OLIVES: A Go-like Model for Stabilizing Protein Structure via Hydrogen Bonding Native Contacts in the Martini 3 Coarse-Grained Force Field

Kasper B Pedersen, Luís Borges-Araújo, Amanda D Stange, Paulo C. T. Souza, Siewert-Jan Marrink, Birgit Schiøtt

OLIVES: A Gō-like Model for Stabilizing Protein Structure via Hydrogen Bonding Native Contacts in the Martini 3 Coarse-Grained Force Field

Kasper B. Pedersen,[*,†] Luís Borges-Araújo,[‡,¶] Amanda D. Stange,[†] Paulo C. T. Souza,[‡,¶] Siewert J. Marrink,[§] and Birgit Schiøtt[*,†,‖]

†*Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark*

‡*Laboratoire de Biologie et Modélisation de la Cellule, CNRS, UMR 5239, Inserm, U1293, Université Claude Bernard Lyon 1, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France*

¶*Centre Blaise Pascal de Simulation et de Modélisation Numérique, Ecole Normale Supérieure de Lyon, 46 Allée d'Italie, 69364, Lyon, France*

§*Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands*

‖*Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark*

E-mail: *kasperbusk@chem.au.dk; *birgit@chem.au.dk

Abstract

Coarse-grained molecular dynamics simulations enable the modeling of increasingly complex systems at millisecond timescales. The transferable coarse-grained force field Martini 3 has shown great promise in modeling a wide range of biochemical processes, yet folded proteins in Martini 3 are not stable without the application of external bias potentials like elastic networks or Gō-like models. We herein develop an algorithm, called OLIVES, which identifies native contacts with hydrogen bond capabilities in coarse-grained proteins and use it to implement a novel Gō-like model for Martini 3. We show that the protein structure instability originates, in part, from the lack of hydrogen bond energy in the coarse-grained force field representation. By using realistic hydrogen bond energies obtained from literature *ab initio* calculations, it is demonstrated that protein stability can be recovered by the reintroduction of a coarse-grained hydrogen bond network and that OLIVES removes the need for secondary structure restraints. OLIVES is validated against known protein complexes, and at the same time addresses the open question of whether there is a need for protein quaternary structure bias in Martini 3 simulations. It is shown that OLIVES can reduce the number of bias terms, hereby speeding up Martini 3 simulations of proteins by up to $\approx$ 30 % on GPU architecture compared to the established GoMARTINI Gō-like model.

Introduction

Coarse-grained molecular dynamics simulations allow for the modeling of increasingly complex systems.[1] A recent example is the modeling of an entire cell[2] in the Martini 3 force field with a full interior of molecular components with proteins comprising as much as 40 % of the intracellular volume.[2,3] However, folded proteins in Martini 3 (version 3.0.0) are not yet stable without the application of external bias potentials,[4] like elastic networks[5] or Gō-like models.[6] Furthermore, additional secondary structure-specific potentials[4] (defined using the DSSP algorithm[7]), need to be employed to limit fold-breaking fluctuations, however, these also restrict the modeling of conformational changes and transient folds, which are often desired in computational studies of proteins.

Hydrogen bonds are essential to life as we know it, and the formation of hydrogen bond networks is a key driver of structure in biochemistry,[8] ubiquitous in critical cell processes such as nucleotide base pairing[9] and protein folding.[10,11] However, the coarse-graining procedure in Martini 3 places hydrogens together with heavy atoms into larger Lennard-Jones (LJ) particles and effectively averages out the energy contribution of directional hydrogen bonding. In this work, we investigate whether the secondary and tertiary structure instability in Martini 3 proteins could originate from missing hydrogen bond energy. We introduce an algorithm called OLIVES which identifies native contacts with hydrogen bond capabilities in coarse-grained structures, as an approximation for the atomistic hydrogen bond network, and uses it to implement a novel Gō-like model for Martini 3 proteins. By using realistic hydrogen bond energies obtained from literature *ab initio* calculations,[12–14] we show that Martini 3 proteins can be stabilized by the introduction of LJ bonds between hydrogen bonding beads in native contacts.

As we embarked on creating a new Gō-like model concept for coarse-grained simulations of proteins, we also established further goals for the model's properties that we think will be advantageous for the Martini 3 community and for supporting future developments. To facilitate the advancement of the Martini 3 protein model, we differentiate between the secondary-

and tertiary networks, such that they could potentially be optimized independently and any imbalances could expose problems in the underlying protein model. Specifically, we wanted to develop a model that stabilizes protein secondary structure without the need for the secondary structure restraints currently used, such as harmonic bonds in β-sheets and high force constant dihedrals in α-helices.[4] Importantly, this could unlock the ability to do more realistic (un)folding simulations of proteins, a goal also addressed in the original work behind GoMARTINI.[6] Having a foldable secondary structure will also be a first step towards enabling protein conformational changes in Martini 3, which often require alternating secondary structures.[15] Additionally, we searched for a model with minimal external bias. Since a Gō-like model can essentially be regarded as a force field correction, the Martini 3 protein model should be built with a minimal set of external bias potentials which will in turn increase the transferability and consistency of Martini 3 force field parameters between proteins and other molecule classes not using a Gō-like model. Naturally, reducing the number of restraints would also have the benefit of making simulations faster. Finally, we wanted to have an easy-to-use model that can be applied by a single script using the command line.

We validate the ability of OLIVES to stabilize known protein complexes situated both in solution and in phospholipid bilayers. By comparing to atomistic simulations of the same complexes, we address the open question of whether there is a need for protein quaternary structure bias in Martini 3 simulations.[16,17] Oligomerization studies are commonly performed using Martini; however, these are often between transmembrane proteins,[18–20] perhaps because Martini started as a phospholipid force field.[21,22] However, oligomerization between soluble proteins has been less studied using Martini 3,[16] and an assessment of the ability of Martini 3 to stabilize known protein partners in both the solution phase and in bilayers is therefore needed. Specifically, because the tertiary structure in Martini is unstable, it is likely that quaternary structure stability is also underestimated since the underlying physical interaction principles are the same.

In the following, it is shown that the OLIVES Gō-like model excellently reproduces in-

tramolecular native contacts[23] in line with the established structure-biased method of the GoMARTINI Gō-like model,[6] even in proteins with a large fraction of hydrophobic contacts not explicitly treated by OLIVES. Furthermore, we find that OLIVES can stabilize the Martini 3 protein model without using DSSP secondary structure restraints. We demonstrate that OLIVES reduces the number of bias potentials, which speeds up simulations by up to ≈ 30 % on GPU hardware compared to the GoMARTINI Gō-like model.[6] We also show that the stability of solution protein complexes cannot be reproduced without using a quaternary bias network, while transmembrane protein complexes are stable in line with previous work.[16,18–20] We conclude the paper with a discussion of the underlying reasons why the stability of Martini 3 solution complexes is underestimated and discuss future directions for the improvement of the Martini 3 protein model.

Methods

The OLIVES Gō-like Model

In order to create a Gō-like model based on hydrogen bond networks, we searched the literature for *ab initio* calculations of hydrogen bond energies[12–14] and use the data to create a hydrogen bond donor/acceptor energy matrix (Table 1). The collected energy values lie in the range of 3.5-8.0 kcal/mol. This is in line with other *ab initio* calculations[24] and literature reviews reporting hydrogen bond energies in the range of 5-6 kcal/mol[10,11] in the gas phase, representative of the protein interior (dielectric constant of ≈3),[25] though the bond energy can be as low as 1.5-2.0 kcal/mol in a water environment.[26] By using an energy matrix for the construction of a hydrogen bond network between different hydrogen bond donor/acceptor moieties, OLIVES increases the chemical specificity of the Gō-like bonds, which in turn is more representative of the local protein environment, compared to uniform energy across all bias potentials as currently used in elastic bond networks[5] and the GoMARTINI[6] Gō-like model. We classify the hydrogen bonding beads in the Martini 3 protein model according to Figure 1. As we were not able to find a complete matrix for all possible combinations of hydrogen bond partners in proteins, we have to simplify a few interactions, consequently, we set the hydroxyl and carboxylic acids acceptor energy to be equivalent to a ketone acceptor.[12] Likewise, we did not find specific values for the lysine primary amine and arginine guanidino groups, which we instead set to equal the energy of imidazole partners.[12] We include π-acceptor energies[13] in the model, although because of the coarse-grained granularity of aromatic rings in Martini 3, we split the interaction energy in two between the outermost beads in aromatic groups (Figure 1, PHE, TYR, TRP).

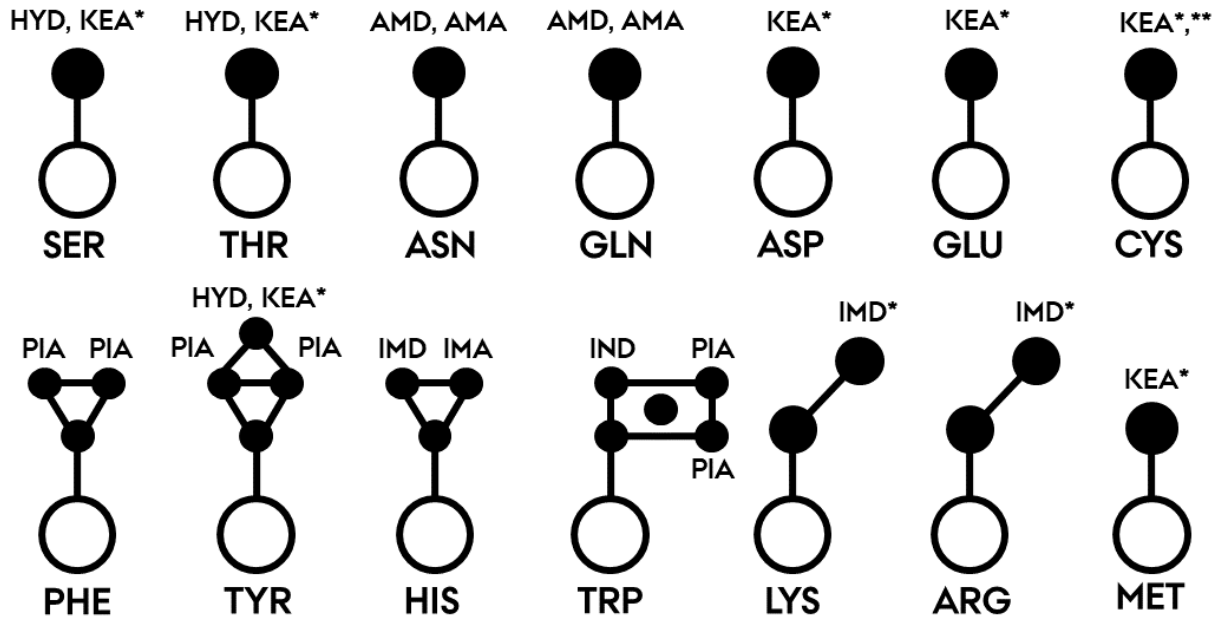


Figure 1: Definition of OLIVES hydrogen bonding beads based on the Martini 3 (v3.0.0) protein model. White circles (○) indicate the residue backbone beads while black circles (●) indicate side chain beads. The beads are classified as follows: (HYD) hydroxyl donor, (AMD) amide donor, (IMD) imidazole donor, (IND) indole donor, (KEA) ketone acceptor, (AMA) amide acceptor, (IMA) imidazole acceptor, and (PIA) π acceptor. The backbone peptide beads are classified as both an amide donor and an amide acceptor. *Missing hydrogen bond energy reference for the moiety and therefore approximated by the closest chemical species in the energy matrix. **Cysteine-cysteine partners are excluded to avoid unrealistic networks near disulfide bridges, but can otherwise also act as a HYD* hydrogen bond donor.

Table 1: Hydrogen bond energy matrix collected from *ab initio* calculations in the literature[12–14] in units of kcal/mol. (HYD) hydroxyl donor, (AMD) amide donor, (IMD) imidazole donor, (IND) indole donor, (KEA) ketone donor, (AMA) amide acceptor, (IMA) imidazole acceptor, (PIA) π acceptor. Note that the majority of the results were obtained by DFT calculations, which have a typical error of 2 kcal/mol.[27] *π acceptor energies are divided by 2 internally, spreading the energy across two beads as illustrated in Figure 1. **Average of two results in the original work by Du et al.[13]

[kcal/mol]	HYD	AMD	IMD	IND
KEA	3.54	4.06	5.78	5.12
AMA	4.51	5.44	6.92	6.27
IMA	6.24	6.55	7.96	7.11
PIA*	4.97**	5.31	6.17	2.10

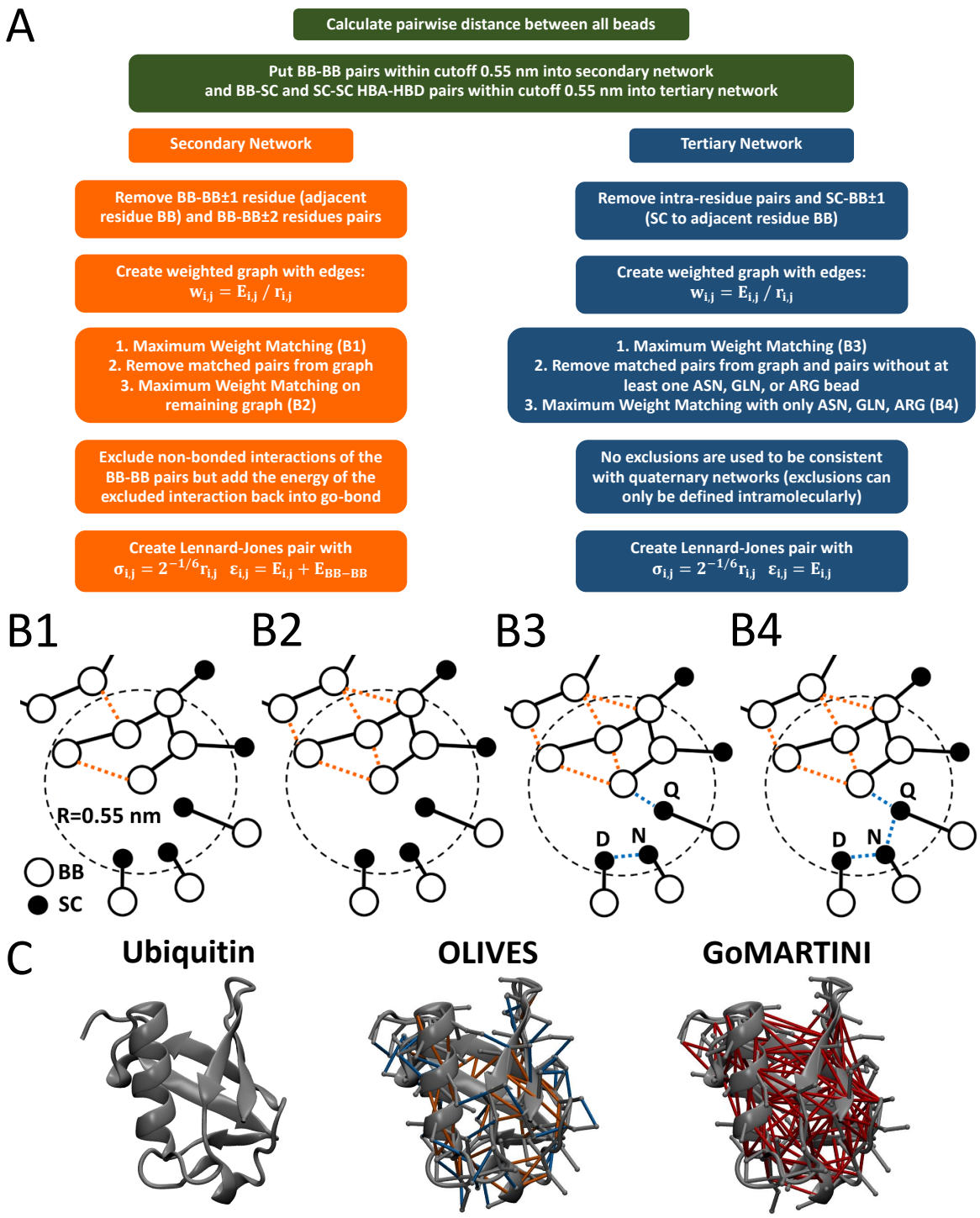


Figure 2: (A) Outline of the OLIVES algorithm for the secondary- (orange) and tertiary (blue) networks. (B) Illustration of the bond pattern created by maximum weight matching in a region with beta-sheet secondary structure. White circles (○) indicate the residue backbone beads while black circles (●) indicate side chain beads. In the illustration, the first and second round of secondary structure bond matching (B1, B2) connects the β-sheet (orange stripes), and the tertiary bond matching (B3, B4) connects the β-sheet to a nearby region via a glutamine residue (blue stripes). (C) The OLIVES network (orange, blue) illustrated for Ubiquitin with the GoMARTINI[6] network (red) shown for reference.

