Of Stalks and Diamonds. Simulation Studies of Membrane Fusion and the Role of Fusion Peptides

Fuhrmans, Marc

Appendix

The MARTINI model

The MARTINI forcefield has been parametrized by extensive calibration of the chemical building blocks of the coarse-grained forcefield against thermodynamic data, in particular oil/water partitioning coefficients. This is similar in spirit to the recent development of the GROMOS forcefield [88]. Processes such as lipid self-assembly, peptide membrane binding and protein-protein recognition depend critically on the degree to which the constituents partition between polar and non-polar environments. The use of a consistent strategy for the development of compatible coarse-grained and atomic level forcefields is of additional importance for its intended use in multiscale applications [107]. The overall aim of this coarse-graining approach is to provide a simple model that is computationally fast and easy to use, yet flexible enough to be applicable to a large range of biomolecular systems.

Currently, the MARTINI forcefield provides parameters for a variety of biomolecules, including many different lipids, cholesterol, carbohydrates, and all amino acids. A protocol for simulating peptides and proteins is also available. Extensive comparison of the performance of the MARTINI model with respect to a variety of experimental properties has revealed that the model performs generally quite well (semi-quantitatively) for a broad range of systems and state points. Properties accurately reproduced include structural (*e.g.* liquid densities [108], area/lipid for many different lipid types [108], accessible lipid conformations [97], or the tilt angle of membrane spanning helices [66]), elastic (*e.g.* bilayer bending modulus [108], rupture tension [46]), dynamic (*e.g.* lipid lateral diffusion rates [108], water transmembrane permeation rate [108], time scales for lipid aggregation [108, 67], cholesterol flip-flop [109], lipid rafts [110]), and thermodynamic (*e.g.* bilayer phase transition temperatures [65, 111], propensity for interfacial versus transmembrane peptide orientation [66], lipid desorption free energy [46]) data.

Basic parametrization

The mapping The MARTINI model [46] is based on a four-to-one mapping, *i.e.* on average four heavy atoms are represented by a single interaction center, with an exception for ring-like molecules. To map the geometric specificity of small ring-like fragments or molecules (*e.g.* benzene, cholesterol, and several of the amino acids) the general four-to-one mapping rule is insufficient. Ring-like molecules

