

## Molecular Dynamics Simulations of *BcZBP*, A Deacetylase from *Bacillus cereus*: Active Site Loops Determine Substrate Accessibility and Specificity

Vasiliki E. Fadouloglou,<sup>†,‡,§,⊥</sup> Athanassios Stavrakoudis,<sup>§</sup> Vassilis Bouriotis,<sup>‡,||</sup>  
 Michael Kokkinidis,<sup>‡,||</sup> and Nicholas M. Glykos<sup>\*,†</sup>

*Department of Molecular Biology and Genetics, Democritus University of Thrace,  
 University campus, GR-68100, Alexandroupolis, Greece, Department of Biology,  
 University of Crete, PO Box 2208, GR-71409, Heraklion, Crete, Greece, Department  
 of Economics, University of Ioannina, GR-45110, Ioannina, Greece, and Institute of  
 Molecular Biology and Biotechnology (IMBB), PO Box 1527,  
 GR-71110, Heraklion, Crete, Greece*

Received May 11, 2009

**Abstract:** *BcZBP* is an LmbE-like, homohexameric, zinc-dependent deacetylase from the opportunistic pathogen *Bacillus cereus* with three, thus far uncharacterized, homologues in *B. anthracis*. Although its specific substrate is still unknown, the enzyme has been shown to preferentially deacetylate *N*-acetylglucosamine and diacetylchitobiose via an active site based on a zinc-binding motif of the type HXDDX<sub>n</sub>H. In the crystal structure, the active site is located at a deep and partially blocked cleft formed at the interface between monomers related by the molecular 3-fold axis, although the major, in structural terms, building block of the enzyme is not the trimer, but the intertwined dimer. Here, we report results from a 50 ns molecular dynamics simulation of *BcZBP* in explicit solvent with full electrostatics and show that (i) the view of the intertwined dimer as the major structural and functional building block of this class of hexameric enzymes is possibly an oversimplification of the rather complex dynamics observed in the simulation, (ii) the most mobile (with respect to their atomic fluctuations) parts of the structure coincide with three surface loops surrounding the active site, and (iii) these mobile loops define the active site's accessibility, and may be implicated in the determination of the enzyme's specificity.

### 1. Introduction

*Bacillus anthracis* has recently attracted significant interest, mainly because of its putative usage as a biological weapon.<sup>1</sup>

\* To whom correspondence should be addressed. Tel. +302551030620. Fax. +302551030620, glykos@mbg.duth.gr. URL: <http://www.mbg.duth.gr/~glykos/>.

<sup>†</sup> Democritus University of Thrace.

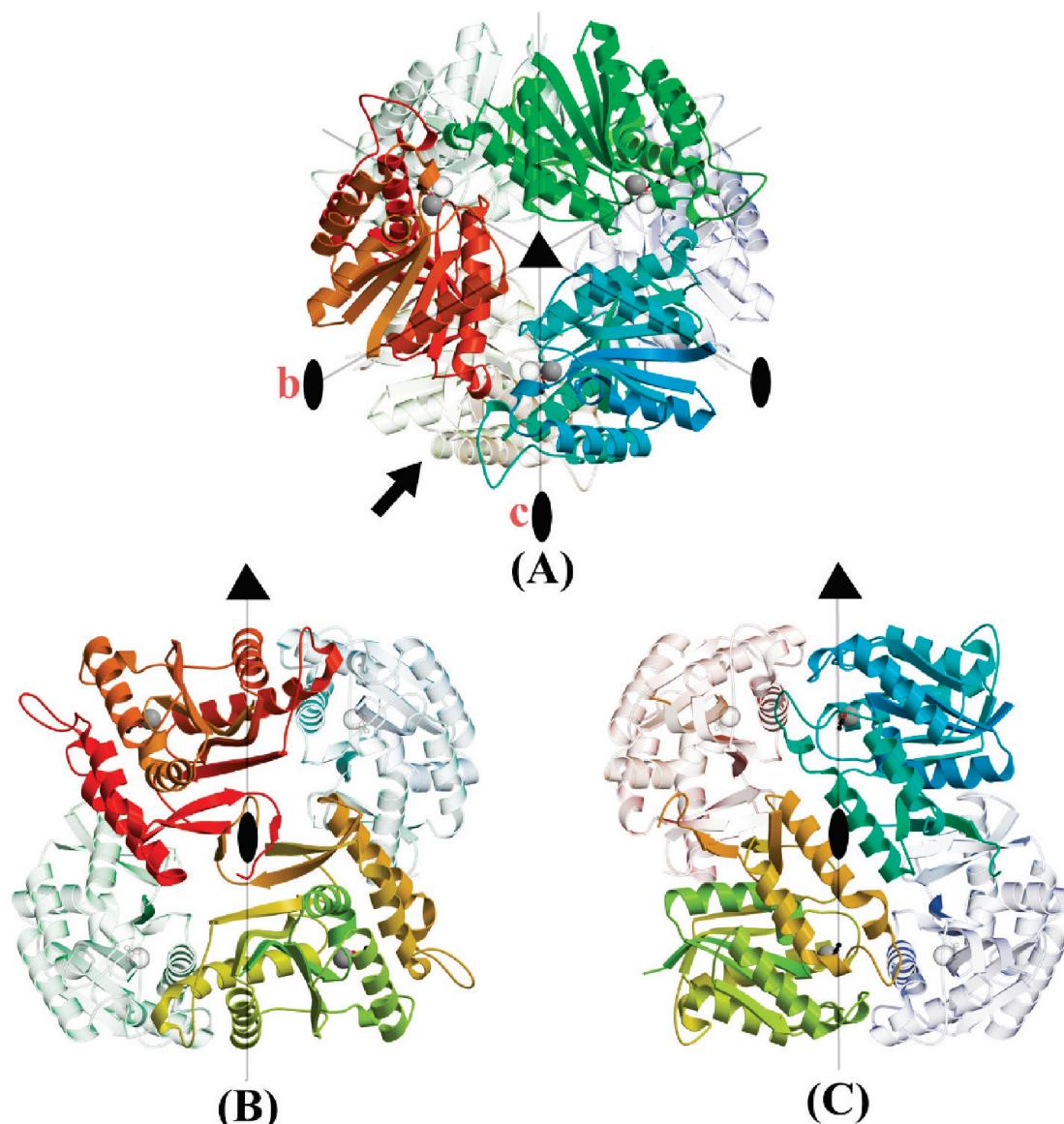
<sup>‡</sup> University of Crete.

<sup>§</sup> University of Ioannina.

<sup>||</sup> Institute of Molecular Biology and Biotechnology (IMBB).

<sup>⊥</sup> Present address: Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, CB2 1GA, Cambridge, England (UK).

Part of this interest was subsequently transferred to more benign, but still closely related to *B. anthracis*, species like *B. cereus*, an opportunistic bacterium causing food poisoning.<sup>2</sup> In a drive to characterize, both functionally and structurally, deacetylases that are shared between these two organisms (and which may be implicated in metabolic pathways of biotechnological and pharmaceutical interest), we have recently reported the characterization, purification, crystallization and crystal structure determination of *BcZBP*, a homohexameric, LmbE-like, zinc-dependent deacetylase from *B. cereus*.<sup>3–5</sup> *BcZBP* is the product of the *bc1534* gene, with three, thus far uncharacterized, homologues in *B. anthracis* sharing sequence identities (at the protein level)<sup>5</sup> of 96%, 28%, and 24%, respectively. Functional studies<sup>5</sup>



**Figure 1.** Crystal structure of *BcZBP*. Panel (A) is a schematic diagram of the hexamer viewed down the molecular 3-fold axis (which coincides with a crystallographic 3-fold of the *R*32 space group). The 3-fold passes through the geometric center of the molecule, is perpendicular to the plane of the paper, and its position is marked with a filled triangle. The 2-fold axes of the hexamer are on the plane of the paper, intersect the 3-fold at the molecular center, and their position is indicated by thin lines marked with filled ovals. The trimer that is farthest from the viewer (and below the plane of the paper) is drawn using transparency to reduce clutter. The arrow points to the monomer (of the lower trimer) that is colored green-orange in panels B and C. The position of the active site in each monomer is marked by the presence of a space-filling representation of the zinc atom. Panels B and C are views of the hexamer along the molecular 2-fold axes which are marked as b and c in panel A. In both cases, the 2-folds are perpendicular to the plane of the paper, pass through the molecular center and their position is marked with a filled oval. The 3-fold axis is on the plane of the paper and is marked with a line ending in a filled triangle. The trimer that was drawn transparent in panel A is now located toward the lower end of the page. To reduce clutter, only four monomers are shown in panels B and C with two of them drawn semitransparent. The dimers shown in panels B and C are referred to in the text as type I and type II dimers, respectively.

