J|A|C|S
ARTICLES

Published on Web 11/17/2010

Energy Landscapes Associated with Macromolecular Conformational Changes from Endpoint Structures

Arianna Fornili,*,‡ Barbara Giabbai,§ Gianpiero Garau,⊥ and Massimo Degano*

Biocrystallography Unit, Division of Immunology, Transplantation, and Infectious Diseases, Scientific Institute San Raffaele, Via Olgettina 58, 20132 Milan, Italy

Received August 24, 2010; E-mail: arianna.fornili@kcl.ac.uk; degano.massimo@hsr.it.

Abstract: Conformational changes modulate macromolecular function by promoting the specific binding of ligands (such as in antigen recognition) or the stabilization of transition states in enzymatic reactions. However, quantitative characterization of the energetics underlying dynamic structural interconversions is still challenging and lacks a unified method. Here, we introduce a novel *in silico* approach based on the combined use of essential dynamics sampling and nonequilibrium free-energy calculations to obtain quantitative data on conformational energy landscapes. This technique allows the unbiased investigation of highly complex rearrangements, and does not require the crucial definition of user-defined collective variables. We show that free-energy values derived from profiles connecting the unliganded and ligand-bound X-ray structures of a bacterial nucleoside hydrolase match the experimental binding constant. This approach also provides first evidence for a rate-limiting character of the conformational transition in this enzyme, and an unexpected role of the protonation state of a single residue in regulating substrate binding and product release.

1. Introduction

Ligand-induced reorganization of tertiary structures is a versatile process that controls and modulates the biological function of macromolecules. Enzymes take advantage of their conformational flexibility to establish specific binding interactions with the substrate and to improve catalytic efficiency. Structural rearrangements are also instrumental in receptor-mediated recognition of ligands, such as in antibody/antigen or peptide/MHC/TCR complexes,[1] to allow the binding of structurally related epitopes. Recent experiments suggest that the intrinsic dynamics of proteins may be biased toward pathways leading to the rearrangements required to perform their specific physiological function.[2] The protein free-energy landscape can be further modulated by the interaction with specific ligands, promoting the preferential sampling of functionally active structures.

The determination of energy landscapes in the conformational space is therefore of central interest in understanding the biochemistry of macromolecules. However, while low-energy structures can be determined experimentally to high resolution using X-ray crystallography and flexible regions can be mapped by NMR, it is yet difficult to observe the high-energy conformations sampled by proteins along their catalytic cycle. To this end, several *in silico* approaches have been developed and successfully tested on prototypical systems,[3−7] even if their application to real cases remains a challenge. A requirement for many such methods is the availability of one or few simple reaction coordinates describing the conformational change. However, the vast majority of structural changes in macromolecules involves a combination of complex events, often occurring cooperatively, ranging from transitions between side chain rotamers to refolding of secondary structure elements, that usually cannot be reduced to a simple coordinate.

Thus, in spite of the increasing availability of experimental structures for different conformational states of the same macromolecule, there is not yet a general method to both identify and energetically characterize the interconversion pathways between them, especially when cooperativity is involved. Here, we show that free-energy profiles of the transition connecting two endpoints can be determined using a combination of targeted essential dynamics sampling (TEDS)[8] and nonequilibrium work (NEW) methods[9−12] without the introduction of a specific reaction coordinate. This approach is general and it does not

‡ Current address: King's College, Randall Division of Cell and Molecular Biophysics, New Hunt's House, SE1 1UL London, U.K.

§ Current address: AdriaCell, Area Science Park, Basovizza SS 14, Km 163.5 ed. Q1, 34012 Trieste, Italy.

⊥ Current address: Istituto Italiano di Tecnologia (IIT), Department of Drug Discovery and Development, Via Morego, 30 16163 Genova, Italy.

(1) Garcia, K. C.; Degano, M.; Pease, L. R.; Huang, M.; Peterson, P. A.; Teyton, L.; Wilson, I. A. *Science* **1998**, *279*, 1166–1172.
(2) Henzler-Wildman, K. A.; Thai, V.; Lei, M.; Ott, M.; Wolf-Watz, M.; Fenn, T.; Pozharski, E.; Wilson, M. A.; Petsko, G. A.; Karplus, M.; Hübner, C. G.; Kern, D. *Nature* **2007**, *450*, 838–844.
(3) Roux, B. *Comput. Phys. Commun.* **1994**, *91*, 275–282.
(4) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
(5) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2007**, *126*, 164106.
(6) Lange, O. F.; Schäfer, L. V.; Grubmüller, H. *J. Comput. Chem.* **2006**, *27*, 1693–1702.
(7) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. *Adv. Chem. Phys.* **2002**, *123*, 1–78.
(8) Amadei, A.; Linssen, A. B.; de Groot, B. L.; van Aalten, D. M.; Berendsen, H. J. *J. Biomol. Struct. Dyn.* **1996**, *13*, 615–625.
(9) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
(10) Crooks, G. E. *Phys. Rev. E* **1999**, *60*, 2721–2726.
(11) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco Jr., I.; Bustamante, C. *Science* **2002**, *296*, 1832–1835.
(12) Shirts, M.; Bair, E.; Hooker, G.; Pande, V. *Phys. Rev. Lett.* **2003**, *91*, 140601.

 10.1021/ja107640u

require a prior knowledge of the process involved. We also assess its accuracy through comparison with experimental data. The method is applied to the characterization of the conformational change of the pyrimidine nucleoside hydrolase (NH) YeiK from *Escherichia coli* upon binding of a competitive inhibitor. The open (loosely bound ligand) and closed (tightly bound ligand) structures of YeiK are derived from X-ray diffraction, with the open form reported here for the first time. The energetic features of the conformational change lead us to propose a pH-mediated regulation mechanism for substrate binding and product release coupled with the enzymatic reaction catalyzed by YeiK.

2. Materials and Methods

2.1. X-ray Diffraction. Recombinant YeiK was crystallized both unliganded or in complex with DAPIR using vapor diffusion. Diffraction data for both crystals were collected at 100 K at beamline ID14-1 of ESRF (Grenoble, F), and the structures were solved by molecular replacement. The crystal structures were refined to 2.2 and 2.0-Å resolution, respectively (PDB codes 3MKM and 3MKN). Details of the structural determination and of the measurement of the inhibition constant for the DAPIR compound are reported as Supporting Information.

