

## Adaptive resolution simulation of an atomistic protein in MARTINI water

Julija Zavadlav,<sup>1</sup> Manuel Nuno Melo,<sup>2</sup> Siewert J. Marrink,<sup>2,a)</sup> and Matej Praprotnik<sup>1,3,b)</sup>

<sup>1</sup>Laboratory for Molecular Modeling, National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

<sup>2</sup>Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

<sup>3</sup>Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, SI-1000 Ljubljana, Slovenia

(Received 30 October 2013; accepted 10 January 2014; published online 6 February 2014)

We present an adaptive resolution simulation of protein G in multiscale water. We couple atomistic water around the protein with mesoscopic water, where four water molecules are represented with one coarse-grained bead, farther away. We circumvent the difficulties that arise from coupling to the coarse-grained model via a 4-to-1 molecule coarse-grain mapping by using bundled water models, i.e., we restrict the relative movement of water molecules that are mapped to the same coarse-grained bead employing harmonic springs. The water molecules change their resolution from four molecules to one coarse-grained particle and vice versa adaptively on-the-fly. Having performed 15 ns long molecular dynamics simulations, we observe within our error bars no differences between structural (e.g., root-mean-squared deviation and fluctuations of backbone atoms, radius of gyration, the stability of native contacts and secondary structure, and the solvent accessible surface area) and dynamical properties of the protein in the adaptive resolution approach compared to the fully atomistically solvated model. Our multiscale model is compatible with the widely used MARTINI force field and will therefore significantly enhance the scope of biomolecular simulations. © 2014 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License. [<http://dx.doi.org/10.1063/1.4863329>]

Molecular simulations provide insight into the physical basis of the structure, dynamics, and function of biological macromolecules, e.g., proteins.<sup>1</sup> Simulations yield complete trajectories of individual particles and can therefore be used to address specific questions about the properties of biomolecular systems that depend on these details. Water as a solvent profoundly affects structural, dynamical, and functional properties of proteins through both direct and hydrodynamic interactions.<sup>2</sup> A protein influences in turn the local structure and dynamics of water. To adequately describe this delicate interplay we have to resort to the atomic resolution in our model. However, due to a large number of atoms these systems are difficult to tackle by atomistic computer simulations. Moreover, the largest portion of the simulation time is spent on the solvent and not on the protein itself. To simplify the model of the system to the largest extent possible, while at the same time keeping the atomistic detail where it is needed, multiscale models of the solvent have been extensively developed.<sup>3</sup>

Many of the multiscale approaches are concerned with interfacing of atomistic and the continuum models of liquids.<sup>4–11</sup> In these hybrid methods, typically molecular dynamics (MD) or a similar approach is used for simulating dynamics on the atomistic scale, whereas the Navier-Stokes equation governs the fluid dynamics on the continuum scale. Alternatively, multiscale schemes have been introduced us-

ing particle-based models only, e.g., atomistic and physically simplified coarse-grained (CG) molecular models.<sup>12</sup> The coupling can be achieved either by a fixed resolution approach, where different resolution domains interact with each other but do not exchange molecules,<sup>13–20</sup> or the adaptive resolution approach, where molecules change their resolution according to their current positions.<sup>21–27</sup> Among the most advanced methods for the latter kind of simulations is the Adaptive Resolution Scheme (AdResS),<sup>21,28–32</sup> which allows for concurrent coupling from quantum all the way to continuum length scales of molecular liquids and soft matter.

In this work, we extend the range of applications of the AdResS scheme toward biological macromolecules. As a proof of principle, we simulated a 56-residue protein G, at full atomistic detail, solvated in water molecules that dynamically change resolution between an atomistic representation – a bundled version<sup>33</sup> of the Simple Point Charge (SPC) model<sup>34</sup> – and a mesoscopic one – the MARTINI CG model.<sup>35,36</sup> The successful interfacing with the widely used MARTINI CG model opens up a range of future applications involving the broad variety of solutes and solvents parameterized for this force field.

