



Coarse grained simulation of the non-myristoylated Recoverin with the Martini force field

Giuseppe Gambini

Contents

1 ABSTRACT	2
2 THEORETICAL INTRODUCTION	2
2.1 The philosophy behind Martini	2
2.2 Mapping and basic parametrization	2
Bonded and non-bonded parameters	
2.3 Validations of the FF in the specific case of proteins	6
Partitioning free energy ▪ Dimerization free energy ▪ Binding of Wimley-White (WW) peptides ▪ Discussion of the free energy data	
2.4 Versatility and pitfalls of the Martini FF	10
2.5 Background information on protein structure and function	10
3 METHODS AND RESULTS	11
3.1 Preparation of the CG system	11
Considerations on the employed parameters ▪ The Martinize2 script ▪ Post-processing of the simulations	
3.2 Structural properties compared with the all-atom simulation	14
Solvent accessible surface area (SASA) and minimum distance of the calcium ion from EF2	
4 CONCLUSIONS	17
5 APPENDIX	17
6 Useful definitions for <i>Martinize2</i>	17
References	18

1 ABSTRACT

Coarse-grained (CG) simulations represent a valuable tool to probe the time and length scales of systems beyond what is feasible with traditional all atom (AA) models, providing a bridge between the atomistic and the mesoscopic scale. Among the different CG models applied in the context of biomolecules, in this report we address the Martini force field (FF), which was originally designed to reproduce the dynamics of lipids and surfactants. In 2008 Monticelli et al.¹ provided an extension of the FF to amino acids. This extension was later refined until the current version of the FF, namely Martini3. In this report, we will focus on the parametrization procedure of amino acids and then discuss a short CG simulation of a Recoverin protein in aqueous environment. Emphasis will be put on the set up of the CG system and on the parameters of the simulation. Finally, a short structural analysis and a comparison with an AA simulation will be shown and discussed. The input and output files and analysis scripts are collected in this [GitHub folder](#).

2 THEORETICAL INTRODUCTION

2.1 The philosophy behind Martini

The Martini FF approach simplifies an high resolution picture of the system by grouping atoms together into coarse-grained beads. A distinguishing feature of Martini is that thermodynamic and partitioning experimental data are used for the calibration of the interactions among beads. This line of reasoning is labeled as *top-down* approach and it is often considered to be the opposite of the *bottom-up* approach, in which the FF parameters are derived from fundamental physics (atomic quantum mechanics or statistical physics) of the atomistic details of the system. Although the *bottom-up* approaches are capable of capturing more of the fine details of the interaction, the *top-down* approach usually provides potentials that are more easily transferable².

The success of Martini lies in the balance between accuracy and computational efficiency achieved thanks to a combination of *top-down* and *bottom-up* strategies. This philosophy was applied in the original lipid force field and later extended to proteins. More specifically, as it will be later discussed in section 2.2.1, nonbonded interactions are based on experimental thermodynamic data available for each amino acid (*top-down* approach), while bonded parameters are derived systematically from distributions of bond lengths, angles, and dihedrals in the Protein Data Bank¹ (statistics-based *bottom-up* approach). Remarkably, the peptide FF is fully compatible with the lipid FF, since it maintains consistency in the potentials between different types of beads.

Martini is particularly well-suited for reproducing the way in which the constituents of a system partition between polar and nonpolar environments. Partitioning constitutes one of the major contributors to the *self-assembly* of biological macromolecules. Here, we give a brief justification of this assessment by listing three relevant examples in the context of proteins:

- **hydrophobic effects:** to maximize entropic contributions coming from the arrangement of the hydrogen bonds network of water, hydrophobic amino acid side chains tend to aggregate and minimize their exposure to the polar solvent. If it were not the case, the water molecules around the isolated hydrophobic side chains would form a thermodynamically unfavourable ordered arrangement, similar to what happens when water molecules surround an isolated oil molecule.
- **protein-membrane binding:** cell membranes are lipid bilayers consisting of hydrophobic tails (nonpolar) and polar head groups. Proteins with hydrophobic domains are more likely to interact with the hydrophobic core of the membrane, meaning that the nonpolar side chains of hydrophobic residues are likely to interact with the hydrophobic lipid tails. Conversely, polar and charged side chains may interact not only with water, but also with the hydrophilic heads of the lipids.
- **protein-protein recognition:** similarly to the protein-membrane binding, the nature of complementary interacting surfaces can stabilize the protein-protein complex.

In addition, the partitioning of the side-chains drives the folding of a protein, although most of the times this process involves complicated interactions at atomic-level and is not fully captured by CG models. Moreover, the secondary structure remains fixed in a Martini simulation, as it represents an input for the generation of the Martini topology (see section 3.1).

2.2 Mapping and basic parametrization

Martini mainly relies on a four-to-one mapping scheme, as on average four heavy atoms and associated hydrogens are mapped into one CG bead. The model features four main types of CG beads, i.e., non polar "C", intermediately polar "N", polar "P" and charged "Q". The level of each feature is specified with a number (1, 2, 3, 4, 5 or 6). On the implementation side, the difference between the bead types lies in the electrostatic and Lennard-Jones (LJ) parameters.

Martini3 supports three possible sizes: regular "R" (usually not specified), small "S" and tiny "T". Additional sublabels ("0", "da", "d", "a") are assigned based on the hydrogen donor or acceptor capabilities of the beads.

The same specific chemical units can be found irrespective of the larger molecule they take part (lipids, proteins, alkanes etc...), allowing for the identification of common types of beads which are systematically mapped and parametrized in Martini.

As accurately described in Alessandri et al.³, the Martini3 mapping is performed according to the following criteria:

- Only non-hydrogen atoms define the mapping.
- The chemical functional groups (for instance amide or carboxylate groups) are kept inside the same bead.
- Fully branched fragments represent an exception to the four-to-one mapping rule for heavy atoms. For instance, tetramethyl methane, which has one central buried carbon atom and other four carbons on the edges of a tetrahedron, is represented with a single bead, as the buried atom is "screened" by the other four.
- While regular "R" beads are used for standard four-to-one mapping, small beads "S" allow a three-to-one mapping and are mostly used for aliphatic ring compounds such as cyclohexane. In addition, the tiny bead T is used for a two-to-one mapping and is typically used in conjugated rings such as aromatic compounds. Three different masses are attributed to these types of beads, namely, 72, 54 and 36 amu respectively for R, S and T.
- Ions are represented by a single CG bead, which represents both the ion and its first hydration shell.
- The single water bead represents four water molecules. Improved screening of interactions can be achieved with the "polarizable water", which still represents four water molecules, but uses three beads.

Below, some examples of chemical building blocks mapped to a single bead:

type	chemical building block	examples	
		2D	name (mapping)
P2	CH ₃ CH ₂ -COOH		propanoic acid (P2)
SP2	CH ₃ -COOH		acetic acid (SP2)
P1	(CH ₃) ₂ CH-OH		isopropanol (P1)

Figure 1. Examples of chemical groups corresponding to one single bead. Two regularly sized polar compounds (labelled with P1 and P2) are represented together with a small sized polar compound SP2.

Regarding the mapping of the beads, an important difference between Martini3 and the previous version Martini2 has to be pointed out: in Martini2 the bond distances were obtained from the atomistic mapping computed with respect to the center of mass (COM) of the heavy atoms only, while in Martini3 they are computed with respect to the center of geometry (COG) of the heavy atoms with hydrogen atoms taken into account. In the specific case of protein backbone beads, the new center of the bead is located in the COG of the backbone atoms plus their associated hydrogens. The COM choice proved unsatisfactory, as it often led to overly high packing densities, while the newly adopted COG strategy better reproduces atomistic volumes, leading to a better molecular packing.

Interactions between the beads are described by effective potentials that incorporate both energetic and entropic contributions resulting from integrating out atomistic details. As in the case of atomistic FF, the empirical potential energy function is given by

$$U(\mathbf{R}) = \sum_i^{\text{bonds}} \frac{k_{b,i}}{2} (b_i - b_{0,i})^2 + \sum_i^{\text{angles}} \frac{k_{\alpha,i}}{2} (\alpha_i - \alpha_{0,i})^2 + \sum_i^{\text{torsions}} \sum_k^M \frac{V_{ik}}{2} [1 + \cos(n_{ik}\theta_{ik} - \theta_{0,ik})] \\ + \sum_i^{\text{impropers}} \frac{k_{\theta,i}^{\text{imp}}}{2} (\theta_i^{\text{imp}} - \theta_{0,i}^{\text{imp}})^2 + \sum_{ij}^{\text{pairs}} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{ij}^{\text{pairs}} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad (1)$$

Because of their intrinsic periodicity, torsional terms (proper dihedrals) are naturally defined by a cosine series of M terms for each dihedral angle. Thus, n_{ik} is a parameter describing the multiplicity for the k th term of the series, $\theta_{0,ik}$ is the corresponding phase angle, and V_{ik} is the energy barrier. The improper dihedral term is used to prevent

out-of-plane distortions of planar groups (peptide bonds, aromatic rings etc).

As in standard atomistic FF, in Martini3, non-bonded interactions between connected beads are always excluded. In standard Martini the water beads do not have partial charges, hence they solely interact via LJ. To remedy this situation in an effective way, the Coulomb potential is screened with a relative dielectric constant $\epsilon_r = 15$. For the polarizable Martini version (with polarizable water) the dielectric constant is set to $\epsilon_r = 2.5$ and the Particle Mesh Ewald (PME) algorithm is suggested for the treatment of the long-range electrostatics.