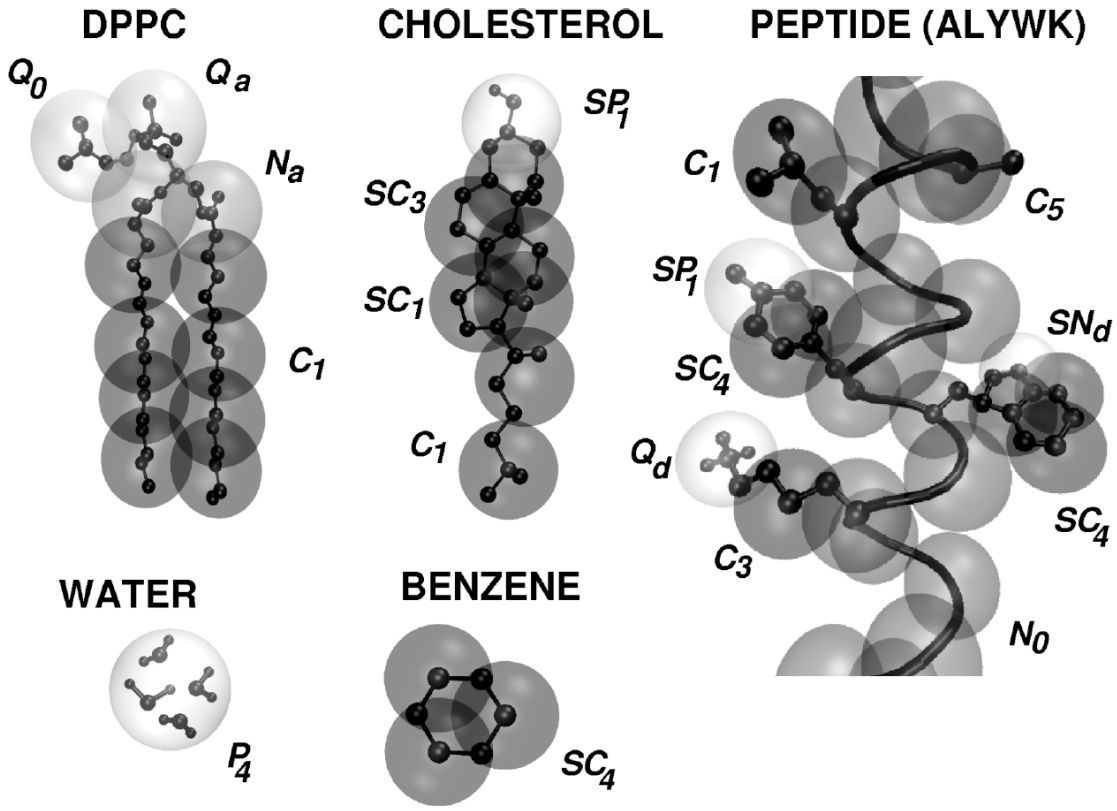

Figure A.1: Mapping between the chemical structure and the coarse-grained model for DPPC, cholesterol, water, benzene and a peptide fragment (with five amino acids highlighted). The coarse-grained bead types which determine their relative hydrophilicity are indicated, with more polar groups shown in lighter shades. The prefix "S" denotes a special class of coarse-grained sites used to model rings.

are therefore mapped with higher resolution (up to two-to-one). The model considers four main types of interaction sites: polar (P), non-polar (N), apolar (C), and charged (Q). Within a main type, subtypes are distinguished either by a letter denoting the hydrogen bonding capabilities (d=donor, a=acceptor, da=both, 0=none), or by a number indicating the degree of polarity (from 1=low polarity, to 5=high polarity). The mapping of representative biomolecules is shown in Fig. A.1.

Non-bonded interactions All particle pairs i and j at distance r_{ij} interact via a Lennard-Jones potential:

$$V_{\mathrm{LJ}} = 4\epsilon_{ij}((\sigma/r_{ij})^{12} - (\sigma/r_{ij})^{6}). \tag{A.1}$$

The strength of the interaction, determined by the value of the well-depth ϵ_{ij} depends on the interacting particle types. The value of ϵ ranges from $\epsilon_{ij} = 5.6$ kJ/mol for interactions between strongly polar groups to $\epsilon_{ij} = 2.0$ kJ/mol for interactions between polar and apolar groups mimicking the hydrophobic effect. The effective

size of the particles is governed by the Lennard-Jones parameter $\sigma = 0.47$ nm for all normal particle types. For the special class of particles used for ring-like molecules slightly reduced parameters are defined to model ring-ring interactions; $\sigma = 0.43$ nm, and ϵ_{ij} is scaled to 75% of the standard value. The full interaction matrix can be found in [46]. In addition to the Lennard-Jones interaction, charged groups (type Q) bearing a charge q interact via a Coulombic energy function with a relative dielectric constant $\epsilon_{\mathrm{rel}} = 15$ for explicit screening:

$$V_{\mathrm{el}} = q_i q_j / (4\pi\epsilon_0 \epsilon_{\mathrm{rel}} r_{ij}). \tag{A.2}$$

Note that the non-bonded potential energy functions are used in their shifted form. The non-bonded interactions are cut off at a distance $r_{\mathrm{cut}} = 1.2$ nm. The Lennard-Jones potential is shifted from $r_{\mathrm{shift}} = 0.9$ nm to r_{cut}. The electrostatic potential is shifted from $r_{\mathrm{shift}} = 0.0$ nm to r_{cut}. Shifting of the electrostatic potential in this manner mimics the effect of a distance-dependent screening.

Bonded interactions Bonded interactions are described by the following set of potential energy functions:

$$\begin{aligned} V_{\mathrm{b}} &= {}^1\!/_2 K_{\mathrm{b}} (d_{ij} - d_{\mathrm{b}})^2 \\ V_{\mathrm{a}} &= {}^1\!/_2 K_{\mathrm{a}} (\cos(\phi_{ijk} - \cos(\phi_{\mathrm{a}}))^2 \\ V_{\mathrm{d}} &= K_{\mathrm{d}} (1 + \cos(\phi_{ijkl} - \phi_{\mathrm{d}})) \\ V_{\mathrm{id}} &= K_{\mathrm{id}} (\phi_{ijkl} - \phi_{\mathrm{id}})^2 \end{aligned} \tag{A.3}$$

acting between bonded sites i,j,k,l with equilibrium distance d_{b}, angle ϕ_{a} and dihedral angles ϕ_{d} and ϕ_{id}. The force constants K are generally weak, inducing flexibility of the molecule at the coarse-grained level resulting from the collective motions at the fine-grained level. The bonded potential V_{b} is used for chemically bonded sites, and the angle potential V_{a} to represent chain stiffness. Proper dihedrals V_{d} are presently only used to impose secondary structure of the peptide backbone, and the improper dihedral angle potential V_{id} is used to prevent out of plane distortions of planar groups. Lennard-Jones interactions between nearest neighbors are excluded.