showed that *BcZBP* preferentially deacetylates *N*-acetylglucosamine and diacetylchitobiose, although its specific substrate remains unknown. The overall structure of the enzyme is shown in Figure 1A. It is a symmetric 32 ( $D_3$  in the Schoenflies notation) homohexamer with a molecular mass of 163 kDa and 234 residues in each monomer (we will hereafter refer to these six chains using the letters A–F). Their arrangement in the hexamer is such that chains A,C,E form the first trimer and chains B,D,F the second. In the

crystal, the molecular 3-fold coincides with a crystallographic 3-fold axis of the *R*32 space group, leaving the equivalent of two monomers (chains A and B, or equivalently C–D or E–F) per crystallographic asymmetric unit. Each monomer folds as a single  $\alpha/\beta$  domain in the form of a four-layered  $\alpha/\alpha/\beta/\alpha$  sandwich (most easily seen in the lower monomer of Figure 1B). The two trimers (of the hexamer) associate strongly, mainly via contacts located around the 2-fold axes. The first of these contacts is shown in Figure 1B and involves

an exchange of two short  $\beta$ -strands between the monomers A–B (and equivalently, C–D and E–F). The second contact (between monomers B–C, D–E, and A–F) is shown in Figure 1C and corresponds to a typical ridges-into-grooves  $\alpha$ -helical association with an interhelix angle of approximately  $25^\circ$  and a mean helix–helix distance of 8 Å. We will hereafter refer to these two types of dimers as “type I” (shown in Figure 1B) and “type II” (shown in Figure 1C) dimers, respectively.

The close, strand-exchange-based association seen in Figure 1B led to the conclusion<sup>5</sup> that the major structural building block of *BcZBP* is the type I dimer, and not, for example, the trimer or the type II dimer shown in Figure 1C. The view of *BcZBP* as a trimer of type I dimers was further reinforced by the relatively loose association of the monomers in the trimers as can be inferred from Figure 1A. Still, this view of *BcZBP* as a trimer of dimers could not be easily reconciled with two pronounced features of the crystal structure: The first feature was a systematic difference between the mean atomic temperature factors of the two trimers, with one trimer having significantly higher thermal parameters than the other. This observation was not consistent with the view of the type I dimer as the major structural and functional building block, mainly because such a dimer comprises monomers belonging to different trimers (it should be noted, however, that the presence of crystallographic symmetry relating the trimers’ monomers, forces any deviations in the overall atomic temperature factors to be at the trimer’s level). The second feature was that the accessibility to the active site appears to be mainly determined from loops originating from neighboring monomers not involved in the formation of the type I dimer. Furthermore, the crystal structure’s active sites were partially blocked from three surface loops of neighboring (in the trimer) monomers, making it difficult to imagine the type I dimer as the enzyme’s functional unit.

Here we present results from a 50 ns molecular dynamics simulation on the *BcZBP* hexamer in explicit solvent and with full electrostatics which was undertaken to characterize the structural and dynamical properties of this enzyme with emphasis on the properties of its hexameric association and its active sites’ accessibility.

## 2. Computational Methods

**2.1. System Preparation.** Starting from the crystallographically determined coordinates of the *BcZBP* hexamer (PDB entry 2IXD) missing side-chain and hydrogen atoms were built with the program PSFGEN from the NAMD distribution<sup>6</sup> and assuming a neutral pH. The histidine residues protonation state was determined according to their chemical environment in the crystal structure. An explicit solvent hexagonal periodic boundary system was prepared using VMD.<sup>7</sup> The unit cell basis vectors (projections along the orthogonal axes) of the periodic cell were (111,0,0), (0,90,52), and (0,0,103) Å with a shortest (initial) solute–solute distance of 30 Å. The solvation system comprised 25938 pre-equilibrated TIP3 water molecules, with the crystallographically determined waters retained throughout the pro-

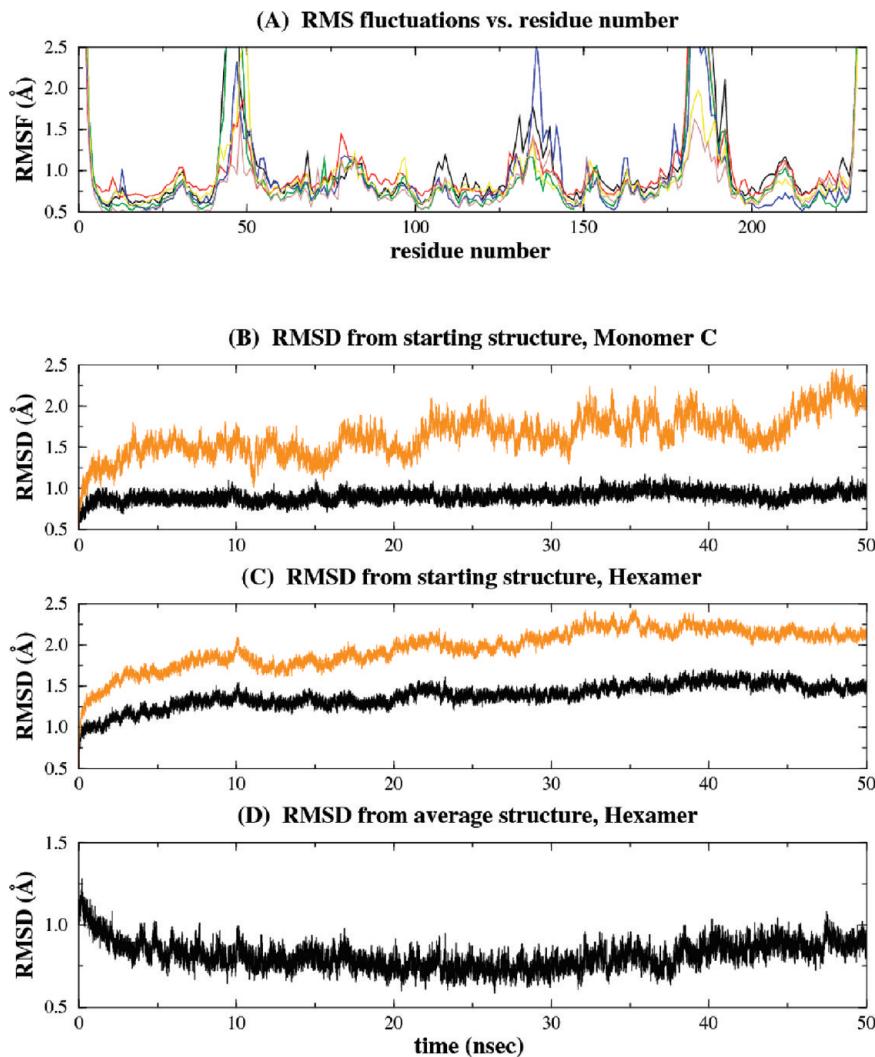
cedure, while those water molecules lying closer than 1.8 Å from the protein surface (or the crystallographic waters) were removed. The net charge of the solute was neutralized through the addition of sodium and chloride ions to a final concentration of ~100 mM corresponding to the addition of 33 sodium and 15 chloride ions. The final system comprised a total 99744 atoms, of which 21834 protein atoms, 6 zinc ions, 42 acetate atoms (located at the active site), and 77814 water atoms. The topology and parameter files used throughout the system preparation were those of the CHARMM27 force field.<sup>8</sup> The zinc ions were modeled using the nonbonded representation<sup>9</sup> implemented by the CHARMM27 force field.

**2.2. Molecular Dynamics Simulation Protocol.** A 50 ns molecular dynamics simulation was performed with the program NAMD<sup>6</sup> using the CHARMM27 force field<sup>8</sup> as follows. The system was first energy minimized for 2000 conjugate gradient steps with the positions of the backbone atoms fixed, and then for another 2000 steps without positional restraints. It was then slowly heated-up to a final temperature of 298 K (with a temperature step  $\Delta T = 20$  K) over a period of 66 ps with the positions of the  $C_\alpha$  atoms harmonically restrained about their energy-minimized positions. Subsequently the system was equilibrated for 200 ps under NpT conditions without any restraints. This was followed by a 50 ns production NpT run with the temperature and pressure controlled using the Nosé-Hoover Langevin dynamics and Langevin piston barostat control methods as implemented by the NAMD program (and maintained at 298 K and 1 atm). The production run was performed with the impulse Verlet-I multiple time step integration algorithm as implemented by NAMD. The inner time step was 2 fs, short-range nonbonded interactions were calculated every one step, and long-range electrostatics interactions were calculated every two timesteps using the particle mesh Ewald method.<sup>10</sup> A cutoff for the van der Waals interactions was applied through a switching function, and SHAKE was used to restrain all bonds involving hydrogen atoms. Trajectories were obtained by saving the atomic coordinates of the whole system every 0.4 ps.