Notice that the functional form of the force field in (1) allows the exploitation of highly efficient MD simulation programs such as Gromacs.

Lastly, the COG mapping choice of Martini3 mentioned above is crucial in the derivation of the bonded parameters $\{b_{0,i}, \alpha_{0,i}, \theta_{0,ik}, \theta_{0,i}^{\text{imp}}\}$, which are thoroughly recalibrated with respect to the previous Martini2.

2.2.1 Bonded and non-bonded parameters

In the Martini parametrization, no rigid distinction exists between the calibration and the validation of results. On a qualitative level, three different levels of complexity can be distinguished:

1. Tests on small molecules mapped to single beads in different solvents. These tests were mainly used for calibration of the bonded parameters and the LJ parameters.
2. A level of complexity is added with simulations of proteins and lipid bilayers. These tests were used as a consistency check for the parameters derived previously.
3. The definitive tests involved simulations of more extended systems such as dimerization PMFs of membrane proteins, protein-protein interactions etc. These were then compared directly with experimental data.

After this small premise, we discuss the actual parameterization. A set of structural data of approximately 2000 PDB proteins was used for the parametrization of bonded interactions. The CG parameters were found by matching the PDB distributions of the bonds, angles and dihedrals with the distributions obtained from CG simulations performed on short test peptides¹. The PDB distributions were obtained by defining "virtual" sites in the atomistic structures as done in table 1 and also on the left of figure 2, where atoms inside the virtual bead are represented as a unique entities. Subsequently, bond lengths, angles and dihedrals between the virtual beads were considered and reported in histograms such as the ones reported in 2. This simplification is fundamental since it leads to a one-to-one correspondence with the parameters of the CG system. This protocol was proposed again for successive versions, among which Martini3.

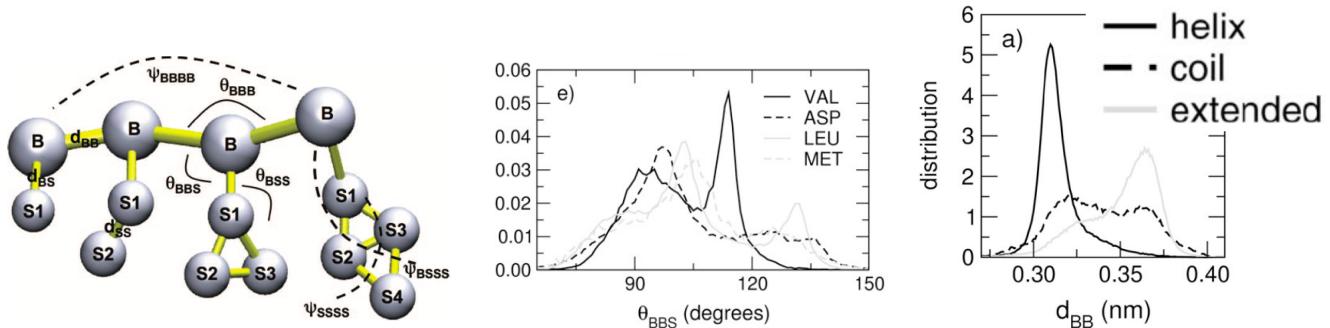


Figure 2. Images taken from Monticelli et al.¹ representing the "virtual" beads superimposed to atomistic structures (on the left), the distributions of backbone-backbone-side chain (BBS) angle distributions for selected amino acids (in the center) and the backbone-backbone bond distance distribution (on the left) over an ensemble of protein structures.

Concerning non-bonded interactions, the depths and the widths ε_{ij} and σ_{ij} reported in the LJ interaction matrix of Martini (in figure 4) have been mainly optimized by looking at partitioning data highlighting the hydrophilic/hydrophobic nature of the beads.

In Martini, all the LJ cross-interactions are explicitly parametrized since there is no combination rule between two dissimilar non-bonded atoms as in atomistic simulations. Affinities of chemical groups toward different solvent phases have been systematically tested for the parametrization. Below in figure 3, we present an example regarding aliphatic (blue squares) and aromatic (green triangles) small molecules inserted in water and in X solvent (hexadecane, octanol and chloroform). The free energy of partitioning ($\Delta G_{\text{transfer}}$) is computed as

$$\Delta G_{\text{transfer}}(\text{wat}/X) = k_B T \ln \left(\frac{\rho_X}{\rho_{\text{wat}}} \right)$$

where ρ_X/ρ_{wat} is the so called *partition coefficient* of the small molecules in the two phases. Free energies of partitioning are computed via *direct counting*: first, an ensemble of configurations of the small molecule in both the water and X solvent is generated, then, the ratio between the number times N_{wat} in which the small molecule is in the water phase and in the X solvent N_X gives an estimate of the partition coefficient. A reliable value for free energy differences is obtained only if the statistics are good enough, meaning that many transitions between the states wat/X and full sampling of each state are required to capture energetic and entropic contributions. Below in 3, a sample of partitioning free energies computed for different small molecules in three different solvents.

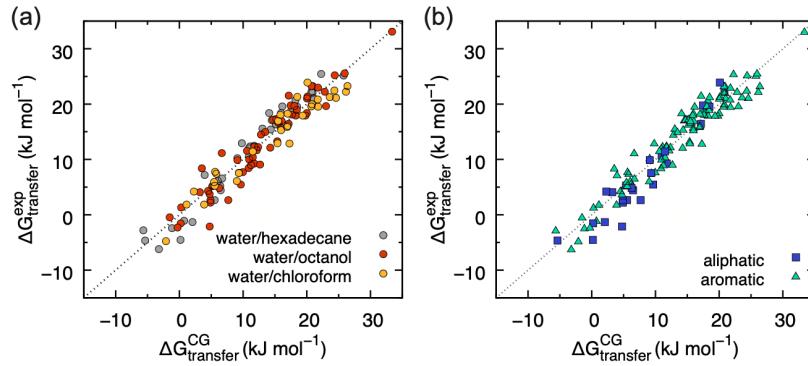


Figure 3. Image taken from Alessandri et al.⁴. The two plot represent the same data: on the left a distinction is made between the phases in which the small molecules are inserted, while on the right different representations are used for aliphatic and aromatic molecules.

Although free energies are reproduced by the CG representation, the loss of internal degrees of freedom for groups of atoms represented by a CG bead inevitably reduces the entropy of the system. This requires a concomitant reduction in the enthalpy⁵, which might produce a wrong temperature behaviour. Consequently, similar to any FF, applications outside the temperature range used for parametrization (~ 270 - 330 K) have to be considered with care⁶.

Below in 4, we present the 18×18 interaction matrix of Martini2, with ten possible levels (in roman letters) with "O" corresponding to the highest depth $\max\{\varepsilon_{ij}\} = 5.6$ kJ/mol and "IX" to the lowest $\min\{\varepsilon_{ij}\} = 2.0$ kJ/mol.

sub	Q				P					N				C				
	da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1
Q	da	O	O	II	O	O	O	I	I	I	I	I	IV	V	VI	VII	IX	IX
	d	O	I	O	II	O	O	O	I	I	III	I	IV	V	VI	VII	IX	IX
	a	O	O	I	II	O	O	O	I	I	I	III	IV	V	VI	VII	IX	IX
	0	II	II	II	IV	I	O	I	II	III	III	III	IV	V	VI	VII	IX	IX
P	5	O	O	O	I	O	O	O	O	I	I	I	IV	V	VI	VI	VII	VIII
	4	O	O	O	O	O	I	I	II	III	III	III	IV	V	VI	VI	VII	VIII
	3	O	O	O	I	O	I	I	II	II	II	II	IV	V	V	VI	VII	VII
	2	I	I	I	II	O	II	II	II	II	II	II	III	IV	IV	V	VI	VII
N	1	I	I	I	III	O	II	II	II	II	II	II	III	IV	IV	IV	V	VI
	da	I	I	I	III	I	III	II	II	II	II	II	IV	IV	V	VI	VI	VI
	d	I	III	I	III	I	III	II	II	II	III	II	IV	IV	V	VI	VI	VI
	a	I	I	III	III	I	III	II	II	II	II	III	IV	IV	V	VI	VI	VI
C	0	IV	IV	IV	IV	IV	IV	IV	IV	III	IV	IV	IV	IV	IV	IV	V	VI
	5	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V
	4	VI	VI	VI	VI	VI	V	IV	IV	V	V	V	IV	IV	IV	IV	V	V
	3	VII	VII	VII	VII	VI	VI	V	V	IV	VI	VI	IV	IV	IV	IV	IV	IV
2	IX	IX	IX	IX	VII	VII	VI	VI	VII	VI	VI	VI	V	V	V	IV	IV	IV
	1	IX	IX	IX	VIII	VIII	VIII	VII	VII	VI	VI	VI	VI	V	V	IV	IV	IV

Figure 4. Table taken from Marrink et al.⁶.

The strongest interaction is between water P4 and charged particles Q. Going down in the P4 column we have progressively weaker interactions giving rise to hydrophobic effects. As expected, the self interactions P-P, N-N, C-C are gradually decreased in this order, reflecting the fact that polar beads tend to stay packed at higher densities due to

their stronger solvation free energy.

The latest version Martini3 presents an extended 24×24 matrix with 20 possible LJ levels.

2.3 Validations of the FF in the specific case of proteins

We present a brief overview of the validations performed for the Martini protein FF. These are provided in the articles by Monticelli et al.¹ (with the Martini2.1 version) and de Jong et al.⁷ (with Martini2.2). The article by de Jong et al. proposes also the version Martini2.2p, introducing polarized particles to remedy some of the shortcomings of polar residues in the previous Martini version of the group of Monticelli. Here, we will discuss some of its most significant results.

