Water promotes the sealing of nanoscale packing defects in folding proteins

2014 J. Phys.: Condens. Matter 26 202101

(http://iopscience.iop.org/0953-8984/26/20/202101)

Fast Track Communications

IOP Publishing

Journal of Physics: Condensed Matter

J. Phys.: Condens. Matter **26** (2014) 202101 (6pp)

doi:10.1088/0953-8984/26/20/202101

Fast Track Communication

Water promotes the sealing of nanoscale packing defects in folding proteins

Ariel Fernández[1,2]

[1] Instituto Argentino de Matemática, National Research Council (CONICET), Saavedra 15, Buenos Aires 1083, Argentina

[2] Collegium Basilea, Institute for Advanced Study, Hochstrasse 51, CH 4053 Basel, Switzerland

E-mail: ariel@afinnovation.com

Received 1 February 2014, revised 26 February 2014
Accepted for publication 3 March 2014
Published 25 April 2014

Abstract

A net dipole moment is shown to arise from a non-Debye component of water polarization created by nanoscale packing defects on the protein surface. Accordingly, the protein electrostatic field exerts a torque on the induced dipole, locally impeding the nucleation of ice at the protein–water interface. We evaluate the solvent orientation steering (SOS) as the reversible work needed to align the induced dipoles with the Debye electrostatic field and computed the SOS for the variable interface of a folding protein. The minimization of the SOS is shown to drive protein folding as evidenced by the entrainment of the total free energy by the SOS energy along trajectories that approach a Debye limit state where no torque arises. This result suggests that the minimization of anomalous water polarization at the interface promotes the sealing of packing defects, thereby maintaining structural integrity and committing the protein chain to fold.

Keywords: biological interface, water polarization, nanoscale dielectrics, protein structure

(Some figures may appear in colour only in the online journal)

1. Introduction

Water sustains the structure of soluble natural proteins, whose native fold must endure the disruptive effects of backbone hydration [1]. On the other hand, the steering role of water in the protein folding process remains a subject of intense scrutiny and debate [2,3]. A standing challenge closely associated with this problem results from the need to understand the multi-scale dielectric behavior of water at biological interfaces with nanoscale detail [4–7]. It is commonly accepted that interfacial biological water fits the Debye picture [8], whereby the water polarization field $\mathbf{P} = \mathbf{P}(\mathbf{r})$ ($\mathbf{r}$ = position vector) aligns with the electrostatic field $\mathbf{E} = \mathbf{E}(\mathbf{r})$. Thus, $\mathbf{P}$ is identified with its projection $\mathbf{P}^{\|}$ along the field $\mathbf{E}$: $\mathbf{P} = \mathbf{P}^{\|} = (\mathbf{P} \cdot \boldsymbol{e})\boldsymbol{e}$ ($\boldsymbol{e} = \mathbf{E}/\|\mathbf{E}\|$). Under this premise, the protein charge distribution $\rho = \rho(\mathbf{r})$ is estimated as $\boldsymbol{\rho} = \nabla \cdot (\varepsilon_o \mathbf{F}^{\|})$, where ε_o is the vacuum permittivity and $\mathbf{F}^{\|} = \mathbf{E} + \varepsilon_o^{-1}\mathbf{P}^{\|}$ is the Debye field.

As recently shown [7], this result is not upheld for solutes that span interfaces with cavities (packing defects) of sub-nanometer local curvature and chemical heterogeneity, especially soluble proteins [9, 10]. Under the nanoscale confinement resulting from packing defects in proteins, water molecules are often statistically deprived of hydrogen bonding partnerships relative to bulk levels and, as a consequence, generate uncompensated partial charges that yield a net dipole moment. The hydrogen-bonding deprivation is a structural defect in the tetrahedral matrix of water–water interactions, precluding net charge cancellation. This defect yields a polarization component, $\mathbf{P}^{\#}$, orthogonal to $\mathbf{E}$ [7]. Thus, the protein charge distribution becomes rigorously $\boldsymbol{\rho} = \nabla \cdot (\varepsilon_o \mathbf{F})$, where $\mathbf{F} = \mathbf{E} + \varepsilon_o^{-1}(\mathbf{P}^{\|} + \mathbf{P}^{\#})$.

In this work we examine the field $\mathbf{P}^{\#}$ on the protein–water interface and assert that water molecules that generate this anomalous polarization are driven out of the nanoscale cavities by the protein electrostatic field. Furthermore, the protein–water interface of a folding protein evolves towards a minimization of the electrostatic energy stored in the anomalous polarization. This result implies that the packing defects

tend to get sealed as the protein folds and water gets expelled from the sites where it creates anomalous polarization. These findings make sense intuitively, given the need to maintain structural integrity while committing the protein chain to fold.

