present study is to shed light on the structure, function and the structure–function relationships of one *B. cereus* protein, a product of the *bc1534* gene, which is highly conserved among the two pathogens and which has, as we show, acetylchitooligosaccharide deacetylase activity. Thus, our work contributes to the understanding of the role

Abbreviations
*Bc*ZBP, *Bacillus cereus* zinc-binding protein; GAB, general-acid-base; GlcNAc, *N*-acetylglucosamine; TLS, translation/libration/screw.

of bacterial *N*-acetylchitooligosaccharide deacetylases and, furthermore, to a better understanding of the properties of *B. anthracis* [3].

The *bc1534* gene of *B. cereus* (ATCC 14579) codes for a soluble polypeptide chain of 234 amino acids (UniProt accession number Q81FP2) and has three homologues in *B. anthracis* (str. A2012) (i.e. *bant_01002171*, *bant_01004539* and *bant_01004184*) [4]. All of them code for uncharacterized proteins that share sequence identities of 96%, 28% and 24%, respectively, with the protein encoded by bc1534 [4,5]. Furthermore, *bc1534* also has a homologue in the *B. cereus* genome, the *bc3461* gene, with 25% identity, at the amino acid level.

The protein encoded by bc1534 is classified as a LmbE-related protein and hereafter will be referred to as *Bc*ZBP (*B. cereus* zinc-binding protein). The N-terminal part of *Bc*ZBP (residues 7–124) belongs to the Pfam02585 family, which comprises deacetylases of the *N*-acetylglucosaminyl-phosphatidylinositol and the 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranoside [6]. Independent of their specific substrates, all members of this family deacetylate the N-acetyl group of the *N*-acetylglucosamine moiety. Only two proteins of the family have been structurally characterized to date: TT1542 from *Thermus thermophilus* [7] and MshB from *Mycobacterium tuberculosis* [8,9], whose structures have been determined at 2.0 and 1.7 Å resolution, respectively. These two proteins share 33% sequence identity and their N-termini are structurally very similar adopting a Rossmann fold motif. Although the biological role and the natural substrate of TT1542 remain unknown, MshB has been characterized as a deacetylase involved in the biosynthetic pathway of mycothiol [10]; its catalytic residues are similar to those found in metalloproteases so that a catalytic mechanism similar to that of metalloproteases [8,11,12] has been proposed for MshB. A further well characterized member of the Pfam02585, the Tk-Dac protein from the archaeon *Thermococcus kodakaraensis* KOD1 has been shown to exhibit diacetylchitobiose deacetylase activity and has been proposed to be involved in a novel, probably common in archea, chitin catabolic pathway [13,14].

Although the functional characterization of the *Bc*ZBP protein is still in progress, preliminary biochemical results which will be presented here, have shown that the enzyme exhibits deacetylase activity on *N*-acetylchitooligosaccharide substrates and that activity depends on the chain length of the substrate. However, both the specific substrate of the enzyme and the biochemical pathway in which *Bc*ZBP is involved remain to be identified.

The crystal structure determination of the *Bc*ZBP protein at a resolution of 1.8 Å, provides important clues towards understanding the enzyme function, including the identification of a zinc-binding site within a potential active site that is similar to active sites of known zinc-dependent deacetylases. Our analysis provides evidence both for the type of the reaction catalyzed and for the catalytic mechanism. Finally, we present possible functional implications for *Bc*ZBP deduced from a structural comparison with sequence homologues.

Results and Discussion

Overview of the structure

As shown in supplementary Fig. S1, the polypeptide chain of *Bc*ZBP folds into a single, compact α/β domain. The overall structure can be divided into two distinct structural motifs shown with different shades of gray in the topology diagram of Fig. 1A. The N-terminal part of the protein (residues 2–149) adopts a Rossmann fold motif. This motif is built-up of five parallel β-strands (β1–β5) forming an open, twisted β-sheet which is surrounded by four α-helices, two on each side of the sheet (α1–α4). A short loop (residues 150–155) links the Rossmann fold motif to the C-terminal part (residues 156–233). The C-terminal part folds into a structure consisting of two hydrogen-bonded antiparallel β-strands (β6–β7), an α-helical hairpin (helices α5, α6) and a C-terminal strand (β8). Helix α5 and strand β7 partially cover one side of the Rossmann fold structure, thereby interacting with helices α1, α2 and strand β5, respectively. The α5-helix has a characteristic, hook-like shape due to a kink which occurs at Tyr175. This kink orients the C-terminal turn of the helix nearly perpendicularly to the rest of the helix. This turn of α5 is connected with α6-helix via a long loop (residues 180–193) which, on the basis of the *B*-values distribution (supplementary Fig. S2C), represents the most flexible part of the structure. Although five of its residues could not be located to the TT1542 model, the electron density map of *Bc*ZBP was interpretable for all residues of this loop in both chains of the asymmetric unit.

Two *Bc*ZBP monomers associate via a local two-fold axis and form the dimer shown in Fig. 1B, which is stabilized by hydrogen bonds and hydrophobic interactions. As the topology diagram of Fig. 1A shows schematically, strand β8, from one monomer, is incorporated into the C-terminus of the other monomer and it is hydrogen bonded to the strand β6. Dimerization is thus mainly established via formation of two,

