MM/PBSA Analysis of Molecular Dynamics Simulations of Bovine β-Lactoglobulin: Free Energy Gradients in Conformational Transitions?

Federico Fogolari,[1]* Elisabetta Moroni,[2] Marcin Wojciechowski,[3] Maciej Baginski,[3] Laura Ragona,[4] and Henriette Molinari[1,4]

[1]*Dipartimento Scientifico e Tecnologico, Università di Verona, Verona, Italy*
[2]*Dipartimento di Fisica G. Occhialini, Università di Milano-Bicocca, and INFN, Sezione di Milano, Milano, Italy*
[3]*Department of Pharmaceutical Technology and Biochemistry, Gdansk University of Technology, Gdansk, Poland*
[4]*Laboratorio NMR, Istituto per lo Studio delle Macromolecole, CNR, Milano, Italy*

ABSTRACT **The pH-driven opening and closure of β-lactoglobulin EF loop, acting as a lid and closing the internal cavity of the protein, has been studied by molecular dynamics (MD) simulations and free energy calculations based on molecular mechanics/Poisson–Boltzmann (PB) solvent-accessible surface area (MM/PBSA) methodology. The forms above and below the transition pH differ presumably only in the protonation state of residue Glu89. MM/PBSA calculations are able to reproduce qualitatively the thermodynamics of the transition. The analysis of MD simulations using a combination of MM/PBSA methodology and the colony energy approach is able to highlight the driving forces implied in the transition. The analysis suggests that global rearrangements take place before the equilibrium local conformation is reached. This conclusion may bear general relevance to conformational transitions in all lipocalins and proteins in general. Proteins 2005;59:91–103.** © 2005 Wiley-Liss, Inc.

Key words: β-lactoglobulin; loop dynamics; MM/PBSA; Poisson-Boltzmann; colony energy

INTRODUCTION

Mobile loops often correspond to protein sites implicated in functions. The characterization of conformational transitions involved in events such as ligand binding, product release, and allosteric changes is therefore of primary importance. Along this line, we have shown for the first time that bovine β-lactoglobulin (BLG), an 18-KDa β-barrel lipocalin composed of 9 antiparallel β-strands defining an internal cavity, modulates its binding properties through the reversible opening–closure of one loop (EF loop, residues 85–90) acting as a lid that closes the protein interior binding site below pH 7 and opens it at higher pH.[1] EF loop conformational transition, first detected by Tanford and Nozaki[2] using optical rotatory dispersion, is triggered by the titration at unusually high pH (~7) of the buried carboxyl group of E89, located within the loop at the protein open end. Molecular details of this conformational change have recently been revealed from X-ray structures by using crystals grown at pH values on either side of the transition,[3] sometimes referred to as the Tanford transition, and explained by electrostatic calculations.[4] The titration of E89 side-chain at high pH (~7) involves a remarkable rearrangement of the H-bond pattern involving L87, E89, N90, E108, N109, S110, and S116 residues, and causes the fold back of the EF loop, with a consequent solvent exposure of glutamic acid side-chain (Fig. 1), changing its solvent accessible area from 4 Å^2 at pH 6.2 to 116 Å^2 at pH 7.1, and 136 Å^2 at pH 8.2.[5]

Our studies[1] on the role of this conformational rearrangement on the functional properties of BLG revealed that this feature is common to all BLGs. Indeed, E89 is conserved in all BLGs from different species, and we have recently shown that its titration acid dissociation constant (pKa) is influenced by the nature and charge of nearby residues, thus regulating the binding pH. A few observations reported for other lipocalins, such as tear lipocalin,[6] bovine retinol-binding protein,[7] nitrophorin,[8] and membrane enzyme PagP,[9] suggest that conformational rearrangements, mainly involving loops at the protein open end, could be a general feature of the lipocalin ligand-binding mechanism, somehow reminiscent of lipases' interfacial activation. Given the important functional role of this conformational transition in lipocalins and, more generally, of conformational transitions in proteins, it is of primary importance to identify a reliable approach for the computation of free energy differences between several conformations of the same biomolecule. Indeed disentangling free energy contributions arising from different interactions allows for better understanding and biomolecular rational design of ligands.

Implicit solvent methods[10,11] that compute conformational free energies offer several advantages over other simulation methods. In particular, (1) averaging over

Grant sponsor: NATO (Collaborative Linkage Grant to F. Fogolari, M. Wojciechowski, and M. Baginski). Grant sponsor: MIUR (Italy) ex 40% 2002. Grant sponsor: FIRB 2001.

*Correspondence to: Federico Fogolari, Dipartimento di Scienze e Tecnologie Biomediche, Universitá di Udine, P.le Kolbe 4, 33100 Udine, Italy. E-mail: fogolari@dstb.uniud.it; http://www.dstb.uniud.it/~fogolari

Received 7 May 2004; Accepted 3 September 2004

Published online 2 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20384

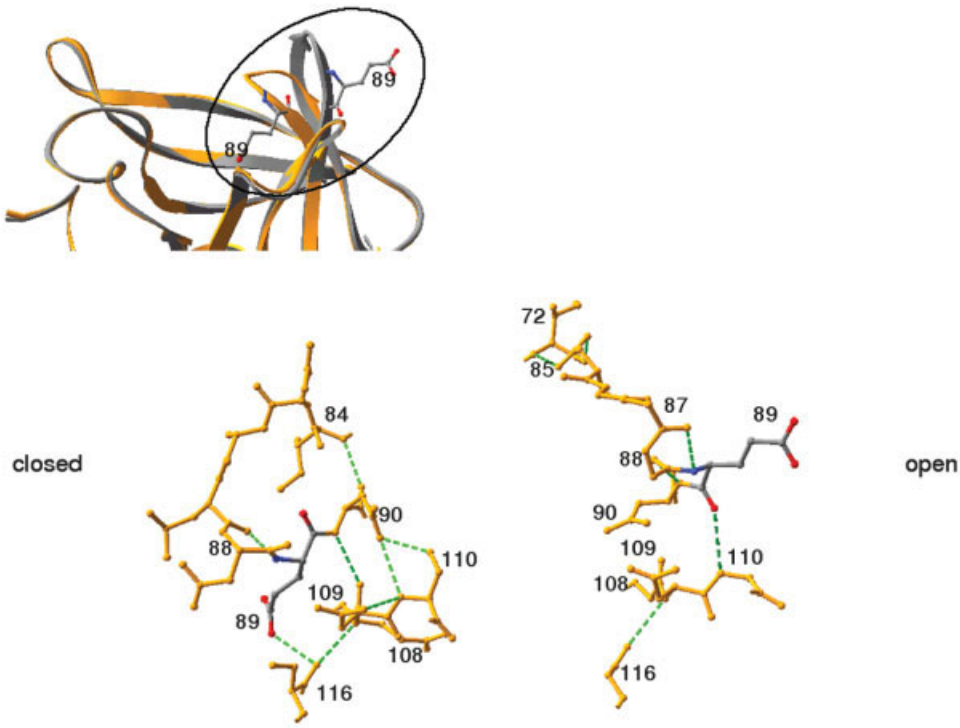

Fig. 1. BLG open end with the EF loop.

solvent degrees of freedom is implicit and therefore, no long-time sampling of solvent dynamics is needed, which is particularly important for ions whose equilibration times may be much longer than commonly performed simulations; and (2) limitations in the size of the simulated system are much less severe.

Among implicit solvent methods, the molecular mechanics/Poisson–Boltzmann solvent-accessible surface area (MM/PBSA) method is rapidly becoming one of the most popular because of its good statistical mechanics foundations and overall good accuracy.[12–16]

Treatment of entropic effects due to solute degrees of freedom is not easy and requires sampling of the conformational neighborhood for each sampled conformation. The heuristic solution proposed by Honig and coworkers[17] has proven its effectiveness in the context of loop predictions. An additional benefit is that the approach is able to smooth fluctuations of the computed energy, adding robustness to free energy estimation.

In a previous study, Eberini et al.[18] performed 1-ns molecular dynamics (MD) simulations for the equilibrium protonated and unprotonated form of BLG, and a 3-ns nonequilibrium MD simulation of the unprotonated form, starting from the conformation adopted by the protonated form. The article by Eberini et al.[18] was focused on the calculation of pKa's and the computation of the reorganization energy following protonation of Glu89.