2. Methods

To compute $\mathbf{P}^{\#} = \mathbf{P}^{\#}(\mathbf{r})$, we introduced a scalar field $g = g(\mathbf{r})$, a time-average descriptor of local water structure, and established a relation between $\mathbf{P}^{\#}(\mathbf{r})$ and $g(\mathbf{r})$ [7]. Specifically, $g(\mathbf{r})$ gives the time-averaged number of hydrogen bonds of a water molecule while it visits a sphere of radius $r = 4\,\text{Å}$ centered at position $\mathbf{r}$ for a minimum timespan $\tau = 1\,\text{ns}$. The (r,τ)-parametrization of the scalar field g has been tuned to guarantee its second-order differentiability. The hydrogen bonds counted by g involve neighboring water molecules and polar groups from the protein. Thus, the polarization $\mathbf{P} = \mathbf{P}(\mathbf{r})$ departs from the relation $\mathbf{P} = \mathbf{P}^{\|}$ whenever $g(\mathbf{r})$ signals distortion (g<4) from the bulk coordination pattern ($g = 4$). Departures from bulk-like coordination create a significant non-Debye polarization component in accord with the proportionality $\mathbf{P}^{\#} \propto \nabla g$, or $\mathbf{P}^{\#} = \xi \nabla g$, where $\xi = (\lambda\varepsilon_o)^{1/2}$ and $\lambda = 9.0\,\text{mJ}\,\text{m}^{-1}$ at 298 K [7]. The net charge $\Gamma^{\#}$ induced by $\mathbf{P}^{\#}$ is then estimated as a function of the curvature $\nabla^2 g$ of the scalar field g as follows:

$$\Gamma^{\#} = -\nabla \cdot \mathbf{P}^{\#} = -\xi \nabla^2 g. \tag{1}$$

A convenient direct measure of water structural distortion is given by the scalar field $\phi = 4 - g$, with $\phi = 0$ indicating no distortion and $\phi = 4$ indicating maximum distortion of the tetrahedral water matrix. Thus, the net dipole moment $\boldsymbol{\mu}^{\#} = \boldsymbol{\mu}^{\#}(\mathbf{r})$ induced by the anomalous polarization $\mathbf{P}^{\#}$ becomes

$$\boldsymbol{\mu}^{\#}(\mathbf{r}) = \int (\mathbf{r}' - \mathbf{r}) \Gamma^{\#}(\mathbf{r}')\, d\mathbf{r}' = \xi \int [(\mathbf{r}' - \mathbf{r}) \nabla^2 \phi(\mathbf{r}')]\, d\mathbf{r}'. \tag{2}$$

The propagation of the distortion of water structure caused by a sub-nanometer cavity on the protein surface and the induced dipole moment arising from the resulting curvature change in ϕ are shown in figure 1(*a*). The illustrative case corresponds to a water-exposed backbone hydrogen bond, a packing defect known as *dehydron* [11]. Due to the confinement of water in the nanoscale cavity of the dehydron, a $\mathbf{P}^{\#}$-component is generated [11]. Dehydrons expose the backbone polar groups amide (>N–H) and carbonyl (>C = O) to hydrogen bonding with the surrounding water molecules, in competition with the intramolecular amide–carbonyl hydrogen bond, thus steering water dipoles away from alignment with $\mathbf{E}$. The confined water molecules relinquish hydrogen bonding opportunities to fit in the nanoscale cavity, departing from the bulk coordination pattern, with a resulting polarization that becomes statistically independent of $\mathbf{E}$. The propagation in the disruption of the water structure is computed using the molecular dynamics protocol described in [7] for a dehydron cavity of typical curvature radius $\theta = 2.63\,\text{Å}$ [12], with $x = 0$ indicating the center of curvature at the dehydron location.

Due to the confinement-induced polarization, an episteric ('around the solid') torque

