Direct mixing of atomistic solutes and coarse-grained water

Mario Orsi,[*] Wei Ding, and Michalis Palaiokostas

School of Engineering & Materials Science, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom

E-mail: m.orsi@qmul.ac.uk

Abstract

We present a new dual-resolution approach for coupling atomistic and coarse-grained models in molecular dynamics simulations of hydrated systems. In particular, a coarse-grained point dipolar water model is used to solvate molecules represented with standard all-atom force fields. A unique characteristic of our methodology is that the mixing of resolutions is direct, meaning that no additional or *ad hoc* scaling factors, intermediate regions, or extra sites are required. To validate the methodology, we compute the hydration free energy of fourteen atomistic small molecules (analogs of amino acid sidechains) solvated by the coarse-grained water. Remarkably, our predictions reproduce the experimental data as accurately as the predictions from state-of-the-art fully atomistic simulations. We also show that the hydration free energy of the coarse-grained water itself is in comparable or better agreement with the experimental value than the predictions from all but one of the most common multisite atomistic models. The coarse-grained water is then applied to solvate a typical atomistic protein containing both α-helix and β-strand elements. Moreover, parallel tempering simulations

---
[*]To whom correspondence should be addressed

are performed to investigate the folding free energy landscape of a representative α helical and a β hairpin structure. For the simulations considered in this work, our dual-resolution method is found to be three to six times more computationally efficient than corresponding fully atomistic approaches.

1 Introduction

The relevance of water as a solvent in countless natural and industrial processes[1,2] is reflected in its common presence in molecular simulations, and in the multitude of hydration models that have been proposed in the literature, over several decades now.[3-9] An important computational aspect in the simulation of explicitly hydrated systems is that the large majority of the computation time is typically spent calculating water-water interactions. It is therefore unsurprising that numerous methods and models have been developed to simplify the treatment of hydration, and hence reduce the corresponding computational cost. In this respect, an increasingly popular approach involves the development of particle-based coarse-grained (CG) models, where one or more water molecules are represented by single interaction sites.[9-17]

An interesting and potentially very useful issue to consider is whether and how CG water models can be used to hydrate molecules described by standard atomistic models. Such a dual-resolution approach is highly desirable, because it allows the CG efficiency to be combined with the accuracy and generality of atomistic force fields - at least in principle. In practice however, complications arise because the CG force fields are not normally compatible with the atomistic ones. In fact, existing dual-resolution hydration schemes rely on one or more of the following *ad hoc* procedures to couple CG water and atomistic solutes: extra parameters or scaling factors to calibrate the atomistic-CG interactions,[18-24] specific parametrization of atomistic-CG interactions,[25-29] additional CG virtual sites,[24,30-32] artificial relative dielectric permittivity between atomistic sites due to lack of CG water electrostatic screening,[29,30] additional layers of atomistic water between the atomistic molecules

and the CG water,[23,32,33] or "adaptive resolution" transition regions.[34]

In this work, we present a new direct approach to CG hydration of atomistic molecules, where the two levels of resolution (atomistic and CG) coexist in the same simulation without requiring any *ad hoc* treatment of the mixed interactions. Our methodology is based on the use of the ELBA CG force field,[35–37] where each water molecule is represented by a point dipole attached to the center of a Lennard-Jones sphere; such a combination of potentials is also known as the Stockmayer model. While the idea of parametrizing a "Lennard-Jones plus point dipole" potential to model water solvation was proposed by Warshel already 35 years ago,[38] the Stockmayer model has been almost exclusively employed to study idealized polar fluids.[39–42] The ELBA model is characterized by a novel parametrization targeted to liquid water, and by an original "shifted-force" variant of the point dipole potential (which, as discussed in more details in Section 2.1, is crucial to the viability and efficiency of the model in molecular dynamics simulations). Recently, ELBA was shown to accurately reproduce many experimental properties of bulk water and the water-vapor interface, including density, diffusion, surface tension, vapor-liquid equilibria, and the critical point; in fact, for several properties, ELBA was found to be as accurate as the best atomistic models commonly used, while proving one to two orders of magnitude more computationally efficient.[37]

It is shown in this article that the ELBA water model can also be used straightforwardly to hydrate molecules described by standard atomistic force fields. Specifically, the mixing of CG and atomistic interactions is obtained simply through shifted-force potentials with mixed parameters determined from the same standard rules employed for interactions among the atomistic sites. Thus, for all the interactions in the system, Lennard-Jones cross terms are determined by the Lorentz-Berthelot combining rules[43] (which involve simple geometric and arithmetic means), and electrostatic cross terms are determined from classical Coulomb expressions, with the relative permittivity set to unity ($\epsilon_r = 1$). The validity of our approach is first tested by computing the free energy of hydration for several atomistically-modeled small molecules (analogs of amino acids) solvated in ELBA water. Many among the most

fundamental (bio)molecular processes, such as self-assembly, ligand binding, transmembrane permeation, and protein folding, are regulated by the free energy of hydration, and hence its accurate reproduction is paramount. The ELBA water is then applied to solvate a full typical protein structure, comprising an α-helix and four β-strands, modeled with a standard all-atom force field. Furthermore, parallel tempering simulations are conducted to study the folding free energy landscape of two typical α helical and β hairpin elements. The presentation of the results is followed by a general critical discussion of our approach, including current limitations and possible future improvements.

2 Models

2.1 Coarse-grained ELBA water

The ELBA water model[35–37] is an original parametrization of the general "Lennard-Jones plus point dipole" (Stockmayer) potential,[38–42,44–46] where each real water molecule is described by a single coarse-grained site (1).

Figure 1: Water coarse-graining. The left image depicts a single water molecule; the oxygen atom (blue) bears a negative charge (red "-" sign), while the two hydrogen atoms (cyan) bear positive charges (red "+" signs). The right image depicts an ELBA coarse-grained water site; the red arrow represents a permanent electrical point dipole.

The total potential energy U_{ij} of an interacting pair of sites i, j is:

$$U_{ij} = U_{ij}^{\text{LJ}} + U_{ij}^{\text{dip}} \tag{1}$$

4

with U_{ij}^{LJ} the Lennard-Jones term and U_{ij}^{dip} the point dipole term. A notable feature of ELBA is the use of "shifted-force" variants of the original potentials, so that both the energy and its gradient (the force) go to zero smoothly at the cutoff.[43,47] This removes cutoff-related artifacts in the particles' motion (especially problematic for orientation dependent potentials such as the point dipole potential[48]), improves simulation stability and energy conservation, and hence permits longer integration timesteps. For the Lennard-Jones term, we use the following shifted-force expression:[49]

$$U_{ij}^{\text{LJ}} = 4\epsilon \left\{ \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right] + \left[6 \left(\frac{\sigma}{r_c} \right)^{12} - 3 \left(\frac{\sigma}{r_c} \right)^{6} \right] \left(\frac{r}{r_c} \right)^{2} - 7 \left(\frac{\sigma}{r_c} \right)^{12} + 4 \left(\frac{\sigma}{r_c} \right)^{6} \right\} \quad (2)$$

where r is the intersite distance, r_c is the cutoff distance, and the constants σ and ϵ have the standard meaning.[43,47] Regarding the point dipole component of Eq. (1), the following shifted-force potential was derived originally[36,37] from the classical electrostatic formula:[43,50]

$$U_{ij}^{\text{dip}} = \frac{p^2}{4\pi\varepsilon_0} \left[1 - 4 \left(\frac{r}{r_c} \right)^3 + 3 \left(\frac{r}{r_c} \right)^4 \right] \left[\frac{1}{r^3}(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{p}}_j) - \frac{3}{r^5}(\hat{\mathbf{p}}_i \cdot \mathbf{r})(\hat{\mathbf{p}}_j \cdot \mathbf{r}) \right] \quad (3)$$

where p is the magnitude of each of the two interacting point dipoles, $\hat{\mathbf{p}}_i$ and $\hat{\mathbf{p}}_j$ are the unit point dipole vectors, $\mathbf{r}$ and r are respectively the pair distance vector and its magnitude, and r_c is the cutoff distance. Analytical expressions for the forces and torques corresponding to the potentials of Eqs. (2) and (3) can be found elsewhere.[37] The parameters used here are the same as in previous work,[36,37] that is: $\epsilon = 0.55\,\text{kcal}\,\text{mol}^{-1}$, $\sigma = 3.05\,\text{Å}$, $p = 2.6\,\text{D}$, $r_c = 12\,\text{Å}$.