**2.3. Trajectory Analysis.** Generation of modified PSF files was performed with X-PLOR.<sup>11</sup> Calculation of the anisotropic fluctuations was performed with the program g\_rmsf from the GROMACS suite of programs.<sup>12</sup> Removal of global rotations-translations, calculation of rms deviations from the experimental structure, calculation of the average trajectory structures, of the rms deviation from the average structures, of the radius of gyration, of the atomic rms fluctuations, the  $C_\alpha$ – $C_\alpha$  distance map (and the corresponding rms deviation from it), the cross-correlation matrix, and the Cartesian<sup>13,14</sup> and dihedral-angle<sup>15,16</sup> principal component analysis were performed with the program Carma,<sup>17</sup> available via <http://www.mbg.duth.gr/~glykos/>.

## 3. Results

**3.1. The *BcZBP* Trajectory is Stable.** The simulation was stable, both with respect to its state variables and the structure of the enzyme, with the notable exception of three

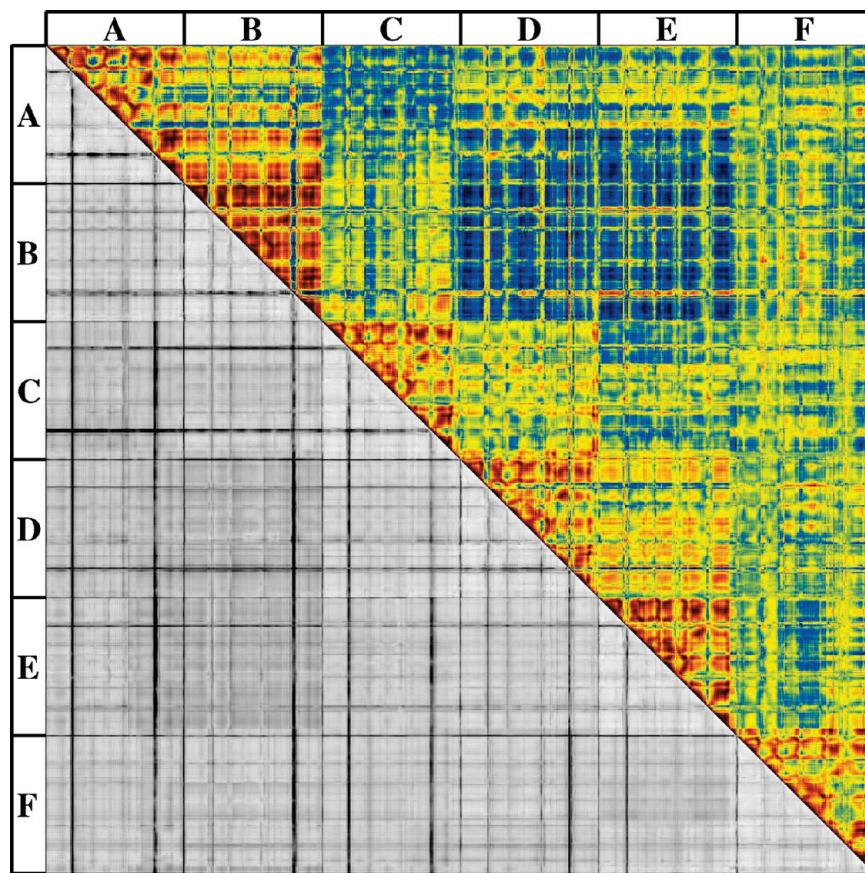


**Figure 2.** Evolution of structure-dependent quantities during the molecular dynamics simulation. Panel A shows the root-mean-squared fluctuations (in Å) of the  $C_\alpha$  atoms around their average positions during the whole length of the simulation. The graph is a superposition of six curves (corresponding to the six monomers) with the horizontal axis corresponding to residue numbers. The three highly mobile active-site loops are centered around residues 48, 135, and 186 (see text for details). For clarity this diagram is truncated to 2.5 Å along the vertical axis, with the fluctuation values reaching out to 4 and 5 Å for the first and third loop respectively. Panel B shows the rms deviation of the  $C_\alpha$  atoms of monomer C versus simulation time. The two graphs shown in this panel were calculated either with the mobile loops included in the calculation (upper curve), or excluded (lower curve). Panel C is the same calculation as for panel B, but using the  $C_\alpha$  atoms of the whole hexamer. Finally, panel D shows the rms deviation of the  $C_\alpha$  atoms of the whole hexamer from their average positions during the length of the simulation (the three mobile loops, see text for details, have been excluded from this calculation).

surface loops which -as will be discussed later- surround the enzyme's active sites. This is shown in Figure 2A, which depicts the root-mean-squared (rms) fluctuations of the  $C_\alpha$  atoms' positions (from their average) for the six monomers and for the whole length of the trajectory. As can be seen from this figure, the majority of atomic fluctuations are significantly less than 1.0 Å with the exception of (a) the first two and last three (N- and C-terminal) residues of each monomer and (b) three surface loops extending from residues 42–51, 129–140, and 180–192. The same conclusions can be drawn from the lower half of Figure 3, which depicts (using a grayscale representation) the rms deviations of the  $C_\alpha$ – $C_\alpha$  distances from their average during the length of the trajectory. The pronounced horizontal and vertical dark lines, which are apparent in this diagram correspond to residues with higher than average mobility with respect to the rest of

the structure, and, not unexpectedly, match closely the loop regions identified from Figure 2A. The agreement between the simulation-derived atomic fluctuations and the crystallographically determined atomic temperature factors (allowing for the fact that only two monomers are crystallographically independent) is quite high, with an average value for the linear correlation coefficient (over all possible pairs of monomers) of 0.70. The highly correlated values of the experimental and simulation-derived atomic fluctuations, together with the stability of the crystallographically determined structure (discussed below), provide further indications for the validity and quality of the simulation protocol.

Examination of the linear correlation coefficients between the per-residue atomic fluctuations of the various monomers (Figure 2A), shows a notable pattern: the correlations between monomers of type I dimers are relatively low at



**Figure 3.** Dynamics of the *BcZBP* hexamer. The upper triangle of the diagram is the normalized variance-covariance (cross-correlation) matrix of the  $C_\alpha$  atoms. The color representation ranges from dark red, through yellow, to dark blue corresponding to correlation values from +1.0 (fully correlated), to 0.0 (uncorrelated), to -1.0 (fully anticorrelated). To increase the contrast of this diagram, a sigmoidal function of the form  $\sigma(x) = 2/[1 + e^{(-3 \cdot x)}] - 1$  has been applied to the raw data. The areas of the diagram corresponding to the various monomers of the hexamer are indicated with the letters A–F at the top and left-hand-side of the matrix. The lower half of the matrix is a grayscale representation of the rms deviation of the  $C_\alpha$ – $C_\alpha$  distances (and for all possible pairs) from their average distances observed during the length of the trajectory (in other words, it is the rms deviation map of the average  $C_\alpha$ – $C_\alpha$  distance map). The grayscale gradient ranges from white (corresponding to an rms deviation of 0.0 Å) to black (corresponding to an rms deviation of 3.0 Å or more). The limits in terms of the individual monomers of the hexamer are shown on the top and left-hand-side of the matrix.

0.70, 0.64, and 0.42, lower than the correlations between type II dimers (at 0.89, 0.75, and 0.46). Moreover, they are both lower than the values observed for monomers belonging to the same trimer (0.97, 0.78, 0.85 for the A–C–E trimer, 0.84, 0.68, 0.82 for the B–D–F trimer). The higher values of the linear correlation coefficient between monomers belonging to the same trimer is consistent with the results discussed in the next section.

