AIP | The Journal of Chemical Physics

Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties

Aoife C. Fogarty, Raffaello Potestio, and Kurt Kremer

Citation: The Journal of Chemical Physics **142**, 195101 (2015); doi: 10.1063/1.4921347
View online: http://dx.doi.org/10.1063/1.4921347
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/142/19?ver=pdfcov
Published by the AIP Publishing

Adaptive resolution simulation of a biomolecule and its hydration shell: Structural and dynamical properties

Aoife C. Fogarty,[a] Raffaello Potestio,[b] and Kurt Kremer[c]
Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

(Received 26 March 2015; accepted 3 May 2015; published online 21 May 2015)

A fully atomistic modelling of many biophysical and biochemical processes at biologically relevant length- and time scales is beyond our reach with current computational resources, and one approach to overcome this difficulty is the use of multiscale simulation techniques. In such simulations, when system properties necessitate a boundary between resolutions that falls within the solvent region, one can use an approach such as the Adaptive Resolution Scheme (AdResS), in which solvent particles change their resolution on the fly during the simulation. Here, we apply the existing AdResS methodology to biomolecular systems, simulating a fully atomistic protein with an atomistic hydration shell, solvated in a coarse-grained particle reservoir and heat bath. Using as a test case an aqueous solution of the regulatory protein ubiquitin, we first confirm the validity of the AdResS approach for such systems, via an examination of protein and solvent structural and dynamical properties. We then demonstrate how, in addition to providing a computational speedup, such a multiscale AdResS approach can yield otherwise inaccessible physical insights into biomolecular function. We use our methodology to show that protein structure and dynamics can still be correctly modelled using only a few shells of atomistic water molecules. We also discuss aspects of the AdResS methodology peculiar to biomolecular simulations. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4921347]

I. INTRODUCTION

From a simulation point of view, biomolecular systems are among the most challenging, because of their innate heterogeneity and the broad range of length and time scales they encompass. Typical length scales of relevance range from μm cells to chemical reactions on the scale of angstroms, while biophysical and biochemical processes cover a time scale spectrum from fs bond vibrations to protein folding or macromolecular diffusion which may take place on the order of seconds or longer.[1–3]

Large systems inevitably pose sampling problems, and values extracted from molecular simulation are only as good as the employed forcefield and the quality of the statistical sampling, a fact which becomes increasingly important the longer the relevant time scales. Of help here can be a coarse-graining approach,[4] in which certain degrees of freedom are eliminated in the model describing the system. However, sufficient detail must be retained in order to correctly describe physical or chemical processes of interest, and in heterogeneous systems, this can lead to a multiresolution scheme in which different parts of the system are simultaneously modelled at different levels of resolution.[5,6]

In fully classical multiscale schemes, to which we restrict ourselves here, the boundaries between resolutions generally correspond to the boundaries between system components, i.e., the biomolecule of interest is modelled at one resolution, while the components of its environment, such as solvent, lipid membranes, and other biomolecules, are each modelled at one or more different resolutions. The resolution of a given molecule then generally remains fixed for the entire duration of the simulation. This approach commonly involves an atomistic (AA) protein in an environment which is coarse-grained (CG),[7–9] implicit,[10,11] or a combination of both.[12] Similarly, one can describe as multiscale an approach in which the biomolecule itself is coarse-grained, while the environment is at an even lower resolution, i.e., implicit solvent[13–15] or continuum.[16] In a few pioneering works, the fixed boundary between resolutions may even occur at a fixed place within a single biomolecule.[17,18] Such fixed-resolution multiscale schemes have successfully been used to reach longer length- and time scales than previously possible, providing insights into processes and phenomena including cellular crowding,[12] protein folding,[14,19–21] ligand binding,[22] and structural dynamics of DNA-protein complexes.[23]

However, the degrees of freedom to be eliminated in such multiscale schemes must be carefully chosen based on the particular system under consideration. There are two main points to bear in mind here. First, particular processes of interest may depend on properties only present in atomistically resolved models. In the case of a protein in aqueous solvent, in particular, one must bear in mind the influence the protein hydration shell is known to have on protein conformation, function, and dynamics.[24–26] Many biomolecular processes are known to depend on properties like the hydrogen-bonding ability of water, which rely on specific chemical detail.[24] In this context, we note that partial unfolding and increased H-bonding between side chains have been observed in simulations of an atomistic protein in coarse-grained water, with the addition of a layer of atomistic hydration water required

[a]Electronic mail: fogarty@mpip-mainz.mpg.de
[b]Electronic mail: potestio@mpip-mainz.mpg.de
[c]Electronic mail: kremer@mpip-mainz.mpg.de

to suppress the occurrence of this problematic behaviour.[27] Second, while coarse-grained water[28,29] and protein[13] models are becoming ever more sophisticated, they suffer from a number of issues:[30–34] a lack of transferability to different contexts or state points, the inability to simultaneously reproduce all structural and thermodynamic properties, and a dynamics which is accelerated compared to the corresponding atomistic model. These problems are reduced or absent when using atomistic models. Both of these two points will determine the resolution at which each system component must be modelled, depending on the properties of interest. Specifically, in the case of a biomolecule in aqueous solution, it becomes desirable to be able to simulate the biomolecule and hydration shell atomistically, yet with the remainder of the solvent modelled at a coarse-grained level for computational efficiency. Simulating a system containing both atomistic and coarse-grained water, with free exchange between different levels of resolution and no restraints on water molecule positions, requires an adaptive resolution approach.[35] Such an approach involves particles whose resolution is not fixed but rather depends on each particle's instantaneous location in predefined atomistic or coarse-grained regions. A hybrid (HY) or transition region between the atomistic and coarse-grained regions allows for smooth coupling between resolutions and free diffusion across their interface. The Adaptive Resolution Scheme (AdResS) methodology has already successfully been used to simulate bulk water[36,37] and a protein in a bundled water model.[38] In the present work, we take the existing AdResS methodology, examine in detail its adaptation for and validity in biomolecular systems, and demonstrate one first possible application of the methodology.

The first study combining the adaptive resolution approach and biomolecular simulation[38] showed that an atomistic protein can successfully be simulated in adaptive resolution water, with the protein remaining stably folded in ns-length simulations. However, the bundled water model used in that study did not allow for an exploration of how protein and water dynamics compared to that observed in experiment and in standard atomistic simulations. Here, we simulate a fully atomistic, 9-kDa protein, solvated in adaptive resolution water such that the hydration shell of the protein is fully atomistic. We first validate our approach, via a detailed comparison of protein and water structure and dynamics to a fully atomistic reference simulation, wherever possible referring also to experimental results. It is important to mention at this point that while the computational speedup obtainable through the AdResS method is advantageous in and of itself, an additional benefit of the method is the opportunity it provides to obtain new physical insight. In the final part of this paper, therefore, we present one example of this, demonstrating how our methodology can be used to provide otherwise inaccessible insight into the level of hydration necessary for protein function.

The structure of the paper is as follows. In Sec. II, we outline the AdResS methodology and in particular its application to biomolecular systems, and in Sec. III, we provide specific details of the simulations used in the present work. In Secs. IV and V, we present a comparison of protein and water structure and dynamics to fully atomistic reference simulations. In Sec. VI, we use the method to study protein hydration level, before concluding the paper.

II. ADAPTIVE RESOLUTION METHODOLOGY

In this section, we present and discuss general aspects of the AdResS methodology as applied to biomolecules, before giving specific simulation details in Sec. III. The AdResS methodology couples two distinct descriptions of a system and in this case models at different resolutions: AA and CG. The system is divided into predefined regions: an atomistic region, a coarse-grained region, and between them a hybrid or transition region in which occurs the coupling between different levels of resolution. To model a fully atomistic protein in AdResS water, the atomistic region is defined as a sphere of radius d_{at} centred on a point $\mathbf{r}_{centre}$ within the protein, with d_{at} greater than the radius of gyration of the protein. During a simulation, this atomistic region therefore moves with the freely diffusing protein. The atomistic region is surrounded by the hybrid region which is a spherical shell of width d_{hy} and immersed in the coarse-grained region which serves as a particle reservoir for the atomistic region and which occupies the remainder of the simulation box. This is illustrated in Figure 1 (left). A solvent particle's instantaneous identity depends on its spatial location within these regions, and particles can move freely between regions.

In the atomistic region, the system is described in full atomistic detail; in the case of a biomolecular simulation, this would typically be pre-existing standard atomistic protein or DNA forcefields and water models. The coarse-grained reservoir of water molecules can be described with any suitable CG potential, or even by hard spheres[39] or an ideal gas[40] at the appropriate density or pressure. The two levels of resolution are coupled via a force-interpolation scheme, in which the intermolecular force between the centres of mass of molecules α and β is given by

$$\mathbf{F}_{\alpha\beta} = \lambda(\mathbf{r}_\alpha)\lambda(\mathbf{r}_\beta)\mathbf{F}^{AA}_{\alpha\beta} + [1 - \lambda(\mathbf{r}_\alpha)\lambda(\mathbf{r}_\beta)]\mathbf{F}^{CG}_{\alpha\beta}, \tag{1}$$

where

$$\mathbf{F}^{AA}_{\alpha\beta} = \sum_{i\in\alpha}\sum_{j\in\beta}\mathbf{F}^{AA}_{ij}, \tag{2}$$

where $\mathbf{F}^{AA}_{ij}$ is the interaction between atoms i and j using the atomistic force-field, $\mathbf{F}^{CG}_{\alpha\beta}$ is the interaction between molecules α and β using the coarse-grained force-field, and λ is a transition function varying smoothly and monotonically between 1 in the AA region and 0 in the CG region, and in the hybrid region taking the form

$$\lambda(\mathbf{r}_\alpha) = \cos^2\left(\frac{\pi}{2d_{hy}}(|\mathbf{r}_{centre} - \mathbf{r}_\alpha| - d_{at})\right). \tag{3}$$

The interaction between two atomistic or two coarse-grained molecules therefore simplifies to $\mathbf{F}^{AA}_{\alpha\beta}$ or $\mathbf{F}^{CG}_{\alpha\beta}$, respectively. Note that since, in general, the interaction between the atomistic protein and the coarse-grained water will not be parametrised, the values of d_{at}, d_{hy}, and the non-bonded interaction cutoff should be chosen such that there is no interaction between these system components.