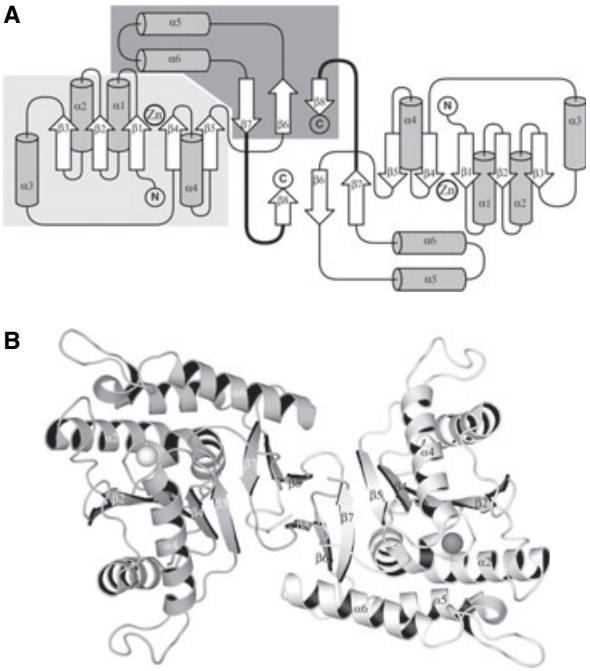

Fig. 1. Structure of the *Bc*ZBP dimer. (A) Topology diagram of the dimer drawn with the program TOPDRAW [34]. The relative orientation of the secondary structure elements is illustrated. The shaded areas highlight a single monomer. Light gray indicates the N-terminal part that folds into a Rossmann motif and dark gray indicates the C-terminal part. The dimer's formation is established by the incorporation of the β8-strand of one monomer into a β-sheet of the other monomer. The position of the zinc ion is indicated by a circle. (B) Schematic diagram of the dimer. Each monomer is shown with a different shade of gray. Zinc ions are presented as spheres. The view is along the local two-fold axis.

mixed, three-stranded β-sheets, each one consisting of the antiparallel β-strands β6, β7 of one monomer and strand-β8 of the other. Solvent-accessible surface [15] calculations show that a substantial area, 1677 $Å^2$ per chain, is buried upon dimer formation.

In the crystal, *Bc*ZBP is a hexamer formed by three dimers that are related through a crystallographic three-fold axis. The resulting trimer of dimers (Fig. 2) is a nearly spherical, compact homohexamer with three monomers in the upper and three monomers in the lower hemisphere; the former being pairwise related the latter by three local two-fold axes lying perpendicularly to the three-fold axis. The maximum dimension of the hexamer along its symmetry axes is approximately 70 Å. A hydrophilic channel with diameter of approximately 20 Å crosses the centre of the hexamer along its three-fold axis from one end to the other. Side chains of mostly Tyr, Lys and Glu residues protrude to the channel which is filled with water molecules. The buried surface area upon hexamer formation is 3894 $Å^2$ per monomer with a corresponding energy gain of approximately 120 kcal/mol [16] as was calculated by the Protein Quaternary Structure server (http://pqs.ebi.ac.uk/). These values indicate a significant stabilization upon hexamerization. In agreement with the crystal structure, gel filtration experiments have shown that the enzyme elutes as a single peak to a volume that is consistent with a spherical hexamer [17]. These results for *Bc*ZBP are consistent with an ultracentrifugation analysis of TT1542 [7], which also indicate a hexameric assembly. It can be thus reasonably assumed that the quartenary structure of the *Bc*ZBP in solution is the hexamer found in the crystals and that this hexamer probably corresponds to the biologically active form of the protein.

In terms of thermal mobility, the *Bc*ZBP hexamer is segregated into two clearly distinguishable halves. This is manifested by the presence of systematic differences between the crystallographic temperature factors of the

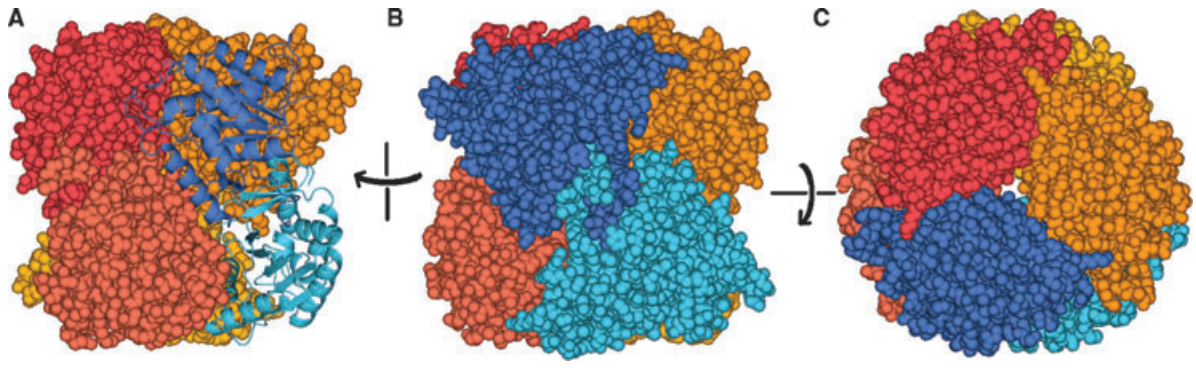

Fig. 2. Assembly of the *Bc*ZBP hexamer. (A, B) Side and (C) top views of the *Bc*ZBP hexamer formed through the association of three dimers. Each dimer has a different color. Different tones of the same color are used to distinguish between monomers of a dimer. (A) Illustrating the orientation of the dimer in the hexamer by representing one of them with a schematic diagram.

two chains as shown in supplementary Fig. S2C. Generally, chain A has higher *B*-values compared to chain B. Thus, into the same hexamer, the one trimer (shown in the supplementary Fig. S2A) is less mobile than the other (supplementary Fig. S2B). Similar differences in mobility are also observed between the chains of the TT1542 dimer.

*Bc*ZBP binds zinc ions through a conserved triad

Almost all α/β structures with a Rossmann fold motif have their active sites at the carboxy edge of the β-sheet [18], within a crevice which is formed between two adjacent loop regions that connect two strands with α-helices on opposite sides of the β-sheet. From the topology diagram of *Bc*ZBP (Fig. 1A), an active site can be predicted in the crevice adjacent to the C-termini of strands β1 and β4. In this crevice of each monomer, a prominent electron density peak (at 13 σ in a 2*Fo–Fc* map) was found. Interestingly, there is no atom in the TT1542 structure corresponding to the position of this peak. X-ray fluorescence analysis of *Bc*ZBP protein crystals, revealed a maximum at the K-edge of zinc (approximately at 9.668 keV; supplementary Fig. S3A). In addition, anomalous difference maps using data collected at the K-edge of zinc (wavelength of 1.282 Å), unambiguously confirmed that this peak corresponds to zinc (supplementary Fig. S3B). As zinc compounds were not used in the purification or crystallization protocols, we conclude that the metal must be intrinsically contained in the protein.