The OLIVES algorithm identifies native contacts with hydrogen bond capabilities in coarse-grained proteins. The algorithm is outlined in Figure 2,A. In the initial step, OLIVES calculates all pairwise distances between beads within a 0.55 nm cutoff in the input coarse-grained structure and uses the classification scheme outlined in Figure 1 to group the bead pairs into hydrogen bond donor/acceptor pairs. These are further divided into two networks which we denote the secondary network and the tertiary network. The secondary network exclusively contains pairs between backbone beads while the tertiary network consists of pairs from side chain beads to backbone beads and side chain beads to side chain beads. To obtain a realistic-looking secondary structure bond network we ignore pairs to the two adjacent backbone beads. A similar strategy is employed in the GoMARTINI[6] model. In the tertiary network, we ignore side chain interactions with the backbone of the adjecent residues. While such patterns do exist in proteins, the OLIVES algorithm identifies fewer false positive hydrogen bond partners using this rule. After this procedure, we are left with two (possibly disjoint) undirected graphs for the secondary and tertiary networks, respectively. These are converted to weighted graphs using weights $w_{ij} = E_{ij}/r_{ij}$ with E_{ij} being the pair interaction energy (Table 1) and r_{ij} is the distance between the beads given by the input structure. OLIVES then constructs a bond network by performing repeated rounds of maximum weight matching[28] (using NetworkX[29] v2.3) pairing hydrogen bond partners such that $\sum w_{ij}$ is maximized. In the secondary network, the backbone beads are allowed to form two hydrogen bonds each, which is obtained by first performing a maximum weight matching and then removing matched pairs from the graph, followed by a second round of matching. The pattern resulting from the maximum weight matching is illustrated in Figure 2,B1-4. In cases where the graph contains an odd number of vertices, some partners are not matched, which is solved by a third round on the small remaining graph. We found this final step to be closing some small gaps in the network important for the formation of bond patterns in helices. Note that maximum weight matching can be performed by forcing maximum cardinality[28] i.e. maximizing the number of edges, which we have tested

but led to generally worse results in terms of protein stability. In the tertiary network, the graph does not distinguish between side chain - backbone and side chain - side chain pairs. The first round of weight matching leads to an almost complete network. However, during testing, we found that it is very important that ASN, GLN, HIS, and ARG side chains are allowed to form two hydrogen bonds. We therefore perform a second round of matching on the remaining graph with the rule that each pair must contain an ASN, GLN, or ARG bead (HIS is not included because it can form bonds at both the epsilon and delta positions in the first round, due to the coarse-grained mapping, Figure 1).

Because the current protein mapping scheme in Martini 3 (v3.0.0) uses regular-sized beads (4-1 mapping) in the backbone,[1] forward mapping of atomistic proteins often leads to a substantial overlap of the backbone LJ spheres resulting in repulsive interactions, especially in secondary structures where the backbone is tightly packed. These non-bonded interactions are excluded when using both elastic networks and the GoMARTINI model, which is a contributing factor to the stabilization of these models. We also find that non-bonded exclusions are needed between backbone beads. However, since exclusions cannot be defined intermolecularly in GROMACS[30,31] we do not use exclusions in the tertiary network for consistency with quaternary networks. This means that the OLIVES interaction energy in the tertiary network is added on top of the non-bonded interactions present in the Martini 3 interaction matrix.[1] Because the non-bonded interactions are excluded from the secondary network, we add the non-bonded energy back onto the energy of the OLIVES bond. The resulting secondary and tertiary network pairings are then collected and used to create LJ pairs with $\sigma_{ij} = 2^{-1/6} r_{ij}$ with $\epsilon_{ij} = E_{ij} + E_{BB-BB}$ or $\epsilon_{ij} = E_{ij}$, for the secondary and tertiary networks, respectively, where E_{BB-BB} is the interaction strength between backbone beads in Martini v3.0.0 (bead type P2). Figure 2,C shows the OLIVES model visualized for Ubiquitin together with the bias network created by the GoMARTINI model for comparison.

Choice of Test Systems

We choose to test four diverse protein complexes to validate the ability of OLIVES to stabilize both secondary, tertiary, and quaternary structures in Martini 3 proteins. The four test systems consist of the solution complex of the Ubiquitin-associated domain (UBA) from Cbl-b Ubiquitin ligase in complex with Ubiquitin[32] (PDB: 2OOB); the solution complex of Barnase–Barstar[33] from *B. amyloliquefaciens* (PDB: 1BRS); the transmembrane homodimer of outer membrane phospholipase A (OMPLA)[34] from *E. coli* (PDB: 1QD6); and the homodimer of a *E. coli* homolog of transmembrane ClC chloride channels[35] (PDB: 1OTS). The test systems will in the remainder be denoted by the following names: UBA domain/Ubiquitin, Barnase/Barstar, OMPLA homodimer, and ClC homodimer.

The test set contains 4 unique protein complexes with 6 distinct protein folds that display a diverse mix of secondary structure content in the solution proteins UBA domain/Ubiquitin and Barnase/Barstar, and almost completely beta structure in the OMPLA homodimer (transmembrane beta-barrel) or only helical structure in the ClC homodimer (transmembrane helix bundle). The rationale behind choosing protein complexes as test systems is that it allows us to obtain more reference sampling with two proteins in each system while also assessing their interface contact stability and thus whether there is a need for stabilizing quaternary structure further in the coarse-grained simulations. The protein interfaces show various interactions that range from being predominantly hydrophobic (UBA domain/Ubiquitin, SI Figure S1 & ClC homodimer, SI Figure S4) and a system containing polar and hydrogen bonding residues (OMPLA homodimer, SI Figure S3) and also a system with a large fraction of charged residues (Barnase/Barstar, SI Figure S2).

Protein Preparation and Atomistic Reference Simulations

All protein structures were downloaded from the RCSB Protein Data Bank.[36] Some of the structures required further processing. In PDB: 2OOB, Chain B (Ubiquitin) we built 4 missing residues in the C-terminal (LRGG) using the CHARMM-GUI webserver,[37,38] as

these residues are close to the binding interface and could be important. In PDB: 1BRS, Chain A (Barnase) we built two missing N-terminal residues (AQ) using CHARMM-GUI, and mutated A40C and A82C, restoring two cysteines from the wild-type sequence. A small loop in PDB: 1BRS, Chain B, residue 64-65 (EN) was built using the Maestro Schrodinger v2021.4 Protein Crosslink Tool. In PDB: 1QD6 a loop of residue 26-29 (HDNP) was missing and built using the Maestro Schrodinger v2021.4 Protein Crosslink Tool. In PDB: 1QD6 S114 was covalently linked to a 1-Hexadecanosulfonic Acid which was removed to restore the serine. Crystal water and ions were removed from all structures except PDB: 1OTS where four structural chloride ions were kept (two in each protein monomer at site S_{cen} and S_{int}[35]). Histidine tautomers were checked for all the structures using the H++ webserver[39] at the pH of the respective crystallization experiments. In PDB: 1OTS, side chains E113 (Chains A and B) and D417 (Chain A) were manually protonated and neutrally charged which is supported by experiment.[40,41]

The UBA domain/Ubiquitin and Barnase/Barstar systems were simulated using the DES-Amber-SF0.9 force field[42] which is optimized for soluble protein-protein complexes. Following protein preparation, the UBA domain/Ubiquitin and Barnase/Barstar systems were solvated with TIP4P-D[43] water in a dodecahedron periodic box with a distance from the protein to the box edge of 2 nm with a salt concentration of 0.15 M NaCl using GROMACS v2021.4.[30,31] A straight non-bonded cutoff of 1 nm was used for both Van der Waals and electrostatic interactions and Particle Mesh Ewald[44] summation was used for electrostatics. The systems were first minimized and equilibrated for 10 ns with restraints on the protein backbone heavy atoms. Production runs were 1 μs with 3 repeats for each system. Frames were saved every 100 ps.

The OMPLA and ClC homodimer systems were simulated using the CHARMM36m force field,[45] a well-tested force field for phospholipids and transmembrane proteins. The OMPLA and ClC homodimer systems were set up in a cubic periodic box using CHARMM-GUI. The membrane of the OMPLA homodimer system was created as a symmetric 1:1

POPC:POPE bilayer, for comparison to the work of Piller et al.,[46] with 191 POPC and 191 POPE phospholipids. The membrane in the ClC homodimer was created as a symmetric 2:1 POPE:POPG bilayer, for comparison to the work of Chadda et al.,[41] with 246 POPE and 123 POPG phospholipids. The OMPLA and ClC homodimer systems were solvated with TIP3P water resulting in a water layer of thickness ≈ 4.5 nm in the z-direction with a salt concentration of 0.15 M NaCl. Van der Waals non-bonded interactions were switched from 1.0 nm to a cutoff of 1.2 nm. Particle Mesh Ewald[44] summation was used for electrostatics with a real-space cutoff of 1.2 nm. The systems were first minimized and equilibrated in a series of small equilibrations, with gradually weaker restraints on protein and lipids (generic CHARMM-GUI settings). The system was then equilibrated for 10 ns with only restraints on the protein backbone heavy atoms. Production runs were 1 μs with 3 repeats for each system. Frames were saved every 100 ps.

Hydrogen bonds were restrained for all systems using LINCS[47] allowing integration steps of 2 fs. Temperature and pressure were kept constant at 310 kelvin and 1 bar using the v-rescaling thermostat[48] ($\tau_t = 1$ ps) and the Parrinello-Rahman barostat[49,50] ($\tau_p = 12$ ps). The c-rescale barostat[51] ($\tau_p = 4$ ps) was used for equilibration. Semi-isotropic pressure coupling was used for the membrane systems. Compressability was set to 4.5×10^{-5}.

Simulations were carried out using GROMACS[30,31] v2021.4 on Nvidia V100 GPUs at the Centre for Scientific Computing Aarhus (CSCAA).[52]

MARTINI 3 Simulations Using OLIVES and GoMARTINI Gō-like Models

To validate the performance of OLIVES we prepared a series of coarse-grained simulations for comparison to the atomistic simulations. For each of the 4 test systems, we created 8 coarse-grained systems: OLIVES (with and without quaternary contacts in the protein complex), OLIVES + DSSP (with secondary structure restraints and with and without quaternary contacts in the protein complex), GoMARTINI (with and without quaternary contacts in

the protein complex), GoMARTINI + DSSP (with secondary structure restraints and with and without quaternary contacts in the protein complex).

We used the Martinize2 tool in the vermouth v0.9.1[4] Python 3 package for coarse-graining of the atomistic protein structures (for the ClC homodimer we used vermouth v0.9.3[4]) and controled the DSSP restraints with the *-ss* flag. The DSSP sequence was found using the DSSP v2.2.0[7] implementation in MDtraj v1.9.7.[53] Additionally, we used the *-scfix* flag to restrain the side chain dihedrals.[54]

OLIVES systems were set up based on the coarse-grained protein output of Martinize2 (see the GitHub link to the OLIVES script below) using a cutoff of 0.55 nm for generating the secondary and tertiary networks. The LJ bond energies were unmodified values from Table 1 as described in Figure 2. As an illustration of how to apply OLIVES in practice, the following command line prompt will convert the (prepared) atomistic protein complex of Ubiquitin/UBA domain (PDB: 2OOB) to a coarse-grained representation using Martinize2, generate a topology (with default name molecule_0.itp), and then apply the OLIVES model:

```
$ martinize2 -f 2OOB.pdb -x 2OOB_CG.pdb -o 2OOB_CG.top
-scfix -cys auto -merge A,B
$ python3 OLIVES_v1.0_M3.0.0.py -c 2OOB_CG.pdb -i molecule_0.itp
```

Note that in this example we merge the topology of chain A and B because we want OLIVES to generate a quaternary network. We leave out the Martinize2 -dssp/-ss flags to avoid generating secondary structure restraints. OLIVES was tested using the Martinize2 -scfix flag, although the side chain conformations are also modulated by the OLIVES LJ potentials. The OLIVES model is applied by calling the OLIVES script after Martinize2, which will automatically insert the model into the provided topology. There is also an option to write out the OLIVES model in a separate topology file and for writing additional information files about the network pairs (which can be used to drive biased simulations due to the close similarity to native contacts). Additional examples together with the OLIVES script can be found in our GitHub repository (`https://github.com/Martini-Force-Field-Initiative/OLIVES`).

GoMARTINI[6] systems were set up by first uploading the atomistic protein starting structures to the rCSU webserver[55] to obtain the OV+rCSU contact map. For systems without a quaternary structure network, protein chains A and B were uploaded individually. To obtain a quaternary protein network we force the webserver to regard the whole complex as one protein by removing the chain A TER record (PDB file format), renaming chain B to A, and renumbering atoms and residues. Then the atomistic protein structures were coarse-grained by Martinize2 using the *–govs-includes* flag, and the GoMARTINI model was then applied using the *create_goVirt.py* script. GoMARTINI settings were left at default values with LJ bond energy of $\epsilon_{go} = 2.25$ kcal/mol (9.414 kJ/mol) and the short and long cutoffs at 0.3 nm and 1.1 nm, respectively.

Following protein coarse-graining, the UBA domain/Ubiquitin and Barnase/Barstar systems were solvated with regular-sized water beads in a dodecahedron periodic box with a distance from the protein to the box edge of 2.5 nm with a salt concentration of 0.15 M NaCl using GROMACS v2021.4. The OMPLA and ClC homodimer systems were set up in a cubic periodic box the *insane.py* script.[56] The membrane of the OMPLA and ClC homodimer systems are identical in composition to the atomistic systems: a symmetric 1:1 POPC:POPE bilayer and a symmetric 2:1 POPE:POPG bilayer, respectively, however, we had to use a larger membrane patch to avoid visible membrane distortions resulting from the inserted protein dimer, which are not observed in our atomistic simulations, and disappear in the large membrane (note that we are using the updated neighbor list settings recommended by Kim et al.,[57] see below). Consequently, the coarse-grained membrane of the OMPLA system consists of 321 POPC and 321 POPE phospholipids, and the membrane of the ClC homodimer systems consists of 401 POPE and 200 POPG phospholipids. The OMPLA and ClC homodimer systems were solvated with regular-sized water beads resulting in a layer of thickness ≈ 5 nm in the z-direction with a salt concentration of 0.15 M NaCl.

The Martini 3 simulations were carried out using GROMACS[30,31] v2021.4 on Nvidia V100 GPUs at CSCAA.[52] We used standard Martini 3 settings[1] with reaction field electrostatics

(cutoff 1.1 nm, $\epsilon_r = 15$, $\epsilon_{rf} = 0$) and for Van der Waals interactions we used potential-shift-verlet with cutoff 1.1 nm. We used some important neighbor-list changes recently recommended by Kim et al.[57] Specifically, we turned off the dual pair list by setting verlet-buffer-tolerance = -1, set rlist = 1.35, and set nstlist = nsttcouple = nstpcouple = 20. Temperature and pressure were kept constant at 310 kelvin and 1 bar using the v-rescaling thermostat[48] ($\tau_t = 1$ ps) and the Parrinello-Rahman barostat[49,50] ($\tau_p = 12$ ps). The c-rescale barostat[51] ($\tau_p = 4$ ps) was used for equilibration. Semi-isotropic pressure coupling was used for the membrane systems. Compressability was set to 3×10^{-4}. All systems were first minimized and equilibrated for 10 ns with restraints on the protein backbone beads. Production runs were 250 ns with 3 repeats for each system. Frames were saved every 100 ps.

Effective Time

In the following analysis, we will use the notion of "effective time" in coarse-grained simulations. Because coarse-grained force fields represent a significantly smoothened energy landscape (i.e. fewer degrees of freedom),[58] kinetics are considerably faster relative to atomistic simulations. A common rule of thumb is to use a speedup factor of 4 when using Martini.[59] Note that this factor is highly system-dependent and even varies at the molecular level. However, we here use a conversion factor of 4 in order to compare across systems. Consequently, we compare 1 μs of atomistic simulation to 250 ns of coarse-grained simulation. To obtain the same number of samples for analysis from the atomistic and coarse-grained simulations, we downsample to 4 ns per frame (250 frames over 1 μs) for the atomistic simulations and 1 ns per frame (250 frames over 250 ns) for the coarse-grained simulations.

Analysis Protocol

We compared the atomistic reference simulations to the coarse-grain simulations using the OLIVES or GoMARTINI models in the following way:

Protein fluctuations were quantified using the residue root mean squared fluctuation around the mean coordinate vector during the second half of the simulations:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - \bar{r})^2} \tag{1}$$

where the sum runs over T=375 trajectory frames (we concatenate the last 125 frames out of 250 frames from each of the three repeats) with residue coordinate vector r_t being either the Cα positions or the backbone bead positions and $\bar{r}$ is the mean coordinate vector over the concatenated 375 frames. To compare the coarse-grained fluctuations to the atomistic counterpart, we report the absolute error to the atomistic RMSF per residue:

$$|\Delta RMSF| = |RMSF_{CG} - RMSF_{AA}| \tag{2}$$

In order to compare the deviation from the starting crystal structure for each monomer, we calculated the time-averaged root mean squared deviation per residue:

$$RMSD = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(r_t - r_0)^2} \tag{3}$$

where the sum runs over T=250 trajectory frames for each repeat separately with coordinate vector r_t being either the Cα positions or the backbone bead positions and r_0 is the coordinate vector of the starting position. For both the RMSF and RMSD calculations, each monomer frame was aligned to the starting crystal structure of the monomer (this is required to determine the mean coordinate vector for the RMSF calculation, so we chose the starting structure for consistency across all test systems).