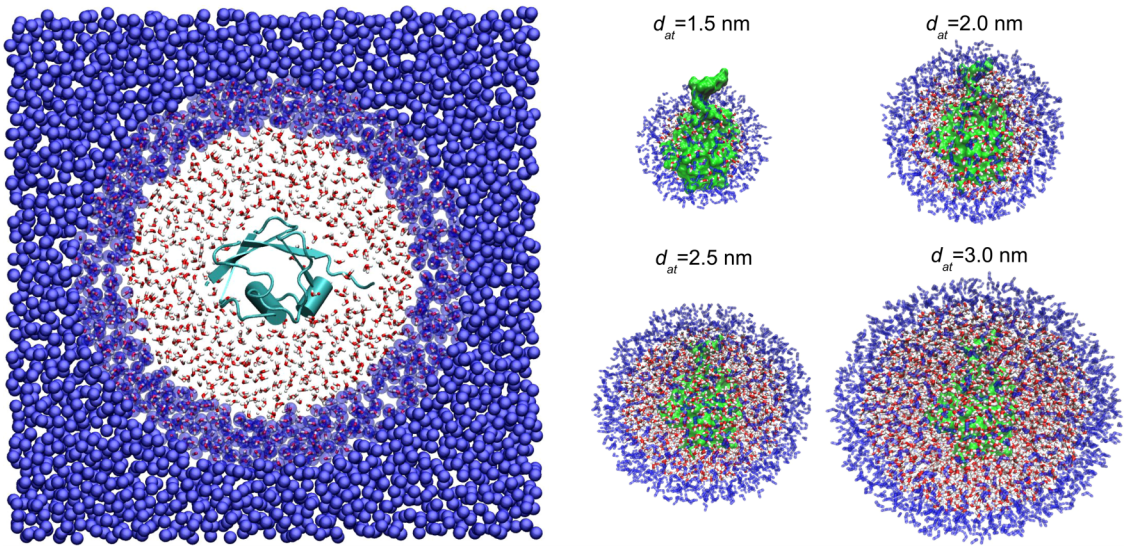

FIG. 1. On the left, we illustrate the AdResS system with the largest atomistic region ($d_{at} = 3.0$ nm), showing the protein in a slab cut from the box of solvent, divided into spherical atomistic (red and white), hybrid (red, white, and blue), and coarse-grained (blue) regions. On the right, we show the protein (green), the atomistic water molecules (red and white), and the hybrid water molecules with $\lambda > 0.5$ (i.e., with more than 50% atomistic character, transparent blue), corresponding to the values given in Table I, for all four AdResS systems.

The thermodynamic inconsistency issues inherent in structure-based coarse-graining have been extensively discussed.[32–34] In particular, the CG potential which correctly reproduces the AA pair distribution function will in general have a different pressure than the reference AA potential. Using such a pair of CG and AA potentials together in the AdResS setup would then lead to a density difference between AA and CG regions. To compensate for this, one can use an external force known as the thermodynamic force, which is obtained iteratively via an expression based on the gradient of the density difference between atomistic and coarse-grained regions.[41] This force acts on the centre of mass of molecules in the hybrid region and ensures thermodynamic equilibrium between AA and CG regions. An alternative option would be to use a CG potential at the same pressure as the AA potential, for example, a potential from the iterative Boltzmann inversion procedure with pressure correction[42] or a hard-sphere or ideal gas potential at the appropriate pressure, in which case it is necessary to check that the difference in compressibilities between the AA and CG models does not lead to a barrier to molecule exchange between regions.[41]

The coupling of models at different levels of resolution via Eq. (1) leads to heat production in the hybrid region, due to the non-conservative nature of the interactions there.[39,43] This excess heat must be removed using a local thermostat; however, coupling to a thermostat leads to a perturbation of system dynamics, and in general terms, in application to biomolecular systems, would perturb processes involving significant thermal changes such as ligand binding and enzymatic reactions. We therefore thermalise only the particles in the hybrid and CG regions, so that these regions function as a thermal bath in which the biologically relevant region is immersed. This setup ensures the desired temperature throughout the atomistic region, while leaving its dynamics unperturbed by direct coupling to a thermostat. Another option would be to employ an energy interpolation scheme[44] instead of the force-based interpolation (Eq. (1)) used here. Such an energy-based scheme allows one to perform energy-conserving microcanonical (i.e., non-thermostatted) simulations; however, such schemes also have a higher computational cost, and the obtention of the necessary thermodynamic compensation functions may be less straightforward.[40]

In this section, we have introduced several different factors which could *a priori* perturb system properties relative to those of a fully atomistic system. Both static and dynamic properties may be disrupted by the use of interpolated forces and coarse-grained forces in the HY and CG regions, respectively, and by the thermodynamic force, applied only in the HY region. The thermostat applied in the HY and CG regions may perturb dynamics properties, as is of course the case for any thermostatted simulation, whether multiscale or fully atomistic. In this work, our aim is the quantitatively correct modelling of physical behaviour in the AA region, while the HY and CG regions serve as a particle reservoir and heat bath. None of the above-mentioned perturbative factors acts directly in the AA region. Their indirect effect is felt in the AA region only in the immediate vicinity of the AA/HY interface, an effect which quickly becomes negligible with increasing size of the AA region, as we explore and quantify in Sec. VI. Beyond this, particles in the AA region remain unperturbed by the above-mentioned factors, as has previously been shown for a range of structural and diffusion properties,[35,37,45] and as we will show in this work for additional static and dynamic properties.

III. COMPUTATIONAL DETAILS

The biomolecular system chosen as a test case was an aqueous solution of ubiquitin, a 9 kDa (76-residue) globular regulatory protein.[46] The starting configuration was obtained from the crystallographic structure with PDB code 1UBQ[47] and solvated in 38 084 water molecules, corresponding to a concentration of ≈1 mM. Note that ubiquitin has no overall net charge, and no counterions are present in the system.

The simulation box was a cube of length ≈10.5 nm, corresponding to a density determined in the NPT ensemble. The protein was described using the AMBER 03 force field[48] and was fully atomistic at all times, with the atomistic region centered on the carbon atom C_{D1} in the side chain of Leu-43, the atom closest to the protein centre of mass. The atomistic region had a radius (d_{at}) of 3 nm in the simulations used for a detailed comparison with fully atomistic simulations. We note here that the radius of gyration of ubiquitin is ≈1.2 nm. We also performed a study of the system behaviour as a function of d_{at} in the range 1.5 nm–3.0 nm. The thickness of the hybrid region (d_{hy}) was 1 nm in all cases. Atomistic water was described using the SPC/E water model.[49] For coarse-grained water, a mapping of one water molecule to one CG bead was used. A CG model in which several water molecules are grouped together in one CG bead would lead to even greater computational efficiency; however, the low lifetime of water clusters means that they would break apart within the hybrid region. With such a CG model, water molecules would need to be reassigned to CG beads on the fly, leading to increased computational complexity and cost, or held together in clusters even within the atomistic region, leading to unphysical behaviour. The interaction potential between our one-molecule CG beads was determined using the iterative Boltzmann inversion method,[42] such that the water centre-of-mass radial distribution functions (rdf's) match in a fully atomistic and fully coarse-grained system, at the desired state point. This coarse-graining was carried out using the VOTCA package[50] and no pressure correction. The thermodynamic force was also determined using the VOTCA package.

Molecular dynamics were carried out using the ESPResSo++ simulation package.[51] A time step of 0.5 fs was used in coherence with the time scale of the fastest process in the AA region, namely, the vibration of hydrogen-containing bonds in the protein. Further computational efficiency could be reached by implementing a multiple time step scheme with a much larger time step in the CG region. Fully atomistic reference simulations were performed in the NVE ensemble. In AdResS, particles in the HY and CG regions were coupled to a Langevin thermostat using $\gamma = 30\ \mathrm{ps}^{-1}$. In the fully atomistic system and in the atomistic region of the AdResS systems, the temperature was 300 K ± 1 K.[52]

Electrostatics were treated using the reaction field method with a dielectric constant of 80 for water-water and water-protein interactions and 4 for intra-protein interactions.[53] The cutoff was 1 nm for all non-bonded interactions, and simulations were performed with periodic boundary conditions and the minimum image convention. The SETTLE algorithm for rigid water was used.[54] Equilibration runs were 1–2 ns long, while production runs had a total length of 10 ns for each system, or 50 ns in all.

Simulation trajectories were analysed using in-house code. For the purposes of comparison between AdResS and fully atomistic systems, analysis was performed on molecules in the atomistic part of the AdResS system and on the corresponding sphere in the all-atom system, i.e., the sphere with the same centre and radius. In many cases, the calculations were limited to subpopulations of water, namely, the first hydration shell and the bulk atomistic water. In those calculations limited to water molecules in the first hydration shell at the time origin, the shell was defined as consisting of all water molecules which were H-bonded to the protein surface or whose oxygen atom was within a certain cutoff distance from the nearest protein heavy atom, the cutoff being taken as the first minimum in the protein heavy atom-water oxygen rdf. H-bond criteria were individually determined for each H-bond donor or acceptor in each protein residue using water-amino acid rdf's. For calculations in the bulk atomistic water part of the system, this was defined as all water molecules more than 0.7 nm from the protein surface and less than $d_{at} - 0.5$ nm from the centre of the atomistic region, at the time origin. Error bars were calculated using block-averaging and reported at a 95% confidence level using Student's t distribution.