2.2 All-atom small molecules and protein

We considered the following fourteen small molecules, analogs of amino acid sidechains (the corresponding amino acid codes are reported between brackets): methane (Ala), propane (Val), n-butane (Ile), isobutane (Leu), methanol (Ser), ethanol (Thr), toluene (Phe), p-cresol (Tyr), methanethiol (Cys), ethylmethylsulfide (Met), acetamide (Asn), propionamide (Gln),

3-methylindole (Trp), 4-methylimidazole (Hid/Hie). Initial atom coordinates were obtained by building the molecules with the *xleap* program from the AmberTools12 package.[51] For each molecule, two models were then generated using two of the most widely used all-atom force fields. The first set of models was described with the CHARMM General Force Field (CGenFF),[52] version 2b7. The *charmm2lammps.pl* tool[53] was used to convert the original *CGenFF* topology and parameter files into LAMMPS input files (which require a different format, as well as conversions to different units for some of the parameters). The second set was represented with the General Amber Force Field (GAFF).[54] The *antechamber*[55] program from the AmberTools12 package[51] was used to assign partial charges with the AM1/BCC method.[56] The resulting files were converted into LAMMPS input files using the *amber2lammps.py* tool.[53] It can be noted that the CGenFF and GAFF force fields are qualitatively similar, in that they comprise the same (or very similar) functional forms for the various potential energy terms. However, they differ quantitatively, in terms of the specific parameters used (charges, Lennard-Jones constants, reference values and rigidity constants of the bonded terms). Also, they adopt different parametrization strategies, especially for the electrostatic terms. Details can be found in the original publications.[52,54]

We then considered protein G,[57] a 56-residue structure which contains both an α-helix and β-strands, and hence represents a good case study.[22,23,58] Protein data bank entry 1PGB[57] was used to obtain initial coordinates for all atoms except hydrogens. The *psfgen* plugin from the VMD program[59] was used to add hydrogens and to assign parameters and topologies from the CHARMM22 All-Hydrogen force field for proteins.[60] The *charmm2lammps.pl* tool from the LAMMPS distribution[53,61] was used to convert CHARMM topology and parameter files into LAMMPS input files for our simulations.

Nonbonded interactions within the atomistic models (small molecules and protein G) were computed using standard approaches. In particular, Lennard-Jones pair interactions were considered up to an atom-based cutoff distance of 12 Å; a switching function[60] was used to make both energies and forces go to zero smoothly between 11 and 12 Å. Lorentz-

Berthelot combining rules[43] were used (as in both the original CHARMM and AMBER force fields[54,62]). For the Coulombic interactions, a cutoff distance of $12\,\text{Å}$ was set for the real space part, while long range interactions were included using the particle-particle particle-mesh (PPPM) solver,[63] with a relative tolerance of 10^{-5}. Intramolecular 1-2 and 1-3 nonbonded interactions were neglected, while 1-4 interactions were treated according to the rules for the corresponding force field.[54,62]

It should be stressed that the CGenFF force field used for the amino acid side chain analogues is the "generalized" version of the CHARMM force field used for the protein (in fact, CGenFF stands for "CHARMM General Force Field"). Not only the two force fields are compatible, but they share most parameters, wherever appropriate. For example, Lennard-Jones constants and partial charges for the side chain analogs are largely the same as in the corresponding protein amino acids.

2.3   Mixed atomistic-CG interactions

Considering an atomistic site i and a CG ELBA water site j, the pair potential energy U_{ij} is:

$$U_{ij} = U_{ij}^{\mathrm{LJ}} + U_{ij}^{\mathrm{qp}} \tag{4}$$

with U_{ij}^{LJ} the Lennard-Jones term and U_{ij}^{qp} the charge-dipole term. For the Lennard-Jones term, we use the same shifted-force potential that models the water-water interactions (Equation 2), but with σ and ϵ now representing the mixed i-j (atom-water) interactions. Such cross terms are assigned with the standard Lorentz-Berthelot rules:[43] $\sigma = (\sigma_i + \sigma_j)/2$, $\epsilon = \sqrt{\epsilon_i \epsilon_j}$. These are the same rules used to assign Lennard-Jones cross terms within purely atomistic interactions. For the electrostatic potential between the atomistic (partial) charges and the CG water dipoles, we use a shifted-force charge-dipole potential:[35]

$$U_{ij}^{\mathrm{qp}} = \frac{q_i}{4\pi\varepsilon_0\varepsilon_r r^3}\left[1 - 3\left(\frac{r}{r_c}\right)^2 + 2\left(\frac{r}{r_c}\right)^3\right]\mathbf{p}_j \cdot \mathbf{r} \tag{5}$$

with q_i the atom partial charge, $\mathbf{p}_j$ the water dipole, ε_0 the vacuum permittivity, ε_r the relative permittivity, $\mathbf{r}$ the pair distance vector, r the magnitude of $\mathbf{r}$, and r_c the cutoff radius. Note that $\varepsilon_r = 1$, as for the electrostatic interactions within standard all-atom models.

3   Simulation protocols

3.1   General molecular dynamics details

Molecular dynamics simulations were run with the program LAMMPS,[53,61] version 16 Aug 2013, modified to include the calculation of solute-solvent potential energies required for the free energy calculations (which are detailed in Section 3.2 below). Complete command scripts and input files are available on the author's website.[64] In all simulations reported in this paper, the integration timestep was 2 fs. The temperature was controlled using a Langevin thermostat[65] with a collision frequency of $1\,\mathrm{ps}^{-1}$. The pressure was controlled isotropically using the barostat by Berendsen et al.[66] with a damping time of 1 ps and an isothermal compressibility of $4.6 \times 10^{-5}\,\mathrm{atm}^{-1}$. For the atomistic solutes, bonds involving hydrogen atoms were constrained using the SHAKE algorithm[67] with a relative tolerance of 10^{-6}. Further simulation details, specific to the different systems simulated, are given below in dedicated sections.

3.2   Alchemical free energy simulations

Each atomistic solute was inserted into an equilibrated box of 1000 ELBA water sites. Decoupling simulations were run by scaling the solute-water potential energy[68] V through the fourth-power function[69] $f(\lambda) = (1 - \lambda)^4$, where λ was varied from 0 (full coupling) to 1 (full decoupling, equivalent to the absence of solute-water interactions). We are aware that common practice involves using soft-core functions,[70] but these are not currently available in LAMMPS.[53,61] However, using the fourth-power function reported above was shown

elsewhere to be a valid approach, at least for computing hydration free energies of small molecules.[69] Each decoupling transformation was carried out in a single simulation lasting 187.5 ns, during which time λ was changed stepwise from 0 to 0.96 through 25 equally spaced values ($\lambda = 0, 0.04, 0.08, \ldots, 0.92, 0.96$). For each λ value, the system was thus simulated for 7.5 ns; the initial 0.5 ns were treated as equilibration and ignored, while the following 7 ns were used to sample the derivative of the solute-solvent potential energy $\partial V / \partial \lambda$. The value of $\partial V / \partial \lambda$ for $\lambda = 1$ was obtained by linear extrapolation.[69] Numerical integration of $\partial V / \partial \lambda$ curves was carried out using the trapezoidal rule; the resulting integral corresponds to the negative of the hydration free energy. Each decoupling simulation was repeated three times, with initial velocities assigned using different random seeds. Since we are not aware of any previous use of LAMMPS to compute hydration free energies by alchemical transformations, we performed a preliminary validation test by reproducing a $\partial V / \partial \lambda$ curve from the literature[69] (Figure S1 in the Supporting Information).

3.3   Hydrated protein simulations

Protein G was solvated with 7598 water molecules, which were initially modeled with the CHARMM[60] version of the TIP3P potential.[71] To neutralize the protein net charge, 4 sodium ions (modeled with standard CHARMM parameters) were added to the system. The operations above were performed with the *charmm2lammps.pl* tool from the LAMMPS distribution.[53,61] To obtain a dual-resolution system, each atomistic water molecule was replaced by an ELBA water site; the relevant files were reprocessed accordingly with an in-house Python script.