Ignoring the three highly mobile loops, the structures of both the individual monomers and of their relative arrangement on the hexamer are well preserved during the simulation. As shown in the upper curve of Figure 2B for a representative monomer (monomer C), the rms deviation from the starting (crystal) structure increases steadily throughout the simulation, reaching values close to 2.5 Å. If the three surface loops surrounding the active site are excluded from the calculation, the results are significantly different (Figure 2B, lower curve): the rms deviation from the crystal structure quickly stabilizes to a value of approximately 0.8 Å and remains stable throughout the simulation. Similarly, Figure 2C compares the behavior of the rms deviation from

the crystal structure with or without the active site loops, but this time considering the  $C_\alpha$  atoms of the whole hexamer. As can be seen from this figure, the effect of excluding the active site loops from the calculation is again pronounced, though less dramatic when compared with the monomer-derived results. This indicates that there is a contribution to the rms deviation arising from the intermonomer association. Again, this is in agreement with the  $C_\alpha$ – $C_\alpha$  distance deviation map (lower half of Figure 3) which clearly shows whole areas with higher than average rms deviations (note, for example, the darker area corresponding to vectors between the  $C_\alpha$  atoms of the B monomer and those of monomers E and D).

In agreement with the results presented above, Figure 2D shows the evolution (as a function of simulation time) of the rms deviation between the trajectory's average structure and each and every of the structures observed during the simulation (considering  $C_\alpha$  atoms only). As can be seen, the structure quickly (within 5 ns) relaxes from the initial crystal structure, and then remains stably close to its average with deviations of approximately 0.6 Å going up to 0.8 Å near

the end of the trajectory. The major differences between the starting (crystal structure) and the trajectory's average structure can be accounted by a concerted relaxation of the intermonomer association as indicated by an increase of the mass-weighted radius of gyration from the starting value of 31.4 Å to a value of 32.1 Å after only 5 ns of simulation time. Nevertheless, the rms deviation between the C<sub>α</sub> atoms' positions in the average and crystal structures remains quite low at 1.2 Å (excluding the active site loops).

**3.2. The Hexameric Association Dynamics Are Complex.** The upper half of Figure 3 shows a pseudocolor representation of the hexamer's normalized variance-covariance (cross-correlation) matrix. Apart from the (expected) trend of strong positive correlations between atoms belonging to the same monomer, the cross-correlation pattern appears to be rather complex and inconclusive, especially about the mode of intermonomer association within the hexamer. If we compare the cross-correlation patterns observed for the type I dimers (A–B, C–D, E–F) with those observed for the type II dimers (B–C, D–E, A–F), we find that none of these two association models is conclusively supported by the simulation: there is strong positive correlation for the (type I) A–B dimer, but negative for the (also type I) E–F dimer. Similarly, there is positive correlation for the type II D–E dimer but negative for the (also type II) B–C dimer.

The case of the E–F dimer warrants additional discussion with respect to the β-strand-exchange dimerization motif: focusing in the area of the matrix corresponding to cross-correlations between the E and F monomers, note the thin band of positive correlation connecting the C-terminus of the F monomer with the whole of E. Similarly, there is a thin band of positive correlation connecting the C-terminus of the E monomer with the whole of F. Apart from these two bands, the rest of the matrix in this area shows either uncorrelated or even anticorrelated motion. What this implies is that, at least for the case examined here, the exchanged strands became integral parts of the monomers that receive them, and that they do not affect the independence of dynamics of the associating monomers. Clearly, even a dimerization motif as strong and explicit as a β-strand-exchange, can be surprisingly malleable with respect to protein dynamics.

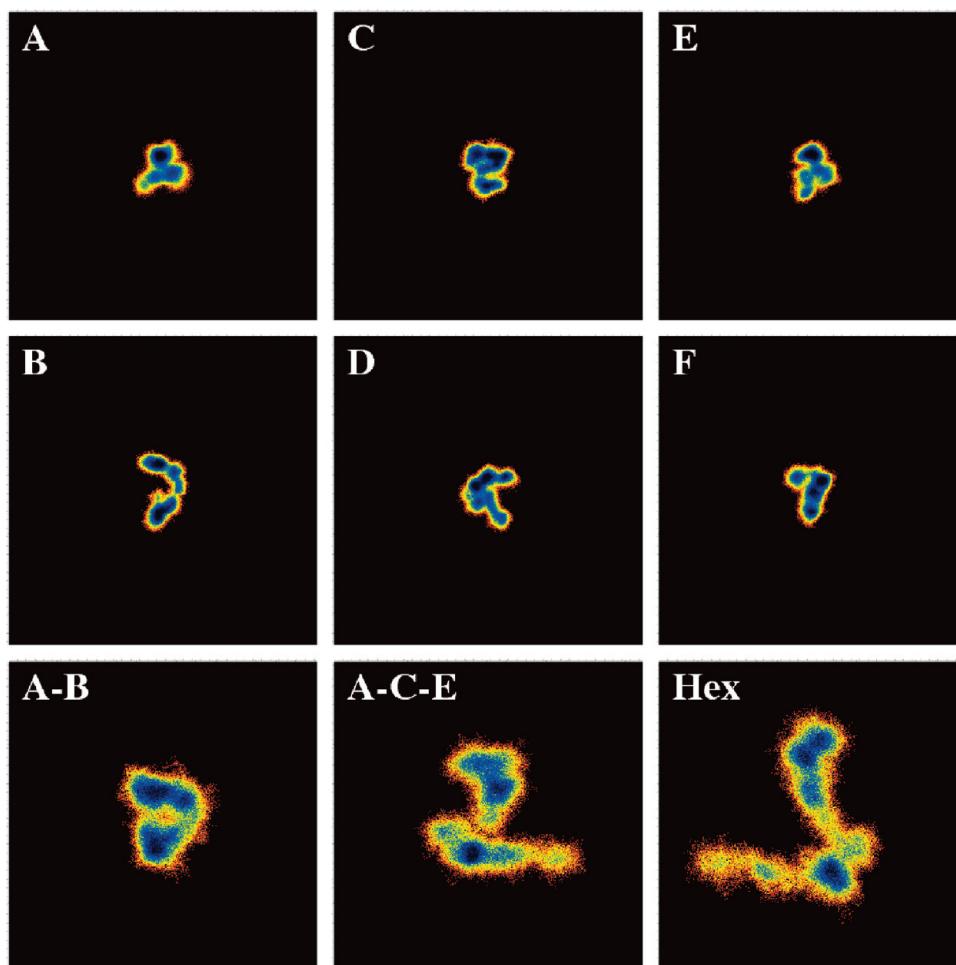
Turning our attention to the trimers, we note what is possibly the most persistent characteristic of the matrix: cross-correlations between monomers related by the molecular 3-fold axis are mostly negative (see the matrix areas defined by the monomer pairs A–C, A–E, and C–E for the first trimer, B–D, B–F, and D–F for the second trimer). This motif of anticorrelations is consistent with a ‘breathing’ motion of the trimers about the molecular 3-fold axis, in agreement with the results from the principal component analysis discussed in the next section.

For completeness we should note that the hexamer dynamics model that best agrees with the variance-covariance matrix is a rather unexpected one, as it involves the interpretation of the data in terms of two dimers (of different types) and of two independent monomers: Referring to Figure 3, the most prominent feature of the cross-correlation matrix is the band of negative correlations connecting the A–B pair

of monomers with the D–E pair. Additionally, the cross-correlations between the monomers A and B on one hand, and monomers D and E on the other, are among the most strongly positive of the matrix. Lastly, monomers C and F appear to be mostly uncorrelated with all other monomers of the hexamer. Taking these indications together, they appear to suggest a 1 + 2 + 2 + 1 model for the hexamer dynamics: four monomers (A, B, D, E) form a dimer of dimers (A–B + D–E), which is capped on either side by two independent monomers (C and F). This arrangement can be indirectly visualized from Figures 1B and C if it is assumed that these two views are related by a rotation of 180° about the 3-fold axis: the two semitransparent monomers correspond to monomers C and F, which cap the two (colored) dimers, one dimer of type I (Figure 1B) and one of the other of type II (Figure 1C). Although this model appears to explain most of the features of the cross-correlation matrix, it is difficult (if not fundamentally impossible) to reconcile with the intramolecular 32 symmetry. Indeed, a trajectory of a stable, symmetric, homo-hexameric protein at equilibrium should, if sufficiently sampled, give a cross-correlation matrix obeying the intramolecular symmetry. We attribute the absence of symmetry from the matrix to the insufficient sampling of our trajectory as discussed below.