2.2. Molecular Dynamics Simulations. All simulations (unbiased Molecular Dynamics (MD), Targeted Essential Dynamics Sampling (TEDS), Targeted MD (TMD), and Thermodynamic Integration (TI)) and the subsequent analyses were performed with GROMACS 3.3.1,[13] using the ff-amber99[14] porting of the AMBER parm99 parameter set,[15] the TIP3P model for water and the General AMBER Force Field (GAFF)[16] for DAPIR. Periodic boundary conditions were imposed. The bonds involving hydrogen bonds were frozen and a 1-fs time step was used. Simulations were performed in the NPT ensemble ($T = 300$ K and $p = 1$ bar). Electrostatic interactions were calculated with the Particle Mesh Ewald (PME) method. A 9-Å cutoff was used for all the nonbonded interactions.

Conformational pathways from the open to the closed structure of YeiK were determined with the TEDS method,[8,17] where the sampling along selected principal components (PCs) of motion,[18] derived from the unbiased MD trajectory of the open form, was enhanced following the least perturbation principle. The overall transition was decomposed into smaller transformations or segments by extracting structures at regular intervals along the path (see Supporting Information). Ensembles of conformations were sampled through MD simulations around each of the selected structures. The system was then driven through each segment with TMD transformations connecting structures belonging to the two endpoint ensembles, both in the forward and reverse directions and using different initial velocities. Each forward work distribution was combined with the corresponding reverse to give the free-energy change for each segment through the Bennett Acceptance Ratio (BAR) formula.[12,19,20] An a posteriori correction was applied when work distributions with outliers were observed, to keep the degree of irreversibility uniform over the TMD transformations (see Supporting Information). In selecting the representative structures from the TEDS trajectories and the velocity of the TMD transformations, we had to take into account that the deviation from reversibility conditions of a given transformation, and hence the accuracy of the free-energy calculation, is influenced by its velocity and by the distance between the endpoints.[21,22] Preliminary calculations were then run to find the conditions that limited the work dispersion, while keeping the computational cost affordable.

Binding of DAPIR to the open and closed conformation of YeiK was calculated using the TI method[23,24] with the double-annihilation scheme. The binding energy of DAPIR was evaluated with respect to that of the two tightly bound water molecules that complete the coordination shell of Ca^{2+} in the unbound state.

Further methodological details, simulation lengths, and full references are given as Supporting Information.

3. Results

3.1. Conformational Changes in NH Enzymes. NHs are Ca^{2+}-dependent enzymes that hydrolyze the N-glycosidic bond in purine and pyrimidine nucleosides to yield free nitrogenous bases as DNA and RNA precursors, or in nicotinamide riboside for NAD^+ biosynthesis.[25] They are targets for the development of specific compounds against purine-auxotrophic pathogens, such as protozoan parasites of the *Trypanosoma* or *Leishmania* genus. Ligand-free NHs assume primarily a loose conformation, characterized by high flexibility of regions neighboring the active site cavity.[26] They then undergo a transition to a highly ordered, closed form upon binding of substrate-mimicking ligands or transition-state-like inhibitors.[27] We identified the molecule 3,4-diaminophenyl-iminoribitol (DAPIR) as a competitive inhibitor of the *E. coli* pyrimidine-specific NH YeiK, with a $K_I = 76 \pm 8$ μM (Materials and Methods, and Supporting Information). To gain further insights into the structural basis of the enzyme specificity, we solved the YeiK crystal structure both unliganded and bound to DAPIR (Supporting Information Table S1). The ligand-free enzyme is characterized by two highly flexible regions, namely, the β3-α3 loop (L1) and the C-terminal end of helix α9 (L2) that are not univocally resolved in electron density maps (Figure 1). When DAPIR is cocrystallized with YeiK, the L1 and L2 regions of three protein subunits in the monoclinic crystals assume a closed conformation that is driven by the formation of specific protein−ligand interactions and likely reflects the arrangement when the substrate is tightly bound to the enzyme before catalysis. The L1 region folds into a short helical segment, and its conserved amino acids Ile81 and His82 establish van der Waals contacts with the hydrophobic ring of DAPIR (Figure 1). The L2 helix undergoes a torsional movement about its axis to bring the side chains of residues Gln227 and Tyr231 within hydrogen bonding distance of the inhibitor amino groups. Notably, one of the four independent YeiK subunits in the asymmetric unit assumes a different, open conformation, perhaps due to stabilizing crystal contacts, where DAPIR is bound to the active site and both the L1 and L2

(13) van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
(14) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472–2493.
(15) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(16) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(17) de Groot, B. L.; Amadei, A.; van Aalten, D. M. F.; Berendsen, H. J. C. *J. Biomol. Struct. Dyn.* **1996**, *13*, 741–751.
(18) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.
(19) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245–268.
(20) Lu, N.; Singh, J. K.; Kofke, D. A. *J. Chem. Phys.* **2003**, *118*, 2977–2984.
(21) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559–3566.
(22) West, D. K.; Olmsted, P. D.; Paci, E. *J. Chem. Phys.* **2006**, *125*, 204910.
(23) Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1991**, *95*, 1175–1188.
(24) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004–2010.
(25) Belenky, P.; Racette, F. G.; Bogan, K. L.; McClure, J. M.; Smith, J. S.; Brenner, C. *Cell* **2007**, *129*, 473–484.
(26) Degano, M.; Gopaul, D. N.; Scapin, G.; Schramm, V. L.; Sacchettini, J. C. *Biochemistry* **1996**, *35*, 5971–5981.
(27) Degano, M.; Almo, S. C.; Sacchettini, J. C.; Schramm, V. L. *Biochemistry* **1998**, *37*, 6277–6285.

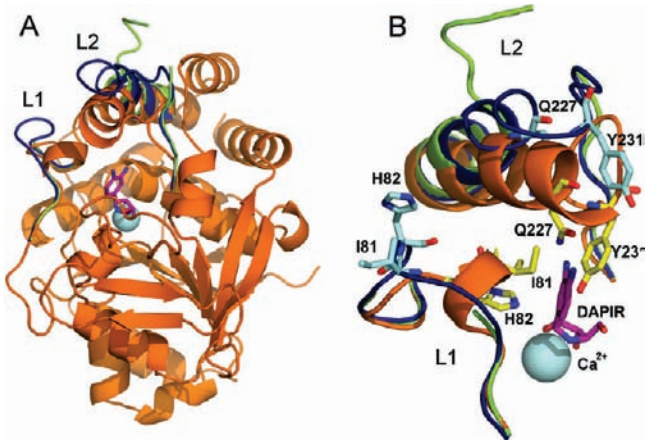

Figure 1. (A) X-ray structure of the closed form of YeiK (orange cartoon) bound to DAPIR (magenta sticks). Loops L1 and L2 from the open (blue) and flexible (green) structures are also shown. Residues 80:83 and 230:233 are not resolved in the electron density maps of the flexible structure. The Ca^{2+} ion is represented as a cyan van der Waals sphere. (B) Close-up view of loops L1 and L2 from the closed (orange), open (blue), and flexible (green) structures of YeiK. Selected residues of the closed and open structures are shown as yellow and cyan sticks, respectively.

regions are traceable in electron density maps, but they do not establish any contacts with the diaminobenzyl ring (Figure 1). This structure represents the first observation of a discrete conformation at the L1 and L2 regions in NH enzymes that is different from the closed form.