For 20 snapshots taken from these trajectories, the pKa's of titratable groups were computed according to a continuum electrostatics/Monte Carlo method. Use of Linear Response Approximation theory allowed the same authors to estimate the so-called reorganization energy upon protonation and the pK^{half} of titratable sites (see definition of pK^{half}[18]). The computed pK^{half} reported in the article was consistent with the pKa computed earlier using the crystallographic structures and 20 NMR structures.[4] Differences are mainly due to the usage of a different

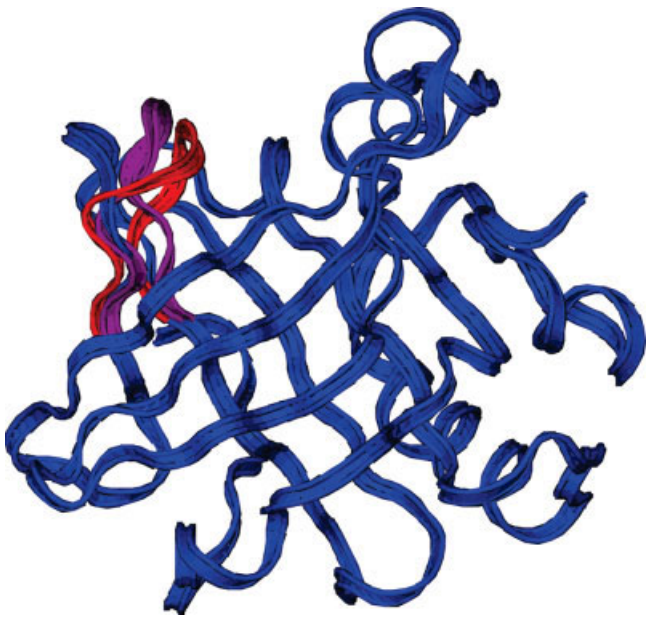

Fig. 7. The snapshot taken at time 2.96 ns from MD_open_COOH (in purple) is shown together with the open (in blue) and closed conformation (in red). The C_α carbons of the open and closed structures have been superimposed on the snapshot structure. Only residues 84–90 are shown.

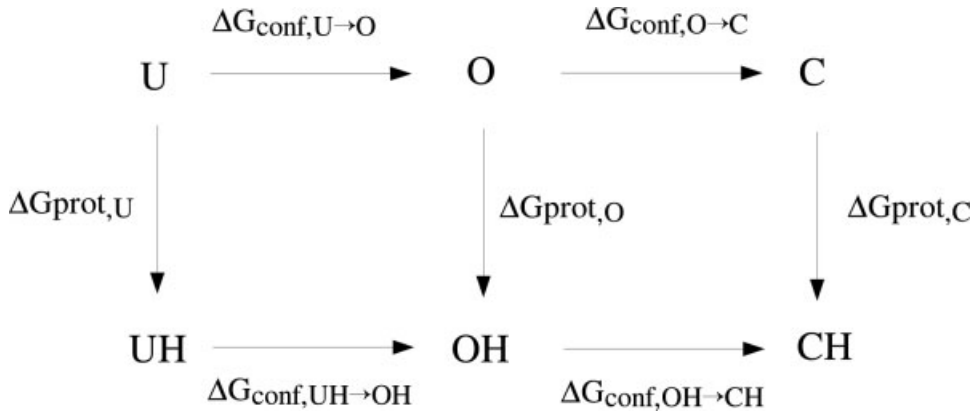

Fig. 2. Thermodynamic cycle corresponding to the protonation of Glu89 in BLG and to the interconversion of unfolded (U), folded and open (O), and folded and closed (C) conformations. The protonated forms are indicated by a H.

dielectric constant that was 20[19] in our study[4] and 8 in the study of Eberini et al.[18]

During the nonequilibrium simulation of the deprotonated form, starting from the equilibrium conformation of the protonated form, a movement of the EF loop of the protein was observed.

In order to better characterize the thermodynamics of this conformational transition, we performed 3-ns equilibrium and nonequilibrium MD simulations of both the protonated and unprotonated forms of BLG starting from the 2 alternative conformations of the EF loop.

From each simulation, 300 snapshots are obtained and analyzed using the MM/PBSA methodology complemented by the colony free energy approach. The free energy is analyzed in terms of different electrostatic and nonelectrostatic contributions.

The results obtained in the nonequilibrium simulations are complementary to those of Eberini et al.[18] in that we observe the reverse movement, but not the one described by these authors.

The analysis of MD simulations, which was performed thoroughly by Eberini et al.,[18] will not be repeated here; rather, we discuss local and global aspects of the transition.

The tendency toward the equilibrium conformation found in the simulation is analyzed in terms of the MM/PBSA free energy gradient. This gradient does not have any counterpart in local trends of backbone torsion angles to adopt the equilibrium conformation.

Since the conformations we are characterizing are linked to protonation of an acidic residue, it is worth noting here the difference between MM/PBSA free energy estimation and standard pKa shift calculations. Following a well-established procedure,[19–21] we use a simple thermodynamic cycle that describes not only the protonation of Glu89 but also the interconversion between 3 conformations (unfolded, open, and closed; Fig. 2). Typically in pKa calculations, free energies of processes corresponding to the vertical arrows in Figure 2 are computed. In the PB methodology, reference is made to the unfolded conformation, where pKa's are assumed to be those of small model compounds.

On the contrary, in MM/PBSA, only free energies of processes corresponding to horizontal arrows in Figure 2 are computed.

In the present article, MM/PBSA free energy estimation is used in conjunction with the colony energy to study the conformational transition of the EF loop in bovine BLG. The results obtained bear general relevance to conformational transitions in proteins.

MATERIALS AND METHODS

MD Simulations and Molecular Models

MD simulations were performed using the program GROMACS[22,23] (version 3.1.4) with the GROMACS all-hydrogens forcefield (ffgmx2).

Initial atomic coordinates were taken from 2 X-ray crystal structures of bovine BLG deposited in the Brookhaven Protein Data Bank[24] (PDB), one obtained at pH 6.2 (PDB ID: 3blg, resolution 2.56 Å), and the other at pH 8.2 (PDB ID: 2blg, resolution 2.46 Å).[3] Every structure of BLG was implemented both with the unprotonated and the protonated forms of Glu89.

The protein was energy-minimized in vacuo using 200 steps of steepest descent and an additional 200 steps using conjugate gradients. A cutoff of 12 Å was used for Coulomb and van der Waals interactions. A preliminary pKa calculation[4,19] was performed for both starting structures. For no acidic residue except Glu89, the pKa was found close to neutrality. Only pKa's for histidines were found to be close to 7.0. Therefore, all titratable residues, except histidines, were assumed to be in their ionized form (i.e., all aspartic and glutamic acid residues and C-terminal carboxyl had a net negative charge, and all lysines, arginines, and N-terminal amino group had a positive charge). The ionization states of histidine residues were determined by the program pdb2gmx of the GROMACS package. The protonation state of histidines is assigned by optimizing the global hydrogen-bond network following a simple geometric criterion.[23]

The net negative charge of the system has been neutralized by placing a corresponding number of sodium counterions in energetically favorable positions. The electrostatic potential was calculated via numerical solutions of the PB equation using the University of Houston Brownian Dynamics (UHBD, version 6.x) program.[25,26]

The PB equation was solved using standard parameters and 150 m*M* ionic strength. The first ion was placed at the lowest potential point of a surface at ~7 Å away from any atom of the protein, in order to allow later for proper hydration of the ions. The surface was generated using the program MSMS,[27] inflating all van der Waals radii of protein atoms to 5.0 Å, and using a probe sphere with a radius of 2.0 Å. The calculations were performed again, and the ions were placed following the same procedure, to start from a rather stable system. Nine sodium ions (or 8, when the structure of BLG was implemented with the protonated form of the Glu89 residue) were added to neutralize the −9 (or −8) net charge of the protein. The rationale for this procedure was to start from a stable configuration of the ions. The protein was put in the middle of a cubic box with periodic boundary conditions, with the distance between the protein and the box set at 2 nm. The protein, whose dimensions are ~40 Å in each direction, and 9 (or 8) counterions were solvated in the cubic box with water. Solvent molecules are removed from the box whenever the distance between any atom of the protein and any atom of the solvent molecule is less than the sum of the Van der Waals radii of both atoms. The simple point charge (SPC) water model was used to describe the solvent. The resulting system contained 2596 (2595 for deprotonated Glu89) protein atoms and ~49,500 solvent atoms, depending on the protein starting conformation and protonation state. After addition of solvent molecules and ions to the system, long-range electrostatic interactions were treated by particle mesh Ewald (PME) method with the following parameters: distance for the Van der Waals cutoff = 12 Å, distance for the Coulomb cutoff = 12 Å, spacing for the fast Fourier transform (FFT) grid = 1.2Å.

The protein and the ions were fixed, and the solvent was energy minimized using 100 steps of steepest descent. A solvent equilibration was carried out by performing an MD simulation for 50 ps using a 1-fs timestep to let the water molecules move to adjust to the conformation of the solute. After solvent equilibration, the system was energy-minimized using 200 steps of steepest descent. The temperatures of protein and solvent with counterions were then separately coupled, through a Berendsen-thermostat, to a bath with temperature 300 K and time constant 0.1 ps. The pressure coupling of the system was based on an exponential relaxation, with time constant 0.5 ps and compressibility 4.5×10^{-5} bar^{-1}. Atom velocities for startup runs were set to zero.

