Chapter 4

Hybrid Simulations: Atomistic Molecules in Coarse-Grained Solvent

This chapter is based on the manuscript:


Klapaucjusz myślał, myślał, wreszcie zmarszczył się i rzekł:
– Dobrze. Niech będzie o miłości i śmierci, ale wszystko to musi być wyrażone językiem wyższej matematyki, a zwłaszcza algebry tensorów. Może być również wyższa topologia i analiza. A przy tym erotycznie silne, nawet zuchwale, i w sferach cybernetycznych.

Klapaucjusz thought, thought, and finally said frowning:
- Right. Let it be about love and death, but all this must be expressed in the language of higher mathematics, especially algebra tensors. It may also be higher topology and analysis. And at the same moment sexually powerful, even audacious, and in the areas of cyberspace.

---

Cyberiada, Wyprawa Pierwsza A, czyli Elektrybał Trurla, Stanisław Lem
4.1 Abstract

Hybrid simulations, in which part of the system is represented at atomic resolution and the remaining part at a reduced, coarse-grained, level offer a powerful way to combine the accuracy associated with the atomistic force fields to the sampling speed obtained with coarse-grained (CG) potentials. In this work we introduce a straightforward scheme to perform hybrid simulations, making use of virtual sites to couple the two levels of resolution. With the help of these virtual sites interactions between molecules at different levels of resolution, i.e. between CG and atomistic molecules, are treated the same way as the pure CG–CG interactions. To test our method, we combine the Gromos atomistic force field with a number of coarse-grained potentials, obtained through several approaches that are designed to obtain CG potentials based on an existing atomistic model, namely iterative Boltzmann inversion, force matching, and a potential of mean force subtraction procedure (SB). We also explore the use of the MARTINI force field for the CG potential. A simple system, consisting of atomistic butane molecules dissolved in CG butane, is used to study the performance of our hybrid scheme. Based on the potentials of mean force for atomistic butane in CG solvent, and the properties of 1:1 mixtures of atomistic and CG butane which should exhibit ideal mixing behavior, we conclude that the MARTINI and SB potentials are particularly suited to be combined with the atomistic force field. The MARTINI potential is subsequently used to perform hybrid simulations of atomistic dialanine peptides in both CG butane and water. Compared to a fully atomistic description of the system, the hybrid description gives similar results provided that the dielectric screening of water is accounted for. Within the field of biomolecules, our method appears ideally suited to study e.g. protein-ligand binding, where the active site and ligand are modeled in atomistic detail and the rest of the protein, together with the solvent, is coarse-grained.

4.2 Introduction

Molecular dynamics (MD) is a well established technique to simulate the structure and dynamics of soft matter in general, and of biomolecular systems in particular. However, the huge computational effort involved in conventional all-atom (AA) MD simulations currently limits accessible time and length scales. To overcome these limitations, coarse-grained (CG) models, in which several atoms are grouped together into effective interaction sites (CG beads), are currently being developed by many groups (for an overview, see ). The sampling efficiency is increased up to several orders of magnitude with these CG models. Nevertheless, this large gain comes at the cost of a reduced accuracy,
4.2. **INTRODUCTION**

due to the inherent simplifications. Thus, in many instances it would be advantageous to combine high resolution AA and low resolution CG descriptions in a multiscale approach. Indeed, several of such hybrid schemes have recently been proposed, falling into four classes: back-mapping methods, interface methods, “constant λ” techniques and fixed resolution methods. A short overview of the development of multiscale techniques and numerous applications is presented in the Introduction chapter of this thesis. In this chapter we introduce a simple hybrid model, which belongs to the fourth, i.e. fixed resolution class of multiscale methods implying we consider a system in which a predefined part is represented in full detail, and the remainder at the CG level. The coupling is provided by the atomistic molecules which carry additional, virtual interaction sites, through which they interact with the CG surroundings using standard CG potentials. The simplicity is the main advantage of our model, making it possible to use it with generic CG and AA forcefields and standard simulation packages. Our approach is therefore more general and requires only few additional parameters. However, mixing of resolutions is not trivial. For instance, it remains to be seen how the interaction between two atomistic molecules is affected by the surrounding of a CG solvent compared to the "normal" atomistic surrounding. The expectation is that "multiscale" CG models, i.e. CG models that are directly derived from the atomistic model, are particularly suited for such a hybrid method. Here we test the performance of our hybrid method, combining the GROMOS 53a5/6 atomistic force field and several CG representations: either a heuristic MARTINI potential or "multiscale" CG potentials obtained using the iterative Boltzmann inversion method (IB), force matching (FM), or the subtraction method (SB). We study the performance of the model on two systems, namely liquid butane and dialanine dissolved in butane or water. We show that the simple hybrid model can connect the two resolution scales, most seamlessly with the SB method. Efficient resolution mixing is also observed with the MARTINI model.