The structure of the active site is illustrated in Fig. 3. The zinc ion is tetrahedrally coordinated by three protein residues and one molecule that has been interpreted, in the electron density map, as acetate (supplementary Fig. S4). A water molecule is also found in the active site, 4.4 Å from the zinc ion and within hydrogen bonding distance from the acetate oxygen atom, which coordinates the metal (supplementary Fig. S4). The close proximity to the zinc ion is highly suggestive of a catalytic water molecule which has been replaced by the acetate moiety. The protein coordinates zinc with the N^{δ} atom of His12, one of the O^{δ} atoms of Asp15 and the N^{ε} atom of His113. His113 protrudes from helix α4 whereas His12 and Asp15 both belong to the loop which joins the β1-strand with the α1-helix and approach the metal from opposite directions. The metal binding residues are all, strictly conserved among the *Bc*ZBP homolgues (data not shown). The zinc-binding motif is of the type **H**XDD(X)$_{98}$**H** (residues in bold are zinc ligands; X is used to represent any residue). Such a motif, with the first two zinc ligands being separated by a short segment of 1–3 residues and the last two ligands being separated by a segment of variable length and with no particular amino acid preferences is frequently found in zinc-hydrolases with deacetylase activity [12,19]. The fourth zinc ligand, the acetate molecule, is found in equivalent positions of the active sites in the protein dimer. This molecule coordinates the metal with its one oxygen atom (Act O^{1}) whereas the other oxygen atom is located within hydrogen bonding distance from the $O^{\delta 2}$ of Asp14 and approximately 2.6 Å from the zinc. Acetates probably originate from the crystallization solution which contains 100 mM CH_3COOH/CH_3COONa as buffer. This relatively high concentration justifies the presence of acetate in the crystal structure. The binding of acetate in the active site is a further indication that the enzyme may be involved in deacetylation because acetate is one of the reaction products.

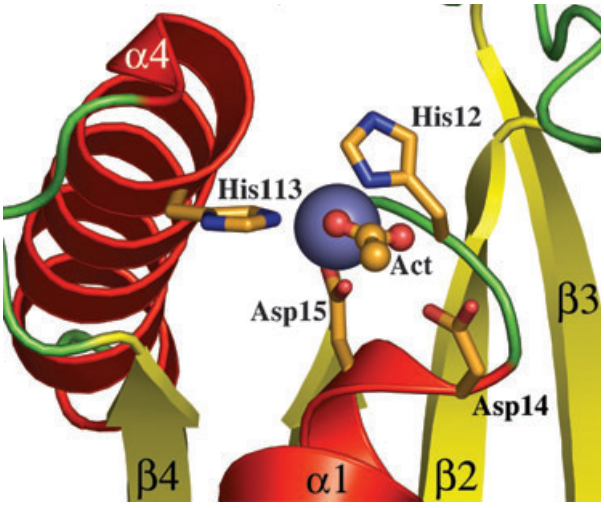

Fig. 3. Structure of the active site. Three protein residues (His12, Asp15, His113) and one acetate molecule (Act) coordinate the zinc ion, which is represented as a blue sphere. The location of Asp14, which adopts an uncommon backbone conformation, is also shown.

Asp14, a residue of the loop joining the β1-strand to the α1-helix, is located close to the metal ion (Fig. 3); its $O^{\delta 2}$ atom is positioned in a distance of 4.2 Å from zinc. The backbone conformation of this residue (ϕ/ψ angles) falls within a 'disallowed' region of the Ramachandran plot. This unusual conformation is the prerequisite for the close proximity of the Asp14 side chain to the potential active site, the zinc ion and the acetate, and suggests a possible role in the enzymatic reaction which will be discussed later.

The structure of the active site, the type of protein ligands, the zinc-binding motif, the presence of a water

Table 1. Deacetylase activity of the *Bc*ZBP protein on *N*-acetylchitooligosaccharides.

Substrate	Deacetylation (%)
GlcNAc	100
$(GlcNAc)_2$	89.82
$(GlcNAc)_3$	10.51
$(GlcNAc)_4$	5.51
$(GlcNAc)_5$	12.19
$(GlcNAc)_6$	3.93

molecule in the active site and the binding of acetate, strongly suggest that *Bc*ZBP acts as a zinc-dependent deacetylase. This is in agreement with our preliminary functional data, which show that the protein does exhibits deacetylase activity. The activity of the enzyme in deacetylating *N*-acetylchitooligosaccharide substrates was tested with several *N*-acetylchitooligomers and the results are summarized in Table 1. There is a clear preference of the enzyme for the two shortest oligomers, i.e. *N*-acetylglucosamine (GlcNAc) and diacetylchitobiose [$(GlcNAc)_2$]. Thus, we suggest that *Bc*ZBP belongs to the class of zinc-dependent hydrolases with deacetylase activity.