Finally, after a 100-ps equilibration, we performed a 3-ns MD simulation of the whole system using a 2-fs timestep. A snapshot of the trajectory was stored every picosecond (500 timesteps) for later analysis. The Lincs algorithm was used[28] in order to restrain bonds at their equilibrium value and to extend the simulation timestep. The simulations were performed partly on a single 2.8 GHz Pentium IV processor personal computer, and partly on a cluster composed of 5 dual-processor nodes based on Intel XeonTM 2.8 GHz, with hyperthreading technology.

MM/PBSA Methodology

The MM/PBSA methodology used here and the tests of the protocol, performed in the context of native loop recognition, have been thoroughly described elsewhere.[28a] Essential features of the methodology are reported hereafter. The free energy of each conformation sampled from MD simulation less the entropic contribution due to the protein degrees of freedom is expressed as a potential of mean force *W*, written as the sum of the intramolecular energy of the protein $U(\bar{r}_1,\ldots,\bar{r}_n)$ and a solvation free energy term that is further split in a polar (electrostatic) and a nonpolar (hydrophobic) term[10,11]:

$$W = U(\bar{r}_1,\ldots,\bar{r}_n) + \Delta G^{\text{polar}}(\bar{r}_1,\ldots,\bar{r}_n) + \Delta G^{\text{nonpolar}}(\bar{r}_1,\ldots,\bar{r}_n). \quad (1)$$

The entropy of the protein is neglected altogether here based on the assumption frequently made in similar studies that the entropy of similar compact conformations is the same.

The solute energy term $U(\bar{r}_1,\ldots,\bar{r}_n)$ has been computed using CHARMM (version 27b2), a classic and well-tested molecular mechanics force field.[29,30] All structures have been energy-minimized in a 2-step fashion. First all Cα coordinates have been fixed and energy-minimized with 20 steepest descent and 30 conjugate gradients steps, in order to remove high energy spots, like van de Waals clashes. During this minimization, a distance-dependent dielectric constant has been employed ($\epsilon = 1r$) with a cutoff of 12 Å. In the second step, the structure has been totally unrestrained and a more accurate solvent model (the generalized Born solvent-accessible model) has been employed using the parameters suggested by Qiu et al.,[31] in order to reduce artifacts ensuing from missing solvent. First 50 steepest descent minimization steps have been performed, followed by 200 conjugate gradients minimization steps. After minimization, the energy was evaluated with no cutoff and with dielectric constant 1.0. The same value was used for consistency in the PBSA solvation energy estimation.

ΔG^{polar} has been computed according to the PB theoretical framework[32–35] as the difference in free energy for the hypothetical charging process of the solute in vacuo and in ionic solvent. All calculations have been performed using the program UHBD.[25,26]

The following protocol has been used: First, the linear PB equation is solved on a large grid, with 40^3 grid points spaced 2.5 Å apart, centered on the protein in order to obtain boundary conditions for the following focusing steps. Then the same equation is solved on smaller and finer grids (33^3 points spaced 0.45 Å apart) centered on each amino acid center of geometry. In each of these focusing steps, the potential at each atomic position of the amino acid is stored and used later to estimate the

electrostatic free energy of the protein, which is expressed for the linear PB equation as $\Delta G = \sum_i \frac{q_i \phi_i}{2}$. The linear PB equation is appropriate for moderately charged biomolecules.[36] In order to speed up calculations, the inner dielectric constant was assigned to space regions within van der Waals radii. Dielectric boundaries have been smoothed according to the procedure of Davis and McCammon.[37] The solvent-accessible surface was generated by inflating all atomic radii by 1.4 Å and considering the inflated van der Waals surface. $\Delta G^{nonpolar}$ is taken to be proportional to the solvent-accessible surface area A (i.e., $\Delta G^{nonpolar} = \gamma A$, where the surface tension coefficient γ has been taken equal to 20 cal mol^{-1} Å^{-2}, which is consistent with the methodology employed here.[38]

Note that MM/PBSA free energies reported hereafter are extremely large, because they are given relative to an unphysical reference state. Through thermodynamic cycles, free energies corresponding to real processes, like conformational transitions, can be obtained.

Colony Energy Approach

The ensemble of snapshots obtained from MD simulations is, in principle, a thermodynamic ensemble. Theoretical and numerical inaccuracies of both MD and PBSA methodologies, however, lead to large fluctuations in the computed MM/PBSA energy. We found it useful to apply here the colony energy approach proposed by Honig and coworkers.[17] The approach was developed in order to account heuristically for entropic effects in native loop conformation predictions. The approach has the outstanding benefit of smoothing energy fluctuations, if the energetic model is accurate and if reasonably extensive conformational sampling is available, leading to a good correlation between energy and root-mean-square deviation (RMSD) from native structure.

It is assumed that each conformation is a sample of a colony of similar conformations. The free energy of each conformation i takes into account both the number and energy of similar structures that are available in the ensemble of all sampled conformations j, according to the following equation:

$$\Delta G_i^C = -kT \ln \frac{\sum_j \exp\left(\frac{-W(r_1^j,\ldots,r_n^j)}{kT} - \frac{\mathrm{RMSD}_{ij}^3}{t^3}\right)}{\sum_j \exp\left(\frac{-W(r_1^j,\ldots,r_n^j)}{kT}\right)}, \quad (2)$$

where the parameter t defines the scale for similarity and directly controls the smoothing of energy fluctuations in the ensemble.

The RMSD considered here is restricted to loop EF region (i.e., residues 84–90) and t has been set to 0.5 Å. This is an acceptable value for defining 2 loops as similar and should not introduce artifacts, because the RMSD between most EF loop conformations is larger than this value.

RESULTS

Comparison of Open (2BLG) and Closed (3BLG) Conformations

The crystal structures obtained at pH 6.2 and 8.2, referred to as the closed and open conformations, respectively, have been compared in order to localize the boundaries of the region involved in the pH-driven transition and in order to check for other differences between the two structures.

These structures have been chosen because they have been obtained by the same group during a single study of the pH-driven transition and should be therefore best suited for comparison.

All backbone atoms have been superimposed, and the contributions of single residues to global RMSD have been sorted. Residues 86–90 exhibit the largest RMSD (1.7–6.5 Å) and overall, all of residues 84–90 display rather large RMSDs. Other regions that have rather large RMSDs, besides terminal residues, are encompassing residues 62–65 and 108–115 (in the range of 0.5–1.0 Å). It is worth noting that segment 108–115 contains residues involved in the formation of H bonds with residues of EF loop. The RMSD in the latter regions parallels the higher temperature factors in both structures. It is interesting that the region involved in the pH-driven transition has also the highest temperature factors measured on the C_α atoms, which implies backbone flexibility (see Fig. 3). Interestingly, the EF loop backbone is not particularly mobile in solution, on a nanosecond to picosecond timescale.[39] However, residue 87 appears to be undergoing conformational exchange suggesting motion on an intermediate timescale, thus explaining the B-factor observed by X-ray in both open and closed BLG conformations.

Backbone torsional angles have been compared in order to assess which and how many residues change their local conformation to bring about the closure of the calyx lid. The largest changes are observed at the ψ angles of residues 86–90 and in the ϕ angles of residues 88, 84, 90, and, to a lesser extent, residue 86. As far as solvent accessibility is concerned, the largest changes are observed, as expected, for residue Glu89, which, upon deprotonation, becomes more solvated, and for hydrophobic Leu87, which exposes its hydrophobic side-chain to solvent. All these features will be considered when we discuss MD trajectories.

Hydrogen-bonding patterns are different, as expected, at the EF loop in the open and closed conformation. In the open conformation, Asp85 OD1 is hydrogen-bonded with Ile72 HN, while in the closed conformation, this is not possible, and Ile72 O is pointing in the direction of Asp85 HN. Glu89 O is hydrogen-bonded with Ser110 HN in the open conformation, while a similar stabilizing hydrogen bond can be established by Asn90 OD1 and Ser110 HN in the closed conformation. With minor rearrangements, the protonated carboxylic group of Glu89 in the closed conformation can be involved in a hydrogen bond with the γ hydroxyl group of Ser116, either as a donor or as an acceptor.