The rest of this article is organized as follows. We set out to describe the methodological basis of the hybrid model. Subsequently, we introduce the principles of the applied methods to derive the CG potentials, and provide a detailed description of the simulation and analysis procedures. We proceed with a presentation of the results on the performance of the hybrid model, first for the pure butane system and next for the peptide/solvent systems. A short conclusive section ends this article.
4.3 Methods

4.3.1 Hybrid model

In the hybrid model, essential parts of the system, for example solute molecules, are represented at a detailed atomistic resolution, while the remaining elements, for example solvent molecules, are modeled using a lower, coarse-grained resolution. An example of a hybrid system consisting of atomistic butane solutes solvated by coarse-grained butanes is depicted in Figure 4.1. Both interactions between solvent–solvent pairs and between solvent–solute pairs of molecules are described with the potentials from a low resolution CG forcefield while the remaining interactions, that is interactions between atomistic molecules, are computed in full detail. To couple an atomistic molecule with a coarse-grained environment so called virtual sites are defined, positioned at the center of mass of the group of atoms they represent (cf. Figure 4.1). This positioning of the virtual sites corresponds to the mapping scheme between CG and atomistic models. The force $F_i$ from CG interactions acting on the virtual site $i$ is spread over the corresponding atoms, according to the equation

$$f_{\gamma} = \frac{m_{\gamma}}{M_i} F_i \quad \forall \gamma \in N_i,$$

(4.1)

where $M_i = \sum_{\gamma \in N_i} m_{\gamma}$ is the total mass of the atomic cluster $N_i$ constituting the virtual site, $m_{\gamma}$ is the mass of the $\gamma^{th}$ atom, and $f_{\gamma}$ is the force acting on it. Important features of our scheme are the following: i) no specific interactions between AA and CG beads need to be parameterized as is done in other boundary methods, making the method easily applicable to any system of interest, ii) no position dependent resolution transformation is required, therefore a well defined Hamiltonian exists without violating Newton’s 3rd law, two conditions not met with other, dynamic exchange, approaches, and iii) the method naturally combines the advantages of FG models (accurate description of the molecule, or part of molecule, of interest) and CG models (explicit treatment of the surroundings at a speed 2-3 orders of magnitude larger compared to AA models). Furthermore, the proposed method is general in the sense that it can, in principle, combine any two force fields provided that the molecule of interest is parametrized at both levels of resolution.

4.3.2 Coarse-grained potentials

The applicability of our hybrid model will depend on the compatibility between the CG and AA force fields. There is, however, no unique method to construct CG potentials from AA potentials. Even when the pair correlations are well described, other properties such as the pressure or solvation free energy...
Figure 4.1: Hybrid model using virtual sites. Coarse grained beads are represented as gray spheres. Atomistic molecules are depicted as yellow sticks and virtual interaction sites are colored red. Atomistic molecules interact with the low resolution environment through virtual sites and with each other as in a standard atomistic system. Coarse grain beads interact both with each other and with the atomistic components via CG potentials.

will be off. In this work we tested four different methods to construct the CG potentials: the iterative Boltzmann (IB) inversion method\(^{22,20,24}\) the force matching (FM) method\(^{17,19}\), a modification of the subtraction procedure (SB) originally described by Villa et al.\(^{132}\), and the MARTINI force field\(^{5,33}\). In each case, the resolution of the coarse-grained models is given by a 4 to 1 center of mass based mapping principle, in analogy to the standard mapping scheme used in the MARTINI model. A pressure correction is applied to the potentials to match the pressure of the reference atomistic system. Although the relation between pressure on the CG and AA level is non-trivial (for instance, one cannot simultaneously reproduce both pressure and compressibility\(^{133,27,134}\)), we decide here on the pressure, in a similar way the adaptive resolution methods do, because we want to combine atomistic and CG models and they should have the same pressure. Details of the IB, FM methods and MARTINI model are presented in the Introduction chapter of the thesis. The subtraction procedure is described here, since modification that we introduced to the method allowed us to obtain potentials useful for the multiscale hybrid model and this modified SB procedure is considered as one of the results of this work.

The subtraction method  To overcome sampling problems in dilute solutions, where \(g(r)\) for solute–solute pairs converges very slowly, a new method to construct CG potentials was introduced recently by Villa and coworkers.\(^{132,135}\) In
this approach, which we will call here the subtraction (SB) method, the PMF between two solute molecules $V_{PMF}^{AA}$ in a solvent box is calculated. Then, to remove the solvent contributions from the PMF the calculation is rerun without direct solute–solute interactions to obtain a correcting potential $V_{PMF, excl}$. This $V_{PMF, excl}$ is subtracted from the $V_{PMF}^{AA}$ and an effective two body solute–solute CG interaction is obtained:

$$V_{SB}^{CG} = V_{PMF}^{AA} - V_{PMF, excl}.$$  (4.2)