Two simulations were performed: 1) a "control" fully atomistic system consisting of the CHARMM protein model solvated in TIP3P water, and 2) a corresponding dual-resolution system consisting of the same CHARMM protein model solvated in ELBA water. Each simulation involved an initial short energy minimization performed with the conjugate gradient algorithm. Each system was then equilibrated through a 1 ns molecular dynamics run,

during which time the protein atoms were restrained to their original coordinates by means of harmonic springs. In particular, the spring rigidity constant was set to $10\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$ during the first $0.5\,\mathrm{ns}$, and was subsequently reduced to $1\,\mathrm{kcal\,mol^{-1}\,\AA^{-2}}$ during the following $0.5\,\mathrm{ns}$. For each system, an unrestrained simulation was then run for $200\,\mathrm{ns}$.

Root mean square deviation (RMSD) and hydrogen bonds were analyzed using respectively the RMSD Trajectory Tool and the Hbonds plugin, both part of the VMD software package.[59] For each simulations, the numerical results presented have been averaged over $191\,\mathrm{ns}$, as the first $10\,\mathrm{ns}$ were considered as equilibration time and discarded. Specifically, each property was averaged over 9550 samples taken every $20\,\mathrm{ps}$; these data were also used to calculate each property's reported standard deviation.

3.4   Parallel tempering simulations

The parallel tempering (or replica exchange) method[72] was applied to investigate the folding free energy of the C-terminal β hairpin of protein G (PDB code: 2GB1) and Trp-cage (PDB code: 1L2Y). The β hairpin and Trp-cage protein structures were solvated with respectively 1901 and 1885 water molecules. Each system was initially equilibrated for $2\,\mathrm{ns}$ at $300\,\mathrm{K}$ and $1\,\mathrm{atm}$; in particular, during the first $1\,\mathrm{ns}$, the protein atoms were restrained to their original positions by applying harmonic springs in the same way as done during the equilibration of protein G (Section 3.3). The systems were then run for $100\,\mathrm{ps}$ in the NVT ensemble, and the final configurations were used as the starting points for the parallel tempering simulations. For both proteins, 60 replicas were simulated in parallel with temperatures spanning the range from 270 to $655\,\mathrm{K}$.[73,74] In particular, each replica was simulated for $3.5\,\mathrm{ns}$. During the first $0.1\,\mathrm{ns}$, exchanges were not attempted. During the remaining $3.4\,\mathrm{ns}$, parallel tempering was performed with exchanges attempted every $0.4\,\mathrm{ps}$, and configurations saved every $0.1\,\mathrm{ps}$. The last $3\,\mathrm{ns}$ of every replica were used for data collection. For the β hairpin, the two reaction coordinates are the number of the native backbone-backbone hydrogen bonds excluding the two near the turn, and the radius of gyration of the side chain atoms of the four hydrophobic

residues (Trp43, Tyr45, Phe52 and Val54).[75] For the Trp-cage, the fraction of native contacts (Q) and the radius of gyration of the α-carbon atoms (Rg) were chosen as the reaction coordinates; a native contact was defined when the distance of a pair of α-carbon atoms from nonadjacent residues is less than 6.5 Å.[76] Free energy landscapes were obtained from histogram analysis.[77]

4 Results

4.1 Small molecule hydration free energies

The two data sets of small molecule hydration free energies obtained in this work, together with a data set from fully atomistic simulations[78] and an experimental data set,[79–81] are reported in 2. The underlying numerical values can be found in Table S1 in the Supporting Information, together with the potential energy derivative curves from the corresponding alchemical free energy calculations (Figures S2 and S3). To describe the predictions from simulation (either all-atom or dual-resolution) reported in 2, their accuracy is assessed in terms of how close they are to the corresponding experimental measurements. In particular, for methane (Ala), p-cresol (Tyr), acetamide (Asn), propionamide (Gln) and 3-methylindole (Trp), it can be seen that our dual-resolution predictions (both CGenFF-ELBA and GAFF-ELBA) are less accurate than the fully atomistic ones (GAFF-TIP3P). However, for butane (Ile), isobutane (Leu), methanol (Ser) and ethanol (Thr) both our CGenFF-ELBA and GAFF-ELBA results are more accurate than the corresponding data from the GAFF-TIP3P fully atomistic systems. For methanethiol (Cys) and 4-methylimidazole (Hid), the CGenFF-ELBA predictions are more accurate than the fully atomistic GAFF-TIP3P data, which in turn are more accurate than the GAFF-ELBA data. For propane (Val), the CGenFF-ELBA result matches the experimental one (within uncertainty), while an almost equal overestimated prediction was obtained from the GAFF-ELBA and GAFF-TIP3P simulations. For ethylmethylsulfide (Met), while none of the simulation prediction is particularly accurate, the
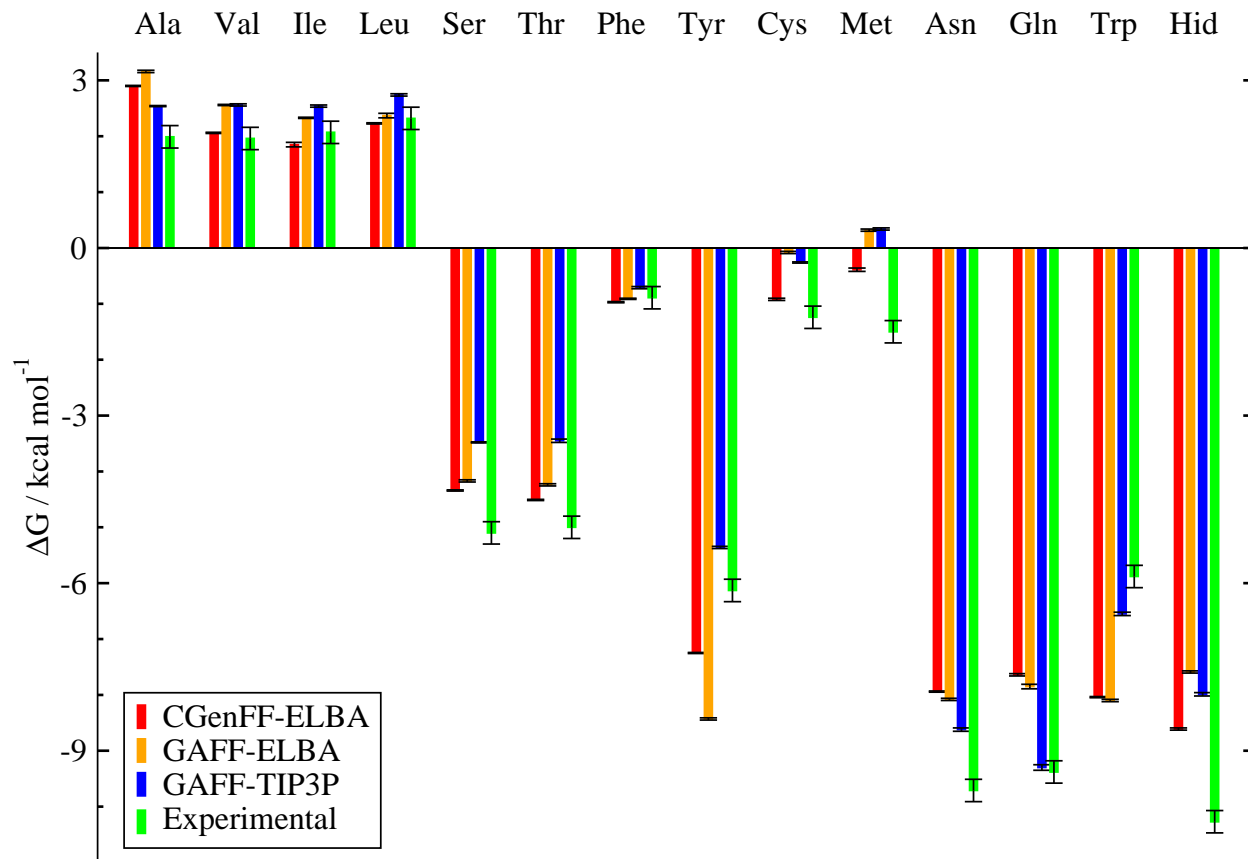
11

Figure 2: Hydration free energies of amino acid sidechain analogs. "CGenFF-ELBA" and "GAFF-ELBA" denote dual-resolution simulation data from this work. "GAFF-TIP3P" denotes fully atomistic simulation data from Mobley et al.[78] Experimental data are from ref.[79–81]

CGenFF-ELBA value correctly reproduces the negative sign of the experimental hydration free energy, as opposed to both the GAFF-ELBA and the GAFF-TIP3P data, which wrongly predict (equal) positive values. The only solute for which all three simulation predictions agree with experiment (within uncertainty) is toluene (Phe).