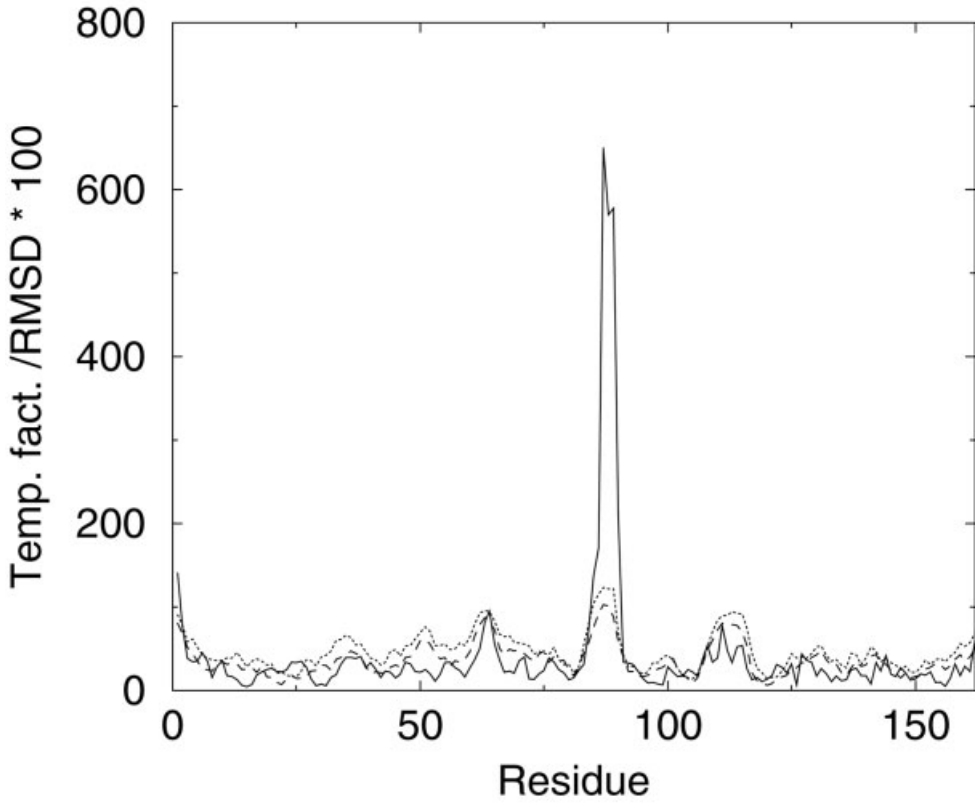

Fig. 3. Temperature factors on C_α carbons (open conformation: dotted line; closed conformation: dashed line) and backbone RMSD (multiplied by a factor of 100) between the open and closed conformations (continuous line).

A hydrogen bond is also established in the closed conformation between Asn90 HN and Glu108 O that cannot be established in the open conformation, while in both structures, Glu108 O is involved in a hydrogen bond with Lys91 HN.

The 2 structures have been compared also with other BLG structures determined by X-ray crystallography. It must be noted that there are some differences in backbone torsion angles among the different structures. We restricted the analyses to residues 84–90. When the structures for apo-forms of BLG are divided into open (PDB ID: 1b8e, 1bsy, 1qg5, 2blg) and closed (PDB ID: 1beb, 1bsq, 1mfh, 3blg) conformation structures, large variations (even larger than 100°) are observed for ϕ and ψ angles of residues 85–90 within the different open or closed conformations. In particular, the consistent unfavorable left-handed α-helix conformation at residues 87 and 88 is not present in structure 3blg, where residues adopt a different conformation.

Notwithstanding the large differences, all closed conformations have very similar overall arrangement at loop EF.

Analysis of MD Trajectories

For the sake of clarity in the following, we will refer to different MD simulation trajectories of bovine BLG as

MD_open_COO- —BLG with deprotonated Glu89 starting from open conformation

MD_close_COO- —BLG with deprotonated Glu89 starting from closed conformation

MD_close_COOH —BLG with protonated Glu89 starting from closed conformation

MD_open_COOH —BLG with protonated Glu89 starting from open conformation

The 2 simulations MD_open_COO- and MD_close_COOH are referred to as equilibrium simulations, while the 2 simulations MD_open_COOH and MD_close_COO- are referred to as nonequilibrium simulations.

First, local mobility was explored in order to assay the effect of starting conformation and/or the effect of Glu89 protonation state. The analysis was performed by global backbone superposition of all snapshots of the trajectory taken at 100-ps intervals, and assigning to each residue the average pairwise RMSD of the backbone atoms of that residue. This analysis has been performed using the program MOLMOL.[40] Regions that exhibit large positional fluctuations are immediately seen as peaks in an RMSD versus residue number plot (Figs. 4 and 5). The largest fluctuations are observed at the EF loop in both simulations, starting from the open conformation. All loop regions are, however, flexible in all simulations.

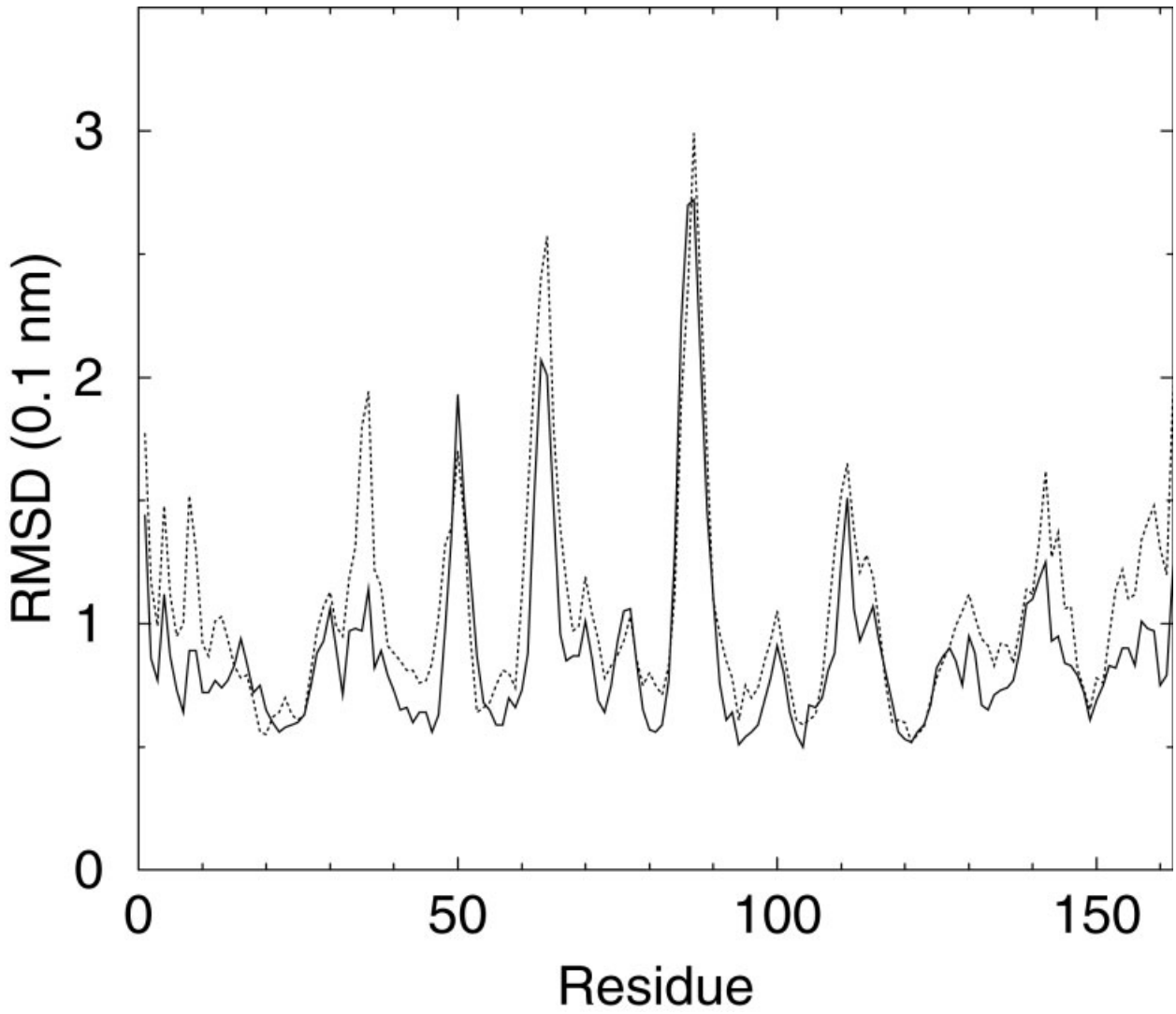

Fig. 4. Average local backbone RMSD between conformations sampled every 100 ps in simulations MD_open_COO- (continuous line) and MD_open_COOH (broken line).

The analysis of global and local RMSDs computed with respect to the starting equilibrium open or closed conformation of BLG provides information on local structural fluctuations and constitutes, by comparison with the crystal structures and the accompanying information, a check of simulation reliability.

The average RMSD from the closed conformation in MD_close_COOH is higher than the average RMSD from the open form in MD_open_COO-, but time sampling is not long enough to allow any safe conclusion to be drawn.

Both RMSD curves are anyhow between 1.0 Å and 2.0 Å, which hints at reliability of the MD simulation protocol applied. The RMSD from closed conformation is similar for both MD_close_COO- and MD_close_COOH, while the protonated form has a higher RMSD from open conformation in MD_open_COOH, rather than the deprotonated form in MD_open_COO-, which is in line with what was expected.

Besides local fluctuations, with respect to global superposition, the analysis of most conserved hydrogen bonds during dynamics may give clues to explain the determinants of stability. There are ~80 hydrogen bonds conserved in at least 20 out of 30 MD snapshots in all trajectories. In trajectory MD_open_COO-, there are almost 90 such bonds, confirming the largest stability observed in the simulation.