We quantify the protein fold integrity using native contacts.[15,23,60] The atomistic starting structure is used to define a set of native contact residue pairs. We defined a residue pair as a native contact if any of their closest heavy atoms are within 0.45 nm of each other[23] (note that

the native contact network is not identical to the OLIVES network). We used the atomistic native contacts as a reference for both the atomistic and coarse-grained simulations to make the comparison consistent. To quantify how native contacts change over the simulation trajectories, we defined the smooth function Q commonly used to track protein folding and conformational changes:[23]

$$Q = \frac{1}{N}\sum_{p=1}^{N}\frac{1}{1+exp[\beta(r_p-\lambda r_{p,0})]} \tag{4}$$

where the sum runs over the N initial native contact residue pairs, r_p is the distance between the closest heavy atoms between residues in a pair or the distance between any beads in the case of coarse-grained structures, $r_{p,0}$ is r_p for the starting structure, β is a smoothening factor set to 50 nm^{-1},[23] and λ is a calibration factor set to 1.8 for the atomistic simulations[23] and 1.5 for the coarse-grained simulations.[61] In order to quantify the stability of the protein complexes, we also defined a set of interface native contacts which is any residue in the first protein with a native contact to a residue in the second protein, calculated from the initial protein complex structure using a cutoff of 0.45 nm. We quantified the interface native contacts according to equation 4.

Results and Discussion

Validating OLIVES for Solution Proteins

UBA domain/Ubiquitin Complex

Our first test system is the Ubiquitin-associated domain (UBA) from Cbl-b Ubiquitin ligase in complex with Ubiquitin[32] (Figure 3,A). The UBA domain consists of three short helices packed around a hydrophobic core. The UBA domain has 44 residues in native contacts out of which 21 are hydrogen bonding (SI Table S1, SI Figure S1,A). Ubiquitin, widespread in the cell and a canonical test system in biochemistry, consists of two helical segments and an

antiparallel β-sheet with 5 strands which together form a barrel-shaped protein fold. The sheet and the longer helix have hydrophobic residues lining the inside of the barrel. The solvent-exposed side consists of polar residues apart from a hydrophobic site on the β-sheet which can bind the UBA domain.[32] Ubiquitin has 76 residues in native contacts out of which 43 are potentially hydrogen bonding (SI Figure S1,B). The interface between Ubiquitin and the UBA domain is mainly hydrophobic where the first helix of the UBA domain has a small hydrophobic patch complementary to Ubiquitin. There are 21 residues in the interfacial native contacts out of which 13 are hydrophobic (SI Figure S1,C). However, the interface is also stabilized by a few hydrogen bonds. Inspecting the atomistic structure complex ASP933(5) from the UBA domain (the index in parenthesis denotes the corresponding residue index in Figure 3,B where the indexing is shifted to start from 1) binds to the backbone amides of ALA46 and GLY47 on Ubiquitin, and the amide carbonyl on GLY47 interacts with LYS950(22) on the UBA domain. OLIVES correctly identifies the hydrogen bond between GLY47 and LYS950(22), but it pairs ASP933(5) with the backbone of PHE45 instead of ALA46. The OLIVES algorithm also forms a few additional bonds not observed in the atomistic structure due to close proximity between hydrogen bond donors and acceptors, that do not have favorable hydrogen bond directionality in the atomistic structure. Specifically, OLIVES identifies a hydrogen bond between HIS68 and the backbone of ASP933(5), and a π-hydrogen bond between PHE18 and the backbone of GLN49. We find this behavior acceptable since the goal of OLIVES is to stabilize coarse-grained structures via native contacts, however, one should have this in mind if inspecting the OLIVES networks closely, and as such, the OLIVES network should be regarded as a coarse-grained representation of a more detailed atomistic hydrogen bond network. The OLIVES network for the UBA domain/Ubiquitin complex is illustrated in Figure 3,A.

Comparing the simulation of the UBA domain/Ubiquitin complex using the OLIVES model to our atomistic reference simulations, we observe very little difference in the fluctuations between using OLIVES with and without the DSSP secondary restrains, showing

that OLIVES effectively captures the secondary structure without the need for additional restraints. We also compare to results using the GoMARTINI model as an established method for stabilization of Martini 3 proteins by a Gō-like model. Inspecting the absolute error between the all-atom and coarse-grained root-mean-square-fluctuations ($|\Delta$RMSF$|$), Figure 3,B, we observe that OLIVES is able to stabilize both proteins on par with the GoMARTINI model. OLIVES has a tendency to deviate slightly more from the atomistic reference compared to the GoMARTINI+DSSP results (SI Figure S5,B). When simulating the complex with GoMARTINI without applying the DSSP secondary structure restraints, however, we observe large fluctuations in the third α-helix of the UBA domain (residues 32-42, Figure 3,B) and a partly unfolding of the Ubiquitin α-helix (residues 24-34, Figure 3,B) which are not compatible with the atomistic reference.

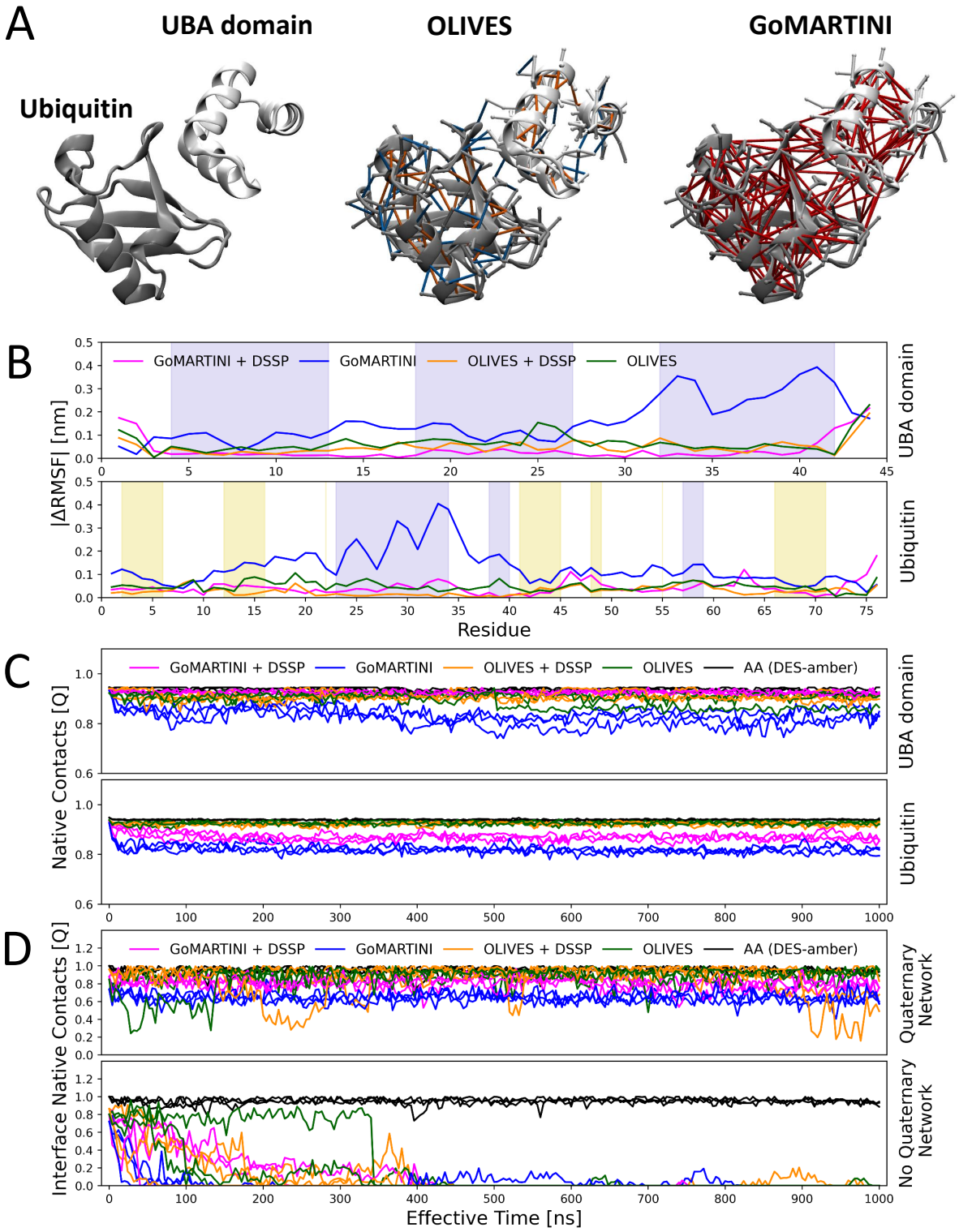


Figure 3: (A) UBA domain (white) in complex with Ubiquitin (grey). The bond networks created by OLIVES (orange/blue) and GoMARTINI (red) are shown. (B) RMSF absolute error to the AA reference for each protein, using a quaternary network. Data without a quaternary network is shown in SI Figure S5. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (C) Intramolecular native contacts for each protein, using a quaternary network. (D) Interface native contacts with and without using a quaternary network.

The intramolecular native contacts within both the UBA domain and Ubiquitin are excellently preserved when using OLIVES compared to the atomistic reference (Figure 3,C). When simulating using GoMARTINI without DSSP restraints, the intramolecular native contacts drop by ≈ 0.1, compared to the atomistic reference.

Inspecting the interface native contacts without the use of a quaternary network (Figure 3,D,bottom panel), it is evident that the UBA domain/Ubiquitin complex interface native contacts break which also can be observed visually in the simulations. This is not compatible with our atomistic simulations where the complex is stable over the entire 1 μs trajectory (Piana et al. have shown in longer atomistic simulations that the complex remains stable for at least 10 μs[42]), and indicates an underestimation of the UBA domain/Ubiquitin complex stability when simulating it without applying a quaternary network. When introducing a quaternary network between the proteins using OLIVES, the interfacial native contacts are conserved, which is interesting considering that the interface is mainly of hydrophobic character with only a few hydrogen bond partners. The GoMARTINI quaternary network also reproduces the interface native contacts, albeit using a more extensive bond network than OLIVES, as seen in Figure 3,A. The interface bias energy (total bias of the Gō-like model bonds using a quaternary network minus the individual protein Gō-like models without a quaternary network), amounts to -16 kcal/mol for the OLIVES model and -54 kcal/mol for the GoMARTINI model, showing that OLIVES stabilizes the complex using three times less bias energy. The experimental dissociation constant of the UBA domain/Ubiquitin complex is ≈ 57 μM,[32] corresponding to dissociation free energy of $\approx$ -5.7 kcal/mol at 293 K, which is difficult to reconcile with bias energies of both -16 kcal/mol for OLIVES or -54 kcal/mol for GoMARTINI. Quantitatively, it is not clear which amount of bias energy would be appropriate to obtain an accurate binding free energy. This could be investigated by calculating the complex dissociation free energy using enhanced sampling methods and comparing it to an atomistic reference, however, the correlation between experimental and simulated dissociation free energies of protein complexes is weak, even in atomistic simulations,[42] and such

a comparison is out of the scope of this introduction to the OLIVES model.

Finally, we have also investigated alternative settings of both the GoMARTINI and OLIVES models for the UBA domain/Ubiquitin system. In one test we changed the long cutoff of GoMARTINI to 0.55 nm and the uniform energy to 5.44 kcal/mol (22.76 kJ/mol), to mimic the OLIVES settings for GoMARTINI. We find that this results in much worse $|\Delta$RMSF$|$ values compared to default GoMARTINI settings (SI Figure S9,A). This can be attributed to the fact that GoMARTINI only forms bonds between backbone beads, and the reduced long cutoff therefore breaks important long-range contacts, which are conserved in the OLIVES model through the side chain interactions in the tertiary network. In another test, we investigated whether the non-uniform energy matrix of OLIVES could be replaced by a uniform bond energy of 5.44 kcal/mol (22.76 kJ/mol). We find that a uniform energy matrix leads to slightly worse $|\Delta$RMSF$|$ results (SI Figure S9,B), which merit the chemical specificity of the energies reported in Table 1.

Barnase/Barstar Complex

The second test system is the Barnase/Barstar complex produced by *B. amyloliquefaciens*, a canonical test system for protein folding and protein-protein recognition[33,62] (Figure 4,A). The Barnase fold has a central anti-parallel β-sheet with 5 strands and several smaller helical segments.[62] Barnase has 110 residues in native contacts out of which 66 are potential hydrogen bonding residues (SI Figure S2,A). The Barstar fold consists of a single parallel β-sheet with 3 strands and 4 helices where helix 2 forms the main interface to Barnase (Figure 4,A). Barstar has 89 residues in native contacts out of which 51 are potentially hydrogen bonding (SI Figure S2,B). The interface between Barnase and Barstar is dominated by hydrogen bond interactions and out of the 35 residues in interfacial native contacts, 26 are potential hydrogen bonding residues (SI Figure S2,C). Buckle et al. have identified 10 essential interface hydrogen bonds between Barnase and Barstar[33] out of which 7 are correctly identified using the OLIVES algorithm. Furthermore, the interface of Barnase/Barstar is stabilized by an

extensive water-mediated hydrogen bond network[33] that currently cannot be represented in Martini 3. Moreover, the regularly sized water beads in Martini 3 are too large to fit in the interface. Some of the water-mediated hydrogen bond partners are identified by OLIVES anyway due to the search cutoff of 0.55 nm, effectively bonding the partners without having the linking water molecule present. Due to the missing water-mediated interactions, the Barnase/Barstar complex represents a challenging test case and the interface likely cannot be stabilized in Martini 3 (v3.0.0) without the introduction of a quaternary network between Barnase and Barstar.

Our simulation results of the Barnase/Barstar complex reveal that RMSF fluctuations between OLIVES and GoMARTINI are very similar and in good agreement with atomistic results (Figure 4,B). In Barnase, OLIVES outperforms GoMARTINI in the loop consisting of residues 60-70 which is too flexible in GoMARTINI. In the first half of Barstar, residues 20-25 and 40-55, OLIVES is slightly more flexible compared to the GoMARTINI+DSSP model where GoMARTINI+DSSP agrees slightly better with the atomistic reference. Again, we observe no difference between having DSSP restraints or not when using OLIVES. The intramolecular native contacts are well preserved using OLIVES, especially in Barnase where the native contacts stay constant whereas native contacts in GoMARTINI drop slightly by 0.05 (Figure 4,C). Observing the interface native contacts (Figure 4,D), it is clear that a quaternary network is required to stabilize the Barnase/Barstar interface. This is in line with the recent work of Lamprakis et al. finding that the dissociation free energy of the Barnase/Barstar complex is underestimated in Martini 3 by as much as 15 kcal/mol compared to experimental reference data.[16,63] The interface native contacts in our simulations are slightly better conserved in OLIVES compared to GoMARTINI despite introducing less interface bias energy of -52 kcal/mol versus -86 kcal/mol when using GoMARTINI. Taken together, this adds evidence to the trend that solution protein complexes are not stable in the current Martini 3 version (v3.0.0).