Considering those molecules containing hydroxyl (OH) groups, that is, methanol (Ser), ethanol (Thr) and p-cresol (Tyr), it can be noted that the hydration free energies from the dual-resolution systems are more negative than those from the all-atom calculations. For those molecules containing carbonyl (C=O) groups, that is, acetamide (Asn) and propionamide (Gln), it can be seen that the hydration free energies from the dual-resolution systems are instead less negative than those from both the all-atom calculations and the experimental measurements.

To quantify and compare the overall accuracy of the three sets of simulation results, we paired each of them, in turn, with the experimental data set, and for each simulation-experimental pair, we computed the mean error (ME), mean absolute error (MAE), root mean square error (RMSE) and correlation coefficient (R^2). The results obtained are reported in 1. In general, more accurate predictions correspond to smaller absolute ME, smaller MAE

Table 1: Statistics for calculated versus experimental free energies (values in kcal/mol)[a]

	$\Delta G_{\text{CGenFF}-\text{ELBA}}$ this work	$\Delta G_{\text{GAFF}-\text{ELBA}}$ this work	$\Delta G_{\text{GAFF}-\text{TIP3P}}$ Mobley et al.[78]
ME	−0.37	−0.58	−0.84
MAE	0.90	1.22	0.94
RMSE	1.13	1.47	1.13
R^2	0.94	0.91	0.97

[a] Mean error ME, mean absolute error MAE, root mean square error RMSE and correlation coefficient R^2. Each estimator was obtained by pairing one of the three simulation data sets considered with the experimental data set from ref.[79–81]

and RMSE, and R^2 closer to 1 (indicating stronger linear correlation). 1 shows that the absolute ME is smaller (hence indicating higher accuracy) for both the dual-resolution data than for the fully atomistic results. The MAE is smallest for the CGenFF-ELBA predictions (indicating highest accuracy), while it is largest for the GAFF-ELBA data. The RMSE is

equal for the CGenFF-ELBA and the fully atomistic predictions, while it is somewhat larger (indicating lower accuracy) for the GAFF-ELBA data. The R^2 values assign the strongest linear correlation with the experimental data to the fully atomistic results (which in this respect are thus most accurate), followed by the CGenFF-ELBA predictions, and lastly by the GAFF-ELBA data. Overall, these data indicate that the dual-resolution CGenFF-ELBA results are at least as accurate as the fully atomistic GAFF-TIP3P data, while the dual-resolution GAFF-ELBA predictions are the least accurate (by an arguably small margin).

4.2   Hydration free energy of ELBA water

As an additional finding, we also computed the hydration free energy of the ELBA water model, with the same method used for the small molecule simulations (Section 3.2). In this case, a system comprising 1001 ELBA water sites was simulated, with one of the sites being designated as solute for the purpose of decoupling the solute-solvent interactions. The result obtained is reported in 2, together with literature values from experiment and from simulation of the most widely used atomistic models. The table also reports the relative error in the simulation results with respect to the experimental value (obviously, the smaller this error, the more accurate the model). It can be seen from 2 that the ELBA result

Table 2: Hydration free energy of water

	ΔG / kcal mol^{-1}	Relative error ^{a}	Reference
Experimental	-6.33	-	Abraham et al.[79]
ELBA	-6.50	2.7 %	This work
SPC	-6.16	2.7 %	Shirts and Pande[82]
SPC/E	-7.05	11.4 %	Shirts and Pande[82]
TIP3P	-6.10	3.6 %	Shirts and Pande[82]
TIP3P-Ew	-5.28	16.6 %	Huggins[83]
TIP4P	-6.11	3.5 %	Shirts and Pande[82]
TIP4P-Ew	-6.98	10.3 %	Shirts and Pande[82]
TIP4P/2005	-6.31	0.3 %	Huggins[83]

^{a} The relative error is defined as $|(\Delta G_{\mathrm{sim}} - \Delta G_{\mathrm{exp}})/\Delta G_{\mathrm{exp}}|$, with ΔG_{sim} and ΔG_{exp} the simulation and experimental values, respectively.

reproduces the experimental measurement at a level of accuracy as high as, or even slightly

higher than, the atomistic models SPC,[84] TIP3P,[71] and TIP4P.[85] ELBA proves markedly more accurate than SPC/E,[86] TIP3P-Ew,[87] and TIP4P-Ew.[88] The only atomistic model that is more accurate than ELBA is TIP4P/2005,[89] which almost matches the experimental value.

4.3 Protein G simulations

The root mean square deviation (RMSD) of the protein backbone atoms from the experimental crystal structure was calculated for both the dual-resolution system and for the control all-atom system over the course of each simulation. The curves obtained are displayed superimposed in 3. It can be seen that both RMSD curves oscillate around roughly similar average
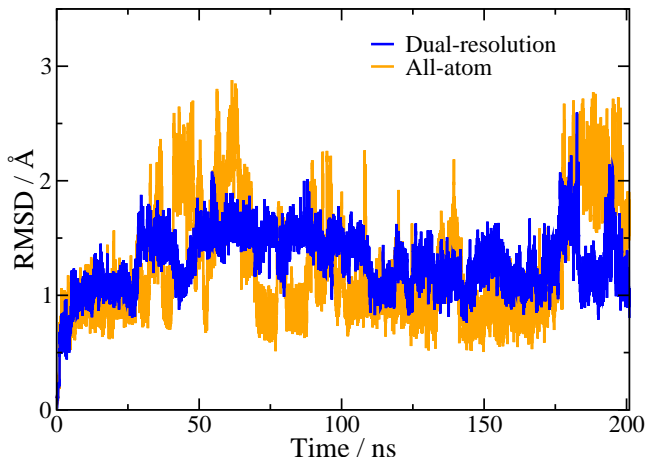
Figure 3: Root-mean-squared deviation (RMSD) of the protein backbone with respect to the crystal structure. "Dual-resolution" curve: CHARMM protein in ELBA water. "All-atom" curve: CHARMM protein in TIP3P water.

values, although it is clear that the spread of the all-atom values is somewhat larger than that of the dual-resolution data. Specifically, for the all-atom system, the protein backbone RMSD has an average of $1.27\,\text{Å}$ and a standard deviation of $0.50\,\text{Å}$. For the dual-resolution system, the protein backbone RMSD has an average of $1.31\,\text{Å}$ and a standard deviation of $0.23\,\text{Å}$. The smaller standard deviation for the dual-resolution system indicates lower flexibility with respect to the all-atom system. Regarding the average, the values for both curves are similar and low, indicating a good degree of preservation of the initial (crystal) structure

in both simulations. This can be further confirmed qualitatively by visualizing a superposition of the time-averaged structures from simulation onto the initial structure from X-ray data, as shown in 4. It can be seen that the overall folding geometry and secondary structure
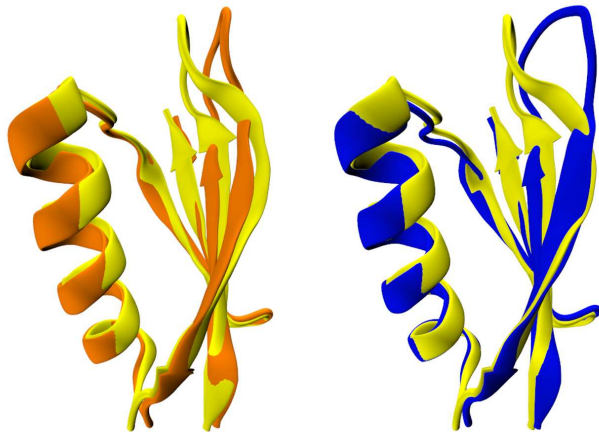

