



# A different approach to dual-scale models

**DOI:**

[10.1016/j.jcp.2020.109465](https://doi.org/10.1016/j.jcp.2020.109465)

**Document Version**

Final published version

[Link to publication record in Manchester Research Explorer](#)

**Citation for published version (APA):**

Gowers, R. J., Carbone, P., & Di Pasquale, N. (2020). A different approach to dual-scale models. *Journal of Computational Physics*, 413, 109465. <https://doi.org/10.1016/j.jcp.2020.109465>

**Published in:**

Journal of Computational Physics

**Citing this paper**

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

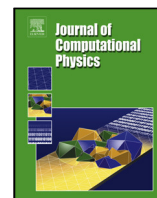
**General rights**

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

**Takedown policy**

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact [uml.scholarlycommunications@manchester.ac.uk](mailto:uml.scholarlycommunications@manchester.ac.uk) providing relevant details, so we can investigate your claim.





## A different approach to dual-scale models

Richard J. Gowers<sup>a</sup>, Paola Carbone<sup>b</sup>, Nicodemo Di Pasquale<sup>c,\*</sup>

<sup>a</sup>College of Engineering and Physical Sciences, University of New Hampshire, Durham, NH, US

present address: NextMove Software Ltd, Innovation Centre, Science Park, Milton Road, Cambridge, CB4 0EY, UK

<sup>b</sup>Department of Chemical Engineering & Analytical Science, University of Manchester, Manchester, M13 9PL, UK

<sup>c</sup>School of Mathematics and Actuarial Science, University of Leicester, University Rd, Leicester, LE1 7RH, UK

### ARTICLE INFO

#### Article history:

Molecular Dynamics, Coarse-graining, Multi-scale modelling

### ABSTRACT

Dual-resolution models which combine atomistic and coarse-grained details are usually built sequentially. First, the coarse-grained model is optimized and subsequently atomistic regions are introduced. Finally, these two different resolutions are connected together. Here, we present a methodology for a direct construction of these dual-resolution models without the inclusion of these two stages. Our model takes advantage of the use of virtual sites and retain the full atomistic structure of the molecules by coarse-graining only some selected intermolecular interactions. The method is tested on different models of octanol where the atomistic details in the OH head groups are retained at an atomistic level, while an iterative Boltzmann inversion CG forcefield is used to model the alkyl tails. The procedure is applied to an All-Atom and United-Atom force-fields model. The dual-resolved versions of each atomistic model are able to retain the hydrogen bonding structure and dynamics, with the united atom model showing slightly better results. Finally, the current computational performance of such a model is explored and compared against the potential theoretical performance.

© 2020 Elsevier Inc. All rights reserved.

## 1. Introduction

The use of coarse-graining techniques to reduce the degrees of freedom of an atomistic model to access longer length and time scales in simulations has become an established and powerful tool in the modern simulator's toolbox. There are two schools of thought in how to derive parameters for these models. The first of these adopts a *bottom up* approach, which uses the atomistic interactions to derive the (effective) potentials among beads in a coarse-grained (CG) description of the molecules [1]. In one of the possible *bottom up* approaches, a set of reference atomistic probability distributions can be optimised using the

\*ndp8@leicester.ac.uk

inverse Boltzmann iteration scheme [3, 4, 5] and the inverse Monte Carlo method [6], which are based on the well-known result that there exists a unique potential to reproduce a given pairwise distribution [2]. Alternatively the potential of mean force between CG sites can directly be targeted using the force matching method [7], or by minimizing the distance between the CG configurational distribution and the atomistic one [8].

Complementary to this, there is the *top down* approach, where the CG interactions are parametrized without explicitly considering an underlying atomistic model [9]. A variety of target quantities can be used to optimise CG interactions in *top down* approach, including density and interfacial tension [10], partition coefficients [11], or a molecular-based equation of state [12].

Many of these methods have matured to the stage that in recent years freely distributed software tools have been created to automate the process of generating potentials such as Martini[13, 14], force matching[15], IBI [15, 4] and Inverse Monte Carlo [16].

Despite their successes, CG simulations have some drawbacks. For example, if hydrogen bonds have to be explicitly accounted for, or we are interested in keeping the information related to the underlying atomistic electrostatic properties, CG models with additional features need to be considered [17, 18, 19, 20, 21]. Secondly, CG methods usually assume that the resulting bonded degrees of freedom are independent from each other, and it is known to be difficult to find a mapping scheme where this is strictly true [24, 25].

These problems are all related with the loss of the atomistic details in CG models and different ways were devised to address it. The most straightforward solution is to reintroduce all the atomistic details in the system, in a process known as backmapping. However, backmapping is still a fundamentally challenging problem even for simple molecules [30, 31]. Building upon the strengths of CG models, other approaches proposed to reincorporate only some of the Fine-Grained (FG) underlying details lost during the coarse-graining procedure. These selected details coming from the fine-grained system are added to the CG model as supplementary features of the beads, such as in the ELBA water model [32, 33]. In this model, each bead is described as a spherical interaction site, which represents the center of mass of the replaced group of atoms and a dipole point. Other approaches include the use of a mixed description of both atoms and beads simultaneously to selectively choose which details to preserve. These models are usually called hybrid, dual-resolution or dual-resolved models. However, the term hybrid is now preferred for continuum/discrete coupling and we will not use it in this work. With two distinct levels of resolution in the system, atoms and beads, the problem of how to describe interactions between them arises. This can be addressed by introducing extra forcefield parameters between atoms and beads [34], or through the use of Virtual Sites (VS) to mediate any interactions between the two domains, atomistic and coarse-grained one [35, 36, 37, 38, 1, 39, 40]. In particular, the description of the electrostatic interactions at an atomistic level in dual-resolved models has been proved a good compromise to retain atomistic details [41], including hydrogen bonds [42].

One of the first attempt of dual-resolved model is the AdRes scheme [43, 44, 45]. In AdRes, the simulation box is split into two parts: an atomistic region and a CG region. Molecules can move freely in and out the two regions. If a CG molecule enters the atomistic region it is remapped to its full atomistic descriptions, whereas the opposite is done when a full atomistic molecule leaves the atomistic region. This set up requires modelling the interactions between molecules at the full atomistic and at the CG level simultaneously and is obtained by using an interaction site located on the center of mass of each atomistic molecule. The dynamic on-the-fly adjustment of the system degrees of freedom that AdRes allows is ideal in those situations where the molecules to be coarse-grained are highly mobile such as water.

More recently we proposed a similar procedure to model polymer melts [36, 42] which keeps the level of model resolution fixed during the whole simulations. The scheme we show here is built on this work, but proposes a different approach. Instead of replacing the atoms with beads, we modify the interactions. While keeping the whole atomistic detail, we replace some of the (computationally expensive) atomistic interactions with the cheaper ones derived from CG models. We apply this methodology to octanol solution as a test case. The hydroxyl group of the octanol and the hydrogen bonds that it creates in the system represents a good illustration of the effectiveness of the methodology. Moreover, octanol represents an important tool in drug discovery [46] and therefore good atomistic and CG models describing this compound become essential.

## 2. Procedure to develop the dual-scaled model

In this paper we describe a procedure to perform molecular dynamics (MD) simulations with the system described simultaneously at both fine-grained and coarse-grained level of detail. Our overall objective is its accurate representation, while minimising the computational cost of the calculations.