Hydrogen bonds involving the EF loop and closed residues are obviously different between simulations starting from open and closed conformation. In particular, the deprotonated form in the open conformation (MD_open_COO-) is stabilized by a Ser110 HN–Glu89 O hydrogen bond, which is found in 18 out of 30 snapshots of the simulation. The same bond is not so conserved for the protonated form (13 out of 30 snapshots of simulation MD_open_COOH), showing that changes induced by protonation start to take place during the 3-ns simulation.

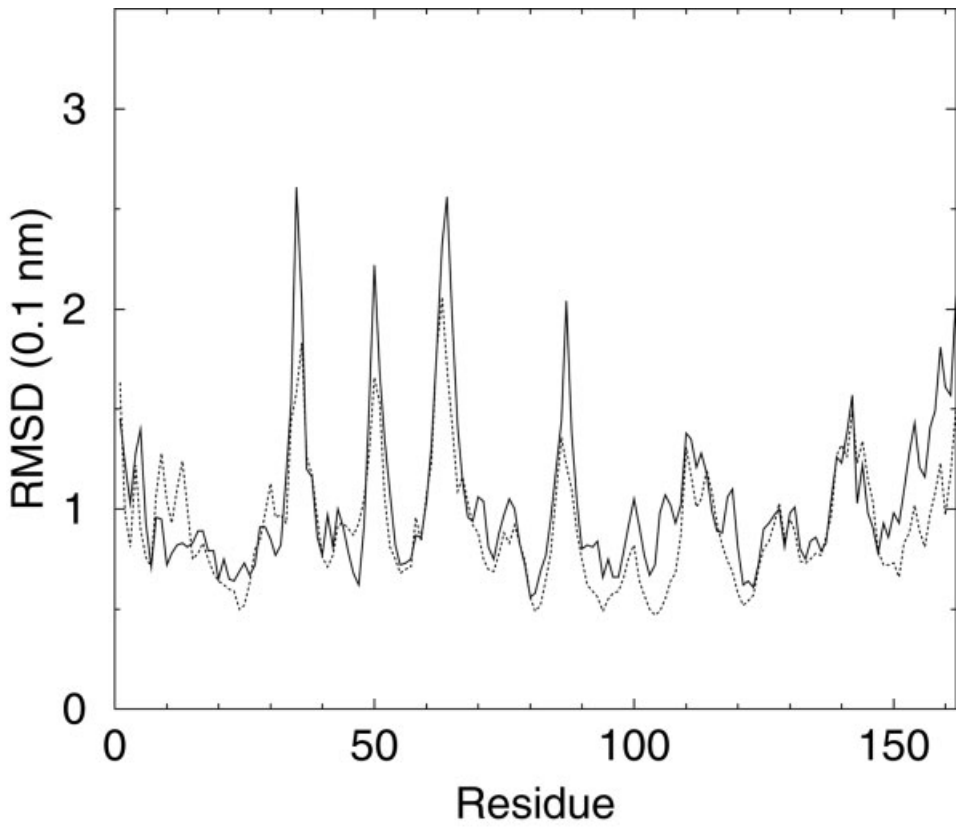

Fig. 5. Average local backbone RMSD between conformations sampled every 100 ps in simulations MD_close_COOH (continuous line) and MD_close_COO- (broken line).

Similarly, the protonated form in the closed conformation is stabilized by a Glu89 HE2–Ser116 O (in 26 out of 30 snapshots of MD_close_COOH) and an Asn109 HD21–Glu89 OE1 (in 24 out of 30 snapshots of MD_close_COOH) hydrogen bonds. The latter hydrogen bond is much less conserved for the deprotonated form in the closed conformation (14 out of 30 snapshots of MD_close_COO-). New hydrogen bonds between Ser116 HG and Glu89 OE1, and Asn88 HD21 and Glu89 OE1 form during simulations of the deprotonated form starting from close conformation (MD_close_COO-) and might play a role in loop opening, which, however, does not take place during simulation. Another important hydrogen bond involves Asn90 HN and Glu108 O. This is found in the protonated closed form and in 20 out of 30 snapshots for MD_close_COOH. In the simulation MD_close_COO-, this hydrogen bond is found only 2 times, showing that deprotonation is able to disrupt the hydrogen-bond network that stabilizes the equilibrium closed conformation for the protonated form.

Most conserved hydrogen bonds are related to secondary structure elements, which are more or less conserved during simulation. For each residue, the percentages of secondary structure (helix, sheet) have been plotted for all simulations. There are not large differences in secondary structure percentages except for the region close to the EF loop, where residue 89 is found to participate in the F strand for 55% of the simulation time in MD_close_COOH.

During the 3-ns simulation time, no transition from one form to the other is observed, although the behavior in RMSD clearly hints at a closure movement in the trajectory of the protonated form starting from the open conformation. In the last part of the trajectory, the RMSD computed on residues 84–90, after global C_α superposition drops from roughly 4.0 Å to 2.5 Å (Fig. 6).

This behavior is to be compared with the analogous curve computed on the trajectory obtained for the deprotonated form starting from the close conformation, where no sign of transition at all is seen.

In the snapshot taken at time 2.96 ns, the open and closed conformations superimposed on snapshot clearly shows that loop 84–90 is halfway between the 2 limit conformations (Fig. 7). However, when backbone torsion angles are examined, no transition from openlike values to closedlike value is found. Similarly, no H-bond involving the EF loop found in the protonated closed conformation is found in the trajectory of the protonated form starting from open conformation.

These observations suggest the idea that this conformational transition does not take place proceeding through gradual rearrangement, but rather overall movement takes place in a short time (here, less than 3 ns), and local conformational rearrangements take place in longer time according to the energy barrier for rotation around backbone torsion angles.

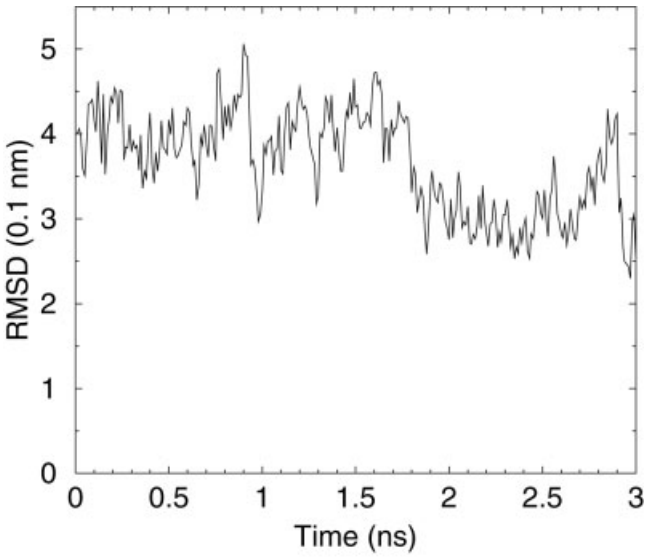

Fig. 6. RMSD between snapshots taken from simulation MD_open-_COOH (continuous line) and closed conformation computed on the C_α carbons of residues 84–90.

One may question how general this suggestion is. We believe that this 2-step procedure could occur quite frequently. A similar behavior has been observed for the reverse transition by Eberini et al.,[18] where overall similarity did not match local similarity, suggesting also in that case that backbone rearrangements follow the overall movement of the loop on a longer timescale than that sampled in the simulation.

A similar observation has been reported for BLG by Ragona et al.,[1] where the establishment of a nativelike hydrophobic cluster is observed in fragment peptides of BLG in solution. The detailed arrangement of the cluster is, however, slightly different from that found in the native structure of BLG. This suggests that local rearrangements must take place in the context of BLG structure, and it is likely that these functional rearrangements, involving backbone degrees of freedom, take place in times longer than those typically sampled by MD simulations.

MM/PBSA Free Energy Analysis

The free energy difference between the 2 conformations for the protonated and deprotonated states and the free energy profile of the pH-driven conformational transition involving the EF loop have been studied using the MM/PBSA model and the colony free energy approach.

Conformational free energy

The open and closed conformations have been treated as described in the Materials and Methods section, and the MM/PBSA free energy corresponding to the protonated and deprotonated forms has been computed. One problem with this straightforward approach is that the computed free energy is contributed also by different conformations of residues that are not related to the transition involving the EF loop. The free energy difference may therefore be dominated by these unwanted contributions. This is at variance with a similar study[28a] on protein loops, where framework residues were kept constant and their contribution to the free energy canceled out in comparison. We should compare only free energies related to conformations obtained by a single procedure (e.g., from MD simulations starting from the same conformer) or be able to cancel out unwanted contributions. This is particularly evident when computing the MM/PBSA free energy for the protonated and deprotonated forms of the open and closed conformations (PDB ID: 2blg and 3blg, respectively). When free energies are computed, we find for both protonated and deprotonated forms that the closed conformation is energetically more favorable, mainly due to electrostatic solvation energy. The overall electrostatic solvation energy in turn is not primarily connected to residues in the EF loop.