It is useful to find a correspondence between the size of the atomistic region and the hydration level or number of hydration shells around the protein, a more intuitive and more easily interpretable value. Since the atomistic region is a smooth sphere, while the protein has an ovaloid shape (aspect ratio approximately 1.5) with a rugged surface, the number of atomistic hydration shells for each surface site will vary, for any given value of d_{at}. We estimated the number of shells as follows: for each H-bond acceptor on the protein surface, we found the distance to the furthest fully atomistic water molecule lying approximately on the vector defined by the atomistic region's centre and the surface site, and converted this to an approximate number of shells using the molecular diameter of a water molecule.

IV. PROTEIN STRUCTURE AND BACKBONE DYNAMICS

The native structure and dynamics of a protein play a crucial role in its function, and their accurate reproduction is an essential test of any biomolecular simulation methodology. It has already been demonstrated that the AdResS approach can simulate a stably folded protein.[38] In that study, the radius of gyration and the root mean square deviation (RMSD) were used to show that the atomistic protein retained its folded state and global structure over ns-length simulations in an adaptive resolution solvent. Here, we concentrate on the more local dynamical properties of the protein. The root mean square fluctuations (RMSFs) along the protein backbone are primarily a probe of local backbone translational dynamics and can be related to experimental temperature or B-factors from protein X-ray crystallography. The RMSF for the backbone C_α atoms is calculated as the square-root of the mass-weighted variance in position for each atom, after removal of translational and rotational displacement of the whole protein. Meanwhile, to probe backbone rotational dynamics, we followed the reorientational relaxation of the NH bonds in backbone amide groups, since this is also experimentally accessible via NMR ^{15}N spin relaxation. We calculated the second-order orientational time-correlation function (tcf) of the backbone NH bonds, defined as

$$C_2(t) = \langle P_2[\mathbf{u}(0) \cdot \mathbf{u}(t)] \rangle , \tag{4}$$

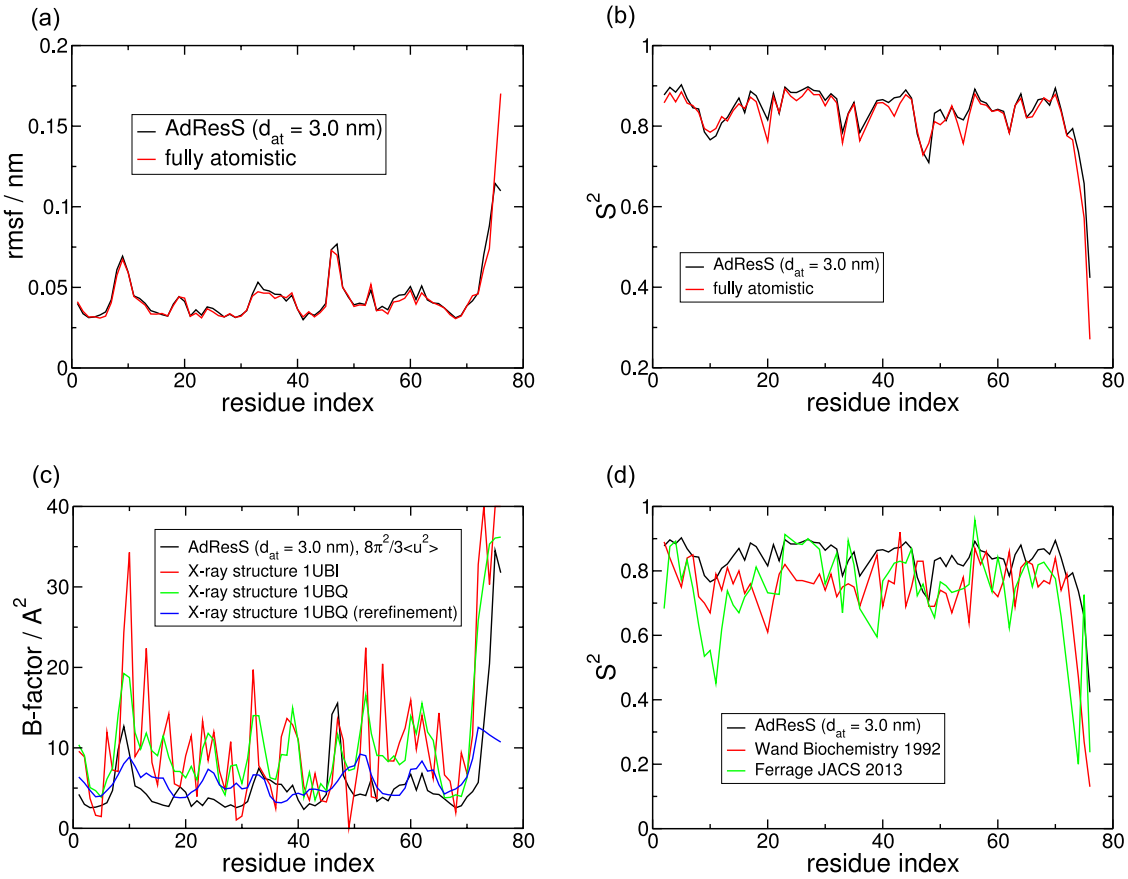

FIG. 2. Comparison between the AdResS ($d_{at} = 3.0$ nm) and fully atomistic systems: (a) Root mean square fluctuations of the protein α-carbons. (b) NMR order parameter S^2 of the protein backbone amide NH bonds. Comparison between AdResS system and experimental results: (c) X-ray crystallographic B-factors[47,64,109] and (d) experimentally determined S^2 order parameters.[65,110]

where **u** is the vector along the NH bond, and P_2 is the second Legendre polynomial. The NMR order parameter S^2 is the infinite-time value of this orientational tcf. Experimentally, S^2 can be obtained in a model-free way from the NMR relaxation parameters.[55] We extracted S^2 for each backbone amide bond from our simulations. In Figure 2, we show the RMSF and S^2 values for each residue. The agreement between the fully atomistic and AdResS profiles shows that the protein backbone translational and rotational dynamics are not perturbed by the solvation of its atomistic hydration shell in a coarse-grained water bath. In Figure 2 we also show a comparison of the above-mentioned quantities to the corresponding experimental results, namely, crystallographic B-factors and S^2 parameters from NMR relaxation experiments. While there is a rough qualitative agreement in each case between simulation and experimental results, the agreement is far from being quantitative, nor do different experimental data sets agree with each other. We include these figures to remind the reader that reproducing fully atomistic simulations is not the be-all and end-all of multiscale simulation. Atomistic protein forcefields, as has been pointed out elsewhere,[56–58] have known weaknesses and may not reproduce experimental reality. Furthermore, experimental data sets themselves may have their own uncertainties. Crystallographic B-factors, for instance, contain errors introduced by static disorder and insufficient data,[59] are collected in the crystalline state and mainly at low temperature, and can only adequately capture small-scale, harmonic, monomodal motions in the protein.[59–61] In addition, rerefinement of the same experimental structure factor data[62,63] can lead to large variations in the resulting B-factors. In Figure 2(c), for example, we compare the B-factors from the original refinement of the structure factor data with PDB code 1UBQ to those from a re-refinement using a different refinement procedure within the PDB_REDO project.[64] Meanwhile, NMR order parameters show a strong field dependence.[65] Our aim here is to show that our AdResS methodology reproduces physical reality as correctly as possible using existing force fields and to the limit at which physical reality can be known using current experimental techniques.

V. WATER STRUCTURE AND DYNAMICS

We now turn to the properties of the aqueous solvent itself. The presence of water is crucial for biomolecular function in most cases,[24] and the hydration shell is known to influence biomolecular structural and dynamic properties.[66] Internal water molecules in globular proteins are conserved across crystallographic structures and form an integral part of the protein secondary structure on a nanosecond time scale.[67] Meanwhile, hydration shell dynamics is known to be an important factor in processes including protein or drug binding to DNA,[24,68] protein folding,[24] heat protection of DNA,[69] and enzyme catalysis.[70] However, the molecular

role of water in biomolecular function has not yet been fully elucidated.

Moreover, the properties of water in the hydration shell of a biomolecule are known to differ from those in bulk water, with the density,[71,72] structure,[73–75] and dynamics[76–80] showing nonbulk-like behaviour. Indeed, the hydration shell can be defined as consisting of those water molecules whose behaviour is distinct from bulk water. The exact spatial extent of the perturbation is controversial[81] and may depend on which properties are being probed. Conversely, the minimum hydration level needed for correct protein function is also open to discussion[82,83] and must again depend on which protein properties are considered. In Sec. VI, we explore this latter question in greater detail. For the moment and for our current purpose, namely, the validation of the adaptive resolution approach, we concentrate our analysis on the water molecules in the first hydration shell, as defined in Sec. III.

A. Structure

We begin by examining the global water structure. Water density as a function of distance from the centre of the atomistic region is presented in Figure 3(a), where the position of the protein is visible as the absence of water density at low distances. The density of the adaptive resolution solvent across the atomistic and coarse-grained regions is identical to that of the fully atomistic system, showing the efficacity of the thermodynamic force in compensating for the pressure imbalance between atomistic and coarse-grained models. This figure also gives a convenient visual indication of the size of the atomistic region relative to the protein size.

Figure 4 shows the water oxygen-oxygen and oxygen-hydrogen radial distribution functions, calculated between atoms in the atomistic region in the AdResS system and in the corresponding sphere cut out of the fully atomistic system. Since our goal is not to quantitatively interpret the rdf's but to compare all-atom and AdResS simulations, they are not corrected for the excluded volume of the protein nor the volume excluded due to the calculation being performed only for particles in the atomistic sphere. The perfect match between rdf's in the fully atomistic and AdResS systems confirms that the water structure is correctly reproduced throughout the atomistic region.