The main difference with other dual-resolution approaches [35, 36, 42, 43] lays in the fact that we do not coarse-grain all or part of the molecules (i.e. we replace atoms with beads), we coarse-grain only the interactions. In other words, the idea is to keep the molecules at the atomistic detail, as in a classic atomistic model, and replace some of the original fine-grained interactions with coarse-grained ones. CG interactions act among specific interactive sites, which we call Virtual Sites, each of which representing a group of atoms. We will now briefly set out the theory behind such a model and then apply this to a system of liquid octanol to test its effectiveness.

Performing a molecular dynamics simulation requires the iterative numerical calculation of the Hamiltonian [47]. The Hamiltonian for a system of  $M^{FG}$  particles in the absence of any external field can be defined as [48]

$$\mathcal{H}(\boldsymbol{\gamma}) = \mathcal{K}(\mathbf{p}) + U(\mathbf{r}) \quad (1)$$

where  $\boldsymbol{\gamma} = (\mathbf{r}, \mathbf{p})$  are the atomistic phase-space variables representing the position  $\mathbf{r}$  and momentum  $\mathbf{p}$  of the  $N$  particles, whereas  $\mathcal{K}$  and  $U$  represents the kinetic and potential energy contributions respectively. The potential energy contribution can be expressed as a sum of non-bonded ( $U_{nb}$ ) and bonded ( $U_{bonded}$ ) terms:

$$U(\mathbf{r}) = U_{bonded}(\mathbf{r}) + U_{nb}(\mathbf{r}) \quad (2)$$

The most computationally expensive part of any MD simulation is the evaluation of nonbonded interactions [49, 50], which can consume around 90% of the simulation runtime. The computational cost of calculating  $U_{nb}$  scales with the number of interactions that must be calculated, meaning that not only large systems are more expensive to simulate, but also systems with a higher number density of particles.

Coarse-grained simulations reduce this computational burden by reducing the degrees of freedom of the system. As discussed, the drawback is that the resulting coarse-grained systems suffers of an irreversible loss in detail with respect the original models. While keeping all the atomistic details is usually not interesting, in some situations the loss of information is too severe, and the atomistic details is needed (at least) in specific regions of the system. In these latter cases, dual-resolved models were developed which couple the coarse-grained description with the atomistic one. In this way we can combine the reduction of the computational burden offered by the coarse-grained models with the detailed description given by the atomistic ones in those regions where we want a higher level of detail.

The potential energy for a dual-resolved system,  $U^H(\mathbf{R}, \mathbf{r})$ , is usually represented by the sum of three terms, the interactions between beads, the interaction between atoms, and the *mixed* interactions,  $U^M(\mathbf{R}, \mathbf{r})$ , between atoms and beads:

$$U^H(\mathbf{R}, \mathbf{r}) = U^{FG}(\mathbf{r}) + U^{CG}(\mathbf{R}) + U^M(\mathbf{R}, \mathbf{r}) \quad (3)$$

While the derivation of the CG interactions is well known and documented in literature and a number of models exists to perform this task, the derivation of the mixed interactions is still under development. If the beads and atoms coexist within the same molecule, some atoms will be bonded to some beads and vice-versa. Previous attempts at using bonds directly between fine and coarse particles showed good agreements with fully atomistic models for simple systems like poly-ethylene and poly-styrene [36], and also for more complicated molecules such as poly-amide, where the double resolution of the model can satisfactorily capture hydrogen bonds formation and breakage [42].

The non-bonded potential between beads and atoms can instead be described by introducing the concept of Virtual Sites (VS)[35, 36]. Each virtual site represents the position the bead would have in a group of atoms if the atoms are coarse-grained away. The idea is to have the “real” beads interact with the VS instead of the atoms through the non-bonded interactions. The force applied to the VS is then distributed on the atoms that compose the VS itself. The VSs act as a proxy for allowing two groups of atoms to interact through a single pairwise potential. It was shown [1] that, when a mapping representing the COM of the group of atoms is used, the VSs do not represent an approximation of the non-bonded interactions, and can be justified directly from the theory. The scheme to obtain the mixed interactions just outlined is

not very efficient, since it includes deriving the CG interactions for all the system and from these the specific mixed interactions for those atoms that are not coarse-grained away. That means that we are wasting time deriving some CG interactions that will not be used at all in our calculations .

In this work we describe a different path to obtain a dual-resolved model which consider only the CG interactions strictly necessary to capture the full behaviour of the system, and most importantly, preserve the atomistic details. The main idea is to leave the system at the full atomistic details and replace instead *some* of the atomistic interactions with CG ones. There is no need to derive a further connection between CG interactions and atomistic description because the VS automatically provide such a contact.

In order to better explain this new derivation, let us start by describing how the total interaction potential term will look like in this case. Equation (2) will be our starting point. Interactions in a full atomistic system can be split in bonded and non-bonded components. Since the formers (either bonds, angles or torsions) are relatively fast to compute, we leave them at their atomistic level. We focus now our attention on the non-bonded interactions which represent the most computational demanding part of any MD calculation. We map specific group of atoms into a bead (a Virtual Site) and we replace the computational expensive step of calculation of non-bonded interactions between atoms, with the cheaper one involving VSs. The total potential energy for our dual-resolved system,  $U^{DR}(\mathbf{r})$ , will therefore be:

$$U^{DR}(\mathbf{r}) := U^{DR}(\mathbf{R}^{VS}(\mathbf{r}), \mathbf{r}) = U_{bonded}(\mathbf{r}) + U_{nb}^{FG}(\mathbf{r}) + U_{nb}^{CG}(\mathbf{R}^{VS}(\mathbf{r})) \quad (4)$$

where we used the symbol  $\mathbf{R}^{VS}(\mathbf{r})$  to emphasise that we are not considering actual beads but virtual sites only, which, in turn, are function of the position of the atoms. The number and types of VSs in the system need to be prescribed in advance, and the latter will define the number of different CG interactions we need to calculate. The additional term  $U_{nb}^{FG}(\mathbf{r})$  in eq. (4) represents the fine-grained non-bonded potentials, which can be kept if we need some (non-bonded) interactions at fine-grained detail. In fact, one of the advantages of this framework is that we can select which non-bonded interactions have to be coarse-grained away, while leaving the others at their atomistic resolution. As an example, in the octanol model that we are going to discuss in detail in the next section (see section 2.1), the interactions involving hydroxyl group of different molecules remain at the atomistic level while the non-bonded interactions involving alkyl groups of the alkyl chain are described at CG level through VS. This ensure an explicit description of the hydrogen bonds without the need to resort to additional models, as we show in the results section (see section 3). In addition, the intramolecular non-bonded interactions, i.e. the non-bonded interactions among atoms belonging to the same molecules, are described at the the atomistic detail.

One thing that needs to be avoided is the double counting of interactions. That means that a group of atoms can interact with other groups of atoms either through atomistic interactions or through the CG interactions mediated by the VS, never both at the same time.

Once we have defined the *splitting* of the interactions as shown in eq. (4) we now need to develop the CG interactions between virtual sites. These are the VS-VS coarse-grained non-bonded interactions that will be added on top of the existing atomistic ones. These can be sourced from pre-existing libraries of CG parameters or, as in this work, they can be derived alongside the dual-resolved model. This represents the central idea of our approach. We use the IBI model to derive the VS interactions that we need (and only those needed), *without* coarse-graining the system. The molecules retain their atomistic identities, with the only difference being that some interactions are mediated by the VS, and these CG interactions are optimized (in this work) with the IBI technique.