In order to circumvent these problems, we decided therefore to generate chimeric conformations where residues 83–91, entailing the EF loop, and all residues that have a heavy atom in van der Waals contact with residues 83–91 at a distance shorter than 1.0 Å are taken from one structure and all other (framework) residues from the other. The average for the original and chimeric structure for each conformation (open and closed) is considered. Thus, conformational effects, other than those involving the EF loop and contacting residues, and artifacts, which can be present in the chimeric forms, should be exactly the same for each conformation and cancel out in comparison. The results are reported in Table I. With this procedure, the open conformation for the deprotonated form has a free energy more favorable by 44 kcal/mol over the closed form. Upon protonation, the situation is the opposite, with the closed conformation favored by 17 kcal/mol over the open form. We can rationalize this analysis in terms of free energy contributions. In general, the difference in the term proportional to the surface area is not largely contributing, giving rise to a difference of 3.3 kcal/mol favoring the closed conformation with respect to the open conformation. The largest differences involve contributions in direct Coulombic energy and electrostatic solvation energy. For the deprotonated form, the solvation energy favors by 105 kcal/mol the open form, while Coulombic energy opposes the open conformation by 55.9 kcal/mol.

For the protonated form, the solvation energy still favors the open conformation, albeit only by 30 kcal/mol, while direct Coulombic interactions favor the closed form by 44 kcal/mol. These figures illustrate how energy differences between conformations are obtained by unbalance of rather large energy contributions.

Free energy of transition

The nonequilibrium simulations of the protonated open form starting from the open form and conversely the simulation of the deprotonated form starting from the closed conformation, if propagated at long times, and if the methodology is reliable, should show eventually the so-called Tanford transition. We could then analyze, using the MM/PBSA approach, the trajectory and completely

TABLE I. Free Energy (in kcal/mol) of the Open and Closed Conformations, in Protonated and Deprotonated Form

Model	RMSD open	RMSD close	MM/PBSA	MM	Vdw	Coul	PB	SA
Open COO-	0.4	4.5	−2995.8	−1292.0	−888.5	−1607.8	−1847.0	143.2
Open chimeric COO-	0.8	4.3	−3027.4	−1297.2	−898.3	−1619.1	−1871.0	140.8
Average open COO-	0.6	4.4	−3011.6	−1294.6	−893.4	−1613.5	−1859.0	142.0
Closed COO-	4.4	0.4	−3020.3	−1262.0	−904.6	−1581.4	−1896.0	137.7
Closed chimeric COO-	4.3	0.9	−2982.0	−1247.7	−892.6	−1568.4	−1874.0	139.7
Average closed COO-	0.4	3.9	−2968.3	−1353.0	−888.8	−1669.4	−1754.0	138.7
Open COOH	0.4	3.9	−2963.8	−1353.0	−888.8	−1669.4	−1754.0	143.2
Open chimeric COOH	0.7	3.8	−2988.2	−1359.0	−899.3	−1679.9	−1770.0	140.8
Average open COOH	0.6	3.9	−2976.0	−1356.0	−894.1	−1674.7	−1762.0	142.0
Closed COOH	3.9	0.3	−3014.4	−1412.0	−909.5	−1728.5	−1740.0	137.6
Closed chimeric COOH	3.8	0.7	−2971.3	−1391.0	−896.4	−1709.1	−1720.0	139.7
Average closed COOH	3.9	0.5	−2992.8	−1401.5	−903.0	−1718.8	−1730.0	138.7
Average COO- Closed→Open	3.9	0.5	−42.7	59.0	−4.6	55.9	−105.0	3.3
Average COOH Open→Closed	3.9	0.5	−16.8	−45.5	−8.9	−44.1	32.0	−3.3

RMSD (in Å) computed on C_α carbons of loop EF.

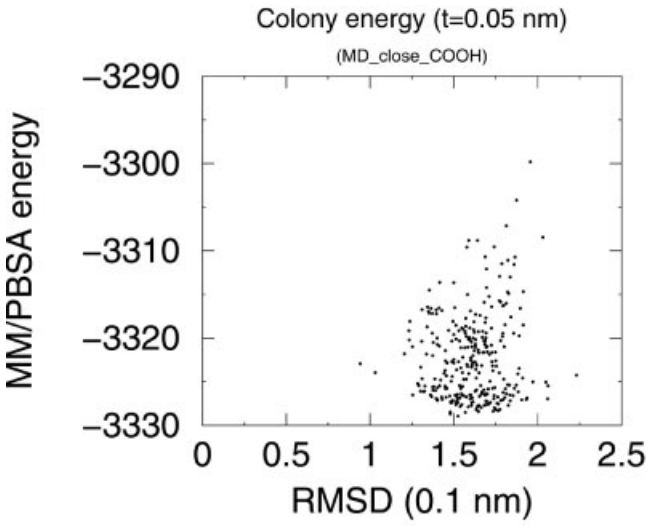

Fig. 8. Colony MM/PBSA free energy with threshold value $t = 0.5$ Å. The RMSD used for computing the colony energy refers to the superposition of C_α carbons of residues 84–90. The plot refers to snapshots taken from simulation MD_close_COOH. The reference structure is the closed conformation.

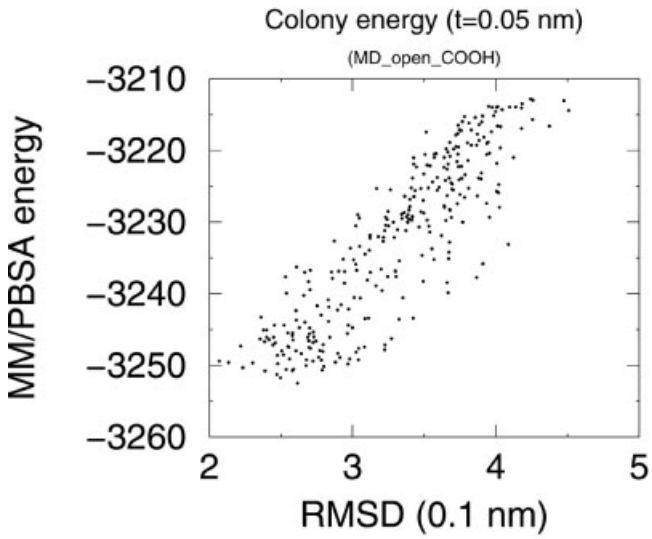

Fig. 9. Colony MM/PBSA free energy with threshold value $t = 0.5$ Å. The RMSD used for computing the colony energy refers to the superposition of C_α carbons of residues 84–90. The plot refers to snapshots taken from simulation MD_open_COOH. The reference structure is the closed conformation.

describe the free energy of the transition. Since the transition is not observed in the short time simulated here (3 ns for each trajectory), we took a different approach.

The simulations for protonated (deprotonated) forms starting from both closed and open forms could not be pooled together for MM/PBSA analysis, in order to cover the whole RMSD range, due to the problems mentioned in the previous section.

We decided therefore to analyze separately the free energy of each MD simulation.

For each simulation, we report here the colony MM/PBSA free energy (see Materials and Methods section) versus the RMSD of C_α carbons of residues 84–90 from the reference conformation, which is the open conformation for deprotonated form simulations and the closed conformation for the protonated form simulations.

Simulations of the protonated form

For simulation MD_close_COOH, no clear correlation between RMSD from reference conformation and MM/PBSA colony energy is seen, as could be expected due to the limited sampling of conformational space in an equilibrium simulation. The original differences in MM/PBSA energy, spanning a range of more than 250 kcal/mol, are greatly hampered by the colony energy approach, with the colony energy spanning at most 30 kcal/mol (see Fig. 8).

A conservative value of the threshold ($t = 0.5$ Å) has been chosen in order not to introduce artifacts connected with the use of the colony energy approach (see Materials and Methods section).