Effective timescale For reasons of computational efficiency the mass of the coarse-grained beads is set to 72 amu (corresponding to four water molecules) for all beads except for beads in ring structures for which the mass is set to 45 amu. Using this setup, typical systems can be simulated with an integration timestep of 30-40 fs. Due to the loss of effective friction caused by the missing atomistic degrees of freedom, however, the coarse-grained dynamics proceed faster than the timestep suggests. Based on comparison of diffusion constants in the MARTINI model and in atomistic models, the effective time sampled using the coarse-grained interactions is 3-8 fold larger. When interpreting the simulation results with the MARTINI model, a standard conversion factor of 4 can be used, which is the effective speed-up factor in the diffusion dynamics of coarse-grained water compared to real water. The same order of acceleration of the overall dynamics

is also observed for a number of other processes, including the permeation rate of water across a membrane [108], the sampling of the local configurational space of a lipid [97], and the aggregation rate of lipids into bilayers [108] or vesicles [67]. However, the speed-up factor might be quite different in other systems or for other processes. Particularly for protein systems no extensive testing of the actual speed-up due to the coarse-grained dynamics have been performed, although protein translational and rotational diffusion was found to be in good agreement with experimental data in simulations of coarse-grained rhodopsin [40]. In general, however, the time scale of the simulations has to be interpreted with care.

Reproducing thermodynamic data: optimizing non-bonded parameters

In order to parametrize the non-bonded interactions of the MARTINI model, a systematic comparison to experimental thermodynamic data was performed. Specifically, the free energy of hydration, the free energy of vaporization, and the partitioning free energies between water and a number of organic phases were calculated for each of the 18 different coarse-grained particle types. Concerning the free energies of hydration and vaporization, the MARTINI model reproduces the correct trend [46]. The actual values are systematically too high, however, implying that the coarse-grained condensed phase is not as stable with respect to the vapor phase as it should be. The same is true with respect to the solid phase. This is a known consequence of using a Lennard-Jones (12-6) interaction potential, which has a limited fluid range. Switching to a different non-bonded interaction potential could, in principle, improve the relative stability of the fluid phase. As long as its applications are aimed at studying the condensed phase and not at reproducing gas/fluid or solid/fluid coexistence regions, the most important thermodynamic property is the partitioning free energy. Importantly, the water/oil partitioning behavior of a wide variety of compounds can be accurately reproduced with the current parametrization of the MARTINI model. Table I shows results obtained for the partitioning between water and a range of organic phases of increasing polarity (hexadecane, chloroform, and octanol) for a selection of the 18 coarse-grained particle types. The free energy of partitioning between organic and aqueous phases, $\Delta G^{\mathrm{oil/aq}}$, was obtained from the equilibrium densities ρ of coarse-grained particles in both phases:

$$\Delta G^{\mathrm{oil/aq}} = kT \ln(\rho_{\mathrm{oil}}/\rho_{\mathrm{aq}}). \tag{A.4}$$

The equilibrium densities can be obtained directly from a long molecular dynamics simulation of the two phase system in which small amounts (around 0.01 mole fraction proved sufficient to be in the limit of infinite dilution) of the target substance are dissolved. With the MARTINI model, simulations can easily be extended into the multi-microsecond range, enough to obtain statistically reliable results to within 1 kJ/mol for most particle types. As can be judged from Table I, comparison to experimental data for small molecules containing four heavy atoms (the basic mapping of the MARTINI model) reveals a close agreement to within 2 kT for almost all compounds and phases; indeed, agreement is within 1 kT

Table I: Oil, chloroform, and octanol/water partitioning free energies for a selection of the 18 coarse-grained particle types (CG), compared to experimental values of the corresponding chemical building blocks (Exp). The experimental data are compiled from various sources (see [46]), the simulation data are obtained using Eq. A.4. All values are expressed in kJ mol^{-1} and obtained at T=300 K.