Figure 4: Superposition of time-averaged structures from 201 ns simulations onto the initial X-ray crystal structure. In both panels, the initial structure is colored yellow. In the left panel, the time-averaged structure (in orange) is from the fully atomistic simulation. In the right panel, the time-averaged structure (in blue) is from the dual-resolution simulation. Images prepared and rendered with the VMD program.[59]

elements are well-preserved, in both systems. The only significant discrepancy involves the loop region that can be seen at the top right of both panels in 4; such loop corresponds to residues 9 to 12 in the crystal structure, hence we will refer to it as the "9-12 loop". A visual inspection of the entire trajectories highlights how the 9-12 loop is rather flexible, and it is especially so for the all-atom system; this can be seen qualitatively through comparison of movies of the backbone dynamics that can be found in the Supporting Information. Further quantitative data on the backbone behavior can be seen in 5, which reports the root-mean-squared fluctuations (RMSF) of the backbone C_α atoms, as a function of residue sequence number, for each of the two simulated systems. It can be noticed that the two sets of RMSF data are mostly similar to each other, especially for the regions corresponding to the secondary structure elements (α-helix and β-strands). However, it is clear that there is a substantial difference regarding the 9-12 loop region. In particular, the RMSF values for the 9-12 loop are markedly higher in the all-atom simulation with respect to the dual-resolution
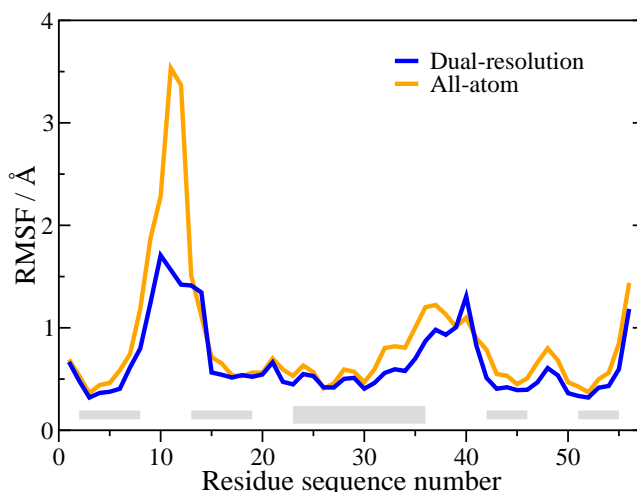
Figure 5: Root-mean-squared fluctuations (RMSF) of the backbone C_α atoms, with respect to the crystal structure, as a function of residue sequence number. "Dual-resolution" curve: CHARMM protein in ELBA water. "All-atom" curve: CHARMM protein in TIP3P water. The horizontal gray segments at the bottom of the diagram indicate the groups of residues forming the α-helix (thicker segment) and the four β-strands (thinner segments) in the crystal structure. Specifically, the α-helix is formed by residues 23-36, and the four β-strands are formed respectively by residues 2-8, 13-19, 42-46, and 51-55.

counterpart; this confirms the previously reported qualitative observation of a more flexible 9-12 loop in the all-atom system compared to the dual-resolution system.

The average intramolecular potential energy was also computed; the results are reported in Figure S4 in the Supporting Information, in terms of the separate contributions from bond stretching, angle bending, dihedral angles, and improper dihedral angles. It can be noticed that the dual-resolution intramolecular energies are in close agreement with the corresponding data for the control all-atom system. This also indicates that the observed structural differences in the protein between the dual-resolution and all-atom simulations are not inconsistent with the atomistic protein force field.

Regarding the intraprotein nonbonded energy, we calculated a value of $-2237\pm19\,\mathrm{kcal\,mol^{-1}}$ for the dual-resolution system, and a value of $-2284\pm23\,\mathrm{kcal\,mol^{-1}}$ for the control all-atom system; the difference between the two mean values obtained is rather small, and of the order of the reported standard deviations.

A further analysis was conducted on the effect of the two different hydration models

considered in terms of hydrogen (H) bonds between protein atoms. In particular, instantaneous values for the total number of intraprotein H bonds were evaluated from individual frames, and final results were computed as averages over time. The total average number of H bonds thus obtained, for each of the two simulations, was divided into the subsets of H bonds within the backbone ("backbone-backbone"), H bonds between backbone and sidechains ("backbone-sidechain") and H bonds between sidechains ("sidechain-sidechain"); corresponding results are reported in 6. Regarding the backbone-backbone H bonds, it can
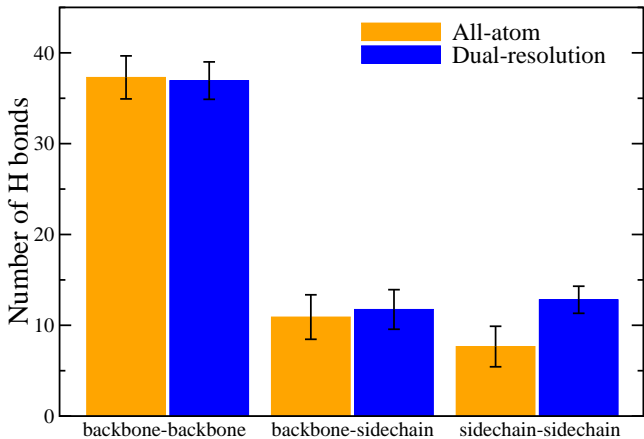
Figure 6: Intraprotein number of hydrogen bonds, averaged over the simulation time, and divided into the subsets of hydrogen bonds within the backbone ("backbone-backbone"), between backbone and sidechains ("backbone-sidechain") and between sidechains ("sidechain-sidechain"). "All-atom" data: CHARMM protein in TIP3P water. "Dual-resolution" data: CHARMM protein in ELBA water.

be seen that their average number is essentially the same in both simulations. The average number of backbone-sidechain H bonds is also very similar for the two systems. However, 6 highlights the presence of a significantly higher average number of sidechain-sidechain H bonds in the dual-resolution simulation with respect to the fully atomistic counterpart. In fact, sidechain-sidechain H bonds in the dual-resolution system are expected to be more favored, because the coarse-grained (ELBA) water, which lacks explicit H bond donor or acceptor sites, cannot compete with the sidechains involved in sidechain-sidechain H bonds to form alternative water-sidechain H bonds. In the all-atom system, the atomistic (TIP3P) water does instead form water-sidechain H bonds, which can replace sidechain-sidechain H

bonds.

A consequence of the observed higher average number of H bonds is the expectation of an overall greater stability (reduced flexibility) in the sidechains of the dual-resolution protein systems compared to those in the all-atom system. To investigate this effect, the behavior of the sidechains was characterized by computing their RMSD with respect to the heavy (non-hydrogen) atoms of the crystal structure; these were also the atoms used in the calculation to superimpose each structure onto the reference (crystal) structure. The data obtained, as a function of simulation time, are reported in 7. It can be seen that both curves oscillate
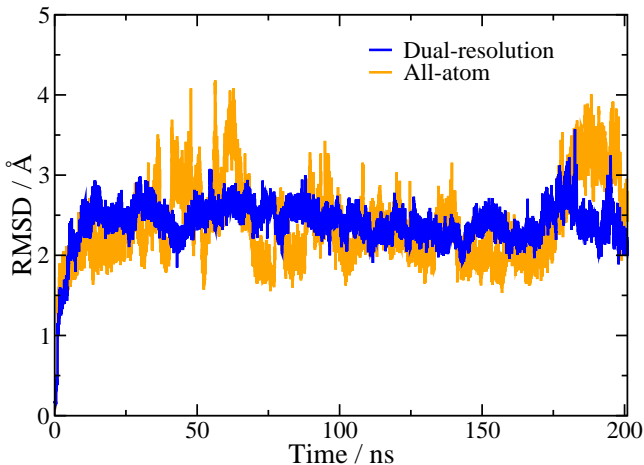

Figure 7: Root-mean-squared deviation (RMSD) of the protein sidechains (excluding hydrogens) with respect to the crystal structure for protein G. "Dual-resolution" curve: CHARMM protein in ELBA water. "All-atom" curve: CHARMM protein in TIP3P water.

around similar average values; however, the all-atom RMSD curve is spread over a larger range of distances. Quantitatively, for the all-atom system, the average protein sidechain RMSD is found to be $2.40 \, \text{Å}$, while the standard deviation is $0.48 \, \text{Å}$; for the dual-resolution system, the average is $2.43 \, \text{Å}$ and the standard deviation is $0.19 \, \text{Å}$. Thus, while the average values are almost identical, the standard deviation in the dual-resolution system is less than half that of the all-atom system, indicating higher stability.

