

Near Native-State Conformational Landscape of Psychrophilic and Mesophilic Enzymes: Probing the Folding Funnel Model

Paolo Mereghetti,[†] Laura Riccardi,[†] Bjørn Olav Brandsdal,[‡] Piercarlo Fantucci,[†] Luca De Gioia,[†] and Elena Papaleo^{*,†}

Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza, 2, 20126 Milan, Italy, and The Norwegian Structural Biology Centre and Centre for Theoretical and Computational Chemistry, Department of Chemistry, University of Tromsø, N-9037 Tromsø, Norway

Received: December 4, 2009; Revised Manuscript Received: April 2, 2010

In recent years, increased interest has been directed to the study of enzyme adaptation to low temperatures. In particular, a peculiar folding funnel model was proposed for the free energy landscape of a psychrophilic α -amylase and other cold-adapted enzymes. In the present contribution, the comparison between the near native-state dynamics and conformational landscape in the essential subspace of different cold-adapted enzymes with their mesophilic counterparts, as obtained by more than 0.1 μ s molecular dynamics simulations at different temperatures, allows the folding funnel model to be probed. Common characteristics were highlighted in the near native-state dynamics of psychrophilic enzymes belonging to different enzymatic families when compared to the mesophilic counterparts. According to the model, a cold-adapted enzyme in its native-state consists of a large population of conformations which can easily interconvert and result in high structural flexibility.

Introduction

The relationship between low thermal stability, high flexibility, and high catalytic activity at low temperatures in cold-adapted enzymes has attracted attention in the last years.^{1–12} Such enzymes have interesting properties not only for fundamental research but also for their biotechnological relevance, being potential targets for industrial applications.^{13–15}

Recently, the accumulation of both sequence and structural data for cold-adapted proteins, as well as their mesophilic and thermophilic homologues, has made it possible to study structural properties related to thermal stability and flexibility of psychrophilic proteins. In particular, nanosecond molecular dynamics simulations were successfully used to investigate the dynamical and structural properties of cold-adapted enzymes.^{16–25}

Dynamical and functional properties of proteins are strongly related to the fluctuations between different local minima on its free energy landscape (FEL). Understanding the shape and how a native protein explores its free energy landscape, in terms of both the potential energy and the entropic contributions, is an important requirement for a complete microscopic description of its function and of the interplay between folding, stability, and dynamic properties.^{26–28} In protein structures, the shape and the ruggedness of the free energy landscape is the result of fulfilling physical and evolutionary constraints.^{29,30}

In an equilibrated thermodynamic system at constant pressure, the free energy can be estimated from the probability density function of one or more reaction coordinates.³¹ Since it is not possible nor helpful to represent the free energy landscape as a function of $3N - 6$ coordinates, an approach to select or identify the most important degrees of freedom is necessary.^{32,33}

For small proteins, several strategies for describing and visualizing the free energy landscape have been developed, such

as disconnectivity graphs (DG).³⁴ However, DG methodologies can only be usefully applied to small protein systems. An alternative to investigate large protein systems close to their native-state, for which a full sampling of the configurational space is not necessary, is a qualitative representation of the FEL, using few reaction coordinates. For example, principal component analysis (PCA)^{35–37} is one of the most popular methods for reducing the dimensionality of a complex system and can be used to derive useful reaction coordinates for free energy landscape analyses.^{36,37}

On the basis of kinetic, biophysical, and structural data, a folding funnel model has been proposed for the free energy landscape of the psychrophilic α -amylase from *Pseudoalteromonas haloptanktis* and other cold-adapted enzymes.³⁸ In this model, a cold-adapted enzyme in its native-state consists of a large population of conformations separated by low energy barriers. This type of energy landscape allows conversion between conformations yielding high structural flexibility. It would be insightful to test this model by comparing the near native-state free energy landscape of cold-adapted enzymes with their mesophilic counterpart, belonging to different families.

In the present contribution, to shed light on the molecular features responsible for cold adaptation in psychrophilic enzyme, we have performed comparative molecular dynamics studies of mesophilic and psychrophilic variants belonging to two different enzymatic families, for which descriptions of protein flexibility^{16,18,24} and biochemical characterization^{39–41} are available in the literature: serine proteases and uracil DNA glycosylases. Due to the size of the selected systems, molecular dynamics simulation is the most suitable computational method to explore the conformational space and allows the consideration of proper dynamical properties. In particular, to ensure an efficient and accurate sampling, multiple molecular dynamics simulations^{42–45} were carried out in explicit solvent at 283 and 310 K, close to the optimal growth temperatures for the organisms, collecting more than 0.1 μ s trajectories. For each enzyme, the resulting ensemble was analyzed estimating and

* To whom correspondence should be addressed. Phone: +39 02 64 48 34 75. E-mail: elena.papaleo@unimib.it.

[†] University of Milano-Bicocca.

[‡] University of Tromsø.

comparing the distribution of the different substates in the proximity of the native-state and providing a qualitative description of free energy landscapes by different reaction coordinates. Moreover, the configurational entropy,^{46–48} estimated by using the Schlitter equation, was also analyzed. Common features were identified in the near native-state conformational distribution of psychrophilic enzymes, probing the folding funnel model proposed for AHA³⁸ and strongly enforcing the current view of the stability/flexibility relationship in cold-adapted enzymes.

Materials and Methods

Molecular Dynamics Simulations. MD simulations were performed with the 3.3 version of the GROMACS software (www.gromacs.org), implemented on a parallel architecture with the GROMOS96 forcefield. The X-ray structures of two mesophilic and psychrophilic proteins were used as starting points for the MD simulations. In particular, a mesophilic (from *Sus scrofa*, mPE, pdb entry 1LVY⁵⁰) and a psychrophilic (from *Salmo salar*, pSE, pdb entry 1ELT⁵¹) elastase, as well as a mesophilic (from *Homo sapiens*, mHUDG, pdb entry 1AKZ⁵²) and a psychrophilic (from *Gaudus morhua*, pCUDG, pdb entry 1OKB⁵³) uracil DNA glycosylase were selected.

Protein structures, including the crystallographic water molecules and calcium ions for elastases, were soaked in a dodecahedral box of SPC water molecules⁵⁴ and simulated by using periodic boundary conditions. Further details on the MD setup and solvent equilibration are in ref 24. For UDGs all the histidines with one exception (His148) were considered as neutral in the simulations, as explained in refs 22 and 23.

Productive MD simulations were performed in the NPT ensemble at 283 and 310 K, using an external bath with a coupling constant of 0.1 ps. Pressure was kept constant (1 bar) by modifying box dimensions⁵⁵ and the time-constant for pressure coupling was set to 1 ps. The LINCS⁵⁶ algorithm was used to constrain bond lengths, allowing the use of a 2 fs time step. van der Waals and Coulomb interactions were truncated at 1.0 nm. Long range electrostatic interactions were calculated by using the Particle-mesh Ewald (PME)⁵⁷ summation scheme. The nonbonded pair list was updated every 10 steps and conformations were stored every 2 ps.

To improve the conformational sampling, ten/twelve 12 ns simulations were carried out at 283 K/310 K, respectively, initializing the MD runs with different initial atomic velocities taken from a Maxwell distribution. In the following, MD trajectories collected for the same system but characterized by different initial velocities are referred to as *replica 1* to *replica 12*.

The root-mean-square deviation (rmsd), which is a crucial parameter in order to evaluate the equilibration of MD trajectories, was computed for main-chain atoms by using the starting structures of the MD simulations as references. The analysis of MD trajectories has been carried out discarding the first 2 ns for each simulation, time which is necessary to reach stable values for main-chain rmsd (data not shown). Investigation of elastase structures during the simulation time shows that the coordination of the calcium ions, which is important for elastase stability,⁶ is maintained throughout the simulations in agreement with data previously reported.¹⁸ Moreover, potential and total energy of the system, as well as the protein gyration radius are constant throughout the simulations (data not shown).

For each protein, the stable regions of the ten *replicas* at the same temperature were joined in a concatenated trajectory,

which is representative of different directions of sampling around the starting structure.

Essential Dynamics/Quasiharmonic Analysis. The all-atoms covariance matrix (\mathbf{C}) was calculated on the equilibrated portions of the trajectories. In particular \mathbf{C} was calculated considering both the concatenated trajectories and the single *replicas* for each system at both 283 and 310 K:

$$\mathbf{C} = \text{cov}(x) = \langle (x - \langle x \rangle)(x - \langle x \rangle)^T \rangle$$

where $\langle \rangle$ indicates the average and x is the vector of the atomic positions.

After removal of the translational and rotational degrees of freedom (fitting each structure onto the initial one), the matrix \mathbf{C} was calculated and diagonalized to obtain the eigenvectors and eigenvalues, which give information about correlated motions throughout the protein.

To define the dimensionality of the essential subspace, the fraction of total motion described by the reduced subspace was computed as the sum of the eigenvalues relative to the included eigenvectors. This describes the amount of variance retained by the reduced representation of the total space.