In simulation MD_open_COOH, there is a clear movement of the loop away from the open starting conformation

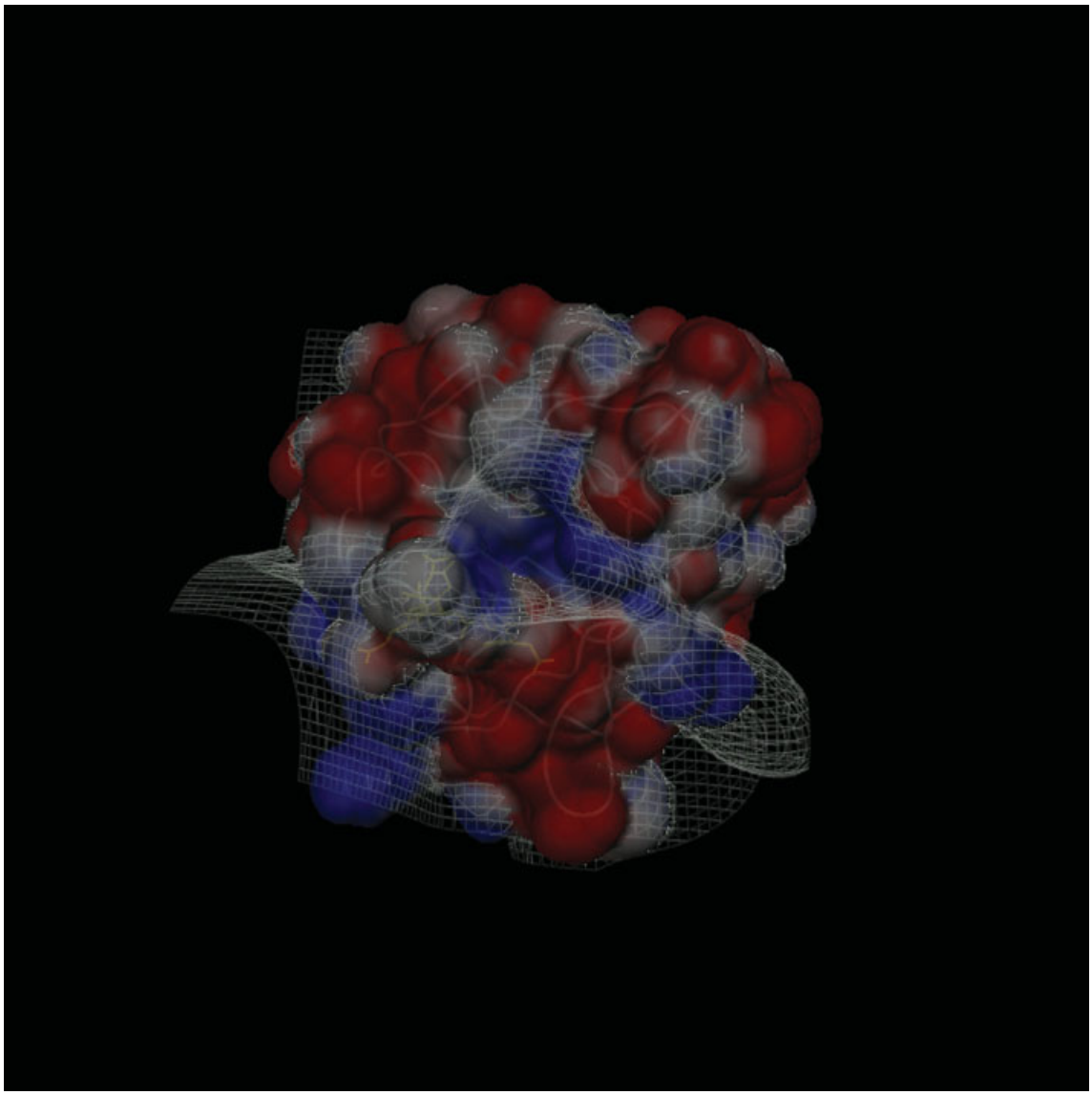

Fig. 10. Electrostatic potential at solvent accessible surface surface. Extreme blue and red colors correspond to +4 and −4 kcal e^{-1} mol^{-1}, respectively. The isopotential surface at 0 kcal e^{-1} mol^{-1} is shown in white. The view is from the top of the binding cavity.

toward the closed conformation. If any effect can be seen in the colony energy versus RMSD plot, it is likely to be more evident here than in simulations where conformations close to the equilibrium are sampled.

In the nonequilibrium simulation MD_open_COOH, the movement from the open conformation toward a closed conformation is paralleled by a free energy decrease, as is apparent from the plot in Figure 9. The correlation between the RMSD from the reference (closed) conformation and the colony energy is apparent and evidenced by the colony energy approach.

It is interesting to note that the low RMSD from the reference conformation does not imply, as discussed above, any local conformational similarity, as detailed by backbone torsional angles analysis. Therefore, the free energy profile in Figure 9 suggests that the real driving force of the transition is at a higher level than local. Backbone degrees of freedom in this picture simply adapt, on a longer timescale, to a transition, which takes place in order to minimize, compatible with local conformation, the free energy.

In order to better characterize the free energy gradient apparent in Figure 9 and reduce fluctuations, we considered the first and last nanosecond of the simulation. For both these two pieces of the trajectory we computed averages and standard deviations of selected energy terms. In particular, we considered MM energy terms involving atoms of the EF loop (residues 84–90) and solvation energies and buried surfaces computed only on EF loop atoms. It should be clear that the electrostatic potential at any atom of the protein depends on all other atoms in the protein that define the low dielectric region. The analysis should, however, point out differences between the first and last part of the simulation.

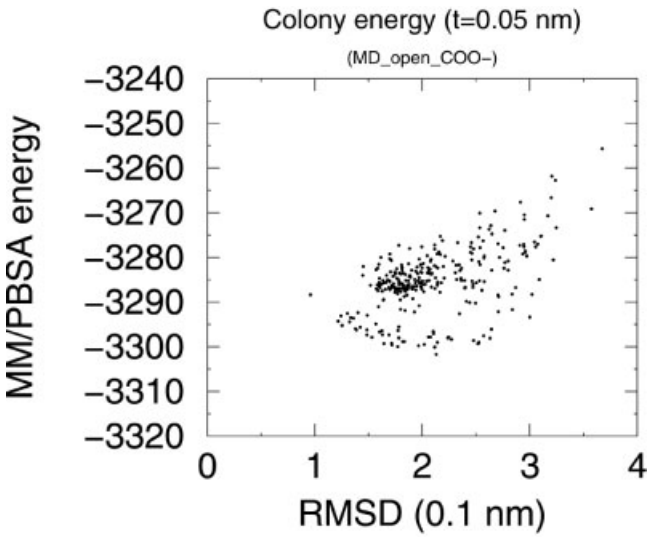

Fig. 11. Colony MM/PBSA free energy with threshold value $t = 0.5$ Å. The RMSD used for computing the colony energy refers to the superposition of C_α carbons of residues 84–90. The plot refers to snapshots taken from simulation MD_open_COO-. The reference structure is the open conformation.

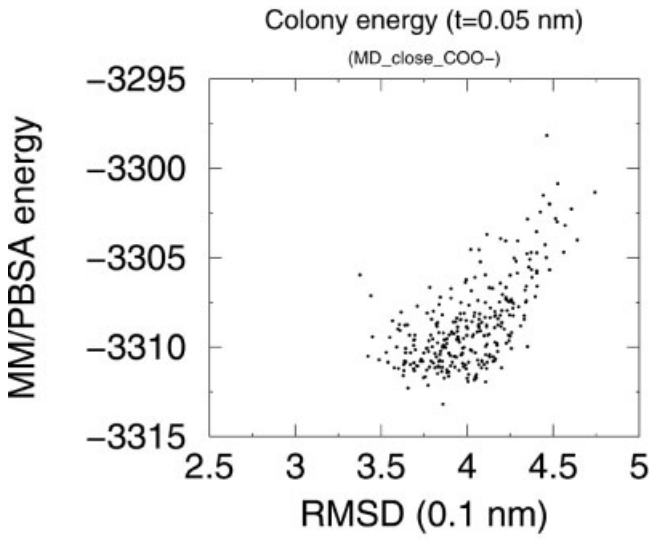

Fig. 12. Colony MM/PBSA free energy with threshold value $t = 0.5$ Å. The RMSD used for computing the colony energy refers to the superposition of C_α carbons of residues 84–90. The plot refers to snapshots taken from simulation MD_close_COO-. The reference structure is the open conformation.

The only relevant difference is connected with direct Coulombic energy and solvation energy. Both terms are more negative in the last part of the simulation compared to the first part, although averages are subjected to large fluctuations. This was at first unexpected, but it is consistent with the constant average surface exposed by residues 84–90 in the first and last part of the simulation.

In view of this result, we considered also the electrostatic potential in the close proximity of the EF loop in the open conformation (PDB ID: 2blg). It is known that cavities are usually associated with rather high potential due to poor solvation.[41] By displaying the zero isopotential curve, we can get an idea of how fast the potential drops to zero or, in other words, how large the electric field is. The zero isopotential surface and the electrostatic potential at the molecular surface are shown in Figure 10. It is seen that the isopotential curve forms a surface that encloses the entrance of the cavity. The electrostatic field is rather intense, because the potential drops from a value of several kcal $e^{-1}\text{mol}^{-1}$ to 0 in a rather short length. This picture is consistent with the analysis indicating large changes in electrostatic energy terms linked with movement of the EF loop. It is likely that such a strong electrostatic field plays a role in EF loop closure, which is later followed by optimal rearrangement.

Simulations of the deprotonated form

For simulation MD_open_COO-, a behavior similar to MD_close_COOH is observed, with a clear equilibrium situation established after roughly 500 ps. The correlation between colony free energy and RMSD is very weak (see Fig. 11).

The analysis of nonequilibrium simulation MD_close_COO- reveals features similar to the other nonequilibrium simulation MD_open_COOH, with an increase of free energy when moving farther away from the equilibrium conformation, even if the correlation is much less evident than for MD_open_COOH simulation (see Fig. 12).