Hexamerization may affect substrate selectivity and specificity

The zinc ion is buried at the bottom of a cavity which is located at the surface of the hexamer. Figure 4 illustrates that the complete active site cavity is formed at the level of the hexamer by two subunits related by the three-fold axis. The main body of each active site, a funnel-like cavity, with a depth of approximately 10 Å and a wide, almost circular opening with diameter of 15 Å, is formed at the level of an individual subunit and it is independent on the association in hexamers (Fig. 4A). Upon hexamer formation, a second subunit associates with the first one and extends the active site into a cavity with a depth of 12 Å and its diameter varies from 8 Å at the bottom to 12 Å near the edge (Fig. 4B). Consequently, the oligomerization of the enzyme ultimately determines the final amino acid composition, shape and size of the active site and may thus influence substrate selectivity and specificity. As shown in Fig. 4, the rim of the complete active site is shaped by both subunits. The one subunit, which carries the main body of the cavity contributes two loops which join strand β2 to helix α2 and helix α5 to helix α6, respectively, whereas the adjacent subunit frames the other side with Arg140 being in a very prominent position at the entry of the cavity. Arg140 adopts two different conformations which block or keep open the

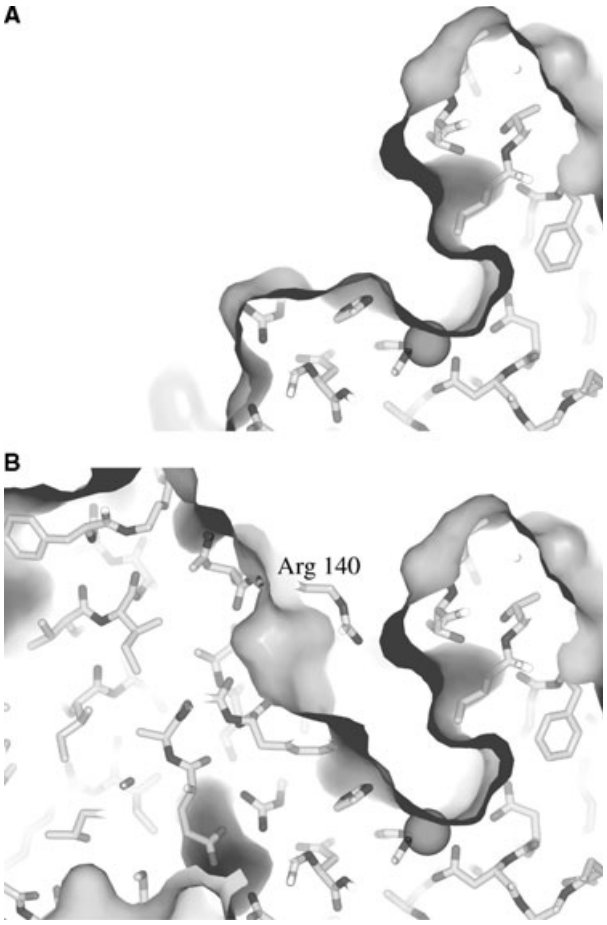

Fig. 4. Oligomerization and active site formation. Sections (6 Å thick) of the protein surface that illustrate the shape and size of the active site. The sphere represents the zinc ion. (A) The main body of the active site cavity is formed inside a single monomer. (B) Upon hexamer formation, a second monomer is packed against the first one, resulting in a longer active site cavity and creating additional constraints (e.g. through Arg140) to active site accessibility. The residue Arg140, which is not used in the calculation of the protein surface, is shown by a stick model.

entry of the active site (Fig. 5) and could play a key role in the interaction of the enzyme with its substrate(s). In analogy to other cases [11,20], the oligomerization of *Bc*ZBP could thus be important for substrate selectivity and specificity by determining the geometry and accessibility of the active site.

Structural comparison of *Bc*ZBP with related proteins

*Bc*ZBP shares significant sequence similarities with the two proteins of known structure from the Pfam02585 family (Fig. 6), namely TT1542 (1UAN.pdb) from *Thermus thermophilus* [7] and MshB (1Q74.pdb) from

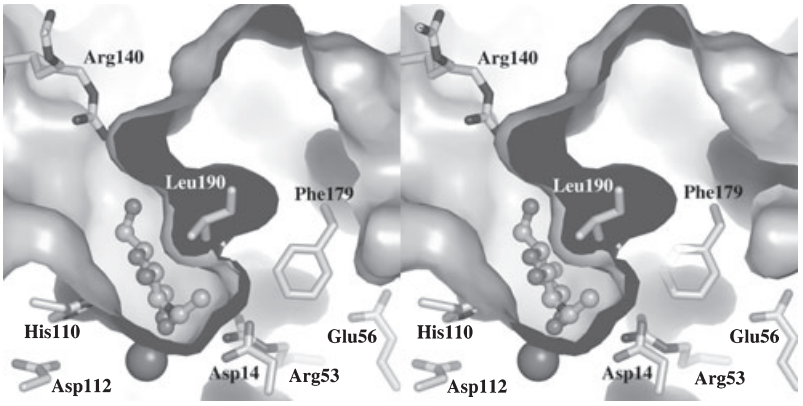

Fig. 5. Model of the *Bc*ZBP–GlcNAc complex. Stereoview of the energy minimized putative *Bc*ZBP–GlcNAc complex. A slice through the active site cavity shows the quality of fit of the *N*-acetylglucosamine molecule (ball-and-stick model) into the bottom of the active site. Catalytically important residues are shown as stick models, the zinc ion as a sphere. The two conformations of Arg140, which is not used in the protein surface calculation, are also shown as stick models.