We use the new multiresolution water to solvate protein G (pdb entry 1PGB) as an example of a well studied protein.<sup>16</sup> The protein is always modeled at full atomistic resolution using the GROMOS 53a6 force field.<sup>37</sup> The solvent's level of representation depends on the distance from the protein's center of mass. For short distances we resort to a bundled version of the SPC model using a topology in which four

<sup>a)</sup>s.j.marrink@rug.nl

<sup>b)</sup>praprotn@cmi.ki.si

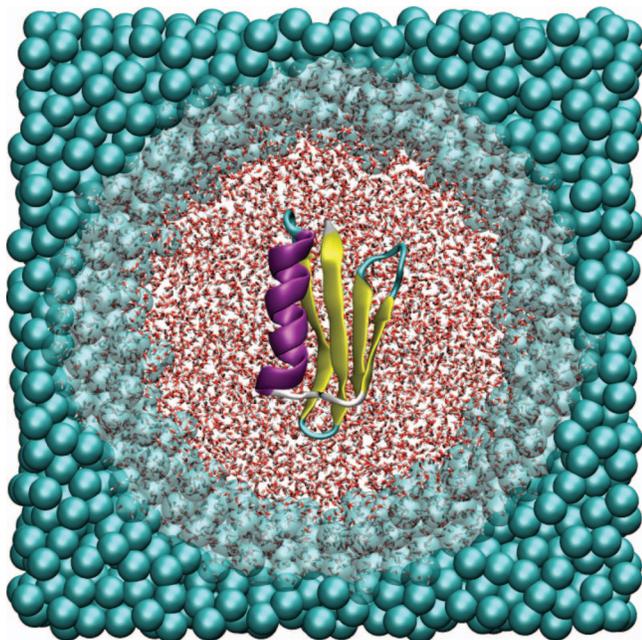


FIG. 1. A schematic cross section of simulation box with spherical adaptive resolution regions. Two levels of resolution are used for solvent molecules. High level of resolution is used for solvent molecules within a certain radius from the protein's center of mass. There the solvent is modeled with the bundled water model. Low level of resolution is used for solvent molecules at larger distances to protein's center of mass, where solvent is modeled with the MARTINI CG water (one bead represents four water molecules). The high resolution region sphere moves with the protein's center of mass. The protein G is thus at all times fully atomistic, but is shown here in cartoon representation (secondary structure) for better clarity.

water molecules are kept together using weak restraints.<sup>33</sup> This allows for exchange of the bundled water with the CG MARTINI water model that represents four water molecules by a single CG bead. At larger distances, the MARTINI water model is employed. The water molecules change their resolution from four molecules to one CG bead and vice versa adaptively according to their current position. A schematic representation of the system is depicted in Fig. 1.

By using the bundled water as our atomistic water model we avoid theoretical difficulties, both of the AdResS<sup>21</sup> as well as the coarse-graining of several nonbonded particles (e.g., solvent molecules) into one CG bead,<sup>38</sup> that arise when the atomistic molecules can drift apart (on a picosecond timescale in the case of water). It has been shown that the changes introduced by the bundling most strongly affect the self-interactions of the water molecules.<sup>33</sup> This results in a slightly modified water structure and consequently different dynamics. On the other hand, interactions with other molecules stay mostly unchanged and with the density and the free energies of hydration of small molecules important properties of SPC water are well preserved. Applications of bundled water to biomolecular systems, e.g., a lipid bilayer and a protein, show that stronger bundling may lead to some artifacts, such as the larger penetration of water into lipid bilayer interfaces and globular proteins, probably because of increased hydration of ionic species.<sup>33</sup> To avoid artifacts close to certain proteins, e.g., incorrect filling of protein water pockets, one could gradually weaken the springs closer to the protein, to such an

extent that the clusters can really deform to adopt to any local shape at no energetic cost. However, even with the current bundling the clusters can take many shapes.

The multiscale MD simulations are carried out, as mentioned above, using the AdResS method where the total force acting on a bundle  $\alpha$  is

$$\begin{aligned} \mathbf{F}_\alpha = & \sum_{\beta \neq \alpha} w(|\mathbf{R}_\alpha - \mathbf{R}|)w(|\mathbf{R}_\beta - \mathbf{R}|)\mathbf{F}_{\alpha\beta}^{ex} \\ & + \sum_{\beta \neq \alpha} [1 - w(|\mathbf{R}_\alpha - \mathbf{R}|)w(|\mathbf{R}_\beta - \mathbf{R}|)]\mathbf{F}_{\alpha\beta}^{cg} \\ & - \mathbf{F}^{TD}(|\mathbf{R}_\alpha - \mathbf{R}|), \end{aligned} \quad (1)$$