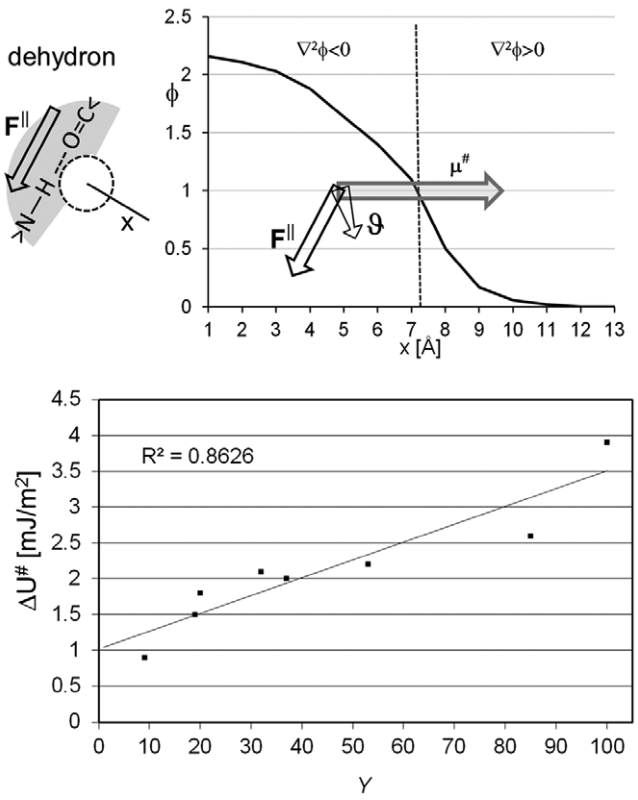

Figure 1. (*a*) Confinement-induced dipole $\boldsymbol{\mu}^{\#}$ and torque $\boldsymbol{\vartheta}$ imposed on it by the Debye field $\mathbf{F}^{\|}$ of the protein. The Debye field is locally generated by the dipole moment of a preformed dehydron (packing defect) pairing an amide (>N–H) and a carbonyl (>C = O) group of the protein backbone. The dipole moment induced by water confinement at the dehydron site is generated by the change in curvature of the water-structure descriptor ϕ. The figure shows the quantitative behavior of $\phi = \phi(x)$ relative to the distance x to the dehydron interface with fixed curvature radius $\theta = 2.63\,\text{Å}$ (dashed circle) with $x = 0$ representing the center of curvature. (*b*) Correlation between the SOS potential energy $\Delta U^{\#}$ (normalized by solvent-accessible surface area) and the dehydron-to-backbone-hydrogen-bond ratio Y for the protein given as percentage. The proteins studied identified by their respective PDB files and ordered by increasing Y-value are 1SRL, 1ESR, 1A8O, 1PIT, 1QGB, 1ATA, 1Q7I, 2PNE. The protein dehydrons are identified from structural coordinates as described in [13].

$$\boldsymbol{\vartheta} = \boldsymbol{\vartheta}(\mathbf{r}) = \boldsymbol{\mu}^{\#}(\mathbf{r}) \times \mathbf{F}^{\|}(\mathbf{r}) \tag{3}$$

is imposed by the Debye field on the interfacial solvent molecules (figure 1(*a*)).

The total reversible re-orientation work $W^{\#}$ needed to align the confinement-induced dipole $\mu^{\#}(\mathbf{r})$ subject to torque $\boldsymbol{\vartheta}(\mathbf{r})$ with the Debye field $\mathbf{F}^{\|}(\mathbf{r})$ is

$$\begin{aligned} W^{\#} &= -\int\int_{\theta(\mathbf{r})=0}^{\theta(\mathbf{r})=\theta^{\#}(\mathbf{r})} \|\mathbf{F}^{\|}(\mathbf{r})\| . \|\boldsymbol{\mu}^{\#}(\mathbf{r})\| \sin\theta(\mathbf{r})\, d\theta(\mathbf{r})\, d\mathbf{r} = \\ &= \int \|\mathbf{F}^{\|}(\mathbf{r})\| . \|\boldsymbol{\mu}^{\#}(\mathbf{r})\| . [1 - \cos\theta^{\#}(\mathbf{r})]\, d\mathbf{r} = \\ &= \int [\|\mathbf{F}^{\|}(\mathbf{r})\| . \|\boldsymbol{\mu}^{\#}(\mathbf{r})\| - \mathbf{F}^{\|}(\mathbf{r}) . \boldsymbol{\mu}^{\#}(\mathbf{r})]\, d\mathbf{r} \end{aligned} \tag{4}$$

where $\theta(\mathbf{r})$ is the angle between dipole and field at position $\mathbf{r}$, and

$$\cos\theta^{\#}(\mathbf{r}) = \mathbf{F}^{\|}(\mathbf{r}).\ \boldsymbol{\mu}^{\#}(\mathbf{r})/\{\|\mathbf{F}^{\|}(\mathbf{r})\|.\ \|\boldsymbol{\mu}^{\#}(\mathbf{r})\|\}. \quad (5)$$

The computation requires the evaluation of the confinement-induced dipole from the local structure of the solvent (equation (2)) after the latter is equilibrated with the protein structure. Thus, the operational equation for re-orientation work is

$$W^{\#} = \xi \int \|\mathbf{F}^{\|}(\mathbf{r})\| \cdot \| \int (\mathbf{r}' - \mathbf{r}) \nabla^2 \phi(\mathbf{r}')\, d\mathbf{r}' \| [1 - \cos\theta^{\#}(\mathbf{r})]\, d\mathbf{r}. \quad (6)$$

The potential energy $\Delta U^{\#} = W^{\#}$ stored in the non-Debye dipole orientation effectively represents a solvent orientation steering (SOS) contribution. Equation (4) quantifies the destabilizing protein–water coupling effect brought about by interfacial induced dipoles that do not align with the Debye field $\mathbf{F}^{\|}$. That is, $W^{\#}$ indicates the work needed to reach a Debye limit.