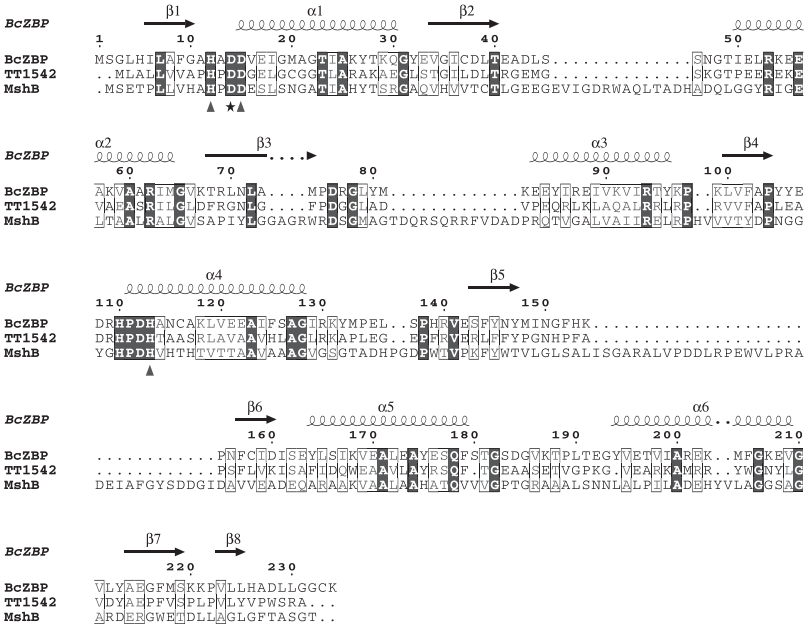

Fig. 6. Sequence alignment. Amino acid sequence comparison of the *Bc*ZBP, TT1542 (38% identity) and MshB (25% identity) proteins. The numbering scheme and the secondary structure elements correspond to the *Bc*ZBP. Alignment was performed with CLUSTALW [27] and plotted with the ESPRIPT program [21]. Strictly conserved residues are highlighted and similar residues are boxed.

Mycobacterium tuberculosis [8,9]. A search with the Dali server [22] (http://www.ebi.ac.uk/dali/) confirms that these proteins are the closest structural relatives of *Bc*ZBP with a *Z*-score of 33.7 and 21.3, respectively.

At the monomer level, *Bc*ZBP and TT1542 superimpose with an rmsd of 1.3 Å for the Ca atoms of 212 residues. The structural comparison is informative in terms of possible functional properties. The most flexible region in both structures is the 14-residue loop joining helices α5 and α6 (residues 180–193 in *Bc*ZBP, supplementary Fig. S2). This loop is positioned next to the active site and varies at sequence level significantly between the two proteins. These features could reflect a role of the loop in the active site (e.g. in substrate recognition).

The α2 helices in the *Bc*ZBP and TT1542 structures are rotated relative to each other around their C-termini; in *Bc*ZBP the N-terminus of the α2 helix is positioned due to this rotation approximately 4 Å closer to the active site compared to the TT1542 helix. Similarly, the preceding loop (residues 40–49) is also shifted by 4 Å relative to the TT1542 loop towards the top of the active site (Fig. 7A); these changes result in a more closely packed environment of the active site compared to TT1542. Superposition of the two structures excluding the 13 shifted residues corresponding to the N-terminus of the α2-helix and to the preceding loop results in a rmsd of 1.0 Å for the Ca atoms. Thus, the movement of the α2 helix accounts for 23% of the rmsd value (i.e. for approximately one quarter of the structural difference between the enzymes). These localized differences in the immediate environment of the predicted active sites of two, otherwise very similar structures could reflect two different enzyme states,

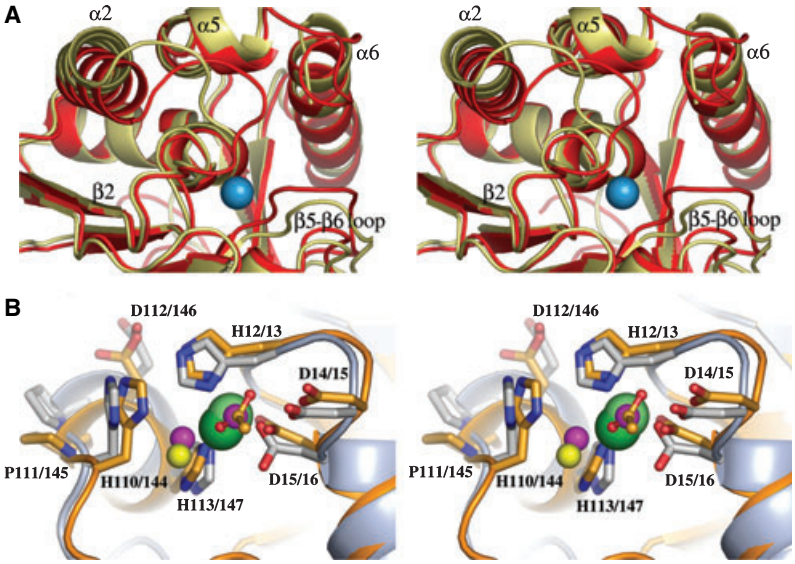

Fig. 7. Structural comparison of *Bc*ZBP with related proteins. (A) Superposition of *Bc*ZBP (red) to TT1542 (yellow). The stereoview illustrates the rotation and shift of the α2-helix and its preceding loop in the *Bc*ZBP structure relative to their counterparts in TT1542. The carboxy-termini are well superimposed whereas the aminotermini are approximately 4 Å apart. The blue sphere represents the zinc ion. (B) Superposition of *Bc*ZBP (orange) to MshB (gray). The stereoview focuses on the active sites and illustrates that they are essentially identical. Residue types are given as the one-letter code. The first number corresponds to *Bc*ZBP and the second to MshB. Zinc ions are presented as large, green spheres. The magenta balls correspond to the two water molecules found into the active site of MshB. The yellow ball represents the active site water molecule of *Bc*ZBP. The acetate molecule of the *Bc*ZBP is represented by a ball-and-stick model.

with TT1542 corresponding to a nonfunctional, zinc-absent structure and *Bc*ZBP to a state following the catalytic reaction, in which the substrate has been processed and removed when the acetate is still bound in the active site. Thus, the conformational switch of the α2-helix could be of functional relevance and be associated with either an 'open' nonfunctional conformation or a 'closed' conformation adopted by the activated enzyme.