A measure of the similarity of a MD trajectory to random diffusion is the cosine content (c_i) of the p_i principal component:⁵⁸

$$c_i = \frac{2}{T} \left(\int \cos(i\pi t) p_i(t) dt \right)^2 \left(\int p_i^2(t) dt \right)^{-1}$$

where T is the total simulation time. c_i is an absolute measure that can be extracted from covariance analysis and ranges between 0 (no cosine) and 1 (a perfect cosine). It has been demonstrated that insufficient sampling can lead to high c_i values, representative of random motions.⁵⁸ The evaluation of the cosine contribution for first eigenvectors is sufficient to give a reliable idea of the protein behavior. When the cosine content of the first few principal components is close to 1, the largest scale motions in the protein dynamics resemble diffusion, and cannot be interpreted in terms of characteristic features of the energy landscape.^{59,60}

The analysis of the sampling convergence can be performed by computing the root-mean-square inner product (RMSIP) as a measure of similarity between subspaces defined by their basis vectors:⁶¹

$$\text{RMSIP} = \frac{1}{D} \sum_{i=1}^D \sum_{j=1}^D (\eta_i^A \eta_j^B)$$

where η_i^A and η_j^B are the eigenvectors of the spaces to be compared. Usually the RMSIP is computed onto the first 10 eigenvectors.^{61,62} The statistical significance of the observed RMSIP value was tested by simulating an empirical distribution of RMSIP data under the null hypothesis of no relationship between both spaces.⁶² In particular, the dependence of the RMSIP on the dimensionality of the spaces can be expressed as $\text{RMSIP}(K) = D^{1/2}/K^{1/2}$, where D and K indicate the number of eigenvectors considered in the computation of RMSIP and the dimensionality of the two spaces, respectively. The expected RMSIP values were obtained by an empirical distribution of RMSIP values considering the 10 eigenvectors of random (normal distributed) orthogonal matrices of different size (from

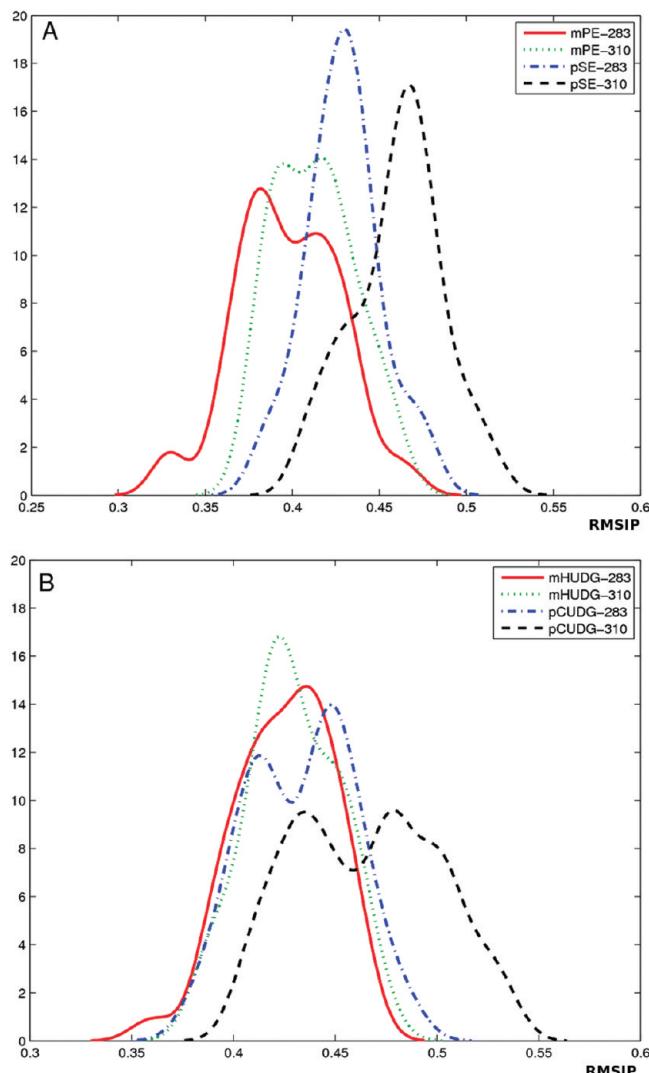


Figure 1. RMSIP distributions computed for *mPE* and *pSE* (A) and *mHUDG* and *pCUDG* (B) comparing all *replicas* relative to the same system at the same temperature. To obtain smoothed distributions, the Gaussian density estimation has been applied.

10 to 500). Comparing these results with the RMSIP distribution of our systems it is possible to assess the significance of our RMSIP.⁶²

Free Energy Landscapes. Given a reaction coordinate q_α , the probability of finding the system in a particular state q_α is proportional to $(e^{-G(q_\alpha)/kT})$, where $G(q_\alpha)$ is the free energy of that state.

The free energy landscape can be computed from $G(q_\alpha) = -kT \ln[P(q_\alpha)]$, where k is the Boltzmann constant, T is the

temperature of the simulation, and $P(q_\alpha)$ is an estimate of the probability density function obtained from a histogram of the data.

To ensure that for the lowest free energy minimum the $\Delta G = 0$ we have subtracted the maximum probability $\Delta G(q_\alpha) = -kT [\ln P(q_\alpha) - \ln P^{\max}(q)]$.

Considering two different *reaction coordinates* q, p , the two-dimensional free energy landscapes were obtained from the joint probability distributions $P(q,p)$ of the considered variables.^{37,46,47,63} In particular, the *reaction coordinates* investigated were the C α rmsd using the starting structures of the simulations as references, the radius of gyration (R_g), and the Cartesian principal components derived by ED analysis, as previously described. During simulations, conformations were saved every 2 ps. Conformations collected with this small time step might be correlated, affecting the equilibrium distribution of conformational states. To exclude this possibility, the autocorrelation function of rmsd and radius of gyration of each *replica* have been computed, founding an average decay of 200 ps. Then, the probability distributions with trajectories sampled every 200 ps were computed and compared, by superposition, to the probability distributions resulting from the original meta-trajectories, sampled at 2 ps. The results are comparable and no relevant difference between the distributions has been detected (data not shown), therefore 2 ps sampled trajectories have been selected for the analysis.

Cluster Analysis. Clustering on Cartesian coordinates was performed by computing the root-mean-square distance matrix calculated for C α atoms between each pair of structure. Therefore, the complete linkage algorithm implemented in Matlab 2007a was applied to this distance matrix obtaining a dendrogram.

Configurational Entropy. The configurational entropy was computed by using the Schlitter equation,^{49,64,65} which provides an approximation to the quantum mechanical (QM) entropy in the classical limit⁶⁶ and yields an upper bound to the true entropy.

$$S_{\text{true}} < S = \frac{1}{2} k_B \ln \det \left[1 + \frac{k_B T e^2}{\hbar^2} M \sigma \right]$$

where k_B is the Boltzmann constant, T is the absolute temperature, e is the Euler's number, $\hbar = h/2\pi$, M is the diagonal matrix of the masses, and σ is the all atoms covariance matrix.

Several approximations are used into the computation of the entropy:^{46,48,67,68} (i) every degree of freedom is treated as a quantum harmonic oscillator; (ii) the equipartition theorem is used to connect the classical variance to the frequency of a quantum harmonic oscillator (this relation holds for $\hbar \omega \ll k_B T$,

TABLE 1: Average RMSIP Values (μ RMSIP) and Standard Deviation (σ RMSIP)

system	μ RMSIP ^a	σ RMSIP ^a	μ random RMSIP ^b	σ random RMSIP ^b	no. of atoms ^b	no. of eigenvalues ^b
mPE, 283 K	0.39	0.023	0.0375	0.005	2349	7047
mPE, 310 K	0.41	0.023	0.0375	0.005	2349	7047
pSE, 283 K	0.43	0.021	0.0382	0.005	2259	6777
pSE, 310 K	0.46	0.025	0.0382	0.005	2259	6777
mHUDG, 283 K	0.42	0.023	0.0376	0.005	2334	7002
mHCDG, 310 K	0.43	0.022	0.0376	0.005	2334	7002
pCUDG, 283 K	0.43	0.026	0.0379	0.005	2297	6891
pCUDG, 310 K	0.46	0.029	0.0379	0.005	2297	6891

^a Average and standard deviation computed on RMSIP distribution for each system. ^b Expected values and standard deviations of empirical RMSIP distributions of unrelated spaces.