The latest Martini3 FF shows improvements over previous versions in areas such as molecular packing, transmembrane helix interactions, protein aggregation, and DNA base pairing⁵. The bead type assignment, partial charges and LJ levels of interactions are again recalibrated in this version. However, the aim of this section is to give the flavour of the parametrization philosophy behind Martini, together with some free energy calculation techniques frequently applied in the calibration of amino acids. For this reason, we will stick to the de Jong version.

2.3.1 Partitioning free energy

The partitioning behavior of a protein, which refers to its tendency to distribute between different phases (e.g., aqueous phase and lipid bilayer), is determined by its amino acid composition and structure. Proteins with hydrophobic regions or amphipathic structures are more likely to partition into membranes. Potential of mean force (PMF) profiles, both in AA and CG representations, were obtained for single amino acids side chains across a DOPC (dioleoyl-phosphatidylcholine) bilayer. A distance between the side chain and the phosphate group of the lipids was used as collective variable in an umbrella sampling calculation: the harmonic bias potential was applied along the distance between the side chain and the phosphates, subsequently, the PMF was achieved with the standard weighted histogram analysis method (WHAM). Atomistic reference simulations were performed with OPLS force field. As expected in the case of hydrophobic residues (Trp and Phe), a barrier is present near the lipid headgroup region (see figure 5). The case of histidine is critical since the charge of its ionizable side chain can vary depending on the pH of the environment. Here both cases are reported. The PMF of charged amino acids (His^+ , Arg, Lys, Glu, Asp) show that the probability of the residues entering the bilayer is negligible. The polar residues (Ser, Thr, Gln, Asn) PMFs are similar to the PMFs of the charged ones, a part from a small barrier near the phosphates and a much smaller free energy scale.

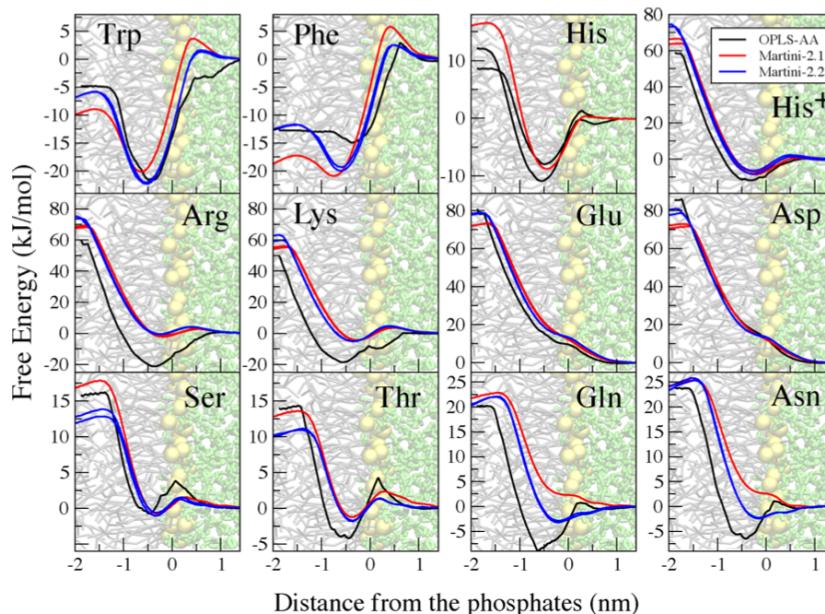


Figure 5. Image taken from de Jong et al.¹ representing the PMFs of charged, polar and hydrophobic amino acids.

2.3.2 Dimerization free energy

To evaluate the quality of the interactions between side chains in Martini, dimerization free energies between residue pairs were computed in water and decane solvent and compared to results from atomistic simulations. The pairs Lys/Glu

and Lys/Lys were chosen to be representative for salt-bridge and charged interactions. The PMF was computed as a function of the side chains' center of mass distance. Starting from a dimeric configuration, the distance between the two COMs (here labeled with r) was progressively constrained to higher values. The PMF was computed with the Jarzinsky equality

$$\text{PMF}(r) = -k_B T \ln \left\langle \exp \left(-\frac{W^{\text{dim}}(r)}{k_B T} \right) \right\rangle$$

over a large number of non-equilibrium simulations (here $W^{\text{dim}}(r)$ represents the work that forces dimerization). Dimerization involves energetic changes including bonding energies, electrostatic interactions, and entropy contributions, which are not easily accounted for through the direct counting technique. For this reason, de Jong et al. resorted to the following dimerization formula for two particle systems⁸:

$$\Delta G^{\text{dim}} = -k_B T \ln K_a = -k_B T \ln \frac{4\pi R_{\max}^3 \int_0^{r_c} r^2 g(r) dr}{3v^\emptyset \int_{r_c}^{R_{\max}} r^2 g(r) dr} \quad g(r) = \exp \left(-\frac{\text{PMF}(r)}{k_B T} \right) = \left\langle \exp \left(-\frac{W^{\text{dim}}(r)}{k_B T} \right) \right\rangle \quad (2)$$

The equilibrium constant of the dimerization reaction K_a is estimated by an integration of the bound and unbound parts of the radial distribution function $g(r)$ obtained from the simulation multiplied by a volume defined by the maximum distance R_{\max} between the COMs. Here v^\emptyset is the molecular standard volume and r_c is a cut-off distance above which the two monomers are considered separated.

Below in figure 6, the PMF obtained for Lys-Glu and Lys-Lys dimers in decane. According to Martini, Glu is mapped to a polar backbone and a charged side chain bead (P, Q), while Lys is mapped to a polar backbone bead and an apolar and a charged bead for the side chain (P, C, Q). From the figures in 6, it is evident that interactions between charged beads in an oil-like medium are grossly underestimated in all the presented Martini FF versions. In the worst case represented by Martini2.1, the imposed dielectric constant of $\epsilon_r = 15$ brings unphysical screening, making Coulomb interactions too weak. The polarizable versions displaying a value of $\epsilon_r = 2.5$ partially mitigate this effect. Other measures have been taken to compensate this weakness of the Martini model: in the Martini2.2P model, the charge is slightly displaced from the center of the bead (while LJ interaction remains on-center)¹, leading to a deeper minimum for unlike charge pairs, which in turn determines an ameliorate PMF (see blue line in the Lys-Glu plot in 6).

Concerning the Lys-Lys PMF, the Coulomb repulsive potential largely determines the atomistic PMF profile, while in the CG case there is a competition between screened Coulomb and attractive LJ interactions. The Martini2.1 version (in green) displays an unphysical minimum at approximately 0.5 nm where screened Coulomb and LJ terms cancel each other. This issue is partially solved in the successive versions (in orange and blue) with $\epsilon_r = 2.5$ and displaced charges, although the effects of the LJ are still evident. These can be seen in the flat region around 0.75 nm and in the steep barrier below 0.5 nm, which is a consequence of the larger dimension of the side chain beads.

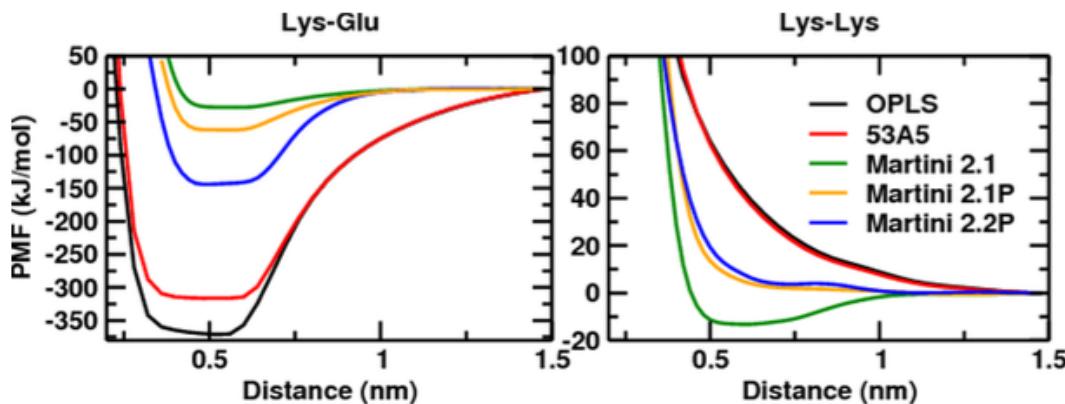


Figure 6. Interactions between charged residues in an apolar medium are severely underestimated because the screening constant does not depend on the environment. The polarizable Martini water model improves the situation. The label 53A5 indicates an extensive parametrization of the united-atom FF Gromos.

¹An additional bead carrying the full positive charge of Lys side chain is introduced. The remaining LJ interactions are still computed with respect to the standard Lys side chain bead, which becomes a sort of "virtual bead".

The values of the dimerization free energy as computed in (2) are reported in the table 8.