**3.3. Principal Component Analysis and Sufficient Sampling.** The relatively low and rather stable rms deviations shown in Figure 2 (both in terms of the starting structures and of the trajectory-average structures) may leave the impression that the protein dynamics (especially at the monomer level, Figure 2B) have been sufficiently sampled during the simulation. As Figure 4 clearly indicates, this is definitely not so: the projections of the C<sub>α</sub> atoms' fluctuations on the planes of their principal components deviate significantly from two-dimensional Gaussians centered at the origin (which is what we would expect from the trajectory of a sufficiently sampled single-state protein structure at equilibrium). Although there is a clear and significant difference between the extend of fluctuations of the monomers and of the higher-order oligomers, even the monomers' principal component projections show fine structure inconsistent with sufficiently sampled dynamics. It could be argued that this fine structure may correspond to functionally important discrete conformational states of the *BcZBP* monomers (with this line of argument being easily expandable to the whole hexamer). But, if this were indeed the case and because of the presence of intramolecular symmetry, we would expect these substates to be correlated between the different monomers. A cursory examination of the monomer diagrams in Figure 4 indicates that this is probably not the case. To resolve the matter in a quantitative way, we calculated the overlap between the eigenvector-defined subspaces for all possible monomer combinations and for the three eigenvectors corresponding to the three largest eigenvalues. The overlap between the subspaces defined by two sets **v** and **w** of *n* eigenvectors is defined as

$$\text{overlap}(\mathbf{v}, \mathbf{w}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (\mathbf{v}_i \cdot \mathbf{w}_j)^2$$

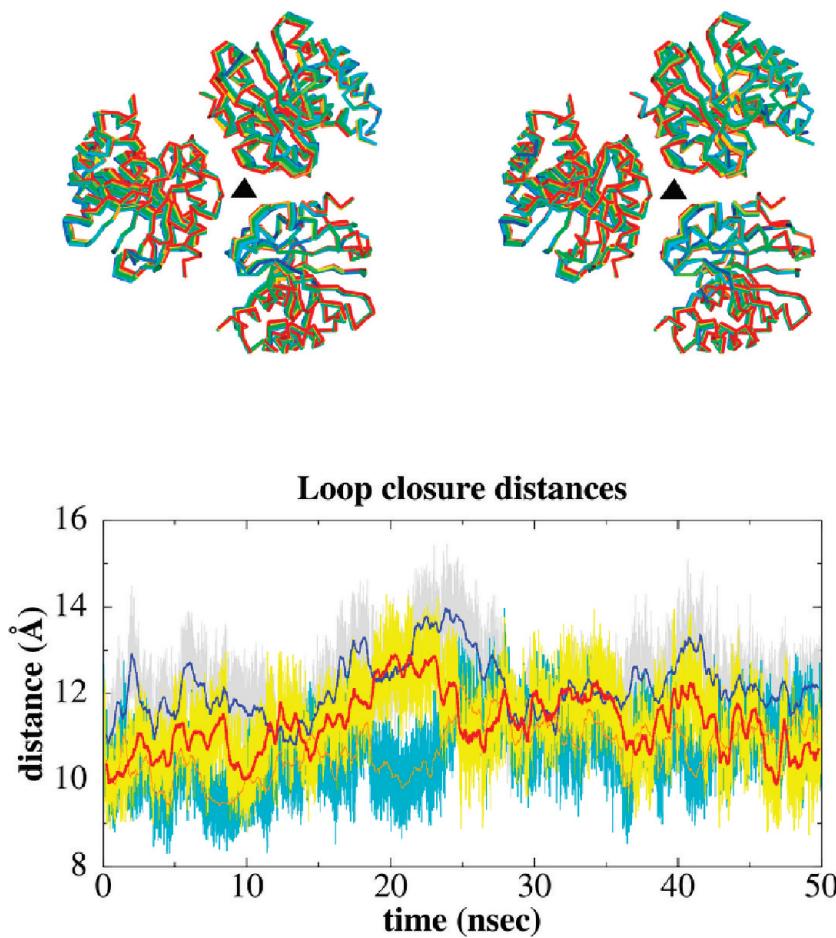


**Figure 4.** Cartesian principal component analysis: first vs second principal component plots for the *BcZBP* monomers (top two rows), the hexamer (lower right), and a representative dimer and trimer (last row). All diagrams shown in this figure are pseudocolor representations of density functions corresponding to the projections of the fluctuations of the C<sub>α</sub> motion (excluding the active-site loops) on the planes of the top two eigenvectors (of the respective molecular species indicated in the figure). For all graphs the origin is on the upper, left-hand side corner, values on all axes range from -35 to 35 Å and the eigenvectors corresponding to the largest eigenvalues are along the vertical axes. The density function shown is  $\Delta G = -k_B T \ln(p/p_{\max})$  where  $k_B$  is Boltzmann's constant,  $T$  is the temperature in Kelvin, and  $p$  and  $p_{\max}$  are probabilities obtained from the distribution of the principal components for each structure (frame) from the corresponding trajectory. As a result of applying this function, the diagrams have units in kcal/mol with corresponding values for the minimum of the A monomer of -3.51 kcal/mol, for B -3.30, C -3.19, D -3.27, E -3.52, F -3.35, for the A-B dimer -2.56, for the A-C-E trimer -2.47, and, finally, for the hexamer -2.37 kcal/mol. Note that these values can only be compared between trajectories of the same molecular species (in this case, only between monomers). Please also note that the  $\Delta G$  values obtained from this procedure are on an arbitrary scale in the sense that they depend on the binning procedure used for calculating the  $p$  and  $p_{\max}$  values. For all diagrams of this figure, the raw data were binned on a square matrix of size  $\sqrt{N}/2$  where  $N$  is the number of frames of the corresponding trajectory.

and takes values from zero (signifying no convergence of the corresponding subspaces) to one (for full overlap of the subspaces). Over all possible monomer combinations, the average overlap value (for the top three eigenvectors) was only 0.05 (with a standard deviation of 0.02). The highest value observed was 0.10 between monomers C and D. What these results clearly show is that the length of the simulation has been inadequate to sample sufficiently even the monomers' dynamics, let alone the whole of the *BcZBP* hexamer. This is in agreement with the indications obtained from the cross-correlation matrix (see last paragraph of the previous section). Similar results to those discussed above have been obtained from a principal component analysis performed in

dihedral ( $\phi, \psi$ ) angle space (which is not sensitive to rigid-body-like motion of protein domains or subdomains).

With the precautions necessitated by the lack of sufficient sampling discussed above, we note the systematic difference between the projections of fluctuations for monomers belonging to different trimers (Figure 4, first row of graphs vs second row): the extent of the atomic fluctuations on the eigenvector planes are correlated at the level of the two trimers, but not at the level of the type I or II dimers (the fluctuations are systematically lower for the A-C-E trimer compared with the B-D-F trimer). Such a systematic difference at the trimers' level was also observed when considering the fluctuations from the average structures in



**Figure 5.** Trimer dynamics: the upper panel stereodiagram (wall-eyed) is a smooth representation of the trimer's C<sub>α</sub> motion (excluding the active-site loops) as calculated from the first principal component only. The intramolecular 3-fold axis is perpendicular to the plane of the paper and its position is noted by the filled triangle. The various structures shown superimposed are color-coded from blue, via green and yellow, to red, and correspond to the structures obtained by applying (on the average structure) the fluctuations corresponding to the first eigenvector weighted by a smoothly varying amplitude (obtained from the principal component analysis, and ranging for this diagram from -19 to 19 Å). The graphs in the lower panel show the variation of the distances between the tips of the three loops closest-to and surrounding the 3-fold axis (the loops immediately next to the filled triangle in the upper panel). The actual distances (as observed in the trajectory) are shown as light-colored backgrounds. The solid lines are averages which were calculated using a 40 ps window.

section 3.1 above, and also during the analysis of the variance-covariance matrix in section 3.2.

These indications concerning correlated dynamics at the trimers' level, prompted us to examine in more detail the molecular motion associated with the trimers' principal components. The eigenvector with the largest contribution to the intermonomeric association dynamics is the first one (data not shown). The top panel of Figure 5 shows a smooth representation of the ACE trimer's motion due to the first eigenvector considering only the C<sub>α</sub> atoms and ignoring the flexible active-site loops. As can be seen from this figure, there are indeed some indications of an anticorrelated motion of the monomers about the molecular 3-fold axis. Such a breathing-like motion of the monomers was also observed when discussing the differences between the crystal structure and the trajectory-average structure (section §3.1). To quantify this statement, we have calculated—as a function of simulation time—the distances between the C<sub>α</sub> atoms of the residues located at the tips of the loops which are closest to the molecular 3-fold axis [Note that these distances were

obtained directly from the molecular dynamics trajectory and not from the principal component-derived motion shown in the top panel of Figure 5]. If the notion of a breathing-like motion was supported by the raw simulation data, then these distances should be correlated. The results from this calculation are shown in the lower graph of Figure 5. As can be seen from this graph, the variation of the loop-closure distances is indeed correlated, but not uniformly: the linear correlation coefficient between the A-C and C-E distances is 0.40, but is reduced to 0.20 for the A-C and A-E monomers, and to 0.12 for the A-E, C-E combination. Taking the results from these calculations together, they seem to be consistent, at least within the limitations posed by the lack of sufficient sampling, with the notion of a breathing-like motion of the monomers about the molecular 3-fold.