4.4 Parallel tempering simulations

The free energy landscapes obtained from the parallel tempering simulations of the the β hairpin are shown in 8. In particular, Figure 8a refers to the fully atomistic run (CHARMM protein with TIP3P water), and Figure 8b represents the results from the dual-resolution system (CHARMM protein with ELBA water). It can be noticed that the two diagrams show
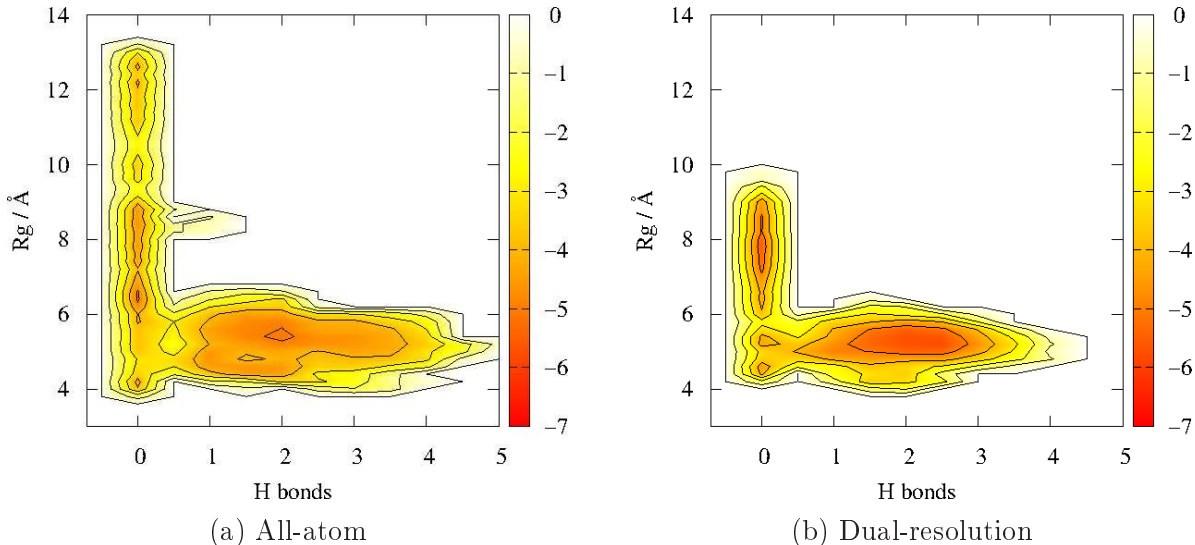

(a) All-atom   (b) Dual-resolution

Figure 8: Free energy landscapes for β hairpin *vs* the two reaction coordinates, that is, number of H bonds and radius of gyration (Rg). The free energy is in units of k_BT, and contours are spaced with $1\,k_B$T intervals. Panel (a): All-atom system. Panel (b): Dual-resolution system.

qualitatively similar "L"-shaped landscapes. To compare the two simulations quantitatively, we can compute the folding free energy. In particular, we consider folded and unfolded states when the number of hydrogen bonds is respectively greater or less than 1;[90] we can then calculate the folding free energy as $\Delta G = -RT \log(P_f/P_u)$ where P_f and P_u denote respectively the probabilities of the folded and unfolded states. The folding free energies obtained are respectively $-0.43\,$kcal/mol and $-0.39\,$kcal/mol for the all-atom and dual-resolution systems, showing reasonable agreement between the two approaches.

Regarding Trp-cage, the free energy landscapes obtained from the parallel tempering simulations are shown in 9. In particular, Figure 9a refers to the fully atomistic run (CHARMM protein with TIP3P water), and Figure 9b represents the results from the dual-resolution

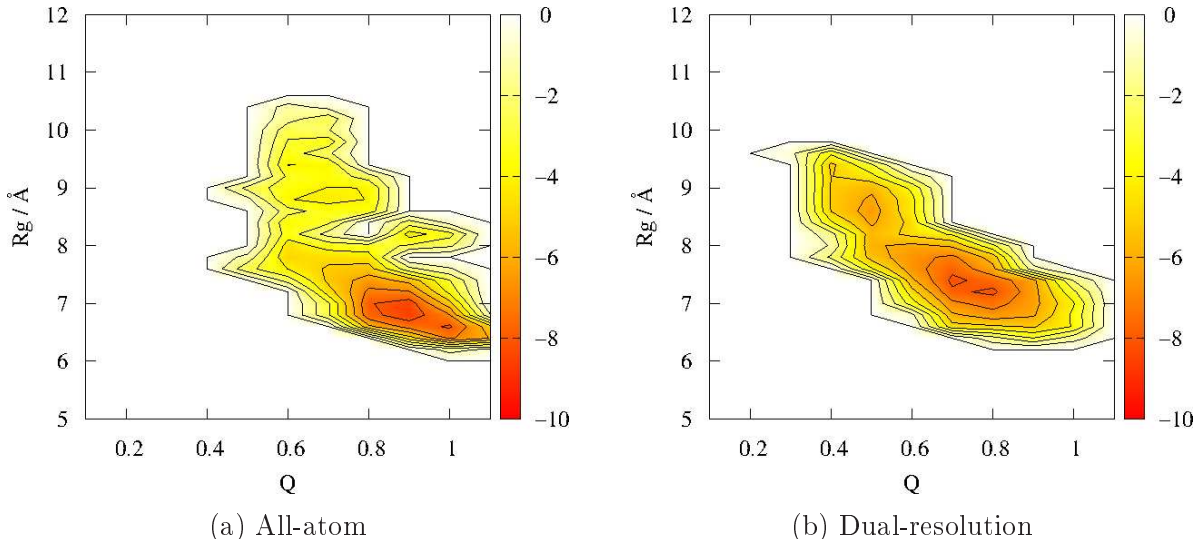system (CHARMM protein with ELBA water). It can be seen that the landscapes share

(a) All-atom   (b) Dual-resolution

Figure 9: Free energy landscapes for Trp-cage *vs* the two reaction coordinates, that is, fraction of native contact (Q) and the radius of gyration of the α-carbon atoms (Rg). The free energy is in units of k_BT, and contours are spaced with $1\,k_BT$ intervals. Panel (a): All-atom system. Panel (b): Dual-resolution system.

qualitatively similar features, especially with respect to the regions of lower free energy, whereas some differences are evident in the upper part of the landscapes, corresponding to regions of higher free energy. With respect to the folding free energy, considering folded states when $Q > 0.6$ and $Rg < 7.8$, we obtain $-1.80\,$kcal/mol for the all-atom system and $-0.66\,$kcal/mol for the dual-resolution system; in this case, there is a factor of three difference in the magnitude of the values. Remarkably, the dual-resolution result is closer to the experimental values of $-0.7\,$kcal/mol[91] and $-0.76\,$kcal/mol.[92]

4.5   Computational efficiency

We estimated the computational efficiency of the dual-resolution methodology in terms of the computational speedup over a standard fully atomistic approach. Such speedup was quantified by conducting comparative tests on representative systems chosen from those simulated in this work. The first system involved a single small molecule (toluene) solvated by 1000 waters, while the second system involved protein G solvated by 7598 waters. For

each of the two systems, we considered a fully atomistic simulation (with TIP3P water) and a corresponding dual-resolution one (with ELBA water). The molecular dynamics details were as reported in Section 3.1. Each simulation was run in serial (on a single processor core) as well as in parallel (using MPI on 4, 8, and 12 processor cores). Full specific details and results for these efficiency tests are reported in Figure S5 in the Supporting Information. In summary, for the hydrated toluene system, the dual-resolution simulations proved ≈ 6 times more efficient than the all-atom ones. For the hydrated protein G system, the dual-resolution speedup factor over all-atom was ≈ 3. The difference between the speedup factors in the two systems can be mostly ascribed to the differing relative content of water. It is intuitively expected that the dual-resolution efficiency will be maximized when the relative water content is highest, as this corresponds to the largest portion of the system being represented at the coarse-grained level of resolution. In fact, in the tests reported here, the water mass percentage weights for the toluene and the protein G systems were respectively $\approx 99 \, \mathrm{wt} \, \%$ and $\approx 95 \, \mathrm{wt} \, \%$. In terms of numbers of atoms, the relative water contents for the toluene and the protein G systems were respectively $\approx 98 \, \%$ and $\approx 96 \, \%$.