The overall picture emerging from the nonequilibrium simulations (i.e., MD_open_COOH and MD_close_COO-) hints at the presence of free energy gradients, which are randomly explored by the structures by fluctuations. The precise movement of the conformation along the free energy gradient requires finer rearrangements in local structure, like in backbone torsion angles, which may need much longer time to take place.

CONCLUSIONS

Three-nanosecond explicit solvent MD simulations have been performed on 4 different models of the apo form of bovine BLG in order to simulate the opening–closure of the EF loop, the so called Tanford transition, which is functional to BLG ligand binding. The simulations employed the open and closed conformations (as determined by X-ray crystallography by Qin et al.[3]) with Glu89 protonated or deprotonated. The nonequilibrium simulations (MD_open_COOH and MD_close_COO-) do not display a clear cut conformational transition. The simulation of the protonated form starting from the open conformation, however, shows a clear rearrangement at the EF loop, which displays an RMSD from the closed conformation that is lower than that of the starting structure. When torsional angles are analyzed, it is seen that the global similarity with the target equilibrium structure is not paralleled by similarity in local conformations. The analysis of the free energy for the starting conformations and the conformations sampled by MD simulations reveals, however, a global free energy gradient toward more favorable equilibrium conformations. This situation is particu-

larly evident for the protonated form. The analysis of the MM/PBSA free energy for the starting structures is able to explain the thermodynamics of the pH-driven conformation of the EF loop. The suggestion that global rearrangements take place before the equilibrium local conformation is reached may bear general relevance to conformational transitions in all lipocalins and proteins in general.

ACKNOWLEDGMENTS

F. Fogolari wishes to thank Drs. G. Tecchiolli and P. Zuccato of Exadron, the HPC Division of the Eurotech Group, for providing hardware and expert technical assistance. Dr. I. Eberini is gratefully acknowledged for useful discussion on the manuscript.

REFERENCES

1. Ragona L, Fogolari F, Catalano M, Ugolini R, Zetta L, Molinari H. EF loop conformational change triggers ligand binding in beta-lactoglobulins. J Biol Chem 2003;278:38840–38846.
2. Tanford C, Nozaki Y. Physico-chemical comparison of beta-lactoglobulins A and B. J Biol Chem 1959;234:2874–2877.
3. Qin B, Bewley M, Creamer L, Baker H, Baker E, Jameson G. Structural basis of the Tanford transition of bovine beta-lactoglobulin. Biochemistry 1998;37:4014–4023.
4. Fogolari F, Ragona L, Licciardi S, Romagnoli S, Michelutti R, Ugolini R, Molinari, H. Electrostatic properties of β-lactoglobulin. Proteins 2001;39:317–330.
5. Taulier N, Chalikian TV. Characterization of pH-induced transitions of beta-lactoglobulin: ultrasonic, densimetric, and spectroscopic studies. J Mol Biol 2001;314:873–889.
6. Gasymov O, Abduragimov A, Yusifov T, Glasgow B. Structural changes in human tear lipocalins associated with lipid binding. Biochim Biophys Acta 1998;1386:145–146.
7. Newcomer M, Ong, D. Plasma retinol binding protein: structure and function of the prototypic lipocalin. Biochim Biophys Acta 2000;1482:57–64.
8. Roberts S, Weichsel A, Qiu Y, Shelnutt J, Walker F, Montfort W. Ligand-induced heme ruffling and bent no geometry in ultra-high-resolution structures of nitrophorin 4. Biochemistry 2001;40: 11327–11337.
9. Hwang PM, Choy W, Lo E, Chen L, Forman-Kay J, Raetz C, Prive G, Bishop R, Kay L. Solution structure and dynamics of the outer membrane enzyme PagP by NMR. Proc Natl Acad Sci USA 2002;99:13560–13565.
10. Roux B, Simonson T. Implicit solvent models. Biophys Chem 1999;78:1–20.
11. Honig B, Sharp K, Yang A-S. Macroscopic models of aqueous solution: biological and chemical applications. J Phys Chem 1993; 97:1101–1109.
12. Dominy BN, Brooks CLI. Identifying native-like protein structures using physics-based potentials. J Comput Chem 2002;23:147–160.
13. Feig M, Brooks CLI. Evaluating CASP4 predictions with physical energy functions. Proteins 2002;49:232–245.
14. Swanson J, Henchman R, McCammon JA. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. Biophys J 2004;86:67–74.
15. Kollman P, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Duan Y, Wang W, Donini O, Cieplak P, Srnivasan J, Case D, Cheatham TE. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
16. Vorobjev YN, Almagro JC, Hermans J. Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model. Proteins 1998;32:399–413.
17. Xiang Z, Cinque SS, Honig B. Evaluating free energies: the colony energy and its application to the problem of loop prediction. Proc Natl Acad Sci USA 2002;99:7432–7437.
18. Eberini M, Baptista A, Gianazza E, Fraternali F, Beringhelli T. Reorganization in apo- and holo-beta-lactoglobulin upon protonation of Glu89: molecular dynamics and pKa calculations. Proteins 2004;54:744–758.
19. Antosiewicz J, McCammon JA, Gilson MK. Prediction of pH-dependent properties of proteins. J Mol Biol 1994;238:415–436.
20. Warshel A. Calculation of enzymatic reactions: calculations of pKa, proton transfer reactions and general acid catalysis reactions in enzymes. Biochemistry 1981;20:3167–3177.
21. Sham YY, Chu ZT, Warshel A. Consistent calculations of pKa's of ionizable residues in proteins: semi-microscopic and microscopic approaches. J Phys Chem B 1997;101:4458–4472.
22. Berendsen H, van der Spoel D, van Drunen R. GROMACS: a message-passing parallel MD implementation. Comp Phys Comm 1995;91:43–56.
23. van der Spoel D, van Buuren AR, Apol E, Meulenhoffi PJ, Sijbers ALTM, Feenstra KA, Lindhal E, van Drunen R, Berendsen HJC. Gromacs user manual, version 3.0. Groningen, The Netherlands: Groningen University; 2001.
24. Berman H, Westbrook J, Feng Z, Gilliland G, Bhat T, Weissig H, Shindyalov I, Bourne P. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
25. Madura JD, Davis ME, Gilson MK, Wade R, Luty BA, McCammon JA. Biological applications of electrostatics calculations and Brownian dynamics simulations. Rev Comp Chem 1994;5:229–267.
26. Madura JD, Briggs JM, Wade R, Davis ME, Luty BA, Ilin A, Antosiewicz JA, Gilson MK, Bagheri B, Ridgway Scott L, McCammon JA. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comp Phys Commun 1995;91:57–95.
27. Sanner M, Spehner J-C, Olson A. Reduced surface: an efficient way to compute molecular surfaces. Biopolymers 1996;38:305–320.
28. Hess B, Bekker H, Berendsen H, Fraaije J. LINCS: a linear constraint solver for molecular simulations. J Comput Chem 1997;18:1463–1472.
28a. Fogolari F, Tosatto S. Application of MM/PBSA colony free energy to loop decoy discrimination: towards correlation between energy and root mean square deviation. Prot Sci in press.
29. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 1983;4: 187–217.
30. MacKerell ADJ, Bashford D, Bellott M, Dunbrack RLJ, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WEI, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
31. Qiu D, Shenkin P, Hollinger F, Still W. The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate Born radii. J Phys Chem 1997;101:3005–3014.
32. Sharp KA, Honig B. Calculating total electrostatic energies with the non-linear Poisson–Boltzmann equation. J Phys Chem 1990; 94:7684–7692.
33. Marcus RA. Calculation of thermodynamic properties of polyelectrolytes. J Chem Phys 1955;23:1057–1068.
34. Zhou HX. Macromolecular electrostatic energy within the nonlinear Poisson–Boltzmann equation. J Chem Phys 1994;100:3152–3162.
35. Fogolari F, Briggs JM. On the variational approach to the Poisson–Boltzmann free energies. Chem Phys Lett 1997;281:135–139.
36. Fogolari F, Zuccato P, Esposito G, Viglino P. Biomolecular electrostatics with the linearized Poisson–Boltzmann equation. Biophys J 1999;76:1–16.
37. Davis ME, McCammon JA. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: an approach to improve accuracy and convergence. J Comp Chem 1991;12:909–912.
38. Fogolari F, Brigo A, Molinari H. Protocol for MM/PBSA MD simulations of proteins. Biophys J 2003;85:159–166.
39. Uhrinova S, Smith MH, Jameson GB, Uhrin D, Sawyer L, Barlow PN. Structural changes accompanying pH-induced dissociation of the beta-lactoglobulin dimer. Biochemistry 2000;39:3565–3574.
40. Koradi R, Billeter M, Wuethrich K. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graphics 1996;14:51–55.
41. Fogolari F, Brigo A, Molinari H. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. J Molec Recogn 2002;15:377–392.