**Algorithm 1** Step-by-step derivation of the dual-resolved model

---

```

1: procedure DUAL-RESOLVED MODEL IBI VERSION (Input: Configuration of the atomistic system)
2:   Mapping of the atomistic model and definition of the position of the VS
3:   Calculation of target VS-VS RDFs
4:   Derivation of first guess VS-VS potentials
5:   while NOT (convergence criteria fulfilled) do
6:     Run atomistic simulations, with interactions as in eq. (4)
7:     Redefinition of the position of VS and atoms
8:     Re-calculations of the RDFs for the current configuration
9:     Update VS potentials
10:  end while
11: end procedure

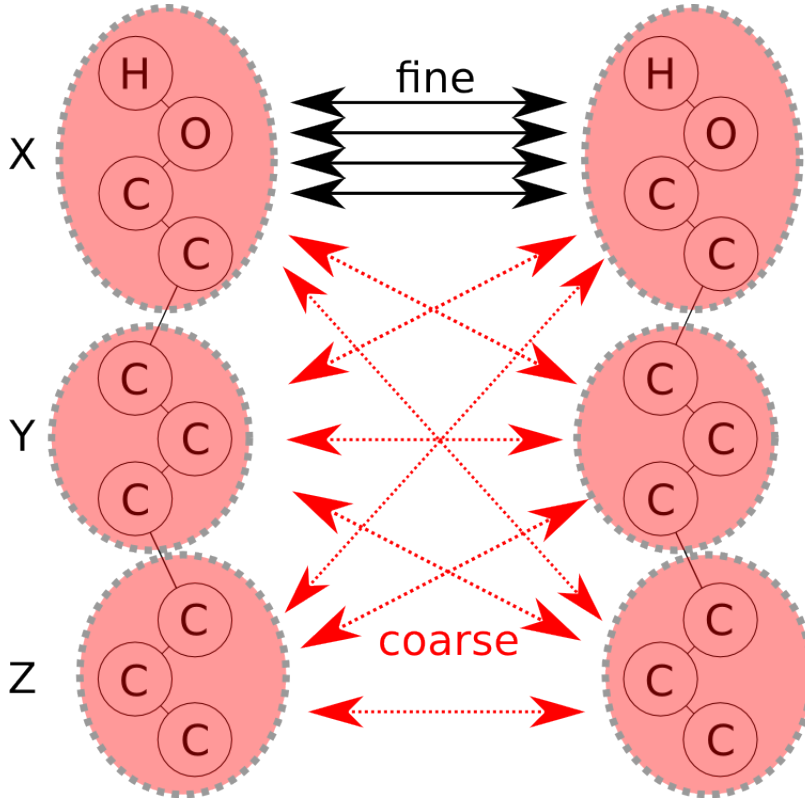
```

---

We reported the step-by-step operations needed to obtain the dual-resolved model in Algorithm 1. As shown in the algorithm reported, the whole process can be fully automatized and possibly included in other packages used to develop CG models (e.g. VOTCA [53]).

### 2.1. Dual-Resolved octanol

The previously outlined methodology is now applied to liquid octanol. Two separate atomistic models of octanol were chosen to serve as reference models, an all-atom (AA) model and a united-atom (UA) model. Apart from the number of interacting sites in each of these models, the main difference between them is the use of partial charges on the atoms. In the AA model, every atom has a small partial charge, whilst in the UA model the partial charges are located only on the hydroxyl group and the carbon attached to it, leaving the majority of the molecule uncharged. We will refer to these two atomistic models as aAA and aUA respectively. To define the VS the same mapping scheme was employed for both atomistic models, with the nine heavy atoms in octanol split into three equally sized CG sites. This mapping scheme is shown in fig. 1, with the three VS being named X, Y and Z, with X being the OH end of the octanol. With each of the three VS in the molecule having different types, there are six different pairs of VS interactions. Of these, it was chosen to model X-X interactions at the fine level (i.e. atomistic), while all other intermolecular interactions are described at the coarse level. The interactions among atoms within the same molecule, i.e. the intramolecular non-bonded interactions, are described at the FG level. A representation of the octanol molecule with the mapping scheme we are using in this work is shown in fig. 1. We will refer to the two dual resolved models as drUA and drAA, the former being the one obtained from the atomistic united-atom model, the latter obtained from the atomistic all-atom model. Since X, Y, Z are not real beads we will refer to them as pseudo-beads.



**Fig. 1:** Mapping scheme for octanol and visualising the level of interaction between two adjacent octanol molecules. Atomistic sites and interactions are shown in solid black lines while coarse are shown in red and dashed.

## 2.2. Deriving CG Potentials

In total five nonbonded potentials are required for the virtual sites, one for each VS pairing, excluding the self-interaction between the pseudo-beads X (see fig. 1). These were derived using the iterative Boltzmann inversion (IBI) method [3]. In this method, the pairwise nonbonded potentials ( $U$ ) are initially guessed according to a Boltzmann inversion of the pair correlation function  $g_{ref}(r)$  as:

$$U(r) \approx -k_B T \ln(g_{ref}(r)) \quad (5)$$

Where  $k_B$  and  $T$  are the Boltzmann constant and temperature respectively. Note that the use of Boltzmann inversion makes the resulting potential tied to a particular thermodynamic point, even if it is not always the case [55]. The structures generated from these initial potentials are generally quite poor. Therefore, the potentials are iteratively refined by adjusting them according to the difference between the reference pair distributions and those produced by the previous iteration step ( $g_i(r)$ ):

$$U_{i+1}(r) = U_i(r) + \alpha k_B T \ln\left(\frac{g_i(r)}{g_{ref}(r)}\right) \quad (6)$$

where  $\alpha$  is a heuristic prefactor to moderate the change to subsequent potentials; in this work we used a value of  $\alpha = 0.3$ . The potentials can be iterated multiple times using this method, with the correction factor approaching zero as the potential converge.

The potentials generated from this method will not reproduce the correct pressure [3], therefore once the forcefield starts to converge on the target distributions, a pressure correction step is added. The pressure correction to the coarse potentials takes the form of a linear ramp, [3]:

$$\Delta U_i(r) = \left(1 - \frac{r}{r_{cut}}\right) A_i \quad (7)$$

The parameter  $A$  was chosen by considering the virial expansion of the pressure, for both the current iteration  $i$ ,  $P_i$ , defined as:

$$P_i V = \frac{\rho}{k_B T} - \frac{2\pi\rho^2}{3r_{cut}} \int_0^{+\infty} r^3 \frac{dU(r)}{dr} dr \quad (8)$$

and the target pressure,  $P_{target}$ :

$$P_{target} V \approx \frac{\rho}{k_B T} - \frac{2\pi\rho^2}{3r_{cut}} \int_0^{+\infty} r^3 \frac{d}{dr} (U(r) + \Delta U(r)) dr \quad (9)$$

where  $\rho$  is the particle density and  $V$  is the volume of the system considered. The parameter  $A_i$  is then defined as [56]:

$$\left[ \frac{-2\pi\rho^2}{3r_{cut}} \int_0^{r_{cut}} r^3 g_i(r) dr \right] A_i = \Delta PV. \quad (10)$$

This adjustment to the potential corrected the pressure within a few iterations, but also disturbed the convergence of the pair potentials, requiring extra iterations to be performed.