In order to specifically probe local water structure in the bulk and in the hydration shell, we use the tetrahedral order parameter, which quantifies how well the four nearest neighbours of a water molecule form a tetrahedron. It has a value of 1 in a perfectly tetrahedral arrangement and an average value of 0 in an ideal gas and is defined as[84]

$$q_{tet} = 1 - \frac{8}{3}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(\cos(\theta_{j,k}) + \frac{1}{3}\right)^2, \tag{5}$$

where $\theta_{j,k}$ is the angle jik formed by a water oxygen i and its nearest neighbours j and k. In the hydration shell, q_{tet} can be calculated considering only other water oxygen atoms as potential nearest neighbours ($q_{tet,w}$) or also considering protein heavy atoms ($q_{tet,p}$). The presence of a biomolecular solute is known to disrupt the tetrahedral structure of water, leading to values of q_{tet} which depend on the distance from the biomolecular surface.[73,75] For our AdResS system, Figure 3(b) shows the distribution of $q_{tet,w}$ values in the bulk water part of the atomistic region and the distribution of $q_{tet,p}$ and $q_{tet,w}$ values in the first hydration shell, as compared to the same distributions in the fully atomistic system. Both distributions are perfectly reproduced, showing that the tetrahedral packing of water in both the bulk and the hydration shell is fully preserved in the AdResS system.

B. Dynamics

We now turn to the dynamics of the water molecules in the atomistic region, in both the protein hydration shell and the bulk water part of the atomistic region. Since the rotational and translational motions of liquid water are highly coupled,[85] we concentrate here on rotational or reorientational dynamics, which can be probed using the second-order orientational time-correlation function, defined as in Eq. (4), where **u** is in this case the vector along the water OH bond. After an initial femtosecond decline due to librational motion, bulk water shows a fast (ps-time scale) monoexponential reorientational decay, while water in biomolecular hydration shells is known

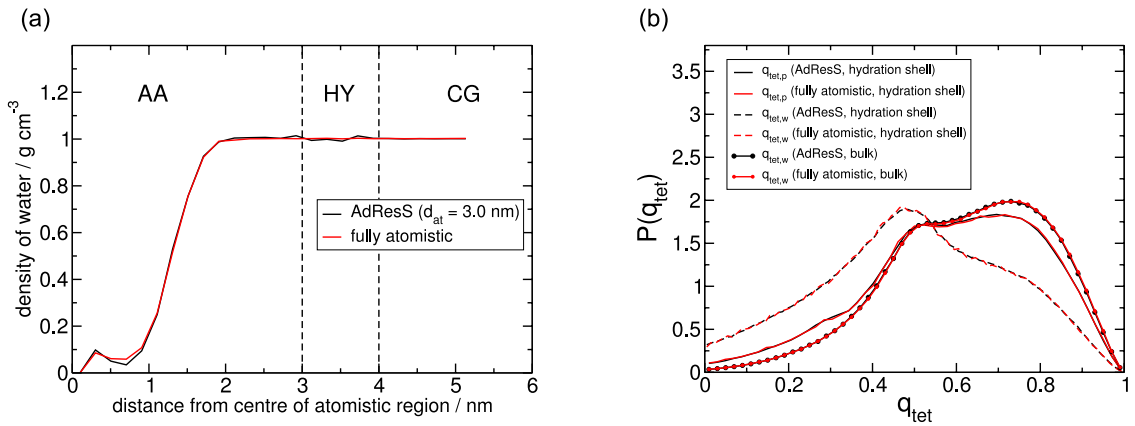

FIG. 3. In the AdResS ($d_{at} = 3.0$ nm) and fully atomistic systems: (a) Density of water as a function of distance from the centre of the atomistic region. The low density values on the far left of the graph correspond to the position of the protein. (b) Probability distribution of the tetrahedral order parameters $q_{tet,p}$ and $q_{tet,w}$ for water molecules in the first hydration shell and $q_{tet,w}$ for water molecules in the bulk atomistic region.

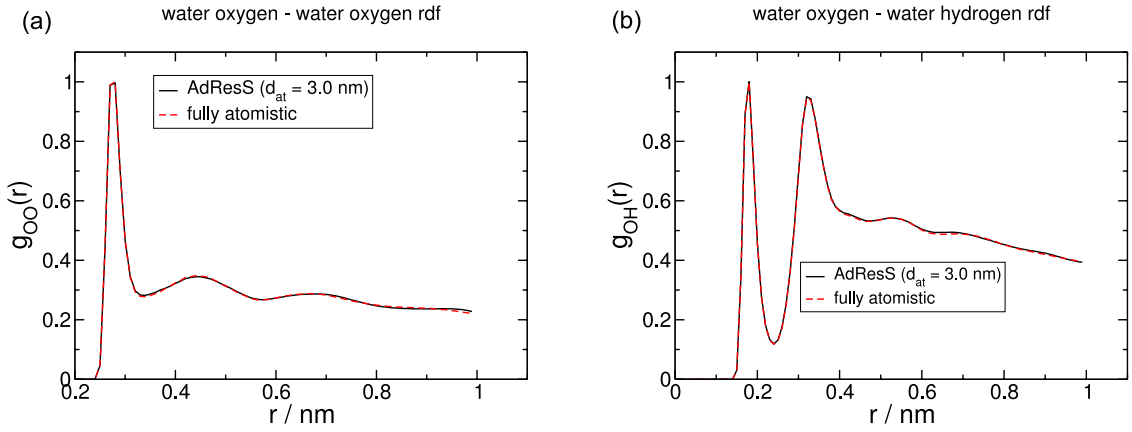

FIG. 4. Radial distribution functions (not normalised for excluded volume), in the AdResS ($d_{at} = 3.0$ nm) and fully atomistic systems, for water molecules in the sphere corresponding to the atomistic region, excluding bonded interactions.

to have a retarded, highly non-monoexponential reorientational relaxation, due to the topological and chemical effects of the biomolecule's heterogeneous surface.[86] Figure 5(a) shows that our AdResS setup correctly reproduces this behaviour, exactly matching the fully atomistic system's reorientational tcf for water in the bulk atomistic region at the time origin and for water in the first protein hydration shell at the time origin.

The non-monoexponential reorientational tcf for hydration shell water is principally due to the heterogeneous nature of the protein surface[78] and can be further understood by means of a spatial resolution of the reorientation dynamics. Following a procedure previously developed to study protein hydration shell dynamics,[86] we decompose the protein surface into sites: H-bond acceptors, H-bond donors, and non-H-bonding carbon atoms (referred to as hydrophobic sites). We then decompose the hydration shell by assigning each OH group to the site to which it is H-bonded or the site into whose hydrophobic cutoff it falls, as defined in Sec. III. We calculate reorientational tcf's for each site by following the reorientational relaxation of water molecules assigned to that site at the time origin. We extract the reorientation times for water next to each surface site via integration of the tcf and construct the probability distribution of reorientation times via weighting by the water population of each protein site. The resulting distributions for the atomistic and AdResS systems are shown in Figure 5(b). Distributions of reorientation times in the hydration shell of globular folded proteins are known to have a characteristic shape, with a peak containing the majority of the water molecules and a long-time tail.[78,86] This shape is quantitatively reproduced for the hydration shell water in the AdResS system, demonstrating that the heterogeneous, nonbulk-like reorientational dynamics of water in the hydration shell is unperturbed by the AdResS methodology.

In addition to the comparison to the fully atomistic simulation, we refer again to experimental results. The experimental NMR correlation time, which corresponds to the integral of the reorientational tcf, is approximately 2.0 ps in bulk water[79,87–90] and in molecular dynamics simulations using the SPC/E model is known to be ≈1.7 ps,[91] coherent with the value we find here (1.63 ± 0.03 ps). NMR relaxation experiments, again, give a slowdown of a factor of ≈2 relative to the bulk for the rotation of individual water molecules in the hydration shell of globular proteins.[79] In our system, the main peak of the distribution of reorientation times (Figure 5(b)) lies at ≈3.4 ps, corresponding to a slowdown factor of 2 for the majority of water molecules in the first hydration shell, in agreement with experiment. The dynamical behaviour of the

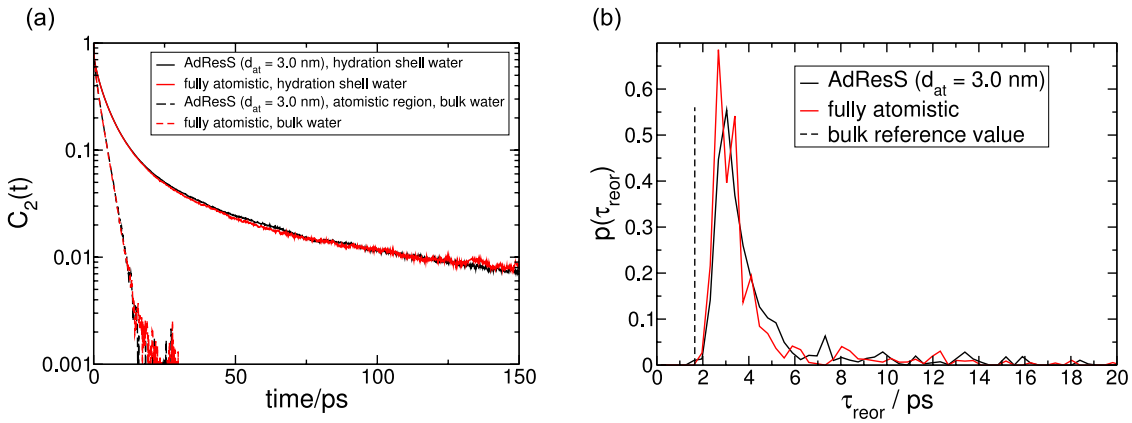

FIG. 5. In the AdResS ($d_{at} = 3.0$ nm) and fully atomistic systems: (a) Reorientation time-correlation function (Eq. (4)) for all water molecules in the first hydration shell at the time origin and for all water molecules in the bulk atomistic part of the system at the time origin. (b) Distribution of reorientation times in the first hydration shell. The dotted black line is the delta function (with an arbitrary height) corresponding to the reorientation time in bulk water.

hydration shell, whether compared to atomistic simulations or experimental results, can therefore be correctly reproduced in the AdResS methodology. We note here the importance of using a weak thermostat (e.g., for the Langevin thermostat, a coupling parameter γ of 0.1 ps^{-1} or less) or preferably the microcanonical simulation ensemble, if the correct dynamics are to be obtained. In the force-interpolation version of AdResS employed here, this is made possible by the use of the thermal bath setup described in Sec. II, where only the hybrid and coarse-grained regions are thermostatted.