In the original SB method by Villa et al. this is done by computing $V_{PMF, excl}^{CG}$ on the CG level. This means a PMF between two CG solute molecules is computed where the direct solute–solute interaction is turned off, while the interactions between the solute particles and the surrounding solvent and the solvent–solvent interactions are CG interactions that had previously been determined, for example by iterative Boltzmann inversion. Note that this subtraction method is not limited to IB. It just means one assumes a set of CG interactions in the environment of the solutes and subtracts the PMF between two noninteracting solutes in this environment from the atomistic target PMF. The resulting $V_{SB}^{CG}$ is an effective solute–solute pair interaction which if added to the set of CG environmental interactions (solvent–solvent and solute–solvent) reproduces the target solute–solute PMF (at infinite dilution). In the present case the subtraction method is modified to target at a CG–CG interaction that is particularly suited in the present hybrid approach. Therefore one subtracts the PMF between two noninteracting solute molecules in atomistic environment from the atomistic target PMF.

$$V_{SB}^{CG} = V_{PMF}^{AA} - V_{PMF, excl}^{AA}.$$ (4.3)

Since this means one subtracts the atomistic multibody effects of the environment from the target PMF, this modification of the procedure permits to obtain an effective two body $V_{SB}^{CG}$ CG interaction in an otherwise atomistic medium.

### 4.3.3 Simulation setup

To construct the CG butane–butane potentials (see below), reference simulations of pure butane were performed at the AA level. This system consisted of 750 butane molecules. Corresponding CG simulations were carried out in systems of the same size. To test the hybrid model, simulations were performed for two types of systems, namely a pure butane system and a mixture of dialanine and either water or butane. For the butane system, the set-up consisted of i) two AA butanes dissolved in a box of 748 CG butane (calculation of association constants) or ii) a box consisting of 375 AA butanes and 375 CG butane molecules (calculation of solvation parameters). For the dialanine containing systems, phase
separation was studied in a box containing 370 AA dialanine molecules mixed with 1120 CG solvent beads. For the calculation of the interaction energy between two dialanine peptides, a system consisting of two dialanines and 1050 solvent beads was additionally simulated. Fully atomistic systems were also simulated for reference.

All simulations were performed using the GROMACS simulation package (version 4.0.5). Atomistic simulations and atomistic parts of hybrid systems were represented with the 53a5/6 parameter set of the GROMOS forcefield. For the pure butane systems, different CG potentials were used as explained below. The systems containing dialanine peptides were only simulated using the MARTINI CG potentials (version 2.0). All of the simulations were performed in a periodic cubic box with dimensions longer than twice the cut-off distance. A cut-off distance for nonbonded interactions was set at 1.4 nm for pure butane systems and 1.2 nm (MARTINI potentials) or 1.4 nm (AA, atomistic part of mixed system) for two component solvent–alanine systems. For the MARTINI potential the standard shift function was applied from 0.9 nm or 1.2 nm for cut-off values 1.2 and 1.4 nm respectively. For the hybrid two-component systems, the electrostatic potential was shifted between 0 and the cut-off distance, and a relative screening with $\varepsilon_r = 15$ was applied in line with the standard screening used in MARTINI. The effect of changing $\varepsilon_r$ on the behavior of these systems was also evaluated.

In all simulations a leapfrog integrator with 2 fs time step was used with a neighbor list updated every 5 steps. Bond distances in atomistic molecules were constrained using the LINCS algorithm. Stochastic temperature coupling, with a time constant of 1.0 ps and a reference temperature of 270 K (butane systems) or 320 K (systems containing water), was applied. A Berendsen barostat was used to equilibrate atomistic and hybrid two component systems at 1 bar before constant volume production runs. Most of the simulations were between 40-100 ns long. To equilibrate the two component systems consisting of butane and dialanine molecules, simulated annealing simulations (SA) were performed before the production runs. During the SA, the temperature was decreased linearly from 600 to 270 K over a period of 100 ns.

**Construction of CG potentials** To calculate IB and FM potentials the VOTCA package was used. First we performed 100 ns long atomistic simulation of about 750 butane molecules under 1 bar pressure in temperature of 270 K. From this simulation reference molecular radial distribution functions and forces were used to generate numerically tabulated potentials with IB and FM method, respectively. In the IB procedure a 200 ps coarse-grained simulation was performed at each iteration. The pressure correction was applied.
Figure 4.2: Root mean square deviation of $g(r)$ and $g_{ref}(r)$ in function of number of iterations.

at every cycle and the convergence (Figure 4.2) was measured as a root mean square deviation $\sqrt{\frac{1}{n} \sum_{i=1}^{n} |g(r_i) - g_{ref}(r_i)|}$ (where summation runs over all distance bins and $n$ is the number of bins) between reference and calculated radial distribution function. After 175 iterations the obtained potential was additionally pressure corrected and used in the subsequent calculations. For each force match calculation 35 frames from the reference simulation were used. The final tabulated force function was calculated as an average of all functions fitted to the fragments of the atomistic trajectory. As the last step a pressure correction of the form presented in Equation 1.9 was iteratively applied, to tune the pressure to the reference atomistic value. Strictly speaking, this type of pressure correction is not formally correct in the case of force matching. Pressure correction for force matching does exist,\cite{27,28} it is however not implemented in the VOTCA package.