### 2.3. Simulation Details

All simulations were obtained using version 5.0.4 of Gromacs [57, 58]. The system is composed by 4096 molecules of octanol, simulated inside a cubic volume with length approximately 10 nm and periodic boundary conditions in the three directions. Parameters for the reference atomistic UA model came from Siwko [59] while parameters for the reference AA model came from the OPLS-AA forcefield [60] and were found using Lipidbook [61]. The parameters for the CG part of the dual resolved version of each model was derived using the IBI method detailed above. A nonbonded interaction cutoff of 0.9 nm was used for the AA atomistic model, 1.2 nm for the UA atomistic model and 1.4 nm for the CG potentials in both models. In all models electrostatics were calculated using the particle-mesh Ewald (PME) method [62]. Production runs were conducted in an NPT ensemble, with the pressure maintained at 1 bar using an isotropic Berendsen barostat [63] with a compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$  and a coupling time of 2 ps. The temperature was maintained at 298 K using the velocity rescale thermostat [64] with a coupling time of 0.1 ps. An integration timestep of 1 fs was used for all simulations, with 50 ns of simulated time.

## 3. Results

Unless otherwise stated, all data were analysed using a combination of the MDAnalysis [65, 66] and datreant [67] Python packages. Plots were created using matplotlib [68].

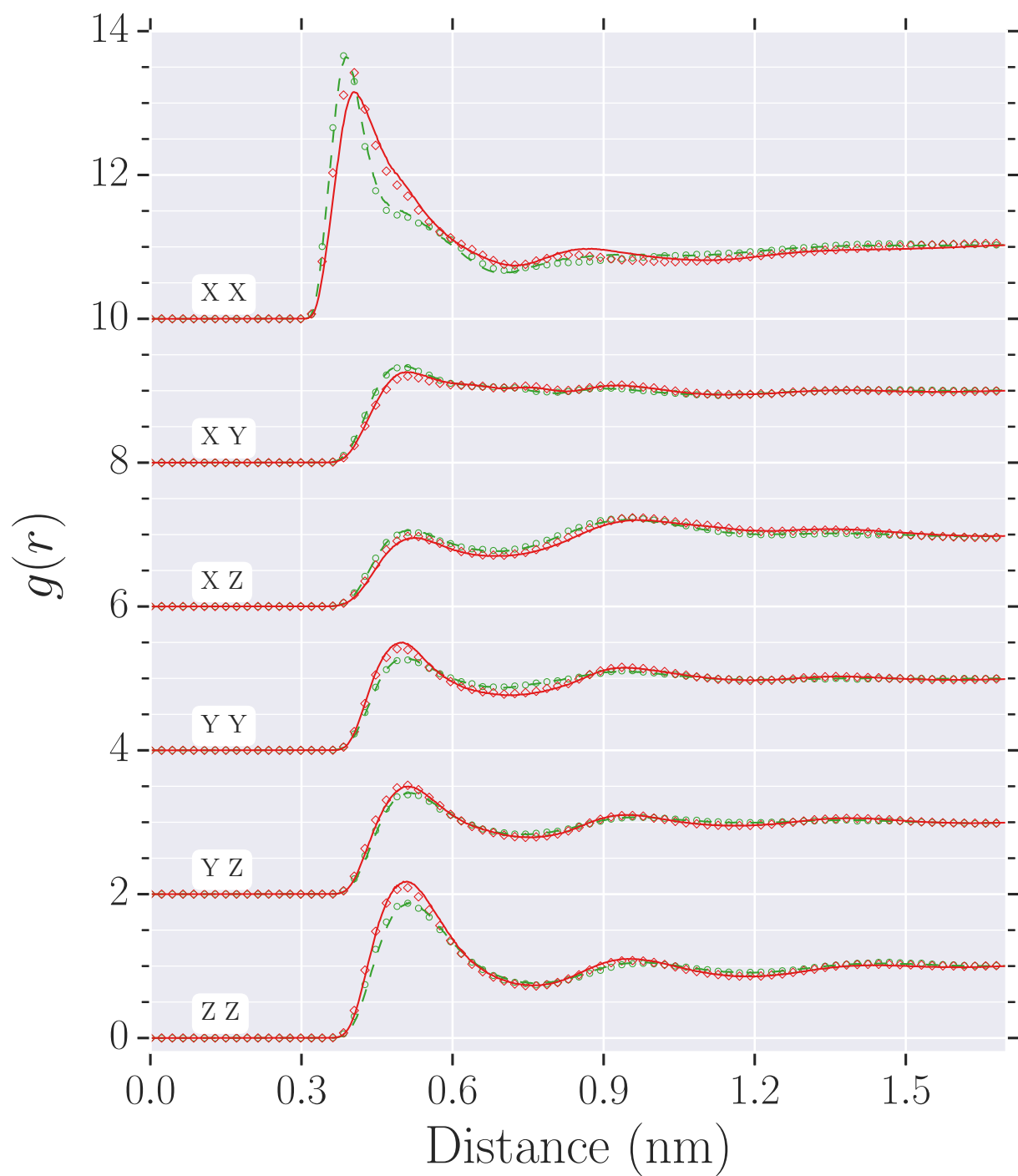
### 3.1. Structural Results

Overall IBI potentials converge after approximately 25 iterations. The final pair distributions for all VS pair combinations are shown in fig. 2. The first thing to notice here is that the IBI method is able to find a solution of the optimization problem (see eq. (6)) also for a model where only parts of the interactions are coarse-grained and part of them remains at the atomistic detail.

For the drAA model, the agreement between the reference and final structure is worse, most notably at a distance of 0.4 nm where a small difference with the reference peak can be observed. This could be attributed to the lack of electrostatic screening caused by the removal of the partial charges in the coarse-grained tail, necessary as these degrees of freedom were removed.

However, when the distribution of the magnitude of the force on each atom is compared between the atomistic and dual resolved models, both the mean and also the distribution around the mean is perfectly matched, as shown in the Supporting Information (see Fig. S1).

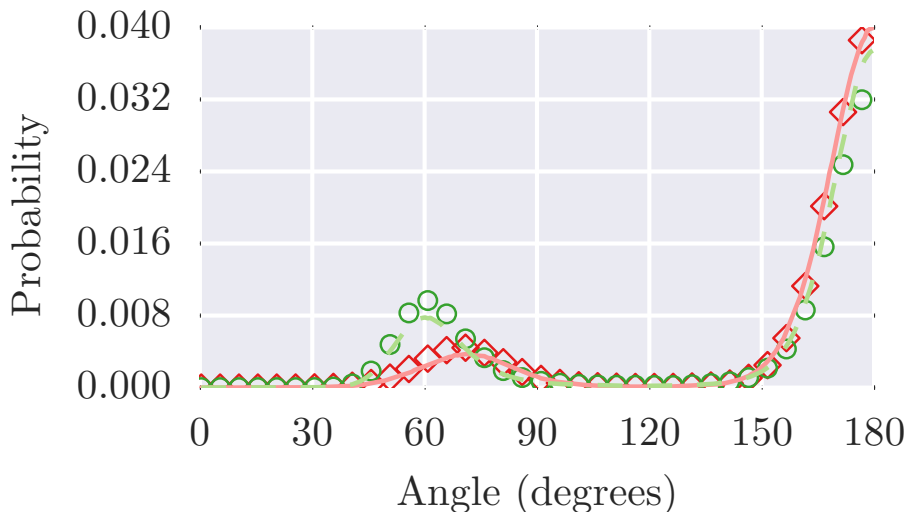




**Fig. 2:** Comparison of pair distributions across the different models tested: atomistic aUA (green dashed line), drUA (green  $\circ$ ), aAA (red solid line), drAA (red  $\diamond$ ). Different pair distributions have been shifted by two units along the y axis.