The discussion above, together with the absence of symmetry from the cross-correlation matrix shown in Figure 3, may create the impression that the symmetry of the hexamer is not well preserved during the simulation. To unequivocally show that this is not the case, we examined

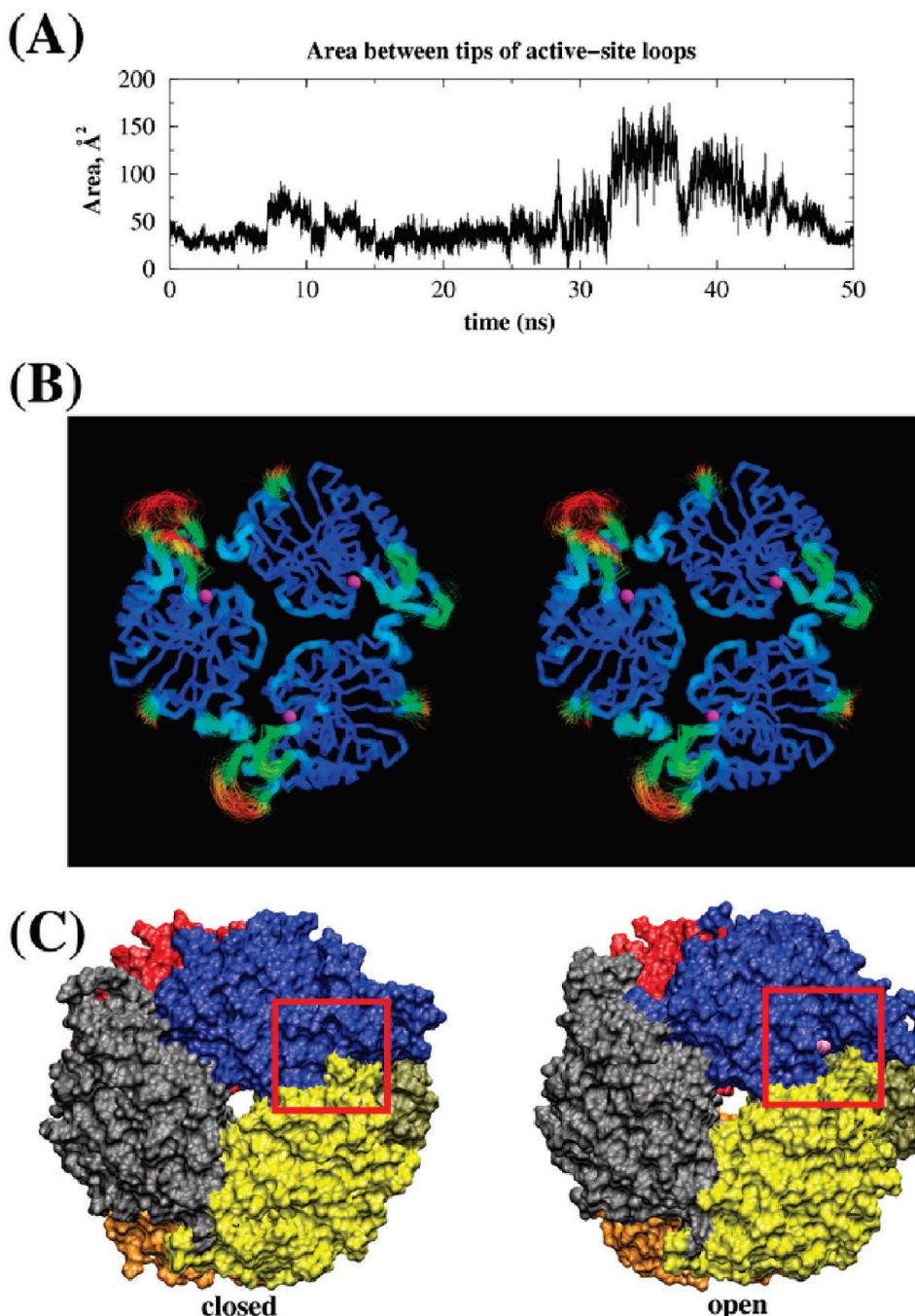
the symmetry and structural conservation of the whole enzyme using for our calculations its average structure calculated over the length of the trajectory. This we did as follows: In the first step, the average structure of monomer A was least-squares superimposed on the average structures of the other five monomers and the rms deviations between their  $C_{\alpha}$  atoms (excluding the flexible loops) were recorded. The values we obtained were 0.57 Å for the A–B monomers, 0.31 Å for A–C, 0.51 Å (A–D), 0.39 Å (A–E) and 0.44 Å between the A and F monomers. This clearly shows that the average structures of the six monomers are practically identical, making unnecessary to calculate superpositions for all their possible pairwise combinations. To examine the preservation of symmetry, we converted the rotation matrices (obtained from the least-squares superposition) to their equivalent sets of polar angles ( $\omega$ ,  $\phi$ ,  $\kappa$ ) and examined the deviations of these angles from the values expected for a hexamer of 32 symmetry (for, example, rotations with  $\Delta\kappa = 120^\circ$  for the 3-fold axes or  $\Delta\kappa = 180^\circ$  for the two-folds). The average deviations of these polar angles from their ideal values (for a perfect 32 hexamer) were as low as 0.9 degrees, showing that the hexameric symmetry is almost perfectly conserved during the simulation. This, however, creates a conceptual problem: if both the symmetry and the structure of the enzyme are so highly conserved, why the variance-covariance matrix is not symmetrical (or, equivalently, why its derived principal components have not converged)? The answer, we believe, is that the variance-covariance matrix (and its principal components) are dominated by the small-scale fluctuations of a very stable (but large) structure, making convergence difficult to achieve within the time scale of our simulation. To show that it is indeed the small-scale fluctuations that dominate the PCA calculation, we recalculated the variance–covariance matrix, but this time we did not normalize it, keeping its units in Å<sup>2</sup>. Taking the average of the absolute values of the matrix (and excluding all intramonomer correlations), we obtained a value of only 0.106 Å<sup>2</sup> with a standard deviation of 0.166 Å<sup>2</sup>. If the flexible loops are excluded from the calculation, then we obtain an average value of 0.077 Å<sup>2</sup> with a standard deviation of only 0.069 Å<sup>2</sup>. These results clearly show that the major contribution to the variance–covariance matrix (and its principal components) is not large-scale correlated motion (which, if present, would mask the absence of convergence for the smaller-scale motions), but small-scale fluctuations about an otherwise stable average structure.

**3.4. Active-Site Loops: Mobility and Accessibility.** The stereodiagram shown in Figure 6B illustrates in a direct and immediate way what has already been mentioned on several occasions in the previous sections: the *BcZBP* structure, both at the monomeric and oligomeric levels, is very well preserved during the simulation with the exception of the three loops that surround the enzyme's active sites. As can be inferred from Figure 6B, both the structural core of the monomers and their relative orientations in the oligomer are very stable as indicated by the excellent superposition of the structures and the low rms fluctuations (indicated by the dark blue color). In contrast, the three loops surrounding the active sites, which are marked by the space-filling model of the

zinc atom, are exceedingly mobile as evidenced both by the divergence of the various structures and the high rms fluctuation values (indicated by the dark red color). For each active site, the three loops surrounding it are contributed by two neighboring (at the trimers' level) monomers: the monomer to which the active site belongs contributes two loops, the first loop comprising residues 42–51 and the second residues 180–192. The neighboring monomer contributes the loop extending from residue 129 to residue 140. We will hereafter refer to these three loops as L<sub>46</sub>, L<sub>185</sub>, and L<sub>135</sub>. Comparison of the loop mobility for the three active sites shown in Figure 6B shows a notable pattern: loops L<sub>46</sub> and L<sub>185</sub> have consistently higher mobility than L<sub>135</sub> (this can also be inferred from the graphs shown in Figure 2A). Additionally, the amount of mobility observed for L<sub>46</sub> and L<sub>185</sub> varies significantly between the various active sites (compare, for example, the two active sites that are located in the upper part of Figure 6B). Although the presence of this variability in the atomic fluctuations may be connected with the limited sampling discussed in section 3.3, the molecular dynamics trajectory per se is remarkably self-consistent with respect to the presence and the amplitude of fluctuations of the hypermobile loops. To quantify this statement we proceeded as follows: The rms atomic fluctuations of all 1386  $C_{\alpha}$  atoms of the protein were calculated for two disjoined trajectory segments extending from 10 to 30 ns (for the first segment), and from 30 to 50 ns (for the second). The value of the linear correlation coefficient between the atomic fluctuations obtained from these two segments was as high as 0.852, clearly indicating that the molecular dynamics trajectory is internally consistent with respect to the presence of the hypermobile loops. Additionally, the fluctuations obtained from the two segments are in excellent agreement with the results obtained from the whole trajectory (and shown in Figure 2A) with corresponding values of the linear correlation coefficient of 0.954 and 0.916.