MshB, a zinc-dependent enzyme from the biosynthetic pathway of mycothiol, deacetylates 1-D-*myo*-inosityl-2-acetamido-2-deoxy-α-D-glucopyranose (GlcNAc-Ins) [10]. The MshB structure, similarly to the *Bc*ZBP monomer, displays a Rossmann fold motif in its N-terminal part (residues 1–184); to the C-terminal parts, the two proteins are structurally unrelated. Although MshB has long loop regions, the Rossmann motifs of MshB and *Bc*ZBP are superimposable with an rmsd of 1.6 Å (for 127 Ca atoms, excluding loops). Consequently, illustrated in Fig. 7B, the active sites of *Bc*ZBP and MshB are essentially identical, with the same residues coordinating a zinc ion. These residues plus an additional conserved motif (Fig. 7B) in the immediate neighborhood of the active site (His110, Pro111, Asp112, His113 in the *Bc*ZBP numbering) adopts the same structural arrangement in both proteins, which is a strong indication of a common functional/structural role. The Rossmann fold motif in both proteins provides the basis for the correct spatial arrangement of catalytically important residues to generate a functional active site. On the other hand, the low degree of conservation in the loop regions near the active site could be associated with differences in the substrates used by the enzymes.

The C-terminal regions of *Bc*ZBP and MshB (i.e. the regions that follow the Rossmann motif) share little structural similarity, with the exception of one β-strand and one α-helix of MshB which are well superimposable to the β6-strand and α5-helix of *Bc*ZBP, respectively. As the intertwining of C-termini is a key feature for the oligomerization of *Bc*ZBP, the differences in C-terminal regions between *Bc*ZBP and MshB could account for the absence of oligomerization in MshB [8,9]. As noted above, the final size and shape of the active site pocket in *Bc*ZBP is established at the level of the hexamer; thus, the quaternary structure differences between the two enzymes may give rise to considerable differences in their interactions with their substrates.

Insights into the probable catalytic mechanism

The predicted active site of *Bc*ZBP is strikingly similar to the active sites of two well characterized

zinc-dependent hydrolases with deacetylase activity. One of them is the MshB mentioned above. The other is LpxC [19], an enzyme which deacetylates UDP-3-*O*-myristoyl-*N*-acetylglucosamine and has no sequence or overall structural similarity to *Bc*ZBP. In general, the reaction mechanisms which are catalyzed by zinc-dependent deacetylases include a nucleophilic attack carried out by a zinc-bound water molecule and a general-acid-base (GAB) catalysis provided by enzyme residues. Two types of GAB catalysis have been identified to date [12] which are based either on a single, bifunctional GAB catalyst or on a GAB catalysts pair. The available biochemical data on MshB and LpxC are not sufficient to unambiguously identify the specific mechanism used by each enzyme, although a GAB pair catalysis agrees better with mutagenesis data for LpxC [12] whereas a single, bifunctional GAB catalysis, similar to the mechanism used by metalloproteases, has been proposed for MshB [8,12].

*Bc*ZBP shares the following common features with the active sites of MshB and LpxC: (a) The enzymes provide identical ligands to the zinc ion (i.e. two His and one Asp residues). (b) A water molecule is found into the active sites, coordinating the zinc ion. (c) A His/Asp pair (His110/Asp112 for *Bc*ZBP, His144/Asp146 for MshB and His265/Asp246 for LpxC) is found close to the active site. It has been proposed that this His/Asp pair could serve as a charge relay during the catalysis. (iv) In close proximity to the active site a carboxylate residue also exists, Glu in the case of LpxC (Glu78), Asp in the cases of MshB and *Bc*ZBP (Asp14 for *Bc*ZBP and Asp15 for MshB). It is believed that this residue could act as a general base catalyst activating the zinc-bound water for nucleophilic attack. Interestingly, in the crystal structures of independently determined related proteins, the ϕ/ψ-values of this Asp residue systematically fell outside the allowed regions of a Ramachandran plot. Asp14 is the only *Bc*ZBP residue that adopts an energetically unfavorable main chain conformation through which a close approach of the side chain to the active site is achieved. Asp15, the equivalent residue in the zinc-bound MshB structure (PDB code 1Q74) adopts a similar strained conformation. On the other hand, in the absence of a zinc ion in the active site, such as in the structures of the zinc-free MshB (PDB code 1Q7T) and TT1542 (PDB code 1UAN), the equivalent Asp residues (Asp15 and Asp12, respectively), adopt a main chain conformation that deviates less from the standard values. It appears that the presence of a functional (i.e. zinc-containing) active site is associated with the extent of the backbone distortion of this residue, so that a certain catalytic role of this Asp appears likely.

Based on the above common features, we suggest that the catalytic mechanism of *Bc*ZBP is probably similar to those proposed for MshB and LpxC. This hypothesis was further explored by modeling the binding of substrate (GlcNAc) in the predicted active site of *Bc*ZBP (Fig. 5). The carboxylate group of GlcNAc was initially placed at the position occupied by the Act O^1 atom; Act and water molecules were removed from the active site. Energy minimization was performed by CNS [30], with the protein atoms fixed. As illustrated by Fig. 5, the model shows that a single *N*-acetylglucosamine moiety is considerably smaller than the active site cavity, however, it fits well in its bottom. The methyl group of the GlcNAc acetyl group was well fitted into a conserved hydrophobic cavity formed by the residues Ile18, Ile149, Leu172, Phe179 and the aromatic ring of Tyr194. The side chains of Tyr194, Asn150 and Asp108 form a hydrophilic patch close to the zinc ion and to the His110/Asp112 pair. This position, which is empty in the modeled complex and partially occupied by the active site water molecule in the *Bc*ZBP crystal structure, could play the role of the 'oxyanion hole' [12]. It has been proposed that this 'hole' accommodates the charged oxygen of the substrate in the intermediate state. In the modeled complex, the sugar is oriented in such a way that the nitrogen of the amide bond faces Asp14 and the Arg53/Glu56 pair and is positioned oppositely to the His110/Asp112 pair.