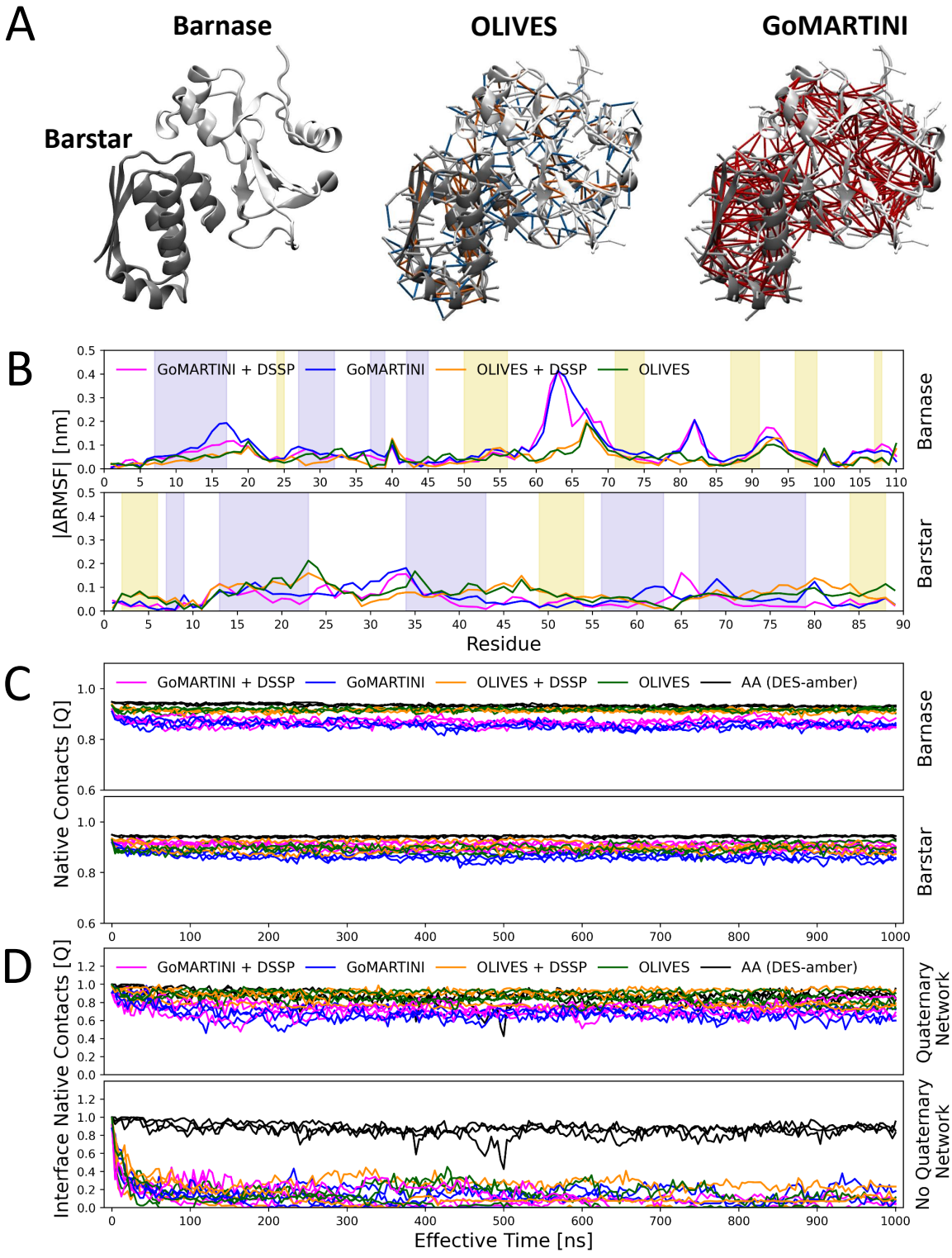


Figure 4: (A) Barnase (white) in complex with Barstar (grey). The bond networks created by OLIVES (orange/blue) and GoMARTINI (red) are shown. (B) RMSF absolute error to the AA reference for each protein, using a quaternary network. Data without a quaternary network is shown in SI Figure S6. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (C) Intramolecular native contacts for each protein, using a quaternary network. (D) Interface native contacts with and without using a quaternary network.

Validating OLIVES for Transmembrane Proteins

Outer Membrane Phospholipase A Homodimer

Having validated OLIVES against solution protein complexes, we now turn to transmembrane proteins, which are commonly simulated using Martini.[18–20] Our first test case is the *E. coli* outer membrane phospholipase A (OMPLA) homodimer (Figure 5,A). OMPLA is a transmembrane β-barrel enzyme that is regulated by reversible dimerization.[34] The interior of the OMPLA barrel is largely polar and out of 257 residues in native contacts in the monomer, 154 are potential hydrogen bonders (SI Figure S3,A). The dimerization of OMPLA activates the enzyme and it has recently been shown by Piller et al. that the dimerization is modulated by membrane asymmetry.[46] Piller et al. show that OMPLA has high activity in a symmetric POPC:POPE bilayer indicating stabilization of the dimer,[46] which makes it an ideal case to test whether the homodimer can be stabilized in Martini 3 simulations without the need for a quaternary network. Furthermore, there have previously been reported problems simulating β-barrel proteins in Martini 2 using elastic networks,[64] and we want to ensure that OLIVES correctly models this protein fold.

The interface between OMPLA has the largest surface area in the transmembrane part,[34] however, there are numerous contacts of polar character both on the extracellular and intracellular parts of the interface. In fact, out of the 83 residues in interfacial native contacts, 45 are polar (SI Figure S3,C) and are mainly found at the bilayer-water boundary. Snijder et al. report three main-chain hydrogen bonds between PHE109(A)–GLY146(B), GLY146(A)–PHE109(B) and LEU32(A)–LEU32(B).[34] OLIVES correctly identifies these backbone-backbone hydrogen bond partners and places them in the quaternary network. Additionally, an essential hydrogen bond is located in the middle of the transmembrane region between GLN94(A)-GLN94(B), which is also correctly identified by OLIVES.

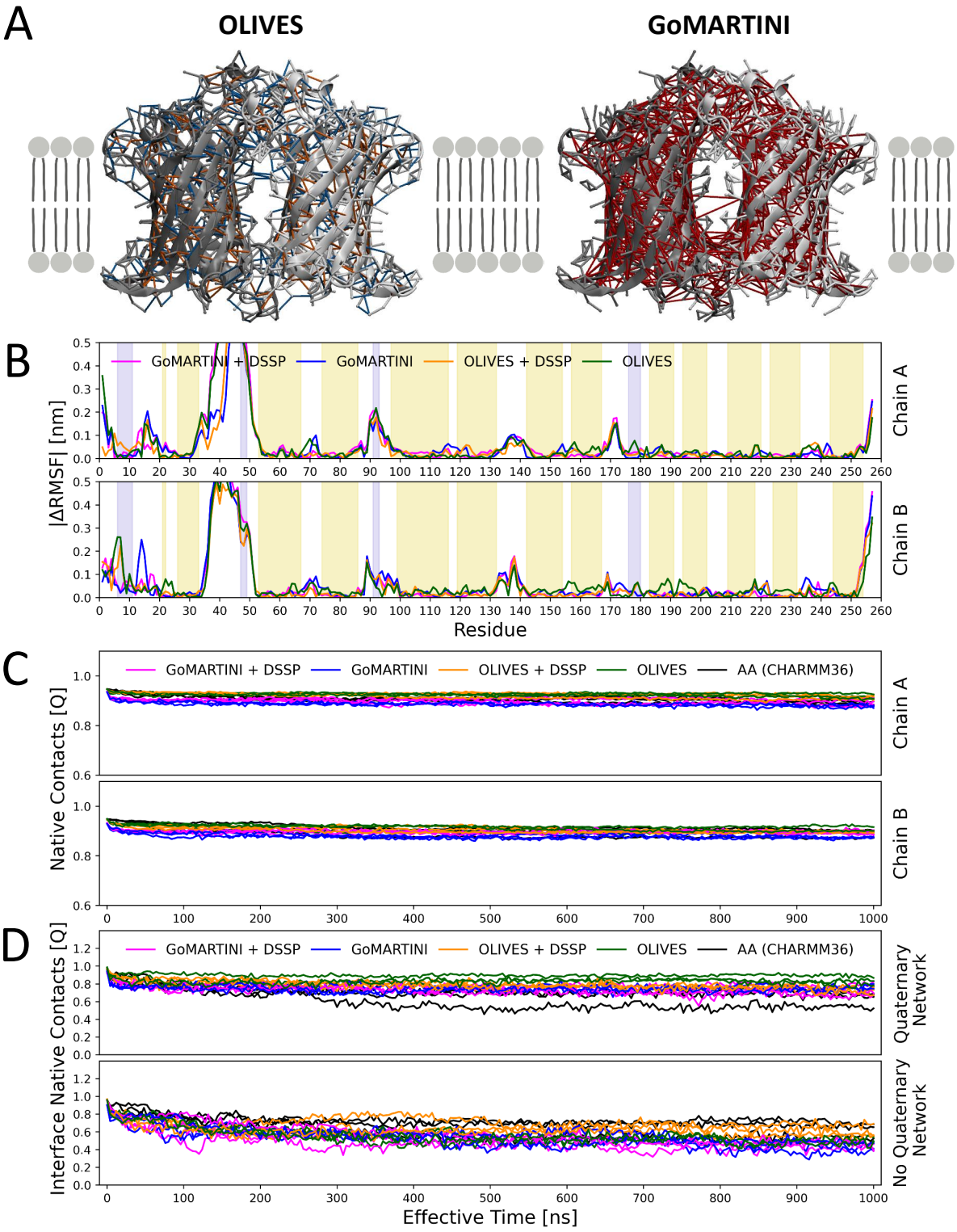


Figure 5: (A) OMPLA homodimer with monomers in white and grey. The bond networks created by OLIVES (orange/blue) and GoMARTINI (red) are shown. (B) RMSF absolute error to the AA reference for each protein, using a quaternary network. Data without a quaternary network is shown in SI Figure S7. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (C) Intramolecular native contacts for each protein, using a quaternary network. (D) Interface native contacts with and without using a quaternary network.

Inspecting the RMSF fluctuations (Figure 5,B), OLIVES and GoMARTINI results are very similar and match the atomistic simulations (i.e. low absolute error), especially in the regions containing β-sheet secondary structure. In two of the atomistic trajectories, an extracellular loop (residues $\approx$ 35-45) unbinds from the bulk protein, and thus becomes flexible, which is not seen in simulations using OLIVES or GoMARTINI, resulting in high $|\Delta\text{RMSF}|$ values.

The native contacts of the OMPLA monomers are reproduced using both OLIVES and GoMARTINI (Figure 5,C) and we do not observe any distortion of the β-barrel, however, a more in-depth study of β-barrels of various sizes is needed to rule out the distortion effects reported by Desikan et al.[64] Interestingly, the interface native contacts (Figure 5,D) are reproduced very well without a quaternary network using either OLIVES and GoMARTINI, showing that the OMPLA dimer is stable and that oligomerization between transmembrane proteins can be captured in Martini 3.

ClC Chloride Channel Homodimer

To further investigate the stability of transmembrane dimers in Martini 3 and to test OLIVES on proteins with diverse secondary structure content, we chose an *E. coli* homolog of the ClC chloride channel family[35] as our final test system. The fold of ClC chloride channel consists of an exclusively α-helical bundle with two structurally similar transmembrane domains exhibiting pseudo two-fold symmetry.[35] The monomer fold is predominantly stabilized by hydrophobic contacts and out of the 441 residues in native contacts, 318 are hydrophobic (SI Figure S4). Compared to the OMPLA protein with β-sheet secondary structure with a polar interior stabilized by hydrogen bond networks, the ClC channel represents a distinct end of the spectrum with a complete α-helix secondary structure content stabilized by mainly hydrophobic interactions. The ClC chloride channel forms a functional homodimer (Figure 6,A) which has been extensively studied both experimentally and computationally by Chadda et al.[41,65] The dimer interface consists of 8 short helices (4 on each monomer) which contribute

to a significant hydrophobic mismatch to the bilayer in the monomer state, thought to drive dimerization.[41] The interface native contacts are also of hydrophobic character with 77 out of 137 residues in contacts being hydrophobic (SI Figure S4).

The simulation results reveal that, in general, the OLIVES and GoMARTINI RMSF are compatible with the atomistic reference (Figure 6,B). In this test case, we can identify a slight difference between running OLIVES with and without DSSP restraints when inspecting residue RMSD from the initial structure (SI Figure S8). In the first few helices, residues 20-80, both OLIVES and GoMARTINI have higher RMSD compared to the simulations with DSSP restraints (SI Figure S8). However, since the native contacts are completely preserved for OLIVES without the DSSP restraints and the absolute errors in the RMSF profile are generally below 0.1 nm (Figure 6,B), this should not be a major concern. However, one could consider applying the DSSP restraints if the study of interest does not require the ability to unfold or change the secondary structure. Inspecting the interface native contacts, we observe that a quaternary network is not required to stabilize the ClC dimer in line with the results for the OMPLA dimer. This is reassuring since Martini is extensively used for oligomerization studies of transmembrane proteins.[16,18–20]

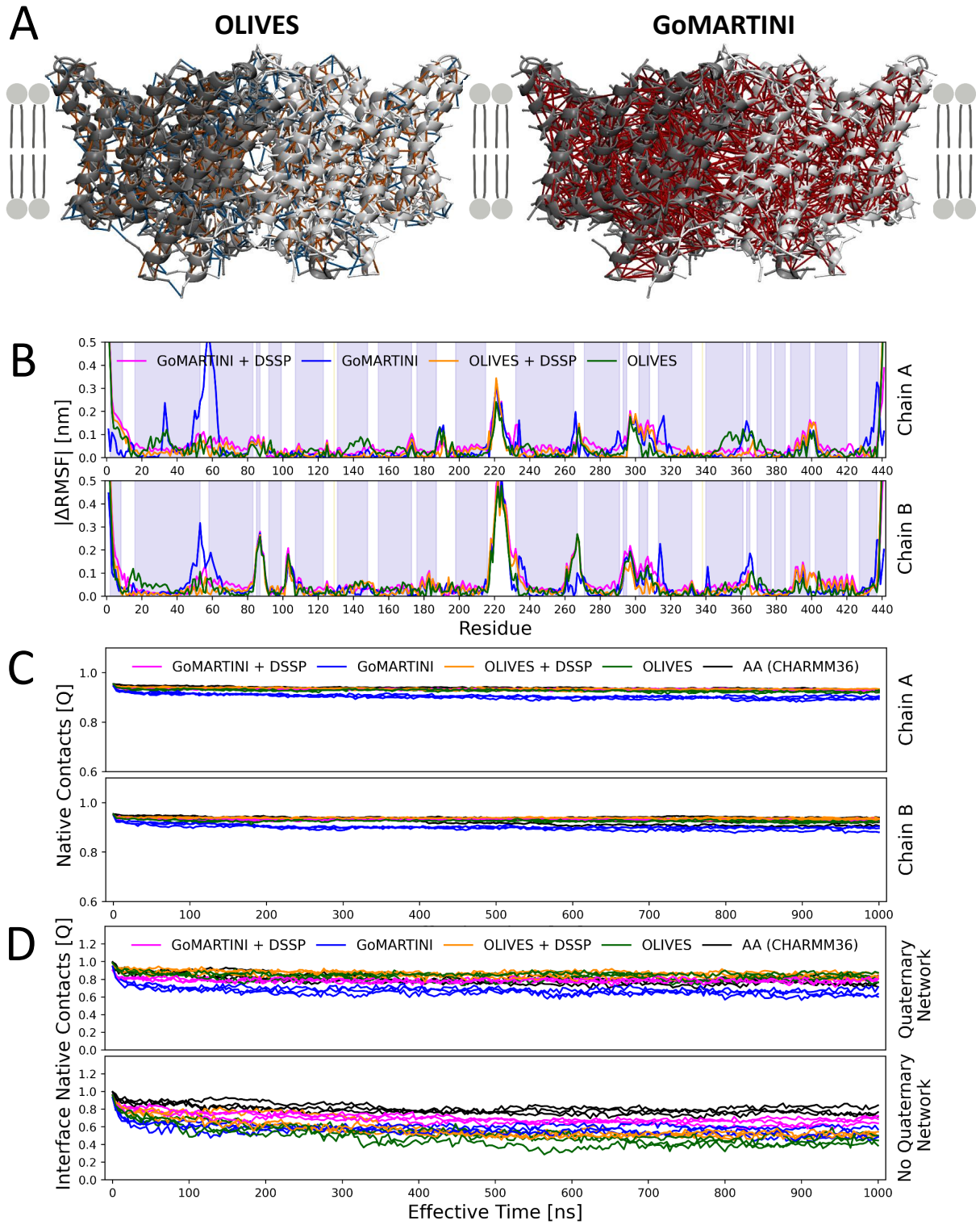


Figure 6: (A) ClC homodimer with monomers in white and grey. The bond networks created by OLIVES (orange/blue) and GoMARTINI (red) are shown. (B) RMSF absolute error to the AA reference for each protein, using a quaternary network. Data without a quaternary network is shown in SI Figure S8. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (C) Intramolecular native contacts for each protein, using a quaternary network. (D) Interface native contacts with and without using a quaternary network.

OLIVES Improves Simulation Speed in Relation to the GoMARTINI Model

The OLIVES model creates fewer bonds in the Gō-like model compared to the GoMARTINI model, which reduces the computational cost of each integration time step. Furthermore, OLIVES is developed to stabilize secondary structure without the need for DSSP-derived restraints, and an additional speed up can be gained if these potentials are also removed. Here, we benchmark the performance of OLIVES on the large 28-mer complex of the yeast 20S proteasome in water (PDB: 1RYP).[66] The 20S proteasome was chosen because of its size to highlight the speed-up advantage of Martini 3 for a system that would have been prohibitively slow to simulate using all-atom simulations. In our simulations, the complex occupies a large fraction of the system volume (14730 protein beads to 45132 water beads), which represents the protein concentration in the cytoplasm.[2,3] Note that the speed-up gain of this system will be different compared to a single transmembrane protein in a bilayer, which is another common Martini use case where the fraction of protein beads usually is lower. The 20S proteasome complex was simulated for 1 μs with 3 repeats using a quaternary network and did not dissociate during the simulation. We compare the models of GoMARTINI+DSSP, OLIVES+DSSP, and OLIVES without DSSP restraints, to illustrate the potential performance gain of OLIVES. Table 2 lists the results for three different parallelization schemes.

Table 2: Performance benchmark of simulating the yeast 20S proteasome in water (PDB: 1RYP).[66] Results are mean and standard deviation over 3 repeats reported in ns/day. * 2 Intel Xeon Platinum 8358 CPU, 32 cores each. ** 1 Intel Xeon Gold 6240 CPU, 18 cores. ***1 Nvidia V100-16GB GPU.