The excellent agreement of protein and water structure and dynamics with both reference simulations and where possible with experiment indicates that the AdResS methodology can be used in order to study a protein's physical properties, including those that depend on specific chemical details of the hydration shell water, and at a reduced computational cost. This established, we now turn to one novel application of the methodology.

VI. SYSTEM BEHAVIOUR AS A FUNCTION OF ATOMISTIC REGION SIZE

A. Hydration level necessary for protein function

In this section, our aim is two-fold: to demonstrate how varying the size of the atomistic region allows us to draw some conclusions about the influence of the hydration water on protein behaviour and function and simultaneously to determine the optimum system size with a balance of computational efficiency and accurate physical behaviour.

One can define the hydration shell of a biomolecule as those water molecules structurally or, more often, dynamically perturbed by the proximity of the biomolecule. However, the spatial extent of this perturbation remains controversial, with experimental and computational determinations ranging from a few angstroms[67,92,93] to several hydration shells[94] or up to 20 Å [95,96] from the protein surface. In fact, the spatial extent may depend on whether individual molecules or collective properties are considered.[97] The most pertinent question in many cases may be the reverse one: how large a hydration shell is needed for protein function. In the case of multiscale simulation, this translates to the minimum number of explicitly modelled water molecules needed. Note that the discussion here focuses on proteins in dilute (mM) solutions as found in many experiments and industrial applications, and not on proteins under the conditions of macromolecular crowding found in cells.

The natural environment of proteins is evidently aqueous solution, and the importance of including solvent effects in biomolecular simulation has long been known. Proteins simulated in vacuum show deviations from the native structure, excessive intraprotein H-bonding, and changes in amplitudes of fluctuation.[98,99] The reasons for this include the absence of the friction, protein-water H-bonds, and dielectric screening provided by explicit water. Similarly, atomistic proteins simulated in coarse-grained solvent show some of these problems, when the CG solvent model used is unable to supply all of the water properties necessary for correct protein behaviour.[27]

It is interesting to consider what happens in between the two extremes of full and zero hydration. The AdResS methodology allows us to vary the size of the region modelled in explicit detail, while maintaining correct solvent density and structure in the hydration shell and correct solvent dynamics and exchange with the bulk. It allows us to vary the protein hydration level and identify the crossover point at which a breakdown of protein behaviour occurs. To this end, in addition to the AdResS system with $d_{at} = 3.0$ nm presented above, we ran trajectories with the value of d_{at} set to 2.5, 2.0, and 1.5 nm. In each case, the positions of the AA, HY, and CG regions with respect to the protein surface can be grasped from Figure 6, which shows the density of water as a function of distance from the centre of the AA region. The systems are also illustrated in Figure 1 (right). For all four systems, details are given in Table I of the correspondence between d_{at} and the number of explicitly modelled hydration shells and water molecules. We calculate this using only fully atomistic water molecules ($\lambda = 1$), and also for all water molecules with $\lambda > 0.5$, as these can be considered more atomistic than coarse-grained. We recall here that there is no atomistic protein–coarse-grained water interaction, so that the environment experienced by the protein will consist of atomistic water, and beyond that, if d_{at} is small enough, atomistic water whose interaction strength decreases following Eq. (3). In the $d_{at} = 1.5$ case, the CG solvent, which exerts no force on the protein, falls within the interaction cutoff of some protein atoms. We also simulate the protein in vacuum, for comparison. Our approach here bears some similarity to that used in a study of the structure formation of toluene around a C60 molecule as a function of atomistic region size.[100]

The protein used in this study, ubiquitin, is a regulatory protein whose function includes conjugation to other proteins and formation of polyubiquitin chains, with the seven lysine residues (distributed across the protein) and the C-terminus (residue 76) playing a functional role.[46] Here, we are interested in ubiquitin as a model for globular proteins in general, including enzymes. We therefore concentrate on general measures of structure and dynamics that will be important for the correct functioning of any globular enzyme or other protein. As an indicator of protein structure, we use the number of intra-protein hydrogen bonds, relative to the number in fully atomistic simulations (i.e., a fully hydrated protein). To probe protein dynamics, we use the error in the RMSF along the protein backbone, relative to the RMSF in the fully atomistic system. This is defined as

$$\left\langle \frac{\delta x}{x} \right\rangle = \frac{1}{N} \sum_{i=1}^{N} \frac{|x_{i,A} - x_{i,S}|}{x_{i,A}}, \tag{6}$$

where $x_{i,A}$ is the RMSF for residue i in the fully atomistic system, $x_{i,S}$ is the corresponding value in AdResS system S, and the sum is over all N residues.

The relative number of intra-protein H-bonds as a function of d_{at} is shown in Figure 7(a). A deviation from the fully hydrated value occurs between $d_{at} = 2.0$ and $d_{at} = 1.5$ nm. The increase is mostly in side chain-side chain and side chain-backbone H-bonds, not in backbone-backbone H-bonds, which are presumably already close to maximised in the fully

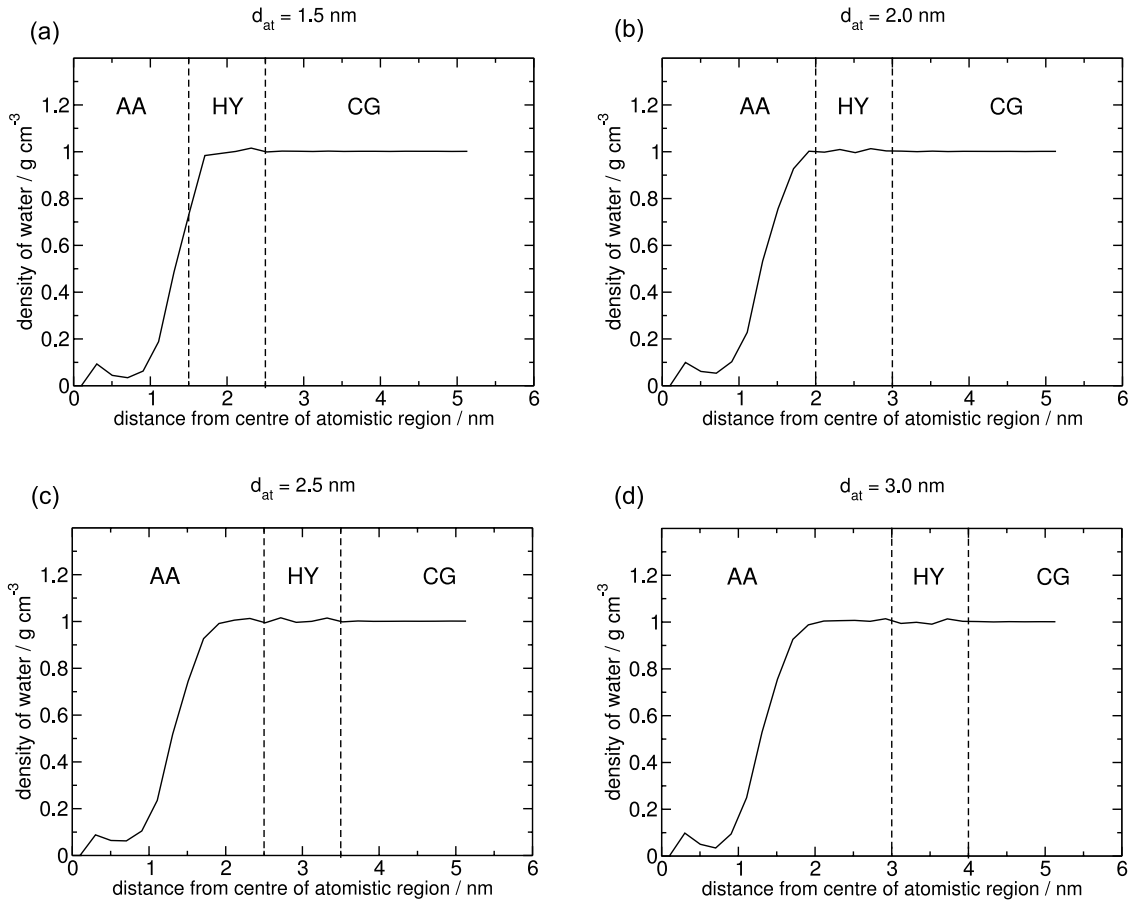

FIG. 6. Density of water as a function of distance from the centre of the atomistic region, for each of the four AdResS systems, showing in each case the positions of the atomistic, hybrid, and coarse-grained regions.

hydrated case. The error in the RMSF as a function of d_{at}, Figure 7(b), shows a similar divergence at the same hydration level. We note that error in the RMSF for individual residues is not correlated with their closeness to the CG region, seen by the protein as a vacuum region. What we see at the point $d_{at} = 1.5$ nm and lower is a change in global protein dynamics, not simply in the local dynamics of those residues which are least hydrated.

The deviation from the behaviour displayed by fully hydrated proteins occurs at a surprisingly low hydration level. At $d_{at} = 2.0$ nm, where some parts of the protein have only one layer of atomistic water, the protein still behaves as though it were solvated in fully atomistic solution. However, this should not be interpreted as a claim that the optimum protein simulation system includes so few water molecules. Rather, we point out that *while certain types of water dynamics may be perturbed out to a long distance by the presence of the protein, this does not necessarily imply the inverse, i.e., that water at that distance plays a significant role in protein structure or dynamics*.

TABLE I. Quantification of the hydration level of the four AdResS systems.