### 3.2. Bonded parameters

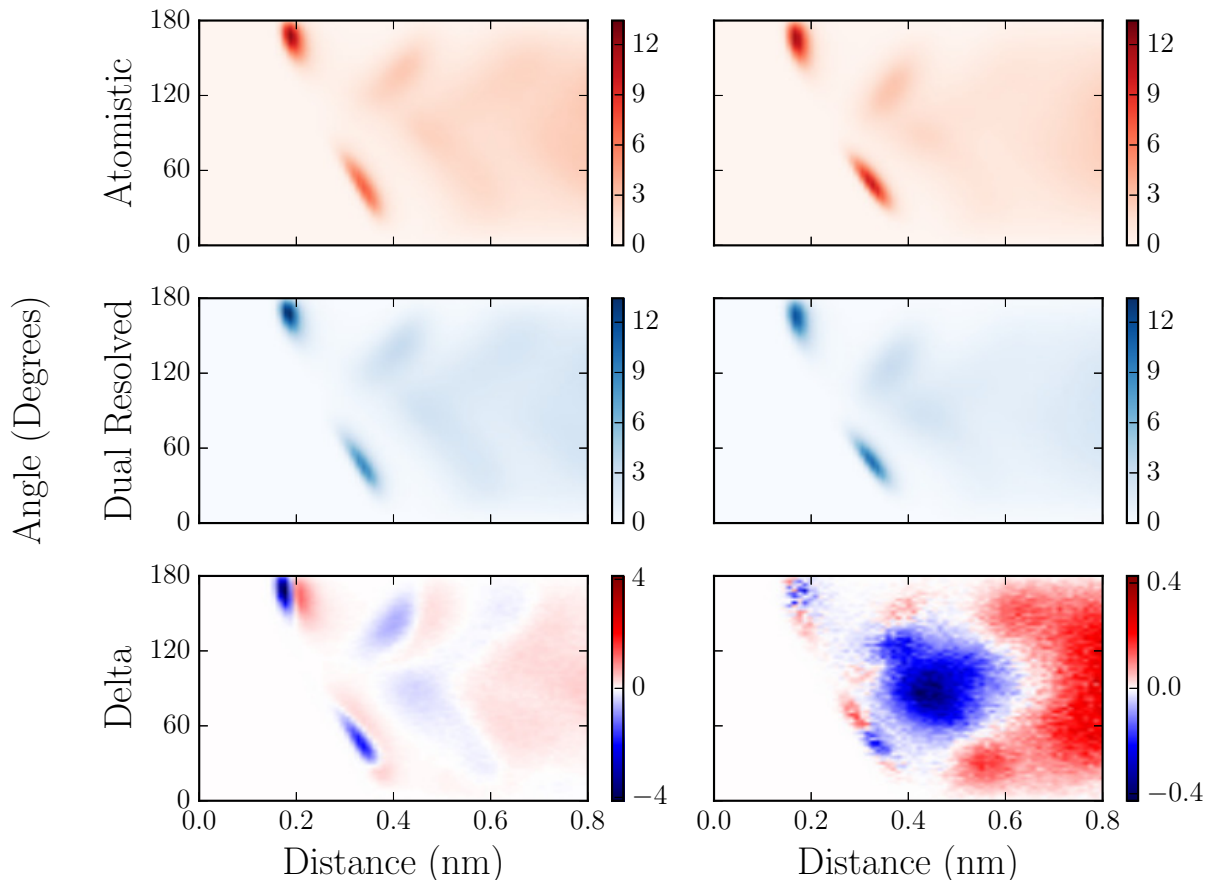
To compare the correctness of the atomistic bonds in the dual resolved case, the average C-C-C-C dihedral angle was measured, as shown in fig. 3. Both the dual resolution models show good agreement with the atomistic model results, with the UA model being the worse of the two. The UA model has a slightly lower population of the trans configuration, indicating that molecules are slightly more coiled compared to the reference atomistic version. The calculated densities are  $828(1) \text{ kg/m}^3$  and  $825(1) \text{ kg/m}^3$  for drAA and drUA respectively where the number in parenthesis represents the error on the last digit. Both models satisfactorily reproduce the experimental density of the octanol,  $822 \text{ kg/m}^3$  in the conditions considered here [69].



**Fig. 3:** Comparison of the backbone dihedral angles across the different models tested: aUA (green dashed line), drUA (green  $\circ$ ), aAA (red solid line), drAA (red  $\diamond$ ).  $180^\circ$  represents a trans configuration. Uses same legend as Figure 2

### 3.3. Hydrogen Bonding Details

The ability for the dual resolved model to correctly represent atomistic details is paramount in assessing its performance, and the most important atomistic detail present in the system is the hydrogen bonding. Hydrogen bonds are formed between a donor oxygen atom sharing its bonded hydrogen atom with an acceptor oxygen atom. These bonds can be characterised by their hydrogen-acceptor distance ( $r_{\text{OH}}$ ) and the O-H-O angle ( $\theta_{\text{OHO}}$ ), giving a probability as a function of these two quantities:  $P(\theta, r_{\text{OH}})$  to directly compare the atomistic and dual resolved versions of each model. These probabilities are plotted in fig. 4. Clearly visible in all models is the population of hydrogen bonds at just under 0.2 nm and around 170 degrees. At distances larger than 0.4 nm there is a repeating chevron pattern, which is evidence of the extended structure in the system caused by hydrogen bonding. Considering the differences between the dual resolved and the atomistic versions of each model, we can clearly see that the aAA model has slightly shorter hydrogen bonds than its dual-resolved counterpart. The corresponding dual resolved description obtained starting from the aUA model is able to more accurately recreate the hydrogen bonding network.



**Fig. 4:** Comparison of the angle and length of potential hydrogen bonds in dual resolved and atomistic models for the AA system (left column) and UA system (right column). The top panels show the atomistic results in red, the dual resolved results are below in blue, while the bottom panels show the difference between these two. Red areas indicate a higher probability in the atomistic model while blue areas indicate a higher probability in the dual resolved model. Note that the scale of the UA map in the bottom plot (Delta) is a tenth of the AA map.

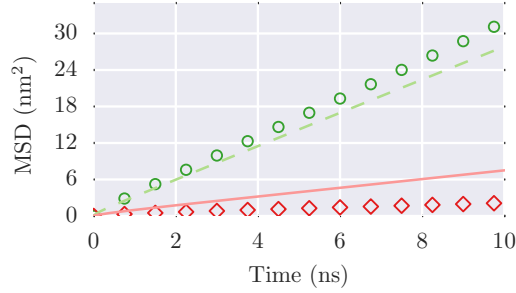
### 3.4. Analysis of the dynamic properties

The self-diffusivity of molecules,  $D$ , as defined according to the Einstein relation can be written as:

$$D = \lim_{t \rightarrow +\infty} \frac{\langle (\mathbf{r}(t) - \mathbf{r}(0))^2 \rangle}{6t} \quad (11)$$

where the angular bracket refers to the ensemble average. The squared quantity in previous equation is the Mean Squared Displacement of a molecule  $MSD(t) = (\mathbf{r}(t) - \mathbf{r}(0))^2$  and is plotted in fig. 5. The measured diffusivity was 0.012, 0.003, 0.47 and  $0.53 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  for the aAA, drAA, aUA and drUA models respectively. The difference in the diffusion coefficient between UA and AA was already reported in literature for the octanol with the OPLS force-field [70], where it was shown that the modification of the partial charges has a significant effect on the dynamics of the molecular models.