Performance Benchmark	No Domain Decomposition	Domain Decomposition	No Domain Decomposition
Processes	1 MPI 64 OpenMP	8 MPI 8 OpenMP	1 MPI 18 OpenMP
Hardware	64 CPU*	64 CPU*	18 CPU** 1 GPU***
	[ns/day]	[ns/day]	[ns/day]
GoMARTINI + DSSP	1619 ± 18	2138 ± 104	2117 ± 15
OLIVES + DSSP	1784 ± 17	2338 ± 8	2410 ± 14
OLIVES	1961 ± 33	2484 ± 8	2869 ± 67

There are multiple ways of parallelizing molecular dynamics simulations in the GROMACS (v2021.4) suite,[67] e.g. the calculation can be distributed across N OpenMP processes with 1 MPI process, where N is the number of CPU cores available and, complementary, one can use domain decomposition[31] with NxM processes where M is the number of MPI processes. The current implementation of OLIVES uses 1-4 LJ interactions (pairs, bond type 1) to implement the Gō-like bonds. However, 1-4 LJ bonds are internally regarded as bonded interactions in GROMACS, which leads to errors if the bond length becomes more than half the length of a domain when using domain decomposition. This would happen if a protein complex dissociates while having a quaternary network defined, however, if OLIVES is used for a monomeric protein or very stable complexes like the 20S proteasome, there should be no issues using domain decomposition. One way to resolve the domain decomposition issue when using quaternary networks is to implement the Gō-like bonds as non-bonded LJ interactions at the force field level between virtual particles co-localized on top of the protein beads, as done in the current GoMARTINI implementation. Though not yet implemented, the virtual site approach is applicable also to OLIVES and could be developed if there is a demand. Another way is to turn domain decomposition off, likely with a computational cost depending on hardware architecture. At our local high-performance computing (HPC) facility at CSCAA,[52] we often do not use domain decomposition given our specific hardware architecture of 1 GPU and 18 CPUs, however, this will be different on other HPC clusters where running domain decomposition could be more efficient. Ideally, GROMACS would implement an LJ bonded type that could be defined between particle pairs but included in the non-bonded list upon topology generation.

A Need for Quaternary Bias Networks in Martini 3 Solution Protein Complexes and Future Directions for Improving Protein-Protein Interactions in Martini 3.

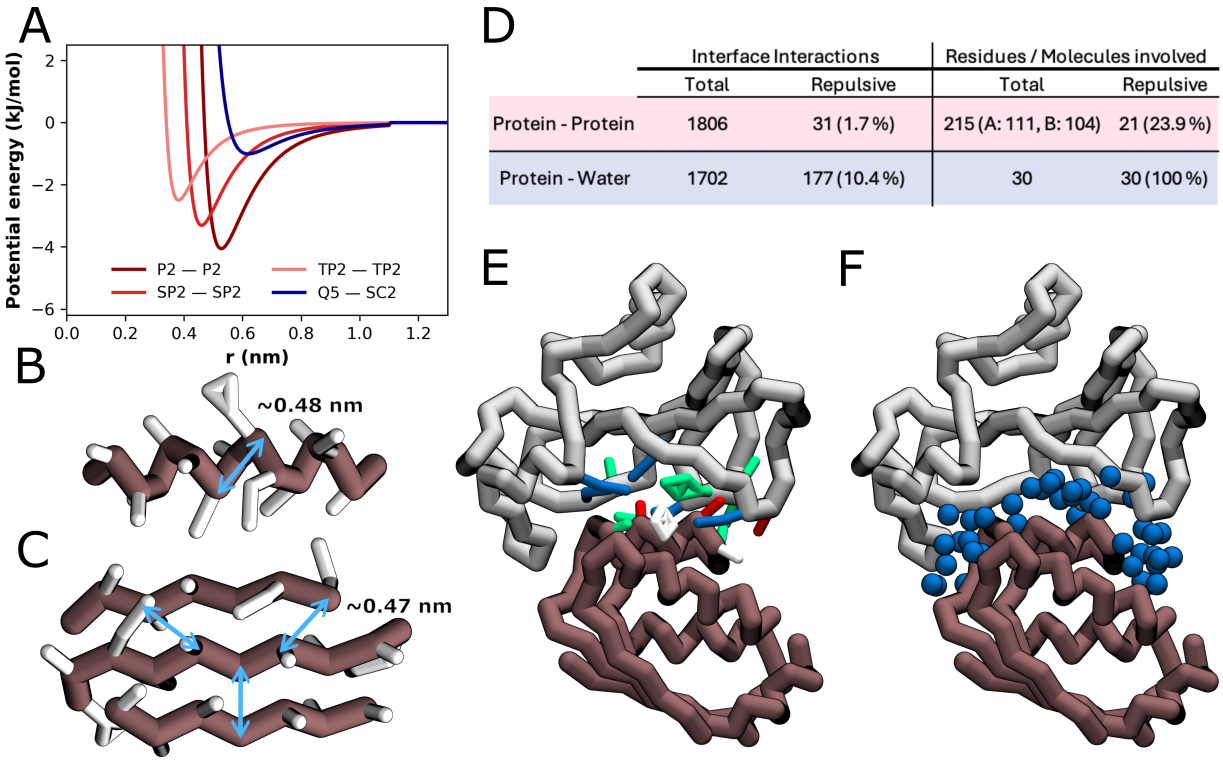
Having simulated both soluble and transmembrane protein complexes, we observe an apparent trend that solution complexes are not stable and transmembrane complexes are. We have simulated 4 additional solution complexes, including the human leukemia inhibitory factor in complex with the gp130 domain (PDB: 1PVH),[68] human cyclophilin A bound to the HIV-1 capsid (PDB: 1AK4),[69] the *E. coli* methionine apo-repressor homodimer (PDB: 1MJM),[70] and mouse nerve growth factor homodimer (PDB: 5LSD),[71] which all follow the same trend of being stable in atomistic simulations but quickly dissociate in Martini 3 simulations without a quaternary bias network, irrespective of the intrachain Gō-like model used. Taken together, this is evidence that protein intermolecular interactions for solution complexes are not stable in the current implementation of Martini 3 proteins. This would be expected from the observation that tertiary structures are not stable.[5,6] However, as we observe transmembrane dimers to be stable, we speculate that protein-protein interactions are somewhat balanced (with respect to the lipids). Instead, imbalances in protein-water interactions could lead to an underestimation of the hydrophobic effect, destabilizing solution complexes like the Ubiquitin/UBA domain system.

Interestingly, there have been recent studies of Martini 3 applied to intrinsically disordered proteins (IDPs), suggesting that either protein-protein interactions are too strong[72] or, alternatively, that protein-water interactions are too weak.[73] It has been clearly shown that either downscaling protein-protein interactions or upscaling protein-water interactions can reconcile simulation results with experimental results of IDPs and proteins with disordered regions.[72,73] Perplexingly, both of these approaches would further destabilize solution protein complexes, which strongly suggests that a flat scaling of either protein-protein or protein-water interactions cannot be a general solution, even though it is useful to the spe-

cific application of simulating IDPs. Further complicating matters, there have also been conflicting reports that protein-water interactions need to be downscaled by 10 % for proper insertion of transmembrane helices into self-assembled phospholipid bilayers and micelles.[17,74] The conflicting studies of flat scaling of protein-water interactions hint that the Martini 3 protein model still has room for improvement, and in the following, we want to direct the attention to some areas where the current description of proteins in Martini 3 could be enhanced, with a particular focus on soluble protein-protein complexes.

(1) The use of regularly sized beads in the protein backbone cannot represent the tight packing required in secondary structures. Backbone-backbone intrachain distances are often already within the steep repulsive regime of the LJ potential of regular backbone beads (interaction between bead type P2-P2, which has a minimum distance at 0.528 nm, Figure 7,A.). The impact of this is obvious when looking at the intrachain packing of secondary structures given the associated numerical and structural instabilities, particularly in α-helices and β-sheets, where most interaction distances are around 0.47 to 0.48 nm (Figure 7, B & C). While this may not seem like a large deviation from the LJ minimum, the steep repulsive regime in the LJ potential results in high forces, even for distances marginally smaller than the minimum. Currently, this issue is mitigated by the use of intramolecular exclusions and secondary structure biases, like the DSSP restrains or the OLIVES secondary network.

However, this effect also impacts protein-protein interactions. Many protein interfaces feature amino acids that are so tightly packed that when mapped to Martini 3 side chains, which are made up of beads of various sizes, may also fall within the repulsive regime of their respective LJ potentials. Indeed this is the case in the Barnase-Barstar complex, which we use as an example here. Analyzing the crystal complex mapped to Martini 3, the interface contains 215 residues with 1806 possible LJ pair interactions (within the Martini non-bonded cutoff of 1.1 nm[1]) out of which 215 are involved in repulsive interactions (Figure 7, D & E). Interactions between specific amino acid residues will be especially susceptible to this issue. In the Martini 3 interaction matrix, "super repulsive" interaction levels[1] have



	Interface Interactions		Residues / Molecules involved	
	Total	Repulsive	Total	Repulsive
Protein - Protein	1806	31 (1.7 %)	215 (A: 111, B: 104)	21 (23.9 %)
Protein - Water	1702	177 (10.4 %)	30	30 (100 %)

Figure 7: Protein-protein & protein-water packing defects in the Martini 3 protein model. (A) LJ potentials between 4 pairs of Martini 3 beads (P2-P2 interaction pair reflects the interaction between protein backbone beads, while Q5-SC2 reflects the interaction between GLU and LEU sidechains). Interactions are cut-off at 1.1 nm to reflect standard Martini 3 non-bonded settings. Martini 3 α-helix (B) and β-sheet (C) structures showcasing short backbone-backbone distances within the LJ repulsive regime of the P2 bead type. Backbone and sidechains are shown in brown and white, respectively. (D) Interaction pair analysis of the Barnase-Barstar interface. A contact is deemed repulsive if the pair distance is shorter than the distance associated with the LJ minimum for that particular pair interaction. Here, interface residues were defined as those within 0.8 nm of the opposite protein partner. Interface water molecules were defined as those within 0.4 nm of both protein chains. Barnase-Barstar protein complex showcasing the interface amino acid residues (E) and water molecules (F) engaged in repulsive interactions. The Martini 3 structure was directly mapped from the experimental crystal structure. Barnase and barstar are represented in white and brown, respectively. Sidechains engaged in repulsive interactions are shown and colored by residue type (basic in blue, acidic in red, polar in green, and hydrophobic in white) and water molecules engaged in repulsive interactions are represented in blue.

their LJ sigma parameter increased with respect to the standard bead size sigma. As a consequence, their LJ minimum is located further apart (Fig 7,A). This impacts interactions between charged (e.g. LYS, ARG, GLU, ASP) and hydrophobic (e.g. ILE, LEU, ALA, VAL) residues (exemplified by the Q5-SC2 interaction, Fig 7,A). The interfaces of many soluble complexes often have some of these residues in close proximity and the Barnase-Barstar complex has one such residue pair at the interface (GLU60-LEU144). Overall, the impact of these issues on protein complex stability is clear, as even in the initial steps of minimization and equilibration, the immediate repulsion between the protein monomers will contribute to the destabilization of the complex.

(2) The current water model in Martini 3 can also pose a substantial obstacle for the accurate depiction of protein-protein interactions. A single regular water bead in Martini 3 represents 4 all-atom water molecules. By itself, this excludes the possibility of water-mediated interactions as the cavities that these bridging water molecules occupy are often too small for a regular Martini water bead. Analyzing the interface water molecules present in the Barnase-Barstar crystal structure, we observe that out of the 30 water molecules in the interface, all would be involved in repulsive LJ interactions if replaced by a regular water bead (Figure 7, D & F). This analysis is not perfectly fair, as it considers each water molecule as a regular Martini bead representing the size of a cluster of 4 water molecules, but it clearly represents how the volume of the regular Martini 3 water beads compares to the volume of the water pockets present at protein complex interfaces.

A solution for this case would be to use smaller Martini water models. Small and tiny sized Martini 3 water beads already exist (which represent 3 to 1 and 2 to 1 all-atom molecules to a water bead, respectively), but there are no clear solutions on how to restrain these smaller water beads to the protein interface. Usage of small and tiny water beads as pure solvents is discouraged, as they are still not fully parameterized with this use case in mind. Moreover, using pure solvents made up of small or tiny beads substantially increases simulation particle count, impacting the simulation speed-up advantage of Martini 3 over conventional all-atom

models. If the issue regarding water bead size is eventually solved, one still has to overcome issues related to the directionality of water interactions. Furthermore, the LJ potentials used in Martini are isotropic, while real water interactions are not. Accounting for this will also play a role in stabilizing water-mediated contacts.

(3) Electrostatics play a crucial role in protein-protein interactions. Many interfaces of soluble protein complexes show substantial charge complementarity in their interface, as is the case of the Barnase-Barstar complex. Standard Martini 3 water beads do not currently carry (partial) charges and therefore do not account for polarization effects. In the current Martini 3 implementation, screening of electrostatic interactions is instead compensated implicitly, assuming a uniform relative dielectric constant. While this approximation is reasonable for bulk water, problems arise at the interfaces between water and other phases and in the vicinity of charged groups. Moreover, because of the applied implicit screening, the interaction strength of polar moieties is underestimated in non-polarizable solvents and in low-dielectric environments, found in the interior of most proteins and de-solvated pockets.[25] It is therefore very likely that the lack of polarizability in the Martini water model is compromising polar and charged interactions responsible for stabilizing soluble protein-protein complexes. This hypothesis becomes more probable considering that, out of all the systems tested in this work, none of the membrane-inserted complexes, where water polarizability is possibly less relevant, dissociated. Taken together, the striking difference in stability between solution and transmembrane complexes hints at interaction imbalances in the protein model with respect to water, and the development of a polarizable water model could possibly help balance protein-water interactions further. Additionally, revisiting the side chain interaction matrix and rebalancing side chain interactions to re-evaluate their partitions and self-interactions may also help resolve these imbalances.

(4) Side chain rotamers play a major role in the interaction geometries within protein-protein interfaces[75–77] and the refinement of side chain torsions has been instrumental in atomistic force fields,[78] which has led to improved protein-protein interactions.[42] The current

Martini 3 approach of restraining side chain conformations based on the starting structure (using the ScFix method[54]) is sub-optimal, and fitting residue-specific side chain dihedral potentials could further improve the Martini 3 protein model. While ScFix is a definite requirement in the current Martini 3 protein model, future developments should aim to remove the need for such biases, albeit this is complicated by distinct environment-dependent rotamer distributions between soluble and transmembrane proteins.[75]

Overall, a combination of several factors, ranging from packing, interaction balance, water model, and electrostatics, may be behind the current issues regarding Martini 3 soluble protein complexes. Future developments regarding the protein model will focus on mitigating some of these limitations,[79] but for now, current iterations of the Martini 3 protein model will require biases applied to the quaternary network, like OLIVES, to better reproduce soluble complexes.

Conclusion

In this work, we have presented a new approach for stabilizing protein structure in Martini 3 simulations called OLIVES. We show that explicit modeling of a coarse-grained hydrogen bond network is enough to stabilize protein structures through a novel Gō-like model implementation using OLIVES. The OLIVES Gō-like model was validated against 4 test systems of protein complexes displaying diverse folds and secondary structure content, and it was shown that, in general, OLIVES enables the simulation of proteins without the use of secondary structure restraints. By reducing the number of Gō-like bonds and secondary structure restraints, OLIVES has the potential to speed up Martini 3 simulations of proteins by up to 30%. We find that solution protein complexes require the use of a quaternary bias network in order to avoid dissociation regardless of which structure bias method is used. Finally, we lay out future directions for improving the current protein model and protein-protein interactions in the Martini 3 coarse-grained force field.

Acknowledgement

The authors would like to thank Sigurd Friis Truelsen, Novozymes A/S, Denmark, for testing OLIVES during development. K.B.P. and B.S. acknowledge funding by the Lundbeck Foundation (R276-2018-671) and the Independent Research Fund Denmark | Technology & Production (0136-00148B). A.D.S. and B.S. acknowledge funding by the Novo Nordisk Foundation (NNF20OC0065431). K.B.P, A.D.S., and B.S. utilized GPU hardware at the Centre for Scientific Computing Aarhus (CSCAA)[52] funded by the Novo Nordisk Foundation (NNF18OC0032608). L.B.A. and P.C.T.S. acknowledge the support of the French National Center for Scientific Research (CNRS) and the funding from a research collaboration agreement with PharmCADD. L.B.A. and P.C.T.S. also acknowledge the support of the Centre Blaise Pascal's IT test platform at ENS de Lyon (Lyon, France) for the computer facilities. The platform operates the SIDUS solution[80] developed by Emmanuel Quemener. S.J.M. acknowledges funding from the European Research Council with the Advanced grant "COMP-O-CELL" (101053661).