To construct the $V_{PMF, excl}^{AA}$ component of the SB potential, a constraint method was applied:

$$V_{PMF}(r) = -\int_{r}^{r_c} \left[ \langle f_c \rangle_s + \frac{2k_BT_s}{s} \right] ds, \quad (4.4)$$

where $\langle f_c \rangle$ is an average force exerted on two particles separated by a fixed distance $s$ and $r_c$ is the interaction cut-off. The second term in the integral removes entropic contributions, that are different for consecutive distances because of larger volumes sampled at larger separations. The LINCS method\cite{124} was used to fix the distances between the two evaluated molecules. The separation for constrained molecules was varied with 0.02 nm increment, from 0.22 nm up to $r_c$. The $V_{PMF}^{AA}$ component was obtained by simply inverting the RDF of the fully atomistic system according to Equation 2. The resulting $V_{SB}^{\text{CG}}$ potential was addi-
tionally pressure corrected to the reference pressure of 1 bar, using the iterative pressure correction procedure (Equation 1.9).

For the current aim of using the MARTINI model in hybrid simulations, the MARTINI potential for the liquid butane system was modified slightly to better represent the molecular \( g_{\text{ref}}(r) \) from the reference atomistic simulation. Particularly, the bead size was increased by 0.25 Å to match the position of the first peak in the \( g_{\text{ref}}(r) \). Subsequently the value of \( \varepsilon \) was increased by 5% to match the atomistic pressure. The modified potential will be named MARTINI* in the rest of the paper and the standard interaction, used in the peptide systems, will be referred to without the “*” sign.

### 4.3.4 Analysis details

**Kirkwood-Buff integrals:** To analyze the behavior of the systems in the hybrid representation we analyzed the preferential solvation parameters \( \delta_{CG} = G_{CG,CG} - G_{CG,AA} \) and \( \delta_{AA} = G_{AA,AA} - G_{CG,AA} \). These parameters are calculated using Kirkwood-Buff integrals\(^{137}\):

\[
G_{i,j} = 4\pi \int_0^{R_{\text{cut}}} [g_{i,j}(r) - 1] r^2 dr, \tag{4.5}
\]

where \( g_{i,j}(r) \) is the radial distribution function between the centers of mass of molecule \( i \) and \( j \) and \( R_{\text{cut}} \) is the distance beyond which the integrand vanishes. The calculated quantities are used to describe the affinities between molecules modeled at different resolutions and to predict (micro)phase separation.

**Association constants:** In order to assess strength of interactions between atomistic molecules in the hybrid model we calculate the association constant \( K_a \), which are directly linked to the above Kirkwood-Buff integrals. \( K_a \) is given as

\[
K_a = 4\pi \int_0^{R_{\text{cut}}} \exp \left( \frac{-V_{\text{PMF}}}{k_B T} \right) r^2 dr, \tag{4.6}
\]

with \( V_{\text{PMF}} \) denoting the potential of mean force between the two molecules and \( R_{\text{cut}} \) being a specified cut-off distance.

**Solvation free energy:** Comparison of thermodynamic properties of the constructed CG potentials was assessed via the free energy of solvation \( \Delta G \), which we calculated using the thermodynamic integration\(^{138}\) procedure. For each calculation 21 simulations of 1 ns length at equally spaced \( \lambda \) values between 0 and 1 were performed. The last 800 ps were used for analysis. A soft core method was used to remove singularities during the calculations. Integration over the \( \lambda \) variable was performed using a trapezoidal rule.
CHAPTER 4. FIXED RESOLUTION HYBRID METHOD

4.4 Results and Discussion

4.4.1 Comparing CG potentials for butane

To test the hybrid model we chose a simple system of liquid butane. The butane molecule, which is modeled as a chain of four united carbon atoms by the GROMOS forcefield, is represented by a single bead on the coarse-grained level. In Figure 4.3A CG potentials for the butane–butane interaction are shown. The CG potentials were obtained either with iterative Boltzmann inversion, force matching, or subtraction method, or with the MARTINI* model. Substantial differences between the potentials are apparent. Notably, IB and FM potentials are shallow and have significant interaction energy contributions from the second neighbors as evidenced by an additional repulsive region (FM) or long range attraction (IB). Furthermore both IB and FM potentials have a softer repulsive part at short distances compared to SB and particularly to the MARTINI* LJ potential. Both the MARTINI* and SB potentials are of comparable shape, especially considering the attractive region beyond 0.7 nm. MARTINI* butanes are however more repulsive at short distances and have a slightly deeper minimum. In spite of these discrepancies, IB, FM and SB potentials, when employed in a coarse-grained simulation of pure butane, generate a very similar $g(r)$ as shown in Figure 4.3B. Note that this result is not trivial as the different methods to construct the CG potentials, while relying on the same atomistic reference sampling, have different targets. Only IB was optimized to reproduce the atomistic RDF. FM is optimized to reproduce a multibody PMF, that means it targets at the pair RDF, but at the same time also at higher order correlation functions and depending on the "basis set" of CG interaction functions (in this case pair potentials) has to make compromises as far as the pair RDF is concerned. The SB methods targets indeed at the pair RDF (in the form of the pair PMF) but since we do not iterate and do not subtract $V_{IB,PMF,excl}^{IB,PMF,excl}$ but $V_{AA,PMF,excl}^{AA,PMF,excl}$ we do not by construction match the RDF. With the MARTINI* potential the phase in $g(r)$ is preserved but the amplitude is too high. This means that molecules are more ordered than in the reference atomistic liquid butane system, a consequence of using of a 12-6 Lennard-Jones interaction functional which results in a very steep repulsion at short distances.