3.2. Dynamic Characterization of Endpoints. We first investigated the stability in aqueous solution of the different L1 and L2 conformations in the YeiK−DAPIR cocrystals and the possibility to observe spontaneous transitions between them. The extent of the sampling of the conformational space around either the closed or open form of YeiK was analyzed in unbiased MD simulations, and the influence of the ligand and of the pH on the protein dynamics was assessed. In particular, the H82 (L1) and H239 (L2) residues were considered either in their neutral or charged state, to analyze the effect of local deviations of pH from the neutral condition possibly occurring during the catalytic cycle.[28−30]

The YeiK closed form was highly stable when bound to ligand, while extraction of DAPIR induced large rearrangements in L2 that sampled clusters of conformations in close agreement with the close, open, and flexible X-ray structures (Figure 2A, Supporting Information Figures S1B and S2). The observation of large conformational changes for L2 within 15 ns would suggest that no significant energy barriers are involved in the process. Although the major collective movement involved L2 almost exclusively, an initial opening of L1 occurred in the first half of the simulation in concert with the motion of L2 (Supporting Information Figures S3A and S3B). The complete transition to the flexible form of L2 in the second half of the trajectory disrupted the contacts with L1 such that instead of opening further it returned to the closed conformation. This suggests the existence of a significant barrier for the rearrangement of L1 and a possible role of the interactions between the two loops in promoting the conformational change.

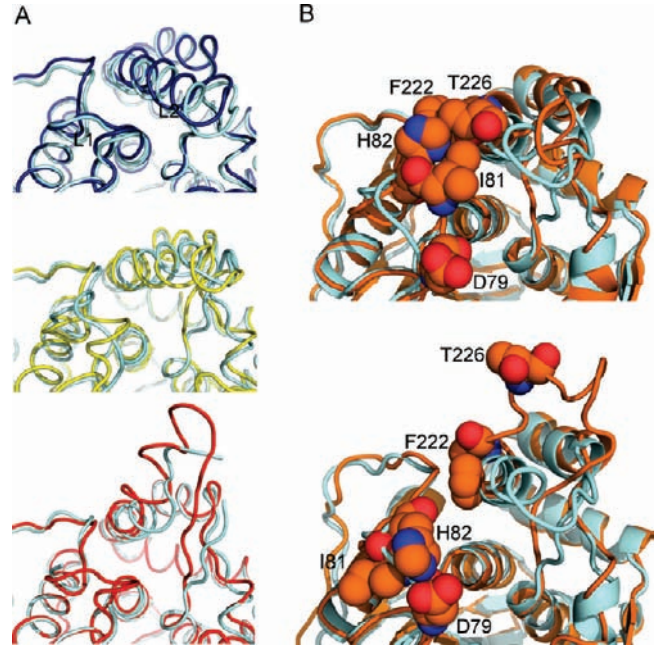

Figure 2. (A) Representative structures of the three most populated clusters in the unbiased MD simulation starting from the closed form of YeiK after ligand extraction (neutral histidines). The cluster analysis has been performed with the method of Daura et al.[31] with a 1-Å cutoff. The clusters containing the 12% (blue), 29% (yellow), and 34% (red) of the overall population are represented in the upper, middle, and lower panels superimposed to the X-ray (cyan) closed (backbone rmsd = 1.12 Å), open (rmsd = 1.53 Å), and flexible (rmsd = 1.55 Å, calculated excluding the missing residues 80:83 and 230:233) structures, respectively. (B) Representative structures (orange) from the MD simulations starting from the X-ray open form (cyan) with neutral (upper) and charged (lower) H82 and H239 residues. Selected residues in L1 and L2 loops are shown as van der Waals spheres.

The dynamics of the open form and in particular of loop L1 was significantly affected by the pH. When H82 and H239 were modeled as uncharged, a partial closure of L1 was detected, favored by transient hydrophobic interactions between the L1 and L2 side chains (Figure 2B). A concerted movement involving the closure of both loops was found among the large-amplitude collective motions (Supporting Information Figure S3C). When the histidines were protonated, reduced contacts between L1 and L2 were observed, probably due to an intraloop salt bridge between H82 and D79 (Figure 2B). No appreciable closure of L1 could be detected, while L2 opened further toward conformations resembling the flexible X-ray structure. This further supports the view that L1−L2 interactions are required to promote the conformational change of L1. The partial closure movement observed in the open conformation with neutral histidines was exploited in the following to achieve a complete transition from the open to the closed form using enhanced-sampling MD techniques.

3.3. Open-to-Closed Transition Path. Among the available methods for the determination of transition pathways,[4−8] we selected the targeted essential dynamics sampling (TEDS).[8,17,32] In this approach, the external perturbation on the system is reduced by adding constraints only when it moves away from the target, while motions that bring the system closer to the target structure result from normal unconstrained MD steps. The sampling toward the X-ray closed structure was biased only in an "essential space" composed by the large-amplitude collective motions of the open conformation, leaving the other degrees of

(28) Giabbai, B.; Degano, M. *Structure* **2004**, *12*, 739–749.
(29) Parkin, D. W.; Schramm, V. L. *Biochemistry* **1995**, *34*, 13961–13966.
(30) Gopaul, D. N.; Meyer, S. L.; Degano, M.; Sacchettini, J. C.; Schramm, V. L. *Biochemistry* **1996**, *35*, 5963–5970.
(31) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
(32) Snow, C.; Qi, G.; Hayward, S. *Proteins* **2007**, *67*, 325–337.