An alternative non-Debye interfacial energy term $\Delta U_{\phi} = \frac{1}{2}\varepsilon_o^{-1}\int\|\mathbf{P}^{\#}(\mathbf{r})\|^2 d\mathbf{r}$ has been introduced in previous work [7] and shown to be equivalent to the interfacial tension via the relation $\mathbf{P}^{\#} \propto -\nabla\phi$. The interfacial contribution ΔU_{ϕ} is solely stored in the orthogonal polarization, while $W^{\#}$ is dependent on both $\mathbf{P}^{\#}$ and the Debye field $\mathbf{F}^{\|}$. This distinction is important in local contexts where the interfacial energy term must distinguish orthogonal water polarization due to geometric confinement within an apolar protein region ($W^{\#} \approx 0$, $\Delta U_{\phi} \neq 0$) from orthogonal water polarization within a polar protein environment ($W^{\#} \neq 0$, $\Delta U_{\phi} \neq 0$). Thus, generally speaking, $\Delta U^{\#} = W^{\#}$ can be said to report more information on the interfacial electrostatics than ΔU_{ϕ} does.

3. Results

A correlation is observed for PDB-reported proteins (figure 1(*b*)) between the $\Delta U^{\#}$ (normalized by solvent-accessible surface area) of a protein and its dehydron ratio Y, defined as the fraction of backbone hydrogen bonds that are dehydrons (water-exposed). The correlation validates the assertion that dehydrons are the promoters of anomalous polarization. The dataset includes the protein with the maximum dehydron ratio $Y = 100\%$: the antifreeze protein from snow flea from PDB entry 2PNE [13]. The huge net torque exerted on the interfacial water molecules of the antifreeze protein is due to the maximum density of packing defects ($Y = 100\%$) on the protein surface and hampers the water orientational reorganization required for ice nucleation.

As we address the protein folding problem, we expect conformational change in the protein to seek minimization of the SOS potential energy as the interface minimizes structure-destabilizing protein–water coupling. Our results uphold this view, suggesting that the search for the Debye limit becomes pivotal to the protein folding process.

To test this hypothesis, we generated folding trajectories driven by a coarse-grained Monte Carlo (MC) scheme [14] incorporating the SOS potential energy in the computations of the canonical free energy change. To cover relevant timescales, the folding dynamics are entrained by coarser 'protodynamics', where the backbone (Φ, Ψ) dihedral torsions are specified 'modulo basins of attraction' in the potential energy surface. Coarse moves are defined by transitions between basins of attraction (R-basins) in the Ramachandran torsional (Φ, Ψ)-map for each residue along the chain. A Ramachandran map represents the potential energy of a residue as a function of its backbone dihedral torsions, and the R-basins are the allowed regions in (Φ, Ψ)-space. Each residue is assigned an R-basin after a coarse move, and the coarse state of the chain becomes a conformational ensemble, with each conformation generated by selecting individual (Φ, Ψ)-coordinates within the assigned R-basins. After each coarse move generated within a MC scheme, the system is allowed to equilibrate with the solvent for 1 ns subject to backbone torsional constraints imposed by remaining within the pre-assigned R-basins. The equilibration within a coarse state is performed as previously described (computational details in legend for figure 2) [7]. The remaining structural coordinates, including side-chain torsional variables and solvent coordinates, are allowed to vary freely during equilibration. The entrainment of the folding process by the stochastic protodynamics is justified within the adiabatic ansatz whereby intra-R-basin equilibration occurs faster than inter-basin transitions [14].