Conclusions

Our present understanding of the biological function of the *Bc*ZBP protein is very limited. The protein exhibits deacetylase activity with the GlcNAc moiety; however, its specific substrate has not yet conclusively identified. On the other hand, the crystal structure of the enzyme reveals some functional properties: (a) The enzyme is a zinc-binding protein. (b) The active site has all the typical features that are expected for a zinc-dependent hydrolase. In addition, it binds acetate which is the product of a deacetylation reaction. (c) The protein forms stable homohexamers both in the crystal form and in solution. Thus, the functional state of the enzyme is probably the hexamer. (d) Hexamer assembly could influence substrate selectivity and specificity because it introduces constraints to active site accessibility and determines the shape of the active site entry. (e) The structure of the active site is essentially identical with the active sites of the MshB and LpxC proteins. The conservation of catalytically important residues implies that *Bc*ZBP could utilize a catalytic mechanism similar, in its general features, to

the mechanisms proposed previously for MshB and LpxC.

Nevertheless, more biochemical, enzymatic and mutagenesis studies will be necessary to test these suggestions. Ongoing mutagenesis analysis focuses on 'key residues' identified by the structural work (e.g. Asp14 and Arg140) and on strand β8 which plays a role on oligomerization and thus probably affects enzyme activity and substrate specificity.

Experimental procedures

Structure determination and refinement

The expression, purification and crystallization of *Bc*ZBP have been reported previously [17]. High resolution diffraction data were collected from a single frozen crystal (100 K) using beamline X12 at the European Molecular Biology Laboratory/Deutsches Elektronen-Synchrotron (Hamburg, Germany). Data processing and scaling were performed with the programs MOSFLM [23] and SCALA [24,25]. Table 2 shows details of data collection, processing and crystallographic refinement. *Bc*ZBP crystallizes with a dimer in the asymmetric unit. The crystals belong to the space group *R*32 with unit cell parameters $a = b = 75.9$, $c = 404.7$ Å (in the hexagonal setting). The structure was determined by the method of molecular replacement using MOLREP [26]. The search model was based on the structure of the TT1542 protein (1UAN.pdb), which has a 38% sequence identity with *Bc*ZBP. After alignment of the *Bc*ZBP and TT1542 sequences with CLUSTALW [27], residues in the TT1542 structure were replaced by alanine, using XFIT from the XTALVIEW package [28], if in the particular position the two sequences were occupied by different amino acids. Molecular replacement using this model and data to a resolution of 3 Å provided a solution with an *R* of 53.0% and a linear correlation coefficient of 0.35. The electron density was calculated by the program GRAPHENT [29]. Crystallographic refinement was performed by the programs CNS [30] and REFMAC5 [31]. Initial cycles of rigid body refinement [31] were followed by several cycles of torsion angles and cartesian molecular dynamics [30]. Side chains and some loop regions were manually built using the program XFIT [28]. The refinement process was completed by positional and translation/libration/screw (TLS) refinement, where each chain of the asymmetric unit was parameterized as an individual TLS group [31]. The final model, with 3720 protein atoms and 471 water molecules, converged to an R/R_{free} of 17.7/20.7%. Residues 2–233 of chain A and 2–231 of chain B had interpretable electron density and were included in the final model. The two chains are almost identical and superimpose [25,32] with an rmsd of 0.242 Å for the Ca atoms of 230 residues.

The atomic coordinates and structure factors have been deposited in the Protein Data Bank [33] with accession code 2IXD.

Table 2. Data collection and refinement statistics. Values in parentheses refer to the outer resolution shell (1.90–1.80 Å).

Data	Value
Data collection and processing	
Wavelength (Å)	1.282
Space group	R32
Unit cell parameters (hexagonal setting)	$a = b = 75.9$, $c = 404.7$
Resolution (Å)	1.80
Number of unique reflections	40471 (4102)
Completeness (%)	92.3 (65.4)
Multiplicity	7.3 (6.3)
R_{sym} (%)	6.2 (50.5)
Mean (I)/σ(I)	18.3 (3.5)
Phasing (molecular replacement)	
Model used	1UAN.pdb (dimer)
Refinement and analysis of molecular model	
Resolution (Å)	55–1.80
R/R_{free} (%)	17.7/20.7 (23.7/26.6)
Atoms modeled (protein/water/act/Zn)	3720/471/8/2
rmsd for bond lengths (Å)	0.006
rmsd for angles (°)	1.429
Residues in the Ramachandran plot	
Most favored region (%)	92.1
Additional allowed regions (%)	7.4
Generously allowed regions (%)	–
Disallowed regions	Asp14 of both chains

Enzyme assays

Polysaccharide deacetylase activity assays were performed using *N*-acetylchitooligosaccharides [$(GlcNAc)_{1-6}$] as substrates. The assay mixture contained 25 mM Hepes-NaOH pH 8.0, 1 mM $CoCl_2$, and 450 nmol $GlcNAc_{1-6}$ incubated with 50–150 μg of enzyme. Activity was measured in a coupled assay, by determining the acetate released by the action of the enzyme on the *N*-acetylchitooligosaccharides using the enzymatic method of Bergmeyer via three coupled enzyme reactions [35].

Acknowledgements

Funding through the General Secretariat for Research and Development programs PYTHAGORAS and PEP-KRITIS is gratefully acknowledged. We thank the European Molecular Biology Laboratory, Hamburg Outstation and the European Union for support through the the EU-I3 access grant from the EU Research Infrastructure Action under the FP6 'Structuring the European Research Area Programme', contract number RII3/CT/2004/5060008.

References

1 Ivanova N, Sorokin A, Anderson I, Galleron N, Candelon B, Kapatral V, Bhattacharyya A, Reznik G, Mikhailova N, Lapidus A *et al.* (2003) Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **423**, 87–91.

2 Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, Tettelin H, Fouts DE, Eisen JA, Gill SR *et al.* (2003) The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **423**, 81–86.

3 Psylinakis E, Boneca IG, Mavromatis K, Deli A, Hayhurst E, Foster SJ, Varum KM & Bouriotis V (2005) Peptidoglycan N-acetylglucosamine deacetylases from *Bacillus cereus*, highly conserved proteins in *Bacillus anthracis*. *J Biol Chem* **280**, 30856–30863.