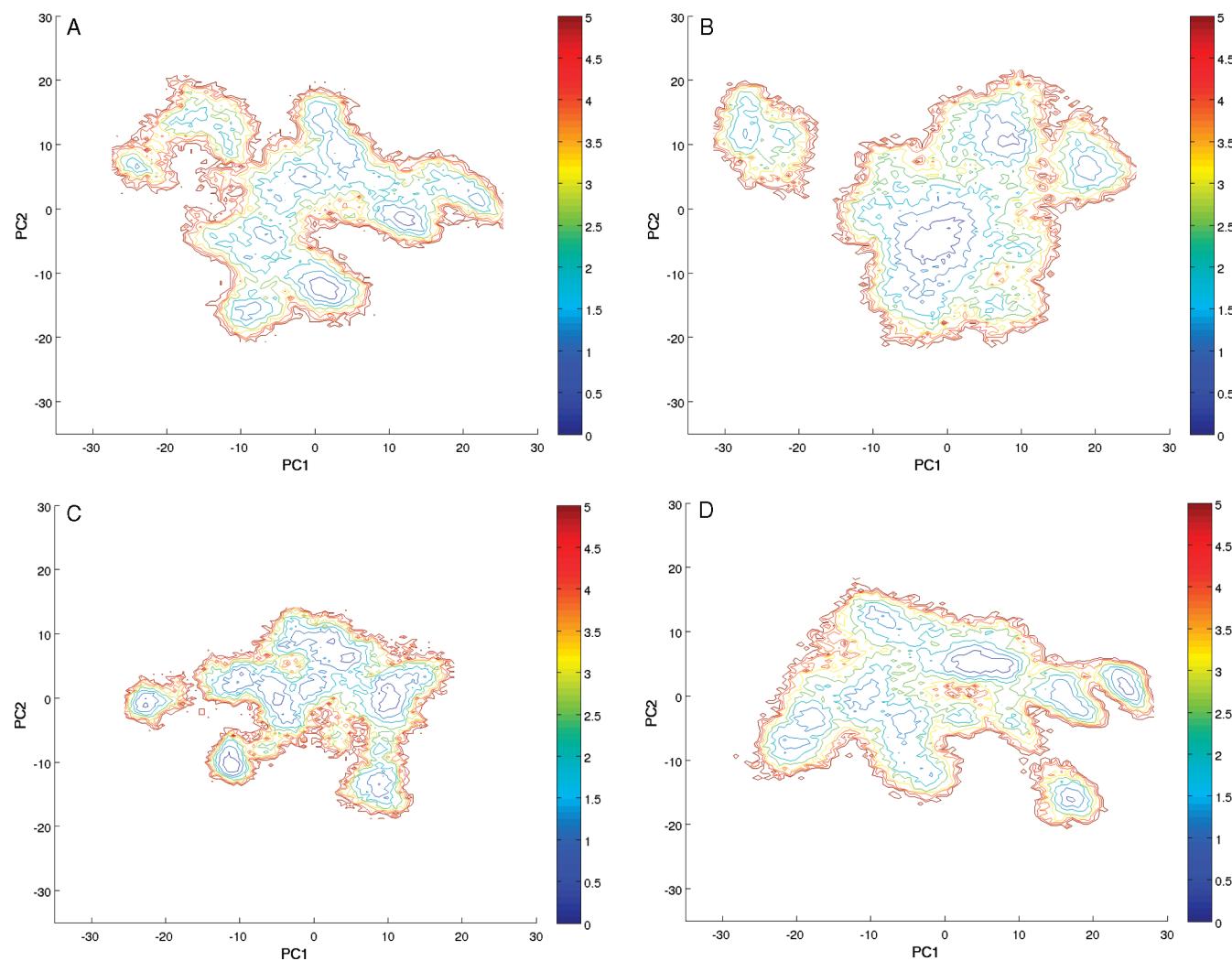


Figure 2. Contour plot representation of the FEL of *mPE* (A, B) and of *pSE* (C, D) at 283 K (A, C) and at 310 K (B, D). The FEL is projected onto the first two principal components of the all-atoms mass-weighted covariance matrix of the concatenated trajectories. The free energy is given in kJ/mol and indicated by the color bar.

which is a good approximation, since the high frequency of motion for which it fails will contribute little to the entropy); and (iii) absence of supralinear correlation between different coordinates.

To obtain a statistically significant comparison among the different systems, the entropy was computed on each *replica*. Then an average value and a standard deviation were obtained. The averages were compared by using a paired *t* test.

Results and Discussion

Evaluation of the Conformational Sampling. MD simulations of mesophilic and psychrophilic elastases and uracil DNA glycosylases were carried out at 283 and 310 K (see Materials and Methods). After concatenation of the equilibrium portions of the trajectories, the resulting MD ensembles consisted of 0.1 μ s trajectories for each system at both temperatures. To gain insights into the configurations visited by the system and to evaluate the conformational sampling, essential dynamics analysis (see Materials and Methods) was carried out, with particular attention to the direction of motion along the first eigenvectors. In fact, the first three eigenvectors are sufficient to describe a consistent part of the total motion, and the subspace defined by them could be used as the three-dimensional (3D) reference subspace to analyze protein dynamics.^{24,35,45} The

projections of simulation frames in the 3D-reference subspace showed a wide sampling of the conformational space with resampling of similar conformations in our simulations, indicating that the essential subspace is well explored when concatenated trajectories are considered (data not shown).

To further evaluate the sampling efficiency, we have also computed the cosine content (c_i) of the principal components of protein motion, which is a measure of the similarity of the trajectories to random diffusion (Figure 1S, Supporting Information). It turned out that single *replica* are often characterized by relatively large c_i in the first eigenvectors, and therefore partly describe a random diffusion motion, while the corresponding concatenated trajectories have lower or null cosine content and therefore adequately represent essential and significant motions.

Another measure used for the convergence assessment is the root-mean-square inner product (RMSIP). For each protein system (*mPE*, *pSE*, *mHUDG*, *pCUDG*) the RMSIP was computed comparing all replicas, obtaining a distributions of RMSIP values (Figure 1). The average values of the distributions are summarized in Table 1. The significance of the RMSIP values obtained from our simulations has been verified by comparing them to the expected RMSIP value of two unrelated spaces (see Materials and Methods). We found that expected RMSIP depends on the dimensionality of the considered space and on

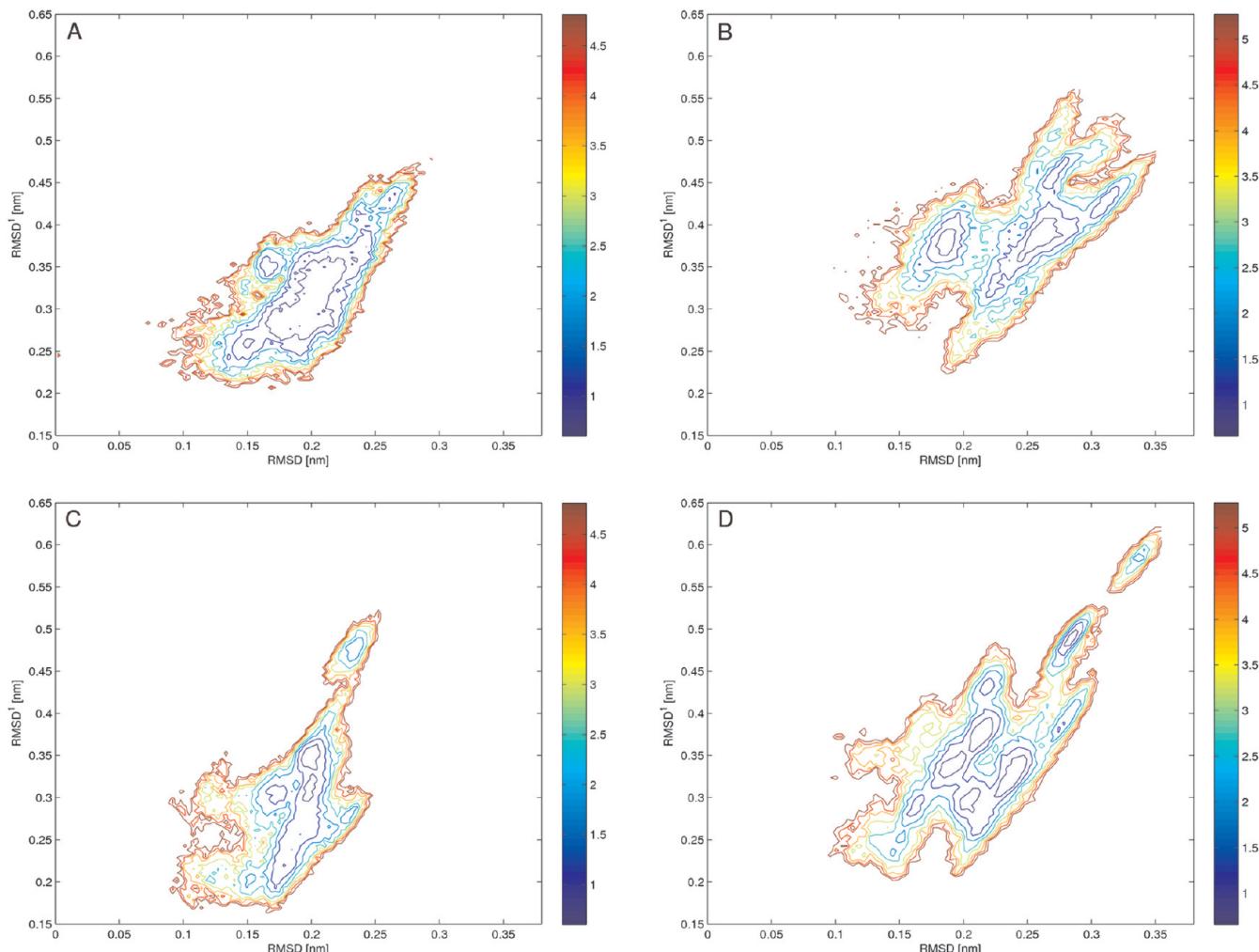


Figure 3. Contour plot representation of the FEL of *mPE* (A, C) and of *pSE* (B, D) at 283 K (A, B) and at 310 K (C, D), using as reaction coordinates the $\text{C}\alpha$ rmsd and the “local” $\text{C}\alpha$ rmsd, computed only considering the residues characterized by high flexibility in the psychrophilic enzyme.^{17,24} The free energy is given in kJ/mol and indicated by the color bar.

the subset of eigenvectors included in the RMSIP calculation (Figure 2S, Supporting Information). Considering the dimensionality of our system (ranging from 2259 to 2349 atoms), we found that if the set of principal components were completely unrelated, the expected RMSIP value would be ~ 0.038 (Table 1). The average RMSIP obtained for our simulations (Table 1) deviates significantly from this random reference and presents a sufficient overlap between different *replicas* of the same system.