It could be argued that the amount of loop mobility observed in the trajectory is not the result of the protein dynamics per se, but arises as an artifact of the nonbonded representation of the zinc ions used for modeling the enzyme's active sites.<sup>18</sup> This is clearly not the case for two reasons. The first reason is that the protein residues involved in zinc coordination (residues 12, 15 and 113) are outside the limits of the mobile loops as described above. The second and more important reason is that the geometries of all six active sites are very highly conserved. To quantify this statement, we calculated the rms deviation from the starting (crystal) structure for all non-hydrogen atoms of all protein residues that are involved in the zinc ion coordination. In the case, for example, of monomer C, the mean rms deviation (and for the whole length of the trajectory) was only 0.47 Å with a standard deviation of 0.05 Å, comparable with the expected coordinate error of the crystal structure. Such low rms deviations for the active site residues clearly indicate that the presence of hypermobile loops is not in any way connected with the model chosen for the representation of the zinc ions.

The amount of loop mobility seen on Figure 6B, immediately possess the question of whether the motion of the



**Figure 6.** Active-sites' loop mobility and accessibility. Panel (A) shows the variation (as a function of simulation time) of the area of a triangle defined by the  $C_\alpha$  atoms of Asp184 and Ser46 from monomer A, and Lys131 from monomer E. These three residues lie at the tips of the three loops surrounding one of the enzyme's active sites. Panel (B) is a stereodiagram (wall-eyed) illustrating the mobility of the loops surrounding the active sites. The diagram corresponds to a superposition of structures obtained directly from the molecular dynamics trajectory (after removal of overall rotations-translations). The view is down the molecular 3-fold axis, and to reduce clutter only one trimer is shown. The position of the active sites is marked by the space-filling models (colored magenta) of the zinc atoms. The structures are colored according to their atomic (per  $C_\alpha$ ) rms fluctuations using a linear gradient ranging from dark blue to dark red. Finally, panel (C) shows two space filling models of the whole hexamer (taken directly from the trajectory) illustrating the loop-dependent opening and closing of one of the active sites (see boxed area of the diagrams).

active-site loops is correlated or not. This would have functionally important implications since the presence of correlated motion would suggest that the loops undergo a concerted movement (as expected, for example, from a periodic opening and closing of the active sites). To tackle this question we once again resorted to the cross-correlation

matrix shown in Figure 3, this time examining only those parts of the matrix that correspond to the  $L_{46}$ ,  $L_{185}$  and  $L_{131}$  loops. To avoid qualitative assessments we proceeded as follows: In the first step, the entries of the matrix corresponding to the cross-correlation values for the  $C_\alpha$  atoms of the three loops were isolated. In the second step, we selected

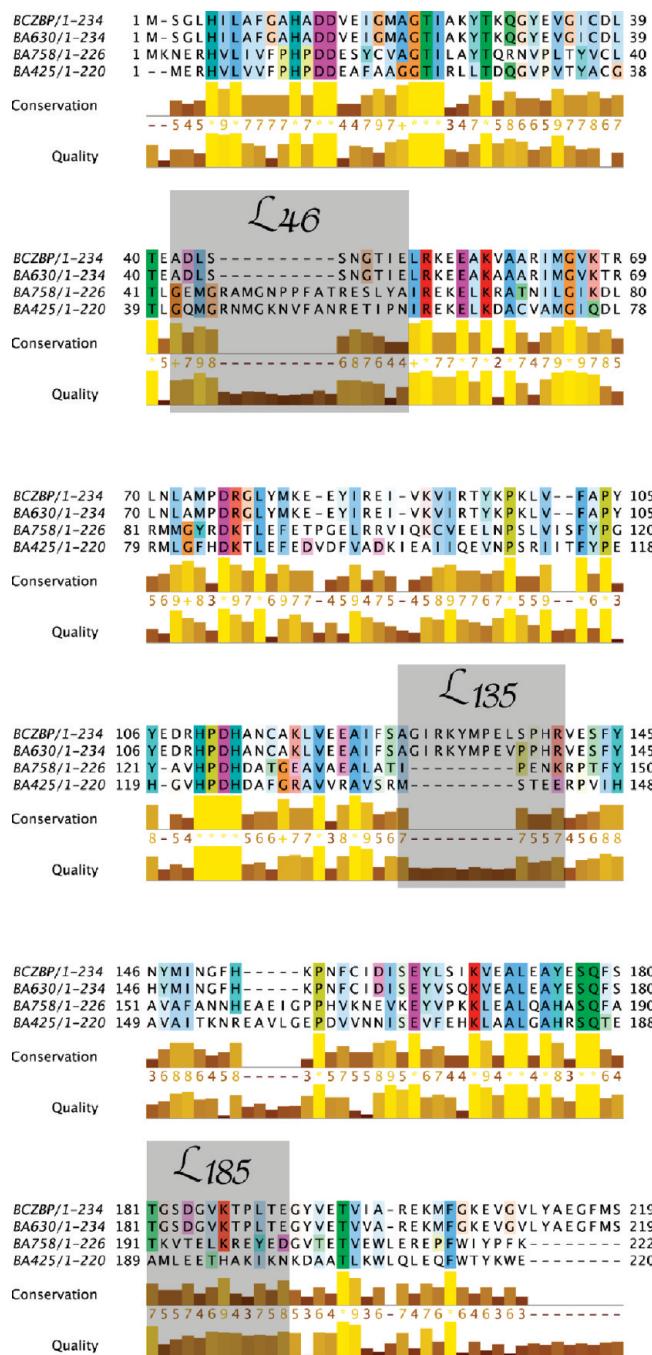
only those entries that corresponded to cross-correlation values between C<sub>α</sub> atoms of different loops (that is, we excluded values between atoms belonging to the same loop). In the final step, we grouped these values with respect to the active sites that the loops belonged to. The net result of this procedure is a set of cross-correlation values between the C<sub>α</sub> atoms of each loop, with each and every C<sub>α</sub> atom of the two other loops that surround the same active site. The numerical results obtained from these calculations were conclusive: the average value of the cross-correlation coefficient between atoms surrounding, for example, the active site located between the A and E monomers is only 0.03 with a standard deviation of 0.14. Similar results have been obtained for all the other active sites of *BcZBP*. The implication of the preceding analysis is that the opening and closing of the enzyme's active sites is a stochastic process, dependent on the random conformational changes of the three loops.

Consistent with the notion of a stochastic process, Figure 6A shows the variation of the area of a triangle defined by the tips of the three loops surrounding the active site located between the A and E monomers. The area defined by the three loops remains more or less stable at ~45 Å<sup>2</sup> for the first 30 ns of the trajectory, and then an opening event is recorded lasting for approximately 17 ns. During this opening event, the area defined by the three loops more than triples, reaching values as high as ~170 Å<sup>2</sup>. To show unequivocally the extend of this significant change in the active sites' accessibility, Figure 6C compares snapshots (recorded directly from the trajectory) taken at 5 ns (diagram on the left, active site closed) and at 37 ns (diagram on the right, active site open). This very notable change in the active sites' accessibility as seen in Figure 6C is amplified even further if it is considered that the plane defined by the tips of the three active site loops is not perpendicular to the viewing axis (and so, the observed amount of opening is only a projection of the real difference in active-site accessibility).