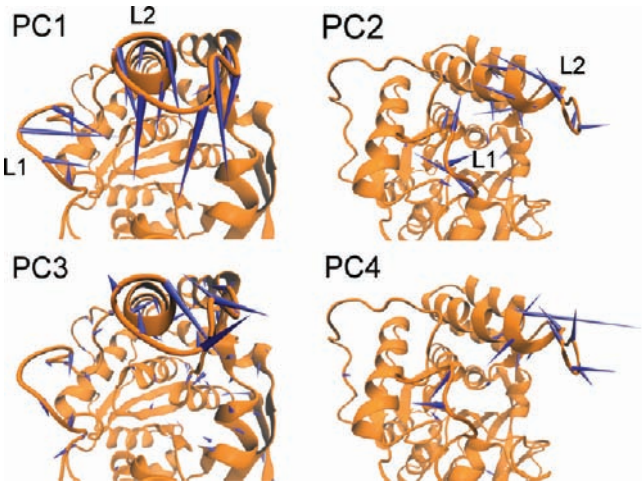

Figure 3. Porcupine representation of the first four PCs extracted from the TEDS trajectory of the ligand-free YeiK. The blue spikes give the direction and relative amplitude of the motion of each Cα atom along the PC. The collective motions can be summarized as a concerted closure of the two loops (PC1), the shortening of L2 region together with a lid-like movement of L1 (PC2), the winding of L2 around its axis (PC3), and the 'gating' of one of L2 turns (PC4).

freedom unperturbed. The collective motions were extracted with a principal component analysis (PCA) of the heavy-atom coordinates of both the backbone and the side chains (2316 atoms). The latter had to be included to obtain a complete closure, consistent with the importance of the interactions between L1 and L2 side chains in determining the transition path (see below). Indeed, the role of side chains in TEDS calculations has already been pointed out by Friedman et al.,[33] when applying TEDS to the cleavage-induced activation of plasmapepsin. Our essential space was composed by the first 300 PCs. This relatively large number is due to the inclusion of a large fraction of the protein atoms in the PCA. Indeed, the essential space only amounts to 4% of the total PCA space, while reproducing more than 90% of the overall fluctuation.

A total of 39 different TEDS simulations were performed either with or without the ligand. Not all the trajectories were successful in reaching the target, since in some of them the closure of L1 was incomplete (Supporting Information Table S2). This is consistent with the least-perturbation approach used in TEDS,[8] where the system is not forced to conformations that it would not sample through a normal step of unbiased MD.[33] When the complete closure was achieved, the system followed similar paths (Supporting Information Figure S4), involving a concerted movement of the two loops. The most complete trajectory, leading to the structure closer to the target, was selected for the geometric and energetic analyses both in the ligand-free and ligand-bound case (the movie of the closure of ligand-bound YeiK is available in mpeg format as Supporting Information Video S1). All the TEDS trajectories pointed to a common mechanism; hence, we do not expect an effect of this choice on the present results. However, multiple pathways would be required to be taken into account when significantly different TEDS trajectories were observed.

The first four PCs extracted from the selected ligand-free TEDS trajectory provided a convenient descriptor of the main features of the transition (Figure 3). The concerted closure of the two loops (PC1) occurred mostly in the first half of the trajectory (Supporting Information Figure S5A), while the consolidation of their secondary structure (PC2-4), was observed mainly in the second half, suggesting that the interloop interactions are required for the loop shaping process.

The L1 residues I81 and H82 presented the most dramatic changes in side chain conformation and environment. Their relocation from the protein surface into the active site pocket occurred in two main stages (Figure 4**a**−**e** and 4**f**−**h**) involving successive rotations of L1 segments around two hinge axes (Figure 5A) to position first H82 in the catalytic pocket, followed by I81. These steps are paralleled by the formation of interloop contacts (Figure 5B). Also, the hydrogen bonding network of L2 underwent a significant reorganization (Figure 5C), going

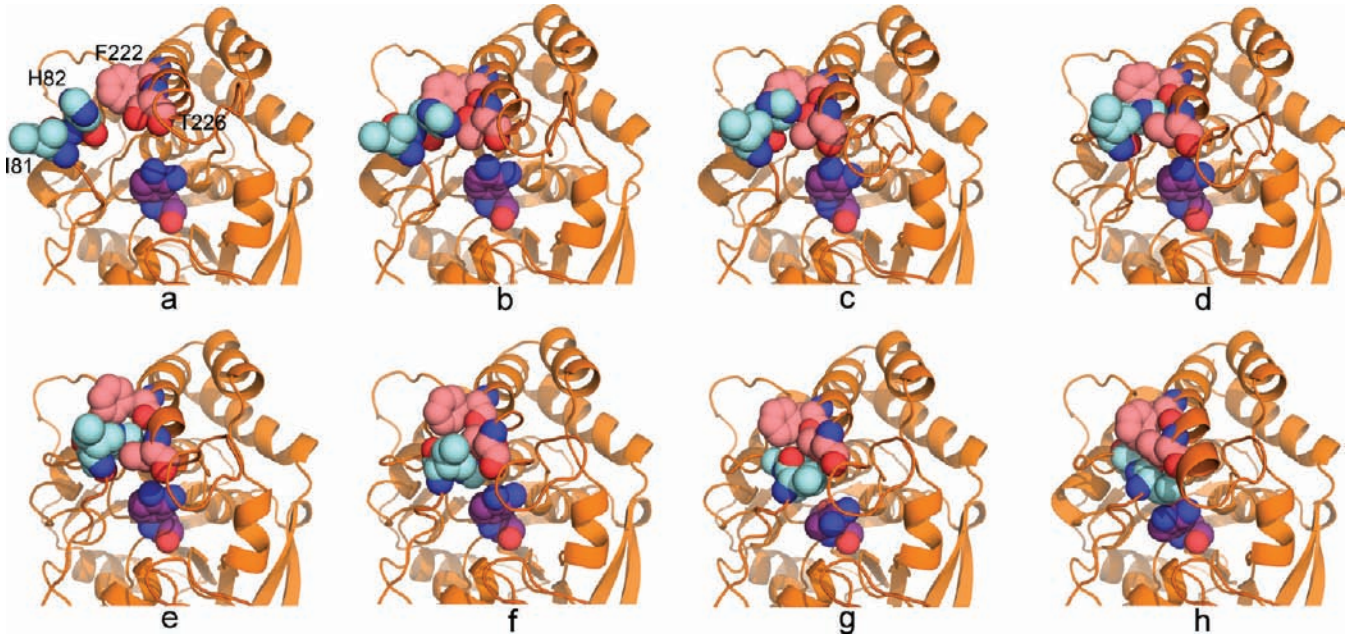

Figure 4. Cartoon snapshots from the TEDS trajectory of ligand-bound YeiK. L1 residues I81 and H82 (cyan), L2 residues F222 and T226 (pink), and the ligand DAPIR (purple) are represented as van der Waals spheres.