where  $\mathbf{F}_{\alpha\beta}^{ex}$  and  $\mathbf{F}_{\alpha\beta}^{cg}$  are the forces between bundles  $\alpha$  and  $\beta$ , obtained from the explicit atomistic and CG potentials, respectively. The sigmoidal function  $w \in [0, 1]$  is used to smoothly couple the high and low resolution regimes, where  $\mathbf{R}_\alpha$ ,  $\mathbf{R}_\beta$ , and  $\mathbf{R}$  are centers of mass of bundles  $\alpha$  and  $\beta$ , and the protein, respectively. The thermodynamic force  $\mathbf{F}^{TD}$  acts on bundles' centers of mass in the hybrid region and ensures the thermodynamic equilibrium of the system.<sup>29,39</sup>  $\mathbf{F}^{TD}$  is computed with an iterative procedure as described in the literature and shown in the supplementary material.<sup>40</sup>

The atomistic region around protein has a spherical shape with the distance between the hybrid domain border and the center of mass of the protein fixed. Hence, the resolution domains follow protein random translation with the center of the atomistic region coinciding with  $\mathbf{R}$  at all times. Our assumption, which is confirmed by the values of the respective diffusion coefficients (see below), is that the protein moves slowly compared to the solvent molecules. This is necessary, so that solvent molecules can equilibrate their degrees of freedom adequately when crossing the borders of domains with different resolution. By varying the size of the atomistic region, we gain insight on the extent of influence the bulk has on the local hydrogen bond network in the hydration shell. To this end, we investigate three sizes of atomistic sphere radii: 3.2, 3.4, and 3.6 nm. In all cases the center of atomistic region sphere moves with the protein's center of mass. As a reference, we use a simulation where the atomistic region extends across the whole simulation box as well as an atomistic simulation of the protein solvated with the SPC water model. Additionally, we have performed simulations of the protein enclosed in an atomistic water droplet in vacuum.

We test two models (models 1 and 2<sup>33</sup>) for bundling of atomistic waters. The two models differ in harmonic spring force constant  $k_s$  and  $C_{12}$  Lennard-Jones parameter between oxygen atoms (model 1:  $k_{s1} = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  and  $C_{12} = 3.25 \times 10^{-6} \text{ kJ mol}^{-1} \text{ nm}^{12}$ , model 2:  $k_{s2} = 4000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  and  $C_{12} = 3.45 \times 10^{-6} \text{ kJ mol}^{-1} \text{ nm}^{12}$ ). Comparing the results for both models, we have not found any noticeable differences. Therefore, we show here only the results for model 1 and direct the reader to the supplementary material for model 2 results.<sup>40</sup> The protein is modeled with the GROMOS 53a6 force field.<sup>37</sup> A cutoff of 1.2 nm is used for the nonbonded interactions. Electrostatic interactions are computed using a reaction field method<sup>41</sup> with dielectric constants of 54<sup>33</sup> and 80 for the bundled and SPC water,

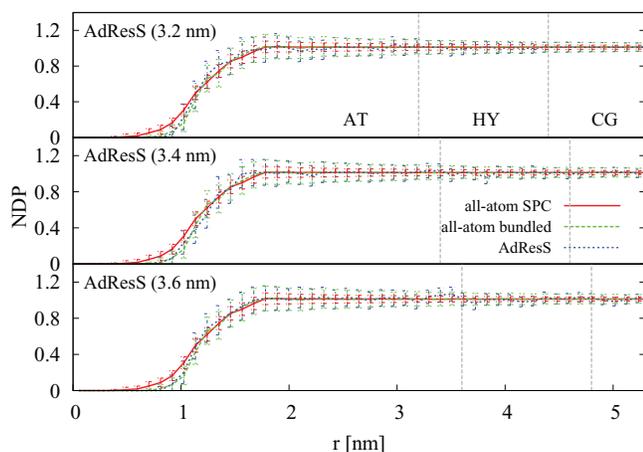


FIG. 2. Solvent density around the center of mass of protein G for AdResS simulations with atomistic region radius sizes of 3.2 nm (top), 3.4 nm (middle), and 3.6 nm (bottom). The plots include error bars.<sup>43</sup> The results are compared to the fully atomistic bundled and SPC solvations. The vertical lines denote the boundaries between resolution domains, i.e., the atomistic (AT), hybrid (HY), and the coarse-grained (CG) domains.