4 Schäffer AA, Aravind L, Madden TL, Shavirin S, Spouge JL, Wolf YI, Koonin EV & Altschul SF (2001) Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res* **29**, 2994–3005.

5 Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM *et al.* (2002) Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* **296**, 2028–2033.

6 Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M & Sonnhammer EL (2002) The Pfam protein families database. *Nucleic Acids Res* **30**, 276–280.

7 Handa N, Terada T, Kamewari Y, Hamana H, Tame JRH, Park S-Y, Kinoshita K, Ota M, Nakamura H, Kuramitsu S *et al.* (2003) Crystal structure of the conserved protein TT1542 from *Thermus thermophilus* HB8. *Protein Sci* **12**, 1621–1632.

8 Maynes JT, Garen C, Cherney MM, Newton G, Arad D, Av-Gay Y, Fahey RC & James MNG (2003) The crystal structure of 1-D-myo-inosityl 2-acetamido-2-deoxy-α-D-glucopyranoside deacetylase (MshB) from *Mycobacterium tuberculosis* reveals a zinc hydrolase with a lactate dehydrogenase fold. *J Biol Chem* **278**, 47166–47170.

9 McCarthy AA, Peterson NA, Knijff R & Baker EN (2004) Crystal structure of MshB from *Mycobacterium tuberculosis*, a deacetylase involved in mycothiol biosynthesis. *J Mol Biol* **335**, 1131–1141.

10 Newton GL, Av-Gay Y & Fahey RC (2000) N-acetyl-1-D-*myo*-inosityl-2-amino-2-deoxy-α-D-glucopyranoside deacetylase (MshB) is a key enzyme in mycothiol biosynthesis. *J Bacteriol* **182**, 6958–6963.

11 Lowther WT & Matthews BW (2002) Metalloaminopeptidases: Common functional themes in disparate structural surroundings. *Chem Rev* **102**, 4581–4607.

12 Hernick M & Fierke CA (2005) Zinc hydrolases: the mechanisms of zinc-dependent deacetylases. *Arch Biochem Biophys* **433**, 71–84.

13 Tanaka T, Fukui T, Atomi H & Imanaka T (2003) Characterization of an exo-β-D-glucosaminidase involved in a novel chitinolytic pathway from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Bacteriol* **185**, 5175–5180.

14 Tanaka T, Fukui T, Fujiwara S, Atomi H & Imanaka T (2004) Concerted action of diacetylchitobiose deacetylase and exo-β-D-glucosaminidase in a novel chitinolytic pathway in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J Biol Chem* **279**, 30021–30027.

15 Chothia C (1975) Structural invariants in protein folding. *Nature* **254**, 304–308.

16 Eisenberg D & McLachlan D (1986) Solvation energy in protein folding and binding. *Nature* **319**, 199–203.

17 Fadouloglou VE, Kotsifaki D, Gazi AD, Fellas G, Meramveliotaki C, Deli A, Psylinakis E, Bouriotis V & Kokkinidis M (2006) Purification, crystallization and preliminary characterization of a putative LmbE-like deacetylase from *Bacillus cereus*. *Acta Crystallogr F* **62**, 261–264.

18 Branden C-I (1980) Relation between structure and function of α/β proteins. *Q Rev Biophys* **13**, 317–338.

19 Whittington DA, Rusche KM, Shin H, Fierke CA & Christianson DW (2003) Crystal structure of LpxC, a zinc-dependent deacetylase essential for endotoxin biosynthesis. *Proc Natl Acad Sci USA* **100**, 8146–8150.

20 Burley SK, David PR, Taylor A & Lipscomb WN (1990) Molecular structure of leucine aminopeptidase at 2.7-Å resolution. *Proc Natl Acad Sci USA* **87**, 6878–6882.

21 Gouet P, Courcelle E, Stuart DI & Metoz F (1999) ESPript: analysis of multiple sequence alignments in postscript. *Bioinformatics* **15**, 305–308.

22 Holm L & Sander C (1998) Touring protein fold space with Dali/FSSP. *Nucleic Acids Res* **26**, 316–319.

23 Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Jnt CCP4/ESF-EACBM Newsl Protein Crystallogr* **26**.

24 Evans PR (1993) Data reduction. In *Proceedings of CCP4 Study Weekend on Data Collection and Processing*, 29–30 January 1993 (Sawyer L, Isaacs N & Bailey S, eds), pp. 114–122.

25 Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D* **50**, 760–763.

26 Vagin A & Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Cryst* **30**, 1022–1025.

27 Thompson JD, Higgins DG & Gibson TJ (1994) ClustalW: improving the sensitivity of progressive

multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choise. *Nucleic Acids Res* **22**, 4673–4680.
28 McRee DE (1999) XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* **125**, 156–165.
29 Glykos NM & Kokkinidis M (2000) GraphEnt: a maximum entropy program with graphics capabilities. *J Appl Cryst* **33**, 982–985.
30 Brünger AT, Adams PD, Clove GM, DeLano WL, Gros P, Grosse-Kunstkeve RW, Jiang J-S, Kuszewski J, Nilges M, Pannu NS *et al.* (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D* **54**, 905–921.
31 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D* **53**, 240–255.
32 Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr A* **32**, 922–923.
33 Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissing & Shindyalov Bourne PE (2000) The protein data bank. *Nucleic Acids Res* **28**, 235–242.
34 Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311–312.
35 Bergmeyer HU (1974) In *Methods of Enzymatic Analysis*, **Vol. 1**, 2nd edn (Bergmeyer HU, ed.), pp. 112–117. Verlag Chemie, Weinheim/Academic Press, Inc., New York, NY.

Supplementary material

The following supplementary material is available online:

Fig. S1. Structure of the *Bc*ZBP monomer.

Fig. S2. Asymmetric B-factors distribution on the highly symmetrical *Bc*ZBP hexamer.

Fig. S3. The active site of the *Bc*ZBP contains a zinc ion.

Fig. S4. Electron density map around the active site.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.