5   Discussion

We have presented a new dual-resolution hydration approach whereby the ELBA coarse-grained model for water is used in combination with all-atom molecular models. A unique feature of our method is that no extra scaling factors, healing regions, or virtual sites are required to mix the two levels of resolution.

To validate the methodology, we computed hydration free energies from dual-resolution systems where the ELBA water was used to solvate a range of small molecules, each described with two of the most common all-atom force fields (Section 4.1). Our simplified hydration model yielded predictions that are overall as accurate as those from fully atomistic simulations (2 and 1). This result is striking, because the hydration of atomistic solutes would

normally be expected to be described more accurately by an atomistic water model than by a coarse-grained one. Similarly notable was the finding that the hydration free energy of ELBA water in itself reproduces the experimental value for real water more accurately than most atomistic water models (2). Overall, these remarkable results are in line with recent work on pure water systems showing ELBA to be as accurate as the best atomistic models in reproducing fundamental properties such as density, diffusion, surface tension, vapor-liquid equilibria, and even the critical point.[37] Regarding a possible explanation for the comparatively high accuracy of ELBA, we believe that an important factor is the magnitude of its permanent dipole moment $(2.6\,\mathrm{D})$, which is significantly closer to that of real liquid water $(2.95\,\mathrm{D}^{[93]})$ compared to those of standard atomistic models (2.18 D for TIP4P,[94,95] 2.27 D for SPC,[96] 2.305 D for TIP4P/2005,[89] 2.35 D for SPC/E[94] and TIP3P[95]).

Dual-resolution simulations were also conducted with ELBA water used to hydrate a typical protein modeled with a standard atomistic force field (Section 4.3). The results obtained were largely consistent with those from a control all-atom simulation in terms of preservation of the experimental structure and energetics. However, the protein in ELBA water was found to be somewhat less flexible than in atomistic water, and was also characterized by an increased average number of hydrogen bonds between sidechains. It is important to note that our protein G simulations cannot demonstrate the overall stability of the force field, for which much longer simulations are required.[97] A further test involved running parallel tempering simulations to calculate folding free energy landscapes of two small α helical and β hairpin structures (Section 4.4). For the β hairpin structure, satisfactory agreement was obtained between the all-atom and dual-resolution results. However, for the α helical structure, the magnitude of the folding free energy obtained from the dual-resolution system was over three times smaller than that from the all-atom system; interestingly, the dual-resolution result was found to be closer to the experimental value. In general, some disagreement in the behavior of protein systems is expected, due to the differences observed between the all-atom and dual-resolution results for the hydration free energies of the amino

acid sidechain analogs; specifically, while the overall accuracy of the simulation approaches is similar (1), significant variations can be observed for most sidechain analogs in terms of individual values of the hydration free energy from the different models (2).

In terms of computational efficiency, we obtained speedup factors for our dual-resolution simulations over all-atom counterparts of ≈ 3 and ≈ 6, depending on the specific system (Section 4.5). While these numbers are already significant, especially considering the comparatively high accuracy of our method, substantial improvements are possible. In particular, since systems of pure ELBA water can be simulated with a 10 fs timestep,[37] the implementation of a multistep approach[98] should markedly increase the dual-resolution speedup. Specifically, while solute-solute and water-solute interactions would be evaluated with a standard 2 fs timestep (as done in this work for all interactions), the water-water interactions (which typically dominate the computational cost) would be evaluated with the 5 times larger timestep of 10 fs. Any current speedup factor would thus increase by up to 5 times; for example, the currently observed speedup factors of ≈ 3 and ≈ 6 would increase up to ≈ 15 and ≈ 30, respectively.

Regarding general limitations of the methodology presented, it is clear that any hydrogen bonding between an atomistic solute and ELBA water is inevitably described at an approximate level. In fact, while the electrostatic interactions between ELBA's dipole and atomistic donors and acceptors are expected to capture some overall features of hydrogen bonding, it is clear that the absence of explicit donor and acceptor sites in ELBA prevents local effects to be represented accurately. An example of the consequences of this limitation was indeed observed in the analysis of the sidechain-sidechain hydrogen bonds. However, it is also interesting and important to stress that the lack of explicit hydrogen bonding sites in ELBA did not prevent the comparatively accurate prediction of the hydration free energy of the amino acid sidechain analogs, and of the ELBA water itself. From a technical standpoint, we should note that the ELBA model and related dual-resolution scheme are currently available only in the LAMMPS simulation program.[53,61] Most other mainstream packages,

such as GROMACS,[99] AMBER,[51] NAMD,[100] or GROMOS,[101] lack the point dipole potential, and related rotational integrator, that ELBA requires; these features could of course be implemented, but major modifications to data structures and core routines would be necessary.

6   Conclusions

We described a novel dual-resolution scheme that couples the ELBA coarse-grained water model with conventional fully atomistic solutes. The approach presented is uniquely simple, since the coarse-grained water interacts directly with the atomistic molecules without the need for extra parameters. The methodology is capable of reproducing the hydration free energy of a diverse range of small organic molecules, and of the coarse-grained water itself, at a level of accuracy rivaling that of standard fully atomistic calculations. The approach presented was also applied to the simulation of a hydrated protein system; while the average structure and energetics were consistent with corresponding all-atom calculations, some differences were noticed regarding flexibility and hydrogen bonding between sidechains. Computationally, our hybrid simulations proved up to six times more efficient than standard fully atomistic counterparts, and future work involving the implementation of multistep integration methods could increase this speedup further.

Supporting Information Available

Curves for the derivative of the solute-solvent potential energy as a function of λ, numerical values of data displayed in 2, individual intramolecular potential energy values for protein G, computational efficiency tests, movies of the protein backbone dynamics from all-atom and dual-resolution simulations. This material is available free of charge via the Internet at `http://pubs.acs.org/`.

References

(1) Chaplin, M. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 861–866.

(2) Ball, P. *Chem. Rev.* **2008**, *108*, 74–108.

(3) Guillot, B. *J. Mol. Liq.* **2002**, *101*, 219–260.

(4) Dill, K. A.; Truskett, T. M.; Vlachy, V.; Hribar-Lee, B. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173–199.

(5) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.

(6) Vega, C.; Abascal, J. L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 19663–19688.

(7) Allison, J. R.; Boguslawski, K.; Fraternali, F.; van Gunsteren, W. F. *J. Phys. Chem. B* **2011**, *115*, 4547–4557.

(8) Darré, L.; Machado, M. R.; Pantano, S. *WIREs: Comput. Mol. Sci.* **2012**, *2*, 921–930.

(9) Hadley, K. R.; McCabe, C. *Mol. Simul.* **2012**, *38*, 671–681.

(10) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198.

(11) Sherwood, P.; Brooks, B. R.; Sansom, M. S. P. *Curr. Opin. Struct. Biol.* **2008**, *18*, 630–640.

(12) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423–12430.

(13) Orsi, M.; Sanderson, W.; Essex, J. W. In *Molecular Interactions - Bringing Chemistry to Life*; Hicks, M. G., Kettner, C., Eds.; Beilstein-Institut: Frankfurt, Germany, 2007; pp 85–205.

(14) Orsi, M.; Haubertin, D. Y.; Sanderson, W. E.; Essex, J. W. *J. Phys. Chem. B* **2008**, *112*, 802–815.

(15) Orsi, M.; Michel, J.; Essex, J. W. *J. Phys.: Condens. Matter* **2010**, *22*, 155106.

(16) Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. *WIREs: Comput. Mol. Sci.* **2013**, *4*, 225–248.

(17) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896–10913.

(18) Michel, J.; Orsi, M.; Essex, J. W. *J. Phys. Chem. B* **2008**, *112*, 657–660.

(19) Orsi, M.; Sanderson, W. E.; Essex, J. W. *J. Phys. Chem. B* **2009**, *113*, 12019–12029.

(20) Orsi, M.; Essex, J. W. *Soft Matter* **2010**, *6*, 3797–3808.

(21) Orsi, M.; Noro, M. G.; Essex, J. W. *J. R. Soc. Interface* **2011**, *8*, 826–841.