To reach folding timescales (>10 μs), the dynamics are steered by the coarse-grained stochastic process with the initial coarse state obtained by random assignment of R-basins to the individual residues along the chain. All thermodynamic quantities (ΔG, ΔH, ΔS) are computed relative to the initial random-coil ensemble. The coarse transition probability $p(t)$ is dependent on the overall free-energy difference between the two consecutive coarse states: $\Delta\Delta G(t) = \Delta\Delta H(t) - T\Delta\Delta S(t)$, where the enthalpy contribution $\Delta\Delta H(t)$ is 'in effect' determined in the NPT ensemble by the energy difference between the two equilibrated conformations belonging to the coarse states at time $(t + \tau)$ and t, respectively. The entropy difference $\Delta\Delta S(t)$ is obtained from the Boltzmann formula: $\Delta\Delta S(t) = k_B \ln[Z(t + \tau)/Z(t)]$, where k_B is the Boltzmann constant and $Z(t)$ is the number of chain/solvent configurations that realize the backbone torsional constraints determined by the coarse state at time t. The Boltzmann ratio $Z(t + \tau)/Z(t)$ between two consecutive coarse states of the protein chain is obtained from the protodynamics computation introduced in [14] as (see [7, 14]):

$$Z(t + \tau)\,/\,Z(t) = \Pi_j[\Omega_j(t + \tau)\,/\,\Omega_j(t)] \times \Pi_i[g_i(t + \tau)\,/\,g_i(t)]\,. \quad (7)$$

The dummy index j ($j = 1,\ldots,57$) indicates residue number along the chain and the index i ($i = 1,\ldots,9332$) labels individual water molecules, $\Omega_j(t)$ is the canonical lake area of the Ramachandran basin of the jth residue at time t [14] and $g_i(t)$ is the average number of hydrogen-bonding partnerships (defined above) for the ith water molecule at time t [7].

As dictated by the MC scheme: $p(t) = \exp[-\Delta\Delta G(t)/k_B T]$ if $\Delta\Delta G(t) > 0$ at $T = 298$ K and $p(t) = 1$ otherwise. If at time t, the coarse move is rejected by the MC procedure, the system