Although structurally the IB, FM and SB potentials generate similar equilibrium ensembles, this does not imply that the thermodynamics is described equally well. To quantify this, we calculated the free energy of solvation $\Delta G$ (transfer of one molecule from vacuum into its solution) for each of the four models. In Table 4.1 the results are shown together with the average potential energies and results for the reference atomistic system. Regardless of the distinct $g(r)$ for MARTINI* with respect to the other potentials, $\Delta G$ for the MARTINI*, IB and SB potentials
4.4. RESULTS AND DISCUSSION

Figure 4.3: Coarse-grained butane–butane potentials. In panel A, CG potentials constructed with the four methods are shown, namely iterative Boltzmann (IB, red), force matching (FM, green), subtraction method (SB, orange), and MARTINI* model (blue). Panel B displays the molecular radial distribution functions for liquid butane, simulated with each of the four CG potentials. A reference radial distribution function from a fully atomistic simulation (black dashed line) is also presented.

is very similar and close to the reference atomistic $\Delta G_{\text{ref}}$. The $\Delta G$ obtained with the FM potential is substantially lower. As we will see later, this is a problem in hybrid simulations as it will result in preferential solvation effects.

### 4.4.2 PMFs for AA butane in CG solvent

In this section we examine the properties of hybrid systems simulated with the different CG potentials. To this aim we calculated potentials of mean force (PMFs) between two atomistic butane molecules solvated in coarse-grained bu-

<table>
<thead>
<tr>
<th>Representation</th>
<th>$\Delta G;(kJ/mol)$</th>
<th>$K_a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA, Gromos 53a5</td>
<td>-13.6 +/- 0.7</td>
<td>17.1</td>
</tr>
<tr>
<td>CG, IB</td>
<td>-11.3 +/- 0.7</td>
<td>18.6</td>
</tr>
<tr>
<td>CG, FM</td>
<td>-7.1 +/- 0.5</td>
<td>18.3</td>
</tr>
<tr>
<td>CG, MARTINI*</td>
<td>-11.6 +/- 1.6</td>
<td>16.6</td>
</tr>
<tr>
<td>CG, SB</td>
<td>-10.3 +/- 1.0</td>
<td>17.3</td>
</tr>
</tbody>
</table>

Table 4.1: Free energy of solvation, $\Delta G$, of a CG butane molecule dissolved into its own bulk. Results obtained with each of the four CG potentials are shown, together with the free energy for an AA system. Association constants $K_a$ of two atomistic butane molecules dissolved in CG butane are also shown.
tane. The constraint method, described in section 4.3.2, was used to compute the PMFs. The setup for each experiment was identical, except for the interactions between CG–CG and CG–AA pairs of molecules which were represented with either one of the four CG potentials (IB, FM, SB, and MARTINI*). Figure 4.4A shows the resulting PMFs including a reference PMF calculated in a fully atomistic butane system. Neither of the CG approaches yields a hybrid model, where the two atomistic molecules show the same pair interaction as if they were in a fully atomistic environment. Overall, the MARTINI* and SB PMFs resemble the reference PMF the most, although the attractive and repulsive regimes are more apparent in both cases. This effect is much more pronounced in case of the IB PMF. Atomistic molecules in the IB solvent are strongly attracted to each other at short distances but effectively repel themselves when outside the first solvation shell. The FM potential exhibits an enhanced attraction over the entire range of the calculated interaction. We can quantify the total interaction strength by means of the association constant $K_a$. Here we integrate the PMF up to a cutoff distance $R_{\text{cut}}$ of 1.6 nm to obtain $K_a$, using Equation 4.6. Results are summarized in Table 4.1. The association constants are similar for all four CG potentials. However, the calculated $K_a$ depends on the integration interval, and differences are larger when a smaller $R_{\text{cut}}$ is chosen (Figure 4.4B), showing largest deviations for the IB and FM potentials. Since the target of the hybrid approach is that the interaction strength between two atomistic molecules in the CG environment should reproduce the one in a fully atomistic system, it appears from these results that the SB and MARTINI* potentials are the most promising candidates for use in such a hybrid model. This assumption is further tested below.