2.3.3 Binding of Wimley-White (WW) peptides

WW peptides are synthetic pentapeptides with both hydrophobic and hydrophilic regions that mimic the properties of transmembrane segments found in proteins. The validation of de Jong et al. was performed with a sequence of five amino acids set to WLXLL, where X denotes a variable residue. The aim was to compute a binding free energy $\Delta G^{WW}(X)$ of these peptides to a lipid POPC bilayer interface. The authors provided the results in terms of a relative binding free energy $\Delta\Delta G^{WW}$, computed with respect to the case X=Ala. Two equivalent ways of computing the relative free energy ΔG^{WW} are discussed below and represented in figure 7:

- The first method consists in subtracting the free energies associated to the vertical lines on the left of figure 7. This implies a so called "alchemical" simulation in which LJ and Coulomb interactions are progressively turned on/off to simulate the switch of the central amino acid of the WW peptide in both bulk water and interface region. Here the notation WLD_XLL represents a peptide in which the side chain of X has been converted to a dummy (D) side chain with LJ and Coulomb interactions turned off. The group of de Jong used free energy perturbation (FEP) to sample the configuration space along the perturbation path and the multistate Bennett acceptance ratio MBAR to estimate the free energy differences between different states. More precisely, if intermediate configurations between WLALL and WLXLL are labelled with sets of switching parameters $\{\{\lambda_1\}, \dots, \{\lambda_M\}\}$ for non-bonded side-chain interactions, then, according to the seminal article by Bennett⁹, the free energy differences

$$\Delta G_{\{\lambda_i\}, \{\lambda_{i+1}\}} = -k_B T \ln \left(\frac{\langle \min(1, \exp[-(U_{\{\lambda_{i+1}\}} - U_{\{\lambda_i\}})/k_B T]) \rangle_{\{\lambda_i\}}}{\langle \min(1, \exp[-(U_{\{\lambda_i\}} - U_{\{\lambda_{i+1}\}})/k_B T]) \rangle_{\{\lambda_{i+1}\}}} \right)$$

where $U_{\{\lambda_i\}} = U(q_1^{\{\lambda_i\}} \dots q_N^{\{\lambda_i\}})$ and the bracket $\langle \dots \rangle_{\{\lambda_i\}}$ notation indicates an ensemble average over the configuration $\{q_1^{\{\lambda_i\}} \dots q_N^{\{\lambda_i\}}\}$.

- The second method consists in subtracting the free energies of the horizontal lines in 7. This involves displacements of the WLXLL and WLALL peptides from the bulk region to the interface. The authors computed the PMFs associated to these displacements from umbrella sampling simulations with the distance between the COM of the POPC bilayer and the COM of the peptide used as collective variable.

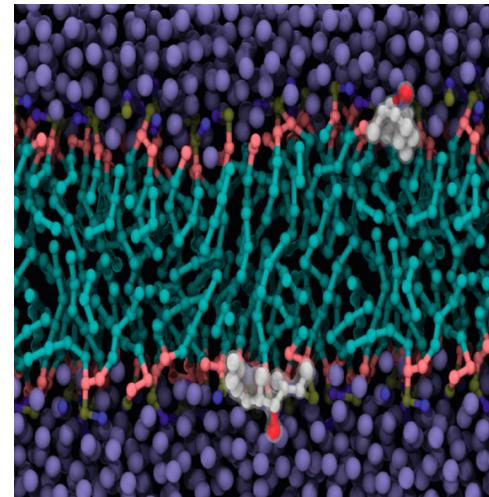
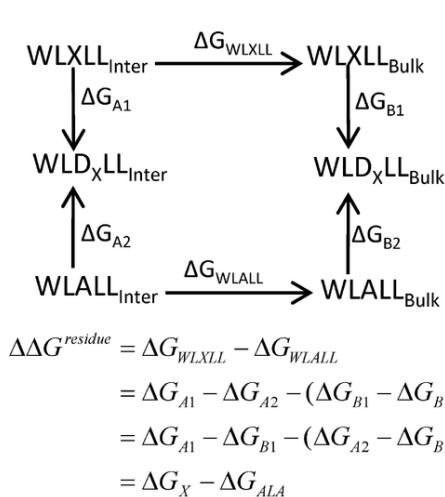


Figure 7. Image taken from Singh and Tieleman¹⁰. On the right, WLALL peptides at the POPC/water interface.

2.3.4 Discussion of the free energy data

The data collected from the three validations described in sections 2.3.1, 2.3.2 and 2.3.3 are reported in a single table. The lines starting with "ref." contain the experimental reference free energies. The second column "type" shows the beads employed for the representation of the side chains. The particle reassignment (in this case from Martini2.1 to Martini2.2) is highlighted in bold.

SC		type (charge) ^b	$\Delta\Delta G^{\text{WW}}{}^c$	$\Delta G^{\text{part}}{}^d$	$\Delta G^{\text{dim}}_{\text{water}}{}^e$	$\Delta G^{\text{dim}}_{\text{oil}}{}^e$
Phe	ref.		5.4 ± 0.3	12	-1.6	-2.9
	CG	SC4-SC4-SC4	12.2 ± 0.1	21	-4.5	-1.3
		SC5-SC5-SC5	7.7 ± 0.1	10	-3.0	-1.7
Trp	ref.		8.5 ± 0.4	9	-3.3	-3.3
	CG	SC4-SP1-SC4-SC4	9.2 ± 0.1	10	-4.7	-3.0
		SC4-SNd-SC5-SC5	9.4 ± 0.1	8	-4.0	-2.7

Figure 8. An extract of a table provided in de Jong et al.⁷

In Martini2.1 the Phe side chain is represented by three apolar (C) beads with "intensity" 4 (the S in front labels the side chain). In the version proposed by de Jong (Martini2.2) the SC4 are replaced with the slightly more polar SC5, leading to a better estimate of $\Delta\Delta G^{\text{WW}}$ and ΔG^{dim} both in water and oil.

In the case of Trp one can notice the presence of an explicit hydrogen bond donor group (SNd) in the reassignment, which better reflects the chemical nature of the Trp side chain and ameliorates the estimate of the dimerization free energy in water, although it worsens the estimate of the relative binding free energy of the WW peptide $\Delta\Delta G^{\text{WW}}$ and the dimerization free energy in oil. Below, another extract of the same table:

Ser	ref.		0.2 ± 0.4	-14	1.6	-5.9
CG	P1		-1.9 ± 0.1	-12	0.0	-2.3
	N0 (0.40)		-0.5 ± 0.3	-14	-0.2	-5.2
Asn	ref.		-1.0 ± 0.4	-28	-0.1	-17.3
	CG	P5	-2.7 ± 0.1	-31	0.3	-4.2
		Nda (0.51)	1.9 ± 0.7	-28	-0.2	-20.6
		Nda (0.46)	2.0 ± 0.4	-23	-0.4	-13.9
		N0 (0.54)	-1.3 ± 0.3	-27	-0.2	-18.1

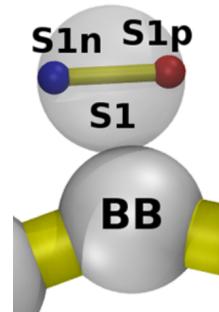


Figure 9. Extract from the same table in 8. Partial charges attributed to the beads are represented in parenthesis. On the right, a zoom on the CG structure of Asn.

In the table above, amino acids with polar side chains are considered. In Martini2.2p, these are improved with the use of polarized particles. For residues like Asn (see 9 on the right), the side chain consists of one virtual particle S1 carrying the LJ interaction and two sites S1n and S1p carrying a negative and a positive partial charge of 0.46 (represented in parenthesis) and no LJ. Importantly, these sites do not interact between each other but only with other beads. The sites S1n and S1p represent an embedded dipole introduced to cure the bad behaviour of the dimerization free energy in apolar solvent (see the last column in 9). If the artificial dipole were not introduced, the two side chain beads involved in the dimerization would only interact via LJ, hence the real interaction would be underestimated. Notice that the particle type in this case changes from P5 (the maximally polar bead) to the intermediate polar type N. This choice affects only the parameters ε_{ij} and σ_{ij} of the LJ interactions reported in the Martini interaction matrix. Since the representation of Asn still remains overall neutral, the introduction of polarizable water does not produce relevant net effects, as can be seen in the second-last column of 9.

As can be seen in 9, the Asn particle reassignment in bold provides a bad estimate of $\Delta\Delta G^{\text{WW}}$ and ΔG^{part} compared to the N0(0.54) choice. The authors explain that the choice of Nda(0.46) led to the appearance of a minimum near the membrane-water interface in the partitioning free energy profile (bottom right plot of 5) similar to the atomistic case, but at the same time, it could not reproduce the partitioning behaviour of Asn alone and in the WW pentapeptide complex.

An important consideration must be pointed out at this point: finding the proper particle assignment for the beads often involves huge compromises. Due to the simplified picture of molecular structures, subtle energetic and entropic

contributions or structural rearrangements may be oversighted by the CG model. Generally speaking, since it is not possible to provide a relevant simplification of the system and retaining all thermodynamic quantities at the same time, trade-offs must be used, as in the treated case. This lack of complete transferability is a common issue in CG models.

2.4 Versatility and pitfalls of the Martini FF

To conclude the theoretical discussion, based on the informations showed in the previous sections, we present a short qualitative discussion of the Martini approach.

The Martini approach from a high-level description of the system allows the simulation of systems containing several classes of biomolecules. In fact, the systematic parametrization of the beads allows for reusing them in different moieties. This represents a huge advantage with respect to physics-based *bottom-up* approaches, which require a reparametrization (or at least a substantial check of the parameters) when the system composition (solvent, ions, molecules) change.

CG models such as Martini are often cheaper due to simpler potential forms and thermodynamically more transferable, at the cost of structural inaccuracies. Nevertheless, the CG potential is actually a free energy, hence transferability is inherently limited and all thermodynamic properties cannot be correct at the same time.