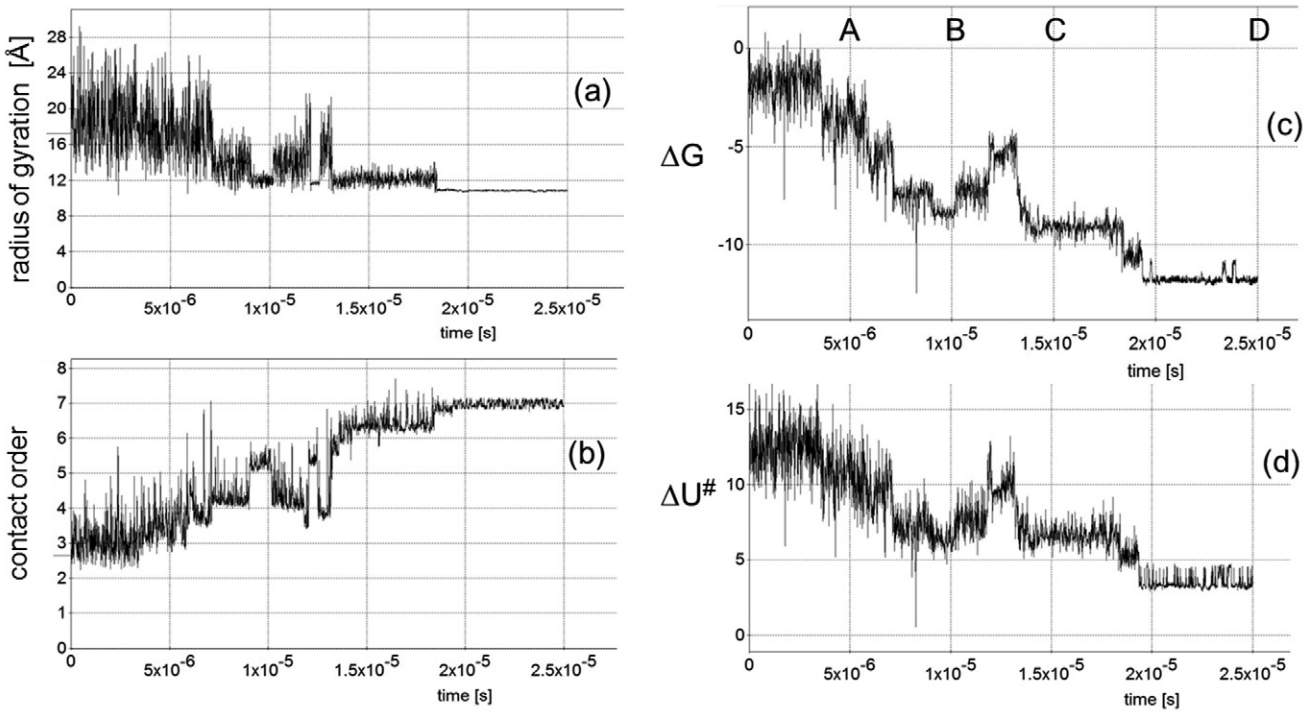

Figure 2. Folding trajectory determined by protodynamics consisting of 2.5×10^4 moves generated by a coarse-grained stochastic process. The coarse moves represent transitions and equilibrations of topological states of the chain for the thermophilic variant of the B1 domain of protein G, whose native fold is reported in PDB entry 1GB4. The folding trajectory was generated adopting the thermodynamic potential $\Delta G = \Delta G_{p,s} + \Delta U^{\#}$ as determinant of the coarse moves and equilibration, where $G_{p,s}$ = additive canonical free energy of uncoupled protein and solvent for a coarse state, and $U^{\#}$ = re-orientation work identified with the SOS potential energy for an equilibrated conformation within a coarse state. (*a*) Time-dependent behavior of the radius of gyration in Å; (*b*) contact order; (*c*) ΔG in kcal mol^{-1}; and (*d*) $\Delta U^{\#}$ in same units. For conformational search and solvent equilibration within each coarse state, a constant pressure of 1 atm and temperature $T = 298$ K are maintained using a Berendsen coupling scheme [17]. An AMBER package [18] was adopted for the MD simulation following each MC step, with charges on the molecules assigned according to the BCC charge model using AM1 optimized geometries and potentials [19]. After protein/solvent equilibration, the protein backbone coordinates are constrained according to the Shake scheme [20] and only side chains are allowed to explore conformation space, generating an ensemble of local hydration patterns used to compute the scalar field $\phi = \phi(\mathbf{r})$.

remains in the coarse state assigned at time t for another 1 ns, during which it undergoes a second round of protein/water equilibration.

We generated four folding trajectories for an autonomously folding protein within the *NPT* (isothermal/isobaric, $T = 298$ K) ensemble. The trajectories were generated with coarse states equilibrated and transitioned using the thermodynamic potential $\Delta G = \Delta G_{p,s} + \Delta U^{\#}$ that incorporates the SOS potential energy, with $G_{p,s}$ = additive canonical free energy of uncoupled protein and solvent for a coarse state specified by an R-basin assignment. To test our hypothesis, we selected the autonomous folding chain of length $N = 57$ for the thermophilic variant of the B1 domain of protein G from species *Streptococcus*. The crystal structure of the native fold is reported in PDB entry 1GB4 and was used only for comparison with the structures generated. The thermophilic variant was chosen because of its higher thermal stability, implying a more severe minimization of the SOS integral over the wild type, thus providing a better testing ground to our hypothesis.

The protodynamics for each trajectory consists of 2.5×10^4 coarse moves. A representative trajectory is reported in figures 2 and 3. All four trajectories converged reproducibly within 20 μs to a destiny steady state that lies within 1.6 Å RMSD of the 3D crystal structure of the folded protein, and belongs to a coarse state identical to that of the crystal structure. By contrast, similar computations were performed within the 25 μs timespan under the same *NPT* ensemble but excluding the SOS contribution ($\Delta G = \Delta G_{p,s}$), and in this case no convergence to a compact state was observed.

A representative folding trajectory incorporating the SOS potential energy contribution to determine coarse-grained transitions and structure equilibration is reported in figures 2 and 3. The convergence of the trajectory at about 19 μs to a compact steady state is evidenced by examining the time dependence of contact order and radius of gyration (figures 2(*a*) and (*b*)). The term $\Delta U^{\#}(t)$ was calculated as the difference in SOS energy between equilibrated conformations within coarse states at times t and 0. The trajectory converges reproducibly at about 19 μs to a steady state (structure D, figure 3) realizing the free energy minimum and endowed with identical topology (coarse state) as the crystal structure.

Four lines of evidence support the dominant role of SOS in the folding process: (1) the rapid convergence to a compact steady state (figures 2(*a*) and (*b*)) when contrasted with the lack of convergence and lack of reproducibility of the folding trajectory when the SOS term is not included; (2) the dynamic subordination of the overall term ΔG by the term

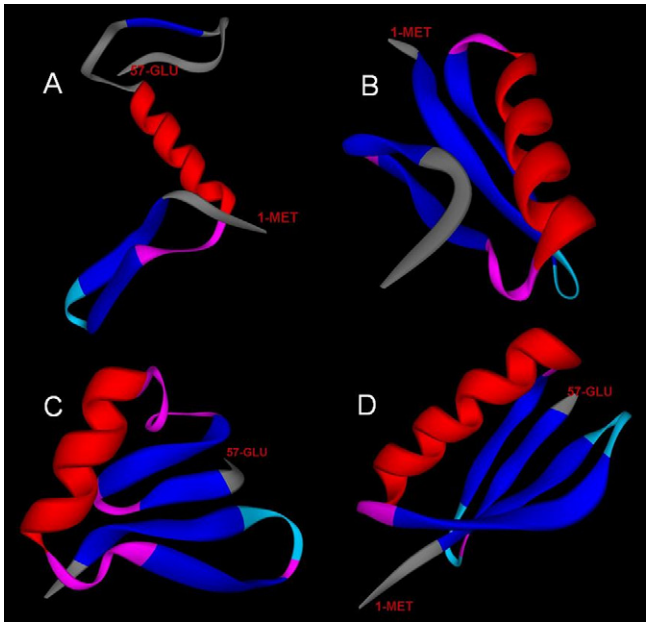

Figure 3. Ribbon representation of intermediate and destiny conformations at 5, 10, 15 and 25 μs along the SOS-folding trajectory, denoted A, B, C and D respectively.