	Including water with $\lambda = 1$		Including water with $\lambda > 0.5$	
d_{at} / nm	No. of water molecules	No. of hydration shells	No. of water molecules	No. of hydration shells
3.0	3466	3–9	5684	4–11
2.5	1852	2–7	3466	3–9
2.0	781	1–5	1852	2–7
1.5	162	0–3	781	1–5

It is of course important to bear in mind that the values presented here are measures of protein behaviour, not protein function, and that accurate reproduction of these measures is necessary but not necessarily sufficient for correct protein function.

In an aside, we note here that a related topic is the use of proteins, specifically enzymes, in organic solvent, an expanding field of research and industrial application.[70] Enzymes can function in organic solvent, albeit with greatly reduced efficiency, and the catalytic activity may be correlated to the amount of trace water present in the system.[101] The behaviour of proteins simulated in pure non-polar organic solvents[102] may be compared to that in many CG forcefields, with similar ideas of insufficient screening or lack of hydrogen bonding ability leading to altered enzyme properties and therefore reduced activity. Molecular dynamics studies have been performed varying the number of water molecules in the hydration layer of enzymes in organic solvent.[103] The difference between those studies and our approach here is that the water between a protein surface and an organic solvent interface is in a situation of extreme confinement, with greatly retarded water dynamics,[86] and not directly relevant to the

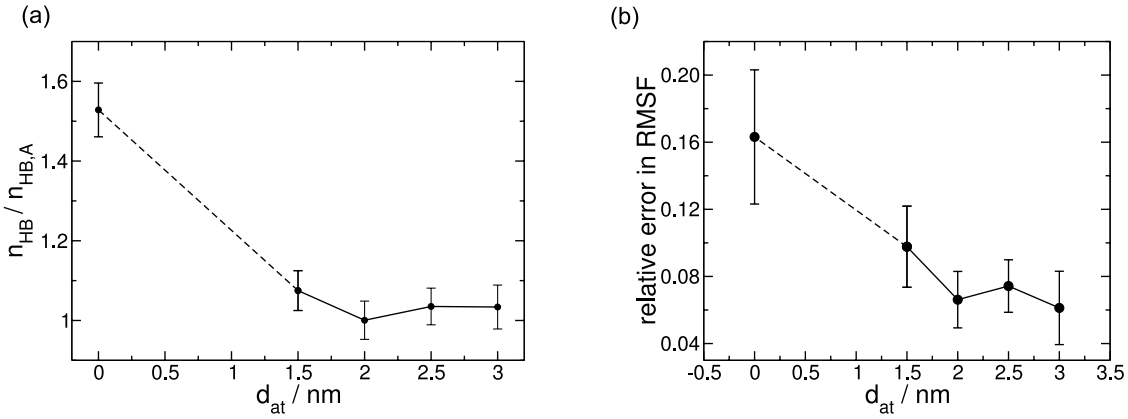

FIG. 7. (a) Number of intra-protein H-bonds relative to the value in the fully atomistic system, as a function of the d_{at} value in each system. (b) Error in RMSF relative to RMSF in the fully atomistic system, as defined in Eq. (6), as a function of the d_{at} value in each system. In each figure, the dotted lines show the transition from AdResS to vacuum systems.

understanding of protein and hydration shell behaviour in bulk aqueous solution.

We note also that there exists a set of simulation methodologies that use a thin layer of explicit atomistic water molecules to model protein hydration, generally with the aim of obtaining greater computational efficiency.[83,104–106] In order to prevent evaporation in such simulations, some sort of restraining force is typically applied to keep water near the protein surface. The bulk solvent may be entirely absent or may be included in a mean field way.[106] Such methodologies naturally provide an improvement in protein structural and dynamical properties over the gas phase[105] and could also be used to study the impact of the size of the explicit hydration shell. However, they are lacking the free particle exchange with bulk water provided by the AdResS methodology and the corresponding exact reproduction of water properties.

B. Water dynamics in the hydration shell

We turn now to the dynamical behaviour of the hydration shell water as a function of the size of the atomistic region. Figure 8(a) shows the reorientational tcf for all molecules in the hydration shell at the time origin, for all systems studied. For d_{at} = 3.0 and 2.5 nm, this is identical to the tcf in the fully atomistic simulations. At d_{at} = 2.0 nm, we see a slowdown in the water dynamics, while the tcf calculated at d_{at} = 1.5 nm seems to display both a speedup and a slowdown of water dynamics. The underlying distributions of reorientation times, shown in Figure 8(c), make this clear. For d_{at} = 1.5 nm, the distribution has broadened, with part of the water population shifted to longer times and part shifted to shorter times, i.e., both a speedup and a slowdown occurring in different parts of the hydration shell. For d_{at} = 2.0 nm, there is simply a shift to longer times, corresponding to the slowdown seen in the tcf. This complex behaviour can be understood by examining how these tcf's change when only fully atomistic water is considered (Figure 8(b)) and by considering how water reorientational dynamics varies as a function of a molecule's position in the AA or HY region (Figure 9).

We begin by discussing the variations in water reorientation time across the AA and HY regions, as shown in Figure 9. In the AA region (left of figure), the reorientation time of course corresponds to that in bulk atomistic reference simulations, as already demonstrated above. Approaching the HY region, a slight slowdown is observed, likely related to the asymmetric nature of the environment in which the water molecules are found. Crossing the HY region from the AA to the CG side, reorientational dynamics then speeds up as the water-water interaction potential becomes more isotropic with decreasing λ value and the transformation of AA into CG water. Indeed, the defined orientation axis used to calculate the correlation function becomes less and less meaningful. In the CG region, of course, coarse-grained water molecules are points with no rotational degrees of freedom.

Returning now to a comparison of Figures 8(a) and 8(b), which show reorientational tcf's, respectively, for all waters in the hydration shell and for fully atomistic waters in the hydration shell, we see that the speedup at short times in Figure 8(a) at d_{at} = 1.5 nm disappears in Figure 8(b) and is due to the inclusion in Figure 8(a) of the quickly reorienting waters within the HY region. In the underlying distributions of reorientation times, corresponding to each set of tcf's, these fast waters are clearly visible in the d_{at} = 1.5 nm distribution in Figure 8(c) and absent in Figure 8(d).

We now turn to the slower decay at longer times in the tcf's for d_{at} = 1.5 and 2.0 nm in Figure 8(a). The slowdown can be quantified by calculating the average reorientation time within the main peak of the distribution for d_{at} = 2.0 nm and the corresponding (identical) values for d_{at} = 2.5 or 3.0 nm (Figure 8(c)). The ratio between these average times gives a value of 1.1 for the slowdown factor. This can be quantitatively understood by recalling that at d_{at} = 2.0 nm most of the hydration shell lies close to the AA/HY interface, where water dynamics is slowed down by a factor of 1.1, as seen in Figure 9 and discussed above. The same explanation holds for that subpopulation of water molecules, which is slowed down at d_{at} = 1.5 nm.

The tcf for atomistic water only, at d_{at} = 1.5 nm (Figure 8(b)) shows an additional slowdown, which can be understood as follows. Exposed water molecules next to convex, protruding parts of the protein surface are known to reorient more quickly than more deeply buried molecules.[78] At low d_{at} values and when including only fully atomistic water in the calculation, those water molecules remaining are

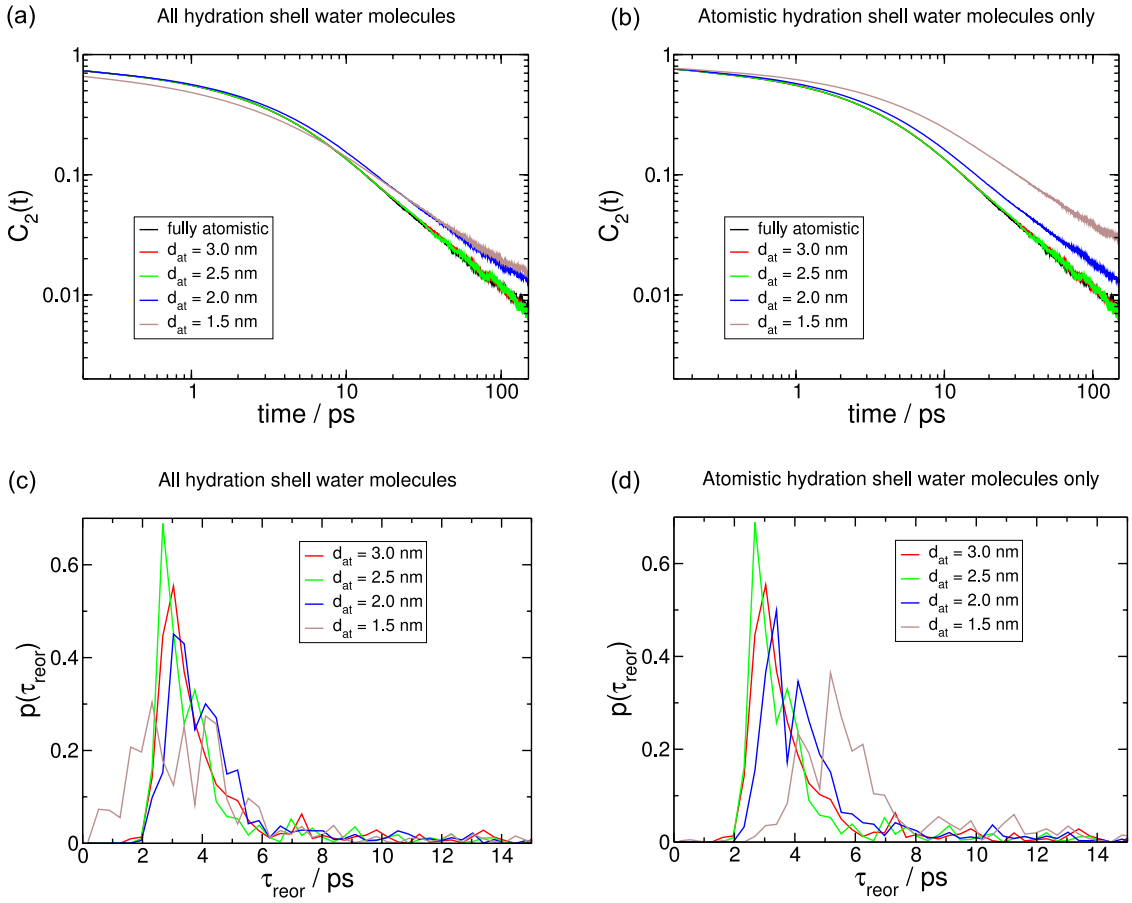

FIG. 8. Hydration shell dynamics as a function of the size of the atomistic region. (a) Reorientational tcf for all water molecules in the hydration shell at the time origin. (b) Reorientational tcf for all fully atomistic ($\lambda = 1$) water molecules in the hydration shell at the time origin. (c) Distribution of reorientation times underlying the tcf's in (a). (d) Distribution of reorientation times underlying the tcf's in (b). For clarity, the fully atomistic reference results, already shown in Figure 5, are not repeated here in (c) and (d).

therefore the slower ones, leading to the perceived slowdown in the calculated tcf. In the corresponding distribution for $d_{at} = 1.5$ nm in Figure 8(d), similarly, only the slower water molecules remain.