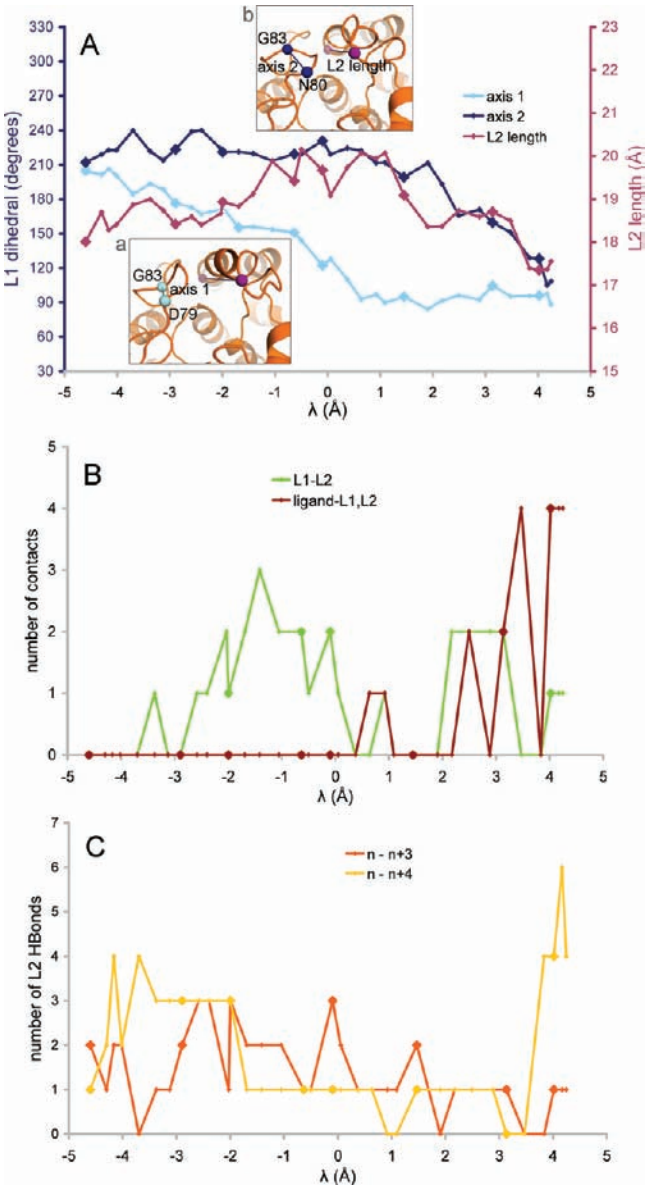

Figure 5. Geometric parameters describing the TEDS ligand-bound trajectory. Highlighted points refer to the selected conformations shown in Figure 4. (A) Dihedral angles A78:Cα-G83:Cα-D79:Cα-N80:Cα (axis 1, light blue) and A78:Cα-G83:Cα-D80:Cα-N81:Cα (axis 2, blue) plotted together with the I219:Cα-N230:Cα distance (L2 length, purple). The progression of the structure i along the trajectory is given by $\lambda_i = \mathrm{rmsd}_{open}(i) - \mathrm{rmsd}_{closed}(i)$, where the rmsd from either the X-ray open or closed conformation is calculated over loops L1−3 (Supporting Information Figure S6D). Two different scales are used for the dihedrals (degrees, left) and the distance (Å, right). The insets a and b illustrate the location of axis 1 and 2, respectively, together with the L2 length for two selected structures. By comparison with Figure 4, it is possible to see that the zones of steepest variation of the axis 1 and axis 2 curve correspond to the entering into the active site pocket of H82 (points **d**−**f**) and I81 (points **f**−**h**), respectively. The shortening of L2 (purple, $\lambda > 1$ Å) occurs simultaneously with the L1 rotation around axis 2 (blue), leading to the correlation between these two movements that is contained in PC2 (Figure 3). (B) Number of contacts between L1 and L2 (green) and between the ligand and the L1 and L2 residues (dark red). Two residues are considered in contact if their minimum distance is within 3.5 Å. (C) Number of $n-n+3$ (orange) and $n-n+4$ (yellow) hydrogen bonds in loop L2 (backbone only). Hydrogen bonds are determined from geometrical parameters (A−D distance <3.5 Å and A···D-H angle < 30°, where A and D are the hydrogen bond acceptor and donor, respectively).

from a pattern rich in $n-n+3$ hydrogen bonds to a partial disruption allowing the folding of L1 and then to the canonical $n-n+4$ bonds.

In the ligand-bound trajectory, interactions between the loops and DAPIR were observed only after the H82 residue has entered the pocket (Figure 5B). The closure of the two loops mostly affected the mobility and structure of the ligand phenyl group, with a significant reorientation and distortion of the ring induced by I81 (Figure 4**f**−**h**). Similar distortions have already been found in MD simulations of pyrimidinic substrates bound to YeiK.[34]

3.4. Free-Energy Profile along the Transition Path. The free-energy profile along the transition was evaluated using the nonequilibrium work (NEW) values required to 'drive' the system through the path. The free-energy values were extracted from work distributions associated with forward and reverse transformations using the Bennett Acceptance Ratio (BAR) formula[12,19,20] to correct for irreversibility effects. The transformations were performed with Targeted MD (TMD) simulations[35,36] connecting representative structures extracted from the TEDS trajectories (Supporting Information Figure S5B). The overall profile was built up by consecutively summing the free-energy change values for each segment (Figure 6). This procedure was applied to both the ligand-free and ligand-bound profiles, with a total simulation time of ∼2 μs.

Despite their ruggedness, similar features can be found in the two profiles (ligand-free and ligand-bound in Figure 6). A decrease of the free energy in the first part (zone A) is followed by a double barrier in both cases (zones B and C), and a final energy increment (zone D). The minimum values on the two sides of the profile can be taken as representative of the open (point **b(c)** of the ligand-free(-bound) profile) and closed (point **h** in both cases) state of the protein. While the open state is predicted to be much more populated in the absence of the ligand (ΔG(open → closed) = +18.7(±2.6) kcal/mol), the stability order is reversed (ΔG = −7.5(±2.5) kcal/mol) upon ligand binding (Table 1).

From the analysis of the conformations (Figure 4) and of the parameters used to describe the transition (Figure 5), it is possible to see that the two barriers correspond to the two-step folding of L1 residues H82 and I81 into the active site. The first "transition state" (point e in Figure 6 for both the ligand-free and ligand-bound profiles) occurs in the middle of the first hinged movement of L1 (Figure 5A) and before ligand-loop contacts are observed (Figure 5B). Indeed, the forward barrier is not affected by the ligand, amounting to ∼18 kcal/mol in both cases (Table 1). The second "transition" state **g**, where the two loops are already in partial contact with DAPIR, is instead stabilized in the presence of the ligand by ∼13 kcal/mol. The largest energy decrease is observed in the segment **g**−**h**, where the entering of I81 (Figure 4**g**−**h** and Figure 5A) and the most part of the shortening and hydrogen bond rearrangement of L2 (Figure 5A,C) take place. Moreover, the ligand-loops contacts are fully recovered (Figure 5B). As a consequence, a further stabilization of ∼9 kcal/mol is observed for the ligand-bound system (Table 1).