Our last consideration concerns the nature of the kinetics in CG models. Due to the reduced number of degrees of freedom, kinetics is artificially enhanced. This may cause problems when simulating structural rearrangements or docking simulations of ligands. As it will be pointed out in section 3.2.1, this might be one of the causes of the sudden detachment of the Ca^{2+} ion in the CG simulation of Recoverin.

2.5 Background information on protein structure and function

In this section we present a brief introduction of the system discussed in the remaining part of the report.

The Recoverin protein belongs to the family of neuronal calcium sensors and it is expressed in the photoreceptor cells of the vertebrate retina. Its functional EF-hand motifs (EF2 and EF3) are crucial for the Ca^{2+} ion binding. Upon an increase in the intracellular Ca^{2+} concentration, a calcium-induced myristoyl switch allows the protein to associate reversibly with a rod outer-segment (ROS) disk membrane and to inhibit there the rhodopsin kinase¹¹, an enzyme involved in visual phototransduction.

The myristoyl switch is a saturated fatty acid covalently attached to the N-terminal glycine residue of Recoverin. Its parametrization in the Martini3 FF has been published only recently in a paper by Koukos et al.¹². Due to the difficulty in dealing with this new parametrization and (mostly) in the insertion of the myristoyl moiety inside the ROS membrane, we decided to simulate a simpler structure that we will henceforth call non-myristoylated Recoverin nmRec. This protein can be simulated in solution without the presence of the membrane since the anchoring provided by the myristoyl group is absent.

In what follows we will describe the set up of the CG simulation and compare (where possible) the CG and AA simulation parameters.

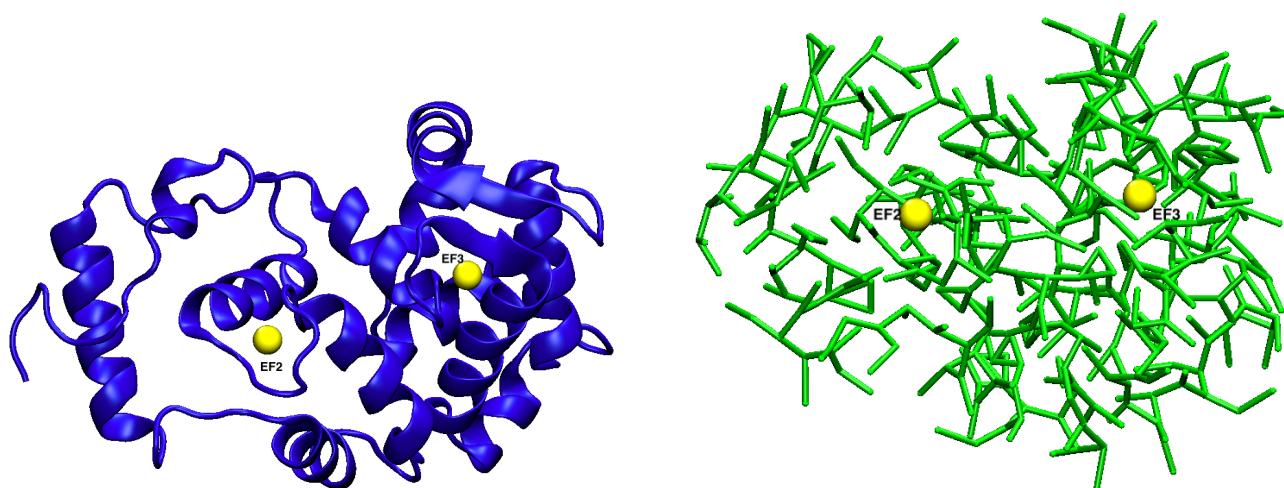


Figure 10. All-atom and coarse-grained representation of nmRec. The two calcium atoms are represented in yellow inside the EF2 and EF3 pockets. The secondary structure is highlighted in the atomistic case, while the CG beads are represented in licorice.

3 METHODS AND RESULTS

3.1 Preparation of the CG system

The procedure outlined in this section largely follows the [tutorial for the coarse-graining of proteins](#) provided on the Martini webpage. The aim of this section is to summarize the preparation steps, for further details about the algorithms and features imposed on the system, please refer to subsections [3.1.1](#) and [3.1.2](#). The input/output files together with the explicit commands employed in the preparation can be found on [this GitHub page](#).

Let us first briefly describe the set up of the atomistic simulation. The initial atomistic structure was taken from the [Protein Data Bank](#) entry 4YI8, namely, the Ca^{2+} bound structure of E153A non-myristoylated bovine recoverin. A mutation was performed using the *PDB reader* tool in Charmm-gui on the mutated Ala residue in order to insert the Glu residue of the wild-type structure (the wild-type version with both EF2 and EF3 Ca^{2+} was not present in the database). Missing residues were also resolved with the same tool. The force field Charmm36 february 2021 version was used to generate the topology. The atomistic system was then solvated with a physiological ion concentration of 150 mM of KCl. The system underwent NVT and NPT equilibration runs with restraints on heavy atoms. After these steps, an unrestrained run of 50 ns was performed at $T=310$ K. The final protein structure sampled from the unrestrained run was used as atomistic input for the CG simulation.

The mapping from the atomistic coordinate file represents the first step in the preparation of the system with Martini. The script *Martinize2.py* was used at this stage. The secondary structure, which can be computed with the [DSSP](#) web-tool was fed to the mentioned python script. Harmonic bonds constituting an elastic network between backbone beads separated by more than two beads² were implemented on the protein again using *Martinize2*. Lower and upper elastic bond cut-off were set to 0.5 nm-0.9 nm, while the spring constant was set to $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. After these parameters were fed in input, *Martinize2* returned the CG structure of the protein in output together with the topology file. Finally, a cubic box with size equal to 12 nm was prepared with the gromacs tool *editconf*.

After this first preparation step, the protein structure was relaxed in vacuum to allow the removal of clashes due to possible bad positioning of the beads. The potential $V(\mathbf{R})$ with $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ is minimized in a $3N$ -dimensional space by means of the steepest descent algorithm aided by a step-size regulation:

$$\mathbf{R}_{n+1} = \mathbf{R}_n + \frac{\mathbf{F}_n}{\max(|\mathbf{F}_n|)} h_n \quad \begin{cases} h_{n+1} = 1.2h_n & \text{if } V(\mathbf{R}_{n+1}) \leq V(\mathbf{R}_n) \\ h_n = 0.2h_n & \text{if } V(\mathbf{R}_{n+1}) > V(\mathbf{R}_n) \end{cases} \quad (3)$$

note that if the potential increases the positions are rejected and recomputed with the new step $0.2h_n$. The algorithm was executed until the maximum force in the system (i.e., $\max(|\mathbf{F}_n|)$) reached a value below $10 \text{ kJ mol}^{-1} \text{ nm}^{-2}$.

After this minimization, the system was then solvated with Martini water beads (representing four atomistic water molecules). A pre-equilibrated box of Martini water provided in the [Martini tutorial](#) was fed to the gromacs tool *solvate* to mimic a bulk-like solvent environment around the protein. Ions were then added according to a replacement criterium. More specifically, neutralizing sodium and chlorine ions were added up to a concentration of 150 mM. As in the vacuum case, a steepest descent minimization was performed with the same prescription of (3) and the same force threshold.

A coarse-grained protein system is unlikely to crash since the complex atomistic details and energetic barriers are smoothed out, determining a smaller "effective friction". For this reason, a single NPT equilibration with duration 2 ns was performed on the system. Harmonic position restraints with a spring constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ were placed on the backbone beads. Berendsen thermostat with ($\tau_T = 2$ ps, $T_{ref} = 310$ K) and barostat with ($\tau_P = 12$ ps, $P_{ref} = 1$ bar) were used for this run.

Lastly, in the unrestrained run we changed the thermostat and the barostat respectively to the v-rescale with ($\tau_T = 1$ ps, $T_{ref} = 310$ K) and Parrinello-Rahman with ($\tau_P = 2$ ps, $P_{ref} = 1$ bar). The v-rescale thermostat is similar to the Berendsen thermostat but with the addition of a stochastic term that leads to a proper sampling of the canonical ensemble:

$$dK = (K_0 - K) \frac{dt}{\tau_T} + \sqrt{\frac{K_0 K}{N_f}} \frac{dW}{\sqrt{\tau_T}}$$

² *Martinize2* does not generate elastic bonds between $i \rightarrow i+1$ and $i \rightarrow i+2$ beads since they are already connected by standard Martini bonds, angles and dihedrals.

Here dW is a Wiener stochastic process, K is the kinetic energy, $K_0 = k_B T N_f / 2$ is the target kinetic energy and N_f is the number of degrees of freedom. The Parrinello-Rahman barostat is an extended Hamiltonian method that introduces the box dimensions as additional degrees of freedom in order to maintain a desired pressure. Here we simply report the equation for the box vector \mathbf{b} :

$$\frac{d\mathbf{b}^2}{dt^2} = V \mathbf{W}^{-1} \mathbf{b}^{-1} (\mathbf{P} - \mathbf{P}_{\text{ref}})$$

where V is the volume of the box and \mathbf{W}^{-1} is the (inverse) mass parameter matrix (related to τ_P) which anisotropically adjusts the volume of the system.