Building block	Type	Hexadecane/water		Chloroform/water		Octanol/water	
		CG	Exp	CG	Exp	CG	Exp
acetamide	P_5	-28	-27	-18	-20	-10	-8
water	P_4	-23	-25	-14	–	-9	-8
propanol	P_1	-11	-10	-2	-2	-1	0
propylamine	N_d	-7	-6	0	1	3	3
methylformate	N_a	-7	-6	0	4	3	0
methoxyethane	N_0	-2	1	6	–	5	3
butadiene	C_4	9	11	13	–	9	11
chloropropane	C_3	13	12	13	–	14	12
butane	C_1	18	18	18	–	17	16

for many of them. Expecting more accuracy of a coarse-grained model might be unrealistic. Note that the multiple non-bonded interaction levels allow for discrimination between chemically similar building blocks, such as saturated versus unsaturated alkanes or propanol versus butanol (which would be modeled as N_{da}) or ethanol (P_2). A more extensive table including all particle types and many more building blocks can be found in [46].

To select particle types for the amino acids, systematic comparison to experimental partitioning free energies is also used. Table II shows the resulting assignment of the amino acid sidechains and the associated partitioning free energies. The simulation data are calculated from equilibrium densities of low concentrations of coarse-grained beads dissolved in a water/butane two phase system, using Eq. A.4. The experimental data [112] refer to partitioning of sidechain analogues between water and cyclohexane. Both the simulation and the experimental data are obtained at 300 K. Where available, the experimental values are reproduced to within 2 kT, a level of accuracy that is difficult to obtain even with atomistic models. Most amino acids are mapped onto single standard particle types in a similar way as was done recently by other groups [113, 114]. Fig. A.1 shows the mapping of a few of them. The apolar amino acids (Leu, Pro, Ile, Val, Cys, and Met) are represented as C type particles, the polar uncharged amino acids (Thr, Ser, Asn, Gln) by the class of P type particles, and the small negatively charged sidechains (Glu, Asp) as Q type. The positively charged amino acids (Arg, Lys) are modeled by a combination of a Q type and an N or C type particle. The bulkier ring-based sidechains are modeled by three (His, Phe, Tyr) or four (Trp) beads of the special class of ring particles. The Gly and Ala residues are only represented by the backbone particle. The type of the backbone particle depends on its secondary structure; when free in solution or in a coil or bend the back-

Table II: Free energy based mapping of the amino acids. The experimental partitioning free energies (Exp) are obtained for cyclohexane/water [112], the simulation results (CG) for butane/water, using Eq. A.4. All values are expressed in kJ mol^{-1} and obtained at T=300 K.

Sidechain	Type	Oil/water	
		CG	Exp
Leu	C_1	22	22
Ile	C_1	22	22
Val	C_2	20	17
Pro	C_2	20	–
Met	C_5	9	10
Cys	C_5	9	5
Ser	P_2	-14	-14
Thr	P_1	-11	-11
Asn	P_5	<-25	-28
Gln	P_4	-23	-25
Asp	Q_a	<-25	–
Asp (uncharged)	P_3	-18	-19
Glu	Q_a	<-25	–
Glu (uncharged)	P_1	-11	-11
Arg	C_5-Q_d	<-25	–
Arg (uncharged)	C_5-P_4	-22	-25
Lys	C_3-Q_d	<-25	–
Lys (uncharged)	C_3-P_1	0	2
His	SC_4-SP_1-SP_1	-19	-20
Phe	SC_4-SP_4-SP_4	19	17
Tyr	SC_4-SC_4-SP_1	-1	-2
Trp	SC_4-SN_d-SC_4-SC_4	12	9

bone has a strong polar character (P type), as part of a helix or beta strand the inter-backbone hydrogen bonds reduce the polar character significantly (N type). Details of the parametrization of the amino acids can be found in [66].