4.4.3 Preferential solvation in mixed AA/CG butane

In order to confirm our observations from the potential of mean force calculations, we simulated liquid butane systems in which half of the butane molecules are represented as coarse-grained particles and half of them atomistically. Equilibrium ensembles were generated with Langevin dynamics as described in the Methods section. As before, we performed calculations for all four CG potentials. To assess the mixing of the AA and CG butane molecules we computed the radial distribution functions $g_{i,j}(r)$ between AA–AA, AA–CG, and CG–CG (i,j) butane pairs. The results are shown in Figure 4.5A-D. In the ideal scenario molecules do not know the resolution of their neighbors and therefore do not have any preference of one representation above the other. Since we use the same CG potential for CG–CG and AA–CG pairs we would expect identical $g_{i,j}(r)$. That is not the case for IB and FM systems, where we see different preferences in the first and second shell or clustering of atomistic molecules, respectively. In the IB system atomistic molecules, which strongly attract each other on a short distance,
favor neighbors of the same type. Interaction on the longer distance, however, is more favorable with the unlike type, since the CG potential is deeper in that position (cf. Fig. 3A). This causes formation of transient structures, e.g. atomistic butane wires or small clusters. The clustering is even more pronounced in the FM system due to lower solvation free energy of AA butane in CG butane than in itself (Table 4.1).

The clustering effect can be quantified with preferential solvation parameters $\delta_i$ that we calculate by integrating $g_{i,j}(r)$ using Eq. 4.5. In Table 4.2 the preferential solvation results are summarized. Indeed we can see an increased preferential solvation for alike molecules in the FM system. In the IB system this is not so clear, since Kirkwood-Buff integrals are calculated over three solvation shells and the differences are averaged out. Therefore we additionally computed preferential solvation parameters $\delta_{CG}^{R_1}$ using a cut-off $R_1 = 0.8$ nm. Large positive numbers for the first shell (and large negative for the rest) support the distance dependent solvation preferences discussed above. In the MARTINI and SB systems the $\delta_{AA}$ and $\delta_{CG}$ values are much closer to zero, implying that there is no significant preferential solvation. This is also clear from the overlapping RDFs for CG–CG and CG–AA pairs (Figure 4.5C,D).

In panel E of Figure 4.5 the $g_{AA,AA}(r)$ from IB, FM, SB and MARTINI* hybrid systems are compared to $g_{ref}(r)$ obtained from a fully atomistic simulation. The RMSD of all four $g_{AA,AA}(r)$ to $g_{ref}(r)$ is shown in Table 4.2. A larger de-
Figure 4.5: Even mixture of AA and CG molecules. A,B,C,D panels: molecular radial distribution functions $g_{AA,CG}(r)$ (long dashed line), $g_{AA,AA}(r)$ (dotted line) and $g_{CG,CG}(r)$ (full line) for IB, FM, MARTINI* and SB potentials, respectively. In panel E atomistic molecular radial distribution functions $g_{AA,AA}(r)$ are shown. Black line represents $g_{ref}(r)$, red line: CG component is described with iterative Boltzmann inversion, green line: force matching, orange line: subtraction method, blue line: MARTINI* model. In panel F carbon radial distribution functions are presented for atomistic part of SB hybrid model (long dashed line), MARTINI* (dotted line) and reference atomistic simulation (full line).
4.4. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>CG potential</th>
<th>$\delta^R_{AA}$</th>
<th>$\delta^R_{CG}$</th>
<th>$\delta^1_{AA}$</th>
<th>$\delta^1_{CG}$</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB</td>
<td>-118</td>
<td>-111</td>
<td>220</td>
<td>202</td>
<td>0.182</td>
</tr>
<tr>
<td>FM</td>
<td>902</td>
<td>684</td>
<td>348</td>
<td>274</td>
<td>0.109</td>
</tr>
<tr>
<td>SB</td>
<td>61</td>
<td>8</td>
<td>49</td>
<td>69</td>
<td>0.101</td>
</tr>
<tr>
<td>MARTINI*</td>
<td>-17</td>
<td>-32</td>
<td>-32</td>
<td>-8</td>
<td>0.099</td>
</tr>
</tbody>
</table>

Table 4.2: Preferential solvation parameters $\delta_i$ obtained from equal mixtures of AA and CG butane molecules. The CG part of the systems is represented with one of the four CG potentials. Results obtained with two cut-off distances, $R_c$ and $\frac{1}{2}R_c$, are presented. The last column shows the RMSD of the molecular $g(r)$ from the hybrid system to the molecular $g_{\text{ref}}(r)$ of a reference atomistic system.

violation is apparent for the IB potential, whereas the three other models show a similar agreement with the reference $g(r)$. Overall, the performance of the SB potential is considered best, since all $g_{i,j}(r)$ are very much alike (Figure 4.5D). Additionally all of them resemble $g_{\text{ref}}(r)$. This would make the SB potential the optimal choice for the hybrid model. Nevertheless, in the MARTINI* hybrid system, where $g_{CG,CG}(r)$ and $g_{AA,AA}(r)$ are different, we do not observe preferential solvation and the atomistic part of the system reproduces the reference molecular radial distribution function $g_{\text{ref}}(r)$. That suggests that the thermodynamic accuracy of the CG potential, more than the structural one, is essential for the proper mixing of resolutions.