(22) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. *Eur. Biophys. J.* **2012**, *41*, 647–661.

(23) Riniker, S.; Eichenberger, A. P.; van Gunsteren, W. F. *J. Phys. Chem. B* **2012**, *116*, 8873–8879.

(24) Wassenaar, T. A.; Ingólfsson, H. I.; Prieß, M.; Marrink, S. J.; Schäfer, L. V. *J. Phys. Chem. B* **2013**, *117*, 3516–3530.

(25) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045–15048.

(26) Masella, M.; Borgis, D.; Cuniasse, P. *J. Comput. Chem.* **2008**, *29*, 1707–1724.

(27) Masella, M.; Borgis, D.; Cuniasse, P. *J. Comput. Chem.* **2011**, *32*, 2664–2678.

(28) Han, W.; Wan, C.-K.; Jiang, F.; Wu, Y.-D. *J. Chem. Theory Comput.* **2010**, *6*, 3373–3389.

(29) Han, W.; Schulten, K. *J. Chem. Theory Comput.* **2012**, *8*, 4413–4424.

(30) Rzepiela, A. J.; Louhivuori, M.; Peter, C.; Marrink, S. J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437–10448.

(31) Di Pasquale, N.; Marchisio, D.; Carbone, P. *J. Chem. Phys.* **2012**, *137*, 164111.

(32) Nagarajan, A.; Junghans, C.; Matysiak, S. *J. Chem. Theory Comput.* **2013**, *9*, 5168–5175.

(33) Gonzalez, H. C.; Darré, L.; Pantano, S. *J. Phys. Chem. B* **2013**, *117*, 14438–14448.

(34) Bevc, S.; Junghans, C.; Kremer, K.; Praprotnik, M. *New J. Phys.* **2013**, *15*, 105007.

(35) Orsi, M.; Essex, J. W. *PLoS One* **2011**, *6*, e28637.

(36) Orsi, M.; Essex, J. W. *Faraday Discuss.* **2013**, *161*, 249–272.

(37) Orsi, M. *Mol. Phys.* **2014**, *112*, 1566–1576.

(38) Warshel, A. *J. Phys. Chem.* **1979**, *83*, 1640–1652.

(39) Neumann, M. *Mol. Phys.* **1983**, *50*, 841–858.

(40) de Leeuw, S. W.; Perram, J. W.; Smith, E. R. *Annu. Rev. Phys. Chem.* **1986**, *37*, 245–270.

(41) Bartke, J.; Hentschke, R. *Phys. Rev. E* **2007**, *75*, 061503.

(42) Johnson, L. E.; Barnes, R.; Draxler, T. W.; Eichinger, B. E.; Robinson, B. H. *J. Phys. Chem. B* **2010**, *114*, 8431–8440.

(43) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 1st ed.; Oxford Science Publications: Oxford, U.K., 1987.

(44) Stockmayer, W. H. *J. Chem. Phys.* **1941**, *9*, 398–402.

(45) Farrell, J. D.; Lines, C.; Shepherd, J. J.; Chakrabarti, D.; Miller, M. A.; Wales, D. J. *Soft Matter* **2013**, *9*, 5407–5416.

(46) van Leeuwen, M. E. *Fluid Phase Equilib.* **1994**, *99*, 1–18.

(47) Rapaport, D. C. *The Art of Molecular Dynamics Simulation*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 2004.

(48) Adams, D. J.; Adams, E. M.; Hills, G. J. *Mol. Phys.* **1979**, *38*, 387–400.

(49) Stoddard, S. D.; Ford, J. *Phys. Rev. A* **1973**, *8*, 1504–1512.

(50) Price, S. L.; Stone, A. J.; Alderton, M. *Mol. Phys.* **1984**, *52*, 987–1001.

(51) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *WIREs: Comput. Mol. Sci.* **2013**, *3*, 198–210.

(52) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.

(53) http://lammps.sandia.gov.

(54) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(55) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graphics Modell.* **2006**, *25*, 247–260.

(56) Jakalian, A.; Bush, B. L.; Jack, D. B. *J. Comput. Chem.* **2000**, *21*, 132–146.

(57) Achari, A.; Hale, S. P.; Howard, A. J.; Clore, G. M.; Gronenborn, A. M.; Hardman, K. D.; Whitlow, M. *Biochemistry* **1992**, *31*, 10449–10457.

(58) Zacharias, M. *Proteins* **2013**, *81*, 81–92.

(59) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

(60) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack Jr., R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W., III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(61) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.

(62) Mackerell, A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

(63) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*, 1st ed.; CRC Press: New York, 1989.

(64) Orsi Group. `http://www.orsi.sems.qmul.ac.uk` (accessed June 10, 2014).

(65) Schneider, T.; Stoll, E. *Phys. Rev. B* **1978**, *17*, 1302–1322.

(66) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(67) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.

(68) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740–5761.

(69) Steinbrecher, T.; Mobley, D. L.; Case, D. A. *J. Chem. Phys.* **2007**, *127*, 214108.

(70) Michel, J.; Essex, J. W. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639–658.

(71) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(72) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(73) Zhou, R. *J. Mol. Graphics Modell.* **2004**, *22*, 451–63.

(74) Zhou, R. In *Protein Folding Protocols*; Bai, Y., Nussinov, R., Eds.; Methods in Molecular Biology; Humana Press: New York, 2006; Vol. 350; pp 205–223.

(75) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068–9073.

(76) Zhou, R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13280–13285.

(77) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.

(78) Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009**, *5*, 350–358.

(79) Abraham, M. H.; Whiting, G. S.; Fuchs, R.; Chambers, E. J. *J. Chem. Soc. Perkin Trans. 2* **1990**, 291–300.

(80) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385–16398.

(81) Wolfenden, R.; Andersson, L.; Cullis, P.; Southgate, C. *Biochemistry* **1981**, *20*, 849–855.

(82) Shirts, M. R.; Pande, V. S. *J. Chem. Phys.* **2005**, *122*, 134508.

(83) Huggins, D. J. *J. Chem. Phys.* **2012**, *136*, 064518.

(84) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Inter-molecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; pp 331–342.

(85) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381–1392.

(86) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(87) Price, D. J.; Brooks III, C. L. *J. Chem. Phys.* **2004**, *121*, 10096–10103.

(88) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

(89) Abascal, J. L. F.; Vega, C. *J. Chem. Phys.* **2005**, *123*, 234505.

(90) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.

(91) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.

(92) Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876–2880.

(93) Gubskaya, A. V.; Kusalik, P. G. *J. Chem. Phys.* **2002**, *117*, 5290–5302.

(94) Kusalik, P. G.; Svishchev, I. M. *Science* **1994**, *265*, 1219–1221.

(95) Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

(96) Kiyohara, K.; Gubbins, K. E.; Panagiotopoulos, A. Z. *Mol. Phys.* **1998**, *94*, 803–808.

(97) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; East-wood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341–346.

(98) Ensing, B.; Nielsen, S. O.; Moore, P. B.; Klein, M. L.; Parrinello, M. *J. Chem. Theory Comput.* **2007**, *3*, 1100–1105.

(99) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(100) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(101) Kunz, A.-P. E.; Allison, J. R.; Geerke, D. P.; Horta, B. A. C.; Hünenberger, P. H.; Riniker, S.; Schmid, N.; van Gunsteren, W. F. *J. Comput. Chem.* **2012**, *33*, 340–353.

Graphical TOC Entry
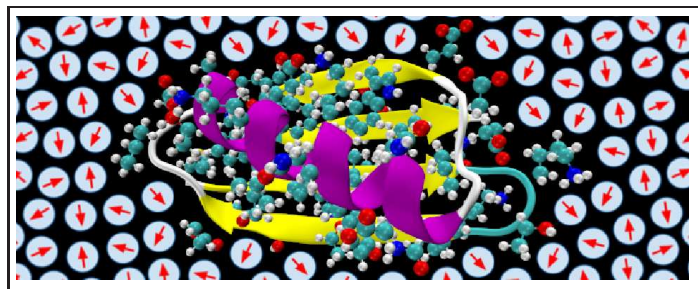
For Table of Contents use only.
Title of paper: Direct mixing of atomistic solutes and coarse-grained water.
Authors: Mario Orsi, Wei Ding, Michalis Palaiokostas.