The above discussion allows us to understand the behaviour of the hydration shell at low d_{at} values. In summary, taking into account the correct reproduction of both protein and hydration shell properties, it seems the minimum advisable value of d_{at} in this system is 2.5 nm. We now compare this value to the dimensions of the protein used here in order to obtain general rules of thumb for the choice of d_{at}. Ubiquitin has an extremely flexible six-residue tail protruding from its globular core, which is important in biological function but which we exclude from the discussion here in order to provide better and more general guidelines applicable to all globular proteins. Not including this tail, the minimum distance from the protein surface to the AA/HY interface is ≈0.5 nm at $d_{at} = 2.5$ nm, while the average thickness of the atomistic layer is ≈1.1 nm. For comparison, the cutoff used for atomistic and coarse-grained non-bonded interactions is 1.0 nm. Another useful measure of protein dimension is the radius of gyration, which for ubiquitin is ≈1.2 nm, meaning that a value of roughly $(R_g + 1.3)$ nm could serve as another rule of thumb for the minimum size of the atomistic region in biomolecular AdResS.

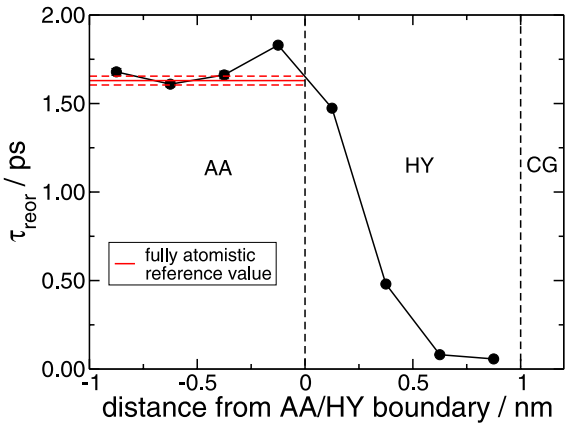

FIG. 9. Reorientation time as a function of a water molecule's location at the time origin. Error bars are within symbol size. The solid red line shows the bulk reorientation time in the fully atomistic simulations, and the dashed red lines are the corresponding error bars.

VII. CONCLUSION

In this paper, we have extended the AdResS methodology to biomolecular systems, demonstrating that for a fully atomistic globular protein with an atomistic hydration shell coupled to a coarse-grained particle reservoir and heat bath, the structural and dynamical properties of both the protein and the atomistic aqueous solvent are correctly reproduced.

We have also determined the minimum possible size of the atomistic region in such a simulation setup.

For the systems studied here, our current implementation of the AdResS methodology provides a speedup of 1.6–2.2, depending on the size of the atomistic region. In principle, a much more significant speedup could be obtained, via, for example, the use of a non-interacting solvent in the coarse-grained region[40] or the implementation of multiple time-stepping. However, it is important to stress that the primary interest of the method is not merely computational speedup, but also the possibility of gaining new physical insight by performing simulations which would otherwise be unfeasible for reasons other than simply the limitations of computational resources. We presented here one example of such an application, demonstrating how our methodology can provide insight into the hydration level at which protein behaviour begins to break down. We showed that, unexpectedly, the structure and dynamics of a globular protein are correctly reproduced even with only a handful of hydration shells. Other examples of applications with biomolecular relevance include the reduction of finite size effects[107] and the possibility of driving system processes via the easy insertion and deletion of particles when coupling to a bath of non-interacting particles.[40]

Further possible extensions to the work presented here include the study of energetic or thermodynamic properties and the implementation of an atomistic region which follows the shape of the biomolecular surface or changes in size as a function of the biomolecule's conformation.[108] Some pioneering works have dealt with systems in which the boundary between atomistic and coarse-grained regions lies within the biomolecule itself.[17,18] One can imagine simulation setups involving not only adaptive resolution solvent but also mixed, adaptive resolution biomolecules.

ACKNOWLEDGMENTS

K.K. and A.C.F. acknowledge research funding through the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement No. 340906-MOLPROCOMP. We are grateful to Torsten Stuehn for assistance with the ESPResSo++ package and to Debashish Mukherji, Tristan Bereau, and Frédéric Leroy for a critical reading of the manuscript.

[1]M. Karplus and J. A. McCammon, Annu. Rev. Biochem. **53**, 263 (1983).
[2]A. Krushelnitsky, D. Reichert, and K. Saalwächter, Acc. Chem. Res. **46**, 2028 (2013).
[3]K. A. Dill and J. L. MacCallum, Science **338**, 1042 (2012).
[4]H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink, Wiley Interdiscip. Rev.: Comput. Mol. Sci. **4**, 225 (2014).
[5]K. Meier, A. Choutko, J. Dolenc, A. P. Eichenberger, S. Riniker, and W. F. van Gunsteren, Angew. Chem., Int. Ed. **52**, 2820 (2013).
[6]V. Tozzini, Acc. Chem. Res. **43**, 220 (2010).
[7]T. A. Wassenaar, H. I. Ingólfsson, M. Prie, S. J. Marrink, and L. V. Schäfer, J. Phys. Chem. B **117**, 3516 (2013).
[8]S. Riniker, A. P. Eichenberger, and W. F. van Gunsteren, Eur. Biophys. J. **41**, 647 (2012).
[9]Q. Shi, S. Izvekov, and G. A. Voth, J. Phys. Chem. B **110**, 15045 (2006).
[10]J. Kleinjung and F. Fraternali, Curr. Opin. Struct. Biol. **25**, 126 (2014).
[11]M. Feig and C. L. Brooks III, Curr. Opin. Struct. Biol. **14**, 217 (2004).
[12]A. V. Predeus, S. Gul, S. M. Gopal, and M. Feig, J. Phys. Chem. B **116**, 8610 (2012).
[13]V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005).
[14]T. Bereau and M. Deserno, J. Chem. Phys. **130**, 235106 (2009).
[15]T. E. Ouldridge, A. A. Louis, and J. P. K. Doye, J. Chem. Phys. **134**, 085101 (2011).
[16]J. K. Sigurdsson, F. L. Brown, and P. J. Atzberger, J. Comput. Phys. **252**, 65 (2013).
[17]M. Neri, C. Anselmi, M. Cascella, A. Maritan, and P. Carloni, Phys. Rev. Lett. **95**, 218102 (2005).
[18]M. R. Machado, P. D. Dans, and S. Pantano, Phys. Chem. Chem. Phys. **13**, 18134 (2011).
[19]W. Han and K. Schulten, J. Chem. Theory Comput. **8**, 4413 (2012).
[20]J. Maupetit, P. Tuffery, and P. Derreumaux, Proteins: Struct., Funct., Bioinf. **69**, 394 (2007).
[21]I. Coluzza, PLoS One **9**, e112852 (2014).
[22]M. Leguèbe, C. Nguyen, L. Capece, Z. Hoang, A. Giorgetti, and P. Carloni, PLoS One **7**, e47332 (2012).
[23]E. Villa, A. Balaeff, and K. Schulten, Proc. Natl. Acad. Sci. U. S. A. **102**, 6783 (2005).
[24]Y. Levy and J. N. Onuchic, Annu. Rev. Biophys. Biomol. Struct. **35**, 389 (2006).
[25]A. M. Klibanov, Trends Biotechnol. **15**, 97 (1997).
[26]A. Kitao, F. Hirata, and N. Gō, Chem. Phys. **158**, 447 (1991).
[27]S. Riniker, A. P. Eichenberger, and W. F. van Gunsteren, J. Phys. Chem. B **116**, 8873 (2012).
[28]C. McCabe and K. R. Hadley, Mol. Simul. **38**, 671 (2012).
[29]J. Lu, Y. Qiu, R. Baron, and V. Molinero, J. Chem. Theory Comput. **10**, 4104 (2014).
[30]M. E. Johnson, T. Head-Gordon, and A. A. Louis, J. Chem. Phys. **126**, 144509 (2007).
[31]H. Wang, C. Junghans, and K. Kremer, Eur. Phys. J. E **28**, 221 (2009).
[32]A. A. Louis, J. Phys.: Condens. Matter **14**, 9187 (2002).
[33]F. H. Stillinger, H. Sakai, and S. Torquato, J. Chem. Phys. **117**, 288 (2002).
[34]G. D'Adamo, A. Pelissetto, and C. Pierleoni, J. Chem. Phys. **138**, 234107 (2013).
[35]M. Praprotnik, L. D. Site, and K. Kremer, J. Chem. Phys. **123**, 224106 (2005).
[36]M. Praprotnik, S. Matysiak, L. D. Site, K. Kremer, and C. Clementi, J. Phys.: Condens. Matter **19**, 292201 (2007).
[37]S. Matysiak, C. Clementi, M. Praprotnik, K. Kremer, and L. D. Site, J. Chem. Phys. **128**, 024503 (2008).
[38]J. Zavadlav, M. N. Melo, S. J. Marrink, and M. Praprotnik, J. Chem. Phys. **140**, 054114 (2014).
[39]H. Wang, C. Hartmann, C. Schütte, and L. D. Site, Phys. Rev. X **3**, 011018 (2013).
[40]K. Kreis, A. C. Fogarty, K. Kremer, and R. Potestio, "Advantages and challenges in coupling an ideal gas to atomistic models in adaptive resolution simulations," Eur. Phys. J.: Spec. Top. e-print arXiv:1412.6810 [cond] (in press).
[41]S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. D. Site, and K. Kremer, Phys. Rev. Lett. **108**, 170602 (2012).
[42]D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[43]A. Agarwal, H. Wang, C. Schütte, and L. D. Site, J. Chem. Phys. **141**, 034102 (2014).
[44]R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, Phys. Rev. Lett. **110**, 108301 (2013).
[45]M. Praprotnik, L. D. Site, and K. Kremer, Annu. Rev. Phys. Chem. **59**, 545 (2008).
[46]C. M. Pickart and M. J. Eddins, Biochim. Biophys. Acta, Mol. Cell Res. **1695**, 55 (2004).
[47]S. Vijay-Kumar, C. E. Bugg, and W. J. Cook, J. Mol. Biol. **194**, 531 (1987).
[48]Y. Duan, C. Wu, S. Chowdhury, M. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman, J. Comput. Chem. **24**, 1999 (2003).
[49]H. Berendsen, J. Grigera, and T. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[50]V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, J. Chem. Theory Comput. **5**, 3211 (2009).
[51]J. D. Halverson, T. Brandes, O. Lenz, A. Arnold, S. Bevc, V. Starchenko, K. Kremer, T. Stuehn, and D. Reith, Comput. Phys. Commun. **184**, 1129 (2013).
[52]Using the Langevin thermostat in the hybrid and coarse-grained regions, we found it necessary to thermostat at 299 K in order to have a temperature in the atomistic region of 300 K. This is due to the fact that particles which gain excess heat in the hybrid region for the reasons outlined above may not have time to become fully thermalised before