Typically, CG particles exhibit much faster diffusion than the corresponding atomistic ones, often by an order of magnitude. This effect is generally attributed to the flatter potential energy landscape the CG particles are moving through. Taking as comparison the diffusion coefficients of the two atomistic models analysed here, our results show that the drUA model has only a modest increase in the diffusion coefficient, while the drAA model is diffusing nearly four times slower. In order to have a better understanding

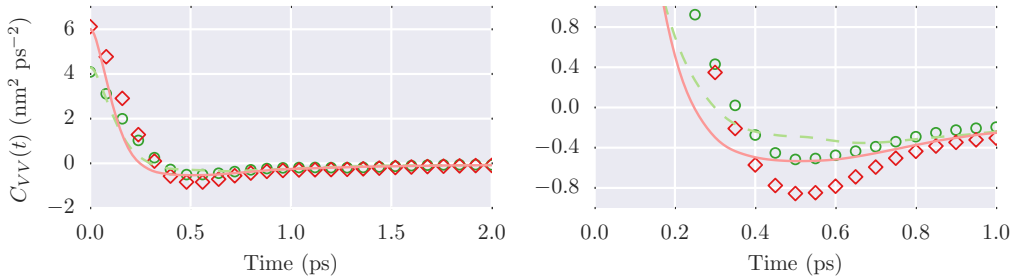


**Fig. 5:** Comparison of the self-diffusivity of octanol molecules across the different models tested: aUA (green dashed line), drUA (green  $\circ$ ), aAA (red solid line), drAA (red  $\diamond$ )

of this effect, we look at the rate at which the velocity of the molecules changes calculating the Velocity Autocorrelation Function, (VACF),  $C_V$ , defined as

$$C_V(t) = \langle \mathbf{v}_{CM}(t_0 + t) \cdot \mathbf{v}_{CM}(t_0) \rangle \quad (12)$$

where  $\mathbf{v}_{CM}$  is the velocity of the center of mass of the molecule and the average is with respect many starting time  $t_0$  and all molecules. The VACF was averaged over all the molecules, and is reported in fig. 6. The large negative well for the drAA model at around 0.5 ps is caused by the short-time vibrations backwards and forwards of the atoms and pseudo-beads that reverse their velocity at the end of each oscillation before diffusing. The drAA model shows the deepest negative well related to this backscattering. These oscillations are usually more prominent in systems with strong intermolecular interactions and that diffuse more slowly. If we look at the RDFs reported in Fig. 2 we can notice that the pseudo-beads of the alkyl chain (Y and Z) are consistently slightly more packed in the drAA model than in the other models (e.g. the first peak of the corresponding RDFs is slightly shifted at shorter distances and show higher intensity). Moreover, the density of the drAA model is the highest among the models presented here (see section 3.2). This effect could explain the reduced diffusion of the drAA model compared to the other ones. However, we want to stress here that the difference in the diffusion coefficients is not as dramatic as for standard mono-resolved atomistic models.



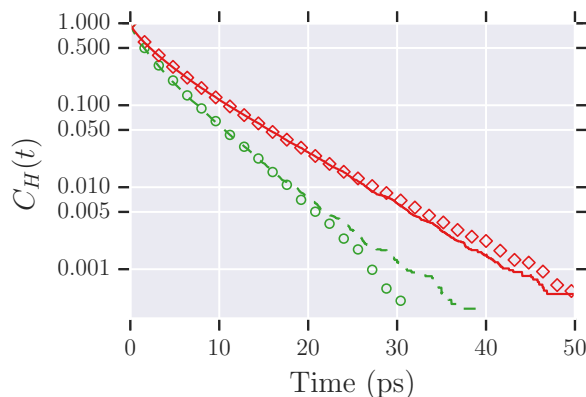
**Fig. 6:** On the left: Velocity Autocorrelation Function (VACF) across the different models tested: aUA (green dashed line), drUA (green  $\circ$ ), aAA (red solid line), drAA (red  $\diamond$ ). On the right: detail of the VACF in the region 0.2 – 1 ps

### 3.5. Hydrogen bond lifetimes

To evaluate the dynamics on a shorter length-scale, the rate at which hydrogen bonds break can be measured. The time autocorrelation of the hydrogen bonds,  $C_H(t)$ , is defined as

$$C_H(t) = \left\langle \frac{\sum_{ij} h_{ij}(t_0)h_{ij}(t_0 + t)}{\sum_{ij} h_{ij}(t_0)^2} \right\rangle \quad (13)$$

where  $h_{ij}(t)$  represents the population of hydrogen bonds at time  $t$ . For the continuous definition of lifetime, once a given hydrogen bond has broken it cannot later reform. Hydrogen bonds were geometrically defined between a donor oxygen, hydrogen and an acceptor oxygen as having a hydrogen acceptor distance of less than 0.3 nm and an O-H-O angle of greater than 130 degrees [71].



**Fig. 7:** Comparison of the continuous hydrogen bond lifetime correlation across the different models tested: aUA (green dashed line), drUA (green  $\circ$ ), aAA (red solid line), drAA (red  $\diamond$ ). Note that the y axis is log scaled.

The time evolution of hydrogen population was extracted by using trajectories of 250 ps, with trajectory frames being saved every 0.01 ps. The lifetimes of hydrogen bonds were then measured and plotted in fig. 7 and from the latter hydrogen bond lifetime,  $\tau_H$ , can then be calculated through numerical integration as:

$$\tau_H = \int_0^{\infty} (C_H(t) - C_H(t \rightarrow \infty)) dt. \quad (14)$$

This yields a value of lifetimes of 4.26 and 4.32 ps for aAA and drAA models respectively, and 2.93 and 2.97 ps for aUA and drUA models respectively. Unlike in previous work [42] where both the small and large scale dynamics had been affected by coarse-graining an atomistic model, here the lifetimes show no significant difference.

## 4. Computational cost of dual resolved models

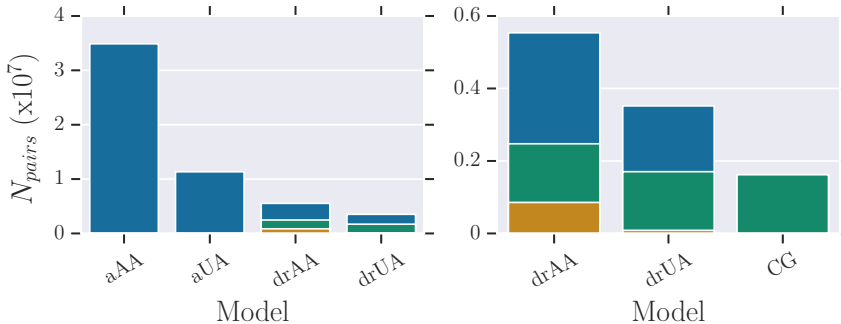
Now that the ability of these models to recreate atomistic results has been discussed, attention can be turned to their computational performance. As stated in the introduction of this paper, one of the main motivations of building this type of models is to reduce the computational cost of performing simulations.