### 3.5. Active-Site Loops: Conservation and Specificity.

Figure 7 shows a multiple sequence alignment of *BcZBP* with its three homologues from *B. anthracis* together with a per residue conservation and quality score as calculated by the program Jalview<sup>20</sup> (note that the alignment shown is the unedited result of a default run of the program T-coffee<sup>19</sup>). The correspondence between the three hypermobile active-site loops and the parts of the alignment with accumulated gaps and low conservation score is striking, especially for the L<sub>46</sub> and L<sub>135</sub> loops. Indeed, if the proteins' termini are excluded from consideration, then almost all low conservation regions from the alignment match exactly the three active-site loops. An exception to this observation is the area centered around *BcZBP*'s Lys154 which also shows very low conservation and quality score. Although this region is also close to the active site, examination of the *BcZBP* structure suggests that it is rather unlikely for this area to be directly involved with the active site accessibility and/or specificity.

Given the observed accumulation of insertions/deletions and the low conservation score of the loops surrounding the active site, it is tempting to speculate that these loops not only contribute to defining the accessibility to the active sites



**Figure 7.** Multiple sequence alignment of *BcZBP* with its three *B. anthracis* homologues. The three hypermobile active site loops are noted with the shaded boxes and are marked as L<sub>46</sub>, L<sub>185</sub>, and L<sub>135</sub> (see text for details). The *B. anthracis* proteins are denoted as: BA630 corresponding to NP\_844007 (gi: 30261630), BA758 corresponding to NP\_846135 (gi: 30263758), and BA425 corresponding to NP\_845802 (gi: 30263425). The multiple sequence alignment shown is the unedited result from a default run of the program T-coffee.<sup>19</sup> The conservation and quality scores are as produced by the program Jalview.<sup>20</sup> The coloring of the amino acids corresponds to the clustal color scheme as implemented by Jalview. The portion of the alignment extending beyond the end of the shortest sequence (BA425) is not shown for clarity.

(see previous section and Figure 6A and C), but may also be implicated in determining the enzymes' substrate specific-

ity. Clearly, the specificity-related clause of the previous sentence may appear as an overinterpretation of the data, especially when it is partly based on something as inherently inconclusive as a multiple sequence alignment. It could be argued, for example, that accumulation of insertions/deletions (and the corresponding low conservation scores) is exactly what we would expect from surface exposed loops with no functional or structural importance (and, thus, with low pressure from natural selection). Although this would be an otherwise valid argument, we find it hard to reconcile this view with the image of the dynamics of these loops as seen in Figure 6B and C. Indeed, it appears highly unlikely that of all surface exposed loops present in the *BcZBP* structure, only those loops surrounding the active sites have no functional importance and thus escape the pressure of natural selection. It should be noted, however, that in the absence of solid experimental evidence in the form of a crystallographically determined structure of an enzyme–substrate complex, it is impossible to take this analysis much further. This is more so given the absence of firm knowledge concerning the specific substrate of *BcZBP*.

## 4. Discussion

We have performed a state-of-the-art molecular dynamics simulation of *BcZBP* in explicit solvent and with full electrostatics. Analysis of the resulting trajectory showed not only that the simulation per se was very stable but also that the overall structure of the enzyme was very well preserved with an average rms deviation at the monomers' level of approximately 0.80 Å. Analysis of the variance-covariance matrix showed that the crystal structure-based view of the enzyme as a trimer of dimers does not convey the complex dynamics observed in the simulation and indicated that even a dimerization motif as strong and explicit as a β-strand-exchange, can show surprising plasticity with respect to protein dynamics. Analysis of the pattern of atomic mobility and fluctuations identified three hyper-mobile regions of the *BcZBP* structure corresponding to the three loops surrounding each of the hexameric enzyme's active sites. Examination of their mobility clearly indicated that at least in the case of the apoenzyme, these loops are directly implicated in determining the active site accessibility. Comparison of the molecular dynamics results with the indications obtained from a multiple sequence alignment with the *B. anthracis* homologues, led to the hypothesis that these three active-site loops may be implicated in determining the enzyme's substrate specificity.

**Acknowledgment.** We should like to thank the Research Center of Scientific Simulation of the University of Ioannina for provision of computational resources.

## References

- (1) Read, T. D.; Peterson, S. N.; Tourasse, N.; Baillie, L. W.; Paulsen, I. T.; Nelson, K. E.; Tettelin, H.; Fouts, D. E.; Eisen, J. A.; Gill, S. R.; Holtzapfle, E. K.; Okstad, O. A.; Helgason, E.; Rilstone, J.; Wu, M.; Kolonay, J. F.; Beanan, M. J.; Dodson, R. J.; Brinkac, L. M.; Gwinn, M.; DeBoy, R. T.; Madpu, R.; Daugherty, S. C.; Durkin, A. S.; Haft, D. H.; Nelson, W. C.; Peterson, J. D.; Pop, M.; Khouri, H. M.; Radune, D.; Benton, J. L.; Mahamoud, Y.; Jiang, L.; Hance, I. R.; Weidman, J. F.; Berry, K. J.; Plaut, R. D.; Wolf, A. M.; Watkins, K. L.; Nierman, W. C.; Hazen, A.; Cline, R.; Redmond, C.; Thwaite, J. E.; White, O.; Salzberg, S. L.; Thomason, B.; Friedlander, A. M.; Koehler, T. M.; Hanna, P. C.; Kolsto, A. B.; Fraser, C. M. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* **2003**, *423*, 81–86.
- (2) Ivanova, N.; Sorokin, A.; Anderson, I.; Galleron, N.; Candelon, B.; Kapatral, V.; Bhattacharyya, A.; Reznik, G.; Mikhailova, N.; Lapidus, A.; Chu, L.; Mazur, M.; Goltsman, E.; Larsen, N.; D'Souza, M.; Walunas, T.; Grechkin, Y.; Pusch, G.; Haselkorn, R.; Fonstein, M.; Ehrlich, S. D.; Overbeek, R.; Kyriides, N. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* **2003**, *423*, 87–91.
- (3) Psylinakis, E.; Boneca, I. G.; Mavromatis, K.; Deli, A.; Hayhurst, E.; Foster, S. J.; Varum, K. M.; Bouriotis, V. Peptidoglycan N-acetylglucosamine deacetylases from *Bacillus cereus*, highly conserved proteins in *Bacillus anthracis*. *J. Biol. Chem.* **2005**, *280*, 30856–30863.
- (4) Fadouloglou, V. E.; Kotsifaki, D.; Gazi, A. D.; Fellas, G.; Meramveliotaki, C.; Deli, A.; Psylinakis, E.; Bouriotis, V.; Kokkinidis, M. Purification, crystallization and preliminary characterization of a putative LmbE-like deacetylase from *Bacillus cereus*. *Acta Crystallogr.* **2006**, *F62*, 261–264.
- (5) Fadouloglou, V. E.; Deli, A.; Glykos, N. M.; Psylinakis, E.; Bouriotis, V.; Kokkinidis, M. Crystal structure of the *BcZBP*, a zinc-binding protein from *Bacillus cereus*: Functional insights from structural data. *FEBS J.* **2007**, *274*, 3044–3054.
- (6) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (7) Humphrey, W.; Dalke, A.; Schulten, K. VMD—Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (8) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics Studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (9) Stote, R. H.; Karplus, M. Zinc binding in proteins and solution: A simple but accurate nonbonded representation. *Proteins* **1995**, *23*, 12–31.
- (10) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald. An Nlog (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (11) Brünger, A. T. Management of Trajectories. In *X-PLOR*; version 3.1, Yale University Press: New Haven, CT, 1992; pp 143–158.
- (12) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7*, 306–317.
- (13) Ichijo, T.; Karplus, M. Collective motions in proteins: A covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **1991**, *11*, 205–217.

- (14) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993**, *17*, 412–425.
- (15) Mu, Y.; Nguyen, P. H.; Stock, G. Energy landscape of a small peptide revealed by dihedral angle principal component analysis. *Proteins* **2005**, *58*, 45–52.
- (16) Altis, A.; Nguyen, P. H.; Hegger, R.; Stock, G. Dihedral angle principal component analysis of molecular dynamics simulations. *J. Chem. Phys.* **2007**, *126*, 244111.
- (17) Glykos, N. M. Carma: A molecular dynamics analysis program. *J. Comput. Chem.* **2006**, *27*, 1765–1768.
- (18) Dal Peraro, M.; Spiegel, K.; Lamoureux, G.; Vivo, M. D.; DeGrado, W. F.; Klein, M. L. Modeling the charge distribution at metal sites in proteins for molecular dynamics simulations. *J. Struct. Biol.* **2007**, *157*, 444–453.
- (19) Notredame, C.; Higgins, D. G.; Heringa, J. T-coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.
- (20) Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J. The Jalview Java alignment editor. *Bioinformatics* **2004**, *20*, 426–427.

CT9002338