Supporting Information Available

The supporting information contains supporting figures. OLIVES is open source and distributed under the Apache License Version 2.0 and is available via GitHub (`https://github.com/Martini-Force-Field-Initiative/OLIVES`)

References

(1) Souza, P. C. T. et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods* **2021**, *18*, 382–388.

(2) Stevens, J. A.; Grünewald, F.; van Tilburg, P. M.; König, M.; Gilbert, B. R.;

Brier, T. A.; Thornburg, Z. R.; Luthey-Schulten, Z.; Marrink, S. J. Molecular dynamics simulation of an entire cell. *Frontiers in Chemistry* **2023**, *11*, 1106495.

(3) Ellis, R. J.; Minton, A. P. Join the crowd. *Nature* **2003**, *425*, 27–28.

(4) Kroon, P. C.; Grünewald, F.; Barnoud, J.; van Tilburg, M.; Souza, P. C. T.; Wassenaar, T. A.; Marrink, S. J. Martinize2 and Vermouth: Unified Framework for Topology Generation. *eLife (https://elifesciences.org/reviewed-preprints/90627)* **2023**,

(5) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition. *Journal of Chemical Theory and Computation* **2009**, *5*, 2531–2543.

(6) Poma, A. B.; Cieplak, M.; Theodorakis, P. E. Combining the MARTINI and structure-based coarse-grained approaches for the molecular dynamics studies of conformational transitions in proteins. *Journal of Chemical Theory and Computation* **2017**, *13*, 1366–1374.

(7) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules* **1983**, *22*, 2577–2637.

(8) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Progress in Biophysics and Molecular Biology* **1984**, *44*, 97–179.

(9) Fonseca Guerra, C.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. Hydrogen bonding in DNA base pairs: reconciliation of theory and experiment. *Journal of the American Chemical Society* **2000**, *122*, 4117–4128.

(10) Fleming, P. J.; Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Science* **2005**, *14*, 1911–1917.

(11) Hubbard, R. E.; Haider, M. K. *Hydrogen bonds in proteins: role and strength*; John Wiley & Sons, Ltd, 2010.

(12) Hao, M.-H. Theoretical calculation of hydrogen-bonding strength for drug molecules. *Journal of Chemical Theory and Computation* **2006**, *2*, 863–872.

(13) Du, Q.-S.; Wang, Q.-Y.; Du, L.-Q.; Chen, D.; Huang, R.-B. Theoretical study on the polar hydrogen-π (Hp-π) interactions between protein side chains. *Chemistry Central Journal* **2013**, *7*, 1–8.

(14) Mundlapati, V. R.; Ghosh, S.; Bhattacherjee, A.; Tiwari, P.; Biswal, H. S. Critical assessment of the strength of hydrogen bonds between the sulfur atom of methionine/-cysteine and backbone amides in proteins. *The Journal of Physical Chemistry Letters* **2015**, *6*, 1385–1389.

(15) Best, R. B.; Chen, Y.-G.; Hummer, G. Slow protein conformational dynamics from multiple experimental structures: the helix/sheet transition of arc repressor. *Structure* **2005**, *13*, 1755–1763.

(16) Lamprakis, C.; Andreadelis, I.; Manchester, J.; Velez-Vega, C.; Duca, J. S.; Cournia, Z. Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. *Journal of Chemical Theory and Computation* **2021**, *17*, 3088–3102.

(17) Claveras Cabezudo, A.; Athanasiou, C.; Tsengenes, A.; Wade, R. C. Scaling Protein–Water Interactions in the Martini 3 Coarse-Grained Force Field to Simulate Transmembrane Helix Dimers in Different Lipid Environments. *Journal of Chemical Theory and Computation* **2023**, *19*, 2109–2119.

(18) Periole, X.; Knepp, A. M.; Sakmar, T. P.; Marrink, S. J.; Huber, T. Structural determinants of the supramolecular organization of G protein-coupled receptors in bilayers. *Journal of the American Chemical Society* **2012**, *134*, 10959–10965.

(19) Periole, X.; Zeppelin, T.; Schiøtt, B. Dimer interface of the human serotonin transporter and effect of the membrane composition. *Scientific Reports* **2018**, *8*, 5080.

(20) Zeppelin, T.; Pedersen, K. B.; Berglund, N. A.; Periole, X.; Schiøtt, B. Effect of palmitoylation on the dimer formation of the human dopamine transporter. *Scientific Reports* **2021**, *11*, 4164.

(21) Marrink, S. J.; De Vries, A. H.; Mark, A. E. Coarse grained model for semiquantitative lipid simulations. *The Journal of Physical Chemistry B* **2004**, *108*, 750–760.

(22) 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 of Physical Chemistry B* **2007**, *111*, 7812–7824.

(23) Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences* **2013**, *110*, 17874–17879.

(24) Kang, Y. K. Which functional form is appropriate for hydrogen bond of amides? *The Journal of Physical Chemistry B* **2000**, *104*, 8321–8326.

(25) Amin, M.; Küpper, J. Variations in proteins dielectric constants. *ChemistryOpen* **2020**, *9*, 691–694.

(26) Sheu, S.-Y.; Yang, D.-Y.; Selzle, H.; Schlag, E. Energetics of hydrogen bonds in peptides. *Proceedings of the National Academy of Sciences* **2003**, *100*, 12683–12687.

(27) Bogojeski, M.; Vogt-Maranto, L.; Tuckerman, M. E.; Müller, K.-R.; Burke, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nature Communications* **2020**, *11*, 5223.

(28) Galil, Z. Efficient algorithms for finding maximum matching in graphs. *ACM Computing Surveys (CSUR)* **1986**, *18*, 23–38.

(29) Hagberg, A.; Swart, P.; S Chult, D. *Conference: Exploring network structure, dynamics, and function using NetworkX (`https://www.osti.gov/biblio/960616`)*; 2008.

(30) Berendsen, H. J.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Computer Physics Communications* **1995**, *91*, 43–56.

(31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(32) Peschard, P.; Kozlov, G.; Lin, T.; Mirza, I. A.; Berghuis, A. M.; Lipkowitz, S.; Park, M.; Gehring, K. Structural basis for ubiquitin-mediated dimerization and activation of the ubiquitin protein ligase Cbl-b. *Molecular Cell* **2007**, *27*, 474–485.

(33) Buckle, A. M.; Schreiber, G.; Fersht, A. R. Protein-protein recognition: Crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. *Biochemistry* **1994**, *33*, 8878–8889.

(34) Snijder, H.; Ubarretxena-Belandia, I.; Blaauw, M.; Kalk, K.; Verheij, H.; Egmond, M.; Dekker, N.; Dijkstra, B. Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **1999**, *401*, 717–721.

(35) Dutzler, R.; Campbell, E. B.; MacKinnon, R. Gating the selectivity filter in ClC chloride channels. *Science* **2003**, *300*, 108–112.

(36) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The protein data bank. *Nucleic Acids Research* **2000**, *28*, 235–242.

(37) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.

(38) Lee, J. et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. *Journal of Chemical Theory and Computation* **2016**, *12*, 405–413.

(39) Gordon, J. C.; Myers, J. B.; Folta, T.; Shoja, V.; Heath, L. S.; Onufriev, A. H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic Acids Research* **2005**, *33*, W368–W371.

(40) Faraldo-Gómez, J. D.; Roux, B. Electrostatics of ion stabilization in a ClC chloride channel homologue from Escherichia coli. *Journal of Molecular Biology* **2004**, *339*, 981–1000.

(41) Chadda, R.; Bernhardt, N.; Kelley, E. G.; Teixeira, S. C.; Griffith, K.; Gil-Ley, A.; Öztürk, T. N.; Hughes, L. E.; Forsythe, A.; Krishnamani, V.; Faraldo-Gómez, J. D.; Robertson, J. L. Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states. *eLife* **2021**, *10*, e63288.

(42) Piana, S.; Robustelli, P.; Tan, D.; Chen, S.; Shaw, D. E. Development of a force field for the simulation of single-chain proteins and protein–protein complexes. *Journal of Chemical Theory and Computation* **2020**, *16*, 2494–2507.

(43) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *The Journal of Physical Chemistry B* **2015**, *119*, 5113–5123.

(44) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N*log (N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(45) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71–73.

(46) Piller, P.; Semeraro, E. F.; Rechberger, G. N.; Keller, S.; Pabst, G. Allosteric modulation of integral protein activity by differential stress in asymmetric membranes. *PNAS Nexus* **2023**, *2*, pgad126.

(47) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997**, *18*, 1463–1472.

(48) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **2007**, *126*, 014101.

(49) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(50) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **1983**, *50*, 1055–1076.

(51) Bernetti, M.; Bussi, G. Pressure control using stochastic cell rescaling. *The Journal of Chemical Physics* **2020**, *153*.

(52) Centre for Scientific Computing Aarhus (CSCAA). `https://phys.au.dk/forskning/faciliteter/cscaa/`.

(53) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528 – 1532.

(54) Herzog, F. A.; Braun, L.; Schoen, I.; Vogel, V. Improved side chain dynamics in MARTINI simulations of protein–lipid interfaces. *Journal of Chemical Theory and Computation* **2016**, *12*, 2446–2458.

(55) Wołek, K.; Gómez-Sicilia, À.; Cieplak, M. Determination of contact maps in proteins:

a combination of structural and chemical approaches. *The Journal of Chemical Physics* **2015**, *143*, 243105.

(56) Wassenaar, T. A.; Ingólfsson, H. I.; Bockmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational lipidomics with insane: a versatile tool for generating custom membranes for molecular simulations. *Journal of Chemical Theory and Computation* **2015**, *11*, 2144–2155.

(57) Kim, H.; Fábián, B.; Hummer, G. Neighbor List Artifacts in Molecular Dynamics Simulations. *ChemRxiv* **2023**,

(58) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-grained protein models and their applications. *Chemical Reviews* **2016**, *116*, 7898–7936.

(59) Marrink, S. J.; Tieleman, D. P. Perspective on the Martini model. *Chemical Society Reviews* **2013**, *42*, 6801–6822.

(60) Karanicolas, J.; Brooks III, C. L. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Science* **2002**, *11*, 2351–2361.

(61) Wang, Y.; Tian, P.; Boomsma, W.; Lindorff-Larsen, K. Monte Carlo sampling of protein folding by combining an all-atom physics-based model with a native state bias. *The Journal of Physical Chemistry B* **2018**, *122*, 11174–11185.

(62) Hartley, R. W. Barnase and barstar: two small proteins to fold and fit together. *Trends in Biochemical Sciences* **1989**, *14*, 450–454.

(63) Schreiber, G.; Fersht, A. R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993**, *32*, 5145–5150.

(64) Desikan, R.; Patra, S. M.; Sarthak, K.; Maiti, P. K.; Ayappa, K. Comparison of coarse-grained (MARTINI) and atomistic molecular dynamics simulations of α and β toxin nanopores in lipid membranes. *Journal of Chemical Sciences* **2017**, *129*, 1017–1030.

(65) Chadda, R.; Cliff, L.; Brimberry, M.; Robertson, J. L. A model-free method for measuring dimerization free energies of CLC-ec1 in lipid bilayers. *Journal of General Physiology* **2018**, *150*, 355–365.

(66) Groll, M.; Ditzel, L.; Löwe, J.; Stock, D.; Bochtler, M.; Bartunik, H. D.; Huber, R. Structure of 20S proteasome from yeast at 2.4 Å resolution. *Nature* **1997**, *386*, 463–471.

(67) Páll, S.; Abraham, M. J.; Kutzner, C.; Hess, B.; Lindahl, E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. Solving Software Challenges for Exascale. Cham, 2015; pp 3–27.

(68) Boulanger, M. J.; Bankovich, A. J.; Kortemme, T.; Baker, D.; Garcia, K. C. Convergent mechanisms for recognition of divergent cytokines by the shared signaling receptor gp130. *Molecular Cell* **2003**, *12*, 577–589.

(69) Gamble, T. R.; Vajdos, F. F.; Yoo, S.; Worthylake, D. K.; Houseweart, M.; Sundquist, W. I.; Hill, C. P. Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell* **1996**, *87*, 1285–1294.

(70) Garvie, C. W.; Phillips, S. E. Direct and indirect readout in mutant Met repressor–operator complexes. *Structure* **2000**, *8*, 905–914.

(71) Paoletti, F.; De Chiara, C.; Kelly, G.; Covaceuszach, S.; Malerba, F.; Yan, R.; Lamba, D.; Cattaneo, A.; Pastore, A. Conformational rigidity within plasticity promotes differential target recognition of nerve growth factor. *Frontiers in Molecular Biosciences* **2016**, *3*, 83.

(72) Thomasen, F. E.; Skaalum, T.; Kumar, A.; Srinivasan, S.; Vanni, S.; Lindorff-Larsen, K. Recalibration of protein interactions in Martini 3. *bioRxiv* **2023**, 2023–05.

(73) Thomasen, F. E.; Pesce, F.; Roesgaard, M. A.; Tesei, G.; Lindorff-Larsen, K. Improving

Martini 3 for disordered and multidomain proteins. *Journal of Chemical Theory and Computation* **2022**, *18*, 2033–2041.

(74) Spinti, J. K.; Nunes, F. N.; Melo, M. N. Room for improvement in the initial martini 3 parameterization of peptide interactions. *Chemical Physics Letters* **2023**, *819*, 140436.

(75) Chamberlain, A. K.; Bowie, J. U. Analysis of side-chain rotamers in transmembrane proteins. *Biophysical Journal* **2004**, *87*, 3460–3469.

(76) Guharoy, M.; Janin, J.; Robert, C. H. Side-chain rotamer transitions at protein-protein interfaces. *PROTEINS: Structure, Function, and Bioinformatics* **2010**, *78*, 3219–3225.

(77) Watkins, A. M.; Bonneau, R.; Arora, P. S. Side-chain conformational preferences govern protein–protein interactions. *Journal of the American Chemical Society* **2016**, *138*, 10386–10389.

(78) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950–1958.

(79) Borges-Araújo, L.; Patmanidis, I.; Singh, A. P.; Santos, L. H. S.; Sieradzan, A. K.; Vanni, S.; Czaplewski, C.; Pantano, S.; Shinoda, W.; Monticelli, L.; Liwo, A.; Marrink, S. J.; Souza, P. C. T. Pragmatic Coarse-Graining of Proteins: Models and Applications. *Journal of Chemical Theory and Computation* **2023**,

(80) Quemener, E.; Corvellec, M. SIDUS—the Solution for Extreme Deduplication of an Operating System. *Linux J.* **2013**, *2013*.

Supporting information

Table S1: Residue classification scheme used in SI Figure S1-S4.

Not Hydrogen Bonding	Hydrogen Bonding	Hydrophobic	Hydrophilic Uncharged	Hydrophilic Charged
G,A,V,L,I,P,F	C,M,N,Q,D,E,T S,K,R,H,Y,W	G,A,V,L,I,M,P F,Y,W	C,N,Q,T,S,H	D,E,K,R

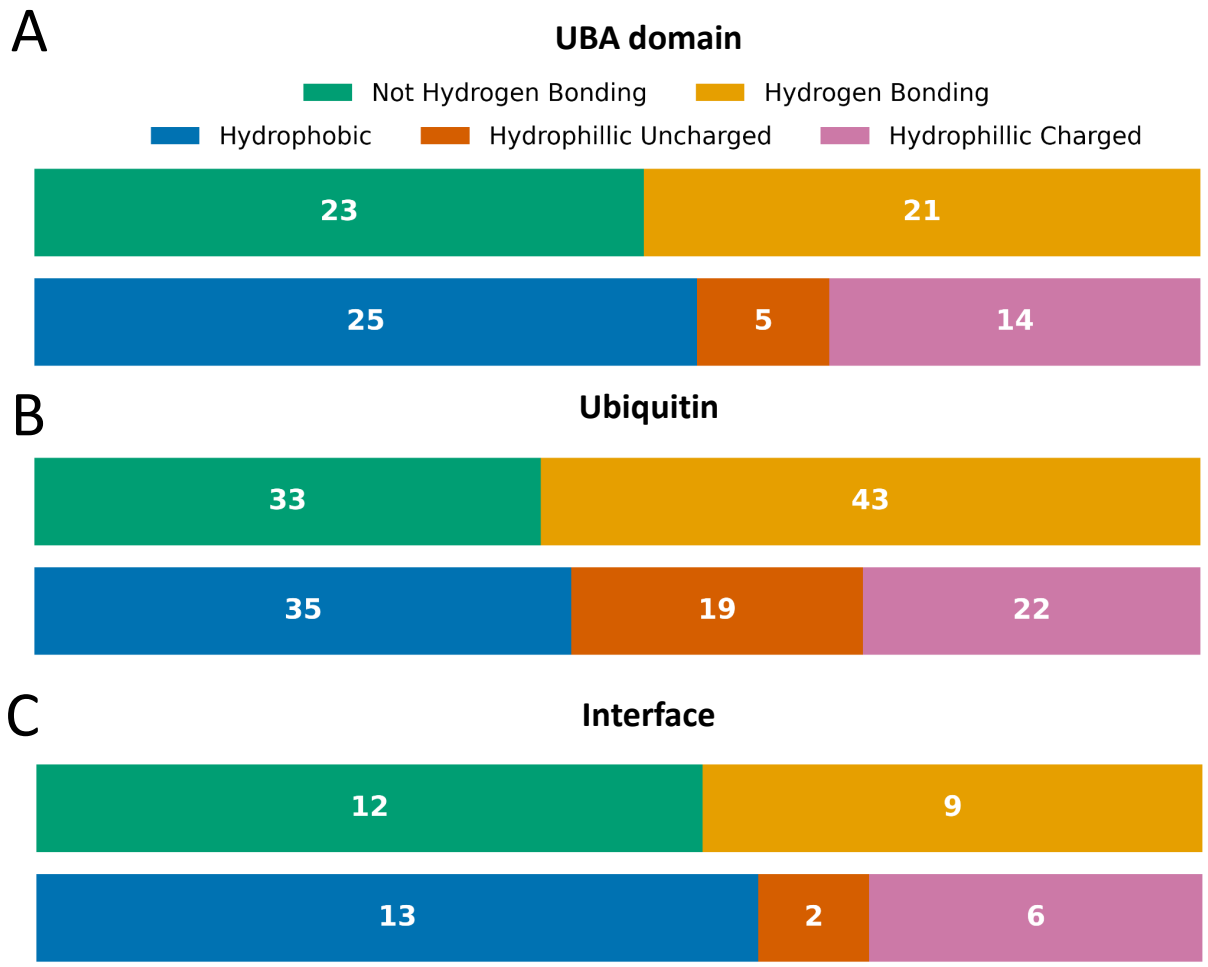


Figure S1: Classification of the residues found in native contacts in the UBA domain/Ubiquitin complex. (A) UBA domain, (B) Ubiquitin, and (C) the interface in the UBA domain/Ubiquitin complex.