A third profile was calculated to investigate the influence of the pH on the process and in particular to test the hypothesis that the closure of L1 is disfavored when the H82 residue is protonated. The resulting curve (ligand-bound($H82^+$) in Figure 6) conserved the double-barrier shape of the other two. However, large energy increments were observed especially when the L1−L2 interactions mediated by H82 are particularly important (Figure 5A,B), in agreement with the unbiased MD simulations

(33) Friedman, R.; Caflisch, A. *Proteins* **2008**, *73*, 814–827.
(34) Fornili, A.; Sironi, M.; Degano, M. *J. Phys. Chem. B* **2007**, *111*, 6297–6302.

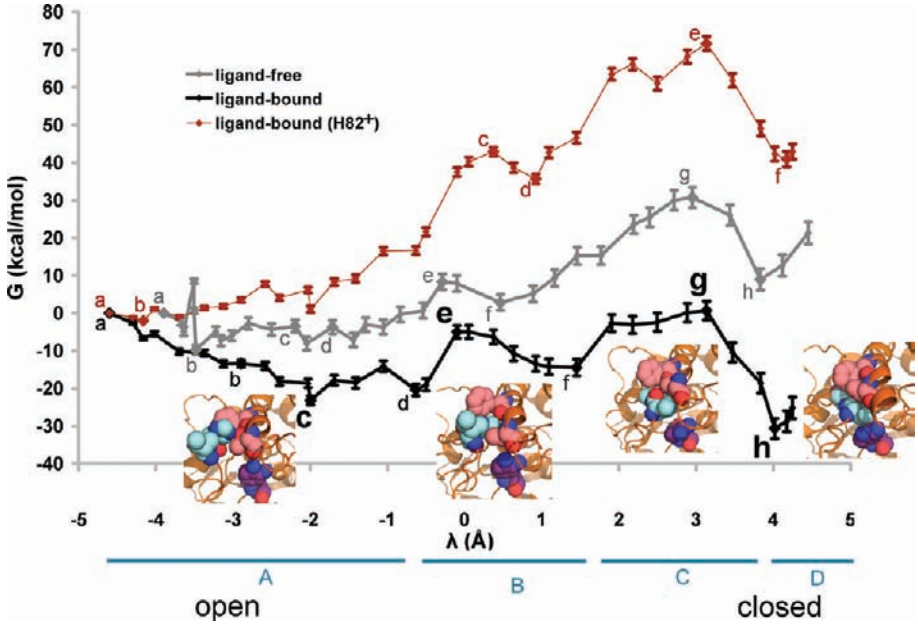

Figure 6. Profiles of free energy (in kcal/mol) calculated along the TEDS trajectories of the ligand-free (thick gray line), ligand-bound with neutral histidines (thick black line), and ligand-bound with charged H82 (thin brown line) YeiK. For each profile, relative free energies are reported where the first conformation (point **a**) is set as the reference state. The standard error of the mean is reported as bars for each point. Highlighted points for the ligand-bound profile refer to the conformations shown in Figure 4. In the insets, close-up views of selected structures from Figure 4 are shown (from left to right: structures **c**, **e**, **g**, and **h**). A broad decomposition of the profiles into four zones (A−D) is also suggested (in light blue).

Table 1. Calculated Free-Energy Differences (in kcal/mol) for the Ligand-Free, Ligand-Bound, and Ligand-Bound with Charged H82 (H82$^+$) Profiles

	profile		
free-energy difference	ligand-free	ligand-bound	ligand-bound (H82$^+$)
open → closed[a]	18.7 (±2.6)	−7.5 (±2.5)	43.0 (±2.0)
barrier B (forward)[b]	18.1 (±1.6)	18.2 (±1.2)	45.0 (±1.2)
barrier B (reverse)[c]	5.5 (±0.6)	9.5 (±1.5)	7.4 (±0.6)
barrier C (forward)[d]	27.9 (±1.8)	15.2 (±1.2)	36.1 (±1.2)
barrier C (reverse)[e]	21.8 (±0.7)	31.3 (±1.3)	30.7 (±0.8)

[a] G(**h**)−G(**b**) (ligand-free), G(**h**)−G(**c**) (ligand-bound), G(**h**)−G(**b**) (ligand-bound(H82$^+$)). Letters in parentheses refer to the points of the transition path highlighted in Figure 6. [b] G(**e**)−G(**b**) (ligand-free), G(**e**)−G(**c**) (ligand-bound), G(**c**)−G(**b**) (ligand-bound(H82$^+$)). [c] G(**e**)−G(**f**) (ligand-free), G(**e**)−G(**f**) (ligand-bound), G(**c**)−G(**d**) (ligand-bound(H82$^+$)). [d] G(**g**)−G(**f**) (ligand-free), G(**g**)−G(**f**) (ligand-bound), G(**e**)−G(**d**) (ligand-bound(H82$^+$)). [e] G(**g**)−G(**h**) (ligand-free), G(**g**)−G(**h**) (ligand-bound), G(**e**)−G(**f**) (ligand-bound(H82$^+$)).

of the open unbound form where L1−L2 contacts were drastically reduced upon H82 protonation (Figure 2). The energy increment with respect to the neutral ligand-bound profile reached a plateau at $\lambda \sim 2$ Å (Figure 6), after the complete relocation of H82 inside the cavity. It has to be pointed out that the ligand-bound(H82$^+$) profile gives the variation of the free energy of the protonated form following a path determined on the unprotonated YeiK structure. This approach, while allowing the direct comparison of the two profiles, could have produced an overestimation of the energy barriers. However, we can still assume that relaxation effects, which were partially taken into account by minimizing each protonated conformation, cannot compensate the magnitude of the increment with respect to the neutral ligand-bound profile.

3.5. Influence of Conformational Change on Ligand Binding Energetics and Comparison with Experiment. Experimental binding free energies ΔG_I of competitive inhibitors can be derived from the inhibition constant K_I. If the enzyme undergoes a significant conformational change, such as the case for YeiK, this value includes two different contributions, that is, the ligand binding ΔG_{bind}^{open} to the open form of the protein and the free-energy change ΔG_{cc}^{bound} associated with the closure of the ligand-bound enzyme (Supporting Information Figure S6B). Hence, the determination of the experimental K_I for DAPIR ("Conformational changes in NH enzymes"), corresponding to a ΔG_I of −5.8(±0.1) kcal/mol, provides a way to validate the accuracy of the ΔG_{cc}^{bound} value that can be extracted from the profiles described above, as long as an estimate of the ΔG_{bind}^{open} term is available.