Although the structural and thermodynamic analysis presented above indicates that the hybrid model performs well, provided that a suitable CG potential is used, our analysis so far has mainly concerned the behavior at the CG level. Whether the atomistic degrees of freedom are also unperturbed is investigated next. In Figure 4.5F, radial distribution functions for carbon atom pairs are presented, for the SB, MARTINI* hybrid systems as well as the purely atomistic reference system. We limited our analysis to the SB and MARTINI* potentials, since the performance of these potential in the hybrid model is best. The discussion, however, is general and concerns all four cases. Despite the similar $g_{AA,AA}(r)$ and $g_{\text{ref}}(r)$ at the CG level (cf. Figure 4.5E), there is a noticeable overstructuring of the carbon pair correlation functions. This effect we attribute to the partly missing interactions of the carbon atoms with their coarse-grained neighbors, which could lead to a change in the alignment of the butane chains in atomistic surrounding compared to the CG (spherical particle) surrounding. As can be seen in Figure 4.5F the effect is rather small; importantly the global structure is not affected (cf. Figure 4.5C,D).
4.4.4 Association of AA dialanine in CG solvents

In addition to the tests performed on pure liquid butane here we validate the mixed model on two component systems composed of dialanine dissolved either in liquid butane or in water. These uncommon systems were chosen here first because of higher complexity than the pure butane case: each time there are two CG bead types and two distinct CG interaction potentials. Moreover, Coulombic interactions are introduced between atomistic molecules. Second we would like to know, if the MARTINI forcefield, without any modifications, could be used as a default CG representation in hybrid simulations. Standard MARTINI potentials are used to represent effective interactions between butane-butane, butane-alanine and alanine-water pairs of molecules. The dialanine molecule is described by two beads on the CG level and therefore two virtual sites are constructed, each representing the center of mass of the single alanine aminoacid. The butane molecule, as previously, is represented with one CG bead. All interactions between dialanine molecules are computed atomistically. For the electrostatic interactions, it is not clear on the forehand whether or not a screening of the charges should be applied. In fully atomistic systems, such a screening is not necessary as the solvent performs the screening explicitly. In the MARTINI model, however, most solvents (including water) are modeled as a simple LJ fluid lacking electrostatic interactions. Coulombic interactions are therefore scaled with a uniform dielectric constant $\varepsilon = 15$ to account for the absence of explicit screening. Here we tested both approaches, i.e. either screening the atomistic peptide-peptide interactions with $\varepsilon_r=15$, or no ($\varepsilon_r=1$) or limited ($\varepsilon_r=2$) screening. For reference, fully atomistic systems were also simulated. See section 4.3.3 for details.

With butane as solvent, clear phase separation is observed between butane and dialanine, as illustrated in Figure 4.6A,B. The density of the peptides across the box is shown in Figure 4.6C. The width of the layer and the density of dialanine molecules inside the layer are similar between the hybrid and fully atomistic systems. Also the sharpness of the interface is comparable in both cases. The results shown in Figure 4.6A-C pertain the situation in which the electrostatic interactions were screened with $\varepsilon_r=15$; similar results were obtained in the absence of screening. Replacing the apolar butane solvent by a polar solvent like water, we also observe phase separation (Figure 4.6 D,E,F) in both atomistic and hybrid models. However, in this case phase separation is only observed with unscreened electrostatics and mixing occurs for $\varepsilon_r=15$ (not shown). Even with unscreened electrostatic interactions, the phase separation in the hybrid model is not as pronounced as in the fully atomistic system (Figure 4.6F). We also calculated the PMF for two atomistic dialanine peptides in CG solvent. Results are shown in Figure 4.7A for butane as a solvent, and 4.7B for water. In a qualitative sense,
Figure 4.6: Mixing of atomistic dialanine peptides in coarse-grained solvents. The solvent is either water (A,B,C) or butane (D,E,F). Snapshots from an atomistic simulation (A,D) are compared to MARTINI hybrid simulations (B,E). Butane is depicted as grey spheres, water as blue spheres, and the peptides with yellow carbon and red oxygen atoms. Panels C and F show the density profiles of the dialanine peptides for atomistic (full line) and hybrid (long dashed line) systems.
the hybrid model reproduces the effect of the solvent on the peptide-peptide interaction: an apolar medium like butane strongly promotes dimerization, whereas in the aqueous environment the interaction strength is at the level of $kT$. However, it turns out that the alanine-alanine interactions are too weak in the hybrid system when using the standard screening $\varepsilon=15$ of the MARTINI model. This is apparent from the comparison of the hybrid PMF to that of the fully atomistic reference system. In the case of butane, this screening is way too strong. Better performance is obtained lowering the dielectric constant to a value between 1 and 2 (Figure 4.7A). In case of water, lowering of the dielectric constant also improves the results with the best agreement to the reference potential at a value in the range between 4 to 6.