In conclusion, all the indexes analyzed by ED indicate that a wide conformational sampling was reached in our simulations and allow the evaluation of the distribution of the conformational substates close to the native-state, by using both a qualitative description of the free energy landscape and structural cluster analysis of the mesophilic and psychrophilic enzymes.

Conformational Landscape of Cold- and Warm-Adapted Enzymes. The study of FEL can give an accurate picture of structural properties around the native-state of a protein.^{27,37,46,47,62} However, in order to understand the behavior of complex systems it is necessary to project it onto low dimensional subspace of physically meaningful coordinates. The choice of the reaction coordinates is a system-related task and is usually driven by the features that have to be depicted. In particular only the degrees of freedom directly related to the properties of interest should be included in the analysis to prevent masking of important information.³⁴

Since many different FEL representations can be obtained by using different combinations of collective variables, several properties have been selected as reaction coordinates in our study: the radius of gyration (R_g), the root-mean-square deviation (rmsd), and the Cartesian principal components. The Cartesian principal components analysis, in particular, is a frequently used method for obtaining collective coordinates, which can be used to project the configurational landscape of proteins.⁶⁹

It is worth pointing out that the FELs obtained as explained above lack barrier information and reflect the overall shape of the free energy surface and not necessarily its details.^{70,71} The absence of barriers might somehow be compensated by the empty space between the sampled regions: poorly sampled or unsampled regions often correspond to high energy regions.⁷⁰ Moreover, because of the large dimensionality reduction due to the projection onto few collective coordinates, these maps may represent an incomplete description of the free energy profile of the protein. This lack of information can be partially complemented by means of a cluster analysis, representing the geometrical relationship in a multidimensional way.⁷¹

In light of the above observations, we carried out the analysis of the distribution of the conformational states and the corresponding FEL using different reaction coordinates, as well as structural clusters analysis considering the Cartesian coordinates (Materials and Methods). Moreover, to verify the conservation of the different shape and properties of the conformational

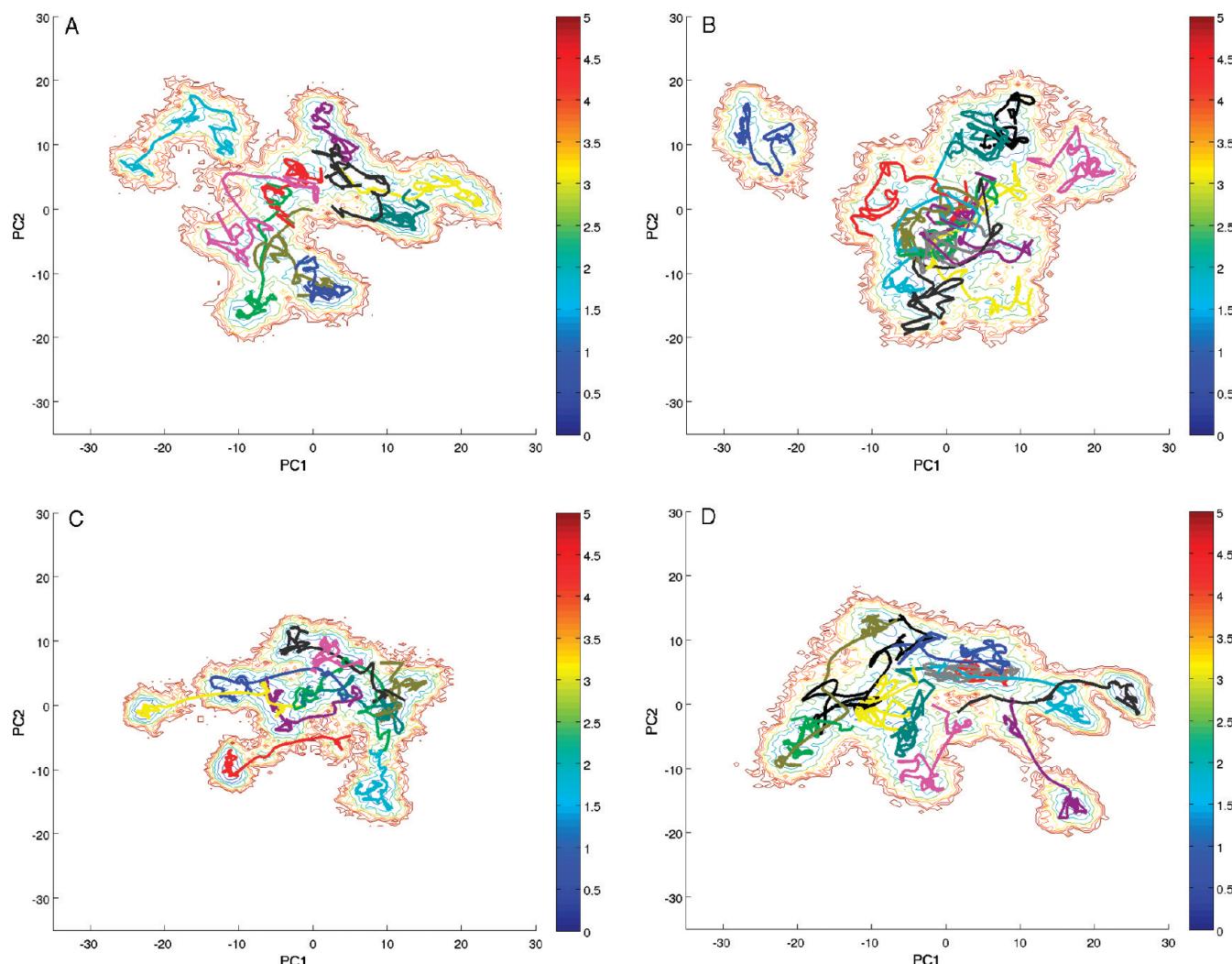


Figure 4. Projection of single trajectories onto the first two principal components of the all-atoms mass-weighted covariance matrix. For the sake of clarity, the trajectories are smoothed by using a Bezier interpolation and drawn over the contour plot representation of the principal component FEL of *mPE* (A, B) and *pSE* (C, D) at 283 K (A, C) and at 310 K (B, D). The free energy is given in kJ/mol and indicated by the color bar.

landscape of psychrophilic and mesophilic enzymes, the FEL at 310 K has been calculated both on a concatenated trajectories including 10 *replicas* (Figure 3S, Supporting Information) and a concatenated trajectory collecting all the available 12 *replicas* (Figures 4b,d and 8d). The two model systems (elastases and uracyl DNA glycosylases) are discussed separately below.

Conformational Landscape of Mesophilic and Psychrophilic Elastases. When rmsd and R_g were used as collective coordinates, the psychrophilic elastase (Supporting Information, Figure 4Sc) shows a more compact FEL compared to the mesophilic homologue at 283 K (Supporting Information, Figure 4Sa). At 310 K, the landscape of the mesophilic enzyme (Supporting Information, Figure 4Sb) splits into two relative minima whereas the psychrophilic enzyme retains mainly a single minimum (Supporting Information, Figure 4Sd).

In fact, global reaction coordinates ($C\alpha$ /main-chain rmsd and R_g) are unable to distinguish among different conformational substates (Supporting Information, Figure 4S) and therefore the free energy landscape of elastases was analyzed in further detail with use of the Cartesian principal components (Figure 2). The same trend of Figure 4S (Supporting Information) can be highlighted for *mPE* (Figure 2a,b), which samples a wider conformational space than *pSE* at 283 K (Figure 2c), and, more

clearly, at 310 K (Figure 2b,d). Moreover it is possible to observe that the psychrophilic enzyme presents a greater number of metastable states at both temperatures (Figure 2c,d), whereas the mesophilic counterpart shows a reduction of the number of relative minima moving from 283 K to 310 K. This is clearly visible from the probability density plots (Supporting Information, Figure 5S), in which the mesophilic enzyme at 310 K depicts two highly populated states, whereas the psychrophilic surface is characterized by several and populated conformational states.

The wider FEL of the mesophilic enzyme could be a consequence of its higher global flexibility,^{17,24} whereas the psychrophilic enzyme shows a smaller surface characterized by many metastable states which may reflect the differences in a specific portion of the protein. Indeed, some loops of *pSE*, which are localized in the proximity of the catalytic site, are associated to a greater flexibility compared to the equivalent regions onto the mesophilic protein.^{17,24} By projecting the free energy landscape onto the rmsd of these regions (Figure 3), it is possible to discriminate different metastable states of the psychrophilic proteins which are not observable considering the global rmsd. This means that the psychrophilic enzyme interconverts between many substates characterized by different conformations of the loops near the catalytic site.