$\Delta U^{\#}$, evidenced in the comparison of figures 2(*c*) and (*d*); (3) the fact that the compact destiny steady state is actually the free energy minimum over all states visited (figure 2(*c*)), in contrast with previous simulations [7] that did not include the defect-induced torque; and (4) the accurate reproduction of the native-state geometry (structure D, figure 3).

A conformational analysis of the SOS-driven trajectory reveals the sequence of folding events conducive to the native fold. (1) All basic secondary structure elements form at about 5 μs (structure A, figures 2(*c*) and (*d*) and 3) but the large fluctuations are indicative of the lack of tertiary structure required to cooperatively stabilize the structure. (2) Fluctuations are significantly suppressed once tertiary structure forms (structure B, figures 2(*c*) and (*d*)) and 3), albeit not entirely correctly, in about 10 μs. The initially formed tertiary structure is *provisional* since the native end-to-end parallel β-sheet (structure D, figure 3) is replaced by a nonnative antiparallel β-sheet (structure B, figure 3). (3) The most significant kinetic barrier occurs within the interval 12–13.5 μs and is due to the structure rearrangement of the antiparallel to a parallel end-to-end β-sheet (structure B rearranging into structure C, figure 3). (4) The correct tertiary structure becomes stabilized at about 15 μs (structure C, figures 2(*c*) and 3). (5) The correct tertiary structure is subject to further refinement for about 4 μs until a coarse steady state is reached at about 19 μs and equilibrated further for 6 μs (structure D, figures 2(*c*) and 3).

Although we do not adhere to preconceived *a priori* notions such as 'folding transition state', it is instructive to compare the folding kinetics that we report in this work with analysis of the measured impact of mutations on folding rates (Φ-analysis) as performed in [15] under the—debatable—assumption that there must be a transition state in the folding pathway. In comparing these approaches we alert the reader that the present study focuses on the thermophilic variant of protein G, whereas the experimental study was done on protein G itself (wild type). This caveat is important because, as the results in [15] demonstrate, the folding process may be fine-tuned by even a single mutation, especially if the latter has a bearing on the stability of the 'transition state' relative to the unfolded state. The results shown in figure 2(*c*) suggest that there is indeed a rate-determining step associated with rearrangement of the nonnative conformation of the second β-hairpin (conformation B, figure 3) into the in-register conformation (conformation C, figure 3) to eventually promote formation of the highly nonlocal parallel β-sheet (conformation D, figure 3). The materialization of the B→C conformational transition commits the protein to fold eventually into the native structure (figure 2). This is in accord with the Φ-analysis of the Baker group [15]: the mutation Asp→Ala (at residue 47) clearly disrupts the side chain–main chain $i,i+2$ interaction in the β-turn of the second β-hairpin and this mutation slows down the folding rate 20-fold. So, with the caveat indicated above, it appears that the formation of the in-register second hairpin conformation is rate-determining and that the transition state for this nonnative-to-native conformational transition is the overall folding transition state. Also, in accord with the Φ-analysis is the observation on the different propensities of the two antiparallel β-sheets (hairpins) along the folding process, with the second hairpin forming first '*albeit* out of register' to form the parallel β-sheet pairing the two extremities of the chain (see [16]), and the transition state signaled by the correct folding of the second hairpin. In our trajectories, the first hairpin forms after the second hairpin, but in the correct register from the start and packed against the α-helix, which may be an artifact resulting from an excess bias towards tertiary structure introduced by incorporation of the SOS. We further enrich the Φ-analysis by showing that the long-range parallel β-sheet is the true enabler of the correct conformation of the second antiparallel β-sheet and is essential to commit the protein to fold into the native structure.

These pathways generated using $W^{\#}$ are more expeditious that those from previous computations [7], where the free energy minimum differed from the native destiny steady state. This suggests that the inclusion of the SOS is the proper way to account for the solvent influence as the folding process seeks a Debye limit with removal of structural defects. However, with the caveat indicated previously, the premature stabilization of the first β-hairpin wrapping the α-helix is at a variance with the findings of the Φ-analysis [15], and implies that the incorporation of SOS may introduce a bias towards formation of tertiary structure.