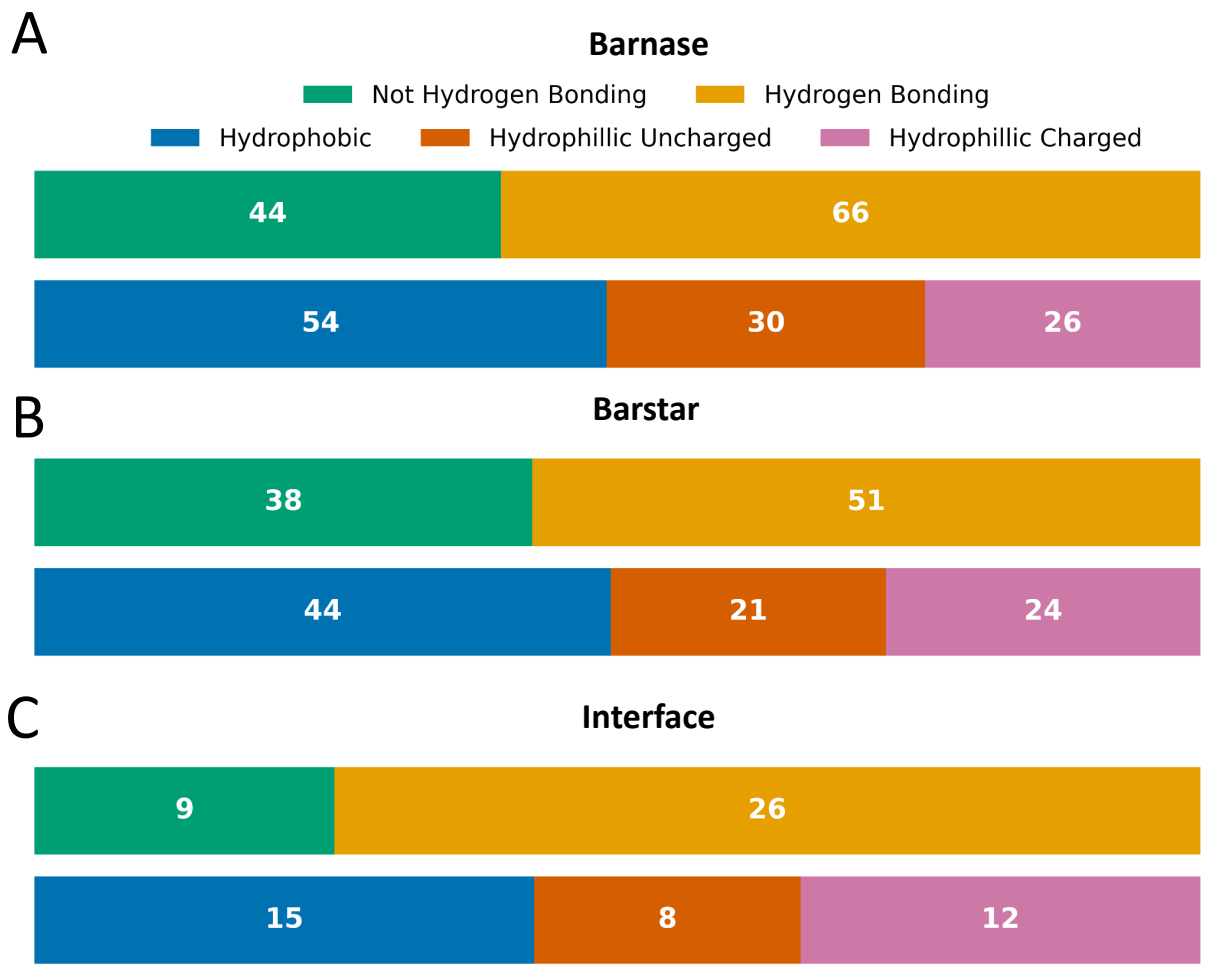


Figure S2: Classification of the residues found in native contacts in the Barnase/Barstar complex. (A) Barnase, (B) Barstar, and (C) the interface in the Barnase/Barstar complex.

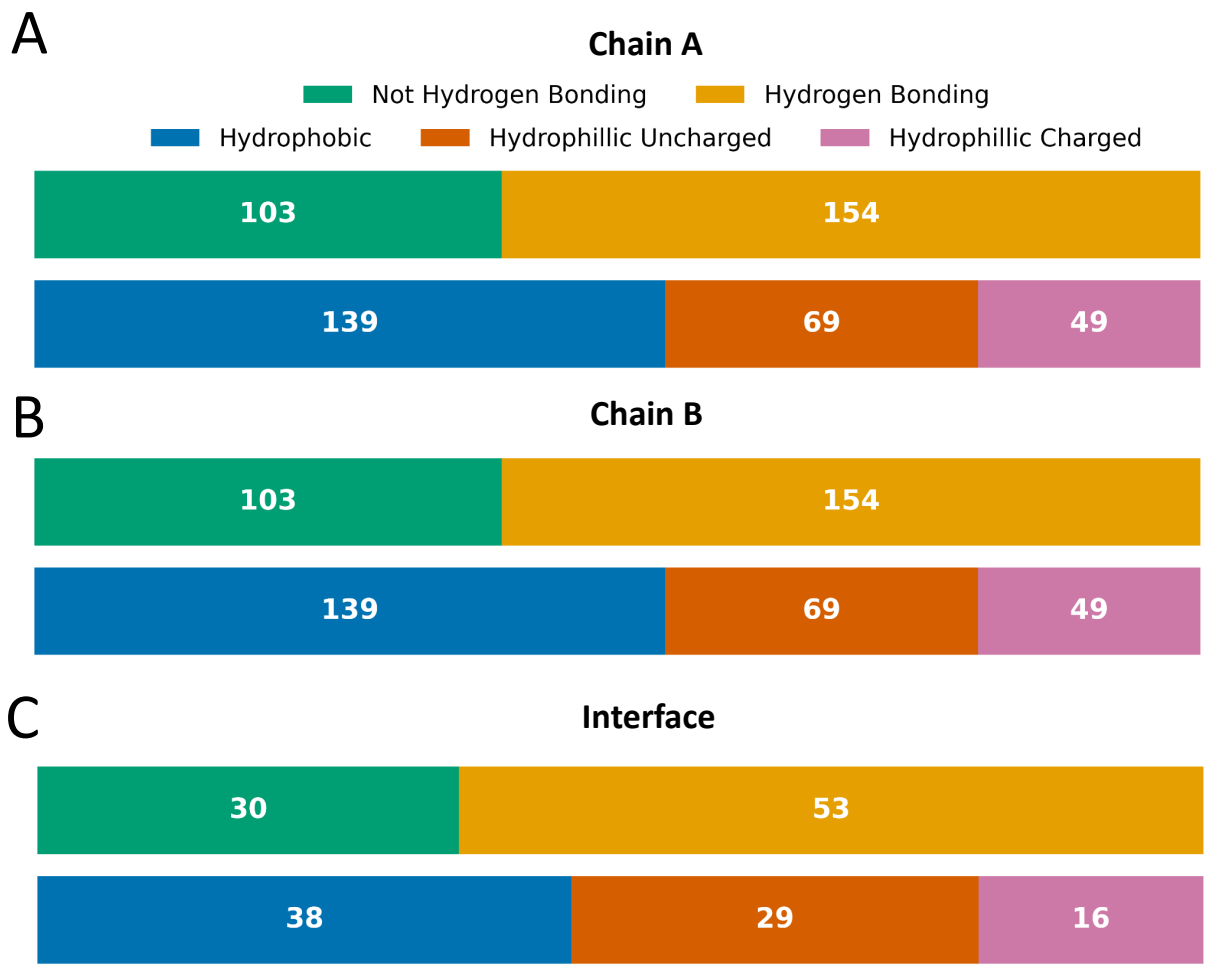


Figure S3: Classification of the residues found in native contacts of the OMPLA homodimer. (A) Chain A, (B) Chain B, and (C) the interface.

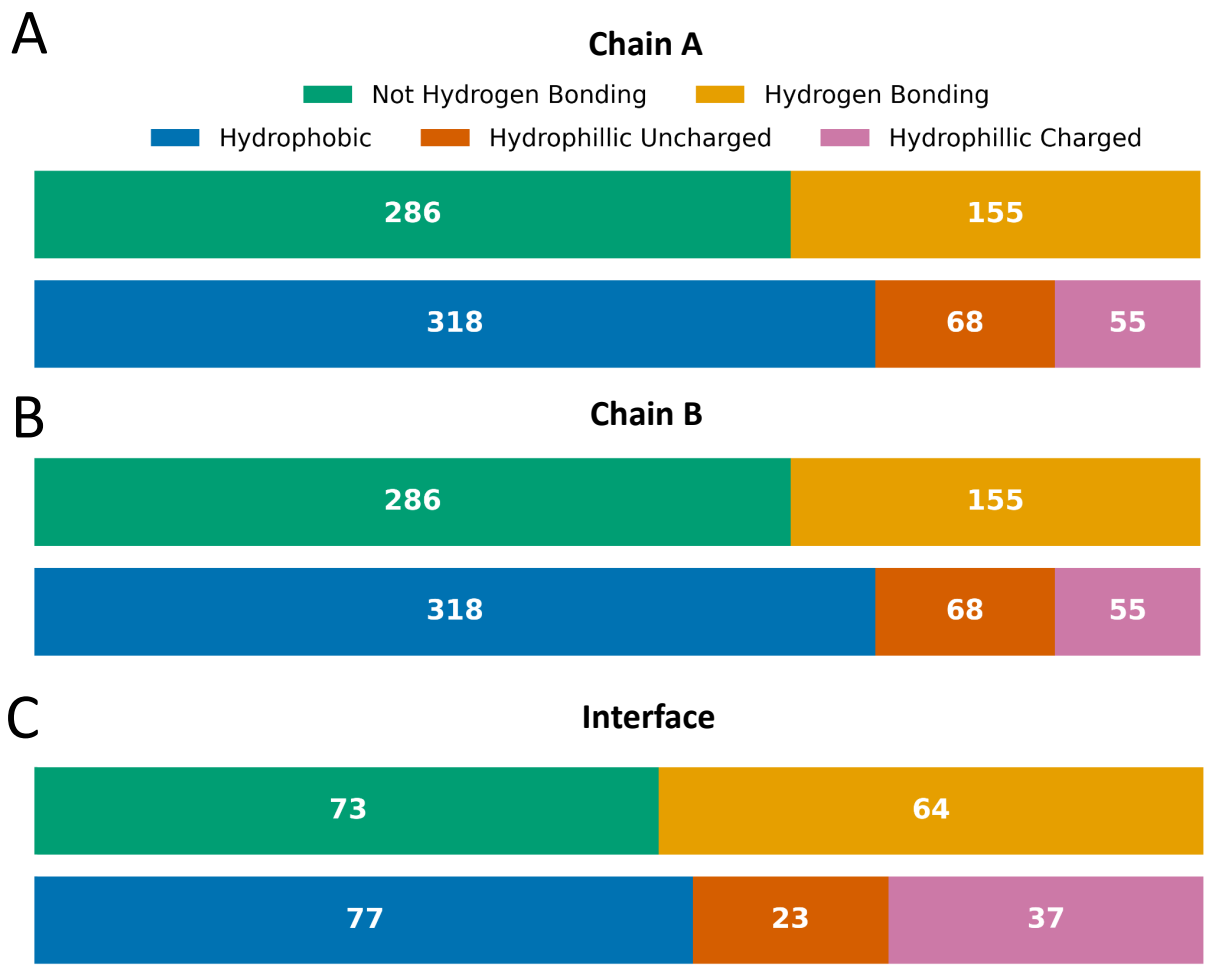


Figure S4: Classification of the residues found in native contacts of the ClC homodimer. (A) Chain A, (B) Chain B, and (C) the interface.

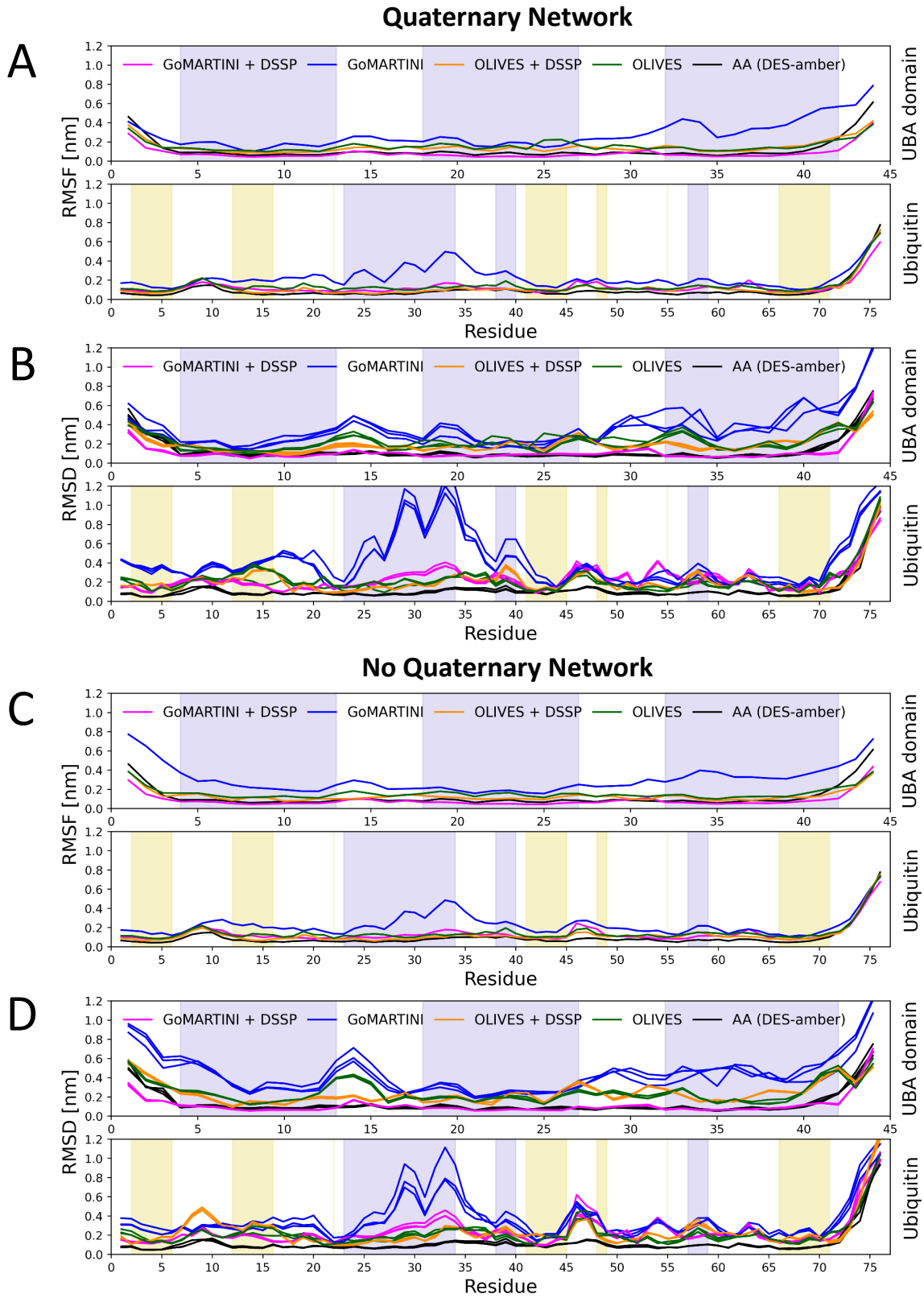


Figure S5: Additional data for the Ubiquition/UBA domain system. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (A) Residue RMSF, using a quaternary network. (B) Residue RMSD, using a quaternary network. (C) Residue RMSF, without using a quaternary network. (D) Residue RMSD, without using a quaternary network.