respectively. All simulations are performed in a cubic simulation box with box size 10.8 nm using periodic boundary conditions. The temperature is for all simulations maintained at 300 K by the Langevin thermostat with a coupling constant of  $25.0 \text{ ps}^{-1}$ . After equilibration, production runs of 15 ns were computed using a 1 fs timestep. All simulations are performed using the ESPResSo++ software package.<sup>42</sup>

First, we check that our adaptive resolution simulations using the new multiscale water model correctly reproduce the structure of the solvent and protein. In Fig. 2, we show the solvent density around the center of mass of the protein for the three atomistic region sizes. The domain with a decreased solvent density around the protein is well within the atomistic region for all cases. All-atom bundled and AdResS simulations give comparable results, while the all-atom SPC simulation gives a slightly shifted curve. This is because we calculate the density profile for bundled water using bundles' centers of mass and not the centers of mass of individual waters in the bundles. The corresponding oxygen (more or less the same as center of mass of individual water molecule) density profiles, where the mentioned shift in the density of water around the protein disappears, are reported in the supplementary material.<sup>40</sup> Note that the solvent density at larger distances from the protein is equal to the bulk density in all adaptive resolution domains.

To show that the multiscale simulation does not affect the structural properties of the protein we have computed the root-mean-square deviation (RMSD) and the root-mean-square fluctuations (RMSF) of the backbone atoms with respect to the crystal structure. The results plotted in Fig. 3 are in agreement with previously published results<sup>16</sup> and indicate that the structure of the protein is stable in all simulations. The average RMSD values (in nm) are  $0.17 \pm 0.05$ ,  $0.12 \pm 0.03$ ,  $0.17 \pm 0.02$ ,  $0.16 \pm 0.01$ , and  $0.19 \pm 0.02$  for the all-atom SPC, all-atom bundled SPC, and AdResS bundled (with the three atomistic region sizes) solvation, respectively.

We have also assessed the protein stability by computing the protein's radius of gyration, the stability of native contacts

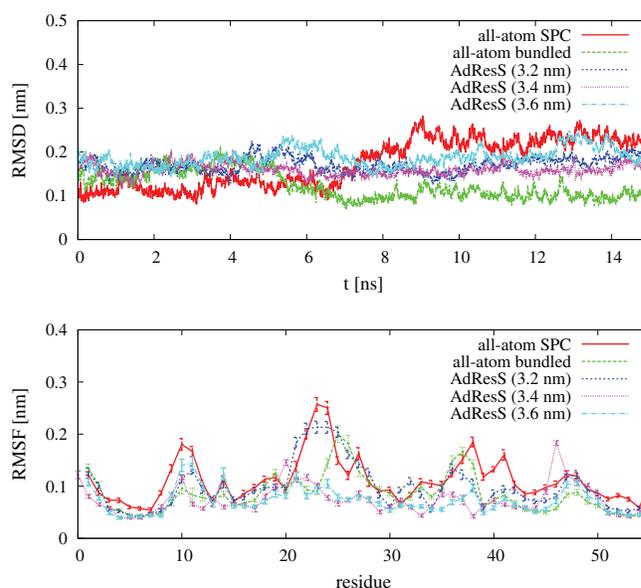


FIG. 3. RMSD (top) and RMSF with error bars (bottom) of the backbone atoms with respect to the crystal structure as a function of time. We compare the results obtained from the fully atomistic simulations using SPC and bundled waters to AdResS simulations with three atomistic region sizes: spheres of radii 3.2 nm, 3.4 nm, and 3.6 nm, respectively.

and secondary structure, and the solvent accessible surface area (see the supplementary material).<sup>40</sup> The results confirm that the protein remains stably folded and resides in the atomistic domain in all our simulations. Should the protein simultaneously extend over multiple adaptive resolution domains one would have to use a different approach as in Ref. 29.

To further validate our new approach, we also compare the dynamic properties of the protein and solvent for the atomistic and multiscale water models. The diffusion coefficients are determined from the mean square displacement using a finite size correction.<sup>44</sup> In particular, the computed values (in units of  $10^{-9} \frac{\text{m}^2}{\text{s}}$ ) for the protein's diffusion coefficient are:  $0.13 \pm 0.04$ ,  $0.08 \pm 0.03$ ,  $0.07 \pm 0.02$  for the all-atom SPC, all-atom bundled, and AdResS bundled (the same for all the atomistic region sizes) solvation, respectively. The obtained coefficients, which within the error bars match, are a lot smaller than the bundled water diffusion coefficient of  $1.8 \pm 0.1$ . This result also justifies our initial assumption of moving the center of the atomistic resolution region along with the protein's center of mass. More detailed results on the multiscale water itself will be published elsewhere.