All-atom MD simulations covering realistic folding timescales are now operational in supercomputers and have enabled the folding of protein G and other fast (μs to ms) folders into their native structures using standard force fields [16]. Such computations are not commensurate with those presented in this work. The latter rely on the theoretical ansatz of breakdown of the Debye approximation, and are steered by a coarse stochastic process only fleshed out at all-atom resolution during the equilibration of coarse states, and treat the solvent semi-empirically through a structural descriptor. However, both approaches reveal a prioritization of

local-propensity structural elements along the folding process, followed by local rearrangements promoted by long-range tertiary interactions. To that extent, we may assert that the interfacial contribution $W^{\#}$ must be subsumed into the full all-atom dynamics and dominate the interfacial energy. In all-atom computations the solvent is treated explicitly and there is no interfacial energy contributing to steer the dynamics. The latter is needed in a treatment using implicit solvent, that in our approach is described via the scalar field $\phi:\boldsymbol{R}^3\rightarrow\boldsymbol{R}$. While the two approaches are not comparable, their respective outputs (folding histories and predicted native topology) are, and in our study case, they produced similar answers with the caveat when comparing a wild type [16] with a thermophilic variant.

4. Conclusions

The steering role of water in the protein folding process remains a subject of intense scrutiny, particularly since the protein folding problem remains an open problem. In this work we showed that this role water can only be understood with the awareness that the Debye polarization scenario actually breaks down at the physically complex protein–water interface. More specifically, the local partial confinement of interfacial water at protein packing defects induces a dipole moment $\mu^{\#}$ that cannot be accounted for within the Debye dielectric picture and is subject to a torque imposed by the Debye electrostatic field. By hampering dipole re-orientation, this torque deters ice nucleation at the interface.

We evaluated the SOS as the reversible work needed to align $\mu^{\#}$ with the Debye field. This work constitutes a measure of the destabilizing protein–solvent coupling, with the Debye limit (no torque) representing the no-conflict state. We also demonstrated that as the protein–water interface changes alongside conformational transition, the minimization of the SOS energy drives the folding process promoting removal of packing defects in the protein. This water-induced seal commits the protein to fold and enables it to maintain its structural integrity. This result sheds light on the long-standing contentions regarding the steering role of water in the protein folding process.

Acknowledgment

The author is a senior investigator at the National Research Council of Argentina (CONICET) and was supported in part through Ariel Fernandez Consultancy, GmbH.

References

[1] Timasheff S N 1995 *Methods Mol. Biol.* **40** 253
[2] Levy Y and Onuchic J N 2006 *Annu. Rev. Biophys. Biomol. Struct.* **35** 389
[3] Thirumalai D, Liu L, O'Brien E P and Reddy G 2013 *Curr. Opin. Struct. Biol.* **23** 22
[4] Schutz C N and Warshel A 2001 *Proteins: Struct. Funct. Bioinform.* **44** 400
[5] Cheng Y K and Rossky P 1998 *Nature* **392** 696
[6] Giovambattista N, Lopez C F, Rossky P and Debenedetti P 2008 *Proc. Natl Acad. Sci. USA* **105** 2274
[7] Fernández A 2013 *J. Chem. Phys.* **139** 085101
[8] Debye P 1929 *Polar Molecules* (New York: Dover)
[9] Giovambattista N, Debenedetti P G and Rossky P J 2007 *J. Phys. Chem.* C **111** 1323
[10] Giovambattista N, Debenedetti P G and Rossky P J 2012 *Annu. Rev. Phys. Chem.* **6** 179
[11] Fernández A 2012 *J. Chem. Phys.* **137** 231101
[12] Schulz E, Frechero M, Appignanesi G and Fernández A 2010 *PLoS One* **5** 12844
[13] Fernández A and Berry R S 2010 *J. Proteome Res.* **9** 2643
[14] Fernández A 2001 *J. Chem. Phys.* **114** 2489
[15] McCallister E L, Alm E and Baker D 2000 *Nature Struct. Biol.* **7** 669
[16] Lindorff-Larsen K, Piana S, Dror R D and Shaw D E 2011 *Science* **334** 517
[17] Berendsen H J, Postma J P, van Gunsteren W F, Di Nola A and Haak J 1984 *J. Chem. Phys.* **81** 3684
[18] Wang J, Wolf R M, Caldwell J W and Kollman P A 2004 *J. Comput. Chem.* **25** 1157
[19] Jakalian A, Jack D B and Bayly C I 2002 *J. Comput. Chem.* **23** 1623
[20] Ryckaert J P, Ciccotti G and Berendsen H J 1977 *J. Comput. Phys.* **23** 327