Reproducing structural data: optimizing bonded parameters

To parametrize the bonded interactions, structural data were used that are either directly derived from the underlying atomistic structure (such as bond lengths of rigid structures), or obtained from comparison to fine-grained simulations. In the latter procedure, the fine-grained simulations are first converted into a mapped coarse-grained simulation by identifying the center-of-mass of the corresponding atoms as the mapped coarse-grained bead. Second, the distribution functions are calculated for the mapped simulation and compared to those obtained from a true coarse-grained simulation. Subsequently the coarse-grained parameters are systematically changed until satisfactory overlap of the distribution functions

is obtained. Using this procedure, simulations of bulk alkanes have been used to determine the optimal values of the standard equilibrium bond distance of 0.47 nm and force constant of $K_\mathrm{b} = 1250$ kJ mol^{-1} nm^{-2}, and equilibrium angle of 180° with force constant of $K_\mathrm{a} = 25$ kJ mol^{-1}. Likewise, standard bonded parameters have been derived for unsaturated alkanes [46], the phospholipid headgroup [108], and for cholesterol [46].

For the bonded interactions involving amino acid sidechains and for the peptide backbone a similar procedure was used. However, instead of deriving target distributions from fine-grained simulations, the distributions were derived directly from the protein databank, using the same mapping protocol. The amount of data is so large that statistically very accurate distributions can be obtained for all the required bonded interactions. These distributions reflect all possible configurations for a large number of different systems under a variety of conditions. Keeping the aim of the MARTINI model in mind, namely to be able to simulate many biomolecules with a single set of parameters, this is the least biased information. Using this procedure, bonded parameters were derived for the backbone (BB) potentials, namely the BB-BB bonded potential, the BB-BB-BB angle potential, and the BB-BB-BB-BB dihedral potential. The last two terms are used to enforce the secondary structure of the backbone, which is therefore an input parameter in the MARTINI model. In the current version, different dihedral and angle parameters are used to distinguish a helix, a strand, or a random coil. It is therefore not possible to study realistic folding-unfolding events at this stage. Furthermore, for each amino acid sidechains (SC) distributions were obtained for the BB-SC bonded potential, the BB-BB-SC angle potential, and for the intra-SC potentials for amino acids containing more than one coarse-grained particle. The complete set of bonded parameters for proteins can be found in [66].

Limitations

The potential range of applications of the MARTINI model is very broad. There are, however, certain important limitations which should be kept in mind. First of all, the model has been parametrized for the fluid phase. Properties of solids, such as crystal packing, are not expected to be accurate. Both the gas and the solid phase appear somewhat too stable with respect to the fluid phase. The thermodynamic behavior of solid/fluid and gas/fluid interfaces should therefore be interpreted with care, at least at the quantitative level. In applications where such interfaces are formed (especially the water/vapor interface in, *e.g.*, rupture of lipid monolayers) these limitations have to be kept in mind.

Furthermore, the parametrization is based on free energies. The inherent entropy loss on coarse-graining is necessarily compensated for by a reduced enthalpy term [97]. The enthalpy/entropy balance of many processes is therefore biased when modeled at the coarse-grained level. Consequently, the temperature dependence is affected, although not necessarily weaker. For instance, the temperature dependence of the hydration free energy for linear alkanes was found to be more pronounced in the coarse-grained representation compared to an all-atom representation [97]. As is true for any forcefield, applications outside the temperature

range used for parametrization (∼270-330 K) have to be considered with care.

Another difficulty encountered in the MARTINI model, and perhaps in most coarse-graining approaches, is the correct modeling of the partitioning of polar and charged compounds into a low dielectric medium. Because of the implicit screening, the interaction strength of polar substances is underestimated in non-polarizable solvents. Applications involving the formation of polar/charged complexes in a non-polar environment are especially prone to be affected. The inability to form a transmembrane water pore upon dragging a lipid across the membrane is an example [46]. Apart from the implicit screening in the coarse-grained model, the neglect of long-range electrostatic forces poses a further limitation. Pairwise interactions beyond 1.2 nm (between two and three coarse-grained beads away) are not taken into account. In principle long-range electrostatic interactions could be added to the coarse-grained model, in similar ways as it is done in atomistic simulations. One has to realize that a modification of the electrostatic interaction scheme will affect other system properties.

Finally, in applications of peptides and proteins one has to be aware that secondary structure transformations are not modeled in the current parametrization. The secondary structure is essentially fixed by the use of a dihedral potential energy function, allowing to discriminate between various secondary structure elements but preventing realistic transitions between them. Processes in which folding and unfolding are playing a substantial role are therefore not suitable for modeling with the current version of the MARTINI forcefield. Movement of secondary structure elements with respect to each other are possible, however, and were shown to be quite realistic in a recent application of the gating of a membrane embedded mechanosensitive channel [115].