As mentioned in theoretical discussion, Martini assumes a uniform screening with relative dielectric constant $\varepsilon_r = 15$. The reaction field method implementation in Gromacs requires the definition of an additional parameter ε_{RF} to make the potential zero at the cut-off distance r_c (commonly set to 1.1 nm in Martini). The expression used by Gromacs for the potential between particles i and j is:

$$V_{RF}(r_{ij}) = \frac{q_i q_j}{4\pi \varepsilon_0 \varepsilon_r r_{ij}} \left[1 + \frac{\varepsilon_{RF} - \varepsilon_r}{2\varepsilon_{RF} + \varepsilon_r} \left(\frac{r_{ij}}{r_c} \right)^3 - \frac{3\varepsilon_{RF}}{2\varepsilon_{RF} + \varepsilon_r} \frac{r_{ij}}{r_c} \right]$$

Below, a table reporting the remaining set-up parameters and algorithms employed:

	type	parameters
integrator	leap-frog integration	$\Delta t = 0.02$ ps
temperature coupling	Protein and non-protein	$T_{\text{ref}} = 310$ K
Pressure coupling in NPT	isotropic	$P_{\text{ref}} = 1$ bar
long-range electrostatics	Reaction field	$\varepsilon_r = 15, \varepsilon_{RF} = \infty$ $r_c = 1.1$ nm
cut-off scheme for non-bonded interactions	Verlet neighbour list	$r_{\text{Coulomb}} = 1.1$ nm $r_{\text{vdw}} = 1.1$ nm
compressibility	/	$\kappa_T = 3 \times 10^{-4}$ bar $^{-1}$
Elastic network	/	$k = 500$ kJ mol $^{-1}$ nm $^{-2}$ $r_{\text{low}} = 0.5$ nm, $r_{\text{up}} = 0.9$ nm

Table 1. Table with the parameters and algorithms used both in the NPT equilibration and in the unrestrained run.

A unrestrained run of 1 μ s was performed for the CG system. Usually studies in the field of CG simulation rely on tenths or hundreds of μ s, here instead we give a demonstration of a protocol rather than providing a detailed analysis on extended trajectories.

3.1.1 Considerations on the employed parameters

The first thing to be noticed is that the integration time step (in table 1) is approximately one order of magnitude larger than the standard Δt in atomistic simulations. The chosen CG time step is able to capture the relevant dynamics and it is also sufficient for the stability of the system. As a side note, a time step of 20 fs would still be too high (or at least critical) for more detailed physics-based bottom-up potentials obtained through iterative Boltzmann inversion (IBI) or force matching (FM).

The coupling parameters τ_T and τ_P controlling temperature and pressure fluctuations around their reference values are approximately one order of magnitude larger than the corresponding atomistic ones (of the order of tenths of ps). Again, the loss of degrees of freedom and the subsequent slowing down of the CG dynamics lower the frequency of the fluctuations around the target kinetic energy and pressure, allowing for larger values of τ_T and τ_P .

From table 1 it can be noticed that shorter cut-off distances for non-bonded interactions can be used without significantly affecting the accuracy of the simulations, while usually AA simulations display cut-offs around 1.2-1.4 nm. The reason for that can be attributed to two concurrent features of CG models:

1. the beads have a larger van der Waals radius (0.264 nm for regular R, 0.230 nm for small S and 0.191 nm for tiny T beads) than atomistic particles.

2. On average there are fewer particles at the radial cut-off distance compared to the atomistic case, hence we are neglecting only few contributions.

The PME approach, which uses the formalism of the lattice sum in the reciprocal space in the treatment of long-range electrostatics, becomes preferable to the reaction field method when the presence of ions, membranes or vesicles becomes relevant. For the present system we do not use this enriched treatment of electrostatics, although, as discussed later in section 3.2.1, the PME could be employed for a more realistic dynamics of the divalent calcium ion.

Lastly, due to the absence of explicit hydrogen bonding directionality, we added an elastic network to the Martini topology in order to maintain the overall 3D structure of the protein, namely, the tertiary structure. Importantly, the elastic network approach does not capture structural changes in the protein, which can instead be sampled using the Go-Martini approach where the elastic bonds are replaced by LJ potentials. The lower and upper cut-off values reported in table 1 are generally sufficient for not affecting the bonded interactions and being spatially contained at the same time, while the value of $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ is largely employed by Martini users.

3.1.2 The Martinize2 script

Informations reported in this section are largely taken from the article by Kroon et al.¹³.

The core of the coarse-graining procedure lies in the action provided by *Martinize2.py*. The script is largely based on the *vermouth* (VERSatile MODular Universal Trasformation Helper) python library, i.e., a collector of graph algorithms that operate on molecular structures and topologies.

Martinize2 enables the preparation of input files for arbitrary (bio)polymers, starting from atomistic structures. The pipeline of the program can be summarized very shortly:

1. Reading input files: this initial parsing operation is fundamental in generating graphs that will be exploited in the rest of the pipeline. Since atom (vertices) identification is governed by their connectivity, it is recommended to use a PDB input file with bonds (edges) specified in the *conect* records.
2. Repairing graphs: each residue is compared against its canonical definition stored in library files¹³. The *maximum common induced subgraph* (see the definitions in the appendix 5) between the residue and its reference is computed by the script. At the end of the process, each residue is repaired if atoms are misplaced and it is labelled with a known *block* taken from the input FF (Martini3 in the treated case). Up to this point no mapping has yet been performed on the system.
3. Resolution transformation: mapping to CG level in *Martinize2* is done with a multistep *subgraph isomorphism* procedure aimed at identifying patterns in the input atomistic molecule that match the definitions of coarse-grained beads. *Subgraph isomorphism* is basically the problem of finding one graph (the pattern) within another graph (the input structure) such that the edge and vertex relationships are preserved (see 5). Edges between the CG beads are generated based on the connectivity of the input molecule.
4. Generating the topology: structures called *links* are used to describe interactions between the mapped residues. A *link* consists of nodes and edges describing the molecular fragment it applies to, as well as the associated changes in MD parameters (bond length, bond angle, and force constants). *Links* determine the structural connectivity of the CG representation, in the specific case of proteins, they define interactions that depend on the protein sequence or secondary structure.
5. Post-processing operations, such as building an elastic network or adding Go sites.
6. Writing output files: topology and structures are written to files and ready to be fed to MD engines such as Gromacs.

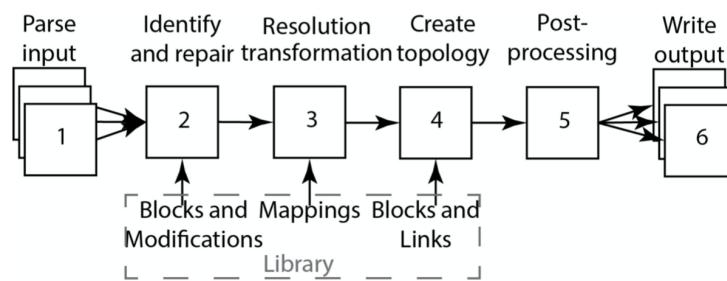


Figure 11. Pipeline of *Martinize2* schematically represented, image taken from Kroon et al.¹³.

3.1.3 Post-processing of the simulations

Unwanted movements of the protein outside the simulation box were removed by means of the `trjconv` tool provided by Gromacs by exploiting the option `-pbc mol` and `-center`. The same tool allowed the alignment of the trajectory using a least squares fit with respect to the backbone atoms (C, C- α and N for each residue) of the structure. The option `-fit rot+trans` option was used for this operation.

3.2 Structural properties compared with the all-atom simulation

In order to quantify the similarity between the protein structures, it is possible to calculate the Root Mean Square Deviation RMSD. This can be computed over an arbitrary selection of atoms and the weights are usually chosen to be the masses of the atoms.

$$\text{RMSD}(t) = \sqrt{\frac{\sum_i^N m_i (\mathbf{r}_i(t) - \mathbf{r}_i^{\text{ref}})^2}{\sum_i^N m_i}} \quad (4)$$

In order to provide a proper comparison between the atomistic and CG models, the definition of the vectors \mathbf{r}_i must be consistent. In this section, we assume that \mathbf{r}_i represent the position of the backbone beads of the i -th residue. Consistently with the Martini3 mapping, in the AA case we define the vector \mathbf{r}_i as the center of geometry of the backbone atoms and hydrogens of the i -th residue.

The RMSD in equation (4) does not give direct information about the fluctuations of specific regions of the protein, but it gives an idea of the overall deviation from the structure from a reference frame. Below, in figure 12, we present the RMSD computed with the backbone beads of the EF2 pocket (in red) and the whole structure (in green). The AA case is also shown on the top right. Henceforth in the report we shall assume that the EF2 pocket consists of the residues from 70 to 90 (colored in red in 12).

Unwanted motions of the protein are removed in both AA and CG trajectories according to the procedure mentioned in 3.1.3.