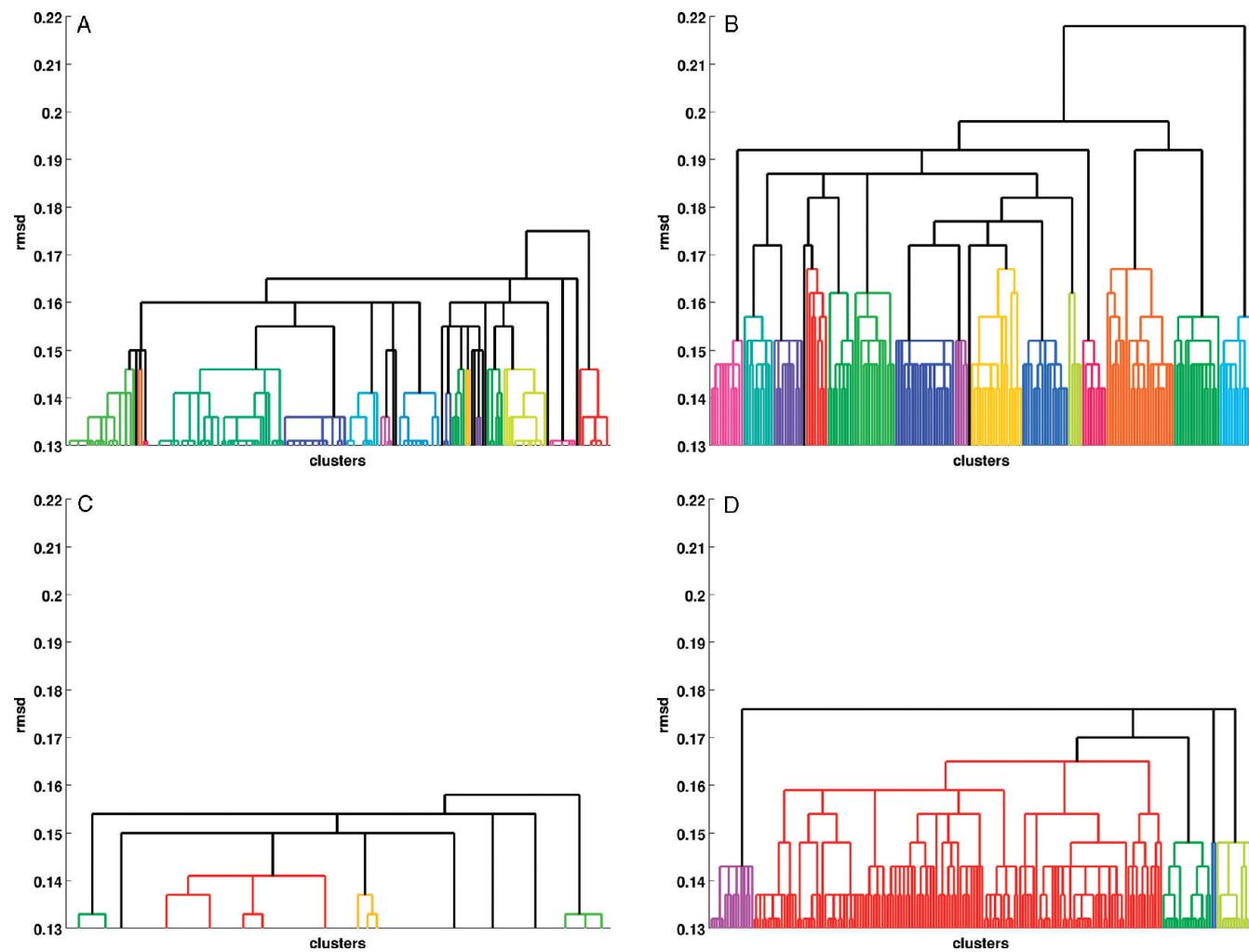


Figure 5. Dendograms of the complete linkage cluster analysis of the C α rmsd distance matrix of *mPE* (A, B) and of *pSE* (C, D) at 283 K, C) and at 310 K (B, D) are shown. Different clusters are highlighted using a color threshold of 0.17 nm at 310 K and 0.15 nm at 283 K.

Despite the scarce information about kinetic barriers among the different metastable states, following the equilibrium portion of the trajectories projected onto the first two eigenvectors, we can observe how easily the enzyme switches between different basins during the trajectory. In most of the trajectories at both 283 and 310 K, the psychrophilic enzyme explores several conformational basins (Figure 4c,d), whereas the mesophilic homologue tends to be trapped in the main basin (Figure 4a,b). To further analyze the differences in flexibility, which are suggested by differences in the FEL wideness of the two enzymes, an estimation of the configurational entropy was carried out.⁶⁵

Detailed results, summarized in Table 2, show that the entropy of *mPE* is higher than that for *pSE* at both temperature. The average entropy differences between the *mPE* and *pSE* are 0.836 kJ/Kmol at 283 K and 0.829 kJ/Kmol at 310 K. The *P* values associated to these differences are much lower than the generally adopted level of significance ($\alpha = 0.01$). Hence, we can conclude that the entropy of the *mPE* is generally greater than the *pSE* value.

In conclusion, the near native conformational landscape, explored during our simulations, in the psychrophilic enzyme seems to be characterized by a narrow shallow basin with a rugged bottom composed by many metastable states. On the other hand, the mesophilic enzyme has a conformational landscape that shows a funnel-like shape, where there is a well-defined basin. These results are confirmed by cluster analysis:

it is clearly visible that there is a pronounced funnel-like structure of the clustering dendrogram of the mesophilic enzyme (Figure 5a-b) compared to the hierarchical pattern of the clustering dendrogram referred to the psychrophilic enzyme (Figure 5c,d).

Conformational Landscape of Mesophilic and Psychrophilic UDGs. Differently from the elastases, the near native FEL of the uracil DNA glycosylases are smoother and flatter (Figure 6, as well as Figure 7S in the Supporting Information). By projecting the near native FEL onto the rmsd and R_g , the landscape results in a single roughly smoothed well (Figure 6S, Supporting Information). Hence, also in this case, these global collective coordinates (rmsd, radius of gyration) are unable to distinguish between different conformational substates and are then useless for the characterization of the shape of the near native FEL. On the other hand, considering the projection onto the first two eigenvectors as collective coordinates, it is possible to distinguish different relative conformational basins in the UDG FEL (Figure 7S, Supporting Information).

*p*CUDG and *m*HUDG are characterized by strongly similar primary sequence identity (87%) and the high similarity is reflected on the near native FEL (Figure 6). The mesophilic enzyme shows a flat and shallow landscape lacking the pronounced funneled-like organization previously shown for the mesophilic elastase (Figure 6a,b; Figure 7Sa,b, Supporting Information). The hierarchical cluster analysis confirms the reduction of the funnel-like distribution of the metastable states