![Figure 4.7](image)

**Figure 4.7:** Potential of mean force between two atomistic dialanines in CG solvent using MARTINI. Results for butane are shown in panel A, for water in panel B, using long dashed lines. The effect of different screening constants $\varepsilon=1$, 2, and 15, is also included. The reference atomistic system is shown with a solid line.

### 4.5 Summary and Conclusions

In this work we designed and tested a straightforward hybrid model to bridge atomistic and coarse-grained representations of an arbitrary system. We focused our efforts on the coarse-grained description, assuming it to be a crucial element for the hybrid model to function. Three procedures to construct CG potentials were tested: iterative Boltzmann (IB) inversion, force matching (FM), and a subtraction (SB) method. In addition, the pre-parameterized MARTINI force field was used. We simulated two systems to explore the performance of the hybrid model: pure liquid butane molecules and dialanine peptides dissolved either in
an apolar solvent, butane, or a polar solvent, water.

Using the first system, i.e. pure butane, we showed that only with MARTINI and SB CG potentials both structural and thermodynamic properties of the hybrid system remain close to those of a fully atomistic reference system. For these potentials, no preferential solvation effects were observed implying that the molecules do not know their resolution, mixing equally well with both AA and CG neighbors. The potentials obtained with IB or FM are less effective, causing distinct preferential solvation effects. We do not claim that better potentials could not be obtained also with IB and FM approaches (for example by applying additional thermodynamic constraints during the parameterization), however, two ingredients of the CG potential appear crucial: i) reproducing the solvation free energy of the atomistic model, and ii) a correct balance between short range and long range forces. In case of the FM potential, the solvation free energy is too low, and in case of the IB potential an additional repulsive regime is present. Both lead to non-ideal mixing of the AA and CG molecules. An accurate representation of the structural pair correlations is not essential, as can be concluded from the good performance of the MARTINI potential. Small deviations of the intra-bead atomistic degrees of freedom is observed irrespective of the CG potential used. It seems unavoidable in this kind of hybrid simulations that the internal atomistic degrees of freedom are somewhat perturbed, especially for systems with a low density of AA particles. Butane may be a worst-case scenario, however, due to its small size. Overall, the best results are obtained with the SB method that was recently developed by Villa et al..\textsuperscript{132,135} The SB CG potentials show an accurate energy of solvation, correct representation of the pair correlation functions, and well-balanced overall shape of the potential, which results in the lack of preferential solvation effects in the hybrid simulations. Whether this is generally the case might depend on the complexity of the multibody interactions present in the system. An indication that butane is possibly less complicated than other systems is that IB, FM, and SB yield very similar pair RDFs. This would be very different, e.g., for water due to important 3-body interactions. Further tests on parameterizing the optimal CG potentials for our hybrid approach are clearly required.

Advantage of the use of the MARTINI force field is that MARTINI potentials are readily available for a large range of biomolecular systems.\textsuperscript{5,36,37} We used the second system, i.e. dialanine peptides mixed with either butane or water, to test how well the hybrid MARTINI model performs with more complicated cases involving electrostatics. From calculations of the PMF between two atomistic dialanine peptides in CG solvent, we conclude that the effect of the solvent is captured in a qualitative sense; the peptides are attractive in the apolar solvent and are rather neutral in polar solvent. Better quantitative agreement with the full
atomistic system is obtained when the dielectric constant is chosen as a free parameter. For the apolar butane solvent, best matching is obtained with $\varepsilon_r$ between 1 and 2, for water at a value around 5. However, in the case of the 1:1 mixture of dialanine and water, a screening of 1 is required to reproduce the phase separation seen in the fully atomistic system. More testing is required to see whether these values are generically valid, and how the effective screening depends on the density of CG versus AA sites. Preliminary results from our group indicate that atomistic water (in the form of bundled SPC$^{90}$) and CG MARTINI water mix homogeneously when using a dielectric constant of 2. Nevertheless, even with the standard $\varepsilon_r=15$ value of the MARTINI force field, we showed here that a system in which a large amount of dialanine peptides is mixed with butane phase separates with the layer thickness and density of molecules the same in atomistic and hybrid representations. In an aqueous environment, the use of the recently parameterized polarizable MARTINI water model$^{38}$ might improve the performance of the hybrid model as it is capable of explicit screening.

Overall we conclude that the proposed hybrid scheme appears to work, at least for the simple systems considered so far. We are currently exploring its range of applications, considering for instance the binding of an enzyme to a protein with only the binding pocket and ligand in full atomic detail, or the association of proteins in which only the surface of the protein is modeled in full detail.