The computational performances of a coarse-grained model was investigated in [72] where the authors consider three contributions: the ratio between atomistic and coarse-grained number of pair interactions, the speed up of the dynamics in coarse-grained models and the increase in the time step achievable in CG simulations. These three effects must be considered carefully in dual-resolved models because: *i*) the time step in a dual-resolved model is the same of the atomistic one (i.e. the smallest degree of freedom is still the vibration of atoms), *ii*) the speed-up of the dynamics is reduced with respect to full CG simulations

due again to the presence of the atoms, which slows down the dynamics of the beads [37]. Therefore, in our case the gain in computational time is dominated only by the reduction of the number of pair interactions in the system. Thus, the most straightforward metric to estimate the difference in computational cost is to evaluate the difference between the total number of pairwise interactions,  $N_{pairs}$  that must be considered in both simulations. The latter can be estimated as

$$N_{pairs} \approx \frac{4}{3}\pi (M^{FG} \rho_{FG} r_{cut,FG}^3 + M^{VS} \rho_{VS} r_{cut,VS}^3) + N_{intra}. \quad (15)$$

We note that this equation reduces to the relation considered in Eq. 8 in [72], when a single “type” of particle is considered, i.e. when there is a single cut-off in the system. In eq. (15),  $M$  refers to the number of particles of each scale in the system,  $\rho$  refers to the number density,  $r_{cut}$  is the interaction cutoff and  $N_{intra}$  is the number of intramolecular pairwise interactions. Here, we assume that the distribution of particles in the system is uniform. With pairwise interactions taking the majority of time in performing simulations [49],  $N_{pairs}$  can act as a rough estimate of the computational cost of a model. The result of eq. (15) is shown



**Fig. 8:** Comparison of the estimated number of nonbonded pairs in each model with the full atomistic model (left) and the full coarse-grained model (right). Atomistic interactions are shown on top in blue, coarse interactions below in green and finally intramolecular atomistic pairs in orange.

in fig. 8, and immediately becomes clear the reduced computational cost of the dual resolved versions for both the all-atoms and united-atoms models. The intramolecular pairs are only visible in the drAA model, and represent 6.2% and 2.6% of the total number of pairs for the AA and UA models respectively. Finally, the right panel of fig. 8 makes a comparison against a pure CG model, showing that the AA and UA models are roughly three times more computationally expensive.

The better performances of dual resolved models with respect to their fine-grained versions, in terms of theoretical scaling behaviour, come from the obvious fact that the former has less pair interactions than the latter. However, the performances of dual resolved models can be increased even further. Once the position of the virtual sites has been determined, the pairwise interactions in the fine and coarse-grained domains can be solved independently. This opens up extra possibilities for parallelisation through solving the two domains on different compute nodes, similar to how PME electrostatics are treated [73]. Some care is required however, as depending on the relative numbers of coarse and fine interactions, this may result in load imbalance between the two domains. Another possibility is the use of a multiple time step scheme [74, 75], where the interactions between VS could be evaluated at a reduced frequency, and indeed this has been shown to be possible for dual resolved models [38].

It is important to point out that the analysis on the speed-up should also take into account other aspects than computational cost, as the reduction in the number of degrees of freedom improves also the efficiency in exploring the conformational space as shown in [42]. However, the analysis presented here represents a good estimate as a lower bound evaluation of the computational efficiency.

## 5. Conclusions

We have presented a framework for constructing mixed resolution models to be simulated with the Gromacs molecular dynamics package but that can be easily extended to other molecular simulation packages which could handle virtual sites. The procedure was applied to two atomistic models of the octanol, united and all-atom. The IBI procedure was conducted including the pressure correction iterations. Overall, the resulting structural agreement between the dual resolved and atomistic versions of each model were satisfactory. The hydrogen bonding structures were also quite well represented in the dual resolved models, with the drUA showing better agreement with its corresponding fine grained version. The structural results are really promising, showing that the combination of IBI and atomistic parameters can be easily combined together, paving the way for application in more complicated systems. Short scale dynamic behaviour was correctly captured in both models, however bulk diffusion was well reproduced only for one of the dual-resolved model. The hybrid model developed from the all atoms force field showed smaller diffusion coefficient than the fully atomistic counterpart. This (fairly small) discrepancy was attributed to the higher density of the drAA system.

Overall the procedure proposed here is relatively easy (the coarse-grained interactions are developed directly using the atomistic model) and preserves the full atomistic structure of the system (including structure and dynamics of the hydrogen bond network) saving in computational cost at least 50% of the computational time.

## Acknowledgements

The authors would like to acknowledge the assistance given by IT Services and the use of the Computational Shared Facility at The University of Manchester.

## References

- [1] N. Di Pasquale, T. Hudson, and M. Icardi, *Physical Review E* **99**, 013303 (2019).
- [2] R. L. Henderson, *Physics Letters* **49A**, 197 (1974).
- [3] D. Reith, M. Putz, and F. Müller-Plathe, *J. Comput. Chem.* **24**, 1624 (2003).
- [4] T. C. Moore, C. R. Iacovella, and C. McCabe, *The Journal of chemical physics* **140**, 06B606.1 (2014).
- [5] P. Carbone and C. Avendaño, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 62 (2014).
- [6] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, *Soft Materials* **1**, 121 (2002).
- [7] W. G. Noid, J. Chu, G. A. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, *Journal of Chemical Physics* **128**, 244114 (2008).
- [8] M. Scott Shell, *Journal of Chemical Physics* **129**, 144108 (2008).
- [9] W. G. Noid, *The Journal of chemical physics* **139**, 09B201.1 (2013).
- [10] W. Shinoda, W. Devane, and M. L. Klein, *Molecular Simulation* **33**, 27 (2007).
- [11] S. J. Marrink, A. H. De Vries, and A. E. Mark, *The Journal of Physical Chemistry B* **108**, 750 (2004).
- [12] C. Avendaño, T. Lafitte, A. Galindo, C. S. Adjiman, G. Jackson, and E. A. Müller, *The Journal of Physical Chemistry B* **115**, 11154 (2011).
- [13] T. Bereau and K. Kremer, *Journal of chemical theory and computation* **11**, 2783 (2015).
- [14] Y. Qi, H. I. Ingólfsson, X. Cheng, J. Lee, S. J. Marrink, and W. Im, *Journal of chemical theory and computation* **11**, 4486 (2015).
- [15] V. Rühle and C. Junghans, *Macromolecular Theory and Simulations* **20**, 472 (2011).
- [16] A. Mirzoev and A. P. Lyubartsev, *Journal of chemical theory and computation* **9**, 1512 (2013).
- [17] P. A. Golubkov, J. C. Wu, and P. Ren, *Physical chemistry chemical Physics* **10**, 2050 (2008).
- [18] D. De Sancho and A. Rey, *Journal of computational chemistry* **28**, 1187 (2007).
- [19] S. Auer and D. Kashchiev, *Physical review letters* **104**, 168105 (2010).
- [20] S. Izvekov and G. A. Voth, *Journal of Chemical Physics* **123**, 134105 (2005).
- [21] Z. Wu, Q. Cui, and A. Yethiraj, *The Journal of Physical Chemistry B* **114**, 10524 (2010).
- [22] T. Shen and S. Gnanakaran, *Biophysical journal* **96**, 3032 (2009).
- [23] A. Milani, M. Casalegno, C. Castiglioni, and G. Raos, *Macromolecular Theory and Simulations* **20**, 305 (2011).
- [24] C. Peter and K. Kremer, *Soft Matter* **5**, 4357 (2009).
- [25] M. Bulacu, N. Goga, W. Zhao, G. Rossi, L. Monticelli, X. Periole, D. P. Tieleman, and S. J. Marrink, *Journal of chemical theory and computation* **9**, 3282 (2013).
- [26] O. Bezkorovaynaya, A. Lukyanov, K. Kremer, and C. Peter, *Journal of computational chemistry* **33**, 937 (2012).
- [27] V. Harmandaris, N. Adhikari, N. van der Vegt, and K. Kremer, *Macromolecules* **39**, 6708 (2006).
- [28] C. Moore, T. C. and Iacovella, R. Hartkamp, A. L. Bunge, and M. C., *The Journal of Physical Chemistry B* **120**, 99449958 (2016).