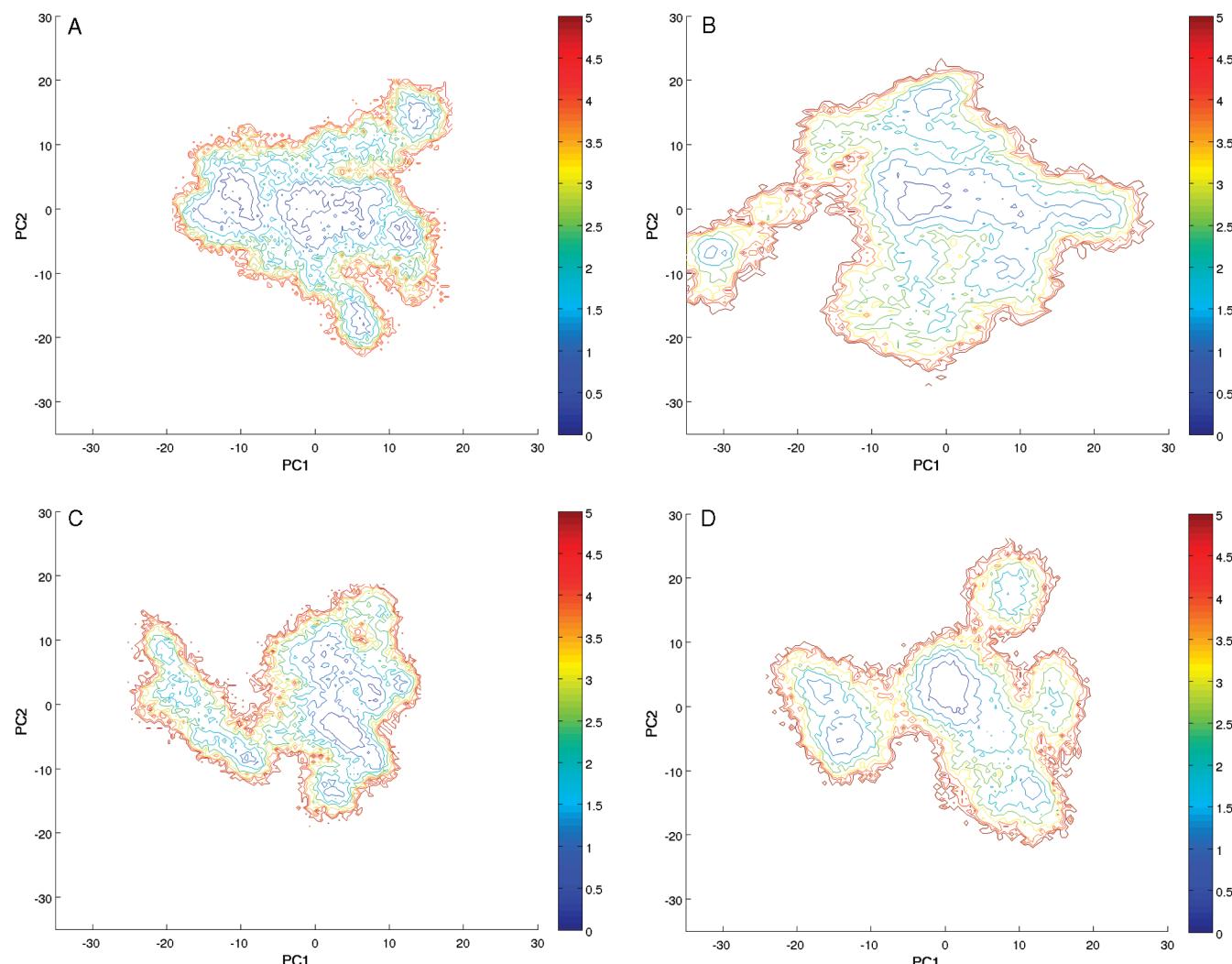


Figure 6. Contour plot representation of the FEL of *m*HUDG (A, B) and of *p*CUDG (C, D) at 283 K (A, C) and at 310 K (B, D). The FEL is projected onto the first two principal components of the all-atoms mass-weighted covariance matrix of the concatenated trajectory. The free energy is given in kJ/mol and indicated by the color bar.

TABLE 2: The Configurational Entropy of Each Replica Estimated by Using the Schlitter Approximation^{46–48a}

T	μ_S^{mPE}	σ_S^{mPE}	μ_S^{pSE}	σ_S^{pSE}	$\mu_S^{mPE} - \mu_S^{pSE}$	P value
283 K	10.763	0.507	9.933	0.135	0.829	9.542×10^{-5}
310 K	12.359	0.161	11.523	0.090	0.836	2.891×10^{-11}
T	μ_S^{mHUDG}	σ_S^{mHUDG}	μ_S^{pCUDG}	σ_S^{pCUDG}	$\mu_S^{mHUDG} - \mu_S^{pCUDG}$	P value
283 K	11.096	0.102	10.841	0.056	0.255	9.401×10^{-7}
310 K	12.824	0.486	12.391	0.381	0.433	0.199×10^{-1}

^a The average and the standard deviation of the entropy (ΔS) is shown here (values are in kJ/Kmol). Moreover the difference between the average values of the psychrophilic and the mesophilic enzyme is shown. The P value for the differences has been computed by using a one-sided paired t-test.

in mesophilic enzyme (Figure 7a,b), while a hierarchical distribution of the states in *p*CUDG is observed (Figure 7c,d), comparable to the cluster distribution in *p*SE.

Following the single trajectories on the FEL representation we can clearly see that they freely move among the different basins, suggesting that the barriers dividing them might be lower than in the mesophilic elastase (Figure 8a,b). On the other hand, the psychrophilic conformational landscape shows several separated metastable states, which are more visible at 310 K

(Figure 8c,d). This landscape partially resembles the psychrophilic elastase landscape with several and separated basins. However, in *p*CUDG the trajectories tend to be trapped more frequently into a single well than in the psychrophilic elastase.

The estimation of the configurational entropy for UDGs indicates a greater conformational entropy of the *m*UDG at both 283 and 310 K (Table 2), in agreement with results for elastases. However, the configurational entropy values at 310 K have to be considered with caution since the associated P value (0.019) is poorly statistically significant (Table 2).

In conclusion, the conformational landscape of the cold-adapted *p*CUDG consists of several and separated different basins. On the other hand, the mesophilic enzyme has a landscape characterized by a slightly pronounced shallow funnel-like shape.

Conclusions

In this study a detailed analysis of conformational landscape near the native-state has been carried out by means of long multiple MD simulations of warm- and cold-adapted enzymes belonging to different families.

On the basis of kinetic, biophysical, and structural data, a folding funnel model was proposed for the free energy landscape of the psychrophilic α -amylase from *Pseudoalteromonas ha-*

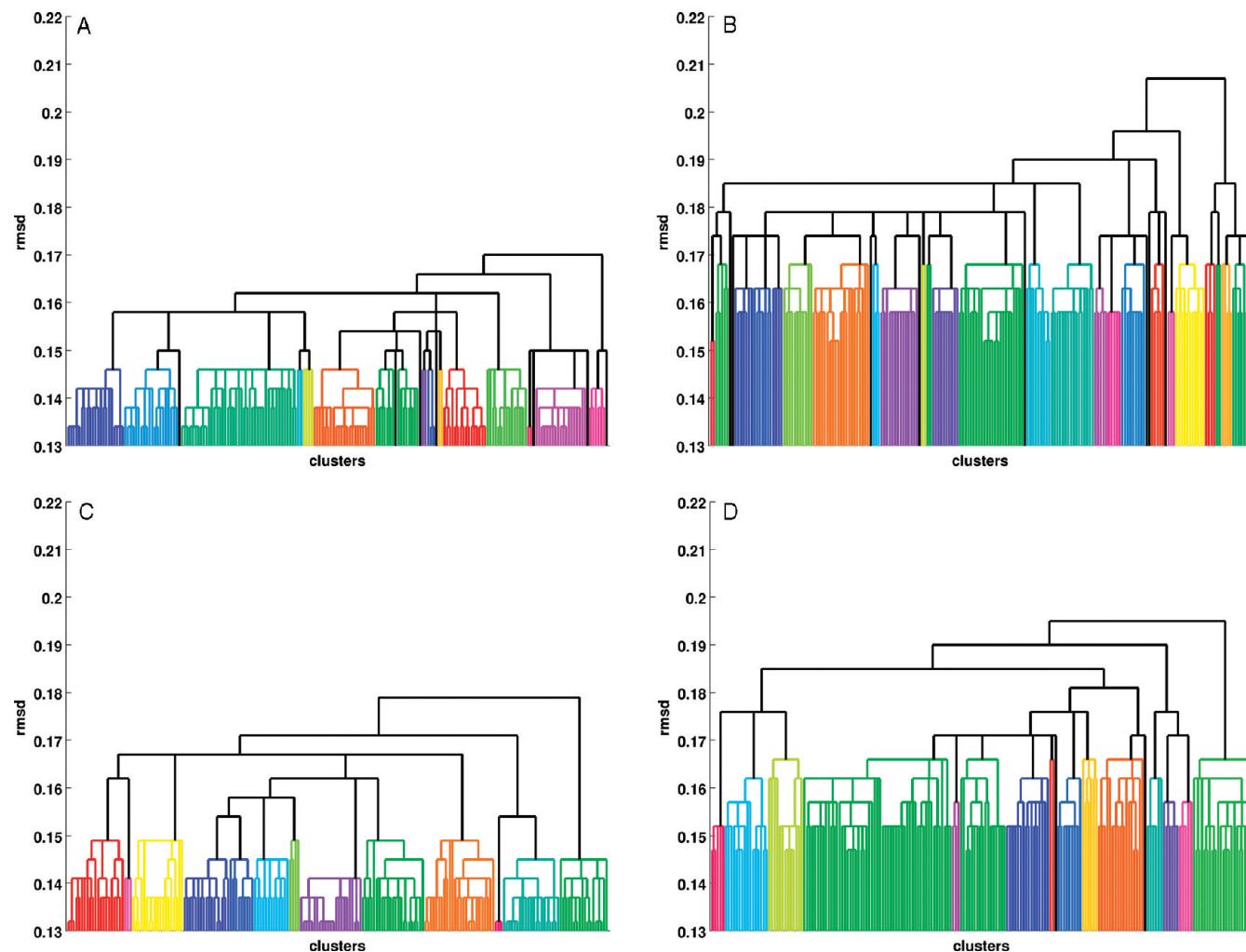


Figure 7. Dendograms of the complete linkage cluster analysis of the $\text{C}\alpha$ rmsd distance matrix of *mUDG* (A, B) and of *pCUDG* (C, D) at 283 K (A, C) and at 310 K (B, D) are shown. Different clusters are highlighted using a color threshold of 0.17 nm at 310 K and 0.15 nm at 283 K.

loptanktis (AHA) and other cold-adapted enzymes³⁸ and recently shown to be consistent with data on zinc metalloproteases of the thermolysin family.⁷² According to the model, a cold-adapted enzyme in its native-state consists of a large population of conformations divided by low energy barriers. This type of energy landscape would promote conversion between conformations and result in high structural flexibility.

In the present contribution, the comparison between conformational landscape close to the native-state of different cold-adapted enzymes with their mesophilic counterparts allows the folding funnel model for AHA proposed by D'Amico et al.³⁸ to be probed.

Using two-dimensional representation of the substates distribution of cold- and warm-adapted enzymes along the first two principal components, we identify more densely populated regions in the conformational subspace. These regions can be considered basins in the conformational landscape, since they suggest a spontaneous tendency of the system to adopt structural conformations associated to those basins.³⁷ While it is important to keep in mind that the totality of the protein dynamics of these systems is not restricted to the basins explored during our simulations, they can represent a significant portion of the native landscape.

In particular, common characteristics were highlighted in the near native-state conformational landscape of psychrophilic enzymes belonging to different enzymatic families when

compared to the mesophilic counterparts. The collected data allow us to propose a simple and qualitative model of the near native state free energy landscape of the psychrophilic and the mesophilic enzymes of the two enzymatic families analyzed here: elastases and uracil DNA glycosylases (Figure 9).

In the case of the elastase enzyme the near native-state free energy landscape is in agreement with the model proposed by D'Amico et al.,³⁸ where the mesophilic enzyme has a funnel-like landscape (Figure 9a), whereas the psychrophilic homologue is characterized by a rugged flat-bottom landscape with low barriers, which favors the interconversion among several metastable states (Figure 9b). Moreover, it is worth mentioning that *pCUDG* and *mHUDG*, even with less pronounced differences, still present characteristics typical of the FEL of cold-adapted and warm-adapted proteins, respectively.

Recently, engineered proteins in conjunction with conformationally gated electron transfer (ET) methods were used to assess the response of the kinetics at the bottom of a folding funnel to global stability.⁷³ It was concluded that the funneled landscape evolved such that loss of global stability lowers barriers at the bottom of a folding funnel, still allowing for efficient folding. They found analogies with the current view on cold-adapted enzymes, which is supported by our computational results, according to which the strongly unstable cold-adapted enzymes present enhanced dynamics, allowing the catalytically active state to still be readily accessible at low temperatures.⁷³

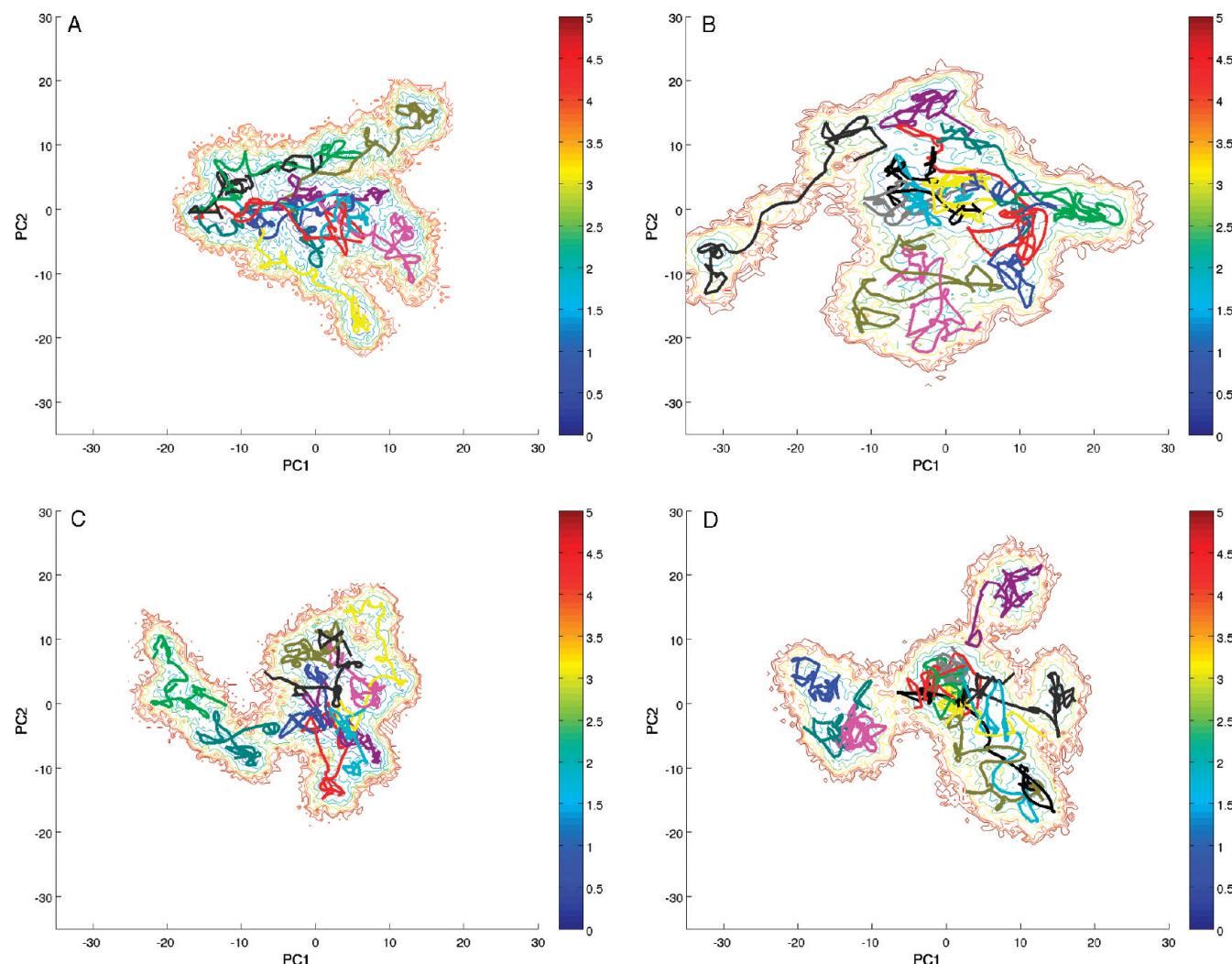


Figure 8. Projection of single trajectories onto the first two principal components of the all-atoms mass-weighted covariance matrix. For the sake of clarity, the trajectories are smoothed by using a Bezier interpolation and drawn over the contour plot representation of the principal component FEL of *m*HUDG (A, B) and of *p*CUDG (C, D) at 283 K (A, C) and at 310 K (B, D). The free energy is given in kJ/mol and indicated by the color bar.

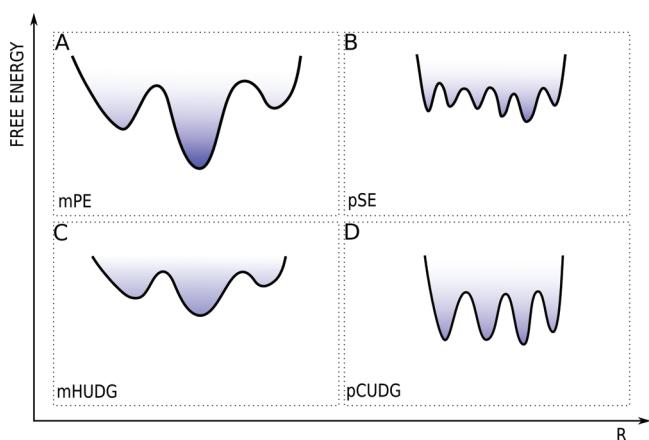


Figure 9. A model of the near-native FEL of the mesophilic (A, C) and psychrophilic (B, D) enzymes. The x-axis represents a generic conformational reaction coordinate (R), whereas the y-axis indicates the free energy.

In this context, our computational data, collected using as model systems elastases and UDGs, which have been deeply characterized and which are among the smaller and more computationally accessible cold-adapted protein systems, strongly enforce the current view on stability and flexibility relationship

in cold-adapted enzymes. The study can be extended in the future to other more computationally challenging known structures of cold-adapted enzymes, such as AHA, which is characterized by a complex multidomain protein structure and which was the model enzyme adopted in the studies of D'Amico et al.³⁸ The evidence of differences in the shape and whole properties of the conformational landscape accessible to psychrophilic and mesophilic enzymes, pointed out in the present investigation, also opens the possibility of further investigations by other computational techniques, such as metadynamics, suitable to capture and describe details of the free energy surface and of energetic barriers dividing the different substates which populate the native-state.

Acknowledgment. The authors thank Marco Pasi and Matteo Roveda for technical help and fruitful comments. This research was supported by P.N.R.A (Progetto Nazionale Ricerche in Antartide), CINECA (Project-562 and Project-609), and CASPUR (Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Università e Ricerca) Standard HPC Grant 2010. The Norwegian Structural Biology Centre (NorStruct) is supported by the Functional Genomics Program (FUGE) of the Research Council of Norway.

Supporting Information Available: Figure 1S showing the cosine content as a function of the eigenvalues. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Siddiqui, K. S.; Cavicchioli, R. *Annu. Rev. Biochem.* **2006**, *75*, 403–433.
- (2) Georlette, D.; Blaise, V.; Collins, T.; D'Amico, S.; Gratia, E.; Hoyoux, A.; Marx, J.; Sonan, G.; Feller, G.; Gerday, C. *FEMS Microbiol. Rev.* **2004**, *28*, 25–42.
- (3) Johns, G. C.; Somero, G. N. *Mol. Biol. Evol.* **2004**, *21*, 314–320.
- (4) Santarossa, G.; Lafranconi, P. G.; Alquati, C.; DeGioia, L.; Alberghina, L.; Fantucci, P.; Lotti, M. *FEBS Lett.* **2005**, *579*, 2383–2386.
- (5) D'Auria, S.; Aurilia, V.; Marabotti, A.; Gonnelli, M.; Strambini, G. *J. Phys. Chem. B* **2009**, *113*, 13171–13178.
- (6) Jayaraman, G.; Srimathi, S.; Bjarnason, J. B. *Biochim. Biophys. Acta* **2006**, *1760*, 47–54.
- (7) Roy, D.; Sengupta, S. *J. Biomol. Struct. Dyn.* **2007**, *24*, 463–470.
- (8) Goldstein, R. A. *Protein Sci.* **2007**, *16*, 1887–1895.
- (9) Fedoy, A.; Yang, N.; Martinez, A.; Leiros, H. S.; Steen, I. H. *J. Mol. Biol.* **2007**, *372*, 130–149.
- (10) Tronelli, D.; Maugini, E.; Bossa, F.; Pascarella, S. *FEBS J.* **2007**, *274*, 4595–4608.
- (11) Bjelic, S.; Brandsdal, B. O.; Aqvist, J. *Biochemistry* **2008**, *47*, 10049–10057.
- (12) Heidarsson, P. O.; Sigurdsson, S. T.; Asgeirsson, B. *FEBS J.* **2009**, *276*, 2725–2735.
- (13) Gerday, C.; Aittaleb, M.; Bentahir, M.; Chessa, J. P.; Claverie, P.; Collins, T.; D'Amico, S.; Dumont, J.; Garsoux, G.; Georlette, D.; Hoyoux, A.; Lonhienne, T.; Meuwis, M. A.; Feller, G. *Trends Biotechnol.* **2000**, *18*, 103–107.
- (14) Guðmundsdóttir, A.; Pálsdóttir, H. M. *Mar. Biotechnol. (N.Y.)* **2005**, *7*, 77–88.
- (15) Tutino, M. L.; di Prisco, G.; Marino, G.; de Pascale, D. *Protein Pept. Lett.* **2009**, *16*, 1172–1180.
- (16) Olufsen, M.; Smalås, A. O.; Moe, E.; Brandsdal, B. O. *J. Biol. Chem.* **2005**, *280*, 18042–18048.
- (17) Papaleo, E.; Fantucci, P.; De Gioia, L. *J. Chem. Theory Comput.* **2005**, *1*, 1286–1297.
- (18) Papaleo, E.; Riccardi, L.; Villa, C.; Fantucci, P.; De Gioia, L. *Biochim. Biophys. Acta* **2006**, *1764*, 1397–1406.
- (19) Adekoya, O. A.; Helland, R.; Willassen, N.; Sylte, I. *Proteins* **2006**, *62*, 435–449.
- (20) Papaleo, E.; Olufsen, M.; De Gioia, L.; Brandsdal, B. O. *J. Mol. Graphics Modell.* **2007**, *26*, 93–103.
- (21) Olufsen, M.; Brandsdal, B. O.; Smalås, A. O. *J. Mol. Graph. Model.* **2007**, *26*, 124–134.
- (22) Olufsen, M.; Papaleo, E.; Smalås, A. O.; Brandsdal, B. O. *Proteins* **2008**, *71*, 1219–1230.
- (23) Olufsen, M.; Smalås, A. O.; Brandsdal, B. O. *J. Mol. Model.* **2008**, *14*, 201–213.
- (24) Papaleo, E.; Pasi, M.; Riccardi, L.; Sambi, I.; Fantucci, P.; De Gioia, L. *FEBS Lett.* **2008**, *582*, 1008–1018.
- (25) Spiwok, V.; Lipovová, P.; Skálová, T.; Dusková, J.; Dohnálek, J.; Hasek, J.; Russell, N. J.; Králová, B. *J. Mol. Model.* **2007**, *13*, 485–497.
- (26) Senet, P.; Maisuradze, G. G.; Foulie, C.; Delarue, P.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 19708–19713.
- (27) Papaleo, E.; Mereghetti, P.; Fantucci, P.; Grandori, R.; De Gioia, L. *J. Mol. Graphics Modell.* **2009**, *27*, 889–899.
- (28) Hori, N.; Chikenji, G.; Berry, R. S.; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 73–78.
- (29) Wolyne, P. G. *Q. Rev. Biophys.* **2005**, *38*, 405–410.
- (30) Gruebele, M. *C. R. Biol.* **2005**, *328*, 701–712.
- (31) Poland, D. *Proteins* **2001**, *45*, 325–336.
- (32) Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005**, *58*, 45–52.
- (33) Altis, A.; Otten, M.; Nguyen, P. H.; Hegger, R.; Stock, G. *J. Chem. Phys.* **2008**, *128*, 245102.
- (34) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.
- (35) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.
- (36) Tavernelli, I.; Cestea, S.; Iorio, E. E. D. *Biophys. J.* **2003**, *85*, 2641–2649.
- (37) Zhuravlev, P. I.; Matarese, C. K.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 8800–8812.
- (38) D'Amico, S.; Marx, J.; Gerday, C.; Feller, G. *J. Biol. Chem.* **2003**, *278*, 7891–7896.
- (39) Berglund, G. I.; Smalås, A. O.; Outzen, H.; Willlassen, N. P. *Mol. Mar. Biol. Biotechnol.* **1998**, *7*, 105–114.
- (40) Lanes, O.; Guddal, P. H.; Gjellesvik, D. R.; Willlassen, N. P. *Comp. Biochem. Physiol., B: Biochem. Mol. Biol.* **2002**, *127*, 399–410.
- (41) Lanes, O.; Leiros, I.; Smalås, A. O.; Willlassen, N. P. *Extremophiles* **2002**, *6*, 73–86.
- (42) Caves, L. S.; Evanseck, J. D.; Karplus, M. *Protein Sci.* **1998**, *7*, 649–666.
- (43) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.
- (44) Loccisano, A. E.; Acevedo, O.; DeChancie, J.; Schulze, B. G.; Evanseck, J. D. *J. Mol. Graphics Modell.* **2004**, *22*, 369–376.
- (45) Pandini, A.; Bonati, L. *Protein Eng., Des. Sel.* **2005**, *18*, 127–137.
- (46) Meirovitch, H. *Curr. Opin. Struct. Biol.* **2007**, *17*, 181–186.
- (47) Nguyen, P. H.; Stock, G.; Mittag, E.; Hu, C. K.; Li, M. S. *Proteins* **2005**, *61*, 795–808.
- (48) Nguyen, P. H. *Chem. Phys. Lett.* **2009**, *468*, 90–93.
- (49) Schlitter, J. *Chem. Phys. Lett.* **1993**, *215*, 617–621.
- (50) Schiltz, M.; Shepard, W.; Fourme, R.; Prangé, T.; de la Fortelle, E.; Bricogne, G. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1997**, *53*, 78–92.
- (51) Berglund, G. I.; Willassen, N. P.; Hordvik, A.; Smalås, A. O. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *51*, 925–937.
- (52) Mol, C. D.; Arvai, A. S.; Slupphaug, G.; Kayli, B.; Alseth, I.; Krokan, H. E.; Tainer, J. A. *Cell* **1995**, *80*, 869–878.
- (53) Leiros, I.; Moe, E.; Lanes, O.; Smalås, A. O.; Willassen, N. P. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2003**, *59*, 1357–1365.
- (54) Hu, Z.; Jiang, J. *J. Comput. Chem.* **2009**, *31*, 371–380.
- (55) Berendsen, H. J. C.; Postma, J. P. M.; Di Nola, A.; Haak, J. R. *J. Phys. Chem.* **1984**, *81*, 3684–3690.
- (56) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (57) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1981**, *98*, 10089.
- (58) Hess, B. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2000**, *62*, 8438–8448.
- (59) Maisuradze, G. G.; Leitner, D. M. *Proteins* **2007**, *67*, 569–578.
- (60) Maisuradze, G. G.; Leitner, D. M. *Chem. Phys. Lett.* **2007**, *421*, 5–10.
- (61) Amadei, A.; Ceruso, M. A.; Di Nola, A. *Proteins* **1999**, *36*, 419–424.
- (62) Pontiggia, F.; Colombo, G.; Micheletti, C.; Orland, H. *Phys. Rev. Lett.* **2007**, *98*, 048102.
- (63) Lange, O. F.; Grubmüller, H. *J. Chem. Phys.* **2006**, *124*, 214903–214918.
- (64) Schafer, H.; Mark, A.; van Gusteren, W. *J. Chem. Phys.* **2000**, *113*, 7809–7817.
- (65) Hsu, S. T.; Peter, C.; van Gusteren, W. F.; Bonvin, A. M. *Biophys. J.* **2005**, *88*, 15–24.
- (66) Andricioaei, I.; Karplus, M. *J. Chem. Phys.* **2001**, *115*, 6289–6292.
- (67) Carlsson, J.; Aqvist, J. *J. Phys. Chem. B* **2005**, *109*, 6448–6456.
- (68) Baron, R.; van Gusteren, W. F.; Hunenberger, P. H. *Trends Phys. Chem.* **2006**, *11*, 87–122.
- (69) Ikeda, K.; Higo, J. *Protein Sci.* **2003**, *12*, 2542–2548.
- (70) Becker, O. M. *Proteins* **1997**, *27*, 213–226.
- (71) Lei, H.; Wu, C.; Liu, H.; Duan, Y. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 4925–4930.
- (72) Xie, B.; Bian, F.; Chen, X.; He, H.; Guo, J.; Gao, X.; Zeng, Y.; Chen, B.; Zhou, B.; Zhang, Y. *J. Biol. Chem.* **2009**, *284*, 9257–0269.
- (73) Bandi, S.; Bowler, B. E. *J. Am. Chem. Soc.* **2008**, *130*, 7540–7541.

JP911523H