they cross into the non-thermostated atomistic region. This artefact does not occur with the use of, e.g., a velocity-rescaling thermostat in the hybrid region. However, thermostats which employ rescaling require the calculation of a global temperature, or at least a temperature over some reasonably large area, which is not straightforward in the AdResS setup. In addition, such thermostats do not sample the canonical ensemble.

[53]L. Li, C. Li, Z. Zhang, and E. Alexov, J. Chem. Theory Comput. **9**, 2126 (2013).
[54]S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).
[55]G. Lipari and A. Szabo, J. Am. Chem. Soc. **104**, 4546 (1982).
[56]E. A. Cino, W.-Y. Choy, and M. Karttunen, J. Chem. Theory Comput. **8**, 2725 (2012).
[57]S. Piana, J. L. Klepeis, and D. E. Shaw, Curr. Opin. Struct. Biol. **24**, 98 (2014).
[58]P. L. Freddolino, S. Park, B. Roux, and K. Schulten, Biophys. J. **96**, 3772 (2009).
[59]J. Kuriyan, G. A. Petsko, R. M. Levy, and M. Karplus, J. Mol. Biol. **190**, 227 (1986).
[60]J. B. Clarage and G. N. Phillips, Jr., Acta Crystallogr., Sect. D: Biol. Crystallogr. **50**, 24 (1994).
[61]A. E. Garca, J. A. Krumhansl, and H. Frauenfelder, Proteins: Struct., Funct., Bioinf. **29**, 153 (1997).
[62]R. P. Joosten, T. Womack, G. Vriend, and G. Bricogne, Acta Crystallogr., Sect. D: Biol. Crystallogr. **65**, 176 (2009).
[63]R. P. Joosten and G. Vriend, Science **317**, 195 (2007).
[64]R. P. Joosten, F. Long, G. N. Murshudov, and A. Perrakis, IUCrJ **1**, 213 (2014).
[65]C. Charlier, S. N. Khan, T. Marquardsen, P. Pelupessy, V. Reiss, D. Sakellariou, G. Bodenhausen, F. Engelke, and F. Ferrage, J. Am. Chem. Soc. **135**, 18665 (2013).
[66]P. Ball, Chem. Rev. **108**, 74 (2008).
[67]B. Halle, Philos. Trans. R. Soc., B **359**, 1207 (2004).
[68]A. Mukherjee, R. Lavery, B. Bagchi, and J. T. Hynes, J. Am. Chem. Soc. **130**, 9747 (2008).
[69]L. Szyc, M. Yang, E. E. E. Nibbering, and T. Elsaesser, Angew. Chem., Int. Ed. **49**, 3598 (2010).
[70]A. M. Klibanov, Nature **409**, 241 (2001).
[71]D. I. Svergun, S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai, Proc. Natl. Acad. Sci. U. S. A. **95**, 2267 (1998).
[72]F. Merzel and J. C. Smith, Proc. Natl. Acad. Sci. U. S. A. **99**, 5378 (2002).
[73]M. Agarwal, H. R. Kushwaha, and C. Chakravarty, J. Phys. Chem. B **114**, 651 (2010).
[74]R. H. Henchman and J. A. McCammon, Protein Sci. **11**, 2080 (2002).
[75]N. Bhattacharjee and P. Biswas, Biophys. Chem. **158**, 73 (2011).
[76]A. R. Bizzarri and S. Cannistraro, Phys. Rev. E **53**, R3040 (1996).
[77]Y. von Hansen, S. Gekle, and R. R. Netz, Phys. Rev. Lett. **111**, 118103 (2013).
[78]F. Sterpone, G. Stirnemann, and D. Laage, J. Am. Chem. Soc. **134**, 4116 (2012).
[79]C. Mattea, J. Qvist, and B. Halle, Biophys. J. **95**, 2951 (2008).
[80]J. T. King, E. J. Arthur, C. L. Brooks, and K. J. Kubarych, J. Phys. Chem. B **116**, 5604 (2012).
[81]A. C. Fogarty, E. Duboue-Dijon, F. Sterpone, J. T. Hynes, and D. Laage, Chem. Soc. Rev. **42**, 5672 (2013).
[82]G. Neumayr, T. Rudas, and O. Steinhauser, J. Chem. Phys. **133**, 084108 (2010).
[83]M. B. Hamaneh and M. Buck, J. Comput. Chem. **30**, 2635 (2009).
[84]J. R. Errington and P. G. Debenedetti, Nature **409**, 318-321 (2001).
[85]G. Stirnemann, F. Sterpone, and D. Laage, J. Phys. Chem. B **115**, 3254 (2011).
[86]A. C. Fogarty and D. Laage, J. Phys. Chem. B **118**, 7715 (2014).
[87]R. Ludwig, F. Weinhold, and T. C. Farrar, J. Chem. Phys. **103**, 6941 (1995).
[88]J. van der Maarel, D. Lankhorst, J. de Bleijser, and J. Leyte, Chem. Phys. Lett. **122**, 541 (1985).
[89]D. Lankhorst, J. Schriever, and J. C. Leyte, Ber. Bunsenges. Phys. Chem. **86**, 215 (1982).
[90]D. Smith and J. Powles, Mol. Phys. **10**, 451 (1966).
[91]D. Laage and J. T. Hynes, J. Phys. Chem. B **112**, 14230 (2008).
[92]P. J. Rossky and M. Karplus, J. Am. Chem. Soc. **101**, 1913 (1979).
[93]M. Marchi, F. Sterpone, and M. Ceccarelli, J. Am. Chem. Soc. **124**, 6787 (2002).
[94]D. R. Martin and D. V. Matyushov, J. Chem. Phys. **141**, 22D501 (2014).
[95]B. Born, S. J. Kim, S. Ebbinghaus, M. Gruebele, and M. Havenith, Faraday Discuss. **141**, 161 (2009).
[96]K. Meister, S. Ebbinghaus, Y. Xu, J. G. Duman, A. DeVries, M. Gruebele, D. M. Leitner, and M. Havenith, Proc. Natl. Acad. Sci. U. S. A. **110**, 1617 (2013).
[97]M. Heyden, J. Chem. Phys. **141**, 22D509 (2014).
[98]M. Levitt and R. Sharon, Proc. Natl. Acad. Sci. U. S. A. **85**, 7557 (1988).
[99]P. J. Steinbach and B. R. Brooks, Proc. Natl. Acad. Sci. U. S. A. **90**, 9135 (1993).
[100]S. Fritsch, C. Junghans, and K. Kremer, J. Chem. Theory Comput. **8**, 398 (2012).
[101]V. Kurkal, R. Daniel, J. L. Finney, M. Tehei, R. Dunn, and J. C. Smith, Biophys. J. **89**, 1282 (2005).
[102]M. Norin, F. Haeffner, K. Hult, and O. Edholm, Biophys. J. **67**, 548 (1994).
[103]R. Wedberg, J. Abildskov, and G. H. Peters, J. Phys. Chem. B **116**, 2575 (2012).
[104]R. Sankararamakrishnan, K. Konvicka, E. L. Mehler, and H. Weinstein, Int. J. Quantum Chem. **77**, 174 (2000).
[105]D. Beglov and B. Roux, Biopolymers **35**, 171 (1995).
[106]V. Lounnas, S. K. Lüdemann, and R. C. Wade, Biophys. Chem. **78**, 157 (1999).
[107]D. Mukherji and K. Kremer, Macromolecules **46**, 9158 (2013).
[108]J. A. Wagoner and V. S. Pande, J. Chem. Phys. **139**, 234114 (2013).
[109]D. Alexeev, S. M. Bury, M. A. Turner, O. M. Ogunjobi, T. W. Muir, R. Ramage, and L. Sawyer, Biochem. J. **299**, 159 (1994), http://www.biochemj.org/bj/299/bj2990159.htm.
[110]D. M. Schneider, M. J. Dellwo, and A. J. Wand, Biochemistry **31**, 3645 (1992).