- [29] T. Tarenzi, V. Calandrini, R. Potestio, A. Giorgetti, and P. Carloni, *Journal of chemical theory and computation* **13**, 5647 (2017).
- [30] A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. De Vries, and S. J. Marrink, *Journal of computational chemistry* **31**, 1333 (2010).
- [31] X. Chen, P. Carbone, G. Santangelo, A. Di Matteo, G. Milano, and F. Müller-Plathe, *Physical Chemistry Chemical Physics* **11**, 1977 (2009).
- [32] M. Orsi, W. Ding, and M. Palaiokostas, *Journal of Chemical Theory and Computation* (2014).
- [33] M. Orsi, *Molecular Physics* **112**, 1566 (2014).
- [34] Q. Shi, S. Izvekov, and G. A. Voth, *J. Phys. Chem. B* **110**, 15045 (2006).
- [35] A. Rzepiela, M. Louhivori, C. Peter, and S. J. Marrink, *Phys. Chem. Chem. Phys.* **13**, 10437 (2011).
- [36] N. Di Pasquale, D. Marchisio, and P. Carbone, *Journal of Chemical Physics* **137**, 164111 (2012).
- [37] N. Di Pasquale and P. Carbone, *Journal of Chemical Physics* **146**, 084905 (2017).
- [38] N. Di Pasquale, R. J. Gowers, and P. Carbone, *Journal of Computational Chemistry* **35**, 1199 (2014).
- [39] J. Jin, Y. Han, and G. A. Voth, *The Journal of chemical physics* **150**, 154103 (2019).
- [40] A. Chaimovich, C. Peter, and K. Kremer, *The Journal of chemical physics* **143**, 243107 (2015).
- [41] T. A. Wassenaar, H. I. Ingólfsson, M. Prieß, S. J. Marrink, and L. V. Schäfer, *The Journal of Physical Chemistry B* **117**, 3516 (2013).
- [42] R. J. Gowers and P. Carbone, *The Journal of Chemical Physics* **142**, 224907 (2015).
- [43] M. Praprotnik, L. Delle Site, and K. Kremer, *Journal of Chemical Physics* **123**, 224106 (14 pages) (2005).
- [44] M. Praprotnik, L. Delle Site, and K. Kremer, *Physical Review E* **73**, 066701 (2006).
- [45] M. Praprotnik, L. Delle Site, and K. Kremer, *Annu. Rev. Phys. Chem.* **59**, 545 (2008).
- [46] A. Bahmani, S. Saaidpour, and A. Rostami, *Scientific reports* **7**, 5760 (2017).
- [47] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, 1987).
- [48] J.-P. Hansen and I. R. McDonald, *Theory of simple liquids* (Elsevier, 1990).
- [49] K. B. Tarmyshov and F. Müller-Plathe, *Journal of chemical information and modeling* **45**, 1943 (2005).
- [50] S. Plimpton, *Journal of computational physics* **117**, 1 (1995).
- [51] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, *J. Am. Chem. Soc.* **118**, 11225 (1996).
- [52] Z. Cao and G. A. Voth, *The Journal of chemical physics* **143**, 243116 (2015).
- [53] V. Ruhle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko, *Journal of chemical theory and computation* **5**, 3211 (2009).
- [54] S. Bevc, C. Junghans, and M. Praprotnik, *Journal of computational chemistry* **36**, 467 (2015).
- [55] P. Carbone, H. Karimi-Varzaneh, X. Y. Chen, and F. Müller-Plathe, *Journal of Chemical Physics* **128**, 64904 (2008).
- [56] H. Wang, C. Junghans, and K. Kremer, *Eur. Phys. J E Soft Matter* **28**, 221 (2009).
- [57] H. J. Berendsen, D. van der Spoel, and R. van Drunen, *Computer physics communications* **91**, 43 (1995).
- [58] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *SoftwareX* **1**, 19 (2015).
- [59] M. E. Siwko, *Disturb or stabilise? Effects of different molecules on biological membranes*, Ph.D. thesis, University of Groningen (2008).
- [60] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, *Journal of Physical Chemistry B* **105**, 6474 (2001).
- [61] J. Domański, P. J. Stansfeld, M. S. Sansom, and O. Beckstein, *The Journal of membrane biology* **236**, 255 (2010).
- [62] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *The Journal of chemical physics* **103**, 8577 (1995).
- [63] H. Berendsen, J. P. M. Postma, W. F. V. Gunsteren, A. Di Nola, and J. R. Haak, *J. Chem. Phys.* **81**, 3684 (1984).
- [64] G. Bussi, D. Donadio, and M. Parrinello, *The Journal of chemical physics* **126**, 014101 (2007).
- [65] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein, *Journal of Computational Chemistry* **32**, 2319 (2011), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21787>.
- [66] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domanski, D. L. Dotson, S. Buchouxk, I. M. Kenney, and O. Beckstein, *Proceedings of the 15th Python in Science Conference (SCIPY 2016)* (2016).
- [67] D. L. Dotson, S. L. Seyler, M. Linke, R. J. Gowers, and O. Beckstein, in *Proceedings of the 15th Python in Science Conference*, edited by Sebastian Benthall and Scott Rostrup (2016) pp. 51 – 56.
- [68] J. D. Hunter, *Computing In Science & Engineering* **9**, 90 (2007).
- [69] A. S. Al-Jimaz, J. A. Al-Kandary, and A.-H. M. Abdul-Latif, *Fluid Phase Equilibria* **218**, 247 (2004).
- [70] T. Kulschewski and J. Pleiss, *Molecular Simulation* **39**, 754 (2013).
- [71] D. Rapaport, *Molecular Physics* **50**, 1151 (1983).
- [72] M. Dinpajoo and M. G. Guenza, *Soft matter* **14**, 7126 (2018).
- [73] S. Páll, M. J. Abraham, C. Kutzner, B. Hess, and E. Lindahl, in *Markidis S., Laure E. (eds) Solving Software Challenges for Exascale. EASC 2014. Lecture Notes in Computer Science*, Vol. 8759 (2015).
- [74] D. D. Humphreys, R. A. Friesner, and B. J. Berne, *The Journal of Physical Chemistry* **98**, 6885 (1994).
- [75] H. Grubmüller and P. Tavan, *Journal of computational chemistry* **19**, 1534 (1998).
- [76] S. Izvekov and G. Voth, *Journal of Chemical Theory and Computation* **2**, 637 (2006).