In conclusion, we have presented a multiscale simulation of a solvated protein employing a hybrid atomistic/mesoscopic water model. The atomistic domain was limited to a sphere embedding the protein, while a mesoscopic MARTINI water model was used farther away. This leads to a significant speedup in comparison to a fully atomistic simulation. The speedup is due to the coarse-grained model, which reduces the number of solvent particles by a factor of 12 and introduces softer interactions. The actual speedup of AdResS simulations depends on the ratio between the atomistic and coarse-grained domain sizes, i.e., up to an order of magnitude using the same integration time step in the atomistic and CG regions.<sup>45</sup> This is important because the hydrodynamic

interactions are long-ranged and we need large systems to avoid finite size effects. Our approach falls in between the multiresolution approaches concurrently coupling atomistic water to a CG water, where one bead represents one water molecule,<sup>21</sup> and hybrid methods interfacing atomistic description with continuum, e.g., Refs. 10 and 11. Hence, it bridges the hydrodynamics from the atomic to mesoscopic scale and enables the study of biophysical phenomena that are beyond the scope of either atomistic or mesoscopic simulations. Future work will include extension of this methodology to polarizable CG water models<sup>46,47</sup> and salt solutions.

We thank M. Karttunen and K. Kremer for useful discussions. M.P. and S.J.M. acknowledge hospitality at KITP where this collaboration was initiated. This research was supported in part by the National Science Foundation under Grant No. NSF PHY11-25915. J.Z. and M.P. acknowledge financial support through Grant Nos. J1-4134 and P1-0002 from the Slovenian Research Agency. M.N.M. acknowledges funding (Veni Grant No. 722.013.010) from the Netherlands Organisation for Scientific Research (NWO).