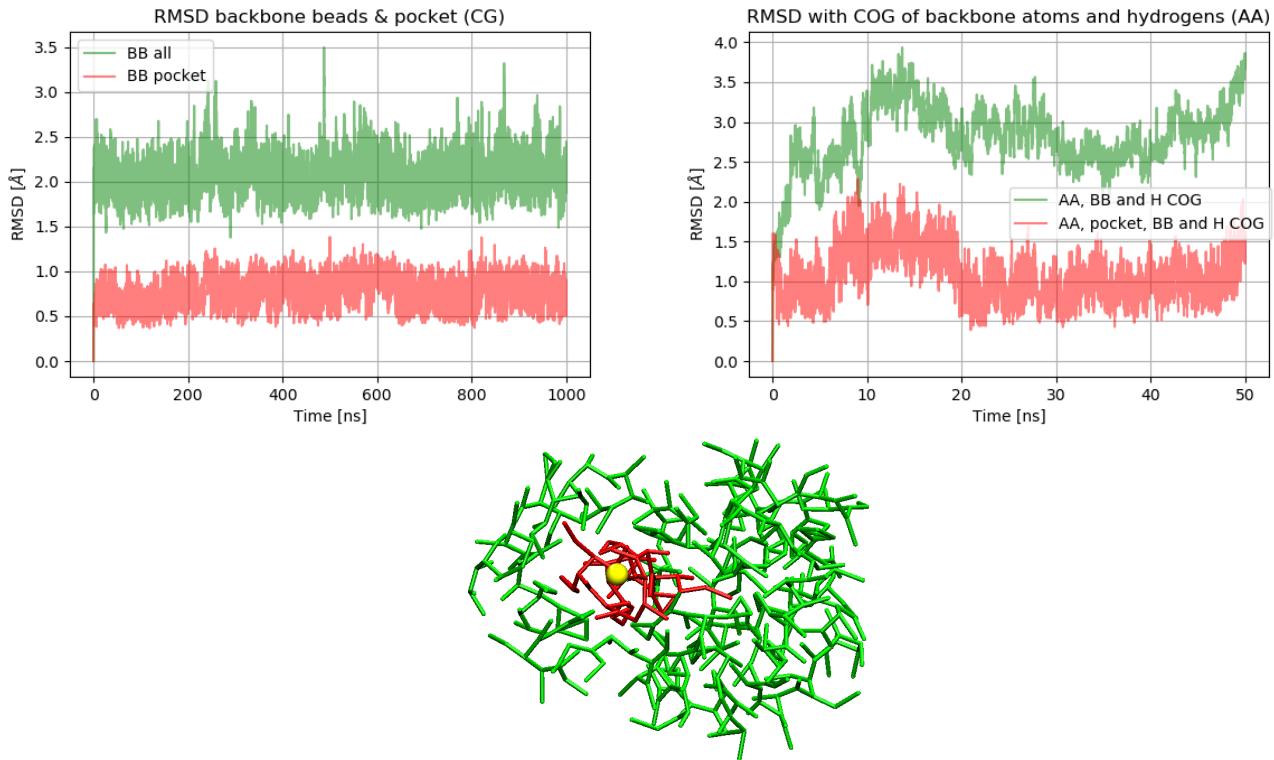


Figure 12. The AA and CG RMSD values are comparable, even though the atomistic RMSD indicates that the system is still out of an equilibrium configuration. Below, a drawing of the EF2 pocket (in red).

Another important property to look at is the Root Mean Square Fluctuation RMSF. This allows the identification of the most mobile and most stable domains of the protein. We computed the RMSF of each backbone bead and atomistic COG using the following expression:

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{j=0}^M (\mathbf{r}_i(j\Delta t) - \langle \mathbf{r}_i \rangle)^2}$$

where the $\langle \mathbf{r}_i \rangle$ represents the average of the position of the i -th backbone bead (or the average of the COG in the AA case) over the trajectory of lenght $T = M\Delta t$. The RMSF constitutes a valuable tool for judging whether the secondary and tertiary structure are sufficiently stable or not.

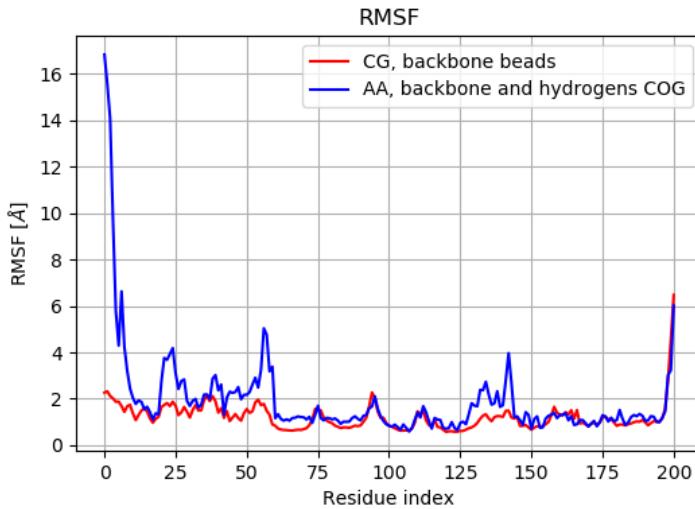


Figure 13. Except the first 9 residues constituting a loop structure subject to large fluctuations, a general agreement is obtained. This fact proves that the tertiary structure, i.e. the three-dimensional organization of the different domains, is conserved during the trajectory.

We conclude this section with the radius of gyration, a measure of the globularity of the protein.

$$R_g(t) = \sqrt{\frac{\sum_i^N m_i (\mathbf{r}_i(t) - \bar{\mathbf{r}}(t))^2}{\sum_i^N m_i}}$$

The notation $\bar{\mathbf{r}}(t)$ indicates an average over the positions of the beads/atoms in the frame at time t . In the atomistic case $\bar{\mathbf{r}}(t)$ was simply taken to be center of geometry of all the backbone atoms and hydrogens of the structure. Below, we present a comparison of the CG and AA $R_g(t)$:

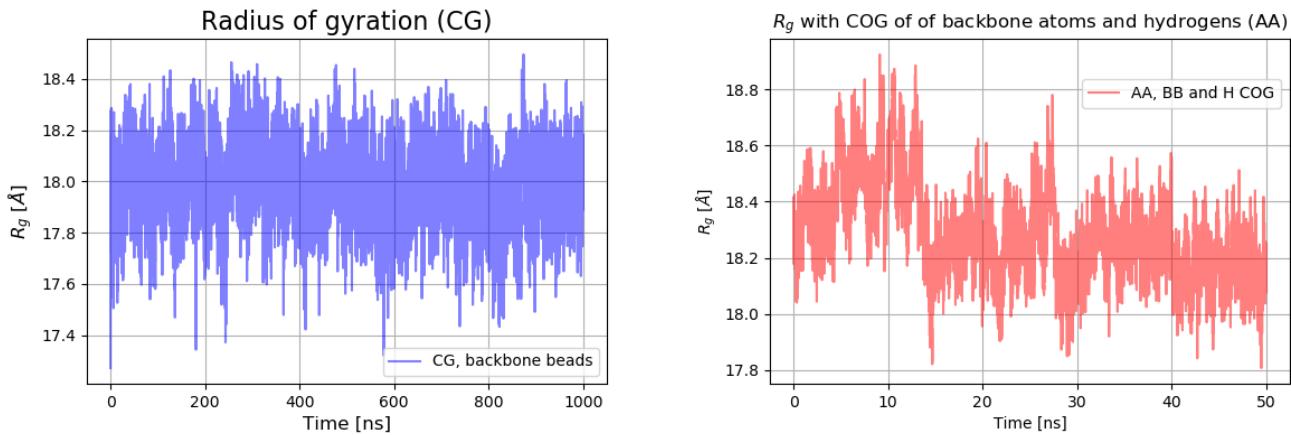


Figure 14. The CG and AA values of R_g are compatible. The CG plot shows that the protein does not expand or shrink significantly during the trajectory.

We remind that quantities like RMSDs, RMSFs and R_g should be computed on multiple CG trajectories employing different values of the elastic network spring constant. Typically, values ranging between $k = 500 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ are tested and optimized against structural data. Here we do not provide this comparison between simulations due to the demonstrative nature of the project.

3.2.1 Solvent accessible surface area (SASA) and minimum distance of the calcium ion from EF2

The atomistic SASA of the binding pocket EF2 was computed using a probe with the standard radius of 0.14 nm (representing the approximate radius of a water molecule), while the CG SASA was computed with a larger probe of 0.191 nm (the vdW radius of T tiny beads). The van der Waals radii of the CG beads fed in input to the *sasa* tool of Gromacs were provided in Alessandri et al.⁴.

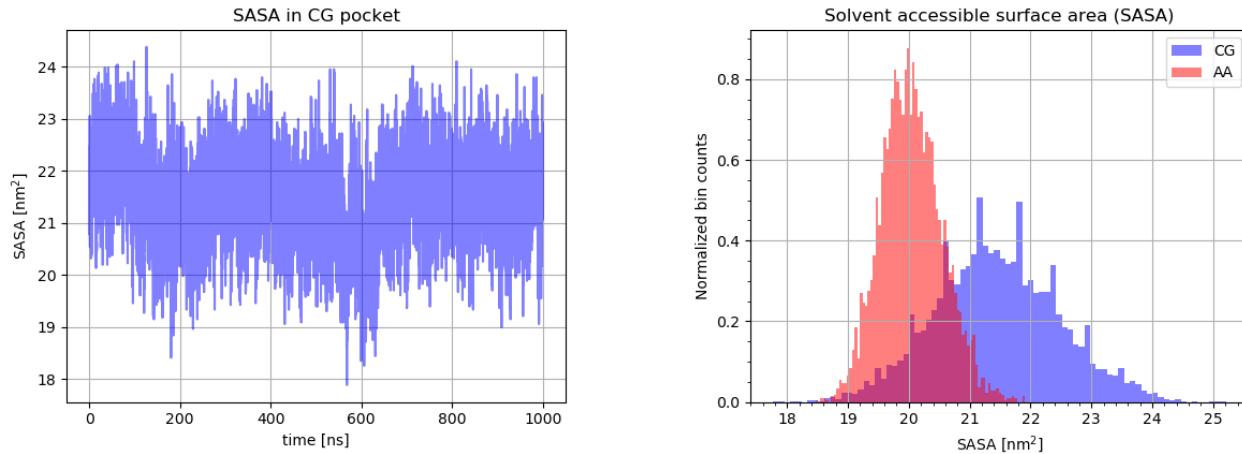


Figure 15. The two distributions of the SASAs along time superimpose. However, a slightly higher surface and a broader spectrum are measured for the CG pocket. The large fluctuations of the CG pocket are due to the sizes of the side chain and backbone beads.