We evaluated the binding energy of DAPIR to the open form of YeiK with the Thermodynamic Integration (TI) method:[23,24] the calculated ΔG_{bind}^{open} amounted to −1.7 (±0.7) kcal/mol (Supporting Information Figure S6C). If the free-energy change ΔG_{cc}^{bound} for the open-to-closed transition in ligand-bound YeiK is added (Table 1), a total of −9 (±3) kcal/mol is obtained. Hence, the agreement between our model and the experimental data, even if not strict, is within the range of the statistical error of the calculated value. It is also to be noted that the conformational reorganization, allowing for more extensive contacts between the protein and the ligand, contributes for a significant part of the free-energy change of the overall binding process, even taking into account the associated uncertainty.

4. Discussion

In this work, we have explored the possibility of setting up a new general procedure for the determination of free-energy profiles associated with a protein conformational change in a 'real' case, that is, for a protein of medium size and a relatively complex process, and starting only from knowledge of the two endpoints. Other approaches in the literature (such as umbrella

(35) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–309.
(36) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11905–11910.

sampling[3] or metadynamics[4]) rely on an accurate selection of a small number of collective variables (CVs) to describe the process. Indeed, the reduction of a process to a few simple coordinates may be problematic for systems beyond the prototypical ones, unless many pre-existing data are available[37,38] or several "trial and error" calculations are performed.[38]

The strategy developed here overcomes this requirement. The path determination was decoupled from the free-energy evaluation, so that we could use a large set of principal components (PCs) to capture the complexity of the transformation. In this stage, we performed TEDS simulations to enhance the sampling in the space defined by selected PCs calculated on the ensemble spanning the starting basin. Thus, the CVs naturally derived from the behavior of the system as observed in an unbiased MD simulation. The relatively large number of PCs involved in the transformation prevented their direct use as reaction coordinates for a free-energy calculation. The free-energy profile along the path was then evaluated at discrete points chosen close enough to capture all the significant features of the transition. The calculation of free-energy differences through a NEW-based method allowed the use of a simple TMD scheme to drive the transformations between the intermediate points. It is known that TMD may introduce artifacts into the path followed by the system.[39] However, here it was used only to determine the relative free-energy of the endpoints, and not to derive a path between them. The system could then be gradually "steered" using only one parameter (eq. S1 in the Supporting Information). It should be noted that in the steering potential harmonic restraints were imposed onto the positions of the atoms involved in the conformational change, but different choices could also have been possible (e.g., harmonic potentials onto the essential PCs would have been particularly suitable). To summarize the strategy adopted, we first determined a minimum-perturbation interconversion pathway with TEDS by using a high-dimensional space of essential coordinates and then we reduced it to a set of "segments" along which the system could be driven by variation of a single parameter to calculate the free energy profile.

Irreversibility effects in the TMD transformations were then taken into account using the BAR formula to extract the free-energy change from the distributions of the nonequilibrium work (NEW) values needed to complete each transformation. The BAR formula was originally developed for equilibrium free-energy methods, but it has been proven to be applicable also to nonequilibrium work (NEW) approaches.[12] Convergence issues can have a significant influence on free-energy values derived from NEW techniques. It has been recently shown that NEW methods relying on the sampling of both forward and reverse work distributions benefit from increased accuracy and faster convergence properties when compared to approaches based on one-sided transformations.[40−42] Indeed, the presence in the BAR formula of a weighting function that enhance the importance of the overlap of the two distributions renders the final value less dependent from the sampling of the tails outside the overlap. A detailed discussion of the convergence properties of different equilibrium and nonequilibrium methods can be found in the cited references.[40−42] The number of trajectories performed in this study averages ∼2100 per profile (Supporting Information Table S2). This is 1 order of magnitude larger than that employed elsewhere for comparable transition velocities (but for a smaller system and for a different type of transformation), where the resulting profiles were in good agreement with the reference data.[21] It has to be noted that the residual error (∼3 kcal/mol) found in the present work for the overall free-energy change makes still difficult a strict comparison with experimental data, and is the result of a tradeoff between the computational effort and the quality of the profile.

To the best of our knowledge, the combination of methods proposed here has never been used before. Other approaches exist that rely on the calculation of free-energy profiles over precalculated trajectories,[43−45] differing from the present one both on the methods used for the trajectory determination (e.g., TMD or Nudged Elastic Band[46]) and on the free-energy calculation strategy, based on equilibrium umbrella sampling calculations along a generic progress variable.

The reliability of the present calculations was checked against experimental data by comparing the free-energy change of the two endpoints of the bound form with the binding affinity derived from the inhibition constant, K_I. The calculated total value (−9 (±3) kcal/mol) compares favorably with the experimental ΔG_I of −5.8(±0.1) kcal/mol, considering that the theoretical data derive from two independent methods, each with its own uncertainty. Indeed, the difference is in the range of the estimated uncertainty of the theoretical value (∼3 kcal/mol). It is also to be taken into account that the ΔG_{cc}^{bound} calculation involves two structures that are well-separated along the transition path (Table 1 and Figure 6). Thus, their relative energy results from the sum of several energy differences evaluated on the conformations in between, so that the final value can be significantly affected by error propagation. A second test involving calculated data only (details are provided as Supporting Information) suggested that the discrepancy between theoretical and experimental data may be due to an overestimation of the stabilization of the closed form in ΔG_{cc}^{bound}.

We applied our protocol to the conformational change of the NH YeiK. Enzymes of the group I NH family have been known to undergo structural rearrangements when interacting with substrates or transition state-like inhibitors. While the biochemical role of this transition in excluding the solvent from the cavity and providing an adequate environment for catalysis has been long appreciated, the details of the mechanism and rate of the transition and the influence on the overall catalysis were yet unresolved issues. The overall picture emerging from the observed transition pathway is that closure occurs in a concerted way. The analysis of the structures along the path suggests that transient interactions between the two loops brought in contact in the initial stages of the closure are required for the subsequent reorganization of the loop secondary structure. A double energetic barrier is found in all the conditions examined, due to the two-stages relocation of L1 residues from the surface of the protein into the active site cavity.