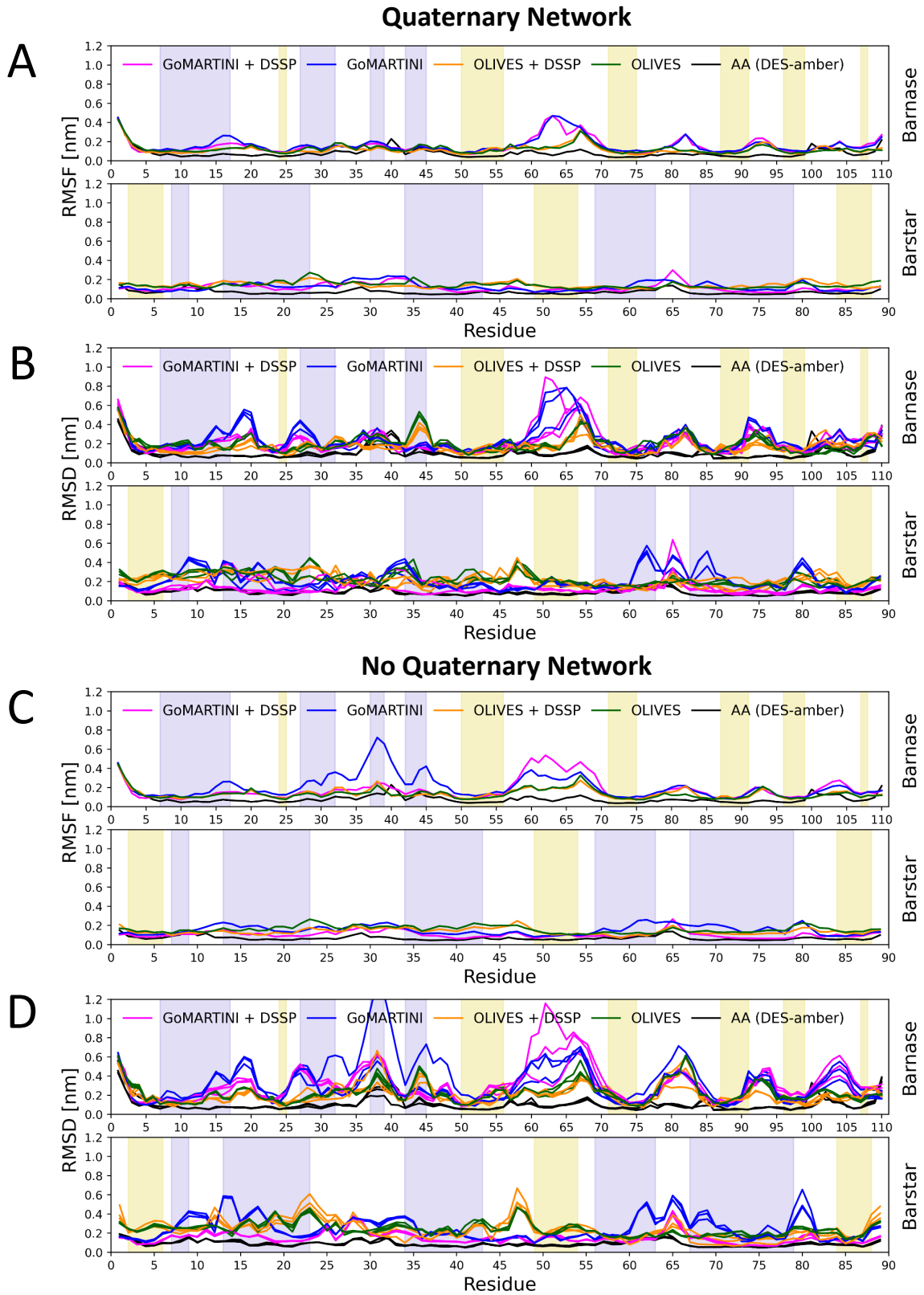


Figure S6: Additional data for the Barnase/Barstar system. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (A) Residue RMSF, using a quaternary network. (B) Residue RMSD, using a quaternary network. (C) Residue RMSF, without using a quaternary network. (D) Residue RMSD, without using a quaternary network.

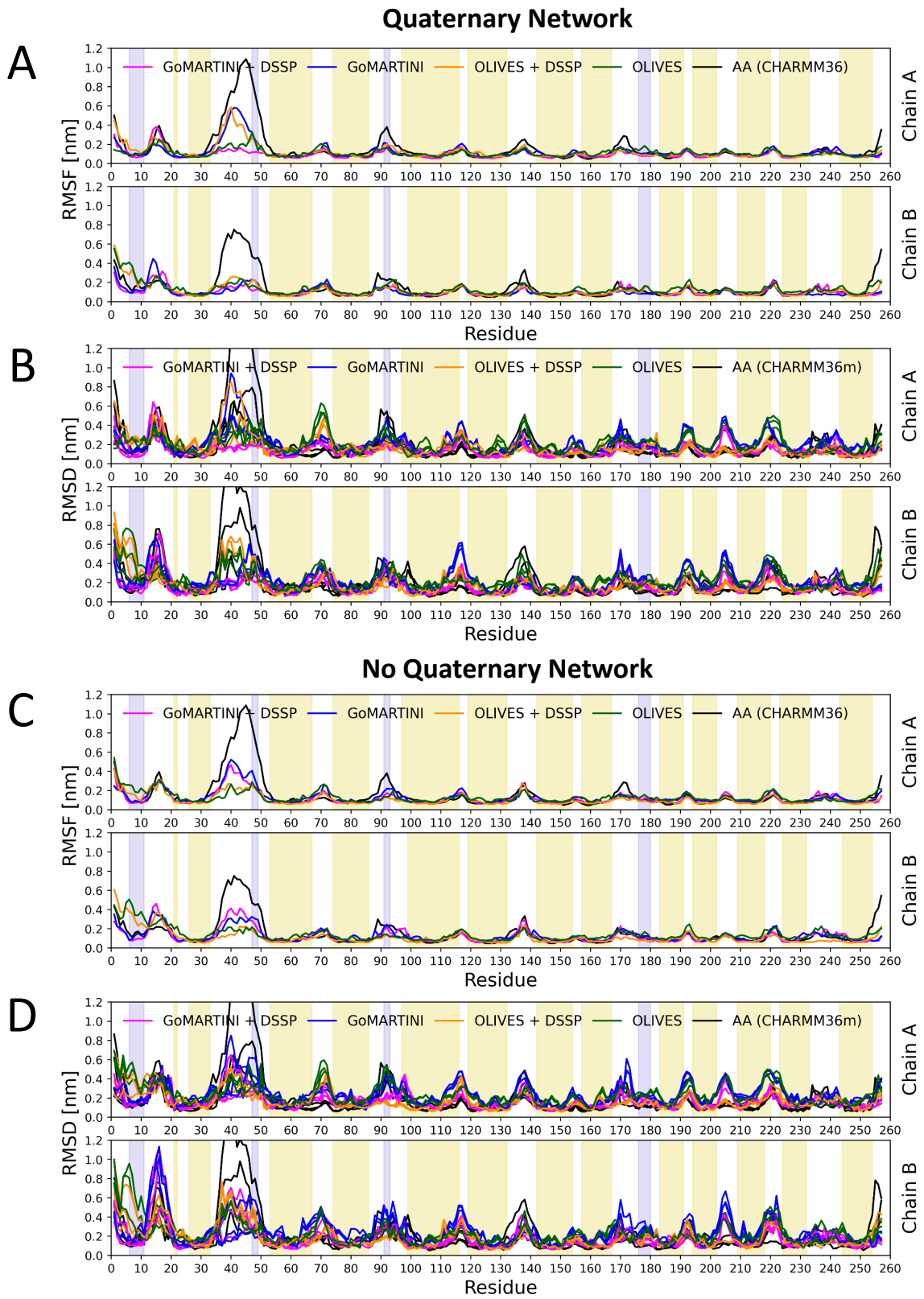


Figure S7: Additional data for the OMPLA homodimer system. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (A) Residue RMSF, using a quaternary network. (B) Residue RMSD, using a quaternary network. (C) Residue RMSF, without using a quaternary network. (D) Residue RMSD, without using a quaternary network.

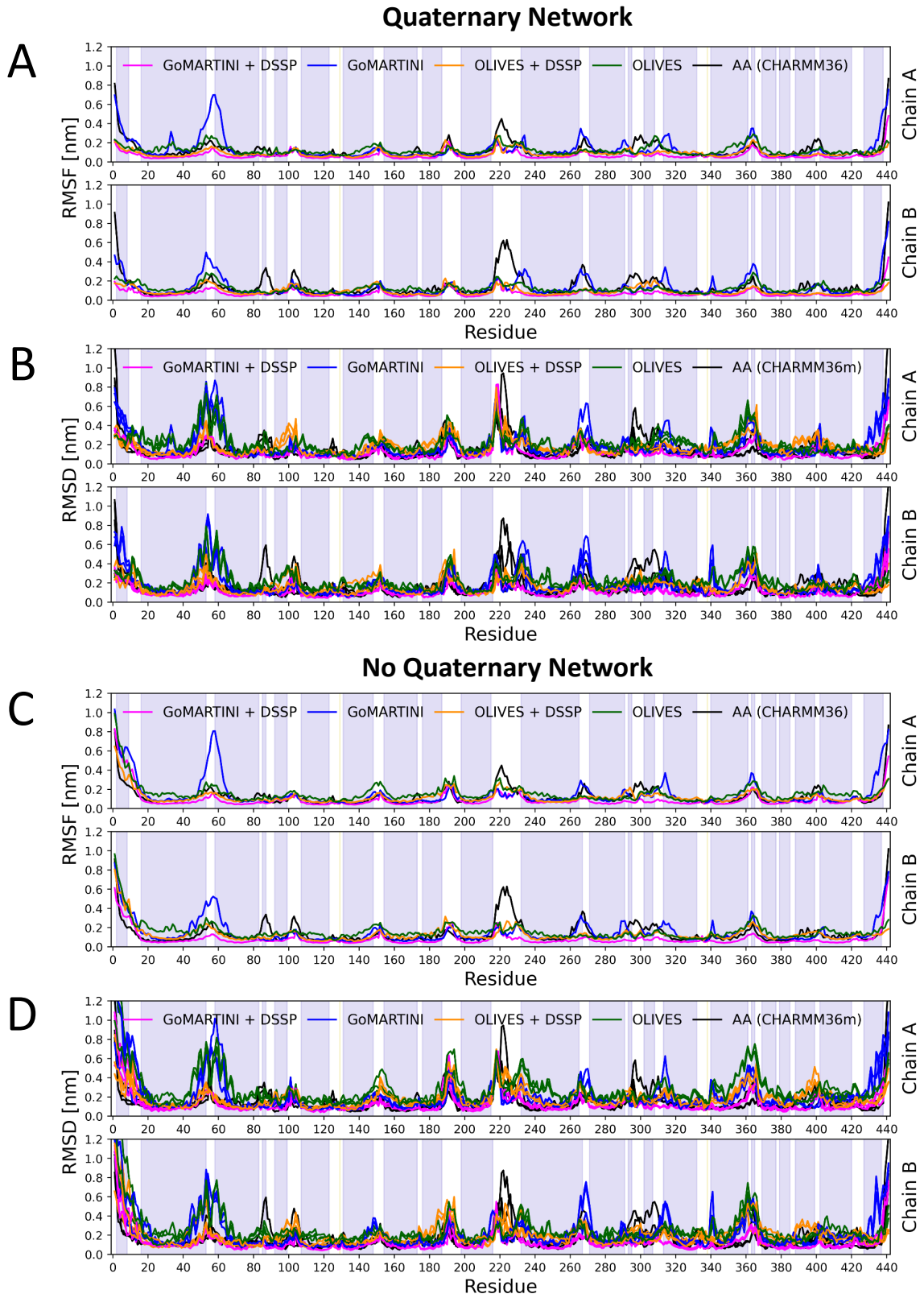


Figure S8: Additional data for the ClC homodimer system. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (A) Residue RMSF, using a quaternary network. (B) Residue RMSD, using a quaternary network. (C) Residue RMSF, without using a quaternary network. (D) Residue RMSD, without using a quaternary network.

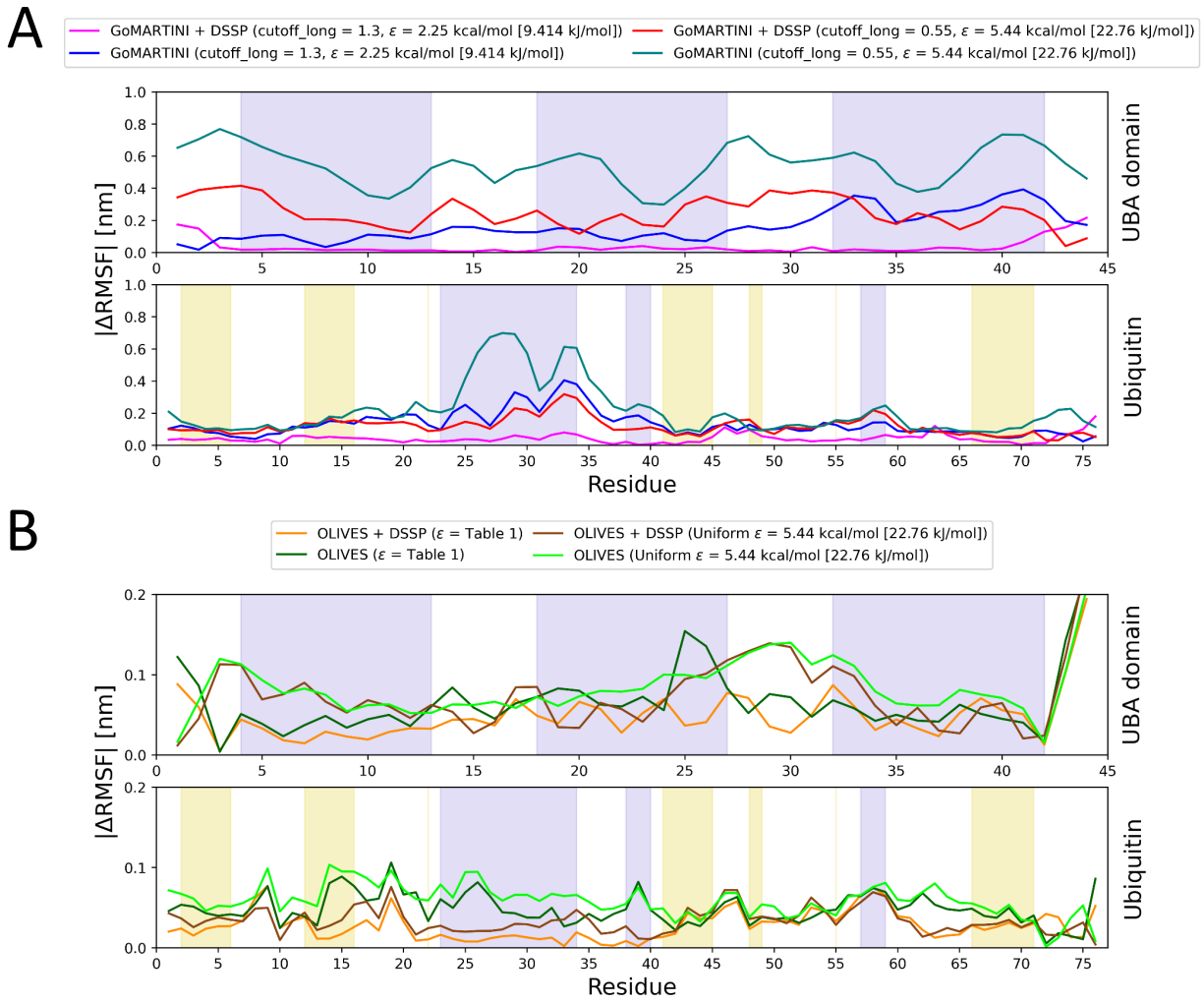


Figure S9: RMSF absolute error for non-default settings of GoMARTINI and OLIVES for the Ubiquitin/UBA domain complex. Purple shaded areas denote helical structure in the starting structure while yellow shade denotes β-structure. (A) Changing the long cutoff and increasing the energy of GoMARTINI to that of the OLIVES cutoff and amide-amide hydrogen bond energy generally lead to drastically worse results (B) Using a uniform energy for all bonds in the OLIVES energy matrix results in a slightly worse RMSF match. Note the different y-axis of (A) and (B).

TOC GRAPHIC

Credit: Stable Diffusion AI