- <sup>1</sup>M. Karplus and J. A. McCammon, *Nat. Struct. Biol.* **9**, 646 (2002).
- <sup>2</sup>G. Chopra, C. M. Summab, and M. Levitt, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 20239 (2008).
- <sup>3</sup>S. C. L. Kamerlin, S. Vicatos, A. Dryga, and A. Warshel, *Annu. Rev. Phys. Chem.* **62**, 41 (2011).
- <sup>4</sup>K. M. Mohamed and A. A. Mohamad, *Microfluid Nanofluid* **8**, 283 (2010).
- <sup>5</sup>G. D. Fabritiis, R. Delgado-Buscalioni, and P. V. Coveney, *Phys. Rev. Lett.* **97**, 134501 (2006).
- <sup>6</sup>X. Nie, M. O. Robbins, and S. Chen, *Phys. Rev. Lett.* **96**, 134501 (2006).
- <sup>7</sup>D. A. Fedosov and G. E. Karniadakis, *J. Comput. Phys.* **228**, 1157 (2009).
- <sup>8</sup>A. Donev, J. B. Bell, A. L. Garcia, and B. J. Alder, *Multiscale Model. Simul.* **8**, 871 (2010).
- <sup>9</sup>R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, *J. Chem. Phys.* **128**, 114110 (2008).
- <sup>10</sup>R. Delgado-Buscalioni, K. Kremer, and M. Praprotnik, *J. Chem. Phys.* **131**, 244107 (2009).
- <sup>11</sup>J. H. Walther, M. Praprotnik, E. M. Kotsalis, and P. Koumoutsakos, *J. Comput. Phys.* **231**, 2677 (2012).
- <sup>12</sup>G. S. Ayton, W. G. Noid, and G. A. Voth, *Curr. Opin. Struct. Biol.* **17**, 192 (2007).
- <sup>13</sup>A. Warshel, *Annu. Rev. Biophys. Biomol. Struct.* **32**, 425 (2003).
- <sup>14</sup>M. Neri, C. Anselmi, M. Cascalla, A. Maritan, and P. Carloni, *Phys. Rev. Lett.* **95**, 218102 (2005).
- <sup>15</sup>A. J. Rzepiela, M. Louhivuori, C. Peter, and S. J. Marrink, *Phys. Chem. Chem. Phys.* **13**, 10437 (2011).
- <sup>16</sup>S. Riniker, A. P. Eichenberger, and W. F. van Gunsteren, *J. Phys. Chem. B* **116**, 8873 (2012).
- <sup>17</sup>W. Han and K. Schulten, *J. Chem. Theory Comput.* **8**, 4413–4424 (2012).
- <sup>18</sup>P. Sokkar, S. M. Choi, and Y. M. Rhee, *J. Chem. Theory Comput.* **9**, 3728 (2013).
- <sup>19</sup>T. A. Wassenaar, H. I. Ingolfsson, M. Priess, S. J. Marrink, and L. V. Schaefer, *J. Phys. Chem. B* **117**, 3516 (2013).
- <sup>20</sup>H. C. Gonzalez, L. Darre, and S. Pantano, *J. Phys. Chem. B* **117**, 14438 (2013).
- <sup>21</sup>M. Praprotnik, L. Delle Site, and K. Kremer, *Annu. Rev. Phys. Chem.* **59**, 545 (2008).
- <sup>22</sup>C. F. Abrams, *J. Chem. Phys.* **123**, 234101 (2005).
- <sup>23</sup>S. O. Nielsen, P. B. Moore, and B. Ensing, *Phys. Rev. Lett.* **105**, 237802 (2010).
- <sup>24</sup>A. Heyden and D. G. Truhlar, *J. Chem. Theory Comput.* **4**, 217 (2008).
- <sup>25</sup>S. Artemova and S. Redon, *Phys. Rev. Lett.* **109**, 190201 (2012).
- <sup>26</sup>R. Potestio, S. Fritsch, P. Español, R. Delgado-Buscalioni, K. Kremer, R. Everaers, and D. Donadio, *Phys. Rev. Lett.* **110**, 108301 (2013).
- <sup>27</sup>R. Potestio, P. Español, R. Delgado-Buscalioni, R. Everaers, K. Kremer, and D. Donadio, *Phys. Rev. Lett.* **111**, 060601 (2013).
- <sup>28</sup>M. Praprotnik, L. Delle Site, and K. Kremer, *J. Chem. Phys.* **123**, 224106 (2005).
- <sup>29</sup>M. Praprotnik, S. Poblete, and K. Kremer, *J. Stat. Phys.* **145**, 946 (2011).
- <sup>30</sup>A. B. Poma and L. Delle Site, *Phys. Rev. Lett.* **104**, 250201 (2010).
- <sup>31</sup>H. Wang, C. Schütte, and L. Delle Site, *J. Chem. Theory Comput.* **8**, 2878 (2012).
- <sup>32</sup>H. Wang, C. Hartmann, C. Schütte, and L. Delle Site, *Phys. Rev. X* **3**, 011018 (2013).
- <sup>33</sup>M. Fuhrmans, B. P. Sanders, S. J. Marrink, and A. H. de Vries, *Theor. Chem. Acc.* **125**, 335 (2010).
- <sup>34</sup>H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, *Intermolecular Forces* (Reidel, 1981), p. 331.
- <sup>35</sup>S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries, *J. Phys. Chem. B* **111**, 7812 (2007).
- <sup>36</sup>S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.* **42**, 6801 (2013).
- <sup>37</sup>C. Oostenbrink, A. Villa, A. E. Mark, and W. F. van Gunsteren, *J. Comput. Chem.* **25**, 1656 (2004).
- <sup>38</sup>H. Bock, K. E. Gubbins, and S. H. Klapp, *Phys. Rev. Lett.* **98**, 267801 (2007).
- <sup>39</sup>S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer, *Phys. Rev. Lett.* **108**, 170602 (2012).
- <sup>40</sup>See supplementary material at <http://dx.doi.org/10.1063/1.4863329> for adaptive resolution simulation of an atomistic protein in MARTINI water.
- <sup>41</sup>M. Neumann, *Mol. Phys.* **50**, 841 (1983).
- <sup>42</sup>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).
- <sup>43</sup>M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
- <sup>44</sup>B. Dunweg and K. Kremer, *J. Chem. Phys.* **99**, 6983 (1993).
- <sup>45</sup>The speedup could be even larger: in the CG region one could use longer time steps (in principle), although that is not implemented yet.
- <sup>46</sup>S. O. Yesylevskyy, L. V. Schafer, D. Sengupta, and S. J. Marrink, *PLoS Comput. Biol.* **6**, e1000810 (2010).
- <sup>47</sup>Z. Wu, Q. Cui, and A. Yethiraj, *J. Phys. Chem. B* **114**, 10524 (2010).