To conclude our analysis, we present the minimum distance of the Ca^{2+} ion from EF2. During the 50 ns of atomistic simulation, the ion remains inside the pocket mainly because of the interaction with six coordinant oxygens. Instead, in the CG simulation, the Ca^{2+} ion detaches from the structure after approximately 250 ns. After this event, it samples an unbound state for approximately 80 ns and then binds again with the protein in a site different from the EF2 pocket. This behaviour is shown in the plot of the minimum distance from EF2 on the left of figure 16. On the right of 16, the distributions of the minimum distances for the AA and CG (only for the first 200 ns) cases are represented.

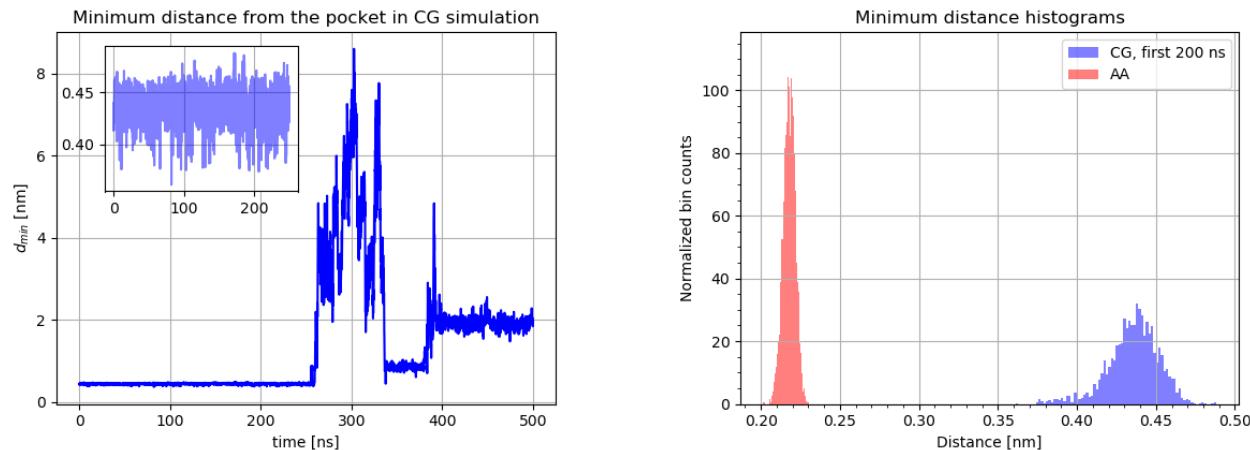


Figure 16. Only the first 500 ns (200 ns) of the CG simulation are represented on the left (right).

Concerning the plot on the left, we point out that the use of the reaction field technique for the electrostatics together with the standard Martini water deeply influence the results for the following reasons:

- the artificial screening of the electrostatic does not account for heterogeneous charge distribution of the system.
- The four-to-one mapping of water beads is not sufficient to recover a proper hydration of the pocket, leading to the unexpected behaviour of the ligand.

Generally speaking, the combination of Particle Mesh Ewald (PME) with polarizable water is more reliable when treating divalent ions.

Regarding the plot on the right of 16 representing the minimum distances of Ca^{2+} from EF2, we note a discrepancy 0.20-0.22 nm between the two distributions. The origin of such a gap mainly lies in the inclusion of the first hydration shell in the Ca^{2+} CG bead. The calcium ion is in fact represented by a small sized bead (S) having a van der Waals radius of 0.23 nm, which inevitably increases the distance of closest approach for the other beads.

4 CONCLUSIONS

Our discussion explored relevant aspects for a CG simulation of a protein with Martini. The parametrization of amino acids, the validation tests conducted on them and the description of the set up of the simulation constitute notable points addressed in the discussion. We did not expanded the discussion on the parametrization of lipids, nucleic acids or other molecules, although a study of these compounds is required for a complete analysis of the FF. Moreover, the validations described for amino acids are not exhaustive because of the wide range of techniques employed by the authors of the force field.

Concerning the practical implementation reported in this report, we noticed a general agreement between AA and CG for simple and basic structural properties, although apparent similarity must always be taken with a grain of salt: first of all, the simulated times (1 μs for CG and 50 ns for AA) are too short for real quantitative assessments, second, the quantities we analyzed are produced from a model that has been optimized for other quantities, especially binding and partitioning free energies. Thus, we think that a study of the myristoylated protein interacting with the membrane (see section 2.5) could be useful for gaining insights on the insertion of the myristoyl group inside the membrane and the subsequent orientation of the protein. We believe that in this way the power of the Martini force field would be fully exploited.

5 APPENDIX

6 Useful definitions for *Martinize2*

Graph: a graph $G = (V, E)$ is a collection of nodes (V) connected by edges (E). *Martinize2* makes use of undirected graphs, where all the edges are bidirectional (i.e., they do not point in a specific direction).

Graph isomorphism: a graph isomorphism m between graphs $H = (W, F)$ and $G = (V, E)$ is a bijective mapping $m : V \mapsto W$ such that any two vertices u and $v \in V$ are adjacent in G if and only if $m(u)$ and $m(v)$ are adjacent in H .

Induced subgraph: an induced subgraph is formed from a subset of the vertices of a starting graph and all of the edges (from the original graph) connecting those vertices in the subset.

Induced subgraph isomorphism: it is a graph isomorphism with two added constraints: $|H| \leq |G|$ (H must have a smaller number of nodes) and equivalent nodes not connected in G are not connected in H .

Maximum common induced subgraph of two graphs G and H : it is an induced subgraph of both G and H , and that has as many vertices as possible.

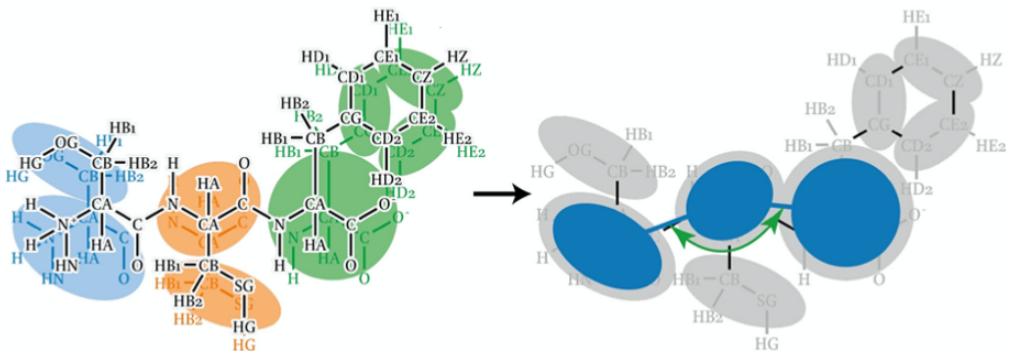


Figure 17. Image taken from Kroon et al.¹³. On the left, a molecular fragment seen at two different resolutions with mappings represented in blue, orange and green. This operation corresponds to the step 3 described in section 3.1.2. On the right, *links* (in blue) are applied between the mapped backbone beads as described in step 4 of 3.1.2.

References

1. Monticelli, L. *et al.* The martini coarse-grained force field: extension to proteins. *J. chemical theory computation* **4**, 819–834 (2008).
2. Ingólfsson, H. I. *et al.* The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **4**, 225–248 (2014).
3. Alessandri, R. *et al.* A practical introduction to martini 3 and its application to protein-ligand binding simulations. In *A Practical Guide to Recent Advances in Multiscale Modeling and Simulation of Biomolecules*, 1–1 (AIP Publishing LLC Melville, New York, 2023).
4. Alessandri, R. *et al.* Martini 3 coarse-grained force field: small molecules. *Adv. Theory Simulations* **5**, 2100391 (2022).
5. Souza, P. C. *et al.* Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. methods* **18**, 382–388 (2021).
6. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The martini force field: coarse grained model for biomolecular simulations. *The journal physical chemistry B* **111**, 7812–7824 (2007).
7. de Jong, D. H. *et al.* Improved parameters for the martini coarse-grained protein force field. *J. chemical theory computation* **9**, 687–697 (2013).
8. De Jong, D. H. *et al.* Determining equilibrium constants for dimerization reactions from molecular dynamics simulations. *J. computational chemistry* **32**, 1919–1928 (2011).
9. Bennett, C. H. Efficient estimation of free energy differences from monte carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
10. Singh, G. & Tieleman, D. P. Using the wimley–white hydrophobicity scale as a direct quantitative test of force fields: the martini coarse-grained model. *J. chemical theory computation* **7**, 2316–2324 (2011).
11. Timr, S., Kadlec, J., Srb, P., Ollila, O. S. & Jungwirth, P. Calcium sensing by recoverin: effect of protein conformation on ion affinity. *The journal physical chemistry letters* **9**, 1613–1619 (2018).
12. Koukos, P. I. *et al.* Martini 3 force field parameters for protein lipidation post-translational modifications. *J. Chem. Theory Comput.* **19**, 8901–8918 (2023).
13. Kroon, P. *et al.* Martinize2 and vermouth: unified framework for topology generation. *elife* **12**. RP90627 (2023).