While the calculated free-energy difference between the open and closed conformations of YeiK could be validated against

(37) Best, R. B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732–6737.
(38) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(39) van der Vaart, A.; Karplus, M. *J. Chem. Phys.* **2005**, *122*, 114903.
(40) Goette, M.; Grubmüller, H. *J. Comput. Chem.* **2009**, *30*, 447–456.
(41) Ytreberg, F. M.; Swendsen, R. H.; Zuckerman, D. M. *J. Chem. Phys.* **2006**, *125*, 184114.
(42) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 144107.
(43) Banavali, N. K.; Roux, B. *J. Am. Chem. Soc.* **2005**, *127*, 6866–6876.
(44) Arora, K.; Brooks, C. L., III. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 18496–18501.
(45) Arora, K.; Brooks, C. L., III. *J. Am. Chem. Soc.* **2009**, *131*, 5642–5647.
(46) Jonsson, H.; Mills, G.; Jacobsen, K. W. In *Classical and Quantum Dynamics in Condensed Phase Simulations*; Berne, B. J., Ciccotti, G., Coker, D. F., Eds.; World Scientific: Singapore, 1998.

the measured K_i value, the lack of experimental data on the rate of the conformational change does not allow a direct assessment of the reliability of the barrier values. Some considerations can, however, be done on the basis of the free-energy profile associated with the closure of the ligand-bound YeiK with neutral H82. The highest energy barrier (~18 kcal/mol) is comparable with that calculated from the steady-state kinetic rate constant ($\Delta G^{\ddagger} \sim 17$ kcal/mol for uridine as evaluated from k_{cat}[28]), suggesting that the conformational change could significantly influence the steady state catalytic rate of YeiK. This is in agreement with the slow onset observed for tight-binding inhibitors of group I NHs.[47] On the basis of our current findings, this behavior could be ascribed to the rate of isomerization between the loosely bound and tightly bound forms of the enzyme. We want to underscore that a strict comparison of the observed energy barrier with the catalytic rate is made difficult by the uncertainty of the theoretical value (Table 1) and by the different conditions of the two measurements. Furthermore, one should also take into account that the rate of the conformational change can be influenced by the diffusion of the system over the free energy landscape.[45] Moreover, the pathway for the conformational change is here derived by minimizing the perturbation onto the system, but still we cannot exclude the existence of alternative pathways associated with lower energy barriers.

The order of magnitude of the free-energy changes along the profile still suggests that factors regulating the conformational change, such as the pH or the identity of the residues involved, can influence the rate of the overall process. The k_{cat} value of group-I NHs displays a bell-shaped pH-dependence with a maximum at pH = 8, requiring at least one deprotonated ($pK_a = 7.1$) and one protonated ($pK_a = 9.1$) amino acid residue for efficient catalysis.[29] The deprotonated residue has been identified as the aspartate corresponding to D12 in YeiK, which extracts a proton from the incoming nucleophilic water.[27] The residue H239 is instead considered to be the general acid that stabilizes the leaving group, possibly through direct protonation.[30] The role of the strictly conserved H82 has not been thoroughly investigated. Our results show that the free-energy profile along the conformational change is significantly influenced by the charge of H82, suggesting that its protonation can trigger and/or regulate the structural transition. Indeed, the charged state of H82 induces larger barriers to the closure, while the closed form becomes thermodynamically unfavored (Figure 6). This finding would support a mechanism that couples the conformational change of YeiK with the enzymatic reaction, where the acquisition of a proton by H82 during the hydrolysis of the substrate in the closed form would "trigger" the opening of the enzyme and the release of the products in solution. The loss of the proton by the open L1 and the binding of a new substrate would then promote the closure of the protein and the beginning of a new catalytic cycle. The proposed regulation mechanism would exploit the specific ability of the histidine residue to be involved in hydrophobic interactions when neutral and in salt bridges when protonated. In the 7−9 pH interval, one can assume that the charge state of the histidine may be modulated by perturbations in its environment induced by the conformational change or by the progress of the reaction. However, in our calculations, we did not explicitly describe the protonation process, so that the driving force for H82 protonation and deprotonation remains unknown. Moreover, the protonated protein was assumed to follow the same pathway as the neutral ligand-bound protein in the profile determination (section 3.4). Further studies on the catalytic reaction mechanism and on site-specific mutants, both *in vitro* and *in silico*, will aid in the definition of the specific role of H82.

5. Conclusions

Here, we have introduced a new computational procedure for the determination of free-energy profiles along conformational pathways connecting two different macromolecular structures. The proposed sequence of calculations involves the characterization of the intrinsic dynamics of the starting point with unbiased MD simulations, the determination of transition pathways using Targeted Essential Dynamics Sampling, and the reconstruction of the free-energy changes between representative points along the pathway through nonequilibrium work distributions. The accuracy of the resulting free-energy profiles has been validated by comparison with experimental data, where we have shown that the agreement is within the uncertainty of the theoretical calculation (3 kcal/mol).

This combination of methods can be used in a variety of cases, especially when the complexity of the transition prevents the use of a few simple collective variables. Indeed, the system we have chosen for its application, the nucleoside hydrolase YeiK from *E. coli*, undergoes a complex rearrangement of two loops upon ligand binding, as shown from the X-ray structures here determined for the apo and inhibitor-bound form of the protein. The transition pathway resulting from our calculations underscores the cooperativity in the closure of the two loops and the involvement of transient inter- and intraloop interactions not detectable from the simple comparison of the two endpoints. Moreover, the calculated free-energy profiles support a coupling of the conformational change with the enzymatic reaction mechanism, where loop closure and opening would be regulated by the charge state of a histidine residue in the enzyme active site.

In conclusion, we believe that this approach may be applied to extract both qualitative and quantitative information in a wide range of reorganizations in macromolecular structures. Knowledge of the energetic contribution of the conformational changes to inhibitor binding could also be of great benefit for the process of drug design, allowing a more comprehensive description of the binding process in the scoring functions used.

Acknowledgment. This work was supported by research grants from NOBEL Cariplo and FIRB-MIUR (to M.D.).

Supporting Information Available: Text and tables with supplementary data on structure determination and simulation parameters and protocols. Plots of the backbone rmsd of YeiK during the unbiased MD simulations. Details on the cluster and principal components analyses performed onto the MD trajectories. Geometrical description of the ligand-free and ligand-bound TEDS trajectories. Details on the calculation of free energies of binding of DAPIR to YeiK: thermodynamic cycles and single contributions to the total free-energy change. Work distributions for a selected TMD transformation. Movie of the open-to-closed transition of the ligand-bound form of YeiK during a representative TEDS simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

JA107640U

(47) Horenstein, B. A.; Schramm, V. L. *Biochemistry* **1993